KATHOLIEKE UNIVERSITEIT
LEUVEN

Arenberg Doctoraatsschool Wetenschap & Technologie
Faculteit Wetenschappen
Department Wiskunde

Counting points in families of nondegenerate curves

Jan Tuitman

Proefschrift voorgedragen
tot het behalen van de
graad van Doctor in de
Wetenschappen

December 2010

Counting points in families of nondegenerate curves

Jan Tuitman

Jury:

Prof.dr. Walter van Assche
Prof.dr. Antoine Chambert-Loir
Prof.dr. Jan Denef (promotor)
Prof.dr. François Loeser (promotor)
Prof.dr. Johannes Nicaise
dr. Fréderik Vercauteren
Prof.dr. Willem Veys

Proefschrift voorgedragen
tot het behalen van de
graad van Doctor in de
Wetenschappen

December 2010

Wettelijk depot
ISBN

Acknowledgements

Abstract

In this thesis we consider the problem of computing the zeta function and the number of rational points of an algebraic curve over a finite field. More precisely, we study this problem for nondegenerate curves, a very general class of curves containing for example all elliptic, hyperelliptic and C_{ab} curves. An algorithm for computing the zeta function of such a curve, similar to Kedlaya's algorithm for hyperelliptic curves, has already been given by Castryck, Denef and Vercauteren. Unfortunately, although this algorithm has a good complexity in both time and space, it has turned out to be unpractical and has therefore not been implemented.

We develop a more practical algorithm, using the deformation method introduced by Lauder. Instead of considering a single curve, we take a family of curves containing an easy fiber, for example defined over a very small field. Now the Frobenius map on the rigid cohomology of a complicated fiber can be computed by first computing it for the easy fiber, using the algorithm of Castryck, Denef and Vercauteren, and then solving a certain p-adic differential equation. The zeta function and number of rational points of the complicated fiber can then be deduced from this Frobenius map.

Something similar has already been done by Hubrechts, and Castryck, Hubrechts and Vercauteren, for hyperelliptic and C_{ab} curves, respectively. However, since they used very rough bounds for the required p-adic precision, they were not able to compute provably correct results. We extend these algorithms to the much more general class of nondegenerate curves, and improve the bounds on the p-adic precision, to the point of often being able to compute provably correct results.

We have completed a first implementation of the algorithm in the computer algebra package MAGMA, and have computed some examples. Although the complexity of our algorithm is roughly the same (and even slighly worse) than that of the algorithm of Castryck, Denef and Vercauteren, it seems to work quite a lot better in practice.

The last chapter is almost completely independent of the rest this thesis. We have found a generalization of the sparse effective nullstellensatz to the case when the Newton polytopes do not coincide. From work of Canny and Emiris it was already known that such a result holds generically. We use some toric geometry and a cohomological vanishing theorem to deduce the right criterion for genericity.

Samenvatting

In dit proefschrift bestuderen we het probleem van het berekenen van de zetafunctie en het aantal rationale punten van een algebraische kromme over een eindig lichaam. We doen dit voor niet-gedegeneerde krommen, een heel algemene klasse van krommen die in het bijzonder alle elliptische, hyperelliptische en C_{ab} krommen omvat. Er bestaat al een algoritme, van Castryck, Denef en Vercauteren, voor het berekenen van de zeta functie van zo'n kromme, analoog aan Kedlaya's algoritme voor hyperelliptische krommen. Hoewel dit algoritme een goede tijd- en geheugencomplexiteit heeft, werkt het niet goed in de praktijk en is daarom nooit geimplementeerd.

Wij ontwikkelen een algoritme dat gebruikt maakt van de deformatie methode, geintroduceerd door Lauder. In plaats van een enkele kromme beschouwen we een familie van krommen, die een eenvoudig vezel bevat, bijvoorbeeld gedefinieerd over een heel klein lichaam. Nu kunnen we de Frobenius afbeelding op de rigide cohomologie van een ingewikkeld vezel bepalen door deze eerst te berekenen voor het eenvoudige vezel met het algoritme van Castryck, Denef en Vercauteren en vervolgens een bepaalde p-adische differentiaalvergelijking op te lossen. De zetafunctie en het aantal rationale punten van het ingewikkelde vezel kunnen vervolgens afgeleid worden uit deze Frobenius afbeelding.

Iets gelijkaardigs is reeds gedaan door Hubrechts, en Castryck, Hubrechts en Vercauteren, voor respectievelijk hyperelliptische en C_{ab} krommen. Echter, de grenzen die zij gebruikten voor de benodigde p-adische precisie waren niet scherp genoeg. Daarom waren de resultaten van hun berekeningen niet bewijsbaar correct. Wij generaliseren deze algoritmen naar de veel algemenere klasse van niet-gedegenereerde krommen en verbeteren de grenzen voor de benodigde p-adische precisie, zodat onze resultaten vaak wel bewijsbaar correct zijn.

We hebben een eerste implementatie gemaakt van het algoritme in het computer algebra pakket MAGMA en wat voorbeelden berekend. Hoewel de complexiteit van ons algoritme grofweg gelijk (maar net iets slechter) is als die van het

algoritme van Castryck, Denef en Vercauteren, lijkt het een stuk beter te werken in de praktijk.

Het laatste hoofdstuk is vrijwel geheel onafhankelijk van de rest van dit proefschrift. We hebben een generalisatie gevonden van de 'sparse effective nullstellensatz', wanneer de Newton polytopen niet gelijk zijn. Iets dergelijks was reeds bekend uit werk van Canny en Emiris in het generieke geval. We gebruiken torische meetkunde, en een stelling die zegt dat bepaalde cohomologieruimten triviaal zijn, om het correcte criterium voor genericiteit af te leiden.

Contents

Chapter 1

Introduction

1.1 Zeta functions

Let $\mathbb{F}_q$ be the finite field of characteristic p with $q = p^n$ elements, and let X be an algebraic variety over $\mathbb{F}_q$. The variety X is given by a finite number of polynomial equations in a finite number of variables. So for every extension field $\mathbb{F}_{q^i}$, one can count the number of solutions, i.e. the number of $\mathbb{F}_{q^i}$-rational points of X:

$$N_i = |X(\mathbb{F}_{q^i})|.$$

This sequence of integers can then be used to the define the following generating function:

$$Z(X,T) = \exp(\sum_{i=1}^{\infty} N_i \frac{T^i}{i})$$

which is called the *zeta function* of X. A priori $Z(X,T)$ is an element of the ring $\mathbb{Q}[[T]]$ of formal power series over $\mathbb{Q}$, but it turns out to be a rational function, i.e. an element of $\mathbb{Q}(t)$. When X is proper and smooth, one can even say much more by the famous *Weil conjectures*:

Theorem 1.1.1. *If X is proper smooth of dimension d, then*

$$Z(X,T) = \frac{p_1 p_3 \cdots p_{2d-1}}{p_0 p_2 p_4 \cdots p_{2d}},$$

where for all i

1. $p_i = \prod_j (1 - \alpha_{i,j} T) \in \mathbb{Z}[T]$,
2. *the transformation* $t \to q^d/t$ *maps the roots of* p_i *to the roots of* p_{2d-i} *(preserving their multiplicities),*
3. $|\alpha_{i,j}| = q^{i/2}$ *for all* j*, and every embedding* $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$.

These conjectures were proved by Dwork [19], Grothendieck and Deligne [16], but they are still called the Weil conjectures. The proof(s) first construct a cohomology theory for algebraic varieties over a finite field and relate the zeta function of a variety to its cohomology spaces by a Lefschetz formula. The Weil conjectures are then a consequence of finiteness, Poincaré duality, and purity respectively, of these cohomology spaces.

Grothendieck and Deligne developed *l-adic cohomology*, which gives cohomology spaces over $\mathbb{Q}_l$, for $l \neq p$, that can be shown to have all the required properties. From a modern perspective, Dwork constructed a kind of p-adic cohomology, with cohomology spaces defined over (a finite extension of) $\mathbb{Q}_p$. He was only able to prove the rationality part of the Weil conjectures, but his theory has since then been generalized and extended by many, and has been shown to have all the required properties to prove the full Weil conjectures as well. This cohomology theory is now called *rigid cohomology*.

1.2 Computing zeta functions

Now one can ask whether it is possible to *compute* zeta functions. It is clearly possible (in theory) to compute N_i for any fixed i, by just trying all of the finitely many possibilities for the coordinates. Since one can usually determine (bounds on) the degrees of the numerator and the denominator of the zeta function [10], one only needs a finite number of N_i to determine it. However, in practice this naive algorithm is of very limited use because its running time is polynomial in q, and therefore exponential in its number of digits $\log q$. So a better way to state the problem is as follows.

Problem. *How to compute* $|X(\mathbb{F}_q)|$ *or* $Z(X, T)$ *efficiently.*

Although this is an interesting problem in itself, in recent years it has received a lot of attention mainly because of a number of applications. We will briefly sketch two of these applications now.

1.2.1 Discrete logarithm problems

If two people, say Alice and Bob, are communicating with each other over an unsecure channel like internet, and they want to prevent others from reading their messages, they have to somehow *encrypt* them. Usually (using symmetric encryption) this means they first have to agree on some secret *key*. However, how can they exchange this key safely? Clearly, sending it over the unsecure channel is not a good idea.

A well known solution to this problem is the so called *Diffie Hellman protocol*. Alice and Bob first choose a commutative group G and an element $g \in G$. This information is public, so can be exchanged over the channel. Then Alice picks an integer n_A, and Bob picks an integer n_B. Now Alice computes $n_A g$, and sends it to Bob. Similarly Bob computes $n_B g$, and sends it to Alice. Both of them can then compute $\mathcal{K} = (n_A n_B)g$: Alice computes it as $n_A(n_B g)$ and Bob computes it as $n_B(n_A g)$. Now they can use $\mathcal{K}$ as their secret key. Only $n_A g$ and $n_B g$ have been sent over the channel, so an eavesdropper that wants to obtain the key will have to solve the following problem.

Problem. *Given $g \in G$, $n_A g$, and $n_B g$, find $(n_A n_B)g$.*

This is known as the *Diffie Hellman problem*. A problem very much related to this is the so called *discrete logarithm problem*:

Problem. *Given $g \in G$ and ng, find (the smallest) n.*

Clearly, a solution to the discrete logarithm problem also implies a solution to the Diffie Hellman problem. However, if G and g are well chosen, then both of these problems are considered to be very hard. Popular choices for G are:

1. $G = \mathbb{F}_q^\times$, the multiplicative group of a finite field,
2. $G = Jac(C)(\mathbb{F}_q)$, the group of rational points of the Jacobian of an algebraic curve C (of genus $1, 2$) over a finite field.

The second of these is considered to be the safest: no subexponential algorithm to solve the Diffie Hellman problem for an elliptic curve, or the Jacobian of an hyperelliptic curve of genus two, is known. However, when $|Jac(C)(\mathbb{F}_q)|$ only has relatively small prime divisors, the problem becomes a lot easier. Therefore, to determine whether a curve is suitable for cryptography, one has to be able to compute the number of rational points on its Jacobian.

Over the last ten years this application has lead to a lot of interest in, and quite some progress on, the problem of computing numbers of points and zeta

functions of algebraic curves and their Jacobians, over finite fields. We should mention that the kind of curves that we consider in this thesis will probably not be used for cryptography, because their genus is higher than 2, and higher genus curves are considered to be (slightly) less safe for cryptography than their low genus counterparts.

1.2.2 Experimental evidence for conjectures

There are a lot of conjectures in arithmetic algebraic geometry for which one can gather evidence by computing zeta functions of varieties over finite fields. Since during the last decade methods and algorithms for computing zeta functions have become a lot better, and can now be applied to much more general classes of varieties, some of these conjectures can now be verified experimentally to a much higher degree of accuracy. We will briefly look at one example of this, about the conjectural distribution of ranks of elliptic curves over $\mathbb{F}_q(t)$ and over $\mathbb{Q}$.

Let $\mathbb{F}_q$ be a finite field, suppose for simplicity of characteristic > 3.

Definition 1.2.1. An *elliptic surface* E over $\mathbb{F}_q$ is an *elliptic curve* over the function field $\mathbb{F}_q(t)$. If we put it into *Weierstrass form*

$$E : y^2 = x^3 + ax + b,$$

with $a, b \in \mathbb{F}_q(t)$, then its *discriminant* is given by

$$\Delta = -16(4a^3 + 27b^2),$$

and its *j-invariant* can be written as

$$j = -1728(4a)^3/\Delta.$$

We will suppose that E extends to a smooth projective surface over $\mathbb{F}_q$, also denoted by E, with a proper connected morphism $\pi : E \to \mathbb{P}^1_{\mathbb{F}_q}$.

Elliptic surfaces over $\mathbb{F}_q$ are similar in many ways to elliptic curves over $\mathbb{Q}$. In particular Néron proved that they satisfy the Mordell-Weil Theorem:

Theorem 1.2.2. *For an elliptic surface E over $\mathbb{F}_q$, with nonconstant j-invariant, the group of rational points $E(\mathbb{F}_q(t))$ is finitely generated.*

The rank of this finitely generated abelian group is called the *Mordell-Weil rank* of E. Like in the case of elliptic curves over $\mathbb{Q}$, there are a lot of conjectures about this Mordell-Weil rank. For example there is a widely held belief that:

Conjecture 1.2.3. *One half of all elliptic surfaces E over $\mathbb{F}_q$ has Mordell-Weil rank greater than 0.*

Remark. This should be understood as an asymptotic statement: the proportion of elliptic surfaces of Mordell-Weil rank greater than 0 tends to $\frac{1}{2}$ as the degree of the discriminant Δ (of a minimal model) of E goes to infinity.

The L-function $L(E,T) \in Z[T]$ of E is a certain polynomial that divides the factor p_2 of the denominator of the zeta function $Z(E,T)$. Computing this L-function is very similar to, and even slightly easier than, computing the zeta function. The number of factors $(1-qT)$ in $L(E,T)$, or equivalently the order of $L(E,q^{-s})$ at $s=1$, is called the *analytic rank* of E. There is no known algorithm for computing the Mordell-Weil rank of an elliptic surface over $\mathbb{F}_q$. However, the *Tate conjecture* (wich many believe to be true) implies:

Conjecture 1.2.4 (Tate)**.** *For an elliptic surface E over $\mathbb{F}_q$ the Mordell-Weil rank is equal to the analytic rank.*

So with this conjecture the Mordell-Weil rank can be computed from the analytic one. By using his fibration method, Lauder [44] was able to compute a lot of L-functions and analytic ranks of elliptic surfaces. His data confirmed that the fraction of elliptic surfaces with analytic rank > 0 tends to $\frac{1}{2}$, while previous experiments had suggested the fraction being closer to $\frac{2}{3}$.

For elliptic curves E over $\mathbb{Q}$ the situation is analogous. Again there is no known algorithm to compute the Mordell-Weil rank. In this case the analytic rank is defined as the order of the L-function $L(E,s)$ of E at $s=1$. It can be obtained by computing the zeta function of the reduction of E at a lot of different primes p. The Birch and Swinnerton-Dyer Conjecture (which many believe to be true) states:

Conjecture 1.2.5 (BSD)**.** *For an elliptic curve E over $\mathbb{Q}$ the Mordell-Weil rank is equal to the analytic rank.*

Again it is expected that the fraction of elliptic curves over $\mathbb{Q}$ having Mordell-Weil rank greater than 0 tends to $\frac{1}{2}$ as the discriminant goes to infinity. The experimental data collected so far [4] suggests that the fraction of elliptic curves of analytic rank > 0 might be closer to $\frac{2}{3}$, but more data is needed.

So another motivation to generalize or improve algorithms for computing zeta functions, is to experimentally check conjectures like the ones in this section to a higher degree of accuracy.

1.3 Point counting algorithms

The problem of computing zeta functions has been approached in at least three different ways.

For elliptic curves, Schoof's l-adic algorithm [47] computes the characteristic polynomial of Frobenius acting on the *Tate module* of the the curve modulo various small primes. This is the only known method that allows one to find the zeta function in time polynomial in $\log q$. It has been extended to higher genus curves [45], with running time still polynomial in $\log q$, but exponential in the genus. However, one has to compute explicit equations for the Jacobian of the curve. Even for a low genus curve, usually hundreds of variables and equations are needed. Hence in practice this algorithm is restricted to elliptic curves.

Another class of algorithms, introduced by Satoh [46], first computes the so called *canonical lift* of the Jacobian of a curve to a p-adic field, and then an approximation of the Frobenius acting on it, to determine the zeta function of the Jacobian. The running time of these algorithms is polynomial in $\log_p q$, but linear in p instead of $\log p$, which limits them to small p. However, for elliptic curves, for small p, and for big enough fields, they are faster than Schoof's algorithm. These algorithms also have a running time exponential in the genus. In practice they are restricted to low genus hyperelliptic curves.

Note that both Schoof's algorithm and the algorithms that use the canonical lift use the structure of the Jacobian as an Abelian variety in an essential way.

The algorithms falling into the third and final class follow the approach of Kedlaya, and approximate the Frobenius action on the rigid cohomology of a curve to deduce its zeta function. From this the zeta function of its Jacobian can in turn be deduced. These algorithms generally again have a running time polynomial in $\log_p q$, but linear in p, and so are restricted to small p. However, recently for hyperelliptic curves the dependence of the running time on p has been improved to $\sqrt{p}$, and much larger primes have been treated [29]. Since these algorithms deal directly with the curve rather than its Jacobian, they run in time polynomial in g, and unlike the other algorithms they can even be extended to higher dimensional nonabelian varieties.

Our algorithm falls into this third class. Therefore, we will now look at these algorithms in some more detail.

Rigid cohomology was first used for computing zeta functions by Kedlaya in the case of hyperelliptic curves in odd characteristic [36]. Soon similar algorithms appeared for hyperelliptic curves in characteristic 2 [18], and for some other classes of curves [25, 17]. The most general class of curves, containing all of the previous cases, for which an algorithm of this type has been worked out, is the class of so called *nondegenerate curves* [12]. We will give the precise definition later, but for now let us remark that 'almost any' equation in two variables defines a nondegenerate curve. Unfortunately, this algorithm for nondegenerate curves by Castryck, Denef and Vercauteren has turned out to be too slow for practical use and has therefore not even been implemented, although its complexity is similar to the one of Kedlaya's original algorithm for hyperelliptic curves.

In [42], Lauder introduced his *deformation* algorithm, based on Dwork deformation, that uses a family of varieties and reduces computing the Frobenius on (the cohomology of) a complicated fiber to computing it on an easier fiber and solving a p-adic differential equation. Both Lauder and Gerkmann then reformulated this method, for smooth projective hypersurfaces, in terms of *relative rigid cohomology*. Gerkmann also made it more practical by improving some of the precision bounds, and he computed quite a lot of examples [27]. Although Lauder's initial motivation for using deformations was to apply this strategy to higher dimensional varieties, Hubrechts [32, 31] showed that even for hyperelliptic curves, in some cases it is already faster (and more memory efficient) to use a deformation. In [13] Castryck, Hubrechts and Vercauteren showed that the same is the case for the somewhat more general class of so called $C_{a,b}$-curves.

The most recent tool for computing zeta functions with rigid cohomology is the *fibration* method, introduced by Lauder in [43]. In the fibration algorithm one also starts with a family of varieties which contains an easy fiber. However, the Frobenius on the cohomology of the easy fiber and a similar p-adic differential equation are now used to deduce the Frobenius action on the rigid cohomology, and hence the zeta function, of the *total space* of the family. This strategy can then be used to inductively compute zeta functions of higher dimensional varieties, by fibering them into a family of smaller dimensional ones. So far this method has only been used in practice for surfaces fibered into low genus hyperelliptic curves [43, 44]. However, recently Lauder's student Walker has computed some slightly more general examples in [55].

1.4 This thesis

We use the deformation method to obtain a more practical algorithm for computing the zeta function of (some) nondegenerate curves. Similar deformation algorithms have already been described for the classes of hyperelliptic and $C_{a,b}$ curves in [32, 31] and [13]. For general nondegenerate curves, there are a number of complications:

1. The 'bad' fibers in a family are not described anymore by an ordinary resultant of polynomials. We will see that we can use a generalized resultant which was defined and studied by Gelfand,Kapranov and Zelevinsky instead.

2. For a complete $C_{a,b}$ curve, one can write down a basis for the cohomology that only depends on a, b. For nondegenerate curves this is not the case. Therefore a basis has to be computed. We first compute a basis for the cohomology of the affine part of the curve lying inside the torus. Then

we compute a cohomological residue map to find the cohomology of the complete curve.

3. All of the p-adic precision bounds used in the algorithm for $C_{a,b}$ curves become worse for general nondegenerate curves. It should be mentioned that, although this is sometimes not very clear from these papers, in [32],[13] all explicit computations were already done with much lower precision than required by the bounds. This is not unreasonable, since the precision bounds are known to be far from optimal, and there are probabilistic ways to test the zeta function. However, this way the output of the algorithm is not provably correct anymore.

   Our goal is to obtain provably correct results, for a more general class of curves, so it is clear that we will have to use better bounds. To improve the bounds, we use some recent results on p-adic differential equations, and instead of using a priori bounds, like in [32, 31] and [13], we try to compute better adapted bounds in any particular case at hand.

The algorithm that we obtain has roughly the same (but slightly worse) complexity as the algorithm of Castryck, Denef and Vercauteren, but turns out to be more practical.

The remainder of this thesis is organized as follows:

In **Chapter 2: Nondegenerate curves**, we collect some material about (families of) nondegenerate curves. First we recall the definition of nondegenerateness over a field. Then we extend this definition to families over more general rings. After recalling some definitions and properties of Koszul complexes, we introduce the resultant of Gelfand, Kapranov and Zelevinsky as the determinant of such a complex. Finally, we show that the sparse effective nullstellensatz for a nondegenerate Laurent polynomial over a field can be extended to nondegenerate families over more general rings.

In **Chapter 3: Cohomology**, we explain all of the (mostly cohomological) theory that will go into our computations. This chapter is mainly intended to collect the theory that we will need in the next chapter, although it does contain some original results. First we recall the definitions of algebraic De Rham cohomology and rigid cohomology over a field and over more general bases. Then we discuss finiteness, base change, and comparison theorems for these cohomology spaces, and we explain the construction of the residue map on the cohomology of a curve. After showing how all of this theory applies to families of nondegenerate curves, we continue with a discussion about the integral structure on the cohomology of a nondegenerate curve, and collect some important results about p-adic differential equations. We end the chapter with a sketch of the algorithm that will be worked out in the next chapter.

In **Chapter 4: The algorithm**, we work out the details of the algorithm sketched at the end of the previous chapter. This chapter is the most important one, and most of the material is original. First we explain how to compute the generalized resultant, and how to find the cohomology of family affine nondegenerate curves. We then show how to compute the cohomology of a family of complete nondegenerate curves, using a cohomological residue map. Moreover, we compute the Gauss-Manin connection on the cohomology of a family of nondegenerate curves. Next we give an overview of the point counting algorithm of Castryck, Denef and Vercauteren for a single nondegenerate curve, and show how it can still be used in the case of a family to obtain some bounds. Then we show how to solve the differential equation for the Frobenius matrix, and we give two different bounds for the error propagation in this computation. Finally, we outline all of the steps of the complete algorithm once more, give different ways of bounding the necessary precision at each step, and briefly discuss the complexity of the algorithm.

In **Chapter 5: Examples**, we give a few explicit examples. We have implemented the complete algorithm in the computer algebra package MAGMA, and we have done some experiments. We use these examples to see how well the algorithm does in practice, and to show how some of the different bounds that we have given compare to each other and to their actual values obtained in the computations.

In **Chapter 6: A refinement of a mixed sparse effective nullstellensatz** we include a paper that is almost unrelated to the rest of this thesis. We have found a generalization of the sparse effective nullstellensatz to the case when the Newton polytopes of the Laurent polynomials do not coincide. By work of Canny and Emiris, it was already known that such a result holds generically. We use some toric geometry and a cohomological vanishing result to deduce the correct criterion for 'genericity'.

Chapter 2

Nondegenerate curves

In this chapter we collect some definitions and results about nondegenerate Laurent polynomials and curves.

2.1 Nondegenerate curves

Let k be a field, and let $\overline{k}$ denote a fixed algebraic closure of k. For a Laurent polynomial

$$f = \sum_{(i,j)\in S} f_{i,j} x^i y^j \in k[\mathbb{Z}^2],$$

with $f_{i,j} \in k^\times$, we call $S \subset \mathbb{Z}^2$ the *support* of f. The convex hull Γ of S in $\mathbb{R}^2$ is called the *Newton polygon* of f. Clearly f defines a function on the two dimensional (split) torus over k, which we denote by $\mathbb{T}_k^2$. For a subset $Q \subset \mathbb{Z}^2$, $k[Q]$ will denote the vector subspace of $k[\mathbb{Z}^2]$ consisting of the elements supported on Q.

Definition 2.1.1. Let $f = \sum_{(i,j)\in S} f_{i,j} x^i y^j \in k[\mathbb{Z}^2]$ be a Laurent polynomial with Newton polygon Γ. For a face γ of Γ, we define $f_\gamma = \sum_{(i,j)\in\gamma\cap\mathbb{Z}^2} f_{i,j} x^i y^j$. Then f is called *nondegenerate* if for all faces γ of Γ (of any codimension) the system of equations

$$f_\gamma = x\frac{\partial f_\gamma}{\partial x} = y\frac{\partial f_\gamma}{\partial y} = 0$$

has no solutions in $\mathbb{T}_k^2$, i.e. in $(\bar{k}^\times)^2$.

The nondegeneracy condition can also be interpreted in a more geometric way. To the polygon Γ one can associate a projective toric surface $X_{\Gamma,k}$ over k [24]. Because it is a toric surface, $X_{\Gamma,k}$ contains $\mathbb{T}_k^2$ as an open dense subvariety, the complement of which can be partitioned into lower dimensional tori which are called the *tori at infinity*. Now f defines a curve C on $\mathbb{T}_k^2$, and we denote the closure of this curve in $X_{\Gamma,k}$ by $\overline{C}$. Then f is nondegenerate, if and only if $\overline{C}$ is a smooth curve intersecting the tori at infinity transversely.

Remark. We will refer to both curves C and $\overline{C}$, associated to a nondegenerate Laurent polynomial f, as *nondegenerate curves*.

Example 2.1.2. For $a, b \in \mathbb{N}$ with $\gcd(a,b) = 1$, let $\Gamma_{a,b}$ denote the polygon (or triangle) with vertices $\{(0,0),(0,a),(b,0)\}$. A complete nondegenerate curve $\overline{C}$ over k with Newton polygon $\Gamma_{a,b}$ is called a $C_{a,b}$*-curve*. In this case the toric surface $X_{\Gamma_{a,b},k}$ is the (singular) weighted projective space $\mathbb{P}_k(1,a,b)$, which contains a copy of $\mathbb{A}^2$, the complement of which only has one single (transverse) intersection with $\overline{C}$. Therefore a $C_{a,b}$ curve can be regarded as a curve in $\mathbb{A}^2$ with one point at infinity. A *hyperelliptic curve* is a $C_{a,b}$ curve with $a = 2$.

Almost all Laurent polynomials are nondegenerate, as can be made precise as follows.

Theorem 2.1.3. *Let Γ be a convex polygon in $\mathbb{R}^2$ with integral vertices and write $S = \Gamma \cap \mathbb{Z}^2$. Then the set of points $(f_{i,j})_{(i,j)\in S} \in \mathbb{A}_k^{|S|}$ for which $\sum f_{i,j}x^iy^j$ is nondegenerate is Zariski dense.*

Proof. See for example [12, Proposition 2.3]. Somewhat surprisingly, in characteristic p the result is not true anymore for polytopes of dimension greater than two. □

There is also an interesting homological characterization of nondegenerateness, that will be useful in what follows. Define the graded ring

$$A_{\Gamma,k} = k[t^d x^i y^j | (i,j) \in d\Gamma],$$

where the grading is given by the degree in t. It is well known [3] that $\mathrm{Proj}(A_{\Gamma,k}) \cong X_{\Gamma,k}$. Let V be the graded $A_{\Gamma,k}$-module $A_{\Gamma,k}(-1) \oplus A_{\Gamma,k}(-1) \oplus A_{\Gamma,k}(-1)$.

Definition 2.1.4. For any Laurent polynomial f supported on Γ one defines the *Koszul complex* $\mathcal{K}^\bullet$:

$$0 \longrightarrow \Lambda^3 V \xrightarrow{\partial_3} \Lambda^2 V \xrightarrow{\partial_2} V \xrightarrow{\partial_1} A_{\Gamma,k} \longrightarrow 0,$$

where if we write $F = tf$, the maps are defined by the following formulas on a set of generators $\{e_t, e_x, e_y\}$ of V:

$$\partial_1 e_t = F, \;\; \partial_1 e_x = x\frac{\partial F}{\partial x}, \;\; \partial_1 e_y = y\frac{\partial F}{\partial y},$$

$$\partial_2(e_t \wedge e_x) = Fe_x - x\frac{\partial F}{\partial x}e_t, \;\; \partial_2(e_t \wedge e_y) = Fe_y - y\frac{\partial F}{\partial y}e_t, \;\; \partial_2(e_x \wedge e_y) = x\frac{\partial F}{\partial x}e_y - y\frac{\partial F}{\partial y}e_x,$$

$$\partial_3(e_t \wedge e_x \wedge e_y) = Fe_x \wedge e_y - x\frac{\partial F}{\partial x}e_t \wedge e_y + y\frac{\partial F}{\partial y}e_t \wedge e_x.$$

We then have the following characterization of nondegenerateness.

Theorem 2.1.5. *Let $f \in k[\mathbb{Z}^2]$ be a Laurent polynomial with Newton polygon Γ. The following are equivalent:*

1. *f is nondegenerate,*
2. *The elements $F, x\frac{\partial F}{\partial x}, y\frac{\partial F}{\partial y}$ (with $F = tf$) form a regular sequence on $A_{\Gamma,k}$,*
3. *The Koszul complex $\mathcal{K}^\bullet$ truncated at the term $A_{\Gamma,k}$ is exact. In other words: $\mathcal{K}^\bullet$ induces a resolution of $A_{\Gamma,k}/(F, x\frac{\partial F}{\partial x}, y\frac{\partial F}{\partial y})$.*

Proof. This is well known, but because it is hard to find in the literature, we include a proof.

$(1 \Rightarrow 2)$: Write $Y_{\Gamma,k} = \mathrm{Spec}(A_{\Gamma,k})$ for the affine toric variety associated to Γ i.e. the cone over $X_{\Gamma,k}$. Let P denote the point of $Y_{\Gamma,k}$ corresponding to the (irrelevant) maximal ideal of A_Γ generated by all elements of degree 1. By decomposing $Y_{\Gamma,k}$ into its orbits one easily shows that f is nondegenerate if and only if $F, x\frac{\partial F}{\partial x}, y\frac{\partial F}{\partial y}$ define P in $Y_{\Gamma,k}$. By a theorem of Hochster [30] $A_{\Gamma,k}$ is a Cohen Macaulay ring. So if $F, x\frac{\partial F}{\partial x}, y\frac{\partial F}{\partial y}$ define P (which has dimension 0) in $Y_{\Gamma,k}$ (which has dimension 3) then by the Cohen Macaulay property, they form a regular sequence.

$(2 \Rightarrow 3)$ It is a standard result in homological algebra [48] that the standard Koszul complex associated to a regular sequence is exact when truncated at its next to last term. Our Koszul complex is different from the standard one by the shift in the grading, but this doesn't affect the exactness.

$(3 \Rightarrow 1)$ If $\mathcal{K}^\bullet$ is exact except at the last term then its only nonzero homology module is $H_0(\mathcal{K}^\bullet) = A_{\Gamma,k}/(F, x\frac{\partial F}{\partial x}, y\frac{\partial F}{\partial y})$. For $i \geq 3$ the i-th graded part of $\mathcal{K}^\bullet$ is given by

$$\begin{array}{ccccccc} 0 & \longrightarrow & k[(i-3)\Gamma] & \xrightarrow{\partial_3} & \oplus_{j=1}^3 k[(i-2)\Gamma] & \xrightarrow{\partial_2} & \dots \\ \dots & \xrightarrow{\partial_2} & \oplus_{j=1}^3 k[(i-1)\Gamma] & \xrightarrow{\partial_1} & k[i\Gamma] & \longrightarrow & 0. \end{array}$$

By a theorem of Ehrhart [21] the number of lattice points lying on (or inside) $n\Gamma$ is given by a polynomial $p(n)$ of degree 2. The dimension over k of the i-th graded part of $H_0(\mathcal{K}^\bullet)$ is then given by

$$\dim H_0(\mathcal{K}^\bullet)_i = p(i) - 3p(i-1) + 3p(i-2) - p(i-3),$$

but the righthandside vanishes for every polynomial p of degree 2. Hence $\dim H_0(\mathcal{K}^\bullet)_i = 0$ for all $i \geq 3$ and so $H_0(\mathcal{K}^\bullet)$ is finite dimensional over k. This in turn implies that $H_0(\mathcal{K}^\bullet)$ must be supported on a finite number of points. Since $F, x\frac{\partial F}{\partial x}, y\frac{\partial F}{\partial y}$ are homogeneous, it must be supported on P, so that $F, x\frac{\partial F}{\partial x}, y\frac{\partial F}{\partial y}$ define P in $Y_{\Gamma,k}$ and f is nondegenerate. □

2.2 Families of nondegenerate curves

So far we have only talked about one curve at a time, but we will need to study curves in families. Let R be a commutative ring. In our applications R will always be a field or a localization of $S[t]$, where S is either a field or a discrete valuation ring, but we find it interesting to state some of the results in this and the following section in more (but varying) generality. Let f be a Laurent polynomial

$$f = \sum f_{i,j} x^i y^j \in R[\mathbb{Z}^2],$$

with $f_{ij} \in R$. Geometrically this corresponds to a family of curves over the base $\mathrm{Spec}(R)$, contained in the two dimensional torus $\mathbb{T}^2_R$ over R. The obvious way to generalize the notion of nondegenerateness is as follows.

Definition 2.2.1. Let $f \in R[\mathbb{Z}^2]$ be a Laurent polynomial with Newton polygon Γ. Then f and its corresponding family of curves are called *nondegenerate*, if for all faces γ of Γ (of any codimension), the elements $f_\gamma, x\frac{\partial f_\gamma}{\partial x}, y\frac{\partial f_\gamma}{\partial y}$ do not have a common zero in $\mathbb{T}^2_R$, i.e. if they generate the unit ideal in $R[\mathbb{Z}^2]$.

Like in the previous section, when R was a field, we define the graded ring $A_{\Gamma,R} = R[t^d x^i y^j | (i,j) \in d\Gamma]$, and the Koszul complex $\mathcal{K}^\bullet$ over $A_{\Gamma,R}$.

Now assume that R is a Noetherian integral domain, and let K denote its field of fractions. Often (only) the generic fiber of a family will be nondegenerate. This means that the corresponding Laurent polynomial $f \in R[\mathbb{Z}^2]$ is nondegenerate as an element of $K[\mathbb{Z}^2]$. We want to find an element $r_f \in R$, such that f becomes nondegenerate over $R[\frac{1}{r_f}]$. For this we can use the *determinant* of the Koszul complex $\mathcal{K}^\bullet$ as in [26]:

Definition 2.2.2. Consider a finite complex $\mathcal{M}^\bullet$ of free R-modules of finite type

$$0 \longrightarrow \mathcal{M}_m \longrightarrow \mathcal{M}_{m-1} \longrightarrow \ldots \longrightarrow \mathcal{M}_1 \longrightarrow \mathcal{M}_0 \longrightarrow 0,$$

and assume that $\mathcal{M}^\bullet$ is generically exact, i.e. that $\mathcal{M}^\bullet \otimes K$ is exact. The *determinant* $\det(\mathcal{M}^\bullet)$ of $\mathcal{M}^\bullet$ is the element of $(R \setminus 0)/R^\times$ defined as follows. Let $[e_1, \ldots, e_s]$ denote an R-basis of $\mathcal{M}_1$. We can assume that the images of $e_1, \ldots, e_r$ in $\mathcal{M}_0$ form a K-basis of $\mathcal{M}_0 \otimes K$, so that $\mathcal{M}_1$ splits into a direct sum

$$\mathcal{M}_1 = \mathcal{M}_1' \oplus \mathcal{M}_1'', \;\; \mathcal{M}_1' = \oplus_{i=1}^r Re_i, \;\; \mathcal{M}_1'' = \oplus_{i=r+1}^s Re_i,$$

and $\det(\mathcal{M}^\bullet)$ can be defined inductively as

$$\det(\mathcal{M}^\bullet) = \det(\mathcal{M}_1' \to \mathcal{M}_0) \det(\mathcal{M}'^\bullet)^{-1},$$

where $\det(\mathcal{M}_1' \to \mathcal{M}_0)$ is the usual determinant of the R-module homomorphism $\mathcal{M}_1' \to \mathcal{M}_0$ (with respect to arbitrary bases), and $\mathcal{M}'^\bullet$ denotes the (shorter, generically exact) complex

$$0 \longrightarrow \mathcal{M}_m \longrightarrow \mathcal{M}_{m-1} \longrightarrow \ldots \longrightarrow \mathcal{M}_2 \longrightarrow \mathcal{M}_1'' \longrightarrow 0,$$

where the homomorphism $\mathcal{M}_2 \to \mathcal{M}_1''$ is the canonical projection.

Remark. One can show that $\det(\mathcal{M}^\bullet)$ is indeed an element of $R \setminus 0$, and does not depend on the choices of bases upto a unit $\in R^\times$ [26, Appendix A].

