p-adic differential equations (version of 7 Jan 08)

Kiran S. Kedlaya

Contents

Preface

1. About the book

This book is an outgrowth of a course taught by the author at MIT during fall 2007, on the subject of p-adic ordinary differential equations. The target audience was graduate students with some prior background in algebraic number theory, including exposure to the p-adic numbers, but not necessarily any background in p-adic analytic geometry (of either the Tate or Berkovich flavors).

Some use was made of the book of Dwork, Gerotto, and Sullivan [**DGS94**], particularly in the early parts of the course. We also drew upon the book of Christol [**Chr83**]; as a result, references to both books litter the text. However, we depart from these by adopting some novel strategies, including the following.

- We limit our use of cyclic vectors. This requires an initial investment in the study of matrix inequalities (Chapter 3), but pays off in significantly stronger results.
- We introduce the notion of a Frobenius descendant (Chapter 9). This complements the older construction of Frobenius antecedents, particularly in dealing with certain "boundary cases" where the antecedent method does not apply.

As a result, we end up with some improvements of existing results, including the following.

- We refine the Frobenius antecedent theorem of Christol-Dwork (Theorem 9.4.2).
- We extend some results of Christol-Dwork, on the variation of the generic radius of convergence, to subsidiary radii (Theorem 10.3.2).
- We extend Young's geometric interpretation of subsidiary generic radii of convergence beyond the range of applicability of Newton polygons (Theorem 10.10.2).
- We give quantitative versions of the Christol-Mebkhout decomposition theorem for differential modules on an annulus which are solvable at a boundary (Theorems 11.2.2 and 11.3.1).
- We improve the bound in the transfer theorem to a disc containing a regular singularity with exponents in $\mathbb{Z}_p$ (Theorem 12.7.1).
- We give improvements on the Christol-Dwork-Robba effective bounds for solutions of p-adic differential equations, in the case of nilpotent regular singularities (Theorem 16.2.4), and in the presence of a Frobenius structure (Theorem 16.3.3). The latter implies a result about logarithmic growth of solutions of differential equations with Frobenius structure (Theorem 16.4.6).
- We state a relative version of the p-adic local monodromy theorem, a/k/a Crew's conjecture (Theorem 18.1.8), and describe in detail how it may be derived either from the p-adic index theory of Christol-Mebkhout or from the slope theory for Frobenius modules of Kedlaya.

Some of the new results are expected to be relevant in theory (in the study of higher-dimensional p-adic differential equations in the context of the "semistable reduction problem" for overconvergent F-isocrystals) or in practice (in the explicit computation of solutions of p-adic differential equations, e.g., for finding zeta functions of explicit varieties).

2. Structure of the book

Within each chapter, we have attempted (not completely successfully) to enforce a uniform basic structure. Each chapter begins with a very brief introduction explaining what is to be discussed. After the main body of material, we typically include a section of afternotes, in which we include detailed references for results in that chapter, fill in historical details, and add additional comments. (This practice is modeled on [**Ful98**].) Note that we have a habit of attributing to various authors slightly stronger versions of their theorems than the ones they originally stated; to avoid complicating the discussion in the text, we resolve these misattributions in the afternotes instead. After the afternotes, we typically include a few exercises; while some of these include the proofs of some results which will be used later, all such results are also stated explicitly in the text, so as to avoid cross-referencing into exercises.

The chapters themselves are grouped into several parts, which we now describe briefly. (Chapter 0, being introductory, does not fit into this grouping.)

Part 1 is preliminary, collecting some basic tools of p-adic analysis. However, it also includes some facts of matrix analysis (the study of the variation of numerical invariants attached to matrices, as a function of the matrix entries) which may not be so familiar.

Part 2 introduces some formalism of differential algebra, such as differential rings and modules, twisted polynomials, and cyclic vectors, and applies these to fields equipped with a nonarchimedean norm.

Part 3 begins the study of p-adic differential equations in earnest, developing some basic theory for differential modules on rings and annuli, including the Christol-Dwork theory of variation of the generic radius of convergence, and the Christol-Mebkhout decomposition theory. We also include a discussion of p-adic exponents, culminating in the statement (without proof) of the Christol-Mebkhout structure theorem for p-adic differential modules on an annulus with intrinsic generic radius of convergence everywhere equal to 1.

Part 4 introduces the concept of a Frobenius structure on a p-adic differential module, to the point of stating the p-adic local monodromy theorem (Crew's conjecture, now a theorem of André, Kedlaya, and Mebkhout) and sketching briefly the proof techniques using either p-adic exponents or Frobenius slope filtrations. We also discuss effective convergence bounds for solutions of p-adic differential equations.

Part 5 consists of a series of brief discussions of several areas of application of the theory of p-adic differential equations. These are somewhat more didactic, and much less formal, than in the other parts; they are meant primarily as suggestions for further reading.

3. Acknowledgments

Thanks to the participants of the MIT course 18.787 (Topics in Number Theory, fall 2007) for numerous comments on the lecture notes which ultimately became this book. Particular thanks are due to Ben Brubaker and David Speyer for giving guest lectures, and to Chris Davis, Hansheng Diao, David Harvey, Raju Krishnamoorthy, Ruochuan Liu, and especially Liang Xiao for providing feedback. Additional feedback was provided by Francesco Baldassarri and Bruno Chiarellotto.

During the preparation of the course and of this book, the author was supported by a National Science Foundation CAREER grant (DMS-0545904) and by a Sloan Research Fellowship.

CHAPTER 0

Introductory remarks

In this chapter, we start with some introductory remarks justifying the study of p-adic differential equations. We then give a summary of one of Dwork's original examples in which the p-adic behavior of a classical differential equation, namely a hypergeometric equation, relates to a manifestly number-theoretic question (the number of points on an elliptic curve over a finite field).

Since this chapter is meant as an introduction only, it is rather full of statements for which we give references instead of proofs (if that). Please be assured that this practice is not typical of the rest of the book, at least not until the later chapters.

1. Why p-adic differential equations?

Although the very existence of a highly-developed theory of p-adic ordinary differential equations is not entirely well-known even within number theory, the subject is actually almost fifty years old. Here are some circumstances, past and present, in which it arises; some of these will be taken up again in Part 5.

Variation of zeta functions. The original circumstance in which p-adic differential equations appeared in number theory was Dwork's work on the variation of zeta functions of algebraic varieties over finite fields. Roughly speaking, solving certain p-adic differential equations can give rise to explicit formulas for number of points on varieties over finite fields.

In contrast to methods involving étale cohomology, methods for studying zeta functions based on p-adic analysis (including also the next item) lend themselves well to numerical computation. Interest in computing zeta functions for varieties where straightforward point-counting is not an option (e.g., curves over tremendously large prime fields) has been driven by applications in computer science, the principal example being cryptography based on elliptic or hyperelliptic curves.

p-adic cohomology. Dwork's work suggested, but did not immediately lead to, a proper analogue of étale cohomology based on p-adic analytic techniques. Such an analogue was eventually developed by Berthelot (based on work of Monsky and Washnitzer, and also ideas of Grothendieck); it is called *rigid cohomology* (see the notes for the origin of the word "rigid"). It is not yet a fully functional analogue of étale cohomology, particularly because there are still open problems related to the construction of a good category of coefficients. These problems are rather closely related to questions concerning p-adic differential equations, and in fact some of the results presented in this course have been (or will be) used for this purpose.

p-adic Hodge theory. The subject of p-adic Hodge theory aims to do for the cohomology of varieties over p-adic fields what ordinary Hodge theory does for the cohomology of varieties over $\mathbb{C}$, namely abstract away the variety and enable a better understanding of the

cohomology of the variety as an object in its own right. In the p-adic case, the cohomology in question is often étale cohomology, which carries the structure of a Galois representation.

The study of such representations, as pioneered by Fontaine, involves a number of exotic auxiliary rings (rings of *p-adic periods*) which serve their intended purposes but are otherwise a bit mysterious. More recently, the work of Berger has connected much of the theory to the study of p-adic differential equations; notably, a key result that was originally intended for use in p-adic cohomology (the *p-adic local monodromy theorem*) turned out to imply an important conjecture about Galois representations (Fontaine's conjecture on potential semistability).

Ramification theory. There are some interesting analogies between properties of differential equations over $\mathbb{C}$ with meromorphic singularities, and wildly ramified Galois representations of p-adic fields. At some level, this is suggested by the parallel formulation of the Langlands conjectures in the number field and function field cases. One can use p-adic differential equations to interpolate between the two situations, by associating differential equations to Galois representations (as in the previous item) and then using differential invariants (irregularity) to recover Galois invariants (Artin and Swan conductor).

For representations of the étale fundamental group of a variety over a field of positive characteristic of dimension greater than 1, it is quite a tough problem to construct meaningful numerical invariants from the Galois point of view. Recent work of Abbes and Saito [**AS02, AS03**] attempts to do this, but the resulting quantities are quite difficult to calculate. One can alternatively use p-adic differential equations to define invariants which are somewhat easier to deal with for some purposes; for instance, one can define a *differential Swan conductor* which is guaranteed to be an integer [**Ked07a**], whereas this is not clear for the Abbes-Saito conductor. One can then equate the two conductors, deducing integrality for the Abbes-Saito conductor; this has been carried out by Chiarellotto and Pulita for one-dimensional representations, and by L. Xiao in the general case.

2. Zeta functions of varieties

Definition 0.2.1. For λ in some field K, let E_λ be the elliptic curve over K defined by the equation

$$E_\lambda : y^2 = x(x-1)(x-\lambda)$$

in the projective plane. Remember that there is one point $O = [0 : 1 : 0]$ at infinity, and that there is a natural group law on $E_\lambda(K)$ under which O is the origin, and three points add to zero if and only if they are collinear (or better, if they are the three intersections of E_λ with some line; this correctly allows for degenerate cases).

Theorem 0.2.2 (Hasse). *Suppose λ belongs to a finite field $\mathbb{F}_q$. Then $\#E_\lambda(\mathbb{F}_q) = q + 1 - a_q(\lambda)$ where $|a_q(\lambda)| \leq 2\sqrt{q}$.*

Proof. See [**Sil91**, Theorem V.1.1]. □

Hasse's theorem was later vastly generalized as follows, originally as a set of conjectures by Weil. (Despite no longer being conjectural, these are still commonly referred to as the *Weil conjectures*.)

Definition 0.2.3. For X an algebraic variety over $\mathbb{F}_q$, the *zeta function* of X is defined as the formal power series

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

another way to write it, which makes it look more like zeta functions you've seen before, is

$$\zeta_X(T) = \prod_x (1 - T^{\deg(x)})^{-1},$$

where x runs over Galois orbits of $X(\overline{\mathbb{F}_q})$, and deg is the size of the orbit. (If you prefer algebro-geometric terminology: x runs over closed points of X, and deg is the degree over $\mathbb{F}_q$.)

Example 0.2.4. For $X = E_\lambda$, one checks (exercise) that

$$\zeta_X(T) = \frac{1 - a_q(\lambda)T + qT^2}{(1-T)(1-qT)}.$$

Theorem 0.2.5 (Dwork, Grothendieck, Deligne, et al). *Let X be an algebraic variety over $\mathbb{F}_q$. Then $\zeta_X(T)$ represents a rational function of T. Moreover, if X is smooth and proper of dimension d, we can write*

$$\zeta_X(T) = \frac{P_1(T) \cdots P_{2d-1}(T)}{P_0(T) \cdots P_{2d}(T)},$$

where each $P_i(T)$ has integer coefficients, satisfies $P_i(0) = 1$, and has all roots in $\mathbb{C}$ on the circle $|T| = q^{-i/2}$.

Proof. The proof of this theorem is a sufficiently massive undertaking that even a reference is not reasonable here; instead, we give [**Har77**, Appendix C] as a metareference. □

Remark 0.2.6. It is worth pointing out that the first complete proof of Theorem 0.2.5 used the fact that you can interpret

$$\#X(\mathbb{F}_{q^n}) = \sum_i (-1)^i \operatorname{Trace}(F^n, H^i_{\mathrm{et}}(X, \mathbb{Q}_\ell)),$$

where for any prime $\ell \neq p$, $H^i_{\mathrm{et}}(X, \mathbb{Q}_\ell)$ is the i-th étale cohomology group of X with coefficients in $\mathbb{Q}_\ell$.

3. Zeta functions and p-adic differential equations

Remark 0.3.1. the interpretation of Theorem 0.2.5 in terms of étale cohomology (Remark 0.2.6) is all well and good, but there are several downsides. One important one is that étale cohomology is not explicitly computable; for instance, it is not straightforward to describe étale cohomology to a computer well enough that the computer can make calculations. (The main problem is that while one can write down étale cocycles, it is very hard to tell whether or not a cocycle is a coboundary.)

Another important downside is that you don't get extremely good information about what happens to ζ_X when you vary X. This is where p-adic differential equations enter the picture. It was observed by Dwork that when you have a family of algebraic varieties defined over $\mathbb{Q}$, the same differential equations appear when you study variation of complex periods, and when you study variation of zeta functions over $\mathbb{F}_p$.

Here is an explicit example due to Dwork.

Definition 0.3.2. Recall that the *hypergeometric series*

$$F(a,b;c;z) = \sum_{i=0}^{\infty} \frac{a(a+1)\cdots(a+i)b(b+1)\cdots(b+i)}{c(c+1)\cdots(c+i)i!} z^i$$

satisfies the *hypergeometric differential equation*

$$z(1-z)y'' + (c-(a+b+1)z)y' - aby = 0.$$

Set in particular

$$\alpha(z) = F(1/2, 1/2; 1; z);$$

over $\mathbb{C}$, α is related to an elliptic integral, for instance, by the formula

$$\alpha(\lambda) = \frac{2}{\pi}\int_0^{\pi/2} \frac{d\theta}{\sqrt{1-\lambda\sin^2\theta}} \qquad (0 < \lambda < 1).$$

(You can extend this to complex λ by being careful about branch cuts.) This elliptic integral can be viewed as a period integral for the curve E_λ, i.e., you're integrating some meromorphic form on E_λ around some loop (homology class).

Let $p \neq 2$ be an odd prime. We now try to interpret $\alpha(z)$ as a function of a p-adic variable rather than a complex variable. Beware that this means z can take *any* value in a field with a norm extending the p-adic norm on $\mathbb{Q}$, not just $\mathbb{Q}_p$ itself. (For the moment, you can imagine z running over a completed algebraic closure of $\mathbb{Q}_p$.)

Lemma 0.3.3. *The series $\alpha(z)$ converges p-adically for $|z| < 1$.*

Proof. Straightforward. □

Dwork discovered that a closely related function admits "analytic continuation".

Definition 0.3.4. Define the *Igusa polynomial*

$$H(z) = \sum_{i=0}^{(p-1)/2} \binom{(p-1)/2}{i}^2 z^i.$$

Modulo p, the roots of $H(z)$ are the values of $\lambda \in \overline{\mathbb{F}_p}$ (which actually all belong to $\mathbb{F}_{p^2}$, for reasons we will not discuss) for which E_λ is a supersingular elliptic curve, i.e., $a_q(\lambda) \equiv 0 \pmod{p}$.

Dwork's analytic continuation result is the following.

Theorem 0.3.5 (Dwork). *There exists a series $\xi(z) = \sum_j P_i(z)/H(z)^i$ converging uniformly for $|z| \leq 1$ and $|H(z)| = 1$, with each $P_i(z) \in \mathbb{Q}_p[z]$, such that*

$$\xi(z) = (-1)^{(p-1)/2}\frac{\alpha(z)}{\alpha(z^p)} \qquad (|z| < 1).$$

Proof. See [**vdP86**, §7]. □

Note that ξ itself satisfies a differential equation, which I won't write out just yet. We will see it again later.

Definition 0.3.6. For $\lambda \in \mathbb{F}_q$, let $\mathbb{Z}_q$ be the unramified extension of $\mathbb{Z}_p$ with residue field $\mathbb{F}_q$. Let $[\lambda]$ be the unique q-th root of 1 in $\mathbb{Z}_q$ congruent to λ mod p (the *Teichmüller lift* of λ).

Theorem 0.3.7 (Dwork). *If $q = p^a$ and $\lambda \in \mathbb{F}_q$ is not a root of $H(z)$, then*

$$T^2 - a_q(\lambda)T + q = (T - u)(T - q/u),$$

where

$$u = \xi([\lambda])\xi([\lambda]^p) \cdots \xi([\lambda]^{p^{a-1}}).$$

That is, the quantity u is the "unit root" of the polynomial $T^2 - a_q(\lambda)T + q$ occurring (up to reversal) in the zeta function.

Proof. See [**vdP86**, §7]. □

4. A word of caution

Remark 0.4.1. Before we embark on the study of p-adic ordinary differential equations, a cautionary note is in order, concerning the rather innocuous-looking differential equation $y' = y$. Over $\mathbb{R}$ or $\mathbb{C}$, this equation is nonsingular everywhere, and its solutions $y = ce^x$ are defined everywhere.

Over a p-adic field, things are quite different. As a power series around $x = 0$, we have

$$y = c\sum_{n=0}^{\infty} \frac{x^n}{n!}$$

and the denominators hurt us rather than helping. In fact, the series only converges for $|x| < p^{-1/(p-1)}$ (assuming that we are normalizing $|p| = p^{-1}$). For comparison, note that the logarithm series

$$\log \frac{1}{1-x} = \sum_{n=1}^{\infty} \frac{x^n}{n}$$

converges for $|x| < 1$.

The conclusion to be taken away is that there is no fundamental theorem of ordinary differential equations over the p-adics! In fact, the hypergeometric differential equation in the previous example was somewhat special; the fact that it had a solution in a disc where it had no singularities was not a foregone conclusion. One of Dwork's discoveries is that this typically happens for differential equations that "come from geometry", such as the Picard-Fuchs equations that arise from integrals of algebraic functions (e.g., elliptic integrals). Another is that one can quantify rather well the obstruction to solving a p-adic differential equation in a nonsingular disc, using similar techniques to those used to study obstructions to solving complex differential equations in singular discs.

Notes

The standard reference for properties of elliptic curves is the book of Silverman [**Sil91**].

I alluded above to the notion of an analytic function, defined as a uniform limit of rational functions with poles prescribed to certain regions. To keep down the background required for the course, I will stick throughout to this approach of defining everything in terms of rings, and not making any attempt to introduce analytic geometry over a nonarchimedean field. However, it must be noted that it is much better in the long run to build this theory in terms of nonarchimedean analytic geometry; for example, it is pretty hopeless to deal with partial differential equations without doing so.

That said, there are several ways to develop a theory of analytic spaces over a nonarchimedean field. The traditional method is Tate's theory of rigid analytic spaces, so-called because one develops everything "rigidly" by imitating the theory of schemes in algebraic geometry, but using rings of convergent power series instead of polynomials. The canonical foundational reference for rigid geometry is the book of Bosch, Güntzer, and Remmert [**BGR84**], but novices may find the text of Fresnel and van der Put [**FvdP04**] or the lecture notes of Bosch [**Bos05**] more approachable. A more recent method, which in some ways is more robust, is Berkovich's theory of nonarchimedean analytic spaces (commonly called *Berkovich spaces*), as introduced in [**Brk90**] and further developed in [**Brk93**]. For both points of view, see also the lecture notes of Conrad [**Con07**].

Dwork's original analysis of the Legendre family of elliptic curves via the associated hypergeometric equation appears in [**Dwo69**, §8]. The treatment in [**vdP86**] is more overtly related to p-adic cohomology.

Exercises

(1) Explain why Theorem 0.2.5 implies Hasse's theorem; this includes verifying the formula for the zeta function of E_λ.

(2) Check that the usual formula
$$\liminf_{n\to\infty} |a_n|^{-1/n}$$
for the radius of convergence of the Taylor series $\sum_{n=0}^\infty a_n x^n$ still works over a nonarchimedean field. (That is, the series converges inside that radius, and diverges outside.)

(3) Check that the exponential series has radius of convergence $p^{-1/(p-1)}$.

(4) Show that a power series which converges for $|x| \leq 1$ may have an integral which only converges for $|x| < 1$, but that its derivative still converges for $|x| \leq 1$. This is backwards from the archimedean situation.

Part 1

Tools of p-adic analysis

CHAPTER 1

Absolute values

In this chapter, we recall some basic facts about absolute values, primarily of the nonarchimedean sort; the treatment is not at all comprehensive, as it is only meant as a review. See [**Rob00**] for a fuller treatment.

Beware that a couple of proofs will forward reference the chapter on Newton polygons.

1. Absolute values on abelian groups

Let us start by recalling some basic definitions from analysis, without yet specializing to the nonarchimedean case.

DEFINITION 1.1.1. Let G be an abelian group. An *semiabsolute value* (or *seminorm*) on G is a function $|\cdot|: G \to [0,+\infty)$ satisfying the following conditions.

(a) We have $|0| = 0$.
(b) For $f, g \in G$, $|f+g| \leq |f| + |g|$.

We say the seminorm $|\cdot|$ is an *absolute value* (or *norm*) if the following additional condition holds.

(a') For $g \in G$, $|g| = 0$ if and only if $g = 0$.

We also express this by saying that G is *separated* under $|\cdot|$. A seminorm on an abelian group G induces a metric topology on G, in which the basic open subsets are the open balls, i.e., sets of the form $\{g \in G : |f-g| < r\}$ for some $f \in G$ and some $r > 0$.

DEFINITION 1.1.2. Let G, G' be abelian groups equipped with seminorms $|\cdot|$, $|\cdot|'$, respectively, and let $\phi : G \to G'$ be a homomorphism. Note that ϕ is continuous for the metric topologies on G, G' if and only if there exists a function $h : (0,+\infty) \to (0,+\infty)$ such that for all $r > 0$,

$$\{g \in G : |g| < h(r)\} \subseteq \{g \in G : |\phi(g)| < r\}.$$

We say that ϕ is *submetric* if $|\phi(g)|' \leq |g|$ for all $g \in G$, and *isometric* if $|\phi(g)|' = |g|$ for all $g \in G$. We say two seminorms $|\cdot|_1, |\cdot|_2$ on G are *topologically equivalent* if they induce the same metric topology, i.e., the identity morphism on G is continuous in both directions.

DEFINITION 1.1.3. Let G be an abelian group equipped with a seminorm. A *Cauchy sequence* in G under $|\cdot|$ is a sequence $\{x_n\}_{n=0}^{\infty}$ such that for any $\epsilon > 0$, there exists an integer N such that for all integers $m, n \geq N$, $|x_m - x_n| < \epsilon$. We say the sequence $\{x_n\}_{n=0}^{\infty}$ is *convergent* if there exists $x \in G$ such that for any $\epsilon > 0$, there exists an integer N such that for all integers $n \geq N$, $|x - x_n| < \epsilon$; in this case, the sequence is automatically Cauchy, and we say that x is a *limit* of the sequence; if G is separated under $|\cdot|$, then limits are unique when they exist. We say G is *complete* under $|\cdot|$ if every Cauchy sequence has a unique limit.

THEOREM 1.1.4. *Let G be an abelian group equipped with an absolute value $|\cdot|$. Then there exists an abelian group G' equipped with an absolute value $|\cdot|'$ under which it is complete, and an isometric homomorphism $\phi : G \to G'$ with dense image.*

This is standard, so we only sketch the proof.

PROOF. Take the set of Cauchy sequences in G, and declare two sequences $\{x_n\}_{n=0}^{\infty}, \{y_n\}_{n=0}^{\infty}$ to be equivalent if the sequence $x_0, y_0, x_1, y_1, \dots$ is also Cauchy. This is easily shown to be an equivalence relation; let G' be the set of equivalence classes. It is then straightforward to construct the group operation (termwise addition) and the norm on G' (the limit of the norms of the terms of the sequence); the map ϕ takes $g \in G$ to the constant sequence $\{g, g, \dots\}$. $\square$

DEFINITION 1.1.5. With notation as in Theorem 1.1.4, we call G' the *completion* of G; it (or rather, the group G' equipped with the absolute value $|\cdot|'$ and the homomorphism ϕ) is functorial in G (and in particular, is unique up to unique isomorphism).

DEFINITION 1.1.6. If R is a ring and $|\cdot|$ is a seminorm on its additive group, we say that $|\cdot|$ is *submultiplicative* **index**submultiplicative (seminorm)—textbf if the following additional condition holds.

(c) For $f, g \in R$, $|fg| \leq |f||g|$.

We say that $|\cdot|$ is *multiplicative* **index**multiplicative (seminorm)—textbf if the following additional condition holds.

(c') For $f, g \in R$, $|fg| = |f||g|$.

The completion of a ring R equipped with a submultiplicative seminorm admits a natural ring structure, because the termwise product of two Cauchy sequences is again Cauchy.

LEMMA 1.1.7. *Let F be a field equipped with a multiplicative norm. Then the completion of F is also a field.*

PROOF. Note that if $\{f_n\}_{n=0}^{\infty}$ is a Cauchy sequence in F, then $\{|f_n|\}_{n=0}^{\infty}$ is a Cauchy sequence in $\mathbb{R}$ by the triangle inequality, and so has a limit since $\mathbb{R}$ is complete. Since F is equipped with a true norm, if $\{f_n\}_{n=0}^{\infty}$ does not converge to 0, then $\{|f_n|\}_{n=0}^{\infty}$ must also not converge to 0. In particular, $|f_n|_{n=0}^{\infty}$ is bounded below, from which it follows easily that $\{f_n^{-1}\}_{n=0}^{\infty}$ is also a Cauchy sequence. This proves that every nonzero element of the completion of F has a multiplicative inverse, as desired. $\square$

PROPOSITION 1.1.8. *Two multiplicative norms $|\cdot|, |\cdot|'$ on a field F are topologically equivalent if and only if there exists $c > 0$ such that $|x|' = |x|^c$ for all $x \in F$.*

PROOF. See [**DGS94**, Lemma I.1.2]. $\square$

2. Valuations and nonarchimedean absolute values

DEFINITION 1.2.1. A *real semivaluation* on an abelian group G is a function $v : G \to \mathbb{R} \cup \{+\infty\}$ with the following properties.

(a) We have $v(0) = +\infty$.

(b) For $f, g \in G$, $v(f + g) \geq \min\{v(f), v(g)\}$.

We say v is a *real valuation* if the following additional condition holds.

(a') For $g \in G$, $v(g) = +\infty$ if and only if $g = 0$.

If v is a real (semi)valuation on G, then the function $|\cdot| = e^{-v(\cdot)}$ is a seminorm on G which is *nonarchimedean*, i.e., it satisfies the strong triangle inequality

(b') For $f, g \in G$, $|f + g| \leq \max\{|f|, |g|\}$.

Conversely, for any nonarchimedean (semi)norm $|\cdot|$, $v(\cdot) = -\log|\cdot|$ is a real valuation. We will apply various definitions made for seminorms to semivaluations in this manner; for instance, if R is a ring and v is a real (semi)valuation on its additive group, we say that v is (sub)multiplicative if the corresponding nonarchimedean (semi)norm is.

Definition 1.2.2. We say a group is *nonarchimedean* if it is equipped with a nonarchimedean norm; we say a ring or field is *nonarchimedean* if it is equipped with a multiplicative nonarchimedean norm. The adjective *ultrametric* is also used, referring to a metric satisfying the strong triangle inequality.

Definition 1.2.3. Let F be a nonarchimedean field. The *multiplicative value group* of a nonarchimedean field F is the image of $F^\times$ under $|\cdot|$, viewed as a subgroup of $\mathbb{R}^+$; we will often denote it simply as $|F^\times|$. The *additive value group* of F is the set of negative logarithms of the multiplicative value group. If these groups are discrete, we say F is *discretely valued.* Define also

$$\begin{aligned} \mathfrak{o}_F &= \{f \in F : v(f) \geq 0\} \\ \mathfrak{m}_F &= \{f \in F : v(f) > 0\} \\ \kappa_F &= \mathfrak{o}_F/\mathfrak{m}_F. \end{aligned}$$

Note that $\mathfrak{o}_F$ is a local ring (the *valuation ring* of F), $\mathfrak{m}_F$ is the maximal ideal of $\mathfrak{o}_F$, and κ_F is a field (the *residue field* of F).

It is worth noting that there are comparatively few archimedean (not nonarchimedean) absolute values on fields.

Theorem 1.2.4 (Ostrowski). *Let F be a field equipped with an absolute value $|\cdot|$. Then $|\cdot|$ fails to be nonarchimedean if and only if the sequence $|1|, |2|, |3|, \dots$ is unbounded. In that case, F is isomorphic to a subfield of $\mathbb{C}$ with the induced absolute value.*

Proof. Exercise, or see [**Rob00**, §2.1.6] and [**Rob00**, §2.2.4], respectively. □

3. Norms on modules

Definition 1.3.1. Let R be a commutative ring equipped with a multiplicative seminorm $|\cdot|$, and let M be an R-module equipped with a seminorm $|\cdot|_M$. We say that $|\cdot|_M$ is *compatible with* $|\cdot|$ (or *with* R) if the following conditions hold.

(a) For $f \in R$, $x \in M$, $|fx|_M = |f||x|_M$.
(b) If $|\cdot|$ is nonarchimedean, then so is $|\cdot|_M$.

Note that (b) is not superfluous; see exercises. Note also that if R is a nonarchimedean field, then two norms $|\cdot|_M, |\cdot|'_M$ are topologically equivalent if and only if there exist $c_1, c_2 > 0$ such that

$$|x|_M \leq c_1|x|'_M, \qquad |x|'_M \leq c_2|x|_M \qquad (x \in M).$$

Definition 1.3.2. Let R be a ring equipped with a multiplicative seminorm $|\cdot|$, and let M be a free R-module. For B a basis of M, define the *supremum norm* of M with respect to B by setting

$$\left|\sum_{b\in B} c_b b\right| = \sup_{b\in B}\{|c_b|\} \qquad (c_b \in R).$$

Theorem 1.3.3. *Let F be a field complete for a norm $|\cdot|$, and let V be a finite dimensional vector space over F. Then any two norms on V compatible with F are equivalent.*

Proof. In the nonarchimedean case, see [**DGS94**, Theorem I.3.2]. Otherwise, apply Theorem 1.2.4 to deduce that $F = \mathbb{R}$ or $F = \mathbb{C}$, then use compactness of the unit ball. $\square$

Definition 1.3.4. For F a nonarchimedean field, a *Banach space* over F is a vector space over F equipped with a norm compatible with F, under which it is complete. For an introduction to nonarchimedean Banach spaces, see [**Sch02**].

4. Examples of nonarchimedean absolute values

Example 1.4.1. For any field F, there is a *trivial absolute value* of F defined by

$$|f|_{\mathrm{triv}} = \begin{cases} 1 & f \neq 0 \\ 0 & f = 0. \end{cases}$$

This absolute value is nonarchimedean, and F is complete under it. The trivial case will always be allowed unless explicitly excluded; it is often a useful input into a highly nontrivial construction, as in the next few examples.

Example 1.4.2. Let F be any field, and let $F((t))$ denote the field of formal Laurent series. The *t-adic valuation* v_t on F is defined as follows: for $f = \sum_i c_i t^i \in F((t))$, $v_t(f)$ is the least i for which $c_i \neq 0$. This exponentiates to give a *t-adic absolute value*, under which $F((t))$ is complete and discretely valued.

Example 1.4.3. For F a nonarchimedean field and $\rho > 0$, the *ρ-Gauss norm* (or the *(t,ρ)-Gauss norm*, in case we need to specify t explicitly) on the rational function field $F(t)$ is defined as follows: for $f = P/Q$ with $P, Q \in F[t]$, write $P = \sum_i P_i t^i$ and $Q = \sum_j Q_j t^j$, and put

$$|f|_\rho = \max_i\{|P_i|\rho^i\}/\max_j\{|Q_j|\rho^j\}.$$

Note that $F(t)$ is discretely valued under the ρ-Gauss norm if and only if either:

(a) F carries the trivial absolute value, in which case the norm is equivalent to the t-adic absolute value no matter what ρ is; or
(b) F carries a nontrivial absolute value, and ρ belongs to the divisible closure of the value group of F.

So far we have not mentioned the principal examples from number theory; let us do so now.

Example 1.4.4. For p a prime number, the *p-adic absolute value* $|\cdot|_p$ on $\mathbb{Q}$ is defined as follows: given $f = r/s$ with $r, s \in \mathbb{Z}$, write $r = p^a m$ and $s = p^b n$ with m, n not divisible by p, then put

$$|f|_p = p^{-a+b}.$$

In particular, we have normalized so that $|p| = p^{-1}$; this convention is usually taken so as to make the *product formula* hold. Namely, for any $f \in \mathbb{Q}$, if $|\cdot|_\infty$ denotes the usual archimedean absolute value, then

$$|f|_\infty \prod_p |f|_p = 1.$$

Completing $\mathbb{Q}$ under $|\cdot|_p$ gives the *field of p-adic numbers* $\mathbb{Q}_p$; it is discretely valued. Its valuation subring is denoted $\mathbb{Z}_p$ and called the *ring of p-adic integers.*

Theorem 1.4.5 (Ostrowski). *Any nontrivial nonarchimedean absolute value on $\mathbb{Q}$ is equivalent to the p-adic absolute value for some prime p.*

Proof. See [**Rob00**, §2.2.4]. □

To equip extensions of $\mathbb{Q}_p$ with absolute values, we use the following result.

Theorem 1.4.6. *Let F be a complete nonarchimedean field. Then any finite extension E of F admits a unique extension of $|\cdot|$ to an absolute value on E.*

Proof. We only prove uniqueness now; existence will be established in Section 3. Let $|\cdot|_1$ and $|\cdot|_2$ be two extensions of $|\cdot|$ to absolute values on E. Then these in particular give norms on E viewed as an F-vector space; by Theorem 1.3.3, these norms are equivalent. That is, there exist $c_1, c_2 > 0$ such that

$$|x|_1 \leq c_1 |x|_2, \qquad |x|_2 \leq c_2 |x|_1 \qquad (x \in E).$$

We now use the extra information that $|\cdot|_1$ and $|\cdot|_2$ are multiplicative, because they really are norms on E as a field in its own right. That is, for any positive integer n, we may substitute x^n in place of x in the previous inequalities, then take n-th roots, to obtain

$$|x|_1 \leq c_1^{1/n} |x|_2, \qquad |x|_2 \leq c_2^{1/n} |x|_1 \qquad (x \in E).$$

Letting $n \to \infty$ gives $|x|_1 = |x|_2$, as desired. □

Remark 1.4.7. Don't forget that the completeness of F is crucial. For instance, the 5-adic absolute value on $\mathbb{Q}$ extends in two different ways to the Gaussian rational numbers $\mathbb{Q}(i)$, depending on whether you want $|2+i| = 5^{-1}, |2-i| = 1$ or vice versa.

Because of the uniqueness in Theorem 1.4.6, it also follows that any algebraic extension E of F, finite or not, inherits a unique extension of $|\cdot|$. However, if $[E : F] = \infty$, then E is not complete, so we may prefer to use its completion instead. For instance, if $F = \mathbb{Q}_p$, we define $\mathbb{C}_p$ to be the completion of an algebraic closure of $\mathbb{Q}_p$. You might worry that this may launch us into an endless cycle of completion and algebraic closure, but fortunately this does not occur.

Theorem 1.4.8. *Let F be an algebraically closed nonarchimedean field. Then the completion of F is also algebraically closed.*

For the proof, see section 3.

5. Unramified extensions

We will again need one forward reference into the next chapter.

LEMMA 1.5.1. *Let F be a complete nonarchimedean field, and let E be a finite extension of F. Then $[E : F] \geq [\kappa_E : \kappa_F]$.*

PROOF. Pick any basis of κ_E over κ_F, and lift it to $\mathfrak{o}_E$. The result is a linearly independent set over F. $\square$

DEFINITION 1.5.2. Let F be a complete nonarchimedean field. A finite extension E of F is *unramified* if κ_E is separable over κ_F and $[E : F] = [\kappa_E : \kappa_F]$.

LEMMA 1.5.3. *Let F be a complete nonarchimedean field, and let U be a finite extension of F. Then for any subextension E of U over F, U is unramified over F if and only if U is unramified over E and E is unramified over F.*

PROOF. Since κ_E sits between κ_U and κ_F, having κ_U separable over κ_F is equivalent to having both κ_U separable over κ_E and κ_E separable over κ_F. By Lemma 1.5.1,

$$[U : F] \geq [\kappa_U : \kappa_F], \qquad [U : E] \geq [\kappa_U : \kappa_E], \qquad [E : F] \geq [\kappa_E : \kappa_F];$$

since $[U : F] = [U : E][E : F]$ and $[\kappa_U : \kappa_F] = [\kappa_U : \kappa_E][\kappa_E : \kappa_F]$, the first of the three inequalities is an equality if and only if the other two are. $\square$

PROPOSITION 1.5.4. *Let F be a complete nonarchimedean field, and let E be a finite extension of E. Then for any separable subextension λ of κ_F over κ_E, there exists a unique unramified extension U of F contained in E with $\kappa_U \cong \lambda$; moreover, U is separable over F.*

PROOF. By the primitive element theorem, one can always write $\lambda \cong \kappa_F[x]/(\overline{P}(x))$ for some monic irreducible separable polynomial $\overline{P}[x] \in \kappa_F[x]$. Choose $t \in \mathfrak{o}_E$ whose image in κ_E corresponds to x in $\kappa_F[x]/(\overline{P}(x))$; then the reduction of $P(x + t)$ into $\kappa_E[x]$ is divisible by x but not by x^2. We may thus apply the slope factorization theorem from the next chapter (Theorem 2.2.1) to deduce that $P(x+t)$ has a root in $\mathfrak{o}_E$. This proves existence and separability of U over F.

To prove uniqueness, let U' be another such extension. Then the previous argument applied to U' in place of E shows that $\mathfrak{o}_{U'}$ contains a root of $P(x+t)$ congruent to 0 modulo $\mathfrak{o}_E$. However, there can only be one such root in E because $\overline{P}$ is a separable polynomial, so in fact $U \subseteq U'$. Again by comparing degrees, we have $U = U'$. $\square$

COROLLARY 1.5.5. *For each finite separable extension λ of κ_F, there exists a unique unramified extension E of F with $\kappa_E \cong \lambda$.*

PROOF. Choose $P(x)$ as in the proof of Proposition 1.5.4. Then $E = F[x]/(P(x))$ is an unramified extension of F with residue field λ. The proof of Proposition 1.5.4 shows that any other unramified extension with residue field λ must contain E; by comparing degrees, we see that this contain must be an equality. $\square$

LEMMA 1.5.6. *Let F be a complete nonarchimedean field, and let E be a finite extension of F. Let U_1, U_2 be unramified subextensions of E over F. Then the compositum $U = U_1U_2$ is also unramified over F, and $\kappa_U = \kappa_{U_1}\kappa_{U_2}$ inside κ_E.*

Proof. Put $U_3 = U_1 \cap U_2$ inside E; by Lemma 1.5.3, U_3 is unramified over F. By Proposition 1.5.4, inside κ_E, $\kappa_{U_1} \cap \kappa_{U_2} = \kappa_{U_3}$. Consequently,

$$\begin{aligned}
[\kappa_U : \kappa_{U_2}] &\leq [U : U_2] \qquad \text{(by Lemma 1.5.1)} \\
&= [U_1 : U_3] \\
&= [\kappa_{U_1} : \kappa_{U_3}] \qquad \text{(because } U_1 \text{ is unramified over } U_3\text{)} \\
&= [\kappa_{U_1}\kappa_{U_2} : \kappa_{U_2}] \\
&\leq [\kappa_U : \kappa_{U_2}] \qquad \text{(because } \kappa_{U_1}\kappa_{U_2} \subseteq \kappa_U\text{)}.
\end{aligned}$$

We deduce first that $\kappa_U = \kappa_{U_1}\kappa_{U_2}$, and second that $[U : U_2] = [\kappa_U : \kappa_{U_2}]$. Hence U is unramified over U_2. Since U_2 is unramified over F by Lemma 1.5.3, U is unramified over F by Lemma 1.5.3 again. □

Definition 1.5.7. Let F be a complete nonarchimedean field, and let E be a finite extension of E. By Lemma 1.5.6, there is a maximal unramified subextension U of E over F; by Proposition 1.5.4, κ_U is the maximal separable subextension of κ_E over κ_F. (We will also say that $\mathfrak{o}_U$ is the "maximal unramified subextension" of $\mathfrak{o}_E$ over $\mathfrak{o}_F$.) We say E is *totally ramified* over F if $U = F$.

Lemma 1.5.8. *Let F be a complete nonarchimedean field, let E be a finite extension of F, and let E' be a subextension of E over F. Let U, U' be the maximal unramified subextensions of E, E' over F. Then $U' = U \cap E'$, and UE' is the maximal unramified subextension of E over E'.*

Proof. We have $U' = U \cap E'$ by Proposition 1.5.4. In particular, κ_U contains $\kappa_{U'}$ and is the maximal separable subextension of κ_E over $\kappa_{E'}$. Consequently, κ_U is contained in $\kappa_U\kappa_{E'}$, so there cannot be a proper unramified extension of UE' contained in E (since there is no room in the residue field extension). That is, UE' is the maximal unramified subextension of E over E'. □

Remark 1.5.9. All of the above carries over to the case where F is not complete, but is *henselian.* The henselian condition (in one of its many equivalent formulations; see [**Nag62**, 43.2]) is that for any monic polynomial $P(x) \in \mathfrak{o}_F[x]$ and any simple root $\overline{r} \in \kappa_F$ of $\overline{P} \in \kappa_F[x]$, there exists a unique root $r \in \mathfrak{o}_F$ of P lifting $\overline{r}$.

6. Tamely ramified extensions

Definition 1.6.1. Let F be a complete nonarchimedean field, let E be a finite extension of F, and let U be the maximal unramified subextension of E over F. We say E is *tamely ramified* over F if the degree $[E : U]$ (the *tame degree*) is not divisible by the residue characteristic of κ_F (this is automatic if κ_F is of characteristic 0), and *wildly ramified* otherwise.

Lemma 1.6.2. *Let F be a complete nonarchimedean field, let E be a finite extension of F, and let E' be a subextension of E over F. Then E is tamely ramified over F if and only if E is tamely ramified over E' and E' is tamely ramified over F.*

Proof. We may assume the characteristic of κ_F is $p > 0$, otherwise there is nothing to check. Let U, U' be the maximal unramified subextensions of E, E' over F. By Lemma 1.5.8, $U' = U \cap E'$, and UE' is the maximal unramified subextension of E over E'. We now write

$$[E : U] = [E : E'U][E'U : U] = [E : UE'][E' : U'],$$

and deduce that $[E : U]$ is coprime to p if and only if $[E : UE']$ and $[E' : U']$ both are. $\square$

Lemma 1.6.3. *Let F be a complete nonarchimedean field, let E be a finite extension of F, and let T_1, T_2 be tamely ramified subextensions of E over F. Then $T = T_1T_2$ is also tamely ramified over F.*

Proof. We may assume the characteristic of κ_F is $p > 0$. Let U, U_1, U_2, U_3 be the maximal unramified subextensions of T, T_1, T_2, T_3 over F. By Lemma 1.6.2, T_1 is tamely ramified over $T_3 = T_1 \cap T_2$. By Lemma 1.5.8, T_3U_2 is the maximal unramified subextension of T_2 over T_3, so $[T_2 : T_3U_2]$ is not divisible by p. Since $[T : U_1U_2]$ divides $[T_2 : T_3U_2]$; it is also not divisible by p. Since $U_1U_2 \subseteq U$, $[T : U]$ divides $[T : U_1U_2]$ and so is also not divisible by p. Hence T is tamely ramified over F. $\square$

Definition 1.6.4. Let F be a complete nonarchimedean field, and let E be a finite extension of E. By Lemma 1.6.3, there is a maximal tamely ramified subextension T of E over F. We say E is *totally wildly ramified* over F if $T = F$.

Remark 1.6.5. In the case of a discrete valuation, we will resume the discussion of ramification theory in Section 3. In the general case, the theory is somewhat subtler; see for instance [**Rib99**] (especially Chapter 6).

7. Spherical completeness

For nonarchimedean fields, there is an important distinction between two different notions of completeness, which does not appear in the archimedean case.

Definition 1.7.1. A metric space is *complete* if any decreasing sequence of closed balls with radii tending to 0 has nonempty intersection. (For an abelian group equipped with a norm, this reproduces our earlier definition.) A metric space is *spherically complete* if any decreasing sequence of closed balls, regardless of radii, has nonempty intersection. (For a topological vector space, the term *linearly compact* is also used.)

Example 1.7.2. The fields $\mathbb{R}$ and $\mathbb{C}$ with their usual absolute value are spherically complete. Also, any complete nonarchimedean field which is discretely valued, e.g., $\mathbb{Q}_p$ or $\mathbb{C}((t))$, is spherically complete. However, any infinite algebraic extension of $\mathbb{Q}_p$ is not spherically complete.

Theorem 1.7.3 (Kaplansky-Krull). *Any nonarchimedean field embeds isometrically into a spherically complete nonarchimedean field. (However, the construction is not functorial.)*

Proof. Since completion is functorial, we may assume we are starting with a complete nonarchimedean field. It was originally shown by Krull [**Kru32**, Theorem 24] that any complete nonarchimedean field admits an extension which is *maximally complete*, in the sense of not admitting any extensions preserving both the value group and the residue field. (In fact, this is not difficult to prove using Zorn's lemma.) The equivalence of this condition with spherical completeness was then proved by Kaplansky [**Kap42**, Theorem 4].

One can also prove the result more directly; for instance, the case of $\mathbb{Q}_p$ is explained in detail in [**Rob00**, §3]. $\square$

Notes

The condition of spherical completeness is quite important in nonarchimedean functional analysis, as it is needed for the Hahn-Banach theorem to hold. (By contrast, the nonarchimedean version of the open mapping theorem requires only completeness of the field.) For expansion of this remark, we recommend [**Sch02**]; an older reference is [**vR78**].

Although the construction of the spherical completion of a nonarchimedean field is not functorial, it is possible to make a canonical construction using generalized power series (Mal'cev-Neumann series); this was described by Poonen [**Poo93**].

For a direct proof of Theorem 1.4.8 in the case of the completed algebraic closure of $\mathbb{Q}_p$, see [**Rob00**, §3.3.3].

Exercises

(1) Prove Ostrowski's theorem (Theorem 1.2.4).
(2) Exhibit an example to show that even for a finite-dimensional vector space V over a complete nonarchimedean field F, the requirement that a norm on $|\cdot|_V$ must satisfy the strong triangle inequality is not superfluous. (That is, a function $|\cdot|_V : V \to [0,\infty)$ can satisfy the ordinary triangle inequality plus conditions (a) and (c) without satisfying the strong triangle inequality.)
(3) Prove that the valuation ring $\mathfrak{o}_F$ of a nonarchimedean field is noetherian if and only if F is discretely valued.
(4) Use Theorem 1.4.6 to prove that for any field F, any nonarchimedean absolute value $|\cdot|$ on F, and any extension of E, there exists *at least one* extension of $|\cdot|$ to an absolute value on E. (Hint: reduce to the cases where E is a finite extension, and where E is a purely transcendental extension.)
(5) Here is a more exotic variation of the t-adic valuation. Choose $\alpha_1, \dots, \alpha_n \in \mathbb{R}$.
    (a) Prove that on the rational function field $F(t_1, \dots, t_n)$, there is a valuation v_α such that $v(f) = 0$ for all $f \in F^*$ and $v(t_i) = \alpha_i$ for $i = 1, \dots, n$.
    (b) Prove that if $\alpha_1, \dots, \alpha_n$ are linearly independent over $\mathbb{Q}$, the valuation v_α is uniquely determined by (a).
    (c) Prove that if $\alpha_1, \dots, \alpha_n$ are not linearly independent over $\mathbb{Q}$, the valuation v_α is not uniquely determined by (a).

CHAPTER 2

Newton polygons

In this chapter, we recall the traditional theory of Newton polygons for polynomials over a nonarchimedean field. In the process, we introduce a general framework which will allow us to consider Newton polygons in a wider range of circumstances.

1. Gauss norms and Newton polygons

Definition 2.1.1. Let R be a commutative ring equipped with a nonarchimedean submultiplicative (semi)norm $|\cdot|$. For $\rho > 0$, define the *ρ-Gauss (semi)norm* $|\cdot|_\rho$ on $R[T]$ by

$$\left|\sum_i P_i T^i\right|_\rho = \max_i\{|P_i|\rho^i\};$$

it is clearly submultiplicative. Moreover, it is also multiplicative if $|\cdot|$ is; see Proposition 2.1.3 below. For $r \in \mathbb{R}$, we define the *r-Gauss (semi)valuation* v_r as the (semi)valuation associated to the e^{-r}-Gauss (semi)norm.

Definition 2.1.2. Let R be a commutative ring equipped with a nonarchimedean submultiplicative (semi)norm $|\cdot|$. For $\rho > 0$ and $P = \sum_i P_i T^i \in R[T]$, define the *width* of P under $|\cdot|_\rho$ as the difference between the maximum and minimum values of i achieving $\max_i\{|P_i|\rho^i\}$.

Proposition 2.1.3. *Let R be a commutative ring equipped with a nonarchimedean multiplicative seminorm $|\cdot|$. For $\rho > 0$ and $P, Q \in R[T]$, the following results hold.*

(a) *We have $|PQ|_\rho = |P|_\rho |Q|_\rho$. That is, $|\cdot|_\rho$ is multiplicative.*
(b) *The width of PQ under $|\cdot|_\rho$ equals the sum of the widths of P and Q under $|\cdot|_\rho$.*

Proof. For $* \in \{P, Q\}$, let j_*, k_* be the minimum and maximum values of i achieving $\max_i\{| *_i |\rho^i\}$. Write

$$PQ = \sum_i (PQ)_i T^i = \sum_i \left(\sum_{g+h=i} P_g Q_h\right) T^i.$$

In the sum $(PQ)_i = \sum_{g+h=i} P_g Q_h$, each summand has norm at most $|P|_\rho |Q|_\rho \rho^{-i}$, with equality if and only if $|P_g| = |P|_\rho \rho^{-g}$ and $|Q_h| = |Q|_\rho \rho^{-h}$. This cannot occur for $i < j_P + j_Q$, and for $i = j_P + j_Q$ it can only occur for $g = j_P, h = j_Q$. Hence

$$|(PQ)_i| < |P|_\rho |Q|_\rho \rho^{-i} \qquad (i < j_P + j_Q)$$
$$|(PQ)_i| = |P|_\rho |Q|_\rho \rho^{-i} \qquad (i = j_P + j_Q).$$

Similarly,

$$\begin{aligned} |(PQ)_i| &< |P|_\rho |Q|_\rho \rho^{-i} \qquad (i > k_P + k_Q) \\ |(PQ)_i| &= |P|_\rho |Q|_\rho \rho^{-i} \qquad (i = k_P + k_Q). \end{aligned}$$

This proves both claims. □

DEFINITION 2.1.4. Let R be a commutative ring equipped with a nonarchimedean submultiplicative seminorm. Given a polynomial $P(T) = \sum_{i=0}^n P_i T^i \in R[T]$, draw the set of points

$$\{(-i, v(P_i)) : i = 0, \dots, n, P_i \neq 0\} \subset \mathbb{R}^2,$$

then form the lower convex hull of these points, i.e., take the intersection of every closed halfplane lying above some nonvertical line containing all the points. The boundary of this region is called the *Newton polygon* of P. The *slopes* of P are the slopes of this polygon, viewed as a multiset with the slope r counting with multiplicity equal to the horizontal width of the segment of the Newton polygon of slope r (or 0 if there is no such segment); the latter can also be interpreted as the width of P under $|\cdot|_{e^{-r}}$. (In case this multiset has cardinality less than $\deg(P)$, we include $+\infty$ with sufficient multiplicity to make up the shortfall.)

PROPOSITION 2.1.5. *Let R be a nonarchimedean commutative ring, and suppose $P(T) = (T - \lambda_1) \cdots (T - \lambda_n)$. Then the slope multiset of P consists of $-\log|\lambda_1|, \dots, -\log|\lambda_n|$.*

PROOF. This is immediate from the multiplicativity of $|\cdot|_{e^{-r}}$. □

2. Slope factorizations and a master factorization theorem

THEOREM 2.2.1. *Let F be a complete nonarchimedean field. Suppose $S \in F[T]$, $r \in \mathbb{R}$, and $m \in \mathbb{Z}_{\geq 0}$ satisfy*

$$v_r(S - T^m) > v_r(T^m).$$

Then there exists a unique factorization $S = PQ$ satisfying the following conditions.

(a) *The polynomal $P \in F[T]$ has degree $\deg(S) - m$, and its slopes are all less than r.*
(b) *The polynomial $Q \in F[T]$ is monic of degree m, and its slopes are all greater than r.*
(c) *We have $v_r(P - 1) > 0$ and $v_r(Q - T^m) > v_r(T^m)$.*

Moreover, for this factorization,

$$\min\{v_r(P - 1), v_r(Q - T^m) - v_r(T^m)\} \geq v_r(S - T^m) - v_r(T^m).$$

It is not so difficult to prove this theorem directly. However, we will be stating a number of similar results as we go along, so rather than giving individual proofs each time, we state a master factorization theorem from which we can deduce Theorem 2.2.1 and all of its variants. It, and the proof given here, are due to Christol [**Chr83**, Proposition 1.5.1].

THEOREM 2.2.2 (Christol). *Let R be a nonarchimedean ring (not necessarily commutative). Suppose the nonzero elements $a, b, c \in R$ and the additive subgroups $U, V, W \subseteq R$ satisfy the following conditions.*

(a) *The spaces U, V are complete under the norm, and $UV \subseteq W$.*
(b) *The map $f(u, v) = av + ub$ is a surjection of $U \times V$ onto W.*

(c) *There exists* $\lambda > 0$ *such that*

$$|f(u,v)| \geq \lambda \max\{|a||v|, |b||u|\} \qquad (u \in U, v \in V).$$

(d) *We have* $ab - c \in W$ *and*

$$|ab - c| < \lambda^2 |c|.$$

Then there exists a unique pair $x \in U, y \in V$ *such that*

$$c = (a+x)(b+y), \qquad |x| < \lambda |a|, \qquad |y| < \lambda |b|.$$

For this x, y, *we also have*

$$|x| \leq \lambda^{-1} |ab - c||b|^{-1}, \qquad |y| \leq \lambda^{-1} |ab - c||a|^{-1}.$$

Before proving this, let us see how it implies Theorem 2.2.1.

Proof of Theorem 2.2.1. We apply Theorem 2.2.2 with the following parameters:

$$\begin{aligned}
R &= F[T] \\
|\cdot| &= |\cdot|_{e^{-r}} \\
U &= \{P \in F[T] : \deg(P) \leq \deg(S) - m - 1\} \\
V &= \{P \in F[T] : \deg(P) \leq m - 1\} \\
W &= \{P \in F[T] : \deg(P) \leq \deg(S) - 1\} \\
a &= 1 \\
b &= T^m \\
c &= S \\
\lambda &= 1,
\end{aligned}$$

then put $P = a + x$ and $Q = b + y$. $\square$

With this motivation in mind, we now proof Theorem 2.2.2.

Proof of Theorem 2.2.2. We define a norm on $U \times V$ by setting

$$|(u,v)| = \max\{|a||v|, |b||u|\}.$$

so that (c) implies

$$\lambda |(u,v)| \leq |f(u,v)| \leq |(u,v)|.$$

In particular, $\lambda \leq 1$, so $|ab - c| < |ab| = |c|$.

Since a, b are nonzero, (c) implies that f is injective. By (b), f is in fact a bijective group homomorphism between $U \times V$ and W. It follows that for all $w \in W$,

$$|f^{-1}(w)| \leq \lambda^{-1} |w|.$$

By (d), we may choose $\mu \in (0, \lambda)$ with $|ab - c| \leq \lambda \mu |c|$. Define

$$B_\mu = \{(u,v) \in U \times V : |(u,v)| \leq \mu |c|\}.$$

For $(u,v) \in B_\mu$, we have

$$|a||v| \leq |(u,v)| \leq \mu |c| = \mu |a||b|,$$

so $|v| \leq \mu|b|$. Similarly $|u| \leq \mu|a|$. As a result,

$$\begin{aligned}|f^{-1}(c - ab - uv)| &\leq \lambda^{-1}|c - ab - uv| \\ &\leq \lambda^{-1}\max\{|c - ab|, |uv|\} \\ &\leq \lambda^{-1}\max\{\lambda\mu|c|, \mu^2|a||b|\} \\ &= \mu|c|.\end{aligned}$$

Consequently, the map $g(u,v) = f^{-1}(c - ab - uv)$ carries B_μ into itself.

We next show that g is contractive. For $(u,v),(t,s) \in B_\mu$,

$$\begin{aligned}|g(u,v) - g(t,s)| &\leq |f^{-1}(ts - uv)| \\ &\leq \lambda^{-1}|ts - uv| \\ &\leq \lambda^{-1}|t(s-v) + (t-u)v| \\ &\leq \lambda^{-1}\max\{\mu|a||s-v|, \mu|t-u||b|\} \\ &\leq \lambda^{-1}\mu|(u-t, v-s)| \\ &= \lambda^{-1}\mu|(u-t) - (v-s)|,\end{aligned}$$

which has the desired effect because $\lambda^{-1}\mu < 1$.

Since g is contractive on B_μ, and $U \times V$ is complete, there is a unique $(x,y) \in U \times V$ fixed by g. That is,

$$ay + xb = f(x,y) = f(g(x,y)) = c - ab - xy$$

and so

$$c = (a+x)(b+y).$$

Moreover, there is a unique such (x,y) in the union of all of the B_μ, and that element belongs to the intersection of all of the B_μ. □

Remark 2.2.3. One can also use Theorem 2.2.2 to recover other instances of Hensel's lemma. For instance, if F is a complete nonarchimedean field, $P(x) \in \mathfrak{o}_F[x]$, and the reduction of $P(x)$ into $\kappa_F[x]$ factors as $\overline{QR}$ with $\overline{Q}, \overline{R}$ coprime, then there exists a unique factorization $P = QR$ in $\mathfrak{o}_F[x]$ with Q, R lifting $\overline{Q}, \overline{R}$.

3. Applications to nonarchimedean field theory

We now go back and apply Theorem 2.2.1 to prove some facts about extensions of nonarchimedean fields which were omitted in the previous chapter.

We first complete the proof of Theorem 1.4.6. For this, we need the following lemma.

Lemma 2.3.1. *Let F be a complete nonarchimedean field. Let $P(T) \in F[T]$ be a polynomial whose slopes are all equal to r. Let $S(T) \in F[T]$ be any polynomial, and write $S = PQ + R$ with* $\deg(R) < \deg(P)$. *Then*

$$v_r(S) = \min\{v_r(P) + v_r(Q), v_r(R)\}.$$

Proof. Exercise. □

Proof of Theorem 1.4.6 (continued). It remains to show that if F is a complete nonarchimedean field, then any finite extension E of F admits an extension of $|\cdot|$ to an absolute value on E. If E' is a field intermediate between F and E, we may first extend the absolute value to E' and then to E. Consequently, it suffices to check the case where $E = F(\alpha)$ for some $\alpha \in E$, that is, $E \cong F[T]/(P(T))$ for some monic irreducible polynomial $P \in F[T]$ (the minimal polynomial of α). Apply Theorem 2.2.1; since $P(T)$ cannot factor nontrivially, we deduce that P must have a single slope r.

We now define an absolute value on E as follows: for $\beta = c_0 + c_1\alpha + \cdots + c_{n-1}\alpha^{n-1}$, with $n = \deg(P) = [E : F]$, put

$$|\beta|_E = \max_i \{|c_i| e^{-ri}\}.$$

That is, take $|\beta|_E$ to be the e^{-r}-Gauss norm of the polynomial $c_0 + c_1 T + \cdots + c_{n-1}T^{n-1}$. The multiplicativity of $|\cdot|_E$ is then a consequence of Lemma 2.3.1. $\square$

We next give the proof of Theorem 1.4.8. For this, we need a crude version of the principle that "the roots of a polynomial over a complete algebraically closed nonarchimedean field vary continuously in the coefficients."

Lemma 2.3.2. *Let F be an algebraically closed nonarchimedean field with completion E, and suppose $P \in E[T]$ is monic of degree d. Then for any $\epsilon > 0$, we can find $z \in F$ such that $|z| \leq |P(0)|^{1/d}$ and $|P(z)| < \epsilon$.*

Proof. If $P(0) = 0$ we may pick $z = 0$, so assume $P(0) \neq 0$. Put $P = T^d + \sum_{i=0}^{d-1} P_i T^i$. For any $\delta > 0$, we can pick a polynomial $Q = T^d + \sum_{i=0}^{d-1} Q_i d^i \in F[T]$ with $|Q_i - P_i| < \delta$ for $i = 0, \dots, d-1$.

Now assume $\delta < \min\{|P_0|, \epsilon, \epsilon/|P_0|\}$, so that $|Q_0| = |P_0|$. By Proposition 2.1.5, we can find a root $z \in F$ of Q_0 with $|z| \leq |Q_0|^{1/d} = |P_0|^{1/d}$. We now have

$$|P(z)| = |(P - Q)(z)| \leq \delta \max\{1, |z|\}^d \leq \delta \max\{1, |P(0)|\} < \epsilon,$$

as desired. $\square$

Proof of Theorem 1.4.8. We must check that the completion E of an algebraically closed nonarchimedean field F is itself algebraically closed. Let $P(T) \in E[T]$ be a monic polynomial of degree d. Define a sequence of polynomials $P_0, P_1, \dots$ as follows. Put $P_0 = P$. Given P_i, apply Lemma 2.3.2 to construct z_i with $|z_i| \leq |P_i(0)|^{1/d}$ and $|P_i(z_i)| < 2^{-i}$, then set $P_{i+1}(T) = P_i(T + z_i)$ so that $P_{i+1}(0) = P_i(z_i)$. If some P_i satisfies $P_i(0) = 0$, then $z_0 + \cdots + z_{i-1}$ is a root of P. Otherwise, we get an infinite sequence $z_0, z_1, \dots$ such that $z_0 + z_1 + \cdots$ converges to a root of P. $\square$

Exercises

(1) Prove Lemma 2.3.1.
(2) State and prove a precise version of the assertion that "the roots of a polynomial over a complete algebraically closed nonarchimedean field vary continuously in the coefficients."

CHAPTER 3

Matrix analysis

In this chapter, we study metric properties of matrices, and matrix invariants, over a field equipped with an absolute value. Although the archimedean and nonarchimedean settings must be handled differently, they exhibit strong similarities, so we present them in parallel, starting with the archimedean case.

NOTATION 3.0.1. Let $\mathrm{Diag}(\sigma_1, \dots, \sigma_n)$ denote the $n \times n$ diagonal matrix D with $D_{ii} = \sigma_i$ for $i = 1, \dots, n$.

1. Singular values and eigenvalues (archimedean case)

HYPOTHESIS 3.1.1. In this section and the next, let A be an $n \times n$ matrix over $\mathbb{C}$.

We are interested in two sets of numerical invariants of A. One of these is the familiar set of eigenvalues.

DEFINITION 3.1.2. Let $\lambda_1, \dots, \lambda_n$ be the list of eigenvalues of A, which we sort so that $|\lambda_1| \geq \cdots \geq |\lambda_n|$.

A second set of numerical invariants of A, which is in many ways better behaved from the point of view of numerical analysis, is the singular values.

DEFINITION 3.1.3. Let A^* denote the conjugate transpose (or *Hermitian transpose*) of A. The matrix A^*A is Hermitian and nonnegative definite, so has nonnegative real eigenvalues. The square roots of these eigenvalues comprise the *singular values* of A; we denote them $\sigma_1, \dots, \sigma_n$ with $\sigma_1 \geq \cdots \geq \sigma_n$. These are not invariant under conjugation, but they are invariant under multiplying A on either side by a unitary matrix.

THEOREM 3.1.4 (Singular value decomposition). *There exist unitary $n \times n$ matrices U, V such that $UAV = \mathrm{Diag}(\sigma_1, \dots, \sigma_n)$.*

PROOF. This is equivalent to showing that there is an orthonormal basis of $\mathbb{C}^n$ which remains orthogonal upon applying A. To construct it, start with a vector $v \in \mathbb{C}^n$ maximizing $|Av|/|v|$, then show that for any $w \in \mathbb{C}^n$ orthogonal to v, Aw is also orthogonal to Av. For further details, see references in the notes. $\square$

COROLLARY 3.1.5. *The singular values of A^{-1} are $\sigma_n^{-1}, \dots, \sigma_1^{-1}$.*

From the singular value decomposition, we may infer a convenient interpretation of σ_i.

COROLLARY 3.1.6. *The number σ_i is the smallest value of λ for which the following holds: for any i-dimensional subspace V of $\mathbb{C}^n$, there exists $v \in V$ nonzero such that $|Av| \leq \lambda |v|$.*

Proof. Theorem 3.1.4 provides an orthonormal basis $v_1, \dots, v_n$ of V such that $Av_1, \dots, Av_n$ is again orthogonal, and $|Av_i| = \sigma_i |v_i|$ for $i = 1, \dots, n$. Let W be the span of $v_i, \dots, v_n$; then for any i-dimensional subspace V of $\mathbb{C}^n$, $V \cap W$ is nonempty, and any $v \in V \cap W$ satisfies $|Av| \leq \sigma_i |v|$. On the other hand, if we take V to be the span of $v_1, \dots, v_i$, then we have $|Av| \geq \sigma_i |v|$ for all $v \in V$. This proves the claim. □

The relationship between the singular values and the eigenvalues is controlled by the following inequality of Weyl [Wey49]. For a vast generalization, see Theorem 3.5.1.

Theorem 3.1.7 (Weyl). *We have*

$$\sigma_1 \cdots \sigma_i \geq |\lambda_1 \cdots \lambda_i| \qquad (i = 1, \dots, n),$$

with equality for $i = n$.

Proof. The equality for $i = n$ holds because $\det(A^*A) = |\det(A)|^2$. We check the inequality first for $i = 1$. Note that if we equip $\mathbb{C}^n$ with the L^2 norm, i.e.,

$$|(z_1, \dots, z_n)| = (|z_1|^2 + \cdots + |z_n|^2)^{1/2},$$

then σ_1 is the operator norm of A, that is,

$$\sigma_1 = \sup_{v \in \mathbb{C}^n - \{0\}} \{|Av|/|v|\}.$$

Since there exists $v \in \mathbb{C}^n - \{0\}$ with $Av = \lambda_1 v$, we deduce that $\sigma_1 \geq |\lambda_1|$.

For the general case, we pass from $\mathbb{C}^n$ to its i-th exterior power $\wedge^i \mathbb{C}^n$, on which A also acts. The maximum norm of an eigenvalue of this action is $|\lambda_1 \cdots \lambda_i|$, and the operator norm is $\sigma_1 \cdots \sigma_i$. Thus the previous inequality gives what we want. □

We mention in passing the following converse of Theorem 3.1.7, due to Horn [Hor54, Theorem 4].

Theorem 3.1.8. *For $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ and $\sigma_1, \dots, \sigma_n \in \mathbb{R}_{\geq 0}$ satisfying*

$$\sigma_1 \cdots \sigma_i \geq |\lambda_1 \cdots \lambda_i| \qquad (i = 1, \dots, n),$$

with equality for $i = n$, there exist an $n \times n$ matrix A over $\mathbb{C}$ with singular values $\sigma_1, \dots, \sigma_n$ and eigenvalues $\lambda_1, \dots, \lambda_n$.

