Algebra & Number Theory

Volume 8

2014

No. 1

msp

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2014 is US $225/year for the electronic version, and $400/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscribers address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

ALGEBRA AND NUMBER THEORY 8:1 (2014)
dx.doi.org/10.2140/ant.2014.8.1

On the Picard number of K3 surfaces over number fields

François Charles

We discuss some aspects of the behavior of specialization at a finite place of Néron–Severi groups of K3 surfaces over number fields. We give optimal lower bounds for the Picard number of such specializations, thus answering a question of Elsenhans and Jahnel. As a consequence of these results, we show that it is possible to compute explicitly the Picard number of any given K3 surface over a number field.

1. Introduction

This paper deals with two questions concerning the arithmetic and the geometry of K3 surfaces. Let X be a polarized K3 surface over a number field k, and let $\mathfrak{p}$ be a finite place of k where X has good reduction. Denote by $X_{\mathfrak{p}}$ the special fiber of a smooth model of X over the ring of integers of $k_{\mathfrak{p}}$. Denote by $\overline{X}$ the base change of X to an algebraic closure of k, and by $\overline{X}_{\mathfrak{p}}$ the base change of $X_{\mathfrak{p}}$ to the residue field of $\mathfrak{p}$. Specialization of divisors induces a specialization map between the Néron–Severi groups of $\overline{X}$ and $\overline{X}_{\mathfrak{p}}$.

Question 1. *What can be said about the specialization map*

$$sp : NS(\overline{X}) \to NS(\overline{X}_{\mathfrak{p}})? \tag{1}$$

A standard argument using the cycle class map and the smooth base change theorem shows that this specialization map is always injective up to torsion. We are here interested in the defect of surjectivity. This problem has been featured prominently in [Bogomolov, Hassett and Tschinkel 2011] and [Li and Liedtke 2012].

The second question is the following. Recall that the Picard number of a variety is by definition the rank of its Néron–Severi group.

Question 2. *Given a projective embedding of X, is it possible to compute the geometric Picard number of X?*

MSC2010: primary 14J28; secondary 14C22, 14G25, 11G35.
Keywords: K3 surfaces, Picard number, Néron–Severi group.

This question is raised by Shioda [1981].

Using the Weil conjectures [Deligne 1974], it is possible to compute the Picard numbers of smooth projective varieties over finite fields, at least in the case the Tate conjecture holds. Indeed, counting points in sufficiently many extensions of the base field, one can compute the characteristic polynomial of the Frobenius acting on the second étale cohomology group and determine the multiplicity of 1 as an eigenvalue. If the Tate conjecture holds, this multiplicity is equal to the Picard number.

In characteristic 0, Question 2 is more difficult. In particular, the first explicit example of a K3 surface over a number field with Picard rank 1 has been recently given by van Luijk [2007]. Van Luijk's method provides a link between both questions. Indeed, it proceeds by computing Picard numbers at sufficiently many finite places in order to get information over the field of definition. In the past few years, the problem of computing Picard numbers of K3 surfaces has been featured for instance in the work of Elsenhans and Jahnel [2008a; 2008b] with recent geometric applications in the work of Hassett and Várilly-Alvarado [2013] and Várilly [Hassett et al. 2011]. Recent work of Hassett, Kresch and Tschinkel [2013] tackles this question in some cases.

With this approach, one of the main problems is finding finite places $\mathfrak{p}$ such that the specialization map (1) is as close to being surjective as possible, that is, such that $\rho(\overline{X}_{\mathfrak{p}})$ is as small as possible.

Note that the situation in this mixed characteristic setting is in stark contrast with the case of equal characteristic 0. Indeed, for K3 surfaces defined over function fields over $\mathbb{C}$ or $\overline{\mathbb{Q}}$, most specializations induce isomorphisms at the level of the Néron–Severi group. This is a consequence of Baire's theorem over $\mathbb{C}$ (see for instance [Voisin 2002, Chapter 13]) and of the Hilbert irreducibility theorem over $\overline{\mathbb{Q}}$ as was first noticed by Terasoma [1985; André 1996, Theorem 5.2(3)]. A different approach to this problem can be found in [Maulik and Poonen 2012].

On the other hand, over finite fields, there are obstructions for the map (1) to be surjective as was first noticed by Shioda [1981; 1983]. Indeed, it is a consequence of the Tate conjecture that the geometric Picard number of a K3 surface over a number field is always even; see for instance [de Jong and Katz 2000]. This striking fact has been recently used in a surprising way by Bogomolov, Hassett and Tschinkel [2011] and Li and Liedtke [2012] to prove that any complex K3 surface with odd Picard rank contains infinitely many rational curves.

In this paper, we describe the Shioda-type obstructions that can prevent the map (1) from being surjective, and we give optimal lower bounds for the Picard number of the specialization. One of our results is that Hodge theory can force the existence of such obstructions even when the Picard number is even; see Theorem 1(2) below.

Let X be a K3 surface over a number field k, and choose a complex embedding of k. Let ρ be the geometric Picard number of X, and for any finite place $\mathfrak{p}$ of k where X has good reduction, let $\rho_{\mathfrak{p}}$ be the geometric Picard number of $X_{\mathfrak{p}}$. Note that we always have

$$\rho_{\mathfrak{p}} \geq \rho.$$

We need to control the Hodge theory of $X_{\mathbb{C}}$. Let T be the orthogonal complement of $NS(X_{\mathbb{C}})$ in the singular cohomology group $H^2(X_{\mathbb{C}}, \mathbb{Q})$ with respect to cup-product. The space T is a sub-Hodge structure of $H^2(X_{\mathbb{C}}, \mathbb{Q})$. Let E be the algebra of endomorphisms of T that respect the Hodge structure. Zarhin [1983] shows that E is either a totally real field or a CM field.

The following result can be considered as a number field analog of the specialization results over function fields mentioned above:

Theorem 1. *Let X, T and E be as above.*

(1) *If E is a CM field or the dimension of T as an E-vector space is even, then there exist infinitely many places $\mathfrak{p}$ of good reduction such that $\rho_p = \rho$. Furthermore, after replacing k by a finite extension, this equality holds for a set of places of density* 1.

(2) *Assume E is a totally real field and the dimension of T as an E-vector space is odd.*

Let $\mathfrak{p}$ be a finite place of k where X has good reduction. Assume that the characteristic of the residue field of $\mathfrak{p}$ is at least 5. *Then*

$$\rho_{\mathfrak{p}} \geq \rho + [E : \mathbb{Q}].$$

The equality $\rho_{\mathfrak{p}} = \rho + [E : \mathbb{Q}]$ is satisfied for infinitely many places $\mathfrak{p}$ of good reduction. Furthermore, after replacing k by a finite extension, this equality holds for a set of places of density 1.

Remark 2. Note that if ρ is odd, the dimension of T over $\mathbb{Q}$ is odd; hence, X satisfies the assumptions of the second part of the theorem.

Remark 3. Elsenhans and Jahnel [2012] ask whether, with notation as in the theorem, there exists $\mathfrak{p}$ such that $\rho_{\mathfrak{p}} - \rho \leq 1$. The result above shows that it is not the case if E is a totally real field of degree at least 2 over $\mathbb{Q}$ such that the dimension of T over E is odd. This is however true in all other cases.

This result shows that the Picard number can be forced to jump in specializations even when the Picard number of X is even. Using the method of Li and Liedtke [2012], we get the following corollary:

Corollary 4. *Let X be either a K3 surface of Picard rank* 2 *with E a totally real field of degree* 4 *or a K3 surface of Picard rank* 4 *with E a totally real field of even degree. Then X contains infinitely many rational curves.*

There exist such K3 surfaces by [van Geemen 2008, Section 3], and they give new examples of K3 surfaces with infinitely many rational curves. Note that complex K3 surfaces of Picard rank different from 2 and 4 are known to contain infinitely many rational curves by [Li and Liedtke 2012].

The second main result of this paper is a solution to Question 2. Recall that the method of van Luijk [2007] to prove that a K3 surface X over $\mathbb{Q}$ has Picard number 1 was to first find two primes p and q of good reduction such that X specializes to a K3 surface of Picard number 2 modulo p and q. If the discriminant of the Néron–Severi lattices modulo p and q differ by a nonsquare factor, van Luijk shows that this implies that X has Picard number 1.

By Remark 3, there are cases where we cannot expect van Luijk's method to work directly for all K3 surfaces of rank 1. However, the second part of Theorem 1 can be used to show that reduction at finite places does indeed give enough information to compute Picard numbers over number fields.

This gives a theoretical explanation to the computations in [van Luijk 2007; Elsenhans and Jahnel 2008a; 2008b; Hassett et al. 2011; Hassett and Várilly-Alvarado 2013].

Theorem 5. *There exists an algorithm that, given a projective K3 surface X over a number field, either returns its geometric Picard number or does not terminate.*

If $X \times X$ satisfies the Hodge conjecture for codimension-2 cycles, then the algorithm applied to X terminates.

Remark 6. In case the algorithm terminates, it is possible to compute the Picard number over the base field by searching for divisors spanning the geometric Néron-Severi group with rational coefficients and taking invariants under the action of the Galois group.

Remark 7. Let X be a K3 surface over $\mathbb{C}$. With the notation of Theorem 1, $X \times X$ satisfies the Hodge conjecture if and only if the field E acts by algebraic correspondences. By [André 1996], this would be a consequence of the standard conjectures. Mukai [2002] has announced a proof in the case E is a CM field.

Remark 8. The proof of the theorem actually shows that the only case where the algorithm would not terminate is, with the notation of Theorem 1, if E is a totally real field that does not act on $H^2(X, \mathbb{Q})$ by algebraic correspondences and T is of odd dimension as a vector space over E.

In particular, the algorithm always terminates for surfaces with $E = \mathbb{Q}$.

While we only consider K3 surfaces in this paper, some of the methods we consider have a wider range of applications. Assuming general conjectures on algebraic cycles, it is a general fact that the Mumford–Tate group associated to the second cohomology group of a variety controls specialization of Néron–Severi groups in a fashion that is similar to the way the monodromy representation appears in [André 1996; Maulik and Poonen 2012]. The multiplicity of the weight 0 in the corresponding representation is what forces the Picard number to jump after specialization. This is related to algorithmic computations of Néron–Severi groups as in our paper.

For K3 surfaces, the work of Zarhin [1983] and Tankeev [1990; 1995] allows us to give precise and unconditional results. The results of our paper conjecturally hold for varieties with $h^{2,0} = 1$. It seems likely that one can prove them unconditionally for holomorphic symplectic varieties by extending the work of Tankeev cited above.

In Section 2, we recall results of Zarhin and Tankeev on the second cohomology group of a K3 surface. This allows us to prove Theorem 1 in Section 3. Section 4 is devoted to discriminant computations that will allow us to prove Theorem 5 in Section 5.

2. Algebraic monodromy groups of K3 surfaces over number fields

The results of this section are mostly contained in the work of Zarhin and Tankeev. After recalling some preliminary material, we describe the algebraic monodromy group of a K3 surface defined over a number field.

2.1. ***Mumford–Tate groups and the Mumford–Tate conjecture.*** Let $\mathbb{S}$ be the Deligne torus, that is, the algebraic group over $\mathbb{R}$ defined as

$$\mathbb{S} = \operatorname{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m.$$

Let H be a finite-dimensional vector space over $\mathbb{Q}$. Giving a Hodge structure on H is equivalent to giving an action of $\mathbb{S}$ on $H_{\mathbb{R}} = H \otimes \mathbb{R}$.

Definition 9. Let H be a rational Hodge structure. The Mumford–Tate group of H is the smallest algebraic subgroup $MT(H)$ of $GL(H)$ such that $MT(H)_{\mathbb{R}}$ contains the image of $\mathbb{S}$ in $GL(H_{\mathbb{R}})$.

We refer to [Deligne et al. 1982, Chapter I] for general properties of Mumford–Tate groups. Since $\mathbb{S}$ is connected, this definition implies that Mumford–Tate groups are connected. Note that the Mumford–Tate group of a polarized Hodge structure is reductive.

Let i and j be nonnegative integers, and consider the Hodge structure

$$V = H^{\otimes i} \otimes (H^*)^{\otimes j}.$$

The Mumford–Tate group $MT(H)$ acts on V. If v is a Hodge class in V, then the line $\mathbb{Q}v$ is globally invariant under the action of $MT(H)$. Conversely, it follows from Chevalley's theorem on affine groups that $MT(H)$ is the largest algebraic subgroup of $GL(H_{\mathbb{C}})$ that leaves all such lines globally invariant [Deligne et al. 1982, Chapter I, Proposition 3.4].

We now turn to the ℓ-adic theory. General results can be found in [Serre 1981]. Let k be a number field, and fix an algebraic closure $\bar{k}$. Let X be a smooth projective variety over k, and denote by $\overline{X}$ the variety $X \times_{\operatorname{Spec} k} \operatorname{Spec} \bar{k}$. Fixing a prime number ℓ, we can consider the étale cohomology group $H^i(\overline{X}, \mathbb{Q}_\ell)$ for some integer i. Let ρ_ℓ denote the continuous representation

$$\rho_\ell : G_k \to GL(H^i(\overline{X}, \mathbb{Q}_\ell))$$

of the absolute Galois group G_k of k. The image of ρ_ℓ is an ℓ-adic Lie group.

Definition 10. With notation as above, let G_ℓ be the Zariski closure of the image of ρ_ℓ in the algebraic group $GL(H^i(\overline{X}, \mathbb{Q}_\ell))$. The algebraic group G_ℓ is called the *algebraic monodromy group* associated to the Galois representation ρ_ℓ.

Note that replacing k by a finite extension replaces G_ℓ by an open subgroup of finite index. In particular, the identity component of the algebraic monodromy group does not depend on the choice of a field of definition for X.

General conjectures on algebraic cycles give important information on Mumford–Tate and algebraic monodromy groups. In particular, the latter are expected to be reductive. The expected relationship between those two groups is described by the Mumford–Tate conjecture as follows; see [Serre 1981].

Conjecture 11. *Let k be a number field, and fix a complex embedding of k. Let X be a smooth projective variety over k.*

Let G_ℓ be the algebraic monodromy group associated to the étale cohomology group $H^i(X_{\mathbb{C}}, \mathbb{Q}_\ell)$ for some prime number ℓ, and let G_ℓ° be its identity component. Then there is a canonical isomorphism

$$G_\ell^\circ \simeq MT(H^i(X_{\mathbb{C}}, \mathbb{Q}))_{\mathbb{Q}_\ell}.$$

The Mumford–Tate conjecture is implied by the conjunction of the Tate and Hodge conjectures. A lot of work has been done in its direction in the case of abelian varieties [Serre 1998; Pink 1998; Vasiu 2008].

In this paper, we will focus on the case of K3 surfaces, where the Mumford–Tate conjecture holds. However, an important part of our method concerning specialization of Néron–Severi groups holds in a general setting if one assumes the Mumford–Tate conjecture.

2.2. *Mumford–Tate groups and algebraic monodromy groups of K3 surfaces.*

The following result is due to Tankeev and is crucial to this paper:

Theorem 12 [Tankeev 1990; 1995]. *The Mumford–Tate conjecture holds for the second cohomology group of K3 surfaces over number fields.*

This result allows us to get a Hodge-theoretic description of the Galois action on the second cohomology group of a K3 surface.

Let us now recall the description due to Zarhin [1983] of the Mumford–Tate group of a K3 surface. Let X be a K3 surface over $\mathbb{C}$, and consider the singular cohomology $H = H^2(X, \mathbb{Q})$ endowed with its weight-2 Hodge structure. The Hodge structure H splits as a direct sum

$$H = NS(X) \oplus T,$$

where $NS(X)$ is the Néron–Severi group of X with rational coefficients and T is the orthogonal of $NS(X)$ in H with respect to the cup-product. The Hodge structure T is called the transcendental part of $H^2(X, \mathbb{Q})$.

The Hodge structure T is simple. By Lefschetz's theorem on $(1, 1)$ classes, T is the smallest sub-Hodge structure of H such that $T \otimes \mathbb{C}$ contains $H^2(X, \mathbb{O}_X)$. By the Hodge index theorem, cup-product on $H^2(X, \mathbb{Q})$ restricts to a polarization $\psi : T \otimes T \to \mathbb{Q}$ on T.

Since $NS(X)$ is spanned by Hodge classes, the Mumford–Tate group of H acts by a character on $NS(X)$ and identifies with the Mumford–Tate group of T. Since T is polarized by ψ, $MT(T)$ is contained in the group of orthogonal similitudes $GO(T, \psi)$.

Let E be the algebra of endomorphisms of the Hodge structure T. Zarhin [1983, Theorem 1.5.1] proves that E is either a totally real field or a CM field. The field E is equipped with an involution induced by the polarization on T, which is either the identity if E is totally real or complex conjugation in case E is CM.

Since E consists of endomorphisms of Hodge structures, the Mumford–Tate group of T commutes with E. By the discussion above, the Mumford–Tate group of T is a subgroup of the centralizer of E in the group $GO(T, \psi)$.

Theorem 13 [Zarhin 1983, Theorem 2.2.1]. *The Mumford–Tate group of T is the centralizer of E in the group of orthogonal similitudes $GO(T, \psi)$.*

Now keep the same notation, and assume X can be defined over a number field k. Fix a prime number ℓ. The action of the absolute Galois group G_k on $H^2(X, \mathbb{Q}_\ell)$ leaves the $\mathbb{Q}_\ell$-span of the Néron–Severi group of X globally invariant as well as its orthogonal complement $T_\ell = T \otimes \mathbb{Q}_\ell$. As above, the identity component of the algebraic monodromy group G_ℓ of $H^2(X, \mathbb{Q}_\ell)$ identifies with the algebraic monodromy group of T_ℓ.

The polarization ψ on T extends to a symmetric bilinear form ψ_ℓ. The representation of G_k in the automorphism group of T_ℓ factors through the group $GO(T_\ell, \psi_\ell)$.

Since Hodge cycles on products of K3 surfaces are absolute Hodge [Deligne et al. 1982, Chapter I, 6.26], the field E corresponding to endomorphisms of the

Hodge structure T acts on T_ℓ and commutes with a finite-index subgroup of G_k. As a consequence, the identity component of G_ℓ commutes with the action of $E \otimes \mathbb{Q}_\ell$.

By Theorem 12, the Mumford–Tate conjecture holds for X. As an immediate corollary of Theorem 13, we get the following description of the identity component of the algebraic monodromy group of X:

Corollary 14. *With notation as above, the identity component of the algebraic monodromy group associated to T_ℓ is the centralizer of $E \otimes \mathbb{Q}_\ell$ in the group of orthogonal similitudes $GO(T_\ell, \psi_\ell)$.*

3. Picard numbers of specializations

This section is devoted to the proof of Theorem 1. We start by the following result, which encompasses the elementary linear algebra needed in Theorem 1.

Let T be a finite-dimensional vector space endowed with a nondegenerate symmetric bilinear form ψ. If f is any linear endomorphism of T, let f' be the adjoint of f with respect to ψ.

Let E be a number field acting on T. Assume that E is stable under $e \mapsto e'$ and that E is either a totally real field with $e = e'$ for all $e \in E$ or a CM field such that $e \mapsto e'$ acts as complex conjugation on E.

Let H be the centralizer of E in the special orthogonal group $SO(T, \psi)$. Let ℓ be a prime number, and let $H_\ell = H \otimes \mathbb{Q}_\ell$.

Proposition 15. (1) *If E is a CM field or the dimension of T as an E-vector space is even, then there exists $h \in H_\ell$ such that h does not have any root of unity as an eigenvalue.*

(2) *If E is a totally real field and the dimension of T as an E-vector space is odd, then the eigenspace of any $h \in H_\ell$ associated to the eigenvalue* 1 *is of dimension at least* $[E : \mathbb{Q}]$. *Furthermore, there exists $h \in H_\ell$ for which this dimension is exactly* $[E : \mathbb{Q}]$ *and such that no root of unity different from* 1 *appears as an eigenvalue of h.*

Proof. Let us first assume that E is a totally real field. By [Zarhin 1983, 2.1], there exists a unique E-bilinear form $\phi : T \times T \to E$ such that $\psi = \mathrm{Tr}_{E/\mathbb{Q}}(\phi)$. With this notation, the centralizer of E in $SO(T, \psi)$ is equal, as a subgroup of $GL(T)$, to the Weil restriction $\mathrm{Res}_{E/\mathbb{Q}}(SO_E(T, \phi))$, where $SO_E(T, \phi)$ denotes the group of orthogonal similitudes of the E-vector space T with respect to ϕ.

Assume furthermore that the dimension of T as a vector space over E is even, and let us show that there is an element $h \in H_\ell$ such that h does not have any root of unity as an eigenvalue.

Considering an orthogonal decomposition of T as an E-vector space endowed with the bilinear form ϕ, we can assume T is of dimension 2 over E. Let h be an orthogonal automorphism of the E-vector space T of determinant 1 that is not of finite order. Then h corresponds to an element of H_ℓ with the desired property.

Now if the dimension of T as a vector space over E is odd, recall that any element of $SO_E(T, \psi)$ admits 1 as an eigenvalue. It follows from the description of H_ℓ as a Weil restriction that any $h \in H_\ell$ has 1 as an eigenvalue and that the corresponding eigenspace is invariant under the action of E. As a consequence, its dimension is at least $[E : \mathbb{Q}]$. One can then argue as in the previous paragraph to conclude the proof of the proposition in this case.

Let us now assume that E is a CM field. Let e be an element of E such that $ee' = 1$ and e is not a root of unity. Then multiplication by e on T corresponds to an element of H_ℓ as in the theorem. □

We now turn to the proof of Theorem 1. From now on, we use the notation there. Let us start with a straightforward lemma.

Lemma 16. *The identity component of the algebraic monodromy group associated to $T_\ell(1)$ is the centralizer of $E \otimes \mathbb{Q}_\ell$ in the special orthogonal group $SO(T_\ell, \psi_\ell)$.*

Proof. The representation of G_k on $T_\ell(1)$ is equal to the representation of G_k on T_ℓ twisted by the cyclotomic character. On the other hand, since the map

$$T_\ell(1) \otimes T_\ell(1) \to \mathbb{Q}_\ell$$

given by cup-product is G_k-equivariant, G_k acts on $T_\ell(1)$ through the orthogonal group $O(T_\ell, \psi_\ell)$.

The lemma then follows from Corollary 14 and the fact that the special orthogonal group is the identity component of the orthogonal group. □

Proof of Theorem 1. We use the notation of the theorem. First note that since specialization of Néron–Séveri groups is injective, the inequality $\rho_{\mathfrak{p}} \geq \rho$ always holds.

Let $F_{\mathfrak{p}}$ be the geometric Frobenius at $\mathfrak{p}$ acting on the étale cohomology group $H^2(\overline{X}_{\mathfrak{p}}, \mathbb{Q}_\ell(1))$, where ℓ is a prime number prime to $\mathfrak{p}$. By the smooth base change theorem, the group $H^2(\overline{X}_{\mathfrak{p}}, \mathbb{Q}_\ell(1))$ identifies with $H^2(\overline{X}, \mathbb{Q}_\ell(1))$, and $F_{\mathfrak{p}}$ leaves both the Néron–Severi group and $T_\ell(1)$ globally invariant.

Let H be the centralizer of $E \otimes \mathbb{Q}_\ell$ in the special orthogonal group $SO(T_\ell, \psi_\ell)$. Let n be the dimension of T as a vector space over $\mathbb{Q}$, and let S be the finite set of complex roots of unity of degree at most n over $\mathbb{Q}$.

Assume first that E is a CM field or E is a totally real field and the dimension of T as a vector space over E is even. By Proposition 15, the set of $h \in H_\ell$ such that h does not have any eigenvalue in S is a dense, Zariski-open subset V_ℓ of H_ℓ.

By Lemma 16 and Chebotarev's density theorem, we can find a finite extension k' of k with the property that if U is the set of finite places $\mathfrak{p}$ of k' such that for any $\mathfrak{p} \in U$, X has good reduction at $\mathfrak{p}$ and the geometric Frobenius $F_{\mathfrak{p}}$ acting on $T_\ell(1)$ does not have any eigenvalue in S, then U has density 1.

Indeed, let k' be a finite extension of k such that the image of absolute Galois group $G_{k'}$ in the automorphism group of $T_\ell(1)$ lands in H_ℓ. By Lemma 16, the image of $G_{k'}$ is Zariski-dense in H_ℓ. By [Serre 1998, I.8, Théorème 2], this implies that the set U of finite places $\mathfrak{p}$ of k' such that for any $\mathfrak{p} \in U$, X has good reduction at $\mathfrak{p}$ and the geometric Frobenius $F_\mathfrak{p}$ acting on $T_\ell(1)$ belongs to V_ℓ, has density 1.

Choose U as above, and let $\mathfrak{p}$ be in U. By the Weil conjectures, the characteristic polynomial of the geometric Frobenius $F_\mathfrak{p}$ has rational coefficients. By definition of S, this implies that it does not have any eigenvalue that is a root of unity.

As a consequence, $F_\mathfrak{p}$ acting on the whole cohomology group $H^2(\overline{X}, \mathbb{Q}_\ell(1))$ admits 1 as an eigenvalue of multiplicity ρ and does not have any other eigenvalue that is a root of unity. It follows that $\rho_\mathfrak{p} \leq \rho$ and finally that $\rho_\mathfrak{p} = \rho$. This proves the first part of Theorem 1.

Now assume that E is a totally real field and that the dimension of T as a vector space over E is odd. By Proposition 15, every element of H_ℓ has 1 as an eigenvalue with multiplicity at least $[E : \mathbb{Q}]$. By definition of the algebraic monodromy group, if $\mathfrak{p}$ is a finite place of k, then some power of the geometric Frobenius belongs to H_ℓ.

By work of Nygaard [1983] and Nygaard and Ogus [1985] in the finite height case and the author in the supersingular case [Charles 2013], K3 surfaces over finite fields of characteristic at least 5 satisfy the Tate conjecture; namely, their geometric Picard number is equal to the dimension of the largest subspace of the second cohomology group where the Frobenius morphism acts by roots of unity. It follows that $\rho_\mathfrak{p} \geq \rho + [E : \mathbb{Q}]$.

By Proposition 15 again, the set of $h \in H_\ell$ such that h admits 1 as an eigenvalue of multiplicity $[E : \mathbb{Q}]$ and does not have any other eigenvalue in S is a dense, Zariski-open subset of H_ℓ.

By Lemma 16 and Chebotarev's density theorem, we can find a finite extension k' of k and a set U of finite places $\mathfrak{p}$ of k' that has density 1 such that for any $\mathfrak{p} \in U$, X has good reduction at $\mathfrak{p}$ and the geometric Frobenius $F_\mathfrak{p}$ acting on $T_\ell(1)$ admits 1 as an eigenvalue of multiplicity $[E : \mathbb{Q}]$ and does not have any other eigenvalue in S.

Choose U as above, and let $\mathfrak{p}$ be in U. Since $X_\mathfrak{p}$ satisfies the Tate conjecture, we can argue as above to finish the proof of Theorem 1. □

Remark 17. Using Frobenius tori as in [Serre 1981] and the fact that Frobenius tori are maximal tori of the Mumford–Tate groups for infinitely many primes, one can work directly in the group of orthogonal similitudes instead of reducing to the special orthogonal group as in Lemma 16.

4. Discriminants of Néron–Severi groups

In this section, we discuss properties of the Néron–Severi lattices of specializations of K3 surfaces. Once again, we use the notation of Theorem 1.

Proposition 18. *Assume that E is a totally real field and that the dimension of T over E is odd. If $\mathfrak{p}$ is a finite place of k such that X has good reduction at $\mathfrak{p}$, denote by $\delta(\mathfrak{p}) \in \mathbb{Q}^*/(\mathbb{Q}^*)^2$ the discriminant of the lattice $NS(\overline{X}_{\mathfrak{p}})$ with respect to the intersection product.*

There exist infinitely many finite places $\mathfrak{p}_i$, $i \in \mathbb{N}$, and infinitely many finite places $\mathfrak{q}_j$, $j \in \mathbb{N}$, of k such that for any integers i and j

(1) *X has good, ordinary reduction at both $\mathfrak{p}_i$ and $\mathfrak{q}_j$,*

(2) *$\rho_{\mathfrak{p}_i} = \rho_{\mathfrak{q}_j} = \rho + [E : \mathbb{Q}]$ and*

(3) *$\delta(\mathfrak{p}_i) \neq \delta(\mathfrak{q}_j)$.*

Remark 19. A special case of this result is that the method developed in [van Luijk 2007] to prove that a given K3 surface over a number field has Picard number 1 always works in the case $E = \mathbb{Q}$. We noted in Remark 3 that it cannot work directly otherwise.

In Section 5, we will adapt the method so as to make it work in every case.

We start with some easy linear algebra.

Lemma 20. *Let ℓ be a prime number, and let V be a free module of finite rank over $\mathbb{Z}_\ell$. Let g be an endomorphism of V such that $g \otimes \mathbb{Q}_\ell$ is a semisimple automorphism of $V \otimes \mathbb{Q}_\ell$, and denote by r the multiplicity of 1 as an eigenvalue of g. Let W be the eigenspace associated to the eigenvalue 1 of g. Let d be a positive integer.*

Then there exists an integer N with the following property. Let h be an endomorphism of V such that $h \otimes \mathbb{Q}_\ell$ is a semisimple automorphism of $V \otimes \mathbb{Q}_\ell$. Assume that r is the multiplicity of 1 as an eigenvalue of h, and let W' be the eigenspace associated to the eigenvalue 1 of h. If h is congruent to g modulo ℓ^N, then $W \otimes \mathbb{Z}/\ell^d\mathbb{Z} = W' \otimes \mathbb{Z}/\ell^d\mathbb{Z}$.

Remark 21. In particular, if V is endowed with a symmetric bilinear form ϕ such that the restriction of ϕ to W is not degenerate and N is sufficiently large, then the discriminants of W and W' are equal in $\mathbb{Q}_\ell^*/(\mathbb{Q}_\ell^*)^2$.

Proof. Write $V = W \oplus \widetilde{W}$, where $\widetilde{W}$ is a g-invariant submodule of V. Since $g \otimes \mathbb{Q}_\ell$ does not fix any nonzero element of $\widetilde{W} \otimes \mathbb{Q}_\ell$, a compactness argument shows that there exists an integer N such that if $g(v) - v \in l^N V$ for some $v \in \widetilde{W}$, then $v \in \ell^k \widetilde{W}$.

Let h be as in the statement of the lemma. By definition of N, if $v \in V$ is fixed by h, then $v \otimes \mathbb{Z}/\ell^k\mathbb{Z} \in W \otimes \mathbb{Z}/\ell^k\mathbb{Z}$. With the notation of the lemma, it follows that $W' \otimes \mathbb{Z}/\ell^k\mathbb{Z} \subset W \otimes \mathbb{Z}/\ell^k\mathbb{Z}$. Since both W and W' are saturated submodules of V of the same rank r, equality follows. □

Proof of Proposition 18. First note that the dimension of T as a vector space over E is at least 3. Indeed, let ω be a generator of $T^{2,0} \subset T \otimes \mathbb{C}$, and let $\sigma : E \to \mathbb{C}$ be the complex embedding of E satisfying

$$e.\omega = \sigma(e)\omega \qquad \text{for all } e \in E.$$

The complex lines $\mathbb{C}\omega$ and $\mathbb{C}\overline{\omega}$ are two distinct one-dimensional subspaces of $T_E \otimes_\sigma \mathbb{C}$, where T_E denotes T endowed with the structure of a vector space over E. As a consequence, the dimension of T as a vector space over E is at least 2 and at least 3 since we assumed it to be odd.

Recall that ψ is the bilinear form on T induced by cup-product. As in Proposition 15, there exists a unique E-bilinear form $\phi : T \times T \to E$ such that $\psi = \mathrm{Tr}_{E/\mathbb{Q}}(\phi)$. Any orthogonal basis of T_E with respect to ϕ induces an orthogonal decomposition of T with respect to ψ

$$T = T_1 \oplus \cdots \oplus T_r,$$

where the T_i are stable under the action of E and of dimension 1 as E-vector spaces.

By the same reasoning as above, since the T_i are one-dimensional over E, there is no integer i such that $T_i \otimes \mathbb{C}$ contains the two-dimensional space $T^{2,0} \oplus T^{0,2}$.

The signature of ψ on T is $(2, \dim(T) - 2)$. By the Hodge index theorem and the remark above, the signature of the restriction of ψ to T_i is either $(0, [E : \mathbb{Q}])$ or $(1, [E : \mathbb{Q}] - 1)$. Since the dimension of T over E is at least 3, both these signatures appear, and this implies that, up to reordering, we can assume that the discriminant of T_1 is negative and the discriminant of T_2 is positive. Let δ and δ' be these two discriminants in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$.

Since $\delta \neq \delta'$ in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$, there exists a prime number ℓ such that the images of δ and δ' in $\mathbb{Q}_\ell^*/(\mathbb{Q}_\ell^*)^2$ are different. If W is any subspace of T_ℓ such that the restriction of ψ_ℓ to W is nondegenerate, let $\delta(W)$ denote the discriminant of W in $\mathbb{Q}_\ell^*/(\mathbb{Q}_\ell^*)^2$

By Lemma 20, Proposition 15 and Chebotarev's density theorem, we can find, for any positive integer d, infinitely many finite places $\mathfrak{p}_i$, $i \in \mathbb{N}$, and infinitely many finite places $\mathfrak{q}_j$, $j \in \mathbb{N}$, of k such that for any integers i and j

(1) X has good, ordinary reduction at both $\mathfrak{p}_i$ and $\mathfrak{q}_j$,

(2) $\rho_{\mathfrak{p}_i} = \rho_{\mathfrak{q}_j} = \rho + [E : \mathbb{Q}]$,

(3) if $F_{\mathfrak{p}_i}$ and $F_{\mathfrak{q}_j}$ denote the geometric Frobeniuses at $\mathfrak{p}_i$ and $\mathfrak{q}_j$ acting on $T_\ell(1)$ and $W_{\mathfrak{p}_i}$ and $W_{\mathfrak{q}_j}$ denote the eigenspaces associated to the eigenvalue 1 of $F_{\mathfrak{p}_i}$ and $F_{\mathfrak{q}_j}$, then $W_{\mathfrak{p}_i} \otimes \mathbb{Z}/\ell^d\mathbb{Z} = T_1 \otimes \mathbb{Z}/\ell^d\mathbb{Z}$ and $W_{\mathfrak{q}_j} \otimes \mathbb{Z}/\ell^d\mathbb{Z} = T_2 \otimes \mathbb{Z}/\ell^d\mathbb{Z}$, respectively, and

(4) the geometric Frobeniuses $F_{\mathfrak{p}_i}$ and $F_{\mathfrak{q}_j}$ acting on $T_\ell(1)$ do not have any eigenvalue different from 1 that is a root of unity.

If d is big enough, this implies that $\delta(W_{\mathfrak{p}_i}) = \delta$ and $\delta(W_{\mathfrak{q}_j}) = \delta'$ in $\mathbb{Q}_\ell^*/(\mathbb{Q}_\ell^*)^2$. Proposition 18 immediately follows by the Tate conjecture for ordinary K3 surfaces. □

Remark 22. The proof above shows that the density of pairs $(\mathfrak{p}, \mathfrak{q})$ as in the proposition is positive.

5. Computing the Picard number over number fields

This section is devoted to a proof of Theorem 5. Given a projective K3 surface over a number field k, we want to compute the Picard number of X using the equations of X in a projective embedding.

There are two steps in our approach. The first one is finding sufficiently many divisors on X, and the second is proving that the rational span of these divisors is the whole Néron–Severi group of X.

In case we want to prove that the K3 surface has Picard number 1, the first step is vacuous as we already have a divisor given by a hyperplane section. In general, the first step is done by going through the Hilbert schemes of curves in the projective space we are working in and doing elimination theory to find curves on X. After a finite number of computations, this will allow us to find divisors on X that span the Néron–Severi group.

The second step will be done by reducing to finite characteristic and using our results above.

However, this is not sufficient. Indeed, the field E of endomorphisms of the transcendental part of the Hodge structure of X plays a role in the behavior of the Picard number after specialization, and in case E is a totally real field strictly containing $\mathbb{Q}$ such that T is of odd dimension over E, this leads to some loss of accuracy in the estimates reduction at finite places can provide.

This problem will be solved by studying codimension-2 varieties in $X \times X$. Assuming the Hodge conjecture for $X \times X$, these determine the field E, which will allow us to conclude.

We start by the following result:

Proposition 23. *Let X be a K3 surface over a number field k. Assume we are given the equations of X in some projective embedding.*

Let T be the transcendental part of $H^2(X, \mathbb{Q})$, and let E be the field of endomorphisms of the Hodge structure T.

Assume that we know that the Picard number of X is greater or equal than some integer ρ and that the degree of E over $\mathbb{Q}$ is greater or equal than some integer d. Then there exists an algorithm with the following properties:

(1) *If the algorithm terminates, it proves that the Picard number of X is ρ.*

(2) *Suppose that the Picard number of X is actually ρ. Then the algorithm terminates unless E is totally real, the dimension of T as a vector space over E is odd and $d < [E : \mathbb{Q}]$.*

Proof. Let ρ' be the actual Picard number of X. We know that $\rho' \geq \rho$. Using the Weil conjectures [Deligne 1974], we can compute the characteristic polynomial of Frobenius at any finite place $\mathfrak{p}$ of k of good reduction; see [van Luijk 2007; Elsenhans and Jahnel 2012]. This allows in particular to compute the numbers $\rho_{\mathfrak{p}}$. Using the Artin–Tate formula [Tate 1966, Conjecture (C)], which holds for surfaces satisfying the Tate conjecture by [Milne 1975], one can also compute the discriminants $\delta(\mathfrak{p})$ as in Proposition 18.

We start computing $\rho_{\mathfrak{p}}$ and $\delta(\mathfrak{p})$ for all places $\mathfrak{p}$ of good reduction.

Let us distinguish three cases. First assume that E is a CM field. By Theorem 1, we can find $\mathfrak{p}$ with $\rho_{\mathfrak{p}} = \rho'$. If it happens that ρ, the lower bound for the Picard number of X that we were given, is equal to the actual Picard number ρ' (that we do not know yet), the computation at $\mathfrak{p}$ together with this lower bound allows us to prove that X has Picard number $\rho = \rho'$.

Now assume that E is totally real and the dimension of T as a vector space over E is even. In that case, Theorem 1 allows us to make the same conclusion.

The last case happens when E is totally real and the dimension of T as a vector space over E is odd. By Proposition 18, the finite field computations give us two finite places $\mathfrak{p}$ and $\mathfrak{q}$ of k where X has good, ordinary reduction, with $\rho_{\mathfrak{p}} = \rho_{\mathfrak{q}} = \rho' + [E : \mathbb{Q}]$ and $\delta(\mathfrak{p}) \neq \delta(\mathfrak{q})$.

Since $\delta(\mathfrak{p}) \neq \delta(\mathfrak{q})$, we know that the specialization maps $NS(\overline{X}) \to NS(\overline{X}_{\mathfrak{p}})$ and $NS(\overline{X}) \to NS(\overline{X}_{\mathfrak{p}})$ are not surjective. This means that $NS(X_{\mathfrak{p}}) \cap T_{\ell}(1)$ is nonzero in $H^2(\overline{X}_{\mathfrak{p}}, \mathbb{Q}_{\ell}(1))$.

Now we know by the analysis in the proof of Theorem 1 that this intersection is stable under the action of E. As a consequence, its dimension is at least $[E : \mathbb{Q}] \geq d$. This gives us the estimation

$$\rho' \leq \rho_{\mathfrak{p}} - d.$$

In case d happens to be equal to the actual degree $[E : \mathbb{Q}]$ and $\rho = \rho'$, these estimates allow us to prove that X has Picard number $\rho = \rho'$. □

Remark 24. In case $\rho = \rho' = 1$ and $E = \mathbb{Q}$, this proves that the method of [van Luijk 2007] always works.

Proof of Theorem 5. Let X, E and T be as above. Let ρ' be the Picard number of X and d' the degree of E over $\mathbb{Q}$. By Proposition 23, we only need to be able to prove that the Picard number of X is at least ρ' and the degree of E over $\mathbb{Q}$ is at least d'.

The assertion on the Picard number is theoretically — although not computationally — easy. One can go through Hilbert schemes of curves in the projective space where X is given and check, using elimination theory, for curves that happen to lie on X. Computing intersection matrices with these divisors on X, one can find divisors that span a ρ'-dimensional subset of the Néron–Severi group of X.

Running these Hilbert scheme computations alongside the computations of Proposition 23 allows for a computation of the Picard number of X unless E is a totally real field strictly containing $\mathbb{Q}$ such that T is of odd dimension over E.

To deal with the latter case, one has to work on $X \times X$. If one assumes the Hodge conjecture for $X \times X$, then elements of E are induced by codimension-2 cycles in $X \times X$. As above, one can use Hilbert schemes to find codimension-2 subschemes in $X \times X$.

Given such a subscheme Z, the action of Z on T can be determined by first computing the characteristic polynomial of the correspondence $H^2(X, \mathbb{Q}) \to H^2(X, \mathbb{Q})$ by computing intersection numbers between T and the various subschemes obtained by composing the correspondence induced by Z with itself.

Factoring the characteristic polynomial, this gives candidates for the algebraic number λ such that $[Z]_*\eta = \lambda\eta$, where η is a nonzero algebraic 2-form on X. An approximate computation can then determine λ. The degree of λ over $\mathbb{Q}$ is a lower bound for $[E : \mathbb{Q}]$.

By the primitive element theorem, it is easy to see that one can find Z such that this computation gives an optimal estimate for the degree of E. Using Proposition 23, this concludes the proof.

In conclusion, an algorithm to compute Picard number of K3 surfaces works as follows. Let X be a K3 surface. Run the three algorithms alongside each other:

(1) Going through Hilbert schemes of a suitable projective space, find divisors on X and compute the dimension of their span in the Néron–Severi group via intersection theory. This gives a lower bound for the Picard number.

(2) Going through Hilbert schemes of a suitable projective space, find codimension-2 cycles in $X \times X$. Using intersection theory again, use these to get a lower bound on the field E of endomorphisms of the transcendental part of $H^2(X, \mathbb{Q})$.

(3) Going through finite places $\mathfrak{p}$ of k, compute the Picard number and the discriminant of the Néron–Severi group of $\overline{X}_{\mathfrak{p}}$ by counting points over finite fields. Using the preceding step, get an upper bound on the Picard number of X.

We showed that the estimates provided by the method solve the problem unconditionally unless E is a totally real field strictly containing $\mathbb{Q}$ and the transcendental part of $H^2(X, \mathbb{Q})$ is of odd dimension over E. In the latter case, the estimates above are sufficiently precise to compute the Picard number if we assume the Hodge conjecture for $X \times X$. □

Remark 25. It seems that the computations of the second step above would be very lengthy to do in practice. We however wanted to point out that they can be done.

Note that the computations terminate much faster in most cases since $E = \mathbb{Q}$ for the majority of K3 complex surfaces in the sense of Baire category.

Acknowledgments

This paper was written during a stay at the University of Bonn. I would like to thank Daniel Huybrechts for his hospitality and many useful comments. I thank Bjorn Poonen and Matthias Schuett for useful remarks on a first version of this text.

References

[André 1996] Y. André, "Pour une théorie inconditionnelle des motifs", *Inst. Hautes Études Sci. Publ. Math.* 83 (1996), 5–49. MR 98m:14022 Zbl 0874.14010

[Bogomolov, Hassett and Tschinkel 2011] F. Bogomolov, B. Hassett, and Y. Tschinkel, "Constructing rational curves on K3 surfaces", *Duke Math. J.* **157**:3 (2011), 535–550. MR 2012d:14061 Zbl 1236.14035

[Charles 2013] F. Charles, "The Tate conjecture for K3 surfaces over finite fields", *Invent. Math.* **194**:1 (2013), 119–145. MR 3103257 Zbl 1282.14014

[Deligne 1974] P. Deligne, "La conjecture de Weil, I", *Inst. Hautes Études Sci. Publ. Math.* 43 (1974), 273–307. MR 49 #5013 Zbl 0287.14001

[Deligne et al. 1982] P. Deligne, J. S. Milne, A. Ogus, and K.-y. Shih, *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Mathematics **900**, Springer, Berlin, 1982. MR 84m:14046 Zbl 0465.00010

[Elsenhans and Jahnel 2008a] A.-S. Elsenhans and J. Jahnel, "K3 surfaces of Picard rank one and degree two", pp. 212–225 in *Algorithmic number theory*, edited by A. J. van der Poorten and A. Stein, Lecture Notes in Comput. Sci. **5011**, Springer, Berlin, 2008. MR 2010h:11102 Zbl 1205.11073

[Elsenhans and Jahnel 2008b] A.-S. Elsenhans and J. Jahnel, "K3 surfaces of Picard rank one which are double covers of the projective plane", pp. 63–77 in *Higher-dimensional geometry over finite fields*, edited by D. Kaledin and Y. Tschinkel, NATO Sci. Peace Secur. Ser. D Inf. Commun. Secur. **16**, IOS, Amsterdam, 2008. MR 2009j:14047 Zbl 1182.14036

[Elsenhans and Jahnel 2012] A.-S. Elsenhans and J. Jahnel, "Kummer surfaces and the computation of the Picard group", *LMS J. Comput. Math.* **15** (2012), 84–100. MR 2911278

[van Geemen 2008] B. van Geemen, "Real multiplication on K3 surfaces and Kuga–Satake varieties", *Michigan Math. J.* **56**:2 (2008), 375–399. MR 2009k:14073 Zbl 1161.14027

[Hassett and Várilly-Alvarado 2013] B. Hassett and A. Várilly-Alvarado, "Failure of the Hasse principle on general K3 surfaces", *J. Inst. Math. Jussieu* **12**:4 (2013), 853–877. MR 3103134

[Hassett et al. 2011] B. Hassett, A. Várilly-Alvarado, and P. Varilly, "Transcendental obstructions to weak approximation on general K3 surfaces", *Adv. Math.* **228**:3 (2011), 1377–1404. MR 2012i:14025 Zbl 1228.14030

[Hassett et al. 2013] B. Hassett, A. Kresch, and Y. Tschinkel, "Effective computation of Picard groups and Brauer-Manin obstructions of degree two K3 surfaces over number fields", *Rend. Circ. Mat. Palermo* (2) **62**:1 (2013), 137–151. MR 3031574

[de Jong and Katz 2000] A. J. de Jong and N. M. Katz, "Monodromy and the Tate conjecture: Picard numbers and Mordell–Weil ranks in families", *Israel J. Math.* **120**:part A (2000), 47–79. MR 2002b:14026 Zbl 1067.14504

[Li and Liedtke 2012] J. Li and C. Liedtke, "Rational curves on K3 surfaces", *Invent. Math.* **188**:3 (2012), 713–727. MR 2917181 Zbl 1255.14026

[van Luijk 2007] R. van Luijk, "K3 surfaces with Picard number one and infinitely many rational points", *Algebra Number Theory* **1**:1 (2007), 1–15. MR 2008d:14058 Zbl 1123.14022

[Maulik and Poonen 2012] D. Maulik and B. Poonen, "Néron–Severi groups under specialization", *Duke Math. J.* **161**:11 (2012), 2167–2206. MR 2957700 Zbl 1248.14011

[Milne 1975] J. S. Milne, "On a conjecture of Artin and Tate", *Ann. of Math.* (2) **102**:3 (1975), 517–533. MR 54 #2659 Zbl 0343.14005

[Mukai 2002] S. Mukai, "Vector bundles on a K3 surface", pp. 495–502 in *Proceedings of the International Congress of Mathematicians, Vol. II* (Beijing, 2002), edited by T. Li, Higher Ed. Press, Beijing, 2002. MR 2004b:14067 Zbl 1046.14016

[Nygaard 1983] N. O. Nygaard, "The Tate conjecture for ordinary K3 surfaces over finite fields", *Invent. Math.* **74**:2 (1983), 213–237. MR 85h:14012 Zbl 0557.14002

[Nygaard and Ogus 1985] N. Nygaard and A. Ogus, "Tate's conjecture for K3 surfaces of finite height", *Ann. of Math.* (2) **122**:3 (1985), 461–507. MR 87h:14014 Zbl 0591.14005

[Pink 1998] R. Pink, "l-adic algebraic monodromy groups, cocharacters, and the Mumford–Tate conjecture", *J. Reine Angew. Math.* **495** (1998), 187–237. MR 98m:11060 Zbl 0920.14006

[Serre 1981] J.-P. Serre, Letter to K. Ribet of January 1, 1981. Printed in his *Œuvres*, vol. IV: 1985–1998, pp. 1–17, Springer, Berlin, 1986.

[Serre 1998] J.-P. Serre, *Abelian l-adic representations and elliptic curves*, Research Notes in Mathematics **7**, A K Peters Ltd., Wellesley, MA, 1998. Revised reprint of the 1968 original. MR 98g:11066 Zbl 0902.14016

[Shioda 1981] T. Shioda, "On the Picard number of a complex projective variety", *Ann. Sci. École Norm. Sup.* (4) **14**:3 (1981), 303–321. MR 83i:14005 Zbl 0498.14018

[Shioda 1983] T. Shioda, "Algebraic cycles on a certain hypersurface", pp. 271–294 in *Algebraic geometry* (Tokyo and Kyoto, 1982), edited by M. Raynaud and T. Shioda, Lecture Notes in Math. **1016**, Springer, Berlin, 1983. MR 85d:14015 Zbl 0579.14005

[Tankeev 1990] S. G. Tankeev, "Surfaces of K3 type over number fields and the Mumford–Tate conjecture", *Izv. Akad. Nauk SSSR Ser. Mat.* **54**:4 (1990), 846–861. In Russian; translated in *Math. USSR-Izv.* **37**:1 (1991), 191–208. MR 91m:11044

[Tankeev 1995] S. G. Tankeev, "Surfaces of K3 type over number fields and the Mumford–Tate conjecture. II", *Izv. Ross. Akad. Nauk Ser. Mat.* **59**:3 (1995), 179–206. In Russian; translated in *Izv. Math.* **59**:3 (1995), 619–646. MR 97d:11104 Zbl 0895.14011

[Tate 1966] J. Tate, "On the conjectures of Birch and Swinnerton-Dyer and a geometric analog", exposé no. 306 in *Séminaire Bourbaki 1965/66*, Benjamin, New York, 1966. Reprinted in *Dix exposés sur la cohomologie des schémas*, Adv. Stud. Studies Pure Math. **3**, 189–214, North-Holland, Amsterdam, 1968, and in *Sém. Bourbaki*, vol. 9, 415–440, Soc. Mat. de France, 1995. Zbl 0199.55604

[Terasoma 1985] T. Terasoma, "Complete intersections with middle Picard number 1 defined over **Q**", *Math. Z.* **189**:2 (1985), 289–296. MR 86f:14010 Zbl 0579.14006

[Vasiu 2008] A. Vasiu, "Some cases of the Mumford–Tate conjecture and Shimura varieties", *Indiana Univ. Math. J.* **57**:1 (2008), 1–75. MR 2010a:14082 Zbl 1173.11039

[Voisin 2002] C. Voisin, *Théorie de Hodge et géométrie algébrique complexe*, Cours Spécialisés **10**, Société Mathématique de France, Paris, 2002. MR 2005c:32024a Zbl 1032.14001

[Zarhin 1983] Y. G. Zarhin, "Hodge groups of K3 surfaces", *J. Reine Angew. Math.* **341** (1983), 193–220. MR 84g:14009

Communicated by János Kollár
Received 2012-02-01 Revised 2012-09-20 Accepted 2012-11-04

francois.charles@univ-rennes1.fr *Institut de Recherche Mathématiques de Rennes, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes, France*
http://perso.univ-rennes1.fr/francois.charles/

ALGEBRA AND NUMBER THEORY 8:1 (2014)
dx.doi.org/10.2140/ant.2014.8.19

Adèle residue symbol and Tate's central extension for multiloop Lie algebras

Oliver Braunling

We generalize the linear algebra setting of Tate's central extension to arbitrary dimension. In general, one obtains a Lie $(n+1)$-cocycle. We compute it to some extent. The construction is based on a Lie algebra variant of Beilinson's adelic multidimensional residue symbol, generalizing Tate's approach to the local residue symbol for 1-forms on curves.

Firstly, recall that to every Lie algebra $\mathfrak{g}$ one can associate its loop Lie algebra $\mathfrak{g}[t^{\pm}]$. Iterating this construction, we obtain *multiloop Lie algebras* $\mathfrak{g}[t_1^{\pm 1}, \dots, t_n^{\pm 1}]$. To begin with, we show that various classes of interesting multiloop Lie algebras can all be embedded into a large (infinite-dimensional) Lie algebra:

Theorem 1. *Let k be a field and $n \geq 1$. There is a universal Lie algebra $\mathfrak{G}$ naturally containing, simultaneously,*

(1) *the abelian Lie algebra $k[t_1^{\pm 1}, \dots, t_n^{\pm 1}]$,*

(2) *Lie algebras of derivations, e.g., spanned by*

$$t_1^{s_1} \cdots t_n^{s_n} \partial_{t_i} \quad (\textit{acting on } k[t_1^{\pm 1}, \dots, t_n^{\pm 1}]),$$

(3) *for any finite-dimensional simple Lie algebra $\mathfrak{g}$, the multiloop algebra*

$$\mathfrak{g}[t_1^{\pm 1}, \dots, t_n^{\pm 1}].$$

The universal Lie algebra $\mathfrak{G}$ has a canonical Lie $(n+1)$-cocycle $\phi \in H^{n+1}(\mathfrak{G}, k)$. For $n = 1$ this cocycle determines a central extension

$$0 \to k \to \widehat{\mathfrak{G}} \to \mathfrak{G} \to 0$$

(known as Tate's central extension) and the pullback of it to one of the above types of subalgebras yields (respectively)

(1) *the Heisenberg algebra,*

This work has been supported by the DFG SFB/TR 45 "Periods, moduli spaces and arithmetic of algebraic varieties" and the Alexander von Humboldt Foundation.

MSC2010: primary 17B56, 17B67; secondary 32A27.

Keywords: adèle, residue symbol, Tate central extension, Kac–Moody, Japanese group.

(2) *the Virasoro algebra*,

(3) *the affine Lie algebra* $\widehat{\mathfrak{g}}$ *associated to* $\mathfrak{g}$.

This will be stated in more detail and proven in Section 6. It is not at all surprising that some Lie algebras can be embedded into larger ones. The interesting fact is that there is such a Lie algebra which carries a canonical cocycle, inducing the ones defining all these classical central extensions. For $n = 1$ the above is well-known — see, for example, [Beilinson et al. 1991, §2.1]. For $n = 1, 2$, see [Frenkel and Zhu 2012]. In the language of the latter, $\mathfrak{G}$ is an example of a "master Lie algebra".

We are interested in the nature of ϕ for $n > 1$ — even if such cocycles cannot be interpreted as a central extension anymore (we get crossed modules, etc.). Indeed, they are meaningful, as we shall see.

A key point of this text is the actual computation of ϕ (with a slight limitation):

Theorem 2. *The cocycle $\phi \in H^{n+1}(\mathfrak{G}, k)$ is given explicitly by*

$$\begin{aligned}&\phi(f_0 \wedge f_1 \wedge \cdots \wedge f_n)\\ &= \operatorname{tr} \sum_{\pi \in \mathfrak{S}_n} \operatorname{sgn} \pi \sum_{\gamma_1, \ldots, \gamma_n \in \{\pm\}} (-1)^{\gamma_1 + \cdots + \gamma_n} (P_1^{-\gamma_1} \operatorname{ad}(f_{\pi(1)}) P_1^{\gamma_1}) \cdots (P_n^{-\gamma_n} \operatorname{ad}(f_{\pi(n)}) P_n^{\gamma_n}) f_0,\end{aligned}$$

whenever $f_0 \otimes f_1 \wedge \cdots \wedge f_n$ is already a $\mathfrak{g}$-valued Lie cycle. The $P_1^+, \ldots, P_n^+$ refer to certain commuting idempotents (see Section 4 for details).

The proof and details regarding the $P_i^\pm$ can be found in Section 6. Effectively, we compute the composition

$$H_n(\mathfrak{g}, \mathfrak{g}) \xrightarrow{I} H_{n+1}(\mathfrak{g}, k) \to k, \tag{0-1}$$

with I a natural map to be explained in Section 2. By the universal coefficient theorem for Lie algebras, $H^{n+1}(\mathfrak{g}, k) \cong H_{n+1}(\mathfrak{g}, k)^*$, referring to the dual space. As such, although ϕ is well-defined, the formula only applies to those cycles admitting a lift under I (as soon as it exists, the choice does not matter). The formula is rather complicated. However, the pullback to particular subalgebras of $\mathfrak{G}$ can be much nicer; for example for multiloop Lie algebras of simple Lie algebras, we get the following:

Theorem 3. *Suppose $\mathfrak{g}/k$ is a finite-dimensional centerless Lie algebra* (*e.g.*, *simple*). *For $Y_0, \ldots, Y_n \in \mathfrak{g}$ we call*

$$B(Y_0, \ldots, Y_n) := \operatorname{tr}_{\operatorname{End}_k(\mathfrak{g})}(\operatorname{ad}(Y_0) \operatorname{ad}(Y_1) \cdots \operatorname{ad}(Y_n))$$

the "generalized Killing form". Then on all Lie cycles admitting a lift under I as in (0-1), *the pullback of ϕ to $\mathfrak{g}[t_1^{\pm}, \dots, t_n^{\pm}]$ is explicitly given by*

$$\phi(Y_0 t_1^{c_{0,1}} \cdots t_n^{c_{0,n}} \wedge \cdots \wedge Y_n t_1^{c_{n,1}} \cdots t_n^{c_{n,n}}) \\ = (-1)^n \sum_{\pi \in \mathfrak{S}_n} \operatorname{sgn} \pi \; B(Y_{\pi(1)}, \dots, Y_{\pi(n)}, Y_0) \prod_{i=1}^{n} c_{\pi(i),i}$$

whenever $\sum_{p=0}^{n} c_{p,i} = 0$ for all $i \in \{1, \dots, n\}$, and vanishes otherwise. Here $c_{i,p} \in \mathbb{Z}$ for all $i = 0, \dots, n$ and $p = 1, \dots, n$.

If $\mathfrak{g}$ is finite-dimensional simple and $n = 1$, then the class ϕ yields the universal central extension of the loop Lie algebra $\mathfrak{g}[t_1, t_1^{-1}]$, the associated affine Lie algebra $\widehat{\mathfrak{g}}$ (without extending by a derivation),

$$0 \to k \to \widehat{\mathfrak{g}} \to \mathfrak{g}[t_1, t_1^{-1}] \to 0.$$

In this case B is obviously just the ordinary Killing form of $\mathfrak{g}$. The above theorem will be proven in Section 8.

Additionally, we should say that these computations have an application outside the theory of Lie algebras. For this we need to return to the roots of the subject. J. Tate [1968] showed that the residue of a rational 1-form $f\, \mathrm{d}g$ at a closed point x on an algebraic curve X/k can be expressed as a certain operator-theoretic trace on an infinite-dimensional space. Arbarello, de Concini and Kac [Arbarello et al. 1989, eq. (2.7)] reformulated this as

$$\operatorname{res}_x f\, \mathrm{d}g = \operatorname{tr}([\pi, g] f). \tag{0-2}$$

On the right-hand side the functions f, g are to be read as multiplication operators acting on the local field $\operatorname{Frac} \widehat{\mathscr{O}}_{X,x} \simeq \kappa(x)((t_1))$, seen as a $\kappa(x)$-vector space, and π denotes some projector on the nonprincipal part, i.e., "we cut off the principal part of the Laurent series." It is natural to ask whether there exists a generalization of this formula to higher residues. We can give such a formula; it will be proven in Section 7:

Theorem 4. *For a multiple Laurent polynomial ring with residue field k, say*

$$R := k[t_1^{\pm}, \dots, t_n^{\pm}],$$

and $f_0, \dots, f_n \in R$ we have

$$\operatorname{res}_{t_1} \dots \operatorname{res}_{t_n} f_0\, \mathrm{d}f_1 \dots \mathrm{d}f_n \\ = (-1)^n \operatorname{tr} \sum_{\pi \in \mathfrak{S}_n} \operatorname{sgn} \pi \sum_{\gamma_1 \dots \gamma_n \in \{\pm\}} (-1)^{\gamma_1 + \cdots + \gamma_n} \\ \times (P_1^{-\gamma_1} \operatorname{ad}(f_{\pi(1)}) P_1^{\gamma_1}) \cdots (P_n^{-\gamma_n} \operatorname{ad}(f_{\pi(n)}) P_n^{\gamma_n}) f_0,$$

where $P_1^{\pm}, \dots, P_n^{\pm}$ are suitable projectors (explained in Section 7; see (7-3)).

(1) *For $n = 1$ and $\pi := P_1^{+}$ the formula reduces to the familiar* (0-2) (as in [Arbarello et al. 1989]).

(2) *If we have $f_i = t_1^{c_{i,1}} \cdots t_n^{c_{i,n}}$ for $i = 0, \dots, n$, the formula reduces to*

$$\operatorname{res} f_0 \, \mathrm{d}f_1 \cdots \mathrm{d}f_n = \det \begin{pmatrix} c_{1,1} & \cdots & c_{n,1} \\ \vdots & \ddots & \vdots \\ c_{1,n} & \cdots & c_{n,n} \end{pmatrix} \qquad \text{if } \sum_{p=0}^{n} c_{p,i} = 0 \text{ for all } i$$

and the residue is zero if the condition on the right-hand side is not satisfied.

(3) *For $n = 1$ and $f_1 = t_1$ this reduces by linearity to the classical definition*

$$\operatorname{res} \alpha t_1^{c_1} \, \mathrm{d}t_1 = \begin{cases} \alpha & \text{if } c_1 = -1, \\ 0 & \text{if } c_1 \neq -1. \end{cases}$$

How do we construct the cocycle ϕ?

There are various ways to approach this construction. Frenkel and Zhu [2012] use distinguished generators of the cohomology ring of infinite matrix algebras, based on computations of Feigin and Tsygan [1983]. This is a very natural approach. However, in this text we use a different approach based on the multidimensional adelic residue of [Beilinson 1980]. Originally, this approach was only used to generalize Tate's approach to the residue symbol to several variables, but it readily generalizes to the problem we are discussing here. This might be interesting also since Beilinson does not give an explicit formula — and it is not totally trivial to extrapolate a formula from the definition.

Theorem 5. *The formula in Theorem 4 arises from the construction of Beilinson* (in Lemma 1 of [Beilinson 1980]), *i.e., it is the composition*

$$\Omega^n_{R/k} \xrightarrow{(-1)^n \varkappa} H^{\mathrm{Lie}}_{n+1}(\mathfrak{G}, k) \xrightarrow{\rho_2} {}^{\wedge}E^{n+1}_{0,n+1} \xrightarrow{(d_{n+1})^{-1}} {}^{\wedge}E^{n+1}_{n+1,1} \xrightarrow{\rho_1} H^{\mathrm{Lie}}_0(\mathfrak{G}, N^{n+1}) \xrightarrow{\mathrm{tr}} k, \quad \text{(0-3)}$$

where

- $\varkappa : f_0 \, \mathrm{d}f_1 \wedge \cdots \wedge \mathrm{d}f_n \mapsto f_0 \wedge \cdots \wedge f_n$,
- N^{n+1} *is a certain $\mathfrak{G}$-module* (see Section 4 for the definition, or T_{*N} in [Beilinson 1980]), *and*
- ρ_1, ρ_2 *are edge maps and d_{n+1} a differential on the $(n+1)$-st page of a certain spectral sequence ${}^{\wedge}E^{\bullet}_{\bullet,\bullet}$* (constructed in Lemma 19, or see [Beilinson 1980, Lemma 1]).

This result is only meaningful to readers familiar with [Beilinson 1980].

The above theorem actually lies at the heart of our approach. We formulate a contracting homotopy for a mild variation of the relevant complexes in [Beilinson 1980] and then, in a slightly tedious computation, make the spectral sequence differential d_{n+1} explicit on the basis of this.

Finally, for applications in algebraic geometry, e.g., the interpretation as a local residue, it is unfortunate to interpret "loop Lie algebra" as $\mathfrak{g}[t, t^{-1}]$. It is better to work with Laurent series, i.e., $\mathfrak{g}((t))$, or even local components of adèles. Tate's original work uses the language of adèles for example. For this reason, we shall axiomatize all these variations through the notion of a "cubically decomposed algebra" (essentially taken from [Beilinson 1980], where it's not given a name).

What is not here. In the present text I only discuss the "linear algebra setting" of Tate's central extension ([Beilinson et al. 1991, §1] for the case $n = 1$). There is also a "differential operator setting" [ibid., §2], which I will treat in a future text. Roughly speaking, $\mathfrak{G}$ will be replaced by much smaller algebras of differential operators on a vector bundle.

Moreover, I do not treat the true multiloop analogue of an affine Kac–Moody algebra in the present text. Already for $n = 1$ I only consider the "plain" affine Lie algebras without extending by a derivation. From the perspective of a triangular decomposition, this is a rather horrible omission: the root spaces are infinite-dimensional! However, as the reader can probably imagine from the computations in Sections 7 and 8 the calculation gets a lot more complicated in the presence of derivations. Thus, this aspect will also be deferred to a future text. The same applies to the analogue of the plain Virasoro algebra. There should also be a nonlinear analogue, distinguished cohomology classes for multiloop groups. The cases $n = 1, 2$ (along with a higher representation theory in categories) are treated in detail by Frenkel and Zhu [2012].

One should also mention that there are completely orthogonal generalizations of Kac–Moody/Virasoro cocycles to multiloop Lie algebras — see, for example, [Frenkel 1987, §9; Neher 2011].

1. Basic framework

For an associative algebra A we shall write A_{Lie} to denote the associated Lie algebra.

Definition 6 [Beilinson 1980]. An (*n-fold*) *cubically decomposed algebra* (over a field k) is the datum $(A, (I_i^{\pm}), \tau)$:

- an associative unital (not necessarily commutative) k-algebra A;
- two-sided ideals I_i^{+}, I_i^{-} such that $I_i^{+} + I_i^{-} = A$ for $i = 1, \dots, n$;

- writing $I_i^0 := I_i^+ \cap I_i^-$ and $I_{\mathrm{tr}} := I_1^0 \cap \cdots \cap I_n^0$, a k-linear map

$$\tau : I_{\mathrm{tr,Lie}}/[I_{\mathrm{tr,Lie}}, A_{\mathrm{Lie}}] \to k.$$

For any finite-dimensional k-vector space V, certain infinite matrix algebras act naturally on the k-vector space of multiple Laurent polynomials $V[t_1^{\pm 1}, \dots, t_n^{\pm 1}]$. This yields an example of this structure — see Section 1.1. There is also an analogue for $V((t_1)) \cdots ((t_n))$, which we leave to the reader to formulate (this links to higher local fields, see [Fesenko and Kurihara 2000]). Local components of Parshin–Beilinson adèles of schemes yield another example, see [Beilinson 1980, §1]. In *loc. cit.* the ideals I_i^+, I_i^- are called X^i, Y^i. The latter gives the multidimensional generalization of the adèle formulation of Tate [1968]. See [Fesenko 2010; Huber 1991; Hübl and Yekutieli 1996; Morrow 2010] for more background on higher-dimensional adèles and their uses.

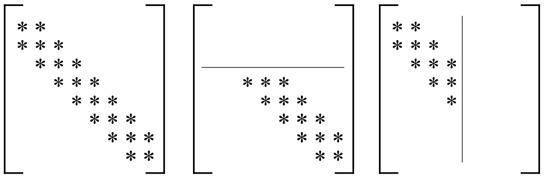
1.1. ***Infinite matrix algebras.*** Fix a field k. Let R be an associative k-algebra, not necessarily unital or commutative. Define an algebra of infinite matrices

$$E(R) := \{\phi = (\phi_{ij})_{i,j\in\mathbb{Z}}, \phi_{ij} \in R \mid \exists K_\phi : |i-j| > K_\phi \Rightarrow \phi_{ij} = 0\}. \tag{1-1}$$

Define a product by $(\phi \cdot \phi')_{ik} := \sum_{j\in\mathbb{Z}} \phi_{ij}\phi'_{jk}$, the usual matrix multiplication formula; this sum only has finitely many nonzero terms and one can choose $K_{\phi\phi'} := K_\phi + K_{\phi'}$. Then $E(R)$ becomes an associative k-algebra. If R is unital, $E(R)$ is also unital. E is a functor from associative algebras to associative algebras; for a morphism $\varphi : R \to S$ there is an induced morphism $E(\varphi) : E(R) \to E(S)$ by using φ entry-by-entry, i.e., $(E(\varphi)\phi)_{ij} := \varphi(\phi_{ij})$. If $I \subseteq R$ is an ideal (which is in particular a nonunital associative ring), $E(I) \subseteq E(R)$ is an ideal. Moreover, for ideals I_1, I_2 one has $E(I_1 \cap I_2) = E(I_1) \cap E(I_2)$ and $E(I_1 + I_2) = E(I_1) + E(I_2)$, as a sum of ideals. Next, define

$$I^+(R) := \{\phi \in E(R) \mid \exists B_\phi : i < B_\phi \Rightarrow \phi_{ij} = 0\},$$
$$I^-(R) := \{\phi \in E(R) \mid \exists B_\phi : j > B_\phi \Rightarrow \phi_{ij} = 0\}$$

and one checks easily that $I^+(R), I^-(R)$ are two-sided ideals in $E(R)$. The following figure attempts to visualize the shape of the matrices in $E(R)$, $I^+(R)$ and $I^-(R)$, respectively:



Define $I^0(R) := I^+(R) \cap I^-(R)$ and one checks that

$$I^0(R) := \{\phi \in E(R) \mid \phi_{ij} = 0 \text{ for all but finitely many } (i, j)\}.$$

There is a trace morphism

$$\operatorname{tr} : I^0(R) \to R, \quad \operatorname{tr} \phi := \sum_{i \in \mathbb{Z}} \phi_{ii}; \tag{1-2}$$

the sum is obviously finite. One easily verifies that $\operatorname{tr}[\phi, \phi'] = \sum_{i,j \in \mathbb{Z}} [\phi_{ij}, \phi'_{ji}]$ and thus $\operatorname{tr}[I^0(R), E(R)] \subseteq [R, R]$. More generally, if $R' \subseteq R$ is a subalgebra,

$$\operatorname{tr}[I^0(R'), E(R)] \subseteq [R', R].$$

We note that this trace does not necessarily vanish on commutators. Moreover, every $\phi \in E(R)$ can be written as $\phi = \phi^+ + \phi^-$ with $\phi^+_{ij} := \delta_{i \geq 0} \phi_{ij}$ (for this R need not be unital, use ϕ_{ij} for $i \geq 0$ and 0 otherwise) and $\phi^- = \phi - \phi^+$. One checks that $\phi^\pm \in I^\pm(R)$. It follows that $I^+(R) + I^-(R) = E(R)$.

Finally, let M be an R-bimodule (over k, i.e., a left-$(A \otimes_k A^{op})$-module; R-bimodules form an abelian category). Analogously to $E(R)$, define

$$E(M) := \{\phi = (\phi_{ij})_{i,j \in \mathbb{Z}}, \phi_{ij} \in M \mid \exists K_\phi : |i - j| > K_\phi \Rightarrow \phi_{ij} = 0\}. \tag{1-3}$$

Again using the matrix multiplication formula, $E(M)$ is an $E(R)$-bimodule. If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of R-bimodules, $0 \to E(M') \to E(M) \to E(M'') \to 0$ is an exact sequence of $E(R)$-bimodules. Note that for an ideal $I \subseteq R$ the object $E(I)$ is well-defined, regardless of whether we regard I as an associative ring as in (1-1) or an R-bimodule as in (1-3).

Now let V be a finite-dimensional k-vector space and R_0 an arbitrary unital subalgebra of $\operatorname{End}_k(V)$. Define $R_i := E(R_{i-1})$ for $i = 1, \dots, n$. Note that via $k \to R_0$, $\alpha \mapsto \alpha \cdot \mathbb{1}_{\operatorname{End}_k(V)}$, k is embedded into the center of R_i. Then $R_n = (E \circ \cdots \circ E)(R_0)$ is a unital associative k-algebra. Its elements may be indexed $\phi = (\phi_{(i_n,j_n),\dots,(i_1,j_1) \in \mathbb{Z}^{2n}} \in R_0)$. By the properties discussed above,

$$I_i^\pm := (\underset{n}{E} \cdots \underset{i+1}{E} \circ \underset{i}{I^\pm} \circ \underset{i-1}{E} \cdots \underset{1}{E})(R_0) \quad (I^\pm \text{ in the } i\text{-th place})$$

is an ideal in R_n (we use centered subscripts only to emphasize the numbering). Moreover,

$$\begin{aligned} I_i^+ + I_i^- &= (E \cdots E \circ I^+ \circ E \cdots E)(R_0) + (E \cdots E \circ I^- \circ E \cdots E)(R_0) \\ &= (E \cdots E \circ E \circ E \cdots E)(R_0) = R_n. \end{aligned}$$

By composing the traces of (1-2) we arrive at a k-linear map τ,

$$\begin{aligned}\tau : I_{\mathrm{tr}} &= I_1^0 \cap \cdots \cap I_n^0 \\ &= (I^0 \circ \cdots \circ I^0)(R_0) \xrightarrow{\mathrm{tr}} \cdots \xrightarrow{\mathrm{tr}} I^0(I^0(R_0)) \xrightarrow{\mathrm{tr}} I^0(R_0) \xrightarrow{\mathrm{tr}} R_0 \xrightarrow{\mathrm{Tr}} k,\end{aligned}$$

where "Tr" (as opposed to "tr") denotes the ordinary matrix trace of $\mathrm{End}_k(V)$ ($\supseteq R_0$). Here we have used that V is finite-dimensional over k. Using inductively the relation

$$\mathrm{tr}[I^0(R'), E(R)] \subseteq [R', R]$$

(valid for subalgebras $R' \subseteq R$), one sees that

$$\begin{aligned}\tau[I_{\mathrm{tr}}, R_n] &= \mathrm{Tr}(\mathrm{tr} \circ \cdots \circ \mathrm{tr} \circ \mathrm{tr})[I^0(I^0(\cdots)), E(E(\cdots))] \\ &\subseteq \mathrm{Tr}(\mathrm{tr} \circ \cdots \circ \mathrm{tr})[I^0(\cdots), E(\cdots)] \subseteq \mathrm{Tr}[R_0, R_0] = 0\end{aligned}$$

since the ordinary trace Tr vanishes on commutators. Hence, τ factors to a morphism $\tau : I_{\mathrm{tr,Lie}}/[I_{\mathrm{tr,Lie}}, R_{\mathrm{Lie}}] \to k$. Summarizing, for every $n \geq 1$, every finite-dimensional k-vector space V and every unital subalgebra $R_0 \subseteq \mathrm{End}_k(V)$, $(R_n, (I_i^{\pm}), \tau)$ is a cubically decomposed algebra.

Finally, note that for any associative algebra R, $E(R)$ is a right-R-submodule of *right*-R-module endomorphisms $\mathrm{End}_R(R[t, t^{-1}])$ of $R[t, t^{-1}]$. Write elements as $a = \sum_{i \in \mathbb{Z}} a_i t^i$, also denoted $a = (a_i)_i$ with $a_i \in R$, and let $\phi = (\phi_{ij})$ act by $(\phi \cdot a)_i := \sum_k \phi_{ik} a_k$. Moreover, each $a \in R[t, t^{-1}]$ determines a right-R-module endomorphism via the multiplication operator $x \mapsto a \cdot x$. We find

$$R[t, t^{-1}] \hookrightarrow E(R) \hookrightarrow \mathrm{End}_R(R[t, t^{-1}]).$$

Multiplication with t^i is represented by a matrix with a diagonal $\ldots, 1, 1, 1, \ldots,$ shifted by i off the principal diagonal. Inductively,

$$R_0[t_1^{\pm 1}, \ldots, t_n^{\pm 1}] \hookrightarrow R_n \hookrightarrow \mathrm{End}_{R_0}(R_0[t_1^{\pm 1}, \ldots, t_n^{\pm 1}]). \tag{1-4}$$

See for example [Jimbo and Miwa 1983, §1; Kac and Raina 1987, Lec. 4] for more information regarding the case $n = 1$ and [Frenkel and Zhu 2012, §3] for a similar procedure when $n = 2$.

2. Modified Chevalley–Eilenberg complexes

Suppose k is a field and $\mathfrak{g}$ a Lie algebra over k. We recall that for any $\mathfrak{g}$-module the conventional Chevalley–Eilenberg complex is given by $C(M)_r := M \otimes \bigwedge^r \mathfrak{g}$ along with the differential

$$\delta := \delta^{[1]} + \delta^{[2]} : C(M)_r \to C(M)_{r-1}, \tag{2-1}$$

$$\delta^{[1]}(f_0 \otimes f_1 \wedge \cdots \wedge f_r) := \sum_{i=1}^{r} (-1)^i [f_0, f_i] \otimes f_1 \wedge \cdots \wedge \widehat{f_i} \wedge \cdots \wedge f_r,$$

$$\delta^{[2]}(f_0 \otimes f_1 \wedge \cdots \wedge f_r) := \sum_{1 \le i < j \le r} (-1)^{i+j+1} f_0 \otimes [f_i, f_j] \wedge f_1 \wedge \cdots \widehat{f_i} \cdots \widehat{f_j} \cdots \wedge f_r$$

for $f_0 \in M$ and $f_1, \dots, f_r \in \mathfrak{g}$. Its homology is (by definition, if one wants) Lie homology with coefficients in M. There is also a cohomological analogue; for details see, for example, [Loday 1992, Chapter 10]. We may view k itself as a $\mathfrak{g}$-module with the trivial structure. There is an obvious morphism

$$I : C(\mathfrak{g})_r \to C(k)_{r+1}, \quad f_0 \otimes f_1 \wedge \cdots \wedge f_r \mapsto (-1)^r \mathbb{1}_k \otimes f_0 \wedge f_1 \wedge \cdots \wedge f_r, \tag{2-2}$$

and one checks easily that this commutes with the respective differentials and thus induces morphisms $H_r(\mathfrak{g}, \mathfrak{g}) \to H_{r+1}(\mathfrak{g}, k)$. The linear dual $\mathfrak{g}^* := \operatorname{Hom}_k(\mathfrak{g}, k)$ is canonically a $\mathfrak{g}$-module via $(f \cdot \varphi)(g) := \varphi([g, f])$ for $\varphi \in \mathfrak{g}^*$ and $f, g \in \mathfrak{g}$. The cohomological analogue of (2-2) is the morphism $I : H^{r+1}(\mathfrak{g}, k) \to H^r(\mathfrak{g}, \mathfrak{g}^*)$ given by

$$(I\phi)(f_1 \wedge \cdots \wedge f_r)(f_0) := (-1)^r \phi(f_0 \wedge f_1 \wedge \cdots \wedge f_r).$$

Remark 7. These maps could be viewed as a Lie-theoretic analogue of map I in Connes' periodicity sequence — see [Loday 1992, §2.2]. We may view $H_{*-1}(\mathfrak{g}, \mathfrak{g})$ as a partial "noncyclic" counterpart of Lie homology. The true Hochschild analogue would be Leibniz homology — see [Loday 1992, §10.6]. For the present purposes, however, we have no use for this analogue.

Let $\mathfrak{j} \subseteq \mathfrak{g}$ be a Lie ideal. As such, it is a $\mathfrak{g}$-module and we may consider $C(\mathfrak{j})_\bullet$. Following [Beilinson 1980] we may work with a "cyclically symmetrized" counterpart: We write $\mathfrak{j} \wedge \bigwedge^{r-1}\mathfrak{g}$ to denote the $\mathfrak{g}$-submodule of $\mathfrak{g} \wedge \bigwedge^{r-1}\mathfrak{g} = \bigwedge^r \mathfrak{g}$ generated by elements $j \wedge f_1 \wedge \cdots \wedge f_{r-1}$ such that $j \in \mathfrak{j}$ and $f_1, \dots, f_{r-1} \in \mathfrak{g}$. If $\mathfrak{j}_i$, $i = 1, 2, \dots$, are Lie ideals, we denote by $\left(\bigoplus_i \mathfrak{j}_i\right) \wedge \bigwedge^{r-1}\mathfrak{g}$ the module $\bigoplus_i \left(\mathfrak{j}_i \wedge \bigwedge^{r-1}\mathfrak{g}\right)$.

Example 8. If $k\langle s, t, u\rangle$ and $k\langle s\rangle$ denote a 3-dimensional abelian Lie algebra along with a 1-dimensional Lie ideal, then $\bigwedge^2 k\langle s, t, u\rangle$ is 3-dimensional with basis $s \wedge t$, $s \wedge u$ and $t \wedge u$. Then $k\langle s\rangle \wedge k\langle s, t, u\rangle$ is 2-dimensional with basis $s \wedge t$, $s \wedge u$.

The k-vector spaces $CE(\mathfrak{j})_r := \mathfrak{j} \wedge \bigwedge^{r-1}\mathfrak{g}$ (for $r \ge 1$) and $CE(\mathfrak{j})_0 := k$ define a subcomplex of $C(k)_\bullet$. In particular, the differential is given by

$$\delta(f_0 \wedge f_1 \wedge \cdots \wedge f_r) := \sum_{0 \le i < j \le r} (-1)^{i+j} [f_i, f_j] \wedge f_0 \wedge \cdots \widehat{f_i} \cdots \widehat{f_j} \cdots \wedge f_r. \tag{2-3}$$

It is well-defined since $\mathfrak{j}$ is a Lie ideal. We get morphisms generalizing I, notably $H_r(\mathfrak{g},\mathfrak{j}) \to H_{r+1}(CE(\mathfrak{j}))$ via $\mathfrak{j} \otimes \bigwedge^r \mathfrak{g} \to \mathfrak{j} \wedge \bigwedge^r \mathfrak{g}$ and analogously $H^{r+1}(CE(\mathfrak{j})) \to H^r(\mathfrak{g},\mathfrak{j}^*)$. We have resisted the temptation to reindex $CE(-)_\bullet$ despite the unpleasant (+1)-shift in (2-2) in order to remain compatible with standard usage in the following sense:

Lemma 9 [Beilinson 1980, Lemma 1(a)]. *$CE(\mathfrak{g})_\bullet$ is a complex of k-vector spaces and is quasi-isomorphic to $k \otimes^{\mathbf{L}}_{U\mathfrak{g}} k$. In particular*

$$H_i(\mathfrak{g},k) = H_i(CE(\mathfrak{g})_\bullet) \quad \textit{and} \quad H^i(\mathfrak{g},k) = H^i(\mathrm{Hom}_k(CE(\mathfrak{g})_\bullet, k)).$$

Proof. As we have explained above, $CE(\mathfrak{g})_\bullet$ agrees with the standard Chevalley–Eilenberg complex and the latter is well-known to represent $k \otimes^{\mathbf{L}}_{U\mathfrak{g}} k$. □

We easily compute

$$\begin{aligned} H_0(\mathfrak{g},\mathfrak{j}) &\xrightarrow[I]{\cong} H_1(CE(\mathfrak{j})) \cong \mathfrak{j}/[\mathfrak{g},\mathfrak{j}], \\ H^1(CE(\mathfrak{j})) &\xrightarrow[I]{\cong} H^0(\mathfrak{g},\mathfrak{j}^*) \cong (\mathfrak{j}/[\mathfrak{g},\mathfrak{j}])^*. \end{aligned} \tag{2-4}$$

In higher degrees the map I ceases to be an isomorphism.

Nonetheless, this computation hints at the principle of computation which we shall use below. Beilinson [1980] uses $CE(-)_\bullet$, whereas we will only be able to do manageable computations with $C(-)_\bullet$. The map I will serve to deduce facts about $CE(-)_\bullet$ while working with $C(-)_\bullet$.

3. Cubically decomposed algebras

Let $(A, (I_i^\pm), \tau)$ be an n-fold cubically decomposed algebra (Definition 6) over a field k; that is, we are given the following data:

- an associative unital (not necessarily commutative) k-algebra A;
- two-sided ideals I_i^+, I_i^- such that $I_i^+ + I_i^- = A$ for $i = 1,\dots,n$;
- writing $I_i^0 := I_i^+ \cap I_i^-$ and $I_{\mathrm{tr}} := I_1^0 \cap \dots \cap I_n^0$, a k-linear map

$$\tau : I_{\mathrm{tr,Lie}}/[I_{\mathrm{tr,Lie}}, A_{\mathrm{Lie}}] \to k.$$

See Section 1 to see how this type of structure arises. As a shorthand, define $\mathfrak{g} := A_{\mathrm{Lie}}$. For any elements $s_1,\dots,s_n \in \{+,-,0\}$ we define the *degree* of the n-tuple $(s_1,\dots,s_n)$ as

$$\deg(s_1 \dots s_n) := 1 + \#\{i \mid s_i = 0\}.$$

Next, following [Beilinson 1980], we construct complexes of $\mathfrak{g}$-modules:

Definition 10 [Beilinson 1980]. For every $1 \leq p \leq n+1$ define

$$
{}^{\wedge}T^p_\bullet := \coprod_{\substack{s_1 \dots s_n \in \{\pm,0\} \\ \deg(s_1 \dots s_n)=p}} \bigcap_{i=1}^{n} \begin{cases} CE(I_i^+)_\bullet & \text{for } s_i = +, \\ CE(I_i^-)_\bullet & \text{for } s_i = -, \\ CE(I_i^+)_\bullet \cap CE(I_i^-)_\bullet & \text{for } s_i = 0, \end{cases} \tag{3-1}
$$

and ${}^{\wedge}T^0_\bullet := CE(\mathfrak{g})_\bullet$.

Each $CE(I_i^\pm)_\bullet$ is a complex and all their differentials are defined by the same formula, (2-3); hence the intersection of these complexes has a well-defined differential and is a complex itself. Same for the coproduct. The complex ${}^{\wedge}T^\bullet_\bullet$ is inspired by a cubical object used by Beilinson [1980].

Example 11. For $n = 2$ we get complexes

$$
\begin{aligned}
{}^{\wedge}T^1_\bullet &= \coprod_{s_1,s_2 \in \{\pm\}} CE(I_1^{s_1})_\bullet \cap CE(I_2^{s_2})_\bullet, \\
{}^{\wedge}T^2_\bullet &= \coprod_{s_1 \in \{\pm\}} CE(I_1^{s_1})_\bullet \cap CE(I_2^+)_\bullet \cap CE(I_2^-)_\bullet \oplus \coprod_{s_2 \in \{\pm\}} CE(I_1^+)_\bullet \cap CE(I_1^-)_\bullet \cap CE(I_2^{s_2})_\bullet, \\
{}^{\wedge}T^3_\bullet &= CE(I_1^+)_\bullet \cap CE(I_1^-)_\bullet \cap CE(I_2^+)_\bullet \cap CE(I_2^-)_\bullet.
\end{aligned}
$$

Note that $CE(I_1^+)_\bullet \cap CE(I_1^-)_\bullet \neq CE(I_1^+ \cap I_1^-)_\bullet$; for example, $I_1^+ \wedge I_1^-$ is a subspace in degree two of the left-hand side, but not of the right-hand side.

Diverging from [Beilinson 1980] we shall primarily use the following slightly different auxiliary construction (which we will later relate to the above one):

Definition 12. For $1 \leq p \leq n+1$ let

$$
{}^{\otimes}T^p_\bullet := \coprod_{\substack{s_1 \dots s_n \in \{\pm,0\} \\ \deg(s_1 \dots s_n)=p}} C(I_1^{s_1} \cap I_2^{s_2} \cap \dots \cap I_n^{s_n})_\bullet \tag{3-2}
$$

and ${}^{\otimes}T^0_\bullet := C(\mathfrak{g})_\bullet$.

So, instead of the modified Chevalley–Eilenberg complex of Section 2 we just use the standard complexes for Lie homology with suitable coefficients. Clearly the morphism $I : C(\mathfrak{g})_r \to C(k)_{r+1}$ descends to morphisms

$$
\begin{gathered}
C(\mathfrak{g})_r \supseteq C(I_i^{s_i})_r \to CE(I_i^{s_i})_{r+1} \subseteq C(k)_{r+1}, \\
\underset{\in I_i^{s_i}}{f_0} \otimes f_1 \wedge \dots \wedge f_r \mapsto (-1)^r \underset{\in I_i^{s_i}}{f_0} \wedge f_1 \wedge \dots \wedge f_r.
\end{gathered}
$$

As we take intersections of Lie ideals on the left $C(I_1^{s_1} \cap \cdots)_\bullet$, as in (3-2), the image lies in the intersection of the individual images, i.e., $CE(I_1^{s_i})_\bullet \cap \cdots$, as in

(3-1). As a result, we obtain morphisms

$$^{\otimes}T^{p}_{\bullet} \xrightarrow{I} {}^{\wedge}T^{p}_{\bullet+1} \quad \text{(for all } p\text{)},$$

and since they are a restriction of the map I to subcomplexes, this is a morphism of complexes, and thus induces maps on homology.

4. The cube complex

Next, we shall define maps $\cdots \to {}^{\otimes}T^{2}_{\bullet} \to {}^{\otimes}T^{1}_{\bullet} \to {}^{\otimes}T^{0}_{\bullet} \to 0$, so that $({}^{\otimes}T_{\bullet})^{\bullet}$ becomes an exact superscript-indexed complex (of subscript-indexed complexes); and the same for ${}^{\wedge}T^{\bullet}_{\bullet}$. We begin by discussing ${}^{\otimes}T^{\bullet}_{\bullet}$.

We define a $\mathfrak{g}$-module $N^0 := \mathfrak{g}$ and for $p \geq 1$

$$N^p := \coprod_{s_1 \dots s_n \in \{+,-,0\}} I_1^{s_1} \cap I_2^{s_2} \cap \cdots \cap I_n^{s_n} \quad (\text{with } \deg(s_1 \dots s_n) = p). \tag{4-1}$$

We shall denote the components $f = (f_{s_1 \dots s_n})$ of elements in N^p with indices in terms of $s_1, \dots, s_n \in \{+, -, 0\}$. Clearly $N^p = 0$ for $p > n+1$. We shall treat all N^p as $\mathfrak{g}$-modules and observe that

$$^{\otimes}T^{p}_{\bullet} = C(N^p)_{\bullet}$$

(by definition!), so by the functoriality and flatness[1] of $C_{\bullet}$ it suffices to construct an exact complex $N^{\bullet}$ out of the N^p and then ${}^{\otimes}T^{p}_{\bullet}$ will be an exact complex in p.

Example 13. For $n = 1$ we have

$$N^2 = I_1^0, \qquad N^1 = I_1^+ \oplus I_1^-$$

and elements would be denoted $f = (f_0) \in N^2$ and $g = (g_+, g_-) \in N^1$. For $n = 2$ we have

$$N^3 = I_1^0 \cap I_2^0, \quad N^2 = \Bigg(\coprod_{s_1 \in \{+,-\}} I_1^{s_1} \cap I_2^0 \Bigg) \oplus \Bigg(\coprod_{s_2 \in \{+,-\}} I_1^0 \cap I_2^{s_2} \Bigg)$$

$$N^1 = \coprod_{s_1, s_2 \in \{+,-\}} I_1^{s_1} \cap I_2^{s_2}.$$

We shall use the shorthand $s_1 \dots \pm \dots s_n$ to indicate that in the i-th place we have $s_i \in \{+, -\}$, whatever i may be at the moment. Similarly, $s_1 \dots 0 \dots s_n$ will

[1] We just tensor N^p with the vector spaces $\bigwedge^i \mathfrak{g}$. Being over a field, this preserves exact sequences.

imply that $s_i = 0$. Define $\mathfrak{g}$-module homomorphisms

$$\begin{aligned}(\partial_i f)_{s_1\ldots\pm\ldots s_n} &:= (-1)^{\#\{j \mid j>i \text{ and } s_j=0\}} f_{s_1\ldots 0\ldots s_n},\\ (\partial_i f)_{s_1\ldots 0\ldots s_n} &:= 0, \\ \partial &:= \sum_{i=1}^{n} \partial_i .\end{aligned} \tag{4-2}$$

One checks easily that $\partial_i^2 = 0$ and $\partial_i\partial_j + \partial_j\partial_i = 0$ for all $i, j = 1, \ldots, n$. As a consequence, $\partial^2 = 0$. The components are given explicitly by

$$(\partial f)_{s_1\ldots s_n} = \sum_{i=1}^{n} (\partial_i f)_{s_1\ldots s_n} = \sum_{\{i \mid s_i=+,-\}} (-1)^{\#\{j \mid j>i \text{ and } s_j=0\}} f_{s_1\ldots 0\ldots s_n}. \tag{4-3}$$

Definition 14. Let $(A, (I_i^{\pm}), \tau)$ be an n-fold cubically decomposed algebra over a field k. A *system of good idempotents* are pairwise commuting elements $P_i^+ \in A$ for $i = 1, \ldots, n$ such that for all i:

(1) $P_i^{+2} = P_i^+$.

(2) $P_i^+ A \subseteq I_i^+$.

(3) $P_i^- A \subseteq I_i^-$ (where we define $P_i^- := \mathbb{1}_A - P_i^+$).

We note that the P_i^- are also pairwise commuting idempotents and $P_i^+ + P_i^- = \mathbb{1}_A$. Next, for $s_i \in \{+, -\}$ define k-vector space homomorphisms

$$\begin{aligned}(\varepsilon_i f)_{s_1\ldots s_i\ldots s_n} &:= (-1)^{s_i} P_i^{s_i} \sum_{\gamma_i \in \{\pm\}} (-1)^{\gamma_i} f_{s_1\ldots\gamma_i\ldots s_n},\\ (\varepsilon_i f)_{s_1\ldots 0\ldots s_n} &:= 0,\end{aligned}$$

where $(-1)^{\pm} = \pm 1$. By direct calculation one verifies the identities $\varepsilon_i^2 = \varepsilon_i$ and $\varepsilon_i\varepsilon_j = \varepsilon_j\varepsilon_i$ for all $i, j = 1, \ldots, n$. Finally, define

$$\begin{aligned}(H_i f)_{s_1\ldots 0\ldots s_n} &:= (-1)^{\#\{j \mid j>i \text{ and } s_j=0\}} \sum_{\gamma_i \in \{\pm\}} P_i^{-\gamma_i} f_{s_1\ldots\gamma_i\ldots s_n},\\ (H_i f)_{s_1\ldots\pm\ldots s_n} &:= 0.\end{aligned}$$

The expression $P_i^{-\gamma_i}$ means P_i^- for $\gamma_i = +$ and P_i^+ for $\gamma_i = -$. One checks that

$$\begin{aligned} H_i^2 = 0 &\quad\text{and}\quad H_iH_j + H_jH_i = 0,\\ \partial_i\varepsilon_j = \varepsilon_j\partial_i &\quad\text{and}\quad H_i\varepsilon_j = \varepsilon_jH_i\end{aligned}$$

for all $i, j = 1, \ldots, n$. Moreover, $\partial_iH_j + H_j\partial_i = 0$ whenever $i \neq j$. In the special case $i = j$ one finds instead that

$$\partial_iH_i + H_i\partial_i = \mathbb{1} - \varepsilon_i .$$

Define $H := H_1 + \varepsilon_1 H_2 + \cdots + \varepsilon_1\varepsilon_2 \cdots \varepsilon_{n-1} H_n$. Using the identities established above, one finds very easily

$$H^2 = 0 \quad \text{and} \quad \partial H + H\partial = \mathbb{1} - \varepsilon_1 \cdots \varepsilon_n. \tag{4-4}$$

The fact $H^2 = 0$ was observed by the anonymous referee; it explains a certain cancellation in the proof of Proposition 24, which had been rather mysterious in an earlier version of this text.

Lemma 15. *An explicit formula for H is given by*

$$\begin{aligned}(Hf)_{s_1 \dots s_n} = (-1)^{\deg(s_1 \dots s_n)} &(-1)^{s_1 + \cdots + s_b} P_1^{s_1} \cdots P_i^{s_b} \\ &\times \sum_{\gamma_1 \dots \gamma_{b+1} \in \{\pm\}} (-1)^{\gamma_1 + \cdots + \gamma_b} P_{b+1}^{-\gamma_{b+1}} f_{\gamma_1 \dots \gamma_{b+1} s_{b+2} \dots s_n},\end{aligned} \tag{4-5}$$

where b denotes the largest index such that $s_1, \dots, s_b \in \{\pm\}$ or $b = 0$ if none (*and so $s_{b+1} = 0$ if $b < n$*; *$b+1$ is the index of the leftmost zero*).