We can now state the following version of a result by Gelfand, Kapranov and Zelevinsky:

Theorem 2.2.3. *Let R be a unique factorization domain, and $f \in R[\mathbb{Z}^2]$ a Laurent polynomial which is nondegenerate as an element of $K[\mathbb{Z}^2]$. We denote the l-th graded part of $\mathcal{K}^\bullet$ by $(\mathcal{K}^\bullet)_l$. Now:*

1. *$\det((\mathcal{K}^\bullet)_l) \in (R \setminus 0)/R^\times$ does not depend on l for $l \geq 3$,*
2. *f is nondegenerate over R if and only if $\det((\mathcal{K}^\bullet)_l) \in R^\times$ for some (and hence all) $l \geq 3$.*

Proof. One can define the *universal family of curves with Newton polygon* Γ by

$$f_\Gamma = \sum_{(i,j)\in\Gamma} a_{i,j} x^i y^j \in \mathbb{Z}[a_{i,j}][\mathbb{Z}^2].$$

Gelfand, Kapranov, and Zelevinsky [26] define the *principal determinant* $E_\Gamma \in \mathbb{Z}[a_{i,j}]$ to be the determinant of the degree l part of the corresponding Koszul complex for $l \gg 0$.

$$E_\Gamma = \det((\mathcal{K}_\Gamma^\bullet)_l) \quad \text{for} \quad l \gg 0.$$

Our Laurent polynomial $f = \sum_{(i,j)\in\Gamma} f_{i,j} x^i y^j$ defines a ringhomomorphism $\phi_f : \mathbb{Z}[a_{i,j}] \to R$ with $\phi_f(a_{i,j}) = f_{i,j}$. Since $\phi_f(f_\Gamma) = f$, and the determinant clearly commutes with base change, we find

$$\phi_f(\det((\mathcal{K}_\Gamma^\bullet)_l)) = \det((\mathcal{K}^\bullet)_l). \tag{2.1}$$

1. In [26, chapter 10, Proposition 1.1], it is (only) shown that $\det((\mathcal{K}_\Gamma^\bullet)_l)$ does not depend on l for $l \gg 0$. Let $\mathcal{F}_\Gamma^\bullet$ denote the complex of sheaves on X_Γ associated to the complex $\mathcal{K}_\Gamma^\bullet$. Note that the complex $(\mathcal{K}_\Gamma^\bullet)_l$ can be interpreted as the global sections of the complex $\mathcal{F}_\Gamma^\bullet \otimes \mathcal{O}(l)$, where $\mathcal{O}(l)$ is defined with respect to the standard projective embedding of X_Γ. From [26, Chapter 3, Theorem 4.2], we see that $l \gg 0$ means that l should be chosen such that the terms of $\mathcal{F}_\Gamma^\bullet \otimes \mathcal{O}(l)$ are acyclic. However, in our case if we tensor with $\mathcal{O}(3)$, then all of these terms become direct sums of line bundles that are generated by their global sections. It is known that a line bundle on a complete toric variety that is generated by its global sections is acyclic [24, section 3.5]. So we see that $\det((\mathcal{K}_\Gamma^\bullet)_l)$ does not depend on l already for $l \geq 3$. By equation (2.1), the same then holds for $\det((\mathcal{K}^\bullet)_l)$.

2. It follows from [26, chapter 10, Proposition 1.1] that a Laurent polynomial $f = \sum_{(i,j)\in\Gamma} f_{i,j} x^i y^j$ over a field k is nondegenerate if and only if $\phi_f(E_\Gamma) \neq 0$. This proves the result when R is a field. Let us now try to extend this to a unique factorization domain R.

   For an irreducible element π of R, we denote the localization of R in the prime ideal (π) by $R_{(\pi)}$. Note that $R_{(\pi)}$ is a discrete valuation ring. If f is nondegenerate over R, then it is so over all $R_{(\pi)}$ as well. Conversely, if f is not nondegenerate over R, then for some face γ of Γ, the closed subscheme of $\mathrm{Spec}(R)[\mathbb{Z}^2]$ defined by $f_\gamma, x\frac{\partial f_\gamma}{\partial x}, y\frac{\partial f_\gamma}{\partial y}$ is nonempty, and projecting this subscheme onto $\mathrm{Spec}(R)$, we can find an irreducible $\pi \in R$ such that f is not nondegenerate over $R_{(\pi)}$. Since an element of R is a unit if and only if it is a unit in every $R_{(\pi)}$, it is then enough to prove the result for the case when R is a discrete valuation ring.

   So now let R be a discrete valuation ring with uniformizer π. If f has the same Newton polygon over the residue field $R/(\pi)$ as over R, then by Lemma 2.3.4 it is nondegenerate over R if and only if it is nondegenerate over $R/(\pi)$. Note that $\det((\mathcal{K}^\bullet)_i) \in R^\times$ if and only if $\det((\mathcal{K}^\bullet)_i) \neq 0$ in the residue field $R/(\pi)$. Since we already know that the theorem holds for a field, this finishes the proof.

□

Hence if we want f to become nondegenerate, then we have to invert exactly $\det((\mathcal{K}^\bullet)_i)$ for any $i \geq 3$.

Definition 2.2.4. Let R be a unique factorization domain and $f \in R[\mathbb{Z}^2]$ a Laurent polynomial which is nondegenerate over the field of fractions K of R. We define the *resultant* $r_f \in (R \setminus 0)/R^\times$ of f as

$$r_f = \det((\mathcal{K}^\bullet)_i)$$

for any $i \geq 3$.

Remark. The element r_f does not have to be squarefree. At first sight it seems that one could take its squarefree part. However, the multiplicities turn out to contain some information about the family (see for example Conjecture 2.3.5).

2.3 An effective nullstellensatz

Let R be a commutative ring. If $f \in R[\mathbb{Z}^2]$ defines a smooth family in $\mathbb{T}^2_R$, then by definition there exist $\alpha, \beta, \gamma \in R[\mathbb{Z}^2]$ such that

$$\alpha f + \beta x \frac{\partial f}{\partial x} + \gamma y \frac{\partial f}{\partial y} = 1.$$

In our computations we will need to find such Laurent polynomials, and we want them to be as 'small' as possible. For f nondegenerate over a field, the following theorem is a special case of [12, Theorem 2.12].

Theorem 2.3.1. *Let k be a field, Γ a two dimensional convex polygon in $\mathbb{R}^2$ with vertices in $\mathbb{Z}^2$, and $f \in k[\mathbb{Z}^2]$ a Laurent polynomial with Newton polygon Γ. Suppose that f is nondegenerate. For every $g \in k[\mathbb{Z}^2]$ with support contained in 3Γ, there exist $\alpha, \beta, \gamma \in k[\mathbb{Z}^2]$ with support contained in 2Γ such that*

$$\alpha f + \beta x \frac{\partial f}{\partial x} + \gamma y \frac{\partial f}{\partial y} = g.$$

Proof. In the proof of Theorem 2.1.5 we have already shown that when $f \in k[\mathbb{Z}^2]$ is nondegenerate, $\dim H_0(\mathcal{K}^\bullet)_i = 0$ for all $i \geq 3$. However,

$$H_0(\mathcal{K}^\bullet)_3 = k[3\Gamma]/(f, x\frac{\partial f}{\partial x}, y\frac{\partial f}{\partial y})k[2\Gamma]$$

which proves the theorem. □

Actually, this theorem can be easily extended to an arbitrary Noetherian ring, as we will now see.

Lemma 2.3.2. *Let M, N be finitely generated modules over a Noetherian ring R, and $\phi : M \to N$ a homomorphism of R-modules. If for every maximal ideal m of R the map*

$$\phi \otimes (R/m) : M \otimes (R/m) \to N \otimes (R/m),$$

is surjective, then ϕ is surjective.

Proof. Consider the exact sequence

$$M \xrightarrow{\phi} N \longrightarrow \text{coker}(\phi) \longrightarrow 0.$$

By the right-exactness of tensor products, for a maximal ideal ideal m of R we get an exact sequence

$$M \otimes (R/m) \xrightarrow{\phi \otimes (R/m)} N \otimes (R/m) \longrightarrow \text{coker}(\phi) \otimes (R/m) \longrightarrow 0.$$

Since $\phi \otimes (R/m)$ is surjective, we see that $\text{coker}(\phi) \otimes (R/m) = 0$.

So $\text{coker}(\phi)$ is a finitely generated R-module that becomes 0 when tensored with (R/m). *Nakayama's Lemma* now implies that $\text{coker}(\phi)$ also becomes 0 when tensored with the local ring R_m. Since this holds for every maximal ideal m of R, we have that $\text{coker}(\phi) = 0$, so that ϕ is surjective. □

We can use this lemma to generalize Theorem 2.3.1:

Theorem 2.3.3. *Let R be a Noetherian ring, Γ a two dimensional convex polygon in $\mathbb{R}^2$ with vertices in $\mathbb{Z}^2$, and $f \in R[\mathbb{Z}^2]$ a Laurent polynomial with Newton polygon Γ. If f is nondegenerate, then for every $g \in R[\mathbb{Z}^2]$ with support contained in 3Γ, there exist $\alpha, \beta, \gamma \in R[\mathbb{Z}^2]$ with support contained in 2Γ, such that*

$$\alpha f + \beta x \frac{\partial f}{\partial x} + \gamma y \frac{\partial f}{\partial y} = g.$$

Proof. We consider the homomorphism of R-modules

$$\phi : R[2\Gamma] \oplus R[2\Gamma] \oplus R[2\Gamma] \to R[3\Gamma],$$

$$\phi(a, b, c) = af + bx \frac{\partial f}{\partial x} + cy \frac{\partial f}{\partial y}.$$

If f is nondegenerate over R, then for any maximal ideal m of R it is nondegenerate over the residue field $k = R/m$ as well, so that

$$\phi \otimes k : k[2\Gamma] \oplus k[2\Gamma] \oplus k[2\Gamma] \to k[3\Gamma]$$

is surjective by Theorem 2.3.1. Now Lemma 2.3.2 finishes the proof. □

We will need the following lemma a couple of times.

Lemma 2.3.4. *Let R be a discrete valuation ring with uniformizer π and residue field $k = R/(\pi)$, Γ a two dimensional convex polygon in $\mathbb{R}^2$ with vertices in $\mathbb{Z}^2$, $f \in R[\mathbb{Z}^2]$ a Laurent polynomial and $\overline{f} \in k[\mathbb{Z}^2]$ its reduction. Suppose that both f and $\overline{f}$ have Newton polygon Γ. Then f is nondegenerate if and only if $\overline{f}$ is nondegenerate.*

Proof. If f is nondegenerate, then clearly $\overline{f}$ is nondegenerate as well. Conversely, if $\overline{f}$ is nondegenerate and the map ϕ is defined as in the proof of Theorem 2.3.3, then $\phi \otimes k$ is surjective by Theorem 2.3.1. So Lemma 2.3.2 implies that ϕ itself is surjective. Let τ be a face of Γ and μ a monomial supported on 3τ. Since ϕ is surjective, there exist $\alpha, \beta, \gamma \in R[2\Gamma]$ such that

$$\alpha f + \beta x \frac{\partial f}{\partial x} + \gamma y \frac{\partial f}{\partial y} = \mu.$$

It is easy to see that

$$\alpha_{2\tau} f_\tau + \beta_{2\tau} x \frac{\partial f_\tau}{\partial x} + \gamma_{2\tau} y \frac{\partial f_\tau}{\partial y} = \mu,$$

so that the Laurent polynomials $f_\tau, x\frac{\partial f_\tau}{\partial x}, y\frac{\partial f_\tau}{\partial y}$ generate the unit ideal in $R[\mathbb{Z}^2]$. □

Finally, we expect that Theorem 2.3.3 can be refined in the following way.

Conjecture 2.3.5. *Let R be a Noetherian unique factorization domain, Γ a two dimensional convex polygon in $\mathbb{R}^2$ with vertices in $\mathbb{Z}^2$, and $f \in R[\mathbb{Z}^2]$ a Laurent polynomial with Newton polygon Γ that is nondegenerate over the field of fractions of R. For every $g \in R[\mathbb{Z}^2]$ with support in 3Γ, there exist $\alpha, \beta, \gamma \in R[\mathbb{Z}^2]$ with support contained in 2Γ such that*

$$\left(\frac{\alpha}{r_f}\right) f + \left(\frac{\beta}{r_f}\right) x \frac{\partial f}{\partial x} + \left(\frac{\gamma}{r_f}\right) y \frac{\partial f}{\partial y} = g.$$

Remark. Note that this is indeed a refinement of Theorem 2.3.3, because it shows that for the denominators of α, β, γ one can take the resultant itself instead of a power of it.

Remark. This is only a conjecture in the sense that *we* don't know how to prove it (yet). It is probably (again) a consequence of the work of Gelfand, Kapranov and Zelevinsky [26]. Note that it is enough to show it for the universal family $f = \sum_{(i,j)\in\Gamma} a_{i,j} x^i y^j \in \mathbb{Z}[a_{i,j}][\Gamma]$.

Chapter 3

Cohomology

In this chapter we recall the definitions, properties, and relations of the cohomology theories we will need later. The cohomology of (families of) nondegenerate curves will have our special attention.

3.1 Algebraic De Rham cohomology

Let k denote a field of characteristic 0.

3.1.1 General definition

For a smooth algebraic variety X over k, let Ω^1_X be the sheaf of Kähler differentials on X. The sheaf of algebraic i-forms is then defined as $\Omega^i_X = \Lambda^i \Omega^1_X$ and one can write down the following *De Rham complex* $\Omega^\bullet_X$:

$$0 \longrightarrow \mathcal{O}_X \xrightarrow{d} \Omega^1_X \xrightarrow{d} \Omega^2_X \xrightarrow{d} \dots.$$

Definition 3.1.1. The *algebraic De Rham cohomology spaces* of X are defined as the hypercohomology spaces

$$H^i_{DR}(X) = \mathbb{H}^i(X, \Omega^\bullet_X),$$

and are finite dimensional vector spaces over k.

3.1.2 Nondegenerate curves

Let $\mathbb{T}_k^2$ be the two dimensional torus over k and $f \in k[\mathbb{Z}^2]$ a nondegenerate Laurent polynomial in two variables x, y with Newton polygon Γ. For simplicity, from now on we will always assume:

Convention. Γ *is two dimensional and contains the origin.*

Let $C \subset \mathbb{T}_k^2$ denote the curve defined by f. Since C is an affine curve the above definition simplifies. Because it is affine we can work with the complex of global sections of its De Rham complex, and since it has dimension 1 there are no nonzero n-forms for $n > 1$. Let $A = k[\mathbb{Z}^2]/(f)$ denote the coordinate ring of C and $\Omega_A^1 = (Adx + Ady)/(df)$ the A-module of its Kähler differentials. The algebraic De Rham cohomology spaces $H_{DR}^0(C)$ and $H_{DR}^1(C)$ are then the cohomology spaces of the complex

$$0 \longrightarrow A \xrightarrow{d} \Omega_A^1 \longrightarrow 0.$$

Since the curve C is connected, $H_{DR}^0(C)$ has rank 1 and is not very interesting. Therefore we are mainly interested in $H_{DR}^1(C) = \Omega_A^1/dA$.

We are now going to recall some definitions and results from [12]. Let $X_{\Gamma,k}$ be the projective toric surface over k associated to Γ and $\overline{C}$ the closure of C in $X_{\Gamma,k}$. The edges $t_1, \ldots, t_k$ of Γ correspond to the 1-dimensional tori at infinity $T_1, \ldots, T_k$ of X_Γ. For each t_i let e_i denote the inward pointing normal vector that is minimally integral i.e. of the form (a_i, b_i) with $a_i, b_i \in \mathbb{Z}$ and $\gcd(a_i, b_i) = 1$. We let (p_i, q_i) denote any point on t_i, write $N_i = \langle (p_i, q_i), e_i \rangle = p_i a_i + q_i b_i$, and define the divisors D_Γ, D on $\overline{C}$ by

$$D_\Gamma = -\sum_{i=1}^{n} N_i (T_i \cap \overline{C}),$$

$$D = \sum_{i=1}^{n} (T_i \cap \overline{C}).$$

Remark. The divisors $(T_i \cap \overline{C})$ are reduced, i.e. all of the points in their support (which do no have to be k-rational) have multiplicity 1, because $\overline{C}$ intersects the T_i tranversely. Moreover, the degree of the divisor $(T_i \cap \overline{C})$ is equal to the *arithmetic length* $\lambda_i = |t_i \cap \mathbb{Z}^2| - 1$ of t_i.

Remark. Since Γ contains the origin, we have that $N_i \leq 0$ for all i, so that D_Γ is an effective divisor.

Theorem 3.1.2.

1. Ω_A^1 *is free of rank* 1 *over* A *with generator* $\omega_f = \frac{dx}{xy\frac{\partial f}{\partial y}}$.

2. $Div(\omega_f) = D_\Gamma - D$.

Proof. (1) follows from (2), for which see [12, Corollary 2.7]. □

For a divisor D on $\overline{C}$ we denote its *Riemann Roch space* by

$$L(D) = \{h \in k(\overline{C}) | Div(h) \geq -D\}.$$

When D is a multiple of D_Γ, these spaces can be described very explicitly:

Theorem 3.1.3.

1. *For $(i,j) \in \mathbb{Z}^2$, we have that*

$$Div(x^i y^j) = \sum_{k=1}^{n} \langle (i,j), e_k \rangle (T_k \cap \overline{C}).$$

2. *For every $n \in \mathbb{N}$, the Riemann Roch space $L(nD_\Gamma)$ coincides with the image of $k[n\Gamma]$ in $H^0(C, \mathcal{O}_C)$.*

Proof. See [12, Corollary 2.6 and Theorem 2.9]. □

One easily verifies that for $g \in k[\mathbb{Z}^2]$

$$dg = xy(\frac{\partial f}{\partial y}\frac{\partial g}{\partial x} - \frac{\partial f}{\partial x}\frac{\partial g}{\partial y})\omega_f.$$

So it is natural to define the differential operator

$$\delta = xy(\frac{\partial f}{\partial y}\frac{\partial}{\partial x} - \frac{\partial f}{\partial x}\frac{\partial}{\partial y}).$$

Theorem 3.1.4.

1. *Any 1-form $\omega \in \Omega^1_A$ is cohomologous to a 1-form $h\omega_f$ with $h \in L(2D_\Gamma)$.*
2. *There is an isomorphism*

$$H^1_{DR}(C) \cong \frac{k[2\Gamma]}{fk[\Gamma] + \delta(k[\Gamma])}.$$

Proof.

1. This is [12, Theorem 3.2]. For later use we include the proof here.

   First suppose that all places P_i in the support of D are k-rational. Write $D_\Gamma = \sum_{k=1}^{r} a_k P_k$. Since Γ contains the origin $a_k \geq 0$ for all k. Note that $D_\Gamma + D = \sum_{k=1}^{r}(a_k+1)P_k$. The divisor $D_\Gamma - D = Div(\omega_f)$ is a canonical divisor, and D is effective, so $\sum_{k=1}^{r} a_k = \deg D_\Gamma > 2g-2$. Suppose that $\omega \in \Omega_A^1$ has a pole of order $b_k + 1 \geq a_k + 1$ at some place P_k. By the Riemann-Roch theorem, there exists a function

   $$h \in L(a_1P_1 + \ldots + b_kP_k + \ldots + a_rP_r) \setminus L(a_1P_1 + \ldots + (b_k-1)P_k + \ldots + a_rP_r).$$

   The differential dh then has a pole of order exactly b_k+1 at P_k, and a pole of order at most a_i+1 at the other P_i. So adding to ω a suitable multiple of dh will reduce the pole order at P_k. If we continue like this, eventually ω will have at most a pole of order a_i+1 at each P_i. It will then be of the form $\omega = h\omega_f$, where h has at most a pole of order $a_i + 1 - (-a_i+1) = 2a_i$ at P_i, i.e. $h \in L(2D_\Gamma)$.

   In general, if l is a finite Galois extension of k such that the P_i are l-rational, then the above argument holds over l. Taking traces (under $Gal(l/k)$), and using that the divisor D_Γ is defined over k, then concludes the proof.

2. This is [12, Corollary 3.3]. By Theorem 3.1.3,

   $$L(2D_\Gamma) \cong \frac{k[2\Gamma]}{fk[\Gamma]},$$

   and by the first part, $H^1_{DR}(C)$ is generated by $L(2D_\Gamma)\omega_f$. Therefore

   $$H^1_{DR}(C) \cong \frac{k[2\Gamma]\omega_f}{fk[\Gamma]\omega_f + d(k[\Gamma])} \cong \frac{k[2\Gamma]}{fk[\Gamma] + \delta(k[\Gamma])}.$$

□

So computing in $H^1_{DR}(C)$ comes down to computing in the quotient of two finite (small) dimensional vector spaces. This is a consequence of working with a nondegenerate Laurent polynomial f. For more general f, one would need to compute Gröbner bases.

Let $g = \dim H^0(\overline{C}, \Omega_{\overline{C}})$ denote the *genus* of $\overline{C}$. It is well known that $\dim H^1_{DR}(\overline{C}) = 2g$. The dimension of H^1 of a nondegenerate curve can also be described in terms of its Newton polygon:

Theorem 3.1.5. *Let $f \in k[\mathbb{Z}^2]$ be a nondegenerate Laurent polynomial with Newton polygon Γ, C the curve defined by f in $\mathbb{T}^2_k$, and $\overline{C}$ its closure in $X_{\Gamma,k}$. If we write*

- $I = |\Gamma^\circ \cap \mathbb{Z}^2|$,
- $B = |\partial\Gamma \cap \mathbb{Z}^2|$,

for the number of points in the interior and on the border of Γ respectively, then we have

1. *dim* $H^1_{DR}(\overline{C}) = 2I$,
2. *dim* $H^1_{DR}(C) = 2I + B - 1$.

Proof.

1. It is enough to show that $g = I$. For a proof of this see for example [12, Corollary 2.8].
2. First note that $|\overline{C} \setminus C| = deg(D) = B$. Since the Euler characteristic is additive, we have that $\chi(\overline{C}) = \chi(C) + B$. However

$$\chi(\overline{C}) = \text{dim}H^0_{DR}(\overline{C}) - \text{dim}H^1_{DR}(\overline{C}) + \text{dim}H^2_{DR}(\overline{C}) = 1 - 2g + 1,$$

$$\chi(C) = \text{dim}H^0_{DR}(C) - \text{dim}H^1_{DR}(C) = 1 - \text{dim}H^1_{DR}(C),$$

so that the result follows.

□

3.2 Rigid cohomology

3.2.1 General definition

Let $\mathbb{F}_q$ be the finite field with $q = p^n$ elements, $\mathbb{Q}_q$ the unique unramified extension of $\mathbb{Q}_p$ of degree n, and $\mathbb{Z}_q$ its ring of integers. We let v_p denote the p-adic valuation on $\mathbb{Q}_q$ and write $|w| = p^{-v_p(w)}$ for the norm of an element $w \in \mathbb{Q}_q$. We first recall very briefly Berthelot's general construction of *rigid cohomology*. More details can be found in [8].

Let X be an algebraic variety over $\mathbb{F}_q$, and $\overline{X}$ a compactification of X, i.e. a complete algebraic variety over $\mathbb{F}_q$ which contains X as an open dense subvariety and denote $\overline{X} \setminus X = Z$. Assume that there exists a closed embedding of $\overline{X}$

into a separated formal scheme of finite type P over $\mathbb{Z}_q$ which is smooth in a neighbourhood of X. Suppose for simplicity that there exists an morphism F_q from P to itself that induces the q-th power Frobenius map on $\overline{X}$. Locally this is always the case, and the construction can be done locally, so there is no loss of generality.

To the formal scheme P one can associate its *generic fiber* $\tilde{P} = P \otimes_{\mathbb{Z}_q} \mathbb{Q}_q$ which is a *rigid analytic space* that comes equiped with a *specialization map* $\mathrm{sp} : \tilde{P} \to P$. Note that as a topological space P can be identified with its special fiber $P \otimes_{\mathbb{Z}_q} \mathbb{F}_q$. For any subscheme $U \subset P \otimes_{\mathbb{Z}_q} \mathbb{F}_q$ the *tube* $]U[\subset \tilde{P}$ is defined as $\mathrm{sp}^{-1}(U)$. Let $j :]X[\to]\overline{X}[$ denote the open immersion.

Definition 3.2.1. A *strict neighbourhood* V of $]X[$ in $]\overline{X}[$ is an admissible open set containing $]X[$, such that locally on P, and for any affinoid W contained in $]\overline{X}[$, there exist functions $f_1, \ldots, f_m$ which cut out Z in $\overline{X}$, and some $\lambda < 1$, such that

$$W \cap \{x \in]\overline{X}[\mid \max_i \{|f_i(x)|\} \geq \lambda\} \subseteq W \cap V.$$

Definition 3.2.2. The rigid cohomology spaces of X are defined as the hypercohomology spaces

$$H^i_{rig}(X) = \mathbb{H}^i(]\overline{X}[, j^\dagger \Omega^\bullet_{]\overline{X}[}),$$

where $j^\dagger$ denotes the direct limit

$$j^\dagger = \varinjlim \iota_* \iota^*$$

over all strict neighbourhoods $\iota : V \hookrightarrow]\overline{X}[$ of $]X[$ in $]\overline{X}[$.

Remark. It can be shown that the $H^i_{rig}(X)$ don't depend on the choices made in their construction, are contravariantly functorial in X, and carry a linear Frobenius automorphism F_q^* induced by F_q [8]. They are known to be finite dimensional vector spaces over $\mathbb{Q}_q$ [9].

Remark. If F_p denotes a lift to P (as a formal scheme over $\mathbb{Z}_p$) of the p-th power Frobenius on $\overline{X}$, then F_p also acts on the $H^i_{rig}(X)$. However, F_p has to act nontrivially on $\mathbb{Q}_q$, say by a map $\sigma = F_p|_{\mathbb{Q}_q}$. The action F_p^* of F_p on $H^i_{rig}(X)$ is then σ-semilinear. In computational applications F_p^* is often easier to compute and $F_q^* = (F_p^*)^n$.

3.2.2 The smooth affine case

For a smooth affine variety this definition can be simplified considerably as we will now see. Actually, historically rigid cohomology was first defined by

Monsky and Washnitzer in this case, and only later generalized by Berthelot to arbitrary separated schemes. Therefore, in the smooth affine case rigid cohomology is also called *Monsky Washnitzer cohomology*.

So now let X denote a smooth affine algebraic variety over $\mathbb{F}_q$, and write $X = \mathrm{Spec}(A)$ where A is an $\mathbb{F}_q$-algebra of finite type. By [22], there exists a smooth $\mathbb{Z}_q$-algebra $\mathcal{A} = \mathbb{Z}_q[x_1, \ldots, x_n]/(f_1, \ldots, f_m)$ that lifts A, i.e. such that $\mathcal{A} \otimes_{\mathbb{Z}_q} \mathbb{F}_q \cong A$. We put $\mathcal{X} = \mathrm{Spec}(\mathcal{A})$, and let $\hat{\mathcal{A}}$ denote the (p-adic) completion of $\mathcal{A}$. Note that X is a closed subset of $\mathrm{Spf}(\hat{\mathcal{A}})$. So for P we can take the closure of $\mathrm{Spf}(\hat{\mathcal{A}})$ in formal projective n-space over $\mathbb{Z}_q$. The special fiber of P is then the closure $\overline{X}$ of X in $\mathbb{P}^n_{\mathbb{F}_q}$, the generic fiber $\tilde{P}$ is the rigid analytic space given by the equations $f_1, \ldots, f_m$ in projective analytic n-space $\mathbb{P}^{n,an}_{\mathbb{Q}_q}$, and the specialization map $sp : \tilde{P} \to \overline{X}$ is simply reduction mod p. The tube $]X[$ is the intersection of P with the open disc of radius 1 around 0 in $\mathbb{P}^{n,an}_{\mathbb{Q}_q}$, and the intersections with the closed discs of radius $\rho > 1$ form a complete system of affinoid strict neighbourhoods.

Definition 3.2.3. The direct limit of the sections of the structure sheaf $\mathcal{O}_{\mathbb{P}^{n,an}_{\mathbb{Q}_q}}$ over the closed discs of radius $\rho > 1$ is called the ring of *overconvergent functions*:

$$\mathbb{Q}_q\langle x_1, \ldots, x_n\rangle^\dagger = \{\sum a_I x^I \mid \exists \rho > 1 \ : \ \lim_{|I|\to\infty} |a_i|\rho^{|I|} = 0\}.$$

With this notation the global sections of the overconvergent De Rham complex $j^\dagger\Omega^\bullet_{]\overline{X}[}$ on $]\overline{X}[$ are

$$H^0(]\overline{X}[, j^\dagger\mathcal{O}_{]\overline{X}[}) = A^\dagger \ = \mathbb{Q}_q\langle x_1, \ldots, x_n\rangle^\dagger/(f_1, \ldots, f_m),$$

$$H^0(]\overline{X}[, j^\dagger\Omega^1_{]\overline{X}[}) = \Omega^1_A{}^\dagger = (A^\dagger dx_1 + \ldots A^\dagger dx_n)/(df_1, \ldots, df_m),$$

$$H^0(]\overline{X}[, j^\dagger\Omega^i_{]\overline{X}[}) = \Omega^i_A{}^\dagger = \Lambda^i\Omega^1_A{}^\dagger.$$

Unlike $\mathcal{A}$, the ring $A^\dagger$ always admits lifts F_p and F_q of the p-th and q-th power Frobenius morphisms on A [53]. Because the direct limit in the definition of $j^\dagger\Omega^\bullet_{]\overline{X}[}$ is taken over affinoid neighbourhoods, the hypercohomology can be replaced by cohomology of the complex of global sections, and hence the rigid cohomology $H^i_{rig}(X)$ is given by the cohomology of the complex

$$0 \longrightarrow A^\dagger \longrightarrow \Omega^1_A{}^\dagger \longrightarrow \Omega^2_A{}^\dagger \longrightarrow \ldots,$$

with the Frobenius map F_q^* induced by F_q.

3.2.3 The Lefschetz formulas

Recall that for an algebraic variety X over $\mathbb{F}_q$

$$Z(X,T) = \exp(\sum_{i=1}^{\infty} N_i \frac{T^i}{i}),$$

with

$$N_i = |X(\mathbb{F}_{q^i})|.$$

The relation between rigid cohomology and the zeta function is given by the following Lefschetz formulas:

Theorem 3.2.4. *Let X be a smooth algebraic variety of dimension d over $\mathbb{F}_q$. We have:*

$$N_k = \sum_{i=0}^{2d} (-1)^k Tr((q^d (F_q^*)^{-1})^k | H^i_{rig}(X)),$$

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

Proof. See [23, Theorem 6.3]. □

Remark. The restriction to smooth X is only used here so that we can use Poincaré duality and avoid having to give the definition of rigid cohomology with compact support to state the more general formulas of [23].

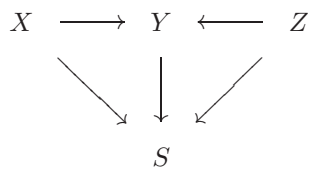
Remark. Since the spaces $H^i_{rig}(X)$ are finite dimensional, the second formula implies that $Z(X,T)$ is a rational function of T. Recently it has been shown that the rest of the Weil conjectures can also be proved within the theory of rigid cohomology [37].

3.3 A comparison theorem

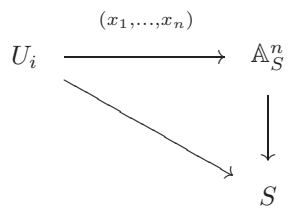
Comparing the definitions of algebraic De Rham cohomology of a lift to characteristic zero of a variety and its rigid cohomology, we note that the only difference is that polynomials have been replaced with overconvergent power series in the De Rham complex. One could hope that this gives the same cohomology. Unfortunately, this cannot be true in general, since the rigid cohomology does not depend on the chosen lift to characteristic zero, while the algebraic De Rham cohomology does. Somehow one has to pick 'the right lift'.

It turns out that any lift that admits a *relative normal crossing compactification* over $\mathbb{Z}_q$ will do. We first recall what a relative normal crossing compactification is. For later use we state the definition in its most general form.

Definition 3.3.1. Let Y be a smooth proper scheme of relative dimension n over some scheme S, Z a relative divisor on Y, and X its open complement.



The divisor Z is called a relative normal crossing divisor if Y can be covered by Zariski open S-schemes U_i, such that for each of them there exists a diagram



and some $m \leq n$, such that the map $(x_1, \ldots, x_n) : U \to \mathbb{A}^n_S$ is étale and $Z \cap U_i \subset U_i$ is defined by $x_1 \ldots x_m = 0$. We then also say that Y is a relative normal crossing compactification of X over S.

Remark. For us X/S will always have (relative) dimension one, so that there cannot be any crossings and Z/S is smooth.

Theorem 3.3.2. *Let X be a smooth $\mathbb{F}_q$-scheme that can be lifted to a smooth $\mathbb{Z}_q$-scheme $\mathcal{X}$ which admits a relative normal crossing compactification. Write $\mathbf{X} = \mathcal{X} \otimes_{\mathbb{Z}_q} \mathbb{Q}_q$. For all i, the canonical map*

$$H^i_{DR}(\mathbf{X}) \to H^i_{rig}(X)$$

is an isomorphism.

Proof. This is a special case of a theorem by Baldassarri and Chiarellotto [2]. Another proof can be obtained by combining some of the comparison theorems of Berthelot in [7, 9] . □

It turns out that in the case we are most interested in such a lift is easily found:

Theorem 3.3.3. *Let $f \in \mathbb{F}_q[\mathbb{Z}^2]$ be a Laurent polynomial which is nondegenerate with Newton polygon Γ, and let $F \in \mathbb{Z}_q[\mathbb{Z}^2]$ be a lift of f which has the same Newton polygon. Denote by $\mathcal{X}_\Gamma$ the projective toric surface over $\mathbb{Z}_q$ associated to Γ. Let $\mathcal{C}$ be the curve over $\mathbb{Z}_q$ defined by F inside of the dense torus of $\mathcal{X}_\Gamma$ and $\overline{\mathcal{C}}$ the closure of $\mathcal{C}$ in $\mathcal{X}_\Gamma$. Then $\mathcal{D} = \overline{\mathcal{C}} \setminus \mathcal{C}$ is a relative normal crossing divisor on $\overline{\mathcal{C}}$.*

Proof. From Lemma 2.3.4 we know that F is nondegenerate as well, since $\mathbb{Z}_q$ is a discrete valuation ring. Note that the construction of $\mathcal{X}_\Gamma$ over $\mathbb{Z}_q$ is the same as over a field, and that its orbits under the action of the torus still correspond to the faces of Γ. The intersections of $\overline{\mathcal{C}}$ with the orbits p_i corresponding to the vertices of Γ are easily seen to be empty. Let t_i denote an edge of Γ, and $e_i = (a_i, b_i)$ the inward pointing normal vector of t_i that is minimally integral. Write $N_i = \langle p_i, e_i \rangle$ for any point p_i on t_i. Choose an integral vector (c_i, d_i) such that

$$\det \begin{pmatrix} -b_i & a_i \\ c_i & d_i \end{pmatrix} = -1.$$

The open affine toric subvariety $\mathcal{V}_i$ of $\mathcal{X}_\Gamma$ corresponding to t_i can then be written as

$$\mathcal{V}_i = \operatorname{Spec} \mathbb{Z}_q[x_i, x_i^{-1}, y_i],$$

with

$$x_i = x^{-b_i} y^{a_i}, \qquad x = x_i^{-d_i} y_i^{a_i},$$

$$y_i = x^{c_i} y^{d_i}, \qquad y = x_i^{c_i} y_i^{b_i}.$$

The toric variety $\mathcal{V}_i$ consists of two orbits, the dense one, and the one corresponding to t_i, which is defined by $y_i = 0$. In the new coordinates (x_i, y_i) we can write

$$F(x, y) = y_i^{N_i}(F_{t_i}(x_i^{-d_i}, x_i^{c_i}) + y_i(\ldots)).$$

Hence on $\mathcal{V}_i$ the closure of $\mathcal{C}$ in $\mathcal{X}_\Gamma$ is defined by

$$F_i(x_i, y_i) = y_i^{-N_i} F(x, y) = F_{t_i}(x_i^{-d_i}, x_i^{c_i}) + y_i(\ldots).$$

Since both sequences $F, x\frac{\partial F}{\partial x}, y\frac{\partial F}{\partial y}$ and $F_{t_i}, x\frac{\partial F_{t_i}}{\partial x}, y\frac{\partial F_{t_i}}{\partial y}$ generate the unit ideal in $\mathbb{Z}_q[\mathbb{Z}^2]$, the schemes

$$\overline{\mathcal{C}} \cap \mathcal{V}_i = \operatorname{Spec} \mathbb{Z}_q[x_i, x_i^{-1}, y_i]/(F_i(x_i, y_i)),$$

$$\mathcal{D} \cap \mathcal{V}_i = \operatorname{Spec} \mathbb{Z}_q[x_i, x_i^{-1}]/(F_i(x_i, 0)),$$

are smooth over $\mathbb{Z}_q$.

Since $\cup_i V_i = \mathcal{X}_\Gamma - \cup_i p_i$, and $\overline{\mathcal{C}}$ doesn't intersect the p_i, we conclude that both $\mathcal{C}$ and $\mathcal{D}$ are smooth over $\mathbb{Z}_q$. In particular $\mathcal{D}$ is a relative normal crossing divisor over $\mathbb{Z}_q$. □

Remark. Note that the same argument works over other rings R as well. If $f \in R[\mathbb{Z}^2]$ is nondegenerate, then f admits a relative normal crossing compactification over R.

3.4 Relative cohomology

Both algebraic De Rham cohomology and rigid cohomology can also be defined for families. In this case they are not just vector spaces, but vector bundles over the base space of the family with some extra structure.

3.4.1 Relative algebraic De Rham cohomology

Let k again denote a field of characteristic 0, and $\pi : X \to S$ a smooth family defined over k. One defines the relative Kähler differentials $\Omega^1_{X/S} = \Omega^1_X/\pi^*\Omega^1_S$, the relative i-forms $\Omega^i_{X/S} = \Lambda^i\Omega^1_{X/S}$, and the relative algebraic De Rham complex $\Omega^\bullet_{X/S}$:

$$0 \longrightarrow \mathcal{O}_X \xrightarrow{d} \Omega^1_{X/S} \xrightarrow{d} \Omega^2_{X/S} \longrightarrow \dots.$$

Definition 3.4.1. The relative algebraic De Rham cohomology sheaves of X/S are defined as the higher derived images of $\Omega^\bullet_{X/S}$ under π

$$H^i_{DR}(X/S) = \mathbb{R}^i\pi_*(\Omega^\bullet_{X/S}).$$

Remark. If X/S is an affine family, then $H^i_{DR}(X/S)$ is just the cohomology of the complex of global sections of $\pi_*\Omega^\bullet_{X/S}$.

Theorem 3.4.2. *If X/S admits a relative normal crossing compactification, then the $H^i_{DR}(X/S)$ are locally free.*

The $H^i_{DR}(X/S)$ come equiped with an integrable connection, which is called the Gauss-Manin connection. Let us first recall the notion of a connection on a vector bundle.

Definition 3.4.3. Let V be a vector bundle on S. A connection on V is a map of vector bundles $\nabla : V \to V \otimes \Omega^1_S$ which satisfies the Leibniz rule

$$\nabla(fs) = f\nabla(s) + s \otimes df$$

for all local sections f of $\mathcal{O}_S$ and s of V.

The Gauss-Manin connection on $H^i_{DR}(X/S)$ can be defined as follows.

Definition 3.4.4. The De Rham complex $\Omega_X^\bullet$ can be equiped with the decreasing filtration

$$F^i = im(\Omega_X^{\bullet -i} \otimes \pi^* \Omega_S^i \to \Omega_X^\bullet).$$

The spectral sequence associated to this filtration has as its first sheet

$$E_1^{p,q} = \Omega_S^p \otimes H_{DR}^q(X/S).$$

The Gauss-Manin connection $\nabla : H^i(X/S) \to H^i(X/S) \otimes \Omega_S^1$ is now defined as the differential $d_1 : E_1^{0,i} \to E_1^{1,i}$ in this spectral sequence.