Equality in Weyl's theorem at an intermediate stage has a structural meaning.

Theorem 3.1.9. *Suppose that for some $i \in \{1, \dots, n-1\}$ we have*

$$\sigma_i > \sigma_{i+1}, \qquad |\lambda_i| > |\lambda_{i+1}|,$$
$$\sigma_1 \cdots \sigma_i = |\lambda_1 \cdots \lambda_i|.$$

Then there exists a unitary matrix U such that $U^{-1}AU$ is block diagonal, with the first block accounting for the first i singular values and eigenvalues, and the second block accounting for the others.

Proof. Let $v_1, \dots, v_n$ be a basis of $\mathbb{C}^n$ such that $v_1, \dots, v_i$ span the generalized eigenspaces with eigenvalues $\lambda_1, \dots, \lambda_i$. and $v_{i+1}, \dots, v_n$ span the generalized eigenspaces with eigenvalues $\lambda_{i+1}, \dots, \lambda_n$. Apply the singular value decomposition to construct an orthonormal basis $w_1, \dots, w_n$ such that $Aw_1, \dots, Aw_n$ are also orthogonal and $|Aw_i| = \sigma_i |w_i|$.

Since $\sigma_i > \sigma_{i+1}$, the only vectors $v \in \wedge^i \mathbb{C}^n$ for which $|Av|/|v|$ achieves its maximum value $\sigma_1 \cdots \sigma_i$ are the nonzero multiples of $w_1 \wedge \cdots \wedge w_i$. However, this is also true for $v_1 \wedge \cdots \wedge v_i$. We conclude that $w_1, \dots, w_i$ span V; this implies that the orthogonal complement of V is spanned by $w_{i+1}, \dots, w_n$, and so is also preserved by A. This yields the desired result. □

THEOREM 3.1.10. *The following are equivalent.*

(a) *There exists a unitary matrix U such that $U^{-1}AU$ is diagonal.*
(b) *The matrix A is normal, i.e., $A^*A = AA^*$.*
(c) *The eigenvalues $\lambda_1, \dots, \lambda_n$ and singular values $\sigma_1, \dots, \sigma_n$ of A satisfy $|\lambda_i| = \sigma_i$ for $i = 1, \dots, n$.*

PROOF. It is clear that (a) implies both (b) and (c). Given (b), we can perform a joint eigenspace decomposition for A and A^*. On any common generalized eigenspace, A has some eigenvalue λ, A^* has eigenvalue $\overline{\lambda}$, and so A^*A has eigenvalue $|\lambda|^2$. This implies (c).

Given (c), Theorem 3.1.9 implies that A can be conjugated by a unitary matrix into a block diagonal matrix in which each block has a single eigenvalue and a single singular value, which coincide. Let B be such a block, with eigenvalue λ, corresponding to a subspace V of $\mathbb{C}^n$. If the common singular value is 0, then $B = 0$. Otherwise, $\lambda \neq 0$ and $\lambda^{-1}B$ is unitary. Hence given orthogonal eigenvectors $v_1, \dots, v_i \in V$ of B, the orthogonal complement in V of their span is preserved by B, so is either zero or contains another eigenvector v_{i+1}. This shows that B is diagonalizable, and thus is itself a scalar matrix. (One can also argue this last step using compactness of the unitary group.) □

In general, we can conjugate any matrix into an almost normal matrix; the "almost" only intervenes when the matrix is not semisimple.

LEMMA 3.1.11. *For any $\eta > 1$, we can choose $U \in \mathrm{GL}_n(\mathbb{C})$ such that for $i = 1, \dots, n$, the i-th singular value of $U^{-1}AU$ is at most $\eta|\lambda_i|$. If A is semisimple (i.e., diagonalizable), we can also take $\eta = 1$.*

PROOF. Put A in Jordan normal form, then rescale so that for each eigenvalue λ, the superdiagonal terms have absolute value at most $(|\eta|-1)|\lambda|$, and all other terms are zero. □

2. Perturbations (archimedean case)

Another inequality of Weyl [Wey12] shows that the singular values do not change much under a small (additive) perturbation.

THEOREM 3.2.1 (Weyl). *Let B be an $n \times n$ matrix, and let $\sigma'_1, \dots, \sigma'_n$ be the singular values of $A + B$. Then*

$$|\sigma'_i - \sigma_i| \leq |B| \qquad (i = 1, \dots, n).$$

It is more complicated to describe what happens to the eigenvalues under a small additive perturbation, but it is not so difficult to quantify what happens to the characteristic polynomial, at least in a crude fashion.

THEOREM 3.2.2. *Let B be an $n \times n$ matrix such that $|B| < \sigma_j$ for some $j \in \{1, \dots, n\}$. Let $P(T) = T^n + \sum_{i=0}^{n-1} P_i T^i$ and $Q(T) = T^n + \sum_{i=0}^{n-1} Q_i T^i$ be the characteristic polynomials* *of A and $A+B$. Then*

$$|P_{n-i} - Q_{n-i}| \leq \left| 2^i \binom{n}{i} \right| \sigma_1 \cdots \sigma_{i-1} |B| \qquad (i = 1, \dots, j).$$

The superfluous enclosure of the integer $2^i \binom{n}{i}$ in absolute value signs is quite deliberate; it will be relevant in the nonarchimedean setting.

Proof. Note that Q_{n-i} is the sum of the $\binom{n}{i}$ principal $i \times i$ minors of $A+B$. Each of these principal minors can be written as a sum of 2^i terms, each of which is the product of a sign, a $k \times k$ minor of A, and an $(i-k) \times (i-k)$ minor of B. One of these terms is P_{n-i} itself; the others all have $k < i$, and so have norm bounded by

$$\sigma_1 \cdots \sigma_k |B|^{i-k} \leq \sigma_1 \cdots \sigma_{i-1} |B|$$

since $|B| < \sigma_j$. This proves the claim. □

We also need to consider multiplicative perturbations.

Proposition 3.2.3. *Let $B \in \mathrm{GL}_n(\mathbb{C})$ satisfy $|B| \leq \eta$. Let $\sigma_1', \dots, \sigma_n'$ be the singular values of AB. Then*

$$\sigma_i' \leq \eta \sigma_i \qquad (i = 1, \dots, n).$$

(The analogous result holds with BA replaced by AB, since transposal does not change singular values.)

Proof. We use the interpretation of singular values given by Corollary 3.1.6. Choose an i-dimensional subspace V of $\mathbb{C}^n$ such that $|BAv| \geq \sigma_i' |v|$ for all $v \in V$. Then choose $v \in V$ nonzero such that $|Av| \leq \sigma_i |v|$. We have

$$\sigma_i' |v| \leq |BAv| \leq |B||Av| \leq \sigma_i |B||v|,$$

proving the claim. □

Proposition 3.2.4. *Let $\sigma_{k,1}, \dots, \sigma_{k,n}$ be the singular values of A^k. Then*

$$\lim_{k \to \infty} \sigma_{k,i}^{1/k} = |\lambda_i| \qquad (i = 1, \dots, n).$$

Proof. Pick $\eta > 1$, and choose U as in Lemma 3.1.11; that is, U is upper-triangular, and each block of eigenvalue λ has some scalar c of norm at most $(|\eta| - 1)|\lambda|$. Let U be the matrix effecting the resulting conjugation.

In a block with eigenvalue λ, the singular values of the k-th power are bounded below by $|\lambda|^k$ and above by $\eta^k |\lambda|^k$. Consequently, we may apply Proposition 3.2.3 to deduce that

$$|\lambda_i|^k |U||U^{-1}| \leq \sigma_{k,i} \leq \eta^k |\lambda_i|^k |U||U^{-1}|.$$

Taking k-th roots and then taking $k \to \infty$, we deduce

$$|\lambda_i| \leq \liminf_{k \to \infty} \sigma_{k,i}^{1/k}, \qquad \limsup_{k \to \infty} \sigma_{k,i}^{1/k} \leq \eta |\lambda_i|.$$

Since $\eta > 1$ was arbitrary, we deduce the desired result. □

3. Singular values and eigenvalues (nonarchimedean case)

We now pass to nonarchimedean analogues.

Hypothesis 3.3.1. Throughout this section and the next, let F be a nonarchimedean field, and let A be an $n \times n$ matrix over F.

Definition 3.3.2. Given a sequence $s_1, \dots, s_n$, we define the *associated polygon* for this sequence to be the polygonal line joining the points

$$(-n+i, s_1 + \cdots + s_i) \qquad (i = 0, \dots, n).$$

This polygon is the graph of a convex function on $[-n, 0]$ if and only if $s_1 \leq \cdots \leq s_n$.

Definition 3.3.3. Let $s_1, \dots, s_n$ be the sequence with the property that for $i = 1, \dots, n$, $s_1 + \cdots + s_i$ is the minimum valuation of an $i \times i$ minor of A; that is, s_i are the *elementary divisors* (or *invariant factors*) of A. The associated polygon is called the *Hodge polygon* of A (see the notes for an explanation of the terminology). Define the *singular values* of A as $\sigma_1, \dots, \sigma_n = e^{-s_1}, \dots, e^{-s_n}$; these are invariant under multiplication on either side by a matrix in $\mathrm{GL}_n(\mathfrak{o}_F)$. One has the relation

$$\sigma_1 = |A|,$$

but this time taking the operator norm defined by the supremum norm on F^n.

We also have an analogue of the singular value decomposition.

Theorem 3.3.4 (Smith normal form). *There exist $U, V \in \mathrm{GL}_n(\mathfrak{o}_F)$ such that UAV is a diagonal matrix whose entries have norms $\sigma_1, \dots, \sigma_n$.*

Proof. It is equivalent to prove that starting with A, one can perform elementary row and column operations defined over $\mathfrak{o}_F$ so as to produce a diagonal matrix. To do this, find the largest entry of A, permute rows and columns to put this entry at the top left, then use it to clear the remainder of the first row and column. Repeat with the matrix obtained by removing the first row and column, and so on. □

Corollary 3.3.5. *The slopes $s_1, \dots, s_n$ of the Hodge polygon satisfy $s_1 \leq \cdots \leq s_n$.*

Proof. The i-th slope s_i is evidently the i-th smallest valuation of a diagonal entry of the Smith normal form. □

Corollary 3.3.6. *The number σ_i is the largest value of λ for which the following holds: for any i-dimensional subspace V of F^n, there exists $v \in V$ nonzero such that $|Av| \leq \lambda |v|$.*

Definition 3.3.7. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A in some algebraic extension of F equipped with an extension of $|\cdot|$, sorted with $|\lambda_1| \geq \cdots \geq |\lambda_n|$. The associated polygon is the *Newton polygon* of A; this is invariant under conjugation by any element of $\mathrm{GL}_n(F)$.

The nonarchimedean analogue of Weyl's inequality is the following.

Theorem 3.3.8 (Newton above Hodge). *We have*

$$\sigma_1 \cdots \sigma_i \geq |\lambda_1 \cdots \lambda_i| \qquad (i = 1, \dots, n),$$

with equality for $i = n$. In other words, the Hodge and Newton polygons have the same endpoints, and the Newton polygon is everywhere on or above the Hodge polygon.

Proof. Again, the case $i = 1$ is clear because σ_1 is the operator norm of A, and the general case follows by considering exterior powers. $\square$

Like its archimedean analogue, Theorem 3.3.8 also has a converse, but in this case we can write the construction down quite explicitly.

Definition 3.3.9. For $P = T^n + \sum_{i=0}^{n-1} P_i T^i$ a monic polynomial of degree n over a ring R, the *companion matrix* of P is defined as the matrix

$$\begin{pmatrix} 0 & \cdots & 0 & -P_0 \\ 1 & \cdots & 0 & -P_1 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & -P_{n-1} \end{pmatrix}$$

with 1's on the subdiagonal, the negated coefficients of P in the right column, and 0's elsewhere. The companion matrix is constructed to have characteristic polynomial equal to P.

Proposition 3.3.10. *Choose $\lambda_1, \dots, \lambda_n \in F^{\mathrm{alg}}$ such that $|\lambda_1| \geq \cdots \geq |\lambda_n|$, and the polynomial $P(T) = (T - \lambda_1) \cdots (T - \lambda_n) = T^n + \sum_{i=0}^{n-1} P_i T^i$ has coefficients in F. Choose $c_1, \dots, c_n \in F$ with $\sigma_i = |c_i|$, such that $\sigma_1 \geq \cdots \geq \sigma_n$, and*

$$\sigma_1 \cdots \sigma_i \geq |\lambda_1 \cdots \lambda_i| \qquad (i = 1, \dots, n),$$

with equality for $i = n$. Then the matrix

$$\begin{pmatrix} 0 & \cdots & 0 & -c_1^{-1} \cdots c_{n-1}^{-1} P_0 \\ c_{n-1} & \cdots & 0 & -c_1^{-1} \cdots c_{n-2}^{-1} P_1 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & c_1 & -P_{n-1} \end{pmatrix}$$

has singular values $\sigma_1, \dots, \sigma_n$ and eigenvalues $\lambda_1, \dots, \lambda_n$.

Proof. The given matrix is conjugate to the companion matrix of P, so its eigenvalues are also $\lambda_1, \dots, \lambda_n$. To compute the singular values, we note that for $i = 1, \dots, n-1$,

$$\begin{aligned} |-c_1^{-1} \cdots c_{n-i-1}^{-1} P_i| &= \sigma_1^{-1} \cdots \sigma_{n-i-1}^{-1} |P_i| \\ &\leq \sigma_1^{-1} \cdots \sigma_{n-i-1}^{-1} |\lambda_1 \cdots \lambda_{n-i}| \\ &\leq \sigma_{n-i}. \end{aligned}$$

Thus we can perform column operations over $\mathfrak{o}_F$ to clear everything in the right column except $-c_1^{-1} \cdots c_{n-1}^{-1} P_0$. By permuting the rows and columns, we obtain a diagonal matrix with entries of norms $\sigma_1, \dots, \sigma_n$. This proves the claim. $\square$

Again, equality has a structural meaning, but the proof requires a bit more work than in the archimedean case, since we no longer have access to orthogonality.

Theorem 3.3.11 (Hodge-Newton decomposition). *Suppose that for some $i \in \{1, \dots, n-1\}$ we have*

$$|\lambda_i| > |\lambda_{i+1}|, \qquad \sigma_1 \cdots \sigma_i = |\lambda_1 \cdots \lambda_i|.$$

(That is, the Newton polygon has a vertex with x-coordinate $-n + i$, and this vertex also lies on the Hodge polygon.) Then there exists $U \in \mathrm{GL}_n(\mathfrak{o}_F)$ such that $U^{-1} A U$ is block upper

triangular, with the top left block accounting for the first i singular values and eigenvalues, and the bottom right block accounting for the others. Moreover, if $\sigma_i > \sigma_{i+1}$, we can ensure that $U^{-1}AU$ is block diagonal.

PROOF. Let $v_1, \dots, v_n$ be a basis of F^n such that $v_1, \dots, v_i$ span the generalized eigenspaces with eigenvalues $\lambda_1, \dots, \lambda_i$, and $v_{i+1}, \dots, v_n$ span the generalized eigenspaces with eigenvalues $\lambda_{i+1}, \dots, \lambda_n$. (This can be constructed because by Theorem 2.2.1 applied to the characteristic polynomial of A, $P(T) = (T - \lambda_1) \cdots (T - \lambda_i)$ and $Q(T) = (T - \lambda_{i+1}) \cdots (T - \lambda_n)$ have coefficients in F; we can thus write $1 = PB + QC$ for some $B, C \in F[T]$, and then $P(A)B(A)$ and $Q(A)C(A)$ give projectors for a direct sum decomposition separating the first i generalized eigenspaces from the others.) Choose a basis $w_1, \dots, w_n$ of $\mathfrak{o}_F^n$ such that $w_1, \dots, w_i$ is a basis of $\mathfrak{o}_F^n \cap (Fv_1 + \cdots + Fv_i)$. Let $e_1, \dots, e_n$ be the standard basis of F^n, and define $U \in \mathrm{GL}_n(\mathfrak{o}_F)$ by $w_j = \sum_i U_{ij} e_i$. Then

$$U^{-1}AU = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

is block upper triangular. By Cramer's rule, each entry of $B^{-1}C$ is an $i \times i$ minor of A divided by the determinant of B. Since $|\det(B)| = \sigma_1 \cdots \sigma_i$, $B^{-1}C$ must thus have entries in $\mathfrak{o}_F$. Writing

$$U^{-1}AU = \begin{pmatrix} B & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I_i & B^{-1}C \\ 0 & I_{n-i} \end{pmatrix},$$

we see that the singular values of B and D together must comprise $\sigma_1, \dots, \sigma_n$. The only way for this to happen, given the constraint that the product of the singular values of B equals $\sigma_1 \cdots \sigma_i$, is to have B accounting for $\sigma_1, \cdots, \sigma_i$ and D accounting for the rest.

This proves the first claim; we may thus assume now that $\sigma_i > \sigma_{i+1}$. In that case, conjugating by the matrix

$$\begin{pmatrix} I_i & -B^{-1}C \\ 0 & I_{n-i} \end{pmatrix}$$

gives a new matrix

$$\begin{pmatrix} B & C_1 \\ 0 & D \end{pmatrix}$$

with $C_1 = B^{-1}CD$. Since

$$|C_1| \leq |B^{-1}||C||D| = \sigma_i^{-1}|C|\sigma_{i+1} < |C|,$$

by repeating this process, we converge to a change of basis over $\mathfrak{o}_F$ which converts A into the block diagonal matrix

$$\begin{pmatrix} B & 0 \\ 0 & D \end{pmatrix}$$

which has the desired form. $\square$

Note that the slopes of the Hodge polygon are forced to be in the additive value group of F, whereas the slopes of the Newton polygon need only lie in the divisible closure of the additive value group. Consequently, it is possible for a matrix to have no conjugates over $\mathrm{GL}_n(F)$ for which the Hodge and Newton polygons coincide. However, the following is true; see also Corollary 3.4.8 below.

LEMMA 3.3.12. *Suppose that one of the following holds.*

(a) *The value group of* $|F^\times|$ *is dense in* $\mathbb{R}_{>0}$, *and* $\eta > 1$.
(b) *We have* $|\lambda_i| \in |F^\times|$ *for* $i = 1, \dots, n$ *(so in particular* $\lambda_i \neq 0$*), and* $\eta \geq 1$.

Then there exists $U \in \mathrm{GL}_n(F)$ *such that the i-th singular value of* $U^{-1}AU$ *is at most* $\eta|\lambda_i|$ *(with equality in case (b)).*

Proof. Case (a) will follow from Corollary 3.4.8 below. Case (b) is directly analogous to Lemma 3.1.11. □

One also has the following variant.

Lemma 3.3.13. *Suppose that* $|F^\times|$ *is discrete. Then there exists* $U \in \mathrm{GL}_n(F)$ *such that for each positive integer* m, $|U^{-1}A^mU|$ *is the least element of* $|F^\times|$ *greater than or equal to* $|\lambda_1^m|$.

Proof. It suffices to check the case where A acts irreducibly on F^n. In this case, choose $v \in F^n$ nonzero; then v is a cyclic vector under A. If $P(T) = T^n + \sum_{i=0}^{n-1} P_i T^i$ is the characteristic polynomial of A, then $A^n v = -\sum_{i=0}^{n-1} P_i A^i v$ by the Cayley-Hamilton theorem. By Theorem 2.2.1, the Newton polygon of P consists of some single slope $r = -\log|\lambda_1|$.

Choose $c_1, \dots, c_n \in F$ such that for $i = 1, \dots, n$, $|c_1 \cdots c_i|$ is the least element of $|F^\times|$ greater than or equal to $|\lambda_1|^i$. Choose U so that $U^{-1}AU$ is the companion matrix in Proposition 3.3.10; one checks that this gives the desired property for $m = 1, \dots, n$. In particular, $U^{-1}A^nU$ has all its Hodge and Newton slopes equal, so the desired conclusion for m implies the desired conclusion for $m + n$. □

4. Perturbations (nonarchimedean case)

Again, we can ask about the effect of perturbations. The analogue of Weyl's second inequality is more or less trivial.

Proposition 3.4.1. *If* B *is a matrix with* $|B| < \sigma_i$, *then the first* i *singular values of* $A + B$ *are* $\sigma_1, \dots, \sigma_i$.

Proof. Exercise. □

We next consider the effect on the characteristic polynomial.

Theorem 3.4.2. *Let* B *be an* $n \times n$ *matrix such that* $|B| < \sigma_j$ *for some* $j \in \{1, \dots, n\}$. *Let* $P(T) = T^n + \sum_{i=0}^{n-1} P_i T^i$ *and* $Q(T) = T^n + \sum_{i=0}^{n-1} Q_i T^i$ *be the characteristic polynomials of* A *and* $A + B$. *Then*

$$|P_{n-i} - Q_{n-i}| \leq \sigma_1 \cdots \sigma_{i-1} |B| \qquad (i = 1, \dots, j).$$

Proof. The proof is as for Theorem 3.2.2, except now the factor $|2^n \binom{n}{i}^2|$ is dominated by 1. □

Question 3.4.3. *Is Theorem 3.4.2 best possible?*

We may also consider multiplicative perturbations.

Proposition 3.4.4. *Let* $B \in \mathrm{GL}_n(F)$ *satisfy* $|B| \leq \eta$. *Let* $\sigma_1', \dots, \sigma_n'$ *be the singular values of* AB. *Then*

$$\sigma_i' \leq \eta \sigma_i \qquad (i = 1, \dots, n).$$

Proof. As for Proposition 3.2.3, but using the Smith normal form instead of the singular value decomposition. $\square$

Corollary 3.4.5. *Suppose that the Newton and Hodge slopes of A coincide, and that $U \in \mathrm{GL}_n(F)$ satisfies $|U| \cdot |U^{-1}| \leq \eta$. Then each Newton slope of $U^{-1}AU$ is at most $\log \eta$ more than the corresponding Hodge slope.*

Here is a weak converse to Corollary 3.4.5. (We leave the archimedean analogue to the reader's imagination.)

Proposition 3.4.6. *Suppose that the Newton slopes of A are nonnegative and that $\sigma_1 \geq 1$. Then there exists $U \in \mathrm{GL}_n(F)$ such that*

$$|U^{-1}AU| \leq 1, \quad |U^{-1}| \leq 1, \quad |U| \leq \sigma_1^{n-1}.$$

Proof. Let $e_1, \dots, e_n$ denote the standard basis vectors. Let M be the smallest $\mathfrak{o}_F$-submodule of F^n containing $e_1, \dots, e_n$ and stable under A. For each i, if $j = j(i)$ is the least integer such that $e_i, Ae_i, \dots, A^j e_i$ are linearly dependent, then we have $A^j e_i = \sum_{h=0}^{j-1} c_h A^h e_i$ for some $c_h \in F$; the polynomial $T^j - \sum_{h=0}^{j-1} c_h T^h$ has roots which are eigenvalues of F, so the nonnegativity of the Newton slopes forces $|c_h| \leq 1$. Hence M is finitely generated, and thus free, over $\mathfrak{o}_F$.

Let $v_1, \dots, v_n$ be a basis of M, and let U be the change-of-basis matrix $v_j = \sum_i U_{ij} e_i$; then $|U^{-1}AU| \leq 1$ because M is stable under A, and $|U^{-1}| \leq 1$ because M contains $e_1, \dots, e_n$. The desired bound on U will follow from the fact that for any $x = c_1 e_1 + \cdots + c_n e_n \in M$, we have

$$\max_i \{|c_i|\} \leq \sigma_1^{n-1}. \tag{3.4.6.1}$$

It suffices to check (3.4.6.1) for $x = A^h e_i$ for $i = 1, \dots, n$ and $h = 0, \dots, j(i) - 1$, as these generate M over $\mathfrak{o}_F$. But it is evident that $|A^h e_1| \leq \sigma_1^h |e_1| = \sigma_1^h$, so we are done. $\square$

Example 3.4.7. The example

$$A = \begin{pmatrix} 1 & c & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}$$

with $|c| > 1$ shows that this bound of Proposition 3.4.6 is sharp; in particular, the bound $|U| \leq \sigma_1^{n-1}$ cannot be improved to $|U| \leq \sigma_1$, as one might initially expect. However, one should be able to get a more precise bound (which agrees with the given bound in this example) by accounting for the other singular values; see Problem 3.4.6.

Corollary 3.4.8. *There exists a continuous function*

$$f_n(\sigma_1, \dots, \sigma_n, \sigma_1', \dots, \sigma_n') : (0, \infty)^{2n} \to (0, \infty)$$

(independent of F) with the following properties.

(a) *If $\sigma_i = \sigma_i'$ for $i = 1, \dots, n$, then $f = 1$.*
(b) *If A has singular values $\sigma_1, \dots, \sigma_n$ and eigenvalues $\lambda_1, \dots, \lambda_n$, none equal to 0, and $\sigma_1', \dots, \sigma_n' \in |F^\times|$ satisfy*

$$\sigma_1 \cdots \sigma_i \geq \sigma_1' \cdots \sigma_i' \geq |\lambda_1 \cdots \lambda_i| \qquad (i = 1, \dots, n),$$

then there exists $U \in \mathrm{GL}_n(F)$ *such that*

$$|U^{-1}| \leq 1, \quad |U| \leq f_n(\sigma_1, \dots, \sigma_n, \sigma'_1, \dots, \sigma'_n)$$

for which $U^{-1}AU$ *has singular values* $\sigma'_1, \dots, \sigma'_n$.

PROOF. This follows by induction on n, using Proposition 3.4.6 (after appropriate rescaling), Proposition 3.4.4, and Theorem 3.3.11. □

For the purposes of this book, it is immaterial what the function f_n is, as long as it is continuous. However, for numerical applications, it may be quite important to identify a good function f; here is a conjectural best possible result. (One can also formulate an archimedean analogue.)

CONJECTURE 3.4.9. *In Corollary 3.4.8, we may take*

$$f_n(\sigma_1, \dots, \sigma_n, \sigma'_1, \dots, \sigma'_n) = \max_i \{(\sigma_1 \cdots \sigma_i)/(\sigma'_1 \cdots \sigma'_i)\}.$$

By imitating the proof of Proposition 3.2.4, we obtain the following.

PROPOSITION 3.4.10. *Let* $\sigma_{k,1}, \dots, \sigma_{k,n}$ *be the singular values of* A^k. *Then*

$$\lim_{k \to \infty} \sigma_{k,i}^{1/k} = |\lambda_i| \qquad (i = 1, \dots, n).$$

5. Horn's inequalities

Although they will not be needed in this course, it is quite natural to mention here some stronger versions of the perturbation inequalities in the archimedean and nonarchimedean cases, introduced by Horn [**Hor62**] in the archimedean case. See the beautiful survey article of Fulton [**Ful00**] for more information.

To introduce the stronger inequalities, we must set up some notation. Put

$$U_r^n = \{(I, J, K) : I, J, K \subseteq \{1, \dots, n\}, \#I = \#J = \#K = r,$$
$$\sum_{i \in I} i + \sum_{j \in J} j = \sum_{k \in K} k + r(r+1)/2\}.$$

For $(I, J, K) \in U_r^n$, write $I = \{i_1 < \cdots < i_r\}$ and similarly for J, K. For $r = 1$, put $T_1^n = U_1^n$. For $r > 1$, put

$$T_r^n = \{(I, J, K) \in U_r^n : \text{for all } p < r \text{ and } (F, G, H) \in T_p^r,$$
$$\sum_{f \in F} i_f + \sum_{g \in G} j_g \leq \sum_{h \in H} k_h + p(p+1)/2\}.$$

For multiplicative perturbations, we obtain the following results, which include the Weyl inequalities (Theorem 3.1.7, Theorem 3.3.8) as well as Propositions 3.2.3 and 3.4.4. It is important for the proofs that one can rephrase the Horn inequalities in terms of Littlewood-Richardson numbers; see [**Ful00**, §3].

THEOREM 3.5.1. *For* $* \in \{A, B, C\}$, *let* $\sigma_{*,1}, \dots, \sigma_{*,n}$ *be a nonincreasing sequence of nonnegative real numbers. Then the following are equivalent.*

(a) *There exist* $n \times n$ *matrices* A, B, C *over* $\mathbb{C}$ *with* $AB = C$ *such that for* $* \in \{A, B, C\}$, $*$ *has singular values* $\sigma_{*,1}, \dots, \sigma_{*,n}$.

(b) *We have* $\prod_{i=1}^n \sigma_{A,i} \prod_{j=1}^n \sigma_{B,j} = \prod_{k=1}^n \sigma_{C,k}$*, and for all* $r < n$ *and* $(I, J, K) \in T_r^n$*, we have*

$$\prod_{k\in K} \sigma_{C,k} \leq \prod_{i\in I} \sigma_{A,i} \prod_{j\in J} \sigma_{B,j}.$$

Proof. See [**Ful00**, Theorem 16]. Note that the first condition in (b) is omitted in the statement given in [**Ful00**], but this is only a typo. □

Theorem 3.5.2. *Let F be a complete nonarchimedean field with additive value group G. For $* \in \{A, B, C\}$, let $\sigma_{*,1}, \dots, \sigma_{*,n}$ be a nonincreasing sequence of elements of $G \cup \{0\}$. Then the following are equivalent.*

(a) *There exist $n\times n$ matrices A, B, C over F with $AB = C$ such that for $* \in \{A, B, C\}$, $*$ has singular values $\sigma_{*,1}, \dots, \sigma_{*,n}$.*

(b) *We have* $\prod_{i=1}^n \sigma_{A,i} \prod_{j=1}^n \sigma_{B,j} = \prod_{k=1}^n \sigma_{C,k}$*, and for all* $r < n$ *and* $(I, J, K) \in T_r^n$*, we have*

$$\prod_{k\in K} \sigma_{C,k} \leq \prod_{i\in I} \sigma_{A,i} \prod_{j\in J} \sigma_{B,j}.$$

Proof. See [**Ful00**, Theorem 7]. □

For additive perturbations, one has an analogous result in the archimedean case; see [**Ful00**, Theorem 15]. I am not aware of an additive result in the nonarchimedean case. Also, in the archimedean case one has analogous results (with slightly different statements) in which one restricts to Hermitian matrices.

Notes

The subject of archimedean matrix inequalities is an old one, with many important applications. A good reference for this is [**Bha97**]; for instance, see [**Bha97**, §I.2] for the singular value decomposition, [**Bha97**, Theorem II.3.6] for the Weyl inequalities (in a much stronger form known as Weyl's majorant theorem), [**Bha97**, Theorem III.4.5] for a strong form of Proposition 3.2.3 (also a consequence of the Horn inequalities), and so on. (A variant of our Theorem 3.2.2 appears as [**Bha97**, Problem I.6.11].)

The strong analogy between archimedean and p-adic numerical analysis appears to be a little-known piece of folklore. As a result, we have been unable to locate a suitable reference.

It should be pointed out that most of what we have done here is the special case for GL_n of a more general theory encompassing the other reductive algebraic groups. This point of view can be seen in [**Ful00**], where GL_n makes some explicit appearances for which other groups can be substituted.

In Theorem 3.1.10, the equivalence of (a) and (b) is standard. We do not have a reference for the equivalence with (c), although it is implicit in most proofs of the equivalence of (a) and (b).

The reader familiar with the notions of elementary divisors or invariant factors may be wondering why the terminology "Hodge polygon" is necessary or reasonable. The answer is that the Hodge numbers of a variety over a p-adic field are reflected by the elementary divisors of the action of Frobenius on crystalline cohomology. The fact that the Newton polygon lies above the Hodge polygon then implies a relation between the characteristic polynomial of Frobenius and the Hodge numbers of the original variety; this relationship was originally conjectured by Katz and proved by Mazur. See [**BO78**] for further discussion of this point, and of crystalline cohomology as a whole.

Much of the work in this chapter can be carried over to the case of a transformation which is only semilinear for some isometric endomorphism of F. This case arises in the study of slope filtrations of Frobenius crystals (F-crystals), as in [**Kat79**]; in fact, the Hodge-Newton decomposition theorem (Theorem 3.3.11) is a direct translation of Katz's corresponding theorem for F-crystals [**Kat79**, Theorem 1.6.1]. The archimedean version (Theorem 3.1.9) is itself a translation of Theorem 3.3.11; we do not know of a reference, although we do not make any claim of originality. Likewise, Proposition 3.4.10 is a direct translation of [**Kat79**, Corollary 1.4.4]; its archimedean analogue (Proposition 3.2.4) is doubtless also known, but we do not have a reference.

The question of how much the characteristic polynomial of a square matrix over a field is affected by a perturbation arises in numerical applications. This is a familiar fact in the archimedean case, but perhaps less so in the nonarchimedean case; numerical applications of the latter include using p-adic cohomology to compute zeta functions of varieties over finite fields. See for instance [**AKR07**, §1.6], [**Ger07**, §3].

Exercises

(1) Prove Proposition 3.4.1.

(2) With notation as in Theorem 3.3.11, suppose $U, V \in \mathrm{GL}_n(\mathfrak{o}_F)$ are congruent to the identity matrix modulo $\mathfrak{m}_F$. Prove that the product of the i largest eigenvalues of UAV again has norm $|\lambda_1 \cdots \lambda_i|$. (Hint: use exterior powers to reduce to the case $i = 1$.) This yields as a corollary [**BC05**, Lemma 5]: if $D \in \mathrm{GL}_n(F)$ is diagonal and $U, V \in \mathrm{GL}_n(\mathfrak{o}_F)$ are congruent to the identity matrix modulo $\mathfrak{m}_F$, then the Newton polygons of D and UDV coincide.

(3) State and prove an archimedean analogue of the previous problem.

(4) Prove the following improved version of Proposition 3.4.6. Suppose that the Newton slopes of A are nonnegative. Then there exists $U \in \mathrm{GL}_n(F)$ such that

$$|U^{-1}AU| \leq 1, \quad |U^{-1}| \leq 1, \quad |U| \leq \prod_{i=1}^{n-1} \max\{1, \sigma_i\}.$$

I do not know of an appropriate archimedean analogue.

Part 2

Differential algebra

CHAPTER 4

Formalism of differential algebra

In this chapter, we introduce some basic formalism of differential algebra.

1. Differential rings and differential modules

DEFINITION 4.1.1. A *differential ring* is a commutative ring R equipped with a derivation $d : R \to R$, i.e., an additive map satisfying the Leibniz rule

$$d(ab) = ad(b) + bd(a) \qquad (a, b \in R).$$

We expressly allow $d = 0$ unless otherwise specified; this will come in handy in some situations. A differential ring which is also a domain, field, etc., will be called a *differential domain, field*, etc.

DEFINITION 4.1.2. A *differential module* over a differential ring (R, d) is a module M equipped with an additive map $D : M \to M$ satisfying

$$D(am) = aD(m) + d(a)m;$$

such a D will also be called a *differential operator* on M relative to d. For example, (R, d) is a differential module over itself; any differential module isomorphic to a direct sum of copies of (R, d) is said to be *trivial*. (If we refer to "the" trivial differential module, though, we mean (R, d) itself.) A *differential ideal* of R is a differential submodule of R itself, i.e., an ideal stable under d.

DEFINITION 4.1.3. For (M, D) a differential module, define

$$H^0(M) = \ker(D), \qquad H^1(M) = \operatorname{coker}(D) = M/D(M).$$

The latter computes Yoneda extensions; see Lemma 4.3.3 below. Elements of $H^0(M)$ are said to be *horizontal* (see notes). Note that $H^0(R) = \ker(d)$ is a subring of R; if R is a field, then $\ker(d)$ is a subfield. We call this the *constant subring/subfield* of R.

REMARK 4.1.4. If R_0 is the constant subring of R, and R' is an R_0-algebra, then there are natural maps $H^i(M) \otimes_{R_0} R' \to H^i(M \otimes_{R_0} R')$. This map is always an isomorphism for $i = 1$, and is an isomorphism for $i = 0$ if R' is flat over R_0.

2. Differential modules and differential systems

DEFINITION 4.2.1. Let R be a differential ring, and let M be a finite free differential module of rank n over R. Let $e_1, \dots, e_n$ be a basis of M. Then for any $v \in M$, we can write $v = v_1 e_1 + \cdots + v_n e_n$ for some $v_1, \dots, v_n \in R$, and then compute

$$D(v) = v_1 D(e_1) + \cdots + v_n D(e_n) + d(v_1)e_1 + \cdots + d(v_n)e_n.$$

If we define the $n \times n$ matrix N over R by the formula

$$D(e_j) = \sum_{i=1}^{n} N_{ij} e_i$$

(we will sometimes call this the *matrix of action of* D on this basis), we then have

$$D(v) = \sum_{i=1}^{n} \left(d(v_i) + \sum_j N_{ij} v_j \right) e_i.$$

That is, if we identify v with the column vector $\mathbf{v} = [v_1, \dots, v_n]$, then

$$D(\mathbf{v}) = N\mathbf{v} + d(\mathbf{v}).$$

Conversely, it is clear that given the underlying finite free R-module, any differential module structure is given by such an equation.

Remark 4.2.2. In other words, differential modules are a coordinate-free version of differential systems. If you are a geometer, you may wish to go further and think of *differential bundles*, i.e., vector bundles equipped with a differential operator. A differential operator on a vector bundle is usually called a *connection*.

3. Operations on differential modules

Definition 4.3.1. For R a differential ring, we regard the differential modules over R as a category in which the morphisms (or *homomorphisms*) from M_1 to M_2 are additive maps $f : M_1 \to M_2$ satisfying $D(f(m)) = f(D(m))$ (we sometimes say these maps are *horizontal*).

The category of differential modules over a differential ring admits certain functors corresponding to familiar functors on the category of modules over an ordinary ring. (Beware that in the following notations, the subscripted R on such symbols as the tensor product will often be suppressed when it is unambiguous.)

Definition 4.3.2. Given two differential modules M_1, M_2, the tensor product $M_1 \otimes_R M_2$ in the category of rings may be viewed as a differential module via the formula

$$D(m_1 \otimes m_2) = D(m_1) \otimes m_2 + m_1 \otimes D(m_2).$$

Similarly, the exterior power $\wedge_R^n M$ may be viewed as a differential module via the formula

$$D(m_1 \wedge \cdots \wedge m_n) = \sum_{i=1}^{n} m_1 \wedge \cdots \wedge m_{i-1} \wedge D(m_i) \wedge m_{i+1} \wedge \cdots \wedge m_n;$$

likewise for the symmetric power $\mathrm{Sym}_R^n M$. The module of R-homomorphisms $\mathrm{Hom}_R(M_1, M_2)$ may be viewed as a differential module via the formula

$$D(f)(m) = D(f(m)) - f(D(m));$$

the homomorphisms from M_1 to M_2 as differential modules are precisely the horizontal elements of $\mathrm{Hom}_R(M_1, M_2)$. If $M_2 \cong R$ is trivial, we write $M_1^\vee$ for $\mathrm{Hom}_R(M_1, R)$ and call it the *dual* of M_1; if M_1 is finite projective (which is the same as finite locally free if R is a noetherian ring), then $\mathrm{Hom}_R(M_1, M_2) \cong M_1^\vee \otimes M_2$ and the natural map $M_1 \to (M_1^\vee)^\vee$ is an isomorphism.

Lemma 4.3.3. *Let M, N be differential modules with M finite projective. Then the group $H^1(M^\vee \otimes N)$ is canonically isomorphic to the Yoneda extension group* $\operatorname{Ext}(M, N)$.

Proof. The group $\operatorname{Ext}(M, N)$ consists of equivalence classes of exact sequences $0 \to N \to P \to M \to 0$ under the relation that this sequence is equivalent to a second sequence $0 \to N \to P' \to M \to 0$ if there is an isomorphism $P \cong P'$ that induces the identity maps on M and N. The addition is to take two such sequences and return the *Baer sum* $0 \to N \to (P \oplus P')/\Delta \to M \to 0$, where $\Delta = \{(n, -n) : n \in N\}$. The identity element is the split sequence $0 \to N \to M \oplus N \to M \to 0$. The inverse of a sequence $0 \to N \to P \to M \to 0$ is the sequence $0 \to N \to P \to M \to 0$ with the map $N \to P$ negated.

Given an extension $0 \to N \to P \to M \to 0$, tensor with $M^\vee$ to get $0 \to M^\vee \otimes N \to M^\vee \otimes P \to M^\vee \otimes P \to 0$, and apply the connecting homomorphism $H^0(M^\vee \otimes M) \to H^1(M^\vee \otimes N)$ from the snake lemma to the trace (the element of $M^\vee \otimes M$ corresponding to the identity map in $\operatorname{Hom}(M, M)$) to get an element of $H^1(M^\vee \otimes N)$. This is the desired map $\operatorname{Ext}(M, N) \to H^1(M^\vee \otimes N)$. To construct its inverse, given an element $H^1(M^\vee \otimes N)$ represented by $x \in M^\vee \otimes N$, form the sequence

$$0 \to N \to \frac{M \oplus N}{(m, \langle m, x\rangle)} \to M \to 0$$

where $\langle \cdot, \cdot \rangle$ represents the natural map $M \times (M^\vee \otimes N) \to N$. □

4. Cyclic vectors

Definition 4.4.1. Let R be a differential ring, and let M be a finite free differential module of rank n over R. A *cyclic vector* for M is an element $m \in M$ such that $m, D(m), \dots, D^{n-1}(m)$ form a basis of M.

Theorem 4.4.2 (Cyclic vector theorem). *Let R be a differential field of characteristic zero with nonzero derivation. Then every finite differential module over R has a cyclic vector.*

For a comment on characteristic p, see the exercises.

Proof. This is a folklore result, that is, it is old enough that giving a proper attribution is difficult. Many proofs are possible; here is the proof from [**DGS94**, Theorem III.4.2].

We start by normalizing the derivation. For $u \in R^\times$, given one differential module (M, D) over (R, d), we get another differential module (M, uD) over (R, ud), and m is a cyclic vector for one if and only if it is a cyclic vector for the other (because the image of m under $(uD)^j$ is in the span of $u, D(u), \dots, D^j(u)$). We may thus assume (thanks to the assumption that the derivation is nontrivial) that there exists an element $x \in R$ such that $d(x) = x$.

Let M be a differential module of dimension n, and choose $m \in M$ so that the dimension μ of the span of $m, D(m), \dots$ is as large as possible. We derive a contradiction under the hypothesis $\mu < n$.

For $z \in M$ and $\lambda \in \mathbb{Q}$, we now have

$$(m + \lambda z) \wedge D(m + \lambda z) \wedge \cdots \wedge D^\mu(m + \lambda z) = 0$$

in the exterior power $\wedge^{\mu+1} M$. If we write this expression as a polynomial in λ, it vanishes for infinitely many values, so must be identically zero. Hence each coefficient must vanish separately, including the coefficient of λ^1, which is

$$\sum_{i=0}^{\mu} m \wedge \cdots \wedge D^{i-1}(m) \wedge D^i(z) \wedge D^{i+1}(m) \wedge \cdots \wedge D^{\mu}(m). \tag{4.4.2.1}$$

Pick $s \in \mathbb{Z}$, substitute $x^s z$ for z in (4.4.2.1), divide by x^s, and set equal to zero. We get

$$\sum_{i=0}^{\mu} s^i \Lambda_i(m, z) = 0 \qquad (s \in \mathbb{Z}) \tag{4.4.2.2}$$

for

$$\Lambda_i(m, z) = \sum_{j=0}^{\mu-i} \binom{i+j}{i} m \wedge \cdots \wedge D^{i+j-1}(m) \wedge D^j(z) \wedge D^{i+j+1}(m) \wedge \cdots \wedge D^{\mu}(m).$$

Again because we are in characteristic zero, we may conclude that (4.4.2.2), viewed as a polynomial in s, has all coefficients equal to zero; that is, $\Lambda_i(m, z) = 0$ for all $m, z \in M$.

We now take $i = \mu$ to obtain

$$(m \wedge \cdots \wedge D^{\mu-1}(m)) \wedge z = 0 \qquad (m, z \in M);$$

since $\mu < n$, we may use this to deduce

$$m \wedge \cdots \wedge D^{\mu-1}(m) = 0 \qquad (m \in M).$$

But that means that the dimension of the span of $m, D(m), \dots$ is always at most $\mu - 1$, contradicting the definition of μ. $\square$

REMARK 4.4.3. If R is not a field, then one obstruction to having a cyclic vector is that M itself might not be a finite free R-module. But even if it is, there is no reason to expect in general that cyclic vectors exist; this will create complications for us later.

5. Differential polynomials

DEFINITION 4.5.1. Let (R, d) be a differential ring. The *ring of twisted polynomials* $R\{T\}$ over R in the variable T is the additive group

$$R \oplus (R \cdot T) \oplus (R \cdot T^2) \oplus \cdots,$$

with noncommuting multiplication given by the formula

$$\left(\sum_{i=0}^{\infty} a_i T^i\right)\left(\sum_{j=0}^{\infty} b_j T^j\right) = \sum_{i,j=0}^{\infty} \sum_{h=0}^{j} \binom{j}{h} a_i d^h(b_j) T^{i+j-h}.$$

In other words, you impose the relation

$$Ta = aT + d(a) \qquad (a \in R)$$

and check that you get a sensible (but not necessarily commutative) ring. We define the *degree* of a twisted polynomial in the usual way, as the exponent of the largest power of T with a nonzero coefficient; the degree of the zero polynomial may be taken to be any particular negative value.

Proposition 4.5.2 (Ore). *For R a differential field, the ring $R\{T\}$ admits a left division algorithm. That is, if $f, g \in R\{T\}$ and $g \neq 0$, then there exist unique $q, r \in R\{T\}$ with $\deg(r) < \deg(g)$ and $f = gq + r$. (There is also a right division algorithm.)*

Proof. Exercise. $\square$

Using the Euclidean algorithm, this yields the following consequence as in the untwisted case.

Theorem 4.5.3 (Ore). *Let R be a differential field. Then $R\{T\}$ is both left principal and right principal; that is, any left ideal (resp. right ideal) has the form $R\{T\}f$ (resp. $fR\{T\}$) for some $f \in R\{T\}$.*

Definition 4.5.4. Note that the opposite ring to $R\{T\}$, i.e., the ring with left and right reversed, is again a twisted polynomial ring, but for the derivation $-d$. Given $f \in R\{T\}$, we define the *formal adjoint* of f as the element f in the opposite ring. This operation looks a bit less formal if you also push the coefficients over to the other side, giving what we will call the *adjoint form* of f. For instance, the adjoint form of $T^3 + aT^2 + bT + c$ is

$$T^3 + T^2 a + T(b - 2d(a)) + d(d(a)) - d(b) + c.$$

Remark 4.5.5. The twisted polynomial ring is rigged up precisely so that for any differential module M over R, we get an action of $R\{T\}$ on M under which T acts like D. In particular, $R\{T\}$ acts on R itself with T acting like d. In fact, the category of differential modules over R is equivalent to the category of left $R\{T\}$-modules. Moreover, if M is a differential module, any cyclic vector $m \in M$ corresponds to an isomorphism $M \cong R\{T\}/R\{T\}P$ for some monic twisted polynomial P, where the isomorphism carries m to the class of 1. (You might want to think of f as a sort of "characteristic polynomial" for M, except that it depends strongly on the choice of the cyclic vector.) Under such an isomorphism, a factorization $P = P_1 P_2$ corresponds to a short exact sequence $0 \to M_1 \to M \to M_2 \to 0$ with

$$M_1 \cong R\{T\}P_2/R\{T\}P \cong R\{T\}/R\{T\}P_1, \qquad M_2 \cong R\{T\}/R\{T\}P_2.$$

6. Differential equations

You may have been wondering when differential equations will appear, those supposedly being the objects of study of this book. If so, your wait is over.

Definition 4.6.1. A *differential equation of order n* over the differential ring (R, d) is an equation of the form

$$(a_n d^n + \cdots + a_1 d + a_0) y = b,$$

with $a_0, \ldots, a_n, b \in R$, and y indeterminate. We say the equation is *homogeneous* if $b = 0$ and *inhomogeneous* otherwise.

Using our setup, we may write this equation as $f(d)y = b$ for some $f \in R\{T\}$. Similarly, we may view systems of differential equations as being equations of the form $f(D)y = b$ where b lives in some differential module (M, D). By the usual method (of introducing extra variables corresponding to derivatives of y), we can convert any differential system into a first-order system $Dy = b$. We can also convert an inhomogeneous system into a homogeneous one by adding an extra variable, with the understanding that we would like the value of that last variable to be 1 in order to get back a solution of the original equation.

Here is a more explicit relationship between adjoint polynomials and solving differential equations. Say you start with the cyclic differential module $M \cong R\{T\}/R\{T\}f$ and you want to find a horizontal element. That means that you want to find some $g \in R\{T\}$ such that $Tg \in R\{T\}f$; we may as well assume that $\deg(g) < \deg(f)$. Then by comparing degrees, we see that in fact $Tg = rf$ for some $r \in R$. Write f in adjoint form as $f_0 + Tf_1 + \cdots + T^n$; then

$$rf \equiv rf_0 - d(r)f_1 + d^2(r)f_2 - \cdots \pm d^n(f) \mod TR\{T\}.$$

In this manner, finding a horizontal element becomes equivalent to solving a differential equation.

7. Cyclic vectors: a mixed blessing

The reader may at this point be wondering why so many points of view are necessary, since the cyclic vector theorem can be used to transform any differential module into a differential equation, and ultimately differential equations are the things one writes down and wants to solve. Permit me to interject here a countervailing opinion.

In ordinary linear algebra (or in other words, when considering differential modules for the trivial derivation), one can pass freely between linear transformations on a vector space and square matrices if one is willing to choose a basis. The merits of doing this depend on the situation, so it is valuable to have both the matricial and coordinate-free viewpoints well in hand. One can then pass to the characteristic polynomial, but not all information is retained (one loses information about nilpotency), and even information that in principle is retained is sometimes not so conveniently accessed. In short, no one would seriously argue that one can dispense with studying matrices because of the existence of the characteristic polynomial.

The situation is not so different in the differential case. The difference between a differential module and a differential system is merely the choice of a basis, and again it is valuable to have both points of view in mind. However, the cyclic vector theorem may seduce one into thinking that collapsing a differential system into a differential polynomial is an operation without drawbacks, and this is far from the case. For instance, determining whether two differential polynomials correspond to the same differential system is not straightforward.

More seriously for our purposes, the cyclic vector theorem only applies over a differential field. Many differential modules are more naturally defined over some ring which is not a field, e.g., those coming from geometry which should be defined over some sort of ring of functions on some sort of geometric space. Working with differential modules instead of differential polynomials has a tremendously clarifying effect over rings.

We find it unfortunate that much of the literature on complex ordinary differential equations, and nearly all of the literature on p-adic ordinary differential equations, is mired in the language of differential polynomials. By instead switching between differential modules and differential polynomials as appropriate, we will be able to demonstrate strategies that lead to a more systematic development of the p-adic theory.

8. Taylor series

Definition 4.8.1. Let R be a *topological* differential ring, i.e., a ring equipped with a topology and a derivation such that all operations are continuous. Assume also that R is a $\mathbb{Q}$-algebra. Let M be a topological differential module over R, i.e., a differential module such that all operations are continuous. For $r \in R$ and $m \in M$, we define the *Taylor series* $T(r,m)$ as the infinite sum

$$\sum_{i=0}^{\infty} \frac{r^i}{i!} D^i(m)$$

whenever the sum converges absolutely (i.e., all rearrangements converge to the same value).

The map $T(r,m)$ is *de facto* additive in m: if $m_1, m_2 \in M$, then

$$T(r,m_1) + T(r,m_2) = T(r, m_1 + m_2)$$

whenever all three terms make sense. Also, the map $T(r,\cdot) : R \to R$ is *de facto* a ring homomorphism: if $s_1, s_2 \in R$, then (by the Leibniz rule)

$$T(r,s_1)T(r,s_2) = T(r,s_1 s_2)$$

whenever all three terms make sense. (Key example: if R is a completion of a rational function field $F(t)$ and $d = d/dt$, then this ring homomorphism is the substitution $t \mapsto t + r$. Note that this can only make sense if $|r| \leq 1$.) More generally, the map $T(r,\cdot)$ on M is *de facto* semilinear for the ring homomorphism $T(r,\cdot)$ on R: if $s \in R$, $m \in M$, then

$$T(r,s)T(r,m) = T(r,sm)$$

whenever all three terms make sense.

Another use for Taylor series is to construct horizontal sections. Note that

$$\begin{aligned} D(T(r,m)) &= \sum_{i=1}^{\infty} d(r) \frac{r^{i-1}}{(i-1)!} D^i(m) + \sum_{i=0}^{\infty} \frac{r^i}{i!} D^{i+1}(m) \\ &= (1 + d(r)) T(r,m) \end{aligned}$$

if everything converges absolutely. In particular, if $d(r) = -1$, then $T(r,m)$ is horizontal.

Notes

The subject of differential algebra is rather well-developed; a classic treatment, though possibly too dry to be relevant, is the book of Ritt [**Rit50**]. As in abstract algebra in general, development of differential algebra was partly driven by *differential Galois theory*, i.e., the study of when solutions of differential equations can be expressed in terms of solutions to ostensibly simpler differential equations. A relatively lively introduction to the latter is [**SvdP03**].

Calling an element of a differential module *horizontal* when it is killed by the derivation makes sense if you consider *connections* in differential geometry. In that setting, the differential operator is measuring the extent to which a section of a vector bundle deviates from some prescribed "horizontal" direction identifying points on one fibre with points on nearby fibres.

Twisted polynomials were introduced by Ore [**Ore33**]. They are actually somewhat more general than we have discussed; for instance, one can also twist by an endomorphism

$\tau : R \to R$ by imposing the relation $Ta = \tau(a)T$. (This enters the realm of the analogue of differential algebra called *difference algebra*, which we will treat in Part 4.) Moreover, one can twist by both an endomorphism and a derivation if they are compatible in an appropriate way, and one can even study differential/difference Galois theory in this setting. A unifying framework for doing so, which is also suitable for considering multiple derivations and automorphisms, is given by André [**And01**].

Differential algebra in positive characteristic has a rather different flavor than in characteristic 0; for instance, the p-th power of the derivation d/dt on $\mathbb{F}_p(t)$ is the zero map. A brief discussion of the characteristic p situation is given in [**DGS94**, §III.1].

Exercises

(1) Prove that if M is a locally free differential module over R of rank 1, then $M^\vee \otimes M$ is trivial (as a differential module).
(2) Check that in characteristic $p > 0$, the cyclic vector theorem holds for modules of rank less than p, but may fail for modules of rank p.
(3) Give a counterexample to the cyclic vector theorem for a differential field of characteristic zero with trivial derivation.
(4) Verify that $R\{T\}$ is indeed a ring; the content in this is to check associativity of multiplication.
(5) Prove the division algorithm (Proposition 4.5.2).

CHAPTER 5

Metric properties of differential modules

In this chapter, we study the metric properties of differential modules over nonarchimedean differential rings.

1. Spectral norms of linear operators

To illustrate what we have in mind, let us review first the difference between the operator norm and spectral norm of a linear operator.

DEFINITION 5.1.1. Let F be a field equipped with a norm $|\cdot|$, let V be a vector space over F equipped with a compatible norm $|\cdot|_V$, and let $T : V \to V$ be a bounded linear transformation. The *operator norm* of T is defined as

$$|T|_V = \sup_{v \in V, v \neq 0} \{|T(v)|_V / |v|_V\};$$

the fact that this is finite is precisely the condition that T be bounded.

The operator norm depends strongly on the norm on V (although the property of being bounded only depends on the equivalence class of the norm). The spectral norm is somewhat less delicate.

DEFINITION 5.1.2. With notation as above, the *spectral norm* of V is defined as

$$|T|_{\mathrm{sp},V} = \lim_{s \to \infty} |T^s|_V^{1/s};$$

the existence of the limit follows from the fact $|T^{m+n}|_V \leq |T^m|_V |T^n|_V$ and the following lemma.

LEMMA 5.1.3 (Fekete). *Let $\{a_n\}_{n=1}^\infty$ be a sequence of real numbers such that $a_{m+n} \geq a_m + a_n$ for all m, n. Then the sequence $\{a_n/n\}_{n=1}^\infty$ either converges to its supremum or diverges to $+\infty$.*

PROOF. Exercise. □

PROPOSITION 5.1.4. *With notation as above, the spectral norm of T depends on the norm $|\cdot|_V$ only up to equivalence.*

PROOF. Suppose $|\cdot|'_V$ is an equivalent norm. We can then choose $c > 0$ such that $|v|'_V \leq c|v|_V$ and $|v|_V \leq c|v|'_V$ for all $v \in V$. We then have $|T(v)|_V/|v|_V \leq c^2 |T(v)|'_V/|v|'_V$ for all $v \in V - \{0\}$. Applying this with T replaced by T^s, this gives $|T^s|_V \leq c^2 |T^s|'_V$, so

$$|T^s|_{\mathrm{sp},V} \leq \lim_{s \to \infty} c^{2/s} (|T^s|'_{\mathrm{sp},V})^{1/s}.$$

Since $c^{2/s} \to 1$ as $s \to \infty$, this gives $|T^s|_{\mathrm{sp},V} \leq |T^s|'_{\mathrm{sp},V}$. The reverse inequality holds by reversing the roles of the norms. □

Remark 5.1.5. Suppose that V is finite dimensional. Pick a basis for V, and equip V with either the L_2 norm or the supremum norm defined by this basis, according as whether F is archimedean or nonarchimedean. Let A be the matrix via which T acts on this basis. Then $|T|_V$ equals the largest singular value of A, whereas $|T|_{\mathrm{sp},V}$ equals the largest norm of an eigenvalue of A.

2. Spectral norms of differential operators

Definition 5.2.1. By a *nonarchimedean differential ring/field*, we mean a nonarchimedean ring equipped with a bounded derivation. For F a nonarchimedean differential field, we can define the operator norm $|d|_F$ and the spectral norm $|d|_{\mathrm{sp},F}$; by hypothesis the former is finite, so the latter is too.

Definition 5.2.2. Let F be a nonarchimedean differential field. By a *normed differential module* over F, we mean a vector space V over F equipped with a norm $|\cdot|_V$ compatible with $|\cdot|_F$, and a derivation D with respect to d which is bounded as an operator on V. Since D is linear over the constant subfield of F, we may consider the operator norm $|D|_V$ and the spectral norm $|D|_{\mathrm{sp},V}$.

Remark 5.2.3. If V is finite dimensional over F and F is complete, then the spectral norm does not depend on the norm on V, since by Theorem 1.3.3 any two norms on V compatible with the norm on F are equivalent.

Lemma 5.2.4. *Let F be a nonarchimedean differential field and let V be a normed differential module over F. Then*

$$|D|_{\mathrm{sp},V} \geq |d|_{\mathrm{sp},F}.$$

Proof. (This proof was suggested by Liang Xiao.) For $a \in F$ and $v \in V$ nonzero, the Leibniz rule gives

$$D^{s-i}(aD^i(v)) = d^{s-i}(a)D^i(v) + \sum_{j=0}^{s-i} \binom{s-i}{j} d^{s-i-j}(a)D^{i+j}(v) \qquad (0 \leq i \leq s).$$

Inverting this system of equations gives an identity of the form

$$d^s(a)v = \sum_{i=0}^{s} c_{s,i} D^{s-i}(aD^i(v))$$

for certain universal constants $c_{s,i} \in \mathbb{Z}$. Consequently,

$$|d^s(a)v|_V \leq \max_{0 \leq i \leq s}\{|D^{s-i}(aD^i(v))|\}. \qquad (5.2.4.1)$$

Given $\epsilon > 0$, we can choose $c = c(\epsilon)$ such that for all $s \geq 0$,

$$|D^s|_V \leq c(|D|_{\mathrm{sp},V} + \epsilon)^s.$$

(The c is only needed to cover small s.) Using (5.2.4.1), we deduce

$$|d^s(a)v|_F \leq c^2(|D|_{\mathrm{sp},V} + \epsilon)^s|v|.$$

Dividing by $|v|_V$ and taking the supremum over $a \in F$, we obtain

$$|d^s|_F \leq c^2(|D|_{\mathrm{sp},V} + \epsilon)^s.$$

Extracting an s-th root and taking limits, we get

$$|d|_{\mathrm{sp},F} \leq |D|_{\mathrm{sp},V} + \epsilon.$$

Since $\epsilon > 0$ was arbitrary, this yields the claim. $\square$

In some cases, it may be useful to compute in terms of a basis of V over F.

LEMMA 5.2.5. *Let F be a complete nonarchimedean differential field, and let V be a finite differential module over F. Fix a basis $e_1, \dots, e_n$ of V, and let D_s be the matrix via which D^s acts on this basis; that is, $D^s(e_j) = \sum_i (D_s)_{ij} e_i$. Then*

$$|D|_{\mathrm{sp},V} = \max\{|d|_{\mathrm{sp},F}, \limsup_{s\to\infty} |D_s|^{1/s}\}. \tag{5.2.5.1}$$

PROOF. (Compare [**CD94**, Proposition 1.3].) Equip V with the supremum norm defined by $e_1, \dots, e_n$; then $|D^s|_V \geq \max_{i,j} |(D_s)_{i,j}|$. This plus Lemma 5.2.4 implies that the left side of (5.2.5.1) is greater than or equal to the the right side.

Conversely, for any $x \in V$, if we write $x = x_1 e_1 + \cdots + x_n e_n$, then

$$D^s(x) = \sum_{i=1}^n \sum_{j=0}^s \binom{s}{j} d^j(x_i) D^{s-j}(e_i),$$

so

$$|D^s|_V^{1/s} \leq \max_{0 \leq j \leq s} \{|d^j|_F^{1/s} |D_{s-j}|^{1/s}\}. \tag{5.2.5.2}$$

Given $\epsilon > 0$, we can choose $c = c(\epsilon)$ such that for all $s \geq 0$,

$$|d^s|_F \leq c(|d|_{\mathrm{sp},F} + \epsilon)^s$$
$$|D_s| \leq c(\limsup_{s\to\infty} |D_s|^{1/s} + \epsilon)^s.$$

Then (5.2.5.2) implies

$$|D^s|_V^{1/s} \leq c^{2/s} \max\{|d|_{\mathrm{sp},F} + \epsilon, \limsup_{s\to\infty} |D_s|^{1/s} + \epsilon\}.$$

As in the previous proof, the factor $c^{2/s}$ tends to 1 as $s \to \infty$. From this it follows that the right side of (5.2.5.1) is greater than or equal to the left side minus ϵ; since $\epsilon > 0$ was arbitrary, we get the same inequality with $\epsilon = 0$. $\square$

REMARK 5.2.6. With slightly more work, one may check that in Lemma 5.2.5, if the maximum is only achieved by the second term, then you can replace the limit superior by a limit.

LEMMA 5.2.7. *Let F be a nonarchimedean differential field.*

(a) *For $0 \to V_1 \to V \to V_2 \to 0$ a short exact sequence of normed differential modules over F,*
$$|D|_{\mathrm{sp},V} = \max\{|D|_{\mathrm{sp},V_1}, |D|_{\mathrm{sp},V_2}\}.$$
(b) *For V a finite normed differential module over F,*
$$|D|_{\mathrm{sp},V^\vee} = |D|_{\mathrm{sp},V}.$$

(c) *For V_1, V_2 normed differential modules over F,*

$$|D|_{\mathrm{sp},V_1 \otimes V_2} \leq \max\{|D|_{\mathrm{sp},V_1}, |D|_{\mathrm{sp},V_2}\},$$

with equality when $|D|_{\mathrm{sp},V_1} \neq |D|_{\mathrm{sp},V_2}$.

PROOF. Everything is straightforward except perhaps the last assertion of (c); we explain how to deduce it from everything else.

Suppose $|D|_{\mathrm{sp},V_1} > |D|_{\mathrm{sp},V_2}$. Then by (b) and the first assertion of (c),

$$\begin{aligned} |D|_{\mathrm{sp},V_1} &= \max\{|D|_{\mathrm{sp},V_1}, |D|_{\mathrm{sp},V_2}\} \\ &\geq \max\{|D|_{\mathrm{sp},V_1 \otimes V_2}, |D|_{\mathrm{sp},V_2^\vee}\} \\ &\geq |D|_{\mathrm{sp},V_1 \otimes V_2 \otimes V_2^\vee}. \end{aligned}$$

Moreover, $V_2 \otimes V_2^\vee$ contains a trivial submodule (the trace), so $V_1 \otimes V_2 \otimes V_2^\vee$ contains a copy of V_1. Hence by (a), $|D|_{\mathrm{sp},V_1 \otimes V_2 \otimes V_2^\vee} \geq |D|_{\mathrm{sp},V_1}$. We thus obtain a chain of inequalities leading to $|D|_{\mathrm{sp},V_1} \geq |D|_{\mathrm{sp},V_1}$; this forces the intermediate equality $|D|_{\mathrm{sp},V_1} = \max\{|D|_{\mathrm{sp},V_1 \otimes V_2}, |D|_{\mathrm{sp},V_2^\vee}\}$. Since $|D|_{\mathrm{sp},V_1} \neq |D|_{\mathrm{sp},V_2} = |D|_{\mathrm{sp},V_2^\vee}$, we can only have $|D|_{\mathrm{sp},V_1} = |D|_{\mathrm{sp},V_1 \otimes V_2}$, as desired. □

COROLLARY 5.2.8. *If V_1, V_2 are irreducible normed differential modules over a nonarchimedean differential field, and $|D|_{\mathrm{sp},V_1} \neq |D|_{\mathrm{sp},V_2}$, then every irreducible submodule W of $V_1 \otimes V_2$ satisfies $|D|_{\mathrm{sp},W} = \max\{|D|_{\mathrm{sp},V_1}, |D|_{\mathrm{sp},V_2}\}$.*

There might be a simple proof improving this to cover irreducible subquotients of $V_1 \otimes V_2$, but I don't know of one. I'll deduce something slightly weaker later (Corollary 5.6.3).

PROOF. Suppose the contrary; we may assume that $|D|_{\mathrm{sp},V_1} > |D|_{\mathrm{sp},V_2}$. The inclusion $W \hookrightarrow V_1 \otimes V_2$ corresponds to a nonzero horizontal section of $W^\vee \otimes V_1 \otimes V_2 \cong (W \otimes V_2^\vee)^\vee \otimes V_1$, which in turn corresponds to a nonzero map $W \otimes V_2^\vee \to V_1$. Since V_1 is irreducible, the map has image V_1; that is, $W \otimes V_2^\vee$ has a quotient isomorphic to V_1.

However, we can contradict this using Lemma 5.2.7. Namely,

$$|D|_{\mathrm{sp},W \otimes V_2^\vee} \leq \max\{|D|_{\mathrm{sp},W}, |D|_{\mathrm{sp},V_2}\} < |D|_{\mathrm{sp},V_1},$$

so each nonzero subquotient of $W \otimes V_2^\vee$ has spectral norm strictly less than $|D|_{\mathrm{sp},V_1}$. □

REMARK 5.2.9. By contrast, when $|D|_{\mathrm{sp},V_1} = |D|_{\mathrm{sp},V_2}$, it is entirely possible for an irreducible submodule W of $V_1 \otimes V_2$ to satisfy $|D|_{\mathrm{sp},W} \neq \max\{|D|_{\mathrm{sp},V_1}, |D|_{\mathrm{sp},V_2}\}$. For instance, take V_1 with $|D|_{\mathrm{sp},V_1} > |d|_{\mathrm{sp},F}$ put $V_2 = V_1^\vee$, and let W be the trace component of $V_1 \otimes V_1^\vee$.