Proof. One shows that

$$(\varepsilon_1 \cdots \varepsilon_i f)_{s_1 \dots s_n} = \begin{cases} (-1)^{s_1 + \cdots + s_i} P_1^{s_1} \cdots P_i^{s_i} & \\ \quad \times \sum_{\gamma_1 \dots \gamma_i \in \{\pm\}} (-1)^{\gamma_1 + \cdots + \gamma_i} f_{\gamma_1 \dots \gamma_i s_{i+1} \dots s_n} & \text{for } s_1, \dots, s_i \in \{\pm\}, \\ 0 & \text{if } 0 \in \{s_1, \dots, s_i\} \end{cases} \tag{4-6}$$

by evaluating $(\varepsilon_j \cdots \varepsilon_i f)$ inductively along $j = i, i-1, \dots, 1$. Plug in $H_{i+1} f$ for f to obtain

$$\begin{aligned}(\varepsilon_1 \cdots \varepsilon_i H_{i+1} f)_{s_1 \dots s_n} = (-1)^{\#\{j \mid j > i+1 \text{ and } s_j = 0\}} &(-1)^{s_1 + \cdots + s_i} P_1^{s_1} \cdots P_i^{s_i} \\ &\times \sum_{\gamma_1 \dots \gamma_{i+1} \in \{\pm\}} (-1)^{\gamma_1 + \cdots + \gamma_i} P_{i+1}^{-\gamma_{i+1}} f_{\gamma_1 \dots \gamma_i \gamma_{i+1} s_{i+2} \dots s_n}\end{aligned}$$

for $s_1, \dots, s_i \in \{\pm\}$ and $s_{i+1} = 0$. Otherwise, (that is, for $0 \in \{s_1, \dots, s_i\}$ or $s_{i+1} \in \{\pm\}$), the respective component is zero. Thus,

$$H_{s_1 \dots s_n} = \sum_{i=1}^{n} (\varepsilon_1 \cdots \varepsilon_i H_{i+1} f)_{s_1 \dots s_n}.$$

The summands with $i > b$ vanish since for them $0 \in \{s_1, \dots, s_i\}$. The summands with $i < b$ vanish since for them $s_{i+1} \in \{\pm\}$. Thus,

$$H_{s_1 \dots s_n} = (\varepsilon_1 \cdots \varepsilon_b H_{b+1} f)_{s_1 \dots s_n}$$

and we use the above explicit formula. Note that $\#\{j \mid j > b+1 \text{ and } s_j = 0\}$ is just one below the total number of slots with value 0 since $s_1, \dots, s_b \in \{\pm\}$ and $s_{b+1} = 0$. Thus, $(-1)^{\#\{j \mid j > i+1 \text{ and } s_j = 0\}} = (-1)^{\deg(s_1 \dots s_n)}$. □

The above maps are defined for N^p in degrees ≥ 1. We extend them to degree zero by defining the maps

$$\hat{\partial} : N^1 \to N^0 \quad \text{and} \quad \hat{H} : N^0 \to N^1$$

as follows:

$$\hat{\partial} f := \sum_{s_1 \dots s_n \in \{+,-\}} (-1)^{s_1 + \dots + s_n} f_{s_1 \dots s_n},$$
$$(\hat{H} f)_{s_1 \dots s_n} := (-1)^{s_1 + \dots + s_n} P_1^{s_1} \cdots P_n^{s_n} f. \tag{4-7}$$

Along with these, we obtain the following crucial fact:

Lemma 16. *Equipped with these morphisms,*

$$N^\bullet = \left[N^{n+1} \overset{\partial}{\underset{H}{\rightleftarrows}} N^n \overset{\partial}{\underset{H}{\rightleftarrows}} \cdots \overset{\partial}{\underset{H}{\rightleftarrows}} N^1 \overset{\hat{\partial}}{\underset{\hat{H}}{\rightleftarrows}} N^0 \right]_{n+1,0} \tag{4-8}$$

is a complex of $\mathfrak{g}$*-modules with differentials* $\partial_\bullet$ (*resp.* $\hat{\partial}$) *and contracting homotopies* $H_\bullet$ (*resp.* $\hat{H}$) *in the category of k-vector spaces.*

Proof. The identities $\partial^2 = 0$ and $\hat{\partial} \circ \partial = 0 : N^2 \to N^0$ are easy to check. Next, we confirm the contracting homotopy. We find $\partial H + H \partial = \mathbb{1} - \varepsilon_1 \cdots \varepsilon_n$ by a telescope cancellation. For $f \in N^i$ with $i \geq 2$ for each component $f_{s_1 \dots s_n}$ there must be at least one i with $s_i = 0$ and thus $\varepsilon_1 \cdots \varepsilon_n \mid_{N^i} = 0$ for $i \geq 2$. It remains to treat $i = 0$, 1. For $i = 1$ we compute

$$\hat{H} \hat{\partial} f = (-1)^{s_1 + \dots + s_n} P_1^{s_1} \cdots P_n^{s_n} \sum_{s_1 \dots s_n \in \{+,-\}} (-1)^{s_1 + \dots + s_n} f_{s_1 \dots s_n} = \varepsilon_1 \cdots \varepsilon_n f$$

(as in (4-6)). Thus, $\partial H + \hat{H} \hat{\partial} = \mathbb{1}$ on N^1. Finally, for $i = 0$ we compute $\hat{\partial} \hat{H} f = f$. □

Corollary 17. $0 \to {}^{\otimes}T_\bullet^{n+1} \to {}^{\otimes}T_\bullet^{n} \to \cdots \to {}^{\otimes}T_\bullet^{0} \to 0$ *with differential* (*and a contracting homotopy*) *induced by* $\partial \otimes \mathrm{id}_{\wedge^\bullet \mathfrak{g}}$ (*and* $H \otimes \mathrm{id}_{\wedge^\bullet \mathfrak{g}}$) *is an exact complex* (*of complexes of k-vector spaces*).

For the corollary, just use that tensoring with $\bigwedge^r \mathfrak{g}$ is exact.

5. The cube complex, II

Next, it would be nice to give a discussion of the ${}^{\wedge}T_\bullet^\bullet$ parallel to the one for ${}^{\otimes}T_\bullet^\bullet$ in the previous section. We can only do this to a limited extent, however.

Lemma 18. *The definition*

$$(\partial f)_{s_1 \dots s_n} = \sum_{\{i \mid s_i = +,-\}} (-1)^{\#\{j \mid j > i \text{ and } s_j = 0\}} f_{s_1 \dots 0 \dots s_n} \tag{5-1}$$

turns ${}^{\wedge}T_\bullet^\bullet$ *into a complex* (*of complexes of k-vector spaces*) *with respect to the superscript index. The morphisms* $I : {}^{\otimes}T_\bullet^p \to {}^{\wedge}T_{\bullet+1}^p$ *yield a morphism of complexes.*

Proof. Easy. Just check that the map ∂ is well-defined and satisfies $\partial^2 = 0$; in fact, exactly the same computation as in (4-2) applies. For the second claim, we just need to show that the map I commutes with the differential of either complex, but this is clear since the differentials are given by the same formula — compare (4-3) with (5-1). □

The complex ${}^{\wedge}T^{\bullet}_{\bullet}$ is the central object in Beilinson's construction [1980]. We will use its analogue ${}^{\otimes}T^{\bullet}_{\bullet}$ as an auxiliary computational device. Firstly, let us explain Beilinson's construction. We need the following entirely homological tool:

Lemma 19. *Suppose we are given an exact sequence*

$$S^{\bullet} = [S^{n+1} \to S^n \to \cdots \to S^0]_{n+1,0}$$

with entries in $\mathbf{Ch}^{+}\mathcal{M}od_k$; *that is, each* $S^i = S^i_{\bullet}$ *is a bounded-below complex of* k*-vector spaces.*[2]

(1) *There is a second-quadrant homological spectral sequence* $(E^r_{p,q}, d_r)$ *converging to zero such that*

$$E^1_{p,q} = H_q(S^p_{\bullet}) \quad (d_r : E^r_{p,q} \to E^r_{p-r,q+r-1}).$$

(2) *There is a first-quadrant cohomological spectral sequence* $(E^{p,q}_r, d^r)$ *converging to zero such that*

$$E^{p,q}_1 = H^q(\operatorname{Hom}_k(S^p_{\bullet}, k)) \quad (d^r : E^{p,q}_r \to E^{p+r,q-r+1}_r).$$

(3) *The following differentials are isomorphisms*:

$$d_{n+1} : E^{n+1}_{n+1,1} \to E^{n+1}_{0,n+1} \quad \text{and} \quad d^{n+1} : E^{0,n+1}_{n+1} \to E^{n+1,1}_{n+1}.$$

(4) *Suppose* $H_p : S^p \to S^{p+1}$ *is a contracting homotopy for* $S^{\bullet}$. *Then*

$$(d_{n+1})^{-1} = H_n \delta_1 H_{n-1} \cdots \delta_{n-1} H_1 \delta_n H_0 = H_n \prod_{i=1,\ldots,n} (\delta_i H_{n-i})$$

(*where the last product depends on the ordering and refers to composition*), *and*

$$(d^{n+1})^{-1} = H_0^* \delta_n^* H_1^* \cdots \delta_1^* H_n^* = H_0^* \prod_{i=n,\ldots,1} (\delta_i^* H_{n+1-i}^*),$$

where we write $f^* = \operatorname{Hom}_k(f, k)$ *as a shorthand.*

The construction is functorial in $S^{\bullet}$; *that is, if* $S^{\bullet} \to S'^{\bullet}$ *is a morphism of complexes as in our assumptions, then there are induced morphisms between their spectral sequences.*

[2] One may alternatively view this as a bicomplex supported horizontally in degrees $[0, n+1]$, bounded from below, and whose rows are exact.

Proof. Parts (1)–(3) are [Beilinson 1980, Lemma 1(a)]. More precisely, for (1) use the bicomplex spectral sequence for

$$E^0_{p,q} = S^p_q \quad \text{and} \quad E_0^{p,q} = \operatorname{Hom}_k(S^p_q, k).$$

If we take differentials "→" for forming the E^0-page, the E^1-page vanishes since $S_\bullet$ is exact (as a complex of complexes) and so the individual sequences of k-vector spaces $S^i_\bullet$ for constant i are exact, so $E^\infty = E^1 = 0$. Then use the bicomplex spectral sequences with differential "↓" on the E^0-page for our claim. It also converges to zero then; (2) is analogous. For (3), the bicomplex is horizontally supported in $[0, n+1]$. For (4), diagram chase. □

We combine Lemma 18 with Lemma 19: Apply the latter to $S^p_q := {}^\wedge T^p_q$; we denote the resulting spectral sequence by ${}^\wedge E^\bullet_{\bullet,\bullet}$. The fact that the (bi)complex of Lemma 19 is supported horizontally in $[n+1, 0]$ homologically (i.e., for ${}^\wedge E^\bullet_{\bullet,\bullet}$) and in $[0, n+1]$ cohomologically (i.e., for ${}^\wedge E^{\bullet,\bullet}_\bullet$) implies that we have edge morphisms

$$\begin{aligned} \rho_1 : {}^\wedge E^{n+1}_{n+1,1} \to {}^\wedge E^1_{n+1,1} \quad &\text{and} \quad \rho_2 : {}^\wedge E^1_{0,n+1} \to {}^\wedge E^{n+1}_{0,n+1}, \\ \wp_1 : {}^\wedge E^{0,n+1}_{n+1} \to {}^\wedge E^{0,n+1}_1 \quad &\text{and} \quad \wp_2 : {}^\wedge E^{n+1,1}_1 \to {}^\wedge E^{n+1,1}_{n+1}. \end{aligned}$$

Next, we identify the objects involved: Using Lemma 9 we compute

$$\begin{aligned} {}^\wedge E^1_{0,n+1} &= H_{n+1}({}^\wedge T^0_\bullet) = H_{n+1}(CE(\mathfrak{g})_\bullet) \cong H_{n+1}(\mathfrak{g}, k), \\ {}^\wedge E^1_{n+1,1} &= H_1({}^\wedge T^{n+1}_\bullet) = H_1\biggl(\bigcap_{i=1,\dots,n} \bigcap_{s_i \in \{\pm\}} CE(I_i^{s_i})_\bullet\biggr) = I_{\mathrm{tr}}/[I_{\mathrm{tr}}, \mathfrak{g}], \\ {}^\wedge E_1^{n+1,1} &= \operatorname{Hom}_k(I_{\mathrm{tr}}/[I_{\mathrm{tr}}, \mathfrak{g}], k) \quad \text{and} \quad {}^\wedge E_1^{0,n+1} = H^{n+1}(\mathfrak{g}, k). \end{aligned}$$

Definition 20 [Beilinson 1980]. Let $(A, (I_i^\pm), \tau)$ be an n-fold cubically decomposed algebra over a field k and $\mathfrak{g} := A_{\mathrm{Lie}}$ its Lie algebra. Define

$$\operatorname{res}_* : H_{n+1}(\mathfrak{g}, k) \to k \qquad \operatorname{res}_* := \tau \circ \rho_1 \circ (d_{n+1})^{-1} \circ \rho_2$$

and

$$\operatorname{res}^* : k \to H^{n+1}(\mathfrak{g}, k) \quad \operatorname{res}^*(1) := (\wp_1 \circ (d^{n+1})^{-1} \circ \wp_2)\tau,$$

where for res* we read τ as an element of $E_1^{n+1,1}$. We will call $\phi := \operatorname{res}^*(1)$ the *Tate extension class.*

In the case $n = 1$ it would also be justified to name this cohomology class after [Kac and Peterson 1981]; it also appears in the works of the Japanese school, e.g., [Jimbo and Miwa 1983].

Remark 21. It follows from the construction of res_* and res^* that

$$\operatorname{res}^*(\alpha)(X_0 \wedge \cdots \wedge X_n) = \alpha \operatorname{res}_* X_0 \wedge \cdots \wedge X_n. \tag{5-2}$$

Now we would like to compute these maps explicitly. Clearly, the most elusive map in the construction is the differential d_{n+1} (resp. d^{n+1}). We can render it explicit using Lemma 19(4) as soon as we have an explicit contracting homotopy available. However, it seems to be quite difficult to construct such a homotopy for the complex ${}^{\wedge}T^{\bullet}$. On the other hand, we *do* have such a contracting homotopy for ${}^{\otimes}T^{\bullet}$ by Lemma 16 and its corollary. Luckily for us, these complexes are closely connected. We may apply Lemma 19 also to $S^p_q := {}^{\otimes}T^p_{q-1}$; this time denote the resulting spectral sequence by ${}^{\otimes}E^{\bullet}_{\bullet,\bullet}$. We easily compute

$$\begin{aligned}
{}^{\otimes}E^1_{0,n+1} &= H_{n+1}({}^{\otimes}T^0_{\bullet-1}) = H_n(C(\mathfrak{g})_{\bullet}) \cong H_n(\mathfrak{g},\mathfrak{g}),\\
{}^{\otimes}E^1_{n+1,1} &= H_1({}^{\otimes}T^{n+1}_{\bullet-1}) = H_0\Bigg(C\bigg(\bigcap_{i=1,\dots,n}\ \bigcap_{s_i\in\{\pm\}} I_i^{s_i}\bigg)_{\bullet}\Bigg) = I_{\mathrm{tr}}/[I_{\mathrm{tr}},\mathfrak{g}],\\
{}^{\otimes}E_1^{n+1,1} &= \mathrm{Hom}_k(I_{\mathrm{tr}}/[I_{\mathrm{tr}},\mathfrak{g}],k) \quad \text{and} \quad {}^{\otimes}E_1^{0,n+1} = H^n(\mathfrak{g},\mathfrak{g}^*).
\end{aligned}$$

We note that some groups even agree with their ${}^{\wedge}T^p_q$-counterpart, as we had already observed in (2-4).

Definition 22. Write ${}^{\otimes}\mathrm{res}_* : H_n(\mathfrak{g},\mathfrak{g}) \to k$ and ${}^{\otimes}\mathrm{res}^*(1) \in H^n(\mathfrak{g},\mathfrak{g}^*)$ for the counterparts of res_*, res^* in Definition 20 using ${}^{\otimes}E$ instead of ${}^{\wedge}E$.

Lemma 23 (Compatibility). *The morphism of bicomplexes* ${}^{\otimes}T^{\bullet}_{\bullet} \xrightarrow{I} {}^{\wedge}T^{\bullet}_{\bullet+1}$ *induces a commutative diagram*

$$\begin{tikzcd}
H_n(\mathfrak{g},\mathfrak{g}) \arrow[r] \arrow[d] \arrow[rrr, bend left, "\text{comes with contracting homotopy}"] & {}^{\otimes}E^{n+1}_{0,n+1} \arrow[d] & {}^{\otimes}E^{n+1}_{n+1,1} \arrow[l, "\cong"', "d_{n+1}"] \arrow[r] \arrow[d] & H_0(\mathfrak{g},\mathfrak{g}) \arrow[d, "\cong"] \\
H_{n+1}(\mathfrak{g},k) \arrow[r] \arrow[rrr, bend right, "\text{Beilinson's residue}"'] & {}^{\wedge}E^{n+1}_{0,n+1} & {}^{\wedge}E^{n+1}_{n+1,1} \arrow[l, "d_{n+1}"', "\cong"] \arrow[r] & H_1(\mathfrak{g},k)
\end{tikzcd}$$

Proof. We had already observed in Lemma 18 that the morphisms I induce a morphism of bicomplexes. The spectral sequences ${}^{\otimes}E^{\bullet}_{\bullet,\bullet}$ and ${}^{\wedge}E^{\bullet}_{\bullet,\bullet}$ both arise from Lemma 19, so by the functoriality of the construction we get an induced morphism of spectral sequences. In particular, all squares

$$\begin{tikzcd}
{}^{\otimes}E^r_{p,q} \arrow[r, "d_r"] \arrow[d] & {}^{\otimes}E^r_{p-r,q+r-1} \arrow[d] \\
{}^{\wedge}E^r_{p,q} \arrow[r, "d_r"'] & {}^{\wedge}E^r_{p-r,q+r-1}
\end{tikzcd}$$

commute, giving the middle square in our claim. The same applies to the edge maps, giving the outer squares. □

Absolutely analogously we obtain a cohomological counterpart

$$\begin{tikzcd} H^1(\mathfrak{g}, k) \arrow[r] \arrow[d, "\cong"'] & H^{n+1}(\mathfrak{g}, k) \arrow[d] \\ H^0(\mathfrak{g}, \mathfrak{g}^*) \arrow[r] & H^n(\mathfrak{g}, \mathfrak{g}^*), \end{tikzcd}$$

where we have a contracting homotopy for the lower row. We leave the details of this formulation to the reader.

6. Concrete formalism

Let $(A, (I_i^{\pm}), \tau)$ be an n-fold cubically decomposed algebra over a field k. In Section 5 we have constructed a canonical morphism

$$\begin{tikzcd} \operatorname{res}_* : H_{n+1}(\mathfrak{g}, k) \arrow[r] & k \\ H_n(\mathfrak{g}, \mathfrak{g}), \arrow[u] & \end{tikzcd}$$

where $\mathfrak{g} := A_{\mathrm{Lie}}$ is the Lie algebra associated to A. By Lemma 23, its values on the image of $H_n(\mathfrak{g}, \mathfrak{g}) \to H_{n+1}(\mathfrak{g}, k)$ can be computed via ${}^{\otimes}\operatorname{res}_*$. In this section we will obtain an explicit formula for the latter morphism.

Given the definition of ${}^{\otimes}\operatorname{res}_*$, Lemma 19(4) tells us that it can be given explicitly in terms of differentials of the ordinary Chevalley–Eilenberg complexes $C(-)_\bullet$ (see Section 2) and contracting homotopies of the cube complex $N^\bullet$ (see Lemma 16 and its corollary), namely

$${}^{\otimes}\operatorname{res}_* = \tau \circ \rho_1 \circ ({}^{\otimes}d_{n+1})^{-1} \circ \rho_2 = \tau \circ \rho_1 H \prod_{i=1,\dots,n} (\delta_i H)\rho_2 \tag{6-1}$$

via the spectral sequence ${}^{\otimes}E^{\bullet}_{\bullet,\bullet}$. The contracting homotopy H depends on the choice of a good system of idempotents; see Definition 14. Different choices will yield formulas that may look different, but as ${}^{\otimes}\operatorname{res}_*$ (just like res_* itself) was defined entirely independently of the choice of any idempotents, all such formulas actually must agree.

Suppose a representative $\theta := f_0 \otimes f_1 \wedge \cdots \wedge f_n$ with $f_0, \dots, f_n \in N^0$ is given (note that N^0 equals $\mathfrak{g}$ as a left-$U\mathfrak{g}$-module by definition, so it is valid to treat all f_i on equal footing). We shall compute ${}^{\otimes}\operatorname{res}_* \theta$ in several steps, starting with $\theta_{0,n} := \rho_2\theta$, then following

$$\begin{tikzcd}
 & & & & 0 \arrow[d, dash] & & & \\
 & & \theta_{1,n} & & \theta_{0,n} \arrow[ll, "H"'] & n & & \\
 & & \vdots & & & \vdots & & q \\
 & \theta_{n,1} \arrow[d] & \theta_{n-1,1} \arrow[l, "H"'] & & & 1 & p & + \arrow[l] \arrow[u] \\
\theta_{n+1,0} & \theta_{n,0} \arrow[l, "H"'] & & & & 0 & & \\
n+1 & n & n-1 & \cdots & 0 & & &
\end{tikzcd} \tag{6-2}$$

as prescribed by (6-1). This graphical arrangement elucidates the position of the term of each step in the computation in the spectral sequence from which (6-1) originates — see Lemma 19. However, for us each $\theta_{*,*}$ will be an E^0-page representative of the respective E^*-page term. Finally ${}^{\otimes}\mathrm{res}_* \theta = \tau\rho_1\theta_{n+1,0}$. We note that ρ_1, ρ_2 are just edge maps, that is, an inclusion of a subobject and a quotient surjection. Hence, as we work with explicit representatives anyway, the operation of these maps is essentially invisible (e.g., in the quotient case it just means that our representative generates a larger equivalence class).

We will need a convenient notation for elements of this complex.

Notation A. We will write $\theta^{w_1 \dots w_p}_{p,q-p|s_1 \dots s_n} \in N^p$ for the summands in any expression of the shape

$$\theta_{p,q-p} = \sum_{\substack{w_1 \dots w_p \\ \in\{1,\dots,n\}}} \sum_{s_1 \dots s_n} \theta^{w_1 \dots w_p}_{p,q-p|s_1 \dots s_n} \otimes f_1 \wedge \cdots \wedge \widehat{f_{w_1}} \wedge \cdots \wedge \widehat{f_{w_p}} \wedge \cdots \wedge f_n, \tag{6-3}$$

where

- $(p, q-p)$ denotes the location of the element in the bicomplex as in (6-2),
- $s_1, \dots, s_n \in \{0, +, -\}$ denotes the component (= direct summand) of N^p as in (4-1), $f_1, \dots, f_n \in \mathfrak{g}$,
- the additional superscripts $w_1, \dots, w_p \in \{1, \dots, n\}$ are used to indicate the omission of wedge factors.

Note that the values $\theta^{w_1 \dots w_p}_{p,q|s_1 \dots s_n}$ are not necessarily uniquely determined since the individual wedge tails need not be linearly independent.

Notation B. We also need a shorthand for the summands in any expression of the shape

$$\theta_{p,q-p-1} = \sum_{\substack{w_1 \dots w_p, w_a, w_b \\ \in\{1,\dots,n\}}} \sum_{s_1 \dots s_n} \theta^{w_1 \dots w_p \| w_a, w_b}_{p,q|s_1 \dots s_n} \otimes [f_{w_a}, f_{w_b}] \wedge f_1 \wedge \cdots \widehat{f_{w_1}} \cdots \widehat{f_{w_a}} \cdots \widehat{f_{w_b}} \cdots \widehat{f_{w_p}} \cdots \wedge f_n. \tag{6-4}$$

Again $s_1, \dots, s_n$ denotes the component in N^p, $w_1, \dots, w_p$ omitted wedge factors. Moreover, w_a and w_b denote two additional omitted wedge factors and simultaneously indicate that $[f_{w_a}, f_{w_b}]$ appears as an additional wedge factor. As for the previous notation, the elements $\theta^{w_1 \dots w_p \| w_a, w_b}_{p,q|s_1 \dots s_n} \in N^p$ are not uniquely determined. We will explain how these expressions arise soon.

Combinatorial preparation: We define for arbitrary $1 \le p \le n$ and $w_1, \dots, w_p \in \{1, \dots, n\}$ a sign function (generalizing the sign of a permutation):

$$\rho(w_1, \dots, w_p) := (-1)^{\sum_{k=1}^{p} \sum_{j<k} \delta_{w_j < w_k}}. \tag{6-5}$$

By abuse of language we do not carry the value p in the notation for ρ as it will always be clear from the number of arguments which variant is used. It is easy to see that $\rho(w_1) = +1$ and $\rho(w_1, w_2) = (-1)^{\delta_{w_1 < w_2}}$. For $p = n$ we have

$$\rho(w_1, \dots, w_n) = \operatorname{sgn} \begin{pmatrix} 1 & \cdots & n \\ w_1 & \cdots & w_n \end{pmatrix}. \tag{6-6}$$

We shall need the inductive formula (which is easy to check by induction)

$$(-1)^{\#\{w_i \mid 1 \le i \le p \text{ s.t. } w_i < w_{p+1}\}} \rho(w_1, \dots, w_p) = \rho(w_1, \dots, w_{p+1}). \tag{6-7}$$

Proposition 24. *Suppose $\theta := f_0 \otimes f_1 \wedge \cdots \wedge f_n$ with $f_i \in N_0 = \mathfrak{g}$. Moreover, suppose $P_1^+, \dots, P_n^+$ is a good system of idempotents as in Definition 14. Then for every $p \ge 0$ the element $\theta_{p+1,q}$ is of the shape as in* (6-3) *and for $\gamma_1 \dots \gamma_{n-p} \in \{+, -\}$ we have*

$$\begin{aligned}
&\theta^{w_1 \dots w_p}_{p+1,q|\gamma_1 \dots \gamma_{n-p} \underbrace{0 \dots 0}_{p}} \\
&= (-1)^{\sum_{u=1}^{p-1}(u+1)} (-1)^{w_1 + \cdots + w_p} \rho(w_1, \dots, w_p) (-1)^{\gamma_1 + \cdots + \gamma_{n-p}} P_1^{\gamma_1} \cdots P_{n-p}^{\gamma_{n-p}} \times \\
&\qquad \sum_{\gamma^*_{n-p+1} \cdots \gamma^*_n \in \{\pm\}} (-1)^{\gamma^*_{n-p+1} + \cdots + \gamma^*_n} \big(P_{n-p+1}^{(-\gamma^*_{n-p+1})} \operatorname{ad}(f_{w_p}) P_{n-p+1}^{\gamma^*_{n-p+1}}\big) \cdots (P_n^{(-\gamma^*_n)} \operatorname{ad}(f_{w_1}) P_n^{\gamma^*_n}) f_0.
\end{aligned}$$

Here $\rho(w_1, \dots, w_p)$ is the sign function defined in (6-5)*. For $p = 0$ the expression $\rho(w_1, \dots, w_p)$ and the whole sum $\big(\sum_{\{\pm\}}(\cdots)\big)$ in $\big(\sum_{\{\pm\}}(\cdots)\big) f_0$ should be read as $+1$ (giving the right-hand side of* (6-8) *below).*

- Note that no terms of the shape as in (6-4) appear. This is not entirely obvious in view of the definition of $\delta^{[2]}$ — see (2-1).
- The formula does not compute $\theta^{w_1 \dots w_p}_{p+1,q|s_1 \dots s_n}$ for arbitrary $s_1 \dots s_n$ of degree $p + 1$. This is due to the fact that we only have further use for the ones treated.
- For $p \le 1$ read $\sum_{u=1}^{p-1}(u + 1)$ as zero.

Proof. We prove this by induction. For $p = 0$ the claim reads

$$\theta_{1,q|\gamma_1\dots\gamma_n} = (-1)^{\gamma_1+\cdots+\gamma_n} P_1^{\gamma_1} \cdots P_n^{\gamma_n} f_0 \tag{6-8}$$

and in view of (4-7) this proves the claim in this case. Now we proceed by induction. Assume the case p is settled, that is, in the notation of (6-3),

$$\theta^{w_1\dots w_p}_{p+1,q|\gamma_1\dots\gamma_{n-p}\underbrace{0\dots0}_{p}}$$

is exactly as in our claim. Next, we need to apply the differential $\delta_q = \delta_q^{[1]} + \delta_q^{[2]}$ of the Chevalley–Eilenberg resolution — see (2-1). The contribution of $\delta_q^{[1]}$ will be relevant, but for $\delta_q^{[2]}$ we shall see that (after applying the next contracting homotopy) the contribution vanishes. We treat each $\delta^{[i]}$, $i = 1, 2$ separately:

(1) Consider $\delta_q^{[1]}$ in (2-1). The sum Σ_i *loc. cit.* maps components indexed by $w_1, \dots, w_p$ to components of $\delta^{[1]}\theta_{p,q}$, indexed by $w_1, \dots, w_p$ and an additional $w_{p+1} \in \{1, \dots, n\} \setminus \{w_1, \dots, w_p\}$ — they correspond to the summands of $\delta^{[1]}\theta_{p,q}$ and to the additional omitted wedge factor, respectively. Moreover, the formula imposes signs $(-1)^{i+1}$, but here i depends on the numbering of the wedges $(\cdots \wedge \cdots \wedge \cdots)$. In the notation of (6-3) the subscript j of f_j does not necessarily indicate the f_j sits in the j-th wedge, due to the possible omission of wedge factors $f_{w_1}, \dots, f_{w_p}$ on the left-hand side of it. To compensate for that in the following computation the term $(-1)^{\#\{w_i | 1 \le i \le p \text{ s.t. } w_i < w_{p+1}\}}$ appears, sign-counting the omission on the left of the new-to-be-omitted w_{p+1} in the component of $\delta^{[1]}\theta_{p+1,q}$. As p remains constant, the indexing $\gamma_1 \dots \gamma_{n-p} 0 \dots 0$ remains unaffected. We get the expression

$$\begin{aligned}
&(\delta^{[1]}\theta_{p+1,q})^{w_1\dots w_p w_{p+1}}_{p+1,q-1|\gamma_1\dots\gamma_{n-p}\underbrace{0\dots0}_{p}} \\
&= (-1)^{\sum_{u=1}^{p-1}(u+1)}(-1)^{w_{p+1}+1}(-1)^{\#\{w_i|1\le i\le p \text{ s.t. } w_i<w_{p+1}\}}\,\mathrm{ad}(f_{w_{p+1}}) \\
&\quad \times (-1)^{w_1+\cdots+w_p}\rho(w_1,\dots,w_p)(-1)^{\gamma_1+\cdots+\gamma_{n-p}}P_1^{\gamma_1}\cdots P_{n-p}^{\gamma_{n-p}} \\
&\quad \times \sum_{\gamma^*_{n-p+1}\cdots\gamma^*_n\in\{\pm\}} (-1)^{\gamma^*_{n-p+1}+\cdots+\gamma^*_n}\big(P_{n-p+1}^{(-\gamma^*_{n-p+1})}\mathrm{ad}\, f_{w_p}\, P_{n-p+1}^{\gamma^*_{n-p+1}}\big)\cdots(P_n^{(-\gamma^*_n)}\mathrm{ad}\, f_{w_1}\, P_n^{\gamma^*_n})f_0.
\end{aligned}$$

Next, we need to apply the contracting homotopy $H : N^{p+1} \to N^{p+2}$. We have $p + 1 \ge 1$, so (4-5) applies. Note that for an index $\gamma_1^\dagger \dots \gamma_{n-p-1}^\dagger 0 \dots 0$ with $\gamma_1^\dagger \dots \gamma_{n-p-1}^\dagger \in \{\pm\}$ and $p+1$ zeros (i.e., an index of degree $p+2$; compare (4-1)), the corresponding index with one fewer 0 has degree $p + 1$. Indices of the latter type have been dealt with above. We obtain

$$
\begin{aligned}
&(H\delta^{[1]}\theta_{p+1,q})^{w_1\ldots w_p w_{p+1}}_{p+2,q-1|\gamma_1^\dagger\ldots\gamma_{n-p-1}^\dagger\underbrace{0\ldots0}_{p+1}}\\
&= (-1)^p(-1)^{\gamma_1^\dagger+\cdots+\gamma_{n-p-1}^\dagger} P_1^{\gamma_1^\dagger}\cdots P_{n-p-1}^{\gamma_{n-p-1}^\dagger}\\
&\qquad\times \sum_{\gamma_1,\ldots,\gamma_{(n-p-1)+1}\in\{\pm\}} (-1)^{\gamma_1+\cdots+\gamma_{n-p-1}} P_{(n-p-1)+1}^{-\gamma_{(n-p-1)+1}} (\delta\theta_{p+1,q})^{w_1\ldots w_{p+1}}_{p+1,q-1|\gamma_1\ldots\gamma_{n-p}\underbrace{0\ldots0}_{p}}.
\end{aligned}
$$

In principle the first factor is $(-1)^{\deg(\cdots)} = (-1)^{p+2}$, but switching to p preserves the correct sign. Next, we expand this using our previous computation and obtain (by noting that many signs are squares and thus $+1$)

$$
\begin{aligned}
&= (-1)^{\sum_{u=1}^{p-1}(u+1)}(-1)^{p+1}(-1)^{\gamma_1^\dagger+\cdots+\gamma_{n-p-1}^\dagger}(-1)^{\#\{w_i\,|\,1\leq i\leq p \text{ s.t. } w_i<w_{p+1}\}}\\
&\quad\times(-1)^{w_1+\cdots+w_{p+1}}\rho(w_1,\ldots,w_p)P_1^{\gamma_1^\dagger}\cdots P_{n-p-1}^{\gamma_{n-p-1}^\dagger}\\
&\quad\times\sum_{\gamma_{n-p}\in\{\pm\}}(-1)^{\gamma_{n-p}}\Bigg(\sum_{\gamma_1\ldots\gamma_{n-p-1}\in\{\pm\}} P_1^{\gamma_1}\cdots P_{n-p-1}^{\gamma_{n-p-1}}\Bigg)P_{n-p}^{-\gamma_{n-p}}\,\mathrm{ad}(f_{w_{p+1}})P_{n-p}^{\gamma_{n-p}}\\
&\quad\times\sum_{\gamma_{n-p+1}^*\ldots\gamma_n^*\in\{\pm\}}(-1)^{\gamma_{n-p+1}^*+\cdots+\gamma_n^*}\\
&\qquad\qquad\times\big(P_{n-p+1}^{(-\gamma_{n-p+1}^*)}\mathrm{ad}(f_{w_p})P_{n-p+1}^{\gamma_{n-p+1}^*}\big)\cdots\big(P_n^{(-\gamma_n^*)}\mathrm{ad}(f_{w_1})P_n^{\gamma_n^*}\big)f_0.
\end{aligned}
$$

The sum in parentheses is the identity since for all i we have $P_i^+ + P_i^- = \mathbb{1}$ by Definition 14. Up to the naming of the indices, and after using (6-7), this is exactly our claim in the case $p+1$ (and this is true despite the fact that we have only considered $\delta^{[1]}$ so far, because we shall next show that the contribution from $H\circ\delta^{[2]}$ vanishes).

(2) Consider $\delta_q^{[2]}$ in (2-1). Using the notation of (6-3) we may write

$$
\theta_{p+1,q} = \bigoplus_{\deg(s_1\ldots s_n)=p+1}\ \sum_{\substack{w_1\ldots w_p\\ \in\{1,\ldots,n\},\\ \text{pairw. diff.}}} \theta^{w_1\ldots w_p}_{p+1,q|s_1\ldots s_n}\otimes f_1\wedge\widehat{f_{w_1}}\cdots\widehat{f_{w_p}}\wedge f_n.
$$

Therefore

$$
\begin{aligned}
&\delta^{[2]}\theta_{p+1,q}\\
&= \bigoplus_{\deg(s_1\ldots s_n)=p+1}\ \sum_{\substack{w_1\ldots w_p\\ \in\{1,\ldots,n\},\\ \text{pairw. diff.}}}\ \sum_{\substack{w_{p+1}<w_{p+2}\\ \in\{1,\ldots,n\}\setminus\{w_1\ldots w_p\}}} (-1)^{w_{p+1}+w_{p+2}}\\
&\qquad\times(-1)^{\#\{w_i\,|\,1\leq i\leq p \text{ s.t. } w_i<w_{p+1}\}}(-1)^{\#\{w_i\,|\,1\leq i\leq p \text{ s.t. } w_i<w_{p+2}\}}\\
&\qquad\times\theta^{w_1\ldots w_p}_{p+1,q|s_1\ldots s_n}\otimes[f_{w_{p+1}},f_{w_{p+2}}]\wedge f_1\wedge\widehat{f_{w_1}}\cdots\widehat{f_{w_{p+1}}}\cdots\widehat{f_{w_{p+2}}}\cdots\widehat{f_{w_p}}\wedge f_n.
\end{aligned}
$$

The two powers of -1 on the middle line of the right-hand side appear since the original summand in $\delta^{[2]}$ carries the sign $(-1)^{i+j}$, so we need to compute the number of the wedge slot correctly, respecting the omitted wedge factors; compare with the discussion in the first part of this proof. We observe that the first wedge factor remains unchanged under $\delta^{[2]}$. Hence, when we apply the contracting homotopy H in this induction step and in the next again, the summand will vanish thanks to $H^2 = 0$; see (4-4). It will not do harm to verify this explicitly: We use the notation of (6-4) and write the above in terms of

$$\begin{aligned}&(\delta^{[2]}\theta_{p+1,q})^{w_1\dots w_p\|w_{p+1},w_{p+2}}_{p+1,q-1|s_1\dots s_n}\\ &\quad=(-1)^{w_{p+1}+w_{p+2}}(-1)^{\#\{w_i \mid 1\le i\le p \text{ s.t. } w_i<w_{p+1}\}}(-1)^{\#\{w_i \mid 1\le i\le p \text{ s.t. } w_i<w_{p+2}\}}\theta^{w_1\dots w_p}_{p+1,q|s_1\dots s_n}.\end{aligned}$$

Next, we apply the map $H : N^{p+1} \to N^{p+2}$ of (4-5). Then for indices $s_1 \dots s_n = \gamma_1^{\dagger} \dots \gamma_{n-p-1}^{\dagger} 0 \dots 0$ and $\gamma_1^{\dagger} \dots \gamma_{n-p-1}^{\dagger} \in \{\pm\}$ (which is of degree $p+2$) we obtain the expression

$$\begin{aligned}&(H\delta^{[2]}\theta_{p+1,q})^{w_1\dots w_p\|w_{p+1},w_{p+2}}_{p+2,q-1|\gamma_1^{\dagger}\dots\gamma_{n-p-1}^{\dagger}\underbrace{0\dots0}_{p+1}}\\ &\qquad\qquad= P_1^{\gamma_1^{\dagger}}\cdots P_{n-p-1}^{\gamma_{n-p-1}^{\dagger}}\sum_{\gamma_1\dots\gamma_{n-p}\in\{\pm\}}(-1)^{(\cdots)}P_{n-p}^{-\gamma_{n-p}}\theta^{w_1\dots w_p}_{p+1,q|\gamma_1\dots\gamma_{n-p}\underbrace{0\dots0}_{p}},\end{aligned}$$

where we have plugged in our previous computation and started to disregard the precise sign. We know the last term of this expression by our induction hypothesis and therefore obtain

$$\begin{aligned}&= P_1^{\gamma_1^{\dagger}}\cdots P_{n-p-1}^{\gamma_{n-p-1}^{\dagger}}\sum_{\gamma_1\dots\gamma_{n-p}\in\{\pm\}}\ \sum_{\gamma^*_{n-p+1}\dots\gamma^*_n\in\{\pm\}}(-1)^{(\dots)}\underline{P_{n-p}^{-\gamma_{n-p}}P_1^{\gamma_1}\cdots P_{n-p}^{\gamma_{n-p}}}\\ &\qquad\qquad\times\big(P_{n-p+1}^{(-\gamma^*_{n-p+1})}\,\mathrm{ad}(f_{w_p})P_{n-p+1}^{\gamma^*_{n-p+1}}\big)\cdots\big(P_n^{(-\gamma^*_n)}\,\mathrm{ad}(f_{w_1})P_n^{\gamma^*_n}\big)f_0.\end{aligned}$$

As the $P_1^{+}, \dots, P_n^{+}$ commute pairwise, the same holds for all $P_1^{\pm}, \dots, P_n^{\pm}$ (by Definition 14). Thus, the underlined expression can be rearranged to

$$P_{n-p}^{-\gamma_{n-p}}P_{n-p}^{\gamma_{n-p}}\cdots.$$

But

$$P_i^{+}P_i^{-} = P_i^{+}(\mathbb{1} - P_i^{+}) = 0$$

because P_i^{+} is an idempotent. The same holds for $P_i^{-}P_i^{+}$. Hence, in all the indices $s_1, \dots, s_n$ relevant for our claim $H\delta^{[2]}\theta_{p+1,q}$ is zero. □

This readily implies the following key computation:

Theorem 25 (main theorem). *Let $(A, (I_i^{\pm}), \tau)$ be an n-fold cubically decomposed algebra over a field k. Then*

$$\begin{aligned}&{}^{\otimes}\mathrm{res}_*(f_0 \otimes f_1 \wedge \cdots \wedge f_n)\\ &\quad = -(-1)^{\frac{(n-1)n}{2}} \tau \sum_{\pi \in \mathfrak{S}_n} \mathrm{sgn}(\pi)\\ &\qquad\quad \times \sum_{\gamma_1 \dots \gamma_n \in \{\pm\}} (-1)^{\gamma_1 + \cdots + \gamma_n} (P_1^{-\gamma_1} \,\mathrm{ad}\, f_{\pi(1)} P_1^{\gamma_1}) \cdots (P_n^{-\gamma_n} \,\mathrm{ad}\, f_{\pi(n)} P_n^{\gamma_n}) f_0,\end{aligned}$$

where $P_1^+, \dots, P_n^+$ is any system of pairwise commuting good idempotents in the sense of Definition 14 (the value does not depend on the choice of the latter). Analogously,

$$({}^{\otimes}\mathrm{res}^* \varphi)(f_1 \wedge \cdots \wedge f_n)(f_0) := \varphi \cdot {}^{\otimes}\mathrm{res}_*(f_0 \otimes f_1 \wedge \cdots \wedge f_n)$$

for every $\varphi \in k$.

We remark that one can also write the above formula as

$$\begin{aligned}&{}^{\otimes}\mathrm{res}_*(f_0 \otimes f_1 \wedge \cdots \wedge f_n) =\\ &-(-1)^{\frac{(n-1)n}{2}} \tau \sum_{\pi \in \mathfrak{S}_n} \mathrm{sgn}(\pi) \sum_{\gamma_1 \dots \gamma_n \in \{\pm\}} (-1)^{\gamma_1 + \cdots + \gamma_n} (P_1^{-\gamma_1} f_{\pi(1)} P_1^{\gamma_1}) \cdots (P_n^{-\gamma_n} f_{\pi(n)} P_n^{\gamma_n}) f_0\end{aligned}$$

since for any expression g we have

$$\begin{aligned}P_i^{-\gamma_i} \,\mathrm{ad}(f_w) P_i^{\gamma_i} g = P_i^{-\gamma_i} [f_w, P_i^{\gamma_i} g] &= P_i^{-\gamma_i} f_w P_i^{\gamma_i} g - P_i^{-\gamma_i} P_i^{\gamma_i} g f_w\\ &= P_i^{-\gamma_i} f_w P_i^{\gamma_i} g \end{aligned} \tag{6-9}$$

since $P_i^{-\gamma_i} P_i^{\gamma_i} = (\mathbb{1} - P_i^{\gamma_i}) P_i^{\gamma_i} = 0$ and $P_i^{\gamma_i}$ is an idempotent.

Proof. Use Proposition 24 with $p = n$. Plugging these components into the shorthand notation of (6-3) we unwind for ${}^{\otimes}\mathrm{res}_*(f_0 \otimes f_1 \wedge \cdots \wedge f_n)$ the formula

$$\begin{aligned}&= -\tau\, (-1)^{\frac{n^2+n}{2}} \sum_{\substack{w_1 \dots w_n\\ =\{1,\dots,n\}}} \rho(w_1, \dots, w_n) (-1)^{w_1 + \cdots + w_n}\\ &\qquad \times \sum_{\gamma_1 \dots \gamma_n \in \{\pm\}} (-1)^{\gamma_1 + \cdots + \gamma_n} (P_1^{-\gamma_1} \,\mathrm{ad}(f_{w_n}) P_1^{\gamma_1}) \cdots (P_n^{-\gamma_n} \,\mathrm{ad}(f_{w_1}) P_n^{\gamma_n}) f_0.\end{aligned}$$

We can clearly replace $w_1, \dots, w_n$ by a sum over all permutations of $\{1, \dots, n\}$. In order to obtain a nice formula (in the above formula the P_i appear in ascending order, while the w_i appear in descending order), we prefer to compose each permutation with the order-reversing permutation $w_i := \pi(n - i + 1)$; hence,

$$\begin{aligned}&{}^{\otimes}\mathrm{res}_*(f_0 \otimes f_1 \wedge \cdots \wedge f_n)\\ &\quad = -\tau(-1)^{\frac{n^2+n}{2}} \sum_{\pi \in \mathfrak{S}_n} \rho(\pi(n), \dots, \pi(1))(-1)^{1+\cdots+n}\\ &\qquad \times \sum_{\gamma_1 \dots \gamma_n \in \{\pm\}} (-1)^{\gamma_1+\cdots+\gamma_n} (P_1^{-\gamma_1} \operatorname{ad}(f_{\pi(1)}) P_1^{\gamma_1}) \cdots (P_n^{-\gamma_n} \operatorname{ad}(f_{\pi(n)}) P_n^{\gamma_n}) f_0.\end{aligned}$$

To conclude, use (6-6) and the (easy) fact that the order-reversing permutation has sign $(-1)^{(n-1)n/2}$, giving the sign of our claim. □

Proof of Theorems 1 and 2. We define $\mathfrak{G} := E^n(k)$, where E is the functor defined in Section 1.1. As already discussed in Section 1.1 this contains $k[t_1^{\pm}, \dots, t_n^{\pm}]$ as a Lie subalgebra, acting as multiplication operators $x \mapsto f \cdot x$. It is also easily checked that the differential operators $t_1^{s_1} \cdots t_n^{s_n} \partial_{t_i}$ can be written as infinite matrices. If $\mathfrak{g}$ is a *finite*-dimensional Lie algebra, observe that $\mathfrak{G} = E^n(k)$ and $E^n(\mathrm{End}_k(\mathfrak{g}))$ are actually isomorphic. If $\mathfrak{g}$ is simple, it is centerless, so the adjoint representation gives an embedding $\mathfrak{g} \hookrightarrow \mathrm{End}_k(\mathfrak{g})$, and thus

$$\mathfrak{g}[t_1^{\pm}, \dots, t_n^{\pm}] \hookrightarrow E^n(\mathrm{End}_k(\mathfrak{g})) \simeq E^n(k) = \mathfrak{G}.$$

This shows that all Lie algebras in the claim are subalgebras of $\mathfrak{G}$. As shown in Section 1.1, $\mathfrak{G}$ is a cubically decomposed algebra, so we define ϕ as in Definition 20, $\phi := \mathrm{res}^*(1)$. Since we work with field coefficients, the universal coefficient theorem for Lie algebras tells us that

$$H^{n+1}(\mathfrak{g}, k) \cong H_{n+1}(\mathfrak{g}, k)^*,$$

that is, knowing the values of a cocycle only on Lie cycles (instead of all of $\bigwedge^{\bullet}\mathfrak{g}$) determines the cocycle uniquely, $\mathrm{res}^*(1)(\alpha) = \mathrm{res}_*\alpha$. However, by Lemma 23 we may evaluate the cocycle on the image of I by using ${}^{\otimes}\mathrm{res}_*$ instead. Using Theorem 25 we get an explicit formula for ${}^{\otimes}\mathrm{res}^*(1)$, proving Theorem 2. Using the explicit formula, it is a direct computation to check that for $n = 1$ the cocycle agrees with the ones mentioned in the claim of Theorem 1. □

7. Application to the multidimensional residue

In this section we will show that the Lie cohomology class of Definition 20 naturally gives the multidimensional (Parshin) residue.

We work in the framework of multivariate Laurent polynomial rings over a field k; see Section 1.1. In other words, as our cubically decomposed algebra we take an infinite matrix algebra $A = E^n(k)$ and $\mathfrak{g} = A_{\mathrm{Lie}}$. Via (1-4) it acts on the k-vector space $k[t_1^{\pm}, \dots, t_n^{\pm}]$. The latter, now interpreted as a ring, also embeds as a *commutative* subalgebra into A. In order to distinguish very clearly between the subalgebra of A and the vector space it acts on, we shall from now on write

$k[\boldsymbol{t}_1^{\pm}, \dots, \boldsymbol{t}_n^{\pm}]$ for the k-vector space. Thus, when we write t_i we always refer to the associated multiplication operator $x \mapsto t_i \cdot x$ in A, e.g., $t_i^m \cdot \boldsymbol{t}_i^l = \boldsymbol{t}_i^{m+l}$.

Following [Beilinson 1980, Lemma 1(b)] we may introduce a (not quite well-defined[3]) "map"

$$\varkappa : \Omega^n_{k[t_1^{\pm},\dots,t_n^{\pm}]/k} \to H_{n+1}(\mathfrak{g}, k), \qquad f_0\, \mathrm{d}f_1 \wedge \cdots \wedge \mathrm{d}f_n \mapsto f_0 \wedge f_1 \wedge \cdots \wedge f_n. \quad (7\text{-}1)$$

As $k[t_1^{\pm}, \dots, t_n^{\pm}]$ is commutative, the f_i commute pairwise and thus $f_0 \wedge \cdots \wedge f_n$ is indeed a Lie homology cycle.

Theorem 26. *The morphism*

$$\operatorname{res}_* \circ \varkappa : \Omega^n_{k[t_1^{\pm},\dots,t_n^{\pm}]/k} \to k$$

(with $\varkappa$ as in (7-1) and res_* as in Definition 20) *for $c_{i,j} \in \mathbb{Z}$ is explicitly given by*

$$t_1^{c_{0,1}} \cdots t_n^{c_{0,n}}\, \mathrm{d}(t_1^{c_{1,1}} \cdots t_n^{c_{1,n}}) \wedge \cdots \wedge \mathrm{d}(t_1^{c_{n,1}} \cdots t_n^{c_{n,n}}) \mapsto -(-1)^{\frac{n^2+n}{2}} \det \begin{pmatrix} c_{1,1} & \cdots & c_{n,1} \\ \vdots & \ddots & \vdots \\ c_{1,n} & \cdots & c_{n,n} \end{pmatrix}$$

whenever $\sum_{p=0}^{n} c_{p,i} = 0$ and is zero otherwise. In particular $-(-1)^{\frac{n^2+n}{2}} (\operatorname{res}_* \circ \varkappa)$ is the conventional multidimensional (*Parshin*) *residue.*

The complicated sign $-(-1)^{\frac{n^2+n}{2}}$ should not concern us too much; it is an artifact of homological algebra. Just by changing our sign conventions for bicomplexes, we could easily switch to an overall opposite sign. Letting $c_{i,j} = \delta_{i=j}$ for $i, j \in \{1, \dots, n\}$ gives the familiar

$$-(-1)^{\frac{n^2+n}{2}} \operatorname{res}_*(a t_1^{c_{0,1}} \cdots t_n^{c_{0,n}} \wedge t_1 \wedge \cdots \wedge t_n) = \delta_{c_{0,1}=-1} \cdots \delta_{c_{0,n}=-1} a$$

for $a \in k$. In particular this assures us that the map res_* gives the correct notion of residue: it is the $(-1, \dots, -1)$-coefficient of the Laurent expansion.

Proof. After unwinding $\varkappa$ it remains to evaluate $\operatorname{res}_*(f_0 \wedge f_1 \wedge \cdots \wedge f_n)$ for $f_i := t_1^{c_{i,1}} \cdots t_n^{c_{i,n}}$ $(i = 0, \dots, n)$. Clearly $f_0 \otimes f_1 \wedge \cdots \wedge f_n$ is a cycle in $H_n(\mathfrak{g}, \mathfrak{g})$, and so by Lemma 23 we may use ${}^{\otimes}\operatorname{res}_*$ instead of res_*. Then Theorem 25 reduces this to the matrix trace

$$\operatorname{res}_*(f_0 \wedge f_1 \wedge \cdots \wedge f_n) = -(-1)^{\frac{(n-1)n}{2}} \sum_{\pi \in \mathfrak{S}_n} \operatorname{sgn}(\pi) \tau M_\pi, \qquad (7\text{-}2)$$

[3] It does not respect the relation $\mathrm{d}(ab) = b\,\mathrm{d}a + a\,\mathrm{d}b$; this artifact already occurs in [Beilinson 1980]. However, this ambiguity dissolves after composing with the residue (as in the theorem) and it is very convenient to treat this as some sort of a map for the moment.

where

$$M_\pi := \sum_{\gamma_1 \dots \gamma_n \in \{\pm\}} (-1)^{\gamma_1+\dots+\gamma_n} (P_1^{-\gamma_1} f_{\pi(1)} P_1^{\gamma_1}) \cdots (P_n^{-\gamma_n} f_{\pi(n)} P_n^{\gamma_n}) f_0.$$

For the evaluation of τM_π fix a permutation π and pick the (pairwise commuting) system of idempotents given by

$$P_j^{+} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} = \delta_{\lambda_j \geq 0} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} \quad (\text{with } \lambda_1, \dots, \lambda_n \in \mathbb{Z}). \tag{7-3}$$

Next, observe that the Laurent polynomial ring $W := k[\boldsymbol{t}_1^{\pm}, \dots, \boldsymbol{t}_n^{\pm}]$ is stable (i.e., $\phi W \subseteq W$) under the endomorphisms $f_0, \dots, f_n$ and the idempotents $P_i^{\pm}$, and therefore under M_π. Hence, it follows that it suffices to evaluate the trace of M_π on the k-vector subspace $k[\boldsymbol{t}_1^{\pm}, \dots, \boldsymbol{t}_n^{\pm}]$. We compute successively

$$\begin{aligned} f_k P_j^{+} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} &= \delta_{\lambda_j \geq 0} \boldsymbol{t}_1^{\lambda_1 + c_{k,1}} \cdots \boldsymbol{t}_n^{\lambda_n + c_{k,n}}, \\ P_j^{-} f_k P_j^{+} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} &= \delta_{0 \leq \lambda_j < -c_{k,j}} \boldsymbol{t}_1^{\lambda_1 + c_{k,1}} \cdots \boldsymbol{t}_n^{\lambda_n + c_{k,n}}, \end{aligned}$$

and analogously for $P_j^{+} f_k P_j^{-}$. We find

$$\begin{aligned} &\sum_{\gamma_j \in \{\pm\}} (-1)^{\gamma_j} (P_j^{-\gamma_j} f_k P_j^{\gamma_j}) \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} \\ &\qquad = (\delta_{0 \leq \lambda_j < -c_{k,j}} - \delta_{-c_{k,j} \leq \lambda_j < 0}) \boldsymbol{t}_1^{\lambda_1 + c_{k,1}} \cdots \boldsymbol{t}_n^{\lambda_n + c_{k,n}}. \end{aligned} \tag{7-4}$$

Subclaim. Writing $w_i := \pi(i)$ we have

$$\begin{aligned} M_\pi \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} = \prod_{i=1}^{n} (\delta_{0 \leq \lambda_i + c_{0,i} + \sum_{p=i+1}^{n} c_{w_p,i} < -c_{w_i,i}} - \delta_{-c_{w_i,i} \leq \lambda_i + c_{0,i} + \sum_{p=i+1}^{n} c_{w_p,i} < 0}) \\ \times \boldsymbol{t}_1^{\lambda_1 + c_{0,1} + \sum_{p=1}^{n} c_{w_p,1}} \cdots \boldsymbol{t}_n^{\lambda_n + c_{0,n} + \sum_{p=1}^{n} c_{w_p,n}}. \end{aligned} \tag{7-5}$$

(*Proof of subclaim.* Define for $i = 1, \dots, n+1$ the truncated sum

$$M_\pi^{(i)} := \Bigg[\sum_{\gamma_i \dots \gamma_n \in \{\pm\}} (-1)^{\gamma_i+\dots+\gamma_n} (P_i^{-\gamma_i} f_{w_i} P_i^{\gamma_i}) \cdots (P_n^{-\gamma_n} f_{w_n} P_n^{\gamma_n}) \Bigg] f_0$$

so that $M_\pi^{(1)} = M_\pi$ and $M_\pi^{(n+1)} = f_0$. We claim that

$$M_\pi^{(i)} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} = \alpha \boldsymbol{t}_1^{\lambda_1 + c_{0,1} + \sum_{p=i}^{n} c_{w_p,1}} \cdots \boldsymbol{t}_n^{\lambda_n + c_{0,n} + \sum_{p=i}^{n} c_{w_p,n}} \tag{7-6}$$

for some factor $\alpha \in \{\pm 1, 0\}$. For $i = n+1$ this is clear since $f_0 = t_1^{c_{0,1}} \cdots t_n^{c_{0,n}}$, in particular $\alpha = 1$. Assuming this holds for $i+1$, for i we get by using (7-4) (with the appropriate values plugged in: $j := i$ and $k := w_i$, and λ_i as in (7-6))

$$\begin{aligned} M_\pi^{(i)} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} &= \sum_{\gamma_i \in \{\pm\}} (-1)^{\gamma_i} (P_i^{-\gamma_i} f_{w_i} P_i^{\gamma_i}) M_\pi^{(i+1)} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} \\ &= (\delta_{0 \le \lambda_i + c_{0,i} + \sum_{p=i+1}^n c_{w_p,i} < -c_{w_i,i}} - \delta_{-c_{w_i,i} \le \lambda_i + c_{0,i} + \sum_{p=i+1}^n c_{w_p,i} < 0}) \\ &\quad \times \alpha \boldsymbol{t}_1^{\lambda_1 + c_{0,1} + \sum_{p=i+1}^n c_{w_p,1} + c_{w_i,1}} \cdots \boldsymbol{t}_n^{\lambda_n + c_{0,n} + \sum_{p=i+1}^n c_{w_p,n} + c_{w_i,n}}. \end{aligned} \tag{7-7}$$

This proves our claim for all i by induction. We observe that the prefactor α in each step just gets multiplied with the expression in (7-7), giving the product in our claim.)

Next, we need to evaluate the trace of M_π as given in (7-5). The endomorphism is nilpotent unless

$$c_{0,1} + \sum_{p=1}^{n} c_{w_p,i} = 0 \quad \text{for all } i. \tag{7-8}$$

We remark that $w_1, \dots, w_n$ is just a permutation of $\{1, \dots, n\}$, so these conditions can be rewritten as $\sum_{p=0}^n c_{p,i} = 0$. In the nilpotent case the trace is clearly zero. Hence, we may assume we are in the case where (7-8) holds. Using these equations and the useful convention $w_{n+1} := 0$, our expression for M_π simplifies to

$$\begin{aligned} &M_\pi \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} \\ &\quad = \prod_{i=1}^{n} (\delta_{0 \le \lambda_i + \sum_{p=i+1}^{n+1} c_{w_p,i} < -c_{w_i,i}} - \delta_{0 \le \lambda_i + c_{w_i,i} + \sum_{p=i+1}^{n+1} c_{w_p,i} < c_{w_i,i}}) \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n}. \end{aligned} \tag{7-9}$$

The endomorphism M_π is visibly diagonal of finite rank and we may reduce the computation of the trace to a (finite-dimensional) stable vector subspace. A finite subset of the $\boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n}$ ($\lambda_1, \dots, \lambda_n \in \mathbb{Z}$) provides a basis. We see in (7-9) that M_π acts diagonally on these basis vectors with eigenvalues ± 1 or 0. Moreover, for each i we either have $c_{w_i,i} \ge 0$ or $c_{w_i,i} < 0$, which shows that each bracket of the shape $(\delta_{0 \le \lambda < -c} - \delta_{-c \le \lambda < 0})$ in (7-9) either attains only values in $\{+1, 0\}$ when we run through all $\lambda_1, \dots, \lambda_n \in \mathbb{Z}$, or only values in $\{-1, 0\}$. This shows that we only need to count (with appropriate sign) the nonzero eigenvalues of M_π in order to evaluate the trace. Note that our finite subset of $\boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n}$ ($\lambda_1, \dots, \lambda_n \in \mathbb{Z}$) indexes a basis, so we need to count the number of such basis vectors with nonzero eigenvalue. We introduce the nonstandard shorthand $\lfloor x \rfloor := \min(0, x)$. Inspecting (7-9) shows that when running through λ_i we have

- $\lfloor -c_{w_i,i} \rfloor$ times the eigenvalue $+1$,
- $\lfloor +c_{w_i,i} \rfloor$ times the eigenvalue -1.

The value of a fixed bracket $(\delta_{0\leq\lambda<-c}-\delta_{-c\leq\lambda<0})$ — when nonzero — is always either $+1$, or always -1. Thus, the number of nonzero eigenvalues is simply the number of elements within the hypercube such that each λ_i lies within the range of length $\lfloor\pm c_{w_i,i}\rfloor$ counted above, and therefore

$$\tau M_\pi=\prod_{i=1}^{n}(\lfloor -c_{w_i,i}\rfloor-\lfloor +c_{w_i,i}\rfloor)=\prod_{i=1}^{n}(-c_{w_i,i})=(-1)^n\prod_{i=1}^{n}c_{\pi(i),i}$$

(because $\lfloor -a\rfloor-\lfloor a\rfloor=-a$ for all $a\in\mathbb{Z}$). We plug this into (7-2) and recognize the usual formula for the determinant. This finishes the proof. □

We are now ready to prove the remaining theorems from the introduction:

Proof of Theorems 4 and 5. We use Theorem 26 to obtain Theorem 4(2). Then Theorem 4(3) follows as a special case. For Theorem 4(1) use the shorthands $\pi=P_1^+=P^+$ (following both the notation of Arbarello, de Concini and Kac and ours). On the one hand we compute

$$\begin{aligned}[]
[\pi,f_1]f_0&=[P,f_1]f_0=Pf_1f_0-f_1Pf_0=[Pf_0,f_1]\\
&=(P^++P^-)[P^+f_0,f_1]\\
&=P^-[P^+f_0,f_1]+P^+[P^+f_0,f_1]\\
&=P^-[P^+f_0,f_1]-P^+[P^-f_0,f_1].
\end{aligned}$$

where for the last equality we used that $[P^+f_0,f_1]+[P^-f_0,f_1]=[f_0,f_1]=0$. On the other hand, we unwind

$$\begin{aligned}
\operatorname{res} f_0\,\mathrm{d}f_1&=(-1)^1\operatorname{tr}\sum_{\gamma_1\in\{\pm\}}(-1)^{\gamma_1}(P_1^{-\gamma_1}\operatorname{ad}(f_{\pi(1)})P_1^{\gamma_1})f_0\\
&=-P^-[f_1,P_1^+f_0]+P^+[f_1,P_1^-f_0]
\end{aligned}$$

and these expressions clearly coincide. Finally Theorem 5 is true since we use the cocycle defined in Definition 20, which is constructed exactly as stated in Theorem 5. □

8. Application to multiloop Lie algebras

Suppose k is a field and $\mathfrak{g}/k$ is a finite-dimensional centerless Lie algebra (e.g., $\mathfrak{g}$ finite-dimensional, semisimple). Then the adjoint representation $\operatorname{ad}:\mathfrak{g}\hookrightarrow\operatorname{End}_k(\mathfrak{g})$ is injective. Thus, we obtain a Lie algebra inclusion

$$i:\mathfrak{g}[\boldsymbol{t}_1^\pm,\dots,\boldsymbol{t}_n^\pm]\hookrightarrow E^n(\operatorname{End}_k(\mathfrak{g}))_{\mathrm{Lie}},$$

where E is the functor described in Section 1.1 (the right-hand side is equipped with the Lie bracket $[a,b]=ab-ba$ based on the associative algebra structure).

Thus, we have the pullback

$$i^* : H^{n+1}(E^n(\mathrm{End}_R(\mathfrak{g}))_{\mathrm{Lie}}, k) \to H^{n+1}(\mathfrak{g}[\boldsymbol{t}_1^{\pm}, \dots, \boldsymbol{t}_n^{\pm}], k),$$

which we may apply to the class $\mathrm{res}^*(1)$ — see Definition 20.

Theorem 27. *Suppose k is a field and $\mathfrak{g}/k$ is a finite-dimensional centerless Lie algebra. For $Y_0, \dots, Y_n \in \mathfrak{g}$ we call*

$$B(Y_0, \dots, Y_n) := \mathrm{tr}_{\mathrm{End}_k(\mathfrak{g})}(\mathrm{ad}(Y_0)\,\mathrm{ad}(Y_1) \cdots \mathrm{ad}(Y_n)) \tag{8-1}$$

the "generalized Killing form". For $n = 1$ and if $\mathfrak{g}$ is semisimple, this is the classical Killing form of $\mathfrak{g}$.

(1) *Then on all Lie cycles admitting a lift under I as in* (0-1), *the pullback $i^*\,\mathrm{res}^*(1) \in H^{n+1}(\mathfrak{g}[\boldsymbol{t}_1^{\pm}, \dots, \boldsymbol{t}_n^{\pm}], k)$ is explicitly given by*

$$(i^*\phi)(Y_0 \boldsymbol{t}_1^{c_{0,1}} \cdots \boldsymbol{t}_n^{c_{0,n}} \wedge \cdots \wedge Y_n \boldsymbol{t}_1^{c_{n,1}} \cdots \boldsymbol{t}_n^{c_{n,n}}) = -(-1)^{\frac{n^2+n}{2}} \sum_{\pi \in \mathfrak{S}_n} \mathrm{sgn}(\pi) B(Y_{\pi(1)}, \dots, Y_{\pi(n)}, Y_0) \prod_{i=1}^{n} c_{\pi(i),i}$$

whenever $\sum_{p=0}^{n} c_{p,i} = 0$ for all $i \in \{1, \dots, n\}$, and it vanishes otherwise.

(2) *If $\mathfrak{g}$ is finite-dimensional and semisimple and $n = 1$, then $i^*\mathrm{res}^*(1) \in H^2(\mathfrak{g}[\boldsymbol{t}_1^{\pm}], k)$ is the universal central extension of the loop Lie algebra $\mathfrak{g}[\boldsymbol{t}_1, \boldsymbol{t}_1^{-1}]$ giving the associated affine Lie algebra $\widehat{\mathfrak{g}}$ (without extending by a derivation),*

$$0 \to k\langle c \rangle \to \widehat{\mathfrak{g}} \to \mathfrak{g}[\boldsymbol{t}_1, \boldsymbol{t}_1^{-1}] \to 0.$$

Proof. (1) By Lemma 23, Theorem 25 and (5-2) the cocycle is explicitly given by

$$\mathrm{res}^*(1)(f_0 \wedge \cdots \wedge f_n) = {}^{\otimes}\mathrm{res}^*(1)(f_0 \otimes f_1 \wedge \cdots \wedge f_n) = \tau \sum_{\pi \in \mathfrak{S}_n} \mathrm{sgn}(\pi) M_\pi,$$

where

$$M_\pi = \sum_{\gamma_1 \cdots \gamma_n \in \{\pm\}} (-1)^{\gamma_1 + \cdots + \gamma_n} (P_1^{-\gamma_1} f_{\pi(1)} P_1^{\gamma_1}) \cdots (P_n^{-\gamma_n} f_{\pi(n)} P_n^{\gamma_n}) f_0.$$

Note that $M_\pi \in E^n(\mathrm{End}_k(\mathfrak{g}))$. As we consider the pullback of the cohomology class along $i : \mathfrak{g}[t_1^{\pm}, \dots, t_n^{\pm}] \hookrightarrow E^n(\mathrm{End}_k(\mathfrak{g}))_{\mathrm{Lie}}$, it suffices to treat elements $f_i := Y_i t_1^{c_{i,1}} \cdots t_n^{c_{i,n}}$ with $c_{i,1}, \dots, c_{i,n} \in \mathbb{Z}$ (for $i = 0, \dots, n$) and $Y_i \in \mathfrak{g}$. Note that by our embedding i an element f_i is mapped to the endomorphism $\mathrm{ad}(Y_i) t_1^{c_{i,1}} \cdots t_n^{c_{i,n}}$ in $E^n(\mathrm{End}_k(\mathfrak{g}))$. Let $\pi \in \mathfrak{S}_n$ be a fixed permutation. In order to compute the trace, it suffices to study the action of M_π on the basis elements $X \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n}$ of $\mathfrak{g}[t_1^{\pm}, \dots, t_n^{\pm}]$, where $\lambda_1, \dots, \lambda_n \in \mathbb{Z}$ and $X \in \mathfrak{g}$ runs through a basis of $\mathfrak{g}$. We denote them with bold letters $\boldsymbol{t}_i$ instead of t_i to distinguish clearly between a basis element and t_i as an endomorphism $t_i : x \mapsto t_i \cdot x$ in $E^n(\mathrm{End}_k(\mathfrak{g}))$. As in the proof of Theorem 26 we compute

$$P_j^- f_k P_j^+ X \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} = \delta_{0 \le \lambda_j < -c_{k,j}} \operatorname{ad}(Y_k) X \boldsymbol{t}_1^{\lambda_1 + c_{k,1}} \cdots \boldsymbol{t}_n^{\lambda_n + c_{k,n}},$$

and as a consequence we find

$$\begin{aligned} &\sum_{\gamma_j \in \{\pm\}} (-1)^{\gamma_j} (P_j^{-\gamma_j} f_k P_j^{\gamma_j}) X \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} \\ &\qquad = (\delta_{0 \le \lambda_j < -c_{k,j}} - \delta_{-c_{k,j} \le \lambda_j < 0}) \operatorname{ad}(Y_k) X \boldsymbol{t}_1^{\lambda_1 + c_{k,1}} \cdots \boldsymbol{t}_n^{\lambda_n + c_{k,n}}. \end{aligned}$$

With an inductive computation entirely analogous to (7-5) we find

$$\begin{aligned} M_\pi X \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n} = \prod_{i=1}^{n} &(\delta_{0 \le \lambda_i + c_{0,i} + \sum_{p=i+1}^{n} c_{w_p,i} < -c_{w_i,i}} - \delta_{-c_{w_i,i} \le \lambda_i + c_{0,i} + \sum_{p=i+1}^{n} c_{w_p,i} < 0}) \\ &\times \operatorname{ad}(Y_{w_1}) \cdots \operatorname{ad}(Y_{w_n}) \operatorname{ad}(Y_0) X \boldsymbol{t}_1^{\lambda_1 + \sum_{p=0}^{n} c_{p,1}} \cdots \boldsymbol{t}_n^{\lambda_n + \sum_{p=0}^{n} c_{p,n}}, \end{aligned}$$

where $w_i := \pi(i)$. Unless $\forall i : \sum_{p=0}^{n} c_{p,i} = 0$ holds, M_π is clearly nilpotent and thus has trace $\tau M_\pi = 0$. This condition is clearly independent of π, showing that $(i^* \operatorname{res}^*(1))(f_0 \wedge \cdots \wedge f_n) = 0$ in this case. From now on assume $\forall i : \sum_{p=0}^{n} c_{p,i} = 0$. Then M_π respects the decomposition

$$\mathfrak{g}[\boldsymbol{t}_1^{\pm}, \ldots, \boldsymbol{t}_n^{\pm}] = \coprod_{\lambda_1 \ldots \lambda_n \in \mathbb{Z}^n} \mathfrak{g} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n}$$

and therefore (as τ is essentially a trace) $\tau M_\pi = \sum_{\lambda_1, \ldots, \lambda_n} \tau M_\pi |_{\mathfrak{g} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n}}$. For each summand of the latter we obtain

$$\begin{aligned} \tau M_\pi \, |_{\mathfrak{g} \boldsymbol{t}_1^{\lambda_1} \cdots \boldsymbol{t}_n^{\lambda_n}} = \prod_{i=1}^{n} &(\delta_{0 \le \lambda_i + c_{0,i} + \sum_{p=i+1}^{n} c_{w_p,i} < -c_{w_i,i}} - \delta_{-c_{w_i,i} \le \lambda_i + c_{0,i} + \sum_{p=i+1}^{n} c_{w_p,i} < 0}) \\ &\times \operatorname{tr}(\operatorname{ad}(Y_{w_1}) \cdots \operatorname{ad}(Y_{w_n}) \operatorname{ad}(Y_0)). \end{aligned}$$

The trace term is independent of $\lambda_1, \ldots, \lambda_n$ (and in the shape of (8-1)), so we may rewrite τM_π as

$$\begin{aligned} \tau M_\pi &= B(Y_{w_1}, \ldots, Y_{w_n}, Y_0) \\ &\quad \times \sum_{\lambda_1, \ldots, \lambda_n} \prod_{i=1}^{n} (\delta_{0 \le \lambda_i + c_{0,i} + \sum_{p=i+1}^{n} c_{w_p,i} < -c_{w_i,i}} - \delta_{-c_{w_i,i} \le \lambda_i + c_{0,i} + \sum_{p=i+1}^{n} c_{w_p,i} < 0}). \end{aligned}$$

For the evaluation of the sum $\sum_{\lambda_1, \ldots, \lambda_n}$ we can apply the same eigenvalue count as in the proof of Theorem 26. This time instead of counting eigenvalues, we count nonzero summands. This yields

$$\tau M_\pi = (-1)^n B(Y_{w_1}, \ldots, Y_{w_n}, Y_0) \prod_{i=1}^{n} c_{w_i,i}$$

and thus our claim.

(2) For $n = 1$ we obtain

$$(i^* \operatorname{res}^*(1))(Y_0 \boldsymbol{t}_1^{c_{0,1}} \wedge Y_1 \boldsymbol{t}_1^{c_{1,1}}) = -c_{1,1} \delta_{c_{0,1}+c_{1,1}=0} B(Y_1, Y_0).$$

This is well-known to be the defining cocycle of the affine Lie algebra $\widehat{\mathfrak{g}}$ (usually with a positive sign, but the class is only well-defined up to nonzero scalar multiple anyway). □

The natural further cases of the Virasoro algebra as well as affine Kac–Moody algebras (i.e., $\widehat{\mathfrak{g}}$ extended by derivations) will be discussed elsewhere. The computations become more involved, but no further ideas are needed.

Acknowledgements

I am very thankful to Ivan Fesenko and Matthew Morrow for many valuable discussions, especially on an adèle interpretation. I thank the research group of Professor Marc Levine for the stimulating scientific environment. I heartily thank the anonymous referee for greatly improving the presentation, especially in Section 4, and observing the fact $H^2 = 0$ in (4-4), which clarifies a crucial cancellation in the proof of Proposition 24.

References

[Arbarello et al. 1989] E. Arbarello, C. De Concini, and V. G. Kac, "The infinite wedge representation and the reciprocity law for algebraic curves", pp. 171–190 in *Theta functions, Part 1* (Bowdoin College, Brunswick, ME, 1987), edited by L. Ehrenpreis and R. C. Gunning, Proc. Sympos. Pure Math. **49**, Amer. Math. Soc., Providence, RI, 1989. MR 90i:22034 Zbl 0699.22028

[Beilinson 1980] A. A. Beilinson, "Residues and adèles", *Funktsional. Anal. i Prilozhen.* **14**:1 (1980), 44–45. MR 81f:14010 Zbl 0509.14018

[Beilinson et al. 1991] A. A. Beilinson, B. L. Feigin, and B. C. Mazur, "Notes on conformal field theory", unpublished, 1991, Available at http://www.math.sunysb.edu/~kirillov/manuscripts.html.

[Feĭgin and Tsygan 1983] B. L. Feĭgin and B. L. Tsygan, "Cohomology of Lie algebras of generalized Jacobi matrices", *Funktsional. Anal. i Prilozhen.* **17**:2 (1983), 86–87. MR 85c:17008 Zbl 0544.17011

[Fesenko 2010] I. Fesenko, "Analysis on arithmetic schemes, II", *J. K-Theory* **5**:3 (2010), 437–557. MR 2011k:14019 Zbl 1225.14019

[Fesenko and Kurihara 2000] I. Fesenko and M. Kurihara (editors), *Invitation to higher local fields*, Geometry & topology monographs **3**, Geometry & topology, Coventry, England, 2000. Papers from the conference held in Münster, August 29–September 5, 1999,. MR 2001h:11005 Zbl 0954.00026

[Frenkel 1987] I. B. Frenkel, "Beyond affine Lie algebras", pp. 821–839 in *Proceedings of the International Congress of Mathematicians* (Berkeley, CA, 1986), vol. 1, 2, edited by A. M. Gleason, Amer. Math. Soc., Providence, RI, 1987. MR 89g:17018 Zbl 0668.17016

[Frenkel and Zhu 2012] E. Frenkel and X. Zhu, "Gerbal representations of double loop groups", *Int. Math. Res. Not.* **2012**:17 (2012), 3929–4013. MR 2972546 Zbl 06088719

[Huber 1991] A. Huber, "On the Parshin–Beĭlinson adèles for schemes", *Abh. Math. Sem. Univ. Hamburg* **61** (1991), 249–273. MR 92k:14024 Zbl 0763.14006

[Hübl and Yekutieli 1996] R. Hübl and A. Yekutieli, "Adèles and differential forms", *J. Reine Angew. Math.* **471** (1996), 1–22. MR 97d:14026 Zbl 0847.14006

[Jimbo and Miwa 1983] M. Jimbo and T. Miwa, "Solitons and infinite-dimensional Lie algebras", *Publ. Res. Inst. Math. Sci.* **19**:3 (1983), 943–1001. MR 85i:58060 Zbl 0557.35091

[Kac and Peterson 1981] V. G. Kac and D. H. Peterson, "Spin and wedge representations of infinite-dimensional Lie algebras and groups", *Proc. Nat. Acad. Sci. U.S.A.* **78**:6, part 1 (1981), 3308–3312. MR 82j:17019 Zbl 0469.22016

[Kac and Raina 1987] V. G. Kac and A. K. Raina, *Bombay lectures on highest weight representations of infinite-dimensional Lie algebras*, Advanced series in mathematical physics **2**, World scientific, Teaneck, NJ, 1987. MR 90k:17013 Zbl 0668.17012

[Loday 1992] J.-L. Loday, *Cyclic homology*, Grundlehren der Mathematischen Wissenschaften [Fundamental principles of mathematical sciences] **301**, Springer, Berlin, 1992. MR 94a:19004 Zbl 0780.18009

[Morrow 2010] M. Morrow, "An explicit approach to residues on and dualizing sheaves of arithmetic surfaces", *New York J. Math.* **16** (2010), 575–627. MR 2012a:14061 Zbl 1258.14031

[Neher 2011] E. Neher, "Extended affine Lie algebras and other generalizations of affine Lie algebras—a survey", pp. 53–126 in *Developments and trends in infinite-dimensional Lie theory*, edited by K.-H. Neeb and A. Pianzola, Progr. Math. **288**, Birkhäuser, Boston, 2011. MR 2011m:17055 Zbl 1261.17023

[Tate 1968] J. Tate, "Residues of differentials on curves", *Ann. Sci. École Norm. Sup.* (4) **1** (1968), 149–159. MR 37 #2756 Zbl 0159.22702

Communicated by Mikhail Kapranov
Received 2012-06-16 Revised 2013-04-14 Accepted 2013-09-09

oliver.braeunling@uni-due.de *Fakultät für Mathematik, Universität Duisburg-Essen,*
Thea-Leymann-Straße 9, 45127 Essen, Germany
http://www.esaga.uni-due.de/oliver.braeunling/

ALGEBRA AND NUMBER THEORY 8:1 (2014)
dx.doi.org/10.2140/ant.2014.8.53

On the number of cubic orders of bounded discriminant having automorphism group C_3, and related problems

Manjul Bhargava and Ariel Shnidman

For a binary quadratic form Q, we consider the action of SO_Q on a 2-dimensional vector space. This representation yields perhaps the simplest nontrivial example of a prehomogeneous vector space that is not irreducible, and of a coregular space whose underlying group is not semisimple. We show that the nondegenerate integer orbits of this representation are in natural bijection with orders in cubic fields having a fixed "lattice shape". Moreover, this correspondence is *discriminant-preserving*: the value of the invariant polynomial of an element in this representation agrees with the discriminant of the corresponding cubic order.

We use this interpretation of the integral orbits to solve three classical-style counting problems related to cubic orders and fields. First, we give an asymptotic formula for the number of cubic orders having bounded discriminant and nontrivial automorphism group. More generally, we give an asymptotic formula for the number of cubic orders that have bounded discriminant and any given lattice shape (i.e., reduced trace form, up to scaling). Via a sieve, we also count cubic *fields* of bounded discriminant whose rings of integers have a given lattice shape. We find, in particular, that among cubic orders (resp. fields) having lattice shape of given discriminant D, the shape is *equidistributed* in the class group Cl_D of binary quadratic forms of discriminant D. As a by-product, we also obtain an asymptotic formula for the number of cubic fields of bounded discriminant having any given quadratic resolvent field.

MSC2010: primary 11R16; secondary 11R45.
Keywords: cubic fields, discriminant.

1. Introduction

An order in a cubic field (or *cubic order* for short) has either 1 or 3 automorphisms. The number of cubic orders with trivial automorphism group and bounded discriminant,[1] and the corresponding number of fields, were computed asymptotically in the classical work of Davenport and Heilbronn [1971]. A corresponding asymptotic formula for the number of cubic fields with an automorphism of order 3 (called *C_3-cubic fields*) was obtained by Cohn [1954], but a formula for C_3-cubic *orders* has not previously been obtained. In this article, we prove the following theorem:

Theorem 1. *The number of cubic orders having automorphism group isomorphic to a cyclic group of order* 3, *and discriminant less than* X, *is*

$$\frac{\pi}{6\sqrt{3}}X^{1/2}+O(X^{1/4}).$$

More generally, we prove asymptotics for the number of cubic orders having any given "lattice shape". To be more precise, a *cubic ring* is a commutative ring with unit that is free of rank 3 as a $\mathbb{Z}$-module. Such a ring R is endowed with a linear map $\mathrm{Tr}: R\to\mathbb{Z}$ called the *trace*, which sends $z\in R$ to the trace of the endomorphism $\times z: R\to R$ defined by multiplication by z. The *discriminant* $\operatorname{Disc} R$ of a cubic ring R with $\mathbb{Z}$-basis $\alpha_1,\alpha_2,\alpha_3$ is defined to be $\det(\mathrm{Tr}(\alpha_i\alpha_j))\in\mathbb{Z}$. A cubic *order* is a cubic ring that is also an integral domain.

For a cubic ring R, the restriction of the trace form $\mathrm{Tr}(z^2)$ to the trace-zero part of $\mathbb{Z}+3R$ is an integer-valued binary quadratic form. If R has nonzero discriminant, then, via a choice of basis, this form can be written as $nQ(x,y)$, where Q is a primitive integral binary quadratic form and n is a positive integer. We define the *shape* of R to be the $\mathrm{GL}_2(\mathbb{Z})$-equivalence class of the binary quadratic form $Q(x,y)$. Since it is often convenient, we will usually refer to Q itself (or an equivalent form) as the shape of R.[2]

If Q is a primitive integral binary quadratic form, then we define $N_3(Q,X)$ to be the number of cubic orders having shape Q and absolute discriminant less than X. It is easy to see that, by definition, the shape Q of a cubic ring cannot be negative definite; hence all quadratic forms Q in this paper are assumed to be either positive definite or indefinite. The following theorem gives an asymptotic formula for $N_3(Q,X)$ as $X\to\infty$.

[1] When referring to the number of cubic orders or fields with a given property, we always mean the number of such objects up to isomorphism.

[2] The shape of R may also be described in terms of the restriction of the trace form $\mathrm{Tr}(z^2)$ to the *projection* of the lattice R onto the plane in $R\otimes\mathbb{Q}$ that is orthogonal to 1. This yields the binary quadratic form $(n/3)Q(x,y)$, and so scaling by $n/3$ again gives the primitive integral binary quadratic form Q, which we call the shape. We prefer to use our definition in terms of $\mathbb{Z}+3R$, as then we can work integrally and do not have to refer to $R\otimes\mathbb{Q}$.

Theorem 2. *Let Q be a primitive integral binary quadratic form with nonsquare discriminant D. Set $\alpha = 1$ if $3 \mid D$ and $\alpha = 0$ otherwise. Set $\beta = 1$ if $D > -4$ and $\beta = 0$ otherwise. Set $\gamma = 1$ if Q is ambiguous*[3] *and $\gamma = 0$ otherwise. Then*

$$N_3(Q, X) = \frac{3^{\alpha+\beta-3/2} \cdot L(1, \chi_D)}{2^\gamma \cdot h(D)\sqrt{|D|}} X^{1/2} + O(X^{1/4}).$$

Here, $L(s, \chi_D)$ is the Dirichlet L-function associated to the primitive quadratic character χ_D of conductor D and $h(D)$ denotes the size of the narrow class group of binary quadratic forms of discriminant D up to $\mathrm{SL}_2(\mathbb{Z})$-equivalence.

A cubic ring has three automorphisms if and only if its shape is equivalent to the quadratic form $Q(x, y) = x^2 + xy + y^2$ (see proof of Theorem 14). Thus, for this choice of Q, the quantity $N_3(Q, X)$ is the number of cubic orders with discriminant less than X that have three automorphisms, and Theorem 1 follows from Theorem 2.

The main term in Theorem 2 is nearly a function of the discriminant D of Q; only the factor of 2^γ depends on the particular equivalence class of Q. With this in mind, we introduce the notion of an *oriented cubic ring*, which is a pair (R, δ) consisting of a cubic ring R and an isomorphism $\delta : \wedge^3 R \to \mathbb{Z}$. We usually refer to an oriented cubic ring (R, δ) simply as R, with the accompanying isomorphism δ being implied. The *shape* of an oriented cubic ring R is defined as before, but now using oriented bases and $\mathrm{SL}_2(\mathbb{Z})$-equivalence.

We define $N_3^{\mathrm{Or}}(Q, X)$ to be the number of isomorphism classes of oriented cubic orders having shape Q and absolute discriminant less than X. Notice that $Q(x, y)$ is ambiguous if and only if its $\mathrm{GL}_2(\mathbb{Z})$-equivalence class coincides with its $\mathrm{SL}_2(\mathbb{Z})$-equivalence class. If Q is not ambiguous, then its $\mathrm{GL}_2(\mathbb{Z})$-class splits into two $\mathrm{SL}_2(\mathbb{Z})$-classes. In other words,

$$N_3^{\mathrm{Or}}(Q, X) = 2^\gamma N_3(Q, X),$$

where γ is defined as in Theorem 2. Thus Theorem 2 is equivalent to the following.

Theorem 3. *Let Q be a primitive integral binary quadratic form with nonsquare discriminant D. Set $\alpha = 1$ if $3 \mid D$ and $\alpha = 0$ otherwise, and set $\beta = 1$ if $D > -4$ and $\beta = 0$ otherwise. Then*

$$N_3^{\mathrm{Or}}(Q, X) = \frac{3^{\alpha+\beta-3/2} \cdot L(1, \chi_D)}{h(D)\sqrt{|D|}} X^{1/2} + O(X^{1/4}).$$

In particular, among oriented cubic orders with shape of discriminant D, the shape is *equidistributed* in the class group Cl_D of binary quadratic forms of discriminant D.

[3] Recall that a quadratic form $Q(x, y) = rx^2 + sxy + ty^2$ is said to be *ambiguous* if there is an automorphism of Q in $\mathrm{GL}_2(\mathbb{Z})$ with determinant -1. Equivalently, Q is ambiguous if it is $\mathrm{SL}_2(\mathbb{Z})$-equivalent to the form $Q' = rx^2 - sxy + ty^2$.

The exponent $\frac{1}{4}$ in the error term in Theorem 3 is optimal. If we count cubic *rings* (instead of just cubic orders) having shape Q, then we find that the main term in Theorem 3 stays the same (that is, the number of cubic rings of a given shape Q that are not orders in cubic fields is negligible), but the error term becomes smaller.

Via a suitable sieve, we use Theorem 3 to determine asymptotics for the number $M_3(Q, X)$ (resp. $M_3^{\mathrm{Or}}(Q, X)$) of *maximal* cubic orders (resp. maximal oriented cubic orders) having shape Q and discriminant bounded by X. Thus $M_3(Q, X)$ is the number of cubic fields with absolute discriminant less than X whose rings of integers have shape Q. As before we have $M_3^{\mathrm{Or}}(Q, X) = 2^{\gamma} M_3(Q, X)$, where $\gamma = 1$ if Q is ambiguous and $\gamma = 0$ otherwise.

Theorem 4. *Let Q be a primitive integral binary quadratic form of discriminant D. Suppose that either D or $-D/3$ is a nonsquare fundamental discriminant. Set $\alpha = 1$ if $3 \mid D$ and $\alpha = 0$ otherwise. Also, set $\beta = 1$ if $D > -4$ and $\beta = 0$ otherwise. Then*

$$\begin{aligned}&M_3^{\mathrm{Or}}(Q, X)\\ &= \frac{3^{\alpha+\beta+1/2}\mu_3(D)}{4\pi^2\sqrt{|D|}} \cdot \frac{L(1, \chi_D)}{h(D)} \prod_{\substack{(D/p)=1\\ p\neq 3}} \left(1 - \frac{2}{p(p+1)}\right) \prod_{\substack{p \mid D\\ p\neq 3}} \left(\frac{p}{p+1}\right) X^{1/2} + o(X^{1/2}),\end{aligned}$$

where

$$\mu_3(D) = \begin{cases} \frac{16}{27} & \text{if } 3 \nmid D, \\ \frac{22}{27} & \text{if } 3 \parallel D, \\ \frac{2}{3} & \text{if } 9 \parallel D. \end{cases}$$

For all other nonsquare values of D, we have $M_3(Q, X) = 0$.

As in Theorem 3, we see that the shapes of rings of integers in oriented cubic fields (when ordered by absolute discriminant) are *equidistributed* in the respective class groups Cl_D of binary quadratic forms of discriminant D. The error term in Theorem 4 can certainly be improved, although we shall not investigate the issue in this paper.

Applying Theorem 4 to the form $Q(x, y) = x^2 + xy + y^2$ yields the following result of Cohn [1954] on the number of abelian cubic fields having bounded discriminant (though our methods are completely different!).

Theorem 5 [Cohn 1954]. *The number of abelian cubic fields having discriminant less than X is*

$$\frac{11\sqrt{3}}{36\pi} \prod_{p\equiv 1(3)} \left(1 - \frac{2}{p(p+1)}\right) \cdot X^{1/2} + o(X^{1/2}).$$

Finally, Theorem 4 can also be used to count the number $N(d, X)$ of cubic fields with absolute discriminant bounded by X and whose quadratic resolvent field is $\mathbb{Q}(\sqrt{d})$. To this end, let $M_3^D(X)$ be the number of cubic fields K with shape of discriminant of D and $|\operatorname{Disc} K| < X$. Then by Theorem 4, we have $M_3^D(X) \sim \frac{1}{2}h(D)M_3^{\mathrm{Or}}(Q, X)$ as $X \to \infty$, for any primitive integral form Q of discriminant D. Regarding $N(d, X)$, we prove:

Theorem 6. *Suppose $d \neq -3$ is a fundamental discriminant. Then*

$$N(d, X) = \begin{cases} M_3^{-3d}(X) & \text{if } 3 \nmid d, \\ M_3^{-3d}(X) + M_3^{-d/3}(X) & \text{if } 3 \mid d. \end{cases}$$

Combining this with Theorem 4, we obtain the following.

Theorem 7. *Let $d \neq -3$ be a fundamental discriminant, and set $D = -d/3$ if $3 \mid d$ and $D = -3d$ otherwise. Set $\alpha = 1$ if $3 \mid D$ and $\alpha = 0$ otherwise. Also set $\beta = 1$ for $D > -4$ and $\beta = 0$ otherwise. Then*

$$N(d, X) = \frac{3^{\alpha+\beta-1/2} \cdot C_0}{\pi^2\sqrt{|D|}} \cdot \prod_{p|D} \frac{p}{p+1} \cdot L(D) \cdot X^{1/2} + o(X^{1/2}),$$

where

$$L(D) = \prod_p \left(1 + \left(\frac{D}{p}\right)\frac{1}{p+1}\right) = L(1, \chi_D) \prod_{(\frac{D}{p})=1} \left(1 - \frac{2}{p(p+1)}\right)$$

and

$$C_0 = \begin{cases} \frac{11}{9} & \text{if } d \not\equiv 0 \pmod 3, \\ \frac{5}{3} & \text{if } d \equiv 3 \pmod 9, \\ \frac{7}{5} & \text{if } d \equiv 6 \pmod 9. \end{cases}$$

We note that this latter result on the asymptotic number of cubic fields having a given quadratic resolvent field was recently obtained independently by Cohen and Morra [2011, Theorem 1.1(2), Corollary 7.6] using very different methods, and with an explicit error term of $O(X^{1/3+\varepsilon})$.

All these results may be extended to the case of square discriminant. The cubic fields with shape of square discriminant are precisely the *pure* cubic fields — that is, those of the form $\mathbb{Q}(\sqrt[3]{m})$ for some integer m — while the orders with such a shape are the orders in pure cubic fields (see Lemma 33). The asymptotic growth in the case of a square discriminant is somewhat larger:

Theorem 8. *Let Q_D be a primitive integral binary quadratic form of square discriminant D. Set $\alpha = 1$ if $3 \mid D$ and $\alpha = 0$ otherwise. Then*

$$N_3^{\mathrm{Or}}(Q_D, X) = \frac{3^{\alpha-3/2}}{2D} X^{1/2} \log X + \frac{3^{\alpha-3/2}}{D}\Big(2\gamma - 1 + \frac{3}{2}\log\frac{D}{3}\Big) X^{1/2} + O(X^{1/4}),$$

where γ is Euler's constant. Also, we have

$$M_3(Q_1, X) = \frac{C}{15\sqrt{3}} X^{1/2}\bigl(\log X - \tfrac{16}{5}\log 3 + 4\gamma + 12\kappa - 2\bigr) + o(X^{1/2})$$

and

$$M_3(Q_9, X) = \frac{C}{40\sqrt{3}} X^{1/2}\bigl(\log X - \tfrac{1}{5}\log 3 + 4\gamma + 12\kappa - 2\bigr) + o(X^{1/2}),$$

where $C = \prod_p (1 - 3/p^2 + 2/p^3)$ and $\kappa = \sum_p (\log p)/(p^2 + p - 2)$. For all other square values of D, we have $M_3(Q_D, X) = 0$.

Finally, $N(-3, X) = M_3(Q_1, X) + 2M_3(Q_9, X)$; hence the density of pure cubic fields is given by

$$N(-3, X) = \frac{7C}{60\sqrt{3}} X^{1/2}\bigl(\log X - \tfrac{67}{35}\log 3 + 4\gamma + 12\kappa - 2\bigr) + o(X^{1/2}). \quad (1)$$

In particular, Theorem 8 shows that, when counting cubic orders with shape of a given square discriminant D, both the first *and* second main terms of the asymptotics become equidistributed in the class group Cl_D. We note that Equation (1) corresponds to Theorem 1.1(1) in [Cohen and Morra 2011], where again an explicit error term of $O(X^{1/3+\varepsilon})$ is proved.

There are two types of pure cubic fields: those with shape of discriminant $D = 1$ and those with shape of discriminant $D = 9$. These turn out to correspond to Dedekind's notion [1899] of pure cubic fields of Types 1 and 2, respectively. In fact, the asymptotics for $N(-3, X)$ can be deduced fairly easily from Dedekind's work, as we explain in the last remark of the paper. In general, Theorem 6 shows that there are two distinct types of cubic fields whenever the discriminant d of the quadratic resolvent algebra is a multiple of 3. In that case, there are cubic fields of shape of discriminant $-d/3$ and of discriminant $-3d$, and these are precisely the cubic fields of Type 1 and Type 2, respectively.