Remark. We can give a more explicit description of ∇ when X/S is affine. If we lift a relative i-cocycle $\omega \in \Omega_{X/S}^i$ to an absolute i-form $\omega' \in \Omega_X^i$ and apply the absolute differential d, in general we get something nonzero which can be represented by an element of $\Omega_S^1 \otimes \Omega_{X/S}^i$. Projecting onto $\Omega_S^1 \otimes H_{DR}^i(X/S)$, we obtain $\nabla(\omega)$. One can check that this gives the same class as the previous definition.

From the relative cohomology we can deduce the cohomology of every fiber of the family by the following *base change theorem*.

Theorem 3.4.5. *Assume X/S admits a relative normal crossing compactification. Let s be a point in S, $X_s = \pi^{-1}(s)$, and $H_{DR}^i(X/S)_s$ the fiber of the vector bundle $H_{DR}^i(X/S)$ in the point s. Then*

$$H_{DR}^i(X/S)_s \cong H_{DR}^i(X_s).$$

Sketch of the proofs. Although Theorems 3.4.2 and 3.4.5 are well known, we have not been able to find a good reference in the literature. Let $\overline{\pi} : Y \to S$ denote a relative normal crossing compactification of X/S, and let D/S be the relative divisor of the complement of X in Y. One can define a *logarithmic De Rham complex* $\Omega_{Y/S}^\bullet(\log D)$ on Y, and by a theorem of Deligne (see section 3.5)

$$\mathbb{R}\overline{\pi}_*(\Omega_{Y/S}^\bullet(\log D)) \cong H_{DR}^i(X/S).$$

Now since $\overline{\pi}$ is proper, this implies that $H_{DR}^i(X/S)$ is a coherent $\mathcal{O}_S$-module. Theorem 3.4.2 then follows because a $\mathcal{D}$-module on S that is $\mathcal{O}_S$-coherent is locally free. Theorem 3.4.5 is a consequence of the proper base change theorem for $\overline{\pi}$. □

We will need one more property of the relative algebraic De Rham cohomology with its Gauss-Manin connection, namely that it is a *regular singular* $\mathcal{D}$-module. For simplicity, we only state this for the case when S is an open dense subscheme of $\mathbb{P}_k^1$.

Theorem 3.4.6. *Let S be an open dense subscheme of the projective line $\mathbb{P}^1_k$, and let $\pi : X \to S$ denote a (smooth) family. The Gauss-Manin connection ∇ on $H^i_{DR}(X/S)$ is regular singular. By this we mean that for every (geometric) point $z \in \mathbb{P}^1_k$, we can choose a basis $[v_1, \ldots, v_n]$ of sections of $H^i_{DR}(X/S)$ on some neighbourhood of z, such that the $n \times n$ matrix N, defined over the function field $k(t)$ by the equations $\nabla v_j = \sum_i N_{ij} v_i$, has at most a simple pole at z.*

Proof. See [28, Theorem 3.1]. □

3.4.2 Relative rigid cohomology

Suppose we have a smooth family $\pi : X \to S$ over $\mathbb{F}_q$. We assume (for simplicity) that this family can be lifted to a smooth family $\pi : \mathcal{X} \to \mathcal{S}$ over $\mathbb{Z}_q$. Let $\overline{\mathcal{X}}, \overline{\mathcal{S}}$ denote compactifications of $\mathcal{X}, \mathcal{S}$. Then there exists a commutative diagram

$$\begin{array}{ccc} \hat{\mathcal{X}} & \xrightarrow{j} & \hat{\overline{\mathcal{X}}} \\ {\scriptstyle\hat{\pi}}\downarrow & & \downarrow{\scriptstyle\hat{\overline{\pi}}} \\ \hat{\mathcal{S}} & \xrightarrow{j} & \hat{\overline{\mathcal{S}}} \end{array}$$

where the horizontal maps are open immersions and the hats denote formal completion over $\mathbb{Z}_q$. Let $\overline{X}, \overline{S}$ be the special fibers of $\overline{\mathcal{X}}, \overline{\mathcal{S}}$. Then $\hat{\overline{\pi}}$ also induces a map $\overline{\hat{\pi}} :]\overline{X}[\to]\overline{S}[$, so that one can define:

Definition 3.4.7. The relative rigid cohomology sheaves of X/S are defined as

$$H^i_{rig}(X/S) = \mathbb{R}^i \overline{\hat{\pi}}_* (j^\dagger \Omega^\bullet_{]\overline{X}[/]\overline{S}[}).$$

Now what kind of object is this? It is a sheaf on the rigid analytic space $]\overline{S}[$, but like its algebraic counterpart it also carries a Gauss-Manin connection. Moreover it has an action of Frobenius. In the rest of this section F will denote (a lift of) either the q-th power or the p-th power Frobenius.

Definition 3.4.8. Let S and $j :]S[\to]\overline{S}[$ be as before. Write $\mathcal{O}^\dagger_S = j^\dagger \mathcal{O}_{]\overline{S}[}$ and ${\Omega^1_S}^\dagger = j^\dagger \Omega^1_{]\overline{S}[}$. An *overconvergent F-isocrystal* on S is a locally free $\mathcal{O}^\dagger_S$-module $\mathcal{E}$ of finite rank equiped with a connection

$$\nabla : \mathcal{E} \to \mathcal{E} \otimes {\Omega^1_S}^\dagger,$$

and an F-semilinear Frobenius map $\mathcal{F} : \mathcal{E} \to \mathcal{E}$ that induces an isomorphism

$$\mathcal{F} : F^* \mathcal{E} \cong \mathcal{E},$$

which is horizontal i.e. commutes with the connection

$$\begin{array}{ccc} \mathcal{E} & \xrightarrow{\nabla} & \mathcal{E} \otimes \Omega_S^{1\,\dagger} \\ {\scriptstyle \mathcal{F}}\downarrow & & \downarrow{\scriptstyle \mathcal{F}\otimes dF} \\ \mathcal{E} & \xrightarrow{\nabla} & \mathcal{E} \otimes \Omega_S^{1\,\dagger} \end{array}$$

Remark. Upto equivalence, the category of overconvergent F-isocrystals on S only depends on S, and has pullbacks under arbitrary morphisms.

Remark. F can denote (a lift of) either the q-th power or the p-th power Frobenius. When we mention overconvergent F-isocrystals, it will either be clear what F is, or it won't matter.

Definition 3.4.9. For an overconvergent F-isocrystal $\mathcal{E}$, where F denotes the q-th power (or p-th power) Frobenius, and an integer k, the k-th *(Tate) twist* $\mathcal{E}(k)$ of $\mathcal{E}$ denotes the same overconvergent F-isocrystal, but with the Frobenius isomorphism $\mathcal{F}$ replaced by $q^{-k}\mathcal{F}$ (or $p^{-k}\mathcal{F}$).

We have the following equivalent of Theorem 3.4.2.

Theorem 3.4.10. *If X/S is a smooth family over $\mathbb{F}_q$ that can be lifted to a smooth family $\mathcal{X}/\mathcal{S}$ over $\mathbb{Z}_q$, which admits a relative normal crossing compactification, then the $H^i_{rig}(X/S)$ are overconvergent F-isocrystals, with the connection ∇ given by the Gauss-Manin connection, and $\mathcal{F}$ the morphism induced by F.*

Again, the cohomology of a fiber can be deduced from the relative cohomology of the family by a base change theorem:

Theorem 3.4.11. *Let X/S be a smooth family over $\mathbb{F}_q$ that can be lifted to a smooth family $\mathcal{X}/\mathcal{S}$ over $\mathbb{Z}_q$, which admits a relative normal crossing compactification. Take $s \in S(\mathbb{F}_q)$ to be a rational point on S, and $X_s = \pi^{-1}(s)$ its fiber. Let $i : s \to S$ denote the closed embedding. We have*

$$i^*(H^i_{rig}(X/S)) \cong H^i_{rig}(X_s)$$

as $\mathbb{Q}_q$-vector spaces with a Frobenius action.

Finally, the comparison theorem between algebraic De Rham cohomology and rigid cohomology also extends to the relative situation:

Theorem 3.4.12. *Let $X \to S$ be a smooth family over $\mathbb{F}_q$ that can be lifted to a smooth family $\mathcal{X}/\mathcal{S}$ over $\mathbb{Z}_q$, which admits a relative normal crossing compactification. For all i, the canonical map*

$$H^i_{DR}(\mathcal{X}/\mathcal{S} \otimes_{\mathbb{Z}_q} \mathbb{Q}_q) \otimes \mathcal{O}_S^\dagger \to H^i_{rig}(X/S)$$

is an isomorphism of locally free sheaves of $\mathcal{O}_S^\dagger$ modules with connection.

Remarks on the proofs. All three of these theorems have already been used by both Gerkmann [27] and Lauder [43], and seem to be widely accepted. For the case when X/S is proper, the proofs can be found in [51, chapter 8]. However, again in the general (normal crossing compactifiable) case a good reference seems to be missing. We expect that for the first two theorems an argument like the one at the end of section 3.4.1, but this time with relative logarithmic De Rham cohomology replaced by *relative log crystalline cohomology*, should work. The third theorem then follows from the first two by applying Theorem 3.3.2 at every fiber. It would also be possible in our applications to only work with proper families. We will, like everyone else, just *assume* that these theorems hold in general. □

3.5 Residues

The cohomology of a (family of) nondegerate curve(s) contained in $\mathbb{T}^2$ is a priori easier to compute than that of its nonsingular projective model, obtained by taking its closure in the corresponding toric surface, because one does not have to deal with open affine covers and hypercohomology. However, usually we are more interested in the cohomology of the (family of) projective curve(s), since it is of lower dimension and therefore allows for more efficient computations. We will need a *cohomological residue map* to relate the two. First we recall the necessary exact sequences.

Let k be a field of characteristic zero, $\pi : C \to S$ a family of smooth affine curves defined over k, which admits a relative normal crossing compactification $\overline{\pi} : \overline{C} \to S$, and let $D = \overline{C} \setminus C$ be the relative divisor of the complement.

Recall that, by definition of a normal crossing compactification, $\overline{C}$ can be covered by open affines U_i/S such that U_i is étale over $\mathbb{A}^1_S$ via a coordinate x, and $D \cap U_i$ is either empty or defined by $x = 0$. Let $j : C \to \overline{C}$ denote the open immersion, and $\Omega^1_{\overline{C}/S}(\log D)$ the sheaf of (relative) differentials on $\overline{C}$ with logarithmic poles along D, i.e. the subsheaf of $\mathcal{O}_{\overline{C}/S}$-modules of $j_*\Omega^1_{C/S}$ generated on U_i by the element $\frac{dx}{x}$ if $U_i \cap D \neq \emptyset$, and by dx otherwise. Again, one constructs a (logarithmic) De Rham complex $\Omega^\bullet_{\overline{C}/S}(\log D)$, with terms $\Omega^i_{\overline{C}}(\log D) = \Lambda^i \Omega^1_{\overline{C}/S}(\log D)$, and the usual differential.

Definition 3.5.1. The relative *log-De Rham* cohomology sheaves of the pair $(\overline{C}, D)/S$ are the higher derived images of its logarithmic De Rham complex.

$$H^i_{DR}((\overline{C}, D)/S) = \mathbb{R}^i\overline{\pi}_*(\Omega^\bullet_{\overline{C}/S}(\log D)).$$

There is a short exact sequence of complexes

$$0 \longrightarrow \Omega^\bullet_{\overline{C}/S} \longrightarrow \Omega^\bullet_{\overline{C}/S}(\log D) \xrightarrow{Res} i_*\Omega^\bullet_{D/S}[+1] \longrightarrow 0,$$

where $i : D \to \overline{C}$ denotes the closed embedding, and the *residue map*

$$Res : \Omega^{\bullet}_{\overline{C}}(\log D)) \to i_* \Omega^{\bullet}_D[+1]$$

is defined by sending a logarithmic differential $g\frac{dx}{x}$ to $g_{|D}$. This short exact sequence then gives rise to the following *long exact sequence*:

$$0 \longrightarrow H^1_{DR}(\overline{C}/S) \longrightarrow H^1_{DR}((\overline{C}, D)/S) \xrightarrow{Res} H^0_{DR}(D/S) \longrightarrow H^2_{DR}(\overline{C}/S).$$

Now by a theorem of Deligne, the algebraic De Rham cohomology of C/S is isomorphic to the log-De Rham cohomology of the pair $(\overline{C}, D)/S$:

Theorem 3.5.2. *The canonical map*

$$h : H^i_{DR}((\overline{C}, D)/S) \to H^i_{DR}(C/S)$$

is an isomorphism.

Proof. For a purely algebraic proof see [1, Theorem 2.2.5]. □

Substituting this into the long exact sequence, we see that $H^1_{DR}(\overline{C}/S)$ injects into $H^1_{DR}(C/S)$, and is the kernel of the cohomological residue map

$$Res : H^1_{DR}((\overline{C}, D)/S) \to H^0_{DR}(D/S).$$

To get the same result for rigid cohomology, suppose that $\mathcal{C}/\mathcal{S}$ is a family of smooth affine curves defined over $\mathbb{Z}_q$, which admits a relative normal crossing compactification $\overline{\mathcal{C}}/\mathcal{S}$ with complementary divisor $\mathcal{D} = \overline{\mathcal{C}} \setminus \mathcal{C}$. Recall that D/S is smooth by the remark after definition 3.3.1.

Theorem 3.5.3. *There exists a commutative diagram*

$$\begin{array}{ccccccc}
0 & \longrightarrow & H^1_{DR}(\overline{\mathcal{C}}/\mathcal{S} \otimes \mathbb{Q}_q) & \longrightarrow & H^1_{DR}(\mathcal{C}/\mathcal{S} \otimes \mathbb{Q}_q) & \xrightarrow{Res} & H^0_{DR}(\mathcal{D}/\mathcal{S} \otimes \mathbb{Q}_q) \\
 & & \downarrow & & \downarrow & & \downarrow \\
0 & \longrightarrow & H^1_{rig}(\overline{\mathcal{C}}/\mathcal{S} \otimes \mathbb{F}_q) & \longrightarrow & H^1_{rig}(\mathcal{C}/\mathcal{S} \otimes \mathbb{F}_q) & \longrightarrow & H^0_{rig}(\mathcal{D}/\mathcal{S} \otimes \mathbb{F}_q)(-1)
\end{array}$$

where the rows are exact, the vertical maps are homomorphisms of (sheaves of) $\mathcal{O}_S$-modules that become isomorphisms when tensored with $\mathcal{O}^{\dagger}_S$, and the second row is Frobenius equivariant.

Proof. The first row is just the above long exact sequence. The bottom row can be constructed similarly and coincides with the excision sequence for relative rigid cohomology which is known to be Frobenius equivariant for the chosen twist. The vertical maps are isomorphisms when tensored with $\mathcal{O}^{\dagger}_S$ by Theorem 3.4.12. □

3.6 Families of nondegenerate curves

3.6.1 Defining a family

Let $\mathbb{F}_q$ be the finite field with $q = p^n$ elements, and $f \in \mathbb{F}_q[t][\mathbb{Z}^2]$ a Laurent polynomial with Newton polygon Γ. Assume that f is nondegenerate over $\mathbb{F}_q(t)$, so that we can define the resultant $r_f \in \mathbb{F}_q[t]$ as in section 2.2. We will always suppose that $r_f(0) \neq 0$.

We write:

$$S = \mathrm{Spec}(\mathbb{F}_q[t, \frac{1}{r_f}]), \qquad X = \mathrm{Spec}(\mathbb{F}_q[t, \frac{1}{r_f}][\mathbb{Z}^2]/(f)).$$

The obvious morphism $X \to S$ then defines a smooth family of curves.

As usual, we let $\mathbb{Q}_q$ denote the unique unramified extension of $\mathbb{Q}_p$ of degree n, and $\mathbb{Z}_q$ its ring of integers. Let $\mathfrak{f} \in \mathbb{Z}_q[\mathbb{Z}^2]$ be a lift of f which preserves both the Newton polygon and the degree in t. We can then lift X/S to $\mathbb{Z}_q$ in the following way:

$$\mathcal{S} = \mathrm{Spec}(\mathbb{Z}_q[t, \frac{1}{r_\mathfrak{f}}]), \qquad \mathcal{X} = \mathrm{Spec}(\mathbb{Z}_q[t, \frac{1}{r_\mathfrak{f}}][\mathbb{Z}^2]/(\mathfrak{f})),$$

where the family $\mathcal{X} \to \mathcal{S}$ is again defined in the obvious way.

Finally, we let

$$\mathbf{S} = \mathrm{Spec}(\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]), \qquad \mathbf{X} = \mathrm{Spec}(\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][\mathbb{Z}^2]/(\mathfrak{f})),$$

and $\mathbf{X} \to \mathbf{S}$ denote the generic fibers of $\mathcal{X}, \mathcal{S}$, and $\mathcal{X} \to \mathcal{S}$, respectively.

$\mathcal{X}/\mathcal{S}$ admits a relative normal crossing compactification $\mathcal{Y}/\mathcal{S}$ by the remark after Theorem 3.3.3. In particular, $\mathbf{X} \to \mathbf{S}$ admits a relative normal crossing compactification, so the relative algebraic De Rham cohomology $H^1_{DR}(\mathbf{X}/\mathbf{S})$ is a vector bundle by Theorem 3.4.2. Since $\mathcal{S}$ is affine, it is enough to consider the $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$-module of its global sections

$$M = H^0(\mathbf{S}, H^1_{DR}(\mathbf{X}/\mathbf{S}))$$

which is free of some finite rank r.

Let $[m_1, \ldots, m_r]$ be a basis for M, and $G \in M_{r,r}(\mathbb{Q}_q[t, \frac{1}{r_\mathcal{F}}])$ the matrix of the Gauss-Manin connection with respect to this basis.

$$\nabla m_j = \sum_{i=1}^{r} G_{ij} m_i \otimes dt.$$

3.6.2 The Frobenius structure

With the definitions and notation from section 3.2 we have

$$\begin{aligned} H^0(]\overline{S}[, \mathcal{O}_S^\dagger) &= \mathbb{Q}_q\langle t, z\rangle^\dagger/(zr_\mathfrak{f} - 1) \\ &= \mathbb{Q}_q\langle t, \frac{1}{r_\mathfrak{f}}\rangle^\dagger. \end{aligned}$$

By Theorem 3.4.10, the relative rigid cohomology $H^1_{rig}(X/S)$ is an overconvergent F-isocrystal on S, say for the p-th power Frobenius. To describe it, we first have to choose a lift F_p to $\mathbb{Q}_q\langle t, \frac{1}{r_\mathfrak{f}}\rangle^\dagger$ of the p-th power Frobenius on $\mathbb{F}_q[t, \frac{1}{r_f}]$. Let σ denote the unique lift of the p-th power map on $\mathbb{F}_q$ to $Gal(\mathbb{Q}_q/\mathbb{Q}_p)$. We then take F_p to be equal to σ on $\mathbb{Q}_q$ and $F_p(t) = t^p$.

The $\mathbb{Q}_q\langle t, \frac{1}{r_\mathfrak{f}}\rangle^\dagger$-module of global sections of $H^1_{rig}(X/S)$

$$\mathcal{M} = H^0(]\overline{S}[, H^1_{rig}(X/S))$$

is again free of rank r, with basis $[m_1, \ldots, m_r]$, by Theorem 3.4.12. Hence the isomorphism $\mathcal{F} : F_p^*\mathcal{M} \to \mathcal{M}$ can be represented by a matrix $A_p \in M_{r,r}(\mathbb{Q}_q\langle t, \frac{1}{r_\mathfrak{f}}\rangle^\dagger)$ such that

$$\mathcal{F}(m_j) = \sum_{i=1}^{r} (A_p)_{ij} m_i.$$

Remark. Note that as a map from $\mathcal{M}$ to itself $\mathcal{F}$ is not linear, but F_p-semilinear.

Recall that the Frobenius isomorphism commutes with the connection

$$\begin{array}{ccc} \mathcal{M} & \xrightarrow{\nabla} & \mathcal{M} \otimes \Omega^{1\,\dagger}_S \\ {\scriptstyle\mathcal{F}}\downarrow & & \downarrow{\scriptstyle\mathcal{F}\otimes dF_p} \\ \mathcal{M} & \xrightarrow{\nabla} & \mathcal{M} \otimes \Omega^{1\,\dagger}_S \end{array}$$

In terms of the matrices G, A_p, this commutativity condition becomes

$$(\frac{dA_p}{dt} + GA_p) \otimes dt = A_p G^{F_p} \otimes d(t^p).$$

So the Frobenius matrix A_p satisfies the first order differential equation:

$$\frac{dA_p}{dt} = pt^{p-1} A_p G^{F_p} - GA_p.$$

3.6.3 Affine fibers

For $\tau \in S(\mathbb{F}_q)$ we write

$$f_\tau = f|_{t=\tau} \in \mathbb{F}_q[\mathbb{Z}^2].$$

The Laurent polynomial f_τ defines a nondegenerate curve $C_\tau \subset \mathbb{T}^2_{\mathbb{F}_q}$. Recall that by Theorem 3.2.4

$$\begin{aligned} Z(C_\tau, T) = & \prod_{i=0}^{2} \det(1 - q(F_q^*)^{-1}T|H^i_{rig}(C_\tau))^{(-1)^{i+1}} \\ = & \frac{\det(1 - q(F_q^*)^{-1}T|H^1_{rig}(C_\tau))}{(1-qT)}, \end{aligned}$$

where the last equality holds because $H^2_{rig}(C_\tau) \cong 0$ (since C_τ is affine), and because F_q^* is the identity on $H^0_{rig}(C_\tau)$.

If $i_\tau : \tau \to S$ denotes the closed embedding, then by Theorem 3.4.11 we have

$$H^1_{rig}(C_\tau) \cong i_\tau^* H^1_{rig}(X/S).$$

Let $\hat{\tau} \in \mathcal{S}(\mathbb{Z}_q)$ denote the *Teichmüller lift* of τ, i.e. $\hat{\tau}$ reduces to τ and is fixed under $F_q = F_p^n$, or in other words $\hat{\tau}^q = \hat{\tau}$. The matrix $A_{p,\tau}$ of the action of F_p on $H^1_{rig}(C_\tau)$, with respect to the basis induced by $[m_1, \ldots, m_r]$, is then obtained by specializing the Frobenius matrix A_p at $t = \hat{\tau}$.

$$A_{p,\tau} = A_p(\hat{\tau}).$$

Since $F_q = F_p^n$, taking into account the F_p-semilinearity, the matrix of F_q^* on $H^1_{rig}(C_\tau)$ is given by

$$A_{q,\tau} = A_{p,\tau} A_{p,\tau}^{F_p} \ldots A_{p,\tau}^{F_p^{n-1}},$$

and

$$Z(C_\tau, T) = \frac{\det(1 - qA_{q,\tau}^{-1}T)}{(1-qT)}.$$

3.6.4 Complete fibers

Instead of working with the affine curves C_τ, it is both more natural, and will turn out to be more efficient, to consider instead their complete (so projective) nonsingular models.

Recall that $\mathcal{Y}/\mathcal{S}$ denotes the relative normal crossing compactification of $\mathcal{X}/\mathcal{S}$, obtained by taking its closure in the projective toric surface over $\mathbb{Z}_q[t, \frac{1}{r_\mathfrak{f}}]$ associated to Γ. Let $\mathbf{Y}/\mathbf{S}$ be the base change of this family to $\mathbb{Q}_q$. We have seen in section 3.5 that $H^1_{DR}(\mathbf{Y}/\mathbf{S})$ injects into $H^1_{DR}(\mathbf{X}/\mathbf{S})$, as a vector bundle with connection, and is the kernel of some residue map. Let $[n_1, \ldots, n_s]$ be a basis for this kernel N. Let us choose the basis $[m_1, \ldots, m_r]$ of M such that $m_i = n_i$ for all $1 \leq i \leq s$.

Let Y/S denote the base change of $\mathcal{Y}/\mathcal{S}$ to $\mathbb{F}_q$, and $\overline{C}_\tau$ the fiber of Y over τ. Clearly $\overline{C}_\tau$ is the complete nonsingular model of C_τ. Again, by section 3.5 $H^1_{rig}(Y/S)$ is a sub-overconvergent F-isocrystal of $H^1_{rig}(X/S)$, and by Theorem 3.4.12 $[m_1, \ldots, m_s]$ is also a basis of the $\mathbb{Q}_q\langle t, \frac{1}{r_\mathfrak{f}}\rangle^\dagger$-module

$$\mathcal{N} = H^0(]\overline{S}[, H^1_{rig}(Y/S)).$$

Since ∇ and $\mathcal{F}$ both respect $\mathcal{N}$, the matrices G, A_p contain $s \times s$ blocks H, B_p such that

$$\frac{dB_p}{dt} = pt^{p-1} B_p H^{F_p} - H B_p. \tag{3.1}$$

If we write

$$B_{p,\tau} = B_p(\hat{\tau}),$$

$$B_{q,\tau} = B_{p,\tau} B_{p,\tau}^{F_p} \ldots B_{p,\tau}^{F_p^{n-1}},$$

then $B_{q,\tau}$ is the matrix of F_q^* on $H^1_{rig}(\overline{C}_\tau)$ with respect to the basis induced by $[m_1, \ldots, m_s]$. Therefore

$$\begin{aligned} Z(\overline{C}_\tau, T) &= \prod_{i=0}^{2} \det(1 - q(F_q^*)^{-1} T | H^i_{rig}(\overline{C}_\tau))^{(-1)^{i+1}} \\ &= \frac{\det(1 - q(F_q^*)^{-1} T | H^1_{rig}(\overline{C}_\tau))}{(1-T)(1-qT)} \\ &= \frac{\det(1 - qB_{q,\tau}^{-1} T)}{(1-T)(1-qT)}. \end{aligned}$$

Remark. If I is the number of interior points of Γ, B is the number of points on its boundary, and g is the genus of the generic fiber of $\mathbf{Y}/\mathbf{S}$, then from section 3.1 we know that

$$\begin{aligned} g &= I, \\ r &= 2g + B - 1, \\ s &= 2g. \end{aligned} \tag{3.2}$$

The curve $\overline{C}_\tau$ is complete, so from Theorem 1.1.1 we know that the multiset of roots of the polynomial $\chi(T) = \det(1 - qB_{q,\tau}^{-1}T)$ is stable under the transformation $T \to q/T$. From this it follows that

$$\det B_{q,\tau} = q^g, \tag{3.3}$$

and

$$Z(\overline{C}_\tau, T) = \frac{\det(1 - B_{q,\tau}T)}{(1-T)(1-qT)}. \tag{3.4}$$

3.7 Integral structure on the cohomology

Let $f \in \mathbb{F}_q[\mathbb{Z}^2]$ be a nondegenerate Laurent polynomial, and let $[e_1, \ldots, e_r]$ denote a basis of the rigid cohomology space $H^1_{rig}(C)$ of the affine curve C that f defines inside of the torus $\mathbb{T}^2_{\mathbb{F}_q}$. The matrix A_p of the action of the p-th power Frobenius map F_p^* on $H^1_{rig}(C)$, with respect to the basis $[e_1, \ldots, e_r]$, then has entries in $\mathbb{Q}_q$. However, it is known that one can pick an *integral basis* $[e_1, \ldots, e_r]$ such that A_p is integral, i.e. has entries in $\mathbb{Z}_q$. We will not go into how to determine such a basis, but we will study how far a given basis is from being integral.

Let $\mathfrak{f}$ denote any lift of f to $\mathbb{Z}_q[\mathbb{Z}^2]$ that preserves the Newton polygon Γ of f, and let $\mathcal{C}$ denote the curve it defines inside of the torus $\mathbb{T}^2_{\mathbb{Z}_q}$. By Theorem 3.3.3, the curve $\mathcal{C}$ admits a relative normal crossing compactification $\overline{\mathcal{C}}$ over $\mathbb{Z}_q$ which can be obtained by taking its closure in the projective toric surface associated to Γ. Moreover, the divisor $\mathcal{D} = \overline{\mathcal{C}} \setminus \mathcal{C}$ is smooth in this case. We can define an *integral logarithmic De Rham complex* $\Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D}))$ as in section 3.5, but now with $\mathbb{Q}_q$ replaced by $\mathbb{Z}_q$. The *logarithmic De Rham cohomology* of the pair $(\overline{\mathcal{C}}, \mathcal{D})$ is then defined as

$$H^1_{DR}((\overline{\mathcal{C}}, \mathcal{D})/\mathbb{Z}_q) = \mathbb{H}^1(\overline{\mathcal{C}}, \Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D}))).$$

There is a natural map of complexes from $\Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D}))$ to the overconvergent De Rham complex $\Omega^{\bullet\,\dagger}_C$ of C, so that we get a map

$$\psi : H^1_{DR}((\overline{\mathcal{C}}, \mathcal{D})/\mathbb{Z}_q) \to H^1_{rig}(C).$$

Theorem 3.7.1. *The image of ψ is a $\mathbb{Z}_q$-lattice in $H^1_{rig}(C)$ that is invariant under F_p^*.*

Proof. This follows from two comparison theorems. First, $H^1_{DR}((\overline{\mathcal{C}}, \mathcal{D})/\mathbb{Z}_q)$ is isomorphic to the *log crystalline cohomology* $H^1_{crys}(\overline{\mathcal{C}}, \mathcal{D})$ of the pair $(\overline{\mathcal{C}}, \mathcal{D})$, by [34, Theorem 6.4]. Second, $H^1_{crys}(\overline{\mathcal{C}}, \mathcal{D}) \otimes_{\mathbb{Z}_q} \mathbb{Q}_q$ is isomorphic to $H^1_{rig}(C)$, as shown in [49, section 2.4]. Since all maps are the natural ones, the composition of these isomorphisms is ψ, and the theorem follows. □

Definition 3.7.2. Λ_{int} is the $\mathbb{Z}_q$-lattice $im(\psi)$ in $H^1_{rig}(C)$.

There is another $\mathbb{Z}_q$-lattice inside of $H^1_{rig}(C)$ that is natural to consider. Again we denote

$$\omega_{\mathfrak{f}} = \frac{dx}{xy\frac{\partial \mathfrak{f}}{\partial y}}.$$

Remark. Note that $\omega_{\mathfrak{f}}$ is a section of $\Omega^1_{\mathcal{C}}$, since there exist $\alpha, \beta, \gamma \in \mathbb{Z}_q[\mathbb{Z}^2]$ such that $\alpha\mathfrak{f} + \beta x\frac{\partial \mathfrak{f}}{\partial x} + \gamma y\frac{\partial \mathfrak{f}}{\partial y} = 1$, and we have $x\frac{\partial \mathfrak{f}}{\partial x}dx + y\frac{\partial \mathfrak{f}}{\partial y}dy = 0$ on $\mathcal{C}$, so that

$$\omega_{\mathfrak{f}} = \frac{dx}{xy\frac{\partial \mathfrak{f}}{\partial y}}(\beta x\frac{\partial \mathfrak{f}}{\partial x} + \gamma y\frac{\partial \mathfrak{f}}{\partial y}) = -\beta\frac{dy}{y} + \gamma\frac{dx}{x}.$$

It also clear that $\omega_{\mathfrak{f}}$ generates $H^0(\mathcal{C}, \Omega^1_{\mathcal{C}})$, since

$$\frac{dx}{x} = y\frac{\partial \mathfrak{f}}{\partial y}\omega_{\mathfrak{f}}, \qquad \frac{dy}{y} = -x\frac{\partial \mathfrak{f}}{\partial x}\omega_{\mathfrak{f}}.$$

From Theorem 3.1.4, and the comparison theorem of section 3.3, we know that $H^1_{rig}(C)$ is generated as a $\mathbb{Q}_q$-vector space by the set (of cohomology classes)

$$V = \{x^i y^j \omega_{\mathfrak{f}} | (i,j) \in 2\Gamma\}.$$

So we can define:

Definition 3.7.3. $\Lambda_{2\Gamma}$ is the $\mathbb{Z}_q$-lattice generated by V in $H^1_{rig}(C)$.

The following result of Kedlaya can be used to relate the lattices Λ_{int} and $\Lambda_{2\Gamma}$.

Theorem 3.7.4. *Let $\omega \in H^0(\mathcal{C}, \Omega^1_{\mathcal{C}})$ be an integral differential. Choose an integer d such that the pole order of ω along each component of $\mathcal{D}$ is at most p^{d+1}. Then the class of $p^d\omega$ in $H^1_{rig}(C)$ is contained in Λ_{int}.*

This follows from the following lemma.

Lemma 3.7.5. *Let $\mathcal{X}$ be a smooth scheme of relative dimension 1 over $\mathbb{Z}_q$, and let $\mathcal{D}$ denote a relative normal crossing divisor on $\mathcal{X}$, with complement $\mathcal{U}$. Suppose that $\omega \in H^0(\mathcal{U}, \Omega^1_{\mathcal{U}})$ is an integral differential, and let d be an integer such that the pole order of ω along each component of $\mathcal{D}$ is at most p^{d+1}. Then $p^d\omega$ represents a class in $H^1_{DR}(\mathcal{U}/\mathbb{Z}_q)$ in the image of $H^1_{DR}((\mathcal{X}, \mathcal{D})/\mathbb{Z}_q)$.*

Proof. The claim is Zariski and étale local on $\mathcal{X}$, because De Rham cohomology can be computed from a Čech resolution. Since $\mathcal{D}$ is a relative normal crossing

divisor, we may therefore reduce to the case that $\mathcal{X}$ is an open subscheme of $\mathbb{A}^1_{\mathbb{Z}_q}$, and $\mathcal{D}$ is either empty or the origin. In case $\mathcal{D}$ is empty there is nothing to prove. When $\mathcal{D}$ is the origin, let x denote the coordinate on $\mathbb{A}^1_{\mathbb{Z}_q}$. We can then write

$$\omega = (\sum_{i \geq -p^{d+1}+1}^{-1} \omega_i x^i \frac{dx}{x}) + \omega',$$

with $\omega_i \in \mathbb{Z}_q$ for all i, and $\omega' \in H^0(\mathcal{X}, \Omega^1_{\mathcal{X}}(\log(\mathcal{D})))$. All of the terms between brackets can be integrated at the cost of dividing by at worst p^d. This implies that the class of $p^d\omega$ in $H^1_{DR}(\mathcal{U}/\mathbb{Z}_q)$ lies in the image of $H^1_{DR}((\mathcal{X}, \mathcal{D})/\mathbb{Z}_q)$. □

Corollary 3.7.6. *For each of the edges t_i of Γ, let e_i denote the inward pointing normal vector that is minimally integral. Denote $N_i = \langle p_i, e_i \rangle$ for some arbitrary point p_i on t_i. Let $N < 0$ be the minimum of the N_i, and d such that $p^{d+1} \geq -N + 1$. Then*

$$p^d \Lambda_{2\Gamma} \subset \Lambda_{int}.$$

Proof. By Theorem 3.1.2 and Theorem 3.1.3, the pole order of an element of V in a component of $\mathcal{D}$ is at most $-N + 1$. □

We now show that we also have an inclusion of lattices going the other way:

Theorem 3.7.7. $\Lambda_{int} \subset \Lambda_{2\Gamma}$.

Proof. Let the divisor $\mathcal{D}_\Gamma$ be defined as D_Γ in section 3.1, but this time over $\mathbb{Z}_q$. Since Γ contains the origin, $\mathcal{D}_\Gamma$ is a nonzero effective divisor. Hence there is an inclusion of complexes

$$\Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \subset \Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma),$$

from which we get a map

$$\phi : \mathbb{H}^1(\overline{\mathcal{C}}, \Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D}))) \to \mathbb{H}^1(\overline{\mathcal{C}}, \Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)),$$

of hypercohomology spaces over $\mathbb{Z}_q$.

Now the kernel of ϕ is a torsion module. The easiest way to see this is to note that after (flat) base change to $\mathbb{Q}_q$, the isomorphism

$$H^1_{DR}((\mathcal{C}, \mathcal{D}) \otimes \mathbb{Q}_q) \cong H^1_{DR}(\mathcal{C} \otimes \mathbb{Q}_q)$$

from Theorem 3.5.2 factorizes over ϕ.

We claim that $H^1(\overline{\mathcal{C}}, \mathcal{O}(\mathcal{D}_\Gamma)) = 0$. Since $\overline{\mathcal{C}}/\mathbb{Z}_q$ is proper, and by Nakayama's lemma, it is enough to show this for both fibers (i.e. the one over $\mathbb{F}_q$ and the

one over $\mathbb{Q}_q$) separately. However, we know from Theorem 3.1.2 that over a field $D_\Gamma - D$ is a canonical divisor. Therefore, by *Serre duality* and since D is effective

$$H^1(\overline{C}, \mathcal{O}_{\overline{C}}(D_\Gamma)) \cong H^0(\overline{C}, \mathcal{O}_{\overline{C}}(-D)) \cong 0.$$

To compute $\mathbb{H}^1(\overline{\mathcal{C}}, \Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}(\mathcal{D}_\Gamma))$, we consider the *hypercohomology spectral sequence*

$$E_1^{pq} = H^q(\overline{\mathcal{C}}, \Omega^p_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)) \Rightarrow \mathbb{H}^{p+q}(\overline{\mathcal{C}}, \Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)).$$

For the first sheet E_1^{pq}, we get

$$\begin{aligned} H^1(\overline{\mathcal{C}}, \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)) &\xrightarrow{d_1} H^1(\overline{\mathcal{C}}, \Omega^1_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)) \\ H^0(\overline{\mathcal{C}}, \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)) &\xrightarrow{d_1} H^0(\overline{\mathcal{C}}, \Omega^1_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)), \end{aligned}$$

where the maps are the exterior derivative maps. Since $H^1(\overline{\mathcal{C}}, \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)) = 0$, from this we compute

$$\begin{aligned} E_2^{10} &= H^0(\overline{\mathcal{C}}, \Omega^1_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma))/d(H^0(\overline{\mathcal{C}}, \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma))), \\ E_2^{01} &= 0, \end{aligned}$$

which implies that

$$\begin{aligned} \mathbb{H}^1(\overline{\mathcal{C}}, \Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)) &= H^0(\overline{\mathcal{C}}, \Omega^1_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma))/d(H^0(\overline{\mathcal{C}}, \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma))), \\ &= \mathbb{Z}_q[2\Gamma]\omega_{\mathfrak{f}}/d(\mathbb{Z}_q[\Gamma]), \end{aligned}$$

so that we get

$$\Lambda_{int} = im\ \mathbb{H}^1(\overline{\mathcal{C}}, \Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D}))) \subset im\ \mathbb{H}^1(\overline{\mathcal{C}}, \Omega^\bullet_{\overline{\mathcal{C}}}(\log(\mathcal{D})) \otimes \mathcal{O}_{\overline{\mathcal{C}}}(\mathcal{D}_\Gamma)) = \Lambda_{2\Gamma},$$

as lattices in $H^1_{rig}(C)$. □

All of this can be used to bound the p-adic valuations of the denominators of the Frobenius matrices that appear in section 3.6. Recall from that section that $[m_1, \ldots, m_r]$ denotes a basis for the relative cohomology module M of an affine family of nondegenerate curves, and $[n_1, \ldots, n_s] = [m_1, \ldots, m_s]$ a basis for the relative cohomology module N of its compactification. Also recall that $\tau \in S(\mathbb{F}_q)$ is a (rational) point on the base space of the family, that C_τ denotes the fiber of the family at τ, and $\overline{C}_\tau$ the compactification of C_τ.