DEFINITION 5.2.10. For V a finite differential module over a nonarchimedean differential field F, let $V_1, \dots, V_l$ be the Jordan-Hölder constituents of V (i.e., the successive quotients in a filtration of V of maximal length; the list of these is unique up to reordering). Define the *full spectrum* of V to be the multiset consisting of $|D|_{\mathrm{sp},V_i}$ with multiplicity $\dim_F V_i$, for $i = 1, \dots, l$.

We will need the following differential version of Proposition 3.4.6 later.

PROPOSITION 5.2.11. *Let F be a complete normed differential field with $|d|_F \leq 1$. Let V be a finite differential module of rank n over F with $|D|_{\mathrm{sp},V} \leq 1$. Fix a norm $|\cdot|_V$ on V, given as the supremum norm for some basis $e_1, \dots, e_n$, for which $|D|_V = c \geq 1$. Then there exists a basis $v_1, \dots, v_n$ of V defining a second supremum norm $|\cdot|'_V$, for which $|D|'_V \leq 1$ and $|x|'_V \leq |x|_V \leq c^{n-1}|x|'_V$ for all $x \in V$.*

PROOF. Let N be the matrix given by $D(e_j) = \sum_i N_{ij} e_i$. By Proposition 3.4.6, there exists $U \in \mathrm{GL}_n(F)$ such that

$$|U^{-1}NU| \leq 1, \quad |U^{-1}| \leq 1, \quad |U| \leq c^{n-1}.$$

By Theorem 3.3.4, we may factor $U = W\Delta X$ with Δ diagonal and $W, X \in \mathrm{GL}_n(\mathfrak{o}_F)$. By changing the original basis $e_1, \dots, e_n$ over $\mathfrak{o}_F$ (so as not to change the original norm $|\cdot|_V$), we can reduce to the case $W = I_n$. We then define $|\cdot|'_V$ as the supremum norm defined by the new basis $\Delta_{11}e_1, \dots, \Delta_{nn}e_n$. This satisfies $|x|'_V \leq |x|_V \leq c^{n-1}|x|'_V$ for all $x \in V$ because $1 \leq |\Delta_{ii}| \leq c^{n-1}$ for all i.

To check $|D|'_V \leq 1$, we note that the matrix of D in the new basis is $\Delta^{-1}N\Delta + \Delta^{-1}d(\Delta)$. Since $\Delta^{-1}N\Delta = X(U^{-1}NU)X^{-1}$ and $X \in \mathrm{GL}_n(\mathfrak{o}_F)$, we have $|\Delta^{-1}N\Delta| \leq 1$. Since $\Delta^{-1}d(\Delta)$ is a diagonal matrix with ii-entry $d(\Delta_{ii})/\Delta_{ii}$, and $|d|_F \leq 1$, we have $|\Delta^{-1}d(\Delta)| \leq 1$. This proves the claim. □

3. A coordinate-free approach

I mention in passing the following more coordinate-free approach to defining the spectral norm; in particular, there is no need to explicitly truncate when using this method.

PROPOSITION 5.3.1 (Baldassarri-di Vizio). *Let F be a nonarchimedean differential field of characteristic 0 with d nontrivial; put $F_0 = \ker(d)$. Let $F\{T\}^{(s)}$ be the set of twisted polynomials of degree at most s; define the norm of $P \in F\{T\}^{(s)}$ as $|P(d)|_F$ (that is, consider $P(d)$ as an operator on F). Let V be a finite differential module over F, and fix a norm on V compatible with $|\cdot|$. Let $L_{F_0}(V)$ be the space of bounded F_0-linear endomorphisms of V, equipped with the operator norm. Let $D_s : F\{T\}^{(s)} \to L_{F_0}(V)$ be the map $P \mapsto P(D)$. Then*

$$|D|_{\mathrm{sp},V} = |d|_{\mathrm{sp},F} \lim_{s \to \infty} |D_s|^{1/s}. \tag{5.3.1.1}$$

PROOF. We have $|D|_{\mathrm{sp},V} \leq |d|_{\mathrm{sp},F} \liminf_{s\to\infty} |D_s|^{1/s}$ because on one hand $|D^s|_V \leq |d^s|_F |D_s|$ by taking $T^s \in F\{T\}^{(s)}$, and on the other hand $\liminf |D_s|^{1/s} \geq 1$ because $1 \in F\{T\}^{(n)}$. In the other direction, we may prove $|D|_{\mathrm{sp},V} \geq |d|_{\mathrm{sp},F} \limsup_{s\to\infty} |D_s|^{1/s}$ by imitating the proof of Lemma 5.2.5. □

4. Newton polygons for twisted polynomials

Twisted polynomials admit a partial analogue of the theory of Newton polygons.

DEFINITION 5.4.1. Let F be a nonarchimedean differential field. For $\rho \geq |d|_F$, define the *ρ-Gauss norm* on the twisted polynomial ring $F\{T\}$ by

$$\left|\sum_i P_i T^i\right| = \max_i\{|P_i|\rho^i\}.$$

For $r \leq -\log|d|_F$, we obtain a corresponding *r-Gauss valuation* $v_r(P) = -\log|P|_{e^{-r}}$.

LEMMA 5.4.2. *For $\rho \geq |d|_F$, the ρ-Gauss norm is multiplicative. Moreover, any polynomial and its formal adjoint have the same ρ-Gauss norm.*

Proof. It suffices to check for $\rho > |d|_F$, as the boundary case may be inferred from continuity of the map $\rho \mapsto |P|_\rho$ for fixed P. The key observation (and the source of the restriction on ρ) is that for $P, Q \in F\{T\}$ and $\rho > |d|_F$,

$$|PQ - QP|_\rho \geq \rho^{-1}|d|_F|P|_\rho|Q|_\rho > |P|_\rho|Q|_\rho.$$

This allows us to deduce multiplicativity on $F\{T\}$ from multiplicativity on $F[T]$. The claim about the adjoint follows similarly. □

Definition 5.4.3. We define the *Newton polygon* of $P = \sum_i P_i T^i \in F\{T\}$ by taking the Newton polygon of the corresponding untwisted polynomial $\sum P_i T^i \in F[T]$, then omitting all slopes greater than or equal to $-\log|d|_F$; this has the usual properties thanks to Lemma 5.4.2. (Note that we cannot include the slope $-\log|d|_F$ itself because we cannot relate the width of PQ under the corresponding Gauss norm to the width of P plus the width of Q.)

As another application of the master factorization theorem (Theorem 2.2.2), we obtain the following.

Theorem 5.4.4. *Let F be a complete nonarchimedean differential field. Suppose $S \in F\{T\}$, $r < -\log|d|_F$, and $m \in \mathbb{Z}_{\geq 0}$ satisfy*

$$v_r(S - T^m) > v_r(T^m).$$

Then there exists a unique factorization $S = PQ$ satisfying the following conditions.

(a) *The polynomal $P \in F\{T\}$ has degree $\deg(S) - m$, and its slopes are all less than r.*
(b) *The polynomial $Q \in F\{T\}$ is monic of degree m, and its slopes are all greater than r.*
(c) *We have $v_r(P - 1) > 0$ and $v_r(Q - T^m) > v_r(T^m)$.*

Moreover, for this factorization,

$$\min\{v_r(P-1), v_r(Q - T^m) - v_r(T^m)\} \geq v_r(S - T^m) - v_r(T^m).$$

In addition, we have the same result if we ask for the factorization in the order $S = QP$ (but the factors themselves may differ).

Proof. The same setup works as in Theorem 2.2.1. □

Corollary 5.4.5. *If $P \in F\{T\}$ is irreducible, then either it has no slopes, or it has all slopes equal to some value less than $-\log|d|_F$.*

5. Twisted polynomials and spectral norms

One can use twisted polynomials over nonarchimedean differential fields to detect only part of the full spectrum of a normed differential module.

Definition 5.5.1. For V a finite differential module over a nonarchimedean differential field F, define the *visible spectrum* of V to be the submultiset of the full spectrum of V consisting of those values greater than $|d|_F$.

REMARK 5.5.2. In the application to regular singularities, we will consider a case where $|d|_F = |d|_{\mathrm{sp},F}$, in which case there is no real loss in restricting to the visible spectrum: the only missing norm is $|d|_F$ itself, and one can infer its multiplicity from the dimension of the module. However, in the applications to p-adic differential equations, we will have $|d|_F > |d|_{\mathrm{sp},F}$, so the restriction to the visible spectrum will cause real problems; these will have to be remedied using pullback and pushforward along a Frobenius map.

THEOREM 5.5.3 (Christol-Dwork). *Let F be a complete nonarchimedean differential field. For $P \in F\{T\}$, put $V = F\{T\}/F\{T\}P$. Let r be the least slope of the Newton polygon of P, or $-\log |d|_F$ if no such slope exists. Then*

$$\max\{|d|_F, |D|_{\mathrm{sp},V}\} = e^{-r}.$$

PROOF. Let $r_1 \leq \cdots \leq r_k$ be the slopes of P, and define $r_{k+1} = \cdots = r_n = -\log |d|_F$. Equip V with the norm

$$\left| \sum_{i=0}^{n-1} a_i T^i \right|_V = \max_i \{|a_i| e^{-r_{n-1} - \cdots - r_{n-i}}\}.$$

As in the proof of Proposition 3.3.10, we then have $|D|_V = e^{-r_1}$, and so $|D|_{\mathrm{sp},V} \leq e^{-r_1}$.

To finish, we must check that if $r_1 < -\log |d|_F$, then $|D|_{\mathrm{sp},V} = e^{-r_1}$. Let δ be the operation

$$\delta \left(\sum_{i=0}^{n-1} a_i T^i \right) = \sum_{i=0}^{n-1} d(a_i) T^i;$$

then $|\delta|_V = |d|_F$, $D - \delta$ is F-linear, and $|D - \delta|_V = |D - \delta|_{\mathrm{sp},V} = e^{-r_1}$. Then for all positive integers s,

$$|(D - \delta)^s|_V = e^{-r_1 s}, \qquad |D^s - (D - \delta)^s|_V \leq e^{-r_1(s-1)} |d|_F < e^{-r_1 s},$$

so $|D^s|_V = e^{-r_1 s}$ and $|D|_{\mathrm{sp},V} = e^{r_1}$ as desired. $\square$

COROLLARY 5.5.4. *Let F be a complete nonarchimedean differential field. For any $P \in F\{T\}$, the visible spectrum of the differential module $F\{T\}/F\{T\}P$ consists of e^{-r} for r running over the slope multiset of the Newton polygon of P.*

PROOF. Write down a maximal factorization of P; it corresponds to a maximal filtration of $F\{T\}/F\{T\}P$. By Corollary 5.4.5, each factor in the factorization has only a single slope, so Theorem 5.5.3 gives what we want. $\square$

6. The visible decomposition theorem

Using twisted polynomials, we can split V into components corresponding to the elements of the visible spectrum.

THEOREM 5.6.1 (Visible decomposition theorem). *Let F be a complete nonarchimedean differential field of characteristic zero with nontrivial derivation, and let V be a finite dimensional differential module over F. Then there exists a decomposition*

$$V = V_0 \oplus \bigoplus_{s > |d|_F} V_s$$

of differential modules, such that every subquotient of V_s has spectral norm s, and every subquotient of V_0 has spectral norm at most $|d|_F$.

PROOF. We induct on $\dim(V)$. Choose a cyclic vector for V (possible by Theorem 4.4.2), because of the hypotheses we imposed on F), yielding an isomorphism $V \cong F\{T\}/F\{T\}P$. Let r be the least slope of P. If $r \geq -\log |d|_F$, we may put $V = V_0$ and be done, so assume $r < -\log |d|_F$. By applying Theorem 5.4.4 once to P, we obtain a short exact sequence $0 \to V_1 \to V \to V_2 \to 0$ in which (by Theorem 5.5.3) every subquotient of V_1 has spectral norm e^{-r}, and every subquotient of V_2 has spectral norm less than e^{-r}. Applying Theorem 5.4.4 again to P but with the factors in the opposite order, we get a short exact sequence $0 \to V_2' \to V \to V_1' \to 0$ where every subquotient of V_1' has spectral norm e^{-r}, and every subquotient of V_2' has spectral norm less than e^{-r}. Moreover, $\dim V_1 = \dim V_1'$ and $\dim V_2 = \dim V_2'$ because P and its formal adjoint have the same Newton polygon (Lemma 5.4.2). Consequently, $V_1 \cap V_2' = 0$, so $V_1 \oplus V_2'$ injects into V; by counting dimensions, this must be an isomorphism. This lets us split $V \cong V_1 \oplus V_2$, and we may apply the induction hypothesis to V_2 to get what we want. $\square$

COROLLARY 5.6.2. *Let F be a complete nonarchimedean differential field, and let V be a finite dimensional differential module over F such that every subquotient of V has spectral norm greater than $|d|_F$. Then $H^0(V) = H^1(V) = 0$.*

PROOF. The claim about H^0 is clear: a nonzero element of $H^0(V)$ would generate a differential submodule of V which would be trivial, and thus would have spectral norm $|d|_{\mathrm{sp},F} \leq |d|_F$. As for H^1, let $0 \to V \to W \to F \to 0$ be a short exact sequence of differential modules. Decompose $W = W_0 \oplus W_1$ according to Theorem 5.6.1, with every subquotient of W_0 having spectral norm at most $|d|_F$, and every subquotient of W_1 having spectral norm greater than $|d|_F$. The map $V \to W_0$ must vanish (its image is a subquotient of both V and W_0), so $V \subseteq W_1$. But $W_1 \neq W$ as otherwise W could not surject onto a trivial module, so $V = W_1$. Hence the sequence splits, proving $H^1(V) = 0$. $\square$

COROLLARY 5.6.3. *If V_1, V_2 are irreducible, $|D|_{\mathrm{sp},V_1} > |d|_F$, and $|D|_{\mathrm{sp},V_1} > |D|_{\mathrm{sp},V_2}$, then every irreducible subquotient W of $V_1 \otimes V_2$ satisfies $|D|_{\mathrm{sp},W} = |D|_{\mathrm{sp},V_1}$.*

PROOF. Decompose $V_1 \otimes V_2 = V_0 \oplus \bigoplus_{s>|d|_F} V_s$ according to Theorem 5.6.1; we have $V_s = 0$ whenever $s > |D|_{\mathrm{sp},V_1}$. If either V_0 or some V_s with $s < |D|_{\mathrm{sp},V_1}$ were nonzero, then $V_1 \otimes V_2$ would have an irreducible submodule of spectral norm less than $|D|_{\mathrm{sp},V_1}$, in violation of Corollary 5.2.8. $\square$

For the study of irregularity, these results are quite sufficient. However, in the p-adic situation, we will have to do better than this in order to further decompose V_0; we will do this using Frobenius antecedents in Chapter 9.

7. Matrices and the visible spectrum

The proof of Theorem 5.5.3 relies on the fact that one can detect the spectral norm of a differential module admitting a cyclic vector, using the characteristic polynomial of the matrix of the action of D on the cyclic basis. For some applications, we need to extend this to some bases not necessarily generated by cyclic vectors; for this, the relationship between singular values and eigenvalues will be crucial.

We state the following lemma over a differential domain rather than a differential field, so that we can use it again later.

LEMMA 5.7.1. *Let R be a complete nonarchimedean differential domain with fraction field F. Let N be a 2×2 block matrix over R with the following properties.*

(a) *The matrix N_{11} has an inverse A over R.*
(b) *We have $|A| \max\{|d|_F, |N_{12}|, |N_{21}|, |N_{22}|\} < 1$.*

Then there exists a block upper triangular unipotent matrix U over R such that $|U_{12}| \leq |A| \max\{|N_{12}|, |N_{21}|, |N_{22}|\}$ and $U^{-1}NU + U^{-1}d(U)$ is block lower triangular.

PROOF. Put

$$\delta = |A| \max\{|N_{12}|, |N_{21}|, |N_{22}|\} < 1, \qquad \epsilon = |AN_{12}| \leq \delta.$$

Let X be the block upper triangular nilpotent matrix with $X_{12} = AN_{12}$, and put $U = I - X$ and

$$N' = U^{-1}NU + U^{-1}d(U).$$

Since $U^{-1} = I + X$, we have $N' = N + XN - NX - XNX - d(X)$. In block form,

$$N' = \begin{pmatrix} N_{11} + X_{12}N_{21} & N_{12} - N_{11}X_{12} + X_{12}N_{22} - X_{12}N_{21}X_{12} + d(X_{12}) \\ N_{21} & N_{22} - N_{21}X_{12} \end{pmatrix}.$$

We claim that

$$\begin{aligned} |N'_{12}| &\leq \epsilon \max\{\delta, |d|_F |A|\} |A|^{-1} \\ |N'_{21}| &\leq \delta |A|^{-1} \\ |N'_{22}| &\leq \delta |A|^{-1}. \end{aligned}$$

The second and third lines hold because

$$|U^{-1}NU - N| = |XN - NX - XNX| \leq \epsilon |A|^{-1}.$$

The first line holds because we can write

$$N'_{22} = X_{12}N_{22} - X_{12}N_{21}X_{12} + d(X_{12}),$$

in which the first two terms have norm at most $\epsilon\delta|A|^{-1}$ and the third has norm at most $|d|_F \epsilon$.

To analyze N'_{11}, we write it as $(I + X_{12}N_{21}A)N_{11}$. Because $|X_{12}N_{21}A| \leq \epsilon < 1$, the first factor is invertible, and it and its inverse both have norm 1. Hence N'_{11} is invertible, $|N'_{11}| = |N_{11}|$, and $|(N'_11)^{-1}| = |A|$.

Since $\epsilon \max\{\delta, |d|_F |A|^{-1}\} < \epsilon$, iterating the construction $N \mapsto N'$ yields obtain a convergent sequence of conjugations whose limit has the desired property. □

We need a refinement of the argument used in Theorem 5.5.3.

LEMMA 5.7.2. *Let F be a complete nonarchimedean differential field. Let V be a finite differential module over F. Let $e_1, \dots, e_n$ be a basis of V, and let N be the matrix of action of D on $e_1, \dots, e_n$. Suppose that $|N| = \sigma > |d|_F$ and $|N^{-1}| = \sigma^{-1}$. Then the full spectrum of V consists entirely of σ.*

PROOF. As in the proof of Theorem 5.5.3, we find that for the supremum norm for $e_1, \dots, e_n$, we have $|D^s w| = \sigma^s |w|$ for all nonnegative integers s. Consequently, for any nonzero differential submodule W of V, we have $|D|_{\mathrm{sp},V} = \sigma$. By Theorem 5.6.1, it follows that every irreducible subquotient of V also has spectral norm σ, as desired. $\square$

LEMMA 5.7.3. *Let F be a complete nonarchimedean differential field. Let V be a finite differential module over F. Let $e_1, \dots, e_n$ be a basis of V, and let N be the matrix of action of D on $e_1, \dots, e_n$. Let $\sigma_1, \dots, \sigma_n$ be the singular values of N and let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of N. Suppose that the following conditions hold for some $i = 1, \dots, n$ and some $\delta \geq |d|_F$.*

(a) *We have $\sigma_i > \delta$.*
(b) *Either $i = n$ or $\sigma_{i+1} \leq \delta$.*
(c) *We have $\sigma_j = |\lambda_j|$ for $j = 1, \dots, i$.*

Then the elements of the full spectrum greater than δ are precisely $\sigma_1, \dots, \sigma_i$.

PROOF. By enlarging F, we may reduce to the case where $\delta = |d|_F$ (this is purely for notational simplicity).

Note that conditions (a), (b), (c), are invariant under a conjugation

$$N \mapsto U^{-1}NU + U^{-1}d(U) = U^{-1}(N + d(U))U$$

for $U \in \mathrm{GL}_n(\mathfrak{o}_F)$, because Theorem 3.4.2 implies that N and $N + d(U)$ have the same norms of eigenvalues greater than $|d|_F$, and conjugating by U changes nothing.

If $\sigma_1 \leq |d|_F$, then we have nothing to check. If $\sigma_1 = \cdots = \sigma_n > |d|_F$, then Lemma 5.7.2 implies the claim. If neither of these cases apply, we may induct on n: choose i with $\sigma_1 = \cdots = \sigma_i > \sigma_{i+1}$, so that necessarily $\sigma_1 > |d|_F$. View N as a 2×2 block matrix with block sizes $i, n-i$. Apply Lemma 5.7.1 to obtain an upper triangular unipotent block matrix U over $\mathfrak{o}_F$ such that $N' = U^{-1}NU + U^{-1}d(U)$ is lower triangular. We may then reduce to checking the claim with N replaced by the two diagonal blocks of N'. $\square$

THEOREM 5.7.4. *Let F be a complete nonarchimedean differential field. Let V be a finite differential module over F. Let $e_1, \dots, e_n$ be a basis of V, and let N be the matrix of action of D on $e_1, \dots, e_n$. Let $\sigma_1, \dots, \sigma_n$ be the singular values of N and let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of N. Define f_n as in Corollary 3.4.8 and put $\theta = f_n(\sigma_1, \dots, \sigma_n, |\lambda_1|, \dots, |\lambda_n|)$. Suppose that the following conditions hold for some $i = 1, \dots, n$ and some $\delta \geq |d|_F \theta$.*

(a) *We have $|\lambda_i| > \delta$.*
(b) *Either $i = n$ or $|\lambda_{i+1}| \leq \delta$.*

Then the elements of the full spectrum greater than δ are precisely $|\lambda_1|, \dots, |\lambda_i|$.

PROOF. There is no harm in enlarging the constant subfield of F so that the additive value group of F becomes equal to $\mathbb{R}$. By Corollary 3.4.8, we can choose a matrix $U \in \mathrm{GL}_n(F)$ such that the following conditions hold.

(a) We have $|U| \leq 1$ and $|U^{-1}| \leq \theta$.
(b) The first i singular values of $U^{-1}NU$ are $|\lambda_1|, \dots, |\lambda_i|$.
(c) Either $i = n$, or the $(i+1)$-st singular value of $U^{-1}NU$ is at least δ.

By Theorem 3.4.2, the new conditions (b) and (c) hold when $U^{-1}NU$ is replaced by $U^{-1}NU + U^{-1}d(U)$. We may thus apply Lemma 5.7.3 to obtain the desired result. $\square$

Notes

Lemma 5.2.5 is tacitly assumed at various places in the literature (including by the present author), but we were unable to locate even an explicit statement, let alone a proof. We again thank Liang Xiao for contributing the proof given here.

Proposition 5.2.11 answers a conjecture of Christol and Dwork [**CD92**, Introduction, Conjecture A]. This conjecture was posed in the context of giving effective convergence bounds, and that is exactly how we will use it here; see Theorem 16.2.1 and its proof.

Proposition 5.3.1 is from as yet unreleased work of Baldassarri and di Vizio (a promised sequel to [**BdV07**]), which gives a development of much of the material we are discussing from the point of view of Berkovich analytic spaces. This point of view will probably be vital for the study of differential modules on higher-dimensional spaces.

Newton polygons for differential operators were considered by Dwork and Robba [**DR77**, §6.2.3]; the first systematic treatment seems to have been made by Robba [**Rob80**]. Our treatment using Theorem 2.2.2 follows [**Chr83**].

The proof of Theorem 5.5.3 given here is close to the original proof of Christol and Dwork [**CD94**, Théorème 1.5], save that we avoid a small gap in the latter. The gap is in the implication $1 \implies 2$; there one makes a finite extension of the differential field, without accounting for the possibility that this might increase $|d|_F$. (It would be obvious that this does not occur if the finite extension were being made in the constant subfield, but that is not the case here.) Compare also [**DGS94**, Lemma VI.2.1].

Exercises

(1) Prove Fekete's lemma (Lemma 5.1.3).

(2) Let A, B be commuting bounded linear operators on a normed vector space V over a nonarchimedean field F. Prove that

$$|A + B|_{\mathrm{sp},V} \leq \max\{|A|_{\mathrm{sp},V}, |B|_{\mathrm{sp},V}\},$$

and that equality occurs when the maximum is achieved only once.

(3) Let V be a normed differential module over a nonarchimedean differential field F. Prove that $|D|_V \geq |d|_F$.

CHAPTER 6

Regular singularities

As an application of the theory developed so far, we reconstruct some of the traditional Fuchsian theory of regular singular points of meromorphic differential equations. While this is assuredly not the most economical development of this theory (because we have had to invest more effort in order to be ready to handle p-adic differential equations), it does provide a simplified illustration of the use of some of the techniques we have amassed.

1. Irregularity

Definition 6.1.1. View $\mathbb{C}((z))$ as a complete nonarchimedean differential field, with the valuation given by the z-adic valuation v_z, and the derivation given by $d = z\frac{d}{dz}$; note that $|d|_{\mathbb{C}((z))} = 1$. Let V be a finite differential module over $\mathbb{C}((z))$, and decompose V according to Theorem 5.6.1. Define the *irregularity* of V as

$$\operatorname{irr}(V) = \sum_{s>1} (-\log s) \dim(V_s).$$

For F a subfield of $\mathbb{C}((z))$ stable under d, and V a finite differential module over F, we define the irregularity of V to be the irregularity of $V \otimes_F \mathbb{C}((z))$. We say that V is *regular* if $\operatorname{irr}(V) = 0$.

Theorem 6.1.2. *For any isomorphism $V \cong F\{T\}/F\{T\}P$, the irregularity of V is equal to the sum of the negations of the slopes of P; consequently, it is always an integer. More explicitly, if $P = T^d + \sum_{i=0}^{d-1} P_i T^i$, then*

$$\operatorname{irr}(V) = \max_i \{-v_z(P_i)\}.$$

Proof. Note that V admits a cyclic vector by Theorem 4.4.2, so the criterion in the theorem always applies. □

Corollary 6.1.3. *Let F be any subfield of $\mathbb{C}((z))$ containing z and stable under d, and let V be a finite differential module over F. Then the following conditions are equivalent.*

(a) *The module V is regular, i.e.,* $\operatorname{irr}(V) = 0$.
(b) *For some isomorphism $V \cong F\{T\}/F\{T\}P$ with P monic, P has coefficients in $\mathfrak{o}_F$.*
(c) *For* any *isomorphism $V \cong F\{T\}/F\{T\}P$ with P monic, P has coefficients in $\mathfrak{o}_F$.*
(d) *There exists a basis of V on which D acts via a matrix over $\mathfrak{o}_F$.*

Proof. By Theorem 6.1.2, (a) implies (c). It is obvious that (c) implies (b), and that (b) implies (d). Given (d), let $|\cdot|_V$ be the supremum norm defined by the chosen basis of V; then $|D|_V \leq 1$, which implies (a). □

Remark 6.1.4. One can also view $\mathbb{C}((z))$ as a differential field with the derivation $\frac{d}{dz}$ instead of $z\frac{d}{dz}$. The categories of differential modules for these two choices of derivation are equivalent in the obvious fashion: given an action of $z\frac{d}{dz}$, we obtain an action of $\frac{d}{dz}$ by dividing by z. If V is a differential module for $z\frac{d}{dz}$ with spectral norm $s > 1$, then the spectral norm of V for $\frac{d}{dz}$ is $s|z|^{-1}$. The notion of irregularity naturally translates over: for instance, if V is a differential module for $\frac{d}{dz}$ isomorphic to $F\{T\}/F\{T\}P$ for some $P = T^n + \sum_{i=0}^{n-1} P_i T^i$, then V is regular if and only if $v_z(P_i) \geq -n + i$ for $i = 1, \dots, n$. For example, for $a, b \in \mathbb{C}$, the differential system corresponding to the hypergeometric differential equation

$$y'' + \frac{(c - (a+b+1)z)}{z(1-z)} y' - \frac{ab}{z(1-z)} y = 0$$

is regular.

2. Exponents in the complex analytic setting

To see why regular singularities are so important in the complex analytic setting (by way of motivation for our p-adic studies), let us consider the monodromy transformation. First, we recall a familiar fact.

Theorem 6.2.1. *Fix $\rho > 0$, and let $R \subset \mathbb{C}[\![z]\!]$ be the ring of power series convergent for $|z| < \rho$. Let N be an $n \times n$ matrix over R. Then the differential system $D(v) = Nv + \frac{d}{dz}(v)$ has a basis of horizontal sections.*

Proof. This can be deduced from the fundamental theorem of ordinary differential equations; however, it will be useful for future reference to give a slightly more detailed explanation.

Note that there exists a unique $n \times n$ matrix U over $\mathbb{C}[\![z]\!]$ such that $U \equiv I_n \pmod{z}$ and $NU + \frac{d}{dz}(U) = 0$; this follows by writing $U = \sum_{i=0}^{\infty} U_i z^i$ and rewriting the equation $NU + \frac{d}{dz} U = 0$ as a recurrence

$$(i+1) U_{i+1} = \sum_{j=0}^{i} N_j U_{i-j} \qquad (i = 0, 1, \dots).$$

An argument of Cauchy [**DGS94**, Appendix III] shows that this series converges in a disc of positive radius.

We now know that any differential system on an open disc admits a basis of horizontal sections on a possibly smaller disc with the same center. Since an open disc is simply connected, and it can be covered with open subsets on which we have a basis of horizontal sections, we obtain a basis of horizontal sections over the entire disc. □

Remark 6.2.2. In the p-adic setting, we will see that the first step of the proof of Theorem 6.2.1 remains valid, but there is no analogue of the second step (analytic continuation), and indeed the whole conclusion becomes false.

Let us now consider a punctured disc and look at monodromy.

Definition 6.2.3. Let $\mathbb{C}\{z\}$ be the subfield of $\mathbb{C}((z))$ consisting of those formal Laurent series which represent meromorphic functions on some neighborhood of $z = 0$ (the choice of the neighborhood may vary with the series). Let V be a finite differential module over $\mathbb{C}\{z\}$;

choose a basis of V and let N be the action of D on this basis. On some disc centered at $z = 0$, the entries of N are meromorphic with no poles away from $z = 0$. On any subdisc not containing 0, by Theorem 6.2.1 we obtain a basis of horizontal sections. If we start with a basis of horizontal sections in a neighborhood of some point away from 0, then analytically continue around a circle proceeding once counterclockwise around the origin, we end up with a new basis of local horizontal sections. The linear transformation from the old basis to the new is called the *monodromy transformation* of V (or its associated differential system). The *exponents* of V are defined (modulo translation by $\mathbb{Z}$) to be the multiset of numbers $\alpha_1, \dots, \alpha_n$ for which $e^{\alpha_1}, \dots, e^{\alpha_n}$ are the eigenvalues of the monodromy transformation.

The monodromy transformation controls our ability to construct global horizontal sections, by the following statement whose proof is evident.

Proposition 6.2.4. *In Definition 6.2.3, any fixed vector under the monodromy transformation corresponds to a horizontal section defined on some punctured disc, rather than the universal covering space of a punctured disc. As a result, the monodromy transformation is unipotent (i.e., the exponents are all zero) if and only if there exists a basis on which D acts via a nilpotent matrix.*

Definition 6.2.5. In Definition 6.2.3, we say that V is *quasi-unipotent* if its exponents are rational; equivalently, V becomes unipotent after pulling back along $z \mapsto z^m$ for some positive integer m. This situation arises in examples "coming from geometry" (i.e., Picard-Fuchs modules), in a sense that we will discuss later.

The relationship between the properties of the monodromy transformation and the existence of horizontal sections of the differential module begs the question: is it possible to extract the monodromy transformation for a differential module, whose definition is purely analytic, from the algebraic data that defines the differential system? In fact, this is only really possible in the case of a regular module; we will see how to do this in the next section.

3. Formal solutions of regular differential equations

Definition 6.3.1. Let K be a field of characteristic 0. Let $N = \sum_{i=0}^{\infty} N_i t^i$ be an $n \times n$ matrix with entries in $K[\![z]\!]$. A *fundamental solution matrix* for N is an $n \times n$ matrix U with $U \equiv I_n \pmod{z}$ such that $U^{-1}NU + U^{-1}z\frac{d}{dz}U = N_0$.

Remark 6.3.2. Note that if U is a fundamental solution matrix for N, then

$$\begin{aligned} U^T N^T U^{-T} + U^T z \frac{d}{dz} U^{-T} &= U^T N^T U^{-T} - U^T U^{-T} (z \frac{d}{dz} U^T) U^{-T} \\ &= U^T N^T U^{-T} - (z \frac{d}{dz} U^T) U^{-T} \\ &= N_0^T. \end{aligned}$$

That is, U^{-T} is a fundamental solution matrix for N^T. Consequently, by proving a general result about U, we also obtain a corresponding result for U^{-T}, and hence for U^{-1}.

To specify when a fundamental solution matrix exists, we need the following definition.

Definition 6.3.3. We say that a square matrix N with entries in a field of characteristic zero has *prepared eigenvalues* if the eigenvalues $\lambda_1, \dots, \lambda_n$ of N satisfy the following conditions:

$$\lambda_i \in \mathbb{Z} \Leftrightarrow \lambda_i = 0$$
$$\lambda_i - \lambda_j \in \mathbb{Z} \Leftrightarrow \lambda_i = \lambda_j.$$

If only the second condition holds, we say that N has *weakly prepared eigenvalues*.

Proposition 6.3.4. *Let K be a field of characteristic 0. Let $N = \sum_{i=0}^{\infty} N_i t^i$ be an $n \times n$ matrix with entries in $K[\![z]\!]$ such that N_0 has weakly prepared eigenvalues $\lambda_1, \dots, \lambda_n$. Then N admits a unique fundamental solution matrix.*

Proof. Rewrite the defining equation as $NU + z\frac{d}{dz}U = UN_0$, then expand $U = \sum_{i=0}^{\infty} U_i t^i$ and write the new defining equation as a recurrence:

$$iU_i = U_i N_0 - N_0 U_i - \sum_{j=1}^{i} N_j U_{i-j} \qquad (i > 0). \tag{6.3.4.1}$$

Viewing the map $X \mapsto XN_0 - N_0 X$ as a linear transformation on the space of $n \times n$ matrices over F, we see that its eigenvalues are the differences $\lambda_j - \lambda_k$ for $j, k = 1, \dots, n$. Likewise, the eigenvalues of $X \mapsto iX - XN_0 + N_0 X$ are $i - \lambda_j + \lambda_k$; for i a positive integer, the condition that the λ's are weakly prepared ensures that $i - \lambda_j + \lambda_k$ cannot vanish (indeed, it cannot be an integer unless it equals i). Consequently, given N and $U_0, \dots, U_{i-1}$, there is a unique choice of U_i satisfying (6.3.4.1); this proves the desired result. $\square$

We then have the following result [**DGS94**, §III.8, Appendix II].

Theorem 6.3.5 (Fuchs). *Let $N = \sum_{i=0}^{\infty} N_i t^i$ be an $n \times n$ matrix with entries in $\mathbb{C}\{z\}$ such that N_0 has weakly prepared eigenvalues $\lambda_1, \dots, \lambda_n$. Then the fundamental solution matrix for N over $\mathbb{C}[\![z]\!]$ also has entries in $\mathbb{C}\{z\}$ (as does its inverse).*

Corollary 6.3.6. *With notation as in Theorem 6.3.5, let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of N_0. Then the eigenvalues of the monodromy transformation (of the system $D(v) = Nv + dv$) are $e^{-2\pi i \lambda_1}, \dots, e^{-2\pi i \lambda_n}$.*

Proof. In terms of a basis via which D acts via N_0, the matrix $\exp^{-N_0 \log(z)}$ provides a basis of horizontal elements. (The case $N_0 = 0$ is Theorem 6.2.1.) $\square$

Remark 6.3.7. The p-adic analogue of Theorem 6.3.5 is much more complicated; see the chapter on p-adic exponents.

In order to enforce the condition on prepared eigenvalues, we use what are classically known as *shearing transformations*.

Proposition 6.3.8 (Shearing transformations). *Let N be an $n \times n$ matrix over $\mathbb{C}[\![z]\!] \cap \mathbb{C}\{z\}$, with constant term N_0. Let α be an eigenvalue of N. Then there exists $U \in \mathrm{GL}_n(\mathbb{C}[z, z^{-1}])$ such that $U^{-1}NU + U^{-1}d(U)$ again has entries in $\mathbb{C}[\![z]\!] \cap \mathbb{C}\{z\}$, and its matrix of constant terms has the same eigenvalues as N_0 except that α has been replaced by $\alpha + 1$. The same conclusion holds with $\alpha - 1$ in place of $\alpha + 1$.*

Proof. Exercise. $\square$

Corollary 6.3.9 (Fuchs). *Let V be a regular finite differential module over $\mathbb{C}\{z\}$. Then any horizontal element of $V \otimes \mathbb{C}((z))$ belongs to V itself; that is, any formal horizontal section is convergent.*

4. Index and irregularity

Definition 6.4.1. Let F be any subfield of $\mathbb{C}((z))$ containing $\mathbb{C}(z)$, and let V be a finite differential module over F. We say V *has index* if $\dim_{\mathbb{C}} H^0(V)$ and $\dim_{\mathbb{C}} H^1(V)$ are both finite; in this case, we define the *index* of V as $\chi(V) = \dim_{\mathbb{C}} H^0(V) - \dim_{\mathbb{C}} H^1(V)$.

Proposition 6.4.2. *For any finite differential module V over $\mathbb{C}((z))$, $H^0(V) = H^1(V) < \infty$, so $\chi(V) = 0$.*

Proof. Exercise. $\square$

In the convergent case, the index carries more information.

Theorem 6.4.3. *Let V be a finite differential module over $\mathbb{C}\{z\}$. Then V has index, and $\chi(V) = -\operatorname{irr}(V)$.*

Proof. See [**Mal74**, Théorème 2.1]. $\square$

Notes

The notion of a regular singularity was introduced by Fuchs in the 19th century, as part of a classification of those differential equations with everywhere meromorphic singularities on the Riemann sphere which had algebraic solutions. Regular singularities are sometimes referred to as *Fuchsian singularities*. Much of our modern understanding of the regularity condition, especially in higher dimensions, comes from the book of Deligne [**Del70**].

The algebraic definition of irregularity is due to Malgrange [**Mal74**]; it had previously been defined in terms of the index of a certain operator. Our approach, incorporating ideas of Robba, is based on [**DGS94**, §3].

A complex analytic interpretation of the Newton polygon, in the manner of the relation between irregularity and index, has been given by Ramis [**Ram84**]. It involves considering subrings of $\mathbb{C}\{z\}$ composed of functions with certain extra convergence restrictions (*Gevrey functions*), and looking at the index of $z\,d/dz$ after tensoring the given differential module with one of these subrings.

Exercises

(1) In this exercise, we prove Fuchs's theorem (Theorem 6.3.5). Let N be an $n \times n$ matrix over $\mathbb{C}[\![z]\!]$. Let U be an $n \times n$ matrix over $\mathbb{C}[\![z]\!]$ congruent to the identity modulo z.
  (a) Show that changing basis by U in the differential system $D(v) = Nv + d(v)$ has the effect of replacing N by $N' = U^{-1}NU + U^{-1}z\frac{dU}{dz}$.
  (b) Show that $N' \equiv N \pmod{z}$.
  (c) Assume that the reduction of N modulo z has weakly prepared eigenvalues. Show that there is a unique choice of U for which N' equals the matrix of constant terms of N.

(d) Suppose that the entries of N converge in the disc $|z| < \rho$. Prove that the entries of the matrix U given in (c) also converge in the disc $|z| < \rho$.

(2) Prove Proposition 6.3.8.

(3) Prove Proposition 6.4.2.

Part 3

p-adic differential equations on discs and annuli

CHAPTER 7

Rings of functions on discs and annuli

In this chapter, we introduce p-adic closed discs and annuli, but in a purely ring-theoretic fashion. This avoids having to introduce any p-adic analytic geometry.

NOTATION 7.0.1. Throughout this chapter (and in all later chapters, unless explicitly contravened), let K be a field complete for a nontrivial nonarchimedean valuation $|\cdot|$. Assume that K has characteristic 0, but the residue field κ_K has characteristic $p > 0$. Also assume that things are normalized so that $|p| = p^{-1}$.

1. Power series on closed discs and annuli

We start by introducing some rings that should be thought of as the analytic functions on a closed disc $|t| \leq \beta$, or a closed annulus $\alpha \leq |t| \leq \beta$. As noted in the introduction, this is more properly done in a framework of p-adic analytic geometry, but we will avoid this framework.

DEFINITION 7.1.1. For $\alpha, \beta > 0$, put

$$K\langle \alpha/t, t/\beta \rangle = \left\{ \sum_{i \in \mathbb{Z}} c_i t^i \in K[\![t, t^{-1}]\!] : \lim_{i \to \pm\infty} |c_i| \rho^i = 0 \quad (\rho \in [\alpha, \beta]). \right\}.$$

That is, consider formal bidirectional power series which converge whenever you plug in a value for t with $|t| \in [\alpha, \beta]$, or in other words, when $\alpha/|t|$ and $|t|/\beta$ are both at most 1; it suffices to check for $\rho = \alpha$ and $\rho = \beta$. Although formal bidirectional power series do not form a ring, the subset $K\langle \alpha/t, t/\beta \rangle$ does form a ring under the expected operations.

DEFINITION 7.1.2. If $\alpha = 0$, the only reasonable interpretation of the previous definition is to require $c_i = 0$ for $i < 0$. When there are no negative powers of t, it is redundant to require the convergence for $\rho < \beta$. In other words,

$$K\langle 0/t, t/\beta \rangle = K\langle t/\beta \rangle = \left\{ \sum_{i=0}^{\infty} c_i t^i \in K[\![t]\!] : \lim_{i \to \infty} |c_i| \beta^i = 0 \right\}.$$

One could also allow $\beta = \infty$ for a similar effect in the other direction. More succinctly put, we identify $K\langle \alpha/t, t/\beta \rangle$ with $K\langle \beta^{-1}/t^{-1}, t^{-1}/\alpha^{-1} \rangle$.

2. Gauss norms and Newton polygons

The rings $K\langle \alpha/t, t/\beta \rangle$ quite a lot like polynomial rings (or Laurent polynomial rings, in case $\alpha \neq 0$) in one variable. The next few statements are all instances of this analogy.

Definition 7.2.1. From the definition of $K\langle \alpha/t, t/\beta\rangle$, we see that it carries a well-defined ρ-Gauss norm

$$\left| \sum_i c_i t^i \right|_\rho = \max_i \{|c_i| \rho^i\}$$

for any $\rho \in [\alpha, \beta]$. For $\rho = \alpha = 0$, this reduces to simply $|c_0|$. (The fact that this is a multiplicative norm follows as in Proposition 2.1.3.) The additive version is this is to take $r \in [-\log \beta, -\log \alpha]$ and put

$$v_r \left(\sum c_i t^i \right) = \min_i \{ v(c_i) + ri \},$$

where $v(c) = -\log |c|$.

Definition 7.2.2. One may define the *Newton polygon* for an element $x = \sum x_i t^i \in K\langle \alpha/t, t/\beta\rangle$ as the lower convex hull of the set

$$\{(-i, v(x_i)) : i \in \mathbb{Z}, x_i \neq 0\},$$

but retaining only those slopes in $[-\log \beta, -\log \alpha]$.

Proposition 7.2.3. *Let $x = \sum_i x_i t^i \in K\langle \alpha/t, t/\beta\rangle$ be nonzero.*

(a) *The Newton polygon of x has finite width.*
(b) *The function $r \mapsto v_r(x)$ on $[-\log \beta, -\log \alpha]$ is continuous, piecewise affine, and convex.*
(c) *The function $\rho \mapsto |x|_\rho$ on $[\alpha, \beta]$ is continuous and log-concave. The log-concavity means that $\rho, \sigma \in [\alpha, \beta]$ and $c \in [0, 1]$, put $\tau = \rho^c \sigma^{1-c}$; then*

$$|x|_\tau \leq |x|_\rho^c |x|_\sigma^{1-c}.$$

(d) *If $\alpha = 0$, then v_r is decreasing on $[-\log \beta, \infty)$; in other words, for all $\rho \in [0, \beta]$, $|x|_\rho \leq |x|_\beta$.*

Part (c) should be thought of as a nonarchimedean analogue of the Hadamard three circle theorem.

Proof. We have (a) because there is a least i for which $|c_i|\alpha^i$ is maximized, and there is a greatest j for which $|c_j|\beta^j$ is maximized. This implies (b) because as in the polynomial case, we may interpret $v_r(x)$ as the y-intercept of the supporting line of the Newton polygon of slope r. This in turn implies (c), and (d) is a remark made earlier. $\square$

When dealing with the ring $K\langle \alpha/t, t/\beta\rangle$, the following completeness property will be extremely useful.

Proposition 7.2.4. *The ring $K\langle \alpha/t, t/\beta\rangle$ is Fréchet complete for the norms $|\cdot|_\rho$ for all $\rho \in I$. That is, if $\{x_n\}_{n=0}^\infty$ is a sequence which is simultaneously Cauchy under $|\cdot|_\rho$ for all $\rho \in I$, then it is convergent. (By Proposition 7.2.3, it suffices to check the Cauchy property at each nonzero endpoint of I.)*

Proof. Exercise. $\square$

For instance, the completeness property is used in the construction of multiplicative inverses.

Lemma 7.2.5. *If $\alpha = 0$ (resp. $\alpha > 0$), a nonzero element $f \in K\langle \alpha/t, t/\beta \rangle$ is a unit if and only if v_r is constant (resp. affine) on $[-\log(\beta), -\log(\alpha)]$.*

Proof. We will just consider the case $\alpha > 0$; the other case is similar (and easier). Put $f = \sum_{i \in \mathbb{Z}} f_i t^i$. Note that the following are equivalent:

(a) there is a single i for which $|f|_\rho = |f_i|\rho^i$ for all $\rho \in [\alpha, \beta]$;
(b) the function $r \mapsto v_r(f)$ on $[-\log(\beta), -\log(\alpha)]$ is affine;
(c) the Newton polygon of f has no slopes in $[-\log(\beta), -\log(\alpha)]$.

By (c), these conditions all hold if f is a unit. Conversely, if these conditions hold, then the series

$$(f_i t_i)^{-1}(1 - (f_i t^i - f)/(f_i t^i))^{-1} = \sum_{j=0}^{\infty} (f_i t^i - f)^j (f_i t_i)^{-j-1}$$

converges by Proposition 7.2.4, and its limit is an inverse of f. □

3. Factorization results

Proposition 7.3.1 (Weierstrass preparation). *Suppose that $f = \sum_{i \in \mathbb{Z}} f_i t^i \in K\langle \alpha/t, t/\beta \rangle$, and that $\rho \in [\alpha, \beta]$ is such that there is a unique $m \in \mathbb{Z}$ maximizing $|f_m|\rho^m$. Then there is a unique factorization $f = f_m t^m g h$ with*

$$g \in K\langle \alpha/t, t/\beta \rangle \cap K[\![t]\!] = K\langle t/\beta \rangle,$$
$$h \in K\langle \alpha/t, t/\beta \rangle \cap K[\![t^{-1}]\!] = K\langle \alpha/t \rangle,$$

$|g|_\rho = |g_0| = 1$, and $|h - 1|_\rho < 1$.

Proof. As in Theorem 2.2.1, this is a consequence of the master factorization theorem (Theorem 2.2.2); the completeness of the ring is provided by Property 7.2.4. □

In light of the finite width property of the Newton polygon, the following should not be a surprise.

Proposition 7.3.2 (More Weierstrass preparation). *For $f \in K\langle \alpha/t, t/\beta \rangle$, there exists a polynomial $P \in K[t]$ and a unit $g \in K\langle \alpha/t, t/\beta \rangle^\times$ such that $f = Pg$. In particular, $K\langle \alpha/t, t/\beta \rangle$ is a principal ideal domain.*

Proof. Using Proposition 7.3.1, we may reduce to two instances of the case $\alpha = 0$, so we restrict to that case hereafter. Put $f = \sum_i f_i t^i$, and choose m maximizing $|f_m|\beta^m$. Let R be the ring of formal sums $\sum_i c_i t^i$ of series with $|c_i|\beta^i$ bounded as $i \to -\infty$ and tending to 0 as $i \to +\infty$. Let e be the inverse of $\sum_{i=0}^m f_i t^i$ in R, and apply Theorem 2.2.2 to factor $ef = gh$ in R, in which g is a unit in $K\langle t/\beta \rangle$ by Lemma 7.2.5. Now $h \sum_{i=0}^m f_i t^i = fg^{-1}$ belongs to

$$K[\![t]\!] \cap t^m K[\![t^{-1}]\!].$$

It is thus a polynomial of degree m, proving the claim. □

We will make frequent and often implicit use of the following patching lemma.

Lemma 7.3.3 (Patching lemma). *Suppose $\alpha \leq \gamma \leq \beta \leq \delta$. Let M_1 be a finite free module over $K\langle \alpha/t, t/\beta \rangle$, let M_2 be a finite free module over $K\langle \gamma/t, t/\delta\rangle$, and suppose we are given an isomorphism*

$$\psi : M_1 \otimes K\langle \gamma/t, t/\beta\rangle \cong M_2 \otimes K\langle \gamma/t, t/\beta\rangle.$$

Then we can find a finite free module M over $K\langle \alpha/t, t/\delta\rangle$ and isomorphisms $M_1 \cong M \otimes K\langle \alpha/t, t/\beta\rangle$, $M_2 \cong M \otimes K\langle \gamma/t, t/\delta\rangle$ inducing ψ. Moreover, M is determined by this requirement up to unique isomorphism.

Proof. We will only explain the case $\alpha > 0$; the case $\alpha = 0$ is similar. Choose bases of M_1 and M_2 and let A be the $n \times n$ matrix defining ψ; then A must be invertible over $K\langle \gamma/t, t/\beta\rangle$. Choose $\rho \in [\gamma, \beta]$; since $\det(A)$ is a unit in $K\langle \gamma/t, t/\beta\rangle$, we can find an invertible $n \times n$ matrix W over $K\langle \gamma/t, t/\beta\rangle$ such that $\det(WA) = 1$. (For instance, take $W = \mathrm{Diag}(\det(A)^{-1}, 1, \dots, 1)$.)

It is then possible (see exercises) to find invertible matrices U, V over $K[t, t^{-1}]$ such that $|UWAV - I_n|_\rho < 1$. By changing the initial choices of bases, we can force ourselves into the case $|A - I_n|_\rho < 1$.

By applying Theorem 2.2.2 in the $n \times n$ matrix ring over $K\langle \gamma/t, t/\beta\rangle$, we can split A as a product of an invertible matrix over $K\langle t/\beta\rangle$ and an invertible matrix over $K\langle \gamma/t\rangle$. Using these to change basis in M_1 and M_2, respectively, we can put ourselves in the situation where $A = I_n$, in which case we may identify the bases of M_1 and M_2. Take M to be the free module over $K\langle \alpha/t, t/\delta\rangle$ with the same basis. □

4. Open discs and annuli

Although we have been talking about closed discs so far, it is quite natural to also consider open discs. One important reason is that the antiderivative of an analytic function on the closed disc of radius β is only defined on the *open* disc of radius β (see exercises for Chapter 8).

Definition 7.4.1. By a finite free module M on the region $|t| \in I$, for $I \subseteq [0, +\infty)$ any interval, we will mean a sequence of finite free modules M_i over $K\langle \alpha_i/t, t/\beta_i\rangle$ with $[\alpha_1, \beta_1] \subseteq [\alpha_2, \beta_2] \subseteq \dots$ an increasing sequence of closed intervals with union I, together with isomorphisms $M_{i+1} \otimes K\langle \alpha_i/t, t/\beta_i\rangle \cong M_i$. Using Lemma 7.3.3, we check that the construction is canonically independent of the choice of the sequence.

Notes

The Hadamard three circles theorem (Proposition 7.2.3(c)) is a special case of the fact that the *Shilov boundary* of the annulus $\alpha \leq |t| \leq \beta$ consists of the two circles $|t| = \alpha$ and $|t| = \beta$. For much amplification of this remark, including a full-blown theory of harmonic functions on Berkovich analytic curves, see [**Thu05**]. For an alternate presentation, restricted to the Berkovich projective line but otherwise more detailed, see [**BR07**].

The patching lemma (Lemma 7.3.3) is a special case of the glueing property of coherent sheaves on affinoid rigid analytic spaces, i.e., the theorems of Kiehl and Tate [**BGR84**, Theorems 8.2.1/1 and 9.4.2/3]. The factorization argument in the proof, however, is older still; it is the nonarchimedean version of what is called a *Birkhoff factorization* over an archimedean field. Similarly, Definition 7.4.1 corresponds to the definition of a locally free coherent sheaf on the corresponding rigid or Berkovich analytic space. Such a sheaf is only guaranteed to be freely generated by global sections in case K is spherically complete [**Ked05a**, Theorem 3.14].

Exercises

(1) Prove Proposition 7.2.4. (Hint: it may be easiest to first construct the limit using a single $\rho \in [\alpha, \beta]$, then show that it must also work for the other ρ.)

(2) Let R be the ring of formal power series over K which converge for $|t| < 1$. Prove that R is not noetherian; this is why I avoided introducing it. (Hint: pick a sequence of points in the open unit disc converging to the boundary, and consider the ideal of functions vanishing on all but finitely many of these points.)

(3) Suppose K is complete for a discrete valuation. Prove that any element of $\mathfrak{o}_K[\![t]\!] \otimes_{\mathfrak{o}_K} K$ (that is, a power series with bounded coefficients) is equal to a polynomial in t times a unit. Then prove that this fails if K is complete for a nondiscrete valuation.

(4) Let A be an $n \times n$ matrix over $K\langle \rho/t, t/\rho \rangle$ such that $|\det(A) - 1|_\rho < 1$. Prove that there exist invertible matrices U, V over $K[t, t^{-1}]$ such that $|UAV - I_n|_\rho < 1$. (Hint: perform approximate Gaussian elimination.) An analogous argument, but in more complicated notation, is [**Ked04a**, Lemma 6.2].

CHAPTER 8

Radius and generic radius of convergence

In this chapter, we consider the radius of convergence of a differential module defined on a closed or open disc. We also introduce some key invariants, the *generic radius of convergence* and *subsidiary radii*, that turn out to be easier to work with than the radius of convergence.

1. Differential modules on rings and annuli

HYPOTHESIS 8.1.1. Throughout this chapter, we will view $K\langle\alpha/t, t/\beta\rangle$ as a differential ring with derivation $d = \frac{d}{dt}$, the formal differentiation in the variable t.

PROPOSITION 8.1.2. *Any finite differential module over $K\langle\alpha/t, t/\beta\rangle$ is torsion-free, and hence free. Consequently, the finite differential modules over $K\langle\alpha/t, t/\beta\rangle$ form an abelian category.*

PROOF. Exercise. $\square$

COROLLARY 8.1.3. *For M a finite differential module over $K\langle\alpha/t, t/\beta\rangle$, any set of horizontal sections which are linearly independent over $K\langle\alpha/t, t/\beta\rangle$ form part of a basis of M.*

PROOF. Let S be such a set; then S determines a morphism from a trivial differential module to M. By Proposition 8.1.2, the image of this map must be a direct summand of M as a module, proving the claim. $\square$

COROLLARY 8.1.4. *For M a finite differential module over $K\langle\alpha/t, t/\beta\rangle$ of rank n admitting a set S of n horizontal sections linearly independent over $K\langle\alpha/t, t/\beta\rangle$, M is trivial and $H^0(M)$ is the K-span of S.*

2. Radius of convergence on a disc

DEFINITION 8.2.1. Let M be a nonzero finite differential module over $K\langle t/\beta\rangle$ (i.e., a nonzero finite differential module on the closed disc of radius β around $t = 0$). Define the *radius of convergence* of M around 0, denoted $R(M)$, to be the supremum of the set of $\rho \in (0, \beta)$ such that $M \otimes K\langle t/\rho\rangle$ has a basis of horizontal elements; we refer to those elements as *local horizontal sections* of M. For M a nonzero finite differential module on the open disc of radius β around $t = 0$, define $R(M)$ as the supremum of $M \otimes R(M \otimes K\langle t/\gamma\rangle)$ over all $\gamma < \beta$. For $\gamma \leq \beta$, note that

$$R(M \otimes K\langle t/\gamma\rangle) = \begin{cases} \gamma & \gamma \leq R(M) \\ R(M) & \gamma > R(M). \end{cases}$$

EXAMPLE 8.2.2. In general, it is possible to have $R(M) < \beta$; that is, there is no p-adic analogue of the global form of the fundamental theorem of ordinary differential equations (as was noted in the introduction). For instance, consider the module $M = K\langle t/\beta\rangle$ with $D(x) =$

x, when $\beta > p^{-1/(p-1)}$; then $R(M) = p^{-1/(p-1)}$ because that is the radius of convergence of the exponential series.

On the other hand, the local form of the fundamental theorem of ordinary differential equations has the following analogue.

PROPOSITION 8.2.3 (p-adic Cauchy theorem). *Let M be a nonzero finite differential module over $K\langle t/\beta \rangle$. Then $R(M) > 0$.*

PROOF. One can give a direct proof of this, but instead we will deduce this from Dwork's transfer theorem (Theorem 8.5.1). We will give a direct proof of a slightly stronger result later (Proposition 16.1.1); see also the notes. $\square$

Here are some easy consequences of the definition of radius of convergence; note the parallels with properties of the spectral norm (Lemma 5.2.7).

LEMMA 8.2.4. *Let M, M_1, M_2 be nonzero finite differential modules over $K\langle t/\beta \rangle$.*

(a) *If $0 \to M_1 \to M \to M_2 \to 0$ is exact, then*

$$R(M) = \min\{R(M_1), R(M_2)\}.$$

(b) *We have*

$$R(M^\vee) = R(M).$$

(c) *We have*

$$R(M_1 \otimes M_2) \geq \min\{R(M_1), R(M_2)\},$$

with equality when $R(M_1) \neq R(M_2)$.

PROOF. For (a), it is clear that $R(M) \leq \min\{R(M_1), R(M_2)\}$; we must check that equality holds. Choose $\lambda < \min\{R(M_1), R(M_2)\}$, so that $M_1 \otimes K\langle t/\lambda \rangle$ and $M_2 \otimes K\langle t/\lambda \rangle$ are both trivial. If we choose a basis of M compatible with the sequence, then the action of D will be block upper triangular nilpotent, and trivializing M amounts to antidifferentiating the entries in the nonzero block. We may not be able to perform this antidifferentiation in $K\langle t/\lambda \rangle$, but we can do it in $K\langle t/\lambda' \rangle$ for any $\lambda' < \lambda$. Since we can make λ and λ' as close to $\min\{R(M_1), R(M_2)\}$ as we like, we find $R(M) \geq \min\{R(M_1), R(M_2)\}$.

For (b), we obtain $R(M^\vee) \geq R(M)$ from the fact that if $M \otimes K\langle t/\lambda \rangle$ is trivial, then so is its dual $M^\vee \otimes K\langle t/\lambda \rangle$. Since M and $M^\vee$ enter symmetrically, we get $R(M^\vee) = R(M)$.

For (c), the inequality is clear from the fact that the tensor product of two trivial modules over $K\langle t/\lambda \rangle$ is also trivial. The last assertion follows from everything else so far as in the proof of Lemma 5.2.7(c). $\square$

EXAMPLE 8.2.5. Let M be the differential module of rank 1 over $K\langle t/\beta \rangle$ defined by $D(v) = \lambda v$ with $\lambda \in K$. Then it is an exercise to show that

$$R(M) = \min\{\beta, |p|^{-1/(p-1)}|\lambda|^{-1}\}$$

3. Generic radius of convergence

In general, the radius of convergence is difficult to compute. To get a better handle on it, we introduce a related but simpler invariant.

Definition 8.3.1. For $\rho > 0$, let F_ρ be the completion of $K(t)$ under the ρ-Gauss norm $|\cdot|_\rho$. Put $d = \frac{d}{dt}$ on $K(t)$; then d extends by continuity to F_ρ, and

$$|d|_{F_\rho} = \rho^{-1}, \qquad |d|_{\mathrm{sp},F_\rho} = \lim_{n\to\infty} |n!|^{1/n}\rho^{-1} = p^{-1/(p-1)}\rho^{-1}.$$

We also make a related construction in case $\rho = 1$.

Definition 8.3.2. Let $\mathcal{E}$ be the completion of $\mathfrak{o}_K((t)) \otimes_{\mathfrak{o}_K} K$ for the 1-Gauss norm. The elements of $\mathcal{E}$ may be identified with formal sums $x = \sum_{i\in\mathbb{Z}} x_i t^i$ satisfying the following conditions.

(a) We have $|c_i| \to 0$ as $i \to -\infty$.
(b) We have $|c_i|$ bounded as $i \to +\infty$.

One again has a 1-Gauss norm $|\cdot|_1$ on $\mathcal{E}$, defined as

$$\left|\sum_i x_i t^i\right| = \sup_i\{|x_i|\}.$$

Beware that if K is discretely valued, the supremum in the Gauss norm is achieved, so $\mathcal{E}$ is a field, and its residue field of $\mathcal{E}$ is equal to $\kappa_K((t))$; however, none of this applies if K is not discretely valued. In any case, $\mathcal{E}$ is complete under $|\cdot|_1$, there is an isometric map $F_1 \to \mathcal{E}$ carrying t to t, and the supremum is achieved for elements of $\mathcal{E}$ in the image of that map; this at least gives an embedding $\kappa_K((t)) \hookrightarrow \kappa_{\mathcal{E}}$.

Definition 8.3.3. Let (V, D) be a nonzero finite differential module over F_ρ or $\mathcal{E}$. We define the *generic radius of convergence* (or for short, the *generic radius*) of V to be

$$R(V) = p^{-1/(p-1)}|D|^{-1}_{\mathrm{sp},V};$$

note that $R(V) > 0$. We will see later (Proposition 8.6.4) that this does indeed compute the radius of convergence of horizontal sections of V on a "generic disc".

We can translate some basic properties of the spectral norm into properties of generic radii of convergence, leading to the following analogue of Lemma 8.2.4. Alternatively, one can first check Proposition 8.6.4 and then simply invoke Lemma 8.2.4 itself around a generic point.

Lemma 8.3.4. *Let V, V_1, V_2 be nonzero finite differential modules over F_ρ.*

(a) *For $0 \to V_1 \to V \to V_2 \to 0$ exact,*

$$R(V) = \min\{R(V_1), R(V_2)\}.$$

(b) *We have*

$$R(V^\vee) = R(V).$$

(c) *We have*

$$R(V_1 \otimes V_2) \geq \min\{R(V_1), R(V_2)\},$$

with equality when $R(V_1) \neq R(V_2)$.

Definition 8.3.5. In some situations, it is more natural to consider the *intrinsic generic radius of convergence*, or for short the *intrinsic radius*, defined as

$$IR(V) = \rho^{-1}R(V) = \frac{|d|_{\mathrm{sp},F_\rho}}{|D|_{\mathrm{sp},V}} \in (0,1].$$

To emphasize the difference, we may refer to the unadorned generic radius of convergence defined earlier as the *extrinsic* generic radius of convergence. (See Proposition 8.6.5 and the notes for some reasons why the intrinsic radius deserves such a name.)

Remark 8.3.6. For I an interval, and for M a nonzero differential module on the annulus $|t| \in I$, it is unambiguous to refer to the generic radius of convergence $R(M \otimes F_\rho)$ of M at radius ρ.

4. Some examples in rank 1

An important class of examples is given as follows.

Example 8.4.1. For $\lambda \in K$, let V_λ be the differential module of rank 1 over F_ρ defined by $D(v) = \lambda t^{-1} v$. It is an exercise to show that $IR(V_\lambda) = 1$ if and only if $\lambda \in \mathbb{Z}_p$.

We can further classify Example 8.4.1 as follows.

Proposition 8.4.2. *We have $V_\lambda \cong V_{\lambda'}$ if and only if $\lambda - \lambda' \in \mathbb{Z}$.*

Proof. Note that $V_\lambda \cong V_{\lambda'}$ if and only if $V_{\lambda-\lambda'}$ is trivial, so we may reduce to the case $\lambda' = 0$. By Example 8.4.1, V_λ is nontrivial whenever $\lambda \notin \mathbb{Z}_p$; by direct inspection, V_λ is trivial whenever $\lambda \in \mathbb{Z}$.

It remains to deduce a contradiction assuming that V_λ is trivial, $\lambda \in \mathbb{Z}_p$, and $\lambda \notin \mathbb{Z}$. There is no harm in enlarging K now, so we may assume that K contains a scalar of norm ρ; by rescaling, we may reduce to the case $\rho = 1$. We now have $f \in F_1^\times$ such that $t\frac{df}{dt} = \lambda f$; by multiplying by an element of $K^\times$, we can force $|f|_1 = 1$.

Let λ_1 be an integer such that $\lambda \equiv \lambda_1 \pmod p$. Then

$$\left| \frac{d(ft^{-\lambda_1})}{dt} \right|_1 = |(\lambda - \lambda_1) f t^{-\lambda_1 - 1}|_1 \leq p^{-1}.$$

Using the embedding $F_1 \hookrightarrow \mathcal{E}$, we may expand $f = \sum_{i \in \mathbb{Z}} f_i t^i$ with $\max_i\{|f_i|\} = 1$. The previous calculation then forces $|f_i| \leq p^{-1}$ unless $i \equiv \lambda_1 \equiv \lambda \pmod p$.

By considering the reduction of f modulo p^n and arguing similarly, we find that $|f_i| \leq p^{-1}$ unless $i \equiv \lambda \pmod{p^n}$ for all n. But since $\lambda \notin \mathbb{Z}$, this means that the image of f in $\kappa_K((t))$ cannot have any terms at all, contradiction. □

5. Transfer theorems

One fundamental relationship between radius of convergence and generic radius of convergence is the following. In the language of Dwork, this is a *transfer theorem*, because it transfers convergence information from one disc to another. (Note that the fact that $R(M) > 0$, which is Proposition 8.2.3, is an immediate corollary.)

Theorem 8.5.1. *For any nonzero finite differential module M over $K\langle t/\beta \rangle$, $R(M) \geq R(M \otimes F_\beta)$. That is, the radius of convergence is at least the generic radius.*

Proof. Suppose $\lambda < \beta$ and $\lambda < p^{-1/(p-1)}|D|_{\mathrm{sp},V}^{-1}$. We claim that for any $x \in M$, the Taylor series

$$y = \sum_{i=0}^{\infty} \frac{(-t)^i}{i!} D^i(x) \tag{8.5.1.1}$$

converges under $|\cdot|_\lambda$. To see this, pick $\epsilon > 0$ such that $\lambda p^{1/(p-1)}(|D|_{\mathrm{sp},V} + \epsilon) < 1$; then there exists $c > 0$ such that $|D^i(x)| \leq c(|D|_{\mathrm{sp},V} + \epsilon)^i$ for all i. The i-th term of the sum defining y thus has norm at most $\lambda^i p^{i/(p-1)} c(|D|_{\mathrm{sp},V} + \epsilon)^i$, which tends to 0 as $i \to \infty$.