Our methods are mostly quite elementary, involving primarily geometry-of-numbers arguments. However, these methods also fit into a larger context. Let G be an algebraic group and V a representation of G. Then the pair (G, V) is called a *prehomogeneous vector space* if G possesses an open orbit; the pair (G, V) is called a *coregular space* if the ring of invariants of G on V is free. A classification of all *irreducible* reductive prehomogeneous vector spaces was attained in [Sato and Kimura 1977], while a similar classification of *simple* (resp. *semisimple* and *irreducible*) coregular spaces was accomplished in [Schwarz 1978] (resp. [Littelmann 1989]). The rational and integer orbits in such spaces tend to have a very rich arithmetic interpretation (see, for example, [Gauss 1801; Delone and Faddeev 1964; Wright and Yukie 1992; Bhargava 2004a; 2004b; 2004c; 2008;

Cremona et al. 2010; Bhargava and Ho 2013]), and have led to several number-theoretic applications, particularly to the study of the density of discriminants of number fields and statistical questions involving elliptic and higher genus curves [Davenport and Heilbronn 1971; Datskovsky and Wright 1988; Bhargava 2005; 2010; Bhargava and Shankar 2010; Bhargava and Gross 2012].

In this paper, we prove Theorems 1–8 and related results by considering the simplest nontrivial example of a coregular space whose underlying group is *not* semisimple (namely, an action of $\mathrm{SO}_Q(\mathbb{C})$ on $\mathbb{C}^2$ for a binary quadratic form Q). It also yields the simplest prehomogeneous vector space that is *not* irreducible (namely, an action of $\mathrm{GO}_Q(\mathbb{C})$ on $\mathbb{C}^2$). We show that the integer orbits on this space, even in this nonirreducible and nonsemisimple scenario, also have a rich and nontrivial number-theoretic interpretation, namely, the integer orbits classify cubic rings whose lattice shape is Q.

In light of these results, we note that there has not been a classification of general reducible prehomogeneous vector spaces nor of general nonsemisimple coregular spaces, akin to the work of Sato and Kimura or Schwarz and Littelmann, respectively, leading to two interesting questions in representation theory. As we hope this paper will illustrate, the solution of these two problems may also have important consequences for number theory.

The problems that we address in this paper are related to problems considered in [Terr 1997], [Mantilla-Soler 2010] and [Zhao 2013]. Terr showed that the shape of cubic rings (ordered by absolute discriminant) is equidistributed amongst lattices, which are viewed as points in Gauss's fundamental domain $\mathcal{F}$. More precisely, the number of cubic rings (of bounded discriminant) having shape lying in some subset $W \subset \mathcal{F}$ is proportional to the area of W (with respect to the hyperbolic measure on $\mathcal{F}$). Terr's work is somewhat orthogonal to our own in that it implies that $N(Q, X) = o(X)$, but it does not say anything more about a single shape $Q(x, y)$. The related problem treated by Mantilla-Soler is to determine when a cubic field is determined by its trace form, which is a finer invariant than the shape. Finally, Zhao carried out a detailed study of the distribution of cubic function fields by discriminant, in which the geometric *Maroni invariant* of trigonal curves plays an important role. In particular, he has suggested an analogue of the Maroni invariant for cubic number fields, which turns out to be closely related to the notion of "shape".

2. Preliminaries

In order to count cubic rings of bounded discriminant, we use a parametrization, due to Delone and Faddeev [1964] and recently refined by Gan, Gross, and Savin [Gan et al. 2002], that identifies cubic rings with integral binary cubic forms.

2.1. ***The Delone–Faddeev correspondence.*** We follow the exposition of [Gan et al. 2002]. Consider the space of all binary cubic forms

$$f(x,y)=ax^3+bx^2y+cxy^2+dy^3$$

with integer coefficients, and let an element $\gamma\in\mathrm{GL}_2(\mathbb{Z})$ act on this space by the twisted action

$$\gamma\cdot f(x,y)=\frac{1}{\det(\gamma)}\cdot f((x,y)\gamma).$$

This action is faithful and defines an equivalence relation on the space of integral binary cubic forms.

The *discriminant* of a binary cubic form $f(x,y)=ax^3+bx^2y+cxy^2+dy^3$ is defined by

$$\mathrm{Disc}\, f=b^2c^2-4ac^3-4b^3d-27a^2d^2+18abcd.$$

The discriminant polynomial is invariant under the action of $\mathrm{GL}_2(\mathbb{Z})$.

Theorem 9 [Gan et al. 2002]. *There is a canonical bijection between isomorphism classes of cubic rings and* $\mathrm{GL}_2(\mathbb{Z})$*-equivalence classes of integral binary cubic forms. Under this bijection, the discriminant of a binary cubic form is equal to the discriminant of the corresponding cubic ring. Furthermore, a cubic ring is an integral domain* (*that is, a cubic order*) *if and only if the corresponding binary cubic form is irreducible over* $\mathbb{Q}$.

Proof. See [Gan et al. 2002, §4; Bhargava et al. 2013, §2]. □

Remark 10. To parametrize *oriented* cubic rings, one must use $\mathrm{SL}_2(\mathbb{Z})$-equivalence in the correspondence of Theorem 9, rather than $\mathrm{GL}_2(\mathbb{Z})$-equivalence. Recall that the shape of an oriented cubic ring is then well-defined up to $\mathrm{SL}_2(\mathbb{Z})$-equivalence.

Not only does Theorem 9 give a bijection between cubic rings and cubic forms, but it also shows that certain properties and invariants of each type of object translate nicely. Next, we describe how the shape of a cubic ring translates into the language of binary cubic forms.

2.2. ***Hessians and shapes.*** Let $f(x,y)=ax^3+bx^2y+cxy^2+dy^3$ be an integral binary cubic form. The integral binary quadratic form

$$H_f(x,y)=-\frac{1}{4}\det\begin{pmatrix}\dfrac{\partial f(x,y)}{\partial x^2} & \dfrac{\partial f(x,y)}{\partial x\partial y}\\[2ex] \dfrac{\partial f(x,y)}{\partial y\partial x} & \dfrac{\partial f(x,y)}{\partial y^2}\end{pmatrix}$$

is called the *Hessian* of f. The Hessian has the following properties [Gan et al. 2002].

Proposition 11. *The Hessian is a* $\mathrm{GL}_2(\mathbb{Z})$*-covariant of integral binary cubic forms*; *that is*, *if two binary cubic forms* f *and* g *are equivalent under the twisted action of* $\mathrm{GL}_2(\mathbb{Z})$, *then the corresponding Hessians* H_f *and* H_g *are also equivalent under the same action but without the twisting factor* (*that is*, *the factor of the determinant*). *For any binary cubic form* f, *we have* $\mathrm{Disc}(H_f) = -3 \cdot \mathrm{Disc}\, f$.

The relevance of the Hessian of a binary cubic form is that it gives the shape of the corresponding cubic ring:

Proposition 12. *Suppose that a cubic ring* (*resp. oriented cubic ring*) R *corresponds to a binary cubic form* $f(x, y)$ *as in Theorem 9. Then the* $\mathrm{GL}_2(\mathbb{Z})$ (*resp.* $\mathrm{SL}_2(\mathbb{Z})$)*-equivalence class of the primitive part of the Hessian* $H_f(x, y)$ *coincides with the shape of* R.

Proof. If we write

$$\gamma = x\alpha + y\beta = \frac{\mathrm{Tr}(\gamma)}{3} + \gamma_0,$$

with $\gamma_0 \in \frac{1}{3}R$ of trace zero, then a computation gives $H_f(x, y) = \frac{3}{2}\mathrm{Tr}(\gamma_0^2)$ [Gan et al. 2002]. □

Example 13. Suppose R is a cubic ring having an order-3 automorphism and corresponding binary cubic form $f(x, y)$. Then the Hessian $H_f(x, y)$ of f must have an order-3 automorphism as well, and so it must be equivalent to an integer multiple of the quadratic form $Q(x, y) = x^2 + xy + y^2$. Conversely, we show in the next section that any ring having the form $Q(x, y)$ as its shape must be a C_3-cubic ring.

3. On cubic orders having automorphism group C_3

In this section we will prove Theorems 1 and 5. As mentioned in the introduction, these theorems are actually special cases of Theorems 2 and 4, respectively. We prove these cases separately because the argument is better motivated and understood after seeing a concrete example. Moreover, the results in this case are interesting in their own right due to their connection with C_3-cubic orders in abelian cubic fields.

3.1. ***The action of*** $\mathbf{SO}_Q(\mathbb{C})$ ***on*** $\mathbb{C}^2$**.** Set $Q(x, y) = x^2 + xy + y^2$, and let $\mathrm{SO}_Q(\mathbb{C})$ denote the subgroup of elements of $\mathrm{SL}_2(\mathbb{C})$ preserving the quadratic form $Q(x, y)$ via its natural (left) action on binary quadratic forms; that is,

$$\mathrm{SO}_Q(\mathbb{C}) = \{\gamma \in \mathrm{SL}_2(\mathbb{C}) : Q(x, y) = Q((x, y)\gamma)\}.$$

We define the *cubic action* of $\mathrm{SO}_Q(\mathbb{C})$ on $\mathbb{C}^2$ by $\gamma \cdot v = \gamma^3 v$ for a column vector $v = (b, c)^t \in \mathbb{C}^2$. The adjoint quadratic form $Q'(b, c) := b^2 - bc + c^2$ of $Q(x, y)$ is an invariant polynomial for this latter action, and it generates the full ring of invariants.

Let $L \subset \mathbb{C}^2$ be the lattice $\{(b, c)^t : b, c \in \mathbb{Z}^2,\ b \equiv c \pmod 3\}$. We will see that L is preserved under $\mathrm{SO}_Q(\mathbb{Z})$, the group of integer matrices in $\mathrm{SO}_Q(\mathbb{C})$.

Theorem 14. *The* $\mathrm{SO}_Q(\mathbb{Z})$*-orbits on nonzero lattice vectors* $(b, c)^t \in L$ *are in natural bijection with* C_3*-cubic oriented rings* R*. Under this bijection,* $\operatorname{Disc} R = Q'(b, c)^2$.

Proof. Let R be a C_3-cubic ring with automorphism σ of order 3, and let

$$f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$$

be a binary cubic form corresponding to R under the Delone–Faddeev correspondence. Also let $H(x, y)$ denote the Hessian of $f(x, y)$. Then σ induces an order-3 automorphism on f and hence on H. Up to $\mathrm{SL}_2(\mathbb{Z})$ equivalence and scaling, there is only one integral binary quadratic form having an $\mathrm{SL}_2(\mathbb{Z})$-automorphism of order 3, namely $Q(x, y) = x^2 + xy + y^2$. Thus, after a change of basis, we may assume that $H(x, y) = nQ(x, y)$ for some nonzero $n \in \mathbb{Z}$. Hence we have

$$b^2 - 3ac = n, \quad bc - 9ad = n, \quad c^2 - 3bd = n. \tag{2}$$

The first equation implies $a = (b^2 - n)/3c$, while the third implies $d = (c^2 - n)/3b$ (assuming b, c nonzero). Substituting these values of a, d into the second equation gives

$$b^2c^2 - (b^2 - n)(c^2 - n) = nbc.$$

Expanding out and dividing by n, we obtain

$$b^2 - bc + c^2 = Q'(b, c) = n.$$

We now have

$$a = \frac{bc - c^2}{3c} = \frac{b - c}{3} \quad \text{and} \quad d = \frac{bc - b^2}{3b} = \frac{c - b}{3}; \tag{3}$$

one easily checks that this gives the unique solution for a and d even when one of b or c is zero. Furthermore, f has integer coefficients precisely when $(b, c)^t \in L$, that is, b, c are integers congruent modulo 3.

Conversely, if $(b, c)^t \in L$ is nonzero, then we can define integers a and d as in (3) and set $f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$; this cubic form has Hessian $H_f(x, y) = nQ(x, y)$, where $n = Q'(b, c)$. A calculation shows that the order-3 transformation

$$S = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \tag{4}$$

is an automorphism of $f(x, y)$, and hence the ring R corresponding to f is a C_3-cubic ring.

Suppose we have binary cubic forms f, f' corresponding to $(b, c)^t$, $(b', c')^t \in L$.

Then f and f' are $\mathrm{SL}_2(\mathbb{Z})$-equivalent if and only if they are $\mathrm{SO}_Q(\mathbb{Z})$-equivalent, since they both have Hessian equal to nQ. Write

$$f(x,y)=\frac{b-c}{3}x^3+bx^2y+cxy^2+\frac{c-b}{3}y^3$$

and $f'=\gamma f$ with $\gamma\in\mathrm{SO}_Q(\mathbb{Z})$. Then an elementary computation shows that

$$(\gamma f)(x,y)=f((x,y)\gamma)=\frac{b'-c'}{3}x^3+b'x^2y+c'xy^2+\frac{c'-b'}{3}y^3,$$

where $(b',c')^t=\gamma^3(b,c)^t$. (The cubic action here is to be expected because the cube roots of the identity generated by $S\in\mathrm{SO}_Q(\mathbb{Z})$ must lie in the kernel of the action of $\mathrm{SO}_Q(\mathbb{Z})$ on L.) It follows that f and f' are $\mathrm{SO}_Q(\mathbb{Z})$-equivalent if and only if $(b,c)^t$ and $(b',c')^t$ are $\mathrm{SO}_Q(\mathbb{Z})$-equivalent under the cubic action. This proves the first part of the theorem. Finally, by Propositions 11 and 12, we know that $\operatorname{Disc} R=\operatorname{Disc} f=-\frac{1}{3}\operatorname{Disc} H$, and we compute $-\frac{1}{3}\operatorname{Disc} H=n^2=Q'(b,c)^2$. □

3.2. ***The number of C_3-cubic orders of bounded discriminant.*** To prove Theorem 1 we need the following lemma, which shows that the reducible forms $f(x,y)$ corresponding to C_3-cubic rings are negligible in number. This will allow us to prove asymptotics for C_3-cubic *orders* rather than just C_3-cubic rings.

Lemma 15. *The number of* $\mathrm{SL}_2(\mathbb{Z})$*-equivalence classes of reducible integral binary cubic forms having Hessian a multiple of* $Q(x,y)=x^2+xy+y^2$, *and discriminant less than* X, *is* $O(X^{1/4})$.

Proof. We give an upper estimate for the number of $\mathrm{SL}_2(\mathbb{Z})$-equivalence classes of reducible forms f of discriminant less than X whose Hessian is a multiple of Q. It will suffice to count first the primitive forms f, and then we will sum over all possible contents for f. Now any such primitive reducible f with Hessian nQ has a linear factor $gx+hy$ with g and h relatively prime integers. Furthermore, the order-3 automorphism (4) permutes the three roots of f in $\mathbb{P}^1$, and hence f must factor into the three primitive linear factors that are obtained by successively applying S to $gx+hy$. Thus we have

$$f(x,y)=(gx+hy)((h-g)x-gy)(-hx+(g-h)y).$$

Computing the discriminant of f, we find

$$\operatorname{Disc} f=(g^2-gh+h^2)^6=Q'(g,h)^6.$$

Thus, if $\operatorname{Disc} f<X$, then $Q'(g,h)<X^{1/6}$, and hence the total number of values for the pair (g,h), and thus f, is at most $O(X^{1/6})$.

In order to count the total number of forms f and not just the primitive ones, we sum over all possible values of the content c of f. Since $\operatorname{Disc}(f/c)=\operatorname{Disc}(f)/c^4$, we thus obtain

$$\sum_{c=1}^{X^{1/4}} O\Big(\Big(\frac{X}{c^4}\Big)^{1/6}\Big) = O(X^{1/4}) \tag{5}$$

as an upper estimate for the total number of $\mathrm{SL}_2(\mathbb{Z})$-equivalence classes of reducible forms f of discriminant less than X whose Hessian is a multiple of Q, as desired. □

Proof of Theorem 1. By Theorem 14 and Lemma 15, it now suffices to count elements $(b, c)^t \in L$, up to $\mathrm{SO}_Q(\mathbb{Z})$-equivalence, subject to the condition $Q'(b, c)^2 = (b^2 - bc + c^2)^2 < X$. The number of integral points inside the elliptic region cut out by the latter inequality is approximately equal to its area $(2\pi/\sqrt{3})X^{1/2}$, with an error of at most $O(X^{1/4})$ [Cohn 1980]. Meanwhile, being the (orientation-preserving) symmetry group of the triangular lattice, $\mathrm{SO}_Q(\mathbb{Z})$ is isomorphic to C_6, the cyclic group of order 6. Since this is the cubic action, the cyclic subgroup $C_3 \subset \mathrm{SO}_Q(\mathbb{Z})$ of order 3 acts trivially. Up to equivalence, we thus obtain

$$\frac{2\pi}{2\sqrt{3}}X^{1/2} + O(X^{1/4})$$

points inside the ellipse. The number of such points with $b \equiv c \pmod 3$ is therefore

$$\frac{\pi}{3\sqrt{3}}X^{1/2} + O(X^{1/4}).$$

This is the number of oriented C_3-cubic rings with discriminant bounded by X. By Lemma 15, the C_3-cubic rings that are not orders will be absorbed by the error term. After dividing by 2 to account for the fact that we counted oriented rings, we obtain the formula in Theorem 1. □

3.3. ***An elementary proof of Cohn's theorem on the number of abelian cubic fields of bounded discriminant.*** Now we wish to count those points $(b, c)^t \in L$ of bounded discriminant corresponding to maximal cubic orders. This is equivalent to counting abelian cubic extensions of $\mathbb{Q}$ of bounded discriminant. Since maximality is a local property, it suffices to determine how many C_3-cubic rings R satisfy the condition that the $\mathbb{Z}_p$-algebra $R \otimes \mathbb{Z}_p$ is maximal for every prime p. The following lemma gives a useful criterion to determine when a cubic ring R is maximal at p.

Lemma 16 [Bhargava et al. 2013, Lemma 13]. *If f is a binary cubic form over $\mathbb{Z}$ (or $\mathbb{Z}_p$), then $R(f)$ is not maximal at p if and only if one of the following conditions holds*:

- $f(x, y) \equiv 0 \pmod p$.
- *f is $\mathrm{GL}_2(\mathbb{Z})$-equivalent to a form $f'(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ such that $a \equiv 0 \pmod{p}^2$ and $b \equiv 0 \pmod p$.*

In particular, if Disc f is nonzero (mod p^2) then $R(f)$ is maximal at p.

With the help of this lemma, we now determine conditions for when the cubic order $R(f)$ corresponding to the binary cubic form

$$f(x, y) = \frac{b-c}{3}x^3 + bx^2y + cxy^2 + \frac{c-b}{3}y^3 \tag{6}$$

is maximal at p. We consider three cases, corresponding to the three possible residue classes of p (mod 3).

First suppose that $p \equiv 2 \pmod 3$. We have Disc $f = Q'(b, c)^2$ by Theorem 14, and $Q'(x, y) = x^2 - xy + y^2$ does not factor (mod p). Then, by the lemma, $R(f)$ is maximal at p as long as $(b, c) \not\equiv (0, 0)$ (mod p). We conclude that for $p \equiv 2$ (mod p), the p-adic density of elements $(b, c)^t \in L$ corresponding to maximal rings (at p) is $1 - 1/p^2$.

Next, suppose $p \equiv 1 \pmod 3$. In order for $Q'(b, c)$ to vanish modulo p, we need $c \equiv \zeta b$ (mod p), where ζ is a primitive sixth root of unity in $\mathbb{Z}/p\mathbb{Z}$. In that case, we obtain

$$f(x, y) \equiv \frac{b}{3}\zeta^{-1}\big(x^3 + 3\zeta x^2y + 3\zeta^2xy^2 + \zeta^3y^3\big) \equiv \frac{b}{3}\zeta^{-1}(x + \zeta y)^3 \pmod p. \tag{7}$$

If $b \equiv 0$ (mod p), then $f(x, y) \equiv 0$ (mod p) and so R is not maximal at p. Otherwise, if we have a pair (b, c) with $c \equiv \zeta b$ (mod p) and $b \not\equiv 0$ (mod p), then we may send the unique multiple root of $f(x, y)$ in $\mathbb{P}^1_{\mathbb{F}_p}$ to the point $(0, 1)$ via a transformation in $\mathrm{GL}_2(\mathbb{Z})$. Then, modulo p, the form $f(x, y)$ is congruent to a multiple of y^3. A proportion of $1/p$ of these forms satisfy $a \equiv 0$ (mod p^2), where a is the coefficient of x^3. By Lemma 16, the p-adic density of points $(b, c)^t \in L$ corresponding to rings maximal at p is therefore

$$1 - \frac{1 + 2(p-1)/p}{p^2} = \frac{p^3 - 3p + 2}{p^3} = \frac{(p-1)^2(p+2)}{p^3},$$

since there are two primitive sixth roots of unity ζ in $\mathbb{Z}/p\mathbb{Z}$ if $p \equiv 1 \pmod 3$.

Finally, if $p = 3$, then we wish to know the density of all $(b, c)^t \in L$ for which the binary cubic form $f(x, y)$ in (6) yields a cubic ring maximal at 3. Clearly, we need $b \equiv -c \pmod 3$ for the discriminant $(b^2 - bc + c^2)^2$ of the corresponding cubic ring to vanish modulo 3. Since already $b \equiv c \pmod 3$, we must have $b \equiv c \equiv 0$ (mod 3) to obtain a ring that is not maximal at 3. Write $b = 3B$ and $c = 3C$. Then we wish to know when the binary cubic

$$g(x, y) = (B - C)x^3 + 3Bx^2y + 3Cxy^2 + (C - B)y^3$$

corresponds to a cubic ring not maximal at 3. Note that $g(x, y) \equiv (B - C)(x - y)^3$ (mod 3) and $g(x, y) \equiv 0 \pmod 3$ if and only if $B \equiv C \pmod 3$. Otherwise, we can send the unique multiple root of $g(x, y)$ in $\mathbb{P}^1_{\mathbb{F}_p}$ to $(0, 1)$, which transforms $g(x, y)$ to a multiple of y^3. As before, a proportion of $\frac{1}{3}$ of such forms will have x^3 coefficient congruent to 0 (mod p^2). By Lemma 16, it follows that a proportion of $\frac{1}{3}\bigl(\frac{1}{3} + \frac{2}{3} \cdot \frac{1}{3}\bigr) = \frac{5}{27}$ of such $g(x, y)$ will correspond to a cubic ring not maximal at 3. We conclude that the density of $(b, c)^t \in L$ that yield a cubic ring maximal at 3 is $\frac{22}{27}$.

We have proven the following proposition.

Proposition 17. *Let $S_{\max}$ denote the set of all $(b, c)^t \in L$ corresponding to rings maximal at p under the bijection of Theorem 14. Then the p-adic density $\mu_p(S_{\max})$ of $S_{\max}$ in L is given by*

$$\mu_p(S_{\max}) = \begin{cases} (p-1)^2(p+2)/p^3 & \textit{if } p \equiv 1 \pmod 3, \\ 1 - 1/p^2 & \textit{if } p \equiv 2 \pmod 3, \\ \frac{22}{27} & \textit{if } p = 3. \end{cases} \tag{8}$$

The proof of Theorem 1 gives the total number $N(L; X)$ of points in L, up to $\mathrm{SO}_Q(\mathbb{Z})$-equivalence, having discriminant at most X. We may similarly determine the number $N(S; X)$ of points, up to $\mathrm{SO}_Q(\mathbb{Z})$-equivalence, having discriminant at most X, where S is any $\mathrm{SO}_Q(\mathbb{Z})$-invariant subset of L defined by finitely many congruence conditions.

Proposition 18. *Let $S \subset L$ be an $\mathrm{SO}_Q(\mathbb{Z})$-invariant subset that is defined by congruence conditions modulo finitely many prime powers. Then the number $N(S; X)$ of points in S, up to $\mathrm{SO}_Q(\mathbb{Z})$-equivalence, having discriminant at most X is given by*

$$N(S; X) = \frac{\pi}{6\sqrt{3}} \prod_p \mu_p(S) \cdot X^{1/2} + O_S(X^{1/4}), \tag{9}$$

where $\mu_p(S)$ denotes the p-adic density of S in L.

The proposition follows from arguments essentially identical to those in the proof of Theorem 1.

The set $S_{\max}$ of elements in L that correspond to maximal cubic rings, however, is defined by infinitely many congruence conditions. To show that (9) still holds for such a set, we require a uniform estimate on the error in (9) when the congruence conditions defining $S_{\max}$ are imposed only at the finitely many primes $\le Y$, as $Y \to \infty$. This is the content of the next result:

Proposition 19. *Let $S_{\max}^{\le Y}$ denote the subset of L corresponding to cubic rings maximal at all primes $\le Y$. Then*

$$N(S_{\max}^{\le Y}; X) - N(S_{\max}; X) = O\Bigl(\frac{X^{1/2}}{Y}\Bigr).$$

Proof. Let W_p denote the subset of elements in L corresponding to cubic rings R that are not maximal at p. Any such ring R is contained in a maximal ring R', where R' also has shape x^2+xy+y^2 (this is because the field containing R must have an order-3 automorphism, and then so does R'). The number of such possible R' (up to isomorphism) with discriminant less than X is $O(X^{1/2})$ by Theorem 1. To count all orders R in such R' having discriminant less than X, we require the following lemma.

Lemma 20 [Datskovsky and Wright 1988]. *If R' is any maximal cubic ring, then the number of orders $R \subset R'$ of index $m = \prod p_i^{e_i}$ is $O_\epsilon\bigl(\prod p_i^{(1+\epsilon)\lfloor e_i/3\rfloor}\bigr)$ for any $\epsilon > 0$.*

The lemma implies that the total number of cubic rings R of discriminant less than X that are not maximal at p and are contained in maximal rings R' of shape x^2+xy+y^2 is at most

$$\biggl(\sum_{e=1}^{\infty}\frac{p^{(1+\epsilon)\lfloor e/3\rfloor}}{p^{2e}}\biggr)\prod_{q\neq p}\biggl(\sum_{e=0}^{\infty}\frac{q^{(1+\epsilon)\lfloor e/3\rfloor}}{q^{2e}}\biggr)O(X^{1/2}) = O\Bigl(\frac{X^{1/2}}{p^2}\Bigr).$$

Since $\displaystyle\sum_{p\geq Y} O\Bigl(\frac{X^{1/2}}{p^2}\Bigr) = O\Bigl(\frac{X^{1/2}}{Y}\Bigr)$, we obtain the desired estimate. □

Thus, by choosing Y large enough, we can make $N\bigl(S_{\max}^{\leq Y}; X\bigr) - N(S_{\max}; X) \leq cX^{1/2}$ for any $c > 0$. We conclude that the number of C_3-cubic fields of discriminant less than X is asymptotic to

$$\frac{\pi}{6\sqrt{3}}\cdot\frac{22}{27}\prod_{p\equiv 1(3)}\frac{(p-1)^2(p+2)}{p^3}\prod_{p\equiv 2(3)}\Bigl(1-\frac{1}{p^2}\Bigr)\cdot X^{1/2}$$
$$= \frac{11\pi}{3^4\sqrt{3}}\cdot\frac{6}{\pi^2}\cdot\frac{9}{8}\prod_{p\equiv 1(3)}\frac{(p-1)(p+2)}{p(p+1)}\cdot X^{1/2},$$

which is

$$\frac{11\sqrt{3}}{36\pi}\prod_{p\equiv 1(3)}\Bigl(1-\frac{2}{p(p+1)}\Bigr)\cdot X^{1/2};$$

and this is the result of Cohn [1954].

4. On cubic orders having a general fixed lattice shape

Let $Q(x, y)$ be a primitive integral binary quadratic form with nonsquare discriminant. In this section we determine asymptotics for the number $N_3^{\mathrm{Or}}(Q, X)$ of oriented cubic orders having absolute discriminant bounded by X and shape Q;

that is, we prove Theorem 3. To accomplish this, we generalize the proofs of the previous section.

We choose to work with *oriented* cubic rings for a couple of reasons. First, this allows us to ignore the determinant -1 automorphisms in $\mathrm{GO}_Q(\mathbb{Z})$, making the proof a bit simpler. Second, Theorem 3 shows that, at least asymptotically, lattice shapes are equidistributed within the (narrow) class group, suggesting that oriented rings are the natural framework for our analysis.

4.1. ***A more general action of*** $\mathbf{SO}_Q(\mathbb{C})$ ***on*** $\mathbb{C}^2$**.** Recall that the shape of a cubic ring R is an equivalence class of binary quadratic forms. We begin by fixing a representative of this class, say $Q(x, y) = rx^2 + sxy + ty^2$. As in the C_3-cubic ring case, we consider a lattice $L = L(Q)$ of elements $(b, c)^t \in \mathbb{Z}^2$, defined by the congruence conditions

$$sb \equiv rc \pmod{3t} \quad \text{and} \quad sc \equiv tb \pmod{3r}.$$

Let $Q'(x, y) = tx^2 - sxy + ry^2$ denote the adjoint quadratic form of $Q(x, y)$. Let $\mathrm{SO}_Q(\mathbb{C})$ denote the subgroup of transformations $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ that fix Q via $(\gamma Q)(x, y) = Q((x, y)\gamma)$. Then we define the *cubic action* of $\mathrm{SO}_Q(\mathbb{C})$ on $\mathbb{C}^2$ just as before, namely $\gamma \cdot v = \gamma^3 v$ for $v \in \mathbb{C}^2$. We will see that $\mathrm{SO}_Q(\mathbb{Z})$, the subgroup of elements in $\mathrm{SO}_Q(\mathbb{C})$ having integer entries, preserves L, and the quadratic form Q' gives an invariant polynomial on L. Define the subset

$$L(Q)^+ = \left\{(b, c)^t \in L(Q) : \frac{Q'(b, c)}{rt} > 0\right\} \subset L(Q).$$

Then we have the following generalization of Theorem 14.

Theorem 21. *Let $Q(x, y) = rx^2 + sxy + ty^2$ be a primitive integral quadratic form with nonsquare discriminant, and let $Q'(x, y) = tx^2 - sxy + ry^2$ denote the adjoint quadratic form of $Q(x, y)$. Then the orbits of the cubic action of $\mathrm{SO}_Q(\mathbb{Z})$ on lattice points $(b, c)^t \in L(Q)^+$ are in natural bijection with the isomorphism classes of oriented cubic rings R having shape Q. Under this bijection, we have*

$$\mathrm{Disc}\, R = -\frac{Q'(b, c)^2 \,\mathrm{Disc}\, Q}{3r^2t^2}.$$

Proof. The proof is similar to that of Theorem 14. Let R be a cubic ring with shape Q. Then, by applying an appropriate $\mathrm{SL}_2(\mathbb{Z})$-transformation, we may assume that the corresponding integral cubic form $f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ has Hessian

$$H(x, y) = n(rx^2 + sxy + ty^2) = nQ(x, y),$$

with n positive. From the definition of the Hessian, we have

$$b^2 - 3ac = nr, \quad bc - 9ad = ns, \quad c^2 - 3bd = nt, \tag{10}$$

for some positive integer n. Assuming b, c are nonzero, these equations imply that

$$a = \frac{b^2 - nr}{3c} \quad \text{and} \quad d = \frac{c^2 - nt}{3b}.$$

Using the middle equation in (10), we find that

$$tb^2 - sbc + rc^2 = ntr. \tag{11}$$

We then get

$$a = \frac{sb - rc}{3t}, \quad d = \frac{sc - tb}{3r} \tag{12}$$

(note that r and t are nonzero because Disc Q is not a square), and one checks that this is the unique solution even if b or c is zero. Notice that $f(x, y)$ has integer coefficients if and only if

$$sb \equiv rc \pmod{3t} \quad \text{and} \quad sc \equiv tb \pmod{3r}, \tag{13}$$

that is, $(b, c)^t \in L$. In this case, we even have $(b, c)^t \in L(Q)^+$, since n is positive. We see that the form $f(x, y)$ is determined once we specify the shape Q and the middle coefficients b and c. Conversely, given any element $(b, c)^t \in L(Q)^+$, we may use the equations in (12) to define a cubic form $f = (a, b, c, d)$ such that $R(f)$ has shape Q and the Hessian of f is nQ for some positive integer n.

Now suppose $f = (a, b, c, d)$ and $f' = (a', b', c', d')$ are two binary cubic forms chosen such that the Hessians are nQ and $n'Q$ with integers $n, n' > 0$; thus, the respective conditions in (12) and (13) hold for the coefficients of f and f'. Then f and f' are $\mathrm{SL}_2(\mathbb{Z})$-equivalent if and only if they are $\mathrm{SO}_Q(\mathbb{Z})$-equivalent. A computation as in the proof of Theorem 14 shows that if $f' = \gamma f$ for $\gamma \in \mathrm{SO}_Q(\mathbb{Z})$, then $(b', c')^t = \gamma^3 (b, c)^t$. It follows that f and f' are $\mathrm{SO}_Q(\mathbb{Z})$-equivalent if and only if $(b, c)^t$ and $(b', c')^t$ are $\mathrm{SO}_Q(\mathbb{Z})$-equivalent under the cubic action.

Thus we have proved the bijection described in the theorem. Further, we have

$$\operatorname{Disc} R = \operatorname{Disc} f = -\frac{n^2 \operatorname{Disc} Q}{3}$$

by Proposition 11, and combining with $Q'(b, c) = nrt$, which was Equation (11), we obtain the desired result. □

Remark 22. If Q is positive definite, then $L(Q)^+$ is the set of nonzero vectors in $L(Q)$. If Q is negative definite, then $L(Q)^+$ is empty. If Q is indefinite, then nonzero elements of L not in $L(Q)^+$ correspond to cubic rings with shape $-Q$.

4.2. *The number of cubic orders of bounded discriminant and given lattice shape.*

We are nearly ready to prove Theorem 3. We consider the cases of definite and indefinite Q separately.

4.2.1. *Definite case.* In this case, we have the following well known lemma.

Lemma 23. *Let $Q(x, y)$ be a definite integral binary quadratic form. The order of* $\mathrm{SO}_Q(\mathbb{Z})$ *is either* 6, 4, *or* 2 *depending on whether the form Q* (*up to equivalence and scaling*) *is* x^2+xy+y^2, x^2+y^2, *or any other definite form.*

We next prove the analogue of Lemma 15 for general definite forms.

Lemma 24. *Let $Q(x, y)$ be a definite integral binary quadratic form of nonsquare discriminant D. Then the number of* $\mathrm{SL}_2(\mathbb{Z})$*-equivalence classes of reducible cubic forms f having shape Q and* $|\operatorname{Disc} f| < X$ *is* $O(X^{1/4})$.

Proof. We give an upper estimate for the number of $\mathrm{SL}_2(\mathbb{Z})$-equivalence classes of forms f of discriminant less than X whose Hessian is a multiple of the fixed definite binary quadratic form $Q(x, y) = rx^2 + sxy + ty^2$ of discriminant D. As in the proof of Lemma 15, it suffices to first count just those f that are primitive.

Now an order-3 automorphism of such a primitive f (over $\bar{\mathbb{Q}}$) of determinant 1 is given by

$$S = \begin{pmatrix} -\dfrac{-\sqrt{D}-s\sqrt{-3}}{2\sqrt{D}} & r\sqrt{\dfrac{-3}{D}} \\ -t\sqrt{\dfrac{-3}{D}} & -\dfrac{-\sqrt{D}+s\sqrt{-3}}{2\sqrt{D}} \end{pmatrix}.$$

The transformation S permutes the roots of f in $\mathbb{P}^1(\bar{\mathbb{Q}})$. So if $gx+hy$ is a linear factor of $f(x, y)$ with g and h coprime integers, then by computing the two other linear factors (obtained by applying S and S^{-1}) and multiplying all three factors together, we obtain a polynomial f_0 that must agree with f up to scaling. Since $\sqrt{D}S$ is an integral matrix of discriminant D, the content of f_0 must divide D^2. Computing the discriminant of f_0, we obtain

$$\operatorname{Disc} f0) = -\frac{27}{D^6}(tg^2 - sgh + rh^2)^6 = -\frac{27}{D^6}Q'(g, h)^6.$$

Since Q' is definite and $\operatorname{Disc} f0) \leq D^8 \operatorname{Disc} f < D^8 X = O(X)$, we see that the total number of choices for (g, h), and thus f, is $O(X^{1/6})$.

Finally, to obtain an upper estimate for the count of all f that are not necessarily primitive, we sum over all possible contents c of f as in (5). We obtain a total of $O(X^{1/4})$ possibilities for f, as desired. □

We denote by $C(Q)$ the cardinality of the subgroup of cubes in $\mathrm{SO}_Q(\mathbb{Z})$. Thus $C(Q) = |\mathrm{SO}_Q|$ unless Q has discriminant -3 (which is the C_3 case already dealt with in Section 3).

Theorem 25. *Let $Q(x,y)=rx^2+sxy+ty^2$ be a positive definite primitive integral quadratic form of discriminant $-D$. Let $\alpha=1$ if $3\mid D$ and $\alpha=2$ otherwise. Then*

$$N_3^{\mathrm{Or}}(Q,X)=\frac{2\pi\sqrt{3}}{3^{\alpha}C(Q)D}X^{1/2}+O(X^{1/4}).$$

Proof. By Theorem 21 and Lemma 24, it suffices to count equivalence classes of elements $(b,c)^t$ in $L(Q)^+$ such that

$$\frac{Q'(b,c)^2D}{3r^2t^2}<X,$$

or equivalently,

$$tb^2-sbc+rc^2<\frac{\sqrt{3}rtX^{1/2}}{\sqrt{D}}.$$

In this case, $L(Q)^+$ is simply the set of nonzero vectors in $L(Q)$. The number of $L(Q)$-points in the elliptic region in $\mathbb{R}^2$ defined by the inequality above is approximately given by the area of this ellipse[4] divided by the area $\mathrm{Vol}(L)$ of a fundamental parallelogram of L. The error is at most[5] $O(X^{1/4})$, where the implied constant depends on the shape of L and thus on Q [Cohn 1980]. So we have

$$N_3^{Q}(X)=\frac{2\pi rt\sqrt{3}X^{1/2}}{\mathrm{Vol}(L)D}+O(X^{1/4}). \tag{14}$$

We further divide by $C(Q)$ to obtain the number of points in L satisfying the inequality up to the cubic action equivalence. Thus the theorem follows from the following lemma.

Lemma 26. *Let r,s,t be integers with no common prime factor and set $D=s^2-4rt$. Also set $\alpha=1$ if $3\mid D$ and $\alpha=2$ otherwise. Then the lattice*

$$L(Q)=\{(b,c)^t\in\mathbb{Z}^2 : sb\equiv rc \pmod{3t} \text{ and } sc\equiv tb \pmod{3r}\}$$

has volume $3^{\alpha}rt$.

Proof. This proof is due to Julian Rosen. Let L' be the lattice in $\mathbb{Z}^2$ generated by $(3t,0)^t$ and $(0,3r)^t$. Then L is the inverse image of L' under the map $\mathbb{Z}^2\to\mathbb{Z}^2$ that sends $(x,y)^t$ to $(sx-ry,tx-sy)^t$. Then $\mathrm{Vol}(L)=\mathrm{Vol}(L')/\mathrm{Vol}(C)$, where C is the lattice spanned by columns of the matrix

$$M=\begin{pmatrix} s & -r & 3t & 0\\ t & -s & 0 & 3r\end{pmatrix}.$$

[4] The area bounded by an ellipse with equation $Q(x,y)=N$ is $2\pi N/\sqrt{D}$.

[5] In fact, work on Gauss's circle problem gives exponents considerably smaller than $\frac{1}{4}$ [Hardy 1915].

Since $\mathrm{Vol}(C)$ is the greatest common divisor of the two-by-two minors of M, which is precisely $3^{2-\alpha}$, we have $\mathrm{Vol}(L) = 9rt/3^{2-\alpha} = 3^{\alpha}rt$ as claimed. □

This concludes the proof of Theorem 25. □

Remark 27. It is natural to sum the main term of Theorem 25 over all shapes of negative discriminant to try to recover Davenport's count [1951] of the number of cubic rings with positive discriminant bounded by X. The method of [Siegel 1944] makes it possible to compute this sum, but it turns out not to equal Davenport's main term $(\pi^2/72)X$. Evidently, the error term in Theorem 25 is contributing to the main term of this sum.

4.2.2. *Indefinite case.* Next we consider counting cubic orders having an indefinite shape $Q(x, y)$. Again, write $Q(x, y) = rx^2 + sxy + ty^2$. The equations (12) are not well-defined if either r or t is zero, so we assume that $D = \mathrm{Disc}\, Q = s^2 - 4rt$ is not a square. We may also assume that $t > 0$.

As in the proof of Theorem 25, we need to count elements $(b, c)^t$ in $L(Q)^+$ up to $\mathrm{SO}_Q(\mathbb{Z})$-equivalence, such that

$$\left|\frac{Q'(b, c)^2 D}{3r^2t^2}\right| < X, \quad \text{or in other words} \quad \left|tb^2 - sbc + rc^2\right| < \left|\frac{\sqrt{3}rtX^{1/2}}{\sqrt{D}}\right|.$$

This inequality cuts out a region in the plane bounded by a hyperbola, and we need to count the orbits of the cubic action of $\mathrm{SO}_Q(\mathbb{Z})$ on $L(Q)^+$ that intersect this region. But $\mathrm{SO}_Q(\mathbb{Z})$ is now an infinite group, so we must construct a fundamental domain for the action at hand (the construction and the ensuing volume computation are taken from [Davenport 2000, Chapter 6]).

In what follows, it is useful to define $\theta = (s + \sqrt{D})/2t$ and $\theta' = (s - \sqrt{D})/2t$. We then have

$$Q'(x, y) = tx^2 - sxy + ry^2 = t(x - \theta y)(x - \theta' y).$$

We also have the following well known facts.

Proposition 28. • *The integral solutions (U, W) of the generalized Pell's equation $u^2 - Dw^2 = 4$ are given by*

$$\tfrac{1}{2}(U + W\sqrt{D}) = \pm\big[\tfrac{1}{2}(U_0 + W_0\sqrt{D})\big]^n,$$

where n is any integer and (U_0, W_0) is a minimal solution.

• *Every element M in $\mathrm{SO}_Q(\mathbb{Z})$ is of the form*

$$M = \begin{pmatrix} \frac{1}{2}(U + sW) & -rW \\ tW & \frac{1}{2}(U - sW) \end{pmatrix}$$

for some solution (U, W) to $u^2 - Dw^2 = 4$.

If $(X, Y)^t \in \mathbb{Z}^2$ and $M \cdot (X, Y)^t = (x, y)^t$ for some $M \in \mathrm{SO}_Q(\mathbb{Z})$, then the second part of the proposition implies

$$\frac{x - \theta' y}{x - \theta y} = \frac{\frac{1}{2}(U + W\sqrt{D})}{\frac{1}{2}(U - W\sqrt{D})} \cdot \frac{X - \theta' Y}{X - \theta Y}$$

for some solution (U, W) to Pell's equation. If we define $\epsilon = \frac{1}{2}(U_0 + W_0\sqrt{D}) > 1$, then the first part gives

$$\tfrac{1}{2}(U + W\sqrt{D}) = \pm\epsilon^m \quad \text{and} \quad \tfrac{1}{2}(U - W\sqrt{D}) = \pm\epsilon^{-m}$$

for some integer m. Thus, in each orbit of $L(Q)^+$ there is a single element $(x, y)^t$ such that

$$1 \le \frac{x - \theta' y}{x - \theta y} < \epsilon^2$$

and $x - \theta y > 0$.

Our goal, then, is to count the number of integer points $(x, y)^t$ obeying the constraints

$$tx^2 - sxy + ry^2 \le N = \left| \frac{\sqrt{3}rtX^{1/2}}{\sqrt{\operatorname{Disc} Q}} \right|, \quad x - \theta y > 0, \quad 1 \le \frac{x - \theta' y}{x - \theta y} < \epsilon^2.$$

This region is a sector emanating from the origin bounded by a hyperbola. Just as in the positive definite case, we can approximate this count by computing the area of this region.

Changing coordinates from x, y to

$$\xi = x - \theta y, \quad \eta = x - \theta' y,$$

this region is

$$\xi\eta \le \frac{N}{t}, \quad \xi > 0, \quad \xi \le \eta < \epsilon^2 \xi.$$

These conditions can be rewritten as

$$0 < \xi \le \left(\frac{N}{t}\right)^{1/2}, \quad \xi \le \eta < \min\left(\epsilon^2 \xi, \frac{N}{t\xi}\right).$$

If we define $\xi_1 = \epsilon^{-1}(N/t)^{1/2}$, then the area of this region is

$$\int_0^{\xi_1} (\epsilon^2 \xi - \xi)\, d\xi + \int_{\xi_1}^{(N/t)^{1/2}} \left(\frac{N}{t\xi} - \xi\right) d\xi.$$

Evaluating the integrals, we obtain

$$(\epsilon^2 - 1)\frac{1}{2}\xi_1^2 + \frac{N}{2t}\log\left(\frac{N}{t}\right) - \left(\frac{N}{t}\right)\log \xi_1 - \frac{1}{2}\left(\frac{N}{t}\right) + \frac{1}{2}\xi_1^2,$$

which simplifies to $(N/t)\log\epsilon$. This is the area in the ξ, η-coordinate system; to get the area in x, y-coordinates, we divide by

$$\frac{\partial(\xi,\eta)}{\partial(x,y)}=\theta-\theta'=\frac{\sqrt{D}}{t}.$$

Thus the desired area is $(N/\sqrt{D})\log\epsilon$. However, recall that we are interested in the orbits of the cubic action of SO_Q on the lattice $L(Q)$, so we must replace ϵ by ϵ^3 in the previous calculation. In the final area estimate, this yields an extra factor of 3 (since $\log\epsilon^3=3\log\epsilon$).

We may now continue as in the rest of the proof of Theorem 25. The proof of Lemma 24 carries over to the indefinite case because we can again bound the number of points of discriminant $< X$ in the fundamental domain in terms of the corresponding area. The final result is:

Theorem 29. *Let $Q(x,y)=rx^2+sxy+ty^2$ be a primitive integral quadratic form having nonsquare discriminant $D>0$. Let $\alpha=1$ if $3\mid D$ and $\alpha=2$ otherwise. Then*

$$N_3^{\mathrm{Or}}(Q,X)=\frac{3\sqrt{3}\log\epsilon}{3^\alpha D}X^{1/2}+O(X^{1/4}).$$

Dirichlet's class number formula [Davenport 2000] states that

$$h(D)=\frac{w\sqrt{|D|}}{2\pi}L(1,\chi_D)$$

for $D<0$, where w is the number of roots of unity in $\mathbb{Q}(\sqrt{D})$, and

$$h(D)=\frac{\sqrt{D}}{\log\epsilon}L(1,\chi_D)$$

for $D>0$.[6] Using these equations, we see that Theorems 25 and 29 can be combined and the result is Theorem 3.

4.3. ***The number of maximal cubic orders of bounded discriminant and given lattice shape.*** As before, let $Q(x,y)=rx^2+sxy+ty^2$ be a quadratic form over $\mathbb{Z}$, with $D=s^2-4rt$ not a square, and let us further assume that Q is primitive.

In this subsection, we use the results of Subsection 4.2 to provide asymptotics for $M_3^{\mathrm{Or}}(Q,X)$, the number of isomorphism classes of *maximal* oriented cubic orders having shape Q. We may ignore those maximal cubic rings that are not orders, because such rings can be written as $\mathbb{Z}\oplus S$, where S is a maximal quadratic ring of discriminant (a rational square multiple of) $-3\,\mathrm{Disc}\,Q$. But there is at most one such maximal quadratic ring and this one exception will be absorbed by the error term. Thus $M_3(Q,X)$ (resp. $M_3^{\mathrm{Or}}(Q,X)$) is essentially the number of cubic

[6] Recall that $h(D)$ is the *narrow* class number.

fields (resp. oriented cubic fields) with ring of integers of shape Q and absolute discriminant less than X.

Just as in the C_3 case in Section 3, we compute the p-adic density of those elements in $L = L(Q)$ corresponding to cubic rings of shape Q that are maximal at p. For every such ring, we can choose a corresponding integral binary cubic form to be

$$f(x, y) = \frac{sb - rc}{3t}x^3 + bx^2y + cxy^2 + \frac{sc - tb}{3r}y^3$$

for some pair $(b, c)^t$ in the lattice L defined by the congruence conditions $sb \equiv rc \pmod{3}t$ and $sc \equiv tb \pmod{3}r$. We proved in the previous section that

$$\operatorname{Disc} f = \frac{D}{3r^2t^2} Q'(b, c)^2, \tag{15}$$

where $Q'(x, y) = tx^2 - sxy + ry^2$.

As it differs from other primes, we first consider primes other than $p = 3$, and then treat $p = 3$ separately. The reader only interested in the results may consult Table 1 on page 77.

In what follows, we denote the reduction (mod p) of the form $Q(x, y)$ by $Q_p(x, y)$.

4.3.1. *The p-adic density for maximality ($p \neq 3$).* We naturally divide into three cases.

Case 1 ($Q_p(x, y)$ has distinct roots in $\mathbb{F}_p$). Using an $\mathrm{SL}_2(\mathbb{Z})$-transformation, we may arrange for $Q_p(x, y)$ to be sxy. The congruence conditions defining L imply that $b \equiv c \equiv 0 \pmod{p}$ in this case. Then

$$f(x, y) \equiv Ax^3 + Dy^3 \pmod{p},$$

where $A, D \in \mathbb{Z}$ are independent parameters. Since $\operatorname{Disc} f \equiv -27A^2D^2 \pmod{p}$, the cubic ring $R(f)$ is maximal unless either A or D is 0 (mod p). By Lemma 16, $R(f)$ is not maximal at p precisely when either one of A or D is 0 (mod p^2) or they simultaneously vanish (mod p). Thus the p-adic density of the set of $(b, c)^t \in L$ that give rise to maximal rings at p is

$$1 - \frac{1}{p^2} - \frac{2(p-1)}{p^3} = \frac{(p-1)^2(p+2)}{p^3}.$$

Case 2 ($Q_p(x, y)$ has distinct roots in $\mathbb{F}_{p^2} - \mathbb{F}_p$). The discriminant D, as well as the outer coefficients r and t, must be nonzero (mod p) in this case. Since Q_p (and hence Q'_p) does not factor modulo p, we see from (15) that p divides $\operatorname{Disc} f$ if and only if $b \equiv c \equiv 0 \pmod{p}$, in which case $f(x, y)$ vanishes (mod p) and so $R(f)$ is not maximal at p. Thus the p-adic density of the set of $(b, c)^t \in L$ that give rise to maximal rings at p in this case is $1 - 1/p^2$.

Case 3 ($Q_p(x, y)$ has a double root in $\mathbb{F}_p$). The quadratic form Q_p has a double root in $\mathbb{F}_p$ if and only if p divides the discriminant D of Q. By sending the double root (mod p) of Q_p to 0 via a transformation in $\mathrm{SL}_2(\mathbb{Z})$, we may assume that Q_p is the form rx^2 (mod p).

Since $t \equiv 0 \pmod{p}$, we see that $c \equiv 0 \pmod{p}$ for all $(b, c)^t \in L$, by the definition of L. Thus in $\mathbb{F}_p$, we have

$$f(x, y) \equiv \frac{sb - rc}{3t} x^3 + bx^2 y \equiv x^2 \biggl(\frac{sb - rc}{3t} x + by \biggr).$$

The coefficient $(sc - tb)/3r$ of y^3 in $f(x, y)$ is 0 (mod p^2) precisely when p^2 divides tb. If $p > 3$, then p^2 divides t if and only if it divides $D = s^2 - 4rt$. Thus for $p > 3$, the p-adic density of the set of $(b, c)^t \in L$ that give rise to maximal rings at p is 0 if $p^2 \mid D$ and $1 - 1/p$ if $p \parallel D$. If $p = 2$, then we write $D = 4m$ and note that $4 \mid t$ if and only if m is congruent to 0 or 1 (mod 4). Thus the density is 0 when m is congruent to 0 or 1 (mod 4) and is $\frac{1}{2}$ when m is congruent to 2 or 3 (mod 4).

4.3.2. *The 3-adic density for maximality.* We again divide into three cases.

Case 1 ($Q_3(x, y)$ has distinct roots in $\mathbb{F}_3$). In this case, it is most convenient to assume that $Q_3(x, y) = x(x + y)$. The congruence conditions defining L imply that $b \equiv c \equiv 0 \pmod{3}$. We then find that

$$f(x, y) \equiv -Nx^3 - My^3 = -(NX + MY)^3$$

for parameters $N, M \in \mathbb{Z}$. After sending the single root of $f(x, y)$ over $\mathbb{F}_3$ to 0, we see that aside from the degenerate form, one third of these forms will have the coefficient of y^3 congruent to 0 (mod 3^2). By Lemma 16, the 3-adic density of the set of $(b, c)^t \in L$ that give rise to maximal rings at 3 is

$$1 - \frac{1 + \frac{1}{3}(3^2 - 1)}{3^2} = \frac{16}{27}.$$

Case 2 ($Q_3(x, y)$ has distinct roots in $\mathbb{F}_9 - \mathbb{F}_3$). Now we may assume that $Q_3(x, y)$ is the form $x^2 + y^2$. Then the conditions defining the lattice L imply that $b \equiv c \equiv 0$ (mod 3) for all $(b, c)^t \in L$. So over $\mathbb{F}_3$ we have

$$f(x, y) \equiv -Cx^3 - By^3 = -(CX + BY)^3,$$

where $b = 3B$ and $c = 3C$. Arguing as in the previous case, we conclude that the 3-adic density of the set of $(b, c)^t \in L$ that give rise to maximal rings at 3 is $\frac{16}{27}$.

Case 3 ($Q_3(x, y)$ has a double root in $\mathbb{F}_3$). We may assume that $Q_3(x, y) = rx^2$, so $c \equiv 0 \pmod{3}$ for all $(b, c)^t \in L$. We first consider the case where $3 \parallel D$, which implies that $3 \parallel t$. Notice that

$$\operatorname{Disc} f = \frac{D}{3r^2t^2} Q'(b,c)^2$$

is divisible by 3 if and only if $b \equiv 0 \pmod 3$. But if $b \equiv 0 \pmod 3$, then $c \equiv 0 \pmod 9$,

$$f(x,y) \equiv \frac{sb-rc}{3t} x^3 \pmod 3,$$

and the coefficient of y^3 vanishes modulo p^2 precisely when $b \equiv 0 \pmod 9$. Thus the 3-adic density of the set of $(b,c)^t \in L$ that give rise to maximal rings at 3 is

$$1 - \tfrac{1}{3}\big(\tfrac{1}{3} + \tfrac{2}{3}\cdot\tfrac{1}{3}\big) = \tfrac{22}{27}.$$

Next we assume that 9 divides D, so that 9 divides t as well. By definition, $sb \equiv rc \pmod{27}$ for all $(b,c)^t \in L$; in particular $c \equiv 0 \pmod 3$. We also have that

$$f(x,y) \equiv \frac{sb-rc}{3t} x^3 + bx^2y \pmod 3.$$

By Lemma 16, it suffices to determine when the coefficient $(sc-tb)/3r$ vanishes modulo 9, that is, when $sc-tb$ vanishes modulo 27. This happens when 3 divides b (because then 9 divides c). If 3 does not divide b, since $sb \equiv rc \pmod{27}$, the congruence $sc \equiv tb \pmod{27}$ is equivalent to $s^2 \equiv rt \pmod{27}$, which is equivalent to $27 \mid D$. Thus if $27 \mid D$, then there are no maximal rings $R(f)$, and if $9 \parallel D$, then the 3-adic density of the set of $(b,c)^t \in L$ that give rise to maximal rings at 3 is $\frac{2}{3}$.

Table 1 shows the p-adic density of points $(b,c)^t \in L$ giving rise to a maximal ring at p. In practice, one determines whether a particular prime fits into Case 1, 2, or 3 based on whether the quadratic residue $\big(\frac{D}{p}\big)$ is -1, 1, or 0, respectively. For $p=2$, this is not well defined, but for convenience we define

$$\left(\frac{D}{2}\right) = \begin{cases} 1 & \text{if } D \equiv 1 \pmod 8, \\ -1 & \text{if } D \equiv 5 \pmod 8. \end{cases}$$

These densities, which we denote by $\mu_p(D)$, are in fact a function of the discriminant D of Q, that is, they are independent of the particular equivalence class of Q.

	$\big(\frac{D}{p}\big) = -1$	$\big(\frac{D}{p}\big) = 1$	$p \parallel D$	$p^2 \parallel D$	$p^3 \parallel D$	$p^4 \mid D$
$p=2$	$\frac{3}{4}$	$\frac{1}{2}$		$\frac{1}{2}$ or 0	$\frac{1}{2}$	0
$p=3$	$\frac{16}{27}$	$\frac{16}{27}$	$\frac{22}{27}$	$\frac{2}{3}$	0	0
$p \geq 5$	$1-1/p^2$	$(p-1)^2(p+2)/p^3$	$1-1/p$	0	0	0

Table 1. Densities $\mu_p(D)$ of $(b,c)^t \in L$ corresponding to rings maximal at p.

In this table, there is an ambiguity in the case where $p = 2$ and $4 \parallel D$. To resolve the ambiguity, one writes $D = 4m$, and if $m \equiv 1 \pmod 4$, then the density is 0, whereas if $m \equiv 3 \pmod 4$, then the density is $\frac{1}{2}$. The table implies the following proposition.

Proposition 30. *If $R(f)$ is a maximal cubic ring with shape $Q(x, y)$ of discriminant D, then either D is a fundamental discriminant or $-D/3$ is a fundamental discriminant.*

We can now use Theorem 25, together with the analogue of the uniformity estimate of Proposition 19 (the proof carries over to any L via Proposition 30), to compute the number of maximal cubic orders (equivalently, cubic fields) of bounded discriminant and with shape $Q(x, y)$. The uniformity estimate allows us (as in Subsection 3.3) to multiply each p-adic density of rings maximal at p, for a given shape $Q(x, y)$, to obtain the proportion of maximal cubic orders of that shape. By the previous proposition, we need only consider quadratic forms with discriminant D squarefree away from 2 or 3.

Recall that $M_3^{\mathrm{Or}}(Q, X)$ denotes the number of oriented maximal cubic orders with shape Q and absolute discriminant less than X, and $N_3^{\mathrm{Or}}(Q, X)$ denotes the number of oriented cubic orders with shape Q and absolute discriminant less than X.

Theorem 31. *Let $Q(x, y)$ be a quadratic form whose discriminant is not a square. Then*

$$M_3^{\mathrm{Or}}(Q, X) = N_3^{\mathrm{Or}}(Q, X) \prod_p \mu_p(D) + o(X^{1/2}).$$

As a corollary, we prove Theorem 4.

Proof of Theorem 4. Using Theorem 3, we compute the main term of $M_3^{\mathrm{Or}}(Q, X)$ (in the following products over primes, $p = 3$ is never included):

$$\frac{3^{\alpha+\beta-3/2} L(1, \chi_D)}{h(D)\sqrt{|D|}} \cdot \mu_3(D) \prod_{(\frac{D}{p})=-1} \left(1 - \frac{1}{p^2}\right) \prod_{(\frac{D}{p})=1} \left(\frac{(p-1)^2(p+2)}{p^3}\right) \prod_{p \mid D} \left(1 - \frac{1}{p_i}\right) X^{1/2}$$

$$= \frac{3^{\alpha+\beta-3/2} L(1, \chi_D)}{h(D)\sqrt{|D|}} \cdot \mu_3(D) \cdot \frac{6}{\pi^2} \cdot \frac{9}{8} \prod_{(\frac{D}{p})=1} \left(1 - \frac{2}{p(p+1)}\right) \prod_{p \mid D} \left(\frac{p_i - 1}{p_i} \cdot \frac{p_i^2}{p_i^2 - 1}\right) X^{1/2}$$

$$= \frac{3^{\alpha+\beta+1/2} L(1, \chi_D)}{4\pi^2 h(D)\sqrt{|D|}} \mu_3(D) \prod_{(\frac{D}{p})=1} \left(1 - \frac{2}{p(p+1)}\right) \prod_{p \mid D} \left(\frac{p_i}{p_i + 1}\right) X^{1/2},$$

with the first equality following from the fact that $\zeta(2) = \pi^2/6$. □

5. On cubic fields having a given quadratic resolvent field

Suppose K is a cubic field whose ring of integers R_K has shape $Q(x, y)$ and $\operatorname{Disc} Q = D$ is not a square. If f is an integral cubic form corresponding to the cubic ring R_K, then $K = \mathbb{Q}(\theta)$, where θ is a root of $f(x, 1)$. Since $\sqrt{\operatorname{Disc} f}$ is the product of differences of the roots of f, we see that the field $\mathbb{Q}(\sqrt{\operatorname{Disc} f})$ is contained in the Galois closure of K. Unless $D = -3$ (which is when K is Galois), this field will be quadratic. The field $\mathbb{Q}(\sqrt{\operatorname{Disc} f})$ is called the *quadratic resolvent field* of K.

Now suppose d is a fundamental discriminant. Proposition 30 and the equation

$$\operatorname{Disc} f = \frac{-Dn^2}{3}$$

(for some integer n) show that K will have $\mathbb{Q}(\sqrt{d})$ as its quadratic resolvent if and only if the shape of R_K has discriminant

$$D = \begin{cases} -3d & \text{if } d \not\equiv 0 \pmod 3, \\ -3d \text{ or } \dfrac{-d}{3} & \text{if } d \equiv 0 \pmod 3). \end{cases}$$

Define $N(d, X)$ to be the number of cubic fields with absolute discriminant bounded by X and with quadratic resolvent field $\mathbb{Q}(\sqrt{d})$. Also define $M_3^D(X)$ to be the number of cubic fields whose rings of integers have shape of discriminant D and whose discriminant is less than X. Then we have

$$N(d, X) = M_3^{-3d}(X) \tag{16}$$

if d is not a multiple of 3 and

$$N(d, X) = M_3^{-3d}(X) + M_3^{-d/3}(X) \tag{17}$$

if d is a multiple of 3. The result of Cohn proved earlier gives the asymptotics for $N(1, X)$, the number of abelian cubic extensions of $\mathbb{Q}$ of bounded discriminant. To generalize this result to general d, we require:

Proposition 32. *Let Q be a primitive integral binary quadratic form of nonsquare discriminant D. Then*

$$M_3^D(X) = \tfrac{1}{2} h(D) M_3^{\mathrm{Or}}(Q, X) + o(X^{1/2}).$$

Proof. Let $H(D)$ be the set of primitive integral binary quadratic forms of discriminant D. If $Q \in H(D)$, set $\gamma(Q) = 1$ if Q is an ambiguous form and set $\gamma(Q) = 0$ otherwise. We have

$$\begin{aligned} M_3^D(X) &= \sum_{Q\in H(D)/\mathrm{GL}_2(\mathbb{Z})} M_3(Q,X)+o(X^{1/2}) \\ &= \sum_{Q\in H(D)/\mathrm{GL}_2(\mathbb{Z})} 2^{-\gamma(Q)} M_3^{\mathrm{Or}}(Q,X)+o(X^{1/2}) \\ &= \sum_{Q\in H(D)/\mathrm{SL}_2(\mathbb{Z})} \tfrac{1}{2} M_3^{\mathrm{Or}}(Q,X)+o(X^{1/2}) = \tfrac{1}{2}h(D)M_3^{\mathrm{Or}}(Q_0,X)+o(X^{1/2}), \end{aligned}$$

where Q varies over representatives of the equivalence classes in the sums and Q_0 is an arbitrary element of $H(D)$. □

Proof of Theorem 7. If d is not a multiple of 3, then we can estimate $N(d,X)$ by combining Theorem 4 (with $D=-3d$), Proposition 32, and (16). In this case, we have $\alpha=1$ and $\mu_3(D)=\frac{22}{27}$. We need only check that the constants in the resulting main term agree with those in Theorem 7. Plugging in $\alpha=1$, $\mu_3(-3d)=\frac{22}{27}$, and $C_0=\frac{11}{9}$, we immediately see that they do.

If d is a multiple of 3, we use (17) instead of (16). The extra summand in (17) makes the calculation slightly more involved. The key is to write all of the constants in the main term of $M_3^{-3d}(X)$ as a function of the discriminant $D=-d/3$. If we define $D_1=-3d=9D$, then our table of p-adic densities of maximalities gives $\mu_3(D_1)=\frac{2}{3}$ and $\mu_3(D)=\frac{16}{27}$. Also, one uses the fact that

$$L(D_1)=\begin{cases} \frac{4}{3}L(D) & \text{if } d\equiv 3 \pmod 9, \\ \frac{4}{5}L(D) & \text{if } d\equiv 6 \pmod 9. \end{cases}$$

After grouping the common factors, one then checks that the remaining numerical constant is equal to the value of C_0 stated in Theorem 7. □

6. On cubic orders with shape of square discriminant

6.1. ***Pure cubic rings.*** We begin by describing when a cubic order has shape of square discriminant.

Lemma 33. *A cubic order has shape of square discriminant if and only if it is contained in a pure cubic field, that is, a field of the form $\mathbb{Q}(\sqrt[3]{m})$ for some $m\in\mathbb{Z}$.*

Proof. Although this is well known, we include a proof here as we could not find one in the literature. Let R be a cubic order with shape of square discriminant D. Then Disc $R=-n^2D/3$ for some integer n. Since the discriminant of an order R differs from the discriminant of the maximal order by a square factor, it suffices to prove the lemma for maximal orders $R=\mathcal{O}_K$. Note that D is a square if and only if the quadratic resolvent field of K is $F=\mathbb{Q}(\sqrt{-3})=\mathbb{Q}(\mu_3)$. If K is a pure cubic field, then certainly its quadratic resolvent field is F.

Conversely, suppose K has quadratic resolvent equal to F, and write M for the Galois closure of $K/\mathbb{Q}$. By Kummer theory, $M = F(\theta)$, where θ is a root of $x^3 - \alpha = 0$, and where $\alpha = a + b\sqrt{-3} \in \mathbb{O}_F$ is cubefree.[7] We are finished if $b = 0$ or if $N_{M/K}(\theta) \notin \mathbb{Z}$, so assume $b \neq 0$ and $N_{M/K}(\theta) = n \in \mathbb{Z}$. In this case, $N_{F/\mathbb{Q}}(\alpha) = a^2 + 3b^2 = n^3$. The six conjugates of θ are the six cube roots of $a \pm b\sqrt{-3}$. Let θ' be the conjugate of θ satisfying $\theta\theta' = n$. Then $\theta + \theta'$ is in K and has minimal polynomial $g = x^3 - 3nx - 2a$. But $\operatorname{Disc} g = 324b^2$ is a square, so K is Galois over $\mathbb{Q}$, a contradiction. □

We consider a primitive integral quadratic form $Q(x, y) = rx^2 + sxy + ty^2$ with discriminant $D = s^2 - 4rt = m^2$, a square in $\mathbb{Z}$. The goal is to estimate the number $N_3^{\mathrm{Or}}(Q, X)$ of oriented cubic orders having discriminant bounded by X and having shape $\tilde{Q}$, the $\mathrm{SL}_2(\mathbb{Z})$-equivalence class of Q. It is not difficult to show that we may take $Q(x, y)$ of the form $rx^2 + sxy$, with $0 \le r < s = \sqrt{D}$. A representative form Q of this type will be called *reduced*. The primitivity of Q implies that $(r, s) = (r, D) = 1$.

Recall that there is a bijective correspondence between isomorphism classes of oriented cubic rings having shape Q and $\mathrm{SL}_2(\mathbb{Z})$-equivalence classes of integral binary cubic forms $f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ whose Hessian $H(x, y)$ is equal to $nQ(x, y)$ for some positive integer n; that is, the cubic form f satisfies the equations

$$b^2 - 3ac = nr, \quad bc - 9ad = ns, \quad c^2 - 3bd = 0. \tag{18}$$

If $r = 0$, then the primitivity of Q forces $s = 1 = D$. In this case, $b = c = 0$, so the space of such cubic forms is precisely those of the form $f(x) = ax^3 + dy^3$, with a and d nonzero. For square $D \neq 1$ we may assume that $r \neq 0$, and one checks that all the variables involved are nonzero. Combining the equations in (18), we find that $3nrd = nsc$, and so

$$c = \frac{3rd}{s}, \quad b = \frac{3r^2d}{s^2} = \frac{cr}{s}.$$

We see that such forms are determined in a linear fashion by the outer coefficients a and d. Using the equations (18) once more, we may write n in terms of the other variables:

$$n = \frac{9d}{D^2}(r^3d - s^3a).$$

We obtain the following formula for the discriminant of the associated cubic ring:

$$\operatorname{Disc} R = \operatorname{Disc} R(f(x, y)) = -\frac{\operatorname{Disc} H(x, y)}{3} = -\frac{Dn^2}{3} = -\frac{27}{D^3}(r^3d^2 - s^3ad)^2.$$

[7] Since $\mathbb{O}_F$ is a UFD, this notion makes sense (up to roots of unity).

The cubic form

$$f = ax^3 + \frac{3r^2d}{s^2}x^2y + \frac{3rd}{s}xy^2 + dy^3$$

is integral if and only if $(a, d)^t \in \mathbb{Z}^2$ lies in the lattice

$$L(Q) = \big\{(a, d)^t \in \mathbb{Z}^2 : 3d \equiv 0 \pmod{D}\big\}.$$

Thus, the set of oriented cubic rings having shape Q is parametrized by the set $L(Q)^+ \subset L(Q)$ of elements $(a, d)^t$ such that $n = 9d(r^3d - s^3a)/D^2 > 0$, modulo $\mathrm{SO}_Q(\mathbb{Z})$-equivalence. Here, $\mathrm{SO}_Q(\mathbb{Z})$ is the subgroup of $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma \cdot Q = Q$. A simple computation shows that $|\mathrm{SO}_Q(\mathbb{Z})| = 2$ when D is a square.

Lemma 24 continues to hold when Q has square discriminant (the proof is similar), so we can again safely ignore the contribution coming from reducible cubic forms when counting cubic rings having shape Q. Thus the main term of $N_3^{\mathrm{Or}}(Q, X)$ is obtained by counting the number of lattice points $(a, d)^t \in L(Q)^+$ (modulo $\mathrm{SO}_Q(\mathbb{Z})$-equivalence) such that

$$0 < |r^3d^2 - s^3ad| < \frac{D^{3/2}X^{1/2}}{3\sqrt{3}} = N$$

and $3d \equiv 0 \pmod{D}$. For $s = 1$ (and hence $r = 0$), this problem amounts to counting lattice points under a rectangular hyperbola, for which there is Dirichlet's estimate

$$2N \log N + 2(2\gamma - 1)N + O(N^{1/2}),$$

where γ is Euler's constant. For $s > 1$, we use a similar estimate to find

$$\frac{2N}{s^3} \log N + \frac{2N}{s^3}(2\gamma - 1) + O(N^{1/2})$$

such lattice points. We then divide by the size of $\mathrm{SO}_Q(\mathbb{Z})$ and the covolume of the lattice of points $(a, d)^t \in \mathbb{Z}^2$ such that $3d \equiv 0 \pmod{D}$. If we set $\alpha = 1$ if $3 \mid D$ and $\alpha = 0$ otherwise, then this volume is $D/3^\alpha$. Putting this all together, we obtain the first part of Theorem 8.

6.2. ***Pure cubic fields.*** As we did with shapes of nonsquare discriminant, we would like to take the product of local maximality densities at each prime p to obtain a formula for the number $M_3(Q, X)$ of maximal cubic orders of shape Q and discriminant bounded by X. However, the sieve we used does not work as well in this case, because the fundamental domain for cubic forms with shape of square discriminant is not convex. Instead, we describe the lattice points that give rise to maximal orders directly.

Proposition 34. *If R is a maximal cubic ring with shape $Q(x, y)$ and $D = \mathrm{Disc}\, Q$ is a square, then either $D = 1$ or $D = 9$.*

Proof. If D is not equal to 1 or 9, then $p^2 \mid D$ for some prime $p \neq 3$ (or $p = 3$ and $p^3 \mid D$) and so any cubic form

$$f(x, y) = ax^3 + \frac{3r^2 d}{s^2} x^2 y + \frac{3rd}{s} xy^2 + dy^3$$

with shape Q has $p^2 \mid d$ because $D \mid 3d$. By Lemma 16, $R(f)$ is not maximal. □

For discriminants $D = 1$ and $D = 9$, the points $(a, d)^t \in \mathbb{Z}^2$ corresponding to maximal cubic rings have a simple description. When $D = 1$, there is just one integral binary quadratic form of discriminant D, up to $\mathrm{SL}_2(\mathbb{Z})$-equivalence, namely $Q_1(x, y) = xy$. The cubic rings of shape Q_1 have associated binary cubic forms $f(x, y) = ax^3 + dy^3$ for $a, d \in \mathbb{Z}$.

Proposition 35. *If $a, d \in \mathbb{Z}$, then the cubic form $f(x, y) = ax^3 + dy^3$ corresponds to a maximal cubic ring if and only if a and d are both squarefree*, $\gcd(a, d) = 1$, *and $a^2 \not\equiv d^2$ in $\mathbb{Z}/9\mathbb{Z}$.*

Proof. First suppose $p \neq 3$ is a prime. Since $\operatorname{Disc} f = -27a^2d^2$, if f is nonmaximal at p, then $ad \equiv 0 \pmod{p}$. By Lemma 16, f is nonmaximal at p if either $a \equiv d \equiv 0 \pmod{p}$ or if one of a or d is congruent to $0 \pmod{p^2}$. This proves the proposition away from $p = 3$.

For $p = 3$, note that $f(x, y) \equiv (ax + dy)^3 \pmod{3}$. If $a \equiv d \equiv 0 \pmod{3}$ or if a or d are divisible by 9, then $R(f)$ is not maximal. Otherwise, we move the root $ax + dy$ to x, and one checks that the coefficient of y^3 is $0 \pmod{9}$ precisely when $a \equiv \pm d \pmod{9}$, which proves the proposition. □

When $D = 9$, there are two $\mathrm{SL}_2(\mathbb{Z})$-equivalence classes of primitive integral quadratic forms whose reduced representatives are $Q_{9,r}(x, y) = rx^2 + 3xy$ for $r = 1, 2$. The cubic rings of shape $Q_{9,r}$ have associated cubic forms

$$f(x, y) = ax^3 + \frac{r^2 d}{3} x^2 y + rdxy^2 + dy^3$$

for integers a and d such that $3 \mid d$.

Proposition 36. *Let f be determined by a and d as above, and set $d' = d/3$ and $a' = r^3 d' - 9a$. Then* $\operatorname{Disc} f = 9(a'd')^2$ *and f is maximal if and only if* $\gcd(a', d') = 1$ *and both a' and d' are squarefree.*

Proof. This can be proved locally at each prime. For $p = 3$, if $R(f)$ is not maximal at 3, then $\operatorname{Disc} f \equiv 0 \pmod{9}$, which occurs if and only if $d' \equiv a' \equiv 0 \pmod{3}$. Conversely, if $d' \equiv 0 \pmod{3}$, then $R(f)$ is not maximal at 3 by Lemma 16.

Next, suppose $p \geq 5$. Again we have $d = 3d'$ for some integer d' and

$$f(x, y) = ax^3 + r^2 d' x^2 y + 3rd' xy^2 + 3d' y^3.$$

If $d' \equiv a \equiv 0 \pmod{p}$, then f is imprimitive at p; hence $R(f)$ is nonmaximal. If $a \equiv 0 \not\equiv d \pmod{p}$, then $\operatorname{Disc} f = -\frac{1}{27}(r^3d^2 - 27ad)^2 \not\equiv 0 \pmod{p}$, so $R(f)$ is maximal and $a' \equiv d' \not\equiv 0 \pmod{p}$. If $d \equiv 0 \not\equiv a \pmod{p}$, then $R(f)$ is maximal precisely when $d \not\equiv 0 \pmod{p^2}$. If both a and d are nonzero $(\text{mod } p)$, then $\operatorname{Disc} f \equiv 0 \pmod{p^2}$ if and only if $9a \equiv r^3d' \pmod{p}$. In this case,

$$f(x, y) \equiv \frac{d'}{9}(rx + 3y)^3 + kpx^3 \pmod{p^2},$$

where k is a parameter in $\mathbb{Z}$. By Lemma 16, $R(f)$ is nonmaximal precisely when $k \equiv 0 \pmod{p}$, that is, $a' \equiv 0 \pmod{p^2}$. This proves the proposition for $p \geq 5$, and also for the case $p = 2$ and $r = 1$. If $p = r = 2$, the proof from the previous paragraph does not work, but one checks easily that the proposition still holds. □

The previous propositions reduce the task of counting cubic fields with shape of square discriminant to estimating the number of lattice points with squarefree and coprime coordinates inside of a rectangular hyperbola. We can perform this computation with a suitable adaptation of Dirichlet's hyperbola method.

Proof of Theorem 8. We have already proved the first part of the theorem, so we consider the second part. It will simplify expressions if we count the number of cubic fields with bounded *conductor*, instead of bounded discriminant. Recall that if the discriminant of a cubic field $K/\mathbb{Q}$ equals df^2 with d a fundamental discriminant, then f is called the conductor of K. Thus for $D = \operatorname{Disc} Q$ equal to 1 or 9, $M_3(Q, X)$ is the number of cubic fields with shape Q and conductor bounded by $N := (X/3)^{1/2}$. First we will estimate $M_3(Q, X)$ when $D = 9$. Since $Q_{9,1}$ and $-Q_{9,2}$ are $\mathrm{GL}_2(\mathbb{Z})$ equivalent, elements of $L(Q_{9,1})^+$ correspond to cubic forms of shape $Q_{9,1}$ and elements of $L(Q_{9,1})^-$ correspond to cubic rings of shape $Q_{9,2}$. Thus we may write Q_9 for either $Q_{9,1}$ or $Q_{9,2}$ in the following. By Proposition 36, and since $\#\mathrm{GO}_{Q_{9,1}}(\mathbb{Z}) = 4$,

$$\begin{aligned} &M_3(Q_9, X) \\ &\quad = \tfrac{1}{2} \cdot \tfrac{1}{4} \cdot \#\big\{(a, b)^t \in \mathbb{Z}^2 : (a, b) = 1, \mu(a)^2 = \mu(b)^2 = 1, a \equiv b \ (\text{mod } 9), ab \leq N\big\} \\ &\quad = \tfrac{1}{2}\#\big\{(a, b)^t \in \mathbb{Z}^2_{\geq 1} : (a, b) = 1, \mu(a)^2 = \mu(b)^2 = 1, a \equiv b \ (\text{mod } 9), ab \leq N\big\}. \end{aligned}$$

We define $A_2(N)$ to be the size of this last set. Following Dirichlet's hyperbola method, we have

$$A_2(N) = 2 \sum_{a \leq \sqrt{N}} \mu(a)^2 \sum_{\substack{b \leq N/a \\ (a,b)=1 \\ a \equiv b \ (9)}} \mu^2(b) - \sum_{a \leq \sqrt{N}} \mu(a)^2 \sum_{\substack{b \leq \sqrt{N} \\ (a,b)=1 \\ a \equiv b \ (9)}} \mu^2(b)$$

$$= 2\sum_{\substack{a\le\sqrt{N}\\ 3\nmid a}} \mu(a)^2\frac{N}{a}\frac{6}{\pi^2}\frac{\phi(a)}{a}\frac{9}{8}\frac{1}{9}\prod_{p\,|\,a}\frac{p^2}{p^2-1} - \sum_{\substack{a\le\sqrt{N}\\ 3\nmid a}} \mu(a)^2\sqrt{N}\frac{6}{\pi^2}\frac{\phi(a)}{a}\frac{1}{9}\frac{9}{8} + O(\sqrt{N})$$

$$= \frac{6N}{4\pi^2}\sum_{\substack{a\le\sqrt{N}\\ 3\nmid a}} \mu(a)^2\prod_{p\,|\,a}\frac{1}{p+1} - \frac{6\sqrt{N}}{8\pi^2}\sum_{\substack{a\le\sqrt{N}\\ 3\nmid a}} \mu(a)^2\prod_{p\,|\,a}\frac{p}{p+1} + O(\sqrt{N}).$$

We use Perron's formula to estimate both of these sums. For the first sum, define the function

$$f(s) = \sum_{\substack{a\ge 1\\ 3\nmid a}} \mu(a)^2\prod_{p\,|\,a}\frac{1}{p+1} = \prod_{p\ne 3}\Bigl(1+\frac{1}{p+1}p^{-s}\Bigr),$$

which converges for $\mathrm{Re}(s) > 0$. Also define $h(s) = \dfrac{f(s)}{\zeta(s+1)}$, which converges for $\mathrm{Re}(s) > -\frac{1}{2}$. By Perron's formula,

$$\sum_{\substack{a\le\sqrt{N}\\ 3\nmid a}} \mu(a)^2\prod_{p\,|\,a}\frac{1}{p+1} = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} h(s)\zeta(s+1)\sqrt{N}^{s}s^{-1}\,ds$$

for any large c. This integral can be estimated by shifting the contour and using Cauchy's formula, which will pick up the residue of $h(s)\zeta(s+1)\sqrt{N}^{s}s^{-1}$ at $s=0$. Using Taylor series, we compute this residue to be

$$h(0)\log(\sqrt{N}) + h'(0) + \gamma h(0).$$

The same technique also works for the second sum in our formula for $A_2(N)$, and the residue turns out to be $h(0)\sqrt{N}$. Altogether, we obtain

$$A_2(N) = \frac{CN}{10}\bigl(\log N + \tfrac{2}{5}\log 3 + 2\gamma + 6\kappa - 1\bigr) + o(N),$$

since $(6/\pi^2)h'(0) = 4C/5$ and $h'(0)/h(0) = 3\kappa + \log 3/5$. Replacing N with $(X/3)^{1/2}$, we obtain the desired formula for $M_3(Q_9, X)$.

To compute $M_3(Q_1, X)$, note that by Proposition 35, we have

$$M_3(Q_1, X) = \frac{\#\mathrm{GO}_{Q_9}(\mathbb{Z})}{\#\mathrm{GO}_{Q_1}(\mathbb{Z})}\cdot\tfrac{1}{2}\bigl(A\bigl(\tfrac{N}{3}\bigr) - 2A_2\bigl(\tfrac{N}{3}\bigr)\bigr) = \tfrac{1}{2}\bigl(A\bigl(\tfrac{N}{3}\bigr) - 2A_2\bigl(\tfrac{N}{3}\bigr)\bigr),$$

where $A(N)$ has the same definition as $A_2(N)$ except without the congruence condition (mod 9). We can compute $A(N)$ exactly as before, except now the Euler factor at 3 will not be missing. Combining this with the formula for $A_2(N/3)$ above gives the estimate for $M_3(Q_1, X)$ stated in the theorem.

By Lemma 33 and Proposition 34, we may add the estimates of $M_3(Q, X)$ for $Q = Q_1$, $Q_{9,1}$, and $Q_{9,2}$ to obtain a formula for $N(-3, X)$, and this gives Theorem 8. □

Remark 37. There is an elementary way to count the density of discriminants of pure cubic fields. First, we note that two integers d, d' greater than one give rise to the same pure cubic field $K_d := \mathbb{Q}(d^{1/3})$ if and only if their quotient or product is a cube in $\mathbb{Q}$. Furthermore, if $d = ab^2$, with d cubefree and a and b squarefree, then Dedekind [1899] computed the discriminant of K_d to be $-3k^2$, where

$$k = \begin{cases} 3ab & \text{if } a^2 \not\equiv b^2 \pmod 9, \\ ab & \text{if } a^2 \equiv b^2 \pmod 9. \end{cases}$$

Thus, counting pure cubic fields of bounded discriminant is a matter of counting lattice points with squarefree and coprime coordinates under a hyperbola. So our general method of computing the number of cubic fields having a fixed quadratic resolvent field and bounded discriminant reduces to the classical method in this special case of pure cubic fields. Furthermore, we see that Dedekind's pure cubic fields of Types 1 and 2 are exactly the pure cubic fields with shape of discriminant 1 and 9, respectively.

Acknowledgments

We are very grateful to Hunter Brooks, Jordan Ellenberg, Wei Ho, Julian Rosen, Peter Sarnak, and Yongqiang Zhao for helpful conversations and comments on an earlier draft of this manuscript. Bhargava was supported by a Packard Foundation Fellowship and National Science Foundation grant DMS-1001828. Shnidman was partially supported by National Science Foundation RTG grant DMS-0943832.

References

[Bhargava 2004a] M. Bhargava, "Higher composition laws, I: A new view on Gauss composition, and quadratic generalizations", *Ann. of Math.* (2) **159**:1 (2004), 217–250. MR 2005f:11062a Zbl 1072.11078

[Bhargava 2004b] M. Bhargava, "Higher composition laws, II: On cubic analogues of Gauss composition", *Ann. of Math.* (2) **159**:2 (2004), 865–886. MR 2005f:11062b Zbl 1169.11044

[Bhargava 2004c] M. Bhargava, "Higher composition laws, III: The parametrization of quartic rings", *Ann. of Math.* (2) **159**:3 (2004), 1329–1360. MR 2005k:11214 Zbl 1169.11045

[Bhargava 2005] M. Bhargava, "The density of discriminants of quartic rings and fields", *Ann. of Math.* (2) **162**:2 (2005), 1031–1063. MR 2006m:11163 Zbl 1159.11045

[Bhargava 2008] M. Bhargava, "Higher composition laws, IV: The parametrization of quintic rings", *Ann. of Math.* (2) **167**:1 (2008), 53–94. MR 2009c:11057 Zbl 1173.11058

[Bhargava 2010] M. Bhargava, "The density of discriminants of quintic rings and fields", *Ann. of Math.* (2) **172**:3 (2010), 1559–1591. MR 2011k:11152 Zbl 1220.11139

[Bhargava and Gross 2012] M. Bhargava and B. Gross, "The average size of the 2-Selmer group of hyperelliptic curves having a rational Weierstrass point", preprint, 2012. arXiv 1206.4746

[Bhargava and Ho 2013] M. Bhargava and W. Ho, "Coregular spaces and genus one curves", preprint, 2013. arXiv 1306.4424

[Bhargava and Shankar 2010] M. Bhargava and A. Shankar, "Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves", preprint, 2010. To appear in *Ann. of Math.* arXiv 1006.1002

[Bhargava et al. 2013] M. Bhargava, A. Shankar, and J. Tsimerman, "On the Davenport–Heilbronn theorems and second order terms", *Invent. Math.* **193**:2 (2013), 439–499. MR 3090184 Zbl 06210492

[Cohen and Morra 2011] H. Cohen and A. Morra, "Counting cubic extensions with given quadratic resolvent", *J. Algebra* **325** (2011), 461–478. MR 2012b:11168 Zbl 1239.11115

[Cohn 1954] H. Cohn, "The density of abelian cubic fields", *Proc. Amer. Math. Soc.* **5** (1954), 476–477. MR 16,222a Zbl 0055.26901

[Cohn 1980] H. Cohn, *Advanced number theory*, Dover, New York, 1980. MR 82b:12001 Zbl 0474.12002

[Cremona et al. 2010] J. E. Cremona, T. A. Fisher, and M. Stoll, "Minimisation and reduction of 2-, 3- and 4-coverings of elliptic curves", *Algebra Number Theory* **4**:6 (2010), 763–820. MR 2012c:11120 Zbl 1222.11073

[Datskovsky and Wright 1988] B. Datskovsky and D. J. Wright, "Density of discriminants of cubic extensions", *J. Reine Angew. Math.* **386** (1988), 116–138. MR 90b:11112 Zbl 0632.12007

[Davenport 1951] H. Davenport, "On the class-number of binary cubic forms, I", *J. London Math. Soc.* **26** (1951), 183–192. MR 13,323e Zbl 0044.27002

[Davenport 2000] H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics **74**, Springer, New York, 2000. MR 2001f:11001 Zbl 1002.11001

[Davenport and Heilbronn 1971] H. Davenport and H. Heilbronn, "On the density of discriminants of cubic fields, II", *Proc. Roy. Soc. London Ser. A* **322**:1551 (1971), 405–420. MR 58 #10816 Zbl 0212.08101

[Dedekind 1899] R. Dedekind, "Ueber die Anzahl der Idealklassen in rein kubischen Zahlkörpern", *J. Reine Angew. Math.* **121** (1899), 40–123. Zbl 30.0198.02

[Delone and Faddeev 1964] B. N. Delone and D. K. Faddeev, *The theory of irrationalities of the third degree*, Translations of Mathematical Monographs **10**, American Mathematical Society, Providence, R.I., 1964. MR 28 #3955 Zbl 0133.30202

[Gan et al. 2002] W. T. Gan, B. Gross, and G. Savin, "Fourier coefficients of modular forms on G_2", *Duke Math. J.* **115**:1 (2002), 105–169. MR 2004a:11036 Zbl 1165.11315

[Gauss 1801] C. F. Gauss, *Disquisitiones Arithmeticae*, Fleischer, Leipzig, 1801. MR 0197380 Zbl 0585.10001

[Hardy 1915] G. H. Hardy, "On the expression of a number as the sum of two squares", *Quart. J. Math* **46** (1915), 263–283. JFM 45.1253.01

[Littelmann 1989] P. Littelmann, "Koreguläre und äquidimensionale Darstellungen", *J. Algebra* **123**:1 (1989), 193–222. MR 90e:20039 Zbl 0688.14042

[Mantilla-Soler 2010] G. Mantilla-Soler, "Integral trace forms associated to cubic extensions", *Algebra Number Theory* **4**:6 (2010), 681–699. MR 2011m:11077 Zbl 1201.11100

[Sato and Kimura 1977] M. Sato and T. Kimura, "A classification of irreducible prehomogeneous vector spaces and their relative invariants", *Nagoya Math. J.* **65** (1977), 1–155. MR 55 #3341 Zbl 0321.14030

[Schwarz 1978] G. W. Schwarz, “Representations of simple Lie groups with regular rings of invariants”, *Invent. Math.* **49**:2 (1978), 167–191. MR 80m:14032 Zbl 0391.20032

[Siegel 1944] C. L. Siegel, “The average measure of quadratic forms with given determinant and signature”, *Ann. of Math.* (2) **45** (1944), 667–685. MR 7,51a Zbl 0063.07007

[Terr 1997] D. C. Terr, *The distribution of shapes of cubic orders*, Ph.D. thesis, University of California, Berkeley, 1997, Available at http://search.proquest.com/docview/304343539. MR 2697241

[Wright and Yukie 1992] D. J. Wright and A. Yukie, “Prehomogeneous vector spaces and field extensions”, *Invent. Math.* **110**:2 (1992), 283–314. MR 93j:12004 Zbl 0803.12004

[Zhao 2013] Y. Zhao, *On sieve methods for varieties over finite fields*, Ph.D. thesis, University of Wisconsin-Madison, 2013, Available at http://gradworks.umi.com/35/93/3593347.html. Zbl 1165.11315

Communicated by Bjorn Poonen
Received 2012-06-18 Revised 2013-08-06 Accepted 2013-11-19

bhargava@math.princeton.edu *Department of Mathematics, Princeton University, Princeton, NJ 08544, United States*

shnidman@umich.edu *Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, 48109, United States*

ALGEBRA AND NUMBER THEORY 8:1 (2014)
dx.doi.org/10.2140/ant.2014.8.89

Polynomial bounds for Arakelov invariants of Belyi curves

Ariyan Javanpeykar
Appendix by Peter Bruin

We explicitly bound the Faltings height of a curve over $\bar{\mathbb{Q}}$ polynomially in its Belyi degree. Similar bounds are proven for three other Arakelov invariants: the discriminant, Faltings' delta invariant and the self-intersection of the dualising sheaf. Our results allow us to explicitly bound these Arakelov invariants for modular curves, Hurwitz curves and Fermat curves in terms of their genus. Moreover, as an application, we show that the Couveignes–Edixhoven–Bruin algorithm to compute coefficients of modular forms for congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$ runs in polynomial time under the Riemann hypothesis for ζ-functions of number fields. This was known before only for certain congruence subgroups. Finally, we use our results to prove a conjecture of Edixhoven, de Jong and Schepers on the Faltings height of a cover of $\mathbb{P}^1_{\mathbb{Z}}$ with fixed branch locus.

1. Introduction and statement of results

We prove that stable Arakelov invariants of a curve over a number field are polynomial in the Belyi degree. We apply our results to give algorithmic, geometric and Diophantine applications.

1.1. ***Bounds for Arakelov invariants of three-point covers.*** Let $\bar{\mathbb{Q}}$ be an algebraic closure of the field of rational numbers $\mathbb{Q}$. Let X be a smooth projective connected curve over $\bar{\mathbb{Q}}$ of genus g. Belyi [1979] proved that there exists a finite morphism $X \to \mathbb{P}^1_{\bar{\mathbb{Q}}}$ ramified over at most three points. Let $\deg_B(X)$ denote the Belyi degree of X, i.e., the minimal degree of a finite morphism $X \to \mathbb{P}^1_{\bar{\mathbb{Q}}}$ unramified over $\mathbb{P}^1_{\bar{\mathbb{Q}}} \setminus \{0, 1, \infty\}$. Since the topological fundamental group of the projective line $\mathbb{P}^1(\mathbb{C})$ minus three points is finitely generated, the set of $\bar{\mathbb{Q}}$-isomorphism classes of curves with bounded Belyi degree is finite.

MSC2010: primary 14G40; secondary 11G30, 11G32, 11G50, 14H55, 37P30.
Keywords: Arakelov theory, Arakelov–Green functions, Wronskian differential, Belyi degree, arithmetic surfaces, Riemann surfaces, curves, Arakelov invariants, Faltings height, discriminant, Faltings' delta invariant, self-intersection of the dualising sheaf, branched covers.

We prove that, if $g \geq 1$, the Faltings height $h_{\mathrm{Fal}}(X)$, the Faltings delta invariant $\delta_{\mathrm{Fal}}(X)$, the discriminant $\Delta(X)$ and the self-intersection of the dualising sheaf $e(X)$ are bounded by a polynomial in $\deg_B(X)$; the precise definitions of these Arakelov invariants of X are given in Section 2.3.

Theorem 1.1.1. *For any smooth projective connected curve X over $\bar{\mathbb{Q}}$ of genus $g \geq 1$,*

$$\begin{aligned} -\log(2\pi)g \leq h_{\mathrm{Fal}}(X) &\leq 13\cdot 10^6 g \deg_B(X)^5, \\ 0 \leq \quad e(X) \quad &\leq 3\cdot 10^7(g-1)\deg_B(X)^5, \\ 0 \leq \quad \Delta(X) \quad &\leq 5\cdot 10^8 g^2 \deg_B(X)^5, \\ -10^8 g^2 \deg_B(X)^5 \leq \delta_{\mathrm{Fal}}(X) &\leq 2\cdot 10^8 g \deg_B(X)^5. \end{aligned}$$

The Arakelov invariants in Theorem 1.1.1 all have a different flavour to them. For example, the Faltings height $h_{\mathrm{Fal}}(X)$ plays a key role in Faltings' proof of his finiteness theorem on abelian varieties; see [Faltings 1983]. On the other hand, the strict positivity of $e(X)$ (when $g \geq 2$) is related to the Bogomolov conjecture; see [Szpiro 1990b]. The discriminant $\Delta(X)$ "measures" the bad reduction of the curve $X/\bar{\mathbb{Q}}$ and appears in the discriminant conjecture of Szpiro [1990a] for semistable elliptic curves. Finally, as was remarked by Faltings [1984, Introduction], Faltings' delta invariant $\delta_{\mathrm{Fal}}(X)$ can be viewed as the minus logarithm of a "distance" to the boundary of the moduli space of compact connected Riemann surfaces of genus g.

We were first led to investigate this problem by work of Edixhoven, de Jong and Schepers on covers of complex algebraic surfaces with fixed branch locus; see [Edixhoven et al. 2010]. They conjectured an arithmetic analogue [Edixhoven et al. 2010, Conjecture 5.1] of their main theorem (Theorem 1.1 in [loc. cit.]). We use our results to prove this conjecture; see Section 6 for a more precise statement.