If we let $\Lambda_{[m_1,\ldots,m_r],\tau}$ be the $\mathbb{Z}_q$-lattice generated by the images of $m_1, \ldots, m_r$ in $H^1_{rig}(C_\tau)$, and we define the $\mathbb{Z}_q$-lattices $\Lambda_{int,\tau}, \Lambda_{2\Gamma,\tau} \subset H^1_{rig}(C_\tau)$ as before, then there exist $d_1, d_2 \in \mathbb{Z}$ such that

$$p^{d_1}\Lambda_{[m_1,\ldots,m_r],\tau} \subset \Lambda_{2\Gamma,\tau}, \qquad p^{d_2}\Lambda_{2\Gamma,\tau} \subset \Lambda_{[m_1,\ldots,m_r],\tau}.$$

Remark. Note that we can find such d_1, d_2 effectively, and we can take them to be independent of τ. We first compute $U_{ijk}, V_{ijk} \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ such that

$$m_k = \sum_{(i,j)\in 2\Gamma} U_{ijk} x^i y^j \omega_\mathfrak{f},$$

$$x^i y^j \omega_\mathfrak{f} = \sum_{k=1}^{r} V_{ijk} m_k,$$

in M. Then we can take d_1, d_2 such that $p^{d_1} U_{ijk}, p^{d_2} V_{ijk} \in \mathbb{Z}_q[t, \frac{1}{r_\mathfrak{f}}]$ for all i, j, k.

Now by Corollary 3.7.6, and Theorem 3.7.7, we have

$$p^{d+d_1} \Lambda_{[m_1,\dots,m_r],\tau} \subset \Lambda_{int,\tau}, \qquad p^{d_2} \Lambda_{int,\tau} \subset \Lambda_{[m_1,\dots,m_r],\tau}. \tag{3.5}$$

Recall that $A_{p,\tau}$ and $A_{q,\tau}$ are the matrices of the actions of F_p^* and F_q^* on $H^1_{rig}(C_\tau)$, with respect to the basis induced by $[m_1, \dots, m_r]$, and $B_{p,\tau}$ and $B_{q,\tau}$ the matrices of the Frobenius actions on the rigid cohomology space $H^1_{rig}(\overline{C}_\tau)$, with respect to the basis induced by $[m_1, \dots, m_s]$. Let v_p denote the p-adic valuation on matrices over $\mathbb{Q}_q$.

Theorem 3.7.8. $v_p(A_{p,\tau}), v_p(A_{q,\tau}) \geq -(d + d_1 + d_2)$.

Proof. We already know that F_p^* and F_q^* map $\Lambda_{int,\tau}$ into itself, so by (3.5)

$$F_p^* \Lambda_{[m_1,\dots,m_r],\tau} \subset p^{-(d+d_1)} \Lambda_{int,\tau} \subset p^{-(d+d_1+d_2)} \Lambda_{[m_1,\dots,m_r],\tau},$$

and similarly for F_q^*, so that the result follows. □

Corollary 3.7.9. $v_p(B_{p,\tau}), v_p(B_{q,\tau}) \geq -(d + d_1 + d_2)$.

Proof. The matrices $B_{p,\tau}$ and $B_{q,\tau}$ are blocks of $A_{p,\tau}$ and $A_{q,\tau}$, respectively. □

We can say something similar about the matrices $B_{p,\tau}^{-1}$ and $B_{q,\tau}^{-1}$:

Theorem 3.7.10. $v_p(pB_{p,\tau}^{-1}), v_p(qB_{q,\tau}^{-1}) \geq -(d + d_1 + d_2)$.

Proof. As a consequence of Poincaré duality for crystalline cohomology, we have that $p(F_p^*)^{-1}$ and $q(F_q^*)^{-1}$ map $\Lambda_{int,\tau} \cap H^1_{rig}(\overline{C}_\tau)$ into itself, so by (3.5)

$$p(F_p^*)^{-1} \Lambda_{[m_1,\dots,m_s],\tau} = p(F_p^*)^{-1}(\Lambda_{[m_1,\dots,m_r],\tau} \cap H^1_{rig}(\overline{C}_\tau)) \subset p^{-(d+d_1)}(\Lambda_{int,\tau} \cap H^1_{rig}(\overline{C}_\tau))$$

$$\subset p^{-(d+d_1+d_2)}(\Lambda_{[m_1,\dots,m_r],\tau} \cap H^1_{rig}(\overline{C}_\tau)) = p^{-(d+d_1+d_2)} \Lambda_{[m_1,\dots,m_s],\tau},$$

and similarly for $q(F_q^*)^{-1}$, so that the result follows. □

Recall from section 3.6 that B_p and $B_q \in M_{s,s}(\mathbb{Q}_q\langle t, \frac{1}{r_\mathfrak{f}}\rangle^\dagger)$ denote the matrices of the p-th and q-th power Frobenius map on the relative rigid cohomology $H^1_{rig}(Y/S)$ of the family of (complete) nondegenerate curves Y/S defined by $\mathfrak{f}$. Let v_p denote the p-adic valuation on matrices over $\mathbb{Q}_q\langle t, \frac{1}{r_\mathfrak{f}}\rangle^\dagger$ (see Definition 3.8.4).

Corollary 3.7.11.

$$v_p(B_p), v_p(pB_p^{-1}), v_p(B_q), v_p(qB_q^{-1}) \geq -(d + d_1 + d_2)$$

Proof. This follows from Corollary 3.7.9 and Theorem 3.7.10, since they hold for all $\tau \in S(\overline{\mathbb{F}}_q)$. □

Remark. The lattices in this section can also be defined globally for the whole family. Let Λ_{int} denote the $\mathbb{Z}_q[t, \frac{1}{r_\mathfrak{f}}]$-lattice in $H^1_{rig}(X/S)$ that is the image of the *relative logarithmic De Rham cohomology* $H^1_{DR}((\mathcal{X}, \mathcal{D})/\mathcal{S})$, and let $\Lambda_{2\Gamma}$ and $\Lambda_{[m_1,\ldots,m_r]}$ denote the $\mathbb{Z}_q[t, \frac{1}{r_\mathfrak{f}}]$-lattices in $H^1_{rig}(X/S)$, generated by the sets $V = \{x^i y^j \omega_\mathfrak{f} | (i,j) \in 2\Gamma\}$ and $\{m_1, \ldots, m_r\}$, respectively. Then for every $\tau \in S(\mathbb{F}_q)$, all of these $\mathbb{Z}_q[t, \frac{1}{r_\mathfrak{f}}]$-lattices specialize to the corresponding $\mathbb{Z}_q$-lattices in $H^1_{rig}(C_\tau)$.

3.8 The differential equation

In section 3.6, we saw that the p-th power Frobenius matrix Φ with respect to a fixed basis of the relative rigid cohomology of a (nice) family of curves satisfies a p-adic differential equation of the form

$$\frac{d\Phi}{dt} = pt^{p-1}\Phi N^{F_p} - N\Phi, \tag{3.6}$$

where N is the matrix of the Gauss-Manin connection ∇ with respect to the chosen basis. In the notation of section 3.6, the matrix Φ could be either A_p or B_p, and the matrix N would then be G or H, respectively.

In our computations we need *effective bounds* on the rate of convergence of the solutions of this (and related) equation(s). In this section we give an overview of the relevant results, which can be found in [39],[35].

Remark. Since we always assume that $r_f(0) \neq 0$ (and hence $r_\mathfrak{f}(0) \neq 0 \bmod p$), our connection matrix N and Frobenius matrix Φ will not have a pole in $t = 0$. Therefore, we can expand these matrices as (formal) power series in t.

First we study the *horizontal sections* of the connection ∇.

Theorem 3.8.1. *Let $N = \sum_{i=0}^{\infty} N_i t^i$ be a $d \times d$ matrix over $\mathbb{Q}_q[[t]]$. There exists a unique $d \times d$ matrix $C = \sum_{i=0}^{\infty} C_i t^i$ over $\mathbb{Q}_q[[t]]$ with $C_0 = I$ satisfying*

$$\frac{dC}{dt} + NC = 0.$$

Proof. Extracting the coefficient of t^k we find

$$(k+1)C_{k+1} = -\sum_{i=0}^{k} N_{k-i} C_i$$

which determines C_{k+1} uniquely in terms of $C_0, \ldots, C_k$. □

Remark. The matrix C is called a *fundamental matrix* of the connection ∇. Note that it is always invertible.

From such a fundamental matrix one can derive a solution to equation (3.6) by the following theorem.

Theorem 3.8.2. *Let $N = \sum_{i=0}^{\infty} N^i t^i$, $C = \sum_{i=0}^{\infty} C_i t^i$ and $\Phi = \sum_{i=0}^{\infty} \Phi_i t^i$ be $d \times d$ matrices over $\mathbb{Q}_q[[t]]$ satisfying the following conditions:*

1. $\frac{dC}{dt} + NC = 0$,
2. $C_0 = I$,
3. $\frac{d\Phi}{dt} = pt^{p-1}\Phi N^{F_p} - N\Phi$.

Then we have

$$\Phi = C\Phi_0 (C^{F_p})^{-1}.$$

Proof. First one easily verifies that

$$\frac{d}{dt}(C^{-1}) = -C^{-1}\frac{dC}{dt}C^{-1}, \qquad \frac{d}{dt}(C^{F_p}) = pt^{p-1}(\frac{dC}{dt})^{F_p}.$$

Using the differential equations for the matrices Φ, C, we compute

$$\begin{aligned}
\frac{d}{dt}(C^{-1}\Phi C^{F_p}) &= C^{-1}\Phi\frac{d}{dt}(C^{F_p}) + C^{-1}\frac{d\Phi}{dt}C^{F_p} + \frac{d}{dt}(C^{-1})\Phi C^{F_p}, \\
&= C^{-1}\Phi pt^{p-1}(\frac{dC}{dt} + NC)^{F_p} - C^{-1}(\frac{dC}{dt} + NC)C^{-1}\Phi C^{F_p}, \\
&= 0.
\end{aligned}$$

Hence $C^{-1}\Phi C^{F_p}$ is a constant matrix, and substituting $t = 0$ we find

$$C^{-1}\Phi C^{F_p} = \Phi_0.$$

□

Remark. By assumption, we have that $r_{\mathfrak{f}}(0) \neq 0 \bmod p$, so if we expand $\frac{1}{r_{\mathfrak{f}}}$ into a power series in t, then the coefficients are integral. Since the matrices N, Φ have entries entries in $\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}]$ and $\mathbb{Q}_q\langle t, \frac{1}{r_{\mathfrak{f}}}\rangle^\dagger$ respectively, the coefficients of their power series expansions in t are *bounded*.

Definition 3.8.3. Let $\mathbb{Q}_q[[t]]_0$ be the subring of $\mathbb{Q}_q[[t]]$ consisting of series with bounded coefficients:

$$\mathbb{Q}_q[[t]]_0 = \mathbb{Z}_q[[t]] \otimes_{\mathbb{Z}_q} \mathbb{Q}_q.$$

Let $|\cdot|_0$ denote the supremum norm on elements of, and more generally on matrices over, the ring $\mathbb{Q}_q[[t]]_0$. Note that the supremum runs both over the powers of t, and over the entries of the matrix. Let $v_{p,0}$ denote the corresponding valuation, i.e. $|\cdot|_0 = p^{-v_{p,0}(\cdot)}$.

Since $\mathbb{Q}_q\langle r, \frac{1}{r_{\mathfrak{f}}}\rangle^\dagger$ is contained in $\mathbb{Q}_q[[t]]_0$, the norm $|\cdot|_0$ is also defined on (matrices over) this ring. Geometrically $|\cdot|_0$ is the supremum norm on the open unit disk around zero.

Usually, one considers the following more natural norm on $\mathbb{Q}_q\langle r, \frac{1}{r_{\mathfrak{f}}}\rangle^\dagger$:

Definition 3.8.4. Let $|\cdot|$ denote the norm on (matrices over) the ring

$$\mathbb{Q}_q\langle t, \frac{1}{r_{\mathfrak{f}}}\rangle^\dagger = \mathbb{Q}_q\langle t, z\rangle^\dagger/(zr_{\mathfrak{f}} - 1)$$

which is induced by the supremum norm on $\mathbb{Q}_q\langle t, z\rangle^\dagger$, and let v_p be the corresponding valuation, i.e. $|\cdot| = p^{-v_p(\cdot)}$.

Geometrically $|\cdot|$ is the supremum norm on the complement of the union of all the open disks of radius 1 around the zeroes of $r_{\mathfrak{f}}$. Since this complement contains the open disk of radius 1 around 0, on the ring $\mathbb{Q}_q\langle r, \frac{1}{r_{\mathfrak{f}}}\rangle^\dagger$ we have that $|\cdot|_0 \leq |\cdot|$, or equivalently $v_{p,0}(\cdot) \geq v_p(\cdot)$.

A first bound on the rate of p-adic convergence of the matrix C is given by the following theorem.

Theorem 3.8.5. *Let $N = \sum_{i=0}^{\infty} N^i t^i$ and $C = \sum_{i=0}^{\infty} C_i t^i$ be $d \times d$ matrices over $\mathbb{Q}_q[[t]]$ satisfying the following conditions:*

1. *the matrix N has entries in $\mathbb{Q}_q[[t]]_0$,*

2. $C_0 = I$,
3. $\frac{dC}{dt} + NC = 0$,
4. *the matrix C converges on the open unit disk.*

Then for every positive integer i,

$$v_p(C_i) \geq -(d-1)(\lfloor \log_p i \rfloor + \min\{0, v_{p,0}(N)\}).$$

Proof. This is a theorem of Dwork and Robba [20]. It is a special case of [39, Theorem 18.2.1]. □

Remark. The condition on the convergence of C on the open unit disk is automatically satisfied here, because of the presence of the Frobenius structure Φ [39, Corollary 17.2.2].

Note that apart from this remark, Theorem 3.8.5 does not use the existence of the matrix Φ at all, it applies to all so called *solvable* p-adic differential equations. One can usually get a sharper bound by taking the Frobenius structure into account:

Theorem 3.8.6. *Let $N = \sum_{i=0}^{\infty} N_i t^i$, $C = \sum_{i=0}^{\infty} C_i t^i$, and $\Phi = \sum_{i=0}^{\infty} \Phi_i t^i$ be $d \times d$ matrices over $\mathbb{Q}_q[[t]]$ satisfying the following conditions:*

1. *the matrix Φ has entries in $\mathbb{Q}_q[[t]]_0$,*
2. *$C_0 = I$ and Φ_0 is invertible,*
3. $\frac{dC}{dt} + NC = 0$,
4. $\frac{d\Phi}{dt} = pt^{p-1}\Phi N^{F_p} - N\Phi$.

Then for every positive integer i,

$$v_p(C_i) \geq (v_p(\Phi_0^{-1}) + v_{p,0}(\Phi))\lceil \log_p i \rceil.$$

Proof. This is a special case of [39, Theorem 18.3.3]. □

Note that upto this point, we have only looked at local solutions of our equations at $t = 0$. We could have done the same thing at any point $t = a$ such that $r_{\mathfrak{f}}(a) \neq 0 \bmod p$. However, near the points where the connection ∇ has a pole, the situation is going to be different.

We already know from Theorem 3.4.6 that the connection ∇ is regular singular, i.e. that for any point of $\mathbb{P}^1_{\mathbb{Q}_q}$ we can make a change of basis, such that the matrix N of ∇ has at most a *simple* pole at this point.

Definition 3.8.7. Let z be a (geometric) point of $\mathbb{A}^1_{\mathbb{Q}_q}$, and suppose that the matrix N has at most a simple pole at z. The *residue matrix* of N at z is then defined as $((t-z)N)|_{t=z}$. The *exponents* of N at z are defined to be the eigenvalues of the residue matrix. For the remaining point $\infty \in \mathbb{P}^1_{\mathbb{Q}_q}$, we may use the same definitions at $z=0$, after applying the coordinate change $t \to t^{-1}$.

Remark. If N is the matrix of a Gauss-Manin connection ∇ with respect to some basis of the relative cohomology, then its exponents are rational numbers. As elements of $\mathbb{Q}$ these exponents depend on the chosen basis, but as elements of $\mathbb{Q}/\mathbb{Z}$ they are invariants of ∇.

The following theorem of Kedlaya, under some conditions, describes the rate of convergence of the Frobenius matrix Φ around the points where the connection ∇ has a pole.

Theorem 3.8.8. *Let U be an open dense subscheme of $\mathbb{P}^1_{\mathbb{Q}_q}$ with complement Z. Suppose that $\mathcal{E}$ is a vector bundle on U equiped with a connection ∇. Let z be a (geometric) point of Z at which ∇ is regular singular with exponents contained in $\mathbb{Q} \cap \mathbb{Z}_p$, and assume that Z does not contain any other points with the same reduction modulo p. Suppose that $[v_1, \dots, v_n]$ is a basis of $\mathcal{E}$ with respect to which the matrix N of ∇ has at most a simple pole at z, and let $\{\lambda, \dots, \lambda_n\}$ denote the exponents of N at z. Put*

$$\mu_1 = \lfloor -p \min_i\{\lambda_i\} + \max_i\{\lambda_i\} \rfloor.$$

Fix a positive integer ν, and define

$$\mu_2 = \begin{cases} 0 & \text{if } N \text{ does not have a pole at } z \\ 0 & \text{if } z \in \{0, \infty\} \\ \nu - 1 & \text{otherwise} \end{cases}$$

Let V denote the rigid analytic subspace of $\mathbb{P}^1_{\mathbb{Q}_q}$ that is the complement of the union of the open disks of radius 1 around the points of Z. Suppose that $\mathcal{E}$ admits a (p-th power) Frobenius structure $\mathcal{F}$ on a strict neighbourhood of V. Let D denote the differential operator such that $\nabla(v) = Dv \otimes dt$. Let $\Phi^{(i)}$ be the matrix of $\frac{1}{i!}\mathcal{F}(D^i)$ with respect to the basis $[v_1, \dots, v_n]$, and denote $\Phi = \Phi^{(0)}$. Suppose that $v_p(\Phi^{(i)}) \geq 0$ for all $i \geq 0$. Then Φ is congruent modulo p^ν to a matrix of rational functions of order greater than or equal to $-(\mu_1 + p\mu_2)$ at z.

This is a slight reformulation (and correction) of [35, Theorem 6.5.10]. The proof proceeds in several steps. We start with the following lemma:

Lemma 3.8.9. *Let $N = \sum_{i=-1}^{\infty} N_i t^i$ be a $d \times d$ matrix such that tN converges on the open unit disk and N_{-1} is a nilpotent matrix. Let $\Phi = \sum_{i=-\infty}^{\infty} \Phi_i t^i$ be a $d \times d$ matrix that converges on some open annulus of outer radius 1. Suppose that N, Φ satisfy equation (3.6). Then $\Phi_i = 0$ for all $i < 0$, so that Φ converges on the whole open unit disk.*

Proof. See [39, Proposition 17.5.1]. □

When the exponents of N at 0 are not necessarily zero, this can be generalized as follows.

Corollary 3.8.10. *Let $N = \sum_{i=-1}^{\infty} N_i t^i$ be a $d \times d$ matrix such that tN converges on the open unit disk and the eigenvalues $\lambda_1, \ldots, \lambda_d$ of N_{-1} are rational numbers with denominators coprime to p. Let $\Phi = \sum_{i=-\infty}^{\infty} \Phi_i t^i$ be a $d \times d$ matrix that converges on some open annulus of outer radius 1. Suppose that N, Φ satisfy equation (3.6). Then $\Phi_i = 0$ whenever*

$$i < p \min_i\{\lambda_i\} - \max_i\{\lambda_i\}.$$

Proof. First we may adjoin $t^{1/k}$ for k coprime to p if necessary, to reduce to the case where $\lambda_1, \ldots, \lambda_d \in \mathbb{Z}$. In that case, by applying so called *shearing transformations*, one can find an invertible $d \times d$ matrix W over $\mathbb{Q}_q(t)$ such that the matrix

$$N' = W^{-1}NW + W^{-1}\frac{dW}{dt}$$

still has (at most) a simple pole at $t = 0$, but now with all exponents equal to 0. Moreover, one can ensure that $t^b W$ and $t^{-a}W^{-1}$ do not have a pole at $t = 0$, for $a = \min_i\{\lambda_i\}$ and $b = \max_i\{\lambda_i\}$. For more details, see [35, Lemma 5.1.6]. If we change basis to the basis given by the colums of W, then

$$N \to N',$$

$$\Phi \to \Phi' = W^{-1}\Phi W^{F_p}.$$

Now Lemma 3.8.9 can be applied to the pair N', Φ', so that $\Phi'_i = 0$ for all $i < 0$. Since $\Phi = W\Phi'(W^{-1})^{F_p}$, this implies that $\Phi_i = 0$ for all $i < pa - b$. □

Recall that in section 3.6 we *chose* the p-th power Frobenius lift $F_p : \mathbb{P}^1_{\mathbb{Q}_q} \to \mathbb{P}^1_{\mathbb{Q}_q}$ satisfying $F_p(t) = t^p$. For this lift a Frobenius structure $\mathcal{F} : F_p^*\mathcal{E} \cong \mathcal{E}$ on a vector bundle with connection $\mathcal{E}$ leads to a differential equation like (3.6). However, we could just as well have chosen a different lift $F'_p : \mathbb{P}^1_{\mathbb{Q}_q} \to \mathbb{P}^1_{\mathbb{Q}_q}$, resulting in a slightly different differential equation and Frobenius structure. The following lemma allows one to change from one Frobenius lift to another.

Lemma 3.8.11. *Let U be an open dense subscheme of $\mathbb{P}^1_{\mathbb{Q}_q}$ with complement Z. Suppose that $\mathcal{E}$ is a vector bundle on U equiped with a connection ∇. Let V denote the rigid analytic subspace of $\mathbb{P}^1_{\mathbb{Q}_q}$ that is the complement of the union of the open disks of radius 1 around the points of Z. Suppose that $\mathcal{E}$ admits a (p-th power) Frobenius structure $\mathcal{F}$ on a strict neighbourhood of V with respect to a Frobenius lift F_p. Let $F'_p : \mathbb{P}^1_{\mathbb{Q}_q} \to \mathbb{P}^1_{\mathbb{Q}_q}$*

be any other Frobenius lift. Then $\mathcal{E}$ also admits a Frobenius structure $\mathcal{F}'$ on a strict neighbourhood of V with respect to F'_p, defined by

$$\Phi' = \mathcal{F}'(v) = \sum_{i=0}^{\infty} \frac{1}{i!}(F'_p(t) - F_p(t))^i \mathcal{F}(D^i(v)).$$

Proof. See [39, Proposition 17.3.1]. □

Now we finally get to the proof of Theorem 3.8.8:

Proof of Theorem 3.8.8. Using the Frobenius lift F'_p with $F'_p(t-z) = (t-z)^p$ and translating z to the origin, we can apply Corollary 3.8.10 to see that the Frobenius matrix Φ' has a pole of order at most μ_1 at z. If we convert back to the original Frobenius lift F_p with $F_p(t) = t^p$, then by using Lemma 3.8.11, noting that this is not necessary when $z \in \{0, \infty\}$, we find that modulo p^ν the Frobenius matrix Φ has pole of order at most $\mu_1 + p\mu_2$ at z. □

Remark. In [35, Theorem 6.5.10], the condition $v_p(\Phi^{(i)}) \geq 0$ is only included for $i = 0$. We have asked Kedlaya about this, and he agrees that it is not enough for the theorem to hold. In our applications $\mathcal{E}$ is the relative cohomology of a family of nondegenerate curves. If we take $[v_1, \ldots, v_n]$ to be a basis for the relative (log)-crystalline cohomology of the family, then $v_p(\Phi^{(i)}) \geq 0$ for all i, so that Theorem 3.8.8 applies. For a more general basis $[v_1, \ldots, v_n]$, the sequence $v_p(\Phi^{(i)})$ is bounded below. So we can apply Theorem 3.8.8 to some twist of $\mathcal{E}$, for which the $v_p(\Phi^{(i)})$ are nonnegative, to obtain a similar bound.

Remark. Suppose that $[v_1, \ldots, v_n]$ is a basis for $\mathcal{E}$ as in Theorem 3.8.8, and let $[w_1, \ldots, w_n]$ denote another basis for $\mathcal{E}$, such that $v_j = \sum_{i=1}^n W_{ij} w_j$, with $W \in M_{n,n}(\mathbb{Q}_q(t))$. Then the matrix Φ' of $\mathcal{F}$ with respect to $[w_1, \ldots, w_n]$ satisfies

$$\Phi' = W\Phi(W^{F_p})^{-1}.$$

This implies that modulo

$$p^{\nu + \min\{v_p(W),0\} + \min\{v_p(W^{-1}),0\}},$$

the matrix Φ' is congruent to a matrix of rational functions of order greater than or equal to

$$-(\mu_1 + p\mu_2) + ord_z(W) + p\ ord_z(W^{-1}).$$

at z. This is not stated correctly either in [35, Theorem 6.5.10].

3.9 Sketch of our algorithm

Remember that our goal is to compute the zeta function of a nondegenerate curve over some large finite field. In this chapter we have recalled a lot of theory which will help us to do so. The next chapter is devoted to the computational details of our algorithm, but in this section we try to give an idea of how we are going to use the theory in our computations.

We use the definitions and the notation from section 3.6.

3.9.1 The deformation method

The idea of the deformation method is as follows.

1. Choose $f \in \mathbb{F}_q[t][\mathbb{Z}^2]$ in such a way that the Frobenius action on $H^1_{rig}(\overline{C}_0)$ (i.e. the matrix $B_{p,0}$) is relatively easy to compute by existing algorithms or special tricks.
2. Write down and solve the differential equation (3.1) for B_p, with the initial condition
$$B_p(0) = B_{p,0}.$$
3. For some $\tau \in S(\mathbb{F}_q)$ compute a Teichmüller lift $\hat{\tau}$, and substitute it into B_p to obtain $B_{p,\tau}$.
4. Compute $B_{q,\tau} = B_{p,\tau} B_{p,\tau}^{F_p} \dots B_{p,\tau}^{F_p^{n-1}}$.
5. Finally, compute the zeta function:
$$Z(\overline{C}_\tau, T) = \frac{\det(1 - qB_{q,\tau}^{-1}T)}{(1-T)(1-qT)}.$$

If we want to compute the zeta function of a concrete nondegenerate curve defined by some $f_1 \in \mathbb{F}_q[\mathbb{Z}^2]$, then the best choice for the family is
$$f = (1-t)f_0 + tf_1,$$
for some 'easy' nondegenerate $f_0 \in \mathbb{F}_p[\mathbb{Z}^2]$ with the same Newton polygon as f_1. Note that the fiber at $t = 0$ is then defined by f_0, and the one at $t = 1$ by f_1. In this case all computations have to be done in the field $\mathbb{Q}_q$, which in general has large degree over $\mathbb{Q}_p$.

However, if we are only interested in computing the zeta functions of some random nondegenerate curves with a given Newton polygon, which is often the case in the applications, then it is better to take a family defined over the prime field, i.e. $f \in \mathbb{F}_p[t][\mathbb{Z}^2]$. In this case the differential equation can be solved over $\mathbb{Q}_p$ which is a lot more efficient. Also, once we have computed B_p, we can substitute many different $\tau \in S(\mathbb{F}_q)$, so that we can 'recycle' the first two steps of the computation.

Remark. The same method can be applied to the affine curves C_τ. However, in this case all of the cohomology spaces will be of dimension r instead of s, and we have to solve an $r \times r$ matrix differential equation instead of an $s \times s$ one. Recall from (3.2) that $r = s + B - 1$, where B is the number of points on the boundary of Γ. So for small Newton polygons, r will usually be significantly bigger than s, and this will result in a less efficient algorithm. Moreover, in applications one is usually only interested in $Z(\overline{C}_\tau, T)$. Therefore, we will work exclusively with the complete curves $\overline{C}_\tau$.

3.9.2 Finite precision

When, for example, we talked about 'computing' $B_{q,\tau}$, what did we actually mean? Note that it is a matrix consisting of elements of $\mathbb{Q}_q$. However, like real numbers, elements of a p-adic field can in general only be approximated with *finite precision*. Therefore, it seems we will only be able to compute $Z(\overline{C}_\tau, T)$ with finite p-adic precision as well. Let us explain briefly why working with sufficient p-adic precision will give us the *exact* zeta function.

Recall from section 3.6 that

$$Z(\overline{C}_\tau, T) \quad = \quad \frac{\chi(T)}{(1-T)(1-qT)},$$

with

$$\chi(T) = \det(1 - B_{q,\tau} T | H^1_{rig}(\overline{C}_\tau)) = \prod_{j=1}^{2g} (1 - \alpha_{1,j} T) \in \mathbb{Z}[T],$$

where

1. the (multi)set of roots of χ is stable under the transformation $T \to q/T$,
2. $|\alpha_{1,j}| = q^{\frac{1}{2}}$ for all j, and for every embedding $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$.

Using this we find

$$\begin{aligned}\chi(T) &= \prod_{i=1}^{2g}(1-\alpha_{1,i}T),\\ &= \prod_{i=1}^{2g}\left(-\frac{qT}{\alpha_{1,i}}\right)(1-\frac{\alpha_{1,i}}{qT}),\\ &= q^gT^{2g}\chi(\frac{1}{qT}).\end{aligned}$$

If we write $\chi(T)=\sum_{i=0}^{2g}a_iT^i$, then this implies that $\chi(T)$ is already determined by $a_0,\ldots,a_g$. Since up to sign a_i is the i-th elementary symmetric polynomial in the $\alpha_{1,i}$, we have

$$|a_i| \leq \binom{2g}{i}q^{\frac{i}{2}}.$$

Therefore, if we know $\chi(T)$ with precision ν (i.e. modulo p^ν) for some ν satisfying $p^\nu \geq 2\binom{2g}{g}q^{\frac{g}{2}}$, then we know it exactly.

However, it is *not* sufficient to work with precision ν when computing the matrices $B_{p,0}, B_p, B_{p,\tau}$ and $B_{q,\tau}$. The problem is the *loss of precision* that occurs when one multiplies with a nonintegral element of $\mathbb{Q}_q$. We can correct for this by carrying more precision.

Lemma 3.9.1. *Let $a_1,\ldots,a_k \in \mathbb{Q}_q$ and $\nu\in\mathbb{N}$. If $a_1',\ldots,a_k' \in \mathbb{Q}_q$ are such that*

$$v_p(a_i-a_i') \geq \nu - \sum_{j\neq i}\min\{0, v_p(a_j)\}$$

for all i, then $v_p(a_1\ldots a_k - a_1'\ldots a_k') \geq \nu$.

So in our computations, we need lower bounds for the valuations of the numbers we are multiplying. Better bounds will imply a lower required precision, and a more efficient algorithm.

Chapter 4

The algorithm

In this chapter we work out the algorithm sketched in section 3.9. We will explain all of the steps of the computation in full detail.

4.1 Computing $r_{\mathfrak{f}}$

Let $\mathfrak{f} \in \mathbb{Z}_q[t][\mathbb{Z}^2]$ be a generically nondegenerate Laurent polynomial over $\mathbb{Z}_q[t]$ with Newton polygon Γ. In section 2.2, we defined its resultant $r_{\mathfrak{f}}$ as the determinant of the i-th graded part $(\mathcal{K}^\bullet)_i$ for $i \geq 3$ of a certain Koszul complex $\mathcal{K}^\bullet$ associated to $\mathfrak{f}$. For $i = 3$, this complex (of $\mathbb{Z}_q[t]$-modules) $(\mathcal{K}^\bullet)_3$ is given by

$$0 \longrightarrow \mathbb{Z}_q[t](e_t \wedge e_x \wedge e_y) \xrightarrow{\partial_3} \mathbb{Z}_q[t][\Gamma](e_x \wedge e_y) \oplus \mathbb{Z}_q[t][\Gamma](e_t \wedge e_y) \oplus \mathbb{Z}_q[t][\Gamma](e_t \wedge e_x)$$

$$\xrightarrow{\partial_2} \mathbb{Z}_q[t][2\Gamma]e_t \oplus \mathbb{Z}_q[t][2\Gamma]e_x \oplus \mathbb{Z}_q[t][2\Gamma]e_y \xrightarrow{\partial_1} \mathbb{Z}_q[t][3\Gamma] \longrightarrow 0$$

where the maps ∂_i are defined as in Definition 2.1.4. Since $\mathfrak{f}$ is generically nondegenerate, $\mathcal{K}^\bullet$ is exact as a complex of $\mathbb{Q}_q(t)$-vector spaces. Now we compute the determinant of this complex like in Definition 2.2.2. Let $[u_1, \ldots, u_k]$, $[v_1, \ldots, v_l]$ and $[w_1, \ldots, w_m]$ denote the monomial bases of the $\mathbb{Z}_q[t]$-modules $\mathbb{Z}_q[t][\Gamma]$, $\mathbb{Z}_q[t][2\Gamma]$ and $\mathbb{Z}_q[t][3\Gamma]$ respectively. We denote $F = t\mathfrak{f}$.

First we consider the map

$$\partial_1 : \mathbb{Z}_q[t][2\Gamma]e_t \oplus \mathbb{Z}_q[t][2\Gamma]e_x \oplus \mathbb{Z}_q[t][2\Gamma]e_y \to \mathbb{Z}_q[t][3\Gamma],$$

where

$$\partial_1 e_t = F, \;\; \partial_1 e_x = x\frac{\partial F}{\partial x}, \;\; \partial_1 e_y = y\frac{\partial F}{\partial y}.$$

The sequence

$$\mathbf{b}_1 = [v_1 e_t, \ldots, v_l e_t, v_1 e_x, \ldots, v_l e_x, v_1 e_y, \ldots, v_l e_y]$$

is a basis for the $\mathbb{Z}_q[t]$-module $\mathbb{Z}_q[t][2\Gamma]e_t \oplus \mathbb{Z}_q[t][2\Gamma]e_x \oplus \mathbb{Z}_q[t][2\Gamma]e_y$ and

$$\mathbf{b}_0 = [w_1, \ldots, w_m]$$

is a basis for $\mathbb{Z}_q[t][3\Gamma]$. Now the sequence $\mathbf{b_1}$ can be partitioned into two sequences $\mathbf{b}'_1$ and $\mathbf{b}''_1$ such that the images of the elements of $\mathbf{b}'_1$ under ∂_1 are a basis of $\mathbb{Q}_q(t)[3\Gamma]$. Note that such a partition can be easily found by repeatedly adding elements to $\mathbf{b}'_1$ such that their images under ∂_1 remain linearly independent. Let M_1 be the matrix of the restriction of ∂_1 to the span of $\mathbf{b}'_1$ with respect to the bases $\mathbf{b}'_1$ and $\mathbf{b}_0$.

Now we consider the next map

$$\partial_2 : \mathbb{Z}_q[t][\Gamma](e_x \wedge e_y) \oplus \mathbb{Z}_q[t][\Gamma](e_t \wedge e_y) \oplus \mathbb{Z}_q[t][\Gamma](e_t \wedge e_x) \to \mathbb{Z}_q[t][2\Gamma]e_t \oplus \mathbb{Z}_q[t][2\Gamma]e_x \oplus \mathbb{Z}_q[t][$$

where

$$\partial_2(e_t \wedge e_x) = Fe_x - x\frac{\partial F}{\partial x}e_t, \;\; \partial_2(e_t \wedge e_y) = Fe_y - y\frac{\partial F}{\partial y}e_t, \;\; \partial_2(e_x \wedge e_y) = x\frac{\partial F}{\partial x}e_y - y\frac{\partial F}{\partial y}e_x.$$

The sequence

$$\mathbf{b}_2 = [u_1(e_x \wedge e_y), \ldots, u_k(e_x \wedge e_y), u_1(e_t \wedge e_y), \ldots, u_k(e_t \wedge e_y), u_1(e_t \wedge e_x), \ldots, u_k(e_t \wedge e_x)]$$

is a basis for $\mathbb{Z}_q[t][\Gamma](e_x \wedge e_y) \oplus \mathbb{Z}_q[t][\Gamma](e_t \wedge e_y) \oplus \mathbb{Z}_q[t][\Gamma](e_t \wedge e_x)$ and can again be partitioned into two sequences $\mathbf{b}'_2$ and $\mathbf{b}''_2$, this time such that the images of the elements of $\mathbf{b}'_2$ under ∂_2 projected onto the $\mathbb{Q}_q(t)$-vector space spanned by $\mathbf{b}''_1$ form a basis for that space. Let M_2 be the matrix of the restriction of ∂_2 to the span of $\mathbf{b}'_2$, composed with the projection onto the span of $\mathbf{b}''_1$, with respect to the bases $\mathbf{b}'_2$ and $\mathbf{b}''_1$.

Finally, we consider the last map

$$\partial_3 : \mathbb{Z}_q[t](e_t \wedge e_x \wedge e_y) \to \mathbb{Z}_q[t][\Gamma](e_x \wedge e_y) \oplus \mathbb{Z}_q[t][\Gamma](e_t \wedge e_y) \oplus \mathbb{Z}_q[t][\Gamma](e_t \wedge e_x).$$

The element $\mathbf{b}_3 = [e_t \wedge e_x \wedge e_y]$ generates $\mathbb{Z}_q[t](e_t \wedge e_x \wedge e_y)$. Its span as a $\mathbb{Q}_q(t)$-vector space is mapped isomorphically to the span of $\mathbf{b}''_2$ by ∂_3 composed with the projection. Let M_3 be the (1×1) matrix of this isomorphism with respect to the bases $\mathbf{b}_3$ and $\mathbf{b}''_2$.

Now by Definition 2.2.2, the resultant $r_\mathfrak{f}$ can be computed as

$$r_\mathfrak{f} = \det(M_3)\det(M_2)^{-1}\det(M_1).$$

Remark. Note that we have explicitly chosen bases for all the modules in this section. Other choices will give the same $r_\mathfrak{f}$ upto an element of $\mathbb{Z}_q^\times$. However, to be able to control the complexity of the computations, one does have to make a choice, and our choice seems the most natural one.

4.2 Linear algebra over $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$

In the following sections we will be doing a lot of computations with (free) $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$-modules. For example we need to find kernels and cokernels, and solve linear systems of equations over $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$. Since $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ is a principal ideal domain, we can reduce most of these problems to computing the Smith normal form of some matrix. Instead of computing these Smith forms over $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$, we do so over $\mathbb{Q}_q[t]$, since this ring is Euclidean. Over a Euclidean ring there are algorithms to compute Smith normal forms which are well known and included in a lot of computer algebra systems. For completeness and for lack of a good reference, we now briefly explain this in more detail.