By differentiating the series expression, we find that

$$\begin{aligned} Dy &= \sum_{i=0}^{\infty} \frac{(-t)^i}{i!} D^{i+1}(x) + \sum_{i=1}^{\infty} \frac{-(-t)^{i-1}}{(i-1)!} D^i(x) \\ &= \sum_{i=0}^{\infty} \frac{(-t)^i}{i!} D^{i+1}(x) - \sum_{i=0}^{\infty} \frac{(-t)^i}{i!} D^{i+1}(x) = 0. \end{aligned}$$

That is, y is a horizontal section of $V \otimes K\langle t/\lambda\rangle$.

If we run this construction over a basis of M, we obtain horizontal sections of $V \otimes K\langle t/\lambda\rangle$ whose reductions modulo t form a basis; they thus form a basis themselves by Nakayama's lemma (and the fact that finite differential modules over $K\langle t/\lambda\rangle$ are free). This proves the claim. $\square$

For future reference, we note that Theorem 8.5.1 extends to the case of nilpotent regular singularities.

Definition 8.5.2. Let M be a finite differential module over the ring $K\langle t/\beta\rangle$ equipped with the derivation $t\frac{d}{dt}$, such that the action of D on M/tM is nilpotent. Let N be the matrix of action of D on some basis, and let U be the fundamental solution matrix (Definition 6.3.1). Define the *radius of convergence* of M to be the supremum of $\rho \in (0, \beta)$ such that U has entries in $K\langle t/\rho\rangle$; this does not depend on the choice of the basis.

Lemma 8.5.3. *Let M be a finite differential module over $K[\![t]\!]$ equipped with the derivation $t\frac{d}{dt}$, such that the action of D on M/tM is nilpotent. Let e be the nilpotency index of the action of D on M/tM. Then for any $x \in M$, the sequence*

$$x_{(i)} = D^{e-1} \frac{(1-D)^e \cdots (i-D)^e}{(i!)^e}(x)$$

converges t-adically to a horizontal section of M.

Proof. The columns of the fundamental solution matrix form a basis $v_1, \dots, v_n$ on which D acts via a nilpotent matrix N_0 over K. Write $x = \sum_{j=1}^n x_j v_j$ and $x_j = \sum_{l=0}^\infty x_{j,l} t^l$. Then the image of $x_{j,0} v_j$ under $D^{e-1}(1-D)^e \cdots (i-D)^e/(i!)^e$ is the same as that under D^{e-1} since $D^e(v_j) = 0$. On the other hand, for $l > 0$, $x_{j,l} t^l v_j$ is annihilated by $(l-D)^e$, hence also by $D^{e-1}(1-D)^e \cdots (i-D)^e/(i!)^e$ whenever $i \geq l$. This proves the claim. $\square$

Theorem 8.5.4. *Let M be a finite differential module over the ring $K\langle t/\beta\rangle$ equipped with the derivation $t\frac{d}{dt}$, such that the action of D on M/tM is nilpotent. Then $R(M) \geq R(M \otimes F_\beta)$.*

PROOF. Let M_1 be any nonzero differential quotient of M, and let e_1 be the index of nilpotency of the action of D on M_1/tM_1. Then for $x \in M_1$, the sequence in Lemma 8.5.3 (applied to M_1) converges in $M_1 \otimes K\langle t/\lambda\rangle$ for any $\lambda \in (0, R(M \otimes F_\beta))$. Moreover, the limit is nonzero if the image of x in M_1/tM_1 not killed by D^{e_1-1}, then the limit is nonzero. Consequently, $M_1 \otimes K\langle t/\lambda\rangle$ admits a nonzero horizontal section.

By iterating this argument, we deduce that for any $\lambda \in (0, R(M \otimes F_\beta))$, $M \otimes K\langle t/\lambda\rangle$ is a successive extension of trivial differential modules. This implies that for any $\lambda' \in (0, \lambda)$, the fundamental solution matrix has entries in $K\langle t/\lambda'\rangle$. This proves the claim. □

6. Geometric interpretation

As promised, here is a construction that explains the name "generic radius of convergence".

DEFINITION 8.6.1. Let L be a complete extension of K. A *generic point* of L of norm ρ is an element $t_\rho \in L$ with $|t_\rho| = \rho$, such that there is no $t \in L^{\mathrm{alg}} \cap K^{\mathrm{alg}}$ with $|t - t_\rho| < \rho$. For instance, one can construct a generic point t_ρ by forming the completion of $K(t_\rho)$ for the ρ-Gauss norm.

DEFINITION 8.6.2. Let L be a complete extension of K. For any $t_\rho \in L$ with $|t_\rho| = \rho$, the substitution $t \mapsto t_\rho + (t - t_\rho)$ induces an isometric map $K[t] \to L\langle (t - t_\rho)/\rho\rangle$. However, if (and only if) t_ρ is a generic point, then the composition of this map with the reduction modulo $t - t_\rho$ is again an isometry; this can be seen by writing the map $K[t] \to L\langle (t-t_\rho)/\rho\rangle$ as

$$f \mapsto \sum_{i=0}^{\infty} \frac{f^{(i)}(t_\rho)}{i!}(t - t_\rho)^i$$

and using the fact that

$$\left|\frac{f^{(i)}(t_\rho)}{i!}\right| = \left|\frac{f^{(i)}}{i!}\right|_\rho \leq \rho^{-i}|f|_\rho = \rho^{-i}|f(t_\rho)|.$$

Hence we get an isometry $F_\rho \to L\langle (t - t_\rho)/\rho\rangle$.

REMARK 8.6.3. In Berkovich's theory of nonarchimedean analytic geometry, the geometric interpretation of the above construction is that the analytic space corresponding to F_ρ is obtained from the closed disc of radius ρ by removing the open disc of radius ρ centered around each point of K^{alg}. As a result, it still contains any open disc of radius ρ that does not meet K^{alg}.

PROPOSITION 8.6.4. *Let V be a nonzero finite differential module over F_ρ, and let V' be the base change of V to the open disc of radius ρ in $t - t_\rho$ over L. Then the generic radius of convergence of V is equal to the radius of convergence of V'.*

PROOF. Let G_λ be the completion of $L(t - t_\rho)$ for the λ-Gauss norm; then the map $F_\rho \to G_\lambda$ is an isometry for any $\lambda \leq \rho$. Consequently, if we compute $|D|_{\mathrm{sp},V}$ in terms of some basis using Lemma 5.2.5, we get the same norms whether we work in F_ρ or G_λ. In other words,

$$|D|_{\mathrm{sp},V\otimes G_\lambda} = \max\{|d|_{\mathrm{sp},G_\lambda}, |D|_{\mathrm{sp},V}\} = \max\{p^{-1/(p-1)}\lambda^{-1}, |D|_{\mathrm{sp},V}\}.$$

On one hand, this implies $R(V) \leq R(V')$ by applying Theorem 8.5.1 to $V \otimes L\langle (t-t_\rho)/\lambda \rangle$ for a sequence of values of λ converging to ρ.

On the other hand, pick any $\lambda < R(V')$; then $V \otimes G_\lambda$ is a trivial differential module, so the spectral norm of D on it is $p^{-1/(p-1)}\lambda^{-1}$. We thus have

$$|D|_{\mathrm{sp},V} \leq p^{-1/(p-1)}\lambda^{-1},$$

so $R(V) \geq \lambda$. This yields $R(V) \geq R(V')$. □

Here is an example illustrating both the use of the geometric interpretation and a good transformation property of the intrinsic normalization.

Proposition 8.6.5. *Let m be a positive integer coprime to p, and let $f_m : F_\rho \to F_{\rho^m}$ be the map $t \mapsto t^m$. Then for any finite differential module V over F_ρ, $IR(V) = IR(f^*(V))$.*

Proof. This follows from the geometric interpretation plus the fact that

$$|t - t_\rho \zeta_m^i| < c\rho \text{ for some } i \in \{0, \dots, m-1\} \Leftrightarrow |t^m - t_\rho^m| < c\rho^m \qquad (c \in (0,1)), \tag{8.6.5.1}$$

whose proof is left as an exercise. □

Remark 8.6.6. A similar construction can be made for $\mathcal{E}$. Let L be the completion of $\mathfrak{o}_K((t_1)) \otimes_{\mathfrak{o}_K} K$ for the 1-Gauss norm. Then the substitution $t \mapsto t_1 + (t - t_1)$ induces an isometry $\mathfrak{o}_K((t)) \to \mathfrak{o}_L[\![t - t_1]\!]$ for the 1-Gauss norm, extending to an isometric embedding of $\mathcal{E}$ into the completion of $\mathfrak{o}_L[\![t - t_1]\!] \otimes_{\mathfrak{o}_L} L$ for the 1-Gauss norm.

7. Subsidiary radii

It is sometimes important to consider not only the generic radius of convergence, but also some secondary invariants.

Definition 8.7.1. Let V be a finite differential module over F_ρ. Let $V_1, \dots, V_m$ be the Jordan-Hölder constituents of V. We define the *subsidiary generic radii of convergence*, or for short the *subsidiary radii*, to be the multiset consisting of $R(V_i)$ with multiplicity $\dim V_i$ for $i = 1, \dots, m$. We also have *intrinsic subsidiary generic radii of convergence*, or *intrinsic subsidiary radii*, obtained by multiplying the subsidiary radii by ρ^{-1}.

Remark 8.7.2. The product of the subsidiary radii is an invariant with properties somewhat analogous to those of the irregularity of a finite differential module over $\mathbb{C}((z))$. We will flesh out this remark later.

Remark 8.7.3. It is not yet clear how to interpret the subsidiary radii as the radii of convergence of anything. We will give this interpretation in a later chapter (Theorem 10.10.2).

Notes

According to [**DGS94**, Appendix III] (which see for more information), the p-adic Cauchy theorem (Proposition 8.2.3) was originally proved by Lutz [**Lut37**]. See Proposition 16.1.1 for a related result.

The notion of restricting a p-adic differential module to a generic disc originates in the work of Dwork [**Dwo73a**], although in retrospect, the base change involved is quite natural in Berkovich's framework of nonarchimedean analytic geometry. (This point of view has been adopted by Baldassarri and di Vizio in [**BdV07**].) Our definition of the generic radius of convergence is taken from Christol and Dwork [**CD94**].

The intrinsic radius of convergence (original terminology) was introduced in [**Ked07e**], where it is called the "toric normalization" in light of Proposition 8.6.5.

The subsidiary radii (original terminology) have not been studied much previously; the one reference we found is the work of Young [**You92**]. We will give Young's interpretation of the subsidiary radii as radii of convergence of certain horizontal sections, in a refined form, as Theorem 10.10.2.

Exercises

(1) Prove Proposition 8.1.2. (Hint: first prove that $K\langle \alpha/t, t/\beta\rangle$ has no nonzero differential ideals. Then given a finite differential module over $K\langle \alpha/t, t/\beta\rangle$, consider the annihilator of the torsion submodule.)

(2) Exhibit an example showing that the cokernel of $\frac{d}{dt}$ on $K\langle \alpha/t, t/\beta\rangle$ is not spanned over K by t^{-1}. That is, antidifferentiation with respect to t is not well-defined.

(3) Prove Example 8.2.5.

(4) Prove Example 8.4.1. (Hint: consider the cases $\lambda \in \mathbb{Z}_p$, $\lambda \in \mathfrak{o}_K - \mathbb{Z}_p$, and $\lambda \notin \mathfrak{o}_K$ separately.)

(5) Give an explicit formula for $IR(V_\lambda)$, in terms of ρ and the minimum distance from λ to an integer.

(6) Prove (8.6.5.1).

(7) Verify that $|t\frac{d}{dt}|_{\mathrm{sp},F_\rho} \neq |t|_\rho |\frac{d}{dt}|_{\mathrm{sp},F_\rho}$. This means that unlike in the case of regular singularities, we cannot be as cavalier about working with $t\frac{d}{dt}$ instead of $\frac{d}{dt}$.

(8) With notation as in Proposition 8.6.5, show that all of the intrinsic subsidiary radii of V and $f_m^*(V)$ match, not just the largest generic radius.

(9) Here is an "off-centered" analogue of Proposition 8.6.5 suggested by Liang Xiao (compare with Theorem 9.8.2). Let m be a positive integer coprime to p. Given $\rho \in (0,1]$, let $f_m : F_\rho \to F_\rho$ be the map $t \mapsto (t+1)^m - 1$. Then for any finite differential module V over F_ρ, $R(V) = R(f_m^*(V))$. (As in the previous exercise, one can also get equality for the other subsidiary radii.)

CHAPTER 9

Frobenius pullback and pushforward

In this chapter, we introduce Dwork's technique of "descent along Frobenius" to analyze the generic radius of convergence and subsidiary radii of a differential module in the range where Newton polygons do not apply.

NOTATION 9.0.1. As in the previous chapter, let K be a complete nonarchimedean field, and let F_ρ be the completion of $K(t)$ for the ρ-Gauss norm for some $\rho > 0$.

1. Why Frobenius?

It may be helpful to review the current state of affairs, to clarify why we need to descend along Frobenius.

Let V be a finite differential module over F_ρ. Then the possible values of the spectral norm $|D|_{\mathrm{sp},V}$ are the real numbers greater than or equal to $|d|_{\mathrm{sp},F_\rho} = p^{1/(p-1)}\rho^{-1}$, corresponding to generic radii of convergence less than or equal to ρ. However, if we want to calculate the spectral norm using the Newton polygon of a twisted polynomial, we cannot distinguish among values less than or equal to the operator norm $|d|_{F_\rho} = \rho^{-1}$. In particular, we cannot use this technique to prove a decomposition theorem for differential modules that separates components of spectral norm between $p^{1/(p-1)}\rho^{-1}$ and ρ^{-1}.

One way one might want to get around this is to consider not d but a high power of d, particularly a p^n-th power. The trouble with this is that iterating a derivation does not give another derivation, but something much more complicated. Instead, we will try to differentiate with respect to t^{p^n} instead of with respect to t. This will have the effect of increasing the spectral norm, so that we can push it into the range where Newton polygons become useful.

2. p-th roots

We first make some calculations in answer to the following question: if two p-adic numbers are close together, how close are their p-th powers, or their p-th roots? (See also [**DGS94**, §V.6] and [**Chr83**, Proposition 4.6.4].)

REMARK 9.2.1. We observed previously (8.6.5.1) that when m is a positive integer coprime to p,

$$|t - \eta\zeta_m^i| < \lambda|\eta| \text{ for some } i \in \{0, \dots, m-1\} \Leftrightarrow |t^m - \eta^m| < \lambda|\eta|^m \qquad (\lambda \in (0,1)).$$

This breaks down for $m = p$, because a primitive p-th root of unity ζ_p satisfies $|1 - \zeta_p| < 1$. The quantities $1 - \zeta_p^m$ for $m = 1, \dots, p-1$ are Galois conjugates, so

$$|1 - \zeta_p| = \left|\prod_{m=1}^{p-1}(1 - \zeta_p^m)\right|^{1/(p-1)} = |p|^{1/(p-1)} = p^{-1/(p-1)}$$

since the product is the derivative of $T^p - 1$ evaluated at $T = 1$.

LEMMA 9.2.2. *Pick $t, \eta \in K$.*

(a) *For $\lambda \in (0,1)$, if $|t-\eta| \leq \lambda|\eta|$, then*

$$|t^p - \eta^p| \leq \max\{\lambda^p, p^{-1}\lambda\}|\eta^p| = \begin{cases} \lambda^p|\eta^p| & \lambda \geq p^{-1/(p-1)} \\ p^{-1}\lambda|\eta^p| & \lambda \leq p^{-1/(p-1)}. \end{cases}$$

(b) *Suppose $\zeta_p \in K$. If $|t^p - \eta^p| \leq \lambda|\eta^p|$, then there exists $m \in \{0, \dots, p-1\}$ such that*

$$|t - \zeta_p^m \eta| \leq \min\{\lambda^{1/p}, p\lambda\}|\eta| = \begin{cases} \lambda^{1/p}|\eta| & \lambda \geq p^{-p/(p-1)} \\ p\lambda|\eta| & \lambda \leq p^{-p/(p-1)}. \end{cases}$$

Moreover, if $\lambda \geq p^{-p/(p-1)}$, we may always take $m = 0$.

We will use repeatedly, and without comment, the fact that

$$\lambda \mapsto \max\{\lambda^p, p^{-1}\lambda\}, \qquad \lambda \mapsto \min\{\lambda^{1/p}, p\lambda\}$$

are strictly increasing functions from $[0,1]$ to itself that are inverse to each other.

PROOF. There is no harm in assuming $\zeta_p \in K$ for both parts. For (a), factor $t^p - \eta^p$ as $t - \eta$ times $t - \eta\zeta_p^m$ for $m = 1, \dots, p-1$, and write

$$t - \eta\zeta_p^m = (t - \eta) + \eta(1 - \zeta_p^m).$$

If $|t - \eta| \geq p^{-1/(p-1)}|\eta|$, then $t - \eta$ is the dominant term, otherwise $\eta(1 - \zeta_p^m)$ dominates. This gives the claimed bounds.

For (b), consider the Newton polygon of

$$t^p - \eta^p - c = \sum_{i=0}^{p-1} \binom{p}{i} \eta^i (t-\eta)^{p-i} - c$$

viewed as a polynomial in $t - \eta$. Suppose $|c| = \lambda|\eta^p|$. If $\lambda \geq p^{-p/(p-1)}$, then the terms $(t-\eta)^p$ and c dominate, and all roots have norm $\lambda^{1/p}|\eta|$. Otherwise, the terms $(t-\eta)^p$, $p(t-\eta)\eta^{p-1}$, and c dominate, so one root has norm $p\lambda|\eta|$ and the others are larger; repeating with η replaced by $\zeta_p^m\eta$ for $m = 0, \dots, p-1$ gives p distinct roots, which accounts for all of them. $\square$

COROLLARY 9.2.3. *Let $T : K[\![t^p - \eta^p]\!] \to K[\![t - \eta]\!]$ be the substitution $t^p - \eta^p \mapsto ((t - \eta) + \eta)^p - \eta^p$.*

(a) *If $f \in K\langle (t^p - \eta^p)/(\lambda|\eta^p|)\rangle$ for some $\lambda \in (0,1)$, then $T(f) \in K\langle (t-\eta)/(\lambda'|\eta|)\rangle$ for $\lambda' = \min\{\lambda^{1/p}, p\lambda\}$.*

(b) *If $T(f) \in K\langle (t-\eta)/(\lambda|\eta|)\rangle$ for some $\lambda \in (p^{-1/(p-1)}, 1)$, then $f \in K\langle (t^p - \eta^p)/(\lambda'|\eta^p|)\rangle$ for $\lambda' = \lambda^p$.*

(c) *Suppose K contains a primitive p-th root of unity ζ_p. For $m = 0, \dots, p-1$, let $T_m : K[\![t^p - \eta^p]\!] \to K[\![t - \zeta_p^m\eta]\!]$ be the substitution $t^p - \eta^p \mapsto ((t - \zeta_p^m\eta) + \zeta_p^m\eta)^p - \eta^p$. If for some $\lambda \in (0, p^{-1/(p-1)}]$ one has $T_m(f) \in K\langle (t - \zeta_p^m\eta)/(\lambda|\eta|)\rangle$ for $m = 0, \dots, p-1$, then $f \in K\langle (t^p - \eta^p)/(\lambda'|\eta^p|)\rangle$ for $\lambda' = p^{-1}\lambda$.*

3. Moving along Frobenius

DEFINITION 9.3.1. Let F'_ρ be the completion of $K(t^p)$ for the ρ^p-Gauss norm, viewed as a subfield of F_ρ, and equipped with the derivation $d' = \frac{d}{dt^p}$. We then have

$$d = \frac{dt^p}{dt} d' = pt^{p-1} d'.$$

Given a finite differential module (V', D') over F'_ρ, we may view $\varphi^* V' = V' \otimes F_\rho$ as a differential module over F_ρ for the derivation $D = pt^{p-1} D' \otimes d$ as a differential

$$D(v \otimes f) = pt^{p-1} D'(v) \otimes f + v \otimes d(f).$$

LEMMA 9.3.2. *Let (V', D') be a finite differential module over F'_ρ. Then*

$$IR(\varphi^* V') \geq \min\{IR(V')^{1/p}, pIR(V')\}.$$

PROOF. For any $\lambda < IR(\varphi^* V')$, any complete extension L of K, and any generic point $t_\rho \in L$ relative to K of norm ρ, $(\varphi^* V') \otimes L\langle (t^p - t_\rho^p)/(\lambda \rho^p) \rangle$ admits a basis of horizontal sections. By Corollary 9.2.3(a), $V' \otimes L\langle (t - t_\rho)/(\min\{\lambda^{1/p}, p\lambda\}\rho) \rangle$ does likewise. □

REMARK 9.3.3. The inequality in Lemma 9.3.2 can be strict; see for instance Definition 9.3.5.

DEFINITION 9.3.4. For V a differential module over F_ρ, define the *Frobenius descendant* of V as the module $\varphi_* V$ obtained from V by restriction along $F'_\rho \to F_\rho$, viewed as a differential module over F'_ρ with differential $D' = p^{-1} t^{-p+1} D$. Note that this operation commutes with duals.

DEFINITION 9.3.5. For $m = 0, \dots, p-1$, let W_m be the differential module over F'_ρ with one generator v, such that

$$D(v) = \frac{m}{p} t^{-p} v.$$

From the Newton polygon associated to v, we read off $IR(W_m) = p^{-p/(p-1)}$ for $m \neq 0$. (You may think of the generator v as a proxy for t^m.)

LEMMA 9.3.6. (a) *For V a differential module over F_ρ, there are canonical isomorphisms*

$$\iota_m : (\varphi_* V) \otimes W_m \cong \varphi_* V \qquad (m = 0, \dots, p-1).$$

(b) *For V a differential module over F_ρ, a submodule U of $\varphi_* V$ is itself the Frobenius descendant of a submodule of V if and only if $\iota_m(U \otimes W_m) = U$ for $m = 0, \dots, p-1$.*

(c) *For V' a differential module over F'_ρ, there is a canonical isomorphism*

$$\varphi_* \varphi^* V' \cong \bigoplus_{m=0}^{p-1} (V' \otimes W_m).$$

(d) *For V a differential module over F_ρ, there is a canonical isomorphism*

$$\varphi^* \varphi_* V \cong V^{\oplus p}.$$

(e) *For V a differential module over F_ρ, there are canonical bijections*

$$H^i(V) \cong H^i(\varphi_* V) \qquad (i = 0, 1).$$

(f) *For V_1, V_2 differential modules over F_ρ, there is a canonical isomorphism*

$$\varphi_* V_1 \otimes \varphi_* V_2 \cong \bigoplus_{m=0}^{p-1} W_m \otimes \varphi_*(V_1 \otimes V_2).$$

Proof. Exercise. $\square$

4. Frobenius antecedents

Unlike Frobenius descendants, Frobenius antecedents can only be constructed in some cases, namely when the intrinsic radius is sufficiently large.

Definition 9.4.1. Let (V, D) be a finite differential module over F_ρ such that $IR(V) > p^{-1/(p-1)}$. A *Frobenius antecedent* of V is a differential module (V', D') over F'_ρ such that $IR(V') > p^{-p/(p-1)}$, together with an isomorphism $V \cong \varphi^* V'$. By Lemma 9.3.2, a necessary condition for existence of a Frobenius antecedent is that $IR(V) > p^{-1/(p-1)}$; Theorem 9.4.2 below implies that this condition is also sufficient.

Theorem 9.4.2 (after Christol-Dwork). *Let (V, D) be a finite differential module over F_ρ such that $IR(V) > p^{-1/(p-1)}$. Then there exists a unique Frobenius antecedent V' of V. Moreover, $IR(V') = IR(V)^p$.*

Proof of Theorem 9.4.2. We may assume $\zeta_p \in K$, as otherwise we may check everything by adjoining ζ_p and then performing a Galois descent at the end.

We first check existence. Since $|D|_{\mathrm{sp},V} < \rho^{-1}$, for any $x \in V$, we may define an action of $\mathbb{Z}/p\mathbb{Z}$ on V using Taylor series:

$$\zeta_p^m(x) = \sum_{i=0}^{\infty} \frac{(\zeta_p^m t - t)^i}{i!} D^i(x).$$

Take V' to be the fixed space for this action; then V' is an F'_ρ-subspace of V, and the map $\phi^* V' \to V$ is an isomorphism by Hilbert's Theorem 90. (You can also show this explicitly by writing down projectors onto the eigenspaces of V for the $\mathbb{Z}/p\mathbb{Z}$-action.) By applying the $\mathbb{Z}/p\mathbb{Z}$-action to a basis of horizontal sections of V in a generic disc $|t - t_\rho| \leq \lambda\rho$, and invoking Corollary 9.2.3(b), we may construct horizontal sections of V' in a generic disc $|t^p - t_\rho^p| \leq \lambda^p \rho^p$. Hence $IR(V') \geq IR(V)^p > p^{-p/(p-1)}$.

To check uniqueness, suppose $V \cong \varphi^* V' \cong \varphi^* V''$ with $IR(V'), IR(V'') > p^{-p/(p-1)}$. By Lemma 9.3.6, we have

$$\varphi_* V \cong \oplus_{m=0}^{p-1} (V' \otimes W_m) \cong \oplus_{m=0}^{p-1} (V'' \otimes W_m).$$

For $m = 1, \dots, p-1$, we have $IR(W_m) = p^{-p/(p-1)}$; since $IR(V') > IR(W_m)$, we have $IR(V' \otimes W_m) = p^{-p/(p-1)}$. Since $IR(V'') > p^{-p/(p-1)}$, the factor $V'' \otimes W_0$ must be contained in $V' \otimes W_0$ and vice versa.

For the last assertion, note that the proof of existence gives $IR(V') \geq IR(V)^p$, whereas Lemma 9.3.2 gives the reverse inequality. $\square$

Corollary 9.4.3. *Let V' be a differential module over F'_ρ such that $IR(V') > p^{-p/(p-1)}$. Then V' is the Frobenius antecedent of $\varphi^* V'$, so $IR(V') = IR(\varphi^* V')^p$.*

Proof. By Lemma 9.3.2, $IR(\varphi^* V') \geq IR(V')^{1/p}$, so $\varphi^* V'$ has a unique Frobenius antecedent by Theorem 9.4.2. Since $IR(V') > p^{-p/(p-1)}$, V' is that antecedent. $\square$

The construction of Frobenius antecedents carries over to discs and annuli as follows.

Theorem 9.4.4. *Let M be a finite differential module over $K\langle \alpha/t, t/\beta \rangle$ (we may allow $\alpha = 0$), such that $IR(M \otimes F_\rho) > p^{-1/(p-1)}$ for $\rho \in [\alpha, \beta]$ (or equivalently, for $\rho = \alpha$ and $\rho = \beta$). Then there exists a unique differential module M' over $K\langle \alpha^p/t^p, t^p/\beta^p \rangle$ such that $M = M' \otimes K\langle \alpha/t, t/\beta \rangle$ and $IR(M' \otimes F'_\rho) > p^{-p/(p-1)}$ for $\rho \in [\alpha, \beta]$; this M' also satisfies $IR(M' \otimes F'_\rho) = IR(M \otimes F_\rho)^p$ for $\rho \in [\alpha, \beta]$.*

Proof. For existence and the last assertion, use the $\mathbb{Z}/p\mathbb{Z}$-action as in the proof of Theorem 9.4.2. (Note that the proof does not apply directly when $\alpha = 0$; we must make a separate calculation on a disc around the origin on which M is trivial.) For uniqueness, apply Theorem 9.4.2 for any single $\rho \in [\alpha, \beta]$. $\square$

5. Frobenius descendants and subsidiary radii

We saw in Lemma 9.3.2 that we can only weakly control the behavior of generic radius of convergence under Frobenius pullback. Under Frobenius pushforward, we can do much better; we can control not only the generic radius of convergence, but also the subsidiary radii.

Theorem 9.5.1. *Let V be a finite differential module over F_ρ with intrinsic subsidiary radii $s_1, \dots, s_n$. Then the intrinsic subsidiary radii of $\varphi_* V$ comprise the multiset*

$$\bigcup_{i=1}^{n} \begin{cases} \{s_i^p,\ p^{-p/(p-1)}\ (p-1 \text{ times})\} & s_i > p^{-1/(p-1)} \\ \{p^{-1} s_i\ (p \text{ times})\} & s_i \leq p^{-1/(p-1)}. \end{cases}$$

In particular,

$$IR(\varphi_* V) = \min\{p^{-1} IR(V), p^{-p/(p-1)}\}.$$

Proof. It suffices to consider V irreducible. First suppose $IR(V) > p^{-1/(p-1)}$. Let V' be the Frobenius antecedent of V (as per Theorem 9.4.2); note that V' is also irreducible. By Lemma 9.3.6, $\varphi_* V \cong \oplus_{m=0}^{p-1} (V' \otimes W_m)$. Since each W_m has rank 1, $V' \otimes W_m$ is also irreducible. Since $IR(V') = IR(V)^p$ and $IR(V' \otimes W_m) = p^{-p/(p-1)}$ for $m \neq 0$, we have the claim.

Next suppose $IR(V) \leq p^{-1/(p-1)}$. We first show that

$$IR(\varphi_* V) \geq p^{-1} IR(V) = \max\{IR(V)^p, p^{-1} IR(V)\}.$$

For t_ρ a generic point of radius ρ and $\lambda \in (0, p^{-1/(p-1)})$, the module $\varphi_* V \otimes L\langle (t^p - t_\rho^p)/(p^{-1}\lambda\rho^p) \rangle$ splits as the direct sum of $V \otimes L\langle (t - \zeta_p^m t_\rho)/(\lambda\rho) \rangle$ over $m = 0, \dots, p-1$. If $\lambda < IR(V)$, by applying Corollary 9.2.3(c), we obtain $IR(\varphi_* V) \geq p^{-1}\lambda$.

Next, let W' be any irreducible subquotient of $\varphi_* V$; then $IR(W') \geq IR(\varphi_* V)$, so Lemma 9.3.2 gives

(9.5.1.1)
$$IR(\varphi^* W') \geq \min\{IR(W')^{1/p}, p IR(W')\} \geq \min\{IR(\varphi_* V)^{1/p}, p IR(\varphi_* V)\} \geq IR(V).$$

On the other hand, $\varphi^* W'$ is a subquotient of $\varphi^* \varphi_* V$, which by Lemma 9.3.6 is isomorphic to $V^{\oplus p}$. Since V is irreducible, each Jordan-Hölder constituent of $\varphi^* W'$ must be isomorphic to

V, yielding $IR(\varphi^* W') = IR(V)$. That forces each inequality in (9.5.1.1) to be an equality; in particular, $IR(W')$ and $IR(\varphi_* V)$ have the same image under the injective map $s \mapsto \min\{s^{1/p}, ps\}$. We conclude that $IR(W') = IR(\varphi_* V) = p^{-1} IR(V)$, proving the claim. $\square$

REMARK 9.5.2. One might be tempted to think that the proof that $IR(\varphi_* V) \geq p^{-1} IR(V)$ in the proof of Theorem 9.5.1 should carry over to the case $IR(V) > p^{-1/(p-1)}$, in which case it would lead to the false conclusion $IR(\varphi_* V) \geq IR(V)^p$. What breaks down in this case is that pushing forward a basis of local horizontal sections of V only gives you $(\dim V)$ local horizontal sections of $\varphi_* V$; what they span is precisely the Frobenius antecedent of V.

COROLLARY 9.5.3. *Let $s_1 \leq \cdots \leq s_n$ be the intrinsic subsidiary radii of V.*

(a) *For i such that $s_i \leq p^{-1/(p-1)}$, the product of the pi smallest intrinsic subsidiary radii of $\varphi_* V$ is equal to $p^{-pi} s_1^p \cdots s_i^p$.*

(b) *For i such that either $i = n$ or $s_{i+1} \geq p^{-1/(p-1)}$, the product of the $pi + (p-1)(n-i)$ smallest intrinsic subsidiary radii of $\varphi_* V$ is equal to $p^{-ni} s_1^p \cdots s_i^p$.*

In particular, the product of the intrinsic subsidiary radii of $\varphi_* V$ is $p^{-np} s_1^p \cdots s_n^p$.

Note that both conditions apply when $s_i = p^{-1/(p-1)}$; this will be important later.

6. Decomposition by spectral norm

We now extend the decomposition by spectral norm across the barrier $|d|_{F_\rho}$. This cannot be done using Frobenius antecedents alone, as they give no information in case $IR(V) = p^{-1/(p-1)}$.

PROPOSITION 9.6.1. *Let V_1, V_2 be irreducible finite differential modules over F_ρ with $IR(V_1) \neq IR(V_2)$. Then $H^1(V_1 \otimes V_2) = 0$.*

PROOF. By dualizing if necessary, we can ensure that $IR(V_2) > IR(V_1)$. If $IR(V_1) < p^{-1/(p-1)}$, then any short exact sequence $0 \to V_2 \to V \to V_1^\vee \to 0$ splits by the original decomposition theorem.

Suppose that $IR(V_1) = p^{-1/(p-1)}$. Let V_2' be the Frobenius antecedent of V_2; it is also irreducible, and $IR(V_2') = IR(V_2)^p > p^{-p/(p-1)}$. By Theorem 9.5.1, each irreducible subquotient W of $\varphi_* V_1$ satisfies $IR(W) = p^{-p/(p-1)}$; hence $H^1(W \otimes V_2') = 0$ by the previous case, so $H^1(\varphi_* V_1 \otimes V_2') = 0$ by the snake lemma.

By Lemma 9.3.6,

$$\begin{aligned} \varphi_* V_1 \otimes \varphi_* V_2 &\cong \oplus_{m=0}^{p-1} (\varphi_* V_1 \otimes W_m \otimes V_2') \\ &\cong (\varphi_* V_1 \otimes V_2')^{\oplus p}. \end{aligned}$$

(The last isomorphism uses the fact that $\varphi_* V_1 \cong \varphi_* V_1 \otimes W_m$.) This yields $H^1(\varphi_* V_1 \otimes \varphi_* V_2) = 0$; since $\varphi_*(V_1 \otimes V_2)$ is a direct summand of $\varphi_* V_1 \otimes \varphi_* V_2$ (again by Lemma 9.3.6), $H^1(\varphi_*(V_1 \otimes V_2)) = 0$. By Lemma 9.3.6 once more, $H^1(V_1 \otimes V_2) = H^1(\varphi_*(V_1 \otimes V_2)) = 0$.

In the general case, $1 \geq IR(V_2) > IR(V_1)$. If $IR(V_1) > p^{-1/(p-1)}$, then Theorem 9.4.2 implies that V_1, V_2 have Frobenius antecedents V_1', V_2', and that any extension $0 \to V_1 \to V \to V_2^\vee \to 0$ itself is the pullback of an extension $0 \to V_1' \to V' \to (V_2')^\vee \to 0$. To show that any extension of the first type splits, it suffices to do so for the second type; that is, we may reduce from V_1, V_2 to V_1', V_2'. By repeating this enough times, we get to a situation where $IR(V_1) \leq p^{-1/(p-1)}$. We may then apply the previous cases. $\square$

From here, the proof of the following theorem is purely formal.

Theorem 9.6.2 (Strong decomposition theorem). *Let V be a finite differential module over F_ρ. Then there exists a decomposition*

$$V = \bigoplus_{s \in (0,1]} V_s$$

where every subquotient W_s of V_s satisfies $IR(W_s) = s$.

Proof. We induct on $\dim V$; we need only consider V not irreducible. Choose a short exact sequence $0 \to U_1 \to V \to U_2 \to 0$ with U_2 irreducible. Split $U_1 = \oplus_{s \in (0,1]} U_{1,s}$ where every subquotient W_s of $U_{1,s}$ satisfies $IR(W_s) = s$. For each $s \neq IR(U_2)$, we have $H^1(U_2^\vee \otimes U_{1,s}) = 0$ by repeated application of Proposition 9.6.1 plus the snake lemma. Consequently, we have

$$V = V' \oplus \bigoplus_{s \neq IR(U_2)} U_{1,s},$$

where $0 \to U_{1,IR(U_2)} \to V' \to U_2 \to 0$ is exact. □

As with the original decomposition theorem, we obtain the following corollaries.

Corollary 9.6.3. *Let V be a finite differential module over F_ρ whose intrinsic subsidiary radii are all less than* 1. *Then $H^0(V) = H^1(V) = 0$.*

Corollary 9.6.4. *With $V = \oplus_{s \in (0,1]} V_s$ as in Theorem 9.6.2, we have $H^i(V) = H^i(V_1)$ for $i = 0, 1$.*

This suggests that the difficulties in computing H^0 and H^1 arise in the case of intrinsic generic radius 1. We will pursue a closer study of this case in Chapter 12.

Corollary 9.6.5. *If V_1, V_2 are irreducible and $IR(V_1) < IR(V_2)$, then every irreducible subquotient W of $V_1 \otimes V_2$ satisfies $IR(W) = IR(V_1)$.*

Proof. Decompose $V_1 \otimes V_2 = \oplus_{s \in (0,1]} V_s$ according to Theorem 9.6.2; we have $V_s = 0$ whenever $s < IR(V_1)$. If some V_s with $s > IR(V_1)$ were nonzero, then $V_1 \otimes V_2$ would have an irreducible submodule of intrinsic radius greater than $IR(V_1)$, in violation of Lemma 8.3.4. □

7. Integrality, or lack thereof

It may be useful to keep in mind the following limited integrality result for the intrinsic generic radius of convergence. (There should be a more refined statement covering also subsidiary radii.)

Theorem 9.7.1. *Let V be a finite differential module over F_ρ with intrinsic subsidiary radii $s_1, \dots, s_n$. Let m be the largest integer such that $s_m = IR(V)$. Then for any nonnegative integer h,*

$$s_1 > p^{p^{-h}/(p-1)} \qquad \Longrightarrow \qquad s_1^m \in |F^\times|^{p^{-h}} \rho^{\mathbb{Z}}.$$

Proof. For $m = 0$, we read this off from a Newton polygon. We reduce from m to $m-1$ by applying φ_* and invoking Theorem 9.5.1. □

The exponent p^{-h} is not spurious; here is an example to illustrate why it cannot be removed.

Example 9.7.2. Pick $\lambda \in K^\times$ and $0 < \alpha \leq \beta$ such that for $\rho \in [\alpha, \beta]$,

$$p^{1/(p-1)} < |\lambda|\rho^{-p} < p^{p/(p-1)}.$$

Let M be the differential module over $K\langle \alpha/t, t/\beta \rangle$ generated by v satisfying $D(v) = -p\pi\lambda t^{-p-1}$. Then $M \cong \varphi^* M'$, where M' is the differential module over $K\langle \alpha^p/t^p, t^p/\beta^p\rangle$ with generator w and $D'(w) = -\pi\lambda(t^p)^{-2}$. We read off

$$|D'|_{M' \otimes F'_\rho} = p^{-1/(p-1)}|\lambda|\rho^{-2p} > \rho^{-p}.$$

Hence we have

$$\begin{aligned} R(M' \otimes F'_\rho) &= |\lambda|^{-1}\rho^{2p} \\ R(M \otimes F_\rho) &= |\lambda|^{-1/p}\rho^2, \end{aligned}$$

where the first equality follows by Theorem 5.5.3 and the second follows from the first by Corollary 9.4.3.

8. Off-centered Frobenius descendants

Since pushing forward along Frobenius does not work well on a disc, we must also consider "off-centered" Frobenius descendants. This can be done rather more generally, but we will stick to one case sufficient for our purposes.

Definition 9.8.1. For $\rho \in (p^{-1/(p-1)}, 1]$, let F''_ρ be the completion of $K((t-1)^p - 1)$ under the ρ^p-Gauss norm, or equivalently, under the restriction of the ρ-Gauss norm on $K(t)$. (One could allow $K((t-\mu)^p - \mu^p)$ for any $\mu \in K$ of norm 1, but there is no loss of generality in rescaling t to reduce to the case $\mu = 1$.) For brevity, write $u = (t-1)^p - 1$. Equip F''_ρ with the derivation

$$d'' = \frac{d}{du} = \frac{1}{du/dt} d.$$

Given a differential module V'' over F''_ρ, we may view $\psi^* V'' = V'' \otimes F_\rho$ as a differential module over F_ρ. Given a differential module V over F_ρ, we may view the restriction $\psi_* V$ of V along $F''_\rho \to F_\rho$ as a differential module over F''_ρ.

We may apply Lemma 9.2.2 with η replaced by $\eta + 1$, keeping in mind that $|\eta + 1| = 1$ for $|\eta| \leq 1$. This has the net effect that everything that holds for φ also holds for ψ, except that intrinsic radius must be replaced by generic radius.

Theorem 9.8.2. *Let (V, D) be a finite differential module over F_ρ such that $R(V) > p^{-1/(p-1)}$. Then there exists a unique differential module (V'', D'') over F''_ρ such that $V \cong \psi^* V''$ and $R(V'') > p^{-p/(p-1)}$. For this V'', one has in fact $R(V'') = R(V)^p$.*

Theorem 9.8.3. *Let V be a finite differential module over F_ρ with extrinsic subsidiary radii $s_1, \dots, s_n$. Then the subsidiary radii of $\psi_* V$ comprise the multiset*

$$\bigcup_{i=1}^n \begin{cases} \{s_i^p,\ p^{-p/(p-1)}\ (p-1 \text{ times})\} & s_i > p^{-1/(p-1)} \\ \{p^{-1}s_i\ (p \text{ times})\} & s_i \leq p^{-1/(p-1)}. \end{cases}$$

Note that one cannot expect Theorem 9.8.3 to hold for $\rho < p^{-1/(p-1)}$, as in that case $p^{-p/(p-1)}$ is too big to appear as a subsidiary radius of $\psi_* V$.

Notes

Lemma 9.2.2 is taken from [**Ked05a**, §5.3] with some typos corrected.

The Frobenius antecedent theorem of Christol-Dwork [**CD94**, Théorème 5.4] is slightly weaker than the one given here: it only applies for $IR(V) > p^{-1/p}$. The discrepancy is created by the introduction of cyclic vectors, which create some regular singularities which can only eliminated under the stronger hypothesis. Much closer to the statement of Theorem 9.4.2 is [**Ked05a**, Theorem 6.13]; the only difference is that uniqueness is only asserted when $IR(V') \geq IR(V)^p$.

The concept of the Frobenius descendant, and the results deduced using it, are original. This includes Theorem 9.5.1, Theorem 9.8.3, and the strong decomposition theorem (Theorem 9.6.2).

Exercises

(1) Prove Lemma 9.3.6.
(2) Prove that for any finite differential module V' over F'_ρ with $IR(V') > p^{-p/(p-1)}$, $H^0(V') = H^0(\varphi^* V')$.
(3) Here is a result of Dwork related to Example 9.7.2. Suppose $\pi \in K$ satisfies $\pi^{p-1} = -p$. Prove that the power series $E(t) = \exp(\pi t - \pi t^p)$ has radius of convergence $p^{(p-1)/p^2}$, even though the series $\exp(\pi t)$ has radius of convergence 1.

CHAPTER 10

Variation of generic and subsidiary radii

In this chapter, we study the variation of the generic radius of convergence, and the subsidiary radii, associated to a differential module on a disc or annulus.

Throughout this chapter, we retain Notation 9.0.1.

1. Harmonicity of the valuation function

For $f \in K\langle \alpha/t, t/\beta \rangle$ and $r \in [-\log \beta, -\log \alpha]$, the function $r \mapsto v_r(f)$ is continuous, piecewise affine, and (by Proposition 7.2.3(c)) concave in r. However, one can make an even more precise statement; for simplicity, we only write this out explicitly for $r = 0$.

Definition 10.1.1. For $\overline{\mu} \in (\kappa_K^{\mathrm{alg}})^\times$, let μ be a lift of $\overline{\mu}$ in some complete extension L of K. Let E be the completion of $\mathfrak{o}_K[t]_{(t)} \otimes_{\mathfrak{o}_K} K$ for the 1-Gauss norm. For $\alpha \leq 1 \leq \beta$, define the substitution

$$T_\mu : K\langle \alpha/t, t/\beta \rangle \to E, \qquad t \mapsto t + \mu.$$

The function $r \mapsto v_r(T_\mu(f))$ on $[0, \infty)$ is continuous and piecewise affine; moreover, its right slope at $r = 0$ does not depend on choice of the field L or of the lift μ of $\overline{\mu}$. We call this slope $s_{\overline{\mu}}(f)$. For $1 < \beta$ (resp. $\alpha < 1$), define $s_\infty(f)$ (resp. $s_0(f)$) to be the left (resp. right) slope of the function $r \mapsto v_r(f)$.

We then have the following harmonicity property.

Proposition 10.1.2. *For $0 \leq \alpha < 1 < \beta$ and $f \in K\langle \alpha/t, t/\beta \rangle$, we have*

$$s_\infty(f) = \sum_{\overline{\mu} \in \kappa_K^{\mathrm{alg}}} s_{\overline{\mu}}(f).$$

Proof. Without loss of generality, we may assume that $|f|_1 = 1$. The quotient of $\mathfrak{o}_{F_1} \cap K\langle \alpha/t, t/\beta \rangle$ by the ideal generated by $\mathfrak{m}_F$ is isomorphic to $\kappa_K[t, t^{-1}]$; let $\overline{f}$ be the image of f in this quotient. Then $s_{\overline{\mu}}$ is the order of vanishing of $\overline{f}$ at $\overline{\mu}$, whereas s_∞ is the negative of the order of vanishing of $\overline{f}$ at ∞. The desired equality then follows from the fact that a rational function has as many zeroes as poles (counted with multiplicity). $\square$

Remark 10.1.3. Note that $s_{\overline{\mu}}(f) \geq 0$ for $\overline{\mu} \neq 0$; thus Proposition 10.1.2 does indeed recover the concavity inequality $s_\infty \geq s_\mu$.

2. Variation of Newton polygons

Before proceeding to differential modules, we study the variation of the Newton polygon of a polynomial over $K\langle \alpha/t, t/\beta \rangle$ when measured with respect to different Gauss valuations. We begin with this both because it motivates the statements of the results for differential modules, and because it will be used heavily in the proofs of those statements.

THEOREM 10.2.1. *Let $P \in K\langle \alpha/t, t/\beta \rangle[T]$ be a polynomial of degree n. For $r \in [-\log \beta, -\log \alpha]$, put $v_r(\cdot) = -\log |\cdot|_{e^{-r}}$. Let $\mathrm{NP}_r(P)$ be the Newton polygon of P under v_r. Let $f_1(P,r), \dots, f_n(P,r)$ be the slopes of $\mathrm{NP}_r(P)$ in increasing order. For $i = 1, \dots, n$, put $F_i(P,r) = f_1(P,r) + \cdots + f_i(P,r)$.*

(a) *(Linearity) For $i = 1, \dots, n$, the functions $f_i(P,r)$ and $F_i(P,r)$ are continuous and piecewise affine in r.*

(b) *(Integrality) If $i = n$ or $f_i(r_0) < f_{i+1}(r_0)$, then the slopes of $F_i(P,r)$ in some neighborhood of $r = r_0$ belong to $\mathbb{Z}$. Consequently, the slopes of each $f_i(P,r)$ and $F_i(P,r)$ belong to $\frac{1}{1}\mathbb{Z} \cup \cdots \cup \frac{1}{n}\mathbb{Z}$.*

(c) *(Superharmonicity) Suppose that $\alpha < 1 < \beta$. For $i = 1, \dots, n$, let $s_{\infty,i}(P)$ and $s_{0,i}(P)$ be the left and right slopes of $F_i(P,r)$ at $r = 0$. For $\overline{\mu} \in \kappa_K^{\mathrm{alg}}$, let $s_{\overline{\mu},i}(P)$ be the right slope of $F_i(T_\mu(P),r)$ at $r = 0$. Then*

$$s_{\infty,i}(P) \geq \sum_{\overline{\mu} \in \kappa_K^{\mathrm{alg}}} s_{\overline{\mu},i}(P),$$

with equality if $i = n$ or $f_i(P,0) < f_{i+1}(P,0)$.

(d) *(Monotonicity) Suppose that P is monic and $\alpha = 0$. For $i = 1, \dots, n$, the slopes of $F_i(P,r)$ are nonnegative.*

(e) *(Concavity) Suppose that P is monic. For $i = 1, \dots, n$, the function $F_i(P,r)$ is concave.*

PROOF. Write $P = \sum_{i=0}^n P_i T^i$ with $P_i \in K\langle \alpha/t, t/\beta \rangle$. The function $v_r(P_i)$ is continuous in r and piecewise affine with slopes in $\mathbb{Z}$; by Proposition 7.2.3(c), it is also concave.

For $s \in \mathbb{R}$ and $r \in [-\log \beta, -\log \alpha]$, put

$$v_{s,r}(P) = \min_i \{v_r(P_i) + is\};$$

that is, $v_{s,r}(P)$ is the y-intercept of the supporting line of $\mathrm{NP}_r(P)$ of slope s. Since $v_{s,r}(P)$ is the minimum of finitely many functions of the pair (r,s), each of which is continuous, piecewise affine, and concave, these are also true of $v_{s,r}(P)$.

Note that $F_i(P,r)$ is the difference between the y-coordinates of the points of $\mathrm{NP}_r(P)$ of x-coordinates $i - n$ and $-n$. That is,

$$F_i(P,r) = \sup_s \{v_{s,r}(P) - (n-i)s\} - v_r(P_n). \tag{10.2.1.1}$$

Moreover, the supremum in (10.2.1.1) is achieved by some s whose denominator is bounded by n. Consequently, $F_i(P,r)$ is continuous and piecewise affine, proving (a).

If $i = n$ or $f_i(P,r_0) < f_{i+1}(P,r_0)$, then the point of $\mathrm{NP}_{r_0}(P)$ of x-coordinate $i - n$ is a vertex, and likewise for r in some neighborhood of r_0. In that case, for r near r_0,

$$F_i(P,r) = v_r(P_{n-i}) - v_r(P_n), \tag{10.2.1.2}$$

proving (b).

Assume that $\alpha < 1 < \beta$. Then Proposition 10.1.2 implies that

$$s_\infty(P_i) = \sum_{\overline{\mu} \in \kappa_K^{\mathrm{alg}}} s_{\overline{\mu}}(P_i) \qquad (i = 0, \dots, n).$$

If $i = n$ or $f_i(P, 0) < f_{i+1}(P, 0)$, then this plus (10.2.1.2) yields that the desired inequality is in fact an equality. Otherwise, let j, k be the least and greatest indices for which $f_j(0) = f_i(0) = f_k(0)$; then $j < i < k$, and the convexity of the Newton polygon implies

$$F_i(P,r) \geq \frac{k-i}{k-j} F_j(P,r) + \frac{i-j}{k-j} F_k(P,r), \tag{10.2.1.3}$$

with equality for $r = 0$. From this plus piecewise affinity, we deduce (c).

Assume that $\alpha = 0$ and that P is monic. Then each $v_r(P_i)$ is a nondecreasing function of r, as then is each $v_{s,r}(P)$. Since $v_r(P_n) = 0$, $F_r(P, r)$ is nondecreasing by (10.2.1.1), proving (d).

To prove (e), one can reduce to working locally around $r = 0$ and then deduce the claim from (c) and (d) (because the latter implies $s_{\overline{\mu},i}(P) \geq 0$ for $\overline{\mu} \neq 0$). However, one can also prove (e) directly as follows. Assume that P is monic, so that $P_n = 1$ and (10.2.1.1) reduces to

$$F_1(P,r) = \sup_s \{v_{s,r}(P) - (n-i)s\}.$$

It is not immediately clear from this that $F_i(P, r)$ is concave, since we are taking the supremum rather than the infimum of a collection of concave functions. To get around this, pick $r_1, r_2 \in [-\log \beta, -\log \alpha]$ and put $r_3 = ur_1 + (1-u)r_2$ for some $u \in [0, 1]$. For $j \in \{1, 2\}$, choose s_j achieving the supremum in (10.2.1.1) for $r = r_j$. Put $s_3 = us_1 + (1-u)s_2$; then using the convexity of $v_{s,r}(P)$ in both s and r, we have

$$\begin{aligned}
F_i(P, r_3) &\geq v_{s_3,r_3}(P) - (n-i)s_3 \\
&\geq u(v_{s_1,r_1}(P) - (n-i)s_1) + (1-u)(v_{s_2,r_2}(P) - (n-i)s_2) \\
&= uF_i(P, r_1) + (1-u)F_i(P, r_2).
\end{aligned}$$

This yields concavity for $F_i(P, r)$, proving (e).

□

REMARK 10.2.2. A more geometric interpretation of the previous proof can be given by writing each $P_i = \sum_j P_{i,j} t^j$ and considering the lower convex hull of the set of points $\{(-i, -j, v(P_{i,j}))\}$ in $\mathbb{R}^3$; we leave elaboration of this point to the reader.

REMARK 10.2.3. It should also be noted that if $i = n$ or $f_i(r_0) < f_{i+1}(r_0)$, then (10.2.1.2) implies that

$$f_1(r_0) + \cdots + f_i(r_0) \in v(K^\times) + \mathbb{Z} r_0.$$

This fact does not analogize to subsidiary radii, because one has to replace $v(K^\times)$ by its p-divisible closure. See Theorem 9.7.1 and Example 9.7.2.

3. Variation of subsidiary radii: statements

In order to state the analogue of Theorem 10.2.1 for subsidiary radii of a differential module on a disc or annulus, we must set some corresponding notation.

NOTATION 10.3.1. Let M be a finite free differential module of rank n over $K\langle \alpha/t, t/\beta \rangle$. For $\rho \in [\alpha, \beta]$, let $R_1(\rho), \ldots, R_n(\rho)$ be the extrinsic subsidiary radii of $M \otimes F_\rho$ in increasing order, so that $R_1(M,\rho) = R(M \otimes F_\rho)$ is the generic radius of convergence of $M \otimes F_\rho$. For $r \in [-\log\beta, -\log\alpha]$, define

$$f_i(M,r) = -\log R_i(M, e^{-r}),$$

so that $f_i(M,r) \geq r$ for all r. Put $F_i(M,r) = f_1(M,r) + \cdots + f_i(M,r)$.

We now have the following results, whose proofs are distributed across the remainder of this chapter (Lemmas 10.6.1, 10.7.1, 10.7.3, 10.8.1). Note that there is an overall sign discrepancy from Theorem 10.2.1, so that concavity becomes convexity and so forth.

Theorem 10.3.2. *Let M be a finite free differential module of rank n over $K\langle\alpha/t, t/\beta\rangle$.*

(a) *(Linearity) For $i = 1,\dots,n$, the functions $f_i(M,r)$ and $F_i(M,r)$ are continuous and piecewise affine.*

(b) *(Integrality) If $i = n$ or $f_i(M,r_0) > f_{i+1}(M,r_0)$, then the slopes of $F_i(M,r)$ in some neighborhood of r_0 belong to $\mathbb{Z}$. Consequently, the slopes of each $f_i(M,r)$ and $F_i(M,r)$ belong to $\frac{1}{1}\mathbb{Z} \cup \cdots \cup \frac{1}{n}\mathbb{Z}$.*

(c) *(Subharmonicity) Suppose that $\alpha < 1 < \beta$ and that $f_i(M,0) > 0$. For $i = 1,\dots,n$, let $s_{\infty,i}(M)$ and $s_{0,i}(M)$ be the left and right slopes of $F_i(M,r)$ at $r = 0$. For $\overline{\mu} \in \kappa_K^{\mathrm{alg}}$, let $s_{\overline{\mu},i}(M)$ be the right slope of $F_i(T_\mu^*(M),r)$ at $r = 0$. Then*

$$s_{\infty,i}(M) \leq \sum_{\overline{\mu}\in\kappa_K^{\mathrm{alg}}} s_{\overline{\mu},i}(M),$$

with equality if either $i = n$ and $f_n(0) > 0$, or $i < n$ and $f_i(M,0) > f_{i+1}(M,0)$.

(d) *(Monotonicity) Suppose that $\alpha = 0$. For $i = 1,\dots,n$, for any point r_0 where $f_i(M,r_0) > r_0$, the slopes of $F_i(M,r)$ are nonpositive in some neighborhood of r_0. (Remember that $f_i(r) = r$ for r sufficiently large.)*

(e) *(Convexity) For $i = 1,\dots,n$, the function $F_i(M,r)$ is convex.*

4. Convexity for the generic radius

As a prelude to tackling Theorem 10.3.2, we give a quick proof of subharmonicity, monotonicity, and convexity (parts (c)-(e) of Theorem 10.3.2) for the function f_1, corresponding to the generic radius of convergence. This argument applies to both discs and annuli, and can be used in place of the full strength of Theorem 10.3.2 for many purposes; indeed, this is true for numerous results which predate Theorem 10.3.2. See the notes for further details.

Proof of Theorem 10.3.2(c), (d), (e) for $i = 1$. Choose a basis of M, and let D_s be the basis via which D^s acts on M. Then recall from Lemma 5.2.5 that

$$R_1(M,\rho) = \min\{\rho, p^{-1/(p-1)} \liminf_{s\to\infty} |D_s|_\rho^{-1/s}\}.$$

For each s, the function $r \mapsto -\log|D_s|_{e^{-r}}^{-1/s}$ is convex in r by Proposition 7.2.3(c). This implies the convexity of

$$f_1(M,r) = \max\{r, \frac{1}{p-1}\log p + \limsup_{s\to\infty}(-\log|D_s|_{e^{-r}}^{-1/s})\}.$$

Similarly, we deduce (c) by applying Proposition 10.1.2 to each D_s. If $\alpha = 0$, then the function $r \mapsto -\log|D_s|_{e^{-r}}^{-1/s}$ is nonincreasing, yielding (d). $\square$

REMARK 10.4.1. To improve upon this result, one might like to try to read off the generic radius of convergence, and maybe even the other subsidiary radii, from the Newton polygon of a cyclic vector. In order to do this, we have to overcome two obstructions.

(a) One can only construct cyclic vectors in general for differential modules over differential fields, not over differential rings.
(b) Some of the subsidiary radii may be greater than $p^{-1/(p-1)}\rho$, in which case Newton polygons will not detect them.

The first problem will be addressed by using a cyclic vector over a fraction field to establish linearity and integrality, then comparing to a carefully chosen lattice to deduce convexity, subharmonicity, and monotonicity. The second problem will be addressed using Frobenius descendants.

5. Finding lattices

One key step in what follows is, given a finite free module over $K\langle \alpha/t, t/\beta \rangle$ and a basis of the extension of the module to a differential field, find a basis of the original module which is close to the original, in the sense that the supremum norms defined by the two bases differ by a small multiplicative factor in either direction. The following lemma produces such bases.

LEMMA 10.5.1 (Lattice lemma). *Let F be a complete extension of K, let R be a complete K-subalgebra of F, and put $R' = R \cap \mathfrak{o}_F$. Let M be a finite free R-module of rank n, and let $|\cdot|_M$ be a norm on $M \otimes F$ compatible with F. Assume that either:*

(a) *$c > 1$ and the value group of K is not discrete; or*
(b) *$c \geq 1$, the value group of K is discrete, and the value groups of K, F, M all coincide.*

Then there exists a norm $|\cdot|'_M$ on $M \otimes F$ such that $\{m \in M : |m|'_M \leq 1\}$ is a finite free R'-module of rank n, and $c^{-1}|m|_M \leq |m|'_M \leq c|m|_M$ for all $m \in M$.

PROOF. We induct on n. Pick any $m_1 \in M$ belonging to a basis of M, so that $M_1 = M/Rm_1$ is also free. Using (a) or (b), we can rescale m_1 by an element of K to force $1 \leq |m_1|_M \leq c^{2/3}$.

Equip M_1 with the quotient norm

$$|x_1|_{M_1} = \inf_{x \in M : x + M_1 = x_1} \{|x|_M\};$$

this is a norm because M_1 is a closed subspace of M. Moreover, in case (b), the infimum is always achieved, so the quotient norm again satisfies (b). Apply the induction hypothesis to choose a basis $m_{2,1}, \dots, m_{n,1}$ of M_1 such that the supremum norm $|\cdot|'_{M_1}$ defined by $m_{2,1}, \dots, m_{n,1}$ satisfies $c^{-1/3}|x_1|_{M_1} \leq |x_1|'_{M_1} \leq c^{1/3}|x_1|_{M_1}$ for all $x_1 \in M_1$. For $i = 2, \dots, n$, choose $m_i \in M$ lifting $m_{i,1}$ such that $|m_i|_M \leq c^{1/3}|m_{i,1}|_{M_1} \leq c^{2/3}$.

Let $|\cdot|'_M$ be the supremum norm defined by $m_1, \dots, m_n$. For $a_1, \dots, a_n \in R'$, we have

$$|a_1 m_1 + \cdots + a_n m_n|_M \leq \max_{1 \leq i \leq n} \{|a_i||m_i|_M\} \leq c^{2/3} \leq c.$$

On the other hand, if $m \in M$ satisfies $|m|_M \leq 1$, we can uniquely write $m = a_1 m_1 + \cdots + a_n m_n$ with $a_i \in R$. By definition of the quotient norm, $|m|_{M_1} \leq 1$, so $|m|'_{M_1} \leq c^{1/3}$. In other words, $|a_2|, \dots, |a_n| \leq c^{1/3}$, so

$$|a_2 m_2 + \cdots + a_n m_n|_M \leq \max_{2 \leq i \leq n} \{|a_i||m_i|_M\} \leq c^{1/3} c^{2/3} = c.$$

Since $|m|_M \leq 1 \leq c$, we have $|a_1 m_1|_M \leq c$. Since $|m_1|_M \geq 1$, we have $|a_1| \leq c$. This proves the desired inequalities. $\square$

REMARK 10.5.2. Although we will only apply Lemma 10.5.1 in the case where the original norm $|\cdot|$ is the supremum norm associated to some basis, it is not convenient to prove it under this extra hypothesis. That is because the construction of the quotient norm does not preserve the property of the norm being generated by a basis.

6. Measuring small radii

In this section, we address concern (a) from Remark 10.4.1.

LEMMA 10.6.1. *For any $i \in \{1, \dots, n\}$ and any r_0 such that $f_i(M, r_0) > r_0 + 1/(p-1)\log p$, Theorem 10.3.2 holds in a neighborhood of r_0.*

PROOF. Put $F = \operatorname{Frac} K\langle \alpha/t, t/\beta \rangle$. Choose a cyclic vector for $M \otimes F$ to obtain an isomorphism $M \otimes F \cong F\{T\}/F\{T\}P$ for some monic twisted polynomial P over F. We may then apply Theorem 10.2.1 to deduce (a), (b), (c).

It remains to prove (d), as we may then deduce (e) from (c) and (d) as noted in the proof of Theorem 10.2.1. To deduce (d), we may work in a neighborhood of a single value r_0 of r. There is no harm in enlarging K, so we may assume $v(K^\times) = \mathbb{R}$. Then we may reduce to the case $r_0 = 0$ by replacing t by λt for some $\lambda \in K^\times$.

Pick $\lambda_1, \dots, \lambda_n \in K$ such that

$$-\log|\lambda_j| = \min\{1/(p-1)\log p - f_j(M, 0), 0\} \qquad (j = 1, \dots, n).$$

By Proposition 3.3.10, the characteristic polynomial of the action of D on the basis B_0 of $M \otimes F_1$ given by

$$\lambda_n^{-1} \cdots \lambda_{n-j+1}^{-1} T^i \qquad (j = 0, \dots, n-1)$$

has eigenvalues of norms $\max\{p^{-1/(p-1)} e^{f_j(M,0)}, 1\}$ for $j = 1, \dots, n$. By Lemma 10.5.1, for any particular $c > 1$, we may construct a basis $m_1, \dots, m_n$ of M such that the supremum norm defined by B_0 differs from the supremum norm defined by the chosen basis of $M \otimes F_1$ by a multiplicative factor of at most c.

Let N be the matrix via which D acts on $m_1, \dots, m_n$. For $c > 1$ sufficiently small, Theorem 5.7.4 implies that for r close to 0, the visible spectrum of $M \otimes F_{e^{-r}}$ is the multiset of those norms of eigenvalues of the characteristic polynomial of N which exceed e^{-r}. We may then deduce (d) from Theorem 10.2.1. $\square$

7. Larger radii

We next address concern (b) from Remark 10.4.1, considering the cases $f_i(M, r_0) > r_0$ and $f_i(M, r_0) = r_0$ separately.

LEMMA 10.7.1. *For any $i \in \{1, \dots, n\}$ and any r_0 such that $f_i(M, r_0) > r_0$, clauses (a), (b), (c), (e) of Theorem 10.3.2 hold in a neighborhood of r_0.*

PROOF. For each nonnegative integer j, we prove the claim for r_0 such that $f_i(M, r_0) > r_0 + p^{-j}/(p-1)\log p$, by induction on j; the base case $j = 0$ is precisely Lemma 10.6.1. As in the proof of Lemma 10.6.1, we may reduce to the case $r_0 = 0$.

Let $R'_1(\rho^p), \dots, R'_n(\rho^p)$ be the subsidiary radii of $\varphi_* M \otimes F'_\rho$ in increasing order. (The normalization is chosen this way because the series variable in F'_ρ is t^p, which has norm ρ^p.) Put $g_i(r) = -\log R'_i(e^{-r})$. By Theorem 9.5.1, the list $g_1(pr), \dots, g_{pn}(pr)$ consists of

$$\bigcup_{i=1}^{n} \begin{cases} \{pf_i(M,r), pr + \frac{p}{p-1}\log p \ (p-1 \text{ times})\} & f_i(M,r) \leq r + 1/(p-1)\log p \\ \{\log p + (p-1)r + f_i(M,r) \ (p \text{ times})\} & f_i(M,r) \geq r + 1/(p-1)\log p. \end{cases}$$

Thus we may deduce (a) from the induction hypothesis.

To check (b), (c), (e), it suffices to handle cases where $i = n$ or $f_i(M,0) > p^{-j}/(p-1)\log p$. (As in the proof of Theorem 10.2.1(c), we may linearly interpolate to establish convexity and subharmonicity in the other cases.) In these cases, we have either $f_i(M,0) > 1/(p-1)\log p$, in which case in some neighborhood of $r = 0$ we have

$$g_1(pr) + \cdots + g_{pi}(pr) = pF_i(M,r) + pi\log p + (p-1)ipr, \tag{10.7.1.1}$$

or $f_{i+1}(M,0) < 1/(p-1)\log p$ or $i = n$, in which case in some neighborhood of $r = 0$ we have

$$g_1(pr) + \cdots + g_{pi+(p-1)(n-i)}(pr) = pF_i(M,r) + pn\log p + (p-1)npr. \tag{10.7.1.2}$$

Moreover, $f_i(M,0) > p^{-j}/(p-1)\log p$ if and only if $g_{pi}(0) > p^{-j+1}/(p-1)\log p$.

If $f_i(M,0) > 1/(p-1)\log p$, apply (10.7.1.1) and the induction hypothesis to write piecewise

$$\begin{aligned} F_i(M,r) &= p^{-1}(g_1(pr) + \cdots + g_{pi}(pr) - pi\log p - (p-1)ipr) \\ &= p^{-1}(m(pr) + *) \\ &= mr + p^{-1}* \end{aligned}$$

for some $m \in \mathbb{Z}$. (Note that $*$ is not guaranteed to be in $p \cdot v(K^\times)$; this explains Example 9.7.2.) If $f_i(M,0) \leq 1/(p-1)\log p$, then $f_{i+1}(M,0) < 1/(p-1)\log p$, so we may apply (10.7.1.2) to write piecewise

$$\begin{aligned} F_i(M,r) &= p^{-1}(g_1(pr) + \cdots + g_{pi+(p-1)(n-i)}(pr) - pn\log p - (p-1)npr) \\ &= p^{-1}(m(pr) + *) \\ &= mr + p^{-1}* \end{aligned}$$

for some $m \in \mathbb{Z}$. This is also applicable when $i = n$. $\square$

Remark 10.7.2. In the proof of Lemma 10.7.1, note the importance of the fact that the domains of applicability of (10.7.1.1) and (10.7.1.2) overlap: if $f_i(M,0) = 1/(p-1)\log p$, then (10.7.1.1) may not remain applicable when we move from $r = 0$ to a nearby value.