1.2. ***Outline of proof.*** To prove Theorem 1.1.1, we will use Arakelov theory for curves over a number field K. To apply Arakelov theory in this context, we will work with *arithmetic surfaces* associated to such curves, i.e., regular projective models over the ring of integers O_K of K. We refer the reader to Section 2.2 for precise definitions and basic properties of Arakelov's intersection pairing on an arithmetic surface. Then, for any smooth projective connected curve X over $\bar{\mathbb{Q}}$ of genus $g \geq 1$, we define the Faltings height $h_{\mathrm{Fal}}(X)$, the discriminant $\Delta(X)$, Faltings' delta invariant $\delta_{\mathrm{Fal}}(X)$ and the self-intersection of the dualising sheaf $e(X)$ in Section 2.3. These are the four Arakelov invariants appearing in Theorem 1.1.1.

We introduce two functions on $X(\bar{\mathbb{Q}})$ in Section 2.3: the canonical Arakelov height function and the Arakelov norm of the Wronskian differential. We show that, to prove Theorem 1.1.1, it suffices to bound the canonical height of some non-Weierstrass point and the Arakelov norm of the Wronskian differential at this point; see Theorem 2.4.1 for a precise statement.

We estimate Arakelov–Green functions and Arakelov norms of Wronskian differentials on finite étale covers of the modular curve $Y(2)$ in Theorem 3.4.5 and Proposition 3.5.1, respectively. In our proof, we use an explicit version of a result of Merkl on the Arakelov–Green function; see Theorem 3.1.2. This version of Merkl's theorem was obtained by Peter Bruin in his master's thesis. The proof of this version of Merkl's theorem is reproduced in the Appendix by Peter Bruin.

In Section 4, we prove the existence of a non-Weierstrass point on X of bounded height; see Theorem 4.5.2. The proof of Theorem 4.5.2 relies on our bounds for Arakelov–Green functions (Theorem 3.4.5), the existence of a "wild" model (Theorem 4.3.2) and a generalisation of Dedekind's discriminant conjecture for discrete valuation rings of characteristic 0 (Proposition 4.1.1), which we attribute to Lenstra.

A precise combination of the above results constitutes the proof of Theorem 1.1.1 given in Section 4.6.

1.3. ***Arakelov invariants of covers of curves with fixed branch locus.*** We apply Theorem 1.1.1 to prove explicit bounds for the height of a cover of curves. Let us be more precise.

For any finite subset $B \subset \mathbb{P}^1(\bar{\mathbb{Q}})$ and integer $d \geq 1$, the set of smooth projective connected curves X over $\bar{\mathbb{Q}}$ such that there exists a finite morphism $X \to \mathbb{P}^1_{\bar{\mathbb{Q}}}$ étale over $\mathbb{P}^1_{\bar{\mathbb{Q}}} - B$ of degree d is finite. In particular, the Faltings height of X is bounded by a real number depending only on B and d. In this section, we give an explicit version of this statement. To state our result, we need to define the height of B.

The (exponential) height $H(\alpha)$ of an element α in $\bar{\mathbb{Q}}$ is defined as $H(\alpha) = \left(\prod_v \max(1, \|\alpha\|_v)\right)^{1/[K:\mathbb{Q}]}$. Here K is a number field containing α and the product runs over the set of normalised valuations v of K. (As in [Khadjavi 2002, Section 2], we require our normalisation to be such that the product formula holds.) For any finite set $B \subset \mathbb{P}^1(\bar{\mathbb{Q}})$, define the height of B as $H_B = \max\{H(\alpha) : \alpha \in B\}$.

Theorem 1.3.1. *Let U be a nonempty open subscheme in $\mathbb{P}^1_{\bar{\mathbb{Q}}}$ with complement $B \subset \mathbb{P}^1(\bar{\mathbb{Q}})$. Let N be the number of elements in the orbit of B under the action of* $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. *Then, for any finite morphism $\pi : Y \to \mathbb{P}^1_{\bar{\mathbb{Q}}}$ étale over U, where Y is a smooth projective connected curve over $\bar{\mathbb{Q}}$ of genus $g \geq 1$,*

$$\begin{aligned}
-\log(2\pi)g \leq h_{\mathrm{Fal}}(Y) &\leq 13\cdot 10^6 g(4NH_B)^{45N^3 2^{N-2}N!}(\deg\pi)^5,\\
0 \leq \quad e(Y) \quad &\leq 3\cdot 10^7 (g-1)(4NH_B)^{45N^3 2^{N-2}N!}(\deg\pi)^5,\\
0 \leq \quad \Delta(Y) \quad &\leq 5\cdot 10^8 g^2(4NH_B)^{45N^3 2^{N-2}N!}(\deg\pi)^5,\\
-10^8 g^2(4NH_B)^{45N^3 2^{N-2}N!}&(\deg\pi)^5\\
\leq \delta_{\mathrm{Fal}}(Y) &\leq 2\cdot 10^8 g(4NH_B)^{45N^3 2^{N-2}N!}(\deg\pi)^5.
\end{aligned}$$

Theorem 1.3.1 is a consequence of Theorem 6.0.4. Note that in Theorem 6.0.4 we consider branched covers of any curve over $\overline{\mathbb{Q}}$ (i.e., not only $\mathbb{P}^1_{\overline{\mathbb{Q}}}$). We use Theorem 1.3.1 to prove [Edixhoven et al. 2010, Conjecture 5.1].

1.4. ***Diophantine application.*** Explicit bounds for Arakelov invariants of curves of genus $g \geq 2$ over a number field K and with bad reduction outside a finite set S of finite places of K imply famous conjectures in Diophantine geometry such as the *effective Mordell conjecture* and the *effective Shafarevich conjecture*; see [Rémond 1999] and [Szpiro 1985a]. We note that Theorem 1.1.1 shows that one "could" replace Arakelov invariants by the Belyi degree to prove these conjectures. We use this philosophy to deal with cyclic covers of prime degree. In fact, in [Javanpeykar and von Känel 2013], we utilise Theorem 1.1.1 and the theory of logarithmic forms to prove the small points conjecture of Szpiro [1985c, p. 284; 1986] for curves that are cyclic covers of the projective line of prime degree; see [Javanpeykar and von Känel 2013, Theorem 3.1] for a precise statement. In particular, we prove Szpiro's small points conjecture for hyperelliptic curves.

1.5. ***Modular curves, Fermat curves, Hurwitz curves and Galois Belyi curves.*** Let X be a smooth projective connected curve over $\overline{\mathbb{Q}}$ of genus $g \geq 2$. We say that X is a Fermat curve if there exists an integer $n \geq 4$ such that X is isomorphic to the planar curve $\{x^n + y^n = z^n\}$. Moreover, we say that X is a Hurwitz curve if $\# \operatorname{Aut}(X) = 84(g-1)$. Also, we say that X is a Galois Belyi curve if the quotient $X/\operatorname{Aut}(X)$ is isomorphic to $\mathbb{P}^1_{\overline{\mathbb{Q}}}$ and the morphism $X \to X/\operatorname{Aut}(X)$ is ramified over exactly three points; see [Clark and Voight 2011, Proposition 2.4] or [Wolfart 1997]. Note that Fermat curves and Hurwitz curves are Galois Belyi curves. Finally, we say that X is a modular curve if $X_{\mathbb{C}}$ is a classical congruence modular curve with respect to some (hence any) embedding $\overline{\mathbb{Q}} \to \mathbb{C}$.

If X is a Galois Belyi curve, we have $\deg_B(X) \leq 84(g-1)$. Zograf [1991] proved that, if X is a modular curve, then $\deg_B(X) \leq 128(g+1)$. Combining these bounds with Theorem 1.1.1 we obtain the following corollary:

Corollary 1.5.1. *Let X be a smooth projective connected curve over $\overline{\mathbb{Q}}$ of genus $g \geq 1$. Suppose that X is a modular curve or Galois Belyi curve. Then*

$$\max\bigl(h_{\mathrm{Fal}}(X), e(X), \Delta(X), |\delta_{\mathrm{Fal}}(X)|\bigr) \leq 2 \cdot 10^{19} g^2 (g+1)^5.$$

Remark 1.5.2. Let $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ be a finite-index subgroup, and let X be the compactification of $\Gamma\backslash\mathbb{H}$ obtained by adding the cusps, where Γ acts on the complex upper half-plane $\mathbb{H}$ via Möbius transformations. Let $X(1)$ denote the compactification of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. The inclusion $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ induces a morphism $X \to X(1)$. For $\overline{\mathbb{Q}} \subset \mathbb{C}$ an embedding, there is a unique finite morphism $Y \to \mathbb{P}^1_{\overline{\mathbb{Q}}}$ of smooth projective connected curves over $\overline{\mathbb{Q}}$ corresponding to $X \to X(1)$. The Belyi degree of Y is bounded from above by the index d of Γ in $\mathrm{SL}_2(\mathbb{Z})$. In particular,

$$\max\bigl(h_{\mathrm{Fal}}(Y), e(Y), \Delta(Y), |\delta_{\mathrm{Fal}}(Y)|\bigr) \leq 10^9 d^7.$$

Remark 1.5.3. Nonexplicit versions of Corollary 1.5.1 were previously known for certain modular curves. Firstly, polynomial bounds for Arakelov invariants of $X_0(n)$ with n squarefree were previously known; see [Ullmo 2000, Théorème 1.1 and Corollaire 1.3; Abbes and Ullmo 1997; Michel and Ullmo 1998, Théorème 1.1; Jorgenson and Kramer 2009]. The proofs of these results rely on the theory of modular curves. Also, similar results for Arakelov invariants of $X_1(n)$ with n squarefree were shown in [Edixhoven and de Jong 2011a; Mayer 2012]. Bounds for the self-intersection of the dualising sheaf of a Fermat curve of prime exponent are given in [Curilla and Kühn 2009; Kühn 2013].

1.6. ***The Couveignes–Edixhoven–Bruin algorithm.*** Corollary 1.5.1 guarantees that, under the Riemann hypothesis for ζ-functions of number fields, the Couveignes–Edixhoven–Bruin algorithm to compute coefficients of modular forms runs in polynomial time; see Theorem 5.0.1 for a more precise statement.

Conventions. By log, we mean the principal value of the natural logarithm. We define the maximum of the empty set and the product taken over the empty set as 1.

2. Arakelov geometry of curves over number fields

We are going to apply Arakelov theory to smooth projective geometrically connected curves X over number fields K. Arakelov [1974] defined an intersection theory on the *arithmetic surfaces* attached to such curves. Faltings [1984] extended Arakelov's work. In this section, we aim at giving the necessary definitions and results for what we need later (and we need at least to fix our notation).

We start with some preparations concerning Riemann surfaces and arithmetic surfaces. In Section 2.3, we define the (stable) Arakelov invariants of X appearing in Theorem 1.1.1. Finally, we prove bounds for Arakelov invariants of X in the height and the Arakelov norm of the Wronskian differential of a non-Weierstrass point; see Theorem 2.4.1.

2.1. ***Arakelov invariants of Riemann surfaces.*** Let X be a compact connected Riemann surface of genus $g \geq 1$. The space of holomorphic differentials $\mathrm{H}^0(X, \Omega^1_X)$ carries a natural hermitian inner product

$$(\omega, \eta) \mapsto \frac{i}{2}\int_X \omega \wedge \bar{\eta}.$$

For any orthonormal basis $(\omega_1, \ldots, \omega_g)$ with respect to this inner product, the Arakelov $(1, 1)$-form is the smooth positive real-valued $(1, 1)$-form μ on X given by

$\mu = (i/2g)\sum_{k=1}^{g} \omega_k \wedge \overline{\omega_k}$. Note that μ is independent of the choice of orthonormal basis. Moreover, $\int_X \mu = 1$.

Let gr_X be the Arakelov–Green function on $(X \times X) \setminus \Delta$, where $\Delta \subset X \times X$ denotes the diagonal; see [Arakelov 1974], [de Jong 2005a], [Edixhoven and de Jong 2011b] or [Faltings 1984]. The Arakelov–Green functions determine certain metrics whose curvature forms are multiples of μ, called *admissible metrics*, on all line bundles $\mathbb{O}_X(D)$, where D is a divisor on X, as well as on the holomorphic cotangent bundle Ω_X^1. Explicitly, for $D = \sum_P D_P P$ a divisor on X, the metric $\|\cdot\|$ on $\mathbb{O}_X(D)$ satisfies $\log \|1\|(Q) = \mathrm{gr}_X(D, Q)$ for all Q away from the support of D, where $\mathrm{gr}_X(D, Q) := \sum_P n_P \,\mathrm{gr}_X(P, Q)$. Furthermore, for a local coordinate z at a point a in X, the metric $\|\cdot\|_{\mathrm{Ar}}$ on the sheaf Ω_X^1 satisfies

$$-\log \|dz\|_{\mathrm{Ar}}(a) = \lim_{b \to a} (\mathrm{gr}_X(a, b) - \log |z(a) - z(b)|).$$

We will work with these metrics on $\mathbb{O}_X(P)$ and Ω_X^1 (as well as on tensor product combinations of them) and refer to them as *Arakelov metrics*. A metrised line bundle $\mathcal{L}$ is called *admissible* if, up to a constant scaling factor, it is isomorphic to one of the admissible bundles $\mathbb{O}_X(D)$. The line bundle Ω_X^1 endowed with the above metric is admissible; see [Arakelov 1974].

For any admissible line bundle $\mathcal{L}$, we endow the determinant of cohomology

$$\lambda(\mathcal{L}) = \det \mathrm{H}^0(X, \mathcal{L}) \otimes \det \mathrm{H}^1(X, \mathcal{L})^\vee$$

of the underlying line bundle with the Faltings metric; see Theorem 1 of [Faltings 1984]. We normalise this metric so that the metric on $\lambda(\Omega_X^1) = \det \mathrm{H}^0(X, \Omega_X^1)$ is induced by the hermitian inner product on $\mathrm{H}^0(X, \Omega_X^1)$ given above.

Let $\mathbb{H}_g$ be the Siegel upper half-space of complex symmetric g-by-g matrices with positive-definite imaginary part. Let τ in $\mathbb{H}_g$ be the period matrix attached to a symplectic basis of $\mathrm{H}_1(X, \mathbb{Z})$, and consider the analytic Jacobian $J_\tau(X) = \mathbb{C}^g/(\mathbb{Z}^g + \tau\mathbb{Z}^g)$ attached to τ. On $\mathbb{C}^g$, one has a theta function $\vartheta(z; \tau) = \vartheta_{0,0}(z; \tau) = \sum_{n \in \mathbb{Z}^g} \exp(\pi i\, {}^t n \tau n + 2\pi i\, {}^t n z)$, giving rise to a reduced effective divisor Θ_0 and a line bundle $\mathbb{O}(\Theta_0)$ on $J_\tau(X)$. The function ϑ is not well-defined on $J_\tau(X)$. Instead, we consider the function

$$\|\vartheta\|(z; \tau) = (\det \Im(\tau))^{1/4} \exp(-\pi\, {}^t y (\Im(\tau))^{-1} y) |\vartheta(z; \tau)| \tag{1}$$

with $y = \Im(z)$. One can check that $\|\vartheta\|$ descends to a function on $J_\tau(X)$. Now consider on the other hand the set $\mathrm{Pic}_{g-1}(X)$ of divisor classes of degree $g-1$ on X. It comes with a canonical subset Θ given by the classes of effective divisors and a canonical bijection $\mathrm{Pic}_{g-1}(X) \xrightarrow{\sim} J_\tau(X)$ mapping Θ onto Θ_0. As a result, we can equip $\mathrm{Pic}_{g-1}(X)$ with the structure of a compact complex manifold, together with a divisor Θ and a line bundle $\mathbb{O}(\Theta)$. Note that we obtain $\|\vartheta\|$ as a function on $\mathrm{Pic}_{g-1}(X)$. It can be checked that this function is independent of the choice of τ. Furthermore, note that $\|\vartheta\|$ gives a canonical way to put a metric on the line bundle $\mathcal{O}(\Theta)$ on $\mathrm{Pic}_{g-1}(X)$.

For any line bundle $\mathcal{L}$ of degree $g-1$, there is a canonical isomorphism from $\lambda(\mathcal{L})$ to $\mathcal{O}(-\Theta)[\mathcal{L}]$, the fibre of $\mathcal{O}(-\Theta)$ at the point $[\mathcal{L}]$ in $\mathrm{Pic}_{g-1}(X)$ determined by $\mathcal{L}$. Faltings [1984, Section 3] proves that, when we give both sides the metrics discussed above, the norm of this isomorphism is a constant independent of $\mathcal{L}$. We will write this norm as $\exp(\delta_{\mathrm{Fal}}(X)/8)$ and refer to $\delta_{\mathrm{Fal}}(X)$ as Faltings' delta invariant of X.

Let $S(X)$ be the invariant of X defined in [de Jong 2005a, Definition 2.2]. More explicitly, by [de Jong 2005a, Theorem 2.5],

$$\log S(X) = -\int_X \log\|\vartheta\|(gP-Q)\cdot\mu(P), \tag{2}$$

where Q is any point on X. It is related to Faltings' delta invariant $\delta_{\mathrm{Fal}}(X)$. In fact, let $(\omega_1,\dots,\omega_g)$ be an orthonormal basis of $\mathrm{H}^0(X,\Omega_X^1)$. Let b be a point on X, and let z be a local coordinate about b. Write $\omega_k = f_k\,dz$ for $k=1,\dots,g$. We have a holomorphic function

$$W_z(\omega) = \det\left(\frac{1}{(l-1)!}\frac{d^{l-1}f_k}{dz^{l-1}}\right)_{1\le k,l\le g}$$

locally about b from which we build the $g(g+1)/2$-fold holomorphic differential $W_z(\omega)(dz)^{\otimes g(g+1)/2}$. It is readily checked that this holomorphic differential is independent of the choice of local coordinate and orthonormal basis. Thus, the holomorphic differential $W_z(\omega)(dz)^{\otimes g(g+1)/2}$ extends over X to give a nonzero global section, denoted by Wr, of the line bundle $\Omega_X^{\otimes g(g+1)/2}$. The divisor of the nonzero global section Wr, denoted by $\mathcal{W}$, is the divisor of Weierstrass points. This divisor is effective of degree g^3-g. We follow [de Jong 2005a, Definition 5.3] and denote the constant norm of the canonical isomorphism of (abstract) line bundles

$$\Omega_X^{g(g+1)/2}\otimes_{\mathcal{O}_X}(\Lambda^g\mathrm{H}^0(X,\Omega_X^1)\otimes_{\mathbb{C}}\mathcal{O}_X)^\vee \to \mathcal{O}_X(\mathcal{W})$$

by $R(X)$. Then

$$\log S(X) = \tfrac{1}{8}\delta_{\mathrm{Fal}}(X) + \log R(X). \tag{3}$$

Moreover, for any non-Weierstrass point b in X,

$$\mathrm{gr}_X(\mathcal{W},b) - \log R(X) = \log\|\mathrm{Wr}\|_{\mathrm{Ar}}(b). \tag{4}$$

2.2. ***Arakelov's intersection pairing on an arithmetic surface.*** Let K be a number field with ring of integers O_K, and let $S = \mathrm{Spec}\,O_K$. Let $p:\mathcal{X}\to S$ be an arithmetic surface, i.e., an integral regular flat projective S-scheme of relative dimension 1 with geometrically connected fibres. For the sake of clarity, let us note that $p:\mathcal{X}\to S$ is a regular projective model of the generic fibre $\mathscr{X}_K \to \operatorname{Spec} K$ in the sense of [Liu 2006a, Definition 10.1.1].

In this section, we will assume the genus of the generic fibre $\mathscr{X}_K$ to be positive. An Arakelov divisor D on $\mathscr{X}$ is a divisor D_{fin} on $\mathscr{X}$ plus a contribution $D_{\text{inf}} = \sum_\sigma \alpha_\sigma F_\sigma$ running over the embeddings $\sigma : K \to \mathbb{C}$ of K into the complex numbers. Here the α_σ are real numbers and the F_σ are formally the "fibres at infinity", corresponding to the Riemann surfaces $\mathscr{X}_\sigma$ associated to the algebraic curves $\mathscr{X} \times_{O_K,\sigma} \mathbb{C}$. We let $\widehat{\operatorname{Div}}(\mathscr{X})$ denote the group of Arakelov divisors on $\mathscr{X}$. To a nonzero rational function f on $\mathscr{X}$, we associate an Arakelov divisor $\widehat{\operatorname{div}}(f) := (f)_{\text{fin}} + (f)_{\text{inf}}$ with $(f)_{\text{fin}}$ the usual divisor associated to f on $\mathscr{X}$ and $(f)_{\text{inf}} = \sum_\sigma v_\sigma(f) F_\sigma$, where $v_\sigma(f) := -\int_{\mathscr{X}_\sigma} \log |f|_\sigma \cdot \mu_\sigma$. Here μ_σ is the Arakelov $(1,1)$-form on $\mathscr{X}_\sigma$. We will say that two Arakelov divisors on $\mathscr{X}$ are linearly equivalent if their difference is of the form $\widehat{\operatorname{div}}(f)$ for some nonzero rational function f on $\mathscr{X}$. We let $\widehat{\operatorname{Cl}}(\mathscr{X})$ denote the group of Arakelov divisors modulo linear equivalence on $\mathscr{X}$.

Arakelov [1974] showed that there exists a unique symmetric bilinear map $(\cdot,\cdot) : \widehat{\operatorname{Cl}}(\mathscr{X}) \times \widehat{\operatorname{Cl}}(\mathscr{X}) \to \mathbb{R}$ with the following properties:

- If D and E are effective divisors on $\mathscr{X}$ without common component, then
$$(D, E) = (D, E)_{\text{fin}} - \sum_{\sigma : K \to \mathbb{C}} \operatorname{gr}_{\mathscr{X}_\sigma}(D_\sigma, E_\sigma),$$
where σ runs over the complex embeddings of K. Here $(D, E)_{\text{fin}}$ denotes the usual intersection number of D and E as in [Liu 2006a, Section 9.1]; i.e.,
$$(D, E)_{\text{fin}} = \sum_{s \in |S|} i_s(D, E) \log \# k(s),$$
where s runs over the set $|S|$ of closed points of S, $i_s(D, E)$ is the intersection multiplicity of D and E at s and $k(s)$ denotes the residue field of s. Note that, if D or E is vertical, the sum $\sum_{\sigma : K \to \mathbb{C}} \operatorname{gr}_{\mathscr{X}_\sigma}(D_\sigma, E_\sigma)$ is zero.
- If D is a horizontal divisor of generic degree n over S, then $(D, F_\sigma) = n$ for every $\sigma : K \to \mathbb{C}$.
- If $\sigma_1, \sigma_2 : K \to \mathbb{C}$ are complex embeddings, then $(F_{\sigma_1}, F_{\sigma_2}) = 0$.

An *admissible line bundle* on $\mathscr{X}$ is the datum of a line bundle $\mathscr{L}$ on $\mathscr{X}$ together with admissible metrics on the restrictions $\mathscr{L}_\sigma$ of $\mathscr{L}$ to the $\mathscr{X}_\sigma$. Let $\widehat{\operatorname{Pic}}(\mathscr{X})$ denote the group of isomorphism classes of admissible line bundles on $\mathscr{X}$. To any Arakelov divisor $D = D_{\text{fin}} + D_{\text{inf}}$ with $D_{\text{inf}} = \sum_\sigma \alpha_\sigma F_\sigma$, we can associate an admissible line bundle $\mathbb{O}_{\mathscr{X}}(D)$. In fact, for the underlying line bundle of $\mathbb{O}_{\mathscr{X}}(D)$, we take $\mathbb{O}_{\mathscr{X}}(D_{\text{fin}})$. Then, we make this into an admissible line bundle by equipping the pull-back of $\mathbb{O}_{\mathscr{X}}(D_{\text{fin}})$ to each $\mathscr{X}_\sigma$ with its Arakelov metric, multiplied by $\exp(-\alpha_\sigma)$. This induces an isomorphism

$$\widehat{\mathrm{Cl}}(\mathcal{X}) \xrightarrow{\sim} \widehat{\mathrm{Pic}}(\mathcal{X}).$$

In particular, the Arakelov intersection of two admissible line bundles on $\mathcal{X}$ is well-defined.

Recall that a metrised line bundle $(\mathcal{L}, \|\cdot\|)$ on $\operatorname{Spec} O_K$ corresponds to an invertible O_K-module, L, say, with hermitian metrics on the $L_\sigma := \mathbb{C} \otimes_{\sigma, O_K} L$. The *Arakelov degree* of $(\mathcal{L}, \|\cdot\|)$ is the real number defined by

$$\widehat{\deg}(\mathcal{L}) = \widehat{\deg}(\mathcal{L}, \|\cdot\|) = \log \#(L/O_K s) - \sum_{\sigma: K \to \mathbb{C}} \log \|s\|_\sigma,$$

where s is any nonzero element of L (independence of the choice of s follows from the product formula).

Note that the relative dualising sheaf $\omega_{\mathcal{X}/O_K}$ of $p : \mathcal{X} \to S$ is an admissible line bundle on $\mathcal{X}$ if we endow the restrictions $\Omega^1_{\mathcal{X}_\sigma}$ of $\omega_{\mathcal{X}/O_K}$ to the $\mathcal{X}_\sigma$ with their Arakelov metric. Furthermore, for any section $P : S \to \mathcal{X}$, we have

$$\widehat{\deg} P^* \omega_{\mathcal{X}/O_K} = (\mathbb{O}_X(P), \omega_{\mathcal{X}/O_K}) =: (P, \omega_{\mathcal{X}/O_K}),$$

where we endow the line bundle $P^*\omega_{\mathcal{X}/O_K}$ on $\operatorname{Spec} O_K$ with the pull-back metric.

Definition 2.2.1. We say that $\mathcal{X}$ is *semistable* (*or nodal*) *over* S if every geometric fibre of $\mathcal{X}$ over S is reduced and has only ordinary double singularities; see [Liu 2006a, Definition 10.3.1]. We say that $\mathcal{X}$ is (*relatively*) *minimal* if it does not contain any exceptional divisor; see [Liu 2006a, Definition 9.3.12].

Remark 2.2.2. Suppose that $\mathcal{X}$ is semistable over S and minimal. The blowing-up $\mathcal{Y} \to \mathcal{X}$ along a smooth closed point on $\mathcal{X}$ is semistable over S but no longer minimal.

2.3. ***Arakelov invariants of curves.*** Let X be a smooth projective connected curve over $\overline{\mathbb{Q}}$ of genus $g \geq 1$. Let K be a number field such that X has a semistable minimal regular model $p : \mathcal{X} \to \operatorname{Spec} O_K$; see Theorems 10.1.8, 10.3.34.a and 10.4.3 in [Liu 2006a]. (Note that we implicitly chose an embedding $K \to \overline{\mathbb{Q}}$.)

The *Faltings delta invariant* of X, denoted by $\delta_{\mathrm{Fal}}(X)$, is defined as

$$\delta_{\mathrm{Fal}}(X) = \frac{1}{[K : \mathbb{Q}]} \sum_{\sigma: K \to \mathbb{C}} \delta_{\mathrm{Fal}}(\mathcal{X}_\sigma),$$

where σ runs over the complex embeddings of K into $\mathbb{C}$. Similarly, we define

$$\|\vartheta\|_{\max}(X) = \left(\prod_{\sigma: K \to \mathbb{C}} \max_{\mathrm{Pic}_{g-1}(\mathcal{X}_\sigma)} \|\vartheta\| \right)^{1/[K:\mathbb{Q}]}.$$

Moreover, we define

$$R(X) = \Big(\prod_{\sigma : K \to \mathbb{C}} R(\mathcal{X}_\sigma) \Big)^{1/[K:\mathbb{Q}]} \quad \text{and} \quad S(X) = \Big(\prod_{\sigma : K \to \mathbb{C}} S(\mathcal{X}_\sigma) \Big)^{1/[K:\mathbb{Q}]}.$$

The *Faltings height* of X is defined by

$$h_{\text{Fal}}(X) = \frac{\widehat{\deg} \det p_* \omega_{\mathcal{X}/O_K}}{[K:\mathbb{Q}]} = \frac{\widehat{\deg} \det R^{\cdot} p_* \mathcal{O}_{\mathcal{X}}}{[K:\mathbb{Q}]},$$

where we endow the determinant of cohomology with the Faltings metric; see Section 2.1. Note that $h_{\text{Fal}}(X)$ coincides with the stable Faltings height of the Jacobian of $\mathcal{X}_K$; see [Szpiro 1985b, Chapter I, Lemma 3.2.1]. Furthermore, we define the *self-intersection of the dualising sheaf* of X, denoted by $e(X)$, as

$$e(X) := \frac{(\omega_{\mathcal{X}/O_K}, \omega_{\mathcal{X}/O_K})}{[K:\mathbb{Q}]},$$

where we use Arakelov's intersection pairing on the arithmetic surface $\mathcal{X}/O_K$. The *discriminant* of X, denoted by $\Delta(X)$, is defined as

$$\Delta(X) = \frac{\sum_{\mathfrak{p} \subset O_K} \delta_{\mathfrak{p}} \log \# k(\mathfrak{p})}{[K:\mathbb{Q}]},$$

where $\mathfrak{p}$ runs through the maximal ideals of O_K and $\delta_{\mathfrak{p}}$ denotes the number of singularities in the geometric fibre of $p : \mathcal{X} \to \operatorname{Spec} O_K$ over $\mathfrak{p}$. These invariants of X are well-defined; see [Moret-Bailly 1990, Section 5.4].

To bound the above Arakelov invariants, we introduce two functions on $X(\overline{\mathbb{Q}})$: the height and the Arakelov norm of the Wronskian differential. More precisely, let $b \in X(\overline{\mathbb{Q}})$ and suppose that b induces a section P of $\mathcal{X}$ over O_K. Then we define the *height of* b, denoted by $h(b)$, to be

$$h(b) = \frac{\widehat{\deg} P^* \omega_{\mathcal{X}/O_K}}{[K:\mathbb{Q}]} = \frac{(P, \omega_{\mathcal{X}/O_K})}{[K:\mathbb{Q}]}.$$

Note that the height of b is the stable canonical height of a point, in the Arakelov-theoretic sense, with respect to the admissible line bundle $\omega_{\mathcal{X}/O_K}$. We define the Arakelov norm of the Wronskian differential at b as

$$\|\mathrm{Wr}\|_{\mathrm{Ar}}(b) = \Big(\prod_{\sigma : K \to \mathbb{C}} \|\mathrm{Wr}\|_{\mathrm{Ar}}(b_\sigma) \Big)^{1/[K:\mathbb{Q}]}.$$

These functions on $X(\overline{\mathbb{Q}})$ are well-defined; see [Moret-Bailly 1990, Section 5.4].

Changing the model for X might change the height of a point. Let us show that the height of a point does not become smaller if we take another regular model over O_K.

Lemma 2.3.1. *Let $\mathcal{Y} \to \operatorname{Spec} O_K$ be an arithmetic surface. Assume that $\mathcal{Y}$ is a model for $\mathcal{X}_K$. If Q denotes the section of $\mathcal{Y}$ over O_K induced by $b \in X(\overline{\mathbb{Q}})$, then*

$$h(b) \leq \frac{(Q, \omega_{\mathcal{Y}/O_K})}{[K:\mathbb{Q}]}.$$

Proof. By the minimality of $\mathcal{X}$, there is a unique birational morphism $\phi : \mathcal{Y} \to \mathcal{X}$; see [Liu 2006a, Corollary 9.3.24]. By the factorisation theorem, this morphism is made up of a finite sequence

$$\mathcal{Y} = \mathcal{Y}_n \xrightarrow{\phi_n} \mathcal{Y}_{n-1} \xrightarrow{\phi_{n-1}} \cdots \xrightarrow{\phi_1} \mathcal{Y}_0 = \mathcal{X}$$

of blowing-ups along closed points; see [Liu 2006a, Theorem 9.2.2]. For $i = 1, \ldots, n$, let $E_i \subset \mathcal{Y}_i$ denote the exceptional divisor of ϕ_i. Since the line bundles $\omega_{\mathcal{Y}_i/O_K}$ and $\phi_i^* \omega_{\mathcal{Y}_{i-1}/O_K}$ agree on $\mathcal{Y}_i - E_i$, there is an integer a such that

$$\omega_{\mathcal{Y}_i/O_K} = \phi_i^* \omega_{\mathcal{Y}_{i-1}/O_K} \otimes_{\mathcal{O}_{\mathcal{Y}_i}} \mathcal{O}_{\mathcal{Y}_i}(aE_i).$$

Applying the adjunction formula, we see that $a = 1$. Since ϕ_i restricts to the identity morphism on the generic fibre, we have a canonical isomorphism of admissible line bundles

$$\omega_{\mathcal{Y}_i/O_K} = \phi_i^* \omega_{\mathcal{Y}_{i-1}/O_K} \otimes_{\mathcal{O}_{\mathcal{Y}_i}} \mathcal{O}_{\mathcal{Y}_i}(E_i).$$

Let Q_i denote the section of $\mathcal{Y}_i$ over O_K induced by $b \in X(\overline{\mathbb{Q}})$. Then

$$\begin{aligned}(Q_i, \omega_{\mathcal{Y}_i/O_K}) &= (Q_i, \phi_i^* \omega_{\mathcal{Y}_{i-1}/O_K}) + (Q_i, E_i)\\ &\geq (Q_i, \phi_i^* \omega_{\mathcal{Y}_{i-1}/O_K})\\ &= (Q_{i-1}, \omega_{\mathcal{Y}_{i-1}/O_K}),\end{aligned}$$

where we used the projection formula in the last equality. Therefore, we conclude that

$$(Q, \omega_{\mathcal{Y}/O_K}) = (Q_n, \omega_{\mathcal{Y}_n/O_K}) \geq (Q_0, \omega_{\mathcal{Y}_0/O_K}) = (P, \omega_{\mathcal{X}/O_K}) = h(b)[K:\mathbb{Q}]. \quad \square$$

2.4. ***Bounding Arakelov invariants in the height of a non-Weierstrass point.*** In this section, we prove bounds for Arakelov invariants of curves in the height of a non-Weierstrass point and the Arakelov norm of the Wronskian differential in this point.

Theorem 2.4.1. *Let X be a smooth projective connected curve over $\overline{\mathbb{Q}}$ of genus $g \geq 1$. Let $b \in X(\overline{\mathbb{Q}})$. Then*

$$\begin{aligned} e(X) &\leq 4g(g-1)h(b),\\ \delta_{\mathrm{Fal}}(X) &\geq -90g^3 - 4g(2g-1)(g+1)h(b).\end{aligned}$$

Suppose that b is not a Weierstrass point. Then

$$\begin{aligned} h_{\mathrm{Fal}}(X) &\leq \tfrac{1}{2}g(g+1)h(b) + \log \|\mathrm{Wr}\|_{\mathrm{Ar}}(b), \\ \delta_{\mathrm{Fal}}(X) &\leq 6g(g+1)h(b) + 12 \log\|\mathrm{Wr}\|_{\mathrm{Ar}}(b) + 4g \log(2\pi), \\ \Delta(X) &\leq 2g(g+1)(4g+1)h(b) + 12 \log \|\mathrm{Wr}\|_{\mathrm{Ar}}(b) + 93g^3. \end{aligned}$$

This theorem is essential to the proof of Theorem 1.1.1 given in Section 4.5. We give a proof of Theorem 2.4.1 at the end of this section.

Lemma 2.4.2. *For a smooth projective connected curve X over $\overline{\mathbb{Q}}$ of genus $g \geq 1$,*

$$\log \|\vartheta\|_{\max}(X) \leq \frac{g}{4} \log \max(1, h_{\mathrm{Fal}}(X)) + (4g^3 + 5g + 1) \log 2.$$

Proof. We kindly thank R. de Jong for sharing this proof with us. We follow the idea of [Graftieaux 2001, Section 2.3.2]; see also [David 1991, Appendice]. Let $\mathcal{F}_g$ be the Siegel fundamental domain of dimension g in the Siegel upper half-space $\mathbb{H}_g$, i.e., the space of complex $(g \times g)$-matrices τ in $\mathbb{H}_g$ such that the following properties are satisfied. Firstly, for every element u_{ij} of $u = \Re(\tau)$, we have $|u_{ij}| \leq \frac{1}{2}$. Secondly, for every γ in $\mathrm{Sp}(2g, \mathbb{Z})$, we have $\det \Im(\gamma \cdot \tau) \leq \det \Im(\tau)$, and finally, $\Im(\tau)$ is Minkowski-reduced; i.e., for all $\xi = (\xi_1, \dots, \xi_g) \in \mathbb{Z}^g$ and for all i such that $\xi_i, \dots, \xi_g$ are nonzero, we have $\xi \Im(\tau)^t \xi \geq (\Im(\tau))_{ii}$ and, for all $1 \leq i \leq g-1$ we have $(\Im(\tau))_{i,i+1} \geq 0$. One can show that $\mathcal{F}_g$ contains a representative of each $\mathrm{Sp}(2g, \mathbb{Z})$-orbit in $\mathbb{H}_g$.

Let K be a number field such that X has a model X_K over K. For any embedding $\sigma : K \to \mathbb{C}$, let τ_σ be an element of $\mathcal{F}_g$ such that $\mathrm{Jac}(X_{K,\sigma}) \cong \mathbb{C}^g/(\tau_\sigma \mathbb{Z}^g + \mathbb{Z}^g)$ as principally polarised abelian varieties, the matrix of the Riemann form induced by the polarisation of $\mathrm{Jac}(X_{K,\sigma})$ being $\Im(\tau_\sigma)^{-1}$ on the canonical basis of $\mathbb{C}^g$. By a result of Bost (see [Graftieaux 2001, Lemme 2.12] or [Pazuki 2012]), we have

$$\frac{1}{[K:\mathbb{Q}]} \sum_{\sigma : K \to \mathbb{C}} \log \det(\Im(\tau_\sigma)) \leq g \log \max(1, h_{\mathrm{Fal}}(X)) + (2g^3 + 2) \log(2). \quad (5)$$

Here we used that the Faltings height of X equals the Faltings height of its Jacobian. Now, let $\vartheta(z; \tau)$ be the Riemann theta function as in Section 2.1, where τ is in $\mathcal{F}_g$ and $z = x + iy$ is in $\mathbb{C}^g$ with $x, y \in \mathbb{R}^g$. Combining (5) with the upper bound

$$\exp(-\pi^t y(\Im(\tau))^{-1} y)|\vartheta(z; \tau)| \leq 2^{3g^3+5g} \quad (6)$$

implies the result. Let us prove (6). Note that, if we write $y = \Im(z) = (\Im(\tau)) \cdot b$ for b in $\mathbb{R}^g$,

$$\exp(-\pi^t g(\Im(\tau))^{-1} y)|\vartheta(z; \tau)| \leq \sum_{n \in \mathbb{Z}^g} \exp(-\pi^t (n+b)(\Im(\tau))(n+b)).$$

Since $\Im(\tau)$ is Minkowski reduced, we have ${}^t m\Im(\tau)m \geq c(g)\sum_{i=1}^{g} m_i^2(\Im(\tau))_{ii}$ for all m in $\mathbb{R}^g$. Here $c(g) = (4/g^3)^{g-1}\left(\frac{3}{4}\right)^{g(g-1)/2}$. Also, $(\Im(\tau))_{ii} \geq \sqrt{3}/2$ for all $i = 1, \ldots, g$ (see [Igusa 1972, Chapter V.4] for these facts). We deduce that

$$\begin{aligned}
&\sum_{n\in\mathbb{Z}^g} \exp(-\pi^t(n+b)(\Im(\tau))(n+b)) \\
&\qquad \leq \sum_{n\in\mathbb{Z}^g} \exp\biggl(-\sum_{i=1}^{g} \pi c(g)(n_i+b_i)^2(\Im(\tau))_{ii}\biggr) \\
&\qquad \leq \prod_{i=1}^{g} \sum_{n_i\in\mathbb{Z}} \exp(-\pi c(g)(n_i+b_i)^2(\Im(\tau))_{ii}) \\
&\qquad \leq \prod_{i=1}^{g} \frac{2}{1-\exp(-\pi c(g)(\Im(\tau))_{ii})} \leq 2^g\biggl(1+\frac{2}{\pi\sqrt{3}c(g)}\biggr)^g.
\end{aligned}$$

This proves (6). □

Lemma 2.4.3. *Let $a \in \mathbb{R}_{>0}$ and $b \in \mathbb{R}_{\leq 1}$. Then, for all real numbers $x \geq b$,*

$$x - a\log\max(1,x) = \tfrac{1}{2}x + \tfrac{1}{2}(x - 2a\log\max(1,x)) \geq \tfrac{1}{2}x + \min\bigl(\tfrac{1}{2}b,\, a - a\log(2a)\bigr).$$

Proof. It suffices to prove that $x - 2a\log\max(1,x) \geq \min(b, 2a - 2a\log(2a))$ for all $x \geq b$. To prove this, let $x \geq b$. Then, if $2a \leq 1$, we have $x - 2a\log\max(1,x) \geq b \geq \min(b, 2a - 2a\log(2a))$. (To prove that $x - 2a\log\max(1,x) \geq b$, we may assume that $x \geq 1$. It is easy to show that $x - 2a\log x$ is a nondecreasing function for $x \geq 1$. Therefore, for all $x \geq 1$, we conclude that $x - 2a\log x \geq 1 \geq b$.) If $2a > 1$, the function $x - 2a\log(x)$ attains its minimum value at $x = 2a$ on the interval $[1, \infty)$. □

Lemma 2.4.4 (Bost). *Let X be a smooth projective connected curve over $\overline{\mathbb{Q}}$ of genus $g \geq 1$. Then*

$$h_{\mathrm{Fal}}(X) \geq -\log(2\pi)g.$$

Proof. See [Gaudron and Rémond 2011, Corollaire 8.4]. (Note that the Faltings height $h(X)$ utilised by Bost, Gaudron and Rémond is bigger than $h_{\mathrm{Fal}}(X)$ due to a difference in normalisation. In fact, we have $h(X) = h_{\mathrm{Fal}}(X) + g\log(\sqrt{\pi})$. In particular, the slightly stronger lower bound $h_{\mathrm{Fal}}(X) \geq -\log(\sqrt{2}\pi)g$ holds.) □

Lemma 2.4.5. *Let X be a smooth projective connected curve over $\overline{\mathbb{Q}}$ of genus $g \geq 1$. Then*

$$\begin{aligned}
&\log S(X) + h_{\mathrm{Fal}}(X) \\
&\qquad \geq \tfrac{1}{2}h_{\mathrm{Fal}}(X) - (4g^3+5g+1)\log 2 + \min\Bigl(-\frac{g}{2}\log(2\pi),\, \frac{g}{4} - \frac{g}{4}\log\Bigl(\frac{g}{2}\Bigr)\Bigr).
\end{aligned}$$

Proof. By the explicit formula (2) for $S(X)$ in Section 2.1 and our bounds on theta functions (Lemma 2.4.2),

$$\log S(X) + h_{\mathrm{Fal}}(X) \geq -\frac{g}{4} \log \max(1, h_{\mathrm{Fal}}(X)) - (4g^3 + 5g + 1) \log 2 + h_{\mathrm{Fal}}(X).$$

Since $h_{\mathrm{Fal}}(X) \geq -g \log(2\pi)$, the statement follows from Lemma 2.4.3 (with $x = h_{\mathrm{Fal}}(X)$, $a = g/4$ and $b = -g \log(2\pi)$). □

Lemma 2.4.6. *Let X be a smooth projective connected curve of genus $g \geq 2$ over $\overline{\mathbb{Q}}$. Then*

$$\frac{(2g-1)(g+1)}{8(g-1)} e(X) + \tfrac{1}{8}\delta_{\mathrm{Fal}}(X) \geq \log S(X) + h_{\mathrm{Fal}}(X).$$

Proof. By [de Jong 2005a, Proposition 5.6],

$$e(X) \geq \frac{8(g-1)}{(g+1)(2g-1)} (\log R(X) + h_{\mathrm{Fal}}(X)).$$

Note that $\log R(X) = \log S(X) - \delta_{\mathrm{Fal}}(X)/8$; see (3) in Section 2.1. This implies the inequality. □

Lemma 2.4.7 (Noether formula). *Let X be a smooth projective connected curve over $\overline{\mathbb{Q}}$ of genus $g \geq 1$. Then*

$$12 h_{\mathrm{Fal}}(X) = e(X) + \Delta(X) + \delta_{\mathrm{Fal}}(X) - 4g \log(2\pi).$$

Proof. This is well-known; see [Faltings 1984, Theorem 6; Moret-Bailly 1989, Théorème 2.2]. □

Proposition 2.4.8. *Let X be a smooth projective connected curve of genus $g \geq 2$ over $\overline{\mathbb{Q}}$. Then*

$$\begin{aligned} h_{\mathrm{Fal}}(X) &\leq \frac{(2g-1)(g+1)}{4(g-1)} e(X) + \tfrac{1}{4}\delta_{\mathrm{Fal}}(X) + 20g^3, \\ -g \log(2\pi) &\leq \frac{(2g-1)(g+1)}{4(g-1)} e(X) + \tfrac{1}{4}\delta_{\mathrm{Fal}}(X) + 20g^3, \\ \Delta(X) &\leq \frac{3(2g-1)(g+1)}{g-1} e(X) + 2\delta_{\mathrm{Fal}}(X) + 248g^3. \end{aligned}$$

Proof. Firstly, by Lemma 2.4.6,

$$\frac{(2g-1)(g+1)}{8(g-1)} e(X) + \tfrac{1}{8}\delta_{\mathrm{Fal}}(X) \geq \log S(X) + h_{\mathrm{Fal}}(X).$$

To obtain the upper bound for $h_{\mathrm{Fal}}(X)$, we proceed as follows. By Lemma 2.4.5,

$$\begin{aligned} &\log S(X) + h_{\mathrm{Fal}}(X) \\ &\qquad \geq \tfrac{1}{2} h_{\mathrm{Fal}}(X) - (4g^3 + 5g + 1) \log 2 + \min\Bigl(-\frac{g}{2} \log(2\pi), \frac{g}{4} - \frac{g}{4} \log\Bigl(\frac{g}{2}\Bigr)\Bigr). \end{aligned}$$

From these two inequalities, we deduce that

$$\tfrac{1}{2}h_{\mathrm{Fal}}(X) \le \frac{(2g-1)(g+1)}{8(g-1)}e(X) + \tfrac{1}{8}\delta_{\mathrm{Fal}}(X) + (4g^3+5g+1)\log 2 \\ + \max\Big(\frac{g}{2}\log(2\pi),\, \frac{g}{4}\log\Big(\frac{g}{2}\Big) - \frac{g}{4}\Big).$$

Finally, it is straightforward to verify the inequality

$$(4g^3+5g+1)\log 2 + \max\Big(\frac{g}{2}\log(2\pi),\, \frac{g}{4}\log\Big(\frac{g}{2}\Big) - \frac{g}{4}\Big) \le 10g^3.$$

This concludes the proof of the upper bound for $h_{\mathrm{Fal}}(X)$.

The second inequality follows from the first inequality of the proposition and the lower bound $h_{\mathrm{Fal}}(X) \ge -g\log(2\pi)$ of Bost (Lemma 2.4.4).

Finally, to obtain the upper bound of the proposition for the discriminant of X, we eliminate the Faltings height of X in the first inequality using the Noether formula and obtain

$$\Delta(X) + e(X) + \delta_{\mathrm{Fal}}(X) - 4g\log(2\pi) \le \frac{3(2g-1)(g+1)}{(g-1)}e(X) + 3\delta_{\mathrm{Fal}}(X) + 240g^3.$$

Faltings [1984, Theorem 5] showed that $e(X) \ge 0$. Therefore, we conclude that

$$\Delta(X) + \delta_{\mathrm{Fal}}(X) - 4g\log(2\pi) \le \frac{3(2g-1)(g+1)}{(g-1)}e(X) + 3\delta_{\mathrm{Fal}}(X) + 240g^3. \quad \square$$

We are now ready to prove Theorem 2.4.1.

Proof of Theorem 2.4.1. The proof is straightforward. The upper bound $e(X) \le 4g(g-1)h(b)$ is well-known; see [Faltings 1984, Theorem 5].

Let us prove the lower bound for $\delta_{\mathrm{Fal}}(X)$. If $g \ge 2$, the lower bound for $\delta_{\mathrm{Fal}}(X)$ can be deduced from the second inequality of Proposition 2.4.8 and the upper bound $e(X) \le 4g(g-1)h(b)$. When $g = 1$, this follows from a result of Szpiro [de Jong 2005b, Proposition 7.2] and the nonnegativity of $h(b)$.

From now on, we suppose that b is a non-Weierstrass point. The upper bound $h_{\mathrm{Fal}}(X) \le \frac{1}{2}g(g+1)h(b) + \log\|\mathrm{Wr}\|_{\mathrm{Ar}}(b)$ follows from Theorem 5.9 in [de Jong 2005a] and (4) in Section 2.1.

We deduce the upper bound $\delta_{\mathrm{Fal}}(X) \le 6g(g+1)h(b) + 12\log\|\mathrm{Wr}\|_{\mathrm{Ar}}(b) + 4g\log(2\pi)$ as follows. Since $e(X) \ge 0$ and $\Delta(X) \ge 0$, the Noether formula implies that

$$\delta_{\mathrm{Fal}}(X) \le 12h_{\mathrm{Fal}}(X) + 4g\log(2\pi).$$

Thus, the upper bound for $\delta_{\mathrm{Fal}}(X)$ follows from the upper bound for $h_{\mathrm{Fal}}(X)$.

The upper bound $\Delta(X) \le 2g(g+1)(4g+1)h(b) + 12\log\|\mathrm{Wr}\|_{\mathrm{Ar}}(b) + 93g^3$ follows from the inequality $\Delta(X) \le 12h_{\mathrm{Fal}}(X) - \delta_{\mathrm{Fal}}(X) + 4g\log(2\pi)$ and the preceding bounds. (One could also use the last inequality of Proposition 2.4.8 to obtain a similar result.) □

3. Bounds for Arakelov–Green functions of Belyi covers

Our aim is to give explicit bounds for the Arakelov–Green function on a Belyi cover of $X(2)$. Such bounds have been obtained for certain Belyi covers using spectral methods in [Jorgenson and Kramer 2006]. The results in [loc. cit.] do not apply to our situation since the smallest positive eigenvalue of the Laplacian can go to zero in a tower of Belyi covers; see [Long 2008, Theorem 4].

Instead, we use a theorem of Merkl to prove explicit bounds for the Arakelov–Green function on a Belyi cover in Theorem 3.4.5. More precisely, we construct a "Merkl atlas" for an arbitrary Belyi cover. Our construction uses an explicit version of [Jorgenson and Kramer 2004] on the Arakelov (1, 1)-form due to Bruin.

We use our results to estimate the Arakelov norm of the Wronskian differential in Proposition 3.5.1.

Merkl's theorem [2011, Theorem 10.1] was used to prove bounds for Arakelov–Green functions of the modular curve $X_1(5p)$ in [Edixhoven and de Jong 2011a].

3.1. ***Merkl's theorem.*** Let X be a compact connected Riemann surface of positive genus, and recall that μ denotes the Arakelov (1, 1)-form on X.

Definition 3.1.1. A *Merkl atlas* for X is a quadruple

$$\big(\{(U_j, z_j)\}_{j=1}^n, r_1, M, c_1\big),$$

where $\{(U_j, z_j)\}_{j=1}^n$ is a finite atlas for X and $\frac{1}{2} < r_1 < 1$, $M \geq 1$ and $c_1 > 0$ are real numbers such that the following properties are satisfied:

(1) Each $z_j U_j$ is the open unit disc.

(2) The open sets $U_j^{r_1} := \{x \in U_j : |z_j(x)| < r_1\}$ with $1 \leq j \leq n$ cover X.

(3) For all $1 \leq j, j' \leq n$, the function $|dz_j/dz_{j'}|$ on $U_j \cap U_{j'}$ is bounded from above by M.

(4) For $1 \leq j \leq n$, write $\mu_{\mathrm{Ar}} = i F_j dz_j \wedge d\overline{z_j}$ on U_j. Then $0 \leq F_j(x) \leq c_1$ for all $x \in U_j$.

Given a Merkl atlas $(\{(U_j, z_j)\}_{j=1}^n, r_1, M, c_1)$ for X, the following result provides explicit bounds for Arakelov–Green functions in n, r_1, M and c_1:

Theorem 3.1.2 (Merkl). *Let $(\{(U_j, z_j)\}_{j=1}^n, r_1, M, c_1)$ be a Merkl atlas for X. Then*

$$\sup_{(X\times X)\setminus\Delta} \mathrm{gr}_X \leq \frac{330n}{(1-r_1)^{3/2}} \log \frac{1}{1-r_1} + 13.2nc_1 + (n-1)\log M.$$

Furthermore, for every index j and all $x \neq y \in U_j^{r_1}$, we have

$$\left|\mathrm{gr}_X(x, y) - \log|z_j(x) - z_j(y)|\right| \leq \frac{330n}{(1-r_1)^{3/2}} \log \frac{1}{1-r_1} + 13.2nc_1 + (n-1)\log M.$$

Proof. Merkl [2011] proved this theorem without explicit constants and without the dependence on r_1. A proof of the theorem in a more explicit form was given by P. Bruin in his master's thesis. This proof is reproduced, with minor modifications, in the Appendix. □

3.2. ***An atlas for a Belyi cover of $X(2)$.*** Let $\mathbb{H}$ denote the complex upper half-plane. Recall that $\mathrm{SL}_2(\mathbb{R})$ acts on $\mathbb{H}$ via Möbius transformations. Let $\Gamma(2)$ denote the subgroup of $\mathrm{SL}_2(\mathbb{Z})$ defined as

$$\Gamma(2) = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) : a \equiv d \equiv 1 \bmod 2 \text{ and } b \equiv c \equiv 0 \bmod 2\right\}.$$

The Riemann surface $Y(2) = \Gamma(2)\backslash\mathbb{H}$ is not compact. Let $X(2)$ be the compactification of the Riemann surface $Y(2) = \Gamma(2)\backslash\mathbb{H}$ obtained by adding the cusps 0, 1 and ∞. Note that $X(2)$ is known as the *compact modular curve associated to the congruence subgroup* $\Gamma(2)$ *of* $\mathrm{SL}_2(\mathbb{Z})$. The modular lambda function $\lambda : \mathbb{H} \to \mathbb{C}$ induces an analytic isomorphism $\lambda : X(2) \to \mathbb{P}^1(\mathbb{C})$; see Section 4.4 for details. In particular, the genus of $X(2)$ is zero. For a cusp $\kappa \in \{0, 1, \infty\}$, we fix an element γ_κ in $\mathrm{SL}_2(\mathbb{Z})$ such that $\gamma_\kappa(\kappa) = \infty$.

We construct an atlas for the compact connected Riemann surface $X(2)$. Let $\dot{B}_\infty$ be the open subset given by the image of the strip

$$\dot{S}_\infty := \left\{x + iy : -1 \leq x < 1,\ y > \tfrac{1}{2}\right\} \subset \mathbb{H}$$

in $Y(2)$ under the quotient map $\mathbb{H} \to \Gamma(2)\backslash\mathbb{H}$ defined by $\tau \mapsto \Gamma(2)\tau$. The quotient map $\mathbb{H} \to \Gamma(2)\backslash\mathbb{H}$ induces a bijection from this strip to $\dot{B}_\infty$. More precisely, suppose that τ and τ' in $\dot{S}_\infty$ lie in the same orbit under the action of $\Gamma(2)$. Then, there exists an element

$$\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma(2)$$

such that $\gamma\tau = \tau'$. If $c \neq 0$, by definition, c is a nonzero integral multiple of 2. Thus, $c^2 \geq 4$. Therefore,

$$\frac{1}{2} < \Im\tau' = \frac{\Im\tau}{|c\tau + d|^2} \leq \frac{1}{4\Im\tau} < \frac{1}{2}.$$

This is clearly impossible. Thus, $c = 0$ and $\tau' = \tau \pm b$. By definition, $b = 2k$ for some integer k. Since τ and τ' lie in the above strip, we conclude that $b = 0$. Thus, $\tau = \tau'$.

Consider the morphism $z_\infty : \mathbb{H} \to \mathbb{C}$ given by $\tau \mapsto \exp(\pi i\tau + \pi/2)$. The image of the strip $\dot{S}_\infty$ under z_∞ in $\mathbb{C}$ is the punctured open unit disc $\dot{B}(0, 1)$. Now, for any τ

and τ' in the strip $\dot{S}_\infty$, the equality $z_\infty(\tau) = z_\infty(\tau')$ holds if and only if $\tau' = \tau \pm 2k$ for some integer k. But then $k = 0$ and $\tau = \tau'$. We conclude that z_∞ factors injectively through $\dot{B}_\infty$. Let $z_\infty : B_\infty \to B(0, 1)$ denote, by abuse of notation, the induced chart at ∞, where $B_\infty := \dot{B}_\infty \cup \{\infty\}$ and $B(0, 1)$ is the open unit disc in $\mathbb{C}$. We translate our neighbourhood B_∞ at ∞ to a neighbourhood for κ, where κ is a cusp of $X(2)$. More precisely, for any τ in $\mathbb{H}$, define $z_\kappa(\tau) = \exp(\pi i \gamma_k^{-1}\tau + \pi/2)$. Let $\dot{B}_\kappa$ be the image of $\dot{S}_\infty$ under the map $\mathbb{H} \to Y(2)$ given by $\tau \mapsto \Gamma(2)\gamma_\kappa\tau$. We define $B_\kappa = \dot{B}_\kappa \cup \{\kappa\}$. We let $z_\kappa : B_\kappa \to B(0, 1)$ denote the induced chart (by abuse of notation).

Since the open subsets B_κ cover $X(2)$, we have constructed an atlas $\{(B_\kappa, z_\kappa)\}_\kappa$ for $X(2)$, where κ runs through the cusps 0, 1 and ∞.

Definition 3.2.1. A *Belyi cover* of $X(2)$ is a morphism of compact connected Riemann surfaces $Y \to X(2)$ that is unramified over $Y(2)$. The points of Y not lying over $Y(2)$ are called *cusps*.

Lemma 3.2.2. *Let $\pi : Y \to X(2)$ be a Belyi cover with Y of genus g. Then, $g \leq \deg \pi$.*

Proof. This is trivial for $g \leq 1$. For $g \geq 2$, the statement follows from the Riemann–Hurwitz formula. □

Let $\pi : Y \to X(2)$ be a Belyi cover. We are going to "lift" the atlas $\{(B_\kappa, z_\kappa)\}$ for $X(2)$ to an atlas for Y.

Let κ be a cusp of $X(2)$. The branched cover $\pi^{-1}(B_\kappa) \to B_\kappa$ restricts to a finite degree topological cover $\pi^{-1}(\dot{B}_\kappa) \to \dot{B}_\kappa$. In particular, the composed morphism

$$\pi^{-1}\dot{B}_\kappa \to \dot{B}_\kappa \xrightarrow[z_\kappa|_{\dot{B}_\kappa}]{\sim} \dot{B}(0, 1)$$

is a finite degree topological cover of $\dot{B}(0, 1)$.

Recall that the fundamental group of $\dot{B}(0, 1)$ is isomorphic to $\mathbb{Z}$. More precisely, for any connected topological cover of $V \to \dot{B}(0, 1)$ of finite degree, there is a unique integer $e \geq 1$ such that $V \to \dot{B}(0, 1)$ is isomorphic to the cover $\dot{B}(0, 1) \to \dot{B}(0, 1)$ given by $x \mapsto x^e$.

For every cusp y of Y lying over κ, let $\dot{V}_y$ be the unique connected component of $\pi^{-1}\dot{B}_\kappa$ whose closure V_y in $\pi^{-1}(B_\kappa)$ contains y. Then, for any cusp y, there is a positive integer e_y and an isomorphism $w_y : \dot{V}_y \xrightarrow{\sim} \dot{B}(0, 1)$ such that $w_y^{e_y} = z_\kappa \circ \pi|_{\dot{V}_y}$. The isomorphism $w_y : \dot{V}_y \to \dot{B}(0, 1)$ extends to an isomorphism $w_y : V_y \to B(0, 1)$ such that $w_y^{e_y} = z_\kappa \circ \pi|_{V_y}$. This shows that e_y is the ramification index of y over κ. Note that we have constructed an atlas $\{(V_y, w_y)\}$ for Y, where y runs over the cusps of Y.

3.3. ***The Arakelov* (1, 1)*-form and the hyperbolic metric.*** Let

$$\mu_{\mathrm{hyp}}(\tau)=\frac{i}{2}\frac{1}{\Im(\tau)^2}\,d\tau\,d\bar{\tau}$$

be the hyperbolic metric on $\mathbb{H}$. A Fuchsian group is a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$. For any Fuchsian group Γ, the quotient space $\Gamma\backslash\mathbb{H}$ is a connected Hausdorff topological space and can be made into a Riemann surface in a natural way. The hyperbolic metric μ_{hyp} on $\mathbb{H}$ induces a measure on $\Gamma\backslash\mathbb{H}$, given by a smooth positive real-valued $(1,1)$-form outside the set of fixed points of elliptic elements of Γ. If the volume of $\Gamma\backslash\mathbb{H}$ with respect to this measure is finite, we call Γ a *cofinite Fuchsian group*.

Let Γ be a cofinite Fuchsian group, and let X be the compactification of $\Gamma\backslash\mathbb{H}$ obtained by adding the cusps. We assume that Γ has no elliptic elements and that the genus g of X is positive. There is a unique smooth function $F_\Gamma : X\to[0,\infty)$ that vanishes at the cusps of Γ such that

$$\mu=\frac{1}{g}F_\Gamma\,\mu_{\mathrm{hyp}}. \tag{7}$$

A detailed description of F_Γ is not necessary for our purposes.

Definition 3.3.1. Let $\pi : Y\to X(2)$ be a Belyi cover. Then we define the cofinite Fuchsian group Γ_Y (or simply Γ) associated to $\pi : Y\to X(2)$ as follows. Since the topological fundamental group of $Y(2)$ equals $\Gamma(2)/\{\pm1\}$, we have $\pi^{-1}(Y(2))=\Gamma'\backslash\mathbb{H}$ for some subgroup $\Gamma'\subset\Gamma(2)/\{\pm1\}$ of finite index. We define $\Gamma\subset\Gamma(2)$ to be the inverse image of Γ' under the quotient map $\Gamma(2)\to\Gamma(2)/\{\pm1\}$. Note that Γ is a cofinite Fuchsian group without elliptic elements.

Theorem 3.3.2 (Jorgenson and Kramer). *For any Belyi cover $\pi : Y\to X(2)$, where Y has positive genus,*

$$\sup_{\tau\in Y}F_\Gamma\le 64\max_{y\in Y}(e_y)^2\le 64(\deg\pi)^2.$$

Proof. This is shown in [Bruin 2013]. More precisely, in the notation of [loc. cit.], Bruin shows that, with $a=1.44$, we have $N_{\mathrm{SL}_2(\mathbb{Z})}(z,2a^2-1)\le 58$. In particular, $\sup_{z\in Y}N_\Gamma(z,z,2a^2-1)\le 58$; see Section 8.2 in [loc. cit.]. Now, we apply Proposition 6.1 and Lemma 6.2 (with $\epsilon=2\deg\pi$) in [loc. cit.] to deduce the sought inequality. □

Remark 3.3.3. Jorgenson and Kramer [2004] prove a stronger (albeit nonexplicit) version of Theorem 3.3.2.

3.4. ***A Merkl atlas for a Belyi cover of* $X(2)$.** In this section, we prove bounds for Arakelov–Green functions of Belyi covers.

Recall that we constructed an atlas $\{(B_\kappa, z_\kappa)\}_\kappa$ for $X(2)$. For a cusp κ of $X(2)$, let

$$y_\kappa : \mathbb{H} \to (0, \infty)$$

be defined by $\tau \mapsto \Im(\gamma_\kappa^{-1}\tau) = \frac{1}{2} - \frac{\log|z_\kappa(\tau)|}{\pi}$. This induces a function $\dot{B}_\kappa \to (0, \infty)$, also denoted by y_κ.

Lemma 3.4.1. *For any two cusps κ and κ' of $X(2)$, we have*

$$\left|\frac{dz_\kappa}{dz_{\kappa'}}\right| \leq 4\exp(3\pi/2)$$

on $B_\kappa \cap B_{\kappa'}$.

Proof. We work on the complex upper half-plane $\mathbb{H}$. We may and do assume that $\kappa \neq \kappa'$. By applying $\gamma_{\kappa'}^{-1}$, we may and do assume that $\kappa' = \infty$. On $B_\kappa \cap B_\infty$, we have

$$dz_\kappa(\tau) = \pi i \exp(\pi i \gamma_\kappa^{-1}\tau + \pi/2) d(\gamma_\kappa^{-1}\tau) \quad \text{and} \quad dz_\infty(\tau) = \pi i \exp(\pi i \tau + \pi/2) d(\tau).$$

Therefore,

$$\frac{dz_\kappa}{dz_\infty}(\tau) = \exp(\pi i(\gamma_\kappa^{-1}\tau - \tau))\frac{d(\gamma_\kappa^{-1}\tau)}{d(\tau)}.$$

It follows from a simple calculation that, for $\gamma_\kappa^{-1} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $c \neq 0$,

$$\left|\frac{dz_\kappa}{dz_\infty}\right|(\tau) = \frac{1}{|c\tau + d|^2}\exp(\pi(y_\infty(\tau) - y_\kappa(\tau))).$$

For τ and $\gamma_\kappa^{-1}\tau$ in B_∞, one has $y_\infty(\tau) > \frac{1}{2}$ and $y_\kappa(\tau) > \frac{1}{2}$. From $|c\tau + d| \geq y_\infty(\tau) = \Im(\tau)$, it follows that

$$y_\kappa(\tau) = \Im(\gamma_\kappa^{-1}(\tau)) = y_\infty\Big(\frac{a\tau + b}{c\tau + d}\Big) = \frac{\Im\tau}{|c\tau + d|^2} \leq \frac{\Im\tau}{(\Im\tau)^2} \leq 2,$$

and similarly, $y_\infty(\tau) \leq 2$. The statement follows. □

Let $\pi : Y \to X(2)$ be a Belyi cover, and let $V = \pi^{-1}(Y(2))$ be the complement of the set of cusps in Y. Recall that we constructed an atlas $\{(V_y, w_y)\}$ for Y. We assume that the genus g of Y is positive, and as usual, we let μ denote the Arakelov $(1, 1)$-form on Y.

Lemma 3.4.2. *Let y be a cusp of $\pi : Y \to X(2)$ with $\kappa = \pi(y)$. Then*

$$i\, dw_y\, d\overline{w}_y = \frac{2\pi^2 y_\kappa^2 |w_y|^2}{e_y^2}\mu_{\text{hyp}} \quad \text{on } \dot{V}_y.$$

Proof. Let $\kappa = \pi(y)$ in $X(2)$. We work on the complex upper half-plane. By the chain rule, we have

$$d(z_\kappa) = d(w_y^{e_y}) = e_y w_y^{e_y - 1}\, dw_y.$$

Therefore,

$$e_y^2|w_y|^{2e_y-2}\,dw_y\,d\overline{w}_y = dz_\kappa\,d\bar{z}_\kappa.$$

Note that $dz_\kappa = \pi i z_\kappa\, d(\gamma_\kappa^{-1})$, where we view $\gamma_\kappa^{-1} : \mathbb{H} \to \mathbb{C}$ as a function. Therefore,

$$e_y^2|w_y|^{2e_y-2}\,dw_y\,d\overline{w}_y = \pi^2|z_\kappa|^2\,d(\gamma_\kappa^{-1})\,d(\overline{\gamma_\kappa^{-1}}).$$

Since $|w_y^{e_y}| = |z_\kappa|$, we have

$$\begin{aligned} i\,dw_y\,d\overline{w}_y &= \frac{i\pi^2|w_y|^2}{e_y^2}\,d(\gamma_\kappa^{-1})\,d(\overline{\gamma_\kappa^{-1}}) \\ &= \frac{2\pi^2y_\kappa^2|w_y|^2}{e_y^2}\frac{i\,d(\gamma_\kappa^{-1})\,d(\overline{\gamma_\kappa^{-1}})}{2y_\kappa^2} = \frac{2\pi^2y_\kappa^2|w_y|^2}{e_y^2}(\mu_{\mathrm{hyp}}\circ\gamma_\kappa^{-1}). \end{aligned}$$

Since μ_{hyp} is invariant under the action of $\mathrm{SL}_2(\mathbb{Z})$, this concludes the proof. □

Proposition 3.4.3. *Let y be a cusp of $\pi : Y \to X(2)$. Write $\mu = iF_y\,dw_y\,d\overline{w}_y$ on V_y. Then F_y is a subharmonic function on V_y and*

$$0 \le F_y \le \frac{128\exp(3\pi)(\deg\pi)^4}{\pi^2 g}.$$

Proof. The first statement follows from [Jorgenson and Kramer 2004, p. 8]; see also [Bruin 2010, p. 58]. The lower bound for F_y is clear from the definition. Let us prove the upper bound for F_y.

For a cusp κ of $X(2)$, let $\dot{B}_\kappa(2) \subset \dot{B}_\kappa$ be the image of the strip $\{x+iy : -1 \le x < 1,\ y > 2\}$ in $Y(2)$ under the map $\mathbb{H} \to Y(2)$ given by $\tau \mapsto \Gamma(2)\gamma_\kappa\tau$. For a cusp y of Y lying over κ, define $\dot{V}_y(2) = \pi^{-1}(\dot{B}_\kappa(2))$ and $V_y(2) = \dot{V}_y(2) \cup \{y\}$. Since the boundary $\partial V_y(2)$ of $V_y(2)$ is contained in $V_y - V_y(2)$, by the maximum principle for subharmonic functions,

$$\sup_{V_y} F_y = \max\Bigl(\sup_{V_y(2)} F_y,\ \sup_{V_y - V_y(2)} F_y\Bigr) = \max\Bigl(\sup_{\partial V_y(2)} F_y,\ \sup_{V_y - V_y(2)} F_y\Bigr) = \sup_{V_y - V_y(2)} F_y.$$

By Lemma 3.4.2, Definition 3.3.1 and (7) in Section 3.3,

$$F_y = F_\Gamma \frac{e_y^2}{2g\pi^2y_\kappa^2|w_y|^2}. \tag{8}$$

Note that $y_\kappa^{-2} < 4$ on V_y. Furthermore,

$$\sup_{V_y - V_y(2)} |w_y|^{-2} \le \sup_{B_\kappa - B_\kappa(2)} |z_\kappa|^{-2} = \exp(-\pi) \sup_{B_\kappa - B_\kappa(2)} \exp(2\pi y_\kappa) \le \exp(3\pi).$$

Thus, the proposition follows from Jorgenson–Kramer's upper bound for F_Γ (Theorem 3.3.2). □

Definition 3.4.4. Define $s_1 = \sqrt{1/2}$. Note that $\frac{1}{2} < s_1 < 1$. For any cusp κ of $X(2)$, let $B_\kappa^{s_1}$ be the open subset of B_κ whose image under z_κ is $\{x \in \mathbb{C} : |x| < s_1\}$. Moreover, define the positive real number r_1 by the equation $r_1^{\deg \pi} = s_1$. Note that $\frac{1}{2} < r_1 < 1$. For all cusps y of $\pi : Y \to X(2)$, define the subset $V_y^{r_1} \subset V_y$ by $V_y^{r_1} = \{x \in V_y : |w_y(x)| < r_1\}$.

Theorem 3.4.5. *Let $\pi : Y \to X(2)$ be a Belyi cover such that Y is of genus $g \geq 1$. Then*

$$\sup_{(Y\times Y)\setminus\Delta} \mathrm{gr}_Y \leq 6378027\frac{(\deg \pi)^5}{g}.$$

Moreover, for every cusp y and all $x \neq x'$ in $V_y^{r_1}$,

$$\left|\mathrm{gr}_Y(x,x') - \log|w_y(x) - w_y(x')|\right| \leq 6378027\frac{(\deg \pi)^5}{g}.$$

Proof. Write $d = \deg \pi$. Let s_1 and r_1 be as in Definition 3.4.4. We define real numbers

$$n := \#(Y - V), \quad M := 4d\exp(3\pi) \quad \text{and} \quad c_1 := \frac{128\exp(3\pi)d^4}{\pi^2 g}.$$

Since n is the number of cusps of Y, we have $n \leq 3d$. Moreover,

$$\frac{1}{1-r_1} \leq \frac{d}{1-s_1}.$$

Note that

$$\frac{330n}{(1-r_1)^{3/2}}\log\frac{1}{1-r_1} + 13.2nc_1 + (n-1)\log M \leq 6378027\frac{d^5}{g}.$$

Therefore, by Theorem 3.1.2, it suffices to show that

$$(\{(V_y, w_y)\}_y, r_1, M, c_1),$$

where y runs over the cusps of $\pi : Y \to X(2)$, constitutes a Merkl atlas for Y.

The first condition of Merkl's theorem is satisfied. That is, $w_y V_y$ is the open unit disc in $\mathbb{C}$.

To verify the second condition of Merkl's theorem, we have to show that the open sets $V_y^{r_1}$ cover Y. For any $x \in V_y$, we have $x \in V_y^{r_1}$ if $\pi(x) \in B_\kappa^{s_1}$. In fact, for any x in V_y, we have $|w_y(x)| < r_1$ if and only if

$$|z_\kappa(\pi(x))| = |w_y(x)|^{e_y} < r_1^{e_y}.$$

Since $r_1 < 1$, we see that $s_1 = r_1^d \leq r_1^{e_y}$. Therefore, if $\pi(x)$ lies in $B_\kappa^{s_1}$, we see that x lies in $V_y^{r_1}$. Now, since $s_1 < \sqrt{3}/2$, we have $X(2) = \bigcup_{\kappa\in\{0,1,\infty\}} B_\kappa^{s_1}$. We conclude that $Y = \bigcup_y V_y^{r_1}$, where y runs through the cusps.

Since we have already verified the fourth condition of Merkl's theorem in Proposition 3.4.3, it suffices to verify the third condition to finish the proof. Let κ and κ' be cusps of $X(2)$. We may and do assume that $\kappa \neq \kappa'$. Now, as usual, we work on the complex upper half-plane. By the chain rule,

$$\left|\frac{dw_y}{dw_{y'}}\right| \leq \frac{d}{|w_y|^{e_y-1}} \sup_{B_\kappa \cap B_{\kappa'}} \left|\frac{dz_\kappa}{dz_{\kappa'}}\right|$$

on $V_y \cap V_{y'}$. Note that $|w_y(\tau)|^{e_y-1} \geq |w_y(\tau)|^{e_y} = |z_\kappa(\tau)|$ for any τ in $\mathbb{H}$. Therefore,

$$\left|\frac{dw_y}{dw_{y'}}\right| \leq \frac{d}{|z_\kappa|} \sup_{B_\kappa \cap B_{\kappa'}} \left|\frac{dz_\kappa}{dz_{\kappa'}}\right| \leq M,$$

where we used Lemma 3.4.1 and the inequality $|z_\kappa| > \exp(-3\pi/2)$ on $B_\kappa \cap B_{\kappa'}$. □

3.5. *The Arakelov norm of the Wronskian differential.*

Proposition 3.5.1. *Let $\pi : Y \to X(2)$ be a Belyi cover with Y of genus $g \geq 1$. Then*

$$\sup_{Y - \operatorname{Supp} \mathcal{W}} \log \|\mathrm{Wr}\|_{\mathrm{Ar}} \leq 6378028 g (\deg \pi)^5.$$

Proof. Let b be a non-Weierstrass point on Y, and let y be a cusp of Y such that b lies in $V_y^{r_1}$. Let $\omega = (\omega_1, \ldots, \omega_g)$ be an orthonormal basis of $\mathrm{H}^0(Y, \Omega^1_Y)$. Then, as in Section 2.1,

$$\log \|\mathrm{Wr}\|_{\mathrm{Ar}}(b) = \log |W_{w_y}(\omega)(b)| + \frac{g(g+1)}{2} \log \|dw_y\|_{\mathrm{Ar}}(b).$$

By Theorem 3.4.5,

$$\frac{g(g+1)}{2} \log \|dw_y\|_{\mathrm{Ar}}(b) \leq 6378027 g (\deg \pi)^5.$$

Let us show that $\log |W_{w_y}(\omega)(b)| \leq g(\deg \pi)^5$. Write $\omega_k = f_k \, dw_y$ on V_y. Note that $\omega_k \wedge \overline{\omega_k} = |f_k|^2 \, dw_y \wedge d\overline{w}_y$. Therefore,

$$\mu = \frac{i}{2g} \sum_{k=1}^{g} \omega_k \wedge \overline{\omega_k} = \frac{i}{2g} \sum_{k=1}^{g} |f_k|^2 \, dw_y \wedge d\overline{w}_y.$$

We deduce that $\sum_{k=1}^{g} |f_k|^2 = 2g F_y$, where F_y is the unique function on V_y such that $\mu = i F_y \, dw_y \wedge d\overline{w}_y$. By our upper bound for F_y (Proposition 3.4.3), for any $j = 1, \ldots, g$,

$$\sup_{V_y} |f_j|^2 \leq \sup_{V_y} \sum_{k=1}^{g} |f_k|^2 = 2g F_y \leq \frac{256 \exp(3\pi)(\deg \pi)^4}{\pi^2}.$$

By Hadamard's inequality,

$$\log|W_{w_y}(\omega)(b)| \le \sum_{l=0}^{g-1} \log\Bigl(\sum_{k=1}^{g}\Bigl|\frac{d^l f_k}{dw_y^l}\Bigr|^2(b)\Bigr)^{1/2}.$$

Let $r_1 < r < 1$ be some real number. By Cauchy's integral formula, for any $0 \le l \le g-1$,

$$\begin{aligned}\Bigl|\frac{d^l f_k}{dw_y^l}\Bigr|(b) &= \Bigl|\frac{l!}{2\pi i}\int_{|w_y|=r}\frac{f_k}{(w_y - w_y(b))^{l+1}}\,dw_y\Bigr|\\ &\le \frac{l!}{(r-r_1)^{l+1}}\sup_{V_y}|f_k| \le \frac{g!}{(1-r_1)^g}\sup_{V_y}|f_k|.\end{aligned}$$

By the preceding estimations, since $g! \le g^g$ and $1/(1-r_1) \le \deg\pi/(1-s_1)$, we obtain that

$$\begin{aligned}&\log|W_{w_y}(\omega)(b)|\\ &\quad\le \sum_{l=0}^{g-1}\log\Bigl(\frac{g!}{(1-r_1)^g}\Bigl(\sum_{k=1}^{g}\sup_{V_y}|f_k|^2\Bigr)^{1/2}\Bigr)\\ &\quad\le \sum_{l=0}^{g-1}\log\Bigl(\frac{g!}{(1-r_1)^g}\Bigl(\sum_{k=1}^{g}\frac{256\exp(3\pi)(\deg\pi)^4}{\pi^2}\Bigr)^{1/2}\Bigr)\\ &\quad= g\log(g!) + g^2\log\Bigl(\frac{1}{1-r_1}\Bigr) + \frac{g}{2}\log\Bigl(\frac{256g\exp(3\pi)}{\pi^2}\Bigr) + 2g\log(\deg\pi)\\ &\quad\le \Bigl(4.5 + \log\Bigl(\frac{1}{1-s_1}\Bigr) + \tfrac{1}{2}\log\Bigl(\frac{256\exp(3\pi)}{\pi^2}\Bigr)\Bigr)g^2\log(\deg\pi)\\ &\quad\le 13g(\deg\pi)^2.\end{aligned}$$

Since $g \ge 1$ and $\pi : Y \to X(2)$ is a Belyi cover, the inequality $\deg\pi \ge 3$ holds. Thus,

$$13g(\deg\pi)^2 \le \frac{13g(\deg\pi)^5}{27} \le g(\deg\pi)^5. \qquad \square$$

4. Points of bounded height

4.1. ***Lenstra's generalisation of Dedekind's discriminant bound.*** Let A be a discrete valuation ring of characteristic 0 with fraction field K. Let ord_A denote the valuation on A. Let L/K be a finite field extension of degree n, and let B be the integral closure of A in L. Note that L/K is separable, and B/A is finite; see [Serre 1979, Proposition I.4.8].

The inverse different $\mathfrak{D}_{B/A}^{-1}$ of B over A is the fractional ideal

$$\{x \in L : \mathrm{Tr}(xB) \subset A\},$$

where Tr is the trace of L over K. The inverse of the inverse different, denoted by $\mathfrak{D}_{B/A}$, is the different of B over A. Note that $\mathfrak{D}_{B/A}$ is actually an integral ideal of L.

The following proposition (which we would like to attribute to H. W. Lenstra, Jr.) is a generalisation of Dedekind's discriminant bound; see [Serre 1979, Proposition III.6.13].

Proposition 4.1.1 (H. W. Lenstra, Jr.). *Suppose that B is a discrete valuation ring of ramification index e over A. Then, the valuation r of the different ideal $\mathfrak{D}_{B/A}$ on B satisfies the inequality*

$$r \leq e - 1 + e \cdot \operatorname{ord}_A(n).$$

Proof. Let x be a uniformiser of A. Since A is of characteristic 0, we may define $y := 1/nx$; note that y is an element of K. The trace of y (as an element of L) is $1/x$. Since $1/x$ is not in A, this implies that the inverse different $\mathfrak{D}_{B/A}^{-1}$ is strictly contained in the fractional ideal yB. (If not, since A and B are discrete valuation rings, we would have that yB is strictly contained in the inverse different.) In particular, the different $\mathfrak{D}_{B/A}$ strictly contains the fractional ideal (nx). Therefore, the valuation $\operatorname{ord}_B(\mathfrak{D}_{B/A})$ on B of $\mathfrak{D}_{B/A}$ is strictly less than the valuation of nx. Thus,

$$\operatorname{ord}_B(\mathfrak{D}_{B/A}) < \operatorname{ord}_B(nx) = e \cdot \operatorname{ord}_A(nx) = e(\operatorname{ord}_A(n) + 1) = e \cdot \operatorname{ord}_A(n) + e.$$

This concludes the proof of the inequality. □

Remark 4.1.2. If the extension of residue fields of B/A is separable, Proposition 4.1.1 follows from the remark following Proposition III.6.13 in [Serre 1979]. (The result in that proposition was conjectured by Dedekind and proved by Hensel when $A = \mathbb{Z}$.) The reader will see that, in the proof of Proposition 4.2.4, we have to deal with imperfect residue fields.

Proposition 4.1.3. *Suppose that the residue characteristic p of A is positive. Let m be the biggest integer such that $p^m \leq n$. Then, for $\beta \subset B$ a maximal ideal of B with ramification index e_β over A, the valuation r_β of the different ideal $\mathfrak{D}_{B/A}$ at β satisfies the inequality*

$$r_\beta \leq e_\beta - 1 + e_\beta \cdot \operatorname{ord}_A(p^m).$$

Proof. To compute r_β, we localise B at β and then take the completions $\widehat{A}$ and $\widehat{B_\beta}$ of A and B_β, respectively. Let d be the degree of $\widehat{B_\beta}$ over $\widehat{A}$. Then, by Lenstra's result (Proposition 4.1.1), the inequality

$$r_\beta \leq e_\beta - 1 + e_\beta \cdot \operatorname{ord}_{\widehat{A}}(d)$$

holds. By definition, $\operatorname{ord}_{\widehat{A}}(d) = \operatorname{ord}_A(d) \leq \operatorname{ord}_A(p^m)$. This concludes the proof. □

4.2. ***Covers of arithmetic surfaces with fixed branch locus.*** Let K be a number field with ring of integers O_K, and let $S = \operatorname{Spec} O_K$. Let D be a reduced effective divisor on $\mathscr{X} = \mathbb{P}^1_S$, and let U denote the complement of the support of D in $\mathscr{X}$.

Let $\mathcal{Y} \to S$ be an integral normal two-dimensional flat projective S-scheme with geometrically connected fibres, and let $\pi : \mathcal{Y} \to \mathcal{X}$ be a finite surjective morphism of S-schemes that is étale over U. Let $\psi : \mathcal{Y}' \to \mathcal{Y}$ be the minimal resolution of singularities [Liu 2006a, Proposition 9.3.32]. Note that we have the following diagram of morphisms:

$$\mathcal{Y}' \xrightarrow{\psi} \mathcal{Y} \xrightarrow{\pi} \mathcal{X} \to S.$$

Consider the prime decomposition $D = \sum_{i \in I} D_i$, where I is a finite-index set. Let D_{ij} be an irreducible component of $\pi^{-1}(D)$ mapping onto D_i, where j is in the index set J_i. We define r_{ij} to be the valuation of the different ideal of $\mathcal{O}_{\mathcal{Y},D_{ij}}/\mathcal{O}_{\mathcal{X},D_i}$. We define the ramification divisor R to be $\sum_{i \in I} \sum_{j \in J_i} r_{ij} D_{ij}$. We define $B := \pi_* R$.

We apply [Liu 2006a, 6.4.26] to obtain that there exists a dualising sheaf $\omega_{\mathcal{Y}/S}$ for $\mathcal{Y} \to S$ and a dualising sheaf ω_π for $\pi : \mathcal{Y} \to \mathcal{X}$ such that the adjunction formula

$$\omega_{\mathcal{Y}/S} = \pi^* \omega_{\mathcal{X}/S} \otimes \omega_\pi$$

holds. Since the local ring at the generic point of a divisor on $\mathcal{X}$ is of characteristic 0, basic properties of the different ideal imply that ω_π is canonically isomorphic to the line bundle $\mathcal{O}_{\mathcal{Y}}(R)$. We deduce the *Riemann–Hurwitz* formula

$$\omega_{\mathcal{Y}/S} = \pi^* \omega_{\mathcal{X}/S} \otimes \mathcal{O}_{\mathcal{Y}}(R).$$

Let $K_{\mathcal{X}} = -2 \cdot [\infty]$ be the divisor defined by the tautological section of $\omega_{\mathcal{X}/O_K}$. Let $K_{\mathcal{Y}'}$ denote the Cartier divisor on $\mathcal{Y}'$ defined by the rational section $d(\pi \circ \psi)$ of $\omega_{\mathcal{Y}'/S}$. We define the Cartier divisor $K_{\mathcal{Y}}$ on $\mathcal{Y}$ analogously; i.e., $K_{\mathcal{Y}}$ is the Cartier divisor on $\mathcal{Y}$ defined by $d\pi$. Note that $K_{\mathcal{Y}} = \psi_* K_{\mathcal{Y}'}$. Also, the Riemann–Hurwitz formula implies the following equality of Cartier divisors:

$$K_{\mathcal{Y}} = \pi^* K_{\mathcal{X}} + R.$$

Let $E_1, \dots, E_s$ be the exceptional components of $\psi : \mathcal{Y}' \to \mathcal{Y}$. Note that the pull-back of the Cartier divisor $\psi^* K_{\mathcal{Y}}$ coincides with $K_{\mathcal{Y}'}$ on

$$\mathcal{Y}' - \bigcup_{i=1}^{s} E_i.$$

Therefore, there exist integers c_i such that

$$K_{\mathcal{Y}'} = \psi^* K_{\mathcal{Y}} + \sum_{i=1}^{s} c_i E_i,$$

where this is an equality of Cartier divisors (*not only* modulo linear equivalence). Note that $(\psi^* K_{\mathcal{Y}}, E_i) = 0$ for all i. In fact, $K_{\mathcal{Y}}$ is linearly equivalent to a Cartier divisor with support disjoint from the singular locus of $\mathcal{Y}$.

Lemma 4.2.1. *For all $i = 1, \dots, s$, we have $c_i \leq 0$.*

Proof. We have the following local statement. Let y be a singular point of $\mathcal{Y}$, and let $E_1, \dots, E_r$ be the exceptional components of ψ lying over y. We define

$$V_+ = \sum_{\substack{i=1 \\ c_i > 0}}^{r} c_i E_i$$

as the sum on the $c_i > 0$. To prove the lemma, it suffices to show that $V_+ = 0$. Since the intersection form on the exceptional locus of $\mathcal{Y}' \to \mathcal{Y}$ is negative-definite [Liu 2006a, Proposition 9.1.27], to prove $V_+ = 0$, it suffices to show that $(V_+, V_+) \geq 0$. Clearly, to prove the latter inequality, it suffices to show that, for all i such that $c_i > 0$, we have $(V_+, E_i) \geq 0$. To do this, fix $i \in \{1, \dots, r\}$ with $c_i > 0$. Since $\mathcal{Y}' \to \mathcal{Y}$ is minimal, we have that E_i is not a (-1)-curve. In particular, by the adjunction formula, the inequality $(K_{\mathcal{Y}'}, E_i) \geq 0$ holds. We conclude that

$$(V_+, E_i) = (K_{\mathcal{Y}'}, E_i) - \sum_{\substack{j=1 \\ c_j < 0}}^{r} c_j (E_j, E_i) \geq 0,$$

where, in the last inequality, we used that, for all j such that $c_j < 0$, we have that $E_j \neq E_i$. □

Proposition 4.2.2. *Let $P' : S \to \mathcal{Y}'$ be a section, and let $Q : S \to \mathcal{X}$ be the induced section. If the image of P' is not contained in the support of $K_{\mathcal{Y}'}$, then*

$$(K_{\mathcal{Y}'}, P')_{\text{fin}} \leq (B, Q)_{\text{fin}}.$$

Proof. By the Riemann–Hurwitz formula, we have $K_{\mathcal{Y}} = \pi^* K_{\mathcal{X}} + R$. Therefore, by Lemma 4.2.1, we get that

$$\begin{aligned}
(K_{\mathcal{Y}'}, P')_{\text{fin}} &= (\psi^* K_{\mathcal{Y}} + \textstyle\sum c_i E_i, P')_{\text{fin}} \\
&= \left(\psi^* \pi^* K_{\mathcal{X}} + \psi^* R + \sum_{i=1}^{s} c_i E_i, P' \right)_{\text{fin}} \\
&\leq (\psi^* \pi^* K_{\mathcal{X}}, P')_{\text{fin}} + (\psi^* R, P')_{\text{fin}}.
\end{aligned}$$

Since the image of P' is not contained in the support of $K_{\mathcal{Y}'}$, we can apply the projection formula for the composed morphism $\pi \circ \psi : \mathcal{Y}' \to \mathcal{X}$ to $(\psi^* \pi^* K_{\mathcal{X}}, P')_{\text{fin}}$ and $(\psi^* R, P')_{\text{fin}}$; see [Liu 2006a, Section 9.2]. This gives

$$(K_{\mathcal{Y}'}, P')_{\text{fin}} \leq (\psi^* \pi^* K_{\mathcal{X}}, P')_{\text{fin}} + (\psi^* R, P')_{\text{fin}} = (K_{\mathcal{X}}, Q)_{\text{fin}} + (\pi_* R, Q)_{\text{fin}}.$$

Since $K_{\mathcal{X}} = -2 \cdot [\infty]$, the inequality $(K_{\mathcal{X}}, Q)_{\text{fin}} \leq 0$ holds. By definition, $B = \pi_* R$. This concludes the proof. □

We introduce some notation. For i in I and j in J_i, let e_{ij} and f_{ij} be the ramification index and residue degree of π at the generic point of D_{ij}, respectively. Moreover, let $\mathfrak{p}_i \subset O_K$ be the maximal ideal corresponding to the image of D_i in $\operatorname{Spec} O_K$. Then, note that e_{ij} is the multiplicity of D_{ij} in the fibre of $\mathcal{Y}$ over $\mathfrak{p}_i$. Now, let $e_{\mathfrak{p}_i}$ and $f_{\mathfrak{p}_i}$ be the ramification index and residue degree of $\mathfrak{p}_i$ over $\mathbb{Z}$, respectively. Finally, let p_i be the residue characteristic of the local ring at the generic point of D_i and, if $p_i > 0$, let m_i be the biggest integer such that $p_i^{m_i} \leq \deg \pi$, i.e., $m_i = \lfloor \log(\deg \pi)/\log(p_i) \rfloor$.

Lemma 4.2.3. *Let i be in I such that $0 < p_i \leq \deg \pi$. Then, for all j in J_i,*

$$r_{ij} \leq 2e_{ij} m_i e_{\mathfrak{p}_i}.$$

Proof. Let ord_{D_i} be the valuation on the local ring at the generic point of D_i. Then, by Proposition 4.1.3, the inequality

$$r_{ij} \leq e_{ij} - 1 + e_{ij} \cdot \operatorname{ord}_{D_i}(p_i^{m_i})$$

holds. Note that $\operatorname{ord}_{D_i}(p_i^{m_i}) = m_i e_{\mathfrak{p}_i}$. Since $p_i \leq \deg \pi$, we have that $m_i \geq 1$. Therefore,

$$r_{ij} \leq e_{ij} - 1 + e_{ij} m_i e_{\mathfrak{p}_i} \leq 2e_{ij} m_i e_{\mathfrak{p}_i}. \qquad \square$$

Let us introduce a bit more notation. Let I_1 be the set of i in I such that D_i is horizontal (i.e., $p_i = 0$) or $p_i > \deg \pi$. Let $D_1 = \sum_{i \in I_1} D_i$. We are now finally ready to combine our results to bound the "nonarchimedean" part of the height of a point.

Proposition 4.2.4. *Let $P' : S \to \mathcal{Y}'$ be a section, and let $Q : S \to \mathcal{X}$ be the induced section. If the image of P' is not contained in the support of $K_{\mathcal{Y}'}$, then*

$$(K_{\mathcal{Y}'}, P')_{\text{fin}} \leq \deg \pi (D_1, Q)_{\text{fin}} + 2(\deg \pi)^2 \log(\deg \pi)[K : \mathbb{Q}].$$

Proof. Note that

$$B = \sum_{i \in I} \Big(\sum_{j \in J_i} r_{ij} f_{ij} \Big) D_i.$$

Let I_2 be the complement of I_1 in I. Let $D_2 = \sum_{i \in I_2} D_i$, and note that $D = D_1 + D_2$. In particular,

$$\begin{aligned}(B, Q)_{\text{fin}} &= \sum_{i \in I} \sum_{j \in J_i} r_{ij} f_{ij} (D_i, Q)_{\text{fin}} \\ &= \sum_{i \in I_1} \sum_{j \in J_i} r_{ij} f_{ij} (D_i, Q)_{\text{fin}} + \sum_{i \in I_2} \sum_{j \in J_i} r_{ij} f_{ij} (D_i, Q)_{\text{fin}}.\end{aligned}$$

Note that, for all i in I_1 and j in J_i, the ramification of D_{ij} over D_i is tame; i.e., the equality $r_{ij} = e_{ij} - 1$ holds. Note that, for all i in I, we have $\sum_{j\in J_i} e_{ij} f_{ij} = \deg \pi$. Thus,

$$\sum_{i\in I_1}\sum_{j\in J_i} r_{ij} f_{ij}(D_i, Q)_{\mathrm{fin}} \leq \sum_{i\in I_1}\sum_{j\in J_i} e_{ij} f_{ij}(D_i, Q)_{\mathrm{fin}} = \deg \pi (D_1, Q)_{\mathrm{fin}}.$$

We claim that

$$\sum_{i\in I_2}\sum_{j\in J_i} r_{ij} f_{ij}(D_i, Q)_{\mathrm{fin}} \leq 2(\deg \pi)^2 \log(\deg \pi)[K : \mathbb{Q}].$$

In fact, since, for all i in I_2 and j in J_i, by Lemma 4.2.3, the inequality

$$r_{ij} \leq 2 e_{ij} m_i e_{\mathfrak{p}_i}$$

holds, we have that

$$\begin{aligned}\sum_{i\in I_2}\sum_{j\in J_i} r_{ij} f_{ij}(D_i, Q)_{\mathrm{fin}} &\leq 2\sum_{i\in I_2} m_i e_{\mathfrak{p}_i}(D_i, Q)_{\mathrm{fin}}\Bigg(\sum_{j\in J_i} e_{ij} f_{ij}\Bigg)\\ &= 2(\deg \pi)\sum_{i\in I_2} m_i e_{\mathfrak{p}_i}(D_i, Q)_{\mathrm{fin}}.\end{aligned}$$

Note that $(D_i, Q) = \log(\#k(\mathfrak{p}_i)) = f_{\mathfrak{p}_i} \log p_i$. We conclude that

$$\begin{aligned}\sum_{i\in I_2} m_i e_{\mathfrak{p}_i}(D_i, Q)_{\mathrm{fin}} &= \sum_{p \text{ prime}}\Bigg(\sum_{i\in I_2,\ p_i = p} e_{\mathfrak{p}_i} f_{\mathfrak{p}_i}\Bigg)\left\lfloor \frac{\log(\deg \pi)}{\log p}\right\rfloor \log(p)\\ &= [K : \mathbb{Q}] \sum_{\mathcal{X}_p \cap |D_2| \neq \varnothing} \left\lfloor \frac{\log(\deg \pi)}{\log p}\right\rfloor \log(p),\end{aligned}$$

where the last sum runs over all prime numbers p such that the fibre $\mathcal{X}_p$ contains an irreducible component of the support of D_2. Thus,

$$(B, Q)_{\mathrm{fin}} \leq (\deg \pi)(D_1, Q)_{\mathrm{fin}} + 2(\deg \pi)[K : \mathbb{Q}] \sum_{\mathcal{X}_p \cap D_2 \neq \varnothing} \left\lfloor \frac{\log(\deg \pi)}{\log p}\right\rfloor \log(p).$$

Note that

$$\sum_{\mathcal{X}_p \cap D_2 \neq \varnothing} \left\lfloor \frac{\log(\deg \pi)}{\log p}\right\rfloor \log(p) \leq \sum_{\mathcal{X}_p \cap D_2 \neq \varnothing} \log(\deg \pi) \leq \deg \pi \log(\deg \pi),$$

where we used that $\mathcal{X}_p \cap D_2 \neq \varnothing$ implies that $p \leq \deg \pi$. In particular,

$$(B, Q)_{\mathrm{fin}} \leq (\deg \pi)(D_1, Q)_{\mathrm{fin}} + 2(\deg \pi)^2 \log(\deg \pi)[K : \mathbb{Q}].$$

By Proposition 4.2.2, we conclude that

$$(K_{\mathcal{Y}'}, P')_{\text{fin}} \leq (\deg \pi)(D_1, Q)_{\text{fin}} + 2(\deg \pi)^2 \log(\deg \pi)[K : \mathbb{Q}]. \qquad \square$$

4.3. ***Models of covers of curves.*** In this section, we give a general construction for a model of a cover of the projective line. Let K be a number field with ring of integers O_K, and let $S = \operatorname{Spec} O_K$.

