Foundations of p-adic differential equations

Kiran S. Kedlaya

Contents

Front matter

Preface

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).

Besides serving as a general introduction to the topic, this book serves the secondary purpose of providing a foundational reference where none was previously available. In particular, some results presented herein turn out to be original, even though the techniques used to prove them have been available for some time. (The best prior introductory book we could find is that of Dwork, Gerotto, and Sullivan [**DGS94**]; as a result, it is cited extensively here.) The original results are expected to have some impact both in theory (in the study of higher-dimensional p-adic differential equations in the context of the "semistable reduction problem" for overconvergent F-isocrystals) and in practice (in the explicit computation of solutions of p-adic differential equations, e.g., for finding zeta functions of explicit varieties).

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.

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 unit 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. Structure of the book

We have attempted to equip each chapter of the book with a similar basic structure. The chapter begins with a short introduction explaining what is to be discussed. After the main body of material, we 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**].) Finally, we attach a few exercises, which in some cases include the proofs of some results which will be used later. (However, all results to be used in the text are stated explicitly in the text, so there are no cross-references into exercises except from other exercises.)

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, who both gave guest lectures during the course, and to Ruochuan Liu and 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 1

An example of Dwork

In this chapter, we 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 is only a summary, we skip many proofs; this practice will be cast aside in the next chapter, not to return until much later in the book.

1. Zeta functions of varieties

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 1.1.1 (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*.) 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 $\mathbb{F}_q$.) For $X = E_\lambda$, one checks (exercise) that

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

THEOREM 1.1.2 (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. □

It is worth pointing out that the first complete proof uses 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 $\mathbb{Q}_\ell$.

2. Zeta functions and p-adic differential equations

All well and good, but there are several downsides of the interpretation in terms of étale cohomology. 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 1.2.1. 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 1.2.2. *The series $\alpha(z)$ converges p-adically for $|z| < 1$.*

PROOF. Straightforward. □

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

DEFINITION 1.2.3. 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 1.2.4 (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 1.2.5. 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 1.2.6 (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]. □

3. A word of caution

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 [**Ber90**] and further developed in [**Ber93**]. 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 1.1.2 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 2

Absolute values

In this unit, 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 [**Rbe00**] for a fuller treatment.

Beware that a couple of proofs will forward reference the unit 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 2.1.1. Let G be an abelian group. An *semi-absolute 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 2.1.2. 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 2.1.3. 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 2.1.4. *Let G be an abelian group equipped with an absolute value $|\cdot|$. Then there exists an abelian group G' equipped with an absolute value $|\cdot|'$, 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 2.1.5. With notation as in Theorem 2.1.4, 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 2.1.6. If R is a ring and $|\cdot|$ is a seminorm on its additive group, we say that $|\cdot|$ is *submultiplicative* if the following additional condition holds.

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

We say that $|\cdot|$ is *multiplicative* 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 2.1.7. *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 2.1.8. *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 2.2.1. 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 2.2.2. 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 2.2.3. 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 2.2.4 (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 [**Rbe00**, §2.1.6] and [**Rbe00**, §2.2.4], respectively. □

3. Norms on modules

DEFINITION 2.3.1. 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 2.3.2. 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 2.3.3. *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 2.2.4 to deduce that $F = \mathbb{R}$ or $F = \mathbb{C}$, then use compactness of the unit ball. □

Definition 2.3.4. 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 2.4.1. 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 2.4.2. 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 2.4.3. 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 2.4.4. 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 2.4.5 (Ostrowski). *Any nontrivial nonarchimedean absolute value on $\mathbb{Q}$ is equivalent to the p-adic absolute value for some prime p.*

Proof. See [**Rbe00**, §2.2.4]. □

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

Theorem 2.4.6. *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 2.3.3, 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 2.4.7. 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 2.4.6, 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 2.4.8. *Let F be an algebraically closed nonarchimedean field. Then the completion of F is also algebraically closed.*

For the proof, see section 3.

5. 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 2.5.1. 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 2.5.2. 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 2.5.3 (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 [**Rbe00**, §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 2.4.8 in the case of the completed algebraic closure of $\mathbb{Q}_p$, see [**Rbe00**, §3.3.3].

Exercises

(1) Prove Ostrowski's theorem (Theorem 2.2.4).
(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 2.4.6 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 3

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 3.1.1. 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

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

it is clearly submultiplicative. Moreover, it is also multiplicative if $|\cdot|$ is; see Proposition 3.1.3 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.

It is easy to check that the ρ-Gauss (semi)norm is submultiplicative. To see that it is multiplicative whenever the original norm is, we make the following observation.

DEFINITION 3.1.2. 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 3.1.3. *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,

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

This proves both claims. $\square$

DEFINITION 3.1.4. 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

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

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 3.1.5. *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 $|\cdot|_{e^{-r}}$. $\square$

2. Slope factorizations and a master factorization theorem

THEOREM 3.2.1. *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 3.2.1 and all of its variants. It, and the proof given here, are due to Christol [**Chr83**, Proposition 1.5.1].

THEOREM 3.2.2 (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 3.2.1.

Proof of Theorem 3.2.1. We apply Theorem 3.2.2 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 3.2.2.

Proof of Theorem 3.2.2. 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. □

3. Applications to nonarchimedean field theory

We now go back and apply Theorem 3.2.1 to prove some facts about extensions of nonarchimedean fields which were omitted in the previous chapter.

We first complete the proof of Theorem 2.4.6. For this, we need the following lemma.

Lemma 3.3.1. *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 2.4.6 (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 3.2.1; 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 3.3.1. $\square$

We next give the proof of Theorem 2.4.8. 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 3.3.2. *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 3.1.5, 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 2.4.8. 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 3.3.2 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 3.3.1.
(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 4

Numerical analysis

For the purposes of this book, numerical analysis is the study of 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. In this unit, we introduce some basic concepts of numerical analysis, first in the archimedean case, then in the nonarchimedean case.

NOTATION 4.0.1. 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 4.1.1. 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 4.1.2. 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 4.1.3. Let A^* denote the conjugate transpose (or *Hermitian transpose*) of A. The matrix A^*A is real symmetric, 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 4.1.4 (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. □

COROLLARY 4.1.5. *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 4.1.6. *The number σ_i is the largest 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 4.1.4 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 4.5.1.

Theorem 4.1.7 (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 4.1.7, due to Horn [Hor54, Theorem 4].

Theorem 4.1.8. *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 4.1.9. *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 eigenvales $\lambda_{i+1}, \dots, \lambda_n$. 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. $\square$

THEOREM 4.1.10. *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 4.1.9 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 λ^{-1} 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.) $\square$

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

LEMMA 4.1.11. *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. $\square$

2. Perturbations (archimedean case)

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

THEOREM 4.2.1 (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 4.2.2. *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 $2^n \binom{n}{i}$ in absolute value signs is quite deliberate; it will be relevant in the nonarchimedean setting.

PROOF. First consider the case $i = j = n$. By continuity, we may assume that $\det(A) \neq 0$. Write

$$\begin{aligned}\det(A+B) &= \det(A)\det(I_n + A^{-1}B) \\ &= \det(A)(1 - R_{n-1} + \cdots \pm R_0),\end{aligned}$$

where $T^n + \sum_{i=0}^{n-1} R_i T^i$ is the characteristic polynomial of $A^{-1}B$. From the expansion of R_{n-i} as a sum of $\binom{n}{i}$ minors of size i, we have $|R_{n-i}| < \binom{n}{i}|A^{-1}B|^i$. Since $|A^{-1}| = \sigma_n^{-1}$, we have $|A^{-1}B| < 1$; we may thus write

$$|\det(A+B) - \det(A)| \leq |2^n| |\det(A)| |A^{-1}B| = |2^n| \sigma_1 \cdots \sigma_{n-1} |B|.$$

For the general case, write the coefficient of T^{n-i} in the characteristic polynomial of a matrix as the sum of $\binom{n}{i}$ minors of size i, then apply the previous case to each of these. $\square$

We also need to consider multiplicative perturbations.

PROPOSITION 4.2.3. *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 4.1.6. 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. $\square$

PROPOSITION 4.2.4. *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 4.1.11; 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 4.2.3 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. $\square$

3. Singular values and eigenvalues (nonarchimedean case)

We now pass to nonarchimedean analogues.

HYPOTHESIS 4.3.1. Throughout this section and the next, let F be a nonarchimedean field, and let A be an $n \times n$ matrix over F.

DEFINITION 4.3.2. 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 4.3.3. 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 4.3.4 (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. $\square$

COROLLARY 4.3.5. *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. $\square$

COROLLARY 4.3.6. *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 4.3.7. 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 archimedean analogue of Weyl's inequality is the following.

Theorem 4.3.8 (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 4.3.8 also has a converse, but in this case we can write the construction down quite explicitly.

Definition 4.3.9. 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

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

its characteristic polynomial is P.

Proposition 4.3.10. *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 4.3.11 (Hodge-Newton decomposition). *Suppose that for some $i \in \{1, \dots, n-1\}$ we have*

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

(That is, the Hodge and Newton polygons share a vertex with x-coordinate $-n+i$.) Then there exists $U \in \mathrm{GL}_n(\mathfrak{o}_F)$ 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 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 eigenvales $\lambda_{i+1}, \dots, \lambda_n$. (This can be constructed because by Theorem 3.2.1 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.) Apply the Smith normal form to construct a basis $w_1, \dots, w_n$ of $\mathfrak{o}_K^n$ such that $|Aw_i| = \sigma_i |w_i|$.