Lemma 10.7.3. *For any $i \in \{1, \dots, n\}$ and any r_0 such that $f_i(M, r_0) = r_0$, Theorem 10.3.2 holds in a neighborhood of r_0.*

Proof. As in the proof of Lemma 10.6.1, it suffices to consider the case $r_0 = 0$. We first check continuity. For this, note that the proofs of Lemma 10.6.1 and 10.7.1 show that for any $c > 0$, the function $\max\{f_i(M,r), r + c\}$ is continuous at $r = 0$. Consequently, for any $\epsilon > 0$, we can find $0 < \delta < \epsilon/2$ such that

$$|\max\{f_i(M,r), r + \epsilon/4\}| < \epsilon/2 \qquad (|r| < \delta).$$

For such r, $-\epsilon < -\delta < f_i(M,r) < \epsilon$; this yields continuity. x We next check piecewise affinity by induction on i. Given that $f_1(M,r), \dots, f_{i-1}(M,r)$ are linear in a one-sided neighborhood of $r = 0$, say $[-\delta, 0]$, and given $f_i(M,0) = 0$, it suffices to check linearity of $f_i(M,r) - r$ in some $[-\delta', 0]$. From what we know already, in a neighborhood of each $r \in [-\delta, 0]$ where $f_i(M,r) - r > 0$, $f_i(M,r) - r$ is convex and piecewise affine with slopes in $\frac{1}{n!}\mathbb{Z}$. Note that none of these slopes can be nonnegative, as otherwise $f_i(M,r) - r$ would thereafter be nondecreasing and could not have limit 0 at $r = 0$. By the same argument, if $f_i(M,r_0) - r_0 = 0$ for some $r_0 \in [-\delta, 0)$, then the slope of $f_i(M,r) - r$ at any point $r \in (r_0, 0)$ with $f_i(M,r) - r > 0$ must simultaneously be positive and negative; since this cannot occur, we must have $f_i(M,r) - r = 0$ for all $r \in [r_0, 0]$.

If $f_i(M,r) - r = 0$ for some $r < 0$, we are then done, as $f_i(M,r) - r$ is constant in a one-sided neighborhood of 0. Otherwise, the slopes of $f_i(M,r) - r$ in $[-\delta, 0)$ form a sequence of discrete values which are negative and nondecreasing. This sequence must then stabilize, so $f_i(M,r) - r$ is linear in a one-sided neighborhood of 0. This proves (a).

To prove (b), note that when $f_i(M,0) = 0$, the input hypothesis can only hold if $i = n$. Suppose we wish to check integrality of the right slope of F_n (the argument for the left slope is analogous). If $f_1(M,r) - r, \dots, f_n(M,r) - r$ are identically zero in a right neighborhood of 0, then we have nothing to check. Otherwise, let j be the greatest integer such that $f_j(M,r) - r$ is not identically zero in a right neighborhood of 0; we then deduce (b) by applying Lemma 10.7.1 with i replaced by j.

Since (c) and (d) make no assertion at $r = 0$ in case $f_i(0) = 0$, it remains to check (e), which we do by induction on i. Given that $F_{i-1}(M,r)$ is convex and that $f_i(M,0) = 0$, it suffices to check that $f_i(M,r) - r$ is convex in a neighborhood of 0. But we already know that $f_i(M,r) - r$ is continuous and piecewise affine near 0, and that it only takes nonnegative values; it must then have nonpositive left slope and nonnegative right slope, and so must be convex near 0. This proves (e). $\square$

8. Monotonicity

To complete the proof of Theorem 10.3.2, we must prove (d) without the restriction $f_i(r_0) > r_0 + 1/(p-1)\log p$. The reason why we do not have (d) as part of Lemma 10.7.1 is that passing from M to $\varphi_* M$ introduces a singularity at $t = 0$, so we cannot hope to infer monotonicity on $\varphi_* M$. To fix this, we must use off-centered Frobenius descendants.

Lemma 10.8.1. *If $\alpha = 0$ and $f_i(M,r_0) > r_0$, then the slope of $f_i(M,r)$ in a right neighborhood of r_0 is nonpositive.*

Proof. We proceed as in the proof of Lemma 10.7.1, but using the off-centered Frobenius ψ instead of φ. Again, we may assume $r_0 = 0$ and that $i = n$ or $f_i(0) > f_{i+1}(0)$ (reducing to the latter case by linear interpolation).

Let $R_1''(\rho^p), \dots, R_n''(\rho^p)$ be the subsidiary radii of $\psi_* M \otimes F_\rho''$ in increasing order. Put $g_i(r) = -\log R_i''(e^{-r})$. By Theorem 9.8.3, if $f_i(M,0) > 1/(p-1)\log p$, then

$$g_1(pr) + \cdots + g_{pi}(pr) = pF_i(M,r) + pi \log p,$$

whereas if $f_{i+1}(M,0) < 1/(p-1)\log p$ or $i = n$, then

$$g_1(pr) + \cdots + g_{pi+(p-1)(n-i)}(pr) = pF_i(M,r) + pn \log p.$$

Moreover, $f_i(M,0) > p^{-j}/(p-1)\log p$ if and only if $g_{pi}(0) > p^{-j+1}/(p-1)\log p$. We may thus proceed as in Lemma 10.7.1 to conclude. $\square$

EXAMPLE 10.8.2. To see in action the discrepancy between the behavior of the centered and off-centered Frobenius descendants, we consider an example suggested by Liang Xiao. (All verifications are left as an exercise.) Take $\beta > 1$, and let M be the differential module over $K\langle t/\beta\rangle$ with a single generator v satisfying $D(v) = t^{p-1}v$. Pick any $\alpha < 1$, so that we may form ϕ_*M on $K\langle \alpha/t^p, t^p/\beta\rangle$. Then ϕ_*M splits as $\oplus_{m=0}^{p-1}(M' \otimes W_m)$, where M' has a single generator v' satisfying $D'(v') = p^{-1}v$. One then computes for $m \neq 0$ and $\overline{\mu} \in \kappa_K^{\mathrm{alg}}$,

$$\begin{aligned}
s_{\infty,1}(M') &= 0\\
s_{\overline{\mu},1}(M') &= 0\\
s_{\infty,1}(M' \otimes W_m) &= 0\\
s_{0,1}(M' \otimes W_m) &= 1\\
s_{-m,1}(M' \otimes W_m) &= -1\\
s_{\overline{\mu},1}(M' \otimes W_m) &= 0 \qquad (\overline{\mu} \neq 0, -m).
\end{aligned}$$

This yields

$$\begin{aligned}
s_{\infty,p}(\phi_*M) &= 0\\
s_{0,p}(\phi_*M) &= p-1\\
s_{\overline{\mu},p}(\phi_*M) &= -1 \qquad (\overline{\mu} \in \mathbb{F}_p^\times)\\
s_{\overline{\mu},p}(\phi_*M) &= 0 \qquad (\overline{\mu} \notin \mathbb{F}_p)
\end{aligned}$$

and in turn

$$\begin{aligned}
s_{\infty,1}(M) &= -p+1\\
s_{0,1}(M) &= 0\\
s_{\overline{\mu},1}(M) &= -1 \qquad (\overline{\mu} \in \mathbb{F}_p^\times)\\
s_{\overline{\mu},1}(M) &= 0 \qquad (\overline{\mu} \notin \mathbb{F}_p).
\end{aligned}$$

9. Radius versus generic radius

As promised, we can recover some information about radius of convergence from the properties of generic radius of convergence.

PROPOSITION 10.9.1. *Let M be a differential module over $K\langle t/\beta\rangle$ for some $\beta > 0$. Then the radius of convergence of M equals e^{-r}, for r the smallest value such that $f_1(r) = r$. Consequently, $f(r') = r'$ for all $r' \geq r$.*

PROOF. By Theorem 8.5.1, the radius of convergence of M is at least the generic radius of convergence of $M \otimes F_{e^{-r}}$, which by hypothesis equals e^{-r}. On the other hand, if $\lambda > e^{-r}$, then by hypothesis $f_1(-\log\lambda) > -\log\lambda$, or in other words $R(M \otimes F_\lambda) < \lambda$. This means that $M \otimes K\langle t/\lambda\rangle$ cannot be trivial, so the radius of convergence cannot exceed λ. This proves the desired result. $\square$

Corollary 10.9.2. *Let M be a differential module over $K\langle t/\beta\rangle$ for some $\beta > 0$. Then the radius of convergence of M belongs to the divisible closure of the multiplicative value group of K.*

Proof. By Theorem 10.3.2 and Theorem 9.7.1, the function $f_1(r)$ is piecewise of the form $ar+b$ with $a \in \mathbb{Q}$ and $b \in p^{-\infty}v(K^\times)$. By Proposition 10.9.1, the radius of convergence of M equals e^{-r} for r the smallest value such that $f_1(r) = r$. To the left of this r, f_1 must be piecewise affine with slope $\neq 1$; by comparing the left and right limits at r, we deduce that $r = ar + b$ for some $a \neq 1$ rational and some $b \in p^{-\infty}v(K^\times)$. Since this gives $r = b/(a-1)$, we deduce the claim. □

One should be able to better control the denominators, as in the following question.

Question 10.9.3. *Let M be a differential module over $K\langle t/\beta\rangle$ for some $\beta > 0$. Does there necessarily exist $j \in \{1, \dots, \mathrm{rank}(M)\}$ such that the j-th power of the radius of convergence of M belongs to the p-divisible closure of the multiplicative value group of K?*

We also have a criterion for when the radius of convergence equals the generic radius.

Corollary 10.9.4. *Let M be a differential module over $K\langle t/\beta\rangle$ for some $\beta > 0$, such that for some $\alpha \in (0,\beta)$, $R(M \otimes F_\rho)$ is constant for $\rho \in [\alpha, \beta]$. Then $R(M) = R(M \otimes F_\rho)$.*

10. Subsidiary radii as radii of convergence

The subsidiary generic radii of convergence can be interpreted as the radii of convergence of a well-chosen basis of local horizontal sections at a generic point. The argument is a variation on Corollary 10.9.4.

Definition 10.10.1. Let M be a differential module of rank n over $K\langle t/\beta\rangle$, or on the open disc of radius β. For $i = 1, \dots, n$, the i-th *subsidiary radius of convergence* of M is the supremum of those $\lambda \in [0, \beta)$ for which there exist $n-i$ linearly independent horizontal sections of $M \otimes K\langle t/\lambda\rangle$. Note that there exists a basis of local horizontal sections $s_1, \dots, s_n$ of M such that s_i has radius of convergence equal to the i-th subsidiary radius: once $s_i, \dots, s_n$ have been chosen, there must be at least a one-dimensional space of choices left for s_{i-1}. Such a basis is sometimes called an *optimal basis* of local horizontal sections.

The following generalizes Proposition 8.6.4.

Theorem 10.10.2 (after Young). *Let (V, D) be a differential module over F_ρ of dimension n with subsidiary (generic) radii $s_1 \leq \cdots \leq s_n$, and let V' be the base change of V to the open disc of radius ρ in $t - t_\rho$ over L. Then the subsidiary radii of V' are also $s_1 \leq \cdots s_n$.*

Proof. We first produce a basis for which $\rho_i = s_i$ for $i = 1, \dots, n$. For this, we may apply Theorem 9.6.2 to decompose V into components each with a single subsidiary radius, and thus reduce to the case $s_1 = \cdots = s_n = s$. By the geometric interpretation of the generic radius (Proposition 8.6.4), each Jordan-Hölder constituent of V admits a basis of local horizontal sections on a generic disc of radius s. By Lemma 5.2.7, the same is true for V itself.

For the remaining inequality, we induct on n. Let m be the largest integer such that $s_1 = s_m$. Let V_1 be the component of V of subsidiary radius s_1, so that $\dim V_1 = m$. We will check that no local horizontal section of V_1 at t_ρ can have radius of convergence strictly greater than s_1.

Put

$$f_i(r) = f_i(V_1 \otimes L\langle (t - t_\rho)/e^{-r}\rangle, r) \qquad (i = 1, \dots, m; r \in (-\log \rho, \infty));$$

then the $f_i(r)$ behave as in Theorem 10.3.2. By the proof of Theorem 10.3.2(d), the $f_i(r)$ are constant in a neighborhood of $r = -\log \rho$. By Theorem 10.3.2(c) and (e) and induction on i,

$$f_i(r) = \begin{cases} -\log s_i & 0 < r \leq -\log s_i \\ r & r \geq -\log s_i. \end{cases}$$

By contrast, if there were a local horizontal section of V_1 at t_ρ which converged on a closed disc of radius λ for some $\lambda \in (s_1, \rho)$, then $V_1 \otimes L\langle (t - t_\rho)/\lambda\rangle$ would have a trivial submodule, and so would have λ as one of its subsidiary radii. This would force $f_n(r) = r$ for $r = -\log \lambda < -\log s_i$, contradiction.

We conclude that any local horizontal section of V that projects nontrivially onto V_1 has radius strictly greater than s_1. We can divide the given basis into m sections that project onto a basis of V_1, and $n - m$ sections that project onto a basis of the complementary component. The first m sections have radius of convergence at most s_1 by above; the others have radii of convergence bounded by $s_{m+1}, \dots, s_n$ by the induction hypothesis. This yields the desired result. □

Notes

The harmonicity property of functions on annuli (Proposition 10.1.2) may be best viewed inside a theory of subharmonic functions on one-dimensional Berkovich analytic spaces. Such a theory has been developed by Thuillier [**Thu05**].

For the function $f_1(M, r) = F_i(M, r)$ representing the generic radius of convergence, Christol and Dwork established convexity [**CD94**, Proposition 2.4] (using essentially the same short proof given here) and continuity at endpoints [**CD94**, Théorème 2.5] (see also [**DGS94**, Appendix I]). The analogous results for the higher $F_i(M, r)$ are original.

When restricted to intrinsic subsidiary radii less than $p^{-1/(p-1)}$, Theorem 10.10.2 is a result of Young [**You92**, Theorem 3.1]. Young's proof is an explicit calculation using twisted polynomials and cyclic vectors.

Exercises

(1) Given an example to show that in Theorem 10.2.1, f_2 need not be concave (even though f_1 and f_2 are concave).
(2) Prove that if K is discretely valued, then $\mathfrak{o}_K\langle t\rangle = F_1 \cap K\langle t\rangle$ is noetherian. It isn't otherwise, because then $\mathfrak{o}_K$ itself is not noetherian.
(3) Prove that each maximal ideal of $\mathfrak{o}_K\langle t\rangle$ is generated by $\mathfrak{m}_K$ together with some $P \in \mathfrak{o}_K[t]$ whose reduction modulo $\mathfrak{m}_K$ is irreducible in $\kappa_K[t]$.
(4) Verify Example 10.8.2.

CHAPTER 11

Decomposition by subsidiary radii

In this chapter, we show that one can sometimes decompose a differential module on a disc according to a separation of the subsidiary radii of convergence.

Besides Notation 9.0.1, we also retain Notation 10.3.1.

1. Metrical detection of units

One can identify the units in $K\langle\alpha/t, t/\beta\rangle$ rather easily in terms of power series coefficients (Lemma 7.2.5). However, for the present application, we need an alternate characterization based on more intrinsic data, namely the Gauss norms.

Definition 11.1.1. For $f \in K\langle\alpha/t, t/\beta\rangle$ with $\alpha \leq 1 \leq \beta$, define the *discrepancy* of f at $r = 0$ as the sum

$$\operatorname{disc}(f, 0) = \sum_{\overline{\mu} \in (\kappa_K^{\mathrm{alg}})^\times} s_{\overline{\mu}}(f);$$

note that $\operatorname{disc}(f, 0) \geq 0$. We define $\operatorname{disc}(f, r)$ for general $r \in [-\log\beta, -\log\alpha]$ by rescaling: assume without loss of generality that K contains a scalar c of norm e^{-r}, let $T_c : K\langle\alpha/t, t/\beta\rangle \to K\langle(\alpha e^r)/t, t/(\beta e^r)\rangle$ be the substitution $t \mapsto c^{-1}t$, then put

$$\operatorname{disc}(f, r) = \operatorname{disc}(T_c(f), 0).$$

Lemma 11.1.2. *For $x \in K\langle t/\beta\rangle$ nonzero, x is a unit if and only if* $s_0(x) = \operatorname{disc}(x, -\log\beta) = 0$.

Proof. We may reduce to the case $\beta = 1$ and $|x|_1 = 1$. In this case, by Lemma 7.2.5, x is a unit if and only if its image modulo $\mathfrak{m}_K$ in $\kappa_K[t]$ is a unit. As noted in Proposition 10.1.2, the order of vanishing of this image at $\overline{\mu} \in \kappa_K^{\mathrm{alg}}$ is precisely $s_{\overline{\mu}}(x)$; this proves the claim. $\square$

For annuli, it is more convenient to prove a weak criterion first.

Lemma 11.1.3. *For $x \in \cup_{\alpha \in (0,\beta)} K\langle\alpha/t, t/\beta\rangle$ nonzero, x is a unit if and only if* $\operatorname{disc}(x, -\log\beta) = 0$.

Proof. We again reduce to the case $\beta = 1$ and $|x|_1 = 1$. In this case, by Lemma 7.2.5, x is a unit if and only if its image modulo $\mathfrak{m}_K$ in $\kappa_K[t, t^{-1}]$ is a unit. We then argue as in Lemma 11.1.2. $\square$

One may then deduce the following.

Lemma 11.1.4. *For $x \in K\langle\alpha/t, t/\beta\rangle$ nonzero, x is a unit if and only if the function $r \mapsto v_r(x)$ is affine on $[-\log\beta, -\log\alpha]$, and* $\operatorname{disc}(x, -\log\alpha) = \operatorname{disc}(x, -\log\beta) = 0$.

Proof. It suffices to check that x is a unit in $K\langle \alpha_i/t, t/\beta_i \rangle$ for a finite collection of closed intervals $[\alpha_i, \beta_i]$ with union $[\alpha, \beta]$. However, Lemma 11.1.3 implies that one can cover a one-sided neighborhood of any element of $[\alpha, \beta]$ with such an interval; compactness of $[\alpha, \beta]$ then yields the claim. $\square$

2. Decomposition over a closed disc

We get different-looking results for decomposition by subsidiary radii, depending on whether we are working on a closed disc or a closed annulus. Let us consider the disc first. First, a general definition.

Definition 11.2.1. Let M be a finite differential module over $K\langle \alpha/t, t/\beta$ with $\alpha \leq 1 \leq \beta$. Define the *i-th discrepancy of M at $r = 0$* as

$$\operatorname{disc}_i(M, 0) = -\sum_{\overline{\mu} \in (\kappa_K^{\mathrm{alg}})^\times} s_{\overline{\mu}, i}(M);$$

it is always nonnegative. Extend the definition to general $r \in [-\log \beta, -\log \alpha]$ as in Definition 11.1.1.

Theorem 11.2.2. *Let M be a finite differential module over $K\langle t/\beta \rangle$ of rank n. Suppose that the following conditions hold for some $i \in \{1, \dots, n-1\}$.*

(a) *We have $f_i(M, -\log \beta) > f_{i+1}(M, -\log \beta)$.*
(b) *The function $F_i(M, r)$ is constant for r in a neighborhood of $-\log \beta$.*
(c) *We have $\operatorname{disc}_i(M, -\log \beta) = 0$.*

Then the decomposition of $M \otimes F_\beta$ separating the first i subsidiary radii lifts to a decomposition of M itself.

Before proving Theorem 11.2.2, we record a trivial but useful observation.

Lemma 11.2.3. *Let R, S, T be subrings of a common ring U with $S \cap T = R$. Let M be a finite free R-module. Then the intersection $(M \otimes S) \cap (M \otimes T)$ inside $M \otimes U$ is equal to M itself.*

This also holds when M is only locally free; see exercises.

Remark 11.2.4. The immediate application of Lemma 11.2.3 is to replace K by a complete extension L in Theorem 11.2.2; inside the completion of $L(t)$ for the 1-Gauss norm, we have

$$F_1 \cap L\langle t \rangle = K\langle t \rangle.$$

Thus obtaining matching decompositions of $M \otimes F_1$ and $M \otimes L\langle t \rangle$ gives a corresponding decomposition of M itself.

We also need a lemma about polynomials over $K\langle t \rangle$.

Lemma 11.2.5. *Let $P = \sum_i P_i T^i$ and $Q = \sum_i Q_i T^i$ be polynomials over $K\langle t \rangle$ satisfying the following conditions.*

(a) *We have $|P - 1|_1 < 1$.*
(b) *For $m = \deg(Q)$, Q_m is a unit and $|Q|_1 = |Q_m|_1$.*

Then P and Q generate the unit ideal in $K\langle t \rangle[T]$.

PROOF. We may assume without loss of generality that $Q_m = 1$. The hypothesis on Q implies that if $R \in K\langle t\rangle[T]$ and S is the remainder upon dividing R by Q, then $|S|_1 \leq |Q|_1$ (compare Proposition 4.5.2). If we then set $\delta = |P-1|_1 < 1$ and let S_i denote the remainder upon dividing $(1-P)^i$ by Q, the series $\sum_{i=0}^{\infty} S_i$ converges and its limit S satisfies $PS \equiv 1 \pmod{Q}$. This proves the claim. □

LEMMA 11.2.6. *Theorem 11.2.2 holds if* $f_i(-\log\beta) > 1/(p-1)\log p - \log\beta$.

PROOF. By invoking Remark 11.2.4 to justifying enlarging K, then rescaling, we may reduce to the case $\beta = 1$. Set notation as in the proof of Lemma 10.6.1. Then for $c > 1$ sufficiently small, the coefficient of T^{n-i} in the characteristic polynomial $Q(T)$ of N is a unit in $K\langle t/\rangle$ by Lemma 11.1.2, and we can apply Theorem 2.2.2 to factor $Q = Q_2Q_1$ so that the roots of Q_1 are the i largest roots of Q under $|\cdot|_1$.

Use the basis $m_1, \dots, m_n$ to identify M with $K\langle t\rangle^n$. Then we obtain a short exact sequence

$$0 \to \ker(Q_1(N)) \to M \to \operatorname{coker}(Q_1(N)) \to 0$$

of free modules over $K\langle t\rangle^n$. (The quotient is torsion-free because by Lemma 11.2.5, Q_1 and Q_2 generate the unit ideal in $K\langle t\rangle[T]$.) Applying Lemma 10.5.1 to both factors (again for $c > 1$ sufficiently small), we construct a basis of M on which D acts via a matrix

$$N_1 = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}$$

in which:

(a) The matrix A_i is invertible and $|A_1^{-1}|_1 |d|_1 < 1$.
(b) The Newton slopes of A_i under v_0 account for the first i subsidiary radii of $M \otimes F_1$.
(c) We have $|B_1|_1, |C_1|_1, |D_1|_1 \leq |A_1^{-1}|_1^{-1}\delta$ for some $\delta < 1$.

By Lemma 5.7.1, M admits a differential submodule accounting for the $n-i$ subsidiary radii of $M \otimes F_{e^{-r}}$ for r near 0. By repeating this argument for $M^\vee$, we obtain the desired splitting. □

To prove Theorem 11.2.2 in general, we must use Frobenius antecedents again.

PROOF OF THEOREM 11.2.2. It suffices to prove that for $\beta = 0$, Theorem 11.2.2 holds if $f_i(0) > p^{-j}/(p-1)\log p$ for each nonnegative integer j; we again proceed by induction on j, with base case $j = 0$ provided by Lemma 11.2.6.

Suppose $f_i(0) > p^{-j}/(p-1)\log p$. Let $M_1' \oplus M_2'$ be the decomposition of $\varphi_* M$ separating the subsidiary radii less than or equal to $e^{-pf_i(0)}$ from the others. This might not be induced by a decomposition of M_1, because some factors of subsidiary radius $p^{-p/(p-1)}$ that are needed in M_2' are instead grouped into M_1'. To fix this, consider instead the decomposition

$$((M_1' \otimes W_0) \cap \dots \cap (M_1' \otimes W_{p-1})) \oplus ((M_2' \otimes W_0) + \dots + (M_2' \otimes W_{p-1}));$$

this is induced by a decomposition of M having the desired properties. □

3. Decomposition over a closed annulus

Over an annulus, one has a decomposition theorem of a somewhat different shape. Fortunately, the proof is essentially the same as for Theorem 11.2.2.

THEOREM 11.3.1. *Let M be a finite differential module over $K\langle \alpha/t, t/\beta \rangle$ of rank n. Suppose that the following conditions hold for some $i \in \{1, \dots, n-1\}$.*

(a) *We have $f_i(M,r) > f_{i+1}(M,r)$ for $-\log \beta \leq r \leq -\log \alpha$.*
(b) *The function $f_1(M,r) + \cdots + f_i(M,r)$ is affine for $-\log \beta \leq r \leq -\log \alpha$.*
(c) *We have $\mathrm{disc}_i(M, -\log \beta) = \mathrm{disc}_i(M, -\log \alpha) = 0$.*

Then there is a decomposition of M inducing, for each $\rho \in [\alpha, \beta]$, the decomposition of $M \otimes F_\rho$ separating the first i subsidiary radii from the others.

We first prove a lemma which looks somewhat more like Theorem 11.2.2.

LEMMA 11.3.2. *Let M be a finite differential module over $K\langle \alpha/t, t/\beta \rangle$ of rank n. Suppose that the following conditions hold for some $i \in \{1, \dots, n-1\}$.*

(a) *We have $f_i(M, -\log \beta) > f_{i+1}(M, -\log \beta)$.*
(b) *We have $\mathrm{disc}_i(M, -\log \beta) = 0$.*

Then for some $\gamma \in [\alpha, \beta)$, there is a decomposition of $M \otimes K\langle \gamma/t, t/\beta \rangle$ inducing the decomposition of $M \otimes F_\beta$ separating the first i subsidiary radii from the others.

PROOF. Using Remark 11.2.4 again, we may enlarge K and then reduce to the case $\beta = 1$. Moreover, it suffices to consider the case where $f_i(0) > 1/(p-1)\log p$, as we may reduce the general case to this one as in the proof of Theorem 11.2.2.

Set notation again as in the proof of Lemma 10.6.1. Then for $c > 1$ sufficiently small and $\gamma \in [\alpha, 1)$ sufficiently large, the coefficient of T^{n-i} in the characteristic polynomial $Q(T)$ of N is a unit in $K\langle \gamma/t, t \rangle$ by Lemma 11.1.3, so we may continue as in the proof of Lemma 11.2.6. □

To prove Theorem 11.3.1 from Lemma 11.3.2, we proceed as in the proof of Lemma 11.1.4.

PROOF OF THEOREM 11.3.1. Note that by subharmonicity (Theorem 10.3.2(d)), conditions (b) and (c) together are equivalent to the condition that $\mathrm{disc}_i(M,r) = 0$ for $-\log \beta < r \leq -\log \alpha$. Consequently, if M satisfies the given hypothesis, then so does $M \otimes K\langle \gamma/t, t/\delta \rangle$ for each closed subinterval $[\gamma, \delta] \subseteq [\alpha, \beta]$.

For each $\rho \in (\alpha, \beta]$, Lemma 11.3.2 implies that for some $\gamma \in [\alpha, \rho)$, $M \otimes K\langle \gamma/t, t/\rho \rangle$ admits a decomposition with the desired property. Similarly, for each $\rho \in [\alpha, \beta)$, for some $\gamma \in (\rho, \beta]$, $M \otimes K\langle \rho/t, t/\gamma \rangle$ admits a decomposition with the desired property.

We now have a collection of intervals $[\gamma_i, \delta_i]$ covering $[\alpha, \beta]$ for which $M \otimes K\langle \gamma_i/t, t/\delta_i \rangle$ admits a decomposition with the desired property. By compactness of $[\alpha, \beta]$, we can reduce to a finite collection of intervals. Since the decomposition of $M \otimes K\langle \gamma_i/t, t/\delta_i \rangle$ is uniquely determined by the induced decomposition over F_ρ for any single $\rho \in [\gamma_i, \delta_i]$, these decompositions agree on overlaps of the covering intervals. By the patching lemma (Lemma 7.3.3), we obtain a decomposition of M itself. □

4. Decomposition over an open disc or annulus

Over open discs, we have similar decomposition theorems but without the discrepancy conditions at endpoints.

THEOREM 11.4.1. *Let M be a finite differential module of rank n over the open disc of radius β. Suppose that the following conditions hold for some $i \in \{1, \dots, n-1\}$ and some $\gamma \in (0, \beta)$.*

(a) *The function $F_i(M,r)$ is constant for $-\log\beta < r \leq -\log\gamma$.*
(b) *We have $f_i(M,r) > f_{i+1}(M,r)$ for $-\log\beta < r \leq -\log\gamma$.*

Then M admits a unique decomposition separating the first i subsidiary radii of $M \otimes F_\rho$ for $\rho \in [\gamma,\beta)$.

Proof. Note that (a) and subharmonicity imply that $\mathrm{disc}_i(M,\delta) = 0$ for $\delta \in [\gamma,\beta)$. Thus for any such δ, we may apply Theorem 11.2.2 to $m \otimes K\langle t/\beta\rangle$; doing so for all such δ (or a sequence ascending to β) yields the desired result. $\square$

Similarly, for open annuli, we obtain a decomposition theorem without a discrepancy condition at endpoints.

Theorem 11.4.2. *Let M be a finite differential module of rank n over the open annulus of inner radius α and outer radius β. Suppose that the following conditions hold for some $i \in \{1,\dots,n-1\}$.*

(a) *The function $F_i(M,r)$ is affine for $-\log\beta < r < -\log\alpha$.*
(b) *We have $f_i(M,r) > f_{i+1}(M,r)$ for $-\log\beta < r < -\log\alpha$.*

Then M admits a unique decomposition separating the first i subsidiary radii of $M \otimes F_\rho$ for any $\rho \in (\alpha,\beta)$.

Remark 11.4.3. One can also obtain a decomposition theorem for a half-open annulus, by covering it with an open annulus and a closed annulus, and patching together the decompositions given by Theorem 11.3.1 and Theorem 11.4.2. Similarly, one can obtain decomposition theorems on more exotic subspaces of the affine line by patching; the reader knowledgeable enough to be interested in such statements should at this point have no trouble formulating and deriving them.

5. Modules solvable at a boundary

Definition 11.5.1. Let M be a finite differential module on the half-open annulus with closed inner radius α and open outer radius β. We say M is *solvable at β* if $R(M \otimes F_\rho) \to \beta$ as $\rho \to \beta^-$, or equivalently, if $IR(M \otimes F_\rho) \to 1$ as $\rho \to \beta^-$. (One can also make a similar definition with the roles of the inner and outer radius reversed; we will not refer to that definition here.)

Lemma 11.5.2. *Let M be a finite differential module on the half-open annulus with closed inner radius α and open outer radius β, which is solvable at β. There exist $b_1 \geq \cdots \geq b_n \in [0,\infty)$ such that for $\rho \in [\alpha,\beta)$ sufficiently close to β, the intrinsic subsidiary radii of $M \otimes F_\rho$ are $(\rho/\beta)^{b_1},\dots,(\rho/\beta)^{b_n}$. Moreover, if $i = n$ or $b_i > b_{i+1}$, then $b_1 + \cdots + b_i \in \mathbb{Z}$.*

Proof. For $r \to (-\log\beta)^+$, $F_i(M,r) - ir$ is a convex function with slopes in a discrete subset of $\mathbb{R}$. Moreover, it is nonnegative and its limit is 0; this implies that the slopes are all positive. However, the slopes lie in a discrete subgroup of $\mathbb{R}$, so they must eventually stabilize. We deduce that each f_i is linear in a neighborhood of $-\log\beta$, and we may infer the desired conclusions from the known properties of the f_i provided by Theorem 10.3.2. $\square$

Definition 11.5.3. Let M be a finite differential module on the half-open annulus with closed inner radius α and open outer radius β, which is solvable at β. The quantities $b_1,\dots,b_n$ defined by Lemma 11.5.2 will be called the *differential slopes* of M at β. (They are also called *ramification numbers*; the reason for this will become clear when we consider quasiconstant differential modules in Chapter 17. See specifically Theorem 17.3.5.)

We now recover a decomposition theorem of Christol-Mebkhout; see the notes for further discussion. We will see several applications of this result later in the book.

THEOREM 11.5.4 (Christol-Mebkhout). *Let M be a finite differential module on the half-open annulus with closed inner radius α and open outer radius β, which is solvable at β. Then for any sufficiently large $\gamma \in [\alpha, \beta)$, the restriction of M to the open annulus with inner radius α and outer radius β splits uniquely as a direct sum $\oplus_{b \in [0,\infty)} M_b$, such that for each $b \in [0, \infty)$, for all $\rho \in [\gamma, \beta)$, the intrinsic subsidiary radii of $M_b \otimes F_\rho$ are all equal to $(\rho/\beta)^b$.*

PROOF. By Lemma 11.5.2, we are in a case where Theorem 11.4.2 may be applied. □

REMARK 11.5.5. For some differential module for which one has fairly explicit series expansions for local horizontal sections, one may be able to establish solvability at a boundary by explicit estimates. However, it is more common for solvability at a boundary to be established by proving the existence of a Frobenius structure; this notion will be introduced in Chapter 15.

Notes

Our results on modules solvable at a boundary are originally due to Christol and Mebkhout [**CM00, CM01**]. In particular, Lemma 11.5.2 for the generic radius is [**CM00**, Théorème 4.2.1], and the decomposition theorem (which implies Lemma 11.5.2 in general) is [**CM01**, Corollaire 2.4–1].

The proof technique of Christol and Mebkhout is significantly different from ours: they construct the desired decomposition by exhibiting convergent sequences for a certain topology on the ring of differential operators. This does not appear to give quantitative results; that is, one does not control the range over which the decomposition occurs, although we are not sure whether this is an intrinsic limitation of the method. (Keep in mind that the approach here crucially uses Frobenius descendants, which were not previously introduced.)

Note also that Christol and Mebkhout work directly with a differential module on an open annulus as a ring-theoretic object; this requires a freeness result of the following form. If K is spherically complete, any finite free module on the half-open annulus with closed inner radius α and open outer radius β is induced by a finite free module over the ring $\cap_{\rho \in [\alpha,\beta)} K\langle \alpha/t, t/\rho \rangle$. (That is, any locally free coherent sheaf on this annulus is freely generated by global sections.) For a proof, see for instance [**Ked05a**, Theorem 3.14]. A result of Lazard [**Laz62**] implies that this property, even when restricted to modules of rank 1, is in fact equivalent to spherical completeness of K.

Exercises

(1) Prove the analogue of Lemma 11.2.3 in which M is only required to be locally free.

CHAPTER 12

p-adic exponents

In this chapter, we discuss (without full proofs) what happens when one tries to analyze p-adic differential modules on annuli for which the intrinsic generic radius of convergence is equal to 1 everywhere; this is precisely the case where the techniques of the previous chapters fail to deliver any information. It turns out that there is a notion of p-adic exponents in this setting, but one must avoid exponents which are closely approximated by integers without being integers themselves (p-adic Liouville numbers). This can already be seen by considering p-adic differential modules on discs with one regular singularity, so we do that first.

1. p-adic Liouville numbers

DEFINITION 12.1.1. For $\lambda \in K$, the *type* of λ, denoted $\mathrm{type}(\lambda)$, is the radius of convergence of the p-adic power series

$$\sum_{m=0, m\neq\lambda}^{\infty} \frac{x^m}{\lambda - m}.$$

This cannot exceed 1, as there are infinitely many m for which $|\lambda - m| = 1$ (namely those not congruent to λ modulo p). Moreover, if $\lambda \notin \mathbb{Z}_p$, then $|\lambda - m|$ is bounded below, so $\mathrm{type}(\lambda) = 1$. We will thus mostly worry about $\lambda \in \mathbb{Z}_p$.

DEFINITION 12.1.2. We say that λ is a *p-adic Liouville number* if either λ or $-\lambda$ has type less than 1, and a *p-adic non-Liouville number* otherwise. The reference to both λ and $-\lambda$ is not superfluous, as they may have different types (exercise).

The following alternate characterization of type may be helpful.

DEFINITION 12.1.3. For $\lambda \in \mathbb{Z}_p$, let $\lambda^{(m)}$ be the unique integer in $\{0, \dots, p^m - 1\}$ congruent to λ modulo p^m.

PROPOSITION 12.1.4. *For $\lambda \in \mathbb{Z}_p$ not a nonnegative integer,*

$$-\frac{1}{\log_p \mathrm{type}(\lambda)} = \liminf_{m\to\infty} \frac{\lambda^{(m)}}{m}. \qquad (12.1.4.1)$$

In particular, λ has type 1 if and only if $\lambda^{(m)}/m \to \infty$ as $m \to \infty$.

PROOF. It suffices to check that for $0 < \eta < 1$, we have

$$\limsup_{m\to\infty} (m + \lambda^{(m)} \log_p \eta) = -\infty \qquad (12.1.4.2)$$

when $\eta < \mathrm{type}(\alpha)$ and

$$\limsup_{m\to\infty} (m + \lambda^{(m)} \log_p \eta) = +\infty \qquad (12.1.4.3)$$

when $\eta > \text{type}(\alpha)$. Namely, (12.1.4.2) implies $m + \lambda^{(m)} \log_p \eta \leq 0$ for all large m, so $\liminf_{m\to\infty} \frac{\lambda^{(m)}}{m} \geq -1/(\log_p \eta)$, whereas (12.1.4.3) implies $m + \lambda^{(m)} \log_p \eta \geq 0$ for infinitely many m, so $\liminf_{m\to\infty} \frac{\lambda^{(m)}}{m} \leq -1/(\log_p \eta)$.

Suppose first that $\text{type}(\lambda) > \eta > 0$; then as $s \to \infty$, $\eta^s/|\lambda - s| \to 0$ or equivalently $v_p(\lambda - s) + s \log_p \eta \to -\infty$. (Here v_p denotes the renormalized valuation with $v(p) = 1$.) Since λ is not a nonnegative integer, we have $\lambda^{(m)} \to \infty$ as $m \to \infty$, so

$$v_p(\lambda - \lambda^{(m)}) + \lambda^{(m)} \log_p \eta \to -\infty.$$

The left side does not increase if we replace $v_p(\lambda - \lambda^{(m)})$ by m, so we may deduce (12.1.4.2).

Suppose next that $\text{type}(\lambda) < \eta < 1$; then we may choose a sequence s_j such that as $j \to \infty$, $v_p(\lambda - s_j) + s_j \log_p \eta \to +\infty$. Put $m_j = v_p(\lambda - s_j)$, so that $s_j \geq \lambda^{(m_j)}$. Then

$$m_j + \lambda^{(m_j)} \log_p \eta \to +\infty,$$

yielding (12.1.4.3). □

The alternate characterization is convenient for such verifications as the fact that rational numbers are non-Liouville (exercise), or this stronger result [**DGS94**, Proposition VI.1.1], whose proof we omit.

Proposition 12.1.5. *Any element of $\mathbb{Z}_p$ algebraic over $\mathbb{Q}$ is non-Liouville.*

We will encounter the p-adic Liouville property in yet another apparently different form. (See exercises for an alternate proof of this lemma.)

Lemma 12.1.6. *For λ not a nonnegative integer, we have an equality of formal power series*

$$\sum_{m=0}^{\infty} \frac{x^m}{\lambda(1-\lambda)(2-\lambda)\cdots(m-\lambda)} = e^x \sum_{m=0}^{\infty} \frac{(-x)^m}{m!} \frac{1}{\lambda - m}.$$

Proof. The coefficient of x^m on the right side is a sum of the form $\sum_{i=0}^{m} c_i/(i-\lambda)$ for some $c_i \in \mathbb{Q}$. It is thus a rational function of λ of the form $P(\lambda)/(\lambda(1-\lambda)\cdots(m-\lambda))$, where P has coefficients in $\mathbb{Q}$ and degree at most m. To check that in fact $P(\lambda) = 1$ identically, we need only check this for $\lambda = 0, \dots, m$.

In other words, to check the original identity, it suffices to check after multiplying both sides by $\lambda - i$ and evaluating at $\lambda = i$, for each nonnegative integer i. On the left side, we obtain

$$\sum_{m=i}^{\infty} \frac{-x^m}{(-1)^{i-1} i!(m-i)!}.$$

On the right side, we obtain

$$e^x \frac{(-x)^i}{i!},$$

which is the same thing. □

Corollary 12.1.7. *If $\lambda \in K$ is not a nonnegative integer, and* $\text{type}(\lambda) = 1$*, then the series*

$$\sum_{m=0}^{\infty} \frac{x^m}{\lambda(1-\lambda)(2-\lambda)\cdots(m-\lambda)}$$

has radius of convergence $p^{-1/(p-1)}$.

2. p-adic regular singularities

We now consider a p-adic analogue of Theorem 6.3.5. Unlike its archimedean analogue, it requires a hypothesis on exponents beyond simply being weakly prepared (which simply meant that no two eigenvalues of the constant matrix differ by a nonzero integer).

Definition 12.2.1. We say that a finite set is *p-adic non-Liouville* if its elements are p-adic non-Liouville number. We say the set has *p-adic non-Liouville differences* if the difference between any two elements of the set is a p-adic non-Liouville number.

Theorem 12.2.2 (p-adic Fuchs theorem). *For $\beta > 0$, let M be a finite differential module on $K\langle t/\beta\rangle$ for the derivation $d = t\frac{d}{dt}$. Let $N = \sum_{i=0}^{\infty} N_i t^i$ be the action of D on some basis. Assume that N_0 has eigenvalues which are weakly prepared and have p-adic non-Liouville differences. Then there exists $\gamma > 0$ such that the fundamental solution matrix for N has entries in $K\langle t/\gamma\rangle$ (as does its inverse).*

Proof. We proceed as in Proposition 16.1.1. Recall (6.3.4.1):

$$N_0U_i - U_iN_0 + iU_i = -\sum_{j=1}^{i} N_jU_{i-j} \qquad (i > 0).$$

Because N_0 has weakly prepared eigenvalues, U is uniquely determined. There is thus no harm in enlarging K to ensure that the eigenvalues $\lambda_1, \dots, \lambda_n$ of N_0 belong to K. Then the map $X \mapsto N_0X - XN_0 + i$ has eigenvalues $\lambda_g - \lambda_h + i$ for $g, h \in \{1, \dots, n\}$. If e is the maximum number of pairwise equal eigenvalues, we obtain the bound

$$|U_i|\beta^i \le \max_{g,h}\{|\lambda_g - \lambda_h + i|^{-2e+1}\}|N|_\beta \max_{j<i}\{|U_j|\beta^j\}.$$

Thus to conclude the theorem, it suffices to verify that for each $h, j \in \{1, \dots, n\}$, the number $\lambda = \lambda_g - \lambda_h$ has the property that

$$\prod_{i=1}^{m} \max\{1, |\lambda - i|^{-1}\}$$

grows at worst exponentially.

If $\lambda \notin \mathbb{Z}_p$, then $|\lambda - i|^{-1}$ is bounded above and the claim is verified. Otherwise, Corollary 12.1.7 and the hypothesis that λ is a p-adic non-Liouville number give the desired estimate. □

By a slight modification of the argument (which we omit), one may obtain the following result of Clark [**Cla66**, Theorem 3].

Theorem 12.2.3 (Clark). *Let M be a finite differential module over $K\langle t/\beta\rangle$ for the derivation $t\frac{d}{dt}$, with a regular singularity at 0 whose exponents are p-adic non-Liouville numbers. Then for any $x \in M$ and $y \in M \otimes K[\![t]\!]$ such that $Dy = x$, we have $Dy \in M \otimes K\langle t/\rho\rangle$ for some $\rho > 0$.*

The p-adic non-Liouville hypothesis in Theorem 12.2.2 turns out not to be superfluous, as demonstated by the following example of Monsky.

Example 12.2.4. Consider the rank 2 differential module over $K\langle t\rangle$ for the derivation $t\frac{d}{dt}$ associated to the differential polynomial differential polynomial $p(1-t)T^2 - tT - a$, where $a \in \mathbb{Z}_p$ is constructed so that

$$\text{type}(a) = 1, \qquad \text{type}(-a) < 1. \tag{12.2.4.1}$$

(The existence of such a is left as an exercise, or see [**DR77**, §7.20].) It can then be shown that the conclusion of Theorem 12.2.2 fails for the basis $1, T$ of M, that is, the fundamental solution matrix does not converge in any disc. (The eigenvalues of N_0 are $0, a$, so the hypothesis of non-Liouville differences is violated by this example.) See [**DR77**, §7] or [**DGS94**, §IV.8] for further discussion.

3. The Robba condition

We are interested in the question: given a finite differential module on an annulus for the derivation $t\frac{d}{dt}$, under what circumstances is it necessarily isomorphic to a differential module which can be defined over a disc?

In order to answer this question, we must identify properties of a differential module on a disc which betray information about the exponents, but which are defined in terms of information away from the center of the disc.

Definition 12.3.1. Let M be a finite differential module on the disc/annulus $|t| \in I$, for I an interval. We say that M satsifies the *Robba condition* if $IR(M \otimes F_\rho) = 1$ for all nonzero $\rho \in I$.

Proposition 12.3.2. *Let M be a finite differential module on the open disc of radius β for the derivation $t\frac{d}{dt}$, satisfying the Robba condition in some annulus. Then the exponents of the action of D on M/tM belong to $\mathbb{Z}_p$.*

Proof. Let $N = \sum_{i=0}^\infty N_i t^i$ be the matrix via which D acts on some basis of M. Suppose N_0 has an eigenvalue $\lambda \notin \mathbb{Z}_p$; there is no harm in enlarging K to force $\lambda \in K$. Choose $v \in M$ such that the image of v in M/tM is a nonzero eigenvector of N_0 of eigenvalue λ. Let D' be the derivation corresponding to $\frac{d}{dt}$ instead of $t\frac{d}{dt}$. Then with notation as in Example 8.2.5, we have for any $\rho < \beta$,

$$\liminf_{s\to\infty} |(D')^s v|^{1/s} > |D'|_{\mathrm{sp},V_\lambda,\rho} > p^{-1/(p-1)}\rho,$$

so $IR(M \otimes F_\rho) < 1$. □

We will establish a partial converse to Proposition 12.3.2 later (Theorem 12.7.1). In the interim, we mention the following easy result.

Proposition 12.3.3. *Let M be a finite differential module on the open disc of radius β for the derivation $t\frac{d}{dt}$, such that the action of D on some basis of M is given by a matrix N_0 over K. Then M satisfies the Robba condition if and only if N_0 has eigenvalues in $\mathbb{Z}_p$.*

Proof. Exercise, or see [**DGS94**, Corollary IV.7.6]. □

4. Abstract p-adic exponents

We now consider the question: given a finite differential module on an annulus for the derivation $t\frac{d}{dt}$ satisfying the Robba condition, if it is isomorphic to a differential module over a disc, how do we read off the exponents of that module by looking only at the original annulus?

The answer to this question is complicated by the fact that the exponents are only well-defined as elements of the quotient $\mathbb{Z}_p/\mathbb{Z}$. This means we cannot hope to identify them using purely p-adic considerations; in fact, we must use *archimedean* considerations to identify them. Here are those considerations.

DEFINITION 12.4.1. We will say that two elements $A, B \in \mathbb{Z}_p^n$ are *equivalent* if there exists a permutation σ of $\{1, \dots, n\}$ such that $A_i - B_{\sigma(i)} \in \mathbb{Z}$ for $i = 1, \dots, n$. This is evidently an equivalence relation.

DEFINITION 12.4.2. We say that $A, B \in \mathbb{Z}_p^n$ are *weakly equivalent* if there exists a constant $c > 0$, a sequence $\sigma_1, \sigma_2, \dots$ of permutations of $\{1, \dots, n\}$, and signs $\epsilon_{i,m} \in \{\pm 1\}$ such that

$$(\epsilon_{i,m}(A_i - B_{\sigma_m(i)}))^{(m)} \leq cm \qquad (i = 1, \dots, n; m = 1, 2, \dots).$$

In other words, the distance from $A_i - B_{\sigma_m(i)}$ to the nearest multiple of p^m is at most cm. Again, this is clearly an equivalence relation, and equivalence implies weak equivalence.

LEMMA 12.4.3. *If $A, B \in \mathbb{Z}_p$ (regarded as 1-tuples) are weakly equivalent, then they are equivalent.*

PROOF. For some $c > 0$, we have

$$|\epsilon_{1,m+1}(\epsilon_{1,m+1}(A - B))^{(m+1)} - \epsilon_{1,m}(\epsilon_{1,m}(A - B))^{(m)}| \leq 2cm + c,$$

and the left side is an integer divisible by p^m. For m large enough, we have $p^m > 2cm + c$ and so

$$\epsilon_{1,m+1}(\epsilon_{1,m+1}(A - B))^{(m+1)} = \epsilon_{1,m}(\epsilon_{1,m}(A - B))^{(m)}.$$

Hence for m large enough, $\epsilon_{1,m}$ is constant and $\epsilon_{1,m}(A - B)$ is a nonnegative integer. $\square$

COROLLARY 12.4.4. *Suppose $A \in \mathbb{Z}_p^n$ is weakly equivalent to hA for some positive integer h. Then $A \in (\mathbb{Z}_p \cap \mathbb{Q})^n$.*

PROOF. We are given that for some $c > 0$, some permutations σ_m, and some signs $\epsilon_{i,m}$,

$$(\epsilon_{i,m}(A_i - hA_{\sigma_m(i)}))^{(m)} \leq cm.$$

The order of σ_m divides $n!$, so we have

$$(\pm(A_i - h^{n!}A_i))^{(m)} \leq n!cm$$

for some choice of sign (depending on i, m). That is, for each i, the 1-tuple consisting of $(h^{n!}-1)A_i$ is weakly equivalent to zero. By Lemma 12.4.3, $(h^{n!}-1)A_i \in \mathbb{Z}$, so $A_i \in \mathbb{Z}_p \cap \mathbb{Q}$. $\square$

PROPOSITION 12.4.5. *Suppose that $A, B \in \mathbb{Z}_p^n$ are weakly equivalent and that B has p-adic non-Liouville differences. Then A and B are equivalent.*

Proof. There is no harm in replacing A by an equivalent tuple in which $B_i - B_j \in \mathbb{Z}$ if and only if $B_i = B_j$.

For some c and σ_m, we have for all m,

$$(\pm(A_i - B_{\sigma_m(i)}))^{(m)} \leq cm$$
$$(\pm(A_i - B_{\sigma_{m+1}(i)}))^{(m+1)} \leq c(m+1)$$

and so

$$(\pm(B_{\sigma_m(i)} - B_{\sigma_{m+1}(i)}))^{(m)} \leq 2cm + c.$$

By hypothesis, the difference $B_{\sigma_m(i)} - B_{\sigma_{m+1}(i)}$ is either zero or a p-adic non-Liouville number which is not an integer; for m large, the previous inequality is inconsistent with the second option, so $B_{\sigma_m(i)} = B_{\sigma_{m+1}(i)}$. That is, for m large we have $\sigma_m = \sigma$ for some fixed σ, so

$$(\pm(A_i - B_{\sigma(i)}))^{(m)} \leq cm \qquad (m = 1, 2, \dots).$$

By Lemma 12.4.3, $A_i - B_{\sigma(i)} \in \mathbb{Z}$, so A and B are equivalent. $\square$

5. Exponents for annuli

Definition 12.5.1. Let M be a finite differential module of rank n over $K\langle \alpha/t, t/\beta \rangle$ satisfying the Robba condition, and fix a basis $e_1, \dots, e_n$ of M. An *exponent* for M is an element $A \in \mathbb{Z}_p^n$ for which there exist a sequence $\{S_m\}_{m=1}^{\infty}$ of $n \times n$ matrices over $K\langle \alpha/t, t/\beta \rangle$ satisfying the following conditions.

(a) For $j = 1, \dots, n$, under the action of ζ_{p^m} on M via Taylor series (which converge because of the Robba condition), the vector $v_{m,j} = \sum_i (S_m)_{ij} e_i$ is carried to $\zeta_{p^m}^{A_j} v_{m,j}$.

(b) For some k, we have $|S_m|_\rho \leq p^{mk}$ for all m and all $\rho \in [\alpha, \beta]$.

(c) Writing $S_m = \sum_{h \in \mathbb{Z}} S_{m,h} t^h$, we have $|S_{m,0}|_\rho \geq 1$ for all $\rho \in [\alpha, \beta]$.

Note that the property of being an exponent does not depend on the choice of the basis (although the choice of the matrices S_m does).

Proposition 12.5.2. *Let M be a finite differential module of rank n over $K\langle \alpha/t, t/\beta \rangle$ satisfying the Robba condition.*

(a) *There exists an exponent for M.*

(b) *Any two exponents for M are weakly equivalent. In particular, if M admits an exponent with non-Liouville differences, then (by Lemma 12.4.3) any other exponent for M is strongly equivalent to it.*

Proof. For (a), see [**Dwo97**, Lemma 3.1, Corollary 3.3]. For (b), see [**Dwo97**, Theorem 4.4]. $\square$

Remark 12.5.3. If M is a differential module of rank n over $K\langle t/\beta \rangle$ for the derivation $t\frac{d}{dt}$, such that the eigenvalues of the action of D on M/tM are in $\mathbb{Z}_p$, then it is easy to check (using shearing transformations) that these eigenvalues form an exponent for $M \otimes K\langle \alpha/t, t/\beta \rangle$ for any $\alpha \in (0, \beta)$.

The following is straightforward to verify.

Lemma 12.5.4. *Let M be a finite differential module of rank n over $K\langle \alpha/t, t/\beta \rangle$ satisfying the Robba condition, and let $\phi : K\langle \alpha/t, t/\beta \rangle \to K\langle \alpha^{1/q}/t, t/\beta^{1/q} \rangle$ be the substitution $t \mapsto t^q$. If A is an exponent of M, then qA is an exponent of $\phi^* M$.*

This gives an important instance where the exponent of a differential module can be controlled.

Corollary 12.5.5. *Let M be a finite differential module on an open annulus with outer radius 1, such that there exists an isomorphism $\phi_K^*\phi^* M \cong M$ on some annulus, where $\phi_K : K \to K$ is an isometry, and ϕ is the substitution $t \mapsto t^q$. Then any exponent for M consists of rational numbers.*

Proof. This holds by Lemma 12.5.4 and Corollary 12.4.4. □

6. The p-adic Fuchs theorem for annuli

Having sufficiently well understood the definition of exponents for a differential module on an open annulus, one then obtains the following theorem. We omit its proof; see notes for further discussion.

Theorem 12.6.1 (Christol-Mebkhout). *Let M be a finite differential module on an open annulus for the derivation $t\frac{d}{dt}$ satisfying the Robba condition, admitting an exponent with non-Liouville differences. Then M is isomorphic to a differential module in which D acts on some basis via a matrix N_0 with coefficients in K, whose eigenvalues represent the exponents of M (and hence are in $\mathbb{Z}_p$). Consequently, M admits a canonical decomposition*

$$M = \bigoplus_{\alpha \in \mathbb{Z}_p/\mathbb{Z}} M_\alpha$$

in which each M_α has exponent identically equal to α.

Remark 12.6.2. The exponent differences condition is difficult to verify in general because of the indirect nature of the definition of exponents. However, if M is a finite differential admits a Frobenius structure, then Corollary 12.5.5 implies that the exponents are rational. This leads to a quasiunipotence result (Theorem 18.4.1) which can be used to establish the p-adic local monodromy theorem (Theorem 18.1.8).

One other consequence of Theorem 12.6.1 that can be stated without reference to exponents is the following.

Corollary 12.6.3. *Let M be a finite differential module on an open annulus for the derivation $t\frac{d}{dt}$ satisfying the Robba condition. Suppose that the restriction M' of M to some smaller open annulus is trivial/unipotent. Then the same is true for M.*

Proof. If M" is unipotent, then its exponent is equivalent to 0. However, any exponent A of M restricts to an exponent of M', so is weakly equivalent to 0 by Proposition 12.5.2. By Lemma 12.4.3, A is equivalent to 0, so Theorem 12.6.1 implies that M is unipotent. Moreover, if M is unipotent and M' is trivial, then M is forced to be trivial also. □

7. Transfer to a regular singularity

As an application of Theorem 12.6.1, we obtain a transfer theorem in the presence of a regular singularity, in the spirit of Theorem 8.5.1 and Theorem 8.5.4 but with a somewhat weaker estimate.

Theorem 12.7.1. *Let M be a finite differential module of rank n over $K[\![t]\!]_0$ for the derivation $t\frac{d}{dt}$, with a regular singularity at $t = 0$ whose exponents are in $\mathbb{Z}_p$ and have non-Liouville differences. Then the fundamental solution matrix of M converges in the open disc of radius $R(M \otimes F_1)^n$. In particular, if M has generic radius of convergence 1, then the fundamental solution matrix of M converges in the open unit disc.*

Proof. By Theorem 12.2.2, the fundamental solution matrix of M converges in a disc of positive radius. From this and Proposition 12.3.3, it follows that $R(M \otimes F_\rho) = \rho$ for $\rho \in (0, 1)$ sufficiently small.

Let λ be the supremum of $\rho \in (0, 1)$ for which $R(M \otimes F_\rho) = \rho$. Note that the function $f_1(r) = -\log R(M \otimes F_{e^{-r}})$ is convex by Theorem 10.3.2, is equal to r for r sufficiently large by the previous paragraph, and is also equal to r for $r = -\log \lambda$ by continuity. Consequently, $f_1(r) = r$ for all $r \geq -\log \lambda$.

Choose $\alpha, \beta \in (0, \lambda)$ with $\alpha < \beta$, such that the fundamental solution matrix of M converges in the open disc of radius β. By Theorem 12.6.1, it also converges in the open annulus of inner radius α and outer radius 1. By patching, we deduce that the fundamental solution matrix converges in the open disc of radius λ.

To conclude, it suffices to give a lower bound for λ. By Theorem 10.3.2, for $r \in [0, -\log \lambda]$, the function f_1 is continuous and piecewise affine, with slopes belonging to $\frac{1}{1}\mathbb{Z} \cup \cdots \cup \frac{1}{n}\mathbb{Z}$. Since the slope for $r > -\log \lambda$ is equal to 1, the slopes for $r \leq -\log \lambda$ cannot exceed 1; moreover, there cannot be a slope equal to 1 in this range, as otherwise it would occur as the left slope at $r = -\log \lambda$, so there would exist $\rho > \lambda$ for which $R(M \otimes F_\rho) = \rho$, contrary to how λ was defined. Consequently, f_1 has all slopes less than or equal to $(n-1)/n$ for $r \in [0, -\log \lambda]$, yielding

$$-\log \lambda = f_1(-\log \lambda) \leq f_1(0) + \frac{n-1}{n}(-\log \lambda).$$

From this we deduce $\lambda \geq R(M \otimes F_1)^n$, as desired. □

Remark 12.7.2. We do not have in mind an example where one does not get convergence on the open disc of radius $R(M \otimes F_1)$.

Notes

The definition of a p-adic Liouville number was introduced by Clark [**Cla66**]; our presentation follows [**DGS94**, §VI.1].

The cited theorem of Clark [**Cla66**, Theorem 3] is actually somewhat stronger than Theorem 12.2.3, as it allows differential operators of possibly infinite order.

Proposition 12.3.2 is originally due to Christol; compare [**DGS94**, Proposition IV.7.7].

The theory of exponents for differential modules on a p-adic annulus satisfying the Robba condition was originally developed by Christol and Mebkhout [**CM97**, §4–5]; in particular, Theorem 12.6.1 appears therein as [**CM97**, Théorème 6.2–4]. A somewhat more streamlined development was later given by Dwork [**Dwo97**], in which Theorem 12.6.1 appears as [**Dwo97**, Theorem 7.1]. (Dwork coyly notes that he did not verify the equivalence between the two constructions; we do not recommend losing any sleep over this.) A useful expository article on the topic is that of Loeser [**Loe96**].

A somewhat more elementary treatment of Theorem 12.7.1 than the one given here is given in [**DGS94**, §6]; it does not rely on the p-adic Fuchs theorem for annuli. However, it gives a weaker result: it only establishes convergence of the fundamental solution matrix in the open disc of radius $R(M \otimes F_1)^{n^2}$. A similar treatment is [**Chr83**, Théorème 6.4.7].

Exercises

(1) Prove that rational numbers are p-adic non-Liouville numbers.
(2) Give another proof of Lemma 12.1.6 (as in [**DGS94**, Lemma VI.1.2]) by first verifying that both sides of the desired equation have the same coefficients of x^0 and x^1, and are killed by the second-order differential operator $\frac{d}{dx}(\frac{d}{dx} - \lambda - x)$.
(3) Show that Theorem 12.2.2 can be deduced from Theorem 12.2.3. (Hint: show that if $H^0(M) \neq 0$, then 0 must occur as an eigenvalue of N_0.)
(4) Prove that there exists $a \in \mathbb{Z}_p$ satisfying (12.2.4.1).
(5) Prove Proposition 12.3.3.

Part 4

Difference algebra and Frobenius structures

CHAPTER 13

Formalism of difference algebra

In this chapter, we set up a bit of formalism for difference algebra, parallel to what we did with differential algebra earlier. This formalism will be used in subsequent chapters to describe Frobenius structures on p-adic differential equations.

1. Difference algebra

DEFINITION 13.1.1. A *difference ring/field* is a ring/field R equipped with an endomorphism ϕ. A *difference module* over R is an R-module M equipped with a map $\Phi : R \to R$ which is additive and ϕ-semilinear; the latter means that

$$\Phi(rm) = \phi(r)\Phi(m) \qquad (r \in R, m \in M).$$

A difference submodule of R itself is also called a *difference ideal*.

DEFINITION 13.1.2. If M is a finite difference module over R freely generated by $e_1, \dots, e_n$, then we can recover the action of Φ from the $n \times n$ matrix A defined by

$$\Phi(e_j) = \sum_i A_{ij} e_i.$$

Namely, if we use the basis to identify M with the space of column vectors of length n over R, then

$$\Phi(v) = A\phi(v).$$

Moreover, if we change to a new basis $e'_1, \dots, e'_n$, and let U be the change-of-basis matrix (defined by $e'_j = \sum_i U_{ij} e_i$), then Φ acts on the new basis via the matrix

$$A' = U^{-1} A \phi(U).$$

We say M is *dualizable* if A is invertible. If M is dualizable, we define the dual $M^\vee$ as the module-theoretic dual $\mathrm{Hom}_R(M, R)$ with Φ-action given on the dual basis by A^{-T} (the inverse transpose). Note that the property of dualizability, and the definition of the dual, do not depend on the choice of the basis; hence they both extend to the case where M is only locally free as an R-module.

DEFINITION 13.1.3. We say that the difference ring R is *inversive* if ϕ is an automorphism. In this case, we can define the *opposite difference ring* R^{opp} to be R again, but now equipped with the endomorphism ϕ^{-1}. If R is inversive and M is locally free, we define the *opposite module* M^{opp} of M as the module-theoretic dual $\mathrm{Hom}_R(M, R)$ equipped with the pullback action (i.e., on the dual basis, use the matrix A^T for the action).

DEFINITION 13.1.4. For M a difference module, write

$$H^0(M) = \ker(\mathrm{id} - \Phi), \qquad H^1(M) = \mathrm{coker}(\mathrm{id} - \Phi).$$

If M_1, M_2 are difference modules with M_1 dualizable, then $H^0(M_1^\vee \otimes M_2)$ computes morphisms from M_1 to M_2, and $H^1(M_1^\vee \otimes M_2)$ computes extensions $0 \to M_2 \to M \to M_1 \to 0$. That is,

$$H^0(M_1^\vee \otimes M_2) = \mathrm{Hom}(M_1, M_2), \qquad H^1(M_1^\vee \otimes M_2) = \mathrm{Ext}(M_1, M_2).$$

2. Twisted polynomials

As in differential algebra, there is a relevant notion of twisted polynomials.

DEFINITION 13.2.1. For R a difference ring, we define the twisted polynomial ring $R\{T\}$ as the set of finite formal sums $\sum_{i=0}^{\infty} r_i T^i$, but with the multiplication this time obeying the rule $Tr = \phi(r)T$. For any $P \in R\{T\}$, the quotient $R\{T\}/R\{T\}P$ is a difference module; if M is a difference module, we say $m \in M$ is a *cyclic vector* if there is an isomorphism $M \cong R\{T\}/R\{T\}P$ carrying m to 1.

DEFINITION 13.2.2. If R is inversive, we again have a formal adjoint construction: given $P \in R\{T\}$, its formal adjoint is obtained by pushing the coefficients to the right side of T. This may then be viewed as an element of the opposite ring of $R\{T\}$, which we may identify with $R^{\mathrm{opp}}\{T\}$.

It is not completely straightforward to analogize the cyclic vector theorem to difference modules; see the exercises for one attempt to do so. Instead, we will use only the following trivial observation.

LEMMA 13.2.3. *Any irreducible finite difference module over a difference field contains a cyclic vector.*

PROOF. If F is a difference field, M is a finite difference module over F, and $m \in M$ is nonzero, then $m, \Phi(m), \dots$ generate a nonzero difference submodule of M. If M is irreducible, this submodule must be all of M. □

DEFINITION 13.2.4. If ϕ is isometric for a norm $|\cdot|$ on F, then we have the usual definition of Newton polygons and slopes for twisted polynomials. If R is inversive, then a twisted polynomial and its adjoint have the same Newton polygon.

Applying the master factorization theorem (Theorem 2.2.2) yields the following.

THEOREM 13.2.5. *Let F be a difference field complete for a norm $|\cdot|$ under which ϕ is isometric. Then any monic twisted polynomial $P \in F\{T\}$ admits a unique factorization*

$$P = P_{r_1} \cdots P_{r_m}$$

for some $r_1 < \cdots < r_m$, where each P_{r_i} is monic with all slopes equal to r_i. (If F is inversive, the same holds with the factors in the opposite order.)

3. Difference-closed fields

DEFINITION 13.3.1. We will say that a difference field F is *weakly difference-closed* if every dualizable finite difference module over F is trivial. We say F is *strongly difference-closed* if F is inversive and weakly difference-closed.

REMARK 13.3.2. Note that the property that F is weakly difference-closed includes the fact that short exact sequences of dualizable finite difference modules over F always split. By contrast, if for instance ϕ is the identity map, then this is never true even if F is algebraically closed, because linear transformations need not be semisimple.