Proposition 4.3.1. *Let $\mathcal{Y} \to \operatorname{Spec} O_K$ be a flat projective morphism with geometrically connected fibres of dimension 1, where $\mathcal{Y}$ is an integral normal scheme. Then, there exists a finite field extension L/K such that the minimal resolution of singularities of the normalisation of $\mathcal{Y} \times_{O_K} O_L$ is semistable over O_L.*

Proof. This follows from [Liu 2006b, Corollary 2.8]. □

The main result of this section reads as follows.

Theorem 4.3.2. *Let K be a number field, and let Y be a smooth projective geometrically connected curve over K. Then, for any finite morphism $\pi_K : Y \to \mathbb{P}^1_K$, there exists a number field L/K such that*

- *the normalisation $\pi : \mathcal{Y} \to \mathbb{P}^1_{O_L}$ of $\mathbb{P}^1_{O_L}$ in the function field of Y_L is finite flat surjective,*
- *the minimal resolution of singularities $\psi : \mathcal{Y}' \to \mathcal{Y}$ is semistable over O_L and*
- *each irreducible component of the vertical part of the branch locus of the finite flat morphism $\pi : \mathcal{Y} \to \mathbb{P}^1_{O_L}$ is of characteristic less than or equal to $\deg \pi$. (The characteristic of a prime divisor D on $\mathbb{P}^1_{O_L}$ is the residue characteristic of the local ring at the generic point of D.)*

Proof. By Proposition 4.3.1, there exists a finite field extension L/K such that the minimal resolution of singularities $\psi : \mathcal{Y}' \to \mathcal{Y}$ of the normalisation of $\mathbb{P}^1_{O_L}$ in the function field of Y_L is semistable over O_L. Note that the finite morphism $\pi : \mathcal{Y} \to \mathbb{P}^1_{O_L}$ is flat. (The source is normal of dimension 2, and the target is regular.) Moreover, since the fibres of $\mathcal{Y}' \to \operatorname{Spec} O_L$ are reduced, the fibres of $\mathcal{Y}$ over O_L are reduced. Let $\mathfrak{p} \subset O_L$ be a maximal ideal of residue characteristic strictly bigger than $\deg \pi$, and note that the ramification of $\pi : \mathcal{Y} \to \mathbb{P}^1_{O_L}$ over (each prime divisor of $\mathbb{P}^1_{O_L}$ lying over) $\mathfrak{p}$ is tame. Since the fibres of $\mathcal{Y} \to \operatorname{Spec} O_L$ are reduced, we see that the finite morphism π is unramified over $\mathfrak{p}$. In fact, since $\mathbb{P}^1_{O_L} \to \operatorname{Spec} O_L$ has reduced (even smooth) fibres, the valuation of the different ideal $\mathcal{D}_{\mathcal{O}_D/\mathcal{O}_{\pi(D)}}$ on $\mathcal{O}_D$ of an irreducible component D of $\mathcal{Y}_{\mathfrak{p}}$ lying over $\pi(D)$ in $\mathcal{X}$ is precisely the multiplicity of D in $\mathcal{Y}_{\mathfrak{p}}$. (Here we let $\mathcal{O}_D$ denote the local ring at the generic point of D and $\mathcal{O}_{\pi(D)}$ the local ring at the generic point of $\pi(D)$.) Thus, each irreducible component of the vertical part of the branch locus of $\pi : \mathcal{Y} \to \mathbb{P}^1_{O_L}$ is of characteristic less or equal to $\deg \pi$. □

4.4. ***The modular lambda function.*** The modular function $\lambda : \mathbb{H} \to \mathbb{C}$ is defined as

$$\lambda(\tau) = \frac{\wp\left(\frac{1}{2} + \frac{\tau}{2}\right) - \wp\left(\frac{\tau}{2}\right)}{\wp\left(\frac{\tau}{2}\right) - \wp\left(\frac{1}{2}\right)},$$

where $\wp$ denotes the Weierstrass elliptic function for the lattice $\mathbb{Z} + \tau\mathbb{Z}$ in $\mathbb{C}$. The function λ is $\Gamma(2)$-invariant. More precisely, λ factors through the $\Gamma(2)$-quotient map $\mathbb{H} \to Y(2)$ and an analytic isomorphism $Y(2) \xrightarrow{\sim} \mathbb{C} \setminus \{0, 1\}$. Thus, the modular function λ induces an analytic isomorphism $X(2) \to \mathbb{P}^1(\mathbb{C})$. Let us note that $\lambda(i\infty) = 0$, $\lambda(1) = \infty$ and $\lambda(0) = 1$.

The restriction of λ to the imaginary axis $\{iy : y > 0\}$ in $\mathbb{H}$ induces a homeomorphism, also denoted by λ, from $\{iy : y > 0\}$ to the open interval $(0, 1)$ in $\mathbb{R}$. In fact, for α in the open interval $(0, 1)$,

$$\lambda^{-1}(\alpha) = i \frac{\mathrm{M}(1, \sqrt{\alpha})}{\mathrm{M}(1, \sqrt{1-\alpha})},$$

where M denotes the arithmetic-geometric mean.

Lemma 4.4.1. *For τ in $\mathbb{H}$, let $q(\tau) = \exp(\pi i \tau)$ and let $\lambda(\tau) = \sum_{n=1}^{\infty} a_n q^n(\tau)$ be the q-expansion of λ on $\mathbb{H}$. Then, for any real number $\frac{4}{5} \leq y \leq 1$,*

$$-\log \left| \sum_{n=1}^{\infty} n a_n q^n(iy) \right| \leq 2.$$

Proof. Note that

$$\sum_{n=1}^{\infty} n a_n q^n = q \frac{d\lambda}{dq}.$$

It suffices to show that $|q d\lambda/dq| \geq \frac{3}{20}$. We will use the product formula for λ. Namely,

$$\lambda(q) = 16q \prod_{n=1}^{\infty} f_n(q) \quad \text{and} \quad f_n(q) := \frac{1 + q^{2n}}{1 + q^{2n-1}}.$$

Write $f_n'(q) = df_n(q)/dq$. Then,

$$q \frac{d\lambda}{dq} = \lambda \left(1 + q \sum_{n=1}^{\infty} \frac{f_n'(q)}{f_n(q)} \right) = \lambda \left(1 + q \sum_{n=1}^{\infty} \frac{d}{dq} \log f_n(q) \right).$$

Note that, for any positive integer n and $\frac{4}{5} \leq y \leq 1$,

$$\left(\frac{d}{dq} \log f_n(q) \right)(iy) \leq 0.$$

Moreover, since $\lambda(i) = \frac{1}{2}$ and $\lambda(0) = 1$, the inequality $\lambda(iy) \geq \frac{1}{2}$ holds for all $0 \leq y \leq 1$. Also, for $\frac{4}{5} \leq y \leq 1$,

$$\Big(-q \sum_{n=1}^{\infty} \frac{d}{dq} \log f_n(q)\Big)(iy) \leq \frac{7}{10}.$$

In fact,

$$\sum_{n=1}^{\infty} \frac{d}{dq} \log f_n(q) = \sum_{n=1}^{\infty} \frac{2nq^{2n-1}}{1+q^{2n}} - \sum_{n=1}^{\infty} \frac{(2n-1)q^{2n-2}}{1+q^{2n-1}}.$$

It is straightforward to verify that, for all $\frac{4}{5} \leq y \leq 1$, the inequality

$$\sum_{n=1}^{\infty} \frac{2nq^{2n-1}(iy)}{1+q^{2n}(iy)} - \sum_{n=1}^{\infty} \frac{(2n-1)q^{2n-2}(iy)}{1+q^{2n-1}(iy)} \\ \geq \frac{100}{109} \sum_{n=1}^{\infty} 2nq^{2n-1}(iy) - \sum_{n=1}^{\infty} (2n-1)q^{2n-2}(iy)$$

holds. Finally, utilising classical formulas for geometric series, for all $\frac{4}{5} \leq y \leq 1$,

$$q(iy) \sum_{n=1}^{\infty} \frac{d}{dq} (\log f_n(q))(iy) \geq q(iy) \Big(\frac{200q(iy)}{109(1-q^2(iy))^2} - \frac{1+q^2(iy)}{(1-q^2(iy))^2} \Big) \geq \tfrac{7}{10}.$$

We conclude that

$$\left| q \frac{d\lambda}{dq} \right| \geq \tfrac{1}{2}\big(1 - \tfrac{7}{10}\big) = \tfrac{3}{20}. \qquad \square$$

4.5. ***A non-Weierstrass point with bounded height.*** The logarithmic height of a nonzero rational number $a = p/q$ is given by

$$h_{\text{naive}}(a) = \log \max(|p|, |q|),$$

where p and q are coprime integers and $q > 0$.

Theorem 4.5.1. *Let $\pi_{\overline{\mathbb{Q}}} : Y \to \mathbb{P}^1_{\overline{\mathbb{Q}}}$ be a finite morphism of degree d, where $Y/\overline{\mathbb{Q}}$ is a smooth projective connected curve of positive genus $g \geq 1$. Assume that $\pi_{\overline{\mathbb{Q}}} : Y \to \mathbb{P}^1_{\overline{\mathbb{Q}}}$ is unramified over $\mathbb{P}^1_{\overline{\mathbb{Q}}} \setminus \{0, 1, \infty\}$. Then, for any rational number $0 < a \leq \frac{2}{3}$ and any $b \in Y(\overline{\mathbb{Q}})$ lying over a,*

$$h(b) \leq 3h_{\text{naive}}(a)d^2 + 6378031\frac{d^5}{g}.$$

Proof. By Theorem 4.3.2, there exist a number field K and a model

$$\pi_K : Y \to \mathbb{P}^1_K$$

for $\pi_{\bar{\mathbb{Q}}} : Y \to \mathbb{P}^1_{\bar{\mathbb{Q}}}$ with the following three properties: the minimal resolution of singularities $\psi : \mathcal{Y}' \to \mathcal{Y}$ of the normalisation $\pi : \mathcal{Y} \to \mathbb{P}^1_{O_K}$ of $\mathbb{P}^1_{O_K}$ in $\mathcal{Y}$ is semistable over O_K, each irreducible component of the vertical part of the branch locus of $\pi : \mathcal{Y} \to \mathbb{P}^1_{O_K}$ is of characteristic less than or equal to $\deg \pi$ and every point in the fibre of π_K over a is K-rational. Also, the morphism $\pi : \mathcal{Y} \to \mathbb{P}^1_{O_K}$ is finite flat surjective.

Let $b \in Y(K)$ lie over a. Let P' be the closure of b in $\mathcal{Y}'$. By Lemma 2.3.1, the height of b is "minimal" on the minimal regular model. That is,

$$h(b) \leq \frac{(P', \omega_{\mathcal{Y}'/O_K})}{[K : \mathbb{Q}]}.$$

Recall the following notation from Section 4.2. Let $\mathcal{X} = \mathbb{P}^1_{O_K}$. Let $K_{\mathcal{X}} = -2 \cdot [\infty]$ be the divisor defined by the tautological section. Let $K_{\mathcal{Y}'}$ be the divisor on $\mathcal{Y}'$ defined by $d(\pi_K)$ viewed as a rational section of $\omega_{\mathcal{Y}'/O_K}$. Since the support of $K_{\mathcal{Y}'}$ on the generic fibre is contained in $\pi_K^{-1}(\{0, 1, \infty\})$, the section P' is not contained in the support of $K_{\mathcal{Y}'}$. Therefore, we get that

$$h(b)[K : \mathbb{Q}] \leq (P', \omega_{\mathcal{Y}'/O_K}) = (P', K_{\mathcal{Y}'})_{\mathrm{fin}} + \sum_{\sigma : K \to \mathbb{C}} (-\log \|d\pi_K\|_\sigma)(\sigma(b)).$$

Let D be the branch locus of $\pi : \mathcal{Y} \to \mathcal{X}$ endowed with the reduced closed subscheme structure. Write $D = 0 + 1 + \infty + D_{\mathrm{ver}}$, where D_{ver} is the vertical part of D. Note that, in the notation of Section 4.2, we have that $D_1 = 0 + 1 + \infty$. Thus, if Q denotes the closure of a in $\mathcal{X}$, by Proposition 4.2.4, we get

$$(P', K_{\mathcal{Y}'})_{\mathrm{fin}} \leq (\deg \pi)(0 + 1 + \infty, Q)_{\mathrm{fin}} + 2(\deg \pi)^2 \log(\deg \pi)[K : \mathbb{Q}].$$

Write $a = p/q$, where p and q are coprime positive integers with $q > p$. Note that

$$\begin{aligned}(0 + 1 + \infty, Q)_{\mathrm{fin}} &= [K : \mathbb{Q}] \log(pq(q-p)) \\ &\leq 3 \log(q)[K : \mathbb{Q}] \\ &= 3h_{\mathrm{naive}}(a)[K : \mathbb{Q}].\end{aligned}$$

We conclude that

$$\frac{(P', K_{\mathcal{Y}'})_{\mathrm{fin}}}{[K : \mathbb{Q}]} \leq 3h_{\mathrm{naive}}(a)(\deg \pi)^2 + 2(\deg \pi)^3.$$

It remains to estimate $\sum_{\sigma : K \to \mathbb{C}} (-\log \|d\pi_K\|_\sigma)(\sigma(b))$. We will use our bounds for Arakelov–Green functions.

Let $\sigma : K \to \mathbb{C}$ be an embedding. The composition

$$Y_\sigma \xrightarrow{\pi_\sigma} \mathbb{P}^1(\mathbb{C}) \xrightarrow{\lambda^{-1}} X(2)$$

is a Belyi cover (Definition 3.2.1). By abuse of notation, let π denote the composed morphism $Y_\sigma \to X(2)$. Note that $\lambda^{-1}(\frac{2}{3}) \approx 0.85i$. In particular, $\Im(\lambda^{-1}(a)) \geq \Im\big(\lambda^{-1}\big(\frac{2}{3}\big)\big) > s_1$. (Recall that $s_1 = \sqrt{1/2}$.) Therefore, the element $\lambda^{-1}(a)$ lies in $\dot{B}^{s_1}_\infty$. Since $V_y^{r_1} \supset V_y \cap \pi^{-1} B^{s_1}_\infty$, there is a unique cusp y of $Y_\sigma \to X(2)$ lying over ∞ such that $\sigma(b)$ lies in $V_y^{r_1}$.

Note that $q = z_\infty \exp(-\pi/2)$. Therefore, since $\lambda = \sum_{j=1}^{\infty} a_j q^j$ on $\mathbb{H}$,

$$\lambda \circ \pi = \sum_{j=1}^{\infty} a_j \exp(-j\pi/2)(z_\infty \circ \pi)^j = \sum_{j=1}^{\infty} a_j \exp(-j\pi/2) w_y^{e_y j}$$

on V_y. Thus, by the chain rule,

$$d(\lambda \circ \pi) = e_y \sum_{j=1}^{\infty} j a_j \exp(-j\pi/2) w_y^{e_y j - 1} d(w_y).$$

By the trivial inequality $e_y \geq 1$, the inequality $|w_y| \leq 1$ and Lemma 4.4.1,

$$\begin{aligned} &-\log \|d(\lambda \circ \pi)\|_{\mathrm{Ar}}(\sigma(b)) \\ &\qquad = -\log \|dw_y\|_{\mathrm{Ar}}(\sigma(b)) - \log \left| e_y \sum_{j=1}^{\infty} j a_j \exp(-j\pi/2) w_y^{e_y j - 1}(\sigma(b)) \right| \\ &\qquad \leq -\log \|dw_y\|_{\mathrm{Ar}}(\sigma(b)) - \log \left| \sum_{j=1}^{\infty} j a_j \exp(-j\pi/2) w_y^{e_y j}(\sigma(b)) \right| \\ &\qquad \leq -\log \|dw_y\|_{\mathrm{Ar}}(\sigma(b)) + 2. \end{aligned}$$

Thus, by Theorem 3.4.5, we conclude that

$$\frac{\sum_{\sigma : K \to \mathbb{C}} (-\log \|d\pi_K\|_\sigma)(\sigma(b))}{[K : \mathbb{Q}]} \leq 6378027 \frac{(\deg \pi)^5}{g} + 2. \qquad \square$$

Theorem 4.5.2. *Let Y be a smooth projective connected curve over $\bar{\mathbb{Q}}$ of genus $g \geq 1$. For any finite morphism $\pi : Y \to \mathbb{P}^1_{\bar{\mathbb{Q}}}$ ramified over exactly three points, there exists a non-Weierstrass point b on Y such that*

$$h(b) \leq 6378033 \frac{(\deg \pi)^5}{g}.$$

Proof. Define the sequence $(a_n)_{n=1}^{\infty}$ of rational numbers by $a_1 = \frac{1}{2}$ and $a_n = n/(2n-1)$ for $n \geq 2$. Note that $\frac{1}{2} \leq a_n \leq \frac{2}{3}$ and that $h_{\mathrm{naive}}(a_n) \leq \log(2n)$. We may and do assume that $\pi : Y \to \mathbb{P}^1_{\bar{\mathbb{Q}}}$ is unramified over $\mathbb{P}^1_{\bar{\mathbb{Q}}} \setminus \{0, 1, \infty\}$. By Theorem 4.5.1, for all $x \in \pi^{-1}(\{a_n\})$,

$$h(x) \leq 3 \log(2n)(\deg \pi)^2 + 6378031 \frac{(\deg \pi)^5}{g}. \tag{9}$$

Since the number of Weierstrass points on Y is at most $g^3 - g$, there exists an integer $1 \leq i \leq (\deg \pi)^2$ such that the fibre $\pi^{-1}(a_i)$ contains a non-Weierstrass point, say b. Applying (9) to b, we conclude that

$$\begin{aligned} h(b) &\leq 3\log(2(\deg\pi)^2)(\deg\pi)^2 + 6378031\frac{(\deg\pi)^5}{g} \\ &\leq 2\frac{(\deg\pi)^5}{g} + 6378031\frac{(\deg\pi)^5}{g}. \end{aligned}$$ □

4.6. For a smooth projective connected curve X over $\bar{\mathbb{Q}}$, we let $\deg_B(X)$ denote the Belyi degree of X.

Proof of Theorem 1.1.1. The inequality $\Delta(X) \geq 0$ is trivial, the lower bound $e(X) \geq 0$ is due to Faltings [1984, Theorem 5] and the lower bound $h_{\mathrm{Fal}}(X) \geq -g\log(2\pi)$ is due to Bost (Lemma 2.4.4).

For the remaining bounds, we proceed as follows. By Theorem 4.5.2, there exists a non-Weierstrass point b in $X(\bar{\mathbb{Q}})$ such that

$$h(b) \leq 6378033\frac{\deg_B(X)^5}{g}.$$

By our bound on the Arakelov norm of the Wronskian differential in Proposition 3.5.1, we have $\log \|\mathrm{Wr}\|_{\mathrm{Ar}}(b) \leq 6378028g \deg_B(X)^5$. To obtain the theorem, we combine these bounds with Theorem 2.4.1. □

5. Computing coefficients of modular forms

Let $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ be a congruence subgroup, and let k be a positive integer. A modular form f of weight k for the group Γ is determined by k and its q-expansion coefficients $a_m(f)$ for $0 \leq m \leq k \cdot [\mathrm{SL}_2(\mathbb{Z}) : \{\pm 1\}\Gamma]/12$. In this section, we follow [Bruin 2011] and give an algorithmic application of the main result of this paper. More precisely, the goal of this section is to complete the proof of the following theorem. The proof is given at the end of this section.

Theorem 5.0.1 (Couveignes, Edixhoven, Bruin). *Assume the Riemann hypothesis for ζ-functions of number fields. There exists a probabilistic algorithm that, given*

- *a positive integer k,*
- *a number field K,*
- *a congruence subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$,*
- *a modular form f of weight k for Γ over K, and*
- *a positive integer m in factored form,*

computes $a_m(f)$ and whose expected running time is bounded by a polynomial in the length of the input.

Remark 5.0.2. We should make precise how the number field K, the congruence subgroup Γ and the modular form f should be given to the algorithm and how the algorithm returns the coefficient $a_m(f)$. We should also explain what "probabilistic" means in this context. For the sake of brevity, we refer the reader to [Bruin 2011, p. 20] for the precise definitions. Following the definitions there, the above theorem becomes a precise statement.

Remark 5.0.3. The algorithm in Theorem 5.0.1 is due to Bruin, Couveignes and Edixhoven. Assuming the Riemann hypothesis for ζ-functions of number fields, it was shown that the algorithm runs in polynomial time for *certain* congruence subgroups; see [Bruin 2011, Theorem 1.1]. Bruin did not have enough information about the semistable bad reduction of the modular curve $X_1(n)$ at primes p such that p^2 divides n to show that the algorithm runs in polynomial time. Nevertheless, our bounds on the discriminant of a curve can be used to show that the algorithm runs in polynomial time for *all* congruence subgroups.

Proof of Theorem 5.0.1. We follow Bruin's strategy [2010, Chapter V.1, p. 165]. He notes that, to assure that the algorithm runs in polynomial time for all congruence subgroups, it suffices to show that, for all positive integers n, the discriminant $\Delta(X_1(n))$ is polynomial in n (or equivalently the genus of $X_1(n)$). The latter follows from Corollary 1.5.1. In fact, the Belyi degree of $X_1(n)$ is at most the index of $\Gamma_1(n)$ in $\mathrm{SL}_2(\mathbb{Z})$. Since

$$[\mathrm{SL}_2(\mathbb{Z}) : \Gamma_1(n)] = n^2 \prod_{p \mid n} (1 - 1/p^2) \le n^2,$$

we conclude that $\Delta(X_1(n)) \le 5 \cdot 10^8 n^{14}$. □

6. Bounds for heights of covers of curves

Let X be a smooth projective connected curve over $\bar{\mathbb{Q}}$. We prove that Arakelov invariants of (possibly ramified) covers of X are polynomial in the degree. Let us be more precise.

Theorem 6.0.4. *Let X be a smooth projective connected curve over $\bar{\mathbb{Q}}$, let U be a nonempty open subscheme of X, let $B_f \subset \mathbb{P}^1(\bar{\mathbb{Q}})$ be a finite set and let $f : X \to \mathbb{P}^1_{\bar{\mathbb{Q}}}$ be a finite morphism unramified over $\mathbb{P}^1_{\bar{\mathbb{Q}}} - B_f$. Define*

$$B := f(X \setminus U) \cup B_f.$$

Let N be the number of elements in the orbit of B under the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, *and let H_B be the height of B as defined in Section 1.3. Define*

$$c_B := (4N H_B)^{45N^3 2^{N-2} N!}.$$

Then, for any finite morphism $\pi : Y \to X$ étale over U, where Y is a smooth projective connected curve over $\overline{\mathbb{Q}}$ of genus $g \geq 1$,

$$\begin{aligned}
-\log(2\pi)g \leq h_{\mathrm{Fal}}(Y) &\leq 13 \cdot 10^6 g c_B (\deg f)^5 (\deg \pi)^5,\\
0 \leq \quad e(Y) \quad &\leq 3 \cdot 10^7 (g-1) c_B (\deg f)^5 (\deg \pi)^5,\\
0 \leq \quad \Delta(Y) \quad &\leq 5 \cdot 10^8 g^2 c_B (\deg f)^5 (\deg \pi)^5,\\
-10^8 g^2 c_B (\deg f)^5 (\deg \pi)^5 \leq \delta_{\mathrm{Fal}}(Y) &\leq 2 \cdot 10^8 g c_B (\deg f)^5 (\deg \pi)^5.
\end{aligned}$$

Proof. We apply Khadjavi's effective version of Belyi's theorem. More precisely, by [Khadjavi 2002, Theorem 1.1.c], there exists a finite morphism $R : \mathbb{P}^1_{\overline{\mathbb{Q}}} \to \mathbb{P}^1_{\overline{\mathbb{Q}}}$ étale over $\mathbb{P}^1_{\overline{\mathbb{Q}}} \setminus \{0, 1, \infty\}$ such that $\deg R \leq (4NH_B)^{9N^3 2^{N-2} N!}$ and $R(B) \subset \{0, 1, \infty\}$. Note that the composed morphism

$$R \circ f \circ \pi : Y \xrightarrow{\pi} X \xrightarrow{f} \mathbb{P}^1_{\overline{\mathbb{Q}}} \xrightarrow{R} \mathbb{P}^1_{\overline{\mathbb{Q}}}$$

is unramified over $\mathbb{P}^1_{\overline{\mathbb{Q}}} \setminus \{0, 1, \infty\}$. We conclude by applying Theorem 1.1.1 to the composition $R \circ f \circ \pi$. □

Note that Theorem 6.0.4 implies Theorem 1.3.1 (with $X = \mathbb{P}^1_{\overline{\mathbb{Q}}}$, B_f the empty set and $f : X \to \mathbb{P}^1_{\overline{\mathbb{Q}}}$ the identity morphism).

In the proof of Theorem 6.0.4, we used Khadjavi's effective version of Belyi's theorem. Khadjavi's bounds are not optimal; see [Liţcanu 2004, Lemme 4.1] and [Khadjavi 2002, Theorem 1.1.b] for better bounds when B is contained in $\mathbb{P}^1(\mathbb{Q})$. Actually, the use of Belyi's theorem makes the dependence on the branch locus enormous in Theorem 6.0.4. It should be possible to avoid the use of Belyi's theorem and improve the dependence on the branch locus in Theorem 6.0.4. This is not necessary for our present purposes.

Remark 6.0.5. Let us mention the quantitative Riemann existence theorem due to Bilu and Strambi [2010]. Bilu and Strambi give explicit bounds for the naive logarithmic height of a cover of $\mathbb{P}^1_{\overline{\mathbb{Q}}}$ with fixed branch locus. Although their bound on the naive height is exponential in the degree, the dependence on the height of the branch locus in their result is logarithmic.

Let us show that Theorem 1.3.1 implies the following:

Theorem 6.0.6 [Edixhoven et al. 2010, Conjecture 5.1]. *Let $U \subset \mathbb{P}^1_{\mathbb{Z}}$ be a nonempty open subscheme. Then there are integers a and b with the following property. For any prime number ℓ and for any connected finite étale cover $\pi : V \to U_{\mathbb{Z}[1/\ell]}$, the Faltings height of the normalisation of $\mathbb{P}^1_{\mathbb{Q}}$ in the function field of V is bounded by $(\deg \pi)^a \ell^b$.*

Proof. We claim that this conjecture holds with $b = 0$ and an integer a depending only on the generic fibre $U_{\mathbb{Q}}$ of U. In fact, let $\pi : Y \to \mathbb{P}^1_{\mathbb{Q}}$ denote the normalisation of $\mathbb{P}^1_{\mathbb{Q}}$

in the function field of V. Note that π is étale over $U_{\mathbb{Q}}$. Let $B = \mathbb{P}^1_{\mathbb{Q}} - U_{\mathbb{Q}} \subset \mathbb{P}^1(\bar{\mathbb{Q}})$, and let N be the number of elements in the orbit of B under the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. By Theorem 1.3.1,

$$h_{\mathrm{Fal}}(Y) := \sum_{X \subset Y_{\bar{\mathbb{Q}}}} h_{\mathrm{Fal}}(X) \leq (\deg \pi)^a,$$

where the sum runs over all connected components X of $Y_{\bar{\mathbb{Q}}} := Y \times_{\mathbb{Q}} \bar{\mathbb{Q}}$, and

$$a = 6 + \log\bigl(13 \cdot 10^6 N (4 N H_B)^{45 N^3 2^{N-2} N!}\bigr).$$

Here we used that $g \leq N \deg \pi$ and

$$13 \cdot 10^6 g (4 N H_B)^{45 N^3 2^{N-2} N!} \leq (\deg \pi)^{1 + \log(13 \cdot 10^6 N (4 N H_B)^{45 N^3 2^{N-2} N!})}.$$

This concludes the proof. □

Let us briefly mention the context in which these results will hopefully be applied. Let S be a smooth projective geometrically connected surface over $\mathbb{Q}$. As is explained in Section 5 of [Edixhoven et al. 2010], it seems reasonable to suspect that there exists an algorithm that, on input of a prime ℓ, computes the étale cohomology groups $\mathrm{H}^i(S_{\bar{\mathbb{Q}},\text{ét}}, \mathbb{F}_\ell)$ with their $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$-action in time *polynomial* in ℓ for all $i = 0, \ldots, 4$.

Appendix: Merkl's method of bounding Green functions

by Peter Bruin

The goal of this appendix is to prove Theorem 3.1.2. Let X be a compact connected Riemann surface, and let μ be a smooth nonnegative $(1, 1)$-form on X such that $\int_X \mu = 1$. Let $*$ denote the star operator on 1-forms on X, given with respect to a holomorphic coordinate $z = x + iy$ by

$$*dx = dy \quad \text{and} \quad *dy = -dx,$$

or equivalently

$$*dz = -i\, d\bar{z} \quad \text{and} \quad *d\bar{z} = i\, dz.$$

The *Green function* for μ is the unique smooth function

$$\mathrm{gr}_\mu : (X \times X) \setminus \Delta \to \mathbb{R},$$

with a logarithmic singularity along the diagonal Δ, such that for fixed $w \in X$ we have, in a distributional sense,

$$\frac{1}{2\pi} d{*}d\, \mathrm{gr}_\mu(z, w) = \delta_w(z) - \mu(z) \quad \text{and} \quad \int_{z \in X \setminus \{w\}} \mathrm{gr}_\mu(z, w) \mu(z) = 0.$$

For all $a, b \in X$, we write $g_{a,b}$ for the unique smooth function on $X \setminus \{a, b\}$ satisfying

$$d{*}dg_{a,b} = \delta_a - \delta_b \quad \text{and} \quad \int_{X\setminus\{a,b\}} g_{a,b}\mu = 0. \tag{1}$$

Then for all $a \in X$, we consider the function $g_{a,\mu}$ on $X \setminus \{a\}$ defined by

$$g_{a,\mu}(x) = \int_{b\in X\setminus\{x\}} g_{a,b}(x)\mu(b). \tag{2}$$

A straightforward computation using Fubini's theorem shows that this function satisfies

$$d{*}dg_{a,\mu} = \delta_a - \mu \quad \text{and} \quad \int_{X\setminus\{a\}} g_{a,\mu}\mu = 0.$$

This implies that $2\pi g_{a,\mu}(b) = \mathrm{gr}_\mu(a, b)$, where gr_μ is the Green function for μ defined above.

We begin by restricting our attention to one of the charts of our atlas, say (U, z). By assumption, z is an isomorphism from U to the open unit disc in $\mathbb{C}$. Let r_2 and r_4 be real numbers with

$$r_1 < r_2 < r_4 < 1,$$

and write

$$r_3 = (r_2 + r_4)/2.$$

We choose a smooth function

$$\tilde{\chi} : \mathbb{R}_{\geq 0} \to [0, 1]$$

such that $\tilde{\chi}(r) = 1$ for $r \leq r_2$ and $\tilde{\chi}(r) = 0$ for $r \geq r_4$. We also define a smooth function χ on X by putting

$$\chi(x) = \tilde{\chi}(|z(x)|) \quad \text{for } x \in U$$

and extending by 0 outside U. Furthermore, we put

$$\chi^{\mathrm{c}} = 1 - \chi.$$

For $0 < r < 1$, we write

$$U^r = \{x \in U : |z(x)| < r\}.$$

For all $a, b \in U^{r_1}$, the function

$$f_{a,b} = \frac{1}{2\pi} \log \left| \frac{(z - z(a))(\overline{z(a)}z - r_4^2)}{(z - z(b))(\overline{z(b)}z - r_4^2)} \right|$$

is defined on $U \setminus \{a, b\}$. Moreover, $f_{a,b}$ is harmonic on $U \setminus \{a, b\}$ since the logarithm of the modulus of a holomorphic function is harmonic. We extend $\chi^{\mathrm{c}} f_{a,b}$ to a smooth function on U by defining it to be zero in a and b.

We consider the open annulus

$$A = U^{r_4} \setminus \overline{U^{r_2}}.$$

Let (ρ, ϕ) be polar coordinates on A such that $z = \rho \exp(i\phi)$. A straightforward calculation shows that in these coordinates the star operator is given by

$$*d\rho = \rho\, d\phi \quad \text{and} \quad *d\phi = -\frac{d\rho}{\rho}.$$

We consider the inner product

$$\langle \alpha, \beta \rangle_A = \int_A \alpha \wedge *\beta$$

on the $\mathbb{R}$-vector space of square-integrable real-valued 1-forms on A. Furthermore, we write

$$\|\alpha\|_A^2 = \langle \alpha, \alpha \rangle_A.$$

Lemma A.1. *For every real harmonic function g on A such that $\|dg\|_A$ exists,*

$$\max_{|z|=r_3} g - \min_{|z|=r_3} g \le \frac{2\sqrt{\pi}}{r_4 - r_2} \|dg\|_A.$$

Proof. By the formula for the star operator in polar coordinates,

$$\begin{aligned} dg \wedge *dg &= (\partial_\rho g\, d\rho + \partial_\phi g\, d\phi) \wedge (\rho \partial_\rho g\, d\phi - \rho^{-1} \partial_\phi g\, d\rho) \\ &= ((\partial_\rho g)^2 + (\rho^{-1} \partial_\phi g)^2) \rho\, d\rho\, d\phi. \end{aligned}$$

Using the mean value theorem, we can bound the left-hand side of the inequality we need to prove by

$$\begin{aligned} \max_{|z|=r_3} g - \min_{|z|=r_3} g &\le \pi \max_{|z|=r_3} |\partial_\phi g| \\ &= \pi |\partial_\phi g|(x) \quad \text{for some } x \text{ with } |z(x)| = r_3. \end{aligned}$$

We write $R = (r_4 - r_2)/2$, and we consider the open disc

$$D = \{z \in U : |z - z(x)| < R\}$$

of radius R around x; this lies in A because $r_3 = (r_4 + r_2)/2$. Let (σ, ψ) be polar coordinates on D such that $z - z(x) = \sigma \exp(i\psi)$. Because g is harmonic, so is $\partial_\phi g$, and Gauss's mean value theorem implies that

$$\partial_\phi g(x) = \frac{1}{\pi R^2} \int_D \partial_\phi g\, \sigma\, d\sigma\, d\psi.$$

On the space of real continuous functions on D, we have the inner product

$$(h_1, h_2) \mapsto \int_D h_1 h_2 \, \sigma \, d\sigma \, d\psi.$$

Applying the Cauchy–Schwarz inequality with $h_1 = \rho^{-1}\partial_\phi g$ and $h_2 = \rho$ gives

$$\begin{aligned}\left|\int_D \partial_\phi g \, \sigma \, d\sigma \, d\psi\right| &\leq \left[\int_D (\rho^{-1}\partial_\phi g)^2 \sigma \, d\sigma \, d\psi\right]^{1/2} \cdot \left[\int_D \rho^2 \sigma \, d\sigma \, d\psi\right]^{1/2} \\ &\leq \left[\int_A (\rho^{-1}\partial_\phi g)^2 \rho \, d\rho \, d\phi\right]^{1/2} \cdot \left[\int_D \sigma \, d\sigma \, d\psi\right]^{1/2} \\ &\leq \left[\int_A dg \wedge *dg\right]^{1/2} [\pi R^2]^{1/2} \\ &= \sqrt{\pi}\, R \|dg\|_A.\end{aligned}$$

Combining the above results finishes the proof. □

Lemma A.2. *For all $a, b \in U^{r_1}$, there exists a smooth function $\tilde{g}_{a,b}$ on X such that*

$$d{*}d\tilde{g}_{a,b} = \begin{cases} d{*}d(\chi^{\mathrm{c}} f_{a,b}) & \text{on } U, \\ 0 & \text{on } X \setminus \overline{U}. \end{cases}$$

It is unique up to an additive constant and fulfils

$$\|d\tilde{g}_{a,b}\|_A \leq \|d(\chi^{\mathrm{c}} f_{a,b})\|_A.$$

Proof. First we note that the expression on the right-hand side of the equality defines a smooth 2-form on X because $d{*}d(\chi^{\mathrm{c}} f_{a,b})(z)$ vanishes for $|z| > r_4$; this follows from the choice of χ and the fact that $f_{a,b}$ is harmonic for $|z| > r_1$. Since moreover $\chi^{\mathrm{c}} f_{a,b} = 0$ on U^{r_2}, we see that the support of this 2-form is contained in the closed annulus $\bar{A}$. By Stokes's theorem,

$$\int_{\bar{A}} d{*}d(\chi^{\mathrm{c}} f_{a,b}) = \int_{\partial \bar{A}} {*}d(\chi^{\mathrm{c}} f_{a,b}).$$

Notice that $f_{a,b}$ is invariant under the substitution $z \mapsto r_4^2/\bar{z}$; this implies that $\partial_\rho f_{a,b}(z) = 0$ for $|z| = r_4$. Furthermore, $\chi^{\mathrm{c}}(z) = 1$ and $d\chi^{\mathrm{c}}(z) = 0$ for $|z| = r_4$, so we see that

$$d(\chi^{\mathrm{c}} f_{a,b})(z) = \chi^{\mathrm{c}}(z)\, df_{a,b}(z) = (\partial_\phi f_{a,b}\, d\phi)(z) \quad \text{if } |z| = r_4.$$

Likewise, since $\chi^{\mathrm{c}} = 0$ and $d\chi^{\mathrm{c}}(z) = 0$ for $|z| = r_2$,

$$d(\chi^{\mathrm{c}} f_{a,b})(z) = \chi^{\mathrm{c}}(z) df_{a,b}(z) = 0 \quad \text{if } |z| = r_2.$$

This means that, for z on the boundary of $\bar{A}$,

$$*d(\chi^{\mathrm{c}} f_{a,b})(z) = \begin{cases} -(\partial_\phi f_{a,b}\, d\rho)(z) & \text{if } |z| = r_4, \\ 0 & \text{if } |z| = r_2. \end{cases}$$

In particular, $*d(\chi^{\mathrm{c}} f_{a,b})$ vanishes when restricted to the submanifold $\partial\bar{A}$ of X. From this, we conclude that

$$\int_{\bar{A}} d*d(\chi^{\mathrm{c}} f_{a,b}) = \int_{\partial\bar{A}} *d(\chi^{\mathrm{c}} f_{a,b}) = 0.$$

This implies that a function $\tilde{g}_{a,b}$ with the required property exists.

To prove the inequality $\|d\tilde{g}_{a,b}\|_A \le \|d(\chi^{\mathrm{c}} f_{a,b})\|_A$, we note that

$$\begin{aligned} \|d(\chi^{\mathrm{c}} f_{a,b})\|_A^2 &= \|d\tilde{g}_{a,b} + d(\chi^{\mathrm{c}} f_{a,b} - \tilde{g}_{a,b})\|_A^2 \\ &= \|d\tilde{g}_{a,b}\|_A^2 + 2\langle d\tilde{g}_{a,b}, d(\chi^{\mathrm{c}} f_{a,b} - \tilde{g}_{a,b})\rangle_A + \|d(\chi^{\mathrm{c}} f_{a,b} - \tilde{g}_{a,b})\|_A^2. \end{aligned}$$

The last term is clearly nonnegative. Furthermore, integration by parts using Stokes's theorem gives

$$\begin{aligned} \langle d\tilde{g}_{a,b}, d(\chi^{\mathrm{c}} f_{a,b} - \tilde{g}_{a,b})\rangle_A &= \int_A d\tilde{g}_{a,b} \wedge *d(\chi^{\mathrm{c}} f_{a,b} - \tilde{g}_{a,b}) \\ &= \int_{\partial\bar{A}} \tilde{g}_{a,b} *d(\chi^{\mathrm{c}} f_{a,b} - \tilde{g}_{a,b}) - \int_A \tilde{g}_{a,b}\, d*d(\chi^{\mathrm{c}} f_{a,b} - \tilde{g}_{a,b}). \end{aligned}$$

The second term vanishes because $d*d\tilde{g}_{a,b} = d*d(\chi^{\mathrm{c}} f_{a,b})$ on A. From our earlier expression for $*d(\chi^{\mathrm{c}} f_{a,b})(z)$ on the boundary of A, we see that

$$\int_{\partial\bar{A}} \tilde{g}_{a,b} *d(\chi^{\mathrm{c}} f_{a,b}) = 0.$$

Finally, because $\partial\bar{A}$ is also the (negatively oriented) boundary of $X \setminus A$ and because $d*d\tilde{g}_{a,b} = 0$ on $X \setminus A$,

$$-\int_{\partial\bar{A}} \tilde{g}_{a,b} *d\tilde{g}_{a,b} = \int_{X\setminus A} d\tilde{g}_{a,b} \wedge *d\tilde{g}_{a,b} \ge 0.$$

Thus, we have

$$\langle d\tilde{g}_{a,b}, d(\chi^{\mathrm{c}} f_{a,b} - \tilde{g}_{a,b})\rangle_A \ge 0,$$

which proves the inequality. □

Lemma A.3. *Let* $\lambda = \max_{r_2 \le r \le r_4} |\chi'(r)|$. *Then*

$$\max_X \tilde{g}_{a,b} - \min_X \tilde{g}_{a,b} \le c_3(r_1, r_2, r_4, \lambda),$$

where

$$c_3(r_1, r_2, r_4, \lambda) =$$
$$4\sqrt{\frac{r_4{+}r_2}{r_4{-}r_2}}\left(\frac{\lambda}{2}\log\frac{(r_1{+}r_4)^2}{(r_2{-}r_1)(r_4{-}r_1)}+\frac{1}{r_2{-}r_1}+\frac{r_1}{r_4(r_4{-}r_1)}\right)+\frac{2}{\pi}\log\frac{(r_1{+}r_4)^2}{(r_2{-}r_1)(r_4{-}r_1)}.$$

Proof. First, we note that

$$\max_X \tilde{g}_{a,b} = \max\Bigl\{\sup_{U^{r_3}} \tilde{g}_{a,b},\ \sup_{X\setminus U^{r_3}} \tilde{g}_{a,b}\Bigr\}, \quad \min_X \tilde{g}_{a,b} = \min\Bigl\{\inf_{U^{r_3}} \tilde{g}_{a,b},\ \inf_{X\setminus U^{r_3}} \tilde{g}_{a,b}\Bigr\}.$$

Furthermore,

$$\sup_{U^{r_3}} \tilde{g}_{a,b} \le \sup_{U^{r_3}}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) + \sup_{U^{r_3}} \chi^{\mathrm{c}} f_{a,b} = \max_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) + \max_{r_2\le|z|\le r_3} \chi^{\mathrm{c}} f_{a,b}$$

because of the maximum principle ($\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}$ is harmonic on U) and because $\chi^{\mathrm{c}}(z) = 0$ for $|z| < r_2$. In the same way, we find

$$\inf_{U^{r_3}} \tilde{g}_{a,b} \ge \min_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) + \min_{r_2\le|z|\le r_3} \chi^{\mathrm{c}} f_{a,b}.$$

We extend $\chi f_{a,b}$ to a smooth function on $X \setminus \{a, b\}$ by putting $(\chi f_{a,b})(x) = 0$ for $x \notin U$. Then $\tilde{g}_{a,b} + \chi f_{a,b}$ is harmonic on $X \setminus \{a, b\}$, and the same method as above gives us

$$\begin{aligned}\sup_{X\setminus U^{r_3}} \tilde{g}_{a,b} &\le \max_{|z|=r_3}(\tilde{g}_{a,b} + \chi f_{a,b}) - \min_{r_3\le|z|\le r_4} \chi f_{a,b}\\ &\le \max_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) + \max_{|z|=r_3} f_{a,b} - \min_{r_3\le|z|\le r_4} \chi f_{a,b}\end{aligned}$$

and

$$\inf_{X\setminus U^{r_3}} \tilde{g}_{a,b} \ge \min_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) + \min_{|z|=r_3} f_{a,b} - \max_{r_3\le|z|\le r_4} \chi f_{a,b}.$$

These bounds imply that

$$\begin{aligned}\max_X \tilde{g}_{a,b} &\le \max_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) + 2\sup_A |f_{a,b}|,\\ \min_X \tilde{g}_{a,b} &\ge \min_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) - 2\sup_A |f_{a,b}|\end{aligned}$$

and hence

$$\max_X \tilde{g}_{a,b} - \min_X \tilde{g}_{a,b} \le \max_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) - \min_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) + 4\sup_A |f_{a,b}|.$$

By Lemmas A.1 and A.2,

$$\begin{aligned}\max_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) - \min_{|z|=r_3}(\tilde{g}_{a,b} - \chi^{\mathrm{c}} f_{a,b}) &\le \frac{2\sqrt{\pi}}{r_4 - r_2}\|d\tilde{g}_{a,b} - d(\chi^{\mathrm{c}} f_{a,b})\|_A\\ &\le \frac{2\sqrt{\pi}}{r_4 - r_2}(\|d\tilde{g}_{a,b}\|_A + \|d(\chi^{\mathrm{c}} f_{a,b})\|_A)\\ &\le \frac{4\sqrt{\pi}}{r_4 - r_2}\|d(\chi^{\mathrm{c}} f_{a,b})\|_A.\end{aligned}$$

We have

$$\begin{aligned}\|d(\chi^{\mathrm{c}} f_{a,b})\|_A &\le \|d(\chi^{\mathrm{c}}) f_{a,b}\|_A + \|\chi^{\mathrm{c}} df_{a,b}\|_A \\ &\le \|\tilde{\chi}'(\rho) f_{a,b}\, d\rho\|_A + \|df_{a,b}\|_A \\ &\le \lambda \|d\rho\|_A \sup_A |f_{a,b}| + \|df_{a,b}\|_A.\end{aligned}$$

Now

$$\|d\rho\|_A^2 = \int_A d\rho \wedge *d\rho = \int_A \rho\, d\rho \wedge d\phi = \pi(r_4^2 - r_2^2).$$

Furthermore, for all $a, b \in U^{r_1}$, we have

$$|f_{a,b}(z)| = \frac{1}{2\pi}\bigl|\log|z - z(a)| + \log|\overline{z(a)}z - r_4^2| - \log|z - z(b)| - \log|\overline{z(b)}z - r_4^2|\bigr|.$$

For all $a \in U^{r_1}$ and all $z \in A$, the triangle inequality gives

$$r_2 - r_1 < |z - z(a)| < r_4 + r_1 \quad \text{and} \quad r_4(r_4 - r_1) < |\overline{z(a)}z - r_4^2| < r_4(r_4 + r_1).$$

From this, we deduce that, for all $a, b \in U^{r_1}$,

$$\sup_A |f_{a,b}| \le \frac{1}{2\pi} \log \frac{(r_1 + r_4)^2}{(r_2 - r_1)(r_4 - r_1)}.$$

Finally, we bound the quantity $\|df_{a,b}\|_A$. Because $f_{a,b}$ is a real function, we have $df_{a,b} = \partial_z f_{a,b}\, dz + \overline{\partial_z f_{a,b}}\, d\bar{z}$. Therefore,

$$\begin{aligned}\|df_{a,b}\|_A^2 &= \int_A df_{a,b} \wedge *df_{a,b} = 2i \int_A |\partial_z f_{a,b}|^2\, dz \wedge d\bar{z} \\ &= 4 \int_0^{2\pi} \int_{r_2}^1 |\partial_z f_{a,b}|^2\, \rho\, d\rho\, d\phi \le 4\pi(1 - r_2^2) \sup_A |\partial_z f_{a,b}|^2.\end{aligned}$$

A straightforward computation gives

$$\partial_z f_{a,b} = \frac{1}{4\pi}\left(\frac{1}{z - z(a)} + \frac{\overline{z(a)}}{\overline{z(a)}z - r_4^2} - \frac{1}{z - z(b)} - \frac{\overline{z(b)}}{\overline{z(b)}z - r_4^2}\right).$$

Our previous bounds for $|z - z(a)|$ and $|\overline{z(a)}z - 1|$ yield

$$\sup_A |\partial_z f_{a,b}| \le \frac{1}{2\pi}\left(\frac{1}{r_2 - r_1} + \frac{r_1}{r_4(r_4 - r_1)}\right).$$

From this, we obtain

$$\|df_{a,b}\|_A \le \sqrt{\frac{r_4^2 - r_2^2}{\pi}}\left(\frac{1}{r_2 - r_1} + \frac{r_1}{r_4(r_4 - r_1)}\right).$$

Combining the bounds for $\sup_A |f_{a,b}|$ and $\|df_{a,b}\|_A$ yields the lemma. □

From now on, we impose the normalisation condition

$$\int_X \tilde{g}_{a,b}\mu = 0$$

on $\tilde{g}_{a,b}$ for all $a, b \in U^{r_1}$; this can be attained by adding a suitable constant to $\tilde{g}_{a,b}$. Then for all $a, b \in U^{r_1}$, the function $g_{a,b}$ defined earlier is equal to

$$g_{a,b} = \tilde{g}_{a,b} + \chi f_{a,b} - \int_X \chi f_{a,b}\mu. \tag{3}$$

Indeed, by the definition of $\tilde{g}_{a,b}$, the right-hand side satisfies (1). Furthermore, for all $a \in U^{r_1}$, we define a smooth function l_a on $X \setminus \{a\}$ by

$$l_a = \begin{cases} (\chi/2\pi)\log|z - z(a)| & \text{on } U, \\ 0 & \text{on } X \setminus \overline{U}; \end{cases}$$

this is bounded from above by $(1/2\pi)\log(r_4 + r_1)$.

Lemma A.4. *For all $a, b \in U^{r_1}$, we have*

$$\max_X |g_{a,b} - l_a + l_b| < c_4(r_1, r_2, r_4, \lambda, c_1),$$

where

$$c_4(r_1, r_2, r_4, \lambda, c_1) = c_3(r_1, r_2, r_4, \lambda) + \frac{1}{2\pi}\log\frac{r_4 + r_1}{r_4 - r_1} + \left(\tfrac{8}{3}\log 2 - \tfrac{1}{4}\right)\frac{c_1}{r_4^2}.$$

Proof. By (3) and the definitions of $f_{a,b}$ and l_a, we get

$$g_{a,b} - l_a + l_b = \tilde{g}_{a,b} - \int_X \chi f_{a,b}\mu + \frac{\chi}{2\pi}\log\left|\frac{\overline{z(a)}z - r_4^2}{\overline{z(b)}z - r_4^2}\right|,$$

where the last term is extended to zero outside U. We bound each of the terms on the right-hand side. From $\int_X \tilde{g}_{a,b}\mu = 0$ and the nonnegativity of μ, it follows that

$$\max_X \tilde{g}_{a,b} \geq 0 \geq \min_X \tilde{g}_{a,b}.$$

Together with the bound for $\max_X \tilde{g}_{a,b} - \min_X \tilde{g}_{a,b}$ from Lemma A.3, this implies

$$\max_X |\tilde{g}_{a,b}| \leq c_3(r_1, r_2, r_4, \lambda, c_1).$$

Because the support of χ is contained in U^{r_4}, the hypothesis (4) of Definition 3.1.1 together with the definition of $f_{a,b}$ gives

$$\int_X \chi f_{a,b}\mu$$

$$= \int_{U^{r_4}} \frac{\chi}{2\pi}\left(\log\left|\frac{z - z(a)}{r_4}\right| + \log\left|\frac{\overline{z(a)}z}{r_4^2} - 1\right| - \log\left|\frac{z - z(b)}{r_4}\right| - \log\left|\frac{\overline{z(b)}z}{r_4^2} - 1\right|\right)\mu.$$

Writing $w = z/r_4$ and $t = z(a)/r_4$, we have

$$\int_{U^{r_4}} \frac{\chi}{2\pi} \log\left|\frac{z - z(a)}{r_4}\right| \mu \le \frac{c_1}{2\pi r_4^2} \int_{\substack{|w|<1\\ |w-t|>1}} \log|w - t|\, i\, dw \wedge d\bar{w}.$$

We note that t satisfies $|t| < r_1/r_4$; for simplicity, we relax this to $|t| \le 1$. Then it is easy to see that the above expression attains its maximum for $|t| = 1$; by rotational symmetry, we can take $t = 1$. We now have to integrate over the crescent-shaped domain $\{w \in \mathbb{C} : |w| < 1 \text{ and } |w - 1| > 1\}$, which is contained in $\{1 + r\exp(i\phi) : 1 < r < 2,\ 2\pi/3 < \phi < 4\pi/3\}$. We get

$$\begin{aligned}\int_{U^{r_4}} \frac{\chi}{2\pi} \log\left|\frac{z - z(a)}{r_4}\right| \mu &< \frac{c_1}{\pi} \int_{2\pi/3}^{4\pi/3} \int_1^2 \log(r)\, r\, dr\, d\phi \\ &= \left(\tfrac{4}{3}\log 2 - \tfrac{1}{2}\right) c_1.\end{aligned}$$

In a similar way, we obtain

$$\begin{aligned}\int_{U^{r_4}} \frac{\chi}{2\pi} \log\left|\frac{z - z(a)}{r_4}\right| \mu &\ge -\frac{c_1}{2r_4^2}, \\ \int_{U^{r_4}} \frac{\chi}{2\pi} \log\left|\frac{\overline{z(a)}z}{r_4^2} - 1\right| \mu &< \left(\tfrac{4}{3}\log 2 - \tfrac{1}{2}\right)\frac{c_1}{r_4^2}, \\ \int_{U^{r_4}} \frac{\chi}{2\pi} \log\left|\frac{\overline{z(a)}z}{r_4^2} - 1\right| \mu &\ge -\frac{c_1}{4r_4^2}.\end{aligned}$$

The same bounds hold for b. Combining everything, we get

$$\left|\int_X \chi f_{a,b} \mu\right| \le \left(\tfrac{8}{3}\log 2 - \tfrac{1}{4}\right)\frac{c_1}{r_4^2}.$$

Finally, we have

$$\begin{aligned}\max_X \frac{\chi}{2\pi} \log\left|\frac{\overline{z(a)}z - r_4^2}{\overline{z(b)}z - r_4^2}\right| &\le \frac{1}{2\pi} \sup_{U^{r_4}} \log\left|\frac{r_4 - \overline{z(a)}z/r_4}{r_4 - \overline{z(b)}z/r_4}\right| \\ &\le \frac{1}{2\pi} \log\frac{r_4 + r_1}{r_4 - r_1},\end{aligned}$$

which finishes the proof. □

We will now apply Lemma A.4, which holds for any chart (U, z) satisfying the hypotheses (1) and (4) of Definition 3.1.1, to our atlas $\{(U_j, z_j) : 1 \le j \le n\}$. Besides including the index j in the notation for the coordinates, we denote by $l_a^{(j)}$ and $\chi^{(j)}$ the functions l_a and χ defined for the coordinate (U_j, z_j). We obtain the following generalisation of Lemma A.4 to the situation where a and b are arbitrary points of X:

Lemma A.5. *For all $a, b \in X$ and all j and k such that $a \in U_j^{r_1}$ and $b \in U_k^{r_1}$,*

$$\sup_X \left| g_{a,b} - l_a^{(j)} + l_b^{(k)} \right| \leq c_5(r_1, r_2, r_4, \lambda, n, c_1, M),$$

where

$$c_5(r_1, r_2, r_4, \lambda, c_1, n, M) = nc_4(r_1, r_2, r_4, \lambda, c_1) + \frac{n-1}{2\pi} \log\left(M \frac{r_4 + r_1}{r_2 - r_1}\right).$$

Proof. We first show that for any two coordinate indices j and k and for all $a \in U_k^{r_1} \cap U_j^{r_1}$,

$$\sup_X \left| l_a^{(k)} - l_a^{(j)} \right| \leq \frac{1}{2\pi} \log\left(M \frac{r_4 + r_1}{r_2 - r_1}\right). \tag{4}$$

To prove this, let $y \in X$. We distinguish three cases to prove that $l_a^{(k)}(y) - l_a^{(j)}(y)$ is bounded from above by the right-hand side of (4); the inequality then follows by interchanging j and k.

Case 1. Suppose $y \in U_j$ with $|z_j(y) - z_j(a)| < (r_2 - r_1)/M$. In this case, we have

$$|z_j(y)| < |z_j(a)| + \frac{r_2 - r_1}{M} < r_2;$$

hence, $a, y \in U_j^{r_2}$. Let $[a, y]^j$ denote the line segment between a and y in the z_j-coordinate, i.e., the curve in $U_j^{r_2}$ whose z_j-coordinate is parametrised by

$$\hat{z}_j(t) = (1-t) z_j(a) + t z_j(y) \quad \text{for } 0 \leq t \leq 1.$$

We claim that this line segment also lies inside $U_k^{r_2}$. Suppose this is not the case; then, because the "starting point" $z_j^{-1}(\hat{z}_j(0)) = a$ does lie in $U_k^{r_2}$, there exists a smallest $t \in (0, 1)$ for which the point

$$y' = z_j^{-1}(\hat{z}_j(t)) \in U_j^{r_2}$$

lies on the boundary of $U_k^{r_2}$. It follows from hypothesis (3) of Definition 3.1.1 that

$$|z_k(y') - z_k(a)| \leq M |z_j(y') - z_j(a)|.$$

On the other hand,

$$\begin{aligned} |z_j(y') - z_j(a)| &= t |z_j(y) - z_j(a)| \\ &< (r_2 - r_1)/M \end{aligned}$$

by assumption, and

$$|z_k(y') - z_k(a)| > r_2 - r_1$$

by the triangle inequality. This implies

$$|z_k(y') - z_k(a)| > M |z_j(y') - z_j(a)|,$$

a contradiction. Therefore, the line segment $[a, y]^j$ lies inside $U_j^{r_2} \cap U_k^{r_2}$. By hypothesis (3) of Definition 3.1.1, we have

$$|z_k(y) - z_k(a)| \leq M|z_j(y) - z_j(a)|.$$

Because $\chi^{(j)}(y) = \chi^{(k)}(y) = 1$, we find

$$l_a^{(k)}(y) - l_a^{(j)}(y) = \frac{1}{2\pi} \log \left| \frac{z_k(y) - z_k(a)}{z_j(y) - z_j(a)} \right| \leq \frac{1}{2\pi} \log M,$$

which is bounded by the right-hand side of (4).

Case 2. Suppose $y \notin U_j$. Then $l_a^{(j)}(y) = 0$, and thus,

$$l_a^{(k)}(y) - l_a^{(j)}(y) = l_a^{(k)}(y) \leq \frac{\log(r_4 + r_1)}{2\pi}.$$

Case 3. Suppose $y \in U_j$ and $|z_j(y) - z_j(a)| \geq (r_2 - r_1)/M$. Then

$$l_a^{(k)}(y) - l_a^{(j)}(y) \leq \frac{\log(r_4 + r_1)}{2\pi} - \frac{\chi^{(j)}(y)}{2\pi} \log \frac{r_2 - r_1}{M},$$

which is also bounded by the right-hand side in (4).

By hypothesis (2) of Definition 3.1.1, the open sets $U_j^{r_1}$ cover X. Furthermore, X is connected. For arbitrary $a, b \in X$ and indices j and k such that $a \in U_j^{r_1}$ and $b \in U_k^{r_1}$, we can therefore choose a finite sequence of indices $j = j_1, j_2, \ldots, j_m = k$ with $m \leq n$ and points $a = a_0, a_1, \ldots, a_m = b$ such that $a_i \in U_{j_i}^{r_1} \cap U_{j_{i+1}}^{r_1}$ for $1 \leq i \leq m - 1$. Using $g_{a,b} = \sum_{i=1}^m g_{a_{i-1},a_i}$, we get

$$\begin{aligned}\sup_X \left|g_{a,b} - l_a^{(j)} + l_b^{(k)}\right| &= \sup_X \left|\sum_{i=1}^{m} \bigl(g_{a_{i-1},a_i} - l_{a_{i-1}}^{(j_i)} + l_{a_i}^{(j_i)}\bigr) + \sum_{i=1}^{m-1} \bigl(l_{a_i}^{(j_{i+1})} - l_{a_i}^{(j_i)}\bigr)\right| \\ &\leq \sum_{i=1}^{m} \sup_X \left|g_{a_{i-1},a_i} - l_{a_{i-1}}^{(j_i)} + l_{a_i}^{(j_i)}\right| + \sum_{i=1}^{m-1} \sup_X \left|l_{a_i}^{(j_{i+1})} - l_{a_i}^{(j_i)}\right|.\end{aligned}$$

The lemma now follows from Lemma A.4 and the inequality (4). □

Proof of Theorem 3.1.2. We choose a continuous partition of unity $\{\phi^j\}_{j=1}^n$ subordinate to the covering $\{U_j^{r_1}\}_{j=1}^n$. Let $a \in X$, and let j be an index such that $a \in U_j^{r_1}$. By the definition of $g_{a,\mu}$, we have

$$\begin{aligned}g_{a,\mu}(x) - l_a^{(j)}(x) &= \int_{b \in X} g_{a,b}(x)\mu(b) - l_a^{(j)}(x) \\ &= \sum_{k=1}^{n} \int_{b \in U_k^{r_1}} \phi^k(b)\bigl(g_{a,b}(x) - l_a^{(j)}(x)\bigr)\mu(b)\end{aligned}$$

$$=\sum_{k=1}^{n}\int_{b\in U_k^{r_1}}\phi^k(b)\big(g_{a,b}(x)-l_a^{(j)}(x)+l_b^{(k)}(x)\big)\mu(b)-\sum_{k=1}^{n}\int_{b\in U_k^{r_1}}\phi^k(b)l_b^{(k)}(x)\mu(b).$$

In a similar way to in the proof of Lemma A.4, one can check that, for every index k and all $x\in X$, we have

$$-\frac{c_1}{2}\le\int_{b\in U_k^{r_1}}\phi^k(b)l_b^{(k)}(x)\mu(b)\le\big(\tfrac{4}{3}\log 2-\tfrac{1}{2}\big)c_1$$

so that

$$\sup_{x\in X}\left|\int_{b\in U_k^{r_1}}\phi^k(b)l_b^{(k)}(x)\mu(b)\right|\le\frac{c_1}{2}.$$

Together with Lemma A.5, this gives the inequality

$$\begin{aligned}\sup_X\big|g_{a,\mu}-l_a^{(j)}\big|&\le c_5(r_1,r_2,r_4,\lambda,c_1,n,M)\sum_{j=1}^{n}\int_{b\in U_j^{r_1}}\phi^j(b)\mu(b)+\sum_{j=1}^{n}\frac{c_1}{2}\\&=c_5(r_1,r_2,r_4,\lambda,c_1,n,M)+\frac{nc_1}{2}.\end{aligned}$$

We also have

$$\sup_X g_{a,\mu}\le\sup_X\big(g_{a,\mu}-l_a^{(j)}\big)+\sup_X l_a^{(j)}\le\sup_X\big(g_{a,\mu}-l_a^{(j)}\big)+\frac{\log(r_4+r_1)}{2\pi}.$$

By varying the choice of r_4 and $\tilde{\chi}$, we can let r_4 tend to 1 and λ to $\dfrac{1}{1-r_2}$. This leads to

$$\begin{aligned}c_3\Big(r_1,r_2,1,\frac{1}{1-r_2}\Big)=4\sqrt{\frac{1+r_2}{1-r_2}}\Big(\frac{1}{2(1-r_2)}\log\frac{(r_1+1)^2}{(r_2-r_1)(1-r_1)}+\frac{1}{r_2-r_1}+\frac{r_1}{1-r_1}\Big)\\+\frac{2}{\pi}\log\frac{(r_1+1)^2}{(r_2-r_1)(1-r_1)},\end{aligned}$$

which implies successively

$$\begin{aligned}c_4\Big(r_1,r_2,1,\frac{1}{1-r_2},c_1\Big)&=c_3\Big(r_1,r_2,1,\frac{1}{1-r_2}\Big)+\frac{1}{2\pi}\log\frac{1+r_1}{1-r_1}+\big(\tfrac{8}{3}\log 2-\tfrac{1}{4}\big)c_1,\\c_5&=nc_4\Big(r_1,r_2,r_4,\frac{1}{1-r_2},c_1\Big)+\frac{n-1}{2\pi}\log\Big(M\frac{1+r_1}{r_2-r_1}\Big).\end{aligned}$$

We take $r_2=0.39+0.61r_1$. Then, for $r_1>\frac{1}{2}$, one can check numerically that

$$c_5\le 52.4\frac{n}{(1-r_1)^{3/2}}\log\frac{1}{1-r_1}+1.60nc_1+\frac{n-1}{2\pi}\log M.$$

From this, the theorem follows. □

Acknowledgements

I thank Peter Bruin, Bas Edixhoven and Robin de Jong. They introduced me to Arakelov theory and Merkl's theorem, and I am grateful to them for many inspiring discussions and their help in writing this article. I also thank Rafael von Känel and Jan Steffen Müller for motivating discussions about this article. I thank Jean-Benoît Bost and Gerard Freixas for discussions on Arakelov geometry, Yuri Bilu for inspiring discussions on Riemann's existence theorem, Jürg Kramer for discussions on Faltings' delta invariant, Hendrik Lenstra and Bart de Smit for their help in proving Proposition 4.1.1, Qing Liu for answering my questions on models of finite morphisms of curves and Karl Schwede for helpful discussions about the geometry of surfaces.

References

[Abbes and Ullmo 1997] A. Abbes and E. Ullmo, "Auto-intersection du dualisant relatif des courbes modulaires $X_0(N)$", *J. Reine Angew. Math.* **484** (1997), 1–70. MR 99e:11077 Zbl 0934.14016

[Arakelov 1974] S. J. Arakelov, "Intersection theory for divisors on an arithmetic surface", *Izv. Akad. Nauk SSSR Ser. Mat.* **38** (1974), 1179–1192. In Russian; translated in *Math. USSR, Izv.* **8**:6 (1974), 1167–1180. MR 57 #12505 Zbl 0355.14002

[Belyi 1979] G. V. Belyi, "On Galois extensions of a maximal cyclotomic field", *Izv. Akad. Nauk SSSR Ser. Mat.* **43**:2 (1979), 267–276. In Russian; translated in *Math. USSR, Izv.* **14**:2 (1980), 247–256. MR 80f:12008 Zbl 0429.12004

[Bilu and Strambi 2010] Y. F. Bilu and M. Strambi, "Quantitative Riemann existence theorem over a number field", *Acta Arith.* **145**:4 (2010), 319–339. MR 2011m:14045 Zbl 1222.11082

[Bruin 2010] P. Bruin, *Modular curves, Arakelov theory, algorithmic applications*, Ph.D. thesis, Universiteit Leiden, 2010, http://www.math.leidenuniv.nl/scripties/proefschrift-peterbruin.pdf.

[Bruin 2011] P. Bruin, "Computing coefficients of modular forms", pp. 19–36 in *Actes de la Conférence "Théorie des Nombres et Applications"*, Presses Univ. Franche-Comté, Besançon, 2011. MR 2894266 Zbl 1267.11036

[Bruin 2013] P. Bruin, "Explicit bounds on automorphic and canonical Green functions of Fuchsian groups", preprint, 2013, http://homepages.warwick.ac.uk/staff/P.Bruin/green.pdf. To appear in *Mathematika*.

[Clark and Voight 2011] P. L. Clark and J. Voight, "Algebraic curves uniformized by congruence subgroups of triangle groups", preprint, 2011, http://math.uga.edu/~pete/triangle-072011.pdf.

[Curilla and Kühn 2009] C. Curilla and U. Kühn, "On the arithmetic self-intersection numbers of the dualizing sheaf for Fermat curves of prime exponent", preprint, 2009. arXiv 0906.3891v1

[David 1991] S. David, "Fonctions thêta et points de torsion des variétés abéliennes", *Compositio Math.* **78**:2 (1991), 121–160. MR 92d:11061 Zbl 0741.14025

[Edixhoven and de Jong 2011a] B. Edixhoven and R. de Jong, "Bounds for Arakelov invariants of modular curves", pp. 217–256 in *Computational aspects of modular forms and Galois representations*, edited by B. Edixhoven and J.-M. Couveignes, Ann. of Math. Stud. **176**, Princeton Univ. Press, 2011. MR 2857095 Zbl 1216.11004

[Edixhoven and de Jong 2011b] B. Edixhoven and R. de Jong, "Short introduction to heights and Arakelov theory", pp. 79–94 in *Computational aspects of modular forms and Galois representations*, edited by B. Edixhoven and J.-M. Couveignes, Ann. of Math. Stud. **176**, Princeton Univ. Press, 2011. MR 2867563 Zbl 1216.11004

[Edixhoven et al. 2010] B. Edixhoven, R. de Jong, and J. Schepers, "Covers of surfaces with fixed branch locus", *Internat. J. Math.* **21**:7 (2010), 859–874. MR 2011i:14032 Zbl 1203.14021

[Faltings 1983] G. Faltings, "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern", *Invent. Math.* **73**:3 (1983), 349–366. MR 85g:11026a Zbl 0588.14026

[Faltings 1984] G. Faltings, "Calculus on arithmetic surfaces", *Ann. of Math.* (2) **119**:2 (1984), 387–424. MR 86e:14009 Zbl 0559.14005

[Gaudron and Rémond 2011] E. Gaudron and G. Rémond, "Théorème des periodes et degrés minimaux d'isogénies", preprint, 2011. To appear in *Comment. Math. Helv.* arXiv 1105.1230v1

[Graftieaux 2001] P. Graftieaux, "Formal groups and the isogeny theorem", *Duke Math. J.* **106**:1 (2001), 81–121. MR 2002f:14055 Zbl 1064.14045

[Igusa 1972] J.-i. Igusa, *Theta functions*, Die Grundlehren der mathematischen Wissenschaften **194**, Springer, New York, 1972. MR 48 #3972 Zbl 0251.14016

[Javanpeykar and von Känel 2013] A. Javanpeykar and R. von Känel, "Szpiro's small points conjecture for cyclic covers", preprint, 2013. arXiv 1311.0043v1

[de Jong 2005a] R. de Jong, "Arakelov invariants of Riemann surfaces", *Doc. Math.* **10** (2005), 311–329. MR 2006j:14030 Zbl 1087.14023

[de Jong 2005b] R. de Jong, "On the Arakelov theory of elliptic curves", *Enseign. Math.* (2) **51**:3-4 (2005), 179–201. MR 2007b:14047 Zbl 1115.14014

[Jorgenson and Kramer 2004] J. Jorgenson and J. Kramer, "Bounding the sup-norm of automorphic forms", *Geom. Funct. Anal.* **14**:6 (2004), 1267–1277. MR 2005m:11071 Zbl 1078.11027

[Jorgenson and Kramer 2006] J. Jorgenson and J. Kramer, "Bounds on canonical Green's functions", *Compos. Math.* **142**:3 (2006), 679–700. MR 2007h:14030 Zbl 1105.14028

[Jorgenson and Kramer 2009] J. Jorgenson and J. Kramer, "Bounds on Faltings's delta function through covers", *Ann. of Math.* (2) **170**:1 (2009), 1–43. MR 2010g:14031 Zbl 1169.14020

[Khadjavi 2002] L. S. Khadjavi, "An effective version of Belyi's theorem", *J. Number Theor.* **96**:1 (2002), 22–47. MR 2003h:11072 Zbl 1078.11046

[Kühn 2013] U. Kühn, "On the arithmetic self-intersection number of the dualizing sheaf on arithmetic surfaces", preprint, 2013. arXiv 0906.2056v2

[Liţcanu 2004] R. Liţcanu, "Propriétés du degré des morphismes de Belyi", *Monatsh. Math.* **142**:4 (2004), 327–340. MR 2005i:11082 Zbl 1078.14034

[Liu 2006a] Q. Liu, *Algebraic geometry and arithmetic curves*, 2nd ed., Oxford Graduate Texts in Mathematics **6**, Oxford University Press, Oxford, 2006. MR 2003g:14001 Zbl 1103.14001

[Liu 2006b] Q. Liu, "Stable reduction of finite covers of curves", *Compos. Math.* **142**:1 (2006), 101–118. MR 2007k:14057 Zbl 1108.14020

[Long 2008] L. Long, "Finite index subgroups of the modular group and their modular forms", pp. 83–102 in *Modular forms and string duality*, edited by N. Yui et al., Fields Inst. Commun. **54**, Amer. Math. Soc., Providence, RI, 2008. MR 2009k:11069 Zbl 1163.11032

[Mayer 2012] H. Mayer, *Self-intersection of the dualizing sheaf of modular curves $X_1(n)$*, Ph.D. thesis, Humboldt-Universität zu Berlin, 2012. arXiv 1212.1294v1

[Merkl 2011] F. Merkl, "An upper bound for Green functions on Riemann surfaces", pp. 203–215 in *Computational aspects of modular forms and Galois representations*, edited by B. Edixhoven and J.-M. Couveignes, Ann. of Math. Stud. **176**, Princeton Univ. Press, 2011. MR 2857094 Zbl 1216.11004

[Michel and Ullmo 1998] P. Michel and E. Ullmo, "Points de petite hauteur sur les courbes modulaires $X_0(N)$", *Invent. Math.* **131**:3 (1998), 645–674. MR 99c:11074 Zbl 0991.11037

[Moret-Bailly 1989] L. Moret-Bailly, "La formule de Noether pour les surfaces arithmétiques", *Invent. Math.* **98**:3 (1989), 491–498. MR 91h:14023 Zbl 0727.14014

[Moret-Bailly 1990] L. Moret-Bailly, "Hauteurs et classes de Chern sur les surfaces arithmétiques", pp. 37–58 in *Séminaire sur les Pinceaux de Courbes Elliptiques* (Paris, 1988), Astérisque **183**, Soc. Math. de France, Paris, 1990. MR 92g:14018a Zbl 0727.14015

[Pazuki 2012] F. Pazuki, "Theta height and Faltings height", *Bull. Soc. Math. France* **140**:1 (2012), 19–49. MR 2903770 Zbl 1245.14029

[Rémond 1999] G. Rémond, "Hauteurs thêta et construction de Kodaira", *J. Number Theory* **78**:2 (1999), 287–311. MR 2000g:11059 Zbl 0947.14016

[Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, New York, 1979. MR 82e:12016 Zbl 0423.12016

[Szpiro 1985a] L. Szpiro, "La conjecture de Mordell (d'après G. Faltings)", pp. 83–103 in *Séminaire Bourbaki* (Paris, 1983–1984), Astérisque **121–122**, Soc. Math. de France, Paris, 1985. MR 87c:11033 Zbl 0591.14027

[Szpiro 1985b] L. Szpiro (editor), *Séminaire sur les pinceaux arithmétiques: la conjecture de Mordell* (Paris, 1983–1984), Astérisque **127**, Soc. Math. de France, Paris, 1985. MR 87h:14017 Zbl 0588.14028

[Szpiro 1985c] L. Szpiro, "Un peu d'effectivité", pp. 275–287 in *Séminaire sur les pinceaux arithmétiques: La conjecture de Mordell* (Paris, 1983–1984), Astérisque **127**, Soc. Math. de France, Paris, 1985. MR 801928 Zbl 1182.11016

[Szpiro 1986] L. Szpiro, "Small points and torsion points", pp. 251–260 in *The Lefschetz centennial conference, I* (Mexico City, 1984), edited by D. Sundararaman, Contemp. Math. **58**, Amer. Math. Soc., Providence, RI, 1986. MR 87k:14029

[Szpiro 1990a] L. Szpiro, "Discriminant et conducteur des courbes elliptiques", pp. 7–18 in *Séminaire sur les Pinceaux de Courbes Elliptiques* (Paris, 1988), Astérisque **183**, Soc. Math. de France, Paris, 1990. MR 91g:11059 Zbl 0742.14026

[Szpiro 1990b] L. Szpiro, "Sur les propriétés numériques du dualisant relatif d'une surface arithmétique", pp. 229–246 in *The Grothendieck Festschrift*, vol. III, edited by P. Cartier et al., Progr. Math. **88**, Birkhäuser, Boston, MA, 1990. MR 92c:14017 Zbl 0759.14018

[Ullmo 2000] E. Ullmo, "Hauteur de Faltings de quotients de $J_0(N)$, discriminants d'algèbres de Hecke et congruences entre formes modulaires", *Am. J. Math.* **122** (2000), 83–115. MR 2000k:11080 Zbl 0991.11033

[Wolfart 1997] J. Wolfart, "The 'obvious' part of Belyi's theorem and Riemann surfaces with many automorphisms", pp. 97–112 in *Geometric Galois actions, 1*, edited by L. Schneps and P. Lochak, London Math. Soc. Lecture Note Ser. **242**, Cambridge Univ. Press, 1997. MR 99a:14036 Zbl 0915.14021

[Zograf 1991] P. Zograf, "A spectral proof of Rademacher's conjecture for congruence subgroups of the modular group", *J. Reine Angew. Math.* **414** (1991), 113–116. MR 92d:11041 Zbl 0709.11031

Communicated by Joseph Silverman
Received 2012-06-22 Revised 2013-02-27 Accepted 2013-04-17

ajavanp@math.leidenuniv.nl *Mathematisch Instituut, Universiteit Leiden, 2717 GA Leiden, Netherlands*

peter.bruin@math.uzh.ch *Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*

ALGEBRA AND NUMBER THEORY 8:1 (2014)
dx.doi.org/10.2140/ant.2014.8.141

The radius of a subcategory of modules

Hailong Dao and Ryo Takahashi

Dedicated to Professor Craig Huneke on the occasion of his sixtieth birthday

We introduce a new invariant for subcategories $\mathscr{X}$ of finitely generated modules over a local ring R which we call the radius of $\mathscr{X}$. We show that if R is a complete intersection and $\mathscr{X}$ is resolving, then finiteness of the radius forces $\mathscr{X}$ to contain only maximal Cohen–Macaulay modules. We also show that the category of maximal Cohen–Macaulay modules has finite radius when R is a Cohen–Macaulay complete local ring with perfect coefficient field. We link the radius to many well-studied notions such as the dimension of the stable category of maximal Cohen–Macaulay modules, finite/countable Cohen–Macaulay representation type and the uniform Auslander condition.

Introduction

Let R be a commutative Noetherian local ring and $\operatorname{mod} R$ the category of finitely generated modules over R. In this paper we introduce and study a new invariant for subcategories $\mathscr{X}$ of $\operatorname{mod} R$ which we call the radius of $\mathscr{X}$. Roughly speaking, it is defined as the least number of extensions necessary to build the whole objects in $\mathscr{X}$ out of a single object in $\operatorname{mod} R$. (For the precise definition, see Definition 2.3 in this paper.) Our definition is inspired by the notion of dimension of triangulated categories that was introduced by Rouquier [2008].

We obtain strong evidences that the concept of radius is intimately linked to both the representation theory and the singularity of R. For example, over a Gorenstein complete local ring R, the category of maximal Cohen–Macaulay modules has radius zero if and only if R has finite Cohen–Macaulay representation type, in other words, R is a simple hypersurface singularity (when R has an algebraically closed coefficient field of characteristic zero). In addition, the category of maximal Cohen–Macaulay modules over a complete local hypersurface (over an algebraically closed

Dao was partially supported by NSF grant DMS 0834050. Takahashi was partially supported by JSPS Grant-in-Aid for Young Scientists (B) 22740008 and by JSPS Postdoctoral Fellowships for Research Abroad.

MSC2010: primary 13C60; secondary 13C14, 16G60, 18E30.

Keywords: radius of subcategory, resolving subcategory, thick subcategory, Cohen–Macaulay module, complete intersection, dimension of triangulated category, Cohen–Macaulay representation type.

field of characteristic not two) of countable Cohen–Macaulay representation type has radius one. We also observe a tantalizing connection to the uniform Auslander condition, which has attracted researchers over the years.

Perhaps most surprisingly, one corollary of our first main result (Theorem 3.3) states:

Theorem I. *Let R be a local complete intersection, and let $\mathscr{X}$ be a resolving subcategory of* $\operatorname{mod} R$. *If the radius of $\mathscr{X}$ is finite, then $\mathscr{X}$ contains only maximal Cohen–Macaulay modules.*

We conjecture that the above result holds for all Cohen–Macaulay local rings. Our second main result below supports this conjecture, which follows from more general results (Theorems 5.7 and 5.11).

Theorem II. *Let R be a Cohen–Macaulay complete local ring with perfect coefficient field. Then the category of maximal Cohen–Macaulay modules over R has finite radius.*

The structure of the paper is as follows. In Section 1 we set the basic notations and definitions. Section 2 contains our key definition (Definition 2.1) of the radius of a subcategory of $\operatorname{mod} R$, as well as some detailed comparisons to similar notions. We also give several results connecting the radius to the singularities of finite and countable Cohen–Macaulay representation type. Sections 3 and 4 consist of the statement and proof of our Theorem I, respectively. We also discuss here thickness of resolving subcategories of maximal Cohen–Macaulay modules over a complete intersection. Section 5 contains the proof of (generalizations of) our Theorem II. Section 6 connects the main results to the uniform Auslander condition and discusses some open questions.

1. Preliminaries

In this section, we recall the definitions of a resolving subcategory, totally reflexive modules and a thick subcategory.

Convention 1.1. Throughout this paper, we assume all rings are commutative Noetherian rings with identity. All modules are finitely generated. All subcategories are full and strict. (Recall that a subcategory $\mathscr{X}$ of a category $\mathscr{C}$ is called *strict* provided that for objects $M, N \in \mathscr{C}$ with $M \cong N$, if M is in $\mathscr{X}$, then so is N.) Hence, the *subcategory* of a category $\mathscr{C}$ consisting of objects $\{M_\lambda\}_{\lambda\in\Lambda}$ always means the smallest strict full subcategory of $\mathscr{C}$ to which M_λ belongs for all $\lambda \in \Lambda$. Note that this coincides with the full subcategory of $\mathscr{C}$ consisting of all objects $X \in \mathscr{C}$ such that $X \cong M_\lambda$ for some $\lambda \in \Lambda$. Let R be a (commutative Noetherian) ring. Denote by $\operatorname{mod} R$ the category of (finitely generated) R-modules and R-homomorphisms. For a Cohen–Macaulay local ring R, we call a maximal Cohen–Macaulay R-module just a Cohen–Macaulay R-module. We denote by $\mathrm{CM}(R)$ the subcategory of $\operatorname{mod} R$ consisting of Cohen–Macaulay R-modules.

The following notation is used throughout this paper.

Notation 1.2. For a subcategory $\mathscr{X}$ of $\operatorname{mod} R$, we denote by $\operatorname{add} \mathscr{X}$ (or $\operatorname{add}_R \mathscr{X}$) the *additive closure* of $\mathscr{X}$, namely, the subcategory of $\operatorname{mod} R$ consisting of direct summands of finite direct sums of modules in $\mathscr{X}$. When $\mathscr{X}$ consists of a single module M, we simply denote it by $\operatorname{add} M$ (or $\operatorname{add}_R M$). For an R-module M, we denote by M^* the R-dual module $\operatorname{Hom}_R(M, R)$. For a homomorphism $f : M \to N$ of R-modules, f^* denotes the R-dual homomophism $N^* \to M^*$ sending $\sigma \in N^*$ to the composition $\sigma \cdot f \in M^*$.

The notion of a resolving subcategory has been introduced by Auslander and Bridger [1969]. It can actually be defined for an arbitrary abelian category with enough projective object. The only resolving subcategories we deal with in this paper are ones of $\operatorname{mod} R$.

Definition 1.3. A subcategory $\mathscr{X}$ of $\operatorname{mod} R$ is called *resolving* if the following hold.

(R1) $\mathscr{X}$ contains the projective R-modules.

(R2) $\mathscr{X}$ is closed under direct summands: if M is an R-module in $\mathscr{X}$ and N is an R-module that is a direct summand of M, then N is also in $\mathscr{X}$.

(R3) $\mathscr{X}$ is closed under extensions: for an exact sequence $0 \to L \to M \to N \to 0$ of R-modules, if L, N are in $\mathscr{X}$, then so is M.

(R4) $\mathscr{X}$ is closed under kernels of epimorphisms: for an exact sequence $0 \to L \to M \to N \to 0$ of R-modules, if M, N are in $\mathscr{X}$, then so is L.

A resolving subcategory is a subcategory such that any two minimal resolutions of a module by modules in it have the same length; see [Auslander and Bridger 1969, Lemma (3.12)]. Note that one can replace the condition (R1) with:

(R1′) $\mathscr{X}$ contains R.

Next we recall the notion of a totally reflexive module.

Definition 1.4. An R-module M is called *totally reflexive* if the natural homomorphism $M \to M^{**}$ is an isomorphism and $\operatorname{Ext}_R^i(M, R) = 0 = \operatorname{Ext}_R^i(M^*, R)$ for all $i > 0$. We denote by $\mathscr{G}(R)$ the subcategory of $\operatorname{mod} R$ consisting of totally reflexive modules.

A totally reflexive module was defined by Auslander [1967], and deeply studied by Auslander and Bridger [1969]. The R-dual of a totally reflexive R-module is also totally reflexive. Every projective module is totally reflexive, i.e., $\operatorname{add} R \subseteq \mathscr{G}(R)$. If R is a Cohen–Macaulay local ring, then every totally reflexive R-module is Cohen–Macaulay, i.e., $\mathscr{G}(R) \subseteq \mathrm{CM}(R)$. When R is a Gorenstein local ring, an R-module is totally reflexive if and only if it is Cohen–Macaulay, i.e., $\mathscr{G}(R) = \mathrm{CM}(R)$. For more details, see [Auslander and Bridger 1969] and [Christensen 2000].

Syzygies, cosyzygies and transposes are key tools in this paper. We recall here their precise definitions.

Definition 1.5. Let $(R, \mathfrak{m})$ be a local ring, and let M be an R-module.

(1) Take a minimal free resolution $\cdots \xrightarrow{\delta_{n+1}} F_n \xrightarrow{\delta_n} F_{n-1} \xrightarrow{\delta_{n-1}} \cdots \xrightarrow{\delta_1} F_0 \to M \to 0$ of M. Then, for each $n \geq 1$, the image of δ_n is called the *n-th syzygy* of M and denoted by $\Omega^n M$ (or $\Omega^n_R M$). For convention, we set $\Omega^0 M = M$.

(2) The cokernel of the R-dual map $\delta_1^* : F_0^* \to F_1^*$ is called the (*Auslander*) *transpose* of M and denoted by $\operatorname{Tr} M$ (or $\operatorname{Tr}_R M$).

(3) Let $0 \to M \to F_{-1} \xrightarrow{\delta_{-1}} \cdots \xrightarrow{\delta_{-(n-1)}} F_{-n} \xrightarrow{\delta_{-n}} F_{-(n+1)} \xrightarrow{\delta_{-(n+1)}} \cdots$ be a *minimal free coresolution* of M, that is, an exact sequence with F_{-n} free and $\operatorname{Im} \delta_{-n} \subseteq \mathfrak{m} F_{-(n+1)}$ for all $n \geq 1$. Then we call the image of δ_{-n} the *n-th cosyzygy* of M and denote it by $\Omega^{-n} M$ (or $\Omega^{-n}_R M$).

Let R be a local ring. Then by [Yoshino 2005, Lemma 3.2] one can replace (R4) with:

(R4′) $\mathscr{X}$ is closed under syzygies: if M is in $\mathscr{X}$, then so is ΩM.

Totally reflexive modules behave well under taking their syzygies, cosyzygies and transposes. Let R be a local ring. Let M be a totally reflexive R-module. The R-dual of a minimal free resolution (respectively, coresolution) of M is a minimal free coresolution (respectively, resolution) of M^*. In particular, a minimal free coresolution of M always exists, and it is uniquely determined up to isomorphism. The n-th syzygy $\Omega^n M$ and cosyzygy $\Omega^{-n} M$ are again totally reflexive for all n. This is an easy consequence of [Christensen 2000, (1.2.9) and (1.4.8)]. The transpose $\operatorname{Tr} M$ is also totally reflexive; see [Auslander and Bridger 1969, Proposition (3.8)]. For an R-module M, the n-th syzygy $\Omega^n M$ for any $n \geq 1$ and the transpose $\operatorname{Tr} M$ are uniquely determined up to isomorphism, since so is a minimal free resolution of M. If M is totally reflexive, then the n-th cosyzygy $\Omega^{-n} M$ for any $n \geq 1$ is also uniquely determined up to isomorphism, since so is a minimal free coresolution of M.

A lot of subcategories of $\operatorname{mod} R$ are known to be resolving. For example, $\mathrm{CM}(R)$ is a resolving subcategory of $\operatorname{mod} R$ if R is Cohen–Macaulay. The subcategory of $\operatorname{mod} R$ consisting of totally reflexive R-modules is resolving by [Auslander and Bridger 1969, (3.11)]. One can construct a resolving subcategory easily by using the vanishing of Tor or Ext. Also, the modules of complexity less than a fixed integer form a resolving subcategory of $\operatorname{mod} R$. For the details, we refer to [Takahashi 2009, Example 2.4].

Now we define a thick subcategory of totally reflexive modules.

Definition 1.6. A subcategory $\mathcal{X}$ of $\mathcal{G}(R)$ is called *thick* if it is closed under direct summands and short exact sequences: for an exact sequence $0 \to L \to M \to N \to 0$ of totally reflexive R-modules, if two of L, M, N are in $\mathcal{X}$, then so is the third.

A typical example of a thick subcategory is obtained by restricting a resolving subcategory to $\mathcal{G}(R)$.

The following proposition is shown by an argument dual to [Yoshino 2005, Lemma 3.2].

Proposition 1.7. *Let R be a local ring. Let $\mathcal{X}$ be a subcategory of $\mathcal{G}(R)$ containing R. Then $\mathcal{X}$ is a thick subcategory of $\mathcal{G}(R)$ if and only if $\mathcal{X}$ is a resolving subcategory of* $\operatorname{mod} R$ *and is closed under cosyzygies: if M is in $\mathcal{X}$, then so is $\Omega^{-1}M$.*

Let $(R, \mathfrak{m})$ be a local ring. We call R a *hypersurface* if the $\mathfrak{m}$-adic completion $\widehat{R}$ of R is a residue ring of a complete regular local ring by a principal ideal. We say that R is a *complete intersection* if $\widehat{R}$ is a residue ring of a complete regular local ring by an ideal generated by a regular sequence.

We recall the definitions of *Gorenstein dimension* and *complete intersection dimension*, which are abbreviated to G-dimension and CI-dimension. These notions have been introduced by Auslander and Bridger [1969] and Avramov, Gasharov and Peeva [Avramov et al. 1997], respectively.

Definition 1.8. Let R be a local ring, and let M be an R-module. The *G-dimension* of M, denoted $\mathsf{Gdim}_R M$, is defined as the infimum of the lengths of totally reflexive resolutions of M, namely, exact sequences of the form $0 \to X_n \to X_{n-1} \to \cdots \to X_1 \to X_0 \to M \to 0$ with each X_i being totally reflexive. The *CI-dimension* of M is defined as the infimum of $\mathsf{pd}_S(M \otimes_R R') - \mathsf{pd}_S R'$, where $R \to R' \leftarrow S$ runs over the quasi-deformations of R. Here, a diagram $R \xrightarrow{f} R' \xleftarrow{g} S$ of homomorphisms of local rings is called a quasi-deformation of R if f is faithfully flat and g is a surjection whose kernel is generated by an S-sequence.

Recall that M is said to have *complexity* c, denoted by $\mathsf{cx}_R M = c$, if c is the least nonnegative integer n such that there exists a real number r satisfying the inequality $\beta_i^R(M) \le ri^{n-1}$ for all $i \gg 0$.

Remark 1.9. For a local ring $(R, \mathfrak{m}, k)$ and a module M over R, the following are known to hold. For the proofs, we refer to [Christensen 2000] and [Avramov et al. 1997].

(1) $\mathsf{Gdim}_R M = \infty$ if and only if M does not admit a totally reflexive resolution of finite length.

(2) $\mathsf{CIdim}_R M = \infty$ if and only if $\mathsf{pd}_S(M \otimes_R R') = \infty$ for every quasi-deformation $R \to R' \leftarrow S$.

(3) One has $M = 0 \Leftrightarrow \mathsf{Gdim}_R M = -\infty \Leftrightarrow \mathsf{CIdim}_R M = -\infty$.

(4) $\mathsf{Gdim}_R\, M \le 0$ if and only if M is totally reflexive.

(5) If $\mathsf{Gdim}_R\, M$ (respectively, $\mathsf{CIdim}_R\, M$) is finite, it is equal to $\mathsf{depth}\ R - \mathsf{depth}_R\, M$.

(6) The inequalities $\mathsf{Gdim}_R\, M \le \mathsf{CIdim}_R\, M \le \mathsf{pd}_R\, M$ hold, and equalities hold to the left of any finite dimension.

(7) If $M \ne 0$, then $\mathsf{Gdim}_R(\Omega^n M) = \sup\{\mathsf{Gdim}_R\, M - n, 0\}$ and $\mathsf{CIdim}_R(\Omega^n M) = \sup\{\mathsf{CIdim}_R\, M - n, 0\}$ hold for all $n \ge 0$.

(8) If R is a Gorenstein ring (respectively, a complete intersection), then $\mathsf{Gdim}_R\, M$ (respectively, $\mathsf{CIdim}_R\, M$) is finite. If $\mathsf{Gdim}_R\, k$ (respectively, $\mathsf{CIdim}_R\, k$) is finite, then R is a Gorenstein ring (respectively, a complete intersection).

(9) If $\mathsf{CIdim}_R\, M < \infty$, then $\mathsf{cx}_R\, M < \infty$.

2. Definition of the radius of a subcategory

This section contains the key definition and establishes several results. More precisely, we will give the definition of the radius of a subcategory of mod R for a local ring R, and compare it with other notions, such as the dimension of a triangulated category defined by Rouquier. We will also explore its relationships with representation types of a Cohen–Macaulay local ring.

Definition 2.1. Let R be a local ring.

(1) For a subcategory $\mathcal{X}$ of mod R we denote by $[\mathcal{X}]$ the additive closure of the subcategory of mod R consisting of R and all modules of the form $\Omega^i X$, where $i \ge 0$ and $X \in \mathcal{X}$. When $\mathcal{X}$ consists of a single module X, we simply denote it by $[X]$.

(2) For subcategories $\mathcal{X}, \mathcal{Y}$ of mod R we denote by $\mathcal{X} \circ \mathcal{Y}$ the subcategory of mod R consisting of the R-modules M which fits into an exact sequence $0 \to X \to M \to Y \to 0$ with $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. We set $\mathcal{X} \bullet \mathcal{Y} = [[\mathcal{X}] \circ [\mathcal{Y}]]$.