Since $\sigma_i > \sigma_{i+1}$, the only vectors $v \in \wedge^i \mathbb{C}^n$ for which $|Av|/|v|$ achieves its minimum value $\sigma_{i+1} \cdots \sigma_n$ are the nonzero multiples of $w_{i+1} \wedge \cdots \wedge w_n$. However, this is also true for $v_{i+1} \wedge \cdots \wedge v_n$. We conclude that $v_{i+1}, \dots, v_n$ and $w_{i+1}, \dots, w_n$ have the same span, so in particular, the span of $w_{i+1}, \dots, w_n$ is stable under A.

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

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

where B accounts for the first i Hodge and Newton slopes of A.

Each $i \times i$ minor of the matrix $(B\,C)$ has valuation at least the sum of the first i Hodge slopes of A, which is the valuation of $\det(B)$. By Cramer's rule, each row of C is a $\mathfrak{o}_F$-linear combination of the rows of B, i.e., CB^{-1} has entries in $\mathfrak{o}_F$. Moreover, the first Hodge slope of D is greater than the last Hodge slope of B, so $|DCB^{-1}| < |C|$. Thus conjugating by the matrix

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

gives a new matrix

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

with $|C_1| < |C|$. Repeating, 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. □

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 4.4.8 below.

LEMMA 4.3.12. *Suppose that one of the following holds.*

(a) *The value group of $|F^*|$ is dense in $\mathbb{R}_{>0}$, and $\eta > 1$.*

(b) *We have $|\lambda_i| \in |F^*|$ 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 4.4.8 below. Case (b) is directly analogous to Lemma 4.1.11. □

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 4.4.1. *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 4.4.2. *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 4.2.2, except now the factor $|2^n \binom{n}{i}^2|$ is dominated by 1. □

QUESTION 4.4.3. *Is Theorem 4.4.2 best possible?*

We may also consider multiplicative perturbations.

PROPOSITION 4.4.4. *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 4.2.3, but using the Smith normal form instead of the singular value decomposition. □

COROLLARY 4.4.5. *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 4.4.5. (We leave the archimedean analogue to the reader's imagination.)

PROPOSITION 4.4.6. *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

$$\max_i \{|c_i|\} \leq \sigma_1^{n-1}. \tag{4.4.6.1}$$

It suffices to check (4.4.6.1) 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. □

Example 4.4.7. 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 4.4.6 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 4.4.6.

Corollary 4.4.8. *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 4.4.6 (after appropriate rescaling), Proposition 4.4.4, and Theorem 4.3.11. One must also separately treat the case where $\sigma_1 = |\lambda_1| = 1$, but 1 occurs as a singular value more times than it occurs as the norm of an eigenvalue; we leave this to the reader. □

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 4.4.9. *In Corollary 4.4.8, 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 4.2.4, we obtain the following.

Proposition 4.4.10. *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

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

For $(I, J, K) \in U^n_r$, write $I = \{i_1 < \cdots < i_r\}$ and similarly for J, K. For $r = 1$, put $T^n_1 = U^n_1$. For $r > 1$, put

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

For multiplicative perturbations, we obtain the following results. which include the Weyl inequalities (Theorem 4.1.7, Theorem 4.3.8) as well as Propositions 4.2.3 and 4.4.4. It is important for the proofs that one can rephrase the Horn inequalities in terms of Littlewood-Richardson numbers; see [**Ful00**, §3].

Theorem 4.5.1. *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 $(I, J, K) \in T^n_r$, 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. $\square$

Theorem 4.5.2. *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

See [**Bha97**, §III] for results in the archimedean case not otherwise cited, such as the fact that a real symmetric matrix has nonnegative real eigenvalues, and the singular value decomposition. (This book was out of the library when I wrote this, so I wasn't able to look up precise references.) We unfortunately cannot recommend a good reference for the strong analogy between archimedean and p-adic numerical analysis; this seems to be a poorly known piece of folklore.

In Theorem 4.1.10, 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 4.3.11) is a direct translation of Katz's corresponding theorem for F-crystals [**Kat79**, Theorem 1.6.1]. The archimedean version (Theorem 4.1.9) is itself a translation of Theorem 4.3.11; we do not know of a reference, although we do not make any claim of originality. Likewise, Proposition 4.4.10 is a direct translation of [**Kat79**, Corollary 1.4.4]; its archimedean analogue (Proposition 4.2.4) 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 4.4.1.

(2) With notation as in Theorem 4.3.11, 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 4.4.6. 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 5

Formalism of differential algebra

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

1. Differential rings and differential modules

DEFINITION 5.1.1. 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 5.1.2. 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 5.1.3. 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 5.3.3 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.

2. Differential modules and differential systems

DEFINITION 5.2.1. 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 $v = v_1e_1 + \cdots + v_ne_n$ for some $v_1, \dots, v_n \in R$, and then compute

$$D(v) = v_1D(e_1) + \cdots + v_nD(e_n) + d(v_1)e_1 + \cdots + d(v_n)e_n.$$

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

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

(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 5.2.2. 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 5.3.1. 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 5.3.2. 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 $R^\vee$ 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 5.3.3. *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* $\mathrm{Ext}(M, N)$.

Proof. The group $\mathrm{Ext}(M, N)$ 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 $\mathrm{Hom}(M, M)$) to get an element of $H^1(M^\vee \otimes N)$. This is the desired map $\mathrm{Ext}(M, N) \to H^1(M^\vee \otimes N)$. 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 5.4.1. 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 $m, D(m), \ldots, D^{n-1}(m)$ form a basis of M.

Theorem 5.4.2 (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 $u, D(u), \ldots, D^j(u)$). 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 $m, D(m), \ldots$ 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

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

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

$$\sum_{i=0}^{\mu} m \wedge \cdots \wedge D^{i-1}(m) \wedge D^i(z) \wedge D^{i+1}(m) \cdots \wedge D^\mu(m). \tag{5.4.2.1}$$

Pick $s \in \mathbb{Z}$, substitute $x^s z$ for z in (5.4.2.1), divide by x^s, and set equal to zero. We get

$$\sum_{i=0}^{\mu} s^i \Lambda_i(m, z) = 0 \qquad (s \in \mathbb{Z}) \tag{5.4.2.2}$$

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 (5.4.2.2), 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 $m, D(m), \ldots$ is always at most $\mu - 1$, contradicting the definition of μ. $\square$

REMARK 5.4.3. 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 5.5.1. 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 5.5.2 (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 5.5.3 (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 5.5.4. 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 5.5.5. 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 5.6.1. 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 5.8.1. 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 $t \mapsto t + r$.) 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 a later unit.) 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 5.5.2).

CHAPTER 6

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 6.1.1. 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

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

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 6.1.2. 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 6.1.3 (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 6.1.4. *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 6.1.5. 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 6.2.1. 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 6.2.2. 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 6.2.3. 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 2.3.3 any two norms on V compatible with the norm on F are equivalent.

Lemma 6.2.4. *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,

$$|d^s(a)v|_V \leq \max_{0 \leq i \leq s}\{|D^{s-i}(aD^i(v))|\}. \tag{6.2.4.1}$$

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 (6.2.4.1), 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 6.2.5. *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*

$$|D|_{\mathrm{sp},V} = \max\{|d|_{\mathrm{sp},F}, \limsup_{s\to\infty} |D_s|^{1/s}\}. \tag{6.2.5.1}$$

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 6.2.4 implies that the left side of (6.2.5.1) 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

$$|D^s|_V^{1/s} \leq \max_{0\leq j\leq s}\{|d^j|_F^{1/s} |D_{s-j}|^{1/s}\}. \tag{6.2.5.2}$$

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

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

Then (6.2.5.2) 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 (6.2.5.1) 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 6.2.6. With slightly more work, one may check that in Lemma 6.2.5, if the maximum is only achieved by the second term, then you can replace the limit superior by a limit.

LEMMA 6.2.7. *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 6.2.8. *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 6.6.3).

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 6.2.7. 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 6.2.9. 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 6.2.10. 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 4.4.6 later.

Proposition 6.2.11. *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 4.4.6, 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 4.3.4, 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 6.3.1 (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*

$$|D|_{\mathrm{sp},V} = |d|_{\mathrm{sp},F} \lim_{s \to \infty} |D_s|^{1/s}. \tag{6.3.1.1}$$

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 6.2.5. □

4. Newton polygons for twisted polynomials

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

DEFINITION 6.4.1. 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 6.4.2. *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 6.4.3. 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 6.4.2. (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 3.2.2), we obtain the following.

THEOREM 6.4.4. *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 3.2.1. □

COROLLARY 6.4.5. *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 6.5.1. 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 6.5.2. 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 6.5.3 (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 4.3.10, 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. □

COROLLARY 6.5.4. *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 6.4.5, each factor in the factorization has only a single slope, so Theorem 6.5.3 gives what we want. □

6. The visible decomposition theorem

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

THEOREM 6.6.1 (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 5.4.2), 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 6.4.4 once to P, we obtain a short exact sequence $0 \to V_1 \to V \to V_2 \to 0$ in which (by Theorem 6.5.3) 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 6.4.4 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 6.4.2). 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 6.6.2. *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 6.6.1, 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 6.6.3. *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 6.6.1; 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 6.2.8. $\square$

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 a later unit.

7. Matrices and the visible spectrum

The proof of Theorem 6.5.3 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 6.7.1. *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 6.5.3.