Definition 4.2.1. Let R denote a principal ideal domain and suppose that A is a matrix over R. The *Smith normal form of* A is a matrix

$$S = \begin{pmatrix} s_1 & 0 & \cdots & 0 & 0 & \dots & 0 \\ 0 & s_2 & \cdots & 0 & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_k & 0 & \dots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

over R with the following properties:

- the only nonzero entries of S are on the diagonal and for all i we have that $s_i \neq 0$ and s_i divides s_{i+1},
- there exist invertible matrices P, Q of the right dimensions over R such that $S = PAQ$.

Such a matrix S always exists. The s_i are unique upto elements of $R^\times$ and are called the *invariant factors* of A.

Now suppose that A is an $m \times n$ matrix with entries in $\mathbb{Q}_q[t]$. Let S be its Smith form, $s_1, \dots, s_k$ the invariant factors, and P, Q invertible matrices such that $S = PAQ$. We will identify the matrix A with the homorphism of modules $\phi : \mathbb{Q}_q[t]^n \to \mathbb{Q}_q[t]^m$ it defines with respect to the standard bases.

4.2.1 Computing a kernel

Note that $\ker(A) = Q\ker(S)$, and that a basis for $\ker(S)$ is given by the last $n-k$ standard basis vectors of $\mathbb{Q}_q[t]^n$. Therefore, a basis for both the $\mathbb{Q}_q[t]$-module $\ker(A)$ and the $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$-module $\ker(A) \otimes_{\mathbb{Q}_q[t]} \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ is given by the last $n-k$ columns of Q.

4.2.2 Computing a cokernel

In general $\operatorname{coker}(A)$ will not be a free $\mathbb{Q}_q[t]$ module. However, we will assume that $\operatorname{coker}(A) \otimes_{\mathbb{Q}_q[t]} \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ is a free $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$-module and we want to find a basis for this module.

Since P, Q are invertible, there is an isomorphism

$$\operatorname{coker}(A) \cong \operatorname{coker}(S)$$

given by multiplication (from the left) by P. By the structure of the matrix S we have

$$\operatorname{coker}(S) = \oplus_{i=1}^{k} (\mathbb{Q}_q[t]/s_i\mathbb{Q}_q[t]) \oplus \mathbb{Q}_q[t]^{(m-k)}.$$

Since $\operatorname{coker}(S) \otimes_{\mathbb{Q}_q[t]} \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ is free, we deduce that $s_i \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]^\times$ for all i. Hence

$$\operatorname{coker}(S) \otimes_{\mathbb{Q}_q[t]} \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}] \cong \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]^{(m-k)},$$

and a basis for $\operatorname{coker}(S) \otimes_{\mathbb{Q}_q[t]} \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ is given by the classes of the last $m-k$ standard basis vectors of $\mathbb{Q}_q[t]^m$. Therefore, a basis for $\operatorname{coker}(A) \otimes_{\mathbb{Q}_q[t]} \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ is given by the classes of the last $m-k$ columns of P^{-1}.

4.2.3 Solving a sytem of linear equations

Now consider the system of linear equations

$$Av = b, \tag{4.1}$$

with $b \in \mathbb{Q}_q[t]^m$. Suppose that there exists a solution $v \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]^n$. Since $S = PAQ$, the above system is equivalent to

$$SQ^{-1}v = Pb.$$

We write $u = Pb$ and let u_i denote the i-th component of u. The existence of a solution $v \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]^n$ to (4.1) implies that

- u_i is divisible by s_i in $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ for $i \leq k$,
- $u_i = 0$ for $i > k$.

If we define $w \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]^n$ by

$$w_i = \begin{cases} \frac{u_i}{s_i} & \text{for } i \leq k \\ 0 & \text{for } i > k \end{cases}$$

then $v = Qw$ is a solution to the linear system (4.1). The general solution can then be obtained by adding an element from the kernel of A which we already know how to compute.

Remark. It is a lot easier to find a solution $v \in \mathbb{Q}_q(t)^n$. Computing a Smith normal form over $\mathbb{Q}_q[t]$ requires quite a heavy computation while we can find a solution $v \in \mathbb{Q}_q(t)^n$ by using Gaussian elimination which is usually much more efficient. So for example when $v \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]^n$ is the only solution that exists in $\mathbb{Q}_q(t)^n$, it is better to compute it over $\mathbb{Q}_q(t)$.

Remark. Note that all the results in this section still hold when we replace $\mathbb{Q}_q$ by any field k, and $r_\mathfrak{f}$ by any polynomial $r \in k[t]$.

4.3 The cohomology of the affine family

In section 3.6, we saw that $\mathfrak{f}$ defines a family $\mathbf{X}/\mathbf{S}$ of affine nondegenerate curves, with

$$\mathbf{S} = \text{Spec}(\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]),$$

$$\mathbf{X} = \text{Spec}(\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][\mathbb{Z}^2]/\mathfrak{f}),$$

and that the $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$-module

$$M = H^0(\mathbf{S}, H^1_{DR}(\mathbf{X}/\mathbf{S})) \cong H^0(\mathbf{X}, \Omega^1_{\mathbf{X}/\mathbf{S}})/dH^0(\mathbf{X}, \mathcal{O}_\mathbf{X})$$

is free of finite rank r. We want to compute a basis for this module. The following theorem will help us to do so. Recall from section 3.1 that

$$\delta = xy(\frac{\partial \mathfrak{f}}{\partial y}\frac{\partial}{\partial x} - \frac{\partial \mathfrak{f}}{\partial x}\frac{\partial}{\partial y}).$$

Theorem 4.3.1.

$$M \cong \frac{\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][2\Gamma]}{\mathfrak{f}\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][\Gamma] + \delta(\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][\Gamma])}.$$

Proof. This is a relative version, and a consequence, of Theorem 3.1.4. Let

$$\phi : \mathbb{Q}_q[t, \tfrac{1}{r_\mathfrak{f}}][2\Gamma] \longrightarrow M$$

denote the homomorphism of $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$-modules that sends a Laurent polynomial $g \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][2\Gamma]$ to the class of $g\omega_\mathfrak{f}$ in M. Let m be a maximal ideal of $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$, and let $k = \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]/m$ denote its residue field. Then $\mathfrak{f}$ is also nondegenerate over k. So by Theorem 3.4.5, and the first part of Theorem 3.1.4, the map $\phi \otimes k : k[2\Gamma] \to M \otimes k$ is surjective. Since this holds for every m, we have that ϕ is surjective by Lemma 2.3.2.

The kernel of ϕ contains both $\mathfrak{f}\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][\Gamma]$ and $\delta(\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][\Gamma])$. Let

$$\psi : \mathfrak{f}\mathbb{Q}_q[t, \tfrac{1}{r_\mathfrak{f}}][\Gamma] + \delta(\mathbb{Q}_q[t, \tfrac{1}{r_\mathfrak{f}}][\Gamma]) \longrightarrow \ker \phi$$

denote the inclusion. Note that $(\ker \phi) \otimes k \cong \ker(\phi \otimes k)$, since M is a free module. So $\psi \otimes k : \mathfrak{f}k[\Gamma] + \delta(k[\Gamma]) \to (\ker \phi) \otimes k$ is surjective, by the second part of Theorem 3.1.4. Since this holds for every m, we have that ψ is surjective by Lemma 2.3.2.

Hence ϕ induces the desired isomorphism. □

Now to find a basis for M, for every monomial $\mu \in \mathbb{Q}_q[t][\Gamma]$ we compute the elements $\delta(\mu), \mathfrak{f}\mu \in \mathbb{Q}_q[t][2\Gamma]$, and their coordinate vectors v_μ, w_μ with respect to the monomial basis $[u_1, \ldots, u_k]$ of $\mathbb{Q}_q[t][2\Gamma]$. Let A denote the matrix over $\mathbb{Q}_q[t]$ with these vectors as its columns. Then

$$M \cong \operatorname{coker}(A) \otimes_{\mathbb{Q}_q[t]} \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}].$$

We can compute a basis for $\operatorname{coker}(A) \otimes_{\mathbb{Q}_q[t]} \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ as in section 4.2. Applying the isomorphism from Theorem 4.3.1, we obtain a basis $[m_1, \ldots, m_r]$ for M.

4.4 The cohomology of the complete family

4.4.1 The residue map

Recall from section 3.6 that $\mathbf{Y}/\mathbf{S}$ denotes the relative normal crossing compactification of the family of affine nondegenerate curves $\mathbf{X}/\mathbf{S}$ defined by f, and that $\mathbf{D}/\mathbf{S}$ denotes the relative divisor of the complement of $\mathbf{X}$ in $\mathbf{Y}$.

We know from section 3.5 that there is an exact sequence

$$0 \longrightarrow N \longrightarrow M \xrightarrow{Res} H^0_{DR}(\mathbf{D}/\mathbf{S}),$$

where $M = H^0(\mathbf{S}, H^1_{DR}(\mathbf{X}/\mathbf{S}))$, and $N = H^0(\mathbf{S}, H^1_{DR}(\mathbf{Y}/\mathbf{S}))$. We are mainly interested in the module N, and we have already found a basis $[m_1, \ldots, m_r]$ for the module M, so we now need to explicitly compute the kernel of the residue map Res on M.

However, at the level of the De Rham complex, Res can only be computed for logarithmic differentials, and in general a cohomology class $m \in M$ can *not* be represented by a differential that simultaneously has logarithmic poles at *all* the points in the support of $\mathbf{D}$. On the other hand, by Theorem 3.5.2, on every affine open $\mathbf{U}/\mathbf{S} \subset \mathbf{Y}/\mathbf{S}$ the cohomology class m can be represented by a relative differential that has a logarithmic pole along $\mathbf{U} \cap \mathbf{D}$. Therefore, we will compute the kernel of the residue map *separately* for every orbit at infinity of the associated toric surface.

Recall from section 3.1 that $t_1, \ldots, t_n$ denote the edges of Γ, and they correspond to the 1-dimensional tori at infinity $\mathbf{T}_1, \ldots, \mathbf{T}_n$ of the toric surface $\mathbf{X}_\Gamma$ over $\mathbf{S}$. Let $\lambda_i = |t_i \cap \mathbb{Z}^2| - 1$ denote the *arithmetic length* of t_i.

We have

$$H^0_{DR}(\mathbf{D}/\mathbf{S}) \cong \oplus_{i=1}^n H^0_{DR}((\mathbf{T}_i \cap \mathbf{Y})/\mathbf{S}),$$

and there exists a decomposition

$$Res = \oplus_{i=1}^n Res_i \quad \text{with:} \quad Res_i : M \to H^0_{DR}((\mathbf{T}_i \cap \mathbf{Y})/\mathbf{S}).$$

As before, we let e_i denote the inward pointing normal vector of t_i which is minimally integral, i.e. of the form (a_i, b_i) with $\gcd(a_i, b_i) = 1$. Let the vertices of Γ be numbered counterclockwise, and let (p_i, q_i) denote the first vertex of t_i. We write $N_i = \langle (p_i, q_i), e_i \rangle$, so that the line that contains t_i is defined by $\langle -, e_i \rangle = N_i$.

If, like in section 3.1, we write

$$\mathbf{D}_\Gamma = -\sum_{i=1}^n N_i(\mathbf{T}_i \cap \mathbf{Y}), \qquad \mathbf{D} = \sum_{i=1}^n (\mathbf{T}_i \cap \mathbf{Y}),$$

now as relative divisors over $\mathbf{S}$, then again

$$Div(\omega_{\mathfrak{f}}) = \mathbf{D}_\Gamma - \mathbf{D} = \sum_{i=1}^{n} -(N_i+1)(\mathbf{T}_i \cap \mathbf{Y}),$$

where the first term is defined because of the normal crossing condition, and the equality can be checked over the function field $k(\mathbf{S}) = \mathbb{Q}_q(t)$ of $\mathbf{S}$.

Moreover, Theorem 3.1.3 still holds:

$$Div(x^i y^j) = \sum_{k=1}^{n} \langle (i,j), e_k \rangle (\mathbf{T}_k \cap \mathbf{Y}),$$

$$L(n\mathbf{D}_\Gamma) = \text{ image of } \mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}][n\Gamma] \text{ in } H^0(\mathbf{X}, \mathcal{O}_{\mathbf{X}}),$$

where $L(n\mathbf{D}_\Gamma)$ is the intersection of the corresponding Riemann Roch space over $k(\mathbf{S})$ with the functions that are regular on $\mathbf{X}$, and the equalities can again be checked over $k(\mathbf{S})$.

So a relative differential on $\mathbf{X}/\mathbf{S}$ is logarithmic along the divisor $\mathbf{D}$, if and only if it can be written as $h\omega_{\mathfrak{f}}$, with $h \in \mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}][\Gamma]$.

Remark. It can also be shown easily that a relative differential on $\mathbf{X}/\mathbf{S}$ is logarithmic along the divisor $\mathbf{T}_i \cap \mathbf{Y}$, if and only if it can be written as $h\omega_{\mathfrak{f}}$, with h supported on the half plane defined by $\langle -, e_i \rangle \geq N_i$.

4.4.2 Computing the kernel

From Theorem 4.3.1, we know that a cohomology class $m \in M$ can be represented by a relative differential $g\omega_{\mathfrak{f}}$, with $g \in \mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}][2\Gamma]$. Now fix an i, with $0 \leq i \leq n$. Since g is supported on 2Γ, in particular it is supported on the half plane defined by $\langle -, e_i \rangle \geq 2N_i$. We want to find a differential $g_i\omega_{\mathfrak{f}}$ that also represents m, but with g_i supported on the half plane defined by $\langle -, e_i \rangle \geq N_i$, so that $g_i\omega_{\mathfrak{f}}$ is logarithmic along $\mathbf{T}_i \cap \mathbf{Y}$.

If we write $s = x^{b_i} y^{-a_i}$, then any monomial supported on t_i can be written as $x^{p_i} y^{q_i} s^k$, with $0 \leq k \leq \lambda_i$. Hence we can define the polynomial $\tilde{\mathfrak{f}}_i = (\frac{\mathfrak{f}_{t_i}}{x^{p_i} y^{q_i}}) \in \mathbb{Q}_q[t][s]$. Now we compute

$$\begin{pmatrix} x\frac{\partial \mathfrak{f}_{t_i}}{\partial x} \\ y\frac{\partial \mathfrak{f}_{t_i}}{\partial y} \end{pmatrix} = x^{p_i} y^{q_i} \begin{pmatrix} p_i & b_i \\ q_i & -a_i \end{pmatrix} \begin{pmatrix} \tilde{\mathfrak{f}}_i \\ s\frac{d\tilde{\mathfrak{f}}_i}{ds} \end{pmatrix}. \tag{4.2}$$

By the nondegeneracy condition for $\mathfrak{f}_{t_i}$, the sequence $\mathfrak{f}, x\frac{\partial \mathfrak{f}_{t_i}}{\partial x}, y\frac{\partial \mathfrak{f}_{t_i}}{\partial y}$ generates the unit ideal in $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][\mathbb{Z}^2]$. So it follows from (4.2) that the sequence $\tilde{\mathfrak{f}}_i, s\frac{d\tilde{\mathfrak{f}}_i}{ds}$ generates the unit ideal in $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][s, s^{-1}]$. Hence there exist $\tilde{\alpha}_i, \tilde{\beta}_i \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][s, s^{-1}]$ such that

$$\tilde{\alpha}_i \tilde{f}_i + \tilde{\beta}_i s\frac{d\tilde{\mathfrak{f}}_i}{ds} = 1.$$

The one-variable Newton polygon of $\tilde{f}_i$ (as a polynomial of s) is equal to $[0, \lambda_i]$. Therefore, the Laurent polynomials $\tilde{\alpha}_i, \tilde{\beta}_i$ can be chosen to be supported on $[0, \lambda_i]$ as well. Over a field this is a consequence of Theorem 6.1.3, and it generalizes to general Noetherian rings in the same way as Theorem 2.3.1.

So $\tilde{\alpha}_i, \tilde{\beta}_i$ can be found by solving a finite system of linear equations over $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$ as in section 4.2. Actually, we can even solve this system over $\mathbb{Q}_q(t)$, since the residue that we are trying to compute is uniquely determined by its restriction to any open set of $\mathbf{S}$.

Let $L_i(\tau)$ denote the line defined by $\langle -, e_i\rangle = \tau$. If $x^j y^k$ is a monomial supported on $L_i(\tau)$, then all monomials supported on $L_i(\tau)$ are of the form $x^j y^k s^l$ with $l \in \mathbb{Z}$, the elements $\mathfrak{f}x^j y^k s^l$ and $\delta(x^j y^k s^l)$ are supported on the half plane defined by $\langle -, e_i\rangle \geq \tau + N_i$, and we compute:

$$(\mathfrak{f}x^j y^k s^l)_{L_i(\tau+N_i)} = \tilde{\mathfrak{f}}_i x^{j+p_i} y^{k+q_i} s^l,$$

$$(\delta(x^j y^k s^l))_{L_i(\tau+N_i)} = \left((-kp_i + jq_i + lN_i)\tilde{\mathfrak{f}}_i - \tau s\frac{d\tilde{\mathfrak{f}}_i}{ds}\right) x^{j+p_i} y^{k+q_i} s^l.$$

for the terms on the boundary of this half plane. So as long as $\tau < 0$, we can write

$$x^{j+p_i} y^{k+q_i} = \delta(-x^j y^k \frac{\tilde{\beta}_i}{\tau}) + \left(\frac{-kp_i + jq_i + N_i s\frac{d\tilde{\beta}_i}{ds}}{\tau} + \tilde{\alpha}_i\right) \mathfrak{f}x^j y^k + h_{j,k},$$

where $h_{j,k}$ is supported on the half plane defined by $\langle -, e_i\rangle \geq \tau + N_i + 1$. The resulting relations in M

$$x^{j+p_i} y^{k+q_i} \omega_\mathfrak{f} = h_{j,k}\omega_\mathfrak{f}$$

can now be used to reduce a relative differential $h\omega_\mathfrak{f}$, with h supported on the half plane defined by $\langle -, e_i\rangle \geq \tau + N_i$, to a differential $h'\omega_\mathfrak{f}$ that represents the same

class in M, but with h' supported on the half plane defined by $\langle -, e_i \rangle \geq \tau + N_i + 1$.

Applying this procedure repeatedly to $g\omega_{\mathfrak{f}}$, eventually we find a relative differential $g_i\omega_{\mathfrak{f}}$ that still represents m, but with g_i supported on the half plane defined by $\langle -, e_i \rangle \geq N_i$, so that $g_i\omega_{\mathfrak{f}}$ is logarithmic along $\mathbf{T}_i \cap \mathbf{Y}$.

We have

$$H^0_{DR}((\mathbf{T}_i \cap \mathbf{Y})/\mathbf{S}) \cong \frac{\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}][s, s^{-1}]}{(\tilde{\mathfrak{f}}_i)}.$$

Now consider the map of $\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}]$-modules

$$\chi_i : M \to \frac{\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}][s, s^{-1}]}{(\tilde{\mathfrak{f}}_i)},$$

defined by $\chi_i(m) = \frac{(g_i)_{L_i(N_i)}}{x^{p_i} y^{q_i}}$. Since $x^{p_i} y^{q_i} \omega_{\mathfrak{f}}$ has order exactly -1, and $g_i\omega_{\mathfrak{f}}$ is logarithmic at the divisor $\mathbf{T}_i \cap \mathbf{Y}$, we have that $Res_i(m) = 0$ if and only if $\chi_i(m) = 0$. So

$$\ker(Res_i) = \ker(\chi_i),$$

and

$$N = \ker(Res) = \ker(\oplus_{i=1}^n Res_i) = \ker(\oplus_{i=1}^n \chi_i).$$

For every i, let $[b_{i,1}, \ldots, b_{i,\lambda_i}]$ denote the basis $[s^0, \ldots, s^{\lambda_i - 1}]$ of the $\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}]$-module $\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}][s, s^{-1}]/(\tilde{\mathfrak{f}}_i)$, and let A be the matrix of $\oplus_{i=1}^n \chi_i$ with respect to the bases $[m_1, \ldots, m_r]$ and $[b_{1,1}, \ldots, b_{1,\lambda_1}, \ldots, b_{n,1}, \ldots, b_{n,\lambda_n}]$. This time the entries of A are not necessarily contained in $\mathbb{Q}_q[t]$, so we first multiply A by a power of $r_{\mathfrak{f}}$ such that its entries lie in $\mathbb{Q}_q[t]$ (note that this does not change the kernel). Now we can first find a basis $[n_1, \ldots, n_s]$ for N over $\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}]$ by computing a basis for the kernel of A, and then complete this basis to a basis for M over $\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}]$ by computing a basis for the cokernel of the inclusion of N in M, as in section 4.2.

4.5 The Gauss-Manin connection

Recall from section 3.4 that the relative cohomology module M carries a natural Gauss-Manin connection

$$\nabla : M \to M \otimes \Omega^1_{\mathbf{S}}.$$

In this section we want to compute the matrix $G \in M_{r,r}(\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}])$ such that for all j we have

$$\nabla m_j = \sum_{i=1}^{r} G_{ij} m_i \otimes dt.$$

4.5.1 Finding α, β, γ

The Laurent polynomial $\mathfrak{f}$ is nondegenerate over $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$, hence by Theorem 2.3.3 there exist $\alpha, \beta, \gamma \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}][2\Gamma]$ such that

$$\alpha\mathfrak{f} + \beta x \frac{\partial \mathfrak{f}}{\partial x} + \gamma y \frac{\partial \mathfrak{f}}{\partial y} = 1. \tag{4.3}$$

We now want to find such α, β, γ. Let

$$\phi : \mathbb{Q}_q[t][2\Gamma] \oplus \mathbb{Q}_q[t][2\Gamma] \oplus \mathbb{Q}_q[t][2\Gamma] \to \mathbb{Q}_q[t][3\Gamma]$$

be the homomorphism of $\mathbb{Q}_q[t]$ modules defined by

$$\phi(a, b, c) = a\mathfrak{f} + bx \frac{\partial \mathfrak{f}}{\partial x} + cy \frac{\partial \mathfrak{f}}{\partial y},$$

let $[\mu_1, \ldots, \mu_n]$ denote the monomial basis of $\mathbb{Q}_q[t][2\Gamma]$, $[\nu_1, \ldots, \nu_m]$ the monomial basis of $\mathbb{Q}_q[t][3\Gamma]$, and let A be the matrix of ϕ with respect to the bases $[\mu_1, \ldots, \mu_n, \mu_1, \ldots, \mu_n, \mu_1, \ldots, \mu_n]$ and $[\nu_1, \ldots, \nu_m]$. By Theorem 2.3.3, the homomorphism ϕ is surjective when tensored with $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]$, so that the linear system of equations

$$Av = b$$

has a solution $v \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]^n$ for every $b \in \mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]^m$. Now if we take b to be the element of $\mathbb{Q}_q[t, \frac{1}{r_\mathfrak{f}}]^m$ that corresponds to the Laurent polynomial $1 \in \mathbb{Q}_q[t][3\Gamma]$, then solving the above linear system as in section 4.2 will give a solution of equation (4.3).

4.5.2 Computing the connection

Since

$$\alpha\mathfrak{f} + \beta x \frac{\partial \mathfrak{f}}{\partial x} + \gamma y \frac{\partial \mathfrak{f}}{\partial y} = 1,$$

we find that

$$\beta x\frac{\partial \mathfrak{f}}{\partial x}+\gamma y\frac{\partial \mathfrak{f}}{\partial y}=1$$

as elements of $H^0(\mathbf{X},\mathcal{O}_{\mathbf{X}})$. Moreover

$$\frac{\partial \mathfrak{f}}{\partial x}dx+\frac{\partial \mathfrak{f}}{\partial y}dy+\frac{\partial \mathfrak{f}}{\partial t}dt=0 \tag{4.4}$$

on $\mathbf{X}$, so that

$$\frac{\partial \mathfrak{f}}{\partial x}dx+\frac{\partial \mathfrak{f}}{\partial y}dy=0$$

as elements of $\Omega^1_{\mathbf{X}/\mathbf{S}}$. Using all of this, we can write

$$\omega_{\mathfrak{f}}=\frac{dx}{xy\frac{\partial \mathfrak{f}}{\partial y}}=\frac{dx}{xy\frac{\partial \mathfrak{f}}{\partial y}}(\beta x\frac{\partial \mathfrak{f}}{\partial x}+\gamma y\frac{\partial \mathfrak{f}}{\partial y})=-\beta\frac{dy}{y}+\gamma\frac{dx}{x},$$

as elements of $\Omega^1_{\mathbf{X}/\mathbf{S}}$. So

$$\nabla m_j=\nabla(g_j\omega_{\mathfrak{f}})=\nabla(-\frac{\beta g_j}{y}dy)+\nabla(\frac{\gamma g_j}{x}dx).$$

Recall the instructions from section 3.4 for computing ∇: apply the absolute differential d, represent the result by an element of $\Omega^1_{\mathbf{S}}\otimes\Omega^1_{\mathbf{X}/\mathbf{S}}$, and finally project back onto $M\otimes\Omega^1_{\mathbf{S}}$. For the first term for example:

$$d(-\frac{\beta g_j}{y}dy)=-\frac{\partial}{\partial x}(\frac{\beta g_j}{y})\,dx\wedge dy+\frac{\partial}{\partial t}(\frac{\beta g_j}{y})\,dy\wedge dt.$$

By equation 4.4, we have

$$dx=\frac{-\frac{\partial \mathfrak{f}}{\partial y}dy-\frac{\partial \mathfrak{f}}{\partial t}dt}{\frac{\partial \mathfrak{f}}{\partial x}}$$

as elements of $\Omega^1_{\mathbf{X}}$. Substituting this, we get

$$d(-\frac{\beta g_j}{y}dy)=\qquad\frac{\partial}{\partial x}(\frac{\beta g_j}{y})\,\frac{\frac{\partial \mathfrak{f}}{\partial t}}{\frac{\partial \mathfrak{f}}{\partial x}}dt\wedge dy-\frac{\partial}{\partial t}(\frac{\beta g_j}{y})\,dt\wedge dy,$$

so that

$$\begin{aligned}\nabla(-\frac{\beta g_j}{y}dy)=&\qquad\left(\frac{\partial}{\partial x}(\frac{\beta g_j}{y})\frac{\partial \mathfrak{f}}{\partial t}\,\frac{dy}{\frac{\partial \mathfrak{f}}{\partial x}}-\frac{\partial}{\partial t}(\frac{\beta g_j}{y})\,dy\right)\otimes dt\\ =&\qquad\left(-\frac{\partial}{\partial x}(\beta g_j)\frac{\partial \mathfrak{f}}{\partial t}\,\frac{dx}{y\frac{\partial \mathfrak{f}}{\partial y}}+\frac{\partial}{\partial t}(\beta g_j)\frac{\partial \mathfrak{f}}{\partial x}\frac{dx}{y\frac{\partial \mathfrak{f}}{\partial y}}\right)\otimes dt\\ =&\qquad\left(-x\frac{\partial}{\partial x}(\beta g_j)\frac{\partial \mathfrak{f}}{\partial t}+\frac{\partial}{\partial t}(\beta g_j)x\frac{\partial \mathfrak{f}}{\partial x}\right)\omega_{\mathfrak{f}}\otimes dt.\end{aligned}$$

By a similar computation

$$\nabla(\frac{\gamma g_j}{x}dx) = \left(-y\frac{\partial}{\partial y}(\gamma g_j)\frac{\partial \mathfrak{f}}{\partial t} + \frac{\partial}{\partial t}(\gamma g_j)y\frac{\partial \mathfrak{f}}{\partial y}\right)\omega_{\mathfrak{f}} \otimes dt.$$

So we obtain

$$\begin{aligned}\nabla m_j = & \quad \nabla(-\frac{\beta g_j}{y}dy) + \nabla(\frac{\gamma g_j}{x}dx) \\ = & \left(-x\frac{\partial}{\partial x}(\beta g_j)\frac{\partial \mathfrak{f}}{\partial t} + \frac{\partial}{\partial t}(\beta g_j)x\frac{\partial \mathfrak{f}}{\partial x} - y\frac{\partial}{\partial y}(\gamma g_j)\frac{\partial \mathfrak{f}}{\partial t} + \frac{\partial}{\partial t}(\gamma g_j)y\frac{\partial \mathfrak{f}}{\partial y}\right)\omega_{\mathfrak{f}} \otimes dt \\ = & \quad h_j\omega_{\mathfrak{f}} \otimes dt,\end{aligned}$$

where h_j is defined by the last equation. Now we have to find the coordinates of this element with respect to the basis $[m_1, \ldots, m_r]$, i.e. compute the matrix $G \in M_{r,r}(\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}])$ such that

$$\nabla m_j = \sum_{i=1}^{r} G_{ij} m_i \otimes dt.$$

Since $\beta, \gamma, g_j \in \mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}][2\Gamma]$ and $\mathfrak{f} \in \mathbb{Q}_q[t][\Gamma]$, the Laurent polynomial h_j is supported on 5Γ. So

$$h_j = \sum_{i=1}^{r} G_{ij} g_i + \delta(k_1) + \mathfrak{f}k_2, \tag{4.5}$$

for some $k_1, k_2 \in \mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}][4\Gamma]$. For every j, this is a linear system of equations over $\mathbb{Q}_q[t]$ for the G_{ij} and the coefficients of k_1, k_2, that can be solved as in section 4.2.

Remark. The matrix $G \in M_{r,r}(\mathbb{Q}_q[t, \frac{1}{r_{\mathfrak{f}}}])$ is uniquely determined by its restriction to any nonempty open set of $\mathbf{S}$. Therefore, the linear systems (4.3) and (4.5) can be solved using linear algebra over $\mathbb{Q}_q(t)$, which is usually more efficient. This is especially important in the case of (4.5), because the rank of $\mathbb{Q}_q[t][5\Gamma]$ is relatively large. Often, it is hard or impossible to solve (4.5) by computing a Smith normal form as in section 4.2.

4.6 The Castryck-Denef-Vercauteren algorithm

In this section, we briefly explain the point counting algorithm for nondegenerate curves of Castryck, Denef and Vercauteren [12]. We do this for two reasons. First, we will use this algorithm to approximate the matrix $A_{p,0}$ of the action F_p^* on the rigid cohomomology space $H^1_{rig}(C_0)$ of the fiber C_0 of our family. Second, we can use the Frobenius lifting techniques and precision loss bounds that are used in this algorithm to give an alternative to Theorem 3.8.8 which applies more generally.

4.6.1 The case of a single curve

As always, let $\mathbb{F}_q$ be the finite field with $q = p^n$ elements. Let $f_0 \in \mathbb{F}_q[\mathbb{Z}^2]$ be a nondegenerate Laurent polynomial with Newton polygon Γ, and $C_0 \subset \mathbb{T}^2_{\mathbb{F}_q}$ the affine curve defined by f_0. Take $\overline{C}_0$ to be the nonsingular projective model of C_0. Choose a lift $\mathfrak{f}_0 \in \mathbb{Z}_q[\mathbb{Z}^2]$ of f_0 with the same Newton polygon Γ. In the notation of section 3.2.2, we have

$$H^0(]\overline{C}_0[, j^\dagger \mathcal{O}_{]\overline{C}_0[}) = A^\dagger = \mathbb{Q}_q\langle \mathbb{Z}^2\rangle^\dagger/(\mathfrak{f}_0).$$

Denote by $\mathbb{Z}_q\langle \mathbb{Z}^2\rangle^\dagger$ the subring of $\mathbb{Q}_q\langle \mathbb{Z}^2\rangle^\dagger$ consisting of the elements with integral coefficients. It is a subring of the p-adic completion of $\mathbb{Z}[\mathbb{Z}^2]$ which we denote by $\mathbb{Z}_q\langle \mathbb{Z}^2\rangle$. Let $\sigma \in Gal(\mathbb{Q}_q/\mathbb{Q}_p)$ be the unique lift of the p-th power map on $\mathbb{F}_q$. We know from Theorem 2.3.1 that there exist Laurent polynomials $\alpha_0, \beta_0, \gamma_0 \in \mathbb{F}_q[2\Gamma]$, such that

$$\alpha_0 f + \beta_0 x\frac{\partial f}{\partial x} + \gamma_0 y\frac{\partial f}{\partial y} = 1.$$

Choose Laurent polynomials $\delta_0, \delta_1, \delta_2 \in \mathbb{Z}_q[2p\Gamma]$ that lift $\alpha_0^p, \beta_0^p, \gamma_0^p$. If we put

$$G(T) = \mathfrak{f}_0^\sigma(x^p(1+\delta_1 T), y^p(1+\delta_2 T)) + \delta_0 \mathfrak{f}_0^p T - \mathfrak{f}_0^p,$$

then $G \in \mathbb{Z}_q[\mathbb{Z}^2][T]$ and

$$G(0) = 0 \bmod p, \qquad \frac{dG}{dT}(0) = 1 \bmod p.$$

So by *Hensel's Lemma*, there exists a unique $Z \in \mathbb{Z}_q\langle \mathbb{Z}^2\rangle$ with $Z = 0 \bmod p$, such that $G(Z) = 0$. Actually, we will now see that this Z is an element of $\mathbb{Z}_q\langle \mathbb{Z}^2\rangle^\dagger$.

Theorem 4.6.1. *Let $Z \in \mathbb{Z}_q\langle \mathbb{Z}^2\rangle$ be defined as above. The terms of Z of p-adic valuation i are supported on $6pi\Gamma$. In particular $Z \in \mathbb{Z}_q\langle \mathbb{Z}^2\rangle^\dagger$.*

Proof. A slightly weaker result first appeared in [12, section 4]. We give an (unpublished) proof of Kedlaya. Given $0 < \epsilon \leq 1$ rational, let $R_{\epsilon,c}$ be the subring of $\mathbb{Z}_q[p^\epsilon]\langle\mathbb{Z}^2\rangle$ of series whose terms of p-adic valuation at most i are supported on $ci\Gamma$. Let $I_{\epsilon,c}$ be the ideal of $R_{\epsilon,c}$ consisting of the series whose terms of valuation at most i are supported on the interior of $ci\Gamma$. Put

$$H(T) = p^{-\epsilon}G(p^\epsilon T).$$

We wish to choose c, ϵ so that

$$H(T) \in R_{\epsilon,c}[T], \qquad H(0) = 0 \bmod I_{\epsilon,c}, \qquad \frac{dH}{dT}(0) = 1 \bmod I_{\epsilon,c}.$$

Note that for $i \geq 2$ the coefficient of T^i in $H(T)$ is supported on $(2i+1)p\Gamma$ and divisible by $p^{(i-1)\epsilon}$. Hence the condition $H(T) \in R_{\epsilon,c}[T]$ imposes the constraint

$$c \geq \frac{(2i+1)p}{(i-1)\epsilon}, \tag{4.6}$$

which is most restrictive for $i = 2$.

Since $H(0) = p^{-\epsilon}G(0)$ is divisible by $p^{1-\epsilon}$ and supported on $p\Gamma$, the condition $H(0) = 0 \bmod I_{\epsilon,c}$ imposes the constraint

$$c > \frac{p}{1-\epsilon}. \tag{4.7}$$

Finally, since $\frac{dH}{dT}(0) - 1 = \frac{dG}{dT}(0) - 1$ is supported on $3p\Gamma$ and divisible by p, the condition $\frac{dH}{dt}(0) = 1 \bmod I_{\epsilon,c}$ imposes the constraint

$$c > 3p.$$

We choose $\epsilon = \frac{5}{6}$ to balance (4.6) (for $i = 2$) and (4.7). For this choice we see that any $c > 6p$ satisfies all constraints. Since $R_{\epsilon,c}$ is complete with respect to $I_{\epsilon,c}$, we can apply Hensel's Lemma to deduce that $p^{-\epsilon}Z \in R_{\epsilon,c}$. Taking the limit as $c \downarrow 6p$ now implies the result. □

Remark. Z can be computed by a Newton iteration. Define two sequences $\{Y_n\}_{n=0}^\infty, \{Z_n\}_{n=0}^\infty$ over $\mathbb{Z}_q[\mathbb{Z}^2]$ as follows. Set

$$Y_0 = 1, \qquad Z_0 = 0.$$

Given Y_n, Z_n, compute $Y_{n+1}, Z_{n+1} \in \mathbb{Z}_q[\mathbb{Z}^2]$ such that

$$Z_{n+1} = Z_n - G(Z_n)Y_n \qquad \bmod p^{2^{n+1}},$$

$$Y_{n+1} = Y_n(2 - Y_n\frac{dG}{dT}(Z_{n+1})) \qquad \bmod p^{2^{n+1}}.$$

Then one verifies that

$$G(Z_n) = 0 \bmod p^{2^n}.$$

So that $Z_n \to Z$ in the p-adic topology as $n \to \infty$.

Since $G(Z) = 0$, and by Theorem 4.6.1, the map

$$x \to x^p(1 + \delta_1 Z), \qquad y \to y^p(1 + \delta_2 Z),$$

defines a lift F_p on $A^\dagger$ of the p-th power Frobenius map on $A = \mathbb{F}_q[\mathbb{Z}^2]/(\mathfrak{f}_0)$. The idea of [12] is to use this lift directly to compute the p-th power Frobenius action F_p^* on $H^1_{rig}(C_0)$.

Remark. Since we want to compute F_p, we are only interested in the image of Z in $A^\dagger$. So when computing Y_{n+1}, Z_{n+1} we can also work mod $\mathfrak{f}_0$. In [12] it is shown that, after applying a suitable coordinate transformation, one may assume that Γ has *unique* top and bottom vertices (p_t, q_t) and (p_b, q_b). In that case $[y^{q_b}, \ldots, y^{q_t - 1}]$ is a basis for $\mathcal{A} = \mathbb{Z}_q[\mathbb{Z}^2]/(\mathfrak{f}_0)$ as a $\mathbb{Q}_q[x, x^{-1}]$-module. Reducing all intermediate results to this basis will result in much more efficient computations. For example, using this reduction the number of coefficients of Z modulo some prime power p^k will only increase linearly with k instead of quadratically.