(3) Let $\mathcal{C}$ be a subcategory of mod R. We define the *ball of radius r centered at* $\mathcal{C}$ as

$$[\mathcal{C}]_r = \begin{cases} [\mathcal{C}] & \text{if } r = 1, \\ [\mathcal{C}]_{r-1} \bullet \mathcal{C} = [[\mathcal{C}]_{r-1} \circ [\mathcal{C}]] & \text{if } r \ge 2. \end{cases}$$

If $\mathcal{C}$ consists of a single module C, then we simply denote $[\mathcal{C}]_r$ by $[C]_r$, and call it the ball of radius r centered at C. We write $[\mathcal{C}]_r^R$ when we should specify that mod R is the ground category where the ball is defined.

Some similar notions have already been introduced. Takahashi [2009, Definition 3.1] defines the subcategory $\mathrm{res}^n\, \mathcal{X}$ of the resolving closure $\mathrm{res}\, \mathcal{X}$ of a given subcategory $\mathcal{X}$ of mod R. This is different from ours in that $\mathrm{res}^n\, \mathcal{X}$ is not closed under syzygies. In [Avramov et al. 2010] the *thickening* $\mathrm{thick}^n\, \mathcal{X}$ of a given subcategory $\mathcal{X}$ of a triangulated category is defined. This cannot be applied directly to a module category.

Proposition 2.2. *Let R be a local ring.*

(1) *Let $\mathcal{X}, \mathcal{Y}$ be subcategories of* $\operatorname{mod} R$. *The following are equivalent for an R-module M:*

  (a) *M belongs to $\mathcal{X} \bullet \mathcal{Y}$.*

  (b) *There exists an exact sequence $0 \to X \to Z \to Y \to 0$ of R-modules with $X \in [\mathcal{X}]$ and $Y \in [\mathcal{Y}]$ such that M is a direct summand of Z.*

(2) *For subcategories $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ of* $\operatorname{mod} R$, *one has $(\mathcal{X} \bullet \mathcal{Y}) \bullet \mathcal{Z} = \mathcal{X} \bullet (\mathcal{Y} \bullet \mathcal{Z})$.*

(3) *Let $\mathcal{C}$ be a subcategory of* $\operatorname{mod} R$, *and let a, b be positive integers. Then one has $[\mathcal{C}]_a \bullet [\mathcal{C}]_b = [\mathcal{C}]_{a+b} = [\mathcal{C}]_b \bullet [\mathcal{C}]_a$.*

Proof. (1) The implication (b) $\Rightarrow$ (a) is obvious. To prove the opposite implication (a) $\Rightarrow$ (b), let M be an R-module in $\mathcal{X} \bullet \mathcal{Y} = [[\mathcal{X}] \circ [\mathcal{Y}]]$. By definition, M is isomorphic to a direct summand of $R^{\oplus p} \oplus \bigoplus_{i=0}^{n} (\Omega^i Z_i)^{\oplus q_i}$, where $p, q_i \geq 0$ and $Z_i \in [\mathcal{X}] \circ [\mathcal{Y}]$. For each $0 \leq i \leq n$ there is an exact sequence $0 \to X_i \to Z_i \to Y_i \to 0$ with $X_i \in [\mathcal{X}]$ and $Y_i \in [\mathcal{Y}]$. Taking syzygies and direct sums, we have an exact sequence

$$0 \to R^{\oplus p} \oplus \bigoplus_{i=0}^{n} (\Omega^i X_i)^{\oplus q_i} \to R^{\oplus p} \oplus \bigoplus_{i=0}^{n} (\Omega^i Z_i)^{\oplus q_i} \oplus R^{\oplus r} \to \bigoplus_{i=0}^{n} (\Omega^i Y_i)^{\oplus q_i} \to 0.$$

The left and right terms are in $[\mathcal{X}]$ and $[\mathcal{Y}]$, respectively. The middle term contains an R-module isomorphic to M. Thus the statement (b) follows.

(2) First, let M be an R-module in $(\mathcal{X} \bullet \mathcal{Y}) \bullet \mathcal{Z}$. By the assertion (1) there is an exact sequence $0 \to W \xrightarrow{f} V \to Z \to 0$ with $W \in \mathcal{X} \bullet \mathcal{Y}$ and $Z \in [\mathcal{Z}]$ such that M is a direct summand of V. By (1) again, we have an exact sequence $0 \to X \to U \to Y \to 0$ with $X \in [\mathcal{X}]$ and $Y \in [\mathcal{Y}]$ such that W is a direct summand of U. Writing $U = W \oplus W'$, we make the following pushout diagram.

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & & \\
 & X \arrow[r, equal] \arrow[d] & X \arrow[d] & & \\
0 \arrow[r] & W \oplus W' \arrow[r, "\left(\begin{smallmatrix} f & 0 \\ 0 & 1 \end{smallmatrix}\right)"] \arrow[d] & V \oplus W' \arrow[r] \arrow[d] & Z \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & Y \arrow[r] \arrow[d] & T \arrow[r] \arrow[d] & Z \arrow[r] & 0 \\
 & 0 & 0 & &
\end{tikzcd}$$

The bottom row implies that T is in $\mathscr{Y} \bullet \mathscr{Z}$, and it follows from the middle column that M belongs to $\mathscr{X} \bullet (\mathscr{Y} \bullet \mathscr{Z})$. Hence we have $(\mathscr{X} \bullet \mathscr{Y}) \bullet \mathscr{Z} \subseteq \mathscr{X} \bullet (\mathscr{Y} \bullet \mathscr{Z})$.

Next, let M be an R-module in $\mathscr{X} \bullet (\mathscr{Y} \bullet \mathscr{Z})$. Then it follows from (1) that there is an exact sequence $0 \to X \to V \xrightarrow{f} W \to 0$ with $X \in [\mathscr{X}]$ and $W \in \mathscr{Y} \bullet \mathscr{Z}$ such that M is a direct summand of V. Applying (1) again, we have an exact sequence $0 \to Y \to U \to Z \to 0$ with $Y \in [\mathscr{Y}]$ and $Z \in [\mathscr{Z}]$ such that W is a direct summand of U. Write $U = W \oplus W'$, and we have a pullback diagram:

$$\begin{tikzcd}
 & & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & X \arrow[r] \arrow[d, equal] & T \arrow[r] \arrow[d] & Y \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & X \arrow[r] & V \oplus W' \arrow[r, "\left(\begin{smallmatrix} f & 0 \\ 0 & 1 \end{smallmatrix}\right)"] \arrow[d] & W \oplus W' \arrow[r] \arrow[d] & 0 \\
 & & Z \arrow[r, equal] \arrow[d] & Z \arrow[d] & \\
 & & 0 & 0 &
\end{tikzcd}$$

We see from the first row that T is in $\mathscr{X} \bullet \mathscr{Y}$, and from the middle column that M is in $(\mathscr{X} \bullet \mathscr{Y}) \bullet \mathscr{Z}$. Therefore $\mathscr{X} \bullet (\mathscr{Y} \bullet \mathscr{Z}) \subseteq (\mathscr{X} \bullet \mathscr{Y}) \bullet \mathscr{Z}$ holds.

(3) It is enough to show the equality $[\mathscr{C}]_a \bullet [\mathscr{C}]_b = [\mathscr{C}]_{a+b}$. We prove this by induction on b. It holds by definition when $b = 1$. Let $b \geq 2$. Then we have

$$[\mathscr{C}]_a \bullet [\mathscr{C}]_b = [\mathscr{C}]_a \bullet ([\mathscr{C}]_{b-1} \bullet \mathscr{C}) = ([\mathscr{C}]_a \bullet [\mathscr{C}]_{b-1}) \bullet \mathscr{C} = [\mathscr{C}]_{a+b-1} \bullet \mathscr{C} = [\mathscr{C}]_{a+b},$$

where the second equality follows from (2), and the induction hypothesis implies the third equality. □

Let $\mathscr{C}$ be a subcategory of $\operatorname{mod} R$ and $r \geq 0$ an integer. By the second and third assertions of Proposition 2.2, without danger of confusion we can write

$$[\mathscr{C}]_r = \overbrace{\mathscr{C} \bullet \cdots \bullet \mathscr{C}}^{r}.$$

Now we can make the definition of the radius of a subcategory.

Definition 2.3. Let R be a local ring, and let $\mathscr{X}$ be a subcategory of $\operatorname{mod} R$. We define the *radius* of $\mathscr{X}$, denoted by $\operatorname{radius} \mathscr{X}$, as the infimum of the integers $n \geq 0$ such that there exists a ball of radius $n + 1$ centered at a module containing $\mathscr{X}$. By definition, $\operatorname{radius} \mathscr{X} \in \mathbb{N} \cup \{\infty\}$.

The definition of the radius of a resolving subcategory looks similar to that of the *dimension* of a triangulated category which has been introduced by Rouquier [2008, Definition 3.2]. The stable category $\underline{\mathrm{CM}}(R)$ of Cohen–Macaulay modules over a Gorenstein local ring R is triangulated by [Buchweitz 1986; Happel 1988], and the dimension of $\underline{\mathrm{CM}}(R)$ in the sense of Rouquier is defined. It might look the same as the radius of $\mathrm{CM}(R)$ in our sense.

However there are (at least) two differences in the definitions:

(1) A defining object for $\dim \underline{\mathrm{CM}}(R)$ is required to be inside the category $\underline{\mathrm{CM}}(R)$, but a defining object for $\operatorname{radius} \mathrm{CM}(R)$ is not, i.e., it is enough to be an object of $\operatorname{mod} R$. More precisely, $\dim \underline{\mathrm{CM}}(R)$ (respectively, $\operatorname{radius} \mathrm{CM}(R)$) is defined as the infimum of the integers $n \geq 0$ such that $\underline{\mathrm{CM}}(R) = \langle G \rangle_{n+1}$ for some object G (respectively, $\mathrm{CM}(R) \subseteq [C]_{n+1}$ for some object C). Then G must be an object of $\underline{\mathrm{CM}}(R)$, while C may not be an object of $\mathrm{CM}(R)$, just being an object of $\operatorname{mod} R$.

(2) Let $\mathscr{X}$ and $\mathscr{Y}$ be subcategories of $\mathrm{CM}(R)$ and $\underline{\mathrm{CM}}(R)$, respectively. Then the subcategory $\langle \mathscr{Y} \rangle$ of $\underline{\mathrm{CM}}(R)$ is closed under taking cosyzygies of Cohen–Macaulay modules in it, but the subcategory $[\mathscr{X}]$ of $\mathrm{CM}(R)$ is not in general. (In fact, this difference is a reason why we can prove Proposition 2.5 below but do not know whether the analogue for dimension holds or not; see Question 2.6 below.)

Thus these two notions are different, but they are still related to each other. In fact, we can show that the following relationship exists between them.

Proposition 2.4. *Let R be a Gorenstein local ring.*

(1) *One has the inequality* $\dim \underline{\mathrm{CM}}(R) \leq \operatorname{radius} \mathrm{CM}(R)$.

(2) *The equality holds if R is a hypersurface.*

Proof. (1) We may assume that $n := \operatorname{radius} \mathrm{CM}(R) < \infty$. Then there exists an R-module C such that $\mathrm{CM}(R)$ is contained in the ball $[C]_{n+1}$.

We claim that $\mathrm{CM}(R) = [\Omega^{-d}\Omega^d C]_{n+1}$ holds, where $d = \dim R$. This claim implies $\underline{\mathrm{CM}}(R) = \langle \Omega^d C \rangle_{n+1}$, which shows $\dim \underline{\mathrm{CM}}(R) \leq n$.

In the following, we show this claim. Since $\Omega^{-d}\Omega^d C$ is a Cohen–Macaulay R-module, and $\mathrm{CM}(R)$ is a resolving subcategory of $\operatorname{mod} R$, the inclusion $\mathrm{CM}(R) \supseteq [\Omega^{-d}\Omega^d C]_{n+1}$ holds. To get the opposite inclusion, it is enough to prove that for every $m \geq 1$ and $M \in [C]_m$ we have $\Omega^{-d}\Omega^d M \in [\Omega^{-d}\Omega^d C]_m$. Let us prove this by induction on m. The case $m = 1$ is obvious, so let $m \geq 2$. According to Proposition 2.2(1), there is an exact sequence $0 \to X \to Y \to Z \to 0$ of R-modules with $X \in [C]_{m-1}$ and $Z \in [C]$ such that M is a direct summand of Y. Taking the d-th syzygies, we have an exact sequence $0 \to \Omega^d X \to \Omega^d Y \oplus R^{\oplus l} \to \Omega^d Z \to 0$

of Cohen–Macaulay R-modules. Since R is Gorenstein, taking the d-th cosyzygies makes an exact sequence

$$0 \to \Omega^{-d}\Omega^{d}X \to \Omega^{-d}\Omega^{d}Y \oplus R^{\oplus k} \to \Omega^{-d}\Omega^{d}Z \to 0$$

of Cohen–Macaulay modules. The induction hypothesis implies $\Omega^{-d}\Omega^{d}Z \in [\Omega^{-d}\Omega^{d}C]$ and $\Omega^{-d}\Omega^{d}X \in [\Omega^{-d}\Omega^{d}C]_{m-1}$. Since $\Omega^{-d}\Omega^{d}M$ is a direct summand of $\Omega^{-d}\Omega^{d}Y$, it belongs to $[\Omega^{-d}\Omega^{d}C]_{m}$.

(2) Let $n := \operatorname{dim}\underline{\mathrm{CM}}(R) < \infty$. We find a Cohen–Macaulay R-module G such that $\underline{\mathrm{CM}}(R) = \langle G\rangle_{n+1}$. We want to prove that $\mathrm{CM}(R) = [G]_{n+1}$ holds, and it suffices to show that for every $m \geq 1$ and $M \in \langle G\rangle_{m}$ we have $M \in [G]_{m}$. Let us use induction on m. The case $m = 1$ follows from the fact that $\Omega N \cong \Omega^{-1}N$ up to free summand for each Cohen–Macaulay R-module N, since R is a hypersurface. When $m \geq 2$, there exists an exact triangle $X \to Y \to Z \to \Sigma X$ in $\underline{\mathrm{CM}}(R)$ with $X \in \langle G\rangle_{m-1}$ and $Z \in \langle G\rangle$ such that M is a direct summand of Y. Then we have an exact sequence $0 \to X \to Y \oplus R^{\oplus h} \to Z \to 0$ of R-modules, and we are done by applying the induction hypothesis. □

In the rest of this section, we will study the relationships between the representation types of a Cohen–Macaulay local ring and the radius of the category of Cohen–Macaulay modules. Recall that a Cohen–Macaulay local ring R is said to be of *finite* (respectively, *countable*) *Cohen–Macaulay representation type* if $\mathrm{CM}(R)$ has only finitely (respectively, countably but not finitely) many indecomposable modules up to isomorphism.

We can describe the property of finite Cohen–Macaulay representation type in terms of a radius.

Proposition 2.5. *Let R be a Gorenstein Henselian local ring. The following are equivalent*:

(1) *One has* $\operatorname{radius}\mathrm{CM}(R) = 0$;

(2) *The ring R has finite Cohen–Macaulay representation type.*

Proof. (2) ⇒ (1): If $M_1, \dots, M_r$ are the nonisomorphic indecomposable Cohen–Macaulay R-modules, then we have $\mathrm{CM}(R) = [M_1 \oplus \cdots \oplus M_r]$.

(1) ⇒ (2): There is an R-module C satisfying $\mathrm{CM}(R) \subseteq [C]$. Setting $d = \operatorname{dim} R$, we have $\mathrm{CM}(R) = [\Omega^{-d}\Omega^{d}C]$. Replacing C with $\Omega^{-d}\Omega^{d}C$, we may assume that $\mathrm{CM}(R) = [C]$ with C being Cohen–Macaulay.

Note that since R is Henselian, the Krull–Schmidt theorem holds, i.e., each R-module uniquely decomposes into indecomposable R-modules up to isomorphism. Let $C_1, \dots, C_n$ be the indecomposable direct summands of C. We may assume that $C = C_1 \oplus \cdots \oplus C_n$. Since R is Gorenstein, taking syzygies preserves indecomposability of nonfree Cohen–Macaulay R-modules. We see that the set of nonisomorphic indecomposable Cohen–Macaulay R-modules is

$$\{R\} \cup \{\, \Omega^i C_j \mid i \geq 0,\ 1 \leq j \leq n \,\}.$$

We may assume that for all $i \geq 0$ and $1 \leq j \neq j' \leq n$ we have $\Omega^i C_j \ncong C_{j'}$, because if $\Omega^i C_j \cong C_{j'}$ for some such i, j, j', then we can exclude $C_{j'}$ from C. Now fix an integer j with $1 \leq j \leq n$. As taking cosyzygies preserves indecomposability of nonfree Cohen–Macaulay R-modules, $\Omega^{-1} C_j$ is isomorphic to $\Omega^a C_b$ for some $a \geq 0$ and $1 \leq b \leq n$. Taking the a-th cosyzygies, we have $\Omega^{-1-a} C_j \cong C_b$, hence $C_j \cong \Omega^{1+a} C_b$. This forces us to have $b = j$, which says that C_j is periodic. Hence there are only finitely many indecomposable Cohen–Macaulay R-modules. □

Question 2.6. Does the equality in Proposition 2.4(1) hold true? If it does, then Proposition 2.5 will say that *a Gorenstein Henselian local ring R has finite Cohen–Macaulay representation type if and only if* $\dim \underline{\mathrm{CM}}(R) \leq 0$. This statement is a partial generalization of Minamoto's theorem [2013, Theorem 0.2], which asserts that the same statement holds for a finite-dimensional self-injective algebra over a perfect field, extending Yoshiwaki's recent theorem [2011, Corollary 3.10].

The next result hints at further relationship between finite radius of $\mathrm{CM}(R)$ and more well-known classification of singularities.

Proposition 2.7. *Let R be a complete local hypersurface over an algebraically closed field of characteristic not two. Assume that R is of countable Cohen–Macaulay representation type. Then* $\operatorname{radius} \mathrm{CM}(R) = 1$.

Proof. It follows from [Araya et al. 2012, Theorem 1.1] that there exists an R-module X such that for every indecomposable module $M \in \mathrm{CM}(R)$ there is an exact sequence $0 \to L \to M \oplus R^n \to N \to 0$ with $L, N \in \{0, X, \Omega X\}$. This shows that $\mathrm{CM}(R) = [X]_2$. Now we see from Proposition 2.5 that the radius of $\mathrm{CM}(R)$ is equal to one. □

3. Finiteness of the radius of a resolving subcategory

In this section we state our guiding conjecture and first main result.

Conjecture 3.1. Let R be a Cohen–Macaulay local ring. Let $\mathscr{X}$ be a resolving subcategory of $\operatorname{mod} R$ with finite radius. Then every R-module in $\mathscr{X}$ is Cohen–Macaulay.

Remark 3.2. The converse of Conjecture 3.1 also seems to be true. We consider this in Section 5.

Let $\mathscr{X}$ be a subcategory of $\operatorname{mod} R$. We denote by $\operatorname{res} \mathscr{X}$ (or $\operatorname{res}_R \mathscr{X}$) the *resolving closure* of $\mathscr{X}$, namely, the smallest resolving subcategory of $\operatorname{mod} R$ containing $\mathscr{X}$. If

$\mathscr{X}$ consists of a single module M, then we simply denote it by $\operatorname{res} M$ (or $\operatorname{res}_R M$). For a prime ideal $\mathfrak{p}$ of R, we denote by $\mathscr{X}_\mathfrak{p}$ the subcategory of $\operatorname{mod} R_\mathfrak{p}$ consisting of all modules of the form $X_\mathfrak{p}$, where $X \in \mathscr{X}$. The first main result of this paper is the following theorem.

Theorem 3.3. *Let R be a commutative Noetherian ring. Let $\mathscr{X}$ be a resolving subcategory of* $\operatorname{mod} R$. *Suppose that there exist a prime ideal $\mathfrak{p}$ of R with $\operatorname{ht}\mathfrak{p} > 0$ and an $R_\mathfrak{p}$-module M with $0 \neq M \in \operatorname{add}_{R_\mathfrak{p}} \mathscr{X}_\mathfrak{p}$ which satisfy one of the following conditions.*

(1) $\mathfrak{p}M = 0$.

(2) $0 < \operatorname{Gdim}_{R_\mathfrak{p}} M = n < \infty$ *and* $\Omega_{R_\mathfrak{p}}^{-2}\Omega_{R_\mathfrak{p}}^{n} M \in \operatorname{add}_{R_\mathfrak{p}} \mathscr{X}_\mathfrak{p}$.

(3) $0 < \operatorname{Gdim}_{R_\mathfrak{p}} M < \infty$ *and* $\operatorname{res}_{R_\mathfrak{p}}(\Omega_{R_\mathfrak{p}}^{n} M)$ *is a thick subcategory of* $\mathscr{G}(R_\mathfrak{p})$ *for some* $n \geq 0$.

(4) $0 < \operatorname{CIdim}_{R_\mathfrak{p}} M < \infty$.

Then $\mathscr{X}$ has infinite radius.

The proof of this theorem will be given in the next section. As a direct consequence of the above theorem, we obtain two cases in which our conjecture holds true.

Corollary 3.4. *Conjecture 3.1 is true if*

(1) *R is a complete intersection, or*

(2) *R is Gorenstein, and every resolving subcategory of* $\operatorname{mod} R$ *contained in* $\operatorname{CM}(R)$ *is a thick subcategory of* $\operatorname{CM}(R)$.

Proof. Conjecture 3.1 trivially holds in the case where R is Artinian, so let $(R, \mathfrak{m}, k)$ be a Cohen–Macaulay local ring of positive dimension. Then we have $\operatorname{ht}\mathfrak{m} > 0$. Let $\mathscr{X}$ be a resolving subcategory of $\operatorname{mod} R$, and suppose that $\mathscr{X}$ contains a non-Cohen–Macaulay R-module M.

(1) We have $0 < \dim R - \operatorname{depth}_R M = \operatorname{depth} R - \operatorname{depth}_R M = \operatorname{CIdim}_R M < \infty$, and Theorem 3.3(4) implies that $\mathscr{X}$ has infinite radius.

(2) We have $0 < n := \dim R - \operatorname{depth}_R M = \operatorname{depth} R - \operatorname{depth}_R M = \operatorname{Gdim}_R M < \infty$. The module $\Omega_R^n M$ is Cohen–Macaulay, and by assumption $\operatorname{res}_R(\Omega_R^n M)$ is a thick subcategory of $\operatorname{CM}(R) = \mathscr{G}(R)$. Theorem 3.3(3) implies that $\mathscr{X}$ has infinite radius. □

4. Proof of Theorem I

This section is devoted to give the proof of Theorem 3.3 (hence of Theorem I from the introduction), which we break up into several parts. Most of them also reveal properties of subcategories of $\operatorname{mod} R$ which are interesting in their own right.

First of all, we make a remark to reduce our theorem to the local case.

Remark 4.1. Let C be an R-module, $n \geq 0$ an integer and $\mathfrak{p}$ a prime ideal of R. Then for a subcategory $\mathscr{X}$ of $\operatorname{mod} R$ the implication

$$\mathscr{X} \subseteq [C]_n^R \quad \Rightarrow \quad \operatorname{add}_{R_\mathfrak{p}} \mathscr{X}_\mathfrak{p} \subseteq [C_\mathfrak{p}]_n^{R_\mathfrak{p}}$$

holds. It follows from [Takahashi 2010, Lemma 4.8] that $\operatorname{add}_{R_\mathfrak{p}} \mathscr{X}_\mathfrak{p}$ is a resolving subcategory of $\operatorname{mod} R_\mathfrak{p}$. (The ring R in (loc. cit.) is assumed to be local, but its proof does not use this assumption, so it holds for an arbitrary commutative Noetherian ring.) Hence, to prove Theorem 3.3, without loss of generality we can assume $(R, \mathfrak{p})$ is a local ring with $\dim R > 0$ and M is an R-module with $0 \neq M \in \mathscr{X}$.

***Proof of Theorem 3.3*(1).** First, we investigate the annihilators of torsion submodules. For an ideal I of R and an R-module M, we denote by $\Gamma_I(M)$ the *I-torsion submodule* of M. Recall that $\Gamma_I(M)$ is by definition the subset of M consisting of all elements that are annihilated by some power of I, and the assignment $M \mapsto \Gamma_I(M)$ defines a left exact additive covariant functor $\Gamma_I : \operatorname{mod} R \to \operatorname{mod} R$.

Lemma 4.2. *Let I be an ideal of R. Let C, M be R-modules and $n \geq 1$ an integer. If M belongs to $[C]_n$, then one has* $\operatorname{Ann}_R \Gamma_I(M) \supseteq (\operatorname{Ann}_R \Gamma_I(R) \cdot \operatorname{Ann}_R \Gamma_I(C))^n$.

Proof. Let us prove the lemma by induction on n.

If $n = 1$, the module M is isomorphic to a direct summand of $\bigl(R \oplus \bigoplus_{i=0}^a \Omega^i C\bigr)^{\oplus b}$ for some $a, b \geq 0$. Hence $\Gamma_I(M)$ is isomorphic to a direct summand of

$$\left(\Gamma_I(R) \oplus \bigoplus_{i=0}^a \Gamma_I(\Omega^i C)\right)^{\oplus b}.$$

For $i \geq 1$, the syzygy $\Omega^i C$ is a submodule of some free module $R^{\oplus c_i}$, and $\Gamma_I(\Omega^i C)$ is a submodule of $\Gamma_I(R)^{\oplus c_i}$, which implies that $\operatorname{Ann}_R \Gamma_I(\Omega^i C)$ contains $\operatorname{Ann}_R \Gamma_I(R)$. Hence we obtain

$$\begin{aligned}\operatorname{Ann}_R \Gamma_I(M) &\supseteq \operatorname{Ann}_R \Gamma_I(R) \cap \left(\bigcap_{i=0}^a \operatorname{Ann}_R \Gamma_I(\Omega^i C)\right) \\ &= \operatorname{Ann}_R \Gamma_I(R) \cap \operatorname{Ann}_R \Gamma_I(C) \supseteq \operatorname{Ann}_R \Gamma_I(R) \cdot \operatorname{Ann}_R \Gamma_I(C).\end{aligned}$$

Let $n \geq 2$. Then M is in $[C]_n = [C]_{n-1} \bullet [C]$, and Proposition 2.2(1) says that there is an exact sequence $0 \to X \to Y \to Z \to 0$ with $X \in [C]_{n-1}$ and $Z \in [C]$ such that M is a direct summand of Y. We have an exact sequence $0 \to \Gamma_I(X) \to \Gamma_I(Y) \to \Gamma_I(Z)$, and therefore we obtain

$$\begin{aligned}\operatorname{Ann}_R \Gamma_I(M) &\supseteq \operatorname{Ann}_R \Gamma_I(Y) \supseteq \operatorname{Ann}_R \Gamma_I(X) \cdot \operatorname{Ann}_R \Gamma_I(Z) \\ &\supseteq (\operatorname{Ann}_R \Gamma_I(R) \cdot \operatorname{Ann}_R \Gamma_I(C))^{n-1} \cdot (\operatorname{Ann}_R \Gamma_I(R) \cdot \operatorname{Ann}_R \Gamma_I(C)) \\ &= (\operatorname{Ann}_R \Gamma_I(R) \cdot \operatorname{Ann}_R \Gamma_I(C))^n,\end{aligned}$$

which is what we want. □

Let $(R, \mathfrak{m})$ be a local ring, and let M be an R-module. We denote by $\ell\ell(M)$ the *Loewy length* of M, which is by definition the infimum of the integers $n \geq 0$ such that $\mathfrak{m}^n M = 0$. Obviously, $\ell\ell(M)$ is finite if and only if M has finite length. There is a relationship between finite radius and Loewy length:

Proposition 4.3. *Let $(R, \mathfrak{m})$ be a local ring, and let $\mathscr{X}$ be a resolving subcategory of* $\operatorname{mod} R$. *If* $\operatorname{radius} \mathscr{X} < \infty$, *then* $\sup_{X \in \mathscr{X}}\{\ell\ell(\Gamma_{\mathfrak{m}}(X))\} < \infty$.

Proof. Put $r = \operatorname{radius} \mathscr{X}$. By definition, there exists an R-module C such that $[C]_{r+1}$ contains $\mathscr{X}$. Let X be a module in $\mathscr{X}$. It follows from Lemma 4.2 that the annihilator $\operatorname{Ann} \Gamma_{\mathfrak{m}}(X)$ contains the ideal $(\operatorname{Ann} \Gamma_{\mathfrak{m}}(R) \cdot \operatorname{Ann} \Gamma_{\mathfrak{m}}(C))^{r+1}$. As $\Gamma_{\mathfrak{m}}(R)$ and $\Gamma_{\mathfrak{m}}(C)$ have finite length, they have finite Loewy length. Set $a = \ell\ell(\Gamma_{\mathfrak{m}}(R))$ and $b = \ell\ell(\Gamma_{\mathfrak{m}}(C))$. Then $\operatorname{Ann} \Gamma_{\mathfrak{m}}(X)$ contains $\mathfrak{m}^{(a+b)(r+1)}$, which means that $\ell\ell(\Gamma_{\mathfrak{m}}(X))$ is at most $(a+b)(r+1)$. Since the number $(a+b)(r+1)$ is independent of the choice of X, we have $\sup_{X \in \mathscr{X}}\{\ell\ell(\Gamma_{\mathfrak{m}}(X))\} \leq (a+b)(r+1) < \infty$. □

The following is the essential part of Theorem 3.3(1).

Theorem 4.4. *Let $(R, \mathfrak{m}, k)$ be a local ring of positive dimension. Let $\mathscr{X}$ be a resolving subcategory of* $\operatorname{mod} R$. *If $\mathscr{X}$ contains k, then $\mathscr{X}$ has infinite radius.*

Proof. We claim that $R/\mathfrak{m}^i$ belongs to $\mathscr{X}$ for all integers $i > 0$. Indeed, there is an exact sequence $0 \to \mathfrak{m}^{i-1}/\mathfrak{m}^i \to R/\mathfrak{m}^i \to R/\mathfrak{m}^{i-1} \to 0$, and the left term belongs to $\mathscr{X}$ as it is a k-vector space. Induction on i shows the claim.

We have $\ell\ell(\Gamma_{\mathfrak{m}}(R/\mathfrak{m}^i)) = \ell\ell(R/\mathfrak{m}^i) = i$, where the second equality follows from the assumption that $\dim R > 0$. Therefore it holds that

$$\sup_{X \in \mathscr{X}}\{\ell\ell(\Gamma_{\mathfrak{m}}(X))\} \geq \sup_{i>0}\{\ell\ell(\Gamma_{\mathfrak{m}}(R/\mathfrak{m}^i))\} = \sup_{i>0}\{i\} = \infty,$$

which implies $\operatorname{radius} \mathscr{X} = \infty$ by Proposition 4.3. □

Now, let us prove Theorem 3.3(1). By Remark 4.1, we may assume that $(R, \mathfrak{p})$ is a local ring with $\dim R > 0$ and that M is a nonzero R-module in $\mathscr{X}$. The assumption $\mathfrak{p}M = 0$ says that M is a nonzero k-vector space, where $k = R/\mathfrak{p}$ is the residue field of R. Since $\mathscr{X}$ is closed under direct summands, k belongs to $\mathscr{X}$. Theorem 4.4 yields $\operatorname{radius} \mathscr{X} = \infty$.

***Proof of Theorem 3.3*(2)(3).** Establishing several preliminary lemmas and propositions is necessary, which will also be used in the proof of Theorem 3.3(4).

We begin with stating an elementary lemma, whose proof we omit.

Lemma 4.5. *Let $(R, \mathfrak{m}, k)$ be a local ring. Let $0 \to L \to M \to N \to 0$ be an exact sequence of R-modules. Then* $\inf\{\operatorname{depth} L, \operatorname{depth} N\} = \inf\{\operatorname{depth} M, \operatorname{depth} N\}$.

For an ideal I of R, we denote by $V(I)$ (respectively, $D(I)$) the closed (respectively, open) subset of $\operatorname{Spec} R$ defined by I in the Zariski topology, namely, $V(I)$ is the set of prime ideals containing I and $D(I) = \operatorname{Spec} R \setminus V(I)$. For an R-module M we denote by $\mathrm{NF}(M)$ the *nonfree locus* of M, that is, the set of prime ideals $\mathfrak{p}$ of R such that the $R_{\mathfrak{p}}$-module $M_{\mathfrak{p}}$ is nonfree. It is well-known that $\mathrm{NF}(M)$ is a closed subset of $\operatorname{Spec} R$.

The next result builds, out of each module in a resolving subcategory and each point in its nonfree locus, another module in the same resolving subcategory whose nonfree locus coincides with the closure of the point. Such a construction has already been given in [Takahashi 2009, Theorem 4.3], but we need in this paper a more detailed version. Indeed, the following lemma yields a generalization of (loc. cit.).

Lemma 4.6. *Let M be an R-module. For every $\mathfrak{p} \in \mathrm{NF}(M)$ there exists $X \in \operatorname{res} M$ satisfying $\mathrm{NF}(X) = V(\mathfrak{p})$ and $\operatorname{depth} X_{\mathfrak{q}} = \inf\{\operatorname{depth} M_{\mathfrak{q}}, \operatorname{depth} R_{\mathfrak{q}}\}$ for all $\mathfrak{q} \in V(\mathfrak{p})$.*

Proof. Note that $V(\mathfrak{p})$ is contained in $\mathrm{NF}(M)$. If $V(\mathfrak{p}) = \mathrm{NF}(M)$, then we can take $X := M \oplus R$. Suppose $V(\mathfrak{p})$ is strictly contained in $\mathrm{NF}(M)$. Then there is a prime ideal $\mathfrak{r}$ in $\mathrm{NF}(M)$ that is not in $V(\mathfrak{p})$. Choose an element $x \in \mathfrak{p} \setminus \mathfrak{r}$. By [ibid., Proposition 4.2], we have a commutative diagram with exact rows

$$\begin{tikzcd}
0 \arrow[r] & \Omega M \arrow[r] \arrow[d, "x"'] & F \arrow[r] \arrow[d] & M \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \Omega M \arrow[r] & N \arrow[r] & M \arrow[r] & 0
\end{tikzcd}$$

where F is free, $V(\mathfrak{p}) \subseteq \mathrm{NF}(N) \subseteq \mathrm{NF}(M)$ and $D((x)) \cap \mathrm{NF}(N) = \varnothing$. The second row shows that N belongs to $\operatorname{res} M$. Since $\mathfrak{r}$ is in $D((x))$, it is not in $\mathrm{NF}(N)$, and we have $V(\mathfrak{p}) \subseteq \mathrm{NF}(N) \subsetneq \mathrm{NF}(M)$.

Now we claim that $\operatorname{depth} N_{\mathfrak{q}} = \inf\{\operatorname{depth} M_{\mathfrak{q}}, \operatorname{depth} R_{\mathfrak{q}}\}$ for all $\mathfrak{q} \in V(\mathfrak{p})$. Indeed, localizing the above diagram at $\mathfrak{q}$ and taking long exact sequences with respect to Ext, we get a commutative diagram with exact rows

$$\begin{tikzcd}
E^{i-1}(M_{\mathfrak{q}}) \arrow[r] \arrow[d, equal] & E^{i}((\Omega M)_{\mathfrak{q}}) \arrow[r] \arrow[d, "x"', "(1)"] & E^{i}(F_{\mathfrak{q}}) \arrow[r] \arrow[d] & E^{i}(M_{\mathfrak{q}}) \arrow[r] \arrow[d, equal] & E^{i+1}((\Omega M)_{\mathfrak{q}}) \arrow[d, "x"', "(2)"] \\
E^{i-1}(M_{\mathfrak{q}}) \arrow[r, "(3)"] & E^{i}((\Omega M)_{\mathfrak{q}}) \arrow[r] & E^{i}(N_{\mathfrak{q}}) \arrow[r] & E^{i}(M_{\mathfrak{q}}) \arrow[r, "(4)"] & E^{i+1}((\Omega M)_{\mathfrak{q}})
\end{tikzcd}$$

for $i \in \mathbb{Z}$, where $E^i(-) = \operatorname{Ext}^i_{R_{\mathfrak{q}}}(\kappa(\mathfrak{q}), -)$. As x is an element of $\mathfrak{q}$, the maps (1), (2) are zero maps, and so are (3), (4). Thus we have a short exact sequence

$$0 \to \operatorname{Ext}^i(\kappa(\mathfrak{q}), (\Omega M)_{\mathfrak{q}}) \to \operatorname{Ext}^i(\kappa(\mathfrak{q}), N_{\mathfrak{q}}) \to \operatorname{Ext}^i(\kappa(\mathfrak{q}), M_{\mathfrak{q}}) \to 0$$

for each integer i. It is easy to see from this that the first equality in the following holds, while the second equality is obtained by applying Lemma 4.5 to the exact sequence $0 \to (\Omega M)_{\mathfrak{q}} \to F_{\mathfrak{q}} \to M_{\mathfrak{q}} \to 0$:

$$\operatorname{depth} N_{\mathfrak{q}} = \inf\{\operatorname{depth}(\Omega M)_{\mathfrak{q}}, \operatorname{depth} M_{\mathfrak{q}}\} = \inf\{\operatorname{depth} R_{\mathfrak{q}}, \operatorname{depth} M_{\mathfrak{q}}\}.$$

Thus the claim follows.

If $V(\mathfrak{p}) = \mathrm{NF}(N)$, then we can take $X := N$. If $V(\mathfrak{p})$ is strictly contained in $\mathrm{NF}(N)$, then the above procedure gives rise to an R-module $L \in \operatorname{res} N$ (so $L \in \operatorname{res} M$) with $V(\mathfrak{p}) \subseteq \mathrm{NF}(L) \subsetneq \mathrm{NF}(N) \subsetneq \mathrm{NF}(M)$ such that

$$\operatorname{depth} L_{\mathfrak{q}} = \inf\{\operatorname{depth} N_{\mathfrak{q}}, \operatorname{depth} R_{\mathfrak{q}}\} = \inf\{\operatorname{depth} M_{\mathfrak{q}}, \operatorname{depth} R_{\mathfrak{q}}\}$$

for all $\mathfrak{q} \in V(\mathfrak{p})$. Since $\operatorname{Spec} R$ is a Noetherian space, iteration of this procedure must stop in finitely many steps, and we eventually obtain such a module X as in the lemma. □

The next lemma will play a crucial role in the proofs of our theorems. The main idea of the proof is similar to that of Lemma 4.2, but a much closer examination is necessary.

Lemma 4.7. *Let $S \to R$ be a homomorphism of rings. Let C, M be R-modules with $M \in [C]^R_n$, and let N be an S-module of injective dimension $m < \infty$. Then*

$$\begin{aligned}\bigcap_{i>0} \operatorname{Ann}_S \operatorname{Ext}^i_S(M, N) &= \bigcap_{i=1}^{m} \operatorname{Ann}_S \operatorname{Ext}^i_S(M, N)\\ &\supseteq \left(\prod_{i=1}^{m} \operatorname{Ann}_S \operatorname{Ext}^i_S(R, N) \cdot \operatorname{Ann}_S \operatorname{Ext}^i_S(C, N)\right)^n.\end{aligned}$$

Proof. For each integer $i \leq 1$ and R-module L, set $\mathfrak{a}^i_L = \operatorname{Ann}_S \operatorname{Ext}^i_S(L, N)$. Note that $\mathfrak{a}^h_L = S$ for all $h > m$ since $\operatorname{Ext}^h_S(L, N) = 0$. It suffices to prove that $\mathfrak{a}^i_M \supseteq \left(\prod_{j=i}^{m} \mathfrak{a}^j_R \mathfrak{a}^j_C\right)^n$. Let us proceed by induction on n.

When $n = 0$, we have $M = 0$, and the above two ideals coincide with S.

Let $n = 1$. Then M is isomorphic to a direct summand of a finite direct sum of copies of $R \oplus \left(\bigoplus_{j=0}^{l} \Omega^j C\right)$. Hence $\operatorname{Ext}^i_S(M, N)$ is isomorphic to a direct summand of a finite direct sum of copies of $\operatorname{Ext}^i_S(R, N) \oplus \left(\bigoplus_{j=0}^{l} \operatorname{Ext}^i_S(\Omega^j C, N)\right)$. Thus we have $\mathfrak{a}^i_M \supseteq \mathfrak{a}^i_R \cap \left(\bigcap_{j=0}^{l} \mathfrak{a}^i_{\Omega^j C}\right)$. For each $j \geq 1$ there is an exact sequence $0 \to \Omega^j C \to R^{\oplus k_j} \to \Omega^{j-1} C \to 0$, which induces an exact sequence

$$\operatorname{Ext}^i_S(R, N)^{\oplus k_j} \to \operatorname{Ext}^i_S(\Omega^j C, N) \to \operatorname{Ext}^{i+1}_S(\Omega^{j-1} C, N).$$

This gives

$$\mathfrak{a}^i_{\Omega^j C} \supseteq \mathfrak{a}^i_R \mathfrak{a}^{i+1}_{\Omega^{j-1} C} \supseteq \mathfrak{a}^i_R \mathfrak{a}^{i+1}_R \mathfrak{a}^{i+2}_{\Omega^{j-2} C} \supseteq \cdots \supseteq \mathfrak{a}^i_R \mathfrak{a}^{i+1}_R \cdots \mathfrak{a}^{i+j-1}_R \mathfrak{a}^{i+j}_C.$$

Regarding $\mathfrak{a}_R^i\mathfrak{a}_R^{i+1}\cdots\mathfrak{a}_R^{i+j-1}\mathfrak{a}_C^{i+j}$ as $\mathfrak{a}_C^i$ when $j=0$, we have

$$\mathfrak{a}_{\Omega^j C}^i \supseteq \mathfrak{a}_R^i\mathfrak{a}_R^{i+1}\cdots\mathfrak{a}_R^{i+j-1}\mathfrak{a}_C^{i+j}$$

for all $j \geq 0$. Thus we obtain

$$\begin{aligned}\mathfrak{a}_M^i &\supseteq \mathfrak{a}_R^i \cap \left(\bigcap_{j=0}^{l} \mathfrak{a}_R^i\mathfrak{a}_R^{i+1}\cdots\mathfrak{a}_R^{i+j-1}\mathfrak{a}_C^{i+j}\right)\\ &\supseteq (\mathfrak{a}_R^i\mathfrak{a}_R^{i+1}\cdots\mathfrak{a}_R^m\mathfrak{a}_R^{m+1}\cdots)(\mathfrak{a}_C^i\mathfrak{a}_C^{i+1}\cdots\mathfrak{a}_C^m\mathfrak{a}_C^{m+1}\cdots)\\ &= (\mathfrak{a}_R^i\mathfrak{a}_R^{i+1}\cdots\mathfrak{a}_R^m)(\mathfrak{a}_C^i\mathfrak{a}_C^{i+1}\cdots\mathfrak{a}_C^m) \supseteq \prod_{j=i}^{m}\mathfrak{a}_R^j\mathfrak{a}_C^j.\end{aligned}$$

Now let us consider the case where $n \geq 2$. We have $M \in [C]_n = [C]_{n-1} \bullet [C]$, and by Proposition 2.2(1), there exists an exact sequence $0 \to X \to Y \to Z \to 0$ with $X \in [C]_{n-1}$ and $Z \in [C]$ such that M is a direct summand of Y. Using the induction hypothesis, we have

$$\mathfrak{a}_M^i \supseteq \mathfrak{a}_Y^i \supseteq \mathfrak{a}_X^i \cdot \mathfrak{a}_Z^i \supseteq \left(\prod_{j=i}^{m}\mathfrak{a}_R^j\mathfrak{a}_C^j\right)^{n-1} \cdot \left(\prod_{j=i}^{m}\mathfrak{a}_R^j\mathfrak{a}_C^j\right) = \left(\prod_{j=i}^{m}\mathfrak{a}_R^j\mathfrak{a}_C^j\right)^{n},$$

which completes the proof of the lemma. □

Here we prepare a lemma, which is an easy consequence of Krull's intersection theorem.

Lemma 4.8. *Let* $(R, \mathfrak{m})$ *be a local ring and* M *an* R*-module. Then* $\operatorname{Ann}_R M = \bigcap_{i>0} \operatorname{Ann}_R(M/\mathfrak{m}^i M)$.

Now we can prove the following proposition, which will be the base of the proofs of our theorems. Actually, all of them will be proved by making use of this proposition.

Proposition 4.9. *Let* $(R, \mathfrak{m}, k)$ *be a local ring. Let* $\mathcal{X}$ *be a resolving subcategory of* $\operatorname{mod} R$*. If* $\mathcal{X}$ *contains a module* M *such that* $0 < \operatorname{pd}_R M < \infty$*, then* $\operatorname{radius}\mathcal{X} = \infty$.

Proof. Applying Lemma 4.6 to $\mathfrak{m} \in \operatorname{NF}(M)$, we find a module $X \in \operatorname{res} M \subseteq \mathcal{X}$ satisfying $\operatorname{NF}(X) = \{\mathfrak{m}\}$ and $\operatorname{depth} X = \inf\{\operatorname{depth} M, \operatorname{depth} R\}$. Since M has finite projective dimension, the depth of M is at most that of R. Hence, we have $\operatorname{depth} X = \operatorname{depth} M$. Note that the subcategory of $\operatorname{mod} R$ consisting of R-modules of finite projective dimension is resolving. Since it contains M, it also contains $\operatorname{res} M$. This implies that X has finite projective dimension, and we have

$$\operatorname{pd}_R X = \operatorname{depth} R - \operatorname{depth} X = \operatorname{depth} R - \operatorname{depth} M = \operatorname{pd}_R M.$$

Thus, replacing M with X, we may assume that M is locally free on the punctured spectrum of R. Taking the n-th syzygy of M where $n = \mathsf{pd}_R M - 1 \geq 0$, we may also assume that the projective dimension of M is equal to 1.

Now $\mathrm{Ext}^1_R(M, R)$ is a nonzero R-module of finite length, and we can choose a socle element $0 \neq \sigma \in \mathrm{Ext}^1_R(M, R)$. It can be represented as a short exact sequence

$$\sigma : 0 \to R \to N \to M \to 0.$$

The module N belongs to $\mathscr{X}$, is locally free on the punctured spectrum of R and has projective dimension at most 1. Hence $\mathsf{pd}_R N = 1$ if and only if $\mathrm{Ext}^1_R(N, R) \neq 0$. Applying the functor $\mathrm{Hom}_R(-, R)$, we get an exact sequence

$$R \xrightarrow{f} \mathrm{Ext}^1_R(M, R) \to \mathrm{Ext}^1_R(N, R) \to 0,$$

where f sends $1 \in R$ to $\sigma \in \mathrm{Ext}^1_R(M, R)$. Hence we obtain an exact sequence

$$0 \to k \to \mathrm{Ext}^1_R(M, R) \to \mathrm{Ext}^1_R(N, R) \to 0.$$

This implies $\mathsf{length}(\mathrm{Ext}^1_R(N, R)) = \mathsf{length}(\mathrm{Ext}^1_R(M, R)) - 1$. Replacing M by N and repeating this process if $\mathsf{length}(\mathrm{Ext}^1_R(N, R)) > 0$, we can assume that $\mathrm{Ext}^1_R(N, R) = 0$. Therefore $\mathrm{Ext}^1_R(M, R) \cong k$. Since $\mathsf{pd}_R M = 1$, we easily get an isomorphism $\mathrm{Tr}\, M \cong k$. Taking the transpose of this isomorphism, we see that $\mathrm{Tr}\, k$ is isomorphic to M up to free summand (see [Auslander and Bridger 1969, Proposition (2.6)(d)]). It follows that $\mathrm{Tr}\, k$ belongs to $\mathscr{X}$.

We claim that $\mathrm{Tr}\, L$ is in $\mathscr{X}$ for any R-module L of finite length. This is shown by induction on $\mathsf{length}\, L$. If $\mathsf{length}\, L > 0$, then there is an exact sequence $0 \to L' \to L \to k \to 0$, and applying [ibid., Lemma (3.9)] (see also [Takahashi 2013, Proposition 3.3(3)]), we have an exact sequence

$$0 = (L')^* \to \mathrm{Tr}\, k \to \mathrm{Tr}\, L \oplus R^{\oplus n} \to \mathrm{Tr}\, L' \to 0,$$

where the equality follows from the fact that R has positive depth. (As R possesses a module of finite positive projective dimension, the depth of R is positive.) The induction hypothesis implies $\mathrm{Tr}\, L' \in \mathscr{X}$, and the above exact sequence shows $\mathrm{Tr}\, L \in \mathscr{X}$, as desired.

Now, assume that we have $\mathsf{radius}\, \mathscr{X} = r < \infty$. We want to deduce a contradiction. There is a ball $[C]^R_{r+1}$ that contains $\mathscr{X}$. Since $\mathscr{X}$ contains $\mathrm{Tr}_R(R/\mathfrak{m}^i)$ for all $i > 0$, the ball $[C]^R_{r+1}$ also contains it. Taking the completions, we have $\mathrm{Tr}_{\widehat{R}}(\widehat{R}/\mathfrak{m}^i \widehat{R}) \in [\widehat{C}]^{\widehat{R}}_{r+1}$ for all $i > 0$. By virtue of Cohen's structure theorem, there exists a surjective homomorphism $S \to \widehat{R}$ such that S is a Gorenstein local ring with $\mathsf{dim}\, S = \mathsf{dim}\, R =: d$. Let $\mathfrak{n}$ denote the maximal ideal of S and note that we have $\widehat{R}/\mathfrak{m}^i \widehat{R} = \widehat{R}/\mathfrak{n}^i \widehat{R}$ for any $i > 0$. Lemma 4.7 gives an inclusion relation

$$\bigcap_{j>0} \operatorname{Ann}_S \operatorname{Ext}_S^j(\operatorname{Tr}_{\widehat{R}}(\widehat{R}/\mathfrak{n}^i\widehat{R}), S) \supseteq \left(\prod_{j=1}^{d} \operatorname{Ann}_S \operatorname{Ext}_S^j(\widehat{R}, S) \cdot \operatorname{Ann}_S \operatorname{Ext}_S^j(\widehat{C}, S)\right)^{r+1}.$$

Fix an integer $i > 0$ and let $a_1, \dots, a_m$ be a system of generators of the ideal $\mathfrak{n}^i$ of S. There is an exact sequence $\widehat{R}^{\oplus m} \xrightarrow{(a_1,\dots,a_m)} \widehat{R} \to \widehat{R}/\mathfrak{n}^i\widehat{R} \to 0$ of $\widehat{R}$-modules. Dualizing this by $\widehat{R}$ induces an exact sequence

$$0 = \operatorname{Hom}_{\widehat{R}}(\widehat{R}/\mathfrak{n}^i\widehat{R}, \widehat{R}) \to \widehat{R} \xrightarrow{\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}} \widehat{R}^{\oplus m} \to \operatorname{Tr}_{\widehat{R}}(\widehat{R}/\mathfrak{n}^i\widehat{R}) \to 0,$$

where the equality follows since $\operatorname{depth} \widehat{R} = \operatorname{depth} R > 0$. This makes an exact sequence

$$\operatorname{Hom}_S(\widehat{R}, S)^{\oplus m} \xrightarrow{(a_1,\dots,a_m)} \operatorname{Hom}_S(\widehat{R}, S) \to \operatorname{Ext}_S^1(\operatorname{Tr}_{\widehat{R}}(\widehat{R}/\mathfrak{n}^i\widehat{R}), S),$$

which yields an injection $\operatorname{Hom}_S(\widehat{R}, S)/\mathfrak{n}^i \operatorname{Hom}_S(\widehat{R}, S) \to \operatorname{Ext}_S^1(\operatorname{Tr}_{\widehat{R}}(\widehat{R}/\mathfrak{n}^i\widehat{R}), S)$. Thus

$$\begin{aligned}
&\operatorname{Ann}_S(\operatorname{Hom}_S(\widehat{R}, S)/\mathfrak{n}^i \operatorname{Hom}_S(\widehat{R}, S)) \\
&\supseteq \operatorname{Ann}_S(\operatorname{Ext}_S^1(\operatorname{Tr}_{\widehat{R}}(\widehat{R}/\mathfrak{n}^i\widehat{R}), S)) \\
&\supseteq \bigcap_{j>0} \operatorname{Ann}_S \operatorname{Ext}_S^j(\operatorname{Tr}_{\widehat{R}}(\widehat{R}/\mathfrak{n}^i\widehat{R}), S) \supseteq \left(\prod_{j=1}^{d} \operatorname{Ann}_S \operatorname{Ext}_S^j(\widehat{R}, S) \cdot \operatorname{Ann}_S \operatorname{Ext}_S^j(\widehat{C}, S)\right)^{r+1},
\end{aligned}$$

and we obtain

$$\begin{aligned}
&\left(\prod_{j=1}^{d} \operatorname{Ann}_S \operatorname{Ext}_S^j(\widehat{R}, S) \cdot \operatorname{Ann}_S \operatorname{Ext}_S^j(\widehat{C}, S)\right)^{r+1} \\
&\qquad \subseteq \bigcap_{i>0} \operatorname{Ann}_S\big(\operatorname{Hom}_S(\widehat{R}, S)/\mathfrak{n}^i \operatorname{Hom}_S(\widehat{R}, S)\big) = \operatorname{Ann}_S \operatorname{Hom}_S(\widehat{R}, S),
\end{aligned}$$

where the equality follows from Lemma 4.8.

Let I be the kernel of the surjection $S \to \widehat{R}$. Since $\dim S = d = \dim \widehat{R}$, the ideal I of S has height zero. Hence there exists a minimal prime ideal $\mathfrak{p}$ of S which contains I. Since we have a ring epimorphism from the Artinian Gorenstein local ring $S_{\mathfrak{p}}$ to $\widehat{R}_{\mathfrak{p}}$, the $\widehat{R}_{\mathfrak{p}}$-module $\operatorname{Hom}_S(\widehat{R}, S)_{\mathfrak{p}} = \operatorname{Hom}_{S_{\mathfrak{p}}}(\widehat{R}_{\mathfrak{p}}, S_{\mathfrak{p}})$ is isomorphic to the injective hull of the residue field of $\widehat{R}_{\mathfrak{p}}$, which is in particular nonzero. This implies that $\mathfrak{p}$ contains the ideal $\operatorname{Ann}_S \operatorname{Hom}_S(\widehat{R}, S)$. Therefore, for some integer $1 \le l \le d$ the ideal $\mathfrak{p}$ contains either $\operatorname{Ann}_S \operatorname{Ext}_S^l(\widehat{R}, S)$ or $\operatorname{Ann}_S \operatorname{Ext}_S^l(\widehat{C}, S)$. If $\mathfrak{p}$ contains $\operatorname{Ann}_S \operatorname{Ext}_S^l(\widehat{R}, S)$, then we have $\operatorname{Ext}_{S_{\mathfrak{p}}}^l(\widehat{R}_{\mathfrak{p}}, S_{\mathfrak{p}}) \neq 0$, which contradicts the fact that $S_{\mathfrak{p}}$ is injective as an $S_{\mathfrak{p}}$-module. Similarly, we have a contradiction when $\mathfrak{p}$ contains $\operatorname{Ann}_S \operatorname{Ext}^l_S(\widehat{C}, S)$. This contradiction proves that $\operatorname{radius} \mathscr{X} = \infty$. □

Now we can show the essential part of Theorem 3.3(2)(3).

Theorem 4.10. *Let R be a local ring. Let $\mathscr{X}$ be a resolving subcategory of* $\operatorname{mod} R$. *One has* $\operatorname{radius} \mathscr{X} = \infty$ *if there exists a module $M \in \mathscr{X}$ with $0 < \operatorname{Gdim}_R M < \infty$ that satisfies either of the following conditions.*

(1) *$\Omega^{-2}\Omega^g M \in \mathscr{X}$, where $g = \operatorname{Gdim}_R M$.*

(2) *$\operatorname{res}(\Omega^n M)$ is a thick subcategory of $\mathscr{G}(R)$ for some $n \geq 0$.*

Proof. According to Proposition 4.9, it suffices to show that $\mathscr{X}$ contains a module of projective dimension one.

(1) We consider a construction whose idea essentially comes from the Auslander–Buchweitz approximation theorem [1989, Theorem 1.1]. There are exact sequences $0 \to \Omega^g M \to R^{\oplus a} \to \Omega^{g-1} M \to 0$ and $0 \to \Omega^g M \to R^{\oplus b} \to \Omega^{-1}\Omega^g M \to 0$, where the latter is possible as $\Omega^g M$ is a totally reflexive module. We make the following pushout diagram.

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & & \\
0 \arrow[r] & \Omega^g M \arrow[r] \arrow[d] & R^{\oplus a} \arrow[r] \arrow[d] & \Omega^{g-1} M \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & R^{\oplus b} \arrow[r] \arrow[d] & N \arrow[r] \arrow[d] & \Omega^{g-1} M \arrow[r] & 0 \\
 & \Omega^{-1}\Omega^g M \arrow[r, equal] \arrow[d] & \Omega^{-1}\Omega^g M \arrow[d] & & \\
 & 0 & 0 & &
\end{tikzcd}$$

As $\Omega^{-1}\Omega^g M$ is totally reflexive, we have $\operatorname{Ext}^1_R(\Omega^{-1}\Omega^g M, R) = 0$. Hence the second column in the above diagram splits, and we get an exact sequence

$$0 \to R^{\oplus b} \to R^{\oplus a} \oplus \Omega^{-1}\Omega^g M \to \Omega^{g-1} M \to 0.$$

There is an exact sequence $0 \to \Omega^{-1}\Omega^g M \to R^{\oplus c} \to \Omega^{-2}\Omega^g M \to 0$, and taking the direct sum with $0 \to R^{\oplus a} \xrightarrow{=} R^{\oplus a} \to 0 \to 0$, we have an exact sequence $0 \to R^{\oplus a} \oplus \Omega^{-1}\Omega^g M \to R^{\oplus(a+c)} \to \Omega^{-2}\Omega^g M \to 0$. Thus the following pushout diagram is obtained.

$$\begin{tikzcd}
 & & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & R^{\oplus b} \arrow[r] \arrow[d, equal] & R^{\oplus a} \oplus \Omega^{-1}\Omega^g M \arrow[r] \arrow[d] & \Omega^{g-1}M \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & R^{\oplus b} \arrow[r] & R^{\oplus(a+c)} \arrow[r] \arrow[d] & L \arrow[r] \arrow[d] & 0 \\
 & & \Omega^{-2}\Omega^g M \arrow[r, equal] \arrow[d] & \Omega^{-2}\Omega^g M \arrow[d] & \\
 & & 0 & 0 &
\end{tikzcd}$$

As $\Omega^{g-1}M$ and $\Omega^{-2}\Omega^g M$ are in $\mathcal{X}$, the module L is also in $\mathcal{X}$. The second row shows that L has projective dimension at most 1. Since $\Omega^{-2}\Omega^g M$ is totally reflexive but $\Omega^{g-1}M$ is not, it follows from the second column that L is nonfree. Therefore the projective dimension of L is equal to 1.

(2) Since by assumption $\operatorname{res}\Omega^n M$ is a subcategory of $\mathcal{G}(R)$, the module $\Omega^n M$ is totally reflexive. Hence $n \geq g := \mathsf{Gdim}_R M$.

We claim that $\operatorname{res}\Omega^n M = \operatorname{res}\Omega^g M$. In fact, since $n - g \geq 0$ and $\Omega^n M = \Omega^{n-g}(\Omega^g M)$, we observe that $\operatorname{res}\Omega^n M$ is contained in $\operatorname{res}\Omega^g M$. There is an exact sequence

$$0 \to \Omega^n M \to F_{n-1} \to F_{n-2} \to \cdots \to F_{g+1} \to F_g \to \Omega^g M \to 0$$

of totally reflexive R-modules with each F_i being free. Since $\operatorname{res}\Omega^n M$ is assumed to be thick in $\mathcal{G}(R)$, decomposing the above exact sequence into short exact sequences of totally reflexive modules, we see that $\Omega^g M$ belongs to $\operatorname{res}\Omega^n M$. Therefore $\operatorname{res}\Omega^g M$ is contained in $\operatorname{res}\Omega^n M$. Thus the claim follows.

Set $\mathcal{X} = \operatorname{res}\Omega^n M = \operatorname{res}\Omega^g M$. Our assumption implies that $\mathcal{X}$ is closed under cosyzygies, whence $\Omega^{-2}\Omega^g M \in \mathcal{X}$. By (1), we conclude that $\mathcal{X}$ contains a module of projective dimension 1. □

Now, using Remark 4.1 and Theorem 4.10, we deduce Theorem 3.3(2)(3).

***Proof of Theorem 3.3*(4).** We use the notion of a module of reducible complexity, which has been introduced by Bergh [2007]. Let us recall the definition.

Definition 4.11. The subcategory $\mathcal{C}_R^r$ of $\operatorname{mod} R$ is defined inductively as follows.

(1) Every module of finite projective dimension belongs to $\mathcal{C}_R^r$.

(2) A module M with $0 < \mathsf{cx}_R\, M < \infty$ belongs to $\mathscr{C}^r_R$ if there exists a homogeneous element $\eta \in \operatorname{Ext}^*_R(M, M)$ with $|\eta| > 0$ which is represented by a short exact sequence $0 \to M \to K \to \Omega^{|\eta|-1} M \to 0$ with $K \in \mathscr{C}^r_R$, $\mathsf{cx}\, K < \mathsf{cx}\, M$ and $\mathsf{depth}\, K = \mathsf{depth}\, M$.

An R-module is said to have *reducible complexity* if it is in $\mathscr{C}^r_R$.

The result below is shown in [Bergh 2007, Proposition 2.2(i)], which is implicitly stated in [Avramov et al. 1997].

Proposition 4.12. *Let R be a local ring. Every R-module of finite CI-dimension has reducible complexity.*

In a resolving subcategory, for any fixed integer $n \geq 0$, existence of modules of CI-dimension n is equivalent to existence of modules of projective dimension n.

Lemma 4.13. *Let R be a local ring. Let $\mathscr{X}$ be a resolving subcategory of* $\operatorname{mod} R$. *Suppose that there is a module $M \in \mathscr{X}$ such that* $\mathsf{CIdim}_R\, M < \infty$. *Then $\mathscr{X}$ contains a module N with* $\mathsf{pd}_R\, N = \mathsf{CIdim}_R\, M$.

Proof. Since M has finite CI-dimension, it has finite complexity. It follows from Proposition 4.12 that M has reducible complexity. If $\mathsf{cx}\, M = 0$, then $\mathsf{pd}\, M < \infty$, and we can take $N := M$. Hence we may assume $\mathsf{cx}\, M > 0$. There exists an exact sequence $0 \to M \to K \to \Omega^{|\eta|-1}_R M \to 0$ with $\mathsf{cx}\, K < \mathsf{cx}\, M$ and $\mathsf{depth}\, K = \mathsf{depth}\, M$, where η is a homogeneous element of $\operatorname{Ext}^*_R(M, M)$. We have $K \in \mathscr{X}$ and

$$\mathsf{CIdim}\, K = \mathsf{depth}\, R - \mathsf{depth}\, K = \mathsf{depth}\, R - \mathsf{depth}\, M = \mathsf{CIdim}\, M.$$

Replacing M with K and iterating this procedure, we can eventually arrive at a module $N \in \mathscr{X}$ with $\mathsf{CIdim}\, N = \mathsf{CIdim}\, M$ and $\mathsf{cx}\, N = 0$. The module N has finite projective dimension, and we have $\mathsf{pd}\, N = \mathsf{CIdim}\, N = \mathsf{CIdim}\, M$. □

Lemma 4.13 and Proposition 4.9 immediately yield the following theorem. This is not only the essential part of Theorem 3.3(4) but also a generalization of Proposition 4.9.

Theorem 4.14. *Let R be a local ring. Let $\mathscr{X}$ be a resolving subcategory of* $\operatorname{mod} R$. *Suppose that there exists a module $M \in \mathscr{X}$ with* $0 < \mathsf{CIdim}_R\, M < \infty$. *Then the radius of $\mathscr{X}$ is infinite.*

Theorem 3.3(4) now follows from Theorem 4.14 and Remark 4.1.

Another proof of Theorem 4.14. In the next theorem, we study the thickness of resolving subcategories of modules of CI-dimension at most zero. This will give another proof of Theorem 4.14.

Theorem 4.15. *Let R be a local ring.*

(1) *Let M be an R-module of CI-dimension at most zero. Then $\Omega^{-1}M$ belongs to* $\operatorname{res} M$.

(2) *Let $\mathscr{X}$ be a resolving subcategory of* $\operatorname{mod} R$. *Suppose that every module in $\mathscr{X}$ has CI-dimension at most zero. Then $\mathscr{X}$ is a thick subcategory of $\mathscr{G}(R)$.*

Proof. (1) Proposition 4.12 implies that M has reducible complexity. Let $K_0 = M$ and let K_{i+1} be a reduction in complexity of K_i for each $i \geq 0$. Then we have a short exact sequence $0 \to K_i \xrightarrow{f_i} K_{i+1} \to \Omega^{t_i-1}K_i \to 0$ with $t_i > 0$ (see also [Avramov et al. 1997, Proposition 7.2]), and eventually we must have $\operatorname{cx}_R K_e = 0$ for some $e \geq 0$. Then K_e has finite projective dimension. As $\operatorname{CIdim}_R M = 0$, we have $\operatorname{depth} K_e = \operatorname{depth} M = \operatorname{depth} R$. Therefore K_e is a free module. Note that the above exact sequence also shows that

$$K_i \in \operatorname{res} M \quad \text{for all } i \geq 0. \tag{4.15.1}$$

If $e = 0$, then M is free and we have $\Omega^{-1}M = 0 \in \operatorname{res} M$. So we may assume $e \geq 1$.

We claim that for each $0 \leq i \leq e-1$ the cokernel C_i of the composite map $f_i \cdots f_1 f_0 : M \to K_{i+1}$ belongs to $\operatorname{res} M$. Let us show this claim by induction on i. When $i = 0$, we have $C_i = \Omega^{t_0-1}M \in \operatorname{res} M$. Let $i \geq 1$. We have the following commutative diagram with exact rows and columns.

$$\begin{tikzcd}
 & & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & M \arrow[r, "f_{i-1}\cdots f_0"] \arrow[d, equal] & K_i \arrow[r] \arrow[d, "f_i"'] & C_{i-1} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & M \arrow[r] & K_{i+1} \arrow[r] \arrow[d] & C_i \arrow[r] \arrow[d] & 0 \\
 & & \Omega^{t_i-1}K_i \arrow[r, equal] \arrow[d] & \Omega^{t_i-1}K_i \arrow[d] & \\
 & & 0 & 0 &
\end{tikzcd}$$

The induction hypothesis implies that C_{i-1} belongs to $\operatorname{res} M$. Since $\Omega^{t_i-1}K_i$ is in $\operatorname{res} M$ by (4.15.1), the right column shows that C_i is also in $\operatorname{res} M$. Thus the claim follows.

Now we have a short exact sequence $0 \to M \xrightarrow{f_{e-1}\cdots f_1 f_0} K_e \to C_{e-1} \to 0$, where K_e is free and C_{e-1} is in $\operatorname{res} M$ by the claim. Since M is totally reflexive and $\mathscr{G}(R)$ is a resolving subcategory of $\operatorname{mod} R$, all the modules in $\operatorname{res} M$ are totally reflexive. Hence all modules appearing in the above exact sequence belong to $\mathscr{G}(R)$. It is easy to verify that there exists an isomorphism $C_{e-1} \cong \Omega^{-1}M \oplus F$ with F being free. Consequently, $\Omega^{-1}M$ belongs to $\operatorname{res} M$.

(2) By assumption, $\mathscr{X}$ is a subcategory of $\mathscr{G}(R)$. Thanks to Proposition 1.7, it is enough to show that $\mathscr{X}$ is closed under cosyzygies. Let M be an R-module in $\mathscr{X}$. Then M is of CI-dimension at most zero, and $\Omega^{-1}M$ belongs to $\operatorname{res} M$ by (1). Since $\mathscr{X}$ is resolving and contains M, it also contains $\operatorname{res} M$. Thus $\Omega^{-1}M$ is in $\mathscr{X}$. □

Theorem 4.15(2) immediately implies:

Corollary 4.16. *Let R be a local complete intersection. The following two are the same.*

- *A resolving subcategory of* $\operatorname{mod} R$ *contained in* $\mathrm{CM}(R)$.
- *A thick subcategory of* $\mathrm{CM}(R)$ *containing R.*

Now let us give another proof of Theorem 4.14. Let R be a local ring, and let $M \in \mathscr{X}$ be an R-module with $0 < \mathsf{CIdim}_R\, M < \infty$. Then we have $c := \mathsf{CIdim}_R\, M = \mathsf{Gdim}_R\, M$, and $\Omega^c M$ has CI-dimension zero. In particular, $\Omega^c M$ is totally reflexive, and hence so is $\Omega^{-1}\Omega^c M$. We have

$$0 = \mathsf{CIdim}_R(\Omega^c M) = \sup\{\mathsf{CIdim}_R(\Omega^{-1}\Omega^c M) - 1, 0\},$$

which especially says that $\Omega^{-1}\Omega^c M$ has finite CI-dimension. Therefore

$$\mathsf{CIdim}_R(\Omega^{-1}\Omega^c M) = \mathsf{Gdim}_R(\Omega^{-1}\Omega^c M) = 0,$$

and Theorem 4.15(1) yields $\Omega^{-2}\Omega^c M \in \operatorname{res}(\Omega^c M) \subseteq \mathscr{X}$. Now Theorem 4.10(1) implies that the radius of $\mathscr{X}$ is infinite.

5. Proof of Theorem II

In this section we prove the main Theorem II from the introduction. In fact, we can prove significantly more general statements (Theorems 5.7 and 5.11). In order to state and prove such results we need to first introduce a couple of definitions related to the concept of radius. In this section, an *n-th syzygy* $\Omega^n M$ of an R-module M means the image of the n-th differential map in a projective resolution of M in $\operatorname{mod} R$. (So it is not unique up to isomorphism but unique up to projective summands.)

Definition 5.1. Let $\mathscr{X}$, $\mathscr{Y}$ be subcategories of $\operatorname{mod} R$. We put $|\mathscr{X}| = \operatorname{add}\mathscr{X}$, and set $\mathscr{X} * \mathscr{Y} = \big||\mathscr{X}| \circ |\mathscr{Y}|\big|$. (The notation "$\circ$" was introduced in Definition 2.1.) For an integer $r > 0$, set

$$|\mathscr{X}|_r = \begin{cases} |\mathscr{X}| & \text{if } r = 1, \\ |\mathscr{X}|_{r-1} * \mathscr{X} & \text{if } r \geq 2. \end{cases}$$

Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be subcategories of $\operatorname{mod} R$. We observe that an object $M \in \operatorname{mod} R$ is in $\mathcal{X} * \mathcal{Y}$ if and only if there is an exact sequence $0 \to X \to E \to Y \to 0$ with $X \in |\mathcal{X}|$ and $Y \in |\mathcal{Y}|$ such that M is a direct summand of E. Also, one has $(\mathcal{X} * \mathcal{Y}) * \mathcal{Z} = \mathcal{X} * (\mathcal{Y} * \mathcal{Z})$ and $|\mathcal{X}|_a * |\mathcal{X}|_b = |\mathcal{X}|_{a+b}$ for all $a, b > 0$.

Definition 5.2. For a subcategory $\mathcal{X}$ of $\operatorname{mod} R$ we define $\operatorname{size} \mathcal{X}$ (respectively, $\operatorname{rank} \mathcal{X}$) to be the infimum of integers $n \geq 0$ such that $\mathcal{X} \subseteq |G|_{n+1}$ (respectively, $\mathcal{X} = |G|_{n+1}$) for some $G \in \operatorname{mod} R$.

It always holds that $\operatorname{size} \mathcal{X} \leq \operatorname{rank} \mathcal{X}$. Since $|\mathcal{X}|_n \subseteq [\mathcal{X}]_n$ for all $n > 0$, one has $\dim \mathcal{X} \geq \operatorname{radius} \mathcal{X} \leq \operatorname{size} \mathcal{X}$. If $\mathcal{X}$ is resolving, then $\dim \mathcal{X} \leq \operatorname{rank} \mathcal{X}$.

For an R-module M, we denote by $M^{\oplus}$ an object in $\operatorname{add}_R M$.

Proposition 5.3. *Let I be an ideal of R and let M be an R/I-module.*

(1) *There is an exact sequence* $0 \to I^{\oplus} \to \Omega_R M \to \Omega_{R/I} M \to 0$.

(2) *One has* $\Omega_R^n M \in \left|\Omega_{R/I}^n M \oplus \left(\bigoplus_{i=0}^{n-1} \Omega_R^i I\right)\right|_{n+1}$ *for all* $n \geq 0$.

Proof. (1) Take a surjection from a free R-module F to M. Then this factors through a surjection $F/IF \to M$. The assertion follows from this.

(2) We induce on n. Let $n > 0$. The induction hypothesis shows

$$\Omega_R^{n-1} \Omega_{R/I} M \in \left|\Omega_{R/I}^n M \oplus \left(\bigoplus_{i=0}^{n-2} \Omega_R^i I\right)\right|_n .$$

By (1) we have an exact sequence $0 \to \Omega_R^{n-1} I^{\oplus} \to \Omega_R^n M \to \Omega_R^{n-1} \Omega_{R/I} M \to 0$. Now the assertion follows. □

For $n \geq 0$ we denote by $\Omega_R^n(\operatorname{mod} R)$ the subcategory of $\operatorname{mod} R$ consisting of n-th syzygies of R-modules. For an ideal I of R, let $\Omega_R^n(\operatorname{mod} R/I)$ be the subcategory of $\operatorname{mod} R$ consisting of n-th syzygies of R-modules annihilated by I.

Corollary 5.4. *Let $d = \dim R < \infty$. Suppose that $R/\mathfrak{p}$ is regular for all $\mathfrak{p} \in \operatorname{Min} R$. Then* $\operatorname{size} \Omega^d(\operatorname{mod} R) < \infty$.

Proof. There is a filtration $R = I_0 \supsetneq I_1 \supsetneq \cdots \supsetneq I_n = 0$ of ideals of R such that for each i one has $I_i/I_{i+1} \cong R/\mathfrak{p}_i$ with $\mathfrak{p}_i \in \operatorname{Spec} R$. Choose a minimal prime $\mathfrak{q}_i$ contained in $\mathfrak{p}_i$. Let M be an R-module. Setting $M_i = I_i M/I_{i+1} M$, we have an exact sequence $0 \to \Omega_R^d(I_{i+1}M) \to \Omega_R^d(I_i M) \to \Omega_R^d M_i \to 0$. Note that each d-th syzygy $R/\mathfrak{q}_i$-module is free. Hence $\Omega_R^d M_i \in |R/\mathfrak{q}_i \oplus L_i|_{d+1}$ by Proposition 5.3(2), where $L_i := \bigoplus_{j=0}^{d-1} \Omega_R^j \mathfrak{q}_i$. Thus $\Omega_R^d M \in \left|\bigoplus_{i=1}^n (R/\mathfrak{q}_i \oplus L_i)\right|_{n(d+1)}$, which implies $\operatorname{size} \Omega^d(\operatorname{mod} R) < n(d+1) < \infty$. □

Corollary 5.5. *Let I be an ideal of R and $n \geq 0$ an integer. Then*

$$\operatorname{size} \Omega_R^n(\operatorname{mod} R/I) < (n+1)(\operatorname{size} \Omega_{R/I}^n(\operatorname{mod} R/I) + 1).$$

In particular, if $\mathsf{size}\,\Omega^n_{R/I}(\operatorname{mod} R/I)$ *is finite, then so is* $\mathsf{size}\,\Omega^n_R(\operatorname{mod} R/I)$.

Proof. This is a consequence of Proposition 5.3(2). □

Lemma 5.6. *Let M be an R-module. Let $x \in R$ be R-regular. Then*

$$\Omega^n_{R/xR}(\Omega_R M/x\Omega_R M) \cong \Omega^{n+1}_R M/x\Omega^{n+1}_R M$$

for any $n \geq 0$.

Proof. We use induction on n. Let $n > 0$. We have

$$\Omega^{n-1}_{R/xR}(\Omega M/x\Omega M) \cong \Omega^n M/x\Omega^n M$$

by the induction hypothesis, and hence

$$\Omega^n_{R/xR}(\Omega M/x\Omega M) \cong \Omega_{R/xR}(\Omega^n M/x\Omega^n M).$$

Note that x is $\Omega^n_R M$-regular. There is an exact sequence

$$0 \to \Omega^{n+1} M \to P \to \Omega^n M \to 0$$

of R-modules with P projective, which gives an exact sequence

$$0 \to \Omega^{n+1} M/x\Omega^{n+1} M \to P/xP \to \Omega^n M/x\Omega^n M \to 0.$$

Hence $\Omega^{n+1} M/x\Omega^{n+1} M \cong \Omega_{R/xR}(\Omega^n M/x\Omega^n M)$. □

Theorem 5.7. *Let $(R, \mathfrak{m})$ be a d-dimensional complete local ring with perfect coefficient field. Then one has* $\mathsf{size}\,\Omega^d(\operatorname{mod} R) < \infty$. *Hence* $\mathsf{radius}\,\Omega^d(\operatorname{mod} R) < \infty$.

Proof. We use induction on d. When $d = 0$, we have $\operatorname{mod} R = |k|_{\ell\ell(R)}$, hence $\mathsf{size}(\operatorname{mod} R) < \ell\ell(R) < \infty$. Let $d \geq 1$. Take a filtration $R = I_0 \supsetneq \cdots \supsetneq I_n = 0$ of ideals such that for each i one has $I_i/I_{i+1} \cong R/\mathfrak{p}_i$ with $\mathfrak{p}_i \in \operatorname{Spec} R$. Suppose that $\mathsf{size}\,\Omega^{d_i}_{R/\mathfrak{p}_i}(\operatorname{mod} R/\mathfrak{p}_i) < \infty$ for all i, where $d_i = \dim R/\mathfrak{p}_i$. Then we have

$$\mathsf{size}\,\Omega^d_{R/\mathfrak{p}_i}(\operatorname{mod} R/\mathfrak{p}_i) < \infty,$$

since $\Omega^d_{R/\mathfrak{p}_i}(\operatorname{mod} R/\mathfrak{p}_i)$ is contained in $\Omega^{d_i}_{R/\mathfrak{p}_i}(\operatorname{mod} R/\mathfrak{p}_i)$. Corollary 5.5 implies $\mathsf{size}\,\Omega^d_R(\operatorname{mod} R/\mathfrak{p}_i) < \infty$. For each R-module M there is an exact sequence $0 \to I_{i+1}M \to I_i M \to I_i M/I_{i+1}M \to 0$, which gives an exact sequence

$$0 \to \Omega^d_R(I_{i+1}M) \to \Omega^d_R(I_i M) \to \Omega^d_R(I_i M/I_{i+1}M) \to 0.$$

As $\Omega^d_R(I_i M/I_{i+1}M)$ is in $\Omega^d_R(\operatorname{mod} R/\mathfrak{p}_i)$, we have $\mathsf{size}\,\Omega^d_R(\operatorname{mod} R) < \infty$. Thus, we may assume R is a domain.

If R is regular, then $\Omega^d(\operatorname{mod} R) = |R|_1$ and $\mathsf{size}\,\Omega^d(\operatorname{mod} R) = 0 < \infty$, so we may assume that R is singular. By [Wang 1994, 5.15] there is an ideal $J \subseteq \mathfrak{m}$ with

$\operatorname{Sing} R = \mathrm{V}(J)$ and $J \operatorname{Ext}_R^{d+1}(\operatorname{mod} R, \operatorname{mod} R) = 0$. Since R is a domain, we find an element $0 \neq x \in J$. The induction hypothesis guarantees

$$\Omega_{R/xR}^{d-1}(\operatorname{mod} R/xR) \subseteq |G|_n^{R/xR}$$

for some R/xR-module G and an integer $n > 0$. Let M be an R-module and put $N = \Omega_R^d M$. Note that x is N-regular as $d > 0$. Hence N is isomorphic to a direct summand of $\Omega_R(N/xN)$ (see [Takahashi 2010, Lemma 2.1]). In view of Lemma 5.6, we have

$$N/xN \cong \Omega_{R/xR}^{d-1}(\Omega_R M/x\Omega_R M) \in \Omega_{R/xR}^{d-1}(\operatorname{mod} R/xR) \subseteq |G|_n^{R/xR}.$$

Hence N/xN is in $|G|_n^R$, which implies $\Omega_R(N/xN) \in |\Omega_R G|_n^R$. Therefore N belongs to $|\Omega_R G|_n^R$, and we obtain $\Omega^d(\operatorname{mod} R) \subseteq |\Omega_R G|_n^R$. It now follows that $\operatorname{size} \Omega^d(\operatorname{mod} R) < \infty$. □

Lemma 5.8. *Let*

$$0 \to M \to C^0 \to C^1 \to \cdots \to C^{n-1} \to N \to 0$$

be an exact sequence in mod R *with* $n \geq 0$. *Then* M *is in* $\left|\Omega^n N \oplus \left(\bigoplus_{i=0}^{n-1} \Omega^i C^i\right)\right|_{n+1}$.

Proof. We induce on n. The case $n = 0$ is trivial, so let $n \geq 1$. There are two exact sequences $0 \to M \to C^0 \to L \to 0$ and $0 \to L \to C^1 \to \cdots \to C^{n-1} \to N \to 0$. The induction hypothesis shows $L \in \left|\Omega^{n-1} N \oplus \left(\bigoplus_{i=0}^{n-2} \Omega^i C^{i+1}\right)\right|_n$. A pullback diagram makes an exact sequence $0 \to \Omega L \to M \oplus R^{\oplus} \to C^0 \to 0$. Since ΩL belongs to $\left|\Omega^n N \oplus \left(\bigoplus_{i=0}^{n-2} \Omega^{i+1} C^{i+1}\right)\right|_n$, we see that M is in $\left|\Omega^n N \oplus \left(\bigoplus_{i=0}^{n-1} \Omega^i C^i\right)\right|_{n+1}$. □

Corollary 5.9. *Let R be a Cohen–Macaulay complete local ring with perfect coefficient field. Then* $\operatorname{size} \operatorname{CM}(R) < \infty$.

Proof. As R is complete, it admits a canonical module ω. Theorem 5.7 implies $\operatorname{size} \Omega^d(\operatorname{mod} R) =: n < \infty$, so we have $\Omega^d(\operatorname{mod} R) \subseteq |G|_{n+1}$ for some R-module G. Let M be a Cohen–Macaulay R-module. Then there exists an exact sequence $0 \to M \to \omega^{\oplus s_0} \to \cdots \to \omega^{\oplus s_{d-1}} \to N \to 0$. It follows from Lemma 5.8 that M is in $|\Omega^d N \oplus W|_{d+1}$, where $W := \bigoplus_{i=0}^{d-1} \Omega^i \omega$. Since $\Omega^d N \in |G|_{n+1}$, we have $M \in |G \oplus W|_{(n+1)(d+1)}$. Thus $\operatorname{size} \operatorname{CM}(R) < (n+1)(d+1) < \infty$. □

Proposition 5.10. *Let R be a Cohen–Macaulay local ring with a canonical module ω.*

(1) $\operatorname{rank} \operatorname{CM}(R) < (\dim R + 1)(\operatorname{size} \operatorname{CM}(R) + 1)$.

(2) $\dim \operatorname{CM}(R) < (\dim R + 1)(\operatorname{radius} \operatorname{CM}(R) + 1)$.

In particular, one has

$$\operatorname{rank} \operatorname{CM}(R) < \infty \Leftrightarrow \operatorname{size} \operatorname{CM}(R) < \infty \Rightarrow \dim \operatorname{CM}(R) < \infty \Leftrightarrow \operatorname{radius} \operatorname{CM}(R) < \infty.$$

Proof. (1) Let $n = \mathsf{size}\,\mathrm{CM}(R)$. We find an R-module G with $\mathrm{CM}(R) \subseteq |G|_{n+1}$. Let $d = \mathsf{dim}\, R$ and $M \in \mathrm{CM}(R)$. Similarly to the proof of Corollary 5.9, there exists $N \in \mathrm{CM}(R)$ such that M is in $|\Omega^d N \oplus W|_{d+1}$, where $W := \bigoplus_{i=0}^{d-1} \Omega^i \omega \in \mathrm{CM}(R)$. Note that $\Omega^d N \in |\Omega^d G|_{n+1}$. Thus we obtain

$$\mathrm{CM}(R) = |\Omega^d G \oplus W|_{(n+1)(d+1)},$$

and $\mathsf{rank}\,\mathrm{CM}(R) < (n+1)(d+1)$.

(2) In the proof of (1), replace "size", "rank" and "$|\ |$" with "radius", "dim" and "$[\ ]$", respectively. □

Theorem 5.11. *Let R be a Cohen–Macaulay local ring admitting perfect coefficient field. Assume that either of the following holds.*

(1) *R is complete.*

(2) *R is excellent with an isolated singularity.*

Then one has $\mathsf{rank}\,\mathrm{CM}(R) < \infty$.

Proof. (1) This assertion follows from Corollary 5.9 and Proposition 5.10(1).

(2) It follows by (1) that $\mathrm{CM}(\widehat{R})$ has finite rank, where $\widehat{R}$ denotes the completion of R. One can prove that $\mathrm{CM}(R)$ also has finite rank, making an argument similar to [Dao and Takahashi 2012, Remark 6.5]:

Putting $n = \mathsf{rank}\,\mathrm{CM}(\widehat{R})$, we have $\mathrm{CM}(\widehat{R}) = |C|_{n+1}$ for some $C \in \mathrm{CM}(\widehat{R})$. Since R is Cohen–Macaulay, excellent and with an isolated singularity, we can apply [Takahashi 2010, Corollary 3.6] to see that C is isomorphic to a direct summand of $\widehat{G}$ for some $G \in \mathrm{CM}(R)$. Hence the equality $\mathrm{CM}(\widehat{R}) = |\widehat{G}|_{n+1}$ holds.

Claim. Let $m > 0$. For any $N \in |\widehat{G}|_m$ there exists $M \in |G|_m$ such that N is isomorphic to a direct summand of $\widehat{M}$.

To show this claim, we use induction on m. As R is an isolated singularity, for all $X, Y \in \mathrm{CM}(R)$ the R-module $\mathrm{Ext}_R^1(X, Y)$ has finite length. Hence there are isomorphisms $\mathrm{Ext}_R^1(X, Y) \cong \mathrm{Ext}_R^1(X, Y)^{\widehat{}} \cong \mathrm{Ext}_{\widehat{R}}^1(\widehat{X}, \widehat{Y})$, which imply that every short exact sequence $0 \to \widehat{Y} \to E \to \widehat{X} \to 0$ of $\widehat{R}$-modules is isomorphic to the completion of some short exact sequence $0 \to Y \to E' \to X \to 0$ of R-modules. The claim follows from this. Using this claim and [Aihara and Takahashi 2011, Lemma 5.7], we observe that $\mathrm{CM}(R) = |G|_{n+1}$ holds. Therefore $\mathsf{rank}\,\mathrm{CM}(R) \le n < \infty$. □

6. Some discussions and open questions

In this section we relate our results to the uniform Auslander condition and discuss some open questions. For a local ring R, Jorgensen and Şega [2004] introduced the *uniform Auslander condition*:

(UAC): There exists an integer n such that for all R-modules M, N with $\operatorname{Ext}_R^i(M, N) = 0$ for all $i \gg 0$ one has $\operatorname{Ext}_R^i(M, N) = 0$ for all $i \geq n$.

It is known that this condition is satisfied if the local ring R is a complete intersection, a Golod ring, a Gorenstein ring with $\operatorname{mult} R = \operatorname{codim} R + 2$, or a Gorenstein ring with $\operatorname{codim} R \leq 4$. Here $\operatorname{mult} R$ denotes the multiplicity of R. These are proved in [Jorgensen and Şega 2004, Proposition 1.4], [Avramov and Buchweitz 2000, Theorem 4.7], [Huneke and Jorgensen 2003, Theorem 3.5] and [Şega 2003, Theorem 3.4], respectively. More information can be found in [Christensen and Holm 2010, Appendix A]. On the other hand, there exists an example of a Gorenstein local ring which does not satisfy (UAC); see [Jorgensen and Şega 2004, Theorem in §0].

The result below says that over a Gorenstein local ring the condition (UAC) is closely related to the thickness of resolving subcategories of Cohen–Macaulay modules.

Proposition 6.1. *Let R be a Gorenstein local ring. Assume every resolving subcategory of* $\operatorname{mod} R$ *contained in* $\operatorname{CM}(R)$ *is a thick subcategory of* $\operatorname{CM}(R)$. *Then R satisfies* (UAC).

Proof. Let $t \geq 0$ be an integer, and let M, N be R-modules with $\operatorname{Ext}_R^i(M, N) = 0$ for all $i > t$. We define a subcategory $\mathcal{X}$ of $\operatorname{mod} R$ to consist of all Cohen–Macaulay R-modules X satisfying $\operatorname{Ext}_R^i(X, N) = 0$ for all $i > t$. Then $\mathcal{X}$ is a resolving subcategory of $\operatorname{mod} R$ contained in $\operatorname{CM}(R)$. By assumption, $\mathcal{X}$ is a thick subcategory of $\operatorname{CM}(R)$. Set $d = \dim R$. Since $\Omega^d M$ is in $\mathcal{X}$, so is $\Omega^{-t}\Omega^d M$. We have

$$\begin{aligned}\operatorname{Ext}_R^i(M, N) &\cong \operatorname{Ext}_R^{i-d}(\Omega^d M, N) \cong \operatorname{Ext}_R^{i-d}(\Omega^t(\Omega^{-t}\Omega^d M), N)\\ &\cong \operatorname{Ext}_R^{i-d+t}(\Omega^{-t}\Omega^d M, N) = 0\end{aligned}$$

for all integers $i > d$. □

There is also a connection between thickness of resolving subcategories of totally reflexive modules and closure under R-duals. Here we say that a subcategory $\mathcal{X}$ of $\operatorname{mod} R$ is *closed under R-duals* if for each module M in $\mathcal{X}$ its R-dual M^* is also in $\mathcal{X}$.

Proposition 6.2. (1) *Let R be local. Let $\mathcal{X}$ be a resolving subcategory of* $\operatorname{mod} R$ *contained in $\mathcal{G}(R)$. If $\mathcal{X}$ is closed under R-duals, then $\mathcal{X}$ is a thick subcategory of $\mathcal{G}(R)$.*

(2) *Let R be a local hypersurface. Then every resolving subcategory of* $\operatorname{mod} R$ *contained in* $\operatorname{CM}(R)$ *is closed under R-duals.*

Proof. (1) According to Proposition 1.7, we have only to show that $\mathcal{X}$ is closed under cosyzygies. Let $X \in \mathcal{X}$. There is an exact sequence $0 \to \Omega(X^*) \to F \to X^* \to 0$, where F is free. Dualizing this by R, we get an exact sequence

$$0 \to X \to F^* \to (\Omega(X^*))^* \to 0.$$

Note that $(\Omega(X^*))^*$ is totally reflexive. We easily see that $(\Omega(X^*))^*$ is isomorphic to $\Omega^{-1}X$ up to free summand. As $\mathscr{X}$ is a resolving subcategory closed under R-duals, $(\Omega(X^*))^*$ belongs to $\mathscr{X}$, and so does $\Omega^{-1}X$.

(2) It follows from [Takahashi 2010, main theorem] that every resolving subcategory of $\operatorname{mod} R$ contained in $\mathrm{CM}(R)$ can be described as $\mathrm{NF}_{\mathrm{CM}}^{-1}(W)$, where W is a specialization-closed subset of $\operatorname{Spec} R$ contained in $\operatorname{Sing} R$. If M is an R-module in $\mathrm{NF}_{\mathrm{CM}}^{-1}(W)$, then we have $\mathrm{NF}(M^*) = \mathrm{NF}(M) \subseteq W$, which shows that $\mathrm{NF}_{\mathrm{CM}}^{-1}(W)$ also contains M^*. □

Now we have reached the following question.

Question 6.3. Let R be a Gorenstein local ring. Let us consider the following five conditions.

(1) R is a complete intersection.

(2) Every resolving subcategory of $\operatorname{mod} R$ contained in $\mathrm{CM}(R)$ is closed under R-duals.

(3) Every resolving subcategory of $\operatorname{mod} R$ contained in $\mathrm{CM}(R)$ is a thick subcategory of $\mathrm{CM}(R)$.

(4) R satisfies (UAC).

(5) Conjecture 3.1 is true for R.

We know that the implications (2) ⇒ (3) and (1) ⇒ (3) ⇒ (4) hold by Propositions 6.1, 6.2(1) and Corollary 4.16. The implication (1) ⇒ (2) is also true if R is a hypersurface by Proposition 6.2(2). Very recently, motivated by the first version of the present paper, Stevenson [2013a] proved that the implication (1) ⇒ (2) holds in the case where R is a quotient of a regular local ring. Corollary 3.4 says that (3) ⇒ (5) holds. How about the other implications among these five conditions?

Remark 6.4. According to a recent preprint by Stevenson [2013b] (see also [Iyengar 2009]), if R is a quotient of a regular local ring by a regular sequence, then one can classify the thick subcategories of $\underline{\mathrm{CM}}(R)$ in terms of “support varieties”. Thus, one can also classify the resolving subcategories of $\operatorname{mod} R$ contained in $\mathrm{CM}(R)$ by using Corollary 4.16 and [Takahashi 2010, Proposition 6.2]. In relation to this, the resolving subcategories over a regular ring can be classified completely. This classification theorem is stated and proved in [Dao and Takahashi 2013].

Acknowledgments

We would like to thank Luchezar Avramov, Jesse Burke, Craig Huneke, Osamu Iyama, Srikanth Iyengar and Greg Stevenson for their valuable comments. This work was done during the visits of Takahashi to University of Kansas in May, July and August, 2011, and July, 2012. He is grateful for their kind hospitality. We also thank the referee for a careful reading and helpful comments.

References

[Aihara and Takahashi 2011] T. Aihara and R. Takahashi, "Generators and dimensions of derived categories", preprint, 2011. arXiv 1106.0205

[Araya et al. 2012] T. Araya, K.-i. Iima, and R. Takahashi, "On the structure of Cohen–Macaulay modules over hypersurfaces of countable Cohen–Macaulay representation type", *J. Algebra* **361** (2012), 213–224. MR 2921619 Zbl 1275.13009

[Auslander 1967] M. Auslander, *Anneaux de Gorenstein, et torsion en algèbre commutative*, Secrétariat mathématique, Paris, 1967. MR 37 #1435 Zbl 0157.08301

[Auslander and Bridger 1969] M. Auslander and M. Bridger, *Stable module theory*, Memoirs of the American Mathematical Society **94**, Amer. Math. Soc., Providence, R.I., 1969. MR 42 #4580 Zbl 0204.36402

[Auslander and Buchweitz 1989] M. Auslander and R.-O. Buchweitz, "The homological theory of maximal Cohen–Macaulay approximations", pp. 5–37 Mém. Soc. Math. France (N.S.) **38**, 1989. MR 91h:13010 Zbl 0697.13005

[Avramov and Buchweitz 2000] L. L. Avramov and R.-O. Buchweitz, "Support varieties and cohomology over complete intersections", *Invent. Math.* **142**:2 (2000), 285–318. MR 2001j:13017 Zbl 0999.13008

[Avramov et al. 1997] L. L. Avramov, V. N. Gasharov, and I. V. Peeva, "Complete intersection dimension", *Inst. Hautes Études Sci. Publ. Math.* 86 (1997), 67–114. MR 99c:13033 Zbl 0918.13008

[Avramov et al. 2010] L. L. Avramov, R.-O. Buchweitz, S. B. Iyengar, and C. Miller, "Homology of perfect complexes", *Adv. Math.* **223**:5 (2010), 1731–1781. MR 2011k:13014 Zbl 1186.13006

[Bergh 2007] P. A. Bergh, "Modules with reducible complexity", *J. Algebra* **310**:1 (2007), 132–147. MR 2008g:13021 Zbl 1117.13016

[Buchweitz 1986] R. O. Buchweitz, "Maximal Cohen–Macaulay modules and Tate–cohomology over Gorenstein rings", preprint, 1986, Available at http://hdl.handle.net/1807/16682.