LEMMA 6.7.2. *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 6.5.3, 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 6.6.1, it follows that every irreducible subquotient of V also has spectral norm σ, as desired. $\square$

LEMMA 6.7.3. *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 4.4.2 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 6.7.2 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 6.7.1 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 6.7.4. *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 4.4.8 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 4.4.8, 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 4.4.2, 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 6.7.3 to obtain the desired result. $\square$

Notes

Lemma 6.2.5 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 6.2.11 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 17.2.1 and its proof.

Proposition 6.3.1 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 3.2.2 follows [**Chr83**].

The proof of Theorem 6.5.3 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 $1 \Longrightarrow 2$; 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 6.1.3).

(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 7

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 7.1.1. 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 6.6.1. 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 7.1.2. *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 5.4.2, so the criterion in the theorem always applies. □

Corollary 7.1.3. *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 7.1.2, (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 7.1.4. 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 7.2.1. *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. $\square$

Remark 7.2.2. In the p-adic setting, we will see that the first step of the proof of Theorem 7.2.1 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 7.2.3. 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 7.2.1 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 7.2.4. *In Definition 7.2.3, 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 7.2.5. In Definition 7.2.3, 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 7.3.1. 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 $U^{-1}NU + U^{-1}z\frac{d}{dz}U = N_0$.

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

DEFINITION 7.3.2. 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 7.3.3. *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:

$$iU_i = U_iN_0 - N_0U_i - \sum_{j=1}^{i} N_iU_{j-i} \qquad (i > 0). \tag{7.3.3.1}$$

Viewing the map $X \mapsto XN_0 - N_0X$ 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 $X \mapsto iX - XN_0 + N_0X$ 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 (7.3.3.1); this proves the desired result. □

Theorem 7.3.4 (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 $\mathbb{C}\{z\}$.*

Proof. See [**DGS94**, §III.8, Appendix II]. □

Corollary 7.3.5. *With notation as in Theorem 7.3.4, 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 7.2.1.) □

Remark 7.3.6. The p-adic analogue of Theorem 7.3.4 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 7.3.7 (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. □

Corollary 7.3.8 (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 7.4.1. 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 7.4.2. *For any finite differential module V over $\mathbb{C}((z))$, $H^0(V) = H^1(V) = 0$.*

PROOF. Exercise. $\square$

In the convergent case, the index carries more information.

THEOREM 7.4.3. *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 7.3.4). 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 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 7.3.7.

(3) Prove Proposition 7.4.2.

Part 3

p-adic differential equations on discs and annuli

CHAPTER 8

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 8.0.1. 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 8.1.1. 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 8.1.2. 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 8.2.1. 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 3.1.3.) 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 8.2.2. 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 8.2.3. *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. □

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

Proposition 8.2.4. *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 8.2.3, it suffices to check the Cauchy property at each nonzero endpoint of I.)*

Proof. Exercise. □

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

LEMMA 8.2.5. *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 8.2.4, and its limit is an inverse of f. □

3. Factorization results

PROPOSITION 8.3.1 (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 3.2.1, this is a consequence of the master factorization theorem (Theorem 3.2.2); the completeness of the ring is provided by Property 8.2.4. □

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

PROPOSITION 8.3.2 (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 8.3.1, 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 3.2.2 to factor $ef = gh$ in R, in which g is a unit in $K\langle t/\beta \rangle$ by Lemma 8.2.5. Now $h \sum_{i=0}^m f_i t^i = f g^{-1}$ 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 8.3.3 (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 3.2.2 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. $\square$

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 9).

Definition 8.4.1. 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 8.3.3, we check that the construction is canonically independent of the choice of the sequence.

Notes

The Hadamard three circles theorem (Proposition 8.2.3(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 8.3.3) 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 8.4.1 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 [**Ked05**, Theorem 3.14].

Exercises

(1) Prove Proposition 8.2.4. (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 [**Ked04**, Lemma 6.2].

CHAPTER 9

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 9.1.1. 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 9.1.2. *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. □

COROLLARY 9.1.3. *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 9.1.2, the image of this map must be a direct summand of M as a module, proving the claim. □

COROLLARY 9.1.4. *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 9.2.1. 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 9.2.2. 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 9.2.3 (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 9.3.4). We will give a direct proof of a slightly stronger result later (Proposition 17.1.1); see also the notes. □

Here are some easy consequences of the definition of radius of convergence; note the parallels with properties of the spectral norm (Lemma 6.2.7).

LEMMA 9.2.4. *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 6.2.7(c). □

EXAMPLE 9.2.5. 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 9.3.1. 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 9.3.2. 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, and the residue field of $\mathcal{E}$ is equal to $\kappa_K((t))$; however, neither of these need hold 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 9.3.3. 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 9.5.4) that this does indeed compute the radius of convergence of horizontal sections of V on a "generic disc".

In the interim, we note the following relationship with the usual radius of convergence. 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 9.2.3, is an immediate corollary.)

Theorem 9.3.4. *For any nonzero finite differential module M over $K\langle t/\rho\rangle$, $R(M) \geq R(M \otimes F_\rho)$. That is, the radius of convergence is at least the generic radius.*

Proof. Suppose $\lambda < \rho$ and $\lambda < p^{-1/(p-1)}|D|^{-1}_{\mathrm{sp},V}$. We claim that for any $x \in M$, the Taylor series

$$y = \sum_{i=0}^{\infty} \frac{(-t)^i}{i!} D^i(x) \tag{9.3.4.1}$$

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. □

We can translate some basic properties of the spectral norm into properties of generic radii, leading to the following analogue of Lemma 9.2.4. Alternatively, one can first check Proposition 9.5.4 and then simply invoke Lemma 9.2.4 itself around a generic point.

LEMMA 9.3.5. *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 9.3.6. 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 9.5.5 and the notes for some reasons why the intrinsic radius deserves such a name.)

REMARK 9.3.7. 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 9.4.1. 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 9.4.1 as follows.

PROPOSITION 9.4.2. *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 9.4.1, 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. $\square$

5. Geometric interpretation

As promised, here is a construction that explains the name "generic radius of convergence".

Definition 9.5.1. 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 9.5.2. 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. Hence we get an isometry $F_\rho \to L\langle (t-t_\rho)/\rho\rangle$.

Remark 9.5.3. 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 9.5.4. *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 6.2.5, 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 9.3.4 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 9.5.5. *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, $IR(V) = IR(V \otimes F_{\rho^m})$.*

PROOF. This follows from the geometric interpretation plus the fact that

$$|t - t_\rho| < c\rho \Leftrightarrow |t^m - t_\rho^m| < c\rho^m \qquad (c \in (0,1)), \tag{9.5.5.1}$$

whose proof is left as an exercise. ☐

REMARK 9.5.6. 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.

6. Subsidiary radii

It is sometimes important to consider not only the generic radius of convergence, but also some secondary invariants.

DEFINITION 9.6.1. 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 *generic radii of subsidiary 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 (generic) radii of subsidiary convergence* obtained by multiplying the subsidiary radii by ρ^{-1}.

REMARK 9.6.2. 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 9.6.3. 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 11.10.1).

Notes

According to [**DGS94**, Appendix III] (which see for more information), the p-adic Cauchy theorem (Proposition 9.2.3) was originally proved by Lutz [**Lut37**]. See Proposition 17.1.1 for a related result.

The notion of restricting a p-adic differential module to a generic disc originates in the work of Dwork [**Dwo73**]. 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 [**Ked07**], where it is called the "toric normalization" in light of Proposition 9.5.5. It also figures prominently in the coordinate-free treatment of generic radii of convergence given in [**BdV07**]; the coordinate-free aspect justifies the modifier "intrinsic".

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 in terms of subsidiary radii, in a refined form, in a later chapter (Theorem 11.10.1).

Exercises

(1) Prove Proposition 9.1.2. (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 9.2.5 by giving an explicit formula for $IR(V_\lambda)$, in terms of ρ and the minimum distance from λ to an integer.
(4) Prove Example 9.4.1. (Hint: for (a), consider the cases $\lambda \in \mathbb{Z}_p$, $\lambda \in \mathfrak{o}_K - \mathbb{Z}_p$, and $\lambda \notin \mathfrak{o}_K$ separately. For (b), use (a) to reduce to the case $\lambda \in \mathbb{Z}_p$.)
(5) Prove (9.5.5.1).

CHAPTER 10

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 10.0.1. 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?

REMARK 10.2.1. We observed previously (9.5.5.1) that when m is a positive integer coprime to p,

$$|t-\eta| < \lambda|\eta| \Leftrightarrow |t^m - \eta^m| < \lambda|\eta|^m \qquad (\lambda \in (0,1)).$$

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 10.2.2. *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 10.2.3. *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 10.3.1. 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 10.3.2. *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 10.2.3(a), $V' \otimes L\langle (t - t_\rho)/(\min\{\lambda^{1/p}, p\lambda\}\rho) \rangle$ does likewise. $\square$

Remark 10.3.3. The inequality in Lemma 10.3.2 can be strict; see for instance Definition 10.3.5.

Definition 10.3.4. 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 10.3.5. 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 polynomial 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 10.3.6. (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 10.4.1. 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 $V \cong \varphi^* V'$ and $IR(V') > p^{-p/(p-1)}$.

Theorem 10.4.2 (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 10.4.2. 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 10.2.3(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 10.3.6, 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 10.3.2 gives the reverse inequality. $\square$

Corollary 10.4.3. *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 $IR(V') = IR(\varphi^* V')^p$.*

S The construction of Frobenius antecedents carries over to discs and annuli as follows.