Let $\mathbf{C}_0 \subset \mathbb{T}^2_{\mathbb{Q}_q}$ be the affine curve defined by $\mathfrak{f}_0$ and $[m_1, \ldots, m_r]$ a basis for $H^1_{DR}(\mathbf{C}_0)$. If, as usual, we denote

$$\omega_{\mathfrak{f}_0} = \frac{dx}{xy\frac{\partial \mathfrak{f}_0}{\partial y}},$$

then by Theorem 3.1.4 we may assume that $m_i = g_i \omega_{\mathfrak{f}_0}$, with $g_i \in \mathbb{Q}_q[2\Gamma]$. Let $\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0 \in \mathbb{Q}_q[2\Gamma]$ be Laurent polynomials such that

$$\boldsymbol{\alpha}_0 \mathfrak{f}_0 + \boldsymbol{\beta}_0 x \frac{\partial \mathfrak{f}_0}{\partial x} + \boldsymbol{\gamma}_0 y \frac{\partial \mathfrak{f}_0}{\partial y} = 1.$$

Like in section 4.5, we have that $\omega_{\mathfrak{f}_0} = -\boldsymbol{\beta}_0 \frac{dy}{y} + \boldsymbol{\gamma}_0 \frac{dx}{x}$, so

$$
\begin{aligned}
&F_p^*(\omega_{\mathfrak{f}_0}) \\
&= -F_p(\boldsymbol{\beta}_0)\left(\frac{\partial F_p(y)}{F_p(y)\partial x}dx + \frac{\partial F_p(y)}{F_p(y)\partial y}dy\right) + F_p(\boldsymbol{\gamma}_0)\left(\frac{\partial F_p(x)}{F_p(x)\partial x}dx + \frac{\partial F_p(x)}{F_p(x)\partial y}dy\right) \\
&= \left(y\frac{\partial \mathfrak{f}_0}{\partial y}\left(F_p(\boldsymbol{\gamma}_0)\frac{x\partial F_p(x)}{F_p(x)\partial x} - F_p(\boldsymbol{\beta}_0)\frac{x\partial F_p(y)}{F_p(y)\partial x}\right)\right. \\
&\quad \left. - x\frac{\partial \mathfrak{f}_0}{\partial x}\left(F_p(\boldsymbol{\gamma}_0)\frac{y\partial F_p(x)}{F_p(x)\partial y} - F_p(\boldsymbol{\beta}_0)\frac{y\partial F_p(y)}{F_p(y)\partial y}\right)\right)\omega_{\mathfrak{f}_0} \\
&= E\omega_{\mathfrak{f}_0}, \qquad (4.8)
\end{aligned}
$$

where E is defined by the last equation. Now

$$F_p^*(m_i) = F_p(g_i)F_p^*(\omega_{\mathfrak{f}_0}) = F_p(g_i)E\omega_{\mathfrak{f}_0}.$$

To compute the matrix $A_{p,0}$ of F_p^* on $H^1_{rig}(C_0)$, to some finite p-adic precision ν, we can proceed roughly as follows:

- calculate F_p by computing Z using the above Newton iteration,
- for every $(i,j) \in 2\Gamma$ compute $F_p(x^iy^j)$,
- calculate E and $F_p(x^iy^j)E$ for all $(i,j) \in 2\Gamma$.

All of these computations have to be done with finite but high enough p-adic precision ν'. From Theorem 4.6.1 it follows (as in Theorem 4.6.6) that the terms of p-adic valuation at most ν' of $F_p(x^iy^j)E$ with $(i,j) \in 2\Gamma$ are supported on $(8p\nu' + 5p)\Gamma$.

- Express the cohomology classes of the differentials $F_p(x^iy^j)E\omega_{\mathfrak{f}_0}$ with $(i,j) \in 2\Gamma$ again as linear combinations of the $x^iy^j\omega_{\mathfrak{f}_0}$ with $(i,j) \in 2\Gamma$, and deduce the matrix $A_{p,0}$.

In this step there will be some loss of precision, because coordinates of the $F_p(x^iy^j)E\omega_{\mathfrak{f}_0}$ with respect to the $x^iy^j\omega_{\mathfrak{f}_0}$ will in general not be integral. The following result can be used to give a bound for this loss of precision.

Theorem 4.6.2. *Let $\omega = h\omega_{\mathfrak{f}_0}$ be a differential, with $h \in \mathbb{Z}_q[m\Gamma]$ for some positive integer m, and let N be defined as in Corollary 3.7.6. If d is chosen such that*

$$p^{d+1} \geq -(m-1)N + 1,$$

then the class of $p^d\omega$ in $H^1_{DR}(\mathbf{C}_0)$ can be represented by a differential $h'\omega_{\mathfrak{f}_0}$ with $h' \in \mathbb{Z}_q[2\Gamma]$.

Proof. We use the definitions and the notation from section 3.7. From Theorem 3.1.2 and Theorem 3.1.3, we know that the pole order of ω at a point at infinity is at most $-(m-1)N+1$. So by Theorem 3.7.4, $p^d\omega$ is contained in the lattice Λ_{int}, and by Theorem 3.7.7 this implies that $p^d\omega \in \Lambda_{2\Gamma}$. □

Remark. A similar (but slightly weaker) bound for this loss of precision, with a much more complicated proof, can be found in [12, Corollary 5.6].

Corollary 4.6.3. *For the matrix $A_{p,0}$ of F_p^* to be correct to precision ν, it is sufficient to take ν' such that*

$$\nu' - \lceil \log_p(-(8p\nu' + 5p - 1)N + 1)\rceil + 1 \geq \nu.$$

Remark. This is a bit of a simplification. First, if we reduce $F_p(g_i)E$ to the basis $y^{q_b}, \ldots, y^{q_t-1}$, it will not be supported on $(8p\nu' + 5p)\Gamma$ anymore. So the loss of precision in the last step will be somewhat bigger. Second, this bound is only sufficient when $[m_1, \ldots, m_r]$ is a basis for the $\mathbb{Z}_q$-lattice $\Lambda_{2\Gamma}$. For more general bases there will again be some additional loss of precision. We will not go into these details, but instead refer to [12].

4.6.2 The case of a family

Now let us consider a family of nondegenerate curves like the one in section 3.6. In this case, the matrix A_p of the p-th power Frobenius map, on the relative rigid cohomology of the family, can in principle still be computed by this direct method. This is not of direct practical use for computations, but it can be used to derive some bounds.

So, like in section 3.6, take $f \in \mathbb{F}_q[t][\mathbb{Z}^2]$, and let $\mathfrak{f} \in \mathbb{Z}_q[t][\mathbb{Z}^2]$ be a lift of f, with the same Newton polygon, and the same degree in t. We now want to 'compute' a lift F_p of the p-th power Frobenius map on the ring $A^\dagger$ of overconvergent functions on the total space of the (affine) family defined by $\mathfrak{f}$.

$$A^\dagger = \mathbb{Q}_q\langle x, y, t, \frac{1}{r_\mathfrak{f}}\rangle^\dagger/(\mathfrak{f}) = \mathbb{Q}_q\langle x, y, t, z\rangle^\dagger/(\mathfrak{f}, zr_\mathfrak{f} - 1).$$

By Conjecture 2.3.5, there exist $\alpha, \beta, \gamma \in \mathbb{F}_q[t][2\Gamma]$ such that

$$\alpha f + \beta x\frac{\partial f}{\partial x} + \gamma y\frac{\partial f}{\partial y} = r_f.$$

Remark. If Conjecture 2.3.5 were false, then we could just replace r_f with some suitable power of itself everywhere in this section.

Choose $\delta_0, \delta_1, \delta_2 \in \mathbb{Z}_q[t][2p\Gamma]$ and $s_f \in \mathbb{Z}_q[t]$, that lift $\alpha^p, \beta^p, \gamma^p$ and r_f^p, and have as few terms as possible. This time if we denote

$$G(T) = \mathfrak{f}^\sigma(t^p)(x^p(1+z^p\delta_1 T), y^p(1+z^p\delta_2 T)) + (z^p\delta_0\mathfrak{f}^p - s_f z^p + 1)T - \mathfrak{f}^p,$$

then $G(T) \in \mathbb{Z}_q[x, y, z, t]$, and again one checks that

$$G(0) = 0 \bmod p, \qquad \frac{dG}{dT}(0) = 1 \bmod p.$$

So, if $\mathbb{Z}_q\langle x, y, z, t\rangle$ denotes the p-adic completion of $\mathbb{Z}_q[x, y, z, t]$, then by Hensel's Lemma, there exists a unique $Z \in \mathbb{Z}_q\langle x, y, z, t\rangle$, such that $G(Z) = 0$.

Note that the proof of Theorem 4.6.1 still holds with $\mathbb{Z}_q$ replaced by $\mathbb{Z}_q\langle t, z\rangle$. We would like to have similar bounds for the degrees in t and z of the terms of Z in terms of their p-adic valuations. We start with the variable t, and let $\deg_t$ denote the degree in t.

Theorem 4.6.4. *If we denote*

$$c_t = \max\{2(\deg_t f + \mu), \deg_t r_f\},$$

where

$$\mu = \max\{\deg_t \alpha, \deg_t \beta, \deg_t \gamma\},$$

then the terms θ of Z of p-adic valuation i satisfy $\deg_t \theta \leq c_t p i$.

Proof. This can be proved in the same way as Theorem 4.6.1. We focus on the necessary changes. Given rational $0 < \epsilon \leq 1$, let $R_{\epsilon,c}$ be the subring of $\mathbb{Z}_q[p^\epsilon]\langle x, y, t, z\rangle$ of series whose terms of p-adic valuation at most i are of $\deg_t \leq ci$. Let $I_{\epsilon,c}$ be the ideal of $R_{\epsilon,c}$ consisting of the series whose terms of valuation at most i are of $\deg_t < ci$. Put

$$H = p^{-\epsilon}G(p^\epsilon T).$$

Again we want to choose H so that

$$H(T) \in R_{\epsilon,c}[T], \qquad H(0) = 0 \bmod I_{\epsilon,c}, \qquad \frac{dH}{dT}(0) = 1 \bmod I_{\epsilon,c}.$$

This time the first condition is satisfied under the constraint

$$c \geq \frac{p(i\mu + \deg_t f)}{(i-1)\epsilon} \tag{4.9}$$

for all $i \geq 2$, and this is again most restrictive for $i = 2$. The second condition imposes the constraint

$$c > \frac{p \deg_t f}{1-\epsilon}, \tag{4.10}$$

and finally, the third condition is satisfied under the constraint

$$c > \max\{p \deg_t \mathfrak{f} + \max\{\delta_0, \delta_1, \delta_2\}, \deg s_f\} = p \max\{\deg_t f + \mu, \deg_t r_f\}.$$

Again we choose ϵ to balance (4.9) (for $i = 2$) and (4.10)

$$\epsilon = \frac{2\mu + \deg_t f}{2(\mu + \deg_t f)},$$

and for this choice we see that all three constraints are satisfied by any $c > c_t p$. So by Hensel's Lemma, we have that $p^{-\epsilon} Z \in R_{\epsilon,c}$ for all $c > c_t p$. Taking the limit as $c \downarrow c_t p$ now implies the result. □

Now we consider the variable z, and let $\deg_z$ denote the degree in z.

Theorem 4.6.5. *The terms θ of Z of p-adic valuation i satisfy $\deg_z \theta \leq 2pi$.*

Proof. The proof is very similar to that of the previous theorem. This time let $R_{\epsilon,c}$ be be the subring of $\mathbb{Z}_q[p^\epsilon]\langle x, y, t, z\rangle$ of series whose terms of p-adic valuation at most i are of $deg_z \leq ci$. Let $I_{\epsilon,c}$ be the ideal of $R_{\epsilon,c}$ consisting of the series whose terms of valuation at most i are of $deg_z < ci$. The conditions on H give the constraints:

$$c \geq \frac{2p}{\epsilon}, \qquad c > 0, \qquad c > p.$$

If we choose $\epsilon = 1$, then all three constraints are satisfied for $c \geq 2p$. So by Hensel's Lemma, $p^{-1} Z \in R_{1,c}$ for all $c \geq 2p$. □

Since $G(Z) = 0$, and by Theorem 4.6.1, Theorem 4.6.4 and Theorem 4.6.5, the map

$$x \to x^p(1 + z^p \delta_1 Z), \qquad y \to y^p(1 + z^p \delta_2 Z), \qquad t \to t^p,$$

defines a lift F_p on $A^\dagger$ of the p-th power Frobenius map on the ring

$$A = \mathbb{F}_q[x, y, t, z]/(z r_f - 1).$$

Theorem 4.6.6. *For all $(i, j) \in 2\Gamma$, the terms θ of $F_p(x^i y^j)E$ of valuation at most ν' satisfy:*

- *θ is supported on $(8p\nu' + 5p)\Gamma$,*
- *$\deg_t \theta \leq (c_t + \mu)p\nu' + \deg_t f$, with c_t, μ as in Theorem 4.6.4,*
- *$\deg_z \theta \leq 3p\nu'$.*

Proof. Using equation (4.8), this follows from Theorem 4.6.1, Theorem 4.6.4, and Theorem 4.6.5. □

In what follows, we use the definitions and the notation from section 3.6. Again we use the presentation

$$\mathbb{Q}_q[r, \frac{1}{r_\mathfrak{f}}] \cong \mathbb{Q}_q[t, z]/(zr_\mathfrak{f} - 1).$$

For a family of nondegenerate curves, we have the following version of Theorem 4.6.2.

Theorem 4.6.7. *There exist $\psi_t, \psi_z \geq 0$ such that if $\omega = h\omega_\mathfrak{f}$ is a (relative) differential on the family $\mathcal{X}/\mathcal{S}$ defined by $\mathfrak{f}$, with $h \in \mathbb{Z}_q[m\Gamma]$, and d is defined as in Theorem 4.6.2, then the class of $p^d\omega$ in $H^1_{DR}(\mathbf{X}/\mathbf{S})$ can be represented by a differential $\omega' = h'\omega_\mathfrak{f}$, with $h' \in \mathbb{Z}_q[t, z][2\Gamma]$ such that*

$$\deg_t h' \leq \psi_t m,$$

$$\deg_z h' \leq \psi_z m.$$

Proof. It is not hard to prove the existence of ψ_t, ψ_z. We may assume that Γ has unique top and bottom vertices (p_t, q_t),(p_b, q_b). Let $\Lambda_{2\Gamma}$ be the $\mathbb{Z}_q[t, \frac{1}{r_\mathfrak{f}}]$-lattice defined in section 3.7, and let $v_1, \ldots, v_r$ be a basis of $\Lambda_{2\Gamma}$. We can first reduce $p^d h$ modulo $\mathfrak{f}$ to an element h'' which only contains monomials $x^i y^j$ with $q_b \leq j < q_t$, and then solve a linear system over $\mathbb{Q}_q(t)$, with coefficients in $\mathbb{Q}_q[t]$, to express h'' as a linear combination h' of $v_1, \ldots, v_r$, modulo the image of δ reduced modulo $\mathfrak{f}$. It is clear that both $\deg_t(h''), \deg_z(h'')$ are bounded by some constant times m. Moreover, the number of variables of the linear system is bounded by a constant times m, and its degree in t is bounded by a constant. Solving the linear system using Cramer's rule therefore shows that $\deg_t(h'), \deg_z(h')$ are both bounded by some constant times m. The integrality follows by applying Theorem 4.6.2 at every fiber of $\mathcal{X}/\mathcal{S}$ over the integers $\overline{\mathbb{Z}}_q$ of $\overline{\mathbb{Q}}_q$. □

Remark. This proof clearly does not produce good ψ_t, ψ_z. For any given particular family one can usually give much better bounds. We also know a much better way to bound ψ_t, ψ_z in general than in the above proof. However, we have decided to leave this out for now, since the proof is really complicated and it still doesn't give good enough χ_1, χ_2 in Theorem 4.6.8.

We can use ψ_t, ψ_z as above to obtain the following bounds on the Frobenius matrix A_p of our family:

Theorem 4.6.8. *Let ν, ν' be such that*

$$\nu' - \lceil \log_p(-(8p\nu' + 5p - 1)N + 1) \rceil + 1 \geq \nu,$$

with N as in Theorem 4.6.2. Let A_p denote the matrix of F_p^* on $H^1_{rig}(X/S)$ with respect to some basis of the $\mathbb{Z}_q[t, \frac{1}{r_\mathfrak{f}}]$-lattice $\Lambda_{2\Gamma}$. Then modulo p^ν, the matrix $r_\mathfrak{f}^{\chi_1} A_p$ consists of polynomials in t of degree at most χ_2, where

$$\chi_1 = 3p\nu' + (8p\nu' + 5p)\psi_z,$$

$$\chi_2 = (c_t + \mu)p\nu' + \deg_t f + (8p\nu' + 5p)\psi_t + \chi_1 \deg r_\mathfrak{f}.$$

Proof. This follows from Theorem 4.6.6 and Theorem 4.6.7. □

Remark. When Theorem 3.8.8 applies, and $r_\mathfrak{f}$ does not have any multiple roots, we roughly get $\chi_1 = p(\nu - 1)$ and $\chi_2 = \chi_1 \deg r_\mathfrak{f}$. So even if $\psi_t = \psi_z = 0$, the χ_1, χ_2 from Theorem 4.6.8 would still be roughly 3 times too big. The best one can expect in general is $\psi_z = 1$, but then χ_1 is already about 11 times too big. So as our estimates for ψ_t, ψ_z get worse, the χ_1, χ_2 from Theorem 4.6.8 very quickly become useless in practice.

4.7 Solving the differential equation

In this section, we will show how to solve the p-adic differential equation

$$\frac{d\Phi}{dt} = pt^{p-1}\Phi N^{F_p} - N\Phi, \qquad \Phi(0) = \Phi_0,$$

from section 3.8, upto some finite p-adic precision ν, i.e. modulo p^ν. We will use the definitions and notation from section 3.8.

4.7.1 Rewriting the equation

Φ is a $d \times d$ matrix of overconvergent functions.

$$\Phi \in M_{d,d}(\mathbb{Q}_q\langle t, \frac{1}{r_\mathfrak{f}}\rangle^\dagger).$$

In particular this means that there exist $\chi_1, \chi_2 \in \mathbb{N}$ such that the matrix $r_\mathfrak{f}^{\chi_1}\Phi$ is congruent to a matrix of polynomials of degree $\leq \chi_2$ modulo p^ν. Note that such χ_1, χ_2 can be found using either Theorem 3.8.8 or Theorem 4.6.8.

Now recall from Theorem 3.8.2 that if $C = \sum_{i=0}^\infty C_i t^i$ denotes the $d \times d$ matrix over $\mathbb{Q}_q[[t]]$ that satisfies

$$\frac{dC}{dt} + NC = 0, \qquad C_0 = I, \tag{4.11}$$

then we have

$$\Phi = C\Phi_0(C^{F_p})^{-1}. \tag{4.12}$$

We could compute the solution Φ by first computing the fundamental matrix C, and then using this equation. However, from the work of Hubrechts [32], and Gerkmann [27], it is known that it is more efficient to compute Φ directly without first computing C. We will now explain how this is done.

We first define a new matrix

$$K = r_{\mathfrak{f}}^{\chi_1}\Phi.$$

This matrix satisfies the differential equation

$$r_{\mathfrak{f}}\frac{dK}{dt} = K(\chi_1\frac{dr_{\mathfrak{f}}}{dt} + r_{\mathfrak{f}}pt^{p-1}N^{F_p}) - r_{\mathfrak{f}}NK, \qquad K(0) = r_{\mathfrak{f}}(0)^{\chi_1}\Phi_0.$$

Next we take a polynomial $\mathfrak{r} \in \mathbb{Z}_q[t]$ such that the matrices $\mathfrak{r}r_{\mathfrak{f}}N$ and $\mathfrak{r}r_{\mathfrak{f}}N^{F_p}$ both consist of polynomials and $\mathfrak{r}(0) \neq 0 \bmod p$. This is possible since N can only have poles at the zeros of $r_{\mathfrak{f}}$ and $r_{\mathfrak{f}}(0) \neq 0 \bmod p$. Now we multiply by $\mathfrak{r}$ to get rid of all denominators

$$\mathfrak{r}r_{\mathfrak{f}}\frac{dK}{dt} = K(\chi_1\mathfrak{r}\frac{dr_{\mathfrak{f}}}{dt} + pt^{p-1}(\mathfrak{r}r_{\mathfrak{f}}N^{F_p})) - (\mathfrak{r}r_{\mathfrak{f}}N)K.$$

This can then be written as

$$a\frac{dK}{dt} + KX + YK = 0, \tag{4.13}$$

where

$$\begin{aligned} a &= \sum_{i=0}^{\infty} a_i t^i &&= \mathfrak{r}r_{\mathfrak{f}}, \\ X &= \sum_{i=0}^{\infty} X_i t^i &&= -(\chi_1\mathfrak{r}\frac{dr_{\mathfrak{f}}}{dt} + pt^{p-1}(\mathfrak{r}r_{\mathfrak{f}}N^{F_p})), \\ Y &= \sum_{i=0}^{\infty} Y_i t^i &&= \mathfrak{r}r_{\mathfrak{f}}N, \end{aligned}$$

are all matrices of polynomials in t. Note that the element a_0 is a p-adic unit. By the definition of χ_1, χ_2, we know that modulo p^ν the entries of the matrix K are polynomials:

$$K = \sum_{i=0}^{\infty} K_i t^i = \sum_{i=0}^{\chi_2} K_i t^i \bmod p^\nu$$

Substituting this expression for K into the equation, and comparing coefficients, we get

$$\sum_{i+j=k} a_i(j+1)K_{j+1} + K_i X_j + K_i Y_j = 0,$$

for all nonnegative integers k. This can be seen as a linear recurrence relation

$$(i+1)a_0 K_{i+1} = f_i(K_i, K_{i-1}, \ldots, K_{i-(\zeta-1)}) \tag{4.14}$$

for the matrices K_i.

4.7.2 Error propagation bounds

Recall that we want to compute the matrix Φ, or equivalently the matrix K, modulo p^ν. However, working with precision ν while solving the recurrence relation (4.14) will not be sufficient, because of *error propagation* in the computation of the K_i. The error in the matrix K_i will also affect the computation of the matrices $K_{i+1}, K_{i+2}, \ldots$ and this will make the error grow. We need to bound the resulting loss of precision.

So suppose that during the computation of the matrices K_i, we work with precision ν', i.e. we reduce all intermediate results modulo $p^{\nu'}$. This way we obtain a matrix

$$\mathcal{K} = \sum_{i=0}^{\infty} \mathcal{K}_i t^i$$

which is an approximation of K. This matrix satisfies a differential equation

$$a\frac{d\mathcal{K}}{dt} + \mathcal{K}X + Y\mathcal{K} = p^{\nu'}\mathcal{E}_K, \tag{4.15}$$

where $\mathcal{E}_K$ is some matrix with entries in $\mathbb{Z}_q[[t]]$. The new term $p^{\nu'}\mathcal{E}_K$ is the error caused by working with finite precision ν'. We want to find bounds for the matrix $\mathcal{K} - K$.

It follows from (4.12) that

$$K = r_{\mathfrak{f}}^{\chi_1} C \Phi_0 (C^{F_p})^{-1}. \tag{4.16}$$

Now let $\tilde{C} = \sum_{i=0}^{\infty} \tilde{C}_i t^i$ be the matrix defined by

$$\mathcal{K} = r_{\mathfrak{f}}^{\chi_1}(C + \tilde{C})\Phi_0 (C^{F_p})^{-1}. \tag{4.17}$$

Substituting this expression into (4.15), and then substracting (4.13), we get

$$a\frac{d}{dt}(r_{\mathfrak{f}}^{\chi_1}\tilde{C}\Phi_0(C^{F_p})^{-1}) + r_{\mathfrak{f}}^{\chi_1}\tilde{C}\Phi_0(C^{F_p})^{-1}X$$
$$+r_{\mathfrak{f}}^{\chi_1}Y\tilde{C}\Phi_0(C^{F_p})^{-1} = p^{\nu'}\mathcal{E}_K. \tag{4.18}$$

Using that

$$\frac{d(C^{F_p})^{-1}}{dt} = -(C^{F_p})^{-1}(\frac{dC^{F_p}}{dt})(C^{F_p})^{-1} = pt^{p-1}(C^{F_p})^{-1}N^{F_p},$$

the first term can be worked out as follows:

$$\begin{aligned} a\frac{d}{dt}(r_{\mathfrak{f}}^{\chi_1}\tilde{C}\Phi_0(C^{F_p})^{-1}) &= a\tilde{C}\Phi_0\frac{d}{dt}(r_{\mathfrak{f}}^{\chi_1}(C^{F_p})^{-1}) + ar_{\mathfrak{f}}^{\chi_1}\frac{d\tilde{C}}{dt}\Phi_0(C^{F_p})^{-1} \\ &= a\tilde{C}\Phi_0(C^{F_p})^{-1}\left(\chi_1 r_{\mathfrak{f}}^{\chi_1-1}\frac{dr_{\mathfrak{f}}}{dt} + r_{\mathfrak{f}}^{\chi_1}pt^{p-1}N^{F_p}\right) \\ &\quad + ar_{\mathfrak{f}}^{\chi_1}\frac{d\tilde{C}}{dt}\Phi_0(C^{F_p})^{-1}. \end{aligned}$$

Substituting this into (4.18), a lot of terms cancel and we are left with

$$ar_{\mathfrak{f}}^{\chi_1}\frac{d\tilde{C}}{dt}\Phi_0(C^{F_p})^{-1} + r_{\mathfrak{f}}^{\chi_1}Y\tilde{C}\Phi_0(C^{F_p})^{-1} = p^{\nu'}\mathcal{E}_K,$$

or

$$\frac{d\tilde{C}}{dt} + N\tilde{C} = p^{\nu'}\mathcal{E}_{\tilde{C}},$$

where

$$\mathcal{E}_{\tilde{C}} = (\frac{1}{r_{\mathfrak{f}}})^{\chi_1}(\frac{1}{a})\mathcal{E}_K C^{F_p}\Phi_0^{-1}.$$

Now from Theorem 3.8.6, we get logarithmic bounds for the p-adic valuations of the coefficients of $C = \sum_{i=0}^{\infty} C_i t^i$:

$$v_p(C_i) \geq \gamma_1\lceil\log_p i\rceil, \qquad \text{with } \gamma_1 = v_p(\Phi_0^{-1}) + v_{p,0}(\Phi).$$

Note that the transformation

$$\Phi \to (\Phi^{-1})^T,$$
$$N \to -N^T,$$
$$C \to -(C^{-1})^T,$$

which corresponds to the duality functor for overconvergent F-isocrystals, preserves equations (4.11) and (4.12).

So we can apply Theorem 3.8.6 to the transformed system. This way we get logarithmic bounds for the p-adic valuations of the coefficients of $C^{-1} = \sum_{i=0}^{\infty}(C^{-1})_i t^i$ as well:

$$v_p((C_i^{-1})) \geq \gamma_2 \lceil \log_p i \rceil, \qquad \text{with } \gamma_2 = v_p(\Phi_0) + v_{p,0}(\Phi^{-1}).$$

For the coefficients of $C^{F_p} = \sum_{i=0}^{\infty}(C^{F_p})_i t^i$ and $(C^{F_p})^{-1} = \sum_{i=0}^{\infty}((C^{F_p})^{-1})_i t^i$ this implies

$$v_p((C^{F_p})_i) \geq \gamma_1 \lceil \log_p \frac{i}{p} \rceil,$$

$$v_p(((C^{F_p})^{-1})_i) \geq \gamma_2 \lceil \log_p \frac{i}{p} \rceil.$$

Note that $\frac{1}{r_{\mathfrak{f}}}, \frac{1}{a} \in \mathbb{Z}_q[[t]]$, since $r_{\mathfrak{f}}(0), a(0) \neq 0 \bmod p$. Hence for the matrix $\mathcal{E}_{\tilde{C}} = \sum_{i=0}^{\infty}(\mathcal{E}_{\tilde{C}})_i t^i$, we get

$$v_p((\mathcal{E}_{\tilde{C}})_i) \geq \gamma_1 \lceil \log_p \frac{i}{p} \rceil + v_p(\Phi_0^{-1}).$$

We have seen that $\tilde{C}$ satisfies the *inhomogeneous* equation

$$\frac{d\tilde{C}}{dt} + N\tilde{C} = p^{\nu'}\mathcal{E}_{\tilde{C}}.$$

Since C is the solution to the corresponding *homogeneous* equation, we can simplify by letting $L = \sum_{i=0}^{\infty} L_i t^i$ denote the matrix such that

$$\tilde{C} = CL.$$

Then we get

$$p^{\nu'}\mathcal{E}_{\tilde{C}} = \frac{d\tilde{C}}{dt} + N\tilde{C} = C\frac{dL}{dt} + (\frac{dC}{dt} + NC)L = C\frac{dL}{dt},$$

so that

$$\frac{dL}{dt} = p^{\nu'}C^{-1}\mathcal{E}_{\tilde{C}}.$$

Since $K_0 = \mathcal{K}_0$, we know that $\tilde{C}_0 = 0$ and hence $L_0 = 0$. Therefore

$$L = \int p^{\nu'}C^{-1}\mathcal{E}_{\tilde{C}},$$

where the integration constant is 0. Only the terms of degree $\leq i-1$ within the integral will contribute to L_i, and the integration of such a term will introduce a denominator of p-adic valuation at most $\lfloor \log_p i \rfloor$. Hence

$$\begin{aligned} v_p(L_i) &\geq -\lfloor \log_p i \rfloor + \nu' + \min_{j \leq i-1} \{v_p((C^{-1})_j)\} + \min_{j \leq i-1} \{v_p((\mathcal{E}_{\tilde{C}})_j)\} \\ &\geq -\lfloor \log_p i \rfloor + \nu' + \gamma_2 \lceil \log_p (i-1) \rceil + \gamma_1 \lceil \log_p \frac{(i-1)}{p} \rceil + v_p(\Phi_0^{-1}). \end{aligned}$$

Since $\tilde{C} = CL$, we have that

$$\begin{aligned} v_p(\tilde{C}_i) &\geq \min_{j \leq i} v_p(C_j) + \min_{j \leq i} v_p(L_j) \\ &\geq \gamma_1 \lceil \log_p i \rceil - \lfloor \log_p i \rfloor + \nu' + \gamma_2 \lceil \log_p (i-1) \rceil + \gamma_1 \lceil \log_p \frac{(i-1)}{p} \rceil + v_p(\Phi_0^{-1}). \end{aligned}$$

The error that we want to bound can be written as

$$\mathcal{K} - K = r_{\mathfrak{f}}^{\chi_1} \tilde{C} \Phi_0 (C^{F_p})^{-1},$$

and putting everything together, we find

$$\begin{aligned} v_p(\mathcal{K}_i - K_i) &\geq \min_{j \leq i} v_p(\tilde{C}_j) + v_p(\Phi_0) + \min_{j \leq i} v_p(((C^{F_p})^{-1})_j) \\ &\geq \gamma_1 \lceil \log_p i \rceil - \lfloor \log_p i \rfloor + \nu' + \gamma_2 \lceil \log_p (i-1) \rceil + \gamma_1 \lceil \log_p \frac{(i-1)}{p} \rceil + v_p(\Phi_0^{-1}) \\ &\quad + v_p(\Phi_0) + \gamma_2 \lceil \log_p \frac{i}{p} \rceil \\ &\geq \nu' + v_p(\Phi_0) + v_p(\Phi_0^{-1}) + (2\gamma_1 + 2\gamma_2 - 1) \lceil \log_p i \rceil - (\gamma_1 + \gamma_2). \end{aligned}$$

So we have shown:

Theorem 4.7.1. *If we want to compute Φ with precision ν, then while solving the recurrence relation for the K_i, it is sufficient to work with precision ν', where*

$$\nu' \geq \nu - v_p(\Phi_0) - v_p(\Phi_0^{-1}) - (2\gamma_1 + 2\gamma_2 - 1) \lceil \log_p \chi_2 \rceil + (\gamma_1 + \gamma_2),$$

and

$$\begin{aligned} \gamma_1 &= v_p(\Phi_0^{-1}) + v_{p,0}(\Phi), \\ \gamma_2 &= v_p(\Phi_0) \quad + v_{p,0}(\Phi^{-1}). \end{aligned}$$

Alternatively, we can use the bounds from Theorem 3.8.5. Recall from that theorem that

$$v_p(C_i), v_p((C^{-1})_i) \geq -(d-1)(\lfloor \log_p i \rfloor + \min\{0, v_{p,0}(N)\}).$$

This also implies that

$$v_p((C^{F_p})_i), v_p(((C^{F_p})^{-1})_i) \geq -(d-1)(\lfloor \log_p \frac{i}{p} \rfloor + \min\{0, v_{p,0}(N)\}).$$

Repeating all of the above estimates with these new bounds, we find:

Theorem 4.7.2. *If we want to compute Φ with precision ν, then while solving the recurrence relation for the K_i, it is also sufficient to work with precision ν', for*

$$\nu' \geq \nu - v_p(\Phi_0) - v_p(\Phi_0^{-1}) + \delta_1 \lfloor \log_p \chi_2 \rfloor - \delta_2,$$

where

$$\delta_1 = (4(d-1)+1),$$

$$\delta_2 = 2(d-1)(1 + 2\min\{0, v_{p,0}(N)\}).$$

Remark. The bounds from Theorem 4.7.1, and Theorem 4.7.2, are a significant improvement of the ones used by Hubrechts [32], and Gerkmann [27], respectively.

Remark. Since the bounds from Theorem 3.8.5 and Theorem 3.8.6 only hold for $i > 0$ (and not for $i = 0$), Theorem 4.7.1 and Theorem 4.7.2 only hold when $\chi_2 \geq 3$. However, since χ_2 always can be chosen ≥ 3, we have decided not to complicate the statements and proofs of these theorems any further.

4.8 The complete algorithm

In this section, we put all of the material in this chapter together, and work out the details of the algorithm sketched in section 3.9.

The input of this algorithm consists of:

1. a generically nondegenerate Laurent polynomial $f \in \mathbb{F}_{\mathfrak{q}}[t][\mathbb{Z}^2]$, where $\mathfrak{q} = p^a$, and p is a (small) prime,
2. an element τ in some field $\mathbb{F}_q$ that contains $\mathbb{F}_{\mathfrak{q}}$, where $q = p^n$, such that the specialization $f_\tau \in \mathbb{F}_q[\mathbb{Z}^2]$ of f at τ is nondegenerate.

Its output will be the zeta function $Z(\overline{C}_\tau, T)$ of the nondegenerate projective curve $\overline{C}_\tau$ defined by f_τ.

We now describe the different steps of the algorithm. The required p-adic precisions ν_0, ν_1 will be given in section 4.9.

4.8.1 Step 1: Computing the cohomology

First we choose a lift $\mathfrak{f} \in \mathbb{Z}_\mathfrak{q}[t][\mathbb{Z}^2]$ that reduces to f mod p. In practice, we will always take this lift to be contained in $\mathcal{O}_K[t][\mathbb{Z}^2]$, where $\mathcal{O}_K$ denotes the ring of integers of some finite extension of $\mathbb{Q}$. All of the computations in this step can then be carried out in an exact way, so that we don't have to worry about the p-adic precision. We have to:

1. compute the resultant $r_\mathfrak{f}$ as in section 4.1,
2. find a basis $[m_1, \ldots, m_r]$ for M as in section 4.3,
3. find a basis $[n_1, \ldots, n_s]$ for N as in section 4.4,
4. compute the matrix H of the Gauss-Manin connection ∇, with respect to the basis $[n_1, \ldots, n_s]$, as in section 4.5.

4.8.2 Step 2: Computing the Frobenius matrix at $t = 0$

We compute the matrix $B_{p,0}$ of F_p^* acting on $H^1_{rig}(\overline{C}_0)$, with respect to the basis induced by $[n_1, \ldots, n_s]$ and with precision ν_0, as in section 4.6.

4.8.3 Step 3: Solving the differential equation

In this step we solve the differential equation

$$\frac{dB_p}{dt} = pt^{p-1} B_p H^{F_p} - H B_p, \qquad B_p(0) = B_{p,0}, \tag{4.19}$$

with precision ν_1, as in section 4.7 (with $\Phi = B_p, N = H, d = 2g$).

1. We compute χ_1, χ_2 using Theorem 3.8.8 and/or Theorem 4.6.8.

2. We either use Theorem 4.7.1, and take

$$\nu_1' = \nu_1 - v_p(B_{p,0}) - v_p(B_{p,0}^{-1}) - (2\gamma_1 + 2\gamma_2 - 1)\lceil \log_p \chi_2 \rceil + (\gamma_1 + \gamma_2). \tag{4.20}$$

or we use Theorem 4.7.2, and take

$$\nu_1' = \nu_1 - v_p(B_{p,0}) - v_p(B_{p,0}^{-1}) + \delta_1 \lfloor \log_p \chi_2 \rfloor - \delta_2. \tag{4.21}$$

3. Finally, we compute the solution $\mathcal{K} = \sum_{i=0}^{\chi_2} \mathcal{K}_i t^i$ to equation 4.13, by solving the corresponding recurrence relation (4.14) for the matrices $\mathcal{K}_i$, working with precision ν_1'.

4.8.4 Step 4: Computing the zeta function

In this final step we compute the zeta function of $\overline{C}_\tau$, by specializing the Frobenius matrix B_p at τ and using the Lefschetz formula. We compute with precision ν_1.

1. First we compute the Teichmüller lift $\hat{\tau} \in \mathbb{Z}_q$ of the element τ.
2. We substitute the element $\hat{\tau}$ into B_p to find the p-th power Frobenius matrix at $t = \tau$:
$$B_{p,\tau} = r_{\mathfrak{f}}(\hat{\tau})^{-\chi_1} \mathcal{K}(\hat{\tau}).$$
3. We compute the norm of $B_{p,\tau}$ which is the q-th power Frobenius matrix at $t = \tau$:
$$B_{q,\tau} = B_{p,\tau} B_{p,\tau}^{F_p} \dots B_{p,\tau}^{F_p^{(n-1)}}.$$
4. Finally, we compute $Z(\overline{C}_\tau, T)$ by using the Lefschetz formula (3.4):
$$Z(\overline{C}_\tau, T) = \frac{\det(1 - B_{q,\tau} T)}{(1-T)(1-qT)}.$$

Remark. For details on how to represent, and compute efficiently with, elements of p-adic fields, we refer to the survey by Vercauteren [14, chapter 12]. There it is also explained how to compute the Teichmüller lift and the norm in step 4 efficiently.

Remark. To use (4.20), we need to know γ_1, γ_2, or $v_p(B_p), v_p(B_p^{-1})$. We can compute $v_p(B_p)$ by solving (4.19) (for B_p) with precision $\nu = 0$. Similarly, we can compute $v_p(B_p^{-1})$ by solving the (dual) equation

$$\frac{d(B_p^{-1})^T}{dt} = H^T (B_p^{-1})^T - p t^{p-1} (B_p^{-1})^T (H^T)^{F_p} \tag{4.22}$$

(for $(B_p^{-1})^T$) with precision $\nu = 0$. To determine the working precision ν' in these computations, we can either use (4.21), or use (4.20) combined with the a priori bounds for $v_p(B_p), v_p(B_p^{-1})$ from Corollary 3.7.11.

Remark. The matrix $\mathcal{K} = \sum_{i=0}^{\chi_2} \mathcal{K}_i t^i$ usually takes up a lot of memory. However, from the recurrence relation (4.14) we see that to compute $\mathcal{K}_{i+1}$, we only need to know the last ζ matrices $\mathcal{K}_{i-(\zeta-1)}, \ldots, \mathcal{K}_i$. So to save memory, we can divide $\mathcal{K}$ up into parts of length $l \geq \zeta$

$$\mathcal{K} = \sum_{i=0}^{\lceil \chi_2/l \rceil} \mathcal{P}_i, \qquad \text{with} \quad \mathcal{P}_i = \sum_{j=kl}^{(k+1)l-1} \mathcal{K}_j t^j,$$

and if we want to compute $\mathcal{K}(\hat{\tau})$, then after computing $\mathcal{P}_{i+1}$ we only need to store $\mathcal{P}_{i+1}$ and $\sum_{j=0}^{i+1} \mathcal{P}_j(\hat{\tau})$. However, if we now want to repeat the computation for another τ in the same family, then we have to start all over again. Therefore, this approach should only be used if otherwise we run out of memory.