[Christensen 2000] L. W. Christensen, *Gorenstein dimensions*, Lecture Notes in Mathematics **1747**, Springer, Berlin, 2000. MR 2002e:13032 Zbl 0965.13010

[Christensen and Holm 2010] L. W. Christensen and H. Holm, "Algebras that satisfy Auslander's condition on vanishing of cohomology", *Math. Z.* **265**:1 (2010), 21–40. MR 2011c:16033 Zbl 1252.16008

[Dao and Takahashi 2012] H. Dao and R. Takahashi, "The dimension of a subcategory of modules", preprint, 2012. arXiv 1203.1955

[Dao and Takahashi 2013] H. Dao and R. Takahashi, "Classification of resolving subcategories and grade consistent functions", 2013. To appear in *Int. Math. Res. Not.*

[Happel 1988] D. Happel, *Triangulated categories in the representation theory of finite-dimensional algebras*, London Mathematical Society Lecture Note Series **119**, Cambridge University Press, 1988. MR 89e:16035 Zbl 0635.16017

[Huneke and Jorgensen 2003] C. Huneke and D. A. Jorgensen, "Symmetry in the vanishing of Ext over Gorenstein rings", *Math. Scand.* **93**:2 (2003), 161–184. MR 2004k:13039 Zbl 1062.13005

[Iyengar 2009] S. B. Iyengar, "Stratifying derived categories associated to finite groups and commutative rings, Kyoto RIMS Workshop on Algebraic Triangulated Categories and Related Topics", Lecture notes, 2009, Available at http://www.math.unl.edu/~siyengar2/Papers/RIMS0709.pdf.

[Jorgensen and Şega 2004] D. A. Jorgensen and L. M. Şega, "Nonvanishing cohomology and classes of Gorenstein rings", *Adv. Math.* **188**:2 (2004), 470–490. MR 2005f:13017 Zbl 1090.13009

[Minamoto 2013] H. Minamoto, "A note on dimension of triangulated categories", *Proc. Amer. Math. Soc.* **141**:12 (2013), 4209–4214. MR 3105864 Zbl 06218150

[Rouquier 2008] R. Rouquier, "Dimensions of triangulated categories", *J. K-Theory* **1**:2 (2008), 193–256. MR 2009i:18008 Zbl 1165.18008

[Şega 2003] L. M. Şega, "Vanishing of cohomology over Gorenstein rings of small codimension", *Proc. Amer. Math. Soc.* **131**:8 (2003), 2313–2323. MR 2004b:13016 Zbl 1017.13008

[Stevenson 2013a] G. Stevenson, "Duality for bounded derived categories of complete intersections", 2013. To appear in *Bull. London Math. Soc.*

[Stevenson 2013b] G. Stevenson, "Subcategories of singularity categories via tensor actions", 2013. To appear in *Compos. Math.*

[Takahashi 2009] R. Takahashi, "Modules in resolving subcategories which are free on the punctured spectrum", *Pacific J. Math.* **241**:2 (2009), 347–367. MR 2010b:13027 Zbl 1172.13005

[Takahashi 2010] R. Takahashi, "Classifying thick subcategories of the stable category of Cohen–Macaulay modules", *Adv. Math.* **225**:4 (2010), 2076–2116. MR 2011h:13014 Zbl 1202.13009

[Takahashi 2013] R. Takahashi, "Classifying resolving subcategories over a Cohen–Macaulay local ring", *Math. Z.* **273**:1-2 (2013), 569–587. MR 3010176 Zbl 1267.13024

[Wang 1994] H.-J. Wang, "On the Fitting ideals in free resolutions", *Michigan Math. J.* **41**:3 (1994), 587–608. MR 96b:13013 Zbl 0822.13007

[Yoshino 2005] Y. Yoshino, "A functorial approach to modules of G-dimension zero", *Illinois J. Math.* **49**:2 (2005), 345–367. MR 2006e:13014 Zbl 1097.13019

[Yoshiwaki 2011] M. Yoshiwaki, "On self-injective algebras of stable dimension zero", *Nagoya Math. J.* **203** (2011), 101–108. MR 2012h:16009 Zbl 1227.16016

Communicated by Kei-Ichi Watanabe
Received 2012-07-14 Revised 2013-08-10 Accepted 2013-09-14

hdao@math.ku.edu
Department of Mathematics, University of Kansas, 405 Snow Hall, 1460 Jayhawk Blvd, Lawrence, KS 66045, United States
http://www.math.ku.edu/~hdao/

takahashi@math.nagoya-u.ac.jp
Department of Mathematics, University of Nebraska, Lincoln, NE 68588-0130, United States

Current address:
Graduate School of Mathematics, Nagoya University, Furocho, Chikusaku, Nagoya 464-8602, Japan
http://www.math.nagoya-u.ac.jp/~takahashi/

mathematical sciences publishers msp

ALGEBRA AND NUMBER THEORY 8:1 (2014)
dx.doi.org/10.2140/ant.2014.8.173

A generalized Bogomolov–Gieseker inequality for the three-dimensional projective space

Emanuele Macrì

A generalized Bogomolov–Gieseker inequality for tilt-stable complexes on a smooth projective threefold was conjectured by Bayer, Toda, and the author. We show that such inequality holds true in general if it holds true when the polarization is sufficiently small. As an application, we prove it for the three-dimensional projective space.

1. Introduction

The notion of tilt-stability, for objects in the derived category of a smooth projective threefold, was introduced in [Bayer et al. 2011b], based on [Bridgeland 2008; Arcara and Bertram 2013]. In [Bayer et al. 2011b, Conjecture 1.3.1] (Conjecture 2.3 of the present paper), we proposed a generalized Bogomolov–Gieseker inequality (BG inequality, for short) for tilt-stable objects. The main application for tilt-stability was to have an auxiliary notion of stability to construct Bridgeland stability conditions. The generalized BG inequality is precisely the missing ingredient to being able to show the existence of Bridgeland stability conditions.

In this note, we prove such inequality in the case of the projective space $\mathbb{P}^3$.

Theorem 1.1. *The generalized Bogomolov–Gieseker inequality for tilt-stable objects in* $\mathrm{D}^{\mathrm{b}}(\mathbb{P}^3)$ *holds.*

This gives the first example when the generalized BG inequality is proved in full generality. As a corollary, by [Bayer et al. 2011b], we can also describe a large open subset of the space of stability conditions on $\mathrm{D}^{\mathrm{b}}(\mathbb{P}^3)$. It would be very interesting to study how moduli spaces of Bridgeland semistable objects vary when varying the stability condition (very much like the situation described in [Arcara et al. 2013; Maciocia and Meachan 2013; Lo and Qin 2011; Minamide et al. 2011; Yanagida and Yoshioka 2012; Bayer and Macrì 2012; Toda 2012b; Yoshioka 2012] for the

MSC2010: primary 14F05; secondary 18E30, 14J30.
Keywords: Bridgeland stability conditions, derived category, Bogomolov–Gieseker inequality.

case of surfaces). The behavior at the "large volume limit point" is described in [Bayer et al. 2011b, Section 6].

The idea of the proof of Theorem 1.1 goes as follows. For a smooth projective threefold X, the notion of tilt-stability depends on two parameters, namely two divisor classes $B, \omega \in \mathrm{NS}_{\mathbb{R}}(X)$ with ω ample. In this paper, we prove a general result, Proposition 2.7: showing the generalized BG inequality for all B and ω can always be reduced to showing it for ω "arbitrarily small", uniformly in B.

For $X = \mathbb{P}^3$, the case in which ω is small was essentially proved in [Bayer et al. 2011b, Theorem 8.2.1]. More precisely, for simplicity, in [Bayer et al. 2011b], only the case $B = 0$ was considered. Proposition 3.1 generalizes that argument to arbitrary B. Together with Proposition 2.7, this completes the proof of Theorem 1.1.

The interest for a general proof of the generalized BG inequality, besides for the existence of Bridgeland stability conditions, relies on its consequences. Indeed, if we assume such inequality to be true, we would have

- a proof of Fujita's conjecture for threefolds [Bayer et al. 2011a],
- a mathematical formulation of Denef and Moore's formula derived in the study of Ooguri, Strominger, and Vafa's conjecture, relating black-hole entropy and topological string [Toda 2013a], and
- the possibility to realize extremal contractions for threefolds as moduli spaces of semistable objects in the derived category [Toda 2013b].

We also mention that in the paper [Polishchuk 2012] the existence of Bridgeland stability conditions on abelian threefolds is tested on a class of objects (called Lagrangian-Invariant objects).

Finally, in [Bayer et al. 2011b], a strict relation between the generalized BG inequality and Castelnuovo's inequality for curves in $\mathbb{P}^3$ was pointed out. In Section 4 of this paper, we show that Theorem 1.1 gives, as an immediate corollary, a weaker version of Castelnuovo's theorem [Hartshorne 1977, IV, 6.4].

A survey on Bridgeland stability conditions and further problems and applications can be found in [Bridgeland 2009; Bayer 2011; Huybrechts 2012; Toda 2012a].

Notation. In this paper, we will always denote by X a smooth projective threefold over the complex numbers and by $\mathrm{D}^{\mathrm{b}}(X)$ its bounded derived category of coherent sheaves. The Chow groups of X modulo numerical equivalence are denoted by $\mathrm{Num}(X)$. In particular, the Néron–Severi group $\mathrm{NS}(X) = \mathrm{Num}^1(X)$. For an abelian group G and a field k ($= \mathbb{Q}, \mathbb{R}, \mathbb{C}$), we denote by G_k the k-vector space $G \otimes k$.

2. The reduction argument

In this section, we give a brief recall on the notion of tilt stability, following [Bayer et al. 2011b]. We show how to reduce the proof of the generalized Bogomolov–Gieseker inequality proposed in [Bayer et al. 2011b, Conjecture 1.3.1] (whose statement is recalled in Conjecture 2.3 below), when ω and B are "parallel", to the case in which the polarization is "sufficiently small".

2A. ***Tilt stability.*** Let X be a smooth projective threefold over $\mathbb{C}$, and let $H \in \mathrm{NS}(X)$ be an ample divisor class. For a pair

$$\begin{aligned} \omega &= \alpha \cdot H, \quad \alpha \in \mathbb{R}_{>0}, \\ B &= \beta \cdot H, \quad \beta \in \mathbb{R}, \end{aligned}$$

we define a slope function $\mu_{\omega,B}$ for coherent sheaves on X in the usual way: for $E \in \mathrm{Coh}(X)$, we set

$$\mu_{\omega,B}(E) = \begin{cases} +\infty & \text{if } \mathrm{ch}_0^B(E) = 0, \\ \dfrac{\omega^2 \,\mathrm{ch}_1^B(E)}{\omega^3 \,\mathrm{ch}_0^B(E)} & \text{otherwise,} \end{cases}$$

where $\mathrm{ch}^B(E) = e^{-B}\,\mathrm{ch}(E)$ denotes the Chern character twisted by B. Explicitly,

$$\begin{aligned} \mathrm{ch}_0^B &= \mathrm{ch}_0, & \mathrm{ch}_2^B &= \mathrm{ch}_2 - B\,\mathrm{ch}_1 + \tfrac{1}{2}B^2\,\mathrm{ch}_0, \\ \mathrm{ch}_1^B &= \mathrm{ch}_1 - B\,\mathrm{ch}_0, & \mathrm{ch}_3^B &= \mathrm{ch}_3 - B\,\mathrm{ch}_2 + \tfrac{1}{2}B^2\,\mathrm{ch}_1 - \tfrac{1}{6}B^3\,\mathrm{ch}_0\,. \end{aligned}$$

A coherent sheaf E is slope-(semi)stable (or $\mu_{\omega,B}$-(semi)stable) if, for all subsheaves $F \hookrightarrow E$, we have

$$\mu_{\omega,B}(F) < (\leq)\ \mu_{\omega,B}(E/F).$$

Due to the existence of Harder–Narasimhan filtrations (HN-filtrations, for short) with respect to slope-stability, there exists a *torsion pair* $(\mathcal{T}_{\omega,B}, \mathcal{F}_{\omega,B})$ defined as follows:

$$\begin{aligned} \mathcal{T}_{\omega,B} &= \{E \in \mathrm{Coh}\,X : \text{any quotient } E \twoheadrightarrow G \text{ satisfies } \mu_{\omega,B}(G) > 0\}, \\ \mathcal{F}_{\omega,B} &= \{E \in \mathrm{Coh}\,X : \text{any subsheaf } F \hookrightarrow E \text{ satisfies } \mu_{\omega,B}(F) \leq 0\}. \end{aligned}$$

Equivalently, $\mathcal{T}_{\omega,B}$ and $\mathcal{F}_{\omega,B}$ are the extension-closed subcategories of $\mathrm{Coh}\,X$ generated by slope-stable sheaves of positive or nonpositive slope, respectively.

Definition 2.1. We let $\mathrm{Coh}^{\omega,B}(X) \subset \mathrm{D}^{\mathrm{b}}(X)$ be the extension-closure

$$\mathrm{Coh}^{\omega,B}(X) = \langle \mathcal{T}_{\omega,B}, \mathcal{F}_{\omega,B}[1] \rangle.$$

The category $\mathrm{Coh}^{\omega,B}(X)$ depends only on ω via H. Hence, to simplify notation, since for us B is also a multiple of H, we denote it by $\mathrm{Coh}^B(X)$. By the general theory of torsion pairs and tilting [Happel et al. 1996], $\mathrm{Coh}^B(X)$ is the heart of a bounded t-structure on $\mathrm{D}^{\mathrm{b}}(X)$.

By using the classical Bogomolov–Gieseker inequality and Hodge index theorem, we can define the following slope function on $\mathrm{Coh}^B(X)$: for $E \in \mathrm{Coh}^B(X)$, we set

$$\nu_{\omega,B}(E) = \begin{cases} +\infty & \text{if } \omega^2 \operatorname{ch}_1^B(E) = 0, \\ \dfrac{\omega \operatorname{ch}_2^B(E) - \frac{1}{2}\omega^3 \operatorname{ch}_0^B(E)}{\omega^2 \operatorname{ch}_1^B(E)} & \text{otherwise.} \end{cases}$$

Definition 2.2. An object $E \in \mathrm{Coh}^B(X)$ is *tilt-(semi)stable* if, for all nontrivial subobjects $F \hookrightarrow E$, we have

$$\nu_{\omega,B}(F) < (\leq)\ \nu_{\omega,B}(E/F).$$

The following is our main conjecture:

Conjecture 2.3 [Bayer et al. 2011b, Conjecture 1.3.1]. *For any $\nu_{\omega,B}$-semistable object $E \in \mathrm{Coh}^B(X)$ satisfying $\nu_{\omega,B}(E) = 0$, we have the following generalized Bogomolov–Gieseker inequality*:

$$\operatorname{ch}_3^B(E) \leq \tfrac{1}{6}\omega^2 \operatorname{ch}_1^B(E). \tag{1}$$

The original definition of tilt-stability in [Bayer et al. 2011b] was given when $\alpha, \beta \in \mathbb{Q}$ (actually it was slightly more general, allowing ω and B to be arbitrary, and ω had a different parametrization $\omega \mapsto \sqrt{3} \cdot \omega$). The extension to $\mathbb{R}$ is the content of the following proposition, which we recall for later use:

Proposition 2.4 [Bayer et al. 2011b, Corollary 3.3.3]. *Let* $\mathrm{St} \subset \mathrm{NS}_{\mathbb{R}}(X) \times \mathrm{NS}_{\mathbb{R}}(X)$ *be the subset of pairs of real classes (ω, B) for which ω is ample. There exists a notion of "tilt-stability" for every $(\omega, B) \in$* St. *For every object E, the set of (ω, B) for which E is $\nu_{\omega,B}$-stable defines an open subset of* St.

Definition 2.5. We define the *generalized discriminant*

$$\overline{\Delta}_H := (H^2 \operatorname{ch}_1^B)^2 - 2H^3 \operatorname{ch}_0^B \cdot (H \operatorname{ch}_2^B).$$

The generalized discriminant is independent of β. Indeed, by expanding the definition, we have

$$\begin{aligned} \overline{\Delta}_H &= (H^2(\operatorname{ch}_1 - \beta \operatorname{ch}_0 H))^2 - 2H^3 \operatorname{ch}_0 \cdot H\big(\operatorname{ch}_2 - \beta H \operatorname{ch}_1 + \tfrac{1}{2}\beta^2 \operatorname{ch}_0 H^2\big) \\ &= (H^2 \operatorname{ch}_1)^2 - 2(H^2 \operatorname{ch}_1) H^3 \beta \operatorname{ch}_0 + \beta^2 (\operatorname{ch}_0)^2 (H^3)^2 - 2H^3 \operatorname{ch}_0 (H \operatorname{ch}_2) \\ &\qquad + 2(H^2 \operatorname{ch}_1) H^3 \beta \operatorname{ch}_0 - \beta^2 (\operatorname{ch}_0)^2 (H^3)^2 \\ &= (H^2 \operatorname{ch}_1)^2 - 2H^3 \operatorname{ch}_0 (H \operatorname{ch}_2). \end{aligned}$$

The following result will be the key ingredient in our proof:

Theorem 2.6 [Bayer et al. 2011b, Corollary 7.3.2]. *For any $\nu_{\omega,B}$-semistable object $E \in \mathrm{Coh}^B(X)$, we have*

$$\overline{\Delta}_H(E) \geq 0.$$

2B. ***Reduction to small ω.*** In this section, we prove our reduction result. We keep the same notation as before, e.g., $\omega = \alpha H$ and $B = \beta H$. To simplify, we will write $\nu_{\alpha,\beta}$ for $\nu_{\omega,B}$, $\mathrm{Coh}^{\beta}(X)$, and so on.

Proposition 2.7. *Assume there exists $\bar{\alpha} \in \mathbb{R}_{>0}$ such that, for all $\alpha < \bar{\alpha}$ and for all $\beta \in \mathbb{R}$, Conjecture 2.3 holds. Then Conjecture 2.3 holds for all $\alpha \in \mathbb{R}_{>0}$ and for all $\beta \in \mathbb{R}$.*

To prove Proposition 2.7, we need first to introduce a bit more of notation. We denote by $\mathbb{H}$ the upper half-plane

$$\mathbb{H} := \{(\beta, \alpha) \in \mathbb{R}^2 : \alpha > 0\}.$$

For a vector

$$v := (\mathrm{ch}_0, \mathrm{ch}_1, \mathrm{ch}_2, \mathrm{ch}_3) \in \mathrm{Num}_{\mathbb{Q}}(X)$$

such that $H^2 \mathrm{ch}_1^{\beta} > 0$, the equation $\nu_{\alpha,\beta}(v) = 0$ defines a curve $\mathscr{C}_v$ in $\mathbb{H}$. Explicitly, we have

$$\mathscr{C}_v : H\,\mathrm{ch}_2 - \beta(H^2\,\mathrm{ch}_1) + \tfrac{1}{2}\beta^2 H^3\,\mathrm{ch}_0 - \tfrac{1}{2}\alpha^2 H^3\,\mathrm{ch}_0 = 0$$

together with the inequality

$$\beta H^3\,\mathrm{ch}_0 < H^2\,\mathrm{ch}_1\,.$$

We can divide into two cases:

$$\mathrm{ch}_0 = 0 \;\rightsquigarrow\; \beta = \frac{H\,\mathrm{ch}_2}{H^2\,\mathrm{ch}_1}, \tag{2}$$

$$\mathrm{ch}_0 \neq 0 \;\rightsquigarrow\; \left(\beta - \frac{H^2\,\mathrm{ch}_1}{H^3\,\mathrm{ch}_0}\right)^2 - \alpha^2 = \frac{\overline{\Delta}_H}{(H^3\,\mathrm{ch}_0)^2}. \tag{3}$$

Hence, if $\overline{\Delta}_H \geq 0$, then the tangent line at a point $(\beta_0, \alpha_0) \in \mathscr{C}$ intersects the line $\alpha = 0$ with an angle $\pi/4 \leq \theta \leq \pi/2$.

Finally, on the curve $\mathscr{C}_v$, we can write the inequality (1) as follows:

$$\begin{aligned} \mathrm{ch}_0 = 0 \;&\rightsquigarrow\; \mathrm{ch}_3 - \frac{(H\,\mathrm{ch}_2)^2}{2(H^2\,\mathrm{ch}_1)} \leq \alpha^2 \frac{H^2\,\mathrm{ch}_1}{6}, \\ \mathrm{ch}_0 \neq 0 \;&\rightsquigarrow\; \beta \frac{\overline{\Delta}_H}{H^3\,\mathrm{ch}_0} \leq \frac{(H\,\mathrm{ch}_2)(H^2\,\mathrm{ch}_1)}{H^3\,\mathrm{ch}_0} - 3\,\mathrm{ch}_3\,. \end{aligned} \tag{4}$$

Indeed, both inequalities in (4) follow directly by rewriting (1) by using (2) and (3), respectively.

Proof of Proposition 2.7. We argue by contradiction. Assume that there exist $\alpha_0 \geq \bar{\alpha}$, $\beta_0 \in \mathbb{R}$, and an object $E_0 \in \mathrm{Coh}^{\beta_0}(X)$ that is ν_{α_0,β_0}-stable, such that $\nu_{\alpha_0,\beta_0}(E_0) = 0$, and that does not satisfy the inequality in Conjecture 2.3.

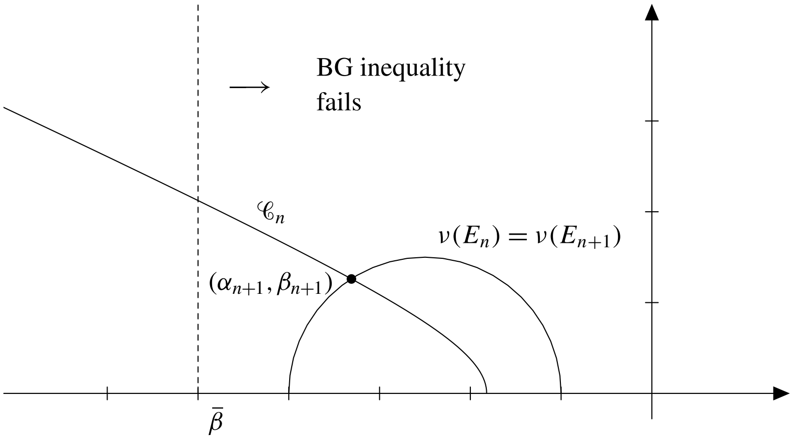


Figure 1. The curve $\mathcal{C}_n$ in the case $\mathrm{ch}_0(E_n) > 0$. The BG inequality is not satisfied when $\beta > \bar{\beta}$, where $\bar{\beta}$ is defined in (5).

Claim 1. *There exist a sequence $(\beta_n, \alpha_n) \in \mathbb{H}$ and a sequence of objects $\{E_n\}_{n\geq 0}$ such that*

- $E_n \in \mathrm{Coh}^{\beta_n}(X) \cap \mathrm{Coh}^{\beta_{n+1}}(X)$ *is ν_{α_n,β_n}-stable,*
- $\nu_{\alpha_n,\beta_n}(E_n) = \nu_{\alpha_{n+1},\beta_{n+1}}(E_n) = 0,$
- $0 < H^2 \mathrm{ch}_1^{\beta_{n+1}H}(E_{n+1}) < H^2 \mathrm{ch}_1^{\beta_{n+1}H}(E_n),$
- E_n *does not satisfy the inequality* (1),
- $\alpha_0 > \alpha_1 > \cdots > \alpha_n > \cdots > 0,$ *and*
- $|\beta_{n+1}| \leq |\beta_0| + \alpha_0.$

Proof. We proceed by induction, the case $n = 0$ being our assumption. Assume that we have constructed E_n with the wanted properties. By Proposition 2.4, the locus in $\mathbb{H}$ where E_n is $\nu_{\alpha,\beta}$-stable is open. Consider the curve $\mathcal{C} := \mathcal{C}_{\mathrm{ch}(E_n)} \subset \mathbb{H}$, and consider the set $U := \{(\beta, \alpha) \in \mathcal{C} : \alpha < \alpha_n\}$. We claim that, for all $(\beta, \alpha) \in U$, the inequality (1) is not satisfied for E_n. Indeed, this can be seen by dividing into three cases, according to whether $\mathrm{ch}_0(E_n)$ is > 0, $= 0$, or < 0 (the case in which $\mathrm{ch}_0(E_n) > 0$ is illustrated in Figure 1). If $\mathrm{ch}_0(E_n) > 0$, then by (4), we must have $\beta > \bar{\beta}$, where

$$\bar{\beta} := \frac{(H\,\mathrm{ch}_2)(H^2\,\mathrm{ch}_1) - 3H^3\,\mathrm{ch}_0\,\mathrm{ch}_3}{\overline{\Delta}_H}. \tag{5}$$

But by assumption, $\beta < H^2\,\mathrm{ch}_1(E_n)/H^3\,\mathrm{ch}_0(E_n)$. Hence, the hyperbola $\mathcal{C}$ is decreasing, which is what we claimed. The case $\mathrm{ch}_0(E_n) < 0$ is analogous, and the case $\mathrm{ch}_0(E_n) = 0$ follows directly again from (4) since, in this case, $H^2\,\mathrm{ch}_1(E_n) > 0$.

Since Conjecture 2.3 holds when $\alpha < \bar{\alpha}$, there must exist $(\beta_{n+1}, \alpha_{n+1}) \in U$ such that E_n is $\nu_{\alpha_{n+1},\beta_{n+1}}$-semistable and is not $\nu_{\alpha,\beta}$-semistable for all $(\beta, \alpha) \in U$ with $\alpha < \alpha_{n+1}$. When $\operatorname{ch}_0(E_n) \neq 0$, the hyperbola $\mathscr{C}$ has asymptotes meeting at the point $(H^2 \operatorname{ch}_1(E_n)/H^3 \operatorname{ch}_0(E_n), 0)$. Hence, for all $(\beta, \alpha) \in U$, we must have $\beta H^3 \operatorname{ch}_0(E_n) < H^2 \operatorname{ch}_1(E_n)$. Therefore, E_n being $\nu_{\alpha_{n+1},\beta_{n+1}}$-semistable, it must belong to the category $\operatorname{Coh}^{\beta_{n+1}}(X)$.

By looking at the $\nu_{\alpha_{n+1},\beta_{n+1}}$-stable factors of E_n (by [Bayer et al. 2011b, Proposition 5.2.2], this makes sense in the category $\operatorname{Coh}^{\beta_{n+1}}(X)$), given the additivity of the Chern character, there must exist an object $E_{n+1} \in \operatorname{Coh}^{\beta_{n+1}}(X)$ that is $\nu_{\alpha_{n+1},\beta_{n+1}}$-stable, such that $\nu_{\alpha_{n+1},\beta_{n+1}}(E_{n+1}) = 0$, and that does not satisfy the inequality (1).

The final inequality, $|\beta_{n+1}| \leq |\beta_0| + \alpha_0$, follows simply by the fact, observed before, that the tangent line at any point in $\mathscr{C}$ intersects the line $\alpha = 0$ with an angle $\pi/4 \leq \theta \leq \pi/2$. See Figure 2. □

We let $\tilde{\alpha} \geq 0$ be the limit of the sequence $\{\alpha_n\}$. By assumption, we would get a contradiction if we prove that $\tilde{\alpha} = 0$. Hence, assume this is not the case, namely $\tilde{\alpha} > 0$. The idea is to find bounds for $\operatorname{ch}_0(E_n)$, $H^2 \operatorname{ch}_1(E_n)$, and $H \operatorname{ch}_2(E_n)$.

Claim 2. *For all $n > 0$, the following inequality holds*:

$$\overline{\Delta}_H(E_n) + (\alpha_n H^3 \operatorname{ch}_0(E_n))^2 < \overline{\Delta}_H(E_0) + (\alpha_0 H^3 \operatorname{ch}_0(E_0))^2.$$

Proof. Again, we proceed by induction. By Claim 1, and by definition of the generalized discriminant, we have

$$\begin{aligned}
&\overline{\Delta}_H(E_{n+1}) + (\alpha_{n+1} H^3 \operatorname{ch}_0(E_{n+1}))^2 \\
&\quad = (H^2 \operatorname{ch}_1^{\beta_{n+1}}(E_{n+1}))^2 - 2H^3 \operatorname{ch}_0(E_{n+1})(H \operatorname{ch}_2^{\beta_{n+1}}(E_{n+1})) + (\alpha_{n+1} H^3 \operatorname{ch}_0(E_{n+1}))^2 \\
&\quad = (H^2 \operatorname{ch}_1^{\beta_{n+1}}(E_{n+1}))^2 - 2H^3 \operatorname{ch}_0(E_{n+1})(\tfrac{1}{2}\alpha_{n+1}^2 H^3 \operatorname{ch}_0(E_{n+1})) + (\alpha_{n+1} H^3 \operatorname{ch}_0(E_{n+1}))^2 \\
&\quad = (H^2 \operatorname{ch}_1^{\beta_{n+1}}(E_{n+1}))^2 \\
&\quad < (H^2 \operatorname{ch}_1^{\beta_{n+1}}(E_n))^2 \\
&\quad = (H^2 \operatorname{ch}_1^{\beta_{n+1}}(E_n))^2 - 2H^3 \operatorname{ch}_0(E_n)(\tfrac{1}{2}\alpha_{n+1}^2 H^3 \operatorname{ch}_0(E_n)) + (\alpha_{n+1} H^3 \operatorname{ch}_0(E_n))^2 \\
&\quad = (H^2 \operatorname{ch}_1^{\beta_{n+1}}(E_n))^2 - 2H^3 \operatorname{ch}_0(E_n)(H \operatorname{ch}_2^{\beta_{n+1}}(E_n)) + (\alpha_{n+1} H^3 \operatorname{ch}_0(E_n))^2 \\
&\quad = \overline{\Delta}_H(E_n) + (\alpha_{n+1} H^3 \operatorname{ch}_0(E_n))^2 \\
&\quad \leq \overline{\Delta}_H(E_n) + (\alpha_n H^3 \operatorname{ch}_0(E_n))^2.
\end{aligned}$$

□

By Claim 2, we deduce, for all $n > 0$, the inequality

$$\overline{\Delta}_H(E_n) + (\tilde{\alpha} H^3 \operatorname{ch}_0(E_n))^2 < \overline{\Delta}_H(E_0) + (\alpha_0 H^3 \operatorname{ch}_0(E_0))^2.$$

Hence, we get immediately

$$\overline{\Delta}_H(E_n) < \overline{\Delta}_H(E_0) + (\alpha_0 H^3 \operatorname{ch}_0(E_0))^2 =: \Gamma_0, \tag{6}$$

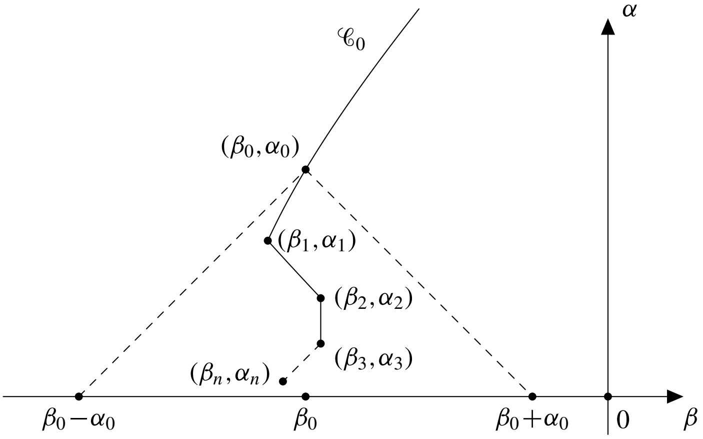


Figure 2. The sequence (β_n,α_n).

and by Theorem 2.6, we have

$$(\mathrm{ch}_0(E_n))^2 < \frac{1}{(\tilde{\alpha}H^3)^2}(\overline{\Delta}_H(E_0) + (\alpha_0 H^3\,\mathrm{ch}_0(E_0))^2) = \Gamma_1. \tag{7}$$

Finally, to bound $H^2\,\mathrm{ch}_1$, assume first that $\mathrm{ch}_0(E_n) \neq 0$. Then, by (3), (6), (7), and Claim 1, we have

$$|H^2\,\mathrm{ch}_1(E_n)| \leq H^3\sqrt{\Gamma_1}\left(|\beta_0| + \alpha_0 + \sqrt{\alpha_0^2 + \frac{\Gamma_0}{(H^3)^2}}\right) =: \Gamma_2. \tag{8}$$

The case in which $\mathrm{ch}_0(E_n) = 0$ follows by Claim 1 by observing that either $\mathrm{ch}_0(E_m) = 0$ for all $0 \leq m \leq n$ or there exists a maximum $0 \leq m < n$ for which $\mathrm{ch}_0(E_m) \neq 0$. In the first case, we have

$$0 < H^2\,\mathrm{ch}_1(E_n) < H^2\,\mathrm{ch}_1(E_0) \tag{9}$$

while in the second

$$0 < H^2\,\mathrm{ch}_1(E_n) < |H^2\,\mathrm{ch}_1(E_m)| + |\beta_m|\,|\mathrm{ch}_0(E_m)| \leq \Gamma_2 + (|\beta_0| + \alpha_0)\Gamma_1. \tag{10}$$

Summing up, by (6), (7), (8), (9), and (10), we found bounds for $\mathrm{ch}_0(E_n)$, $H^2\,\mathrm{ch}_1(E_n)$, and $H\,\mathrm{ch}_2(E_n)$ for all n. But this shows that these classes are finite, and so there must exist an object E that does not satisfy the inequality in Conjecture 2.3 for all α close to 0, which contradicts our assumption. □

3. The case of the projective space

In this section, we expand [Bayer et al. 2011b, Section 8.2] to show that, in the case of $X = \mathbb{P}^3$, the assumptions in Proposition 2.7 are satisfied. This will complete the proof of Theorem 1.1. To simplify notation, we directly identify $\mathrm{Num}_{\mathbb{R}}(\mathbb{P}^3)$ with $\mathbb{R}^{\oplus 4}$, and we take $\omega = \alpha$, $B = \beta \in \mathbb{R}$, and $\alpha > 0$. The tilted slope becomes, up to an irrelevant multiplicative constant,

$$\nu_{\alpha,\beta} = \frac{\mathrm{ch}_2^{\beta} - \frac{1}{2}\alpha^2 \,\mathrm{ch}_0}{\mathrm{ch}_1^{\beta}} = \frac{\mathrm{ch}_2 - \beta\, \mathrm{ch}_1 + \left(\frac{1}{2}\beta^2 - \frac{1}{2}\alpha^2\right) \mathrm{ch}_0}{\mathrm{ch}_1 - \beta\, \mathrm{ch}_0}.$$

Proposition 3.1. *For all $\alpha < \frac{1}{3}$ and for all $\beta \in \mathbb{R}$, Conjecture 2.3 holds.*

The proof is an adaptation of [Bayer et al. 2011b, Section 8.2], where only the case $\beta = 0$ was considered. The idea is to use the existence of Bridgeland's stability conditions on $\mathrm{D}^{\mathrm{b}}(\mathbb{P}^3)$ associated to strong exceptional collections of sheaves (see [Bridgeland 2007, Example 5.5; Macrì 2007, Section 3.3)]. Here, we will use the full strong exceptional collection $\mathfrak{E}$ on $\mathrm{D}^{\mathrm{b}}(\mathbb{P}^3)$ given by

$$\mathfrak{E} := \{\mathcal{O}_{\mathbb{P}^3}(-1), \mathcal{Q}, \mathcal{O}_{\mathbb{P}^3}, \mathcal{O}_{\mathbb{P}^3}(1)\},$$

where $\mathcal{Q} := T_{\mathbb{P}^3}(-2)$ is given by

$$0 \to \mathcal{O}_{\mathbb{P}^3}(-2) \to \mathcal{O}_{\mathbb{P}^3}(-1)^{\oplus 4} \to \mathcal{Q} \to 0.$$

We consider the region V given by

$$V := \left\{(\beta, \alpha) \in \mathbb{H} : 0 \geq \beta > -\tfrac{2}{3},\ 0 < \alpha < \tfrac{1}{3}\right\}.$$

Lemma 3.2. *Assume that Conjecture 2.3 holds for all $(\beta, \alpha) \in V$. Then it holds for all $\alpha < \frac{1}{3}$ and for all $\beta \in \mathbb{R}$.*

Proof. Assume, for a contradiction, there exist $\alpha_0 < \frac{1}{3}$ and $\beta_0 \in \mathbb{R}$ and $E \in \mathrm{D}^{\mathrm{b}}(\mathbb{P}^3)$ that does not satisfy Conjecture 2.3. By acting with the autoequivalence $\otimes\mathcal{O}_{\mathbb{P}^3}(1)$ and with the local dualizing functor $\mathbb{D}(\,\cdot\,) := \boldsymbol{R}\mathcal{H}\mathrm{om}(\,\cdot\,, \mathcal{O}_X[1])$, we can assume (see [Bayer et al. 2011b, Proposition 5.1.3]) that $0 > \beta_0 \geq -\frac{1}{2}$, which contradicts our assumption. □

The next result will allow us to use the exceptional collection $\mathfrak{E}$ for doing computations. We postpone the proof to the end of the section.

Lemma 3.3. *For all $(\beta, \alpha) \in V$, we have $\mathcal{Q}[1] \in \mathrm{Coh}^{\beta}(\mathbb{P}^3)$ and $\nu_{\alpha,\beta}^{\min}(\mathcal{Q}[1]) > 0$.*

We divide the region V into three parts:

$$V_1 := \{(\beta, \alpha) \in V : \beta < -\alpha\},$$
$$V_2 := \{(\beta, \alpha) \in V : \beta > -\alpha\},$$
$$V_3 := \{(\beta, \alpha) \in V : \beta = -\alpha\}.$$

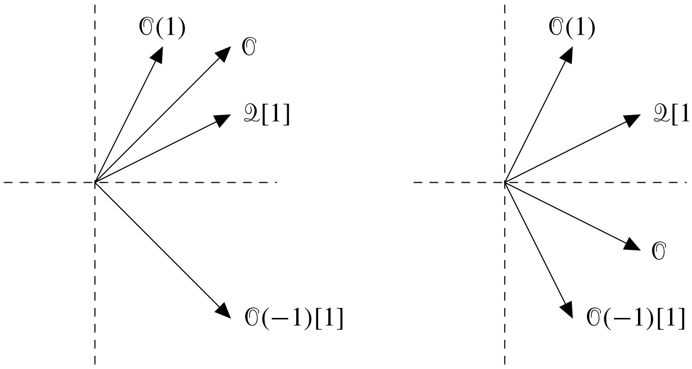


Figure 3. The slopes in $\mathrm{Coh}^\beta(\mathbb{P}^3)$ of the exceptional objects when $(\beta,\alpha)\in V_1$ (left) and $(\beta,\alpha)\in V_2$ (right). The tilt to $\mathcal{A}^{\alpha,\beta}$ corresponds to considering the upper half-plane. The two-dimensional picture is obtained by plotting denominator and numerator of $\nu_{\alpha,\beta}$. It is therefore oriented counterclockwise.

We first examine V_1 and V_2. On V_1, we have

$$\begin{aligned}
\nu_{\alpha,\beta}(\mathcal{O}) &= \frac{1}{2}\cdot\frac{\beta^2-\alpha^2}{-\beta} > 0,\\
\nu_{\alpha,\beta}(\mathcal{O}(-1)) &= \frac{1}{2}\cdot\frac{(\beta+1)^2-\alpha^2}{-\beta-1} < 0,\\
\nu_{\alpha,\beta}(\mathcal{O}(1)) &= \frac{1}{2}\cdot\frac{(\beta-1)^2-\alpha^2}{1-\beta} > 0,\\
\nu_{\alpha,\beta}(\mathcal{Q}) &= \frac{3}{2}\cdot\frac{(\beta+\frac{2}{3})^2-\alpha^2-\frac{4}{9}}{-2-3\beta} > 0.
\end{aligned}$$

On V_2, we get the same expressions, but now $\nu_{\alpha,\beta}(\mathcal{O}) < 0$ (see Figure 3).

We now tilt one more time $\mathrm{Coh}^\beta(\mathbb{P}^3)$, as explained in [Bayer et al. 2011b, Definition 3.2.5]. As in Section 2A, we can define a torsion pair

$$\begin{aligned}
\mathcal{T}'_{\omega,B} &= \{E\in\mathrm{Coh}^\beta(\mathbb{P}^3) : \text{any quotient } E \twoheadrightarrow G \text{ satisfies } \nu_{\alpha,\beta}(G) > 0\},\\
\mathcal{F}'_{\omega,B} &= \{E\in\mathrm{Coh}^\beta(\mathbb{P}^3) : \text{any subsheaf } F \hookrightarrow E \text{ satisfies } \nu_{\alpha,\beta}(F) \le 0\}.
\end{aligned}$$

We let $\mathcal{A}^{\alpha,\beta}\subset\mathrm{D}^{\mathrm{b}}(\mathbb{P}^3)$ be the extension-closure

$$\mathcal{A}^{\alpha,\beta} := \langle\mathcal{T}'_{\alpha,\beta}, \mathcal{F}'_{\alpha,\beta}[1]\rangle.$$

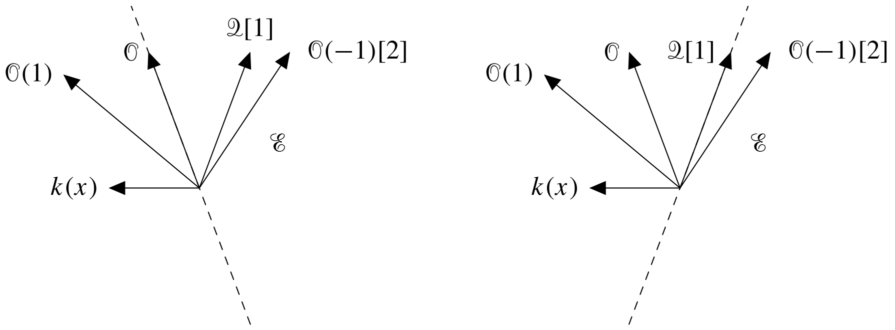


Figure 4. The slopes in $\mathcal{A}^{\alpha,\beta}$ of the exceptional objects and the skyscraper sheaves when $(\beta,\alpha)\in V_1$ (left) and $(\beta,\alpha)\in V_2$ (right). The category $\mathcal{E}$, obtained by tilting to the right along the dotted line, is the extension-closed subcategory generated by $\mathcal{O}(-1)[2]$, $\mathcal{Q}[1]$, $\mathcal{O}$, and $\mathcal{O}(1)[-1]$. It is equivalent to the category of modules over the finite-dimensional algebra determined by the dual exceptional collection to $\mathfrak{E}$.

By the previous computation, by [Bayer et al. 2011b, Proposition 7.4.1], and by Lemma 3.3, we have

$$\begin{aligned} &\{\mathcal{O}_{\mathbb{P}^3}(-1)[2], \mathcal{Q}[1], \mathcal{O}_{\mathbb{P}^3}, \mathcal{O}_{\mathbb{P}^3}(1)\}\subset\mathcal{A}^{\alpha,\beta} \qquad &&\text{for } (\beta,\alpha)\in V_1,\\ &\{\mathcal{O}_{\mathbb{P}^3}(-1)[2], \mathcal{Q}[1], \mathcal{O}_{\mathbb{P}^3}[1], \mathcal{O}_{\mathbb{P}^3}(1)\}\subset\mathcal{A}^{\alpha,\beta} \qquad &&\text{for } (\beta,\alpha)\in V_2. \end{aligned}$$

On the category $\mathcal{A}^{\alpha,\beta}$, we consider the following function (a posteriori, this will be a slope function):

$$\lambda_{\alpha,\beta} := \begin{cases} +\infty & \text{if } \mathrm{ch}_2^\beta - \frac{1}{2}\alpha^2\,\mathrm{ch}_0^\beta = 0,\\ \dfrac{\mathrm{ch}_3^\beta - \frac{1}{6}\alpha^2\,\mathrm{ch}_1^\beta}{\mathrm{ch}_2^\beta - \frac{1}{2}\alpha^2\,\mathrm{ch}_0^\beta} & \text{otherwise.} \end{cases}$$

We have

$$\begin{aligned} \lambda_{\alpha,\beta}(\mathcal{O}) &= -\tfrac{1}{3}\beta,\\ \lambda_{\alpha,\beta}(\mathcal{O}(-1)) &= -\tfrac{1}{3}\beta - \tfrac{1}{3},\\ \lambda_{\alpha,\beta}(\mathcal{O}(1)) &= -\tfrac{1}{3}\beta + \tfrac{1}{3},\\ \lambda_{\alpha,\beta}(\mathcal{Q}) &= \frac{\left(\frac{2}{3}-\beta^2-\frac{1}{2}\beta^3\right)+\frac{1}{6}\alpha^2(3\beta+2)}{2\beta+\frac{3}{2}\beta^2-\frac{3}{2}\alpha^2}. \end{aligned}$$

On V_1, we deduce that $\lambda_{\alpha,\beta}(Q) < \lambda_{\alpha,\beta}(\mathcal{O}(1))$ while, on V_2, $\lambda_{\alpha,\beta}(Q) < \lambda_{\alpha,\beta}(\mathcal{O})$ (see Figure 4).

By [Bayer et al. 2011b, Proposition 8.1.1] (and mimicking the proof of [Bayer et al. 2011b, Theorem 8.2.1]), this shows that Conjecture 2.3 holds for all $(\beta, \alpha) \in V_1 \cup V_2$.

To deal with the region V_3 (namely, the case $\alpha = -\beta$), we consider a slightly modified function on $\mathscr{A}^{\alpha,\beta}$

$$\lambda_{\alpha,\beta} := \begin{cases} +\infty & \text{if } \operatorname{ch}_2^\beta - \frac{1}{2}\alpha^2 \operatorname{ch}_0^\beta = 0, \\ \dfrac{\operatorname{ch}_3^\beta - \frac{1}{6}\alpha^2 \operatorname{ch}_1^\beta - \epsilon \operatorname{ch}_1^\beta}{\operatorname{ch}_2^\beta - \frac{1}{2}\alpha^2 \operatorname{ch}_0^\beta} & \text{otherwise,} \end{cases}$$

where $\epsilon > 0$. In this case, we still have

$$\{\mathscr{O}_{\mathbb{P}^3}(-1)[2], \mathscr{Q}[1], \mathscr{O}_{\mathbb{P}^3}[1], \mathscr{O}_{\mathbb{P}^3}(1)\} \subset \mathscr{A}^{\alpha,\beta},$$

and

$$\begin{aligned} \lambda_{\alpha,\beta}(\mathscr{O}) &= +\infty, \\ \lambda_{\alpha,\beta}(\mathscr{O}(-1)) &= -\tfrac{1}{3}\beta - \tfrac{1}{3} + 2\epsilon \frac{\beta+1}{2\beta+1}, \\ \lambda_{\alpha,\beta}(\mathscr{O}(1)) &= -\tfrac{1}{3}\beta + \tfrac{1}{3} + 2\epsilon \frac{\beta-1}{1-2\beta}, \\ \lambda_{\alpha,\beta}(\mathscr{Q}) &= \frac{1-\beta^2}{3\beta} + \epsilon \frac{3\beta+2}{2\beta}. \end{aligned}$$

We deduce that, for all $0 > \beta > -\frac{1}{3}$, there exists $\epsilon(\beta) > 0$ such that

$$\lambda_{\alpha,\beta}(\mathscr{O}(1)) > \lambda_{\alpha,\beta}(\mathscr{O}(-1)) \quad \text{and} \quad \lambda_{\alpha,\beta}(\mathscr{O}(1)) > \lambda_{\alpha,\beta}(\mathscr{Q})$$

when $(\beta, \alpha) \in V_3$ and $\epsilon < \epsilon(\beta)$. Again, by [Bayer et al. 2011b, Proposition 8.1.1], if we fix β and let $\epsilon \to 0$, this shows that Conjecture 2.3 holds also for all $(\beta, \alpha) \in V_3$. By Lemma 3.2, this would complete the proof of Proposition 3.1 once Lemma 3.3 is proved.

Proof of Lemma 3.3. Since $\mathscr{Q} \in \operatorname{Coh}(\mathbb{P}^3)$ is slope-stable with Chern character $\operatorname{ch}(\mathscr{Q}) = (3, -2, 0, \frac{2}{3})$, we have, by definition, $\mathscr{Q}[1] \in \operatorname{Coh}^\beta(\mathbb{P}^3)$ for all $\beta \geq -\frac{2}{3}$. Moreover, for $0 \geq \beta > -\frac{2}{3}$ and for all $\alpha > 0$, we have $\nu_{\alpha,\beta}(\mathscr{Q}[1]) > 0$.

Assume, for a contradiction, there exists $(\beta_0, \alpha_0) \in V$ such that $\nu^{\min}_{\alpha_0,\beta_0}(Q[1]) \leq 0$. Let $N_0 \in \operatorname{Coh}^{\beta_0}(\mathbb{P}^3)$ be the tilt-stable quotient $\mathscr{Q}[1] \twoheadrightarrow N_0$ in $\operatorname{Coh}^{\beta_0}(\mathbb{P}^3)$ such that $\nu_{\alpha_0,\beta_0}(N_0) \leq 0$. By taking the long exact sequence in cohomology, $N_0 \cong M_0[1]$, where $M_0 \in \operatorname{Coh}(\mathbb{P}^3)$ is a torsion-free sheaf.

Consider the curves $\mathscr{B}_0$, given by $\nu_{\alpha,\beta}(\mathscr{Q}[1]) = \nu_{\alpha,\beta}(N_0)$, and $\mathscr{C}_0$, given by $\nu_{\alpha,\beta}(N_0) = 0$ in the region $\beta > \operatorname{ch}_1(M_0)/\operatorname{ch}_0(M_0)$. Since the vector $(3, -2, 0)$ is primitive, $\mathscr{B}_0$ must be a semicircle in $\mathbb{H}$. Consider the unique point of intersection $(x, y) \in \mathscr{C}_0 \cap \mathscr{B}_0$. Since $\nu_{\alpha,\beta}(\mathscr{Q}[1]) > 0$, for $0 \geq \beta > -\frac{2}{3}$, we have $x > 0$. In particular, $\mathscr{B}_0 \cap \{\beta = 0\} \neq \varnothing$. See Figure 5.

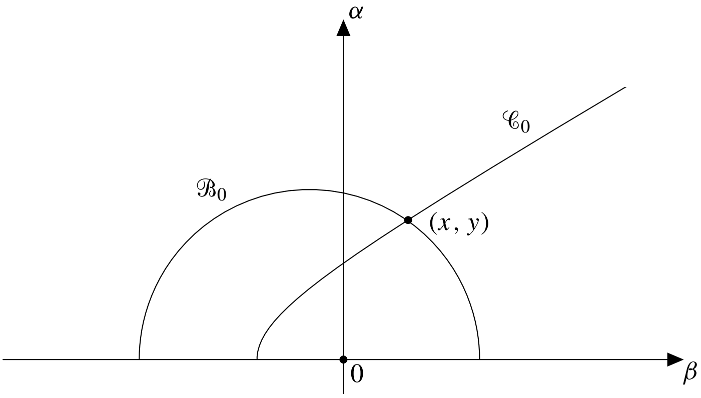


Figure 5. The curves $\mathcal{B}_0$ and $\mathcal{C}_0$.

By Bertram's nested wall theorem of [Maciocia 2012] (whose proof works as well in our context due to Theorem 2.6), we know that *pseudo-walls* for $\mathcal{Q}[1]$ are *nested* semicircles; namely, either $\mathcal{Q}[1]$ is tilt-stable outside $\mathcal{B}_0$ and unstable in the interior, or there exists another semicircle $\mathcal{B}_1$ with the same property and $\mathcal{B}_1$ contains $\mathcal{B}_0$ in its interior. In both cases, by the previous argument, the semicircles $\mathcal{B}_0$ and $\mathcal{B}_1$ intersect the half-line $\beta = 0$. Hence, there exists $\alpha_1 > 0$ such that $\mathcal{Q}[1]$ is not $\nu_{\alpha_1,0}$-stable. This contradicts Lemma 3.4 below. □

Lemma 3.4. *For all $\alpha > 0$, $\mathcal{Q}[1]$ is $\nu_{\alpha,0}$-stable.*

Proof. First of all, we observe that $\mathcal{Q}[1]$ is PGL(4)-invariant. By uniqueness of Harder–Narasimhan filtrations, if $\mathcal{Q}[1]$ is not tilt-stable, then its HN factors have to be PGL(4)-invariant as well.

Consider the category $\mathrm{Coh}^{\beta=0}(\mathbb{P}^3)$. The function $f_0 := \mathrm{ch}_1$ is additive and takes nonnegative integral values on $\mathrm{Coh}^0(\mathbb{P}^3)$. Since $f_0(\mathcal{Q}[1]) = 2$, if there exists an exact sequence in $\mathrm{Coh}^0(\mathbb{P}^3)$

$$0 \to P \to \mathcal{Q}[1] \to N \cong M[1] \to 0 \tag{11}$$

that is destabilizing with N tilt-semistable, then $f_0(P) = f_0(N) = 1$ and both P and N must be tilt-stable. To prove this claim, we first observe that $\mathcal{Q}[1]$ cannot have any subobject P with $f_0(P) = 0$. Indeed, in such a case, by definition, P belongs to the category generated by extensions by $F[1]$, where F is a μ-stable torsion-free sheaves with $\mu(F) = 0$, and by torsion sheaves supported in dimension ≤ 1. Therefore, $\mathrm{Hom}(P, \mathcal{Q}[1]) = 0$. Hence, a subobject P of $\mathcal{Q}[1]$ can have either $f_0(P) = 1$ or $f_0(P) = 2$. But if $f_0(P) = 2$, then the sequence is not destabilizing. The same argument shows that P and N are also tilt-stable.

The long exact sequence in cohomology gives

$$0 \to \mathcal{H}^{-1}(P) \to \mathcal{Q} \to M \to \mathcal{H}^0(P) \to 0$$

with $\mathcal{H}^{-1}(P)$ and M torsion-free with $\mu_{\alpha,0}^{\max} \le 0$. Since (11) is destabilizing and both P and N are tilt-stable with $f_0 = 1$, we must have $\mu_{\alpha,0}^{\max}(M)$, $\mu_{\alpha,0}^{\max}(\mathcal{H}^{-1}(P)) < 0$. This shows that there are only two possibilities:

(a) either $\mathrm{ch}_1(M) = \mathrm{ch}_1(\mathcal{H}^{-1}(P)) = -1$,

(b) or $\mathcal{H}^{-1}(P) = 0$.

For case (a), we must have $\mathrm{ch}_1(\mathcal{H}^0(P)) = 0$, and so $\mathcal{H}^0(P)$ is a torsion sheaf supported on a one-dimensional subscheme. By the PGL(4)-invariance, $\mathcal{H}^0(P) = 0$. Finally, since $\mathcal{Q}$ is slope-stable, we must have $\mathrm{ch}_0(\mathcal{H}^{-1}(P)) = 1$, and so $\mathcal{H}^{-1}(P) \cong \mathcal{I}_C(-1)$ for $C \subset \mathbb{P}^3$ a one-dimensional subscheme of degree $d \ge 0$. Again, by the PGL(4)-invariance, $C = 0$. Summarizing, we proved that in case (a), $P \cong \mathcal{O}_{\mathbb{P}^3}(-1)[1]$. But then, the equation $\nu_{\alpha,0}(\mathcal{Q}[1]) = \nu_{\alpha,0}(P)$ has no solutions, and so (11) cannot be destabilizing.

For case (b), we have $P \in \mathrm{Coh}(\mathbb{P}^3)$ and an exact sequence in $\mathrm{Coh}(\mathbb{P}^3)$

$$0 \to \mathcal{Q} \to M \to P \to 0$$

with $\mathrm{ch}_1(M) = -1$, $\mathrm{ch}_1(P) = 1$, and $\mathrm{ch}_0(M) \ge 3$. We now use Theorem 2.6 once more. Indeed, since N must be tilt-stable, we have

$$\mathrm{ch}_2(M) \le \frac{1}{2\,\mathrm{ch}_0(M)},$$

and so $\mathrm{ch}_2(M) \le 0$. As a consequence, the equation $\nu_{\alpha,0}(\mathcal{Q}[1]) = \nu_{\alpha,0}(P)$ has no solutions $\alpha > 0$, and so (11) cannot be destabilizing also in this case. □

4. An application

In this section, we briefly discuss an application of Theorem 1.1 and some examples.

In [Bayer et al. 2011b, Example 7.2.4], we pointed out a relation between Conjecture 2.3 and Castelnuovo's inequality for curves in $\mathbb{P}^3$. In particular, by using Castelnuovo's inequality, we showed that Conjecture 2.3 holds for ideal sheaves of curves with respect to some tilt-stability. It is interesting to observe that a sort of converse holds: from Theorem 1.1, we can deduce a certain inequality for curves in $\mathbb{P}^3$, which is much weaker than Castelnuovo's one but already nontrivial.

Corollary 4.1. *Let C be a pure one-dimensional scheme in $\mathbb{P}^3$ of degree $d \ge 2$. Let $h := \mathrm{ch}_3(\mathcal{I}_C) - 2d$. Then*

$$h \le \frac{2d^2 - 5d}{3}. \tag{12}$$

Moreover, if C is integral and not contained in a plane, then

$$h \leq \frac{d^2-4d}{3}. \tag{13}$$

We recall that, for an ideal sheaf $\mathcal{I}_C$ of an integral curve $C \subset \mathbb{P}^3$ of degree d and arithmetic genus g, $h = g-1$. Hence, the inequality (13) compares with [Hartshorne 1977, IV, 6.4].

To prove Corollary 4.1, we introduce some more notation. We define the two semicircles

$$\mathcal{B}_1 : \alpha^2 + \left(\beta + \frac{2d+1}{2}\right)^2 = \left(\frac{2d-1}{2}\right)^2,$$
$$\mathcal{B}_2 : \alpha^2 + \left(\beta + \frac{d+2}{2}\right)^2 = \left(\frac{d-2}{2}\right)^2.$$

They correspond to the loci

$$\nu_{\alpha,\beta}(\mathcal{I}_C) = \nu_{\alpha,\beta}(\mathcal{O}_{\mathbb{P}^3}(-1)) \quad \text{and} \quad \nu_{\alpha,\beta}(\mathcal{I}_C) = \nu_{\alpha,\beta}(\mathcal{O}_{\mathbb{P}^3}(-2)),$$

respectively. More generally, for an object $A \in \mathrm{D}^{\mathrm{b}}(\mathbb{P}^3)$ such that $(\mathrm{ch}_0(A), \mathrm{ch}_1(A), \mathrm{ch}_2(A))$ is not a multiple of $(1, 0, -d)$, we denote by $\mathcal{B}_A$ the semicircle with equation $\nu_{\alpha,\beta}(\mathcal{I}_C) = \nu_{\alpha,\beta}(A)$.

Finally, as in Section 2B, we denote by $\mathcal{C}$ the branch of the hyperbola $\nu_{\alpha,\beta}(\mathcal{I}_C) = 0$ in $\mathbb{H}$; explicitly,

$$\mathcal{C} : \beta^2 - \alpha^2 = 2d, \quad \beta < 0.$$

Proof of Corollary 4.1. For the first part of the statement, we would like to show that on the exterior part of the semicircle $\mathcal{B}_1$ in $\mathbb{H} \cap \{-2d < \beta < -1\}$ the ideal sheaf $\mathcal{I}_C$ is $\nu_{\alpha,\beta}$-stable.

First of all, we consider the half-line $\beta = -1$ and the category $\mathrm{Coh}^{\beta=-1}(\mathbb{P}^3)$. The function $f_{-1} := \mathrm{ch}_1 + \mathrm{ch}_0$ is additive and takes nonnegative integral values on $\mathrm{Coh}^{-1}(\mathbb{P}^3)$. Since $f_{-1}(\mathcal{I}_C) = 1$, then $\mathcal{I}_C$ must be $\nu_{\alpha,-1}$-stable for all $\alpha > 0$.

We now consider the half-line $\beta = -2$ and the category $\mathrm{Coh}^{-2}(\mathbb{P}^3)$. By [Bridgeland 2008, Proposition 14.2] (whose proof generalizes to our case), we know that, for $\alpha \gg 0$, $\mathcal{I}_C$ is $\nu_{\alpha,-2}$-stable. Assume that $\mathcal{I}_C$ is not $\nu_{\alpha,-2}$-semistable for all $\alpha > 0$. Then, by Proposition 2.4, there exists $\alpha_0 > 0$ such that $\mathcal{I}_C$ is $\nu_{\alpha,-2}$-stable for $\alpha > \alpha_0$, is $\nu_{\alpha,-2}$-semistable at $\alpha = \alpha_0$, and is not semistable for $\alpha < \alpha_0$. Then α_0 must be in the intersection of the half-line $\beta = -2$ with a semicircle $\mathcal{B}_A$ for some $A \in \mathrm{Coh}^{-2}(\mathbb{P}^3)$ such that $A \hookrightarrow \mathcal{I}_C$ in $\mathrm{Coh}^{-2}(\mathbb{P}^3)$. By looking at the long exact sequence in cohomology, we deduce that $A \in \mathrm{Coh}(\mathbb{P}^3)$ and $\mathrm{ch}_0(A) \geq 1$ and it is torsion-free. Moreover, since the function $f_{-2} := \mathrm{ch}_1 + 2\,\mathrm{ch}_0$ is additive and takes nonnegative integral values on $\mathrm{Coh}^{-2}(\mathbb{P}^3)$ and $f_{-2}(\mathcal{I}_C) = 2$, we must have

$f_{-2}(A) = 1$, namely
$$\frac{\mathrm{ch}_1(A)}{\mathrm{ch}_0(A)} = -2 + \frac{1}{\mathrm{ch}_0(A)}.$$

Let $(-2, \alpha_1)$ be the intersection point in $\mathbb{H}$ between $\beta = -2$ and $\mathcal{B}_1$ (the intersection is nonempty since $d \geq 2$). We claim that $\alpha_0 \leq \alpha_1$. Indeed, if $\mathrm{ch}_0(A) = 1$, then $\mathrm{ch}_1(A) = -1$. Hence, $A \cong \mathcal{I}_W(-1)$ for some subscheme W of dimension 1. Therefore, $\alpha_0 \leq \alpha_1$. If $\mathrm{ch}_0(A) \geq 2$, then $-2 < \mathrm{ch}_1(A)/\mathrm{ch}_0(A) < -1$. By Bertram's nested wall theorem of [Maciocia 2012], we know that either $\mathcal{B}_A = \mathcal{B}_1$ or they are disjoint. Since $\mathcal{B}_A \cap \{\beta = \mathrm{ch}_1(A)/\mathrm{ch}_0(A)\} = \varnothing$, this immediately implies that $\alpha_0 \leq \alpha_1$, as we wanted.

By using the nested wall theorem again, since we proved that, on the line $\beta = -1$, the ideal sheaf $\mathcal{I}_C$ is stable and, on the line $\beta = -2$, the first wall is $\mathcal{B}_1$, this shows that on the exterior part of the semicircle $\mathcal{B}_1$ in $\mathbb{H} \cap \{-2d < \beta < -1\}$ the ideal sheaf $\mathcal{I}_C$ is $\nu_{\alpha,\beta}$-stable, which is what we wanted. To get the inequality (12), we only need to compute the intersection point $\mathcal{C} \cap \mathcal{B}_1$. Theorem 1.1 yields then directly (12).

The proof of (13) is very similar. We consider the half-line $\beta = -3$, the category $\mathcal{A}_{-3} := \mathrm{Coh}^{-3}(\mathbb{P}^3)$, and $A \hookrightarrow \mathcal{I}_C$ in $\mathrm{Coh}^{-3}(\mathbb{P}^3)$. By looking at the function $f_{-3} := \mathrm{ch}_1 + 3\,\mathrm{ch}_0$, we must have either $f_{-3}(A) = 1$ or $= 2$. If $\mathrm{ch}_0(A) \geq 3$, then by using again [Maciocia 2012], we can deduce that $\mathcal{B}_A$ is contained in the interior of $\mathcal{B}_2$. If $\mathrm{ch}_0(A) = 2$, we distinguish two possibilities according to whether $f_{-3}(A) = 1$ or $= 2$. If $= 1$, then we can argue as before and deduce that $\mathcal{B}_A$ is contained in the interior of $\mathcal{B}_2$. If $= 2$, then $\mathrm{ch}_1(A) = -4$, and so by Theorem 2.6, $\mathrm{ch}_2(A) \leq 4$. If $\mathrm{ch}_2(A) = 4$, then $\mathcal{B}_A = \mathcal{B}_2$. If $\mathrm{ch}_2(A) < 4$, then $\mathcal{B}_A$ is again contained in the interior of $\mathcal{B}_2$.

Finally, if $\mathrm{ch}_0(A) = 1$, then either $A \cong \mathcal{I}_W(-2)$ or $A \cong \mathcal{I}_W(-1)$ with W a closed subscheme of dimension 1. The first case can be dealt as before. To exclude the second case, we use the assumption that C is integral and not contained in a plane. Indeed, in such a case, we must have $C \subset W$, and so $A \hookrightarrow \mathcal{I}_C$ does not destabilize.

As before, to get the inequality (13), we only need to compute the intersection point $\mathcal{C} \cap \mathcal{B}_2$ and apply Theorem 1.1. □

Example 4.2. For the case $d = 1$, the situation is slightly degenerate. Indeed, in such a case, $\mathcal{I}_C$ is $\nu_{\alpha,\beta}$-semistable for all $(\beta, \alpha) \in \mathbb{H}$ for which
$$\alpha^2 + (\beta + \tfrac{3}{2})^2 \geq \tfrac{1}{4}.$$
Hence, in particular, it is semistable for all $(\beta, \alpha) \in \mathcal{C}$. Theorem 1.1 gives then $h \leq -\frac{2}{3}$, namely $g\ (= 0) \leq \frac{1}{3}$.

Example 4.3. If the curve C in Corollary 4.1 is contained in a surface $F \subset \mathbb{P}^3$ of degree $k > 0$, then there is a strong form for Castelnuovo's theorem, as proved by Harris [1980; Hartshorne 1978]. But in this case, we cannot directly conclude such inequality by using stability since it is not true that the first wall when $\mathcal{I}_C$ is destabilized coincides with the locus

$$\nu_{\alpha,\beta}(\mathcal{I}_C)=\nu_{\alpha,\beta}(\mathcal{O}_{\mathbb{P}^3}(-k)), \quad \text{namely } \mathcal{O}_{\mathbb{P}^3}(-k)\hookrightarrow \mathcal{I}_C.$$

The simplest example (see [Hartshorne 1977, V, 4.13.1]) is when C is smooth with $k=3$, $d=7$, and $g=5$. In such a case, a destabilizing quotient is given instead by

$$\mathcal{I}_C \twoheadrightarrow \mathcal{O}_{\mathbb{P}^3}(-5)[1].$$

This gives the (well-known) existence of a nontrivial extension, $\mathcal{G}\in \mathrm{Coh}(\mathbb{P}^3)$ of rank 2, which must be stable. It may be interesting to study the general situation and see which kind of new stable objects arise as destabilizing factors of $\mathcal{I}_C$.

Acknowledgements

I would like to thank Arend Bayer and Ciaran Meachan for very useful discussions and Roman Bezrukavnikov for pointing out the idea of using Theorem 1.1 to show a version of Castelnuovo's theorem. I am also very grateful to the referee for the careful reading of the manuscript and for several suggestions that improved readability. This paper was completed during my stay at the Mathematical Institute of the University of Bonn, whose warm hospitality is gratefully acknowledged. I was partially supported by the NSF grant DMS-1160466 and, during the visit to Bonn, by the grant SFB/TR 45.

References

[Arcara and Bertram 2013] D. Arcara and A. Bertram, "Bridgeland-stable moduli spaces for K-trivial surfaces", *J. Eur. Math. Soc.* **15**:1 (2013), 1–38. MR 2998828 Zbl 1259.14014

[Arcara et al. 2013] D. Arcara, A. Bertram, I. Coskun, and J. Huizenga, "The minimal model program for the Hilbert scheme of points on $\mathbb{P}^2$ and Bridgeland stability", *Adv. Math.* **235** (2013), 580–626. MR 3010070 Zbl 1267.14023

[Bayer 2011] A. Bayer, "A tour to stability conditions on derived categories", notes, 2011, Available at http://www.maths.ed.ac.uk/~abayer/dc-lecture-notes.pdf.

[Bayer and Macrì 2012] A. Bayer and E. Macrì, "Projectivity and birational geometry of Bridgeland moduli spaces", preprint, 2012. To appear in *J. Am. Math. Soc.* arXiv 1203.4613v1

[Bayer et al. 2011a] A. Bayer, A. Bertram, E. Macrì, and Y. Toda, "Bridgeland stability conditions on threefolds, II: An application to Fujita's conjecture", preprint, 2011. To appear in *J. Alg. Geom.* arXiv 1106.3430v1

[Bayer et al. 2011b] A. Bayer, E. Macrì, and Y. Toda, "Bridgeland stability conditions on threefolds, I: Bogomolov–Gieseker type inequalities", preprint, 2011. To appear in *J. Alg. Geom.* arXiv 1103.5010v1

[Bridgeland 2007] T. Bridgeland, "Stability conditions on triangulated categories", *Ann. of Math.* (2) **166**:2 (2007), 317–345. MR 2009c:14026 Zbl 1137.18008

[Bridgeland 2008] T. Bridgeland, "Stability conditions on $K3$ surfaces", *Duke Math. J.* **141**:2 (2008), 241–291. MR 2009b:14030 Zbl 1138.14022

[Bridgeland 2009] T. Bridgeland, "Spaces of stability conditions", pp. 1–21 in *Algebraic geometry, Part 1* (Seattle, 2005), edited by D. Abramovich et al., Proc. Sympos. Pure Math. **80**, Amer. Math. Soc., Providence, RI, 2009. MR 2010f:14011 Zbl 1169.14303

[Happel et al. 1996] D. Happel, I. Reiten, and S. O. Smalø, "Tilting in abelian categories and quasitilted algebras", *Mem. Amer. Math. Soc.* **120**:575 (1996), 1–88. MR 97j:16009 Zbl 0849.16011

[Harris 1980] J. Harris, "The genus of space curves", *Math. Ann.* **249**:3 (1980), 191–204. MR 81i:14022 Zbl 0449.14006

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[Hartshorne 1978] R. Hartshorne, "Stable vector bundles of rank 2 on $\mathbb{P}^3$", *Math. Ann.* **238**:3 (1978), 229–280. MR 80c:14011 Zbl 0411.14002

[Huybrechts 2012] D. Huybrechts, "Introduction to stability conditions", preprint, 2012. arXiv 1111.1745v2

[Lo and Qin 2011] J. Lo and Z. Qin, "Mini-walls for Bridgeland stability conditions on the derived category of sheaves over surfaces", preprint, 2011. arXiv 1103.4352v1

[Maciocia 2012] A. Maciocia, "Computing the walls associated to Bridgeland stability conditions on projective surfaces", preprint, 2012. To appear in *Asian J. Math.* arXiv 1202.4587v1

[Maciocia and Meachan 2013] A. Maciocia and C. Meachan, "Rank-1 Bridgeland stable moduli spaces on a principally polarized abelian surface", *Int. Math. Res. Not.* **2013**:9 (2013), 2054–2077. MR 3053413

[Macrì 2007] E. Macrì, "Stability conditions on curves", *Math. Res. Lett.* **14**:4 (2007), 657–672. MR 2008k:18011 Zbl 1151.14015

[Minamide et al. 2011] H. Minamide, S. Yanagida, and K. Yoshioka, "Some moduli spaces of Bridgeland's stability conditions", preprint, 2011. arXiv 1111.6187v3

[Polishchuk 2012] A. Polishchuk, "Phases of Lagrangian-invariant objects in the derived category of an abelian variety", preprint, 2012. arXiv 1203.2300v1

[Toda 2012a] Y. Toda, "Introduction and open problems of Donaldson–Thomas theory", pp. 289–318 in *Derived categories in algebraic geometry* (Tokyo, 2011), edited by Y. Kawamata, Eur. Math. Soc., Zürich, 2012. MR 3050708 Zbl 1256.14001

[Toda 2012b] Y. Toda, "Stability conditions and birational geometry of projective surfaces", preprint, 2012. To appear in *Compos. Math.* arXiv 1205.3602v1

[Toda 2013a] Y. Toda, "Bogomolov–Gieseker-type inequality and counting invariants", *J. Topol.* **6**:1 (2013), 217–250. MR 3029426

[Toda 2013b] Y. Toda, "Stability conditions and extremal contractions", *Math. Ann.* **357**:2 (2013), 631–685. MR 3096520 Zbl 06228499

[Yanagida and Yoshioka 2012] S. Yanagida and K. Yoshioka, "Bridgeland's stabilities on abelian surfaces", preprint, 2012. arXiv 1203.0884v1

[Yoshioka 2012] K. Yoshioka, "Bridgeland's stability and the positive cone of the moduli spaces of stable objects on an abelian surface", preprint, 2012. arXiv 1206.4838v1

Communicated by Yujiro Kawamata
Received 2012-07-30 Revised 2013-06-10 Accepted 2013-06-11

macri.6@math.osu.edu *Department of Mathematics, The Ohio State University, 231 West 18th Avenue, Columbus, OH 43210-1174, United States*

ALGEBRA AND NUMBER THEORY 8:1 (2014)
dx.doi.org/10.2140/ant.2014.8.191

(φ, Γ)-modules over noncommutative overconvergent and Robba rings

Gergely Zábrádi

We construct noncommutative multidimensional versions of overconvergent power series rings and Robba rings. We show that the category of étale (φ, Γ)-modules over certain completions of these rings is equivalent to the category of étale (φ, Γ)-modules over classical overconvergent or Robba rings as the case may be (hence also to the category of p-adic Galois representations of $\mathbb{Q}_p$). In the case of Robba rings, the assumption of étaleness is not necessary, so there exists a notion of trianguline objects in this sense.

1. Introduction

In recent years it has become increasingly clear that some kind of p-adic version of the local Langlands correspondence should exist. In fact, Colmez [2010c; 2010b] constructed such a correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$. His construction is done in several steps using (φ, Γ)-modules (the category of which is well known [Fontaine 1990] to be equivalent to the category of p-adic Galois representations of $\mathbb{Q}_p$). We briefly recall Colmez's correspondence here. Let K be a finite extension of $\mathbb{Q}_p$ with ring of integers o_K and uniformizer p_K.

The "Montreal functor" associates to a smooth o_K-torsion representation of the standard Borel subgroup $B_2(\mathbb{Q}_p)$ of $\mathrm{GL}_2(\mathbb{Q}_p)$ an o_K-torsion (φ, Γ)-module over Fontaine's ring $\mathbb{O}_{\mathscr{E}}$. If we are given a unitary Banach space representation Π over the field K of the group $\mathrm{GL}_2(\mathbb{Q}_p)$, then it admits an o_K-lattice $L(\Pi)$ that is invariant under $\mathrm{GL}_2(\mathbb{Q}_p)$. Hence $L(\Pi)/p_K^r$ is a smooth o_K-torsion representation that we restrict now to $B_2(\mathbb{Q}_p)$. The (φ, Γ)-module associated to Π is the projective limit (as $r \to \infty$) of the (φ, Γ)-modules associated to $L(\Pi)/p_K^r$ via the Montreal functor. This is generalized in [Schneider and Vignéras 2011] to general reductive groups over $\mathbb{Q}_p$.

The reverse direction, how one adjoins a unitary continuous p-adic representation to a 2-dimensional (φ, Γ)-module D over Fontaine's ring, is even more subtle. One first constructs a unitary p-adic Banach space representation $\Pi(D)$ to each

MSC2010: primary 11F70; secondary 20G25, 16U20.
Keywords: overconvergent, Robba ring, noncommutative, (φ, Γ)-module.

2-dimensional *trianguline* (φ, Γ)-module D over $\mathscr{E} = \mathcal{O}_{\mathscr{E}}[p^{-1}]$ using a kind of parabolic induction. This Banach space is well described as a quotient of the space of p-adic functions satisfying certain properties by a certain $\mathrm{GL}_2(\mathbb{Q}_p)$-invariant subspace (see [Colmez 2010a; Breuil 2004] for details); however, a priori it is not clear whether or not it is nontrivial. On the other hand, there is a general construction of a (somewhat bigger) $\mathrm{GL}_2(\mathbb{Q}_p)$-representation $D \boxtimes_\delta \mathbb{P}^1$ that is in fact the space of global sections of a $\mathrm{GL}_2(\mathbb{Q}_p)$-equivariant sheaf $U \mapsto D \boxtimes_\delta U$ ($U \subseteq \mathbb{P}^1$ open) on the projective space $\mathbb{P}^1(\mathbb{Q}_p) \cong \mathrm{GL}_2(\mathbb{Q}_p)/B_2(\mathbb{Q}_p)$ for any (not necessarily 2-dimensional) (φ, Γ)-module D and any unitary character $\delta\colon \mathbb{Q}_p^\times \to o_K^\times$. This sheaf has the following properties: (i) the center of $\mathrm{GL}_2(\mathbb{Q}_p)$ acts via δ on $D \boxtimes_\delta \mathbb{P}^1$; (ii) we have $D \boxtimes_\delta \mathbb{Z}_p \cong D$ as a module over the monoid

$$\begin{pmatrix} \mathbb{Z}_p \setminus \{0\} & \mathbb{Z}_p \\ 0 & 1 \end{pmatrix}$$

(where we regard $\mathbb{Z}_p$ as an open subspace in $\mathbb{P}^1 = \mathbb{Q}_p \cup \{\infty\}$). (See [Schneider et al. 2012] for a generalization of this construction to general reductive groups.) Then Colmez shows that in case D is 2-dimensional and trianguline, there exists a unitary character δ (namely $\delta = \chi^{-1} \det D$, where χ is the cyclotomic character and $\det D$ is the character associated to the 1-dimensional (φ, Γ)-module $\bigwedge^2 D$ via Fontaine's equivalence composed with class field theory) such that a certain subspace $D^\natural \boxtimes_\delta \mathbb{P}^1$ (for the definition see [Colmez 2010b]) of $D \boxtimes_\delta \mathbb{P}^1$ is isomorphic to the dual of the Banach space representation $\Pi(\check{D})$ associated earlier to the dual (φ, Γ)-module $\check{D}$ — therefore showing in particular that the previous construction is nonzero. This subspace makes sense also when D is not trianguline (nor of rank 2), but a priori only known to be $B_2(\mathbb{Q}_p)$-invariant. Also, whenever D is indecomposable and 2-dimensional, then the above δ is unique [Paškūnas 2013], and whenever D is absolutely irreducible and at least 3-dimensional, then there does not exist such a character δ (so that the subspace $D^\natural \boxtimes_\delta \mathbb{P}^1$ is $\mathrm{GL}_2(\mathbb{Q}_p)$-invariant) [Paškūnas 2013]. Since the construction of $D \mapsto D^\natural \boxtimes_\delta \mathbb{P}^1$ behaves well in families (see Chapter II in [Colmez 2010c]) and the trianguline Galois representations are Zariski-dense in the deformation space of 2-dimensional (φ, Γ)-modules with given reduction mod p [Kisin 2010], Colmez [2010c] shows that this subspace is not only $B_2(\mathbb{Q}_p)$, but also $\mathrm{GL}_2(\mathbb{Q}_p)$-invariant for general 2-dimensional (φ, Γ)-modules. For $\delta = \chi^{-1} \det D$ (in this case we omit the subscript δ from the notation), we have a short exact sequence

$$0 \to \Pi(\check{D}) \to D \boxtimes \mathbb{P}^1 \to \Pi(D) \to 0,$$

where $\Pi(D)$ is the unitary Banach-space representation associated to D via the p-adic Langlands correspondence.

Colmez [2010c, Chapters V and VI] also identifies the space $\Pi(D)^{\mathrm{an}}$ of locally analytic and the space $\Pi(D)^{\mathrm{alg}}$ of locally algebraic vectors in the Banach-space representation $\Pi(D)$. These play a crucial role in the proof of the compatibility of the p-adic and classical local Langlands correspondences. In fact, we have $\Pi(D)^{\mathrm{an}} = (D^{\dagger} \boxtimes \mathbb{P}^1)/K \cdot (D^{\natural} \boxtimes \mathbb{P}^1)$, where $D^{\dagger} \boxtimes \mathbb{P}^1$ is the subspace of elements $x \in D \boxtimes \mathbb{P}^1$ such that both $\mathrm{Res}_{\mathbb{Z}_p}^{\mathbb{P}^1}(x)$ and

$$\mathrm{Res}_{\mathbb{Z}_p}^{\mathbb{P}^1}\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} x\right)$$

lie in the subspace of overconvergent elements $D^{\dagger} \subset D \cong D \boxtimes \mathbb{Z}_p$. $D^{\dagger}$ is an étale (φ, Γ)-module over the ring $\mathscr{E}^{\dagger}$ of overconvergent power series with coefficients in K such that $D \cong \mathscr{E} \otimes_{\mathscr{E}^{\dagger}} D^{\dagger}$ [Cherbonnier and Colmez 1998].

Let now G be the group of $\mathbb{Q}_p$-points of a connected $\mathbb{Q}_p$-split reductive group and let $P = TN$ be a Borel subgroup of G. Further denote by Φ^+ the set of positive roots with respect to P and $\Delta \subset \Phi^+$ the set of simple roots. The above noted generalizations of Colmez's work [Schneider and Vignéras 2011; Schneider et al. 2012] both use a certain microlocalization $\Lambda_{\ell}(N_0)$ (constructed originally in [Schneider and Venjakob 2010]) of the Iwasawa algebra $\Lambda(N_0)$ of a compact open subgroup N_0 of N. This can be thought of as the noncommutative analogue of Fontaine's ring $\mathbb{O}_{\mathscr{E}}$. On the other hand, Colmez's p-adic Langlands correspondence heavily relies on the theory of trianguline (φ, Γ)-modules. A (φ, Γ)-module over the Robba ring is a free module $D^{\dagger}_{\mathrm{rig}}$ over $\mathscr{R}$ together with commuting semilinear actions of the operator φ and the group Γ such that φ takes a basis of the free module to another basis. Such a (φ, Γ)-module $D^{\dagger}_{\mathrm{rig}}$ is said to be étale (or of slope 0) if there is a basis of $D^{\dagger}_{\mathrm{rig}}$ such that the matrix of φ in this basis is an invertible matrix over the subring $\mathbb{O}^{\dagger}_{\mathscr{E}} \subset \mathscr{R}$ of overconvergent Laurent series. An étale (φ, Γ)-module over $\mathscr{R}$ is *trianguline* if it admits a filtration of (not necessarily étale) (φ, Γ)-modules over $\mathscr{R}$ with subquotients of rank 1 possibly after a finite base change $E \otimes_K \cdot$. The fact that the Robba ring and the ring of overconvergent Laurent series play such a role in the construction of the p-adic Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$ and also in the identification of the locally analytic vectors is the motivation for the construction of noncommutative analogues of these rings — as they will most probably be needed for a future correspondence for reductive groups other than $\mathrm{GL}_2(\mathbb{Q}_p)$.

The motivation of this paper is twofold. On the one hand, we reinterpret the ring $\Lambda_{\ell}(N_0)$ as follows. Instead of localizing and completing the Iwasawa algebra $\Lambda(N_0)$, one may construct $\Lambda_{\ell}(N_0)$ as the projective limit of certain skew group rings over $\mathbb{O}_{\mathscr{E}}$. The only assumptions on the ring $R = \mathbb{O}_{\mathscr{E}}$ such that this new construction of $\Lambda_{\ell}(N_0)$ can be carried out are that R admits an inclusion $\chi\colon \mathbb{Z}_p \to R^{\times}$ of the additive group $\mathbb{Z}_p$ into its group of invertible elements and an étale action of an operator φ that is compatible with χ. The noncommutative ring that is constructed is a completed skew group ring $R[\![H_1, \ell]\!]$ of a closed normal subgroup H_1 of a pro-p group H_0 such that $\ell\colon H_0 \twoheadrightarrow \mathbb{Z}_p$ is a homomorphism with kernel H_1 (hence $H_0/H_1 \cong \mathbb{Z}_p$). The main result in this direction is Proposition 3.1, showing that the category of φ-modules over R is equivalent to the category of φ-modules over the completed skew group ring $R[\![H_1, \ell]\!]$. This can be applied also to the ring $R = \mathcal{O}_{\mathcal{E}}^{\dagger}$ of overconvergent Laurent series with coefficients in o_K and the Robba ring $\mathcal{R}$. The other motivation (probably the more important one) is the construction of the right noncommutative analogues of $\mathcal{O}_{\mathcal{E}}^{\dagger}$ and $\mathcal{R}$. The elements of the rings $\mathcal{R}[\![H_1, \ell]\!]$ and $\mathcal{O}_{\mathcal{E}}^{\dagger}[\![H_1, \ell]\!]$, however, are not necessarily convergent in any open annulus since they are obtained by taking an inverse limit. Therefore we shall we construct the rings $\mathcal{R}(H_1, \ell)$ and $\mathcal{R}^{\mathrm{int}}(H_1, \ell)$ (pages 221 ff. and 231–232, with N_1 for H_1) as direct limits of certain microlocalizations of the distribution algebra. The elements of these rings are convergent in a region of the form

$$\big\{\rho_2 < |b_\alpha| < 1,\ |b_\beta| < |b_\alpha|^r \text{ for } \beta \in \Phi^+ \setminus \{\alpha\}\big\}$$

for some $p^{-1} < \rho_2 < 1$ and $1 \leq r \in \mathbb{Z}$. We will show (pages 233 ff.) that $\mathcal{R}[\![N_1, \ell]\!]$ and $\mathcal{O}_{\mathcal{E}}^{\dagger}[\![N_1, \ell]\!]$ are in a certain sense completions of $\mathcal{R}(N_1, \ell)$ and $\mathcal{R}^{\mathrm{int}}(N_1, \ell)$, respectively. Note that, although the natural map $j_{\mathrm{int}}\colon \mathcal{R}^{\mathrm{int}}(N_1, \ell) \to \mathcal{O}_{\mathcal{E}}^{\dagger}[\![N_1, \ell]\!]$ is injective, the map $j\colon \mathcal{R}(N_1, \ell) \to \mathcal{R}[\![N_1, \ell]\!]$ is not. Both the ring $\mathcal{R}(N_1, \ell)$ and its integral version admit an étale action of the monoid $T_+ = \{t \in T \mid tN_0t^{-1} \subseteq N_0\}$. However, it is an open question whether the categories of étale T_+-modules over these rings are equivalent to the étale T_+-modules over their completions.

In my opinion, the right noncommutative analogue of the ring $\mathcal{R}$ (resp. $\mathcal{O}_{\mathcal{E}}^{\dagger}$) is $\mathcal{R}(N_1, \ell)$ (resp. $\mathcal{R}^{\mathrm{int}}(N_1, \ell)$) in the context of $\mathbb{Q}_p$-split reductive groups G over $\mathbb{Q}_p$, as both rings admit an étale action of the monoid T_+ and their elements converge in certain polyannuli. However, it might still be useful to also consider the rings $\mathcal{R}[\![H_1, \ell]\!]$ and $\mathcal{O}_{\mathcal{E}}^{\dagger}[\![H_1, \ell]\!]$, as they can help us compare the category of usual (φ, Γ)-modules with the category of T_+-modules over $\mathcal{R}(N_1, \ell)$ (resp. over $\mathcal{R}^{\mathrm{int}}(N_1, \ell)$) using the equivalence of categories in Proposition 3.1. Only one variable is inverted in these rings, in contrast to the rings constructed in [Zábrádi 2012]. The reasons for this are the following: (i) this way $\mathcal{R}^{\mathrm{int}}(N_1, \ell)$ is a subring of $\Lambda_\ell(N_0)$; (ii) the equivalence of categories in Proposition 3.1 holds for rings in which only one variable is inverted; and (iii) all the usual (φ, Γ)-modules are overconvergent, that is, they descend to $\mathcal{O}_{\mathcal{E}}^{\dagger}$ already in one variable. However, if $\mathbb{Q}_p$ is replaced by a finite unramified extension F, then one might have to consider Lubin–Tate (φ, Γ_F)-modules (with $\Gamma_F \cong o_F^{\times}$) instead so that the monoid $\varphi^{\mathbb{N}}\Gamma_F$ is isomorphic to $o_F \setminus \{0\}$. These (φ, Γ_F)-modules are not overconvergent in general but they might still correspond to objects over certain multivariable Robba rings (in which all the variables are inverted). For a first result in this direction see [Berger 2013]. It is plausible to expect that for general reductive groups G over F one has to invert exactly $|F:\mathbb{Q}_p|$ ($\mathbb{Q}_p$-)variables that correspond to the root subgroup $N_\alpha \cong F \cong \mathbb{Q}_p^{|F:\mathbb{Q}_p|}$ for a given simple root α.

2. Completed skew group rings

Let R be a commutative ring (with identity) with the following properties:

(i) There exists a group homomorphism $\chi\colon \mathbb{Z}_p \hookrightarrow R^\times$.

(ii) The ring R admits an étale action of the p-Frobenius φ that is compatible with χ. More precisely, there is an injective ring homomorphism $\varphi\colon R \hookrightarrow R$ such that $\varphi(\chi(x)) = \chi(px)$ and

$$R = \bigoplus_{i=0}^{p-1} \chi(i)\varphi(R).$$

In particular, R is free of rank p over $\varphi(R)$.

We remark first of all that one may iterate (ii) c times for any positive integer c to obtain

$$R = \bigoplus_{i=0}^{p^c-1} \chi(i)\varphi^c(R). \tag{1}$$

Indeed, by induction, we may assume that (1) holds for $c-1$ and obtain

$$\begin{aligned} R = \bigoplus_{k=0}^{p^{c-1}-1} \chi(k)\varphi^{c-1}(R) &= \bigoplus_{k=0}^{p^{c-1}-1} \chi(k)\varphi^{c-1}\Bigl(\bigoplus_{j=0}^{p-1} \chi(j)\varphi(R)\Bigr) \\ &= \bigoplus_{k=0}^{p^{c-1}-1} \bigoplus_{j=0}^{p-1} \chi(k+p^{c-1}j)\varphi^c(R), \end{aligned}$$

since φ^{c-1} takes direct sums to direct sums as it is injective. Now the claim follows from noting that any integer $0 \le i \le p^c - 1$ can be uniquely written in the form $i = k + p^{c-1}j$ with $0 \le k \le p^{c-1}-1$ and $0 \le j \le p-1$.

For any $x \in \mathbb{Z}_p$ we have $\chi(p^c x) = \varphi^c(\chi(x)) \in \varphi^c(R)^\times$. Hence $\chi(i)\varphi^c(R) = \chi(i + p^c x)\varphi^c(R)$ and we may replace each value of i in the formula (1) by any element in the coset $i + p^c\mathbb{Z}_p$.

Definition 2.1. We call a ring R with the above properties (i) and (ii) a φ-ring over $\mathbb{Z}_p$ or often just a φ-ring.

For example, if $K/\mathbb{Q}_p$ is a finite extension with ring of integers o and uniformizer p_K, then the Iwasawa algebra $o[\![T]\!]$ is a φ-ring with the homomorphism

$$\chi\colon \mathbb{Z}_p \to o[\![T]\!], \quad 1 \mapsto 1+T$$

and Frobenius $\varphi(T) = (T+1)^p - 1$. Similarly, with the same χ and φ, Fontaine's ring $\mathcal{O}_{\mathscr{E}}$, its field of fractions $\mathscr{E}$, the Robba ring $\mathscr{R}$ and the rings $\mathscr{E}^\dagger$, $\mathcal{O}_{\mathscr{E}}^\dagger$ of overconvergent power series are also φ-rings. (For the definitions of $\mathcal{O}_{\mathscr{E}}$ and $\mathscr{E}$ see the paragraph before Lemma 2.13, and for those of $\mathscr{R}$, $\mathcal{O}_{\mathscr{E}}^\dagger$, and $\mathscr{E}^\dagger$ see (12) and subsequent paragraph.)

Lemma 2.2. *For any positive integer c we have a ring isomorphism*

$$\varphi^c(R)[X]/\big(X^{p^c} - \chi(p^c)\big) \xrightarrow{\sim} R, \quad X \mapsto \chi(1).$$

Proof. Since the polynomial ring $\varphi^c(R)[X]$ is a free object in the category of commutative $\varphi^c(R)$-algebras, we may extend the natural inclusion homomorphism $f\colon \varphi^c(R) \hookrightarrow R$ given by (ii) to a ring homomorphism $\tilde{f}\colon \varphi^c(R)[X] \to R$ by any free choice for the value $\tilde{f}(X)$, in particular such that $\tilde{f}(X) := \chi(1) \in R$ and, of course, $\tilde{f}_{|\varphi^c(R)} := f$. We need to show that $\tilde{f}$ is surjective with kernel equal to the ideal generated by $X^{p^c} - \chi(p^c)$. Note that $\chi(p^c) = \varphi^c(\chi(1))$ lies in $\varphi^c(R)$, so the claim makes sense.

By (i), the map χ is a group homomorphism, so $\chi(r) = \chi(1)^r = \tilde{f}(X)^r = \tilde{f}(X^r)$ lies in the image of $\tilde{f}$ for any positive integer r. Hence we obtain the surjectivity from (1) by noting that $\varphi^c(R)$ also lies in the image of $\tilde{f}$.

Using again $\chi(r) = \tilde{f}(X^r)$ with the choice of $r = p^c$, we see immediately that $X^{p^c} - \chi(p^c)$ lies in the kernel of $\tilde{f}$. Moreover,

$$\varphi^c(R)[X]/\big(X^{p^c} - \chi(p^c)\big)$$

is a free module of rank p^c over $\varphi^c(R)$ with generators the classes of $\{X^r\}_{r=0}^{p^c-1}$ in the quotient. On the other hand, R is also a free module of rank p^c with generators $\{\chi(r)\}_{r=0}^{p^c-1}$, by (1), and these two sets of generators correspond to each other under the map $\tilde{f}$; hence the isomorphism. □

Let H_0 be a pro-p group of finite rank (therefore a compact p-adic Lie group by Corollary 4.3 and Theorem 8.18 in [Dixon et al. 1999]) without elements of order p admitting a continuous surjective group homomorphism $\ell\colon H_0 \twoheadrightarrow \mathbb{Z}_p$ with kernel $H_1 := \mathrm{Ker}(\ell)$. We further assume the following:

(A) H_0 also admits an injective group endomorphism $\varphi\colon H_0 \hookrightarrow H_0$ with finite cokernel and compatible with ℓ in the sense that $\ell(\varphi(h)) = \varphi(\ell(h)) = p\ell(h)$. In particular, we have $\varphi(H_1) \subseteq H_1$.

(B) $\bigcap_{n\geq 1} \varphi^n(H_0) = \{1\}$ and the subgroups $\varphi^n(H_0)$ form a system of neighborhoods of 1 in H_0.

We remark first of all that by a theorem of Serre [Dixon et al. 1999, Theorem 1.17], any finite index subgroup in H_0 is open. Hence the homomorphism φ is automatically continuous and the subgroups $\varphi^n(H_0)$ are open.

H_1 is a closed subgroup of H_0; hence it is also a pro-p group of finite rank. By assumption (B), we also have in particular that the subgroups $\varphi^n(H_1)$ form a system of open neighborhoods of 1 in H_1. The subgroups $\varphi^n(H_1)$ may not be normal in either H_1 or H_0. Hence for $k > 1$ we define the normal subgroup $H_k \lhd H_0$ as the normal subgroup of H_0 generated by $\varphi^{k-1}(H_1)$. Since H_1 is normal in H_0 we automatically have $H_k \subseteq H_1$ for any $k \geq 1$. Also, since the p-adic Lie group H_1 has a system of neighborhoods of 1 containing only characteristic subgroups, the H_k also form a system of neighborhoods of 1 in H_1. On the other hand, we have by definition that $\varphi(H_k) \subseteq H_{k+1} \subseteq H_k$ for each $k \geq 1$. In particular, we have an induced φ action on the quotient group H_0/H_k. This is, of course, no longer injective.

Since the group $\mathbb{Z}_p$ is topologically generated by one element, we may find a splitting $\iota: \mathbb{Z}_p \hookrightarrow H_0$ for the group homomorphism ℓ. We fix this splitting ι, too. Assume further that:

(C) the group homomorphism ι is φ-equivariant, that is, we have $\iota(\varphi(x)) = \varphi(\iota(x))$ for all $x \in \mathbb{Z}_p$.