THEOREM 10.4.4. *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 10.4.2. (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 10.4.2 for any single $\rho \in [\alpha, \beta]$. □

5. Frobenius descendants and subsidiary radii

We saw in Lemma 10.3.2 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 10.5.1. *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 10.4.2); note that V' is also irreducible. By Lemma 10.3.6, $\varphi_* V \cong \oplus_{m=0}^{p-1}(V' \otimes W_m)$. 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 10.2.3(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 10.3.2 gives

(10.5.1.1)

$$IR(\varphi^* W') \geq \min\{IR(W')^{1/p}, pIR(W')\} \geq \min\{IR(\varphi_* V)^{1/p}, pIR(\varphi_* V)\} \geq IR(V).$$

On the other hand, $\varphi^* W'$ is a subquotient of $\varphi^* \varphi_* V$, which by Lemma 10.3.6 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 (10.5.1.1) 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. □

Remark 10.5.2. One might be tempted to think that the proof that $IR(\varphi_* V) \geq p^{-1} IR(V)$ in the proof of Theorem 10.5.1 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 10.5.3. *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 10.6.1. *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 10.5.1, 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 10.3.6,

$$\begin{aligned} \varphi_* V_1 \otimes \varphi_* V_2 &\cong \oplus_{m=0}^{p-1} (\varphi_* V_1 \otimes W_m \otimes V_2') \\ &\cong (\varphi_* V_1 \otimes V_2')^{\oplus p}. \end{aligned}$$

(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 10.3.6), $H^1(\varphi_*(V_1 \otimes V_2)) = 0$. By Lemma 10.3.6 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 10.4.2 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. □

From here, the proof of the following theorem is purely formal.

Theorem 10.6.2 (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 10.6.1 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. $\square$

As with the original decomposition theorem, we obtain the following corollaries.

Corollary 10.6.3. *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 10.6.4. *With $V = \oplus_{s \in (0,1]} V_s$ as in Theorem 10.6.2, 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 a later unit.

Corollary 10.6.5. *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 10.6.2; 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 a result from a previous unit. $\square$

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 10.7.1. *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 10.5.1. $\square$

The exponent p^{-h} is not spurious; here is an example to illustrate why it cannot be removed.

EXAMPLE 10.7.2. 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
$$R(M' \otimes F'_\rho) = |\lambda|^{-1}\rho^{2p}$$
$$R(M \otimes F_\rho) = |\lambda|^{-1/p}\rho^2,$$
where the first equality follows by Theorem 6.5.3 and the second follows from the first by Corollary 10.4.3.

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, as follows.

DEFINITION 10.8.1. 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 10.2.2 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 10.8.2. *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 10.8.3. *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 10.8.3 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 10.2.2 is taken from [**Ked05**, §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 10.4.2 is [**Ked05**, 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 10.5.1, Theorem 10.8.3, and the strong decomposition theorem (Theorem 10.6.2).

Exercises

(1) Prove Lemma 10.3.6.
(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 10.7.2. 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 11

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 10.0.1.

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 8.2.3(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 11.1.1. 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 11.1.2. *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 11.1.3. Note that $s_{\overline{\mu}}(f) \geq 0$ for $\overline{\mu} \neq 0$; thus Proposition 11.1.2 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 11.2.1. *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) *(Concavity) Suppose that P is monic. For $i = 1, \dots, n$, the function $F_i(P,r)$ is concave.*
(d) *(Superharmonicity) Suppose that P is monic and 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*
$$s_{\infty,i}(P) \geq \sum_{\overline{\mu} \in \kappa_K^{\mathrm{alg}}} s_{\overline{\mu},i}(P).$$
(e) *(Monotonicity) Suppose that P is monic and that $\alpha = 0$. For $i = 1, \dots, n$, the slope of $F_i(P,r)$ is nonnegative.*

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 8.2.3(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,

$$F_i(P,r) = \sup_s \{v_{s,r}(P) - (n-i)s\} - v_r(P_n). \tag{11.2.1.1}$$

Moreover, the supremum in (11.2.1.1) 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,

$$F_i(P,r) = v_r(P_{n-i}) - v_r(P_n), \tag{11.2.1.2}$$

proving (b).

Assume that P is monic, so that $P_n = 1$ and (11.2.1.1) 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 (11.2.1.1) 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 (c).

Assume that $\alpha < 1 < \beta$. Then Proposition 11.1.2 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 $f_i(0) < f_{i+1}(0)$, then this plus (11.2.1.2) 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

$$F_i(P, r) \geq \frac{k-i}{k-j} F_j(P, r) + \frac{i-j}{k-j} F_k(P, r), \tag{11.2.1.3}$$

with equality for $r = 0$. From this plus piecewise affinity, we deduce (d).

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 (11.2.1.1), proving (e). $\square$

Remark 11.2.2. 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 11.2.3. It should also be noted that if $i = n$ or $f_i(r_0) < f_{i+1}(r_0)$, then (11.2.1.2) 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 10.7.1 and Example 10.7.2.

3. Variation of subsidiary radii: statements

In order to state the analogue of Theorem 11.2.1 for subsidiary radii of a differential module on a disc or annulus, we must set some corresponding notation.

Notation 11.3.1. 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 $R_1(\rho), \dots, R_n(\rho)$ 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 11.6.1, 11.7.1, 11.7.3, 11.8.1). Note that there is an overall sign discrepancy from Theorem 11.2.1, so that concavity becomes convexity and so forth.

Theorem 11.3.2. *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) *(Convexity) For $i = 1, \dots, n$, the function $F_i(M, r)$ is convex.*
(d) *(Subharmonicity) Suppose that $\alpha < 1 < \beta$ and that $f_i(0) > 0$. 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*
$$s_{\infty,i}(M) \leq \sum_{\overline{\mu} \in \kappa_K^{\mathrm{alg}}} s_{\overline{\mu},i}(M).$$
(e) *(Monotonicity) Suppose that $\alpha = 0$. For $i = 1, \dots, n$, for any point r_0 where $f_i(r_0) > r_0$, 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.)*

4. Convexity for the generic radius

As a prelude to tackling Theorem 11.3.2, we give a quick proof of convexity, subharmonicity, and monotonicity (parts (c)-(e) of Theorem 11.3.2) 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 11.3.2 for many purposes; indeed, this is true for numerous results which predate Theorem 11.3.2. See the notes for further details.

Proof of Theorem 11.3.2(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 6.2.5 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 8.2.3(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 (d) by applying Proposition 11.1.2 to each D_s. If $\alpha = 0$, then the function $r \mapsto -\log |D_s|_{e^{-r}}^{-1/s}$ is nonincreasing, yielding (e). □

Remark 11.4.1. 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 11.5.1 (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. □

REMARK 11.5.2. Although we will only apply Lemma 11.5.1 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 11.4.1.

LEMMA 11.6.1. *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 11.3.2 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 11.2.1 to deduce (a) and (b).

To deduce (c), (d), and (e), 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 4.3.10, 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 11.5.1, 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 6.7.4 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 (c), (d), (e) from Theorem 11.2.1. $\square$

7. Larger radii

We next address concern (b) from Remark 11.4.1, considering the cases $f_i(M, r_0) > r_0$ and $f_i(M, r_0) = r_0$ separately.

LEMMA 11.7.1. *For any $i \in \{1, \dots, n\}$ and any r_0 such that $f_i(M, r_0) > r_0$, clauses (a)-(d) of Theorem 11.3.2 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 11.6.1. As in the proof of Lemma 11.6.1, 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 10.5.1, 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)-(d), 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 11.2.1(d), 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

$$g_1(pr) + \cdots + g_{pi}(pr) = pF_i(M,r) + pi\log p + (p-1)ipr, \tag{11.7.1.1}$$

or $f_{i+1}(0) < 1/(p-1)\log p$ or $i=n$, in which case in some neighborhood of $r=0$ we have

$$g_1(pr) + \cdots + g_{pi+(p-1)(n-i)}(pr) = pF_i(M,r) + pn\log p + (p-1)npr. \tag{11.7.1.2}$$

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 (11.7.1.1) and the induction hypothesis to write piecewise

$$\begin{aligned}F_i(r) &= p^{-1}(g_1(pr) + \cdots + g_{pi}(pr) + pi\log p + (p-1)ipr)\\ &= p^{-1}(m(pr) + *)\\ &= mr + p^{-1}*\end{aligned}$$

for some $m \in \mathbb{Z}$. (Note that $*$ is not guaranteed to be in $p \cdot v(K^\times)$; this explains Example 10.7.2.) 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 (11.7.1.2) to write piecewise

$$\begin{aligned}F_i(r) &= p^{-1}(g_1(pr) + \cdots + g_{pi+(p-1)(n-i)}(pr) + pn\log p + (p-1)npr)\\ &= p^{-1}(m(pr) + *)\\ &= mr + p^{-1}*\end{aligned}$$

for some $m \in \mathbb{Z}$. $\square$

Remark 11.7.2. In the proof of Lemma 11.7.1, note the importance of the fact that the domains of applicability of (11.7.1.1) and (11.7.1.2) overlap: if $f_i(M,0) = 1/(p-1)\log p$, then (11.7.1.1) may not remain applicable when we move from $r=0$ to a nearby value.

Lemma 11.7.3. *For any $i \in \{1,\dots,n\}$ and any r_0 such that Theorem 11.3.2 holds in a neighborhood of r_0.*

Proof. As in the proof of Lemma 11.6.1, it suffices to consider the case $r_0 = 0$. We first check continuity. For this, note that the proofs of Lemma 11.6.1 and 11.7.1 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 11.7.1 with i replaced by j.

We next check (c) 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 (c).