4.9 Precision

Let g be the genus of the generic fiber of the family defined by $\mathfrak{f}$. In section 3.9, we saw that it is enough to compute $\chi(T)$ with p-adic precision ν_2, satisfying

$$p^{\nu_2} \geq 2\binom{2g}{g} q^{\frac{g}{2}}. \tag{4.23}$$

Remark. Actually, the factor $2\binom{2g}{g}$ can be somewhat improved [38]. For our algorithm, this will only change the required precision very slightly.

4.9.1 Bounding ν_1

Since the matrices $B_{p,\tau}$ and $B_{q,\tau}$ are in general not integral, when computing the norm and the characteristic polynomial in step 4, there will be a loss of precision. To what precision ν_1 do we have to know $B_{p,\tau}$ to compute $\chi(T)$ with precision ν_2?

Theorem 4.9.1. *If we know $B_{p,\tau}$ with precision*

$$\nu_1 \geq \nu_2 + (g+1)(d + d_1 + d_2),$$

then we can compute $\chi(T)$ with precision ν_2.

Proof. We use the definitions and the notation from section 3.7. Let $[v_1, \ldots, v_s]$ be a $\mathbb{Z}_q$-basis of $\Lambda_{int,\tau} \cap H^1_{rig}(\overline{C}_\tau)$, and let $L_{p,\tau} \in M_{s,s}(\mathbb{Q}_q)$ be the matrix such that $F_p^* v_j = \sum_{i=1}^s (L_{p,\tau})_{ij} v_i$. Note that this matrix has entries in $\mathbb{Z}_q$. Let $W \in M_{s,s}(\mathbb{Q}_q)$ be the matrix such that $v_j = \sum_{i=1}^s W_{ij} m_i$. From (3.5), we see that

$$v_p(W) \geq -d_2, \qquad\qquad v_p(W^{-1}) \geq -(d + d_1).$$

One can check that

$$B_{p,\tau} = W L_{p,\tau} (W^{-1})^{F_p}.$$

Now let $\tilde{B}_{p,\tau} \in M_{s,s}(\mathbb{Q}_q)$ be a matrix such that $v_p(\tilde{B}_{p,\tau} - B_{p,\tau}) \geq \nu_1$. If $\tilde{L}_{p,\tau}$ denotes the matrix defined by $\tilde{B}_{p,\tau} = W \tilde{L}_{p,\tau} (W^{-1})^{F_p}$, then

$$\tilde{L}_{p,\tau} - L_{p,\tau} = W^{-1} (\tilde{B}_{p,\tau} - B_{p,\tau}) W^{F_p},$$

so that we have

$$v_p(\tilde{L}_{p,\tau} - L_{p,\tau}) \geq \nu_2 + g(d + d_1 + d_2).$$

Since the matrices $L_{p,\tau}, \tilde{L}_{p,\tau}$ are integral, there will be no loss of precision when computing their norms $\tilde{L}_{q,\tau}, L_{q,\tau}$:

$$v_p(\tilde{L}_{q,\tau} - L_{q,\tau}) \geq \nu_2 + g(d + d_1 + d_2).$$

If $\tilde{B}_{q,\tau}$ denotes the norm of $\tilde{B}_{p,\tau}$, then one can check that

$$\tilde{B}_{q,\tau} - B_{q,\tau} = W (\tilde{L}_{q,\tau} - L_{q,\tau}) W^{-1},$$

so that

$$v_p(\tilde{B}_{q,\tau} - B_{q,\tau}) \geq \nu_2 + (g-1)(d + d_1 + d_2),$$

and $\tilde{B}_{q,\tau}$ is correct to precision $\geq \nu_2 + (g-1)(d + d_1 + d_2)$.

Now every coefficient a_i of

$$\chi(T) = \sum_{i=0}^{2g} a_i T^i = \det(1 - B_{q,\tau} T)$$

is a sum of products of at most i entries of $B_{q,\tau}$, each of which has p-adic valuation $\geq -(d + d_1 + d_2)$ by Theorem 3.7.9. Hence by Lemma 3.9.1, the maximum loss of precision in computing a_i is $(i-1)(d + d_1 + d_2)$. Since we only have to compute $a_0, \ldots, a_g$, the polynomial $\tilde{\chi}(T)$ that we compute from the matrix $\tilde{B}_{q,\tau}$ will therefore be correct to precision $\geq \nu_2$. □

Remark. Note that if $[m_1, \ldots, m_s]$ is an integral basis, i.e. if it is a basis for $\Lambda_{int,\tau} \cap H^1_{rig}(\overline{C}_\tau)$, then there will be *no* loss of precision in step 4 of the algorithm, and we can take

$$\nu_1 = \nu_2.$$

4.9.2 Bounding ν_0

We have seen that for the solution B_p to the differential equation (4.19) to be correct to precision ν_1, it is sufficient to work with precision ν_1' during its computation. However, to what precision ν_0 do we have to know the (initial value) matrix $B_{p,0}$? It is enough to take $\nu_0 = \nu_1'$, but one can do a lot better.

1. *Estimating* the loss of precision

Theorem 4.9.2. *Let χ_2 be as in step 3, and H as in step 1 of the algorithm. Also, let γ_1, γ_2 be defined as in Theorem 4.7.1 (with $\Phi = B_p$). If we know $B_{p,0}$ either with precision*

$$\nu_0 \geq \nu_1 - (\gamma_1 + \gamma_2)\lceil \log_p \chi_2 \rceil + \gamma_2, \tag{4.24}$$

or with precision

$$\nu_0 \geq \nu_1 + 2(2g-1)(\lfloor \log_p \chi_2 \rfloor - \min\{0, v_{p,0}(H)\}) - (2g-1),$$

then we can compute B_p with precision ν_1.

Proof. We use the definitions and the notation from section 4.7. Suppose that $\tilde{B}_{p,0} \in M_{s,s}(\mathbb{Q}_q)$ is a matrix such that $v_p(\tilde{B}_{p,0} - B_{p,0}) \geq \nu_0$. Let $\tilde{B}_p$ denote the solution to the differential equation (4.19), with initial condition $\tilde{B}_p(0) = \tilde{B}_{p,0}$. We write

$$K = \sum_{i=0}^{\infty} K_i t^i = r_{\mathfrak{f}}^{\chi_1} B_p, \qquad \tilde{K} = \sum_{i=0}^{\infty} \tilde{K}_i t^i = r_{\mathfrak{f}}^{\chi_1} \tilde{B}_p.$$

Recall that all of the elements $r_{\mathfrak{f}}$, $\frac{1}{r_{\mathfrak{f}}}$, $r_{\mathfrak{f}}(0)$,$r_{\mathfrak{f}}(0)^{-1}$ are p-adically integral. Hence computing B_p (starting from $B_{p,0}$), with precision ν_1, is equivalent to computing K (starting from K_0), with precision ν_1. From equation (4.16), it follows that

$$K = r_{\mathfrak{f}}^{\chi_1} C B_{p,0} (C^{F_p})^{-1}, \qquad \tilde{K} = r_{\mathfrak{f}}^{\chi_1} C \tilde{B}_{p,0} (C^{F_p})^{-1}.$$

So we obtain

$$\tilde{K} - K = r_{\mathfrak{f}}^{\chi_1} C (\tilde{B}_{p,0} - B_{p,0}) (C^{F_p})^{-1}. \tag{4.25}$$

Now we can either use the bounds

$$v_p(C_i) \geq \gamma_1 \lceil \log_p i \rceil, \qquad v_p(((C^{F_p})^{-1})_i) \geq \gamma_2 \lceil \log_p \frac{i}{p} \rceil,$$

or alternatively (recall that $d = 2g$)

$$v_p(C_i) \geq -(2g-1)(\lfloor \log_p i \rfloor + \min\{0, v_{p,0}(H)\}),$$

$$v_p(((C^{F_p})^{-1})_i) \geq -(2g-1)(\lfloor \log_p \frac{i}{p} \rfloor + \min\{0, v_{p,0}(H)\}),$$

from section 4.7, to deduce that

$$v_p(\tilde{K}_i - K_i) \geq \nu_0 + (\gamma_1 + \gamma_2)\lceil \log_p i \rceil - \gamma_2,$$

and

$$v_p(\tilde{K}_i - K_i) \geq \nu_0 - 2(2g-1)(\lfloor \log_p i \rfloor + \min\{0, v_{p,0}(H)\}) + (2g-1),$$

respectively. To compute K with precision ν_1, only the matrices K_i for $i \leq \chi_2$ are needed. So with ν_0 as in the theorem, the matrix $\tilde{K}$ (and hence also the matrix $\tilde{B}_p$) will be correct to precision ν_1. □

Remark. These bounds for ν_0 are roughly halfway between ν_1 and the corresponding bounds for ν_1'. In the papers [13, 32], the authors take $\nu_0 = \nu_1'$, so our bounds are quite an improvement compared to this.

Remark. If the basis $[m_1, \ldots, m_s]$ for $H^1_{rig}(Y/S)$ is not chosen too badly, then the bound for ν_0 in terms of γ_1, γ_2 is usually the best one. When $[m_1, \ldots, m_s]$ is an integral basis, i.e. a basis for the $\mathbb{Z}_q[t, \frac{1}{r_\dagger}]$-lattice $\Lambda_{int} \cap H^1_{rig}(Y/S)$, then

$$v_p(B_p) = 0, \qquad\qquad v_p(B_p^{-1}) = -1,$$

so that $\gamma_1 = \gamma_2 = -1$, and we can take

$$\nu_0 \geq \nu_1 + 2\lceil \log_p \chi_2 \rceil - 1 \tag{4.26}$$

Experimentally this bound turns out to be near optimal. It also seems to hold for non-integral bases. Heuristically this can be explained, because we can transform from an arbitrary basis $[m_1, \ldots, m_s]$ to an integral one, solve the differential equation with respect to that basis, and then transform back. However, we have not yet been able to turn this idea into a bound for ν_0, that is as good as (4.26).

2. *Computing* the loss of precision

For a general basis $[m_1, \ldots, m_s]$ of $H^1_{rig}(Y/S)$, the bounds from Theorem 4.9.2 are usually still quite far from optimal. We now explain how we can *compute* a sharp bound.

Let $E^{(\sigma\tau)} \in M_{s,s}(\mathbb{Q}_q)$ be the matrix defined by

$$E^{(\sigma\tau)}_{uv} = \begin{cases} 1 \text{ if } (u,v) = (\sigma,\tau) \\ 0 \text{ otherwise} \end{cases}$$

Since $v_p(\tilde{B}_{p,0} - B_{p,0}) \geq \nu_0$, we can write

$$r_{\mathfrak{f}}(0)^{\chi_1}(\tilde{B}_{p,0} - B_{p,0}) = p^{\nu_0} \sum_{\sigma,\tau \leq s} \epsilon_{\sigma\tau} E^{(\sigma\tau)},$$

with $\epsilon_{\sigma\tau} \in \mathbb{Z}_q$ for all σ, τ. Substituting this into (4.25), we obtain

$$\tilde{K} - K = p^{\nu_0} \sum_{\sigma,\tau \leq s} \epsilon_{\sigma\tau} r_{\mathfrak{f}}^{\chi_1} C r_{\mathfrak{f}}(0)^{-\chi_1} E^{(\sigma\tau)} (C^{F_p})^{-1} = p^{\nu_0} \sum_{\sigma,\tau \leq s} \epsilon_{\sigma\tau} L^{(\sigma\tau)},$$

where

$$L^{(\sigma\tau)} = \sum_{i=0}^{\infty} L_i^{(\sigma\tau)} t^i = r_{\mathfrak{f}}^{\chi_1} C r_{\mathfrak{f}}(0)^{-\chi_1} E^{(\sigma\tau)} (C^{F_p})^{-1}.$$

Comparing with equation (4.16), we see that the matrices $L^{(\sigma\tau)}$ satisfy the same differential equation (4.13) as K

$$a\frac{dL^{(\sigma\tau)}}{dt} + L^{(\sigma\tau)} X + Y L^{(\sigma\tau)} = 0, \tag{4.27}$$

with the initial condition

$$L^{(\sigma\tau)}(0) = E^{(\sigma\tau)}.$$

Since the $\epsilon_{\sigma\tau}$ are integral, we have

$$v_p(\tilde{K}_i - K_i) \geq \nu_0 + \min_{\sigma,\tau \leq s} \{v_p(L_i^{\sigma\tau})\}.$$

So we need to bound the valuations of the $L_i^{(\sigma\tau)}$. Now instead of using the logarithmic bounds on the coefficients of the matrices $C, (C^{F_p})^{-1}$, like in the proof of Theorem 4.9.2, we can just compute the $L_i^{(\sigma\tau)}$ as in section 4.7, with precision $\nu = 0$, to obtain their valuations. This way we can compute

$$\Delta = \min_{i \leq \chi_2} \min_{\sigma,\tau \leq s} \{v_p(L_i^{\sigma\tau})\},$$

and it is then sufficient to take

$$\nu_0 \geq \nu_1 + \Delta.$$

Remark. To compute this bound, we have to solve $s^2 = 4g^2$ differential equations, each one of which (in practice) takes about the same time to solve as the differential equation (4.19). It is much better (contrary to what we are used to) to instead compute the matrices C, C^{-1} by solving the differential equations:

$$\frac{dC}{dt} + NC = 0,$$

$$\frac{d(C^{-1})^T}{dt} - N^T(C^{-1})^T = 0,$$

so that we can compute

$$\Delta = \min_{i \leq \chi_2}\{v_p(C_i)\} + \min_{i \leq \chi_2}\{v_p(((C^{-1})^{F_p})_i)\}.$$

Then by (4.25) it is again sufficient to take

$$\nu_0 \geq \nu_1 + \Delta.$$

We will not go into more detail on how to solve the equations for C, C^{-1}, or how to determine the required working precision. This is all very similar to what we have seen in section 4.7.

4.10 Complexity

In this section, we discuss the complexity of the algorithm. For simplicity, we suppose that p and $\deg_t f$ are fixed, so that we can ignore these parameters.

Recall that our family is defined over some finite field $\mathbb{F}_\mathfrak{q}$ with $\mathfrak{q} = p^a$ elements, and that we want to compute the zeta function of the (complete) fiber $\overline{C}_\tau$ of this family at some point $\tau \in S(\mathbb{F}_q)$, where $\mathbb{F}_q$ is a finite field with $q = p^n$ elements which contains $\mathbb{F}_\mathfrak{q}$. Let g denote the genus of a generic fiber of the family.

We will use the *soft-Oh* symbol $\tilde{\mathcal{O}}$ from complexity theory. It ignores factors that are logarithmic in the input size. So by $h_1 = \tilde{\mathcal{O}}(h_2)$ we mean that there exist constants $\kappa_0, \kappa_1, \kappa_2$, such that

$$h_1 \leq \kappa_0 \log(g)^{\kappa_1} \log(n)^{\kappa_2} h_2.$$

Remark. The bounds in this section are only intended to give a good indication of the complexity of the algorithm, and should not be taken too seriously. Some numbers that come up in the estimates are always very small in practice, or can be chosen that way, but it is hard to either bound them, or explain in full detail why and how they can always be chosen to be small. We will not spend too much time on this. In our opinion, asymptotic bounds for $g, n \to \infty$ are not very useful anyway, when the algorithm can only be used in practice for $g \leq 10, n \leq 100$.

Using *Fast Fourier Transform* methods, field operations on elements of $\mathbb{Q}_{p^k}$, with precision l, can be done in time and space $\tilde{\mathcal{O}}(kl)$ [5].

From (4.23), we see that we can take $\nu_2 = \tilde{\mathcal{O}}(ng)$. We assume that we can always take d_1, d_2 to be 0. Since $d = \tilde{O}(1)$, by Theorem 4.9.1 we have $\nu_1 = \tilde{\mathcal{O}}(ng)$ as well. Finally, by Corollary 3.7.11, γ_1, γ_2 are both $\tilde{\mathcal{O}}(1)$, hence ν_1', ν_0 are still $\tilde{\mathcal{O}}(ng)$ by Theorem 4.7.1 and Theorem 4.9.2. So from now on we will assume that we are computing with p-adic precision $\tilde{\mathcal{O}}(ng)$.

Step 1

The Smith normal form of a matrix A of size m and degree $\mathfrak{d}$ over $\mathbb{Q}_\mathfrak{q}[t]$ can be computed in time $\tilde{\mathcal{O}}(m^6\mathfrak{d}^3(ang))$ and space $\tilde{\mathcal{O}}(m^4\mathfrak{d}^2(ang))$. Moreover, the resulting Smith form S, and the multiplier matrices P, Q (and P^{-1}, Q^{-1}), can be taken to be of degree $\tilde{\mathcal{O}}(m\mathfrak{d})$. For all of this see [54].

By the degree of an element of $\mathbb{Q}_q(t)$ we mean the maximum of the degrees of its numerator and denominator. The matrices of which we need to compute the Smith form over $\mathbb{Q}_q[t]$, or the inverse over $\mathbb{Q}_q(t)$, are all of size $\tilde{\mathcal{O}}(g)$ and can be taken to be of degree $\tilde{\mathcal{O}}(1)$. So all computations in this step can be done in time $\tilde{\mathcal{O}}(ang^7)$ and space $\tilde{\mathcal{O}}(ang^5)$. The degrees of $\alpha, \beta, \gamma, r_\mathfrak{f}, H$ are all $\tilde{\mathcal{O}}(g)$.

Step 2

Slightly adaptating the complexity analysis from [12], we get that the matrix $B_{p,0}$ can be computed in time $\tilde{\mathcal{O}}(an^2g^{6.5})$ and space $\tilde{\mathcal{O}}(an^2g^4)$, for 'most common' Newton polygons.

Step 3

We suppose that Theorem 3.8.8 applies. If we ignore the contributions from the change of basis matrices W, W^{-1}, and from the exponents $\lambda_1, \ldots, \lambda_s$ of the monodromy at the singularities, both of which are always very small, but not easy to bound in terms of a, g, then we see that

$$\chi_1 = \tilde{\mathcal{O}}(ng), \qquad \chi_2 = \tilde{\mathcal{O}}(\chi_1 \deg(r_\mathfrak{f})) = \tilde{\mathcal{O}}(ng^2).$$

The depth ζ of the recursion relation for the matrices $\mathcal{K}_i$ is $\tilde{\mathcal{O}}(g)$. So at each step we have to compute $\tilde{\mathcal{O}}(g)$ products of matrices of size $2g$. This can be done in time $\tilde{\mathcal{O}}(ang^5)$. Hence step 3 can be done in time $\tilde{\mathcal{O}}(an^2g^7)$ and space $\tilde{\mathcal{O}}(an^2g^5)$.

Step 4

This step is the same for all deformation algorithms. It is known (see for example [32]) that it can be done in time $\tilde{\mathcal{O}}(n^3g^5)$ and space $\tilde{\mathcal{O}}(n^2g^3)$.

Taking the maximum over all steps, we find a time complexity of $\tilde{\mathcal{O}}(n^3g^7)$, and a space complexity of $\tilde{\mathcal{O}}(an^2g^5)$.

The algorithm of Castryck, Denef and Vercauteren has a time complexity of $\tilde{\mathcal{O}}(n^3g^{6.5})$, and a space complexity of $\tilde{\mathcal{O}}(n^3g^4)$ [12]. So the complexity of our algorithm is slightly worse.

Chapter 5

Examples

We have implemented the complete algorithm of chapter 4 in the computer algebra package MAGMA, and have used the code to compute some first examples.

Remark. All computations were done with MAGMA V2.15-13 running on a Pentium IV 2.4 Ghz. The stated amounts of memory exclude the 7 MB it takes to load MAGMA.

5.1 A family of genus 3 in characteristic 3

We consider the family of $C_{4,3}$ curves of genus 3 over $\mathbb{F}_3$ defined by

$$f = y^4 + x^3 + x + 1 + txy \in \mathbb{F}_3[t][\mathbb{Z}^2],$$

and let $\mathfrak{f}$ denote the lift of f to $\mathbb{Z}_3[t][\mathbb{Z}^2]$, given by the 'same' equation.

Step 1

We have

$$r_{\mathfrak{f}} = 24800000t^{12} - 8546080000t^8 + 623027552256t^4 - 15494111297536,$$

and the roots of this polynomial are all different modulo 3. A basis $[m1, \ldots, m_{13}]$ for M is given by

$$[x^3y^2, x^2y^2 - xy^2, x^2y^3, x^3y, x^2y, x^3y^3, y^4, x^4, x^5, x^6, y^3, y^6, y^5]\omega_{\mathfrak{f}},$$

where $[m1, \ldots, m_6]$ is a basis for N. For this basis, we have $d_1 = d_2 = 0$, and $d = 2$. We compute the matrix H of ∇, with respect to $[m1, \ldots, m_6]$, and find that $r_{\mathfrak{f}}H$ consists of polynomials in t (of degree at most 13). Hence H has at most a simple pole at the zeroes of $r_{\mathfrak{f}}$. The exponents of the monodromy at these points are all 0. If we apply the basis transformation $m'_j = \sum_{i=1}^{6} W_{ij} m_i$, where

$$W = \begin{pmatrix} 0 & 0 & -t^2 & 0 & 0 & -4/11t \\ 0 & 1/7t^2 & 0 & 0 & 0 & 0 \\ t & 0 & 0 & 0 & 0 & 0 \\ 0 & -3/7t & t & 0 & -1/4t & -2/11 \\ 0 & 5/28t & -1/12t^5 & -3/4t^2 & 0 & -3/11 \\ 0 & 0 & t^3 & 0 & 0 & 0 \end{pmatrix},$$

then the matrix H' of ∇, with respect to $[m_1, \ldots, m_6]$, has a simple pole at ∞, and the exponents of the monodromy are $-\frac{17}{5}, -\frac{16}{5}, -3, -\frac{14}{5}, -\frac{13}{5}, -1$. So we can use Theorem 3.8.8 to compute χ_1, χ_2 in step 3. The matrix W has a pole of order 5, and W^{-1} does not have a pole, at ∞. All the computations in step 1 together took about 10 seconds, and 7 Mb of memory.

Step 2

We compute the matrix $B_{3,0}$ with precision ν_0, using our implementation of the Castryck-Denef-Vercauteren algorithm. The following table contains the required time and memory of this step as a function of the precision ν_0.

ν_0	time (in s)	memory (in MB)
24	10	5
56	59	18
119	710	84
247	15090	244

We note that $v_3(B_{3,0}) = 0, v_3(B_{3,0}^{-1}) = -1$.

Step 3

We first determine $v_3(B_3), v_3(B_3^{-1})$. We have the a priori bounds $v_3(B_3) \geq -2$, $v_3(B_3^{-1}) \geq -3$. Solving the differential equation (4.19) and its dual with precision $\nu = 0$, we find

$$v_3(B_3) = 0, \qquad v_3(B_3^{-1}) = -1.$$

Substituting these values for $v_3(B_3^{-1}), v_3(B_3^{-1})$ into (4.20) and (4.24), we see that when solving the differential equation in step 3, we can take

$$\nu'_1 = \nu_1 + 5\lceil \log_3 \chi_2 \rceil - 2, \qquad \nu_0 = \nu_1 + 2\lceil \log_3 \chi_2 \rceil - 1.$$

We now solve the differential equation (4.19) with these bounds. The following table contains the required time and memory of this computation as a function of the precision ν_1.

ν_1	ν_1'	ν_0	time (in s)	memory (in MB)
13	42	24	14	3
43	77	56	41	10
104	143	119	97	19
230	274	247	220	50

Step 4

Since $v_3(B_3) = 0$, there will not be any additional loss of precision. The following table contains the required time for step 4, for a random $\tau \in S(\mathbb{F}_q)$ with $q = 3^n$, as a function of n. We have also included the necessary precision ν_1.

n	ν_1	time (in s)
6	13	0
26	43	1
67	104	12
151	230	233

The (extra) memory for this step is negligible.

Putting everything together, the time and memory requirements for computing the zeta function of the complete fiber $\overline{C}_\tau$ of the family, at a random point $\tau \in S(\mathbb{F}_q)$ with $q = 3^n$, are as follows.

n	time (in s)	memory (in Mb)
6	34	15
26	111	35
67	829	110
151	15552	301

5.2 A family of genus 4 in characteristic 2

We consider the family of nondegenerate curves of genus 4 over $\mathbb{F}_2$ defined by

$$f = 1 + x^3(1 + y + y^2) + xy^3 + tx^2y \in \mathbb{F}_2[t][\mathbb{Z}^2],$$

and let ƒ denote the lift of f to $\mathbb{Z}_2[t][\mathbb{Z}^2]$, given by the 'same' equation.

Step 1

We have

$$r_{\mathfrak{f}} = 6834375t^{13} + 14175000t^{11} + 395482500t^{10} + 12150000t^9 + 10153998000t^8$$
$$+3536751600t^7 + 49209373692t^6 - 42328551753t^5 + 66550389507t^4 - 25218745497t^3$$
$$+363703706658t^2 + 137127821832t + 21036078189,$$

and the roots of this polynomial are all different modulo 2. A basis $[m1, \ldots, m_{14}]$ for M is given by

$$[3x^5y^4 + (3t^2 - 6)x^5y + (-t^4 + 3t^2 - 2)x^3y^4, 3x^4y^2 + 2tx^3y^4, 27x^6y + (-7t^3 - 36t + 18)x^3y^4,$$
$$3x^4y - tx^3y^4, 3x^5y^2 + 6x^5y + (-t^2 - 6)x^3y^4, 3x^5y + (-t^2 - 1)x^3y^4 + 3x^3y^2, 6tx^5y + 3x^4y^4$$
$$+ (-2t^3 - 6t)x^3y^4, 5x^4y^4 + 10x^4y^3 - 10tx^3y^4, x^5, x^2y^6, -3x^5y + (t^2 - 1)x^3y^4, x^4, x^6, 7x^3y^4]\omega_{\mathfrak{f}}$$

where $[m1, \ldots, m_8]$ is a basis for N. For this basis, we have $d_1 = 0$, $d_2 = 2$, and $d = 2$. We compute the matrix H of ∇, with respect to $[m1, \ldots, m_8]$, and find that $r_{\mathfrak{f}}H$ consists of polynomials in t (of degree at most 16). The exponents of the monodromy at these points are all 0. If we apply the basis transformation $m'_j = \sum_{i=1}^{8} W_{ij} m_i$, where

$$W = t^{-3} \begin{pmatrix}
\frac{1}{3}t^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{3}t^5 & \frac{-4}{9}t^3 & -2 & 0 & \frac{2}{3}t^3 & 0 & 0 \\
0 & 0 & \frac{1}{27}t^4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{3}t^5 & 0 & t^3 & 0 & 0 \\
0 & \frac{2}{3}t^4 & \frac{4}{27}t^5 & \frac{2}{3}t^4 & \frac{1}{3}t^4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{3}t & 0 & 0 \\
\frac{-1}{6}(t^4 - 3t^2 + 2) & 0 & \frac{-11}{54}(t^4 + \frac{6}{11}t) & \frac{-1}{2}t^3 & \frac{-1}{6}(t^3 + 54t) & \frac{-1}{6}t^3 & \frac{1}{6}t & 0 \\
\frac{1}{10}(t^4 - 3t^2 + 2) & 0 & \frac{-1}{30}(3t^4 - 2t) & \frac{-1}{2}t^3 & \frac{17}{10}(t^3 + \frac{54}{17}t) & \frac{1}{10}t^3 & \frac{-1}{30}(t^3 + 3t) & \frac{-1}{15}t^3
\end{pmatrix}$$

then the matrix H' of ∇, with respect to $[m_1, \ldots, m_8]$, has a simple pole at ∞, and the exponents of the monodromy are $-4, -\frac{10}{3}, -\frac{8}{3}, -1, -\frac{4}{5}, -\frac{3}{5}, -\frac{2}{5}, -\frac{1}{5}$. So we can use Theorem 3.8.8 to compute χ_1, χ_2 in step 3. The matrix W has a pole of order 2, and W^{-1} has a pole of order 3, at ∞. All the computations in step 1 together took about 70 seconds, and 18 Mb of memory.

Step 2

We compute the matrix $B_{2,0}$ with precision ν_0, using our implementation of the Castryck-Denef-Vercauteren algorithm. The following table contains the required time and memory of this step as a function of the precision ν_0.

ν_0	time (in s)	memory (in MB)
23	149	34
54	531	43
117	2546	56
244	16712	96

We note that $v_2(B_{2,0}) = -2, v_2(B_{2,0}^{-1}) = -3$.

Step 3

We first determine $v_2(B_2), v_2(B_2^{-1})$. We have the a priori bounds $v_2(B_2) \geq -4$, $v_2(B_2^{-1}) \geq -5$. Solving the differential equation (4.19) and its dual with precision $\nu = 0$, we find

$$v_2(B_2) = -4, \qquad v_2(B_2^{-1}) = -3.$$

Substituting these values for $v_2(B_2^{-1}), v_2(B_2^{-1})$ into (4.20) and (4.24), we see that when solving the differential equation in step 3, we can take

$$\nu_1' = \nu_1 + 25\lceil \log_2 \chi_2 \rceil - 12, \qquad \nu_0 = \nu_1 + 12\lceil \log_2 \chi_2 \rceil - 6.$$

However, in practice ν_1' and ν_0 can be chosen *much* smaller. In the case of ν_0 this makes a big difference. Therefore, we first compute ν_0 for each ν_1 as in section 4.9. We then solve the differential equation (4.19) with this ν_0, and ν_1' as above. The following table contains the required time and memory of this computation as a function of the precision ν_1.

ν_1	ν_1'	ν_0	time (in s)	memory (in MB)
7	220	23	27	7
34	297	54	69	15
95	383	117	166	37
221	534	244	379	82

The experimentally determined ν_0 satisfy

$$\nu_0 \approx \nu_1 + 2\lceil \log_2 \chi_2 \rceil - 1.$$

Step 4

Since $v_2(B_2) = -4$ and $g = 4$, we have to take $\nu_1 \geq \nu_2 + 20$. The following table contains the required time for step 4, for a random $\tau \in S(\mathbb{F}_q)$ with $q = 2^n$, as a function of n. We have also included the necessary precisions ν_1, ν_2.

n	ν_1	ν_2	time (in s)
-	-	-	-
3	34	14	0
33	94	74	7
96	220	200	98

The (extra) memory for this step is negligible.

Putting everything together, the time and memory requirements for computing the zeta function of the complete fiber $\overline{C}_\tau$ of the family, at a random point $\tau \in S(\mathbb{F}_q)$ with $q = 2^n$, are as follows.

n	time (in s)	memory (in Mb)
-	-	-
3	670	76
33	2789	111
96	17259	151

5.3 A family of genus 4 in characteristic 5

We consider the family of nondegenerate curves of genus 4 over $\mathbb{F}_5$ defined by

$$f = y^6 + x^3 + 1 + txy \in \mathbb{F}_5[t][\mathbb{Z}^2],$$

and let $\mathfrak{f}$ denote the lift of f to $\mathbb{Z}_5[t][\mathbb{Z}^2]$, given by the 'same' equation. Note that this family is not C_{ab}, because 3 and 6 are not coprime.

Step 1

We have

$$r_{\mathfrak{f}} = (t^6 - 432)^3,$$

A basis $[m_1, \ldots, m_{19}]$ for M is given by

$$[x^2y^3, x^2y^4, 9tx^2y^8 - 9xy^7 - t^3xy^4, x^2y^5, xy^5, xy^3, xy^2, x^2y^7,$$

$$x^3, y^7, y^8, y^9, y^{10}, y^{11}, y^{12}, x^2, xy^4, 9x^2y^8 - t^2xy^4, x^4]\omega_{\mathfrak{f}}$$

where $[m1, \ldots, m_8]$ is a basis for N. For this basis, we have $d_1 = 0$,$d_2 = 1$, and $d = 1$. We compute the matrix H of ∇, with respect to $[m1, \ldots, m_8]$, and find that $(t^6 - 432)H$ consists of polynomials in t (of degree at most 7). Hence H has at most a simple pole at the zeroes of $r_{\mathfrak{f}}$. The exponents of the monodromy at these points are all 0. If we apply the basis transformation $m'_j = \sum_{i=1}^{8} W_{ij} m_i$, where

$$W = \begin{pmatrix} t & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & t & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/27t & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1/3t & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1/3t^4 & 0 & 15/4t^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/18t^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

then the matrix H' of ∇, with respect to $[m_1, \ldots, m_8]$, has a simple pole at ∞, and the exponents of the monodromy are $-3, -\frac{8}{3}, -\frac{8}{3}, -\frac{8}{3}, -\frac{7}{3}, -\frac{7}{3}, -\frac{7}{3}, -1$. So we can use Theorem 3.8.8 to compute χ_1, χ_2 in step 3. The matrix W has a pole of order 4, and W^{-1} does not have a pole, at ∞. All the computations in step 1 together took about 11 seconds, and 4 Mb of memory.

Step 2

We compute the matrix $B_{5,0}$ with precision ν_0, using our implementation of the Castryck-Denef-Vercauteren algorithm. The following table contains the required time and memory of this step as a function of the precision ν_0.

ν_0	time (in s)	memory (in MB)
27	141	7
59	297	11
122	625	13
250	1353	35

We note that $v_5(B_{5,0}) = 0, v_5(B_{5,0}^{-1}) = -1$.

Step 3

We first determine $v_5(B_5), v_5(B_5^{-1})$. We have the a priori bounds $v_5(B_5) \geq -2$, $v_5(B_5^{-1}) \geq -3$. Solving the differential equation (4.19) and its dual with precision $\nu = 0$, we find

$$v_5(B_5) = 0, \qquad v_5(B_5^{-1}) = -1.$$

Substituting these values for $v_5(B_5^{-1}), v_5(B_5^{-1})$ into (4.20) and (4.24), we see that when solving the differential equation in step 3, we can take

$$\nu_1' = \nu_1 + 5\lceil \log_5 \chi_2 \rceil - 2, \qquad \nu_0 = \nu_1 + 2\lceil \log_5 \chi_2 \rceil - 1.$$

We now solve the differential equation (4.19) with these bounds. The following table contains the required time and memory of this computation as a function of the precision ν_1.

ν_1	ν_1'	ν_0	time (in s)	memory (in MB)
20	38	27	56	10
50	73	59	138	13
112	140	122	320	26
239	267	250	683	65

Step 4

Since $v_5(B_5) = 0$, there will not be any additional loss of precision. The following table contains the required time for step 4, for a random $\tau \in S(\mathbb{F}_q)$ with $q = 5^n$, as a function of n. We have also included the necessary precision ν_1.

n	ν_1	time (in s)
8	20	0
23	50	1
54	112	7
117	238	76

The (extra) memory for this step is negligible.

Putting everything together, the time and memory requirements for computing the zeta function of the complete fiber $\overline{C}_\tau$ of the family, at a random point $\tau \in S(\mathbb{F}_q)$ with $q = 5^n$, are as follows.

n	time (in s)	memory (in Mb)
8	208	21
23	447	28
54	963	43
117	2123	104

Chapter 6

A mixed sparse effective nullstellensatz

For $k \geq n+1$ Laurent polynomials $f_1, \ldots, f_k$ in n variables which are allowed to have different Newton polytopes, we give a generically satisfied criterion, under which we can find Laurent polynomials $h_1, \ldots, h_k$, with particularly small Newton polytopes, such that $h_1 f_1 + \ldots + h_k f_k = 1$. This gives a refinement of a result of Canny and Emiris.

This chapter has now been published [52].

6.1 Introduction

6.1.1 The effective nullstellensatz problem

Let K be a field and $\overline{K}$ an algebraic closure of K. If $f_1, \ldots, f_k$ denote k polynomials or Laurent polynomials in n variables over K, that do not have a common zero in $\overline{K}^n$ or $(\overline{K}^\times)^n$ respectively, then *Hilbert's nullstellensatz* says that they generate the unit ideal, i.e. that there exist n-variable polynomials over K, or Laurent polynomials over K respectively, denoted by $g_1, \ldots, g_k$ such that

$$f_1 g_1 + \ldots + f_k g_k = 1. \tag{6.1}$$

The *effective nullstellensatz problem* is to find upper bounds on the (total) degrees or Newton polytopes of the g_i, such that there exist a solution to (6.1) satisfying these bounds. Note that such bounds allow one to write down a finite dimensional

linear system over K for the coefficients of the g_i. If one wants to find concrete g_i satisfying (6.1), then solving this linear system is usually more efficient than using methods that involve computing a Gröbner basis.

Over the last twenty years there has been a lot of work on the effective nullstellensatz problem. In [40] Kollar gave sharp bounds on the degrees of the g_i in terms of those of the f_i, when the f_i are polynomials of degree at least three. Ten years later Sombra [50] gave bounds for the (joint) Newton polytope of the g_i in terms of the one of the f_i, for both polynomials and Laurent polynomials, which are better than Kollar's bound when the f_i are *sparse*, i.e. have a small Newton polytope compared to their degrees. Unfortunately these bounds are not sharp. More recently in [33] Jelonek has given an elementary proof of Kollar's bounds and extended them to the degree two case. Sombra has informed us that Jelonek's results can also be used to improve his bounds for the Newton polytope of the g_i.

All of these results hold for *arbitrary* (Laurent) polynomials. This amount of generality does come at a cost. Even Kollar's bounds, which are known to be sharp, are still very big in practice. However, for a sufficiently *generic* set of f_i one can find *much* better bounds. In the rest of this paper we will only consider such generic results.

6.1.2 Generic effective nullstellensätze

Let us first introduce some terminology and notation. We identify the algebra of Laurent polynomials in n variables $x_1, \ldots, x_n$ over K with the group algebra $K[\mathbb{Z}^n]$ by identifying the monomial $x^i = x_1^{i_1} \ldots x_n^{i_n}$ with the point $i = (i_1, \ldots, i_n) \in \mathbb{Z}^n$. For a set $S \subset \mathbb{Z}^n$ we will denote by $K[S] \subset K[\mathbb{Z}^n]$ the vector space generated by S. The *support* of an element $h \in K[\mathbb{Z}^n]$ is the set of exponent vectors corresponding to monomials that appear with a nonzero coefficient in h and the *Newton polytope* of h is the convex hull of this support. For any subset σ of $\mathbb{R}^n$ we denote by h_σ the Laurent polynomial obtained from h by setting all coefficients of monomials whose exponent vectors lie outside of σ equal to zero. Note that a *face* of a polytope in $\mathbb{R}^n$ can have codimension ranging from 0 to n and that a *facet* is a face of codimension 1. We recall that for convex polytopes $\Gamma_1, \ldots, \Gamma_k$ the *Minkowski sum* is defined as $\Gamma_1 + \ldots + \Gamma_k = \{p_1 + \ldots + p_k | \; p_i \in \Gamma_i\}$ and is again a convex polytope. Finally, note that if $\Gamma_1, \ldots, \Gamma_k$ are convex polytopes in $\mathbb{R}^n$ and we denote $S_i = \Gamma_i \cap \mathbb{Z}^n$, then choosing $f_1, \ldots, f_k \in K[\mathbb{Z}^n]$ with each f_i supported on Γ_i, is equivalent to picking a K-rational point on $\prod_{i=1}^k \mathbb{A}_K^{|S_i|}$.