LEMMA 13.3.3. *The difference field F is weakly difference-closed if and only if the following conditions hold.*

(a) *Every nonconstant monic twisted polynomial $P \in F\{T\}$ factors as a product of linear factors.*
(b) *For every $c \in F^\times$, there exists $x \in F^\times$ with $\phi(x) = cx$.*
(c) *For every $c \in F^\times$, there exists $x \in F^\times$ with $\phi(x) - x = c$.*

PROOF. We first suppose that F is weakly difference-closed. To prove (a), it suffices to check that if $P \in F\{T\}$ is nonconstant monic with nonzero constant term, then P factors as P_1P_2 with P_2 linear. The nonzero constant term implies that $M = F\{T\}/F\{T\}P$ is a dualizable finite difference module over F, so must be trivial by the hypothesis that F be weakly difference-closed. In particular, there exists a short exact sequence $0 \to M_1 \to M \to M_2 \to 0$ with M_2 trivial; this corresponds to a factorization $P = P_1P_2$ with P_2 linear.

To prove (b), note that $F\{T\}/F\{T\}(T - c^{-1})$ must be trivial, which means there exists $x \in F^\times$ such that $Tx - x = y(T - c^{-1})$ for some $y \in F$. Then $y = \phi(x)$ and $yc^{-1} = x$, proving the claim.

To prove (c), form the ϕ-module V corresponding to the matrix $\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}$. By construction, we have a short exact sequence $0 \to V_1 \to V \to V_2 \to 0$ with V_1, V_2 trivial; since V must also be trivial, this extension must split. That means that we can find $x \in F$ with $\phi(x) - x = c$, proving the claim.

Conversely, suppose that (a), (b), (c) hold. Every nonzero dualizable finite difference module over F admits an irreducible quotient. This quotient admits a cyclic vector by Lemma 13.2.3, and so admits a quotient of dimension 1 by (a). That quotient in turn is trivial by (b). By induction, we deduce that every dualizable finite difference module over F admits a filtration whose successive quotients are trivial of dimension 1. This filtration splits by (c). □

PROPOSITION 13.3.4. *Let F be a separably (resp. algebraically) closed field of characteristic $p > 0$ equipped with a power of the absolute Frobenius. Then F is weakly (resp. strongly) difference-closed.*

PROOF. For $P = \sum_{i=0}^m P_i T^m \in F\{T\}$ with $m > 0$, $P_m = 1$, and $P_0 \neq 0$, the polynomial $Q(x) = \sum_{i=0}^m P_i x^{q^i}$ has degree $q^m \geq 2$, and $x = 0$ occurs as a root only with multiplicity 1. Moreover, the formal derivative of P is a constant polynomial, so has no common roots with P; hence P is a separable polynomial. Since F is separably closed, there must exist a nonzero root x of Q; this implies criteria (a) and (b) of Lemma 13.3.3. To deduce (c), note that for $c \in F^\times$, the polynomial $x^q - x - c$ is again separable, so has a root in F. □

4. Difference algebra over a complete field

HYPOTHESIS 13.4.1. For the rest of this chapter, let F be a difference field complete for a norm $|\cdot|$ with respect to which ϕ is isometric. We do not assume that F is inversive; if not, then we can embed into F into an inversive difference field by forming the completion F' of the direct limit of the system

$$F \xrightarrow{\phi} F \xrightarrow{\phi} \cdots.$$

We sometimes call F' the *ϕ-perfection* of F.

As in the differential case, we would like to classify finite difference modules over F by the spectral norm of Φ. The following basic properties will help, as long as we are mindful of the discrepancies between the differential and difference cases.

LEMMA 13.4.2. *Let V, V_1, V_2 be nonzero finite difference modules over F.*

(a) *For $0 \to V_1 \to V \to V_2 \to 0$ a short exact sequence,*

$$|\Phi|_{\mathrm{sp},V} = \max\{|\Phi|_{\mathrm{sp},V_1}, |\Phi|_{\mathrm{sp},V_2}\}.$$

(b) *We have*

$$|\Phi|_{\mathrm{sp},V_1 \otimes V_2} = |\Phi|_{\mathrm{sp},V_1} |\Phi|_{\mathrm{sp},V_2}.$$

(c) *We have*

$$|\Phi|_{\mathrm{sp},V} = |\Phi|_{\mathrm{sp},V \otimes F'}.$$

PROOF. Exercise. □

The relationship between V and the dual $V^\vee$ is more complicated.

LEMMA 13.4.3. *If $V \cong F\{T\}/F\{T\}P$ and P has only one slope r in its Newton polygon, then*

$$|\Phi|_{\mathrm{sp},V} = e^{-r}.$$

If F is inversive, then also

$$|\Phi^{-1}|_{\mathrm{sp},V} = e^{-r}.$$

PROOF. By replacing F with F', we may reduce to the case where F is inversive. Put $n = \deg(P)$, and define a norm on V by

$$|a_0 + \cdots + a_{n-1}T^{n-1}| = \max_i\{|a_i|e^{-ri}\};$$

then

$$|\Phi|_V = e^{-r}, \qquad |\Phi^{-1}|_V = e^r.$$

We deduce that

$$|\Phi|_{\mathrm{sp},V} \leq e^{-r}, \qquad |\Phi^{-1}| \leq e^r;$$

since

$$1 = |\Phi|_{\mathrm{sp},V}|\Phi^{-1}|_{\mathrm{sp},V} \leq e^{-r}e^r,$$

we obtain the desired equalities. □

COROLLARY 13.4.4. *For any nonzero finite difference module V over F, either $|\Phi|_{\mathrm{sp},V} = 0$, or there exists an integer $m \in \{1, \dots, \dim_F V\}$ such that $|\Phi|^m_{\mathrm{sp},V} \in |F^\times|$.*

DEFINITION 13.4.5. Let V be a nonzero finite difference module over F. We say that V is *pure of norm s* if all of the Jordan-Hölder constituents of V have spectral norm s. Note that V is pure of norm 0 if and only if $\Phi^{\dim_F V} = 0$. If V is pure of norm 1, we also say that V is *étale* or *unit-root*.

PROPOSITION 13.4.6. *Let V be a nonzero finite difference module over F. Then V is pure of norm $s > 0$ if and only if*

$$|\Phi|_{\mathrm{sp},V\otimes F'} = s, \qquad |\Phi^{-1}|_{\mathrm{sp},V\otimes F'} = s^{-1}. \tag{13.4.6.1}$$

PROOF. If V is pure of norm s, then (13.4.6.1) holds by Lemma 13.4.3. Conversely, if (13.4.6.1) holds and W is a subquotient of V, then

$$|\Phi|_{\mathrm{sp},W\otimes F'} \leq |\Phi|_{\mathrm{sp},V\otimes F'}, \qquad |\Phi^{-1}|_{\mathrm{sp},W\otimes F'} \leq |\Phi^{-1}|_{\mathrm{sp},V\otimes F'}.$$

We thus have

$$1 \leq |\Phi|_{\mathrm{sp},W\otimes F'}|\Phi^{-1}|_{\mathrm{sp},W\otimes F'} \leq ss^{-1} = 1,$$

which forces $|\Phi|_{\mathrm{sp},W} = |\Phi|_{\mathrm{sp},W\otimes F'} = s$. □

COROLLARY 13.4.7. *Let V_1, V_2 be nonzero finite difference modules over F which are pure of respective norms s_1, s_2. Then $V_1 \otimes_F V_2$ is pure of norm $s_1 s_2$.*

PROOF. If $s_1 s_2 = 0$, then it is easy to check that $V_1 \otimes V_2$ is pure of norm 0. Otherwise, one direction of Proposition 13.4.6 yields

$$|\Phi|_{\mathrm{sp},V_1\otimes V_2\otimes F'} = |\Phi|_{\mathrm{sp},V_1\otimes F'}|\Phi|_{\mathrm{sp},V_2\otimes F'} = s_1 s_2,$$
$$|\Phi^{-1}|_{\mathrm{sp},V_1\otimes V_2\otimes F'} = |\Phi^{-1}|_{\mathrm{sp},V_1\otimes F'}|\Phi^{-1}|_{\mathrm{sp},V_2\otimes F'} = s_1^{-1} s_2^{-1},$$

so the other direction of Proposition 13.4.6 implies that $V_1 \otimes V_2$ is pure of norm $s_1 s_2$. □

COROLLARY 13.4.8. *Let V be a nonzero finite difference module over F. Then for any positive integer d, V is pure of norm s if and only if V becomes pure of norm s^d when viewed as a difference module over (F, ϕ^d).*

PROPOSITION 13.4.9. *Let V be a nonzero finite difference module over F. Suppose that either:*

(a) $|\Phi|_{\mathrm{sp},V} < 1$, *or*
(b) *F is inversive and* $|\Phi^{-1}|_{\mathrm{sp},V} < 1$.

Then $H^1(V) = 0$.

PROOF. In case (a), given $v \in V$, the series

$$w = \sum_{i=0}^{\infty} \Phi^i(v)$$

converges to a solution of $w - \Phi(w) = v$. In case (b), the series

$$w = -\sum_{i=0}^{\infty} \Phi^{-i-1}(v)$$

does likewise. □

COROLLARY 13.4.10. *If V_1, V_2 are nonzero finite differential modules over F which are pure of respective norms s_1, s_2, and either:*

(a) $s_1 < s_2$; *or*
(b) *F is inversive and* $s_1 > s_2$;

then any exact sequence $0 \to V_1 \to V \to V_2 \to 0$ splits.

Proof. If $s_2 > 0$, then by Corollary 13.4.7, $V_2^\vee \otimes V_1$ is pure of norm s_1/s_2, so Proposition 13.4.9 gives the desired splitting. Otherwise, we must be in case (b), so we can pass to the opposite ring to make the same conclusion. $\square$

If F is inversive, we again get a decomposition theorem.

Theorem 13.4.11. *Suppose that F is inversive. Let V be a finite difference module over F. Then there exists a unique direct sum decomposition*

$$V = \bigoplus_{s \geq 0} V_s$$

of difference modules, in which each V_s is pure of norm s. (Note that V is dualizable if and only if $V_0 = 0$.)

Proof. This follows at once from Corollary 13.4.10. $\square$

Remark 13.4.12. Note that in case ϕ is the identity map on F, Theorem 13.4.11 simply reproduces the decomposition of V in which the generalized eigenspaces for all eigenvalues of a given modulus are grouped together.

If F is not inversive, we only get a filtration instead of a decomposition.

Theorem 13.4.13. *Let V be a finite difference module over F. Then there exists a unique filtration*

$$0 = V_0 \subset V_1 \subset \cdots \subset V_l = V$$

of difference modules, such that each successive quotient V_i/V_{i-1} is pure of some norm s_i, and $s_1 > \cdots > s_l$. (Note that V is dualizable if and only if $V = 0$ or $s_l > 0$.)

Proof. Start with any filtration of V with irreducible successive quotients, and let s_1 be the largest norm which appears. By Corollary 13.4.10, we can change the filtration to move the first appearance of s_1 one step earlier; consequently, we can put all appearances of s_1 before all other slopes. Group these together to form V_1, then repeat to construct the desired factorization. Uniqueness follows by tensoring with F' and invoking the uniqueness in Theorem 13.4.11. $\square$

The following alternate characterization of pureness may be useful in some situations.

Proposition 13.4.14. *Let V be a finite difference module over F, and choose $\lambda \in F^\times$. Then V is pure of norm $|\lambda|$ if and only if there exists a basis of V on which Φ acts via λ times an element of $\mathrm{GL}_n(\mathfrak{o}_F)$.*

Proof. If such a basis exists, then Proposition 13.4.6 implies that V is pure of norm $|\lambda|$. Conversely, if V is irreducible of spectral norm $|\lambda|$, then Lemma 13.4.3 provides a basis of the desired form. Otherwise, we proceed by induction on $\dim_F V$. Suppose we are given a short exact sequence $0 \to V_1 \to V \to V_2 \to 0$ in which V_1, V_2 admit bases of the desired form. Let $e_1, \dots, e_m \in V$ form such a basis for V_1, and let $e_{m+1}, \dots, e_n \in V$ lift such a basis for V_2. Then for $\mu \in F$ of sufficiently small norm,

$$e_1, \dots, e_m, \mu e_{m+1}, \dots, \mu e_n$$

will form a basis of V of the desired form. $\square$

Remark 13.4.15. Note that whenever V is pure of positive norm, we can apply Proposition 13.4.14 after replacing Φ by some power of it, thanks to Corollary 13.4.4.

5. Hodge and Newton polygons

Definition 13.5.1. Let V be a finite difference module over F equipped with a norm defined as the supremum norm for some basis $e_1, \dots, e_n$. Let A be the basis via which Φ acts on this basis; define the *Hodge polygon* of V as the Hodge polygon of the matrix A. Given the choice of the norm on V, this definition is independent of the choice of the basis: we can only change basis by a matrix $U \in \mathrm{GL}_n(\mathfrak{o}_F)$, which replaces A by $U^{-1}A\phi(U)$, and ϕ being an isometry ensures that $\phi(U) \in \mathrm{GL}_n(\mathfrak{o}_F)$ also. As in the linear case, we list the Hodge slopes $s_{H,i}, \dots, s_{H,n}$ in increasing order.

Definition 13.5.2. Let V be a finite difference module over F. Define the *Newton polygon* of V to have slopes $s_{N,1}, \dots, s_{N,n}$ such that r appears with multiplicity equal to the dimension of the quotient in Theorem 13.4.13 of norm e^{-r}.

Lemma 13.5.3. *Let V be a finite difference module over F. We have*

$$s_{H,1} + \cdots + s_{H,i} = -\log |\Phi|_{\wedge^i V} \qquad (i = 1, \dots, n)$$
$$s_{N,1} + \cdots + s_{N,i} = -\log |\Phi|_{\mathrm{sp}, \wedge^i V} \qquad (i = 1, \dots, n).$$

Proof. The first assertion follows from the corresponding fact in the linear case. The second assertion reduces to the fact that if V is irreducible of dimension n and spectral norm s, then $\wedge^i V$ has spectral norm s^i for $i = 1, \dots, n$; this follows by imitating the proof of Lemma 13.4.3. □

Corollary 13.5.4 (Newton above Hodge). *We have*

$$s_{N,1} + \cdots + s_{N,i} \geq s_{H,1} + \cdots + s_{H,i} \qquad (i = 1, \dots, n)$$

with equality for $i = n$.

Theorem 13.5.5. *Let V be a finite difference module over F equipped with a basis. If for some $i \in \{1, \dots, n-1\}$ we have*

$$s_{N,i} > s_{N,i+1}, \qquad s_{N,1} + \cdots + s_{N,i} = s_{H,1} + \cdots + s_{H,i},$$

then we can change basis by a matrix in $\mathrm{GL}_n(\mathfrak{o}_F)$ so that the matrix of action of Φ becomes block upper triangular, with the top left block accounting for the first i Hodge and Newton slopes of M. Moreover, if F is inversive and $s_{H,i} > s_{H,i+1}$, we can ensure that the matrix of action of Φ is block diagonal.

Proof. As in Theorem 3.3.11. □

Remark 13.5.6. Beware that the Newton polygon, unlike the Hodge polygon, cannot be directly read off from the matrix via which Φ acts on some basis; see exercises for a counterexample. On the other hand, this works if the matrix of Φ is a companion matrix; this is a restatement of the following fact.

Proposition 13.5.7. *If $V \cong F\{T\}/F\{T\}P$, then the Newton polygon of V coincides with that of P.*

Proof. This reduces to Lemma 13.4.3. □

PROPOSITION 13.5.8. *Suppose that F is discrete. Let K be a complete difference subfield of F with the same value group, and let R be a complete difference subring of F containing K. Let M be a finite free difference module over R, and suppose $c \in \mathbb{R}$ is such that the least Newton slope of M is at least c. Then there exists a basis of M with respect to which for each positive integer m, the least Hodge slope of Φ^m is the least element of $v(F^\times)$ greater than or equal to cm.*

PROOF. Construct a suitable basis of $M \otimes F$ using Lemma 3.3.13, then apply Lemma 10.5.1. $\square$

6. The Dieudonné-Manin classification theorem

DEFINITION 13.6.1. For $\lambda \in F$ and d a positive integer, let $V_{\lambda,d}$ be the difference module over F with basis $e_1, \dots, e_d$ such that

$$\Phi(e_1) = e_2, \quad \dots, \quad \Phi(e_{d-1}) = e_d, \quad \Phi(e_d) = \lambda e_1.$$

LEMMA 13.6.2. *Suppose $\lambda \in F^\times$ and the positive integer d are such that there is no $i \in \{1, \dots, d-1\}$ such that $|\lambda|^{i/d} \in |F^\times|$. Then $V_{\lambda,d}$ is irreducible.*

PROOF. Note that

$$\Phi^d e_i = \phi^{i-1}(\lambda) e_i \qquad (i = 1, \dots, n).$$

Hence by Proposition 13.4.14, $V_{\lambda,d}$ is pure of norm $\lambda^{1/d}$, as then is any submodule. But if the submodule were proper and nonzero, we would have a violation of Corollary 13.4.4. $\square$

THEOREM 13.6.3. *Let F be a complete discretely valued field equipped with an isometric endomorphism ϕ, such that κ_F is strongly difference-closed. Then every dualizable finite difference module over F can be split (non-uniquely) as a direct sum of submodules, each of the form $V_{\lambda,d}$ for some λ, d. Moreover, for π any fixed uniformizer of F, we can force each λ to be a power of π.*

PROOF. We first check that if V is pure of norm 1, then V is trivial. We must show that for any $A \in \mathrm{GL}_n(\mathfrak{o}_F)$, there exists a convergent sequence $U_1, U_2, \dots \in \mathrm{GL}_n(\mathfrak{o}_F)$ such that

$$U_m^{-1} A \phi(U_m) \equiv I_n \pmod{\pi^m}.$$

Specifically, we will insist that $U_{m+1} \equiv U_m \pmod{\pi^m}$. Finding U_1 amounts to trivializing a dualizable difference module of dimension n over κ_F. For $m > 1$, given U_m, we must have $U_{m+1} = U_m(I_n + \pi^m X_m)$ for some m, and

$$(I_n + \pi^m X_m)^{-1}(U_m^{-1} A \phi(U_m))(I_n + \pi^m X_m) \equiv I_n \pmod{\pi^{m+1}}.$$

Since already $U_m^{-1} A \phi(U_m) \equiv I_n \pmod{\pi^m}$, this amounts to solving

$$-X_m + \pi^{-m}(U_m^{-1} A \phi(U_m) - I_n) + \phi(X_m) \equiv 0 \pmod{\pi},$$

which we solve by applying criterion (c) from Lemma 13.3.3.

By similar (but easier) arguments, we also show that:

- ϕ is surjective on $\mathfrak{o}_F$, so F is inversive;
- if V is trivial, then $H^1(V) = 0$.

In particular, we may apply Theorem 13.4.11 to reduce the desired result to the case where V is pure of norm $s > 0$.

Let d be the smallest positive integer such that $s^d = |\pi^m|$ for some integer m. Then the first paragraph implies that $\pi^{-m}\Phi^d$ fixes some nonzero element of V; this gives us a nonzero map from $V_{\pi^m,d}$ to V. By Lemma 13.6.2, this map must be injective. Repeating this argument, we write V as a successive extension of copies of $V_{\pi^m,d}$. However, $V_{\pi^m,d}^\vee \otimes V_{\pi^m,d}$ is pure of norm 1, so has trivial H^1 as above. Thus V splits as a direct sum of copies of $V_{\pi^m,d}$, as desired. □

By Proposition 13.3.4, Theorem 13.6.3 has the following immediate corollary.

COROLLARY 13.6.4. *Let F be a complete discretely valued field, normalized so that the additive value group is $\mathbb{Z}$, such that κ_F is algebraically closed of characteristic $p > 0$. Let $\phi : F \to F$ be an isometric automorphism lifting a power of the absolute Frobenius on κ_F. Then every dualizable finite difference module over F can be split (non-uniquely) as a direct sum of difference submodules, each of the form $V_{\lambda,d}$ for some $\lambda \in F^\times$ and some positive integer d coprime to the valuation of λ. Moreover, for π any fixed uniformizer of F, we can force each λ to be a power of π.*

REMARK 13.6.5. The case of Corollary 13.6.4 in which k is an algebraically closed field of characteristic p, $W(k)$ is the ring of p-typical Witt vectors (i.e., the unique complete discrete valuation ring with residue field k and maximal ideal (p)), $F = \mathrm{Frac}(W(k))$, and ϕ is the Witt vector Frobenius is the *Dieudonné-Manin theorem*, i.e., the classification theorem of rational Dieudonné modules over an algebraically closed field.

Notes

The parallels between difference and differential algebra are quite close, enough so that a survey of references for difference algebra strongly resembles its differential counterpart. An older, rather dry reference is [**Coh65**]; a somewhat more lively modern reference, which develops difference Galois theory under somewhat restrictive conditions, is [**SvdP97**]. We again mention [**And01**] as a useful unifying framework for difference and differential algebra.

Proposition 13.3.4 can be found in SGA7 [**DK73**, Exposé XXII, Corollaire 1.1.10], wherein Katz attributes it to Lang. Indeed, it is a special case of the nonabelian Artin-Schreier theory associated to an algebraic group over a field of positive characteristic (in our case GL_n), via the *Lang torsor*; see [**Lan56**].

In the special case of the difference field $\mathrm{Frac}(W(k))$, with k perfect of characteristic $p > 0$, a number of the results in this chapter appear (in marginally less generality) in [**Kat79**]. For instance, Corollary 13.5.4 reproduces Mazur's [**Kat79**, Theorem 1.4.1], while Proposition 13.5.8 generalizes [**Kat79**, Theorem 2.6.1] and answers the question posed by Katz in the following remark.

For the original classification of rational Dieudonné modules over an algebraically closed field, see Manin's original paper [**Man63**] or the book of Demazure [**Dem72**].

As in Chapter 3, one can interpret what we have done here as the special case for GL_n of a construction for any reductive algebraic group. This point of view was originally introduced by Kottwitz [**Kot85, Kot97**], but a full development of the analogy is the subject of recent and ongoing (as of this writing) work of Kottwitz [**Kot03**] and Csima [**Csi07**].

Exercises

(1) Let F be a difference field of characteristic zero containing an element x such that $\phi(x) = \lambda x$ for some λ fixed by ϕ. Prove that every finite difference module for M admits a cyclic vector. (Hint: under these hypotheses, one can readily imitate the proof of Theorem 4.4.2.)

(2) Let F be the completion of $\mathbb{Q}_p(t)$ for the 1-Gauss norm, viewed as a difference field for ϕ equal to the substitution $t \mapsto t^p$. Let V be the difference module corresponding to the matrix
$$A = \begin{pmatrix} 1 & t \\ 0 & p \end{pmatrix}.$$
Prove that there is a nonsplit short exact sequence $0 \to V_1 \to V \to V_2 \to 0$ with V_1, V_2 pure of slopes s_1, s_2 with $s_1 < s_2$.

(3) Here is a beautiful example from [**Kat79**, §1.3] (attributed to B. Gross). Let p be a prime congruent to 3 modulo 4, put $F = \mathbb{Q}_p(i)$ with $i^2 = -1$, and let ϕ be the automorphism $i \mapsto -i$ of F over $\mathbb{Q}_p$. Define a difference module M of rank 2 over F using the matrix
$$A = \begin{pmatrix} 1-p & (p+1)i \\ (p+1)i & p-1 \end{pmatrix}.$$
Compute the Newton polygons of A and M and verify that they do not coincide. (Hint: find another basis of M on which Φ acts diagonally.)

(4) Prove that every difference field can be embedded into a strongly difference-closed field. (This requires your favorite equivalent of the axiom of choice, e.g., Zorn's lemma.) Variant: prove that every complete nonarchimedean difference field can be embedded into a strongly difference-closed complete nonarchimedean difference field.

CHAPTER 14

Frobenius modules

In this chapter, we restrict the formalism of difference algebra to the special case of Frobenius lifts.

Hypothesis 14.0.6. Throughout the remaining chapters in this part, we assume that our complete nonarchimedean field K is *discretely valued*. This is necessary to avoid a number of technical complications, some of which we will point out as we go along.

1. A multitude of rings

One can talk about Frobenius structures on a variety of rings; here are some of the ones we will use.

Definition 14.1.1. We have already defined

$$K\langle \alpha/t, t/\beta \rangle = \left\{ \sum_{i \in \mathbb{Z}} c_i t^i : c_i \in K, \lim_{i \to -\infty} |c_i| \alpha^i = 0, \lim_{i \to +\infty} |c_i| \beta^i = 0 \right\}.$$

We will also need

$$K[\![t]\!]_0 = \left\{ \sum_{i=0}^{\infty} c_i t^i : c_i \in K, \sup_i \{|c_i|\} < \infty \right\}$$

$$K\{\{t\}\} = \left\{ \sum_{i=0}^{\infty} c_i t^i : c_i \in K, \lim_{i \to \infty} |c_i| \rho^i = 0 \quad (\rho \in (0,1)) \right\}.$$

We will allow the following hybrids:

$$K\langle \alpha/t, t]\!]_0 = \left\{ \sum_{i \in \mathbb{Z}} c_i t^i : c_i \in K, \lim_{i \to -\infty} |c_i| \alpha^i = 0, \sup_i \{|c_i|\} < \infty \right\}$$

$$K\langle \alpha/t, t\}\} = \left\{ \sum_{i \in \mathbb{Z}} c_i t^i : c_i \in K, \lim_{i \to -\infty} |c_i| \alpha^i = 0, \lim_{i \to +\infty} |c_i| \rho^i = 0 \quad (\rho \in (0,1)) \right\}.$$

Definition 14.1.2. For later use, we give special notations to certain rings appearing in this framework. We already have defined $\mathcal{E}$ to be the completion of $\mathfrak{o}_K((t)) \otimes_{\mathfrak{o}_K} K$ for the 1-Gauss norm; that is, $\mathcal{E}$ consists of formal sums $\sum c_i t^i$ which have bounded coefficients and satisfy $|c_i| \to 0$ as $i \to -\infty$. Since we are assuming that K is discretely valued, this is a complete nonarchimedean field with residue field $\kappa_K((t))$. We next set

$$\mathcal{E}^\dagger = \bigcup_{\alpha \in (0,1)} K\langle \alpha/t, t]\!]_0;$$

that is, $\mathcal{E}^\dagger$ consists of formal sums $\sum c_i t^i$ which have bounded coefficients and converge in some range $\alpha \leq |t| < 1$. We also put

$$\mathcal{R} = \bigcup_{\alpha \in (0,1)} K\langle \alpha/t, t\}\}$$

that is, $\mathcal{R}$ consists of formal sums $\sum c_i t^i$ which converge in some range $\alpha \leq |t| < 1$, but need not be bounded. The ring $\mathcal{R}$ is commonly known as the *Robba ring* with coefficients in K.

REMARK 14.1.3. Beware that since $\mathcal{R}$ consists of series with unbounded coefficients, the 1-Gauss norm $|\cdot|_1$ is not defined on all of $\mathcal{R}$. We will conventionally write $|x|_1 = \infty$ if $x \in \mathcal{R}$ has unbounded coefficients.

2. Frobenius lifts

DEFINITION 14.2.1. Let q be a power of p. Let R be one of the following rings:

- $K\langle t \rangle$, $K\llbracket t \rrbracket_0$, or $K\{\{t\}\}$;
- the union of $K\langle \alpha/t, t\rangle$, $K\langle \alpha/t, t\rrbracket_0$, or $K\langle \alpha/t, t\}\}$ over all $\alpha \in (0,1)$;
- F_1, the completion of $K(t)$ for the 1-Gauss norm;
- $\mathcal{E}$, the completion of $K\llbracket t \rrbracket_0[t^{-1}]$ for the 1-Gauss norm.

By a *q-power Frobenius lift* on R, we will mean a map $\phi : R \to R$ of the form

$$\sum_i c_i t^i \mapsto \sum_i \phi_K(c_i) u^i,$$

where:

- the map $\phi_K : K \to (R \cap K\llbracket t \rrbracket_0)$ is an isometry whose composition with reduction modulo t is also an isometry;
- the element $u \in R$ satisfies $|u - t^q|_1 < 1$. (If $R = \mathcal{R}$, this forces $u \in \mathcal{E}^\dagger$.)

We say ϕ is *scalar-preserving* if ϕ_K carries K into K. We say ϕ is *absolute* if ϕ_K carries K into K and lifts the q-power Frobenius on κ_K.

LEMMA 14.2.2. *For any Frobenius lift ϕ on $K\llbracket t \rrbracket_0$, there exists a unique $\lambda \in \mathfrak{m}_K$ such that $\phi(t - \lambda) \equiv 0 \pmod{t - \lambda}$.*

PROOF. Exercise. $\square$

DEFINITION 14.2.3. We define the *center* of a Frobenius lift ϕ on $K\llbracket t \rrbracket_0$ to be the element $\lambda \in \mathfrak{m}_K$ given by Lemma 14.2.2. We say ϕ is *zero-centered* if its center is equal to 0, i.e., if $\phi(t) \equiv 0 \pmod{t}$.

REMARK 14.2.4. In case K is not discretely valued, one could also allow u such that $|u - t^q|_1 = 1$ but each coefficient of $u - t^q$ has norm less than 1. However, this creates certain technical complications which we do not wish to deal with here. In any case, we will restrict to K discretely valued at some point to avoid even further messiness.

3. Generic versus special Frobenius

DEFINITION 14.3.1. Let M be a finite free difference module over $K[\![t]\!]_0$, for ϕ a Frobenius lift. Define the *generic Newton polygon* of M to be the Newton polygon of $M \otimes \mathcal{E}$. Define the *special Newton polygon* of M to be the Newton polygon of $M/(t-\lambda)M$, for λ the center of ϕ.

THEOREM 14.3.2 (Grothendieck, Katz). *Let M be a finite free difference module over $K[\![t]\!]_0$, for ϕ a Frobenius lift. Then the special Newton polygon lies on or above the generic Newton polygon, with the same endpoints.*

PROOF. Choose a basis of M, and use it to define supremum norms on $M \otimes \mathcal{E}$ and $M/(t-\lambda)M$. Then it is evident that for any positive integer n, the Hodge polygon of Φ^n acting on $M \otimes \mathcal{E}$ lies on or above the Hodge polygon of Φ^n acting on $M/(t-\lambda)M$, with the same endpoints. If we divide all slopes by n and take limits as $n \to \infty$, then an analogue of Proposition 3.4.10 implies that the generic/special Hodge slopes converge to the generic/special Newton slopes. □

As in the comparison of Hodge and Newton polygons, one gets a decomposition result in case the special and generic Newton polygons touch somewhere.

THEOREM 14.3.3. *Let M be a finite free difference module of rank n over $K[\![t]\!]_0$, for ϕ a zero-centered Frobenius lift. Let $s_{g,1} \leq \cdots \leq s_{g,n}$ and $s_{s,1} \leq \cdots \leq s_{s,n}$ be the generic and special Newton slopes, respectively. Suppose that for some $i \in \{1, \dots, n-1\}$, we have*

$$s_{s,i} > s_{s,i+1}, \qquad s_{g,1} + \cdots + s_{g,i} = s_{s,1} + \cdots + s_{s,i}.$$

Then there is a unique difference submodule N of M with M/N free whose generic and special Newton slopes are $s_{g,1}, \dots, s_{g,i}$ and $s_{s,1}, \dots, s_{s,i}$, respectively.

PROOF. Uniqueness follows from the uniqueness in $M \otimes \mathcal{E}$, as in Theorem 13.4.13. For existence, we first replace ϕ by a suitable power to ensure that all of the slopes are integral multiples of $-\log p$; we then apply Lemma 10.5.1 to change basis in M to ensure that the generic Hodge slopes of M are also equal to $s_{g,1}, \dots, s_{g,i}$.

If $s_{s,H,1}, \dots, s_{s,H,n}$ denote the special Hodge slopes in this basis, then we have

$$s_{s,1} + \cdots + s_{s,i} \geq s_{s,H,1} + \cdots + s_{s,H,i}$$

by Corollary 13.5.4, but also

$$s_{s,H,1} + \cdots + s_{s,H,i} \geq s_{g,1} + \cdots + s_{g,i}$$

by Theorem 14.3.2, and $s_{g,1} + \cdots + s_{g,i} = s_{s,1} + \cdots + s_{s,i}$ by hypothesis. Consequently, $s_{s,1} + \cdots + s_{s,i} = s_{s,H,1} + \cdots + s_{s,H,i}$; that is, for this basis, the condition of Theorem 13.5.5 is also satisfied by M/tM.

We can thus choose a basis of M on which the action of Φ is via a block matrix

$$A = \begin{pmatrix} B & C \\ D & E \end{pmatrix}$$

in which modulo t, B accounts for the first i Hodge and Newton slopes of M/tM, E accounts for the remaining slopes of M/tM, and D vanishes. By Cramer's rule, DB^{-1} has entries in $\mathfrak{o}_K[\![t]\!]$. Conjugating by the block lower triangular unipotent matrix U with off-diagonal block DB^{-1}, we obtain

$$U^{-1}AU = \begin{pmatrix} B + CDB^{-1} & C \\ E\phi(DB^{-1}) - DB^{-1}C\phi(DB^{-1}) & E - DB^{-1}C \end{pmatrix}.$$

Repeating this operation, we get a sequence of block matrices

$$A_l = \begin{pmatrix} B_l & C_l \\ D_l & E_l \end{pmatrix}$$

in which $D_l B_l^{-1}$ has entries in $\mathfrak{o}_K[\![t]\!]$, and converges to zero for the $(t, \mathfrak{m}_K)$-adic topology (because $\phi(t) \equiv 0 \pmod{t^q, \mathfrak{m}_K}$). This proves the claim. □

THEOREM 14.3.4. *Let M be a finite free difference module of rank n over $K[\![t]\!]_0$, for ϕ a zero-centered Frobenius lift. Suppose that the generic and special Frobenius slopes of M are all equal to a single value r. Then there is a canonical isomorphism $M \cong (M/tM) \otimes_K K[\![t]\!]_0$ of differential modules.*

PROOF. First suppose that $r = 0$. By Lemma 10.5.1, we can choose a basis for which the generic Hodge slopes are all equal to 0. Let A be the matrix of action of Φ on this basis. We wish to construct an $n \times n$ matrix $U = \sum_{i=0}^{\infty} U_i t^i$ over $\mathfrak{o}_K[\![t]\!]$ with $U_0 = I_n$ such that $U^{-1}A\phi(U) = A_0$, or equivalently $U = A\phi(U)A_0^{-1}$. Since the map $U \mapsto A\phi(U)A_0^{-1}$ is contractive for the $(t, \mathfrak{m}_K)$-adic topology on $I_n + tM_{n\times n}(\mathfrak{o}_K[\![t]\!])$, it has a unique fixed point, which gives the desired isomorphism.

If $r \in v(K^\times)$, we may apply the above argument after twisting by a scalar. Otherwise, we may replace ϕ by a power, then twist and apply the above argument. □

Notes

Much of the existing literature makes the restriction that Frobenius lifts must be absolute, or at least scalar-preserving. The generality we consider here is relevant for some applications (e.g., to families of Galois representations in p-adic Hodge theory), so it is prudent to allow it as much as possible.

Theorem 14.3.2 in the absolute case is a local formulation of a geometric result of Grothendieck. The proof given is from [**Kat79**, Theorem 2.3.1]; the analogue of Proposition 3.4.10 used therein is [**Kat79**, Corollary 1.4.4].

Theorem 14.3.4 is an adaptation of [**Kat79**, Theorem 2.7.1].

Exercises

(1) Suppose that κ_K is perfect and that $\phi : \mathcal{E} \to \mathcal{E}$ is a q-power Frobenius lift inducing the absolute q-power Frobenius on $\kappa_K((t))$. Prove that ϕ is absolute in the sense of Definition 14.2.1, i.e., $\phi(K) = K$. (Hint: use Witt vector functoriality; otherwise put, for $x \in K$, show that $\phi(x^{p^n}) = \phi(x)^{p^n}$ is p-adically close to an element of K.)

(2) Prove Lemma 14.2.2. (Hint: view ϕ as a map taking $\lambda \in \mathfrak{m}_K$ to the reduction of $\phi(t)$ modulo $t - \lambda$. Then show that this map is contractive.)

CHAPTER 15

Frobenius structures on differential modules

In this chapter, we define the concept of a Frobenius structure on a differential module on a disc or annulus.

1. Frobenius structures

DEFINITION 15.1.1. Let R be a ring as in Definition 14.2.1. For M a finite free differential module over R, a *Frobenius structure* on M with respect to a Frobenius lift ϕ on R is an isomorphism $\Phi : \phi^* M \cong M$ of differential modules. In more explicit terms, we must equip M with the structure of a dualizable difference module over (R, ϕ), such that

$$D(\Phi(m)) = \frac{d\phi(t)}{dt}\Phi(D(v)) \qquad (m \in M).$$

In even more explicit terms, if A, N are the matrices via which Φ, D act on some bases, they must satisfy

$$NA + \frac{dA}{dt} = \frac{d\phi(t)}{dt} A\phi(N). \tag{15.1.1.1}$$

REMARK 15.1.2. We may also speak about Frobenius structures on finite free differential modules for the derivation $t\frac{d}{dt}$; the analogue of (15.1.1.1) is

$$NA + t\frac{dA}{dt} = \frac{t}{\phi(t)}\frac{d\phi(t)}{dt} A\phi(N). \tag{15.1.2.1}$$

However, if R is a subring of $K[\![t]\!]$, then (15.1.2.1) only makes sense if $\phi(t) = t^q u$ for $u \in R^\times$, in which case taking constant terms in (15.1.2.1) yields $N_0 A_0 = q u_0 A_0 \phi(N_0)$. This gives $|N_0|_{\mathrm{sp}} = q^{-1}|N_0|_{\mathrm{sp}}$, so N_0 must be nilpotent.

It is not easy to directly construct Frobenius structures except in a few simple examples. However, they frequently manifest on Picard-Fuchs modules; see Chapter 19.

2. Frobenius structures and generic radius of convergence

LEMMA 15.2.1. *Let ϕ be a Frobenius lift on $\mathcal{E}^\dagger$. Then there exists $\epsilon \in (0,1)$ such that for $\beta, \gamma \in [\epsilon, 1)$ with $\beta \leq \gamma$, ϕ carries $K\langle \beta/t, t/\gamma\rangle$ to $K\langle \beta^{1/q}/t, t/\gamma^{1/q}\rangle$, and*

$$|f|_\beta = |\phi(f)|_{\beta^{1/q}}.$$

PROOF. Since $|\phi(t)t^{-q} - 1|_1 < 1$, by continuity we can choose ϵ so that $|\phi(t)t^{-q} - 1|_{\rho^{1/q}} < 1$ for $\rho \in [\epsilon, 1]$; this inequality implies $|\phi(t^i)t^{-qi} - 1|_{\rho^{1/q}} < 1$ for all $i \in \mathbb{Z}$. For such an ϵ, the claim is easily verified: if $f = \sum_i f_i t^i$, then for $\beta \in [\epsilon, 1)$,

$$|f|_\beta = \left|\sum_i \phi_K(f_i)t^{qi}\right|_{\beta^{1/q}} > \left|\sum_i \phi_K(f_i)(t^{qi} - \phi(t^i))\right|_{\beta^{1/q}}.$$

$\square$

Remark 15.2.2. By virtue of Lemma 15.2.1, we can talk about Frobenius structures with respect to ϕ on finite differential modules on the half-open annulus with closed inner radius α and open outer radius 1, whether or not they are not represented by finite free modules over $K\langle \alpha/t, t\}\}_0$.

One of Dwork's early discoveries is that the presence of a Frobenius structure forces solvability at the boundary.

Proposition 15.2.3. *Let M be a finite differential module on the half-open annulus with closed inner radius α and open outer radius 1, equipped with a Frobenius structure. Then*

$$\lim_{\rho \to 1^-} IR(M \otimes F_\rho) = 1,$$

that is, M is solvable at the outer boundary. More precisely, for $\rho \in (0,1)$ sufficiently close to 1,

$$IR(M \otimes F_{\rho^{1/q}}) \geq IR(M \otimes F_\rho)^{1/q}.$$

Proof. By imitating the proof of Lemma 9.3.2, we may show that for $\rho \in (0,1)$ sufficiently close to 1,

$$IR(M \otimes F_{\rho^{1/q}}) \geq \min\{IR(M \otimes F_\rho)^{1/q}, qIR(M \otimes F_\rho)\}.$$

The function $f(s) = \min\{s^{1/q}, qs\}$ on $(0,1]$ is strictly increasing, and any sequence of the form $s, f(s), f(f(s)), \dots$ converges to 1. This proves the first claim; for the second claim, note that once s is sufficiently close to 1, $f(s) = s^{1/q}$. $\square$

The following corollary is sometimes called "Dwork's trick".

Corollary 15.2.4 (Dwork). *Let M be a finite differential module on the open unit disc, equipped with a Frobenius structure. Then M admits a basis of horizontal sections.*

Proof. By Proposition 15.2.3, for each $\lambda < 1$, there exists $\rho \in (\lambda, 1)$ such that $R(M \otimes F_\rho) > \lambda$. By Dwork's transfer theorem (Theorem 8.5.1), $M \otimes K\langle t/\lambda\rangle$ admits a basis of horizontal sections. Taking λ arbitrarily close to 1 yields the claim. $\square$

Remark 15.2.5. Corollary 15.2.4 admits the following geometric interpretation. By Proposition 8.2.3, the horizontal sections converge on some disc of positive radius ρ. Pulling back by Frobenius gives a new space of horizontal sections on the disc of radius $\min\{\rho^{1/p}, p\rho\}$, but this space must coincide with the original space. Repeating the construction, we eventually stretch the horizontal sections out over the entire open unit disc.

One also has a nilpotent analogue of Dwork's trick, by using Theorem 8.5.4 in place of Theorem 8.5.1.

Corollary 15.2.6. *Let M be a finite differential module on the open unit disc for the derivation $t\frac{d}{dt}$, equipped with a Frobenius structure as in Remark 15.1.2. Then M has radius of convergence 1.*

A nice application of Dwork's trick is the following.

PROPOSITION 15.2.7. *Let M be a finite differential module over $K\llbracket t \rrbracket_0$ with $R(M) = 1$. (For instance, this holds if M admits a Frobenius structure, by Dwork's trick.) Then $H^0(M) = H^0(M \otimes \mathcal{E}^\dagger)$.*

PROOF. By hypothesis, there exists a horizontal basis $e_1, \dots, e_n$ of $M \otimes K\{\{t\}\}$. If $v \in H^0(M \otimes \mathcal{E}^\dagger)$, then when we write $v = \sum_{i=1}^n v_i e_i$ with $v_i \in \mathcal{R}$, we must have $d(v_i) = 0$ for $i = 1, \dots, n$. This forces $v_i \in K$ for $i = 1, \dots, n$, so

$$v \in (M \otimes \mathcal{E}^\dagger) \cap (M \otimes K\{\{t\}\}) = M \otimes (\mathcal{E}^\dagger \cap K\{\{t\}\}) = M \otimes K\llbracket t \rrbracket_0 = M.$$

□

It should be noted that Dwork's trick also holds in the absence of a differential structure; see [**Ked05c**, Proposition 4.3].

THEOREM 15.2.8. *Let M be a finite free difference module over $K\{\{t\}\}$ for an absolute Frobenius lift. Then there exists a noncanonical isomorphism of difference modules $M \cong (M/tM) \otimes_K K\{\{t\}\}$.*

3. Independence from the Frobenius lift

Another key property of Frobenius structures is that the exact shape of the Frobenius lift is immaterial.

PROPOSITION 15.3.1. *Let ϕ_1, ϕ_2 be two Frobenius lifts on R. Let M be a finite free differential module over R equipped with a Frobenius structure for ϕ_1. Then there is a functorial way to equip M with a Frobenius structure for ϕ_2.*

PROOF. The Frobenius structure for ϕ_2 is defined by

$$\Phi_2(m) = \sum_{i=0}^{\infty} \frac{(\phi_2(t) - \phi_1(t))^i}{i!} \Phi_1 \left(\frac{d^i}{dt^i}(m) \right).$$

By Proposition 15.2.3 and the fact that $|\phi_2(t) - \phi_1(t)|_1 < 1$, this series converges under $|\cdot|_\rho$ for $\rho \in (0,1)$ sufficiently close to 1 (if this makes sense for R), and also under $|\cdot|_1$ (if this makes sense for R). □

COROLLARY 15.3.2. *Let ϕ_1, ϕ_2 be two Frobenius lifts on R. Then there is a canonical equivalence between the categories of finite free differential modules over R equipped with Frobenius structure with respect to ϕ_i for $i = 1, 2$; this equivalence is the identity functor on the underlying difference modules.*

Notes

We cannot resist viewing Dwork's trick (Corollary 15.2.4) as an instance of a general principle articulated beautifully by Coleman [**Col82**, §III]:

> Rigid analysis was created to provide some coherence in an otherwise totally disconnected p-adic realm. Still, it is often left to Frobenius to quell the rebellious outer provinces.

CHAPTER 16

Effective convergence bounds

In this chapter, we discuss some effective bounds on the solutions of p-adic differential equations with nilpotent singularities; we put this chapter here partly to illustrate the improvement one gets in the bounds by accounting for a Frobenius structure. Just like their archimedean counterparts, these are important for carrying out rigorous numerical calculations.

1. Nilpotent singularities in the p-adic setting

For applications in geometry, it is important to have effective bounds not just for nonsingular differential equations, but also for some regular singular differential equations. However, in the p-adic case, the p-adic behavior of the exponents creates many headaches. The case where the exponents are all zero is an important middle ground.

PROPOSITION 16.1.1. *Let $N = \sum_{i=0}^{\infty} N_i t^i$ be an $n \times n$ matrix over $K\langle t/\beta \rangle$ corresponding to the differential system $D(v) = Nv + d(v)$, where $d = t\frac{d}{dt}$. Assume that N_0 is nilpotent with nilpotency index m; that is, $N_0^m = 0$ but $N_0^{m-1} \neq 0$. Assume also that $|N_0| \leq 1$. Then the fundamental solution matrix $U = \sum_{i=0}^{\infty} U_i t^i$ over $K[\![t]\!]$ (as in Proposition 6.3.4) satisfies*

$$|U_i|\beta^i \leq |i!|^{-2m+1} \max\{|N_j|\beta^j : 0 \leq j \leq i\} \qquad (i = 1, 2, \dots). \tag{16.1.1.1}$$

Consequently, U has entries in $K\langle t/(p^{-(2m-1)/(p-1)}\beta)\rangle$ (as does its inverse).

Note that this reproves the p-adic Cauchy theorem (Proposition 8.2.3).

PROOF. Recall (6.3.4.1):

$$N_0 U_i - U_i N_0 + iU_i = -\sum_{j=1}^{i} N_j U_{i-j} \qquad (i > 0) >$$

The map $f(X) = NX - XN$ on $n \times n$ matrices is nilpotent with nilpotency index $2m - 1$: this is most easily seen by writing

$$f^i(X) = \sum_{j=0}^{i} a_j N^j X N^{i-j}$$

for some $a_j \in \mathbb{Z}$, then noting that each term vanishes for $i = 2m-1$ because $\min\{j, i-j\} \geq m$. Hence the map $X \mapsto iX + f(X)$ has inverse

$$X \mapsto \sum_{j=0}^{2m-2} (-1)^j i^{-j-1} f^j(X).$$

This gives the claim by induction on i. $\square$

2. Effective bounds for solvable modules

We now give an improved version of Proposition 16.1.1 under the hypothesis that U has entries in $K\langle t/\beta\rangle$. The hypothesis is only qualitative, in that it implies that $|U_i|\beta^i \to 0$ as $i \to \infty$ but does not give a specific bound on $|U_i|$ for any particular i. Somewhat surprisingly, this hypothesis plus any explicit bound on N together imply a rather strong explicit bound on $|U_i|$. We first suppose the bound on N is of a specific form.

Theorem 16.2.1. *Let $N = \sum_{i=0}^{\infty} N_i t^i$, $U = \sum_{i=0}^{\infty} U_i t^i$ be $n \times n$ matrices over $K[\![t]\!]$ such that*

(a) *N has entries in $K\langle t/\beta\rangle$;*
(b) *$U_0 = I_n$;*
(c) *$U^{-1}NU + U^{-1}t\frac{d}{dt}(U) = N_0$;*
(d) *N_0 is nilpotent;*
(e) *U and U^{-1} have entries in $K\langle t/\beta\rangle$.*

Then for every nonnegative integer i,

$$|U_i|\beta^i \leq p^{(n-1)\lfloor \log_p i \rfloor} \max\{1, |N|_\beta^{n-1}\}.$$

The first step in the proof of Theorem 16.2.1 is to change basis to reduce $|N|_\beta$; however, we pay the price of decreasing β slightly.

Lemma 16.2.2. *With notation as in Theorem 16.2.1, for any $\lambda < 1$, $\mu > 1$, there exists an invertible $n \times n$ matrix $X = \sum_{i=0}^{\infty} U_i t^i$ over $K\langle t/(\lambda\beta)\rangle$ such that*

$$\begin{aligned} |X^{-1}NX + U^{-1}t\frac{d}{dt}(X)|_{\lambda\beta} &\leq 1 \\ |X^{-1}|_{\lambda\beta} &\leq \mu \\ |X|_{\lambda\beta} &\leq |N|_\beta^{n-1}\mu. \end{aligned}$$

Proof. Let M be the differential module over $K\langle t/\beta\rangle$ for the operator $t\frac{d}{dt}$, with a basis on which D acts via N, and let $|\cdot|$ be the supremum norm defined by this basis. Since the fundamental solution matrix for M converges in the closed disc of radius β, M has generic radius of convergence β. In particular,

$$|D|_{\mathrm{sp},M} \leq \left| t\frac{d}{dt} \right|_{F_\beta} = 1.$$

By Proposition 5.2.11 plus the lattice lemma (Lemma 10.5.1), for any desired $\epsilon > 0$, we may find $V \in \mathrm{GL}_n(K\langle t/(\lambda^{1/2}\beta)\rangle)$ such that for $N' = V^{-1}NV + V^{-1}t\frac{d}{dt}(V)$,

$$\begin{aligned} |N'|_\beta &\leq 1 + \epsilon \\ |V^{-1}|_\beta &\leq 1 + \epsilon \\ |V|_\beta &\leq |N|_\beta^{n-1}(1 + \epsilon). \end{aligned}$$

Since the constant coefficient N'_0 of N' is nilpotent, it has spectral norm 0. By Proposition 3.4.6, there exists $W \in \mathrm{GL}_n(K)$ with

$$|W^{-1}| \leq 1, \qquad |W| \leq (1+\epsilon)^{n-1}, \qquad |W^{-1}N'_0W| \leq 1.$$

We now take $X = VW$, so that $X^{-1}NX + X^{-1}t\frac{d}{dt}(X) = W^{-1}N'W$. We then have

$$
\begin{aligned}
|(W^{-1}N'W)_0| &\leq 1 \\
|W^{-1}N'W|_\beta &\leq (1+\epsilon)^n \\
|X^{-1}|_\beta &\leq 1+\epsilon \\
|X|_\beta &\leq |N|_\beta^{n-1}(1+\epsilon)^n.
\end{aligned}
$$

For ϵ such that $(1+\epsilon)^n \leq \max\{\lambda^{-1}, \mu\}$, we have the desired inequalities. □

Using Lemma 16.2.2, we prove Theorem 16.2.1 by using Frobenius antecedents to reduce the index from i to $\lfloor i/p \rfloor$. One can improve upon this argument if one has a Frobenius structure on the differential module; see Lemma 16.3.2.

LEMMA 16.2.3. *With notation as in Theorem 16.2.1, suppose that $|N|_\beta \leq 1$. Then for any $\lambda < 1$, $\mu > 1$, there exist $n \times n$ matrices N', U' over $K\langle t/(\lambda^p\beta^p)\rangle$ satisfying the hypotheses of Theorem 16.2.1, such that*

$$
\begin{gathered}
|N'|_{\lambda\beta} \leq p \\
\max\{|U_j|(\lambda\beta)^j : 0 \leq j \leq i\} \leq \max\{|U'_j|(\lambda\beta)^{pj} : 0 \leq j \leq i/p\}.
\end{gathered}
$$

PROOF. Define the invertible $n \times n$ matrix $V = \sum_{i=0}^\infty V_i t^i$ over $K[\![t]\!]$ as follows. Start with $V_0 = I_n$. Given $V_0, \dots, V_{i-1}$, if $i \equiv 0 \pmod p$, put $V_i = 0$. Otherwise, put $W = \sum_{j=0}^{i-1} V_j t^j$ and $N_W = W^{-1}NW + W^{-1}t\frac{d}{dt}(W)$, and let V_i be the unique solution of the matrix equation

$$N_0 V_i - V_i N_0 + iV_i = -(N_W)_i.$$

By induction on i, $|V_i|\beta^i \leq 1$ for all i, so V is invertible over $K\langle t/(\lambda^{1/2}\beta)\rangle$.

Let $\phi : K[\![t]\!] \to K[\![t]\!]$ denote the substitution $t \mapsto t^p$. Put $N'' = V^{-1}NV + V^{-1}t\frac{d}{dt}(V)$; then N'' has entries in $K[\![t^p]\!]$, and $|\phi^{-1}(N'')|_{\lambda^{p/2}\beta^p} \leq 1$. Put $U'' = V^{-1}U$, so that $|U''|_{\lambda^{1/2}\beta} = 1$; then

$$(U'')^{-1}N''U'' + (U'')^{-1}t\frac{d}{dt}(U'') = N''_0 = N_0,$$

which forces U'' also to have entries in $K[\![t^p]\!]$. We may then take $N' = p^{-1}\phi^{-1}(N'')$ and $U' = \phi^{-1}(U'')$. □

We now put everything together.

PROOF OF THEOREM 16.2.1. We prove the claim by induction on i, in three stages. First, if $i < p$ and $|N|_\beta \leq 1$, then the desired estimate follows from Proposition 16.1.1. Second, for any given i, the desired estimate for general N follows from the estimate for the same i in the case $|N|_\beta \leq 1$, by Lemma 16.2.2. (More precisely, for any $\lambda < 1, \mu > 1$, replace the pair N, U by $X^{-1}NX + X^{-1}t\frac{d}{dt}(X), X^{-1}UX_0$; then take the limit as $\lambda, \mu \to 1$.) Third, if $|N|_\beta \leq 1$, then the desired estimate for any given i follows from the corresponding estimate for general N with i replaced by $\lfloor i/p \rfloor$, by Lemma 16.2.3 (again applying the argument for any $\lambda < 1, \mu > 1$, then taking the limit as $\lambda, \mu \to 1$). □

We will often apply Theorem 16.2.1 through the following corollary (deduced by taking β to be an arbitrary value less than 1).

THEOREM 16.2.4. *Let* $N = \sum_{i=0}^{\infty} N_i t^i$, $U = \sum_{i=0}^{\infty} U_i t^i$ *be* $n \times n$ *matrices over* $K[\![t]\!]$ *such that:*

(a) $|N|_1 < \infty$ *(i.e.,* $|N_i|$ *is bounded over all* i*);*
(b) $U_0 = I_n$*;*
(c) $U^{-1}NU + U^{-1}t\frac{d}{dt}(U) = N_0$*;*
(d) N_0 *is nilpotent;*
(e) *for all* $\beta < 1$, U *and* U^{-1} *have entries in* $K\langle t/\beta \rangle$.

Then for every nonnegative integer i,

$$|U_i| \leq p^{(n-1)\lfloor \log_p i \rfloor} |N|_1^{n-1}.$$

EXAMPLE 16.2.5. It is easy to make an example that shows that one cannot significantly improve the bound of Theorem 16.2.1 without extra hypotheses. (There is a tiny improvement possible; see notes.) For instance, one can use the functions

$$f_i = \frac{1}{i!}(\log(1+t))^i \qquad (i = 0, \dots, n-1)$$

which satisfy the differential system

$$\frac{d}{dt} f_0 = 0, \quad \frac{d}{dt} f_i = \frac{1}{1+t} f_{i-1} \qquad (i = 1, \dots, n-1),$$

in which the coefficients have 1-Gauss norm at most 1.

3. Frobenius structures

Although Theorem 16.2.4 is close to optimal under its hypotheses, it can be improved in case the differential module in question admits a Frobenius structure.

HYPOTHESIS 16.3.1. In this section, fix a power q of p, and let ϕ be a scalar-centered q-power Frobenius lift on $K[\![t]\!]_0$.

The key here is to imitate the proof of Theorem 16.2.1 with the differential equation replaced by a certain Frobenius equation.

LEMMA 16.3.2. *Let* $U = \sum_{i=0}^{\infty} U_i t^i$, $A = \sum_{i=0}^{\infty} A_i t^i$ *be* $n \times n$ *matrices over* $K[\![t]\!]$ *such that:*

(a) $|A|_1 < \infty$*;*
(b) $U_0 = I_n$ *and* A_0 *is invertible;*
(c) $U^{-1}A\phi(U) = A_0$.

Then

$$\max\{|U_j| : 0 \leq j \leq i\} \leq |A|_1 |A_0^{-1}| \max\{|U_j| : 0 \leq j \leq i/q\}.$$

Consequently, for every nonnegative integer i,

$$|U_i| \leq (|A|_1 |A_0^{-1}|)^{\lceil \log_q i \rceil}.$$

PROOF. Note that (c) can be rewritten as

$$U = A\phi(U)A_0^{-1}.$$

This gives the first inequality. To deduce the second inequality, we proceed as in the proof of Theorem 16.2.1, except that we iterate $\lceil \log_q i \rceil$ times to get to the case $i = 0$ (rather than iterating $\lfloor \log_q i \rfloor$ times to get to the case $0 < i < p$). □

THEOREM 16.3.3. *Let* $N = \sum_{i=0}^\infty N_i t^i$, $U = \sum_{i=0}^\infty U_i t^i$, $A = \sum_{i=0}^\infty A_i t^i$ *be* $n \times n$ *matrices over* $K[\![t]\!]$ *such that:*

(a) $|A|_1 < \infty$;
(b) $U_0 = I_n$ *and* A_0 *is invertible;*
(c) $U^{-1}NU + U^{-1}t\frac{d}{dt}(U) = N_0$;
(d) $NA + t\frac{d}{dt}(A) = qA\phi(N)$.

Then $U^{-1}A\phi(U) = A_0$, *and for every nonnegative integer* i,

$$|U_i| \leq (|A_0^{-1}||A|_1)^{\lceil \log_q i \rceil}.$$

PROOF. As noted in Remark 15.1.2, the commutation relation (d) implies that $N_0 A_0 = qA_0\phi(N_0)$, which forces N_0 to be nilpotent. Put $B = U^{-1}A\phi(U) = \sum_{i=0}^\infty B_i t^i$. Then $B_0 = A_0$, and $N_0 B + t\frac{d}{dt}(B) = qB\phi(N_0)$. Hence

$$N_0 B_i + iB_i = qB_i\phi(N_0) = B_i A_0^{-1} N_0 A_0,$$

or

$$N_0(B_i A_0^{-1}) + i(B_i A_0^{-1}) = (B_i A_0^{-1})N_0. \tag{16.3.3.1}$$

As in the proof of Proposition 16.1.1, the operator $X \mapsto N_0 X - XN_0 + iX$ on $n \times n$ matrices is invertible for $i \neq 0$, so (16.3.3.1) implies $B_i = 0$ for $i > 0$.

We conclude that indeed $U^{-1}A\sigma(U) = A_0$, so we may conclude by applying Lemma 16.3.2 to reduce to the case $i < q$, then applying Theorem 16.2.4. □

REMARK 16.3.4. By combining Theorem 16.2.4 with Theorem 16.2.1 (applying the latter for $i < q$), we can obtain the bound

$$|U_i| \leq |N|_1^{n-1} p^{(n-1)\lfloor \log_p i - (\log_p q)\lfloor \log_q i \rfloor \rfloor} (|A_0^{-1}||A|_1)^{\lfloor \log_q i \rfloor}.$$

REMARK 16.3.5. In applications to Picard-Fuchs modules, the difference between the bounds given by Theorem 16.2.4 and Theorem 16.3.3 can be quite significant. For instance, given a Picard-Fuchs module arising from a family of curves of genus g, the bound of Theorem 16.2.4 contains the factor $p^{(2g-1)\lfloor \log_p i \rfloor}$, but the bound of Theorem 16.3.3 replaces the factor of $2g - 1$ by 1. In general, it should be possible to use Theorem 16.3.3 (and perhaps also Theorem 16.3.6) to explain various instances in which a calculation of n terms of a power series involves a precision loss of $p^{O(\log(n))}$, even though the accumulated factors of p by which one divides throughout the calculation amount to $p^{O(n)}$. (A typical example of this is [**Ked01**, Lemma 3].)

We record also a sharper form of Theorem 16.3.3 for use in the discussion of logarithmic growth in the next section.

THEOREM 16.3.6. *Let* v *be a column vector of length* n *over* $K[\![t]\!]$, *let* $A = \sum_{i=0}^\infty A_i t^i$ *be an* $n \times n$ *matrix over* $K[\![t]\!]$, *and let* $\lambda \in K$ *be such that:*

(a) $|A|_1 < \infty$;
(b) A_0 *is invertible;*
(c) $A\sigma(v) = \lambda v$.

Then

$$\max\{|v_j| : 0 \leq j \leq i\} \leq |A|_1 |\lambda^{-1}| \max\{|v_j| : 0 \leq j \leq i/q\}.$$

Consequently, for every nonnegative integer i,

$$|v_i| \leq |v_0| (|A|_1 |\lambda^{-1}|)^{\lceil \log_q i \rceil}.$$

Proof. Rewrite (c) as $v = \lambda^{-1} A \sigma(v)$ and proceed as in Lemma 16.3.2. □

4. Logarithmic growth

Definition 16.4.1. For $\delta \geq 0$, let $K[\![t]\!]_\delta$ be the subset of $K[\![t]\!]$ consisting of those $f = \sum_{i=0}^\infty f_i t^i$ for which

$$|f|_\delta = \sup_i \left\{ \frac{|f_i|}{(i+1)^\delta} \right\} < \infty;$$

note that $K[\![t]\!]_\delta$ forms a Banach space under the norm $|\cdot|_\delta$. (The notation for $\delta = 0$ is consistent with our earlier usage.) However, $K[\![t]\!]_\delta$ is not a ring for $\delta > 0$; rather, we have

$$K[\![t]\!]_{\delta_1} \cdot K[\![t]\!]_{\delta_2} \subset K[\![t]\!]_{\delta_1 + \delta_2}.$$

Also, $K[\![t]\!]_\delta$ is stable under $\frac{d}{dt}$, but antidifferentiation carries it into $K[\![t]\!]_{\delta+1}$. Put

$$K[\![t]\!]_{\delta+} = \bigcap_{\delta' > \delta} K[\![t]\!]_{\delta'}.$$

We also consider a logarithmic version:

$$K[\![t]\!][\log t]_\delta = \bigoplus_{i=0}^{\lfloor \delta \rfloor} K[\![t]\!]_{\delta - i} (\log t)^i.$$

For another useful characterization of $K[\![t]\!]_\delta$, see the exercises.

Definition 16.4.2. For $f \in K[\![t]\!][\log t]$, we say that f has *order of log-growth δ* if $f \in K[\![t]\!][\log t]_\delta$ but $f \notin K[\![t]\!][\log t]_{\delta'}$ for any $\delta' < \delta$. We say f has *order of log-growth $\delta+$* if $f \notin K[\![t]\!][\log t]_\delta$ but $f \in K[\![t]\!][\log t]_{\delta'}$ for any $\delta' > \delta$. We have similar definitions for vectors or matrices over $K[\![t]\!][\log t]$, and for elements of $M \otimes K[\![t]\!][\log t]$ if M is a finite free module over $K[\![t]\!]_0$ (by computing in terms of a basis, the choice of which will not affect the answer).

We then deduce the following from Theorem 16.2.4.

Proposition 16.4.3. *Let M be a differential module of rank n over $K[\![t]\!]_0$ for the operator $t\frac{d}{dt}$, which is nilpotent at the origin. Then $M \otimes K[\![t]\!]_{n-1}[\log t]$ is trivial.*

Corollary 16.4.4. *Let M be a differential module of rank n over $K[\![t]\!]_0$ for the operator $t\frac{d}{dt}$, which is nilpotent at the origin with index of nilpotency e. Then any element of $H^0(M \otimes K[\![t]\!][\log t])$ has order of log-growth at most $n - 1 + e$.*

Remark 16.4.5. In Corollary 16.4.4, it should be possible to reduce $n - 1 + e$ to n.

In the presence of a Frobenius structure, one obtains a much sharper bound.

Theorem 16.4.6. *Let M be a differential module of rank n over $K[\![t]\!]_0$ for the operator $t\frac{d}{dt}$, equipped with a Frobenius structure for a q-power Frobenius lift as in Remark 15.1.2. Then any element $v \in H^0(M \otimes K[\![t]\!][\log t])$ satisfying $\Phi(v) = \lambda v$ for some $\lambda \in K$ has order of log-growth at most $(-\log|\lambda| - s_0)/(\log q)$, where s_0 is the least generic Newton slope of M.*

Proof. By replacing the Frobenius lift by some power, we can reduce to the case where s_0 is a multiple of $-\log p$. We can then twist into the case $s_0 = 0$. By Proposition 13.5.8, we can choose a basis of M such that the least generic Hodge slope of M is also 0. Then the claim follows immediately from Theorem 16.3.6. $\square$

Remark 16.4.7. Refining a conjecture of Dwork, Chiarellotto and Tsuzuki [**CT06**] have conjectured that if M is indecomposable, then Theorem 16.4.6 is optimal. That is, in the notation of Theorem 16.4.6, v should have order of log-growth exactly $(-\log|\lambda|-s_0)/(\log q)$); Chiarellotto and Tsuzuki have proven this for $\mathrm{rank}(M) \leq 2$ [**CT06**, Theorem 7.2]. It should be possible to extend their proof to all cases where $-\log|\lambda|$ is less than or equal to s_1 (the least Newton slope of M greater than s_0, not counting multiplicity), but it is less clear what happens in general.

Remark 16.4.8. By contrast, if M does not carry a Frobenius structure, then the order of log-growth of a horizontal section behaves much less predictably. For instance, it need not be rational, and it can have the form $\delta+$ instead of δ [**CT06**, §5.2].

5. Nonzero exponents

So far, we only have considered regular differential systems with all exponents equal to zero. Concerning nonzero exponents, we limit ourselves to two remarks.

Remark 16.5.1. Suppose the eigenvalues of N_0 are rational numbers with least common denominator dividing m. One can then apply Theorem 16.2.1 after making the substitution $t \mapsto t^m$, resulting in the bound

$$|U_i|\beta^i \leq p^{(n-1)\lfloor \log_p(im) \rfloor} \leq p^{(n-1)\lfloor \log_p m \rfloor} p^{(n-1)\lfloor \log_p i \rfloor}.$$

Note that as i varies, the difference between the bound in this case and in the nilpotent case is only a constant multiplicative factor.

Remark 16.5.2. Suppose that the eigenvalues of N_0 all belong to $\mathbb{Z}_p$. (One might want to consider this remark instead of Remark 16.5.1 even if the eigenvalues are rational, in case one does not have an *a priori* bound on their denominators.) One can then prove an effective bound by imitating the proof of Theorem 16.2.1, but using shearing transformations to force the exponents to be multiples of p before forming the Frobenius antecedent. However, the best known bound using this technique is worse than in Remark 16.5.1; it has the form $p^{(n^2+cn)\lfloor \log_p i \rfloor}$ for some constant c. See [**DGS94**, Theorem V.9.1] for more details.