Finally, note that (d) and (e) make no assertion at $r = 0$ in case $f_i(0) = 0$, so we are done. □

8. Monotonicity

To complete the proof of Theorem 11.3.2, we must prove (e) without the restriction $f_i(r_0) > r_0 + 1/(p-1)\log p$. The reason why we do not have (e) as part of Lemma 11.7.1 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 11.8.1. *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 11.7.1, 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 10.8.3, 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 11.7.1 to conclude. □

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 11.9.1. *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 9.3.4, 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. □

Corollary 11.9.2. *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 11.3.2 and Theorem 10.7.1, 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 11.9.1, 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 11.9.3. *Let M be a differential module over $K\langle t/\beta \rangle$ for some $\beta > 0$. Does there necessarily exist $j \in \{1, \dots, \operatorname{rank}(M)\}$ 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 11.9.4. *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 generic radii of subsidiary 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 11.9.4.

Theorem 11.10.1 (after Young). *Let (V, D) be a differential module over F_ρ of dimension n with subsidiary radii $s_1 \leq \cdots \leq s_n$. Choose a basis $e_1, \dots, e_n$ of local horizontal sections of V at a generic point η. For $i = 1, \dots, n$, let ρ_i be the radius of convergence of e_i, and suppose that $\rho_1 \leq \cdots \leq \rho_n$. Then $\rho_i \leq s_i$ for $i = 1, \dots, n$; moreover, there exists a choice of basis for which $\rho_i = s_i$ for $i = 1, \dots, n$.*

Proof. We first produce a basis for which $\rho_i = s_i$ for $i = 1, \dots, n$. For this, we may apply Theorem 10.6.2 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 generic radius (Proposition 9.5.4), each Jordan-Hölder constituent of V admits a basis of local horizontal sections on a generic disc of radius s. By Lemma 6.2.7, 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 a generic point η can have radius of convergence strictly greater than s_1.

Put

$$f_i(r) = f_i(V_1 \otimes L\langle (t-\eta)/e^{-r}\rangle, r) \qquad (i = 1, \dots, m; r \in (-\log \rho, \infty 0));$$

then the $f_i(r)$ behave as in Theorem 11.3.2. By the proof of Theorem 11.3.2(d), the $f_i(r)$ are constant in a neighborhood of $r = -\log \rho$. By Theorem 11.3.2(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 η which converged on a closed disc of radius λ for some $\lambda \in (s_1, \rho)$, then $V_1 \otimes L\langle (t-\eta)/\lambda\rangle$ 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. $\square$

Remark 11.10.2. In Theorem 11.10.1, a basis of local horizontal sections for which $\rho_i = s_i$ for $i = 1, \dots, n$ is sometimes called an *optimal basis*.

Notes

The harmonicity property of functions on annuli (Proposition 11.1.2) 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 11.10.1 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 11.2.1, 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 $\kappa_K[t]$.

CHAPTER 12

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 10.0.1, we also retain Notation 11.3.1.

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 8.2.5). However, for the present application, we need an alternate characterization based on more intrinsic data, namely the Gauss norms.

DEFINITION 12.1.1. 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

$$\mathrm{disc}(f, 0) = \sum_{\overline{\mu} \in (\kappa_K^{\mathrm{alg}})^\times} s_{\overline{\mu}}(f);$$

note that $\mathrm{disc}(f, 0) \geq 0$. We define $\mathrm{disc}(f, r)$ 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

$$\mathrm{disc}(f, r) = \mathrm{disc}(T_c(f), 0).$$

LEMMA 12.1.2. *For $x \in K\langle t/\beta\rangle$ nonzero, x is a unit if and only if $s_0(x) = \mathrm{disc}(x, -\log \beta) = 0$.*

PROOF. We may reduce to the case $\beta = 1$ and $|x|_1 = 1$. In this case, by Lemma 8.2.5, 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 11.1.2, 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. □

For annuli, it is more convenient to prove a weak criterion first.

LEMMA 12.1.3. *For $x \in \cup_{\alpha \in (0,\beta)} K\langle \alpha/t, t/\beta\rangle$ nonzero, x is a unit if and only if $\mathrm{disc}(x, -\log \beta) = 0$.*

PROOF. We again reduce to the case $\beta = 1$ and $|x|_1 = 1$. In this case, by Lemma 8.2.5, 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 12.1.2. □

One may then deduce the following.

LEMMA 12.1.4. *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 $\mathrm{disc}(x, -\log \alpha) = \mathrm{disc}(x, -\log \beta) = 0$.*

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 12.1.3 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 12.2.1. 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

$$\mathrm{disc}_i(M, 0) = -\sum_{\overline{\mu} \in (\kappa_K^{\mathrm{alg}})^\times} s_{\overline{\mu}, i}(M);$$

it is always nonnegative. Extend the definition to general $r \in [-\log \beta, -\log \alpha]$ as in Definition 12.1.1.

Theorem 12.2.2. *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 $\mathrm{disc}_i(M, -\log \beta) = 0$.*

Then the decomposition of $M \otimes F_\beta$ separating the first i subsidiary radii lifts to a decomposition of M itself.

Before proving Theorem 12.2.2, we record a trivial but useful observation.

Lemma 12.2.3. *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 12.2.4. The immediate application of Lemma 12.2.3 is to replace K by a complete extension L in Theorem 12.2.2; 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 12.2.5. *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 5.5.2). 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. $\square$

Lemma 12.2.6. *Theorem 12.2.2 holds if* $f_i(-\log \beta) > 1/(p-1)\log p - \log \beta$.

Proof. By invoking Remark 12.2.4 to justifying enlarging K, then rescaling, we may reduce to the case $\beta = 1$. Set notation as in the proof of Lemma 11.6.1. 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 12.1.2, and we can apply Theorem 3.2.2 to factor $Q = Q_2 Q_1$ 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 12.2.5, Q_1 and Q_2 generate the unit ideal in $K\langle t \rangle[T]$.) Applying Lemma 11.5.1 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 6.7.1, 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. $\square$

To prove Theorem 12.2.2 in general, we must use Frobenius antecedents again.

Proof of Theorem 12.2.2. It suffices to prove that for $\beta = 0$, Theorem 12.2.2 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 12.2.6.

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

$$((M_1' \otimes W_0) \cap \cdots \cap (M_1' \otimes W_{p-1})) \oplus ((M_2' \otimes W_0) + \cdots + (M_2' \otimes W_{p-1}));$$

this is induced by a decomposition of M having the desired properties. $\square$

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 12.2.2.

THEOREM 12.3.1. *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 $\operatorname{disc}_i(M, -\log\beta) = \operatorname{disc}_i(M, -\log\alpha) = 0$.*

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 12.2.2.

LEMMA 12.3.2. *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 $\operatorname{disc}_i(M, -\log\beta) = 0$.*

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 12.2.4 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 12.2.2.

Set notation again as in the proof of Lemma 11.6.1. 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 12.1.3, so we may continue as in the proof of Lemma 12.2.6. □

To prove Theorem 12.3.1 from Lemma 12.3.2, we proceed as in the proof of Lemma 12.1.4.

PROOF OF THEOREM 12.3.1. Note that by subharmonicity (Theorem 11.3.2(d)), conditions (b) and (c) together are equivalent to the condition that $\operatorname{disc}_i(M,r) = 0$ 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 12.3.2 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 8.3.3), 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 12.4.1. *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 12.2.2 to $m \otimes K\langle t/\beta \rangle$; doing so for all such δ (or a sequence ascending to β) yields the desired result. □

Similarly, for open annuli, we obtain a decomposition theorem without a discrepancy condition at endpoints.

Theorem 12.4.2. *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 12.4.3. 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 12.3.1 and Theorem 12.4.2. 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 12.5.1. 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 12.5.2. *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 11.3.2. □

Definition 12.5.3. 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 12.5.2 will be called the *differential slopes* of M at β.

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 12.5.4 (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 12.5.2, we are in a case where Theorem 12.4.2 may be applied. □

REMARK 12.5.5. 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 14.

6. Notes

Our results on modules solvable at a boundary are originally due to Christol and Mebkhout [**CM00, CM01**]. In particular, Lemma 12.5.2 for the generic radius is [**CM00**, Théorème 4.2.1], and the decomposition theorem (which implies Lemma 12.5.2 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 [**Ked05**, 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 12.2.3 in which M is only required to be locally free.

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. $\square$

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 3.2.2) 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 $x \in F$ 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 $P_1 P_2$ 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 $P = P_1 P_2$ 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). $\square$

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. $\square$

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

$$F \xrightarrow{\phi} F \xrightarrow{\phi} \cdots .$$

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 $\Phi^{\dim_F V} = 0$.

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. □

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. □

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. □

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. □

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 $i = n$.

REMARK 13.5.5. 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 13.5.6. *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. □

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. □

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.

7. 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.

In the special case of the difference field $\mathrm{Frac}(W(k))$, with k perfect of characteristic $p > 0$, most of the results of this section appear in [**Kat79**]. However, it is awkward to give direct references since we have organized our presentation rather differently.

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**].

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**].

8. 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 the cyclic vector theorem for differential modules.)

(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 p, 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.)

CHAPTER 14

Frobenius structures on differential equations

CHAPTER 15

Quasiunipotent differential modules

1. Galois representations and differential modules

2. Swan conductors

(The attribution of this theorem is complicated; see the notes.)

THEOREM 15.2.1.

3. The unit-root monodromy theorem

We omit the proof of the following theorem here. For detailed references, see the notes.