In their work on computing *sparse resultants*, Canny and Emiris obtained the following *generic effective nullstellensatz*.

Theorem 6.1.1. *[11, Theorem 8.1] Let $\Gamma_0, \ldots, \Gamma_n$ be convex polytopes in $\mathbb{R}^n$ with vertices in $\mathbb{Z}^n$. Put $\Gamma = \sum_{i=0}^n \Gamma_i$ and suppose that $dim(\Gamma) = n$. There exists a Zariski dense open subset U of $\prod_{i=0}^n \mathbb{A}_K^{|S_i|}$ with the following property. If $f_0, \ldots, f_n \in K[\mathbb{Z}^n]$ are such that the support of each f_i is contained in Γ_i, and they correspond to a point of U, then for any $g \in K[\mathbb{Z}^n]$ with support in Γ there exist $h_0, \ldots, h_n \in K[\mathbb{Z}^n]$, where each h_i has support contained in $\sum_{j \neq i} \Gamma_j$, such that $g = h_0 f_0 + \ldots + h_n f_n$.*

Remark. Note that by applying the theorem to a monomial g and then dividing out by it, one finds $h_0, \ldots, h_n$ with equally small (but shifted) Newton polytopes such that $h_0 f_0 + \ldots + h_n f_n = 1$. So Theorem 6.1.1 is indeed an effective nullstellensatz type result.

To be able to compare the approach of Canny and Emiris with ours, and because we have formulated their result in a slightly different way than they did, we will now explain very briefly how Theorem 6.1.1 follows from their work. We refer to [11] and [15, section 7.6] for more details.

Sketch of the argument:

Again let $\Gamma_0, \ldots, \Gamma_n$ be convex polytopes in $\mathbb{R}^n$ with vertices in $\mathbb{Z}^n$ and suppose that $\Gamma = \sum_{i=0}^n \Gamma_i$ is of dimension n. Put $S_i = \Gamma_i \cap \mathbb{Z}^n$, let $W \subset \prod_{i=0}^n \mathbb{A}_K^{|S_i|}$ denote the set of points corresponding to Laurent polynomials $f_0, \ldots, f_n$ that have a common zero in $\overline{K}^\times$, and let Z be the Zariski closure of W in $\prod_{i=0}^n \mathbb{A}_K^{|S_i|}$. From the work of Gelfand, Kapranov and Zelevinsky [26], it is known that Z is the zero locus of a single irreducible polynomial $R_{\Gamma_0,\ldots,\Gamma_n}$ in the coefficients of the f_i which is called the *mixed sparse resultant* of $\Gamma_0, \ldots, \Gamma_n$. The goal of Canny and Emiris in [11] is to compute this polynomial $R_{\Gamma_0,\ldots,\Gamma_n}$.

First they *lift* the polytopes $\Gamma_0, \ldots, \Gamma_n \subset \mathbb{R}^n$ to $\mathbb{R}^{n+1}$ by picking random vectors $l_0, \ldots, l_n \in \mathbb{Z}^n$ and considering the polytopes

$$\hat{\Gamma}_i = \{(v, l_i \cdot v) |\ v \in \Gamma_i\} \subset \mathbb{R}^{n+1}.$$

Put $\hat{\Gamma} = \hat{\Gamma}_0 + \ldots + \hat{\Gamma}_n$. We say that a facet of $\hat{\Gamma}$ is a *lower facet* if its inward pointing normal has a positive last coordinate. If the l_i are sufficiently generic, and we will assume that they are, then the projection $\mathbb{R}^{n+1} \to \mathbb{R}^n$ onto the first n coordinates carries the lower facets of $\hat{\Gamma}$ to the cells of a so called *coherent mixed subdivision* of Γ. Every cell R of this mixed subdivision is the projection of a unique lower facet $\hat{R}$ of $\hat{\Gamma}$, and one can show that such an $\hat{R}$ can be *uniquely* written as

$$\hat{R} = \hat{F}_0 + \ldots + \hat{F}_n,$$

where $\hat{F}_i$ is a face of $\hat{\Gamma}_i$. Let F_i denote the projection of $\hat{F}_i$. Then F_i is also a face of Γ_i. Now the sum

$$R = F_0 + \ldots + F_n$$

is called *coherent* and one can show that $dim(R) = dim(F_0) + \ldots + dim(F_n)$. Since R has dimension n, this implies that at least one of the F_i is a vertex.

Now Canny and Emiris take $\mathcal{E} = \mathbb{Z}^n \cap (\Gamma + \delta)$ for some small $\delta \in \mathbb{R}^n$ which is chosen in such a way that for every $\alpha \in \mathcal{E}$ there is a cell R of the mixed subdivision such that α lies in the *interior* of $R + \delta$. For $i = 0, \ldots, n$ put

$$T_i = \{\alpha \in \mathcal{E} | \text{ if } \alpha \in R + \delta \text{ and } R = F_0 + \ldots + F_n \text{ is coherent}$$
$$\text{then } i \text{ is the largest index such that } F_i \text{ is a vertex } \},$$

so that $T_0 \sqcup \ldots \sqcup T_n = \mathcal{E}$. For $\alpha \in T_i$ let $v(\alpha) \in \mathbb{Z}^n$ be such that $F_i = \{v(\alpha)\}$. Again take $f_0, \ldots, f_n \in K[\mathbb{Z}^n]$, with each f_i supported on Γ_i. One can show that if $\alpha \in T_i$, then $x^{\alpha - v(\alpha)} f_i$ is supported on $\mathcal{E}$. Therefore we get a well-defined linear map $\phi : K[\mathcal{E}] \to K[\mathcal{E}]$ that sends $\alpha \in T_i$ to $x^{\alpha - v(\alpha)} f_i$. The matrix M of ϕ with respect to the basis given by the elements of $\mathcal{E}$ is called a *Newton matrix*. Now suppose that this matrix M is *nonsingular* i.e. that $\det(M) \neq 0$. Then ϕ is surjective, so that any g with support in $\Gamma + \delta$ can be written as

$$g = \sum_{i=0}^{n} (\sum_{\alpha \in T_i} h_{\alpha,i} x^{\alpha - v(\alpha)}) f_i \ ,$$

for some $h_{\alpha,i} \in K$. Note that if $\alpha \in T_i$, then $\alpha - v(\alpha) \in \sum_{j \neq i} \Gamma_j + \delta$. So the Laurent polynomial $h_i = \sum_{\alpha \in T_i} h_{\alpha,i} x^{\alpha - v(\alpha)}$ is supported on $\sum_{j \neq i} \Gamma_j + \delta$.

For any sequence $f_0, \ldots, f_n$ as above we have defined a Newton matrix M. By taking f_i with general coefficients, $\det(M)$ can be regarded as a regular function on $\prod_{i=0}^{n} \mathbb{A}_K^{|S_i|}$, i.e. as a polynomial in the coefficients of the f_i. Canny and Emiris prove that the resultant $R_{\Gamma_0, \ldots, \Gamma_n}$ is a divisor of this polynomial. They show that $det(M)$ is *not* identically zero by explicitly checking that it is nonzero for some well chosen point on $\prod_{i=0}^{n} \mathbb{A}_K^{|S_i|}$. Therefore $\det(M) \neq 0$ defines a Zariski dense subset of $\prod_{i=0}^{n} \mathbb{A}_K^{|S_i|}$.

This would already imply Theorem 6.1.1 if δ were equal to 0, but unfortunately this is not the case. If we choose δ small enough, then we will always have that $\mathbb{Z}^n \cap (\Gamma + \delta) \subset \mathbb{Z}^n \cap \Gamma$, and $\mathbb{Z}^n \cap (\sum_{j \neq i} \Gamma_j + \delta) \subset \mathbb{Z}^n \cap (\sum_{j \neq i} \Gamma_j)$ for all i. However, in general these inclusions will be strict and we get something slightly weaker than Theorem 6.1.1. This can be solved by considering multiple δ. It is not hard to see that there exist $\delta_1, \ldots, \delta_j$ such that $\cup_{i=1}^{j} (\mathbb{Z}^n \cap (\Gamma + \delta_i)) = \mathbb{Z}^n \cap \Gamma$. If we denote the corresponding Newton matrices by $M_1, \ldots, M_j$, then we can take U to be the Zariski dense subset of $\prod_{i=0}^{n} \mathbb{A}_K^{|S_i|}$ where *none* of the $\det(M_i)$ vanish. □

The bounds in Theorem 6.1.1 apply to almost all sequences $f_0, \ldots, f_n$ for which each f_i has support in Γ_i. However, what is missing is a good criterion to

determine whether the theorem applies to some *particular* sequence $f_0, \ldots, f_n$. Although one can in principle construct all of the Newton matrices and compute their determinants, this is cumbersome and the condition that one obtains this way will in general depend on the choices made in the construction. In the *unmixed* case, i.e. when all of the Γ_i coincide, Castryck, Denef and Vercauteren obtained the following result.

Theorem 6.1.2. *[12] Let Γ be a convex polytope in $\mathbb{R}^n$ with vertices in $\mathbb{Z}^n$ and suppose that $dim(\Gamma) = n$. Let $f_0, \ldots, f_n \in K[\mathbb{Z}^n]$ have support in Γ. Suppose that for every face γ of Γ (of any codimension), the system $(f_0)_\gamma = \ldots = (f_n)_\gamma = 0$ does not have a solution in $(\overline{K}^\times)^n$. Then for any $g \in K[\mathbb{Z}^n]$ with support in $(n+1)\Gamma$ there exist $h_0, \ldots, h_n \in K[\mathbb{Z}^n]$ with support in $n\Gamma$ such that $g = h_0 f_0 + \ldots + h_n f_n$.*

Remark. As we will see in the next section, the condition in Theorem 6.1.2 is indeed generically satisfied, i.e. there exists a Zariski dense subset U of $\prod_{i=0}^n \mathbb{A}_K^{|S|}$, such that for every point of U the corresponding $f_0, \ldots, f_n$ satisfy the condition. So in the unmixed case Theorem 6.1.2 is a refinement of Theorem 6.1.1.

Remark. In the recent preprint [56] Wulcan has given a proof of Theorem 6.1.2 when $K = \mathbb{C}$ by using a combination of toric geometry and complex analysis. Apart from the complex analysis her methods are somewhat similar to ours.

In this paper we extend Theorem 6.1.2 to the general *mixed* case. We show that Theorem 6.1.1 applies to a particular sequence $f_0, \ldots, f_n$, if this sequence is what we call *nondegenerate* with respect to $\Gamma_0, \ldots, \Gamma_n$, a notion which will be defined in the next section. In the unmixed case this nondegenerateness condition coincides with the condition in Theorem 6.1.2. Our main result can be stated as follows.

Theorem 6.1.3. *Let $\Gamma_1, \ldots, \Gamma_k$ be convex polytopes in $\mathbb{R}^n$ with vertices in $\mathbb{Z}^n$, denote $\Gamma = \sum_{i=1}^k \Gamma_i$, and suppose that $dim(\Gamma) = n$. Let $f_1, \ldots, f_k \in K[\mathbb{Z}^n]$ be such that for all i the support of f_i is contained in Γ_i. Suppose that $f_1, \ldots, f_k$ are nondegenerate with respect to $\Gamma_1, \ldots, \Gamma_k$. For all $g \in K[\mathbb{Z}^n]$ with support in Γ there exist $h_1, \ldots, h_k \in K[\mathbb{Z}^n]$, where h_i has support in $\sum_{j \neq i} \Gamma_i$, such that $g = h_1 f_1 + \ldots + h_k f_k$. Conversely if for all $g \in K[\mathbb{Z}^n]$ with support in Γ there exist $h_1, \ldots, h_k \in K[\mathbb{Z}^n]$, where h_i has support in $\sum_{j \neq i} \Gamma_i$, such that $g = h_1 f_1 + \ldots + h_k f_k$, then the f_i are nondegenerate with respect to the Γ_i.*

Remark. This extends Theorem 6.1.2 and is a refinement of Theorem 6.1.1. We remove the restriction that the number of Laurent polynomials has to be $n+1$. While the proof of Theorem 6.1.1 extends easily to the case of $\geq n+1$ Laurent polynomials, the proof of Theorem 6.1.2 does not. So even in the unmixed case we improve on Theorem 6.1.2. We will see that the nondegenerateness condition is generically satisfied for $k \geq n+1$, so that we indeed obtain a refinement of Theorem 6.1.1. Finally, our result shows that the nondegenerateness condition is a necessary one, so that it really gives a criterion for when Theorem 6.1.1 applies.

Remark. Experts probably expected something like Theorem 6.1.3 to be true, but even Theorem 6.1.2 did not appear in the literature before [12]. Our proof of Theorem 6.1.3 is short, elegant, and works in positive characteristic as well. We use techniques from [26] and some toric geometry.

6.2 Nondegenerateness

For a vector $\mathbf{v} \in \mathbb{R}^n$ and a convex polytope $\Gamma \subset \mathbb{R}^n$ we write $m(\mathbf{v}, \Gamma) = \min_{\mathbf{x}\in\Gamma}(\mathbf{v} \cdot \mathbf{x})$. The *first meet locus* of $\mathbf{v}$ with Γ is then defined as $F(\mathbf{v}, \Gamma) = \{x \in \Gamma | \mathbf{v} \cdot \mathbf{x} = m(\mathbf{v}, \Gamma)\}$ and is a face of Γ. Let $\Gamma_1, \ldots, \Gamma_k$ be convex polytopes in $\mathbb{R}^n$ with vertices in $\mathbb{Z}^n$ and let $f_1, \ldots, f_k \in K[\mathbb{Z}^n]$ be such that for all i the support of f_i is contained in Γ_i. Put $\Gamma = \Gamma_1 + \ldots + \Gamma_k$ and suppose that $dim(\Gamma) = n$.

Definition 6.2.1. We say that the f_i are *nondegenerate* with respect to the Γ_i, if for any vector $\mathbf{v} \in \mathbb{R}^n$ the system

$$(f_1)_{F(\mathbf{v},\Gamma_1)} = \ldots = (f_k)_{F(\mathbf{v},\Gamma_k)} = 0$$

does not have a solution in $(\overline{K}^{\times})^n$.

Remark. In Lemma 6.3.1 we will see that geometrically this means that $f_1, \ldots, f_k$ don't have a common zero in the projective toric variety associated to Γ.

Clearly the condition only has to be verified for a finite number of vectors, but we can be a bit more precise. Let $\Sigma : \prod_{i=1}^k \Gamma_i \to \Gamma$ denote the addition map. Then for any vector $\mathbf{v} \in \mathbb{R}^n$ we have that

$$\Sigma^{-1}(F(\mathbf{v}, \Gamma)) = \prod_{i=1}^{k} F(\mathbf{v}, \Gamma_i). \tag{6.2}$$

So two vectors have a different first meet locus with some Γ_i if and only if they have a different first meet locus with Γ. Therefore it suffices to take a set of vectors representing all possible first meet loci of Γ.

When $k \geq n + 1$, it is to be expected that the nondegenerateness condition is generically satisfied, because then a system of k equations in n variables will in general not have a solution. Let us check that this is indeed the case.

Lemma 6.2.2. *If $k \geq n + 1$, then there exists a Zariski dense subset $U \subset \prod_{i=1}^k \mathbb{A}_K^{|S_i|}$, such that for every point of U the corresponding $f_1, \ldots, f_k$ are nondegenerate with respect to the Γ_i.*

Proof. For a vector $\mathbf{v} \in \mathbb{R}^n$, let $Y_{\mathbf{v}} \subset \prod_{i=1}^k \mathbb{A}_K^{|S_i|} \times \mathbb{G}_m^n$ denote the algebraic subvariety defined by the k equations

$$(\sum_{(j_1,\ldots,j_n)\in S_i} a_{i,j_1,\ldots,j_n} x_1^{j_1} \ldots x_n^{j_n})_{F(\mathbf{v},\Gamma_i)} = 0,$$

where the $a_{i,j_1,\ldots,j_n}$ with $(j_1,\ldots,j_n) \in S_i$ denote the standard coordinates on $\mathbb{A}_K^{|S_i|}$. Let π_1 and π_2 denote the projections of $\prod_{i=1}^k \mathbb{A}_K^{|S_i|} \times \mathbb{G}_m^n$ onto its factors. In each fiber of π_2 the variety $Y_{\mathbf{v}}$ cuts out a linear subvariety of codimension at least $n+1$. This implies that $Y_{\mathbf{v}}$ has codimension at least $n+1$ in $\prod_{i=1}^k \mathbb{A}_K^{|S_i|} \times \mathbb{G}_m^n$ and that the codimension of $\pi_1(Y_{\mathbf{v}})$ in $\prod_{i=1}^k \mathbb{A}_K^{|S_i|}$ is at least 1. However, $\pi_1(Y_{\mathbf{v}})$ corresponds exactly to those $f_1,\ldots,f_k$ for which the nondegenerateness condition with respect to $\mathbf{v}$ is *not* satisfied. Now we take a set $\mathbf{v}_1,\ldots,\mathbf{v}_m$ of vectors representing all first meet loci of Γ and write $Y = \cup_{i=1}^m \pi_1(X_{\mathbf{v}_i})$. It is clear that Y is of codimension at least 1 in $\prod_{i=1}^k \mathbb{A}_K^{|S_i|}$ and that we can take U to be its complement. □

When $k < n+1$, we expect that generically $f_1,\ldots f_k$ will not be nondegenerate with respect to the Γ_i, again by counting numbers of equations and variables. Actually, they *never* are.

Lemma 6.2.3. *If $k < n+1$, then $f_1,\ldots,f_k$ are not nondegenerate with respect to $\Gamma_1,\ldots,\Gamma_k$.*

Proof. First assume that $k = n$. If the nondegenerateness condition for $f_1,\ldots,f_k$ is satisfied for all vectors $\mathbf{v} \neq 0 \in \mathbb{R}^n$, then the number of solutions of $f_1 = \ldots = f_k = 0$ in $(\overline{K}^\times)^n$ is equal to the *mixed volume* $V(\Gamma_1,\ldots,\Gamma_k)$ of the polytopes $\Gamma_1,\ldots,\Gamma_k$ by ([6],[41]). However, since $dim(\Gamma) = n$ this mixed volume is positive, so the nondegenerateness condition for $\mathbf{v} = 0$ is not satisfied. If for some $k < n$ there were nondegenerate $f_1,\ldots,f_k$, then by adding some zeros (and their Newton polytopes) we would get nondegenerate $f_1,\ldots,f_n$, which as we have seen is not possible. □

6.3 Proof of the main result

In this final section we prove Theorem 6.1.3.

Let X be the projective toric variety associated to Γ. Recall that it can be constructed as follows. For any face γ of Γ let σ_γ be the cone generated by the minimal inward pointing normal vectors with integer coordinates $\mathbf{n}_\mathcal{G}$ of all facets $\mathcal{G}$ of Γ that contain γ. This set of cones forms a *fan* and X is the toric variety

associated to this fan.

Each Γ_i determines an invertible sheaf $\mathcal{L}_i$ on X in the following way. We can define the function

$$\psi_{\Gamma_i}(\mathbf{x}) = m(\mathbf{x}, \Gamma_i) = \min_{\mathbf{y} \in \Gamma_i \cap \mathbb{Z}^n} \mathbf{x} \cdot \mathbf{y}$$

on the fan of X. Let γ be a face of Γ and $\mathcal{G}_1, \ldots, \mathcal{G}_m$ the set of facets of Γ that contain γ. Note that $\mathcal{G}_j = F(\mathbf{n}_{\mathcal{G}_j}, \Gamma)$ for all j. Since $\cap_{j=1}^m F(\mathbf{n}_{\mathcal{G}_j}, \Gamma) \neq \emptyset$ equation (6.2) implies that $\cap_{j=1}^m F(\mathbf{n}_{\mathcal{G}_j}, \Gamma_i) \neq \emptyset$. So on σ_γ the minimum in the definition of ψ_{Γ_i} is realized by any $\mathbf{y} \in \cap_{j=1}^m F(\mathbf{n}_{\mathcal{G}_j}, \Gamma_i)$. Hence ψ_{Γ_i} is linear on the cones of the fan of X and defines an invertible sheaf $\mathcal{L}_i$ on X. Since ψ_{Γ_i} is *convex*, this invertible sheaf is generated by its global sections, which can be identified with $K[\Gamma_i]$. For all of this see ([24], section 3.4). Another way to define the $\mathcal{L}_i$ can be found in ([26], p. 254). It is clear that f_i defines a global section s_i of $\mathcal{L}_i$. Writing $\mathcal{V} = \bigoplus_{i=1}^k \mathcal{L}_i$, we obtain a section $s = \oplus s_i$ of the locally free sheaf $\mathcal{V}$.

Lemma 6.3.1. *$f_1, \ldots, f_k$ are nondegenerate with respect to $\Gamma_1, \ldots, \Gamma_k$ if and only if s is nowhere zero on X.*

Proof. This can be seen by decomposing X into its orbits. Every such orbit τ corresponds to a face of Γ and hence to a (unique) first meet locus $F(\mathbf{v}, \Gamma)$ for some (non-unique) vector $\mathbf{v} \in \mathbb{R}^n$. On τ the zero locus of s_i is defined by $(f_i)_{F(\mathbf{v}, \Gamma_i)}$ divided by any monomial supported on $F(\mathbf{v}, \Gamma_i)$. Indeed such a monomial generates $\mathcal{L}_i$ on some neighbourhood of τ. We see that the f_i are nondegenerate with respect to the Γ_i if and only if the s_i do not have a common zero on any τ, so on X. □

To the pair $(\mathcal{V}, s)$ one can associate the following *(dual) Koszul complex* ([26], p.51) $\mathcal{C}^\bullet$ on X, where the maps are contraction with s, i.e. dual to exterior product with s:

$$\mathcal{C}^\bullet : 0 \longrightarrow \Lambda^k \mathcal{V}^\vee \xrightarrow{(\lrcorner s)_{k-1}} \ldots \longrightarrow \Lambda^2 \mathcal{V}^\vee \xrightarrow{(\lrcorner s)_1} \mathcal{V}^\vee \xrightarrow{(\lrcorner s)_0} O_X \longrightarrow 0$$

Lemma 6.3.2. *The complex $\mathcal{C}^\bullet$ is exact if and only if the section s is nowhere zero on X.*

Proof. Exterior multiplication with a nonzero vector on the exterior algebra of a vector space is exact. This can be seen easily by choosing a basis containing this vector. The same is then true for contraction with such a vector. So if s does not vanish on X, then $\mathcal{C}^\bullet$ will be fiberwise exact. Hence it is also exact as a complex of (locally free) sheaves. If on the other hand s has a zero, then $(\lrcorner s)_0$ will have a nontrivial cokernel. □

So if $f_1, \ldots, f_k$ are nondegenerate with respect to $\Gamma_1, \ldots, \Gamma_k$ then the complex $\mathcal{C}^\bullet$ is exact.

We cannot in general conclude that the complex of global sections of an exact complex $\mathcal{C}^\bullet$ of sheaves on X is exact as well. However, if the terms of the complex don't have any higher cohomology, i.e. if $H^i(X, \mathcal{C}^j) = 0$ for all j and all $i > 0$, then by standard homological algebra the complex of global sections will still be exact, because its cohomology is isomorphic to the cohomology of the zero sheaf, which is zero.

We define $\mathcal{O}_\Gamma = \mathcal{L}_1 \otimes \ldots \otimes \mathcal{L}_k$. The global sections of this invertible sheaf can be identified with $K[\Gamma]$. Now $\mathcal{O}_\Gamma$ is ample, because the corresponding function ψ_Γ is *strictly convex*. If we take the tensor product of an exact complex of sheaves of $\mathcal{O}_X$-modules with $\mathcal{O}_\Gamma$, then the resulting complex will be still be exact, because $\mathcal{O}_\Gamma$ is invertible. On a general projective variety *Serre's vanishing theorem* states that if we do this often enough, i.e. tensor with a high enough power of the ample invertible sheaf $\mathcal{O}_\Gamma$, then the higher cohomology of the terms of the complex will vanish and the complex of global sections will also be exact. We want to show that for $\mathcal{C}^\bullet$ tensoring once with $\mathcal{O}_\Gamma$ is enough. For this we need the following well known lemma.

Lemma 6.3.3. *If an invertible sheaf $\mathcal{L}$ on a projective toric variety X is generated by its global sections, then its higher cohomology vanishes, i.e. $H^i(X, \mathcal{L}) = 0$ for all $i > 0$.*

Proof. See for example ([24], p.74). □

Note that $\Lambda^2 \mathcal{L}_i^\vee = 0$ for all i, since $\mathcal{L}_i$ is invertible. Hence

$$\Lambda^m \mathcal{V}^\vee \cong \Lambda^m (\bigoplus_{i=1}^k \mathcal{L}_i^\vee) \cong \bigoplus_{I \subset \{1,\ldots,k\}, |I|=m} \bigotimes_{i \in I} \mathcal{L}_i^\vee.$$

So by just tensoring once with $\mathcal{O}_\Gamma = \mathcal{L}_1 \otimes \ldots \otimes \mathcal{L}_k$, the duals of the $\mathcal{L}_i$ will disappear from this decomposition. Since the $\mathcal{L}_i$ are generated by global sections, so are the $(\bigotimes_{i \in I} \mathcal{L}_i^\vee) \otimes \mathcal{O}_\Gamma$ and they don't have any higher cohomology. Hence the sheaves $\Lambda^m \mathcal{V}^\vee \otimes \mathcal{O}_\Gamma$ don't have higher cohomology either. We conclude that the complex of global sections of $\mathcal{C}^\bullet \otimes \mathcal{O}_\Gamma$ is exact. Now we take a closer look at the last map in this complex. Note that

$$\mathcal{V}^\vee \otimes \mathcal{O}_\Gamma \cong \bigoplus_{i=1}^k \bigotimes_{j \neq i} \mathcal{L}_j,$$

so that

$$H^0(X, \mathcal{V}^\vee \otimes \mathcal{O}_\Gamma) \cong \bigoplus_{i=1}^k K[\sum_{j \neq i} \Gamma_j].$$

On the other hand we have that

$$H^0(X, \mathcal{O}_\Gamma) \cong K[\Gamma].$$

The map $(\lrcorner s)_0$ between these spaces is given by multiplication with f_i on the i-th factor. Since the complex of global sections is exact, this map is surjective. This finishes the first part of the proof of Theorem 6.1.3.

For the converse, suppose that any $g \in K[\Gamma]$ can be written as $g = h_1 f_1 + \ldots + h_k f_k$, with $h_i \in K[\sum_{j \neq i} \Gamma_j]$ for all i. We have seen that this means that the map $H^0(X, \mathcal{V}^\vee \otimes \mathcal{O}_\Gamma) \to H^0(X, \mathcal{O}_\Gamma)$ is surjective. However, since $\mathcal{O}_\Gamma$ is generated by its global sections, this implies that the map $\mathcal{V}^\vee \otimes \mathcal{O}_\Gamma \to \mathcal{O}_\Gamma$ is surjective as well. Tensoring with $\mathcal{O}(\Gamma)^\vee$, we find that $\mathcal{V}^\vee \to \mathcal{O}$ is surjective, so that the section s is nowhere zero and $f_1, \ldots f_k$ are nondegenerate with respect to $\Gamma_1, \ldots \Gamma_k$ by Lemma 6.3.1. This ends the proof of Theorem 6.1.3.

Remark. We can also also prove the first part of Theorem 6.1.3 more along the lines of Canny and Emiris as follows. We suppose (like they do) that $k = n + 1$, number the f_i starting from $i = 0$, and use the notation from section 6.1.2. Recall that after choosing vectors $l_0, \ldots, l_n$ and δ, they define a Newton matrix M, that $\det(M)$ can be considered as a regular function on $\prod_{i=1}^n \mathbb{A}_K^{|S_i|}$, i.e. as a polynomial in the coefficients of the f_i, and that the resultant $R_{\Gamma_0,\ldots,\Gamma_n}$ is a divisor of this polynomial. Now they vary the partition $\mathcal{E} = T_0 \sqcup \ldots \sqcup T_n$, for example by replacing 'largest' by 'smallest' in the definition of T_i. This way they construct $n + 1$ Newton matrices $M_0, \ldots, M_n$, and prove that

$$R_{\Gamma_0,\ldots,\Gamma_n} = \gcd(\det(M_0), \ldots, \det(M_n)).$$

From this we see that if $R_{\Gamma_0,\ldots,\Gamma_n}(f_0, \ldots, f_n)$ is nonzero for a particular sequence $f_0, \ldots, f_n$, then the specialization of at least one of the Newton matrices $M_0, \ldots, M_n$ at this sequence will be nonsingular. By the same argument as in section 6.1.2 (again using multiple δ) this implies that the conclusion of Theorem 6.1.1 holds for $f_0, \ldots, f_n$. From the work of Gelfand,Kapranov and Zelevinsky [26] it is known that $R_{\Gamma_0,\ldots,\Gamma_n}(f_0, \ldots, f_n)$ is nonzero if and only if $f_0, \ldots, f_n$ are nondegenerate with respect to $\Gamma_0, \ldots, \Gamma_n$. So we obtain a new proof of the first part of Theorem 6.1.3 in this case. However, the second part of Theorem 6.1.3 does *not* seem to follow this way. Also while our first proof was rather short and self contained, this second proof depends both implicitly on some some quite heavy results from [26] about the structure (irreducibility,degree) of $R_{\Gamma_0,\ldots,\Gamma_n}(f_0, \ldots, f_n)$, and explicitly on a lot of arbitrary choices of parameters.

Conclusion

In this thesis we have presented an algorithm to compute the zeta function of a nondegenerate curve over a finite field using deformations in rigid cohomology.

In particular we have:

1. shown that the deformation method can be applied to the class of nondegenerate curves,
2. explained how the rigid cohomology of a family of affine or projective nondegenerate curves and its Gauss-Manin connection can be computed,
3. proved some new bounds on the p-adic precision required for obtaining provably correct results,
4. written a first implementation and computed some examples.

We have only recently completed our implementation which consists of (around) 1500 lines of MAGMA code. Therefore, we have not yet computed a lot of examples. What can be concluded from what we have computed, is that the algorithm does work in ranges of practical interest, and that the heaviest step is usually still applying the algorithm of Castryck, Denef and Vercauteren to the easy fiber. It is therefore probably better to use a fibration algorithm instead to compute the Frobenius matrix of the easy fiber, fibering it into a family of points.

Bibliography

[1] T. Abbot, K. Kedlaya, and D. Roe, *Bounding Picard numbers of surfaces using p-adic cohomology*, Arithmetic, Geometry and Coding Theory (AGCT 2005), Séminaires et Congrès 21, Societé Mathématique de France **2009**, 125–159.

[2] F. Baldassarri and B. Chiarellotto, *Algebraic versus rigid cohomology with logarithmic coefficients*, Barsotti Symposium in Algebraic Geometry (Abano Terme, 1991), Perspect. Math. **15** (1994), 11–50.

[3] V. Batyrev and D. Cox, *On the Hodge structure of projective hypersurfaces in toric varieties*, Duke Math Journal **75** (1994), 293–338.

[4] B.Bektemirov, B. Mazur, W. Stein, and M. Watkins, *Average ranks of elliptic curves: tension between data and conjecture*, Bulletin of the American Mathematical Society **44** (2007), 233–254.

[5] D.J. Bernstein, *Fast multiplication and its applications. Algorithmic number theory: lattices, number fields, curves and cryptography*, Mathematical Sciences Research Institute Publications **44** (2008), 325–384.

[6] D.N. Bernstein, *The number of roots of a system of equations*, Functional analysis and its applications **9** (1975), 1–4.

[7] P. Berthelot, *Cohomologie cristalline des schémas de caractéristique $p > 0$*, Lecture notes in Math., Springer Verlag **407** (1974).

[8] ———, *Géométrie rigide et cohomologie des variétés algébriques de caractéristique p*, Introductions aux cohomologies p-adiques (Luminy 1984), Mém. Soc. Math. France **23** (1986), 7–32.

[9] ———, *Finitude et pureté cohomologique en cohomologie rigide (with an appendix by A.J. de Jong)*, Invent. Math. **128** (1997), 329–372.

[10] E. Bombieri, *On exponential sums in finite fields*, Invent. Math. **88** (1966), 71–105.

[11] J. Canny and I.Z. Emiris, *A subdivision-based algorithm for the sparse resultant*, Journal of the ACM **47** (2000), 417–451.

[12] W. Castryck, J. Denef, and F. Vercauteren, *Computing zeta functions of nondegenerate curves*, International Mathematics Research Papers **2006** (2006), 1–57.

[13] W. Castryck, H. Hubrechts, and F. Vercauteren, *Computing zeta functions in families of C_{ab} curves using deformation*, In A. van der Poorten, A. Stein (Ed.) Algorithmic number theory - ANTS VIII, Lecture Notes in Computer, Springer (2008).

[14] H. Cohen, G. Frey, R. Avanzi, C. Doche, T. Lange, K. Nguyen, and F. Vercauteren, *Handbook of elliptic and hyperelliptic curve cryptography*, Chapman Hall/CRC, 2006.

[15] D. Cox, J. Little, and D. O'Shea, *Using algebraic geometry*, Springer, 2004.

[16] P. Deligne, *La conjecture de Weil I*, Publications mathématiques de l'IHÉS **43** (1974), 273–307.

[17] J. Denef and F. Vercauteren, *Computing zeta functions of C_{ab} curves*, Finite Fields and Their Applications **12** (2006), 78–102.

[18] ———, *An extension of Kedlaya's algorithm to hyperelliptic curves in characteristic 2*, Journal of Cryptology **19** (2006), 1–25.

[19] B. Dwork, *On the rationality of the zeta function of an algebraic variety*, American Journal of Mathematics **82** (1960), 631–648.

[20] B. Dwork and P. Robba, *Effective p-adic bounds for homogeneous linear differential equations*, Trans. Amer. Math. Soc. **259** (1980), 559–577.

[21] E. Ehrhart, *Sur un problème de géométrie diophantienne linéaire. I. Polyèdres et reseaux*, J. Reine Angew. Math. **226** (1967), 1–29.

[22] R. Elkik, *Solutions d'équations à coefficients dans un anneau Hensélien*, Ann. Sci. École Norm. Sup. **6** (1973-1974), 553–603.

[23] J.Y. Étesse and B. le Stum, *Fonctions L associées aux F-isocristaux surconvergents, I. Interprétation cohomologique*, Math. Ann. **296** (1993), 557–576.

[24] W. Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies Princeton University Press, 1993.

[25] P. Gaudry and N. Gürel, *An extension of Kedlaya's algorithm to superelliptic curves*, Advances in Cryptology, Asiacrypt 2001, Lecture Notes in Computer Science **2248** (2001), 480–494.

[26] I.M. Gelfand, M.M. Kapranov, and A.V. Zelevinsky, *Discriminants, resultants and multidimensional determinants*, Birkhäuser, 1994.

[27] R. Gerkmann, *Relative rigid cohomology and deformation of hypersurfaces*, International Mathematical Research Papers **2007** (2007).

[28] P. Griffiths, *Periods of integrals on algebraic manifolds: summary of main results and discussion of open problems*, Bulletin of the American Mathematical Society **76** (1970), 228–296.

[29] D. Harvey, *Kedlaya's algorithm in larger characteristic*, International Mathematics Research Notices **2007** (2007).

[30] M. Hochster, *Rings of invariants of tori, Cohen-Macaulay rings generated by monomials, and polytopes*, Annals of Mathematics **96** (1972).

[31] H. Hubrechts, *Point counting in families of hyperelliptic curves in characteristic 2*, LMS J. Comput. Math **10** (2007), 207–234.

[32] ———, *Point counting in families of hyperelliptic curves*, Foundations of Computational Mathematics **8** (2008), 137–169.

[33] Z. Jelonek, *On the effective nullstellensatz*, Inventiones Mathematicae **162** (2005), 1–17.

[34] K. Kato, *Logarithmic structures of Fontaine-Illusie, Algebraic analysis, geometry, and number theory*, Johns Hopkins Univ. Press, Baltimore, 1989.

[35] K. Kedlaya, *Effective p-adic cohomology for cyclic cubic threefolds*.

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

[37] ———, *Fourier transforms and p-adic Weil II*, Compositio Mathematica **142** (2006), 1426–1450.

[38] ———, *Search techniques for root-unitary polynomials, Computational Arithmetic Geometry*, Contemp. Math. **463** (2008), 71–81.

[39] ———, *p-adic Differential Equations*, Cambridge University Press, to appear.

[40] J. Kollar, *Sharp effective nullstellensatz*, Journal of the AMS **1** (1988), 963–975.

[41] A.G. Kushnirenko, *Newton polytopes and the Bezout theorem*, Functional analysis and its applications **10** (1976), 233–235.

[42] A. Lauder, *Counting solutions to equations in many variables over finite fields*, Foundations of Computational Mathematics **4** (2004), 221–267.

[43] ———, *A recursive method for computing zeta functions of varieties*, LMS J. Comput. Math. **9** (2006), 222–269.

[44] ———, *Ranks of elliptic curves over function fields*, LMS J. Comput. Math. **11** (2008), 172–212.

[45] J. Pila, *Frobenius maps of abelian varieties and finding roots of unity in finite fields*, Math. Comp. **55** (1990), 745–763.

[46] T. Satoh, *The canonical lift of an ordinary elliptic curve over a finite field and its point counting*, Journal of the Ramanujan Mathematical Society **15** (2000), 247–270.

[47] R. Schoof, *Counting points on elliptic curves over finite fields*, 18ièmes journées arithmétiques, Bordeaux 1993, Journal de Théorie des Nombres de Bordeaux **7** (1995), 219–254.

[48] J.P. Serre, *Local algebra*, Monographs in Mathematics. Springer-Verlag, Berlin, 2000.

[49] A. Shiho, *Crystalline fundamental groups. II. Log convergent cohomology and rigid cohomology*, J. Math. Sci. Univ. Tokyo **9** (2002), 1–163.

[50] M. Sombra, *A sparse effective nullstellensatz*, Advances in applied mathematics **22** (1999), 271–295.

[51] B. Le Stum, *Rigid cohomology*, Cambridge university press, 2007.

[52] J. Tuitman, *A refinement of a mixed sparse effective nullstellensatz*, International Mathematical Research Notices **2010** (2010).

[53] M. van der Put, *The cohomology of Monsky and Washnitzer*, Mémoires de la Société Mathématique de France **23** (1983), 33–59.

[54] G. Villard, *Generalized subresultants for computing the Smith normal form of polynomial matrices*, Journal of Symbolic Computation **20** (1995), 269–286.

[55] G. Walker, *Computing zeta functions of varieties via fibration*, Phd thesis, Oxford, 2010.

[56] E. Wulcan, *Sparse effective nullstellensätze via residue currents*, http://arxiv.org/abs/0903.3618 (2009).

Arenberg Doctoral School of Science, Engineering & Technology
Faculty .....
Department ...
Research group ...
Address ...