Notes

In the case of no singularities ($N_0 = 0$), the effective bound of Theorem 16.2.4 is due to Dwork and Robba [**DR80**], with a slightly stronger bound: one may replace $p^{(n-1)\lfloor \log_p i \rfloor}$ with the maximum of $|j_1 \cdots j_{n-1}|^{-1}$ over $j_1, \dots, j_{n-1} \in \mathbb{Z}$ with $1 \leq j_1 < \cdots < j_{n-1} \leq i$. See also [**DGS94**, Theorem IV.3.1].

The general case of Theorem 16.2.1 is due to Christol and Dwork [**CD91**], except that their bound is significantly weaker: it is roughly $p^{c(n-1)\lfloor \log_p i \rfloor}$ with $c = 2 + 1/(p-1)$. The discrepancy comes from the fact that the role of Proposition 5.2.11 is played in [**CD91**] by an effective version of the cyclic vector theorem, which does not give optimal bounds. As usual, use of cyclic vectors also introduces singularities which must then be removed, leading to some technical difficulties. See also [**DGS94**, Theorem V.2.1]. (The poor estimate in the case of exponents in $\mathbb{Z}_p$ does not appear to be due to use of cyclic vectors.)

In the case of no singularities, Proposition 16.4.3 was first proved by Dwork; it appears in [**Dwo73a**] and [**Dwo73b**]. (See also [**Chr83**].) The nilpotent case appears to be original; as noted above, the effective bounds in [**CD91**] are not strong enough to imply this. Theorems 16.3.3 and 16.3.6 are original, but they owe a great debt to the proof of [**CT06**, Theorem 7.2]; the main difference is that we prefer to argue in terms of matrices rather than cyclic vectors.

The theory of logarithmic growth in the p-adic setting (which may be viewed as loosely analogous to its archimedean counterpart, as in [**Del70**]) emerged from some close analysis made by Dwork [**Dwo73a, Dwo73b**] of the finer convergence behavior of solutions of certain p-adic differential equations. The subject languished until the recent work of Chiarellotto and Tsuzuki [**CT06**]; inspired by this, André [**And07**] proved a conjecture of Dwork [**Dwo73b**, Conjecture 2] analogizing the specialization property of Newton polygons (Theorem 14.3.2) to logarithmic growth.

Exercises

(1) Prove that for $\delta \geq 0$,

$$K[\![t]\!]_\delta = \{f \in K[\![t]\!] : \limsup_{\rho \to 1^-} \frac{|f|_\rho}{(-\log \rho)^\delta} < \infty.\}$$

(Hint: the inequality

$$\sup_i \{(i+1)^\delta \rho^i\} \leq \rho^{-1} \left(\frac{\delta}{e}\right)^\delta (-\log \rho)^{-\delta}$$

may be helpful.)

CHAPTER 17

Quasiconstant differential modules

In this chapter, we construct a class of examples of differential modules on open annuli which are solvable at a boundary, and relate these to Galois representations of local fields in positive characteristc. We also assert (without proof) a numerical relationship between wild ramification and convergence of solutions of p-adic differential equations.

HYPOTHESIS 17.0.1. Throughout this chapter, assume that K is a complete *discretely valued* nonarchimedean field of characteristic 0 and residual characteristic p.

NOTATION 17.0.2. For E/F a Galois extension of fields, write $G_{E/F}$ for $\mathrm{Gal}(E/F)$. If $E = F^{\mathrm{sep}}$, write G_F instead, to mean the absolute Galois group.

1. Some key rings

DEFINITION 17.1.1. Recall that we defined the ring $\mathcal{E}$ as the completion of $\mathfrak{o}_K((t)) \otimes_{\mathfrak{o}_K} K$ for the 1-Gauss norm

$$\left| \sum_{i \in \mathbb{Z}} c_i t^i \right|_1 = \sup_i \{|c_i|\}.$$

Besides the p-adic topology, it is natural to consider also the *weak topology* on $\mathcal{E}$, in which a sequence converges to 0 if it does so in the t-adic topology on $\mathcal{E}/\mathfrak{m}_K^m \mathfrak{o}_{\mathcal{E}}$ for each $m \in \mathbb{Z}$. Note that $\mathcal{E}$ is complete for both topologies.

REMARK 17.1.2. Because K carries a discrete valuation, the supremum defining the Gauss norm of a nonzero element $x = \sum x_i t^i \in \mathcal{E}$ is achieved by some i. If j is the least such index, then the sum

$$x_j^{-1} t^{-j} \sum_{l=0}^{\infty} (1 - x_j^{-1} t^{-j} x)^l$$

converges in the weak topology (but not in the p-adic topology!) to an inverse of x. That is, $\mathcal{E}$ is a discrete complete nonarchimedean field with residue field $\kappa_K((t))$.

We now try to analogize from $\mathcal{E}$ to $\mathcal{E}^\dagger$, which is no longer complete.

LEMMA 17.1.3. (a) *The ring $\mathcal{E}^\dagger$ is a field.*
(b) *Under the norm $|\cdot|_1$, the valuation ring $\mathfrak{o}_{\mathcal{E}^\dagger}$ is a local ring with maximal ideal $\mathfrak{m}_K \mathfrak{o}_{\mathcal{E}^\dagger}$.*
(c) *The field $\mathcal{E}^\dagger$, equipped with $|\cdot|_1$, is henselian (see Remark 1.5.9).*

This last property implies that finite separable extensions of $\kappa_{\mathcal{E}^\dagger} = \kappa_K((t))$ lift functorially to finite étale extensions of $\mathfrak{o}_{\mathcal{E}^\dagger}$ (and to unramified extensions of $\mathcal{E}^\dagger$). In particular, the maximal unramified extension $\mathcal{E}^{\dagger,\mathrm{unr}}$ carries an action of $G_{\kappa_K((t))}$.

Proof. The proof of (a) uses the same construction as for $\mathcal{E}$, except that the series converges under $|\cdot|_\alpha$ for some $\alpha < 1$. From this, (b) is straightforward. The proof of (c) is to reduce to working in some $K\langle\alpha/t, t\rrbracket_0$ and use the fact that the latter ring is complete for the Fréchet topology generated by $|\cdot|_\alpha$ and $|\cdot|_1$. $\square$

2. Finite representations and differential modules

Definition 17.2.1. Let $\mathcal{E}_L^\dagger$ be the finite unramified extension of $\mathcal{E}^\dagger$ corresponding to L; then $G_{\kappa_K((t))}$ acts on $\mathcal{E}_L^\dagger$ with fixed field $\mathcal{E}^\dagger$.

Remark 17.2.2. By the Cohen structure theorem, L can always be written as a power series field $\lambda((u))$, and similarly for $\mathcal{E}_L^\dagger$. But if L induces an inseparable residue field extension, then you can't ensure that κ_K can be contained in λ. I recommend not worrying about this unless you really have to.

Definition 17.2.3. Let V be a finite dimensional vector space over K, and let $\tau : G_{\kappa_K((t))} \to \mathrm{GL}(V)$ be a continuous homomorphism for the *discrete* topology on $\mathrm{GL}(V)$. That is, τ factors through $G_{L/\kappa_K((t)))}$ for some finite separable extension L of $\kappa_K((t))$. Let us view $V \otimes_K \mathcal{E}_L^\dagger$ as a $G_{\kappa_K((t))}$-module with the action on the first factor coming from τ and the action on the second factor as above. Put

$$D^\dagger(V) = (V \otimes_K \mathcal{E}_L^\dagger)^{G_{\kappa_K((t))}}.$$

Lemma 17.2.4. *The space $D^\dagger(V)$ is an $\mathcal{E}^\dagger$-vector space of dimension* $\dim_K(V)$. *Equivalently, the natural map $D^\dagger(V) \otimes_{\mathcal{E}^\dagger} \mathcal{E}_L^\dagger \to V \otimes_K \mathcal{E}_L^\dagger$ is an isomorphism.*

Proof. This is a consequence of the nonabelian version of Hilbert's Theorem 90: for any finite Galois extension E/F of fields, the nonabelian cohomology set $H^1(G_{E/F}, \mathrm{GL}_n(E))$ is trivial. $\square$

Definition 17.2.5. Note that $\frac{d}{dt}$ extends uniquely to $\mathcal{E}_L^\dagger$, and hence to $D^\dagger(V)$ by taking the action on V to be trivial. Since the action of $\frac{d}{dt}$ commutes with the Galois action, we also obtain an action on $D^\dagger(V)$. That is, $D^\dagger(V)$ is a differential module over $\mathcal{E}^\dagger$. By the same token, $D^\dagger(V)$ admits a Frobenius structure for any Frobenius lift ϕ on $\mathcal{E}^\dagger$. This Frobenius structure is unit-root, because V admits a Galois-stable lattice L, and ϕ acts on $(V \otimes_K \mathcal{E}_L^\dagger)^{G_{\kappa_K((t))}}$.

Definition 17.2.6. Note that there is a sense in which it makes sense to compute the subsidiary radii of $D^\dagger(V) \otimes F_\rho$ for $\rho \in (0, 1)$ sufficiently close to 1. Namely, realize $D^\dagger(V)$ as a differential module over $K\langle\alpha/t, t\rrbracket_0$ for some α and compute there. Beware that any two such realizations for a given α need only become isomorphic over $K\langle\beta/t, t\rrbracket_0$ for some $\beta \in [\alpha, 1)$. However, statements about the germ at 1 of the function $\rho \mapsto R(D^\dagger(V) \otimes F_\rho)$ are unambiguous.

Proposition 17.2.7. *The generic radius of convergence of $D^\dagger(V) \otimes \mathcal{E}$ is equal to* 1. *Consequently (by continuity of generic radius of convergence), $D^\dagger(V)$ is solvable at* 1.

Proof. This follows from the existence of a Frobenius structure on $D^\dagger(V)$. $\square$

Remark 17.2.8. Note that the kernel of d on $\mathcal{E}_L^\dagger$ is the integral closure K' of K in $\mathcal{E}_L^\dagger$ (exercise). Consequently, the space of horizontal sections of $D^\dagger(V) \otimes_{\mathcal{E}^\dagger} \mathcal{E}_L^\dagger$ is equal to $V \otimes_K K'$. This suggests that we cannot recover all of V from $D^\dagger(V)$, at least if we only use the differential structure; instead, we only recover the restriction of V to the inertia subgroup of $G_{\kappa_K((t))}$, which we can identify with $G_{\kappa_K^{\mathrm{sep}}((t))}$.

The previous remark suggests the following construction.

Definition 17.2.9. Let $\tau : G_{\kappa_K((t))} \to \mathrm{GL}(V)$ be a homomorphism which is now continuous for the p-adic topology on V, rather than the discrete topology. One can form a differential module over $\mathcal{E}$ by taking

$$D(V) = (V \otimes_K \widehat{\mathcal{E}^{\mathrm{unr}}})^{G_{\kappa_K((t))}}$$

but this in general does not descend to $\mathcal{E}^\dagger$. Suppose, however, that the image of $G_{\kappa_K((t)),1} \cong G_{\kappa_K^{\mathrm{sep}}((t))}$ (the inertia subgroup) is finite; that is, τ has *finite local monodromy*. Let $\mathcal{E}^\dagger_{\kappa_K^{\mathrm{sep}}((t))}$ be the ring defined in the same fashion as $\mathcal{E}^\dagger$ but using $\widehat{K^{\mathrm{unr}}}$ on the coefficients; let $G_{\kappa_K((t))}$ act on this ring via the quotient by its inertia subgroup. We can then define

$$D^\dagger(V) = (V \otimes_K (\mathcal{E}^\dagger_{\kappa_K^{\mathrm{sep}}((t))})^{\mathrm{unr}})^{G_{\kappa_K((t))}}$$

and this will be a differential module over $\mathcal{E}^\dagger$ of the right dimension, again carrying a unit-root Frobenius structure.

Example 17.2.10 (Dwork). Assume that K contains an element π with $\pi^{p-1} = -p$; then K contains a unique p-th root of unity ζ_p satisfying $1 - \zeta_p \equiv \pi \pmod{\pi^2}$ (exercise). Let $L = \kappa_K((t))[z]/(z^p - z - \overline{f})$ be an Artin-Schreier extension, and let V be the Galois representation corresponding to the character of $G_{L/\kappa_K((t))}$ taking the automorphism $z \mapsto z+1$ to ζ_p. We can then explicitly describe $D^\dagger(V)$ as follows: if we pick any lift $f \in \mathfrak{o}_{\mathcal{E}^\dagger}$ of $\overline{f}$, then there is a generator v of $D^\dagger(V)$ with

$$D(v) = \pi \frac{df}{dt}.$$

Moreover, for any Frobenius lift ϕ, the Frobenius action on v is given by

$$\Phi(v) = \exp(\pi f - \pi\phi(f))v.$$

3. Ramification and differential slopes

There is a close relationship between $R(D^\dagger(V) \otimes F_\rho)$ and wild ramification of the representation V. To explain this, we must recall a bit of classical ramification theory for local fields, from [**Ser79**, Chapter IV].

Definition 17.3.1. Let F be a complete discrete nonarchimedean field whose residue field κ_F is perfect. (For more on what happens when the perfectness hypothesis is lifted, see the notes.) Let E be a finite Galois extension of F. The *lower numbering filtration* of $G_{E/F}$ is defined as follows: for $i \geq -1$ an integer.

$$G_{E/F,i} = \ker(G_{E/F} \to \mathrm{Aut}(\mathfrak{o}_E/\mathfrak{m}_E^{i+1})).$$

In particular, $G_{E/F,-1} = G_{E/F}$, and $G_{E/F,0}$ is the Galois group of E over its maximal unramified subextension (the *inertia group*). For $i \geq -1$ real, we define $G_{E/F,i} = G_{E/F,\lceil i \rceil}$.

The lower numbering filtration behaves nicely with respect to subgroups of $G_{E/F}$ but not quotients; it thus cannot be defined on the absolute Galois group G_F.

LEMMA 17.3.2. *The quotient $G_{E/F,0}/G_{E/F,1}$ is cyclic of order prime to p. For $i > 0$, the quotient $G_{E/F,i}/G_{E/F,i+1}$ is an elementary abelian p-group.*

COROLLARY 17.3.3. *The fixed field of $G_{E/F,1}$ is the maximal tamely ramified subextension of E over F, and $G_{E/F,1}$ is a p-group.*

The *upper numbering filtration* of $G_{E/F}$ is defined by the relation $G_{E/F}^{\phi_{E/F}(i)} = G_{E/F,i}$, where

$$\phi_{E/F}(i) = \int_0^i [G_{E/F,0} : G_{E/F,t}]^{-1}\, dt.$$

Note that the indices where the filtration jumps are now rational numbers, but not necessarily integers. In any case, Proposition 17.3.4 below implies that there is a unique filtration G_F^i on G_F which induces the upper numbering filtration on each $G_{E/F}$ (that is, $G_{E/F}^i$ is the image of G_F^i under the surjection $G_F \to G_{E/F}$).

PROPOSITION 17.3.4 (Herbrand). *Let E' be a Galois subextension of E/F, and put $H = \mathrm{Gal}(E/E')$, so that H is normal in $G_{E/F}$ and $G_{E/F}/H = G_{E'/F}$. Then $G_{E'/F}^i = (G_{E/F}^i H)/H$; that is, the upper numbering filtration is compatible with forming quotients of $G_{E/F}$.*

In the equal characteristic case, we can relate the ramification filtration to generic radius of convergence for suitable differential modules, as follows. We will not give a proof here; see the notes for attribution and references. (Also see the notes for discussion of what happens for imperfect residue field case, and in mixed characteristic.)

THEOREM 17.3.5. *Assume that κ_K is perfect. Let V be a finite dimensional vector space over K, and let $\tau : G_{\kappa_K((t))} \to \mathrm{GL}(V)$ be a continuous homomorphism for the p-adic topology on $\mathrm{GL}(V)$, with finite local monodromy. Then for $\rho \in (0,1)$ sufficiently close to 1,*

$$R(D^\dagger(V) \otimes F_\rho) = \rho^b, \qquad b = \max\{i \geq 1 : G^i_{\kappa_K((t))} \not\subseteq \ker(\tau)\}.$$

COROLLARY 17.3.6. *Let $V_1, \dots, V_m$ be the constituents of V, and let $\tau_j : G_{\kappa_K((t))} \to \mathrm{GL}(V_j)$ be the corresponding homomorphisms. For $\rho \in (0,1)$ sufficiently close to 1, the multiset of subsidiary radii of $D^\dagger(V) \otimes F_\rho$ consists of $\max\{i \geq 1 : G_{\kappa_K((t)),i} \not\subseteq \ker(\tau_j)\}$ with multiplicity $\dim(V_j)$, for $j = 1, \dots, m$.*

REMARK 17.3.7. One interpretation of Theorem 17.3.5 is that the decomposition of V by ramification numbers matches up with the Christol-Mebkhout decomposition of $D^\dagger(V) \otimes \mathcal{R}$ provided by Theorem 11.5.4. While the latter was inspired by analogues for meromorphic connections in the complex analytic setting, the analogy with wild ramification was anticipated somewhat before it was incarnated by Theorem 17.3.5.

REMARK 17.3.8. Using the integrality properties of subsidiary radii, we may deduce that for $\rho \in (0,1)$ sufficiently close to 1, the the product of the subsidiary radii is an integral power of ρ; this amounts to verifying the *Hasse-Arf theorem* for V (integrality of the Artin conductor).

4. Unit-root Frobenius structures

DEFINITION 17.4.1. If M is a differential module over $\mathcal{E}^\dagger$, we say M is *quasiconstant* if $M \otimes \mathcal{E}_L^\dagger$ admits a basis of horizontal sections for some L.

The goal of this section is to prove the following theorem of Tsuzuki [**Tsu98a**]; see notes for further discussion.

THEOREM 17.4.2 (Tsuzuki). *Let M be a finite differential module over $\mathcal{E}^\dagger$ admitting a unit-root Frobenius structure for some absolute scalar-preserving Frobenius lift. Then M is quasiconstant.*

Note that we must assume that the Frobenius lift is absolute; this is needed in some parts of the argument but not others, as we will attempt to illustrate.

In the course of proving Theorem 17.4.2, we will need to change Frobenius lifts; this is facilitated by the following observation.

LEMMA 17.4.3. *Let M be a finite differential module over $\mathcal{E}^\dagger$ equipped with a unit-root Frobenius structure for some scalar-preserving Frobenius lift. Then for any $c > 0$, there exists a basis of M on which D acts via a matrix N with $|N|_1 < c$.*

PROOF. If N is the matrix of action of D on the basis $e_1, \dots, e_n$, then the matrix of action of D on the basis $\phi(e_1), \dots, \phi(e_n)$ is $(d\phi(t)/dt)\phi(N)$. Since $|d\phi(t)/dt|_1 < 1$, iterating this construction eventually gives a basis of the desired form. $\square$

DEFINITION 17.4.4. Let M be a finite differential module over $\mathcal{E}^\dagger$ admitting a unit-root Frobenius structure for some Frobenius lift. For $c \in [0, 1)$, we say that a basis $e_1, \dots, e_n$ of M is *c-constant* if Φ acts on this basis via a matrix $A = \sum_i A_i t^i$ satisfying $|A|_1 = |A^{-1}| = 1$ and $|A - A_0|_1 \leq c$.

LEMMA 17.4.5. *Let M be a finite differential module over $\mathcal{E}^\dagger$ admitting a unit-root Frobenius structure for some Frobenius lift ϕ_1. Suppose that $e_1, \dots, e_n$ is a c-constant basis. Then the Frobenius structure corresponding to any other Frobenius lift ϕ_2 is also unit-root, and $e_1, \dots, e_n$ is again a c-constant basis.*

PROOF. This follows from the proof of Proposition 15.3.1, after first applying Lemma 17.4.3 to make the matrix of action of D sufficiently small. $\square$

REMARK 17.4.6. During the proof of Theorem 17.4.2, we will need to apply Remark 4.1.4 in the case where K' is a copy of K viewed as a K-algebra via a power of ϕ, then making finite separable extensions of $\kappa_{K'}((t))$ and deducing the conclusion of Theorem 17.4.2. In order to recover the desired conclusion over K, one must note that any finite separable extension of $\kappa_{K'}((t))$ is induced by a finite separable extension of $\kappa_K((t))$: simply write down a defining polynomial over $\kappa_{K'}((t))$, then apply ϕ until the coefficients of this polynomial end up in $\kappa_K((t))$.

REMARK 17.4.7. As a followup to the previous remark, note that if κ_K is not perfect, then not every finite separable extension of $\kappa_K((t))$ can be written as a field of formal Laurent series; for instance, if $c \in \kappa_K - \kappa_K^p$, then

$$\kappa_K((t))[z]/(z^p - z - ct^{-p})$$

cannot be so written. However, if we are allowed to replace K by some K' as in Remark 17.4.6, then this always becomes possible. To check this, it suffices to check for a totally wildly ramified extension: any finite separable extension of $\kappa_K((t))$ can be written as an unramified extension followed by a tamely ramified extension followed by a totally wildly ramified extension, and there is no problem with either of the first two types. Moreover, any totally wildly ramified extension is a tower of Artin-Schreier extensions (by Corollary 17.3.3 and some elementary group theory), so it suffices to check for an extension L of the form $\kappa_K((t))[z]/(z^p - z - x)$ for some $x = \sum_i x_i t^i \in \kappa_K((t))$. The terms x_i for $i > 0$ do not change the extension, so we may assume they all vanish. Moreover, we can replace x by $x + y^p - y$ for some y so that the nonzero i for which $x_i \neq 0$ all have the same p-adic valuation h. (If there are no such i, then L is unramified and we are done.) Then for suitable K', the compositum of L and $\kappa_{K'}((t))$ is generated over $\kappa_{K'}((t))$ is generated by a root of the polynomial $z^p - z - \phi^{-h}(x)$, in which $\phi^{-h}(x)$ has t-adic valuation not divisible by p. Such an extension can be written as a power series field.

Our next goal is to show that if M admits a c-constant basis for c sufficiently small, then M is constant. For this, it will be convenient to eliminate positive powers of t.

Lemma 17.4.8. *Let M be a finite differential module over $\mathcal{E}^\dagger$ admitting a unit-root Frobenius structure for some scalar-preserving Frobenius lift ϕ for which $\phi(t) = t^q$. Suppose that M admits a c-constant basis for some $c < 1$. Then there exists a c-constant basis $e_1, \dots, e_n$ of M on which Φ, tD act via matrices A, N with entries in $\mathcal{E}^\dagger \cap K[\![t^{-1}]\!]$.*

Proof. We first put A in the desired form; this can be done using Theorem 2.2.2, or explicitly as follows. Start with any basis, write the action of Φ as $A = \sum_i A_i t^i$, then replace A by $U^{-1} A \phi(U)$ for

$$U = I - \sum_{i>0} A_i t^i.$$

If we repeat this process, the resulting sequence of basis changes converges in the weak topology and puts A in the desired form.

To see that N is also in the desired form, consider the commutation relation

$$N = -t \frac{d}{dt}(A) A^{-1} + q A \phi(N) A^{-1}.$$

This implies that the coefficient of t^i in N vanishes for i positive and not divisible by q, then for i positive and not divisible by q^2, and so on. Hence N also has entries in $\mathcal{E}^\dagger \cap \mathfrak{o}_K[\![t^{-1}]\!]$. $\square$

We next treat an important special case of Theorem 17.4.2 which forms the heart of the proof.

Lemma 17.4.9. *Let M be a finite differential module over $\mathcal{E}^\dagger$ admitting a unit-root Frobenius structure for a scalar-preserving Frobenius lift. Suppose that M admits a c-constant basis for some $c < p^{-1/(p-1)}$. Then M is constant.*

Proof. By Lemma 17.4.5, we may assume that $\phi(t) = t^q$. By forming a restriction of scalars, we may assume that in fact $q = p$. By Remark 17.4.6, at any time we may replace K by a copy K' of K viewed as a K-algebra via a power of ϕ.

By Lemma 17.4.8, we may assume that there is a c-constant basis on which Φ, tD act via matrices A, N over $\mathcal{E}^\dagger \cap \mathfrak{o}_K[\![t^{-1}]\!]$. Then A and N together represent a finite differential module over $\cup_{\alpha\in(0,1)} K\langle \alpha/t\rangle$ equipped with a unit-root Frobenius structure. Moreover, the commutation relation $NA + t\frac{d}{dt}(A) = qA\phi(N)$ and the fact that $|A_0| = |A_0^{-1}| = |A|_1 = |A^{-1}|_1 = 1$ force $N_0 = 0$ and $|N|_1 \leq c$; consequently, there exists $\alpha \in (0,1)$ for which $|N|_\alpha \leq 1$.

We now proceed as in the proof of Lemma 16.2.3. As in that proof, we construct a matrix V with entries in $\mathcal{E}^\dagger \cap K[\![t^{-1}]\!]$ with $|V - I_n|_1 \leq c$ and $|V - I_n|_\alpha \leq 1$, for which $A' = V^{-1}A\phi(V)$ and $N' = V^{-1}NV + V^{-1}t\frac{d}{dt}(V)$ have entries in $\mathcal{E}^\dagger \cap K[\![t^{-p}]\!]$. After replacing K by some K' as above, we may take inverse images to obtain $A'' = \phi^{-1}(A'), N'' = p^{-1}\phi^{-1}(N')$. Since A'' and N'' again satisfy a commutation relation, we have $|N''|_1 \leq c$; that plus the estimate $|N'|_\alpha \leq 1$ yields $|N''|_\beta \leq 1$ for

$$\log \beta = \frac{\log \alpha}{p}\left(\frac{\log p}{\log c} + 1\right).$$

Since $c < p^{-1/(p-1)}$, we have $(\log p)/(\log c) + 1 > -p$, so $\beta < \alpha^{1+\epsilon}$ for some fixed $\epsilon > 0$.

As a result, we can make α arbitrarily small; in particular, we can force $\alpha < p^{-1/(p-1)}$. Now let M' be the differential module on the disc of radius α^{-1} in the coordinate t^{-1} with action of $t^{-1}\frac{d}{t^{-1}} = -t\frac{d}{dt}$ given by $-N$. The fact that $|N|_\alpha \leq 1$ implies that the spectral norm of $\frac{d}{dt^{-1}}$ on M' is at most α. Hence the generic radius of convergence at radius α^{-1} is at least $p^{-1/(p-1)}\alpha^{-1} > 1$, so Theorem 8.5.1 proves that the local horizontal sections at infinity converge on a disc of radius at least 1. We may then restrict these to a basis of horizontal sections of M. $\square$

Lemma 17.4.10. *Let V be a finite dualizable difference module over $\kappa_K((t))$. Then there exists a positive integer m coprime to p such that $V \otimes \kappa_K((t^{1/m}))$ can be written as a successive quotient of difference modules defined over κ_K.*

Proof. We may assume V is irreducible. Pick $v \in V$ nonzero, and write $\Phi^n(v) = \sum_{i=0}^{n-1} c_i \Phi^i(v)$. By rescaling v by a suitable power of $t^{1/(q-1)}$, we can ensure that the minimum t-adic valuation of the c_i is 0. By Theorem 2.2.2, if c_0 has positive valuation, then the twisted polynomial $T^n - \sum_{i=0}^{n-1} c_i T^i$ factors nontrivially in $\kappa_K((t^{1/(q-1)}))\{T\}$, so V has become reducible, and we may reduce to cases of lower dimension. Otherwise, we have a basis of V on which Φ acts via an invertible matrix over $\kappa_K[\![t^{1/(q-1)}]\!]$, in which case it is an exercise (as in Lemma 17.4.8) to change basis so that Φ acts via an invertible matrix over κ_K. $\square$

Lemma 17.4.11. *Let M be a finite differential module over $\mathcal{E}^\dagger$ admitting a unit-root Frobenius structure for some scalar-preserving Frobenius lift. Then for some positive integer m coprime to p and some finite extension K' of K with the same residue field (admitting an extension of ϕ), $M \otimes \mathcal{E}^\dagger[t^{1/m}]$ admits a c-constant basis for some $c \in (0,1)$.*

Proof. Pick any basis of M on which Φ acts via a matrix over $\mathfrak{o}_{\mathcal{E}^\dagger}$, then apply Lemma 17.4.10. We get a new basis on which Φ acts a matrix over $\mathfrak{o}_{\mathcal{E}^\dagger}$ whose reduction modulo $\mathfrak{m}_K$ is block upper triangular. To fix this, we need to conjugate by a block diagonal matrix whose entries have norms sufficiently close to 1; we can accomplish this after adjoining $p^{1/h}$ for some h coprime to both p and the absolute ramification index of K. $\square$

Remark 17.4.12. In the case of an absolute Frobenius lift, which is all that is covered by Theorem 17.4.2, the proof given above is a bit of overkill; instead, one may simply invoke Proposition 13.3.4.

Here is the most crucial use of the absoluteness of the Frobenius lift.

LEMMA 17.4.13. *Let M be a finite differential module over $\mathcal{E}^\dagger$ admitting a unit-root Frobenius structure for some absolute scalar-preserving Frobenius lift, and a c-constant basis for some $c \in (0,1)$. For m a positive integer, let K_m be a copy of K viewed as a K-algebra via ϕ^m. Then for some m, there exists a finite separable extension L of $\kappa_{K_m}((t))$ with the following properties.*

(a) *The residue field κ_L is equal to the integral closure of κ_{K_m} in L.*
(b) *There exists a uniformizer $u \in \kappa_L$ so that $\phi^h(u) = u^{q^h}$ for some positive integer h.*
(c) *There exists a c'-constant basis of $M \otimes \mathcal{E}^\dagger_L$ for some $c' \in (0, c)$.*

PROOF. Let $e_1, \dots, e_n$ be a c-constant basis, and suppose Φ, tD act on this basis via the matrices A, N. We may assume $|A - A_0|_1 = c$; choose $\lambda \in K$ with $|\lambda| = c$. We may also assume that each index i for which $|A_i| = c$ is negative and not divisible by p (for this we may need to pass from K to K_m).

It suffices to exhibit an $n \times n$ matrix U over L, for suitable L, such that

$$UA - (\phi(\lambda)/\lambda)A\phi(U) \equiv \lambda^{-1}(A - A_0) \pmod{\mathfrak{m}_K}.$$

This amounts to adjoining the roots of some separable polynomials to $\kappa_K((t))$; we can enforce (a) and (b) as in Remark 17.4.7. □

PROOF OF THEOREM 17.4.2. By Lemma 17.4.9, it suffices to exhibit a c-constant basis of $M \otimes \mathcal{E}^\dagger_L$ for some finite separable extension L of $\kappa_K((t))$ and some $c < p^{-1/(p-1)}$. By Lemma 17.4.11, we can produce such for some $c \in (0,1)$ (after replacing K by a finite extension, but this is harmless by Remark 4.1.4); by Lemma 17.4.13, we can reduce c by a discrete multiplicative factor. Thus we eventually obtain the desired basis and complete the proof. □

Notes

A more detailed survey of most of the material in this chapter is the article [**Ked05a**].

Ramification theory for a complete discrete nonarchimedean field with imperfect residue field is substantially more complicated than in the case of a perfect residue field. On the other hand, it is of interest in the study of finite covers of schemes of dimension greater than 1. A satisfactory theory for abelian extensions was introduced by Kato [**Kat89**]. A generalization to nonabelian extensions was later introduced by Abbes and Saito [**AS02, AS03**].

Theorem 17.3.5 was originally stated in its present form by Matsuda [**Mat02**, Corollary 8.8]; a reformulation in the formalism of Tannakian categories was given by André [**And02**, Complement 7.1.2]. However, thanks to the p-adic global index theorem of Christol and Mebkhout [**CM00**, Theorem 8.4–1], [**CM01**, Corollaire 5.0–12], this could have already been deduced from a Grothendieck-Ogg-Shafarevich formula for unit-root overconvergent F-isocrystals in rigid cohomology; such a formula was proved by Tsuzuki [**Tsu98b**, Theorem 7.2.2] (by Brauer induction, as is possible in the ℓ-adic case) and Crew [**Cre00**, Theorem 5.4] (using the Katz-Gabber theory of canonical extensions, as also is possible in the ℓ-adic case). For a proof by direct computations and Brauer inductions (not using the Christol-Mebkhout theory), see [**Ked05a**, Theorem 5.23].

In the case of an imperfect residue field, it was originally suggested by Matsuda [**Mat04**] to formulate an analogue of Theorem 17.3.5 relating the Abbes-Saito conductor to a suitable differential analogue. That differential analogue was described by Kedlaya [**Ked07a**]; the comparison with the Abbes-Saito conductor has been established by Chiarellotto and Pulita [**CP07**] for one-dimensional representations, and by Xiao [**Xia07**] in the general case. This has the side effect of establishing integrality of the Abbes-Saito conductor in equal characteristic, which is not evident from the original construction.

In mixed characteristic, the appropriate analogue of the functor $V \mapsto D(V)$ is provided by the theory of (ϕ, Γ)-modules; see Chapter 21. It is distinctly less clear what sort of analogue of Theorem 17.3.5 should exist in mixed characteristic. Even in the case of a perfect residue field, where one is asking for a differential interpretation of the usual conductor, only a partial answer is known, by a result of Marmora [**Mar04**].

Theorem 17.4.2 is due to Tsuzuki [**Tsu98a**, Theorem 5.1.1] in case κ_K is algebraically closed; instead of proceeding as we have, one may instead deduce the general case from that special case. An alternate exposition was given by Christol [**Chr01**], but Christol assumes his Frobenius is always in the standard for $t \mapsto t^p$, and neglects to point out that this is not stable under making extensions of $\kappa_K((t))$ (which we remedy by performing a change of Frobenius). Yet another exposition may be inferred from [**Ked07b**, Theorem 4.5.2], where a stronger result is proved (for purposes that we will not discuss here). We will discuss the analogue of Theorem 17.4.2 for a nonabsolute scalar-preserving Frobenius lift in the next chapter.

Exercises

(1) Prove that for any finite separable extension L of $\kappa_K((t))$, the kernel of d on $\mathcal{E}^\dagger_L$ is the integral closure of K in $\mathcal{E}^\dagger_L$.
(2) Suppose that K contains an element π with $\pi^{p-1} = -p$. Prove that K contains a unique p-th root of unity ζ_p satisfying $1 - \zeta_p \equiv \pi \pmod{\pi^2}$. (Hint: reduce to Hensel's lemma.)
(3) Let ϕ be an endomorphism of $\kappa_K((t))$ of the form $\phi(\sum_i c_i t^i) = \sum_i \phi_K(c_i) t^{qi}$, where ϕ_K is an endomorphism of κ_K and $q > 1$. Prove that for any invertible matrix $A = \sum_{i=0}^\infty A_i t^i$ over $\kappa_K[\![t]\!]$, there is a unique matrix $U = \sum_{i=0}^\infty U_i t^i$ over $\kappa_K[\![t]\!]$ with $U_0 = I_n$ such that $U^{-1} A \phi(U) = A_0$.

CHAPTER 18

The p-adic local monodromy theorem

In this chapter, we assert the p-adic local monodromy theorem, and sketch its proof. The sketchiness arises because one must invoke one or the other of two facts too difficult to be proved in this book: the p-adic index theorem of Christol-Mebkhout, or a slope filtration theorem of Kedlaya for Frobenius modules over the Robba ring.

1. Statement of the theorem

REMARK 18.1.1. Recall that we have defined the Robba ring to be

$$\mathcal{R} = \cup_{\alpha \in (0,1)} K\langle \alpha/t, t\}\};$$

that is, $\mathcal{R}$ consists of formal sums $\sum c_i t^i$ which converge in some range $\alpha \leq |t| < 1$, but need not have bounded coefficients. Unlike its subring $\mathcal{E}^\dagger$, $\mathcal{R}$ is not a field; for instance, the element

$$\log(1+t) = \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} t^i$$

is not invertible (because its Newton polygon has infinitely many slopes). More generally, we have the following easy fact.

LEMMA 18.1.2. *We have $\mathcal{R}^\times = (\mathcal{E}^\dagger)^\times$.*

DEFINITION 18.1.3. Because $\mathcal{R}$ consists of series with possibly unbounded coefficients, it does not carry a natural p-adic topology. The most useful topology on $\mathcal{R}$ is the *LF topology*, which is the direct limit of the Fréchet topology on each $K\langle \alpha/t, t\}\}$ defined by the $|\cdot|_\rho$ for $\rho \in [\alpha, 1)$.

In fact, the ring $\mathcal{R}$ is not even noetherian (this is related to an earlier exercise), but the following useful facts are true; see notes.

PROPOSITION 18.1.4. *For an ideal I of $\mathcal{R}$, the following are equivalent.*

(a) *The ideal I is closed in the LF topology.*
(b) *The ideal I is finitely generated.*
(c) *The ideal I is principal.*

PROPOSITION 18.1.5. *Any finite free module on the half-open annulus with closed inner radius α and open outer radius 1 is represented by a finite free module over $K\langle \alpha/t, t\}\}$, and so corresponds to a finite free module over $\mathcal{R}$. (The first part generalizes to half-open and open annuli with arbitrary boundary radii.)*

DEFINITION 18.1.6. For L a finite separable extension of $\kappa_K((t))$, put

$$\mathcal{R}_L = \mathcal{R} \otimes_{\mathcal{E}^\dagger} \mathcal{E}_L^\dagger.$$

We say a finite differential module M over $\mathcal{R}$ is *quasiconstant* if there exists L such that $M \otimes \mathcal{R}_L$ is trivial. We say M is *quasiunipotent* if it is a successive extension of quasiconstant modules; it is equivalent to ask that $M \otimes \mathcal{E}^\dagger_L$ be unipotent (i.e., an extension of trivial differential modules) for some L (exercise).

Quasiunipotent differential modules have many useful properties. For instance, by Proposition 17.2.7, they are all solvable at 1. Another important property is the following.

PROPOSITION 18.1.7. *Let M be a quasiunipotent differential module over $\mathcal{R}$. Then the spaces $H^0(M), H^1(M)$ are finite dimensional, and there is a perfect pairing*

$$H^0(M) \times H^1(M^\vee) \to H^1(M \otimes M^\vee) \to H^1(\mathcal{R}) \cong K\frac{dt}{t}.$$

PROOF. This can be reduced to the unipotent case, for which it is an exercise. □

The following important theorem asserts that many naturally occurring differential modules, including Picard-Fuchs modules, are quasiunipotent. In the case of an absolute Frobenius lift, it is due independently to André [**And02**], Kedlaya [**Ked04a**], and Mebkhout [**Meb02**]. See the notes for further discussion.

THEOREM 18.1.8 (p-adic local monodromy theorem). *Let M be a finite differential module over $\mathcal{R}$ admitting a Frobenius structure for some scalar-preserving Frobenius lift. Then M is quasiunipotent.*

REMARK 18.1.9. One can show that if the conclusion of Theorem 18.1.8 holds for a given M after replacing K by a finite extension K', then it holds without that replacement. Namely, let L be a finite separable extension of $\kappa_K((t))$. Let λ be the maximal separable subextension of κ_L/κ_K. Let K_1 be the unramified extension of K with $\kappa_{K_1} = \lambda$. Suppose that M is nonzero and that K' is a finite extension of K_1 such that $M \otimes_{\mathcal{R}} \mathcal{R}_L \otimes_{K_1} K$ is quasiunipotent. (Note that our hypotheses so far ensure that $\mathcal{R}_L \otimes_{K_1} K$ is again a Robba ring.) By Remark 4.1.4,

$$H^0(M \otimes_{\mathcal{R}} \mathcal{R}_L) \otimes_{K_1} K' = H^0(M \otimes_{\mathcal{R}} \mathcal{R}_L \otimes_{K_1} K');$$

since the latter is nonzero, $H^0(M \otimes_{\mathcal{R}} \mathcal{R}_L) \neq 0$. Hence $M \otimes_{\mathcal{R}} \mathcal{R}_L$ contains a nontrivial constant submodule; by repeating the argument, we deduce that $M \otimes_{\mathcal{R}} \mathcal{R}_L$ is unipotent, and so M is quasiunipotent.

With a slightly more involved argument, replacing Remark 4.1.4 by a suitable use of linear compactness, one can make the same argument for any complete extension K' of K. See the proof of [**Ked04a**, Proposition 6.11] for a brutally elementary description of the technique.

REMARK 18.1.10. Using Remark 18.1.9, we may safely make finite separable extensions L of $\kappa_K((t))$ without worrying about whether $\mathcal{R}_L$ will still be a Robba ring; namely, we may apply the technique of Remark 17.4.7 to fix this by making a suitable extension of K.

2. An example

It may be worth seeing what Theorem 18.1.8 says in an explicit example. This example was originally considered by Dwork [**Dwo74**]; the analysis given here is due to Tsuzuki [**Tsu98c**, Example 6.2.6], and was cited in the introduction of [**Ked04a**].

EXAMPLE 18.2.1. Let p be an odd prime, put $K = \mathbb{Q}_p(\pi)$ with $\pi^{p-1} = -p$. Let M be the differential module of rank 2 over $\mathcal{R}$ with the action of D on a basis e_1, e_2 given by

$$N = \begin{pmatrix} 0 & t^{-1} \\ \pi^2 t^{-2} & 0 \end{pmatrix}.$$

Then M admits a Frobenius structure; this was shown by explicit calculation in [**Dwo74**], but can also be derived by consideration of a suitable Picard-Fuchs module. Define the tamely ramified extension L of $\kappa_K((t))$ by adjoining u such that $4u^2 = t$, and put

$$u_\pm = 1 + \sum_{n=1}^{\infty} (\pm 1)^n \frac{(2n)!^2}{(32\pi)^n n!^3} u^n \in K\{\{u\}\}.$$

Define the matrix

$$U = \begin{pmatrix} u_+ & u_- \\ u\frac{d}{du}(u_+) + (\frac{1}{2} - \pi u^{-1})u_+ & u\frac{d}{du}(u_-) + (\frac{1}{2} + \pi u^{-1})u_- \end{pmatrix}$$

and use it to change basis; then the action of $\frac{d}{du}$ on the new basis e_+, e_- of $M \otimes \mathcal{R}_L$ is via the matrix

$$\begin{pmatrix} -\frac{1}{2}^y -1 + \pi y^{-2} & 0 \\ 0 & -\frac{1}{2}y^{-1} - \pi y^{-2}. \end{pmatrix}.$$

That is, $M \otimes \mathcal{R}_L$ splits into two differential submodules of rank 1. To render these quasi-constant, we must adjoin to L to L a square root of y (to eliminate the terms $-\frac{1}{2}y^{-1}$) and a root of the polynomial $z^p - z = y^{-1}$ (which by Example 17.2.10 eliminates the terms $\pm\pi y^{-2}$).

By further analysis (carried out in [**Tsu98c**, Example 6.2.6]), one determines that in this example, the generic Newton slopes are 0 and $\log p$, while the special Newton slopes are both $\frac{1}{2}\log p$.

3. The monodromy theorem in rank 1

In Theorem 18.1.8, the case $\mathrm{rank}(M) = 1$ occupies a special role. On one hand, it is somewhat more approachable than the general case. On the other hand, depending on the approach one takes to proving Theorem 18.1.8 in full, one is either obliged or strongly recommended to understand the rank 1 case first.

One approach to the rank 1 case, using only properties of differential modules and some explicit calculations, has been described by Mebkhout [**Meb02**, Théorème 2.0–1]; we give only a sketch here.

LEMMA 18.3.1 (Mebkhout). *Let M be a differential module of rank 1 over $\mathcal{R}$ which is solvable at 1. Then M can be written as the tensor product of a quasiconstant module and a module on which tD acts via a scalar in K.*

SKETCH OF PROOF. By Lemma 11.5.2, there exists a nonnegative integer b such that $IR(M \otimes F_\rho) = \rho^b$ for $\rho \in (0,1)$ sufficiently close to 1. We proceed by induction on b; it suffices to find a quasiunipotent module M' of rank 1 for which $IR(M \otimes M' \otimes F_\rho) > \rho^b$ for $\rho \in (0,1)$ sufficiently close to 1. These can be produced by considering $D(V)$ for characters V of $G_{\kappa_K((t))}$; see [**Meb02**, Théorème 2.0–1] for the full calculation. $\square$

REMARK 18.3.2. Using Lemma 18.3.1, in order to prove Theorem 18.1.8 in rank 1, it suffices to prove that any rank 1 module on which tD acts via a scalar in K is quasiconstant. If λ is this scalar, then the presence of a q-power Frobenius action forces $q\lambda = \lambda$, so $\lambda \in \mathbb{Z}_p \cap \mathbb{Q}$ and adjoining a $(q-1)$-st root of t suffices to render the module constant.

A second approach, which uses the Frobenius action more directly, can be inferred from some results of Crew [**Cre87**].

LEMMA 18.3.3. *Let M be a differential module of rank 1 over $\mathcal{R}$ equipped with a Frobenius structure. Then for any generator v of M, we have $\Phi(v) = av$ and $tD(v) = nv$ for some $a, n \in \mathcal{E}^\dagger$.*

PROOF. Since a must be a unit in $\mathcal{R}$, it is forced to be in $\mathcal{E}^\dagger$. To check $n \in \mathcal{E}^\dagger$, we may switch to a Frobenius lift ϕ with $\phi(t) = t^q$. The compatibility between the Frobenius and differential structures in this case amounts to the equality $na + t\frac{da}{dt} = qa\phi(n)$, or

$$n - q\phi(n) = \frac{t}{a}\frac{da}{dt}. \tag{18.3.3.1}$$

The right side belongs to $\mathcal{E}^\dagger$, so we can choose $c > 0$ for which $|\frac{t}{a}\frac{da}{dt}|_1 < c$. Write $n = \sum_{i \in \mathbb{Z}} n_i t^i$; we claim that $|n_i| \leq c$ for all $i > 0$. We establish this by induction on i. Compare the coefficients of t^i on both sides of (18.3.3.1): if i is not divisible by q this yields $|n_i| \leq c$ directly, otherwise it yields $|n_i - q\phi(n_{i/q})| \leq c$. Since we have by the induction hypothesis $|q\phi(n_{i/q})| < |\phi(n_{i/q})| = |n_{i/q}| \leq c$, in either case we deduce $|n_i| \leq c$.

Since $|n_i|$ is automatically bounded for $i \leq 0$, we deduce that $|n_i|$ is bounded over all i, whence $n \in \mathcal{E}^\dagger$. □

LEMMA 18.3.4 (Crew). *Let M be a differential module of rank 1 over $\mathcal{E}^\dagger$. Then there exists a positive integer m such that $M^{\otimes m}$ admits a generator on which tD acts via a scalar in K.*

PROOF. Let v be a generator of M, and define $n \in \mathcal{E}^\dagger$ by $tD(v) = nv$; then for any m, $v^{\otimes m}$ is a generator of $M^{\otimes m}$ and $(tD)(v^{\otimes m}) = mnv$. Write $n = \sum_{i \in \mathbb{Z}} n_i t^i$ and put $v = \sum_{i \neq 0} (n_i/i) t^i \in \mathcal{E}^\dagger$, so that $t\frac{dv}{dt} = n - n_0$.

If we take m to be a sufficiently large power of p, we will then have $|mv|_\rho < p^{-1/(p-1)}$ for ρ in some range of the form $[\alpha, 1]$. For such m, we can form $u = \exp(-mv) \in \mathcal{E}^\dagger$, and $m(n-n_0)u + t\frac{du}{dt} = 0$. Consequently, the generator $w = uv^{\otimes m}$ of M satisfies $D(w) = n_0 w$. □

COROLLARY 18.3.5. *Let M be a differential module of rank 1 over $\mathcal{R}$ admitting a Frobenius structure. Then there exists a positive integer m such that $M^{\otimes m}$ is constant.*

PROOF. By Lemma 18.3.3, M is the base extension of a differential module over $\mathcal{E}^\dagger$. By Lemma 18.3.4, there exists a positive integer m_0 such that $M^{\otimes m_0}$ admits a generator on which tD acts via a scalar in K. As in Remark 18.3.2, this implies that $M^{\otimes m}$ is constant for $m = m_0(q-1)$. □

One now easily reduces the rank 1 case of Theorem 18.1.8 to the following lemma.

LEMMA 18.3.6. *Let M be a differential module of rank 1 over $\mathcal{R}$ admitting a Frobenius structure, such that $M^{\otimes p}$ is constant. Then M is quasiconstant.*

Proof. We may assume $\phi(t) = t^q$ and $\zeta_p \in K$ (the latter thanks to Remark 18.1.9). Let v be a generator of M. As in Lemma 18.3.3, we have $\Phi(v) = av$ and $tD(v) = nv$ for some $a, n \in \mathcal{E}^\dagger$ satisfying (18.3.3.1). Note that we may twist M by rescaling the Φ-action by a factor in $K^\times$ without changing n. We may thus ensure $|a|_1 = 1$, and then (18.3.3.1) forces $|n|_1 \leq 1$ (since $|n - q\phi(n)|_1 = |n|_1$ once we know $n \in \mathcal{E}^\dagger$).

The fact that $M^{\otimes p}$ is constant means that there exists $u \in \mathcal{E}^\dagger$ such that

$$pn = \frac{t}{u}\frac{du}{dt}, \qquad \frac{a^p}{\lambda} = \frac{\phi(u)}{u}$$

for some $\lambda \in K$. After replacing K by a finite extension, we may reduce to the case $\lambda = 1$.

We now make a series of changes of the choice of the generator v. Let $\overline{u}, \overline{a}$ denote the image of u, a in $\kappa_K((t))$. Then $\overline{u} = \overline{\phi(u)}/\overline{a}^p$ implies that the t-adic valuations v_u, v_a of $\overline{u}, \overline{a}$ satisfy $v_u = pv_u - pv_a$, or $(p-1)v_u = pv_a$. Consequently, v_a is divisible by $p-1$, so we may change generators to force $v_a = v_u = 0$. We may then force the reduction of $\overline{a}$ modulo t to be equal to 1 (by twisting M), then change generators to force $\overline{a} = \overline{u} = 1$.

Write $a = \sum_i a_i t^i$ and $u = \sum_i u_i t^i$. After replacing K by a finite extension, we may change generators to force $u_0 = 1$. Put $c = |u-1|_1$, so that $|\phi(u)/u - 1|_1 = c$ also. Suppose $c > p^{-p/(p-1)}$; then $|a^p - 1|_1 = c$, so $|a-1|_1 = c^{1/p}$ and $|a^p - 1 - (a-1)^p| \leq p|a-1|_1 = pc^{1/p} < c$. In other words, modulo quantities of norm less than c, $(a-1)^p$ is congruent to $\phi(u)/u - 1$, which in turn is congruent to $\sum_i \phi(u_i)t^{qi} - \sum_i u_i t^i$.

We deduce that modulo quantities of norm less than c, u_i vanishes if i is not divisible by p, and otherwise it is congruent to a p-th power. In other words, u is congruent to a p-th power, with which we can change generators to force $|u-1|_1 < c$.

We may repeat this process until $|u-1|_1 \leq p^{-p/(p-1)}$. If $|u-1|_1 < p^{-p/(p-1)}$, then u has a p-th root and so M is already constant. Otherwise, let π satisfy $\pi^{p-1} = -p$; then one checks (using Example 17.2.10) that $M \otimes \mathcal{R}_L$ is constant for L equal to the Artin-Schreier extension

$$\kappa_K((t))[z]/\left(z^p - z - \sum_{i \not\equiv 0 \pmod{p}} \overline{n_i/i\pi}\right).$$

$\square$

Remark 18.3.7. In the case of an absolute Frobenius lift, a third approach is to simply observe (using Lemma 18.3.3) that the rank 1 case of Theorem 18.1.8 is a special case of Tsuzuki's theorem (Theorem 17.4.2.

4. The differential approach

With the rank 1 case of Theorem 18.1.8 under control, it is time to consider the general case. There are two general strategies available to prove this. The first of these, which we consider in this section, is that used by André [**And02**] and Mebkhout [**Meb02**]; it is to array as many results as possible about differential modules over $\mathcal{R}$ which are solvable at 1, and make minimal use of the Frobenius structure. One advantage of this approach is that it requires essentially no change for the case where the Frobenius lift is not absolute.

We will not be able to give a complete proof of Theorem 18.1.8 using this approach; the reason is that the one result we need which involves Frobenius (outside of the rank 1 case, as addressed in the previous section) is beyond the scope of this course.

Theorem 18.4.1 (Christol-Mebkhout). *Let M be a finite differential module over $\mathcal{R}$ admitting a Frobenius structure for some scalar-preserving Frobenius lift. Suppose that $IR(M \otimes F_\rho) = 1$ for $\rho \in (0,1)$ sufficiently close to 1 (that is, M satisfies the Robba condition near 1). Then there exists a positive integer m coprime to p such that $M \otimes \mathcal{R}[t^{1/m}]$ is unipotent.*

Proof. This follows from the p-adic Fuchs theorem for annuli, which was stated but not proved earlier (Theorem 12.6.1). □

Using this result as a black box, one may prove Theorem 18.1.8 as follows.

Lemma 18.4.2. *Let V be an indecomposable differential module over F_ρ of rank n, such that $IR(V) \in |F^\times|$ and $IR(V) < p^{-1/(p-1)}$. Then for any positive integer j coprime to p, $IR(\wedge^j V) = IR(V)$.*

Proof. By Proposition 5.2.11, we may construct a basis $e_1, \dots, e_n$ of V such that the corresponding supremum norm satisfies $|D|_V = |D|_{\mathrm{sp},V}$. Let N be the matrix via which D acts on this basis; then the eigenvalues $\lambda_1, \dots, \lambda_n$ of N all have norm $|D|_{\mathrm{sp},V}$. Let $e_1^*, \dots, e_n^*$ be the dual basis of $V^\vee$, and use $e_i^* \otimes e_j$ for $i, j = 1, \dots, n$ as a basis of $V^\vee \otimes V$; then D acts on this basis via the matrix $I_N \otimes N - N \otimes I_n$, with eigenvalues $\lambda_j - \lambda_i$ for $i, j = 1, \dots, n$.

Let V' be the differential module of rank 1 with the action of D on a generator v given by $D(v) = -\lambda_1 v$. Since V is indecomposable, so is $V' \otimes V$. However, by Theorem 5.6.1, this means that the intrinsic subsidiary radii of $V' \otimes V$ must all be equal to each other; in particular, either all of them are equal to $IR(V)$ or none of them are. By Theorem 5.7.4, the number of intrinsic subsidiary radii of $V' \otimes V$ equal to $IR(V)$ is equal to the number of $j \in \{1, \dots, n\}$ for which $|\lambda_j - \lambda_1| = |\lambda_1|$. Since $j = 1$ does not have this property, we deduce that $|\lambda_j - \lambda_1| < |\lambda_1|$ for $j = 2, \dots, n$ also.

If we now form a basis for $\wedge^j V$ using exterior products of $e_1, \dots, e_n$, then D will act on this basis via a matrix whose eigenvalues are the j-fold sums of $\lambda_1, \dots, \lambda_n$. By what we have shown (and the fact that j is coprime to p), these sums all have norm $|\lambda_1|$; by Theorem 5.7.4 once more, $|D|_{\mathrm{sp},\wedge^j V} = |D|_{\mathrm{sp},V}$ as desired. □

Lemma 18.4.3. *Assume that Theorem 18.1.8 is true for modules of rank 1. Let M be a finite differential module over $\mathcal{R}$ admitting a Frobenius structure for some scalar-preserving Frobenius lift, of rank j coprime to p. Then there exists a finite separable extension L of $\kappa_K((t))$ and a finite extension K' of K such that $M \otimes_{\mathcal{R}} \mathcal{R}_L \otimes_K K'$ is either unipotent or decomposable.*

Proof. Invoking the hypothesis that Theorem 18.1.8 is known for the rank 1 module $\wedge^j M$, we may choose L such that $(\wedge^j M) \otimes \mathcal{R}_L$ is trivial as a differential module. For notational simplicity, we assume hereafter that $L = \kappa_K((t))$.

If M itself is decomposable, we are done. Otherwise, by Theorem 11.5.4, there exists $b \geq 0$ such that for ρ sufficiently close to 1, the intrinsic subsidiary radii of $M \otimes F_\rho$ are all equal to ρ^b. If $b = 0$, then we can render M unipotent by Theorem 18.4.1, so assume $b > 0$. By Theorem 11.5.4 plus Theorem 18.4.1, for ρ sufficiently close to 1 and K' a finite totally extension of K,

$$H^0((M^\vee \otimes M) \otimes_K K') = H^0((M^\vee \otimes M \otimes F_\rho) \otimes_K K').$$

(Namely, we may replace $M^\vee \otimes M$ by its component of ramification number 0, whose structure we can describe using Theorem 18.4.1.)

Pick some ρ in the prime-to-p divisible closure of $|K^\times|$ for which:

- $\wedge^j M \otimes F_\rho$ is trivial;
- $IR(M \otimes F_\rho) = \rho^b$;
- $H^0((M^\vee \otimes M) \otimes_K K') = H^0(((M^\vee \otimes M) \otimes F_\rho) \otimes_K K')$ for any K';
- $\rho^b \neq p^{-p^{-m}/(p-1)}$ for any nonnegative integer m.

By replacing M by a suitable Frobenius antecedent, we may reduce to the case $\rho^b < p^{-1/(p-1)}$. Since ρ^b is still in the divisible closure of $|K^\times|$, we can choose K' so that $p^{-1/(p-1)}, |\rho^b| \in |(K')^\times|$. If $M \otimes_K K'$ were indecomposable (i.e., if the not necessarily commutative K'-algebra $H^0((M^\vee \otimes M) \otimes_K K')$ had no nontrivial idempotents), then so would be $(M \otimes F_\rho) \otimes_K K'$; however, Lemma 18.4.2 would then imply

$$\rho^b = IR(M \otimes F_\rho) = IR(\wedge^j (M \otimes F_\rho)) = 1,$$

contradiction. Hence $M \otimes_K K'$ is decomposable, as desired. $\square$

Remark 18.4.4. The statement of Lemma 18.4.3 is not so surprising if one thinks of linear representations of a finite p-group on an algebraically closed field, as these must have dimensions which are powers of p.

Proof of Theorem 18.1.8. As noted in Remark 18.1.9, it suffices to check that M becomes unipotent after finitely many operations of the following forms: replace $\kappa_K((t))$ by a finite separable extension, or replace K by a finite extension. We will refer to these operations simply as "replacing" in what follows.

We proceed by induction on $n = \operatorname{rank}(M)$, with the base case $n = 1$ known by the arguments of the previous section. If n is greater than 1 and prime to p, by Lemma 18.4.3, M becomes decomposable after replacing, so the induction hypothesis applies.

If n is divisible by p, we note that $M^\vee \otimes M$ decomposes as the direct sum of a trivial submodule (the *trace submodule*) and a complement (the *trace zero submodule*). The latter of these has rank $n^2 - 1$, which is not divisible by p. Hence after replacing, $M^\vee \otimes M$ acquires a subobject of rank 1, which by the induction hypothesis can be rendered trivial by replacing again.

If M is reducible, we may again invoke the induction hypothesis. Otherwise, $D = H^0(M^\vee \otimes M)$ is a division algebra of finite dimension over K. It is a standard algebra result that there exists a finite extension K' of K such that $D \otimes_K K'$ is isomorphic to a matrix algebra $M_{n \times n}(K')$. (It suffices to check this for $K' = K^{\mathrm{alg}}$ instead. The key observation is that for any element of $D \otimes_K K^{\mathrm{alg}}$, one can subtract some element of K^{alg} to get a noninvertible element; namely, pick any eigenvalue of the action of $D \otimes_K K^{\mathrm{alg}}$ on itself by left multiplication.) Consequently, after replacing, M becomes isomorphic to a direct sum of rank 1 modules, so again the induction hypothesis applies. $\square$

5. The difference approach: absolute case

The second strategy available for proving Theorem 18.1.8 is that used in the absolute case by Kedlaya [**Ked04a**]. It is to first analyze the structure of difference modules over the Robba ring closely, then only later make reference to differential modules.

We will not be able to give a complete proof of Theorem 18.1.8 using this approach either; the reason is that the main result of [**Ked04a**] (Theorem 18.5.1) below is beyond the scope of this course, aside from the discussion in Remark 18.5.2.

Theorem 18.5.1 (Slope filtration theorem). *Let M be a finite free difference module over $\mathcal{R}$. Then there exists a unique filtration $0 = M_0 \subset \cdots \subset M_l = M$ by difference submodules with the following properties.*

(a) *Each successive quotient M_i/M_{i-1} is finite free, and is the base extension of a difference module over $\mathcal{E}^\dagger$ which is pure of some norm s_i.*
(b) *We have $s_1 > \cdots > s_l$.*

Remark 18.5.2. Theorem 18.5.1 was first proved by Kedlaya in [**Ked04a**] for an absolute Frobenius lift; a second, somewhat simpler proof (establishing a stronger result which we will not describe) appears in [**Ked05b**]. The form given above is from [**Ked07c**]; although the proof there is even more streamlined than the previous ones, it is still too involved to discuss in detail here. Instead, we give a brief summary of the two stages of the proof.

(a) Construct a suitable "difference closure" $\tilde{\mathcal{R}}$ of $\mathcal{R}$, over which one has an analogue of the Dieudonné-Manin classification (Corollary 13.6.4). To obtain the classification, one first shows that every difference module over $\tilde{\mathcal{R}}$ admits a filtration with pure quotients, then rearranges this filtration until it is no longer possible, at which point everything splits. (This is very loosely analogous to Grothendieck's classification of vector bundles on the projective line; indeed, much of the theory of slope filtrations is analogous to the theory of vector bundles on smooth projective curves.)
(b) Use faithfully flat descent to show that given a difference module over $\mathcal{R}$, the slope filtration obtained over $\tilde{\mathcal{R}}$ by applying (a) descends to $\mathcal{R}$. One must also show that the property of being pure of a given norm also descends. (In earlier proofs, a more complicated Galois descent was used instead.)

In the case of an absolute Frobenius lift, one can use Theorem 18.5.1 to reduce Theorem 18.1.8 to Theorem 17.4.2. However, one must make the following two verifications.

Lemma 18.5.3. *Let M be a finite differential module over $\mathcal{R}$ equipped with a Frobenius structure. Then the steps of the filtration of Theorem 18.5.1 are differential modules, not just difference modules.*

Proof. It suffices to check for M_1, as then we may quotient by M_1 and repeat the argument. Note that the composition $M_1 \xrightarrow{D} M \to M/M_1$ is $\mathcal{R}$-linear, since $D(rv) = rD(v) + d(r)v$ and the second term gets killed in the quotient. Hence it suffices to check that $M_1 \to M/M_1$ is the zero map.

This in turn follows from the following fact: if N_1, N_2 are finite free difference modules over $\mathcal{R}$ which are pure of slopes s_1, s_2 with $s_1 > s_2$, then $\mathrm{Hom}(N_1, N_2) = 0$. (Namely, this implies that the map $M_1 \to M/M_{l-1}$ vanishes, so we get a map $M_1 \to M_{l-1}$; but the induced map $M_1 \to M_{l-1}/M_{l-2}$ vanishes, et cetera.) By replacing N_1, N_2 with $\mathcal{R}, N_1^\vee \otimes N_2$, we may rewrite the claim as follows: if N is a finite free difference module over $\mathcal{R}$ pure of some slope $s < 1$, then $H^0(N) = 0$.

By Proposition 13.5.8, we may choose a basis for N on which Φ acts via a matrix A over $\mathcal{E}^\dagger$ of norm at most 1. Let N_0 be the $\mathcal{E}^\dagger$-span of this basis; then $H^0(N) = H^0(N_0)$ by

Lemma 18.5.6 below. However, $H^0(N_0) = 0$ because we are now working over a difference field, where there cannot exist any maps between pure modules of different norms. $\square$

This lemma generalizes Lemma 18.3.3 to arbitrary rank.

Lemma 18.5.4. *Let M be a finite free unit-root difference module over $\mathcal{E}^\dagger$ such that $M \otimes \mathcal{R}$ admits a compatible differential structure. Then this structure is induced by a corresponding differential structure on M itself.*

Proof. Let N, A be the matrices via which D, ϕ act on a basis of M. Write the commutation relation between N, A in the form $N - pt^{p-1}A\phi(N)A^{-1} = \frac{d}{dt}(A)A^{-1}$. We deduce from Lemma 18.5.6 below that N has entries in $\mathcal{E}^\dagger$. $\square$

Remark 18.5.5. The proof of Lemma 18.5.3 uses the fact that if M is a finite free difference module over $\mathcal{R}$ which is pure of norm $s < 1$, then $H^0(M) = 0$. The same is not in general true if $s > 1$, illustrating another way that $\mathcal{R}$ fails to behave like a difference field. For instance, if ϕ is the Frobenius lift for which $\phi(1+t) = (1+t)^p$, and $M = \mathcal{R}v$ with $\Phi(v) = p^{-1}v$, then

$$\log(1+t)v \in H^0(M);$$

this example is in fact critical for p-adic Hodge theory.

Lemma 18.5.6. *Let A be an $n \times n$ matrix over $\mathfrak{o}_{\mathcal{E}^\dagger}$, and suppose $v \in \mathcal{R}^n$, $w \in (\mathcal{E}^\dagger)^n$ satisfy $v - A\phi(v) = w$. Then $v \in (\mathcal{E}^\dagger)^n$.*

Proof. Exercise, or see [**Ked07c**, Proposition 1.2.6]. $\square$

6. The difference approach: general case

Unfortunately, it seems difficult to generalize the proof of Theorem 17.4.2 to the case of a nonabsolute Frobenius lift (the main difficulty being Lemma 17.4.13). However, it is nonetheless possible to give a proof of Theorem 18.1.8 that uses slope filtrations (via Theorem 18.5.1) in lieu of p-adic exponents (via Theorem 12.6.1). That is because one can use Theorem 18.5.1 to give a second proof of Theorem 18.4.1, and then proceed as in the differential approach.

To be more specific, using Theorem 18.5.1 and Lemma 18.5.3, it suffices to check Theorem 18.4.1 in the case of a unit-root Frobenius structure. This amounts to the following lemma.

Lemma 18.6.1. *Let M be a differential module over $\mathcal{R}$ with $IR(M \otimes F_\rho) = 1$ for ρ sufficiently close to 1, equipped with a unit-root Frobenius structure. Then for some positive integer m coprime to p, $M \otimes \mathcal{R}[t^{1/m}]$ is constant.*

Proof. By Lemma 18.5.4, we can write $M \cong M_0 \otimes \mathcal{R}$ for some finite differential module M_0 over $\mathcal{E}^\dagger$ equipped with a unit-root Frobenius. By Lemma 17.4.5, we may replace the Frobenius lift with one satisfying $\phi(t) = t^q$. By Lemma 17.4.11, for some positive integer m coprime to p, we can choose a basis $e_1, \dots, e_n$ of $M \otimes \mathcal{R}[t^{1/m}]$ on which Φ acts via a matrix A with $|A - A_0|_1 < 1$. For notational simplicity, we will assume $m = 1$. By Lemma 17.4.8, we can rechoose this basis so that A has entries in $\mathcal{E}^\dagger \cap K[\![t^{-1}]\!]$.