THEOREM 15.3.1 (Tsuzuki). *Assume that K is discretely valued. Let M be a differential module over $\cup_{\alpha>0} K\langle \alpha/t, t \rrbracket_0$ equipped with a unit-root Frobenius structure. Then M is quasiunipotent.*

Notes

Theorem 15.2.1 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 [**Tsu98a**, Theorem 7.2.2] (by Brauer induction, as is possible in the ℓ-adic case) and Crew [**Crew**, 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 [**Ked05**, Theorem 5.23].

In the case of an imperfect residue field, it was originally suggested by Matsuda [**Mat04**] to formulate an analogue of Theorem 15.2.1 as follows. The analogue of the Swan conductor is the logarithmic conductor of Abbes-Saito [**AS02, AS03**]; it generalizes a definition for one-dimensional representations introduced by Kato [**Kto89**]. A suitable analogue of the differential conductor was described by Kedlaya [**Ked07a**]. The equality between the two was established by Chiarellotto and Pulita [**CP07**] for one-dimensional representations, and by Xiao [**Xia07**] in the general case.

Tsuzuki's original proof of Theorem 15.3.1 is [**Tsu98**, Theorem 5.1.1] (he assumes κ_K is algebraically closed, but it is easy to reduce to this case). Since then, several variants have appeared, although the general technique is essentially the same as in Tsuzuki's original work. Christol [**Chr01**] gave a reinterpretation in terms of Frobenius antecedents, but there is a minor gap in exposition: Christol assumes at all points that Frobenius has the standard form $t \mapsto t^p$, but as this property is not stable under extensions, one must perform a change of Frobenius after making an extension. A proof in a similar spirit (proving a slightly more general result) appears in [**Ked07d**, Theorem 4.5.2].

CHAPTER 16

The slope filtration theorem

In this chapter, we state the slope filtration for Frobenius modules over the Robba ring, and briefly discuss its applications in the theory of p-adic differential equations.

Part 5

Refined convergence information

CHAPTER 17

Effective bounds for nilpotent singularities

In this chapter, we discuss some effective bounds on the solutions of p-adic differential equations wih nilpotent singularities. 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 17.1.1. *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 7.3.3) satisfies*

$$|U_i| \leq |i!|^{-2m+1} \max\{|N_j| : 0 \leq j \leq i\} \qquad (i = 1, 2, \dots). \tag{17.1.1.1}$$

Consequently, U has entries in $K\langle t/(p^{-(2m-1)/(p-1)}\beta)\rangle$.

Note that this reproves the p-adic Cauchy theorem (Proposition 9.2.3).

PROOF. Recall (7.3.3.1):

$$N_0 U_i - U_i N_0 + iU_i = -\sum_{j=0}^{i-1} N_j U_{i-j}.$$

The map $f(X) = NX - XN$ on $n \times n$ matrices is nilpotent with nilpotency index $2m-1$. Hence the inverse of 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. □

We also need a nilpotent version of Dwork's transfer theorem (Theorem 9.3.4).

PROPOSITION 17.1.2. *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. Then for any $\rho < R(M \otimes F_\beta)$, there exists a basis of $M \otimes F_\rho$ on which D acts via a nilpotent matrix with entries in K.*

Proof. Pick any $\gamma \in (\rho, R(M \otimes F_\beta))$. Let e be the nilpotency index of the action of D on M/tM. We now verify that for any $x \in M$, the series

$$y(x) = \sum_{i=0}^{\infty} D^{e-1} \frac{(1-D)^e \cdots (i-D)^e}{(i!)^e} x,$$

which converges in $M \otimes K\langle t/\gamma \rangle$ because $\gamma < R(M \otimes F_\beta)$, in fact represents a horizontal section of M. It suffices to check this in $M \otimes K[\![t]\!]$, where we can write everything in terms of a basis $e_1, \dots, e_n$ of M on which D acts via a nilpotent matrix N_0. Then D acts on $t^i e_1, \dots, t^i e_n$ via $N + i$, so if $i > 0$ and $j \geq i$, $(1-D)^e \cdots (j-D)^e$ annihilates $t^i e_1, \dots, t^i e_n$ because it contains the factor $(i-D)^e$.

We also note from the above argument that if we choose $x \in M$ whose image in M/tM is not killed by D^{e-1}, then $y(x) \neq 0$. Thus $M \otimes K\langle t/\gamma \rangle$ admits a nonzero horizontal section y.

To conclude, put $n = \operatorname{rank} M$, and note that for $i = 1, \dots, n$, we have as above that $H^0(\wedge^i M \otimes K\langle t/\gamma \rangle) \neq 0$. From this, it is an exercise to check that $M \otimes K\langle t/\gamma \rangle$ is a successive extension of trivial differential modules, from which one deduces that $M \otimes K\langle t/\rho \rangle$ has a basis of the desired form. □

2. Effective bounds for solvable modules

We now give an improved version of Proposition 17.1.1 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 17.2.1. *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 $K\langle t/\beta \rangle$ and $|N|_\beta \leq 1$;*
(b) *$U_0 = I_n$;*
(c) *$U^{-1} N U + U^{-1} t \frac{d}{dt}(U) = N_0$;*
(d) *N_0 is nilpotent;*
(e) *U and U^{-1} have entries in $K\langle t/\beta \rangle$.*

Then for every nonnegative integer i,

$$|U_i|\beta^i \leq p^{(n-1)\lfloor \log_p i \rfloor}.$$

The argument amounts to using Frobenius antecedents to reduce the index from i to $\lfloor i/p \rfloor$, the key being the following lemma. (We will see in a later unit how to improve this lemma in the presence of a Frobenius structure.)

Lemma 17.2.2. *With notation as in Theorem 17.2.1, 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 17.2.1, such that*

$$\max\{|U_j|(\lambda\beta)^j : 0 \leq j \leq i\} \leq \mu p^{n-1} \max\{|U'_j|(\lambda\beta)^{pj} : 0 \leq j \leq i/p\}.$$

One then deduces Theorem 17.2.1 by induction on i, with base case the fact that $|U_i| \leq 1$ for $i \leq p-1$ by Proposition 17.1.1.

Proof of Lemma 17.2.2. Throughout this proof, let d be the derivation $t\frac{d}{dt}$. We first consider the case $|N|_\beta \leq 1$. 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 $N_W = W^{-1}NW + W^{-1}d(W)$, and let V_i be the unique solution of the matrix equation

$$N_0 V_i - V_i N_0 + i V_i = -(N_W)_i.$$

By induction on i, $|V_i|\beta^i \leq 1$ for all i.

By construction, $N'' = V^{-1}NV + V^{-1}d(V)$ has entries in $K[\![t^p]\!] \cap K\langle t/(\lambda^{1/2}\beta)\rangle = K\langle t^p/(\lambda^{p/2}\beta^p)\rangle$, and $|N''|_{\lambda^{1/2}\beta} \leq 1$. (Although it is also true that $|N''|_\beta \leq 1$, we can't ensure that the coefficient of t^i in N'' converges to zero under $|\cdot|_\beta$ as $i \to \infty$, so the entries of N'' might not belong to $K\langle t/\beta\rangle$.) Put $U'' = V^{-1}U$, so that $|U''|_{\lambda^{1/2}\beta} = 1$; then

$$(U'')^{-1}N''U'' + (U'')^{-1}d(U'') = N''_0 = N_0,$$

which forces U'' also to have entries in $K\langle t^p/(\lambda^{p/2}\beta^p)\rangle$. Let $\phi : K[\![t]\!] \to K[\![t]\!]$ denote the substitution $t \mapsto t^p$. We would like to take $N' = p^{-1}\phi^{-1}(N'')$ and $U' = \phi^{-1}(U'')$, but we only have $|N'|_{\lambda^{p/2}\beta^p} \leq p$, so we have to modify things somewhat.

Let M' be the differential module over $K\langle t/(\lambda^{p/2}\beta^p)\rangle$, with a basis on which D acts via $p^{-1}\phi^{-1}(N'')$, and let $|\cdot|$ be the supremum norm defined by this basis. Since M' has a basis of horizontal sections given by the column vectors of $\phi^{-1}(U'')$, by the nilpotent transfer theorem (Proposition 17.1.2), $M' \otimes F_{\lambda^{p/2}\beta^p}$ has generic radius of convergence $\lambda^{p/2}\beta^p$. In particular,

$$|D|_{\mathrm{sp}, M' \otimes F_{\lambda^{p/2}\beta^p}} \leq |d|_{F_{\lambda^{p/2}\beta^p}} = \lambda^{-p/2}\beta^{-p}.$$

By Proposition 6.2.11 (applied after an appropriate rescaling) plus the lattice lemma (Lemma 11.5.1), for any desired $\epsilon > 0$, we may find $W \in \mathrm{GL}_n(K\langle t/(\lambda^{p/2}\beta^p)\rangle)$ such that

$$\begin{aligned}
|W^{-1}|_{\lambda^{p/2}\beta^p} &\leq 1 + \epsilon \\
|W|_{\lambda^{p/2}\beta^p} &\leq p^{n-1}(1+\epsilon) \\
|W^{-1}p^{-1}\phi^{-1}(N'')W + W^{-1}d(W)|_{\lambda^{p/2}\beta^p} &\leq 1 + \epsilon.
\end{aligned}$$

We would now like to take

$$N' = W^{-1}p^{-1}\phi^{-1}(N'')W + W^{-1}d(W), \qquad U' = W^{-1}\phi^{-1}(U''),$$

but there are two problems still: this N' has norm bounded by $1+\epsilon$ rather than 1, and this U' need not have constant coefficient I_n.

The constant coefficient of the N' proposed above is $p^{-1}W^{-1}N_0W$, which is nilpotent; in particular, it has spectral norm 0. So we can find $X \in \mathrm{GL}_n(K)$ with

$$|X^{-1}| \leq 1, \qquad |X| \leq (1+\epsilon)^{n-1}, \qquad |p^{-1}X^{-1}W^{-1}N_0WX| \leq 1.$$

We instead take

$$\begin{aligned}
N' &= (WX)^{-1}p^{-1}\phi^{-1}(N'')(WX) + (WX)^{-1}d(WX), \\
U' &= (WX)^{-1}\phi^{-1}(U'')W_0X_0.
\end{aligned}$$

(The multiplication by W_0X_0 on the right side of U' ensures that U' has constant coefficient I_n.) We then have

$$|N'_0| \leq 1$$
$$|N'|_{\lambda^{p/2}\beta^p} \leq (1+\epsilon)^n.$$

If we take ϵ small enough that $(1+\epsilon)^n \leq \lambda^{-p/2}$, then these force $|N'|_{(\lambda\beta)^p} \leq 1$. Similarly,

$$|(WX)^{-1}|_{\lambda^{p/2}\beta^p} \leq 1$$
$$|WX|_{\lambda^{p/2}\beta^p} \leq p^{n-1}(1+\epsilon)^n$$

and if we take ϵ small enough that $(1+\epsilon)^n \leq \mu$, then these force $|(WX)^{-1}|_{(\lambda\beta)^p}|WX|_{(\lambda\beta)^p} \leq \mu p^{n-1}$. This yields the desired result. □

Using the technique appearing at the end of the proof of Theorem 17.2.1, one can loosen the bound on N as follows; we omit details.

THEOREM 17.2.3. *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,

$$|U_i|\beta^i \leq p^{(n-1)\lfloor \log_p i\rfloor} \max\{1, |N|_\beta^{n-1}\}.$$

We will often apply Theorem 17.2.1 through the following corollary (deduced by taking β to be an arbitrary value less than 1).

THEOREM 17.2.4. *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,

$$|U_i| \leq p^{(n-1)\lfloor \log_p i\rfloor}|N|_1^{n-1}.$$

It is easy to make an example that shows that one cannot, in general, improve this bound by much. 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 17.2.4 is close to optimal under its hypotheses, it can be improved in case the differential module in question admits a Frobenius structure. it is useful in practice to improve the bounds under additional hypotheses. One natural extra hypothesis is a Frobenius structure, since this is often what is needed to establish the hypothesis that a differential system defined on the open unit disc has solutions throughout the disc.

HYPOTHESIS 17.3.1. In this section, fix a power q of p, and let $\sigma : K[\![t]\!] \to K[\![t]\!]$ be a ring homomorphism of the form $\sum_i c_i t^i \mapsto \sum_i \sigma_K(c_i) t^{qi}$ with $\sigma_K : K \to K$ isometric.

The key here is to imitate the proof of Theorem 17.2.1 with the differential equation replaced by a certain Frobenius equation.

LEMMA 17.3.2. *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) $U^{-1} A \sigma(U) = A_0$.

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

$$U = A\sigma(U) A_0^{-1}.$$

This gives the first inequality, which implies the second as in the proof of Theorem 17.2.1, except that we iterate $\lceil \log_q i \rceil$ times to get to the case $i = 0$ (rather than to the case $0 < i < p$). $\square$

THEOREM 17.3.3. *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) $|N|_1 \leq 1$ *and* $|A|_1 \leq \infty$;
(b) $U_0 = I_n$ *and* A_0 *is invertible;*
(c) $U^{-1} N U + U^{-1} t \frac{d}{dt}(U) = N_0$;
(d) $NA + t\frac{d}{dt}(A) = pA\sigma(N)$.

Then $U^{-1} A \sigma(U) = A_0$, and for every nonnegative integer i,

$$|U_i| \leq p^{(n-1)((\log_p q)-1)} (|A_0^{-1}||A|_1)^{\lfloor \log_q i \rfloor}.$$

PROOF. We first note that

$$N_0 A_0 = pA_0 \sigma(N_0).$$

Choose an extension of σ_K to an isometric endomorphism of K^{alg}. Then if $\lambda_0 \in K^{\mathrm{alg}}$ were an eigenvalue of N_0, then so would be $\lambda_i = p\sigma(\lambda_{i-1})$ for $i = 1, 2, \dots$. Since $|\lambda_i| = p^{-1}|\lambda_{i-1}|$, the λ_i would all be distinct, which is impossible because N_0 cannot have infinitely many eigenvalues. From this contradiction, we deduce that N_0 is nilpotent.

Put $B = U^{-1}A\sigma(U) = \sum_{i=0}^{\infty} B_i t^i$. Then $B_0 = A_0$, and $N_0 B + t\frac{d}{dt}(B) = pB\sigma(N_0)$. Hence

$$N_0 B_i + iB_i = pB_i\sigma(N_0) = B_i A_0^{-1} N_0 A_0,$$

or

$$N_0(B_i A_0^{-1}) + i(B_i A_0^{-1}) = (B_i A_0^{-1})N_0. \tag{17.3.3.1}$$

As in the proof of Proposition 17.1.1, the operator $X \mapsto N_0 X - X N_0 + iX$ on $n \times n$ matrices is invertible for $i \neq 0$, so (17.3.3.1) 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 17.3.2 to reduce to the case $i < q$, then applying Theorem 17.2.4. □

We record a sharper form of Lemma 17.3.2 for use later in the discussion of logarithmic growth; the proof is analogous.

Proposition 17.3.4. *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 17.3.2. □

4. 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 17.4.1. Suppose the eigenvalues of N_0 are rational numbers with least common denominator dividing m. One can then apply Theorem 17.2.1 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 this bound and the bound in the nilpotent case is only a multiplicative constant factor.

Remark 17.4.2. Suppose that the eigenvalues of N_0 all belong to $\mathbb{Z}_p$. (One might want to consider this remark instead of Remark 17.4.1 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 17.2.1, but using shearing transformations to force the exponents to be multiples of p before each application of Lemma 17.2.2. However, the best known bound using this technique is worse than in Remark 17.4.1; it has the form $p^{(n^2+cn)\lfloor \log_p m \rfloor}$ 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 17.2.4 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 $1 \leq j_1 < \cdots < j_{n-1} \leq i$. (We suspect it is possible to recover this bound also using the proof technique of Theorem 17.2.1.) See also [**DGS94**, Theorem IV.3.1].

The general case of Theorem 17.2.1 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 Lemma 17.2.2 is played in [**CD91**] by an effective version of the cyclic vector theorem, which does not give the best possible element. 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]. (Note that this discrepancy is not the source of the poor estimate in the case of rational exponents without a bound on denominators.)

CHAPTER 18

Logarithmic growth

In this chapter, we investigate finer convergence questions around a differential module on a closed disc (possibly with nilpotent singularities) for which the local horizontal sections have maximal radius of convergence. Many of these questions were originally raised by Dwork, and only some have been answered.

1. Logarithmic growth for functions

DEFINITION 18.1.1. For $\delta \geq 0$, let $K[\![t]\!]_\delta$ be the subset of $K[\![t]\!]$ consisting of those $f = \sum_{i=0}^{\infty} f_i t^i \in K[\![t]\!]$ 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 $|\cdot|_\delta$. 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

$$K[\![t]\!]_{\delta+} = \bigcap_{\delta' > \delta} K[\![t]\!]_{\delta'}.$$

A useful alternate characterization is the following.

LEMMA 18.1.2. *For $\delta \geq 0$,*

$$K[\![t]\!]t_\delta = \{f \in K[\![t]\!] : \limsup_{\rho \to 1^-} \frac{|f|_\rho}{(-\log \rho)^\delta} < \infty.$$

PROOF. This is relatively straightforward once one observes the inequality

$$\sup_i \{(i+1)^\delta \rho^i\} \leq \rho^{-1} \left(\frac{\delta}{e}\right)^\delta (-\log \rho)^{-\delta}.$$

□

2. Logarithmic growth for differential modules

Before proceeding, we must recall the following consequence of Theorem 17.2.4.

PROPOSITION 18.2.1. *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.*

Definition 18.2.2. 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. Define the *log-growth filtration* on M to be the ascending filtration on $M \otimes K\{\{t\}\}[\log t]$ in which M_δ is the span of $H^0(M \otimes K[\![t]\!]_{\delta+})$. Define the *log-growth polygon* of M as the slope polygon in which δ occurs as a slope with multiplicity

$$\operatorname{rank} M_\delta - \operatorname{rank}(\cup_{\delta' < \delta} M_{\delta'}).$$

Remark 18.2.3. The log-growth polygon can behave rather pathologically. For instance, the slopes need not be rational [**CT06**, §5.2]. Also, if δ is a jump, then M_δ may or may not equal the span of $H^0(M \otimes K[\![t]\!]_\delta)$.

3. Generic log-growth

Definition 18.3.1. Recall that $\mathcal{E}$ was defined as the completion of $\mathfrak{o}_K((t)) \otimes_{\mathfrak{o}_K} K$ under the 1-Gauss norm. Let M be a finite differential module over $\mathcal{E}$; for η a generic point of norm 1 lying in a complete extension L of K, we can then base change M to $L[\![t-\eta]\!]_0$. We call the log-growth polygon of the resulting module the *generic log-growth polygon* of M.

Generic log-growth polygons behave much more simply than their nongeneric counterparts; witness for instance the following theorem of Christol [**Chr83**, Théorème 4.3.5].