Let N be the matrix of action of tD on $e_1, \dots, e_n$. As in the proof of Lemma 17.4.9, use A and N to define a differential module M' over $K\langle \alpha/t \rangle$ for some $\alpha \in (0, 1)$ such that

$IR(M \otimes F_\alpha) = 1$. Then Theorem 8.5.1 implies that M' has a basis of horizontal sections on the open disc of radius α^{-1} in the t^{-1}-line, so M is constant. $\square$

REMARK 18.6.2. It would be interesting to know whether one can prove Theorem 18.4.1 without using either p-adic exponents or slope filtrations, but instead simply using the fact that the hypothesis forces M to extend across the entire punctured open unit disc (by pasting together Frobenius antecedents).

7. Applications of the monodromy theorem

The original area of application of the p-adic local monodromy theorem (in its original form, with only an absolute Frobenius lift) was in the subject of rigid cohomology; the name comes from the fact that it plays a role analogous to the ℓ-adic local monodromy theorem of Grothendieck in the subject of étale cohomology. See Chapter 20 for further discussion. In particular, Crew [**Cre98**] showed that Theorem 18.1.8 implies the finite dimensionality of the rigid cohomology of a curve with coefficients in an overconvergent F-isocrystal; this was later generalized to arbitrary varieties by Kedlaya [**Ked06a**].

Another area of application of the p-adic local monodromy theorem (in both the absolute and nonabsolute forms) is in p-adic Hodge theory. See Chapter 21 for further discussion.

Theorem 18.1.8 is also needed for the proofs of some more mundane facts about differential modules; it can often be used in lieu of Dwork's trick (Corollary 15.2.4) when working over an annulus instead of a disc. Here is a typical example; see notes for further discussion.

THEOREM 18.7.1. *Let M be a finite differential module over $R = K[\![t]\!]_0$ or $\mathcal{E}^\dagger$ admitting a Frobenius structure for an absolute Frobenius lift. Then*

$$H^0(M) = H^0(M \otimes_R \mathcal{E}).$$

PROOF. For the case $R = \mathcal{E}^\dagger$, it is shown in [**Ked04b**] that any F-invariant horizontal section of $M \otimes_R \mathcal{E}$ belongs to M. Here is a quick sketch of the argument. One first uses a technique of de Jong [**dJ98a**] to show that if $v \in H^0(M \otimes_R \mathcal{E})$, then the induced F-equivariant horizontal map $\psi : M^\vee \to \mathcal{E}$ has the property that $\psi^{-1}(\mathcal{E}^\dagger) \neq 0$, and that the generic slopes of $M^\vee/\psi^{-1}(\mathcal{E}^\dagger)$ has all slopes negative. (This argument uses only the Frobenius structure, not the differential structure.) One then uses Theorem 18.1.8 (plus some additional considerations) to show that the short exact sequence

$$0 \to \ker(\psi) \to \psi^{-1}(\mathcal{E}^\dagger) \to \psi^{-1}(\mathcal{E}^\dagger)/\ker(\psi) \to 0$$

must split. This yields $\psi(M) = \mathcal{E}^\dagger$, forcing $v \in H^0(M)$.

To check the claim at hand in the case $R = \mathcal{E}^\dagger$, we may enlarge K to have algebraically closed residue field; then Corollary 13.6.4 implies that $H^0(M \otimes_R \mathcal{E})$ is spanned by one-dimensional fixed subspaces for the Frobenius action. The previous argument shows that any generator of one of these subspaces belongs to M, proving the claim.

For the case $R = K[\![t]\!]_0$, we may use the previous argument to reduce to checking that $H^0(M) = H^0(M \otimes_R \mathcal{E}^\dagger)$. Since $K[\![t]\!]_0 = K\{\{t\}\} \cap \mathcal{E}^\dagger$ inside $\mathcal{R}$, this is equivalent to checking that $H^0(M \otimes_R K\{\{t\}\}) = H^0(M \otimes_R \mathcal{R})$. However, this is evident because $M \otimes_R K\{\{t\}\}$ is a trivial differential module by Dwork's trick (Corollary 15.2.4). $\square$

Notes

Proposition 18.1.4 is the essential content of a paper of Lazard [**Laz62**]. Note that it depends on K being spherically complete, and is false otherwise; however, we have assumed in this part that K is discretely valued, so we are safe.

The p-adic local monodromy theorem (Theorem 18.1.8) was originally formulated and proved only in the absolute case. In this case, it is often referred to in the literature as "Crew's conjecture", because it emerged from the work of Crew [**Cre98**] on finite dimensionality of rigid cohomology with coefficients in an overconvergent F-isocrystal. Crew's original conjecture was somewhat more limited still, as it only concerned modules such that the differential and Frobenius structures were both defined over $\mathcal{E}^\dagger$; this form was restated in a more geometric form by de Jong [**dJ98b**]. A closer analysis of Crew's conjecture was then given by Tsuzuki [**Tsu98c**], who explained (using Theorem 17.4.2) how Theorem 18.1.8 in the absolute case would follow from a slope filtration theorem [**Tsu98c**, Theorem 5.2.1].

The relevance of allowing a nonabsolute Frobenius lift in Theorem 18.1.8 is that this situation occurs in the context of relative p-adic Hodge theory. Namely, Berger and Colmez [**BC07**] use the full strength of Theorem 18.1.8 to prove an analogue of Fontaine's conjecture on potential semistability of de Rham representations (Corollary 21.4.5) for a family of de Rham representations parametrized by an affinoid base space.

The analysis made by Mebkhout in the rank 1 case is actually somewhat more general than Lemma 18.3.1; it applies to any irreducible module whose highest ramification number is an integer. In this case, one still shows that the ramification number can be reduced by twisting with a suitable quasiconstant module; however, the new ramification number is not itself guaranteed to be an integer. Consequently, Mebkhout must account for this possibility elsewhere in the proof of Theorem 18.1.8.

The analysis made by Crew in the rank 1 case does not include Lemma 18.3.3; that is because Crew only considered cases where the conclusion of this lemma was already known. It also does not include Lemma 18.3.6; instead, Crew (who only considers the case of an absolute Frobenius lift) makes an argument similar to that used in the proof of Theorem 17.4.2.

The differential approach to Theorem 18.1.8 presented here is close to that given by Mebkhout [**Meb02**]. It is less apparently similar to that of André [**And02**], who uses an approach couched in the language of Tannakian categories.

Some readers may be disappointed that we did not include a fuller treatment of the slope filtration theorem (Theorem 18.5.1). Our defenses are that the material used to prove the theorem is somewhat orthogonal to the other material in the book, and that we have little to add to the treatment in [**Ked07c**].

The case of Theorem 18.7.1 with $R = \mathcal{E}^\dagger$ was originally conjectured by Tsuzuki [**Tsu02**, Conjecture 2.3.3]. The case with $R = K[\![t]\!]_0$ is an older result of de Jong [**dJ98a**]; the arguments in [**Ked04b**] are closely modeled on those of [**dJ98a**], with the key addition being the substitution of Theorem 18.1.8 for Dwork's trick.

In the case of a unit-root Frobenius structure, Theorem 18.7.1 was known prior to the availability of Theorem 18.1.8. It figures in the work of Cherbonnier-Colmez [**CC98**], which we will discuss in Chapter 21 (see Remark 21.2.6); it was also established by Tsuzuki [**Tsu96**, Proposition 4.1.1].

Exercises

(1) Show that if M is unipotent, then there exists a basis of M on which the derivation corresponding to $t\frac{d}{dt}$ acts via a matrix over K. Then prove Proposition 18.1.7 in case M is unipotent.

(2) Prove Lemma 18.5.6. (Hint: reduce to the case where $|A|_\rho \leq 1$ for $\rho \in [\alpha, 1)$. Then show that $|v|_\rho$ is bounded for $\rho \in [\alpha, 1)$, by comparing $|v|_\rho$ with $|v|_{\rho^{1/q}}$ using Lemma 15.2.1.)

Part 5

Areas of application

CHAPTER 19

Picard-Fuchs modules

In this chapter, we revisit the territory of Chapter 0, briefly discussing how Picard-Fuchs modules give rise to differential equations with Frobenius structures, and what this has to do with zeta functions.

1. Picard-Fuchs modules

Definition 19.1.1. Let t be a coordinate on $\mathbb{P}^1_K$. Let $f : X \to \mathbb{P}^1_K$ be a proper, flat, generically smooth morphism of algebraic varieties. Let $S \subset \mathbb{P}^1_K$ be a zero-dimensional subscheme containing ∞ (for convenience) and all points over which f is not smooth. The *Picard-Fuchs modules* on $\mathbb{P}^1_K \setminus S$ associated to f are finite locally free differential modules M_i for $i = 0, \dots, 2\dim(f)$ over $R = \Gamma(\mathbb{P}^1_K \setminus S, \mathcal{O})$ with respect to the derivation $\frac{d}{dt}$; it also has regular singularities at each point of S. For $\lambda \notin S$, the fibre of M_i at λ can be canonically identified with the i-th de Rham cohomology of the fibre $f^{-1}(\lambda)$.

Although the classical construction of the Picard-Fuchs module is analytic (it involves viewing f as an analytically locally trivial fibration and integrating differentials against moving homology classes), there is an algebraic construction due to Katz and Oda [**KO68**], involving a Leray spectral sequence for the algebraic de Rham cohomology of the total space.

As originally noticed by Dwork by explicitly calculating some examples, Picard-Fuchs modules often carry Frobenius structures. A systematic explanation of this is given by p-adic cohomology; here is an explicit statement.

Theorem 19.1.2. *With notation as above, suppose that f extends to a proper morphism $\mathfrak{X} \to \mathbb{P}^1_{\mathfrak{o}_K}$ such that the intersection of $\mathbb{P}^1_k$ with the nonsmooth locus is contained in the intersection of $\mathbb{P}^1_k$ with the Zariski closure of S (i.e., the morphism is smooth over all points of $\mathbb{P}^1_k$ which are not the reductions of points in S). Let M_i be the i-th Picard-Fuchs module for f, and let $\phi : \mathbb{P}^1_{\mathfrak{o}_K} \to \mathbb{P}^1_{\mathfrak{o}_K}$ be a Frobenius lift (e.g., $t \mapsto t^p$) that acts on $\mathfrak{o}_K$ as a lift of the absolute Frobenius. Then for some $\alpha \in (0,1)$, there exists an isomorphism $\phi^*(M_i) \cong M_i$ over a ring R which is the Fréchet completion of $\Gamma(\mathbb{P}^1_K \setminus S, \mathcal{O})$ for (for $\rho \in [\alpha, 1)$) the ρ^{-1}-Gauss norm and the Gauss norms $|t - \lambda| = \rho$ for $\lambda \in S$.*

Remark 19.1.3. Geometrically, the Frobenius structure is defined on the complement in $\mathbb{P}^1_K$ of a union of discs around the points of S, each of radius less than 1 (where a disc of radius less than 1 around ∞ corresponds to the complement of a disc of radius greater than 1 around 0). In particular, by working in a unit disc not containing any points of S, we obtain a differential module with Frobenius structure over $K[\![t]\!]_0$. In a unit disc containing one or more points of S, we only obtain a differential module with Frobenius structure over $\cup_{\alpha>0} K\langle \alpha/t, t]\!]_0$. (If the disc contains exactly one point of S and the exponents at that point are all 0, we can also get a differential module with Frobenius structure over $K[\![t]\!]_0$ for the derivation $t\frac{d}{dt}$, provided that ϕ fixes that point.)

EXAMPLE 19.1.4. For example, for the Legendre family of elliptic curves $y^2 = x(x - 1)(x - \lambda)$, we take $S = \{0, 1, \infty\}$ and obtain a module corresponding to the hypergeometric equation discussed in the introduction. For $p \neq 2$, that equation admits a Frobenius structure by Theorem 19.1.2. (For $p = 2$, we cannot make the reduction modulo p generically smooth without changing the defining equation.)

2. Relationship with zeta functions

The Frobenius structure on a Picard-Fuchs equation can be used to compute zeta functions. (The condition on λ allows for a unique choice in each residue disc, by Lemma 14.2.2.)

THEOREM 19.2.1. *Retain notation as in Theorem 19.1.2, and assume now that $\kappa_K = \mathbb{F}_q$ with $q = p^a$, and that ϕ is a q-power Frobenius lift on $\mathbb{P}^1_{\mathfrak{o}_K}$. Suppose that $\lambda \in \mathfrak{o}_K$ satisfies $\phi(t-\lambda) \equiv 0 \pmod{t-\lambda}$, and suppose that f extends smoothly over the residue disc containing λ. Then*

$$\zeta(f^{-1}(\overline{\lambda}), T) = \prod_{i=0}^{2\dim(f)} \det(1 - T\Phi, (M_i)_\lambda)^{(-1)^{i+1}}.$$

This suggests an interesting strategy for computing zeta functions, advanced by Alan Lauder.

REMARK 19.2.2. Suppose you have in hand the differential module, plus the matrix of the action of Φ on some individual $(M_i)_\lambda$. If you view the equation

$$NA + \frac{dA}{dt} = \frac{d\phi(t)}{dt} A\phi(N)$$

as a differential equation with initial condition provided by $(M_i)_\lambda$, you can then solve for A, and then evaluate at another λ.

More explicitly, let's say for simplicity that $\lambda = 0$ is the starting value. In the open unit disc around 0, you can compute U such that

$$U^{-1}NU + U^{-1}\frac{dU}{dt} = 0$$

and then write down

$$A = UA_0\phi(U^{-1}).$$

This only gives you a power series representation around 0 with radius of convergence 1, which does not give you any way to specialize to, say, $\lambda = 1$.

However, Theorem 19.1.2 implies that the entries of A can be written as uniform limits of rational functions with limited denominators. Once you recover a sufficiently good rational function approximation to A, you can specialize at $\lambda = 1$. For more detailed references discussing this technique, see the notes.

REMARK 19.2.3. One can recover the example of Dwork discussed in Chapter 0 from Theorem 19.2.1. What is going on in that example is that one is separating the Picard-Fuchs module, which has rank 2, into a unit-root component and a component of slope $\log p$. For this to be possible, one must be in the situation of Theorem 14.3.4; this fails precisely at the residue discs at which the Igusa polynomial vanishes, which is why one must invert the Igusa polynomial in the course of the computation.

Notes

The differential operator on a Picard-Fuchs module is also called a *Gauss-Manin connection*. Lauder's strategy (also called the *deformation method*) was introduced in [**Lau04**]; it has been worked out in detail for hyperelliptic curves by Hubrechts [**Hub07**]. (Hubrechts implemented the resulting algorithm in version 2.14 of the computer algebra system `Magma`.) A version for hypersurfaces has been described by Gerkmann [**Ger07**].

CHAPTER 20

Rigid cohomology

In this chapter, we introduce a bit of the theory of rigid p-adic cohomology, as developed by Berthelot and others. In particular, we illustrate the role played by the p-adic local monodromy theorem in a fundamental finiteness problem in the theory.

1. Isocrystals on the affine line

In this section, we recall Crew's interpretation [**Cre98**] of overconvergent F-isocrystals on the affine line and their cohomology.

DEFINITION 20.1.1. Let k be a perfect (for simplicity) field of characteristic $p > 0$. Let K be a complete discrete (again for simplicity) nonarchimedean field of characteristic zero with $\kappa_K = k$. An *overconvergent F-isocrystal* on the affine line over k (with coefficients in K) is a finite differential module with Frobenius structure on the ring $\mathcal{A} = \cup_{\beta>1} K\langle t/\beta\rangle$, for some absolute Frobenius lift ϕ; as in Proposition 15.3.1, the resulting category is independent of the choice of the Frobenius lift.

DEFINITION 20.1.2. Let M be an overconvergent F-isocrystal on the affine line over k. Let $\mathcal{R}$ be a copy of the Robba ring with series parameter t^{-1}, so that we can identify $\mathcal{A}$ as a subring of $\mathcal{R}$. Define

$$\begin{aligned}
H^0(\mathbb{A}^1_k, M) &= \ker(D, M)\\
H^1(\mathbb{A}^1_k, M) &= \operatorname{coker}(D, M)\\
H^0_{\mathrm{loc}}(\mathbb{A}^1_k, M) &= \ker(D, M \otimes_{\mathcal{A}} \mathcal{R})\\
H^1_{\mathrm{loc}}(\mathbb{A}^1_k, M) &= \operatorname{coker}(D, M \otimes_{\mathcal{A}} \mathcal{R})\\
H^1_c(\mathbb{A}^1_K, M) &= \ker(D, M \otimes_{\mathcal{A}} (\mathcal{R}/\mathcal{A}))\\
H^2_c(\mathbb{A}^1_K, M) &= \operatorname{coker}(D, M \otimes_{\mathcal{A}} (\mathcal{R}/\mathcal{A})).
\end{aligned}$$

By taking kernels and cokernels in the short exact sequence

$$0 \to M \to M \otimes_{\mathcal{A}} \mathcal{R} \to M \otimes_{\mathcal{A}} (\mathcal{R}/\mathcal{A}) \to 0$$

and applying the snake lemma, we get an exact sequence

$$0 \to H^0(\mathbb{A}^1_k, M) \to H^0_{\mathrm{loc}}(\mathbb{A}^1_k, M) \to H^1_c(\mathbb{A}^1_k, M) \to H^1(\mathbb{A}^1_k, M) \to H^1_{\mathrm{loc}}(\mathbb{A}^1_k, M) \to H^2_c(\mathbb{A}^1_k, M) \to 0.$$

REMARK 20.1.3. Crew shows [**Cre98**] that in this construction, H^i computes the *rigid cohomology* of M, H^i_c computes the *rigid cohomology with compact supports*, and H^i_{loc} computes some sort of local cohomology at ∞.

Crew's main result in this setting is the following.

Theorem 20.1.4 (Crew). *The spaces $H^i(\mathbb{A}^1_k, M)$, $H^i_c(\mathbb{A}^1_k, M)$, $H^i_{\mathrm{loc}}(\mathbb{A}^1_K, M)$ are all finite dimensional over K. Moreover, the Poincaré pairings*

$$H^i(\mathbb{A}^1_k, M) \times H^{2-i}_c(\mathbb{A}^1_k, M^\vee) \to H^2_c(\mathbb{A}^1_k, \mathcal{A}) \cong K$$
$$H^i_{\mathrm{loc}}(\mathbb{A}^1_k, M) \times H^{1-i}_{\mathrm{loc}}(\mathbb{A}^1_k, M^\vee) \to H^1_{\mathrm{loc}}(\mathbb{A}^1_K, \mathcal{A}) \cong K$$

are perfect.

The key ingredient is the fact that $M \otimes \mathcal{R}$ is quasiunipotent by the p-adic local monodromy theorem (Theorem 18.1.8), which implies finiteness of $H^i_{\mathrm{loc}}(\mathbb{A}^1_k, M)$. This implies the finite dimensionalities except for $H^1_c(\mathbb{A}^1_k, M)$ and $H^1(\mathbb{A}^1_k, M)$; however, these are related by a map with finite dimensional kernel and cokernel. Moreover, they carry incompatible topologies: the former is a Fréchet space, while the latter is dual to a Fréchet space. This incompatibility can only be resolved by both spaces being finite dimensional.

2. Consequences in rigid cohomology

The previous construction extends, with some work, to a theory of rigid cohomology with/without compact supports on arbitrary varieties over k, with coefficients in overconvergent F-isocrystals. For constant coefficients, it was shown by Berthelot [**Brt97a, Brt97b**] that this theory has all of the desired properties of a Weil cohomology: finite dimensionality, Poincaré duality, Künneth formula, cycle class maps, etc. Using a relative version of Theorem 20.1.4, one can extend these to nonconstant coefficients [**Ked06a**].

The analogy with étale cohomology with ℓ-adic coefficients is tempting, and indeed motivates most of the preceding development, but remains somewhat imperfect. Most notably, overconvergent F-isocrystals in rigid cohomology are analogous only to lisse (smooth) ℓ-adic sheaves, whereas for most serious computations one needs also constructible sheaves (or some appropriate derived category thereof). There is a proposed theory of arithmetic $\mathcal{D}$-modules that would play the appropriate p-adic role, but this theory remains underdeveloped; see [**Brt02**].

Nonetheless, in the interim, one can still carry many good properties of ℓ-adic cohomology to the p-adic setting, e.g., Laumon's Fourier-theoretic reinterpretation of Deligne's second proof of the Weil conjectures [**Ked06b**]. It is hoped that one can go further, establishing some properties in p-adic cohomology that are only conjectural in ℓ-adic cohomology, such as Deligne's weight-monodromy conjecture.

3. Machine computations

In recent years, interest has emerged in explicitly computing the zeta functions of algebraic varieties defined over finite fields. Some of this interest has come from cryptography, particularly the use of Jacobians of elliptic (and later hyperelliptic) curves over finite fields as "black box abelian groups" for certain public-key cryptography schemes (Diffie-Hellman, ElGamal).

For elliptic curves, a good method for doing this was proposed by Schoof [**Sch85**]. It amounts to computing the trace of Frobenius on the ℓ-torsion points, otherwise known as the étale cohomology with $\mathbb{F}_\ell$-coefficients, for enough small values of ℓ to determine uniquely the one unknown coefficient of the zeta function within the range prescribed by the Hasse-Weil bound.

It turns out to be somewhat more difficult to execute Schoof's scheme for curves of higher genus, as discovered by Pila [**Pil90**]. One is forced to work with higher division polynomials in order to compute torsion of the Jacobian of the curve; the interpretation in terms of étale cohomology is of little value because the definition of étale cohomology is not intrinsically computable. (It is easy to write down cohomology classes, but it is difficult to test two such classes for equality.)

It was noticed by several authors that rigid cohomology is intrinsically more computable, and so lends itself better to this sort of task. Specifically, Kedlaya [**Ked01**] proposed an algorithm using rigid cohomology (in its guise for smooth affine varieties, known as Monsky-Washnitzer cohomology) for computing the zeta function of a hyperelliptic curve over a finite field of small odd characteristic. The limitation to odd characteristic was lifted by Denef and Vercauteren [**DV06**]; the limitation to small characteristic was somewhat remedied by Harvey [**Har07**], who improved the dependence on the characteristic p from $O(p)$ to $O(p^{1/2+\epsilon})$.

More recently, interest has emerged in considering also higher-dimensional varieties, partly come from potential applications in the study of mirror symmetry for Calabi-Yau varieties. In this case, étale cohomology is of no help at all, since there is no geometric interpretation of H^i_{et} for $i > 1$ analogous to the interpretation for $i = 1$ in terms of the Jacobian. Rigid cohomology should still be computable, but relatively little progress has been made in making these computations practical (one exception being the treatment of smooth surfaces in projective 3-space in [**AKR07**]). It may be necessary to combine these techniques with Lauder's deformation method (see Remark 19.2.2) for best results.

Notes

Until recently, while there were some useful survey articles about rigid cohomology (e.g., Berthelot's [**Brt86**]), and some fragmentary foundational materials (e.g., Berthelot's [**Brt96**]), there was no comprehensive introductory text on the subject. That state of affairs has been remedied by the appearance of the book of le Stum [**leS07**]; this book may be particularly helpful for those interested in machine calculations.

Crew's work, and subsequent work which builds on it (e.g., [**Ked06a**]), makes essential use of nonarchimedean functional analysis, as is evident in the discussion of Theorem 20.1.4. We recommend Schneider's book [**Sch02**] as a friendly introduction to this topic.

As a companion to our original paper on hyperelliptic curves [**Ked01**], we recommend Edixhoven's course notes [**Edi06**]; some discussion is also included in [**FvdP04**, Chapter 7]. We gave a high-level summary of the general approach in [**Ked04c**].

CHAPTER 21

p-adic Hodge theory

In this chapter, we describe an analogue of the construction of Chapter 17 for p-adic representations of the absolute Galois group of a mixed characteristic local field. Beware that our presentation is historically inaccurate; see the notes.

HYPOTHESIS 21.0.1. Throughout this chapter, let K be a finite extension of $\mathbb{Q}_p$, let V be a finite dimensional $\mathbb{Q}_p$-vector space, and let $\tau : G_K \to \mathrm{GL}(V)$ be a continuous homomorphism for the p-adic topology on V.

1. A few rings

DEFINITION 21.1.1. Put $K_n = K(\zeta_{p^n})$ and $K_\infty = \cup_n K_n$. Let $F = \operatorname{Frac} W(\kappa_K)$ and F' be the maximal subfields of K and K_∞, respectively, which are unramified over $\mathbb{Q}_p$. Put $H_K = G_{K_\infty}$ and $\Gamma_K = G_{K_\infty/K} = G_K/H_K$.

DEFINITION 21.1.2. Put $\mathfrak{o} = \mathfrak{o}_{\mathbb{C}_p}$. Let $\tilde{\mathbf{E}}^+$ be the inverse limit of the system

$$\cdots \to \mathfrak{o}/p\mathfrak{o} \to \mathfrak{o}/p\mathfrak{o}$$

in which each map is the p-power Frobenius (which is a ring homomorphism). More explicitly, the elements of $\tilde{\mathbf{E}}^+$ are sequences $(x_0, x_1, \dots)$ of elements of $\mathfrak{o}/p\mathfrak{o}$ for which $x_{n+1}^p = x_n$ for all n. In particular, for any nonzero $x \in \tilde{\mathbf{E}}^+$, the quantity $p^n v_p(x_n)$ is the same for all n for which $x_n \neq 0$; we call this quantity $v(x)$, and put conventionally $v(0) = +\infty$. Choose $\epsilon = (\epsilon_0, \epsilon_1, \dots) \in \tilde{\mathbf{E}}^+$ with $\epsilon_0 = 1$ and $\epsilon_1 \neq 1$.

The following observations were made by Fontaine and Wintenberger [**FW79**].

PROPOSITION 21.1.3. *The following are true.*

(a) *The ring $\tilde{\mathbf{E}}^+$ is a domain in which $p = 0$, with fraction field $\tilde{\mathbf{E}} = \tilde{\mathbf{E}}^+[\epsilon^{-1}]$.*
(b) *The function $v : \tilde{\mathbf{E}}^+ \to [0, +\infty]$ extends to a valuation on $\tilde{\mathbf{E}}$, under which $\tilde{\mathbf{E}}$ is complete and $\mathfrak{o}_{\tilde{\mathbf{E}}} = \tilde{\mathbf{E}}^+$.*
(c) *The field $\tilde{\mathbf{E}}$ is the algebraic closure of $\kappa_K((\epsilon - 1))$. (The embedding of $\kappa_K((\epsilon - 1))$ into $\tilde{\mathbf{E}}$ exists because $v(\epsilon - 1) = p/(p-1) > 0$.)*

DEFINITION 21.1.4. Let $\tilde{\mathbf{A}}$ be the ring of Witt vectors of $\tilde{\mathbf{E}}$, i.e., the unique complete discrete valuation ring with maximal ideal p and residue field $\tilde{\mathbf{E}}$. The uniqueness follows from the fact that $\tilde{\mathbf{E}}$ is algebraically closed, hence perfect. In particular, the p-power Frobenius on $\tilde{\mathbf{E}}$ lifts to an automorphism ϕ of $\tilde{\mathbf{A}}$.

DEFINITION 21.1.5. Each element of $\tilde{\mathbf{A}}$ can be uniquely written as a sum $\sum_{n=0}^\infty p^n [x_n]$, where $x_n \in \tilde{\mathbf{E}}$ and $[x_n]$ denotes the *Teichmüller lift* of x_n (the unique lift of x_n that has a p^m-th root in $\tilde{\mathbf{A}}$ for all positive integers m); note that $\phi([x]) = [x^p] = [x]^p$. We may thus equip $\tilde{\mathbf{A}}$ with a *weak topology*, in which a sequence $x_m = \sum_{n=0}^{\infty} p^n [x_{m,n}]$ converges to zero if for each n, $v(x_{m,n}) \to \infty$ as $m \to \infty$. Let $\mathbf{A}_{\mathbb{Q}_p}$ be the completion of $\mathbb{Z}_p[([\epsilon]-1)^{\pm}]$ in $\tilde{\mathbf{A}}$ for the weak topology; as a topological ring, it is isomorphic to the ring $\mathfrak{o}_{\mathcal{E}}$ defined over the base field $\mathbb{Q}_p$ with its own weak topology. It is also ϕ-stable because $\phi([\epsilon]) = [\epsilon]^p$.

Definition 21.1.6. Let $\mathbf{A}$ be the completion of the maximal unramified extension of $\mathbf{A}_{\mathbb{Q}_p}$, viewed as a subring of $\tilde{\mathbf{A}}$. Put

$$\mathbf{A}_K = \mathbf{A}^{H_K},$$

where the right side makes sense because we have made all the rings so far in a functorial fashion, so that they indeed carry a G_K-action. Note that $\mathbf{A}_K$ can be written as a ring of the form $\mathfrak{o}_{\mathcal{E}}$, but with coefficients in K' rather than in $\mathbb{Q}_p$.

Definition 21.1.7. For any ring denoted with a boldface A so far, define the corresponding ring with $\mathbf{A}$ replaced by $\mathbf{B}$ by tensoring over $\mathbb{Z}_p$ with $\mathbb{Q}_p$. For instance, $\tilde{\mathbf{B}} = \tilde{\mathbf{A}} \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is the fraction field of $\tilde{\mathbf{A}}$.

2. (ϕ, Γ)-modules

We are now ready to describe the mechanism, introduced by Fontaine, for converting Galois representations into modules over various rings equipped with much simpler group actions.

Definition 21.2.1. Recall that V is a finite-dimensional vector space equipped with a continuous G_K-action. Put

$$D(V) = (V \otimes_{\mathbb{Q}_p} \mathbf{B})^{H_K};$$

by Hilbert's Theorem 90, $D(V)$ is a finite dimensional $\mathbf{B}_K$-vector space, and the natural map $D(V) \otimes_{\mathbf{B}_K} \mathbf{B} \to V \otimes_{\mathbb{Q}_p} \mathbf{B}$ is an isomorphism. Since we only took H_K-invariants, $D(V)$ retains a semilinear action of $G_K/H_K = \Gamma_K$; it also inherits an action of ϕ from $\mathbf{B}$. That is, $D(V)$ is a *(ϕ, Γ)-module* over $\mathbf{B}_K$, i.e., a finite free $\mathbf{B}_K$-module equipped with semilinear ϕ and Γ_K-actions which commute with each other. It is also *étale*, which is to say the ϕ-action is étale (unit-root); as in Definition 17.2.5, this is because one can find a G_K-invariant lattice in V.

Theorem 21.2.2 (Fontaine). *The functor D, from the category of continuous representations of G_K on finite dimensional $\mathbb{Q}_p$-vector spaces to the category of étale (ϕ, Γ)-modules over $\mathbf{B}_K$, is an equivalence of categories.*

Proof. From $D(V)$, one can recover

$$V = (D(V) \otimes_{\mathbf{B}_K} \mathbf{B})^{\phi=1}.$$

□

Theorem 21.2.2 was refined by Cherbonnier and Colmez as follows [**CC98**].

Definition 21.2.3. Let $\mathbf{B}^{\dagger}_{\mathbb{Q}_p}$ be the image of $\mathcal{E}^{\dagger}$ under the identification of $\mathcal{E}$ (with coefficients in $\mathbb{Q}_p$) with $\mathbf{B}_{\mathbb{Q}_p}$ sending t to $[\epsilon]-1$. Let $\mathbf{B}^{\dagger}_K$ be the integral closure of $\mathbf{B}^{\dagger}_{\mathbb{Q}_p}$ in $\mathbf{B}_K$. Again, $\mathbf{B}^{\dagger}_K$ carries actions of ϕ and Γ_K.

Definition 21.2.4. Let $\mathbf{A}^\dagger$ be the set of $x = \sum_{n=0}^\infty p^n[x_n] \in \mathbf{A}$ such that $\liminf_{n\to\infty}\{v(x_n)/n\} > -\infty$. Define

$$D^\dagger(V) = (V \otimes_{\mathbb{Q}_p} \mathbf{B}^\dagger)^{H_K};$$

it is an étale (ϕ, Γ)-module over $\mathbf{B}_K^\dagger$.

The following is the main result of [**CC98**].

Theorem 21.2.5 (Cherbonnier-Colmez). *The functor $D^\dagger$, from the category of continuous representations of G_K on finite dimensional $\mathbb{Q}_p$-vector spaces to the category of étale (ϕ, Γ)-modules over $\mathbf{B}_K^\dagger$, is an equivalence of categories.*

Remark 21.2.6. By Theorem 21.2.2, it suffices to check that the base extension functor from étale (ϕ, Γ)-modules over $\mathbf{B}_K^\dagger$ to étale (ϕ, Γ)-modules over $\mathbf{B}_K$ is an equivalence. The full faithfulness of this functor is elementary; it follows from Lemma 18.5.6. The essential surjectivity is much deeper; it amounts to the fact that the natural map

$$D^\dagger(V) \otimes_{\mathbf{B}_K^\dagger} \mathbf{B}^\dagger \to V \otimes_{\mathbb{Q}_p} \mathbf{B}^\dagger$$

is an isomorphism. Verifying this requires developing an appropriate analogy to Sen's theory of decompletion; this analogy has been developed into a full abstract Sen theory by Berger and Colmez [**BC07**].

A further variant was proposed by Berger [**Brg02**].

Definition 21.2.7. Using the identification $\mathbf{B}_{\mathbb{Q}_p}^\dagger \cong \mathcal{E}^\dagger$, put

$$\mathbf{B}_{\mathrm{rig},K}^\dagger = \mathbf{B}_K^\dagger \otimes_{\mathbf{B}_{\mathbb{Q}_p}^\dagger} \mathcal{R}.$$

Note that $\mathbf{B}_{\mathrm{rig},K}^\dagger$ admits continuous extensions (for the LF-topology) of the actions of ϕ and Γ_K. Define

$$D_{\mathrm{rig}}^\dagger(V) = D^\dagger(V) \otimes_{\mathbf{B}_K^\dagger} \mathbf{B}_{\mathrm{rig},K}^\dagger;$$

it is an étale (ϕ, Γ)-module over $\mathbf{B}_{\mathrm{rig},K}^\dagger$.

Theorem 21.2.8 (Berger). *The functor $D_{\mathrm{rig}}^\dagger$, from the category of continuous representations of G_K on finite dimensional $\mathbb{Q}_p$-vector spaces to the category of étale (ϕ, Γ)-modules over $\mathbf{B}_{\mathrm{rig},K}^\dagger$, is an equivalence of categories.*

Remark 21.2.9. The principal content in Theorem 21.2.8 is that the base extension functor from étale ϕ-modules over $\mathcal{E}^\dagger$ to étale ϕ-modules over $\mathcal{R}$ is fully faithful; this is elementary (see exercises). The essential surjectivity of the functor is almost trivial, since étaleness of the ϕ-action is defined over the Robba ring by base extension from $\mathcal{E}^\dagger$. One needs only check that the Γ_K-action also descends to any étale lattice, but this is easy (it is similar to Lemma 18.5.4).

3. Galois cohomology

Since the functor D and its variants lose no information about Galois representations, it is unsurprising that they can be used to recover basic invariants of a representation, such as Galois cohomology.

Definition 21.3.1. Assume for simplicity that Γ_K is procyclic; this only eliminates the case where $p = 2$ and $\{\pm 1\} \subset \Gamma$, for which see Remark 21.3.2 below. Let γ be a topological generator of Γ. Define the *Herr complex* over $\mathbf{B}_K$ associated to V as the complex (with the first nonzero term placed in degree zero)

$$0 \to D(V) \to D(V) \oplus D(V) \to D(V) \to 0$$

with the first map being $m \mapsto ((\phi - 1)m, (\gamma - 1)m)$ and the second map being $(m_1, m_2) \to (\gamma - 1)m_1 - (\phi - 1)m_2$. (The fact that this is a complex follows from the commutativity between ϕ and γ.) Similarly, define the Herr complex over $\mathbf{B}_K^\dagger$ or $\mathbf{B}_{\mathrm{rig},K}^\dagger$ by replacing $D(V)$ by $D^\dagger(V)$ or $D_{\mathrm{rig}}^\dagger(V)$, respectively.

Remark 21.3.2. A more conceptual description, which also covers the case where Γ_K need not be profinite, is that one takes the total complex associated to

$$0 \to C^\cdot(\Gamma_K, D(V)) \stackrel{\phi-1}{\to} C^\cdot(\Gamma_K, D(V)) \to 0.$$

One might think of this as the "monoid cohomology" of $\Gamma_K \times \phi^{\mathbb{Z}_{\geq 0}}$ acting on $D(V)$.

Theorem 21.3.3. *The cohomology of the Herr complex computes the Galois cohomology of V.*

Proof. For $\mathbf{B}_K$, the desired result was established by Herr [**Her98**]. The argument proceeds in two steps. One first takes cohomology of the Artin-Schreier sequence

$$0 \to \mathbb{Q}_p \to \mathbf{B} \stackrel{\phi-1}{\to} \mathbf{B} \to 0$$

after tensoring with V. This reduces the claim to the fact that the inflation homomorphisms

$$H^i(\Gamma_K, D(V)) \to H^i(G_K, V \otimes_{\mathbb{Q}_p} \mathbf{B})$$

are bijections; this is proved by adapting a technique introduced by Sen.

For $\mathbf{B}_K^\dagger$ and $\mathbf{B}_{\mathrm{rig},K}^\dagger$, the desired result was established by Liu [**Liu07**]; this proceeds by comparison with the original Herr complex rather than by imitating the above argument, though one could probably do that also. $\square$

Remark 21.3.4. As is done in [**Her98, Liu07**], one can make Theorem 21.3.3 more precise. For instance, the construction of Galois cohomology is functorial; there is also an interpretation in the Herr complex of the cup product in cohomology.

Remark 21.3.5. One can also use the Herr complex to recover Tate's fundamental theorems about Galois cohomology (finite dimensionality, Euler-Poincaré characteristic formula, local duality). This was done by Herr in [**Her01**].

4. Differential equations from (ϕ, Γ)-modules

One of the original goals of p-adic Hodge theory was to associate finer invariants to p-adic Galois representations, so as for instance to distinguish those representations which arose in geometry (i.e., from the étale cohomology of varieties over K). This was originally done using a collection of "period rings" introduced by Fontaine; more recently, Berger's work has demonstrated that one can reproduce these constructions using (ϕ, Γ)-modules. Here is a brief description of an example that shows the relevance of p-adic differential equations to this study. We will make reference to Fontaine's rings $\mathbf{B}_{\mathrm{dR}}, \mathbf{B}_{\mathrm{st}}$ without definition, for which see [**Brg04**].

Definition 21.4.1. Let $\chi : \Gamma_K \to \mathbb{Z}_p^\times$ denote the cyclotomic character; that is, for all nonnegative integers m and all $\gamma \in \Gamma_K$,

$$\gamma(\zeta_{p^m}) = \zeta_{p^m}^{\chi(\gamma)}.$$

For $\gamma \in \Gamma_K$ sufficiently close to 1, we may compute

$$\nabla = \frac{\log(\gamma)}{\log \chi(\gamma)}$$

as an endomorphism of $D(V)$, using the power series for $\log(1+x)$. The result does not depend on γ.

Remark 21.4.2. If one views Γ_K as a one-dimensional p-adic Lie group over $\mathbb{Z}_p$, then ∇ is the action of the corresponding Lie algebra.

Definition 21.4.3. Note that ∇ acts on $\mathbf{B}^\dagger_{\mathrm{rig},K}$ with respect to

$$f \mapsto [\epsilon] \log[\epsilon] \frac{df}{d[\epsilon]}.$$

As a result, it does not induce a differential module structure with respect to $\frac{d}{dt}$ on $D(V)$, but only on $D(V) \otimes \mathbf{B}^\dagger_{\mathrm{rig},K}[(\log[\epsilon])^{-1}]$. We say that V is *de Rham* if there exists a differential module with Frobenius structure M over $\mathbf{B}^\dagger_{\mathrm{rig},K}$ and an isomorphism

$$D(V) \otimes \mathbf{B}^\dagger_{\mathrm{rig},K}[(\log[\epsilon])^{-1}] \to M \otimes \mathbf{B}^\dagger_{\mathrm{rig},K}[(\log[\epsilon])^{-1}]$$

of differential modules with Frobenius structure.

One then has the following results of Berger [**Brg02**].

Theorem 21.4.4 (Berger). (a) *The representation V is de Rham if and only if it is de Rham in Fontaine's sense, i.e., if*

$$D_{\mathrm{dR}}(V) = (V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}})^{G_K}$$

satisfies

$$D_{\mathrm{dR}}(V) \otimes_K \mathbf{B}_{\mathrm{dR}} \cong V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{dR}}.$$

(b) *Suppose that V is de Rham. Then V is semistable in Fontaine's sense, i.e.,*

$$D_{\mathrm{st}}(V) = (V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{st}})^{G_K}$$

satisfies

$$D_{\mathrm{st}}(V) \otimes_F \mathbf{B}_{\mathrm{st}} \cong V \otimes_{\mathbb{Q}_p} \mathbf{B}_{\mathrm{st}},$$

if and only if there exists M as in Definition 21.4.3 which is unipotent.

Applying Theorem 18.1.8 then yields the following corollary, which was previously a conjecture of Fontaine [**Fon94**, 6.2].

Corollary 21.4.5 (Berger). *Every de Rham representation is potentially semistable, i.e., becomes semistable upon restriction to G_L for some finite extension L of K.*

REMARK 21.4.6. The term "de Rham" is meant to convey the fact that if $V = H^i_{\mathrm{et}}(X \times_K K^{\mathrm{alg}}, \mathbb{Q}_p)$ for X a smooth proper variety over K, then V is de Rham and you can use the aforementioned constructions to recover $H^i_{\mathrm{dR}}(X, K)$ functorially from V (solving Grothendieck's "problem of the mysterious functor"). See [**Brg04**] for more of the story.

5. Beyond Galois representations

The category of arbitrary (ϕ, Γ)-modules over $\mathbf{B}^\dagger_{\mathrm{rig},K}$ turns out to have its own representation-theoretic interpretation; it is equivalent to the category of *B-pairs* introduced by Berger [**Brg07a**]. One can associate "Galois cohomology" to such objects using the Herr complex; it has been shown by Liu [**Liu07**] that the analogues of Tate's theorems (see Remark 21.3.5) still hold. These functors can be interpreted as the derived functors of $\mathrm{Hom}(D^\dagger_{\mathrm{rig}}(V_0), \cdot)$ for V_0 the trivial representation [**Ked07f**, Appendix].

One may wonder why one should be interested in (ϕ, Γ)-modules over $\mathbf{B}^\dagger_{\mathrm{rig},K}$ if ultimately one has in mind an application concerning only Galois representations. One answer is that converting Galois representations into (ϕ, Γ)-modules exposes extra structure that is not visible without the conversion.

DEFINITION 21.5.1 (Colmez). We say V is *trianguline* if $D^\dagger_{\mathrm{rig}}(V)$ is a successive extension of (ϕ, Γ)-modules of rank 1 over $\mathbf{B}^\dagger_{\mathrm{rig},K}$. The point is that these need not be étale, so V need not be a successive extension of representations of dimension 1.

The trianguline representations have the dual benefits of being relatively easy to classify, and somewhat commonplace. On one hand, Colmez [**Col07**] classified the two-dimensional trianguline representations of $G_{\mathbb{Q}_p}$; the classification includes a parameter (the $\mathcal{L}$-invariant) relevant to p-adic L-functions. On the other hand, a result of Kisin [**Kis03**] shows that the Galois representations associated to many classical modular forms are trianguline.

Notes

Our presentation here is largely a summary of Berger's [**Brg04**], which we highly recommend.

A variant of the theory of (ϕ, Γ)-modules was introduced by Kisin [**Kis06**], using the Kummer tower $K(p^{1/p^n})$ instead of the cyclotomic tower $K(\zeta_{p^n})$. This leads to certain advantages, particularly when studying crystalline representations. Kisin's work is based on an earlier paper of Berger [**Brg07b**]; both of these use slope filtrations (as in Theorem 18.5.1) to recover a theorem of Colmez-Fontaine classifying semistable Galois representations in terms of certain linear algebraic data.

After [**Brg02**] appeared, Fontaine succeeded in giving a direct proof of Corollary 21.4.5 (i.e., not going through p-adic differential equations). We do not have a reference for this.

Exercises

(1) (Compare [**Tsu96**, Proposition 2.2.2].) Let A be an $n \times n$ matrix over $\mathfrak{o}_{\mathcal{E}^\dagger}$, and suppose $v \in \mathcal{E}^n$, $w \in (\mathcal{E}^\dagger)^n$ satisfy $Av - \phi(v) = w$. Then $v \in (\mathcal{E}^\dagger)^n$. This gives a direct proof of some cases of Theorem 18.7.1, in the spirit of Lemma 18.5.6. (Hint: reduce to the case where $|A|_\rho \leq 1$ for some $\rho \in (0,1)$ for which $|w|_\rho < \infty$. Then use $|w|_\rho$ to bound the terms of $v = \sum_i v_i t^i$ for which $|v_i| \geq c$.)

Bibliography

[AS02] A. Abbes and T. Saito, Ramification of local fields with imperfect residue fields, *Amer. J. Math.* **124** (2002), 879–920.

[AS03] A. Abbes and T. Saito, Ramification of local fields with imperfect residue fields, II, *Doc. Math.* Extra Vol. (2003), 5–72.

[AKR07] T.G. Abbott, K.S. Kedlaya, and D. Roe, Bounding Picard numbers of surfaces using p-adic cohomology, arXiv:math/0601508v2 (2007); to appear in *Arithmetic, Geometry and Coding Theory (AGCT 2005)*, Societé Mathématique de France.

[And01] Y. André, Différentelles non commutatives et théorie de Galois différentielle ou aux différences, *Ann. Sci. École Norm. Sup.* (4) **34** (2001), 685–739.

[And02] Y. André, Filtrations de type Hasse-Arf et monodromie p-adique, *Invent. Math.* **148** (2002), 285–317.

[And07] Y. André, Dwork's conjecture on the logarithmic growth of solutions of p-adic differential equations, *Compos. Math.*, to appear.

[BdV07] F. Baldassarri and L. di Vizio, Continuity of the radius of convergence of p-adic differential equations on Berkovich analytic spaces, preprint.

[BR07] M. Baker and R. Rumely, *Potential Theory on the Berkovich Projective Line*, book in progress, available at `http://www.math.gatech.edu/~mbaker/papers.html`.

[Brg02] L. Berger, Représentations p-adiques et équations différentielles, *Invent. Math.* **148** (2002), 219–284.

[Brg04] L. Berger, An introduction to the theory of p-adic representations, in *Geometric Aspects of Dwork Theory. Vol I, II*, de Gruyter, Berlin, 2004, 255–292.

[Brg07a] L. Berger, Construction de (φ, Γ)-modules: représentations p-adiques et B-paires, *Algebra and Num. Theory*, to appear.

[Brg07b] L. Berger, Équations différentielles p-adiques et (ϕ, N)-modules filtrés, *Astrérisque*, to appear.

[BC07] L. Berger and P. Colmez, Familles de représentations de de Rham et monodromie p-adique, *Astérisque*, to appear.

[Brk90] V.G. Berkovich, *Spectral Theory and Analytic Geometry over Non-Archimedean Fields*, Math. Surveys and Monographs 33, Amer. Math. Soc., Providence, 1990.

[Brk93] V.G. Berkovich, Étale cohomology for non-Archimedean analytic spaces, *Publ. Math. IHÉS* **78** (1993), 5–161.

[Brt86] P. Berthelot, Géométrie rigide et cohomologie des variétés algébriques de caractéristique p, Introductions aux cohomologies p-adiques (Luminy, 1984), *Mém. Soc. Math. France (N.S.)* **23** (1986), 7–32.

[Brt96] P. Berthelot, Cohomologie rigide et cohomologie rigide à support propre. Première partie, Prépublication IRMAR 96-03, available at `www.math.univ-rennes1.fr/~berthelo`.

[Brt97a] P. Berthelot, Finitude et pureté cohomologique en cohomologie rigide (with an appendix in English by A.J. de Jong), *Invent. Math.* **128** (1997), 329–377.

[Brt97b] P. Berthelot, Dualité de Poincaré et formule de Künneth en cohomologie rigide, *C.R. Acad. Sci. Paris* **325** (1997), 493–498.

[Brt02] P. Berthelot, Introduction à la théorie arithmétique des $\mathcal{D}$-modules, Cohomologies p-adiques et applications arithmétiques, II, *Astérisque* **279** (2002), 1–80.

[BO78] P. Berthelot and A. Ogus, *Notes on Crystalline Cohomology*, Princeton Univ. Press, Princeton, 1978.

[Bha97] R. Bhatia, *Matrix Analysis*, Graduate Texts in Math. 169, Springer-Verlag, New York, 1997.

[Bos05] S. Bosch, Lectures on formal and rigid geometry, preprint (2005) available at `http://wwwmath1.uni-muenster.de/sfb/about/publ/bosch.html`.

[BGR84] S. Bosch, U. Güntzer, and R. Remmert, *Non-Archimedean Analysis*, Grundlehren der Math. Wiss. 261, Springer-Verlag, Berlin, 1984.

[BC05] K. Buzzard and F. Calegari, Slopes of overconvergent 2-adic modular forms, *Compos. Math.* **141** (2005), 591–604.

[CC98] F. Cherbonnier and P. Colmez, Représentations p-adiques surconvergentes, *Invent. Math.* **133** (1998), 581–611.

[CT06] B. Chiarellotto and N. Tsuzuki, Logarithmic growth and Frobenius filtrations for solutions of p-adic differential equations, preprint (2006).

[CP07] B. Chiarellotto and A. Pulita, Arithmetic and Differential Swan conductors of rank one representations with finite local monodromy, preprint (2007).

[Chr83] G. Christol, *Modules Différentiels et Equations Différentielles p-adiques*, Queen's Papers in Pure and Applied Math. 66, Queen's Univ., Kingston, 1983.

[Chr01] G. Christol, About a Tsuzuki theorem, in *p-adic Functional Analysis (Ioannina, 2000)*, Lecture Notes in Pure and Appl. Math., 222, Dekker, New York, 2001, 63–74.

[CD91] G. Christol and B. Dwork, Effective p-adic bounds at regular singular points, *Duke Math. J.* **62** (1991), 689–720.

[CD92] G. Christol and B. Dwork, Differential modules of bounded spectral norm, in *p-adic Methods in Number Theory and Algebraic Geometry*, Contemp. Math. 133, Amer. Math. Soc., Providence, 1992.

[CD94] G. Christol and B. Dwork, Modules différentielles sur les couronnes, *Ann. Inst. Fourier* **44** (1994), 663–701.

[CM97] G. Christol and Z. Mebkhout, Sur le théorème de l'indice des équations différentielles, II, *Annals of Math.* **146** (1997), 345–410.

[CM00] G. Christol and Z. Mebkhout, Sur le théorème de l'indice des équations différentielles, III, *Annals of Math.* **151** (2000), 385–457.

[CM01] G. Christol and Z. Mebkhout, Sur le théorème de l'indice des équations différentielles, IV, *Invent. Math.* **143** (2001), 629–672.

[Cla66] D. Clark, A note on the p-adic convergence of solutions of linear differential equations, *Proc. Amer. Math. Soc.* **17** (1966), 262–269.

[Coh65] R.M. Cohn, *Difference Algebra*, John Wiley & Sons, New York-London-Sydney, 1965.

[Col82] R.F. Coleman, Dilogarithms, regulators and p-adic L-functions, *Invent. Math.* **69** (1982), 171–208.

[Col07] P. Colmez, Représentations triangulines de dimension 2, preprint (2007), available online at `http://www.institut.math.jussieu.fr/~colmez/publications.html`.

[Con07] B. Conrad, Several approaches to non-archimedean geometry, lecture notes from the 2007 Arizona Winter School, available online at `http://swc.math.arizona.edu`.

[Cre87] R. Crew, F-isocrystals and p-adic representations., in *Algebraic geometry, Bowdoin, 1985 (Brunswick, Maine, 1985)*, Proc. Sympos. Pure Math., 46, Part 2, Amer. Math. Soc., Providence, 1987, 111–138.

[Cre98] R. Crew, Finiteness theorems for the cohomology of an overconvergent isocrystal on a curve, *Ann. Sci. Éc. Norm. Sup.* **31** (1998), 717–763.

[Cre00] R. Crew, Canonical extensions, irregularities, and the Swan conductor, *Math. Ann.* **316** (2000), 19–37.

[Csi07] N.E. Csima, Newton-Hodge filtration for self-dual F-crystals, arXiv:0706/2530v1 (2007).

[dJ98a] A.J. de Jong, Homomorphisms of Barsotti-Tate groups and crystals in positive characteristic, *Invent. Math.* **134** (1998), 301–333.

[dJ98b] A.J. de Jong, Barsotti-Tate groups and crystals, *Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998)*, *Doc. Math.* Extra Vol. II (1998), 259–265.

[Del70] P. Deligne, *Equations Différentielles à Points Singuliers Réguliers*, Lecture Notes in Math. 163, Springer-Verlag, Berlin, 1970.

[DK73] P. Deligne and N. Katz (eds.), *Groupes de Monodromie en Géométrie Algébrique II, Séminaire de Géométrie Algébrique du Bois-Marie 1967–1969 (SGA 7 II)* Lecture Notes in Math. 340, Springer-Verlag, Berlin-New York, 1973.

[Dem72] M. Demazure, *Lectures on p-divisible Groups*, Lecture Notes in Math. 302, Springer-Verlag, New York, 1972.

[DV06] J. Denef and F. Vercauteren, An extension of Kedlaya's algorithm to hyperelliptic curves in characteristic 2, *J. Cryptology* **19** (2006), 1–25.

[Dwo69] B. Dwork, p-adic cycles, *Publ. Math. IHÉS* **37** (1969), 27–115.

[Dwo73a] B. Dwork, On p-adic differential equations, II. The p-adic asymptotic behavior of solutions of ordinary linear differential equations with rational function coefficients, *Ann. of Math.* (2) **98** (1973), 366–376.

[Dwo73b] B. Dwork, On p-adic differential equations, III. On p-adically bounded solutions of ordinary linear differential equations with rational function coefficients, *Invent. Math.* **20** (1973), 35–45.

[Dwo74] B. Dwork, Bessel functions as p-adic functions of the argument, *Duke Math. J.* **41** (1974), 711–738.

[Dwo97] B.M. Dwork, On exponents of p-adic differential modules, *J. reine angew. Math.* **484** (1997), 85–126.

[DGS94] B. Dwork, G. Gerotto, and F. Sullivan, *An Introduction to G-Functions*, Annals of Math. Studies 133, Princeton University Press, Princeton, 1994.

[DR77] B. Dwork and P. Robba, On ordinary linear p-adic differential equations, *Trans. Amer. Math. Soc.* **231** (1977), 1–46.

[DR80] B. Dwork and P. Robba, Effective p-adic bounds for solutions of homogeneous linear differential equations, *Trans. Amer. Math. Soc.* **259** (1980), 559–577.

[Edi06] B. Edixhoven, Point counting after Kedlaya, course notes (2006) available at `http://www.math.leidenuniv.nl/~edix/oww/mathofcrypt/carls_edixhoven/kedlaya.pdf`.

[Fon94] J.-M. Fontaine, Représentations p-adiques semi-stables, Périodes p-adiques (Bures-sur-Yvette, 1988), *Astérisque* **23** (1994), 113–184.

[FW79] J.-M. Fontaine and J.-P. Wintenberger, Le "corps de normes" de certaines extensions algébriques de corps locaux, *C.R. Acad. Sci. Paris Sér. A-B* **288** (1979), A367–A370.

[FvdP04] J. Fresnel and M. van der Put, *Rigid Analytic Geometry and its Applications*, Progress in Mathematics 218, Birkhäuser, Boston, 2004.

[Ful98] W. Fulton, *Intersection Theory*, second edition, Springer-Verlag, Berlin, 1998.

[Ful00] W. Fulton, Eigenvalues, invariant factors, highest weights, and Schubert calculus, *Bull. Amer. Math. Soc. (N.S.)* **37** (2000), 209–249.

[Ger07] R. Gerkmann, Relative rigid cohomology and deformation of hypersurfaces, *Int. Math. Res. Papers* **2007** (2007), article ID rpm003 (67 pages).

[Har77] R. Hartshorne, *Algebraic Geometry*, Graduate Texts in Math. 52, Springer-Verlag, New York, 1977.

[Har07] D. Harvey, Kedlaya's algorithm in larger characteristic, *Int. Math. Res. Notices* **2007** (2007), article ID rnm095 (29 pages).

[Her98] L. Herr, Sur la cohomologie galoisienne des corps p-adiques, *Bull. Soc. Math. France* **126** (1998), 563–600.

[Her01] L. Herr, Une approche nouvelle de la dualité locale de Tate, *Math. Ann.* **320** (2001), 307–337.

[Hor54] A. Horn, On the eigenvalues of a matrix with prescribed singular values, *Proc. Amer. Math. Soc.* **5** (1954), 4–7.

[Hor62] A. Horn, Eigenvalues of sums of Hermitian matrices, *Pacific J. Math.* **12** (1962), 225–241.

[Hub07] H. Hubrechts, Point counting in families of hyperelliptic curves, *Found. Comp. Math.*, to appear.

[Kap42] I. Kaplansky, Maximal fields with valuations, *Duke Math. J.* **9** (1942), 303–321.

[Kat79] N.M. Katz, Slope filtration of F-crystals, Journées de Géométrie Algébrique de Rennes (Rennes, 1978), Vol. I, *Astrisque* **63** (1979), 113–163.

[Kat89] K. Kato, Swan conductors for characters of degree one in the imperfect residue field case, *Contemp. Math.*, vol. 83, Amer. Math. Soc., Providence, 1989, 101–131.

[KO68] N.M. Katz and T. Oda, On the differentiation of de Rham cohomology classes with respect to parameters, *J. Math. Kyoto Univ.* **8** (1968), 199–213.

[Ked01] K.S. Kedlaya, Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology, *J. Ramanujan Math. Soc.* **16** (2001), 323-338; errata, *ibid.* **18** (2003), 417-418.

[Ked04a] K.S. Kedlaya, A p-adic local monodromy theorem, *Annals of Math.* **160** (2004), 93–184.

[Ked04b] K.S. Kedlaya, Full faithfulness for overconvergent F-isocrystals, in A. Adolphson et al (eds.), *Geometric Aspects of Dwork Theory (Volume II)*, de Gruyter, Berlin, 2004, 819–835.

[Ked04c] K.S. Kedlaya, Computing zeta functions via p-adic cohomology, *Algorithmic Number Theory (ANTS VI)*, Lecture Notes in Comp. Sci. 3076, Springer-Verlag, 2004, 1–17.

[Ked05a] K.S. Kedlaya, Local monodromy of p-adic differential equations: an overview, *Int. J. Number Theory* **1** (2005), 109–154; errata at `http://math.mit.edu/~kedlaya/papers`.

[Ked05b] K.S. Kedlaya, Slope filtrations revisited, *Documenta Math.* **10** (2005), 447-525; errata, *ibid.* **12** (2007), 361–362.

[Ked05c] K.S. Kedlaya, Frobenius modules and de Jong's theorem, *Math. Res. Lett.* **12** (2005), 303–320.

[Ked06a] K.S. Kedlaya, Finiteness of rigid cohomology with coefficients, *Duke Math. J.* **134** (2006), 15–97.

[Ked06b] K.S. Kedlaya, Fourier transforms and p-adic "Weil II", *Compos. Math.* **142** (2006), 1426–1450.

[Ked07a] K.S. Kedlaya, Swan conductors for p-adic differential modules, I: A local construction, *Algebra and Number Theory* **1** (2007), 269–300.

[Ked07b] K.S. Kedlaya, The p-adic local monodromy theorem for fake annuli, arXiv:math/0507496v4 (2006), to appear in *Rend. Sem. Math. Padova.*

[Ked07c] K.S. Kedlaya, Slope filtrations for relative Frobenius, arXiv:math/0609272v2 (2007); to appear in *Astérisque.*

[Ked07d] K.S. Kedlaya, Swan conductors for p-adic differential modules, II: Global variation, arXiv:0705.0031v1 (2007).

[Ked07e] K.S. Kedlaya, Semistable reduction for overconvergent F-isocrystals, III: local semistable reduction at monomial valuations, arXiv:math/0609645v2 (2007).

[Ked07f] K.S. Kedlaya, Some new directions in p-adic Hodge theory, arXiv:0709.1970v1 (2007).

[Kis03] M. Kisin, Overconvergent modular forms and the Fontaine-Mazur conjecture, *Invent. Math.* **153** (2003), 373–454.

[Kis06] M. Kisin, Crystalline representations and F-crystals, in *Algebraic Geometry and Number Theory*, Progress in Math. 253, Birkhäuser, Boston, 2006, 459–496.

[Kot85] R.E. Kottwitz, Isocrystals with additional structure, *Comp. Math.* **56** (1985), 201–220.

[Kot97] R.E. Kottwitz, Isocrystals with additional structure. II, *Comp. Math.* **109** (1997), 255–339.

[Kot03] R.E. Kottwitz, On the Hodge-Newton decomposition for split groups, *Int. Math. Res. Notices* **26** (2003), 1433–1447.

[Kru32] W. Krull, Allgemeine Bewertungstheorie, *J. für Math.* **167** (1932), 160–196.

[Lan56] S. Lang, Algebraic groups over finite fields, *Amer. J. Math.* **78** (1956), 555–563.

[Lau04] A.G.B. Lauder, Deformation theory and the computation of zeta functions, *Proc. London Math. Soc.* **88** (2004), 565–602.

[Laz62] M. Lazard, Les zéros d'une fonction analytique d'une variable sur un corps valué complet, *Publ. Math. IHÉS* **14** (1962), 47–75.

[leS07] B. le Stum, *Rigid Cohomology*, Cambridge Tracts in Math. 172, Cambridge Univ. Press, 2007.

[Liu07] R. Liu, Cohomology and duality for (ϕ, Γ)-modules over the Robba ring, *Int. Math. Res. Notices*, to appear.

[Loe96] F. Loeser, Exposants p-adiques et théorèmes d'indice pour les équations différentielles p-adiques (d'après G. Christol et Z. Mebkhout), Séminaire Bourbaki, Vol. 1996/97, *Astérisque* **245** (1997), 57–81.

[Lut37] E. Lutz, Sur l'équation $y^2 = x^3 + Ax + B$ sur les corps $\mathfrak{p}$-adiques, *J. reine angew Math.* **177** (1937), 238–247.

[Mal74] B. Malgrange, Sur les points singuliers des équations différentielles, *Enseign. Math.* **20** (1974), 147–176.

[Man63] Yu. I. Manin, The theory of commutative formal groups over fields of finite characteristic (Russian), *Usp. Math.* 18 (1963), 3–90; English translation, *Russian Math. Surveys* **18** (1963), 1–80.

[Mar04] A. Marmora, Irrégularité et conducteur de Swan p-adiques, *Doc. Math.* **9** (2004), 413–433.

[Mat02] S. Matsuda, Katz correspondence for quasi-unipotent overconvergent isocrystals, *Comp. Math.* **134** (2002), 1–34.

[Mat04] S. Matsuda, Conjecture on Abbes-Saito filtration and Christol-Mebkhout filtration, in *Geometric Aspects of Dwork Theory. Vol. I, II*, de Gruyter, Berlin, 2004, 845–856.

[Meb02] Z. Mebkhout, Analogue p-adique du théorème de Turrittin et le théorème de la monodromie p-adique, *Invent. Math.* **148** (2002), 319–351.

[Nag62] M. Nagata, *Local Rings*, Interscience Tracts in Pure and Applied Math. 13, John Wiley & Sons, New York, 1962.

[Ore33] O. Ore, Theory of non-commutative polynomials, *Annals of Math.* **34** (1933), 480–508.

[Pil90] J. Pila, Frobenius maps of abelian varieties and finding roots of unity in finite fields, *Math. Comp.* **55** (1990), 745–763.

[Poo93] B. Poonen, Maximally complete fields, *Enseign. Math.* (2) **39** (1993), 87–106.

[Ram84] J.-P. Ramis, Théorèmes d'indices Gevrey pour les équations différentielles ordinaires, *Mem. Amer. Math. Soc.* **48** (1984).

[Rib99] P. Ribenboim, *The theory of classical valuations*, Springer-Verlag, New York, 1999.

[Rit50] J.F. Ritt, *Differential Algebra*, Colloq. Pub. XXXIII, Amer. Math. Soc., New York, 1950.

[Rob80] P. Robba, Lemmes de Hensel pour les opérateurs différentiels. Application à la réduction formelle des équations différentielles, *Enseign. Math.* (2) **26** (1980), 279–311.

[Rob00] A.M. Robert, *A Course in p-adic Analysis*, Graduate Texts in Math. 198, Springer-Verlag, New York, 2000.

[Sch02] P. Schneider, *Nonarchimedean Functional Analysis*, Springer-Verlag, Berlin, 2002.

[Sch85] R. Schoof, Elliptic curves over finite fields and the computation of square roots mod p, *Math. Comp.* **44** (1985), 483–494.

[Ser79] J.-P. Serre, *Local Fields*, Graduate Texts in Math. 67, Springer-Verlag, 1979.

[Sil91] J.H. Silverman, *The Arithmetic of Elliptic Curves*, second printing, Graduate Texts in Math. 106, Springer-Verlag, New York, 1991.

[SvdP97] M.F. Singer and M. van der Put, *Galois Theory of Difference Equations*, Lecture Notes in Math. 1666, Springer-Verlag, Berlin, 1997.

[SvdP03] M.F. Singer and M. van der Put, *Galois Theory of Linear Differential Equations*, Grundlehren der Math. Wiss. 328, Springer-Verlag, Berlin, 2003.

[Thu05] A. Thuillier, Théorie du potentiel sur les courbes en géométrie analytique non archimédienne. Applications à la theorie d'Arakelov, thesis, Université de Rennes 1, 2005.

[Tsu96] N. Tsuzuki (as T. Nobuo), The overconvergence of morphisms of étale ϕ-∇-spaces on a local field, *Compos. Math.* **103** (1996), 227–239.

[Tsu98a] N. Tsuzuki, Finite local monodromy of overconvergent unit-root F-isocrystals on a curve, *Amer. J. Math.* **120** (1998), 1165–1190.

[Tsu98b] N. Tsuzuki, The local index and the Swan conductor, *Comp. Math.* **111** (1998), 245–288.

[Tsu98c] N. Tsuzuki, Slope filtration of quasi-unipotent overconvergent F-isocrystals, *Ann. Inst. Fourier (Grenoble)* **48** (1998), 379–412.

[Tsu02] N. Tsuzuki, Morphisms of F-isocrystals and the finite monodromy theorem for unit-root F-isocrystals, *Duke Math. J.* **111** (2002), 385–419.

[vdP86] M. van der Put, The cohomology of Monsky and Washnitzer, Introductions aux cohomologies p-adiques (Luminy, 1984), *Mém. Soc. Math. France* **23** (1986), 33–59.

[vR78] A.C.M. van Rooij, *Non-Archimedean Functional Analysis*, Monographs and Textbooks in Pure and Applied Math. 51, Marcel Dekker, New York, 1978.

[Wey12] H. Weyl, Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung), *Math. Ann.* **71** (1912), 441–479.

[Wey49] H. Weyl, Inequalities between the two kinds of eigenvalues of a linear transformation, *Proc. Nat. Acad. Sci. USA* **35** (1949), 408–411.

[Xia07] L. Xiao, On Abbes-Saito's ramification filtrations and p-adic differential equations, I, in preparation.

[You92] P.T. Young, Radii of convergence and index for p-adic differential operators, *Trans. Amer. Math. Soc.* **333** (1992), 769–785.