Theorem 18.3.2 (Christol). *Let M be a finite differential module over E with $R(M) = 1$. Then M admits a filtration $0 = M_0 \subset \cdots \subset M_l = M$ such that each successive quotient is spanned by bounded horizontal sections.*

4. Generic versus special

Definition 18.4.1. 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 we can associate two log-growth polygons to M: the *special log-growth polygon*, obtained by restricting to the open unit disc, and the *generic log-growth polygon*, obtained by tensoring with E and applying the previous definition.

The following result was conjectured by Dwork [**Dwo73b**, Conjecture 2] and proved by André [**And07**].

Theorem 18.4.2 (André). *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. Assume that M is solvable at 1. Then the special log-growth polygon of M lies above the generic log-growth polygon, provided that the two are aligned to have the same* right *endpoint.*

Remark 18.4.3. In Theorem 18.4.2, it is not known in general whether the same inequality holds when the *left* endpoints are aligned. This would follow if the sum of the slopes of the two polygons were known to be equal, but it is not even clear whether this should be expected in general.

5. Log-growth and Frobenius

In the presence of a Frobenius structure, Dwork predicted a close relationship between log-growth and Frobenius slopes.

Conjecture 18.5.1. *Let M be an indecomposable differential module of rank n over $K[\![t]\!]_0$ for the operator $t\frac{d}{dt}$, equipped with a Frobenius structure. Assume that the lowest generic Frobenius slope of M is equal to* 0*. Then for $v \in V(M)$ and $\delta \geq 0$, $v \in M \otimes K[\![t]\!]_\delta[\log t]$ if and only if v belongs to $M' \otimes K\{\{t\}\}[\log t]$, where M' is the union of the steps of the Frobenius slope filtration of slope at most δ.*

This is known for $\mathrm{rank}(M) = 1$ trivially, and for $\mathrm{rank}(M) = 2$ by work of Chiarellotto and Tsuzuki [**CT06**, Theorem 7.2].

Using our effective bounds for convergence, we can prove part of Conjecture 18.5.1.

Theorem 18.5.2. *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. Assume that the lowest generic Frobenius slope of M is equal to* 0*. Then for each step M_i of the Frobenius slope filtration of slope s_i, and each $v \in V(M_i)$, $v \in M_i \otimes K[\![t]\!]_\delta[\log t]$ whenever $\delta \leq s_i$.*

Proof. This follows from Proposition 17.3.4, once we choose a basis of M so that the least generic Hodge slope of M is also equal to 0. $\square$

Corollary 18.5.3. *With notation as in Theorem 18.5.2, the special log-growth polygon of M lies on or below the Frobenius slope polygon of M/tM, provided that the two are aligned to have the same* left *endpoint.*

Notes

In the case of no singularities, Proposition 18.2.1 was first proved by Dwork; it appears in [**Dwo73**] and [**Dwo73b**]. (See also [**Chr83**].) The nilpotent case appears to be original; as noted in Chapter 17, the Christol-Dwork effective bounds in [**CD91**] are not strong enough to imply this.

Theorem 18.5.2 is original. It may be possible to use a similar technique to prove the other direction of Conjecture 18.5.1, but we are not sure about this. In the rank 2 case, Chiarellotto and Tsuzuki in [**CT06**] instead make a careful analysis of a cyclic vector for the Frobenius structure.

Part 6

p-adic exponents

CHAPTER 19

p-adic exponents

1. A cautionary note

Let M be a differential module over $K\langle t/\beta \rangle$ for which $IR(M \otimes F_\beta) = 1$. Then by Theorem 9.3.4, for any $\rho \in [0, \beta)$, $M \otimes K\langle t/\rho \rangle$ is trivial, so we have an isomorphism

$$M \otimes K\langle t/\rho \rangle \cong K\langle t/\rho \rangle^{\oplus n}$$

of differential modules.

Now let M be a differential module over $K\langle \alpha/t, t/\beta \rangle$ for which $IR(M \otimes F_\rho) = 1$ for all $\rho \in [\alpha, \beta]$. This most favorable situation was originally thought to be analogous to the situation of regular singularities in the complex setting. In particular, it was believed that for any $\alpha < \gamma \leq \delta < \beta$, it would be possible to write

$$M \otimes K\langle \gamma/t, t/\delta \rangle \cong M_{\lambda_1} \oplus \cdots \oplus M_{\lambda_n}$$

for some $\lambda_1, \dots, \lambda_n \in \mathbb{Z}_p$, where M_λ is the differential module of rank defined by $D(v) = \lambda t^{-1} v$ (as in the previous section).

This hope was dashed when a counterexample was exhibited by Monsky; it is the rank 2 differential module associated to the differential polynomial $p(1-x)T^2 - xT - a$, where $a \in \mathbb{Z}_p$ is constructed so that

$$\liminf_{m \to +\infty} |a + m|^{1/m} < 1, \qquad \liminf_{m \to +\infty} |a - m|^{1/m} = 1. \tag{19.1.0.1}$$

(The existence of such a is left as an exercise, or see [DR77, §7.20].) I plan to expand on this in a further unit; in the meantime, see [DR77, §7] for further discussion.

What this suggests is the hypothesis on the intrinsic radius needs to be supplemented with some extra hypotheses in order to get the decomposition we want. We will see one such hypothesis later (the existence of a Frobenius structure); in the interim, see the notes for further discussion.

Notes

Logarithmic growth is also commonly studied in the archimedean case; see [**Del70**].

A theory of exponents for differential modules on an annulus with intrinsic radius of convergence 1 everywhere was developed by Christol and Mebkhout [CM97, §4–5]; an alternate development was later given by Dwork [Dwo97] (see also [DGS94, §6]). The exponents are elements of the quotient $\mathbb{Z}_p/\mathbb{Z}$; this makes the construction somewhat complicated, as one must use *archimedean* considerations to identify a p-adic number, in a manner we will not elaborate further here. When the differences between exponents satisfy a p-adic non-Liouville condition (that is, they cannot be approximated unusually well by integers), one obtains a decomposition into modules of rank 1 [CM97, §6]. This is automatic in case of a Frobenius structure, as then the set of exponents is invariant under the operation $x \to x^p$

and so all of the exponents are forced to be rational numbers, whose differences are always non-Liouville. See [Loe97] (or a promised later unit) for a detailed exposition.

Exercises

(1) Prove that there exists $a \in \mathbb{Z}_p$ satisfying (19.1.0.1).

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`.

[Ber90] V.G. Berkovich, *Spectral theory and analytic geometry over non-Archimedean fields*, Math. Surveys and Monographs 33, Amer. Math. Soc., Providence, 1990.

[Ber93] V.G. Berkovich, Étale cohomology for non-Archimedean analytic spaces, *Publ. Math. IHÉS* **78** (1993), 5–161.

[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.

[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.

[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`.

[Crew] R. Crew, Canonical extensions, irregularities, and the Swan conductor, *Math. Ann.* **316** (2000), 19–37.

[Del70] P. Deligne, *Equations Différentielles à Points Singuliers Réguliers*, Lecture Notes in Math. 163, Springer-Verlag, Berlin, 1970.

[DK73] SGA7.

[Dem72] M. Demazure, *Lectures on p-divisible Groups*, Lecture Notes in Math. 302, Springer-Verlag, New York, 1972.

[Dwo69] B. Dwork, p-adic cycles, *Publ. Math. IHÉS* **37** (1969), 27–115.

[Dwo73] 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.

[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.

[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.

[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.

[Kap42] I. Kaplansky, Maximal fields with valuations, *Duke Math. J.* **9** (1942), 303–321.

[Kat79] N.M. Katz, Slope filtration of F-crystals, Journes de Gomtrie Algbrique de Rennes (Rennes, 1978), Vol. I, *Astrisque* **63** (1979), 113–163.

[Kto89] 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.

[Ked04] K.S. Kedlaya, A p-adic local monodromy theorem, *Annals of Math.* **160** (2004), 93–184.

[Ked05] 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.*, to appear.

[Ked07] K.S. Kedlaya, Semistable reduction for overconvergent F-isocrystals, III: local semistable reduction at monomial valuations, arXiv:math/0609645v2 (2007).

[Ked07a] K.S. Kedlaya, Swan conductors for p-adic differential modules, I: A local construction, *Alg. and Number Theory*, to appear; arXiv:math/0611835v3 (2007).

[Ked07c] K.S. Kedlaya, Slope filtrations for relative Frobenius, arXiv:math/0609272v2 (2007); to appear in *Astérisque*.

[Ked07b] K.S. Kedlaya, Swan conductors for p-adic differential modules, II: Global variation, arXiv:0705.0031v1 (2007).

[Ked07d] K.S. Kedlaya, The p-adic local monodromy theorem for fake annuli, arXiv:math/0507496v4 (2006), to appear in *Rend. Sem. Math. Padova*.

[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.

[Laz62] M. Lazard, Les zéros d'une fonction analytique d'une variable sur un corps valué complet, *Publ. Math. IHÉS* **14** (1962), 47–75.

[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.

[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.

[Ore33] O. Ore, Theory of non-commutative polynomials, *Annals of Math.* **34** (1933), 480–508.

[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).

[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.

[Rbe00] 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.

[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.

[Tsu98] N. Tsuzuki, Finite local monodromy of overconvergent unit-root F-isocrystals on a curve, *Amer. J. Math.* **120** (1998), 1165–1190.

[Tsu98a] N. Tsuzuki, The local index and the Swan conductor, *Comp. Math.* **111** (1998), 245–288.

[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.