We define the skew group ring $R[H_1/H_k, \ell, \iota]$ as follows. We put

$$R[H_1/H_k, \ell, \iota] := \bigoplus_{h \in H_1/H_k} Rh \tag{2}$$

as left R-modules. Since H_1 is a normal subgroup in H_0, we also have

$$H_1/H_k \lhd H_0/H_k.$$

Therefore we obtain a conjugation action of $\mathbb{Z}_p$ on H_1/H_k given by

$$\rho: \mathbb{Z}_p \to \mathrm{Aut}(H_1/H_k), \quad z \mapsto \big(h \mapsto \iota(z)h\iota(z)^{-1}, h \in H_1/H_k\big).$$

Since H_1/H_k is a finite p-group, $\mathrm{Aut}(H_1/H_k)$ is finite and we have an integer $c_k \geq 1$ such that $p^{c_k}\mathbb{Z}_p \subseteq \mathrm{Ker}(\rho)$. The multiplication is defined so that $\varphi^{c_k}(R)$ commutes with elements h in H_1/H_k and $\chi(i)$ acts on H_1/H_k via $\iota \circ \chi^{-1}$ and conjugation. More precisely, for $r_1, r_2 \in R$ and $h_1, h_2 \in H_1/H_k$ we may write

$$r_2 = \sum_{i=0}^{p^{c_k}-1} \chi(i)\varphi^{c_k}(r_{i,2}) \tag{3}$$

and put

$$(r_1h_1)(r_2h_2) := \sum_{i=0}^{p^{c_k}-1} r_1\chi(i)\varphi^{c_k}(r_{i,2})\big((\iota(i)^{-1}h_1\iota(i))h_2\big) \in \bigoplus_{h \in H_1/H_k} Rh. \tag{4}$$

In case $r_2 = 1$ we have $(r_1h_1)h_2 = r_1(h_1h_2)$, and in case $h_1 = 1$ we have $r_1(r_2h_2) = (r_1r_2)h_2$. By the choice of c_k, $\iota(p^{c_k}\mathbb{Z}_p)$ lies in the center of H_0/H_k. So we may use any set of representatives of $\mathbb{Z}_p/p^{c_k}\mathbb{Z}_p$ instead of $\{0, 1, \dots, p^{c_k} - 1\}$ in (3) in order to compute (4). Indeed, if $i \equiv i' \pmod{p^{c_k}}$, then

$$\chi(i)\varphi^{c_k}(r_{i,2}) = \chi(i')\varphi^{c_k}\left(\chi\left(\frac{i-i'}{p^{c_k}}\right)r_{i,2}\right)$$

and $\iota(i)^{-1}h_1\iota(i) = \iota(i')^{-1}h_1\iota(i')$.

Lemma 2.3. *The multiplication* (4) *equips* $R[H_1/H_k, \ell, \iota]$ *with a ring structure.*

Proof. There exists an easy, but rather long computation showing this. However, there is another, more conceptual description of the ring $R[H_1/H_k, \ell, \iota]$ pointed out by Torsten Schoeneberg that proves this lemma without any computations. Let S be the group ring $S := \varphi^{c_k}(R)[H_1/H_k]$ and σ be the automorphism of S trivial on $\varphi^{c_k}(R)$ and acting by conjugation with $\iota(1)$ on H_1/H_k, that is, for $h \in H_1/H_k$ put $\sigma(h) := \iota(1)^{-1}h\iota(1)$. Now define the skew polynomial ring $S[X, \sigma]$ by the relation $aX = X\sigma(a)$ for $a \in S$. By the definition of σ, the subring $\varphi^{c_k}(R)$ lies in the center of $S[X, \sigma]$; therefore, so does $\chi(p^{c_k}) = \varphi^{c_k}(\chi(1)) \in \varphi^{c_k}(R)$. On the other hand, we have

$$aX^{p^{c_k}} = X^{p^{c_k}}\sigma^{p^{c_k}}(a) = X^{p^{c_k}}a$$

for all $a \in S$, since $\sigma^{p^{c_k}}$ is the conjugation by the central element $\iota(1)^{p^{c_k}} = \iota(p^{c_k})$ of H_0/H_k on H_1/H_k and is trivial by definition on $\varphi^{c_k}(R)$; hence $\sigma^{p^{c_k}} = \mathrm{id}_S$. This shows that $X^{p^{c_k}} - \chi(p^{c_k})$ is central and that

$$S[X,\sigma]\bigl(X^{p^{c_k}} - \chi(p^{c_k})\bigr) = \bigl(X^{p^{c_k}} - \chi(p^{c_k})\bigr)S[X,\sigma]$$

is a two-sided ideal in $S[X, \sigma]$. So we may form the quotient ring and compute (as left $\varphi^{c_k}(R)$-modules)

$$\begin{aligned} S[X,\sigma]/\bigl(X^{p^{c_k}} - \chi(p^{c_k})\bigr) &\cong \Bigl(\bigoplus_{r=0}^{\infty}\bigoplus_{h\in H_1/H_k} X^r\varphi^{c_k}(R)h\Bigr)\Big/\bigl(X^{p^{c_k}} - \chi(p^{c_k})\bigr) \\ &\cong \bigoplus_{h\in H_1/H_k}\bigl(\varphi^{c_k}(R)[X]/(X^{p^{c_k}} - \chi(p^{c_k}))\bigr)h \\ &\cong \bigoplus_{h\in H_1/H_k} Rh, \end{aligned}$$

using Lemma 2.2 in the middle. On the component $h = 1$ in the above direct sum, the identification is even multiplicative as Lemma 2.2 gives an isomorphism of rings, not just $\varphi^{c_k}(R)$-modules. Hence $S[X,\sigma]/\bigl(X^{p^{c_k}} - \chi(p^{c_k})\bigr)$ contains R as a subring and the isomorphism above is an isomorphism of left R-modules. The transport of ring structure gives back the definition (4) of multiplication on the right side. Indeed, we have

$$\begin{aligned}(r_1h_1)(r_2h_2) &= \sum_{i=0}^{p^{c_k}-1} r_1h_1\chi(i)\varphi^{c_k}(r_{i,2})h_2 = \sum_{i=0}^{p^{c_k}-1} r_1h_1X^i\varphi^{c_k}(r_{i,2})h_2 \\ &= \sum_{i=0}^{p^{c_k}-1} r_1X^i\sigma^i(h_1)\varphi^{c_k}(r_{i,2})h_2 = \sum_{i=0}^{p^{c_k}-1} r_1\chi(i)\varphi^{c_k}(r_{i,2})\big((\iota(i)^{-1}h_1\iota(i))h_2\big),\end{aligned}$$

since $\chi(i)$ corresponds to X^i under the isomorphism in Lemma 2.2. □

We further have a natural action of φ on $R[H_1/H_k, \ell, \iota]$ coming from the φ-action on both R and H_1/H_k by putting $\varphi(rh) := \varphi(r)\varphi(h)$ for $r \in R$ and $h \in H_1/H_k$.

Lemma 2.4. *The map* $\varphi\colon R[H_1/H_k, \ell, \iota] \to R[H_1/H_k, \ell, \iota]$ *defined above is a ring homomorphism.*

Proof. The additivity is clear, so it suffices to check the multiplicativity. Using (4) we compute

$$\begin{aligned}\varphi((r_1h_1)(r_2h_2)) &= \sum_{i=0}^{p^{c_k}-1} \varphi\big(r_1\chi(i)\varphi^{c_k}(r_{i,2})\big)\varphi\big((\iota(i)^{-1}h_1\iota(i))h_2\big) \\ &= \sum_{i=0}^{p^{c_k}-1} \varphi(r_1)\chi(pi)\varphi^{c_k+1}(r_{i,2})\big((\iota(pi)^{-1}\varphi(h_1)\iota(pi))\varphi(h_2)\big) \\ &= \sum_{i=0}^{p^{c_k}-1} \varphi(r_1)\varphi^{c_k+1}(r_{i,2})\varphi(h_1)\chi(pi)\varphi(h_2) \\ &= \varphi(r_1)\varphi(h_1)\sum_{i=0}^{p^{c_k}-1} \chi(pi)\varphi^{c_k+1}(r_{i,2})\varphi(h_2) \\ &= \varphi(r_1h_1)\varphi\bigg(\sum_{i=0}^{p^{c_k}-1} \chi(i)\varphi^{c_k}(r_{i,2})h_2\bigg) = \varphi(r_1h_1)\varphi(r_2h_2). \qquad □\end{aligned}$$

The map χ and the inclusion of the group H_1/H_k in the multiplicative group of $R[H_1/H_k, \ell, \iota]$ are compatible in the sense that they glue together to a φ-equivariant group homomorphism $\chi_k\colon H_0 \to R[H_1/H_k, \ell, \iota]^\times$ (with kernel $\operatorname{Ker}\chi_k = H_k$), making the diagram

$$\begin{tikzcd} \mathbb{Z}_p \arrow[r, "\iota"'] \arrow[d, "\chi"'] & H_0 \arrow[d, "\chi_k"'] \\ R \arrow[r, "\iota_{R,k}"'] & R[H_1/H_k, \ell] \end{tikzcd} \tag{5}$$

commutative, where $\iota_{R,k}$ is the natural inclusion of R in $R[H_1/H_k,\ell]$. Indeed, $H_0 \cong \iota(\mathbb{Z}_p) \ltimes H_1$, so we put

$$\chi_k(\iota(i)h) := \chi(i)(hH_k)$$

for $i \in \mathbb{Z}_p$, $h \in H_1$ and compute

$$\begin{aligned}\chi_k\big(\iota(i_1)h_1\iota(i_2)h_2\big) &= \chi_k\big(\iota(i_1+i_2)\iota(i_2)^{-1}h_1\iota(i_2)h_2\big)\\ &= \chi(i_1+i_2)\big(\iota(i_2)^{-1}h_1\iota(i_2)h_2\big)H_k\\ &= \chi(i_1)(h_1H_k)\chi(i_2)(h_2H_k) = \chi_k(\iota(i_1)h_1)\chi_k(\iota(i_2)h_2),\end{aligned}$$

showing that χ_k is indeed a group homomorphism. The commutativity of the diagram (5) is clear by definition. Moreover, χ_k is φ-equivariant, since we have

$$\begin{aligned}\chi_k\circ\varphi(\iota(i)h) = \chi_k(\iota(pi)\varphi(h)) &= \chi(pi)\varphi(h)H_k\\ &= \varphi(\chi(i)hH_k) = \varphi\circ\chi_k(\iota(i)h).\end{aligned}$$

Lemma 2.5. *The above definition of $R[H_1/H_k,\ell,\iota]$ does not depend on the choice of the section ι up to natural isomorphism.*

Proof. Let $\iota'\colon\mathbb{Z}_p \hookrightarrow H_0$ be another section of ℓ. The integer c_k depends on ι, but we also have another integer c'_k such that $\iota'(p^{c'_k})$ acts trivially by conjugation on H_1/H_k, that is, $\iota'(p^{c'_k})$ lies in the center of H_0/H_k. On the other hand, we may choose $m_k \geq 0$ so that $H_1^{p^{m_k}} \subseteq H_k$, since H_1/H_k is a finite p-group. From $\ell\circ\iota = \mathrm{id}_{\mathbb{Z}_p} = \ell\circ\iota'$, we see that $\iota^{-1}\iota'(\mathbb{Z}_p) \subseteq \operatorname{Ker}(\ell) = H_1$, and hence for any $x\in\mathbb{Z}_p$ we have

$$\begin{aligned}&\iota^{-1}(p^{m_k+\max(c_k,c'_k)}x)\iota'(p^{m_k+\max(c_k,c'_k)}x)\\ &\qquad= \iota^{-1}(p^{\max(c_k,c'_k)}x)^{p^{m_k}}\iota'(p^{\max(c_k,c'_k)}x)^{p^{m_k}}\\ &\qquad= \big(\iota^{-1}(p^{\max(c_k,c'_k)}x)\iota'(p^{\max(c_k,c'_k)}x)\big)^{p^{m_k}} \in H_1^{p^{m_k}} \subseteq H_k.\end{aligned}$$

Therefore for $m \geq m_k + \max(c_k, c'_k)$, the map $\iota'_k\colon R \hookrightarrow R[H_1/H_k,\ell,\iota]$ given by

$$\iota'_k\left(\sum_{i=0}^{p^m-1}\chi(i)\varphi^m(r_i)\right) := \sum_{i=0}^{p^m-1}\chi(i)\varphi^m(r_i)\big(\iota(i)^{-1}\iota'(i)\big) \tag{6}$$

extends to an isomorphism

$$\iota'_k\colon R[H_1/H_k,\ell,\iota'] \to R[H_1/H_k,\ell,\iota], \quad rh \mapsto \iota'_k(r)h,$$

of φ-rings. Indeed, the map ι'_k is clearly additive and bijective. We claim that it is multiplicative and φ-equivariant. We first show the latter statement and compute

$$\begin{aligned}\varphi \circ \iota'_k\Bigl(\sum_{i=0}^{p^m-1} \chi(i)\varphi^m(r_i)\Bigr) &= \sum_{i=0}^{p^m-1} \varphi\bigl(\chi(i)\varphi^m(r_i)\bigl(\iota(i)^{-1}\iota'(i)\bigr)\bigr)\\ &= \sum_{i=0}^{p^m-1} \chi(pi)\varphi^{m+1}(r_i)\bigl(\iota(pi)^{-1}\iota'(pi)\bigr)\\ &= \iota'_k\Bigl(\sum_{i=0}^{p^m-1} \chi(pi)\varphi^{m+1}(r_i)\Bigr) = \iota'_k \circ \varphi\Bigl(\sum_{i=0}^{p^m-1} \chi(i)\varphi^m(r_i)\Bigr).\end{aligned}$$

Since $m \geq \max(c_k, c'_k)$, the subring $\varphi^m(R)$ lies in the center of both $R[H_1/H_k, \ell, \iota]$ and $R[H_1/H_k, \ell, \iota']$. Therefore — in view of the associativity (Lemma 2.3) — we may compute the multiplication (4) by expanding elements of R to degree m. So we write

$$r_1 = \sum_{j=0}^{p^m-1} \chi(j)\varphi^m(r_{j,1}), \quad r_2 = \sum_{i=0}^{p^m-1} \chi(i)\varphi^m(r_{i,2}).$$

We may compute (6) using any set of representatives of $\mathbb{Z}_p/p^m\mathbb{Z}_p$ (for example $\{j, j+1, \dots, j+p^m-1\}$ instead of $\{0, 1, \dots, p^m-1\}$) since $\iota^{-1}\iota'(p^m\mathbb{Z}_p) \subseteq H_k$. Hence we obtain

$$\begin{aligned}&\iota'_k\bigl((r_1h_1)(r_2h_2)\bigr)\\ &= \iota'_k\Bigl(\sum_{i=0}^{p^m-1} r_1\chi(i)\varphi^m(r_{i,2})\bigl(\iota'(i)^{-1}h_1\iota'(i)h_2\bigr)\Bigr)\\ &= \iota'_k\Bigl(\sum_{i,j=0}^{p^m-1} \chi(j)\varphi^m(r_{j,1})\chi(i)\varphi^m(r_{i,2})\bigl(\iota'(i)^{-1}h_1\iota'(i)h_2\bigr)\Bigr)\\ &= \sum_{i=0}^{p^m-1} \iota'_k\Bigl(\sum_{j=0}^{p^m-1} \chi(i+j)\varphi^m(r_{j,1}r_{i,2})\Bigr)\iota'(i)^{-1}h_1\iota'(i)h_2\\ &= \sum_{i,j=0}^{p^m-1} \chi(i+j)\varphi^m(r_{j,1}r_{i,2})\iota(i+j)^{-1}\iota'(i+j)\iota'(i)^{-1}h_1\iota'(i)h_2\\ &= \sum_{i,j=0}^{p^m-1} \chi(i+j)\varphi^m(r_{j,1}r_{i,2})\iota(i+j)^{-1}\iota'(j)h_1\iota'(i)h_2\\ &= \sum_{i,j=0}^{p^m-1} \chi(j)\varphi^m(r_{j,1})\chi(i)\varphi^m(r_{i,r})\iota(i)^{-1}\bigl(\iota(j)^{-1}\iota'(j)h_1\bigr)\iota(i)\bigl(\iota(i)^{-1}\iota'(i)h_2\bigr)\\ &= \Bigl(\sum_{j=0}^{p^m-1} \chi(j)\varphi^m(r_{j,1})\bigl(\iota(j)^{-1}\iota'(j)h_1\bigr)\Bigr)\Bigl(\sum_{i=0}^{p^m-1} \chi(i)\varphi^m(r_{i,2})\bigl(\iota(i)^{-1}\iota'(i)h_2\bigr)\Bigr)\\ &= \Bigl(\sum_{j=0}^{p^m-1} \varphi^m(r_{j,1})\iota'(j)h_1\Bigr)\Bigl(\sum_{i=0}^{p^m-1} \varphi^m(r_{i,2})\iota'(i)h_2\Bigr) = \iota'_k(r_1h_1)\iota'_k(r_2h_2). \qquad \square\end{aligned}$$

In view of this lemma we omit ι from the notation from now on. This construction is compatible with the natural surjective homomorphisms $H_1/H_{k+1} \twoheadrightarrow H_1/H_k$; therefore the rings $R[H_1/H_k, \ell]$ form an inverse system for the induced maps. So we may define the completed skew group ring $R[\![H_1, \ell]\!]$ as the projective limit

$$R[\![H_1, \ell]\!] := \varprojlim_k R[H_1/H_k, \ell].$$

We denote by I_k the kernel of the canonical surjective homomorphism from $R[\![H_1, \ell]\!]$ to $R[H_1/H_k, \ell]$.

Whenever R is a topological ring, we equip $R[\![H_1, \ell]\!]$ with the projective limit topology of the product topologies on each $\bigoplus_{h \in H_1/H_k} Rh$.

The augmentation map $H_1 \twoheadrightarrow 1$ induces a ring homomorphism

$$\ell := \ell_R \colon R[\![H_1, \ell]\!] \twoheadrightarrow R.$$

This also has a section $\iota := \iota_R = \varprojlim \iota_{R,k} \colon R \hookrightarrow R[\![H_1, \ell]\!]$ (whenever clear we omit the subscript $_R$), that is, $\ell_R \circ \iota_R = \mathrm{id}_R$. By (5), the group homomorphism $\chi \colon \mathbb{Z}_p \to R^\times$ extends to a group homomorphism $\chi_{H_0} \colon H_0 \to R[\![H_1, \ell]\!]^\times$, making the diagram

$$\begin{tikzcd}
\mathbb{Z}_p \arrow[r, "\iota"'] \arrow[d, "\chi"'] & H_0 \arrow[d, "\chi_{H_0}"'] \\
R \arrow[r, "\iota_R"'] & R[\![H_1, \ell]\!]
\end{tikzcd}$$

commutative.

The operator φ acts naturally on this projective limit. If R is a topological ring and φ acts continuously on R, then φ also acts continuously on each $R[H_1/H_k, \ell]$ by taking the limit also on $R[\![H_1, \ell]\!]$. For an open subgroup H' of a profinite group H we use the notation $J(H/H')$ for a set of representatives of the left cosets of H' in H. Similarly, we use $J(H' \setminus H)$ for a set of representatives of the right cosets $H' \setminus H$.

Lemma 2.6. (a) *Let $L \leq K \leq H$ be groups. Then the set $J(H/K)J(K/L)$ (resp. $J(L \setminus K)J(K \setminus H)$) is a set of representatives for the cosets H/L (resp. for $L \setminus H$).*

(b) *Let $K \leq H$ be groups and $N \lhd H$ a* normal *subgroup. Then $J((K \cap N) \setminus N)$ is also a set of representatives for $K \setminus KN$.*

Proof. These are well known facts in group theory; however, for the convenience of the reader, we recall their proofs here. In (b) we need N to be a normal subgroup so that KN is a subgroup of H. Moreover, $J(K \setminus KN)$ might not lie in N in general.

(a) Let $h_1, h_2 \in J(H/K)$ and $k_1, k_2 \in J(K/L)$. Suppose we have $h_1k_1L = h_2k_2L$. Then we also have $h_1^{-1}h_2 \in K$ and so $h_1 = h_2$, whence $k_1^{-1}k_2 \in L$ and so $k_1 = k_2$.

So the elements of the set $J(H/K)J(K/L)$ are in distinct left cosets of L. On the other hand, if $hL \in H/L$ is a left coset, then we may first choose $h_1 \in J(H/K)$ so that $h_1^{-1}h \in K$ and then $k_1 \in J(K/L)$ so that $k_1^{-1}h_1^{-1}h \in L$; that is, $hL = h_1k_1L$.

(b) If $n_1 \neq n_2 \in J((K \cap N) \setminus N)$ are distinct, then $Kn_1 \neq Kn_2$, as $n_1n_2^{-1}$ does not lie in $K \cap N$, but it lies in N. On the other hand, if $kn \in KN$, then we may find $n_1 \in J((K \cap N) \setminus N)$ such that $nn_1^{-1} \in K \cap N$, and hence $knn_1^{-1} \in K$. □

Proposition 2.7. *The map $\varphi\colon R[\![H_1,\ell]\!] \to R[\![H_1,\ell]\!]$ is injective. Also*

$$R[\![H_1,\ell]\!] = \bigoplus_{h \in J(\varphi(H_0)\setminus H_0)} \varphi(R[\![H_1,\ell]\!])h.$$

In particular, $R[\![H_1,\ell]\!]$ is a free (left) module of rank $[H_0 : \varphi(H_0)]$ over itself via φ.

Proof. Step 1. Let k be an integer and denote by A_k the kernel of the map $\varphi\colon R[H_1/H_k,\ell] \to R[H_1/H_k,\ell]$ so that we have a short exact sequence of abelian groups

$$0 \to A_k \to R[H_1/H_k,\ell] \xrightarrow{\varphi} \varphi(R[H_1/H_k,\ell]) \to 0.$$

We show that the sequence A_k satisfies the trivial Mittag-Leffler condition. From this the injectivity of φ follows, and we obtain

$$\varprojlim_k \varphi(R[H_1/H_k,\ell]) \cong \varphi(\varprojlim_k R[H_1/H_k,\ell]) = \varphi(R[\![H_1,\ell]\!]). \tag{7}$$

Take a fixed positive integer k. Since $\varphi\colon H_1 \to H_1$ is an open map (bijective and continuous between the compact sets H_1 and $\varphi(H_1)$, and hence a homeomorphism) and the subgroups H_l form a system of neighborhoods, we find an integer $l > k$ such that $H_k \supseteq \varphi^{-1}(H_l)$. In view of Lemma 2.6 we put

$$J(H_1/H_l) := J(H_1/\varphi^{-1}(H_l))J(\varphi^{-1}(H_l)/H_l)$$

for $J(H_1/\varphi^{-1}(H_l))$ and $J(\varphi^{-1}(H_l)/H_l)$ arbitrarily fixed sets of representatives for the cosets of $H_1/\varphi^{-1}(H_l)$ and of $\varphi^{-1}(H_l)/H_l$, respectively.

Now let $\sum_{h \in J(H_1/H_l)} r_h\chi_l(h)$ be an element in A_l and denote by $f_{k,l}$ the natural surjection from $R[H_1/H_l,\ell] \twoheadrightarrow R[H_1/H_k,\ell]$. We have

$$\begin{aligned}0 &= \varphi\Big(\sum_{h\in J(H_1/H_l)} r_h\chi_l(h)\Big)\\ &= \sum_{h_1\in J(H_1/\varphi^{-1}(H_l))}\ \sum_{h_2\in J(\varphi^{-1}(H_l)/H_l)} \varphi(r_{h_1h_2})\chi_l(\varphi(h_1h_2))\\ &= \sum_{h_1}\sum_{h_2}\varphi(r_{h_1h_2})\chi_l(\varphi(h_1))\\ &= \sum_{h_1\in J(H_1/\varphi^{-1}(H_l))} \varphi\Big(\sum_{h\in J(H_1/H_l)\cap h_1\varphi^{-1}(H_l)} r_h\Big)\chi_l(\varphi(h_1)).\end{aligned}$$

For $h_1 \neq h_1' \in J(H_1/\varphi^{-1}(H_l))$ we have $\varphi(h_1)H_l \neq \varphi(h_1')H_l$. Since $R[H_1/H_l,\ell]$ is defined as a direct sum, we obtain

$$\varphi\Big(\sum_{h\in J(H_1/H_l)\cap h_1\varphi^{-1}(H_l)} r_h\Big) = 0, \quad \text{whence} \sum_{h\in J(H_1/H_l)\cap h_1\varphi^{-1}(H_l)} r_h = 0$$

for any fixed $h_1 \in J(H_1/\varphi^{-1}(H_l))$, as φ is injective on R. On the other hand, we have

$$\begin{aligned}f_{k,l}\Big(\sum_{h\in J(H_1/H_l)} r_h\chi_l(h)\Big) &= \sum_{h_1\in J(H_1/H_k)}\Big(\sum_{h\in J(H_1/H_l)\cap h_1H_k} r_h\Big)\chi_k(h_1)\\ &= \sum_{h_1\in J(H_1/H_k)} 0\chi_k(h_1) = 0,\end{aligned}$$

as h_1H_k is a disjoint union of cosets of $\varphi^{-1}(H_l)$ by the choice of l. This shows that $f_{k,l}(A_l) = 0$ as claimed. Therefore (7) follows as discussed above.

Step 2. Since $\varphi(H_0) \cap H_1$ is open in H_1, there exists an integer $k_0 \geq 2$ such that for $k \geq k_0$ we have $H_k \subseteq \varphi(H_0)$. (We may not be able to take $k_0 = 2$ because H_k is the *normal* subgroup generated by $\varphi(H_1)$, which does have elements outside $\varphi(H_0)$ in general.) We claim now the decomposition

$$R[H_1/H_k,\ell] = \bigoplus_{h\in J(\varphi(H_0)\backslash H_0)} \varphi(R[H_1/H_k,\ell])\chi_k(h) \tag{8}$$

for $k \geq k_0$. Since H_k is a normal subgroup of H_0 contained in $\varphi(H_0)$, the elements $\chi_k(h)$ above are distinct.

For the proof of (8) we apply Lemma 2.6(b) in the situation $K := \varphi(H_0)$, $N := H_1$, and $H := H_0$ to be able to choose

$$J(\varphi(H_0) \setminus \varphi(H_0)H_1) := J((\varphi(H_0) \cap H_1) \setminus H_1).$$

Also, by the injectivity of φ on H_0/H_1, we see that $\varphi(H_0)\cap H_1 = \varphi(H_1)$. On the other hand, $\iota(\{0,1,\dots,p-1\})$ is a set of representatives for the cosets $H_1\varphi(H_0)\setminus H_0$. Therefore (using Lemma 2.6(a) with $L := \varphi(H_0)$, $K := \varphi(H_0)H_1$, and $H := H_0$) we may choose

$$\begin{aligned} J(\varphi(H_0)\setminus H_0) &:= J(\varphi(H_0)\setminus \varphi(H_0)H_1)J(\varphi(H_0)H_1\setminus H_0)\\ &= J(\varphi(H_1)\setminus H_1)\iota(\{0,1,\dots,p-1\}). \end{aligned}$$

We are going to use this specific set $J(\varphi(H_0)\setminus H_0)$ in order to compute the right side of (8). Let $\sum_{h\in J(H_1/H_k)} r_h\chi_k(h)$ be an arbitrary element in $R[H_1/H_k,\ell]$. By the étaleness of the action of φ on R (noting that R is commutative), we may uniquely decompose

$$r_h = \sum_{i=0}^{p-1}\chi(i)\varphi(r_{i,h}) = \sum_{i=0}^{p-1}\varphi(r_{i,h})\chi(i).$$

On the other hand, we write $\iota(i)h\iota(i)^{-1} = \varphi(u_{i,h})v_{i,h}$ with unique $u_{i,h}\in H_1$ and $v_{i,h}\in J(\varphi(H_1)\setminus H_1)$. Therefore we have

$$\begin{aligned} \sum_{h\in J(H_1/H_k)} r_h\chi_k(h) &= \sum_{h\in J(H_1/H_k)}\sum_{i=0}^{p-1}\varphi(r_{i,h})\chi(i)\chi_k\bigl(\iota(i)^{-1}\varphi(u_{i,h})v_{i,h}\iota(i)\bigr)\\ &= \sum_{h\in J(H_1/H_k)}\sum_{i=0}^{p-1}\varphi(r_{i,h}\chi_k(u_{i,h}))\chi_k(v_{i,h}\iota(i))\\ &\qquad\in \sum_{h\in J(\varphi(H_0)\setminus H_0)}\varphi(R[H_1/H_k,\ell])\chi_k(h), \end{aligned}$$

as $\chi(i)=\chi_k(\iota(i))$ and $\chi_k\circ\varphi=\varphi\circ\chi_k$, by (5).

It remains to show that the sum in (8) is indeed direct. For this we may expand any element $x_{i,h}\in R[H_1/H_k,\ell]$ as

$$x_{i,h} = \sum_{m\in J(H_1/H_k)} r_{i,h,m}\chi_k(m)$$

and compute

$$\begin{aligned} &\sum_{i=0}^{p-1}\sum_{h\in J(\varphi(H_1)\setminus H_1)}\varphi(x_{i,h})\chi_k(h\iota(i))\\ &= \sum_{i,h,m}\varphi(r_{i,h,m})\chi_k(\varphi(m)h\iota(i))\\ &= \sum_{i,h,m}\chi(i)\varphi(r_{i,h,m})\chi_k\bigl(\iota(i)^{-1}\varphi(m)h\iota(i)\bigr) \end{aligned}$$

$$= \sum_{i,h} \sum_{m_0 \in J(\varphi(H_1)/H_k)} \chi(i) \Bigg(\sum_{m \in J(H_1/H_k) \cap \varphi^{-1}(m_0 H_k)} \varphi(r_{i,h,m}) \Bigg) \chi_k\big(\iota(i)^{-1} m_0 h \iota(i)\big). \quad (9)$$

Assume now that the left side of (9) is 0. The set $J(\varphi(H_1)/H_k)J(\varphi(H_1)\setminus H_1)$ is a set of representatives of $H_k \backslash H_1$ because H_k is normal in H_1, whence $\varphi(H_1)/H_k = H_k \setminus \varphi(H_1)$. This shows that the elements $m_0 h$ are distinct in H_1/H_k on the right side of (9). The conjugation by $\iota(i)$ is an automorphism of H_1/H_k; therefore the elements $\iota(i)^{-1} m_0 h \iota(i)$ are also distinct for any fixed $i \in \{0, 1, \dots, p-1\}$. On the other hand, by the étaleness of φ on R and by (2), we obtain

$$R[H_1/H_k, \ell] = \bigoplus_{i=0}^{p-1} \bigoplus_{h_1 \in H_1/H_k} \chi(i)\varphi(R)h.$$

Hence we have

$$\sum_{m \in J(H_1/H_k) \cap \varphi^{-1}(m_0 H_k)} \varphi(r_{i,h,m}) = 0$$

for any fixed m_0, i, and h. In particular, we also have

$$\begin{aligned} \varphi(x_{i,h}) &= \sum_{m \in J(H_1/H_k)} \varphi(r_{i,h,m}) \chi_k(\varphi(m)) \\ &= \sum_{m_0 \in J(\varphi(H_1)/H_k)} \Bigg(\sum_{m \in J(H_1/H_k) \cap \varphi^{-1}(m_0 H_k)} \varphi(r_{i,h,m}) \Bigg) \chi_k(m_0) = 0, \end{aligned}$$

showing that the sum in (8) is direct.

Step 3. The result follows by taking the projective limit of (8) using (7). □

Remark 2.8. This lemma also holds if we interchange left and right, that is,

$$R[\![H_1, \ell]\!] = \bigoplus_{h \in J(H_0/\varphi(H_0))} h\varphi(R[\![H_1, \ell]\!]).$$

Let S be a (not necessarily commutative) ring (with identity) with the following properties:

(i) There exists a group homomorphism $\chi\colon H_0 \hookrightarrow S^\times$.

(ii) The ring S admits an étale action of the p-Frobenius φ that is compatible with χ. More precisely, there is an injective ring homomorphism $\varphi\colon S \hookrightarrow S$ such that $\varphi(\chi(x)) = \chi(\varphi(x))$ and

$$S = \bigoplus_{h \in \varphi(H_0)\backslash H_0} \varphi(S)\chi(h) = \bigoplus_{h \in H_0/\varphi(H_0)} \chi(h)\varphi(S).$$

In particular, S is free of rank $|H_0 : \varphi(H_0)|$ as a left as well as a right module over $\varphi(S)$.

Definition 2.9. We call a ring S with the properties (i) and (ii) a φ-ring over H_0.

Corollary 2.10. *The map $R \mapsto R[\![N_1, \ell]\!]$ is a functor from the category of φ-rings over $\mathbb{Z}_p$ to the category of φ-rings over H_0.*

Remark 2.11. We have $\varphi(I_k) \subseteq I_{k+1}$ for all $k \geq 1$.

Proof. Take $x \in I_k$ and write $x + I_{k+1} \in R[H_1/H_{k+1}, \ell]$, as

$$x + I_{k+1} = \sum_{h \in J(H_1/H_{k+1})} r_h \chi_{k+1}(h).$$

Since $x \in I_k$, we have

$$0 = \sum_{h \in J(H_1/H_{k+1})} r_h \chi_k(h) = \sum_{h_1 \in J(H_1/H_k)} \sum_{h \in J(H_1/H_{k+1}) \cap h_1 H_k} r_h \chi_k(h_1),$$

and hence $\sum_{h \in J(H_1/H_{k+1}) \cap h_1 H_k} r_h = 0$ for any fixed $h_1 \in J(H_1/H_k)$. So we compute

$$\begin{aligned}\varphi(x) + I_{k+1} &= \sum_{h_1 \in J(H_1/H_k)} \sum_{h \in J(H_1/H_{k+1}) \cap h_1 H_k} \varphi(r_h)\varphi(\chi_k(h)) \\ &= \sum_{h_1 \in J(H_1/H_k)} \sum_{h \in J(H_1/H_{k+1}) \cap h_1 H_k} \varphi(r_h)\chi_k(\varphi(h_1)) \\ &= \sum_{h_1 \in J(H_1/H_k)} 0 \chi_k(\varphi(h_1)) = 0,\end{aligned}$$

since $\varphi(H_k) \subseteq H_{k+1}$, whence $\varphi(h_1) = \varphi(h)$ above. □

Recall that Fontaine's ring $\mathbb{O}_{\mathscr{E}} := \varprojlim_n (o[\![T]\!][T^{-1}])/p_K^n$ is defined as the p-adic completion of the ring of formal Laurent series over o. It is a complete discrete valuation ring with maximal ideal $p_K \mathbb{O}_{\mathscr{E}}$, residue field $k(\!(T)\!)$, and field of fractions $\mathscr{E} = \mathbb{O}_{\mathscr{E}}[p_K^{-1}]$. We show that the completed skew group ring $\mathbb{O}_{\mathscr{E}}[\![H_1, \ell]\!]$ is isomorphic to the previously constructed microlocalized ring $\Lambda_\ell(H_0)$ of the Iwasawa algebra $\Lambda(H_0)$ ([Schneider and Venjakob 2010]; see also [Schneider and Vignéras 2011, Section 8; Schneider et al. 2012; Zábrádi 2011]). ($H_0 = N_0$ in the notations of [Schneider and Vignéras 2011; Schneider et al. 2012; Zábrádi 2011].) For the convenience of the reader we recall the definition here. Let $\Lambda(H_0) := o[\![H_0]\!]$ be the Iwasawa algebra of the pro-p group H_0. It is shown in [Coates et al. 2005] that $S := \Lambda(H_0) \setminus (p_K, H_1 - 1)$ is a left and right Ore set in $\Lambda(H_0)$ so that the localization $\Lambda(H_0)_S$ exists. The ring $\Lambda_\ell(H_0)$ is defined as the $(p_K, H_1 - 1)$-adic completion of $\Lambda(H_0)_S$ (the so-called "microlocalization"). Since $\varphi \colon H_0 \to H_0$ is a continuous group homomorphism, it induces a continuous ring homomorphism $\varphi \colon \Lambda(H_0) \to \Lambda(H_0)$ of the Iwasawa algebra. Since $\varphi(S) \subset S$, φ extends to a ring homomorphism $\varphi\colon \Lambda(H_0)_S \to \Lambda(H_0)_S$ and, by continuity, to its completion $\Lambda_\ell(H_0)$ (see Section 8 of [Schneider and Vignéras 2011] for more details).

Remark 2.12. Let R be a φ-ring containing (as a φ-subring) the Iwasawa algebra $o[\![T]\!] \cong \Lambda(\mathbb{Z}_p)$. Then using (1) we compute

$$\begin{aligned} R[H_1/H_k,\ell] &\cong \bigl(R \otimes_{\Lambda(\mathbb{Z}_p)} \Lambda(\mathbb{Z}_p)\bigr)[H_1/H_k,\ell] \cong R \otimes_{\Lambda(\mathbb{Z}_p)} \bigl(\Lambda(\mathbb{Z}_p)[H_1/H_k,\ell]\bigr) \\ &\cong R \otimes_{\Lambda(\mathbb{Z}_p),\iota} \Lambda(H_0/H_k) \\ &\cong \bigl(\varphi^{c_k}(R) \otimes_{\varphi^{c_k}(\Lambda(\mathbb{Z}_p))} \Lambda(\mathbb{Z}_p)\bigr) \otimes_{\Lambda(\mathbb{Z}_p),\iota} \Lambda(H_0/H_k) \\ &\cong \varphi^{c_k}(R) \otimes_{\varphi^{c_k}(\Lambda(\mathbb{Z}_p)),\iota} \Lambda(H_0/H_k), \end{aligned}$$

for any $k \geq 1$.

Lemma 2.13. *We have a φ-equivariant ring-isomorphism $\mathcal{O}_{\mathcal{E}}[\![H_1,\ell]\!] \cong \Lambda_\ell(H_0)$.*

Proof. The ring $\Lambda_\ell(H_0)$ is complete and Hausdorff with respect to the filtration by the ideals generated by $(H_k - 1)$, since these ideals are closed with intersection zero in the pseudocompact ring $\Lambda_\ell(H_0)$ (compare Theorem 4.7 in [Schneider and Venjakob 2010]). So it remains to show that $\Lambda_\ell(H_0/H_k)$ is naturally isomorphic to the skew group ring $\mathcal{O}_{\mathcal{E}}[H_1/H_k,\ell]$. First we show that $\Lambda(H_0/H_k) \cong \Lambda(\mathbb{Z}_p)[H_1/H_k,\ell]$. Both sides are free modules of rank $|H_1/H_k|$ over $\Lambda(\mathbb{Z}_p)$ with generators $h \in H_1/H_k$, so there is an obvious isomorphism between them as $\Lambda(\mathbb{Z}_p)$-modules. Moreover, $\varphi^{c_k}(\Lambda(\mathbb{Z}_p))$ lies in the center of both rings. However, the obvious map above is also multiplicative, since the multiplication on $\Lambda(\mathbb{Z}_p)[H_1/H_k,\ell]$ is uniquely determined by (4), so that (5) is satisfied and $\varphi^{c_k}(\Lambda(\mathbb{Z}_p))$ lies in the center.

Now by Remark 2.12, we have

$$\mathcal{O}_{\mathcal{E}}[H_1/H_k,\ell] \cong \varphi^{c_k}(\mathcal{O}_{\mathcal{E}}) \otimes_{\varphi^{c_k}(\Lambda(\mathbb{Z}_p)),\iota} \Lambda(H_0/H_k)$$

for any $k \geq 1$.

Since $\iota(\varphi^{p^{c_k}}(\Lambda(\mathbb{Z}_p)))$ lies in the center of $\Lambda(H_0/H_k)$, the right side above is the localization of $\Lambda(H_0/H_k)$, inverting the central element $\varphi^{p^{c_k}}(T)$ and taking the p-adic completion afterwards (that is, "microlocalization" at $\varphi^{p^{c_k}}(T)$). However, in a p-adically complete ring, T is invertible if and only if $\varphi^{p^{c_k}}(T)$ is too. Indeed, we have

$$T \mid \varphi^{p^{c_k}}(T) = (T+1)^{p^{c_k}} - 1 = \sum_{i=1}^{p^{c_k}} \binom{p^{c_k}}{i} T^i \in T^{p^{c_k}}\bigl(1 + po[\![T]\!][T^{-1}]\bigr).$$

Hence we obtain

$$\varphi^{c_k}(\mathcal{O}_{\mathcal{E}}) \otimes_{\varphi^{c_k}(\Lambda(\mathbb{Z}_p)),\iota} \Lambda(H_0/H_k) \cong \Lambda_\ell(H_0/H_k),$$

as both sides are the microlocalization of $\Lambda(H_0/H_k)$ at T. □

3. Equivalence of categories

Let S be a φ-ring over any pro-p group H_0 satisfying (A), (B), and (C) (for now it would suffice to assume that S has an injective ring-endomorphism $\varphi\colon S \to S$). We define a φ-module over S to be a free S-module D of finite rank together with a semilinear action of φ such that the map

$$1 \otimes \varphi\colon S \otimes_{S,\varphi} D \to D, \quad r \otimes d \mapsto r\varphi(d), \tag{10}$$

is an isomorphism. For rings S in which p is not invertible (such as $S = \mathcal{O}_{\mathscr{E}}$ and $\mathcal{O}_{\mathscr{E}}^{\dagger}$), this is the definition of an *étale* φ-module. However, for rings in which p is invertible (such as the Robba ring $\mathscr{R}$), this is the usual definition of a φ-module. We use this definition for both $S = R$ and $S = R[\![H_1, \ell]\!]$ — the former being a φ-ring over $\mathbb{Z}_p$ and the latter being a φ-ring over H_0. We denote the category of φ-modules over R (resp. over $R[\![H_1, \ell]\!]$) by $\mathfrak{M}(R, \varphi)$ (resp. by $\mathfrak{M}(R[\![H_1, \ell]\!], \varphi)$). These are clearly additive categories. However, they are not abelian in general, as the kernel and cokernel might not be a free module over R (resp. over $R[\![H_1, \ell]\!]$).

For modules M over $R[\![H_1, \ell]\!]$, saying that (10) (with $D = M$) is an isomorphism is equivalent to saying that each element $m \in M$ is uniquely decomposed as

$$m = \sum_{u \in J(H_0/\varphi^k(H_0))} u\varphi^k(m_{u,k})$$

for $k = 1$, or equivalently, for all $k \geq 1$.

There is an obvious functor in both directions induced by ℓ_R and ι_R that we denote by

$$\mathbb{D} := R \otimes_{R[\![H_1,\ell]\!],\ell} \cdot\colon \mathfrak{M}(R[\![H_1, \ell]\!], \varphi) \to \mathfrak{M}(R, \varphi),$$
$$\mathbb{M} := R[\![H_1, \ell]\!] \otimes_{R,\iota} \cdot\colon \mathfrak{M}(R, \varphi) \to \mathfrak{M}(R[\![H_1, \ell]\!], \varphi).$$

The following is a generalization of Theorem 8.20 in [Schneider et al. 2012]. The proof is also similar, but we include it here for the convenience of the reader.

Proposition 3.1. *The functors $\mathbb{D}$ and $\mathbb{M}$ are quasi-inverse equivalences of categories.*

Proof. We first note that since $\ell \circ \iota = \mathrm{id}_R$, we also have $\mathbb{D} \circ \mathbb{M} \cong \mathrm{id}_{\mathfrak{M}(R,\varphi)}$. So it remains to show that $\mathbb{D}$ is full and faithful.

For the faithfulness of $\mathbb{D}$, let $f\colon M_1 \to M_2$ be a morphism in $\mathfrak{M}(R[\![H_1, \ell]\!], \varphi)$ such that $\mathbb{D}(f) = 0$, which means that $f(M_1) \subseteq I_1 M_2$. Let $m \in M_1$. For any $k \in \mathbb{N}$, we write $m = \sum_{u \in J(H_0/\varphi^k(H_0))} u\varphi^k(m_{u,k})$ and

$$f(m) = \sum_{u \in J(H_0/\varphi^k(H_0))} u\varphi^k f(m_{u,k}) \in \varphi^k(I_1 M_2) \subseteq I_{k+1} M_2,$$

by Remark 2.11. Therefore $f(M_1) \subseteq I_{k+1}M_2$ for any $k \geq 0$, and therefore $f = 0$ since M_2 is a finitely generated free module over $R[\![H_1,\ell]\!]$, and $\bigcap_{k\geq 0} I_{k+1} = 0$ since $R[\![H_1,\ell]\!] \cong \varprojlim R[\![H_1,\ell]\!]/I_k$.

Now we prove that for any object M in $\mathfrak{M}(R[\![H_1,\ell]\!],\varphi)$ we have an isomorphism $\mathbb{M} \circ \mathbb{D}(M) \to M$. We start with an arbitrary finite $R[\![H_1,\ell]\!]$-basis $(\epsilon_i)_{1\leq i\leq d}$ of M (where d is the rank of M). As R-modules we have

$$M = \Big(\bigoplus_{1\leq i\leq d} \iota(R)\epsilon_i\Big) \oplus \Big(\bigoplus_{1\leq i\leq d} I_1\epsilon_i\Big).$$

Clearly, the $R[\![H_1,\ell]\!]$-linear map from M to $\mathbb{M}(\mathbb{D}(M))$ sending ϵ_i to $1 \otimes (1 \otimes \epsilon_i)$ is bijective. It is φ-equivariant if and only if $\bigoplus_{1\leq i\leq d} \iota(R)\epsilon_i$ is φ-stable, which is, of course, not true in general. We always have

$$\varphi(\epsilon_i) = \sum_{1\leq j\leq d} (a_{i,j} + b_{i,j})\epsilon_j, \text{ where } a_{i,j} \in \iota(R),\ b_{i,j} \in I_1.$$

If the $b_{i,j}$ are not all 0, we will find elements $x_{i,j} \in I_1$ such that

$$\eta_i := \epsilon_i + \sum_{1\leq j\leq d} x_{i,j}\epsilon_j$$

satisfies

$$\varphi(\eta_i) = \sum_{1\leq j\leq d} a_{i,j}\eta_j \text{ for } i \in I.$$

The conditions on the matrix $X := (x_{i,j})_{1\leq i,j\leq d}$ are

$$\varphi(\mathrm{id} + X)(A + B) = A(\mathrm{id} + X)$$

for the matrices $A := (a_{i,j})_{1\leq i,j\leq d}$, $B := (b_{i,j})_{1\leq i,j\leq d}$. The coefficients of A belong to the commutative ring $\iota(R)$. The matrix $A + B$ is invertible because the $R[\![H_1,\ell]\!]$-endomorphism f of M defined by

$$f(\epsilon_i) = \varphi(\epsilon_i) \text{ for } 1 \leq i \leq d$$

is an automorphism of M as M lies in $\mathfrak{M}(R[\![H_1,\ell]\!],\varphi)$. Therefore the matrix $A = \ell(A + B)$ is also invertible. We have reduced the proof to solving the equation

$$A^{-1}B + A^{-1}\varphi(X)(A + B) = X$$

in the indeterminate X. We are looking for the solution X in the form of an infinite sum

$$X = A^{-1}B + \cdots \\ + \big(A^{-1}\varphi(A^{-1})\cdots\varphi^{k-1}(A^{-1})\varphi^k(A^{-1}B)\varphi^{k-1}(A+B)\cdots\varphi(A+B)(A+B)\big) \\ + \cdots.$$

The coefficients of $A^{-1}B$ belong to the two-sided ideal I_1 of $R[\![H_1,\ell]\!]$, and the coefficients of the k-th term of the series

$$\big(A^{-1}\varphi(A^{-1})\cdots\varphi^{k-1}(A^{-1})\varphi^{k}(A^{-1}B)\varphi^{k-1}(A+B)\cdots\varphi(A+B)(A+B)\big)$$

belong to $\varphi^{k}(I_1)\subseteq I_{k+1}$. Hence the series converges, since

$$R[\![H_1,\ell]\!]\cong\varprojlim_{k} R[\![H_1,\ell]\!]/I_k.$$

Its limit X is the unique solution of the equation. The coefficients of every term in the series belong to I_1 and I_1 is closed in $R[\![H_1,\ell]\!]$, and hence $x_{i,j}\in I_1$ for $1\le i$, $j\le d$.

We still need to show that the set $(\eta_i)_{1\le i\le d}$ is an $R[\![H_1,\ell]\!]$-basis of M. Similarly to the above equation, we may find a matrix Y with coefficients in I_1 such that

$$(A+B)(\mathrm{id}+Y)=\varphi(\mathrm{id}+Y)A.$$

Therefore we obtain

$$(A+B)(\mathrm{id}+Y)(\mathrm{id}+X)=\varphi((\mathrm{id}+Y)(\mathrm{id}+X))(A+B),$$

which means that the map

$$(\mathrm{id}+Y)(\mathrm{id}+X)\colon M\to M,\quad \epsilon_i\mapsto(\mathrm{id}+Y)(\mathrm{id}+X)\epsilon_i$$

is a φ-equivariant map such that $\mathbb{D}((\mathrm{id}+Y)(\mathrm{id}+X))=\mathrm{id}$, so $(\mathrm{id}+Y)(\mathrm{id}+X)=\mathrm{id}$ by the faithfulness of $\mathbb{D}$. By a similar computation, we also obtain

$$A(\mathrm{id}+X)(\mathrm{id}+Y)=\varphi((\mathrm{id}+X)(\mathrm{id}+Y))A,$$

showing that $(\mathrm{id}+X)(\mathrm{id}+Y)$ is a φ-equivariant endomorphism of $\mathbb{M}\circ\mathbb{D}(M)$ reducing to the identity modulo I_1. Hence $(\mathrm{id}+Y)$ is a two-sided inverse to the map $(\mathrm{id}+X)$, and in particular $(\eta_i)_{1\le i\le d}$ is an $R[\![H_1,\ell]\!]$-basis of M. So we obtain an isomorphism in $\mathfrak{M}(R[\![H_1,\ell]\!],\varphi)$,

$$\Theta : M\to M(\mathbb{D}(M)),\ \Theta(\eta_i)=1\otimes(1\otimes\eta_i)\text{ for }1\le i\le d,$$

such that $\mathbb{D}(\Theta)$ is the identity morphism of $\mathbb{D}(M)$.

Now if $f\colon\mathbb{D}(M_1)\to\mathbb{D}(M_2)$, then for

$$\mathbb{M}(f)\colon M_1\cong\mathbb{M}\circ\mathbb{D}(M_1)\to\mathbb{M}\circ\mathbb{D}(M_2)\cong M_2,$$

we have $\mathbb{D}\circ\mathbb{M}(f)=f$; therefore $\mathbb{D}$ is full. □

Remark 3.2. There is a small mistake in Lemma 1 of [Zábrádi 2011]. The map ω is in fact not a p-valuation, since assertion (iii) states that $\omega(g^p)=\omega(g)+1$ is false. It is only true in the weaker form $\omega(g^p)\ge\omega(g)+1$. However, this does not influence the validity of the rest of the paper, as $N_{0,n}:=\{g\in N_0\mid\omega(g)\ge n\}$ is still a subgroup satisfying Lemma 2. Alternatively, it is possible to modify ω so that one truly obtains a p-valuation. I would like to take this opportunity to thank Torsten Schoeneberg for pointing this out to me.

Remark 3.3. In the case of $R = \mathcal{O}_{\mathscr{E}}$ we may end the proof of Proposition 3.1 by saying that $\mathrm{id} + X$ is invertible since X lies in $I_1^{d\times d}$ and $\mathcal{O}_{\mathscr{E}}[\![H_1,\ell]\!] \cong \Lambda_\ell(H_0)$ is I_1-adically complete. However, in the general situation $R[\![H_1,\ell]\!]$ may not be complete I_1-adically. The reason for this is that the ideals $(I_k)_{k\geq 1}$ are only cofinal with the ideals I_1^k whenever R is killed by a power of p. Therefore if R is not p-adically complete, we do not have $R[\![H_1,\ell]\!] \cong \varprojlim R[\![H_1,\ell]\!]/I_1^k$ in general. In the case of $R = \mathcal{O}_{\mathscr{E}}$, Proposition 3.1 holds for not necessarily free modules as well. See [Schneider et al. 2012] for the proof of this.

Remark 3.4. The matrix Y in the proof of Proposition 3.1 is given by a convergent sum of the terms

$$-(A{+}B)^{-1}\varphi((A{+}B)^{-1})\cdots\varphi^{k-1}((A{+}B)^{-1})\varphi^k((A{+}B)^{-1}B)\varphi^{k-1}(A)\cdots\varphi(A)A$$

for $k \geq 0$, and a direct computation also shows that $(\mathrm{id} + Y)(\mathrm{id} + X) = \mathrm{id} = (\mathrm{id} + X)(\mathrm{id} + Y)$.

Reductive groups over $\mathbb{Q}_p$ and Whittaker functionals. Let p be a prime number and let $\mathbb{Q}_p \subseteq K$ be a finite extension with ring of integer o_K, uniformizer p_K, and residue field $k = o_K/p_K$. This field will only play the role of coefficients; the reductive groups will all be defined over $\mathbb{Q}_p$. Following [Schneider and Vignéras 2011], let G be the $\mathbb{Q}_p$-rational points of a $\mathbb{Q}_p$-split connected reductive group over $\mathbb{Q}_p$. In particular, G is a locally $\mathbb{Q}_p$-analytic group. We also assume that the center of G is connected. We fix a Borel subgroup $P = TN$ in G with maximal split torus T and unipotent radical N. Let Φ^+ denote, as usual, the set of positive roots of T with respect to P and let $\Delta \subseteq \Phi^+$ be the subset of simple roots. For any $\alpha \in \Phi^+$ we have the root subgroup $N_\alpha \subseteq N$. We recall that $N = \prod_{\alpha\in\Phi^+} N_\alpha$ (set-theoretically) for any total ordering of Φ^+. Let $T_0 \subseteq T$ be the maximal compact subgroup. We fix a compact open subgroup $N_0 \subseteq N$ that is totally decomposed; in other words, $N_0 = \prod_\alpha (N_0 \cap N_\alpha)$ for any total ordering of Φ^+. Hence $P_0 := T_0 N_0$ is a group. We introduce the submonoid $T_+ \subseteq T$ of all $t \in T$ such that $tN_0t^{-1} \subseteq N_0$, or equivalently, such that $|\alpha(t)| \leq 1$ for any $\alpha \in \Delta$. Obviously, $P_+ := N_0 T_+ = P_0 T_+ P_0$ is then a submonoid of P.

We fix once and for all isomorphisms of algebraic groups

$$\iota_\alpha \colon N_\alpha \overset{\cong}{\to} \mathbb{Q}_p$$

for $\alpha \in \Delta$, such that

$$\iota_\alpha(tnt^{-1}) = \alpha(t)\iota_\alpha(n)$$

for any $n \in N_\alpha$ and $t \in T$. We normalize these isomorphisms so that $\iota_\alpha(N_0 \cap N_\alpha) = \mathbb{Z}_p \subset \mathbb{Q}_p$. Since $\prod_{\alpha\in\Delta} N_\alpha$ is naturally a quotient of $N/[N, N]$ we may view any homomorphism

$$\ell\colon \prod_{\alpha\in\Delta} N_\alpha \to \mathbb{Q}_p$$

as a functional on N. We fix once and for all a homomorphism ℓ such that $\ell(N_0) = \mathbb{Z}_p$. Let $X^*(T) := \mathrm{Hom}_{\mathrm{alg}}(T, \mathbb{G}_m)$ (resp. $X_*(T) := \mathrm{Hom}_{\mathrm{alg}}(\mathbb{G}_m, T)$) be the group of algebraic characters (resp. cocharacters) of T. Since we assume that the center of G is connected, the quotient $X^*(T)/\bigoplus_{\alpha\in\Delta} \mathbb{Z}\alpha$ is free. Hence we find a cocharacter ξ in $X_*(T)$ such that $\alpha \circ \xi = \mathrm{id}_{\mathbb{G}_m}$ for any α in Δ. It is injective and uniquely determined up to a central cocharacter. We fix such a ξ. It satisfies

$$\xi(\mathbb{Z}_p \setminus \{0\}) \subseteq T_+$$

and

$$\ell(\xi(a)n\xi(a^{-1})) = a\ell(n) \tag{11}$$

for any a in $\mathbb{Q}_p^\times$ and n in N, since ℓ is a linear functional on the space $\prod_{\alpha\in\Delta} N_\alpha$ and therefore can be written as a linear combination of the isomorphisms $\iota_\alpha\colon N_\alpha \to \mathbb{Q}_p$.

For example, if $G = \mathrm{GL}_n(\mathbb{Q}_p)$, T is the group of diagonal matrices, and N is the group of unipotent upper triangular matrices, then we could choose

$$\xi\colon \mathbb{G}_m(\mathbb{Q}_p) = \mathbb{Q}_p^\times \to T = (\mathbb{Q}_p^\times)^n, \quad \xi(x) := \begin{pmatrix} x^{n-1} & & & \\ & x^{n-2} & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}.$$

Put $\Gamma := \xi(\mathbb{Z}_p^\times)$ and $s := \xi(p)$. The element s acts by conjugation on the group N_0 such that $\bigcap_k s^k N_0 s^{-k} = \{1\}$. We denote this action by $\varphi := \varphi_s$. This is compatible with the functional ℓ in the sense $\ell \circ \varphi = p\ell$ (see Section 2) by (11). Therefore we may apply the theory of the preceding sections to any φ-ring R with the homomorphism $\ell\colon N_0 \to \mathbb{Z}_p$ and $N_1 := \mathrm{Ker}(\ell_{|N_0})$. We are going to apply the theory of Section 2 in the setting $H_0 := N_0$ and $H_1 := N_1$.

In [Schneider and Vignéras 2011] and [Zábrádi 2011] ℓ is assumed to be generic—we do not assume this here, though. For any $\alpha \in \Delta$ the restriction of ℓ to a fixed N_α is either zero or an isomorphism of N_α with $\mathbb{Q}_p$, and we put $a_\alpha := \ell(\iota_\alpha^{-1}(1))$. By the assumption $\ell(N_0) = \mathbb{Z}_p$ we obtain $a_\alpha \in \mathbb{Z}_p$ for all $\alpha \in \Delta$, and $a_\alpha \in \mathbb{Z}_p^\times$ for at least one α in Δ. We put $T_{+,\ell} := \{t \in T_+ \mid tN_1t^{-1} \subseteq N_1\}$. The monoid $T_{+,\ell}$ acts on the group $\mathbb{Z}_p$ via $\ell\colon N_0 \to \mathbb{Z}_p$, too.

A (φ, Γ)-ring R is by definition a φ-ring (in the sense of Section 2) together with an action of $\Gamma \cong \mathbb{Z}_p^\times$ commuting with φ and satisfying $\gamma(\chi(x)) = \chi(\xi^{-1}(\gamma)x)$. For example, $\mathbb{O}_{\mathscr{E}}, \mathbb{O}_{\mathscr{E}}^\dagger, \mathscr{E}^\dagger, \mathscr{R}$ are (φ, Γ)-rings. The endomorphism ring $\mathrm{End}(\mathbb{Z}_p)$ of the

p-adic integers (as a topological abelian group) is isomorphic to $\mathbb{Z}_p$. On the other hand, the multiplicative monoid $\mathbb{Z}_p \setminus \{0\}$ is isomorphic to $\varphi^{\mathbb{N}}\Gamma$. Now having an action of φ and Γ on R we obtain an action of $T_{+,\ell}$ on R since the map $\ell\colon N_0 \to \mathbb{Z}_p$ induces a monoid homomorphism $T_{+,\ell} \to \mathbb{Z}_p \setminus \{0\} \cong \varphi^{\mathbb{N}}\Gamma$. We denote the kernel of this monoid homomorphism by $T_{0,\ell}$. Similarly, we have a natural action of $T_{+,\ell}$ on the ring $R[\![N_1,\ell]\!]$ by conjugation. Indeed, if $t \in T_{+,\ell}$, then since T is commutative, we have

$$t\varphi^k(N_1)t^{-1} = ts^kN_1s^{-k}t^{-1} = s^ktN_1t^{-1}s^{-k} = \varphi^k(tN_1t^{-1}) \subseteq \varphi^k(N_1),$$

whence $tN_kt^{-1} \subseteq N_k$. Hence t acts naturally on the skew group ring $R[N_1/N_k,\ell]$, and by taking the limit, we also obtain an action on $R[\![N_1,\ell]\!]$. We denote the map on both R and $R[\![N_1,\ell]\!]$ induced by the action of $t \in T_{+,\ell}$ by φ_t.

Now a $T_{+,\ell}$-module over R (resp. over $R[\![N_1,\ell]\!]$) is a finitely generated free R-module D (resp. $R[\![N_1,\ell]\!]$-module M) with a semilinear action of $T_{+,\ell}$ (denoted by $\varphi_t\colon D \to D$, resp. $\varphi_t\colon M \to M$ for any $t \in T_{+,\ell}$) such that the restriction of the $T_{+,\ell}$-action to $s \in T_{+,\ell}$ defines a φ-module over R (resp. over $R[\![N_1,\ell]\!]$). We denote the category of $T_{+,\ell}$-modules over R (resp. over $R[\![N_1,\ell]\!]$) by $\mathfrak{M}(R,T_{+,\ell})$ (resp. by $\mathfrak{M}(R[\![N_1,\ell]\!],T_{+,\ell})$).

Lemma 3.5. *Let M be in $\mathfrak{M}(R[\![N_1,\ell]\!],T_{+,\ell})$ and D be in $\mathfrak{M}(R,T_{+,\ell})$. Then the maps*

$$1\otimes\varphi_t\colon R[\![N_1,\ell]\!]\otimes_{R[\![N_1,\ell]\!],\varphi_t} M \to M, \quad r\otimes m \mapsto r\varphi_t(m)$$

and

$$1\otimes\varphi_t\colon R\otimes_{R,\varphi_t} D \to D, \quad r\otimes d \mapsto r\varphi_t(d)$$

are isomorphisms for any $t \in T_{+,\ell}$.

Proof. We only prove the statement for M (the statement for D is entirely analogous). First note that the subgroups $s^kN_0s^{-k}$ (resp. $s^kN_1s^{-k}$) form a system of neighborhoods of 1 in N (resp. in $\mathrm{Ker}(\ell)$). On the other hand, if t is in $T_{+,\ell}$, then

$$t\,\mathrm{Ker}(\ell_{|N})t^{-1} = t\left(\bigcup_{k\in\mathbb{Z}} s^kN_1s^{-k}\right)t^{-1} = \bigcup_{k\in\mathbb{Z}} s^ktN_1t^{-1}s^{-k} = \mathrm{Ker}(\ell_{|N}),$$

since tN_1t^{-1} has finite index in N_1. Now since $t^{-1}N_0t$ and $t^{-1}N_1t$ are compact, we find $k_0 > 0$, so that $t^{-1}N_0t \subseteq s^{-k_0}N_0s^{k_0}$ and $t^{-1}N_1t \subseteq s^{-k_0}N_1s^{k_0}$, whence $s^{k_0}t^{-1}$ lies in $T_{+,\ell}$. Since M is a φ-module over $R[\![N_1,\ell]\!]$, the map

$$1\otimes\varphi_{s^{k_0}}\colon R[\![N_1,\ell]\!]\otimes_{R[\![N_1,\ell]\!],\varphi_{s^{k_0}}} M \to M, \quad r\otimes m \mapsto r\varphi_{s^{k_0}}(m)$$

is an isomorphism. Under the identifications

$$\begin{aligned}R[\![N_1,\ell]\!]\otimes_{R[\![N_1,\ell]\!],\varphi_t}\bigl(R[\![N_1,\ell]\!]\otimes_{R[\![N_1,\ell]\!],\varphi_{s^{k_0}t^{-1}}}M\bigr)\\ \cong R[\![N_1,\ell]\!]\otimes_{R[\![N_1,\ell]\!],\varphi_{s^{k_0}}}M\\ \cong R[\![N_1,\ell]\!]\otimes_{R[\![N_1,\ell]\!],\varphi_{s^{k_0}t^{-1}}}\bigl(R[\![N_1,\ell]\!]\otimes_{R[\![N_1,\ell]\!],\varphi_t}M\bigr)\end{aligned}$$

we have

$$(1\otimes\varphi_t)\circ(1\otimes(1\otimes\varphi_{s^{k_0}t^{-1}}))=1\otimes\varphi_{s^{k_0}}=(1\otimes\varphi_{s^{k_0}t^{-1}})\circ(1\otimes(1\otimes\varphi_t)),$$

so $1\otimes\varphi_t$ is surjective by the equality on the left and injective by the equality on the right. □

Remark 3.6. The action of $T_{0,\ell}$ on a $T_{+,\ell}$-module D over R is linear, since $T_{0,\ell}$ acts trivially on R. Therefore this action extends (uniquely) to the subgroup $T_\ell\le T$ generated by the monoid $T_{0,\ell}$.

Proof. By the étaleness of the action of φ_t for $t\in T_{0,\ell}$ we see immediately that φ_t is an automorphism of D since $\varphi_t\colon R\to R$ is the identity map. Therefore φ_t has a (left and right) inverse (as a linear transformation of the R-module D), which we denote by $\varphi_{t^{-1}}$. The remark follows if we note that T_ℓ consists of the quotients of elements of $T_{0,\ell}$. □

In the case when $\ell=\ell_\alpha$ given by the projection of $\prod_{\beta\in\Delta}N_\beta$ to N_α for some fixed simple root $\alpha\in\Delta$, it is clear that $T_{+,\ell}=T_+$, as N_β is T_+-invariant for each $\beta\in\Phi^+$ and $\mathrm{Ker}(\ell)=\prod_{\alpha\ne\beta\in\Phi^+}N_\beta$. Therefore $T_\ell\cong(\mathbb{Q}_p^\times)^{n-1}$, where $n=\dim T$. This is the case in which a G-equivariant sheaf on G/P is constructed in [Schneider et al. 2012] associated to any object D in $\mathfrak{M}(\mathcal{O}_{\mathcal{E}},T_{+,\ell})$. So an object in $\mathfrak{M}(\mathcal{O}_{\mathcal{E}},T_{+,\ell})$ is nothing other than a (φ,Γ)-module over $\mathcal{O}_{\mathcal{E}}$ with an additional linear action of the group T_ℓ (once we fixed the cocharacter ξ). In case of $G=\mathrm{GL}_2(\mathbb{Q}_p)$ this additional action is just an action of the center $Z=T_\ell$ of G. In the work of Colmez [2010c; 2010b] on the p-adic Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$, the action of Z on an irreducible 2-dimensional étale (φ,Γ)-module D is given by the determinant (that is, the action of $\mathbb{Q}_p^\times\cong Z$ on $\bigwedge^2 D$). It is unclear at this point whether the action of T_ℓ can be chosen canonically (in a similar fashion) for a given n-dimensional irreducible étale (φ,Γ)-module D.

As a corollary of Proposition 3.1, we obtain:

Proposition 3.7. *The functors $\mathbb{D}=R\otimes_{R[\![N_1,\ell]\!],\ell}\cdot$ and $\mathbb{M}=R[\![N_1,\ell]\!]\otimes_{R,\iota}\cdot$ are quasi-inverse equivalences of categories between $\mathfrak{M}(R[\![N_1,\ell]\!],T_{+,\ell})$ and $\mathfrak{M}(R,T_{+,\ell})$.*

Proof. Since we clearly have $\mathbb{D}\circ\mathbb{M}\cong\mathrm{id}_{\mathfrak{M}(R,T_{+,\ell})}$ and the faithfulness of $\mathbb{D}$ is a formal consequence of Proposition 3.1, it suffices to show that the isomorphism

$\Theta\colon M \to \mathbb{M} \circ \mathbb{D}(M)$ is $T_{+,\ell}$-equivariant whenever M lies in $\mathfrak{M}(R[\![N_1, \ell]\!], T_{+,\ell})$. Let $t \in T_{+,\ell}$ be arbitrary and for an $m \in M$ write $m = \sum_{u \in J(N_0/\varphi_s^k(N_0))} u\varphi_s^k(m_{u,k})$. Since $\mathbb{D}(\Theta) = \mathrm{id}_{\mathbb{D}(M)}$, we have $(\Theta \circ \varphi_t - \varphi_t \circ \Theta)(M) \subseteq I_1 \mathbb{M} \circ \mathbb{D}(M)$. We compute

$$\begin{aligned}(\Theta \circ \varphi_t - \varphi_t \circ \Theta)(m) &= \sum_{u \in J(N_0/\varphi_s^k(N_0))} \varphi_t(u)\varphi_s^k \circ (\Theta \circ \varphi_t - \varphi_t \circ \Theta)(m_{u,k}) \\ &\subseteq \varphi_s^k(I_1 \mathbb{M} \circ \mathbb{D}(M)) \subseteq I_{k+1} \mathbb{M} \circ \mathbb{D}(M)\end{aligned}$$

for all $k \geq 0$, showing that Θ is φ_t-equivariant. □

4. The case of overconvergent and Robba rings

The locally analytic distribution algebra. Let p be a prime and put $\epsilon_p = 1$ if p is odd and $\epsilon_p = 2$ if $p = 2$. If H is a compact locally $\mathbb{Q}_p$-analytic group, then we denote by $D(H, K)$ the algebra of K-valued locally analytic distributions on H. Recall that $D(H, K)$ is equal to the strong dual of the locally convex vector space $C^{\mathrm{an}}(H, K)$ of K-valued locally $\mathbb{Q}_p$-analytic functions on H with the convolution product.

Recall that a topologically finitely generated pro-p group H is uniform if it is powerful (that is, $H/\overline{H^{p^{\epsilon_p}}}$ is abelian) and $|P_i(H) : P_{i+1}(H)| = |H : P_2(H)|$ for all $i \geq 1$, where $P_1(H) = H$ and $P_{i+1}(H) = \overline{P_i(H)^p[P_i(H), H]}$ (see [Dixon et al. 1999] for more details). Now if H is uniform, it has a bijective global chart

$$\mathbb{Z}_p^d \to H, \quad (x_1, \dots, x_d) \mapsto h_1^{x_1} \cdots h_d^{x_d},$$

where $h_1, \dots, h_d$ is a fixed (ordered) minimal set of topological generators of H. Putting $b_i := h_i - 1 \in \mathbb{Z}[G]$, $\boldsymbol{b^k} := b_1^{k_1} \cdots b_d^{k_d}$ for $\boldsymbol{k} = (k_i) \in \mathbb{N}^d$, we can identify $D(H, K)$ with the ring of all formal series

$$\lambda = \sum_{\boldsymbol{k} \in \mathbb{N}^d} d_{\boldsymbol{k}} \boldsymbol{b^k},$$

with $d_{\boldsymbol{k}}$ in K such that the set $\{|d_{\boldsymbol{k}}|\rho^{\epsilon_p|\boldsymbol{k}|}\}_{\boldsymbol{k}}$ is bounded for all $0 < \rho < 1$. Here the first $|\cdot|$ is the normalized absolute value on K and the second one denotes the degree of $\boldsymbol{k}$, that is, $\sum_i k_i$. For any ρ in $p^{\mathbb{Q}}$ with $p^{-1} < \rho < 1$, we have a multiplicative norm $\|\cdot\|_\rho$ on $D(H, K)$ [Schneider and Teitelbaum 2003] given by

$$\|\lambda\|_\rho := \sup_{\boldsymbol{k}} |d_{\boldsymbol{k}}|\rho^{\epsilon_p|\boldsymbol{k}|}.$$

The family of norms $\|\cdot\|_\rho$ defines the Fréchet topology on $D(H, K)$. The completion with respect to the norm $\|\cdot\|_\rho$ is denoted by $D_{[0,\rho]}(H, K)$.

Microlocalization. Let G be the group of $\mathbb{Q}_p$-points of a $\mathbb{Q}_p$-split connected reductive group with a fixed Borel subgroup $P = TN$. We also choose a simple root α for the Borel subgroup P and let $\ell = \ell_\alpha$ be the functional given by the projection

$$\ell_\alpha \colon N \to N/[N,N] \to \prod_{\beta \in \Delta} N_\beta \to N_\alpha \xrightarrow{\iota_\alpha} \mathbb{Q}_p \, .$$

Therefore we have $T_{+,\ell} = T_+$, as N_β is T_+-invariant for each $\beta \in \Phi^+$. We assume further that N_0 is *uniform.*

Let us begin by recalling the definition of the classical Robba ring for the group $\mathbb{Z}_p$. The distribution algebra $D(\mathbb{Z}_p, K)$ of $\mathbb{Z}_p$ can clearly be identified with the ring of power series (in variable T) with coefficients in K that are convergent in the p-adic open unit disc. Now put

$$\mathscr{A}_{[\rho,1)} := \text{the ring of all Laurent series } \sum_{n\in\mathbb{Z}} a_n T^n \text{ that converge for } \rho \le |T| < 1.$$

For $\rho \le \rho'$ we have a natural inclusion $\mathscr{A}_{[\rho,1)} \hookrightarrow \mathscr{A}_{[\rho',1)}$, so we can form the inductive limit

$$\mathscr{R} := \varinjlim_{\rho \to 1} \mathscr{A}_{[\rho,1)} \tag{12}$$

defining the Robba ring. $\mathscr{R}$ is a (φ, Γ)-ring over $\mathbb{Z}_p$ with the maps $\chi \colon \mathbb{Z}_p \to \mathscr{R}^\times$ and $\varphi \colon \mathscr{R} \to \mathscr{R}$ such that $\chi(1) = 1+T$, $\varphi(T) = (T+1)^p - 1$, and $\gamma(T) = (1+T)^{\xi^{-1}(\gamma)} - 1$ for $\gamma \in \Gamma$.

Recall that the ring

$$\mathcal{O}_{\mathscr{E}}^\dagger := \left\{ \sum_{n\in\mathbb{Z}} a_n T^n \;\middle|\; a_n \in o_K \text{ and there exists a } \rho < 1 \text{ such that } \lim_{n \to -\infty} |a_n| \rho^n = 0 \right\}$$

is called the ring of overconvergent power series. It is a subring of both $\mathcal{O}_{\mathscr{E}}$ and $\mathscr{R}$. We put $\mathscr{E}^\dagger := K \otimes_{o_K} \mathcal{O}_{\mathscr{E}}^\dagger$, which is also a subring of the Robba ring. These rings are also (φ, Γ)-rings.

The rings $\mathcal{O}_{\mathscr{E}}^\dagger[\![N_1, \ell]\!]$ and $\mathscr{R}[\![N_1, \ell]\!]$ constructed in the previous sections are only overconvergent or Robba in the variable b_α for the fixed simple root α. In all the other variables b_β, they behave like the Iwasawa algebra $\Lambda(N_1)$, since we took the completion with respect to the ideals generated by $(N_k - 1)$. Moreover, in the projective limit $\mathcal{O}_{\mathscr{E}}^\dagger[\![N_1, \ell]\!] \cong \varprojlim_k \mathcal{O}_{\mathscr{E}}^\dagger[N_1/N_k, \ell]$, the terms are not forced to share a common region of convergence. In this section we construct the rings $\mathscr{R}^{\mathrm{int}}(N_1, \ell)$ and $\mathscr{R}(N_1, \ell)$ with better analytic properties.

We start by constructing a ring $\mathfrak{R}_0 = \mathfrak{R}_0(N_0, K, \alpha)$ as a certain microlocalization of the distribution algebra $D(N_0, K)$. We fix the topological generator n_α of $N_0 \cap N_\alpha$ such that $\ell_\alpha(n_\alpha) = 1$. This is possible since we normalized $\iota_\alpha \colon N_\alpha \xrightarrow{\sim} \mathbb{Q}_p$ so that $\iota_\alpha(N_0 \cap N_\alpha) = \mathbb{Z}_p$. We also fix topological generators n_β of $N_0 \cap N_\beta$ for each

$\alpha \neq \beta \in \Phi^+$. Since N_0 is uniform of dimension $|\Phi^+|$, the set $A := \{n_\beta \mid \beta \in \Phi^+\}$ is a minimal set of topological generators of the group N_0. Moreover, $A \setminus \{n_\alpha\}$ is a minimal set of generators of the group $N_1 = \mathrm{Ker}(\ell) \cap N_0$. Further, we put $b_\beta := n_\beta - 1$. For any real number $p^{-1} < \rho < 1$ in $p^{\mathbb{Q}}$, the formula $\|b_\beta\|_\rho := \rho$ (for all $\beta \in \Phi^+$) defines a multiplicative norm on $D(N_0, K)$. The completion of $D(N_0, K)$ with respect to this norm is a Banach algebra that we denote by $D_{[0,\rho]}(N_0, K)$. Let now $p^{-1} < \rho_1 < \rho_2 < 1$ be real numbers in $p^{\mathbb{Q}}$. We take the generalized microlocalization (see the Appendix of [Zábrádi 2012]) of the Banach algebra $D_{[0,\rho_2]}(N_0, K)$ at the multiplicatively closed set $\{(n_\alpha - 1)^i\}_{i \geq 1}$ with respect to the pair of norms (ρ_1, ρ_2). This provides us with the Banach algebra $D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$. Recall that the elements of this Banach algebra are equivalence classes of Cauchy sequences $((n_\alpha - 1)^{-k_n} x_n)_n$ (with $x_n \in D_{[0,\rho_2]}(N_0, K)$) with respect to the norm $\|\cdot\|_{\rho_1,\rho_2} := \max(\|\cdot\|_{\rho_1}, \|\cdot\|_{\rho_2})$.

Letting ρ_2 tend to 1, we define

$$D_{[\rho_1,1)}(N_0, K, \alpha) := \varprojlim_{\rho_2 \to 1} D_{[\rho_1,\rho_2]}(N_0, K, \alpha).$$

This is a Fréchet–Stein algebra (the proof is completely analogous to that of Theorem 5.5 in [Zábrádi 2012], but it is not a formal consequence of that). However, we will not need this fact in the sequel, so we omit the proof. Now the partial Robba ring $\mathfrak{R}_0 := \mathfrak{R}_0(N_0, K, \alpha) := \varinjlim_{\rho_1 \to 1} D_{[\rho_1,1)}(N_0, K, \alpha)$ is defined as the injective limit of these Fréchet–Stein algebras. We equip $\mathfrak{R}_0$ with the inductive limit topology of the Fréchet topologies of $D_{[\rho_1,1)}(N_0, K, \alpha)$. By the following parametrization, the partial Robba ring can be thought of as a skew Laurent series ring on the variables b_β ($\beta \in \Phi^+$) with certain convergence conditions such that only the variable b_α is invertible. In [Zábrádi 2012], a "full" Robba ring is constructed such that all the variables b_β are invertible. We denote the corresponding "fully" microlocalized Banach algebras by $D_{[\rho_1,\rho_2]}(N_0, K)$. In all these rings, we will often omit K from the notation if it is clear from the context.

Remark 4.1. The microlocalization of quasiabelian normed algebras (Appendix of [Zábrádi 2012]) is somewhat different from the microlocalization constructing $\Lambda_\ell(N_0)$, where first a localization (with respect to an Ore set) is constructed and then the completion is taken. The set we are inverting here does not satisfy the Ore property, so the localization in the usual sense does not exist. However, we may complete and localize at the same time in order to obtain a microlocalized ring directly.

In order to be able to work with these rings we will show that their elements can be viewed as Laurent series. The discussion below is completely analogous to the discussion before Proposition A.24 in [Zábrádi 2012]. However, for the convenience of the reader, we explain the method specialized to our case here. We introduce the affinoid domain

$$A_\alpha[\rho_1,\rho_2]:=\{(z_\beta)_{\beta\in\Phi^+}\in\mathbb{C}_p^{\Phi^+}\mid\rho_1\le|z_\alpha|\le\rho_2,\,0\le|z_\beta/z_\alpha|\le 1\text{ for }\alpha\ne\beta\in\Phi^+\}.$$

This has the affinoid subdomain

$$X^{\Phi^+}_{[\rho_1,\rho_2]}:=\{(z_\beta)_{\beta\in\Phi^+}\in\mathbb{C}_p^{\Phi^+}\mid\rho_1\le|z_{\beta_1}|=\dots=|z_{\beta_{|\Phi^+|}}|\le\rho_2\}$$

(where $\{\beta_1,\dots,\beta_{|\Phi^+|}\}=\Phi^+$) as defined in [Zábrádi 2012, Proposition A.24].

Lemma 4.2. *The ring $\mathbb{O}_K(A_\alpha[\rho_1,\rho_2])$ of K-analytic functions on $A_\alpha[\rho_1,\rho_2]$ is the ring of all Laurent series*

$$f(\boldsymbol{Z})=\sum_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}}d_{\boldsymbol{k}}\boldsymbol{Z}^{\boldsymbol{k}}$$

with $d_{\boldsymbol{k}}\in K$ and such that $\lim_{\boldsymbol{k}\to\infty}|d_{\boldsymbol{k}}|\rho^{\boldsymbol{k}}=0$ for any $\rho_1\le\rho\le\rho_2$. (Here $\boldsymbol{Z}^{\boldsymbol{k}}:=\prod_{\beta\in\Phi^+}Z_\beta^{k_\beta}$ and $\rho^{\boldsymbol{k}}:=\rho^{\sum_{\beta\in\Phi^+}k_\beta}$, and $\boldsymbol{k}\to\infty$ means that $\sum_{\beta\in\Phi^+}|k_\beta|\to\infty$.) *This is the subring of $\mathbb{O}_K(X^{\Phi^+}_{[\rho_1,\rho_2]})$ consisting of elements in which the variables Z_β appear only with nonnegative exponent for all $\alpha\ne\beta\in\Phi^+$.*

Proof. Since $X^{\Phi^+}_{[\rho_1,\rho_2]}\subseteq A_\alpha[\rho_1,\rho_2]$, we clearly have $\mathbb{O}_K(A_\alpha[\rho_1,\rho_2])\subseteq\mathbb{O}_K(X^{\Phi^+}_{[\rho_1,\rho_2]})$. Also, the power series in $\mathbb{O}_K(A_\alpha[\rho_1,\rho_2])$ converge for $z_\beta=0$ ($\beta\ne\alpha$), and hence these variables appear with nonnegative exponent. On the other hand, if we have a power series $f(\boldsymbol{Z})\in\mathbb{O}_K(X^{\Phi^+}_{[\rho_1,\rho_2]})$ such that the variables Z_β have nonnegative exponent for all $\alpha\ne\beta\in\Phi^+$, then it also converges in the region $A_\alpha[\rho_1,\rho_2]$, as we have in this case the trivial estimate

$$\Bigl|\prod_{\beta\in\Phi^+}z_\beta^{k_\beta}\Bigr|\le|z_\alpha|^{\sum_{\beta\in\Phi^+}k_\beta}.$$ □

Since $\rho^{\boldsymbol{k}}\le\max(\rho_1^{\boldsymbol{k}},\rho_2^{\boldsymbol{k}})$ for any $\rho_1\le\rho\le\rho_2$ and any $\boldsymbol{k}\in\mathbb{Z}^{(\alpha)}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}$, the convergence condition on f is equivalent to

$$\lim_{\boldsymbol{k}\to\infty}|d_{\boldsymbol{k}}|\rho_1^{\boldsymbol{k}}=\lim_{\boldsymbol{k}\to\infty}|d_{\boldsymbol{k}}|\rho_2^{\boldsymbol{k}}=0.$$

The spectral norm on the affinoid algebra $\mathbb{O}_K(A_\alpha[\rho_1,\rho_2])$ (for the definition of these notions see [Fresnel and van der Put 2004]) is given by

$$\begin{aligned}\|f\|_{A_\alpha[\rho_1,\rho_2]}&=\sup_{\rho_1\le\rho\le\rho_2}\max_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}}|d_{\boldsymbol{k}}|\rho^{\boldsymbol{k}}\\&=\max\Bigl(\max_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}}|d_{\boldsymbol{k}}|\rho_1^{\boldsymbol{k}},\max_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}}|d_{\boldsymbol{k}}|\rho_2^{\boldsymbol{k}}\Bigr).\end{aligned}$$

Setting $\boldsymbol{b}^{\boldsymbol{k}}:=\prod_{\beta\in\Phi^+}b_\beta^{k_\beta}$ for some fixed ordering of Φ^+ and for any $\boldsymbol{k}=(k_\beta)_{\beta\in\Phi^+}$

in $\mathbb{Z}^{\{\alpha\}} \times \mathbb{N}^{\Phi^+ \setminus \{\alpha\}}$, we claim that

$$f(\boldsymbol{b}) := \sum_{\boldsymbol{k} \in \mathbb{Z}^{\{\alpha\}} \times \mathbb{N}^{\Phi^+ \setminus \{\alpha\}}} d_{\boldsymbol{k}} \boldsymbol{b}^{\boldsymbol{k}}$$

converges in $D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$ for $f \in \mathbb{O}_K(A_\alpha[\rho_1, \rho_2])$. As a consequence of Proposition A.21 and Lemma A.7.iii in [Zábrádi 2012], we have

$$\|\boldsymbol{b}^{\boldsymbol{k}}\|_{\rho_1,\rho_2} = \max(\rho_1^{\boldsymbol{k}}, \rho_2^{\boldsymbol{k}})$$

for any $\boldsymbol{k} \in \mathbb{Z}^{\{\alpha\}} \times \mathbb{N}^{\Phi^+ \setminus \{\alpha\}}$. Hence

$$\begin{aligned} \lim_{\boldsymbol{k} \to \infty} \|d_{\boldsymbol{k}} \boldsymbol{b}^{\boldsymbol{k}}\|_{\rho_1,\rho_2} &= \lim_{\boldsymbol{k} \to \infty} \max\bigl(|d_{\boldsymbol{k}}|\rho_1^{\boldsymbol{k}}, |d_{\boldsymbol{k}}|\rho_2^{\boldsymbol{k}}\bigr) \\ &= \max\bigl(\lim_{\boldsymbol{k} \to \infty} |d_{\boldsymbol{k}}|\rho_1^{\boldsymbol{k}}, \lim_{\boldsymbol{k} \to \infty} |d_{\boldsymbol{k}}|\rho_2^{\boldsymbol{k}}\bigr) \\ &= 0. \end{aligned}$$

Therefore

$$\mathbb{O}_K(A_\alpha[\rho_1, \rho_2]) \to D_{[\rho_1,\rho_2]}(N_0, K, \alpha), \quad f \mapsto f(\boldsymbol{b}),$$

is a well defined K-linear map. In order to investigate this map we introduce the filtration

$$F^i D_{[\rho_1,\rho_2]}(N_0, K, \alpha) := \bigl\{e \in D_{[\rho_1,\rho_2]}(N_0, K, \alpha) : \|e\|_{\rho_1,\rho_2} \le |p|^i\bigr\} \quad \text{for } i \in \mathbb{R},$$

on $D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$. Since K is discretely valued and $\rho_1, \rho_2 \in p^{\mathbb{Q}}$, this filtration is quasi-integral in the sense of [Schneider and Teitelbaum 2003, §1]. The corresponding graded ring $\operatorname{gr}^{\cdot} D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$, by Proposition A.21 in [Zábrádi 2012], is commutative. We let $\sigma(e) \in \operatorname{gr}^{\cdot} D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$ denote the principal symbol of any element $e \in D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$.

Proposition 4.3. (i) $\operatorname{gr}^{\cdot} D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$ *is a free* $\operatorname{gr}^{\cdot} K$*-module with basis*

$$\bigl\{\sigma(\boldsymbol{b}^{\boldsymbol{k}}) : \boldsymbol{k} \in \mathbb{Z}^{\{\alpha\}} \times \mathbb{N}^{\Phi^+ \setminus \{\alpha\}}\bigr\}.$$

(ii) *The map*

$$\mathbb{O}_K(A_\alpha[\rho_1, \rho_2]) \xrightarrow{\cong} D_{[\rho_1,\rho_2]}(N_0, K, \alpha), \quad f \mapsto f(\boldsymbol{b}),$$

is a K-linear isometric bijection.

Proof. Since $\bigl\{b_\alpha^{-l}\mu : l \ge 0, \mu \in D_{[0,\rho_1]}(N_0, K)\bigr\}$ is dense in $D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$, every element in the graded ring $\operatorname{gr}^{\cdot} D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$ is of the form $\sigma(b_\alpha^{-l}\mu)$. Suppose that $\mu = \sum_{\boldsymbol{k} \in \mathbb{N}_0^d} d_{\boldsymbol{k}} \boldsymbol{b}^{\boldsymbol{k}}$. Then $b_\alpha^{-l}\mu = \sum_{\boldsymbol{k} \in \mathbb{N}_0^d} d_{\boldsymbol{k}} b_\alpha^{-l} \boldsymbol{b}^{\boldsymbol{k}}$ and, using [Zábrádi

2012, Lemma A.7.iii], we compute

$$\begin{aligned}\|b_\alpha^{-l}\mu\|_{\rho_1,\rho_2} &= \max\bigl(\|b_\alpha\|_{\rho_1}^{-l}\|\mu\|_{\rho_1}, \|b_\alpha\|_{\rho_2}^{-l}\|\mu\|_{\rho_2}\bigr)\\ &= \max\bigl(\max_{\boldsymbol{k}\in\mathbb{N}_0^d}|d_{\boldsymbol{k}}|\rho_1^{\boldsymbol{k}-l}, \max_{\boldsymbol{k}\in\mathbb{N}_0^d}|d_{\boldsymbol{k}}|\rho_2^{\boldsymbol{k}-l}\bigr)\\ &= \max_{\boldsymbol{k}\in\mathbb{N}_0^d}|d_{\boldsymbol{k}}|\max\bigl(\rho_1^{\boldsymbol{k}-l},\rho_2^{\boldsymbol{k}-l}\bigr) = \max_{\boldsymbol{k}\in\mathbb{N}_0^d}|d_{\boldsymbol{k}}||b_\alpha^{-l}\boldsymbol{b}^{\boldsymbol{k}}|_{\rho_1,\rho_2}.\end{aligned}$$

It follows that $\operatorname{gr}^\cdot D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$ as a $\operatorname{gr}^\cdot K$-module is generated by the principal symbols $\sigma(b_\alpha^{-l}\boldsymbol{b}^{\boldsymbol{k}})$ with $\boldsymbol{k}\in\mathbb{N}_0^d$, $l\geq 0$. But it also follows that, for a fixed $l \geq 0$, the principal symbols $\sigma(b_\alpha^{-l}\boldsymbol{b}^{\boldsymbol{k}})$ with $\boldsymbol{k}$ running over $\mathbb{N}_0^d$ are linearly independent over $\operatorname{gr}^\cdot K$. By Proposition A.21 in [Zábrádi 2012], we may permute the factors in $\sigma(b_\alpha^{-l}\boldsymbol{b}^{\boldsymbol{k}})$ arbitrarily. Hence $\operatorname{gr}^\cdot D_{[\rho_1,\rho_2]}(N_0, K, \alpha)$ is a free $\operatorname{gr}^\cdot K$-module with basis $\{\sigma(\boldsymbol{b}^{\boldsymbol{k}}) : \boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}\}$.

On the other hand, we of course have

$$\begin{aligned}\|f(\boldsymbol{b})\|_{\rho_1,\rho_2} &\leq \max_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}}|d_{\boldsymbol{k}}||\boldsymbol{b}^{\boldsymbol{k}}|_{\rho_1,\rho_2} = \max_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}}|d_{\boldsymbol{k}}|\max(\rho_1^{\boldsymbol{k}},\rho_2^{\boldsymbol{k}})\\ &= \max\bigl(\max_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}}|d_{\boldsymbol{k}}|\rho_1^{\boldsymbol{k}}, \max_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}}|d_{\boldsymbol{k}}|\rho_2^{\boldsymbol{k}}\bigr) = |f|_{A_\alpha[\rho_1,\rho_2]}.\end{aligned}$$

This means that if we introduce on $\mathcal{O}_K(A_\alpha[\rho_1,\rho_2])$ the filtration defined by the spectral norm, then the asserted map respects the filtrations, and by the above reasoning, it induces an isomorphism between the associated graded rings. Hence, by completeness of these filtrations, it is an isometric bijection. □

Now we turn to the construction of $\mathscr{R}(N_1,\ell)$. The problem with (naïve) microlocalization is that the ring $\mathfrak{R}_0$ is not finitely generated over $\varphi(\mathfrak{R}_0)$. The reason for this is that φ improves the order of convergence for a power series in $\mathfrak{R}_0$. In the case $G\neq \mathrm{GL}_2(\mathbb{Q}_p)$, the operator $\varphi=\varphi_s$ acts by conjugation on N_β by raising to the $\beta(s)$-th power. Whenever $\beta\in\Phi^+\setminus\Delta$ is not a simple root, $\beta(s)=p^{m_\beta}>\alpha(s)=p$, where m_β is the degree of the map $\beta\circ\xi\colon\mathbb{G}_m\to\mathbb{G}_m$.

Lemma 4.4. *We have* $\|b_\beta\|_\rho=\|b_\alpha\|_\rho=\rho$ *and*

$$\|\varphi(b_\beta)\|_\rho=\max_{0\leq j\leq m_\beta}(\rho^{p^j}p^{j-m_\beta})<\max(\rho^p,p^{-1}\rho)=\|\varphi(b_\alpha)\|_\rho$$

for any $p^{-1}<\rho<1$. *In general,* $\|\varphi_t(b_\beta)\|_\rho=\max_{0\leq j\leq \mathrm{val}_p(\beta(t))}\bigl(\rho^{p^j}p^{j-\mathrm{val}_p(\beta(t))}\bigr)$.

Proof. We compute

$$\begin{aligned}\|\varphi_t(b_\beta)\|_\rho = \bigl\|(1+b_\beta)^{\beta(t)}-1\bigr\|_\rho &= \Bigl\|\sum_{i=1}^{\infty}\binom{\beta(t)}{i}b_\beta^i\Bigr\|_\rho\\ &= \max_{0\leq j\leq\mathrm{val}_p(\beta(t))}\bigl(\rho^{p^j}p^{j-\mathrm{val}_p(\beta(t))}\bigr).\end{aligned}$$

Here we use the trivial estimate

$$\mathrm{val}_p\binom{n}{k} = \mathrm{val}_p\left(\frac{n}{k}\binom{n-1}{k-1}\right) \geq \mathrm{val}_p(n) - \mathrm{val}_p(k)$$

for $n := \beta(t) \in \mathbb{Z}_p$ and $k \in \mathbb{N}$. We see immediately that whenever $m_\beta > 1$, we have $\rho^{p^j} p^{j-m_\beta} < \rho^p$ for $1 \leq j \leq m_\beta$ and $p^{-m_\beta}\rho < p^{-1}\rho$. □

Now choose an ordering $<$ on Φ^+ such that (i) $m_{\beta_1} < m_{\beta_2}$ implies $\beta_1 > \beta_2$ and (ii) $\alpha > \beta$ for any $\alpha \neq \beta, \beta_1, \beta_2 \in \Phi^+$. Then by Proposition 4.3, any element in $\mathfrak{R}_0$ has a skew Laurent-series expansion

$$f(\boldsymbol{b}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^{\{\alpha\}} \times \mathbb{N}^{\Phi^+ \setminus \{\alpha\}}} c_{\boldsymbol{k}} \boldsymbol{b}^{\boldsymbol{k}}$$

such that there exists $p^{-1} < \rho < 1$ with $|c_{\boldsymbol{k}}|_p \rho_1^{\sum k_\beta} \to 0$ as $\sum |k_\beta| \to \infty$ for all $\rho < \rho_1 < 1$. By Lemma 4.4 and the discussion above, we clearly have:

Example 4.5. Let $\beta \in \Phi^+ \setminus \Delta$ be a nonsimple root. Then the series $\sum_{n=1}^{\infty} b_\beta^n b_\alpha^{-n}$ does not belong to $\mathfrak{R}_0(N_0)$. However, the series $\sum_{n=1}^{\infty} \varphi(b_\beta^n b_\alpha^{-n})$ converges in each $D_{[\rho_1,\rho_2]}(N_0, \alpha)$ (for arbitrary $p^{-1} < \rho_1 < \rho_2 < 1$); hence it defines an element in $\mathfrak{R}_0(N_0)$. Therefore we cannot have a continuous left inverse ψ to φ on $\mathfrak{R}_0(N_0)$, as otherwise $\psi\left(\sum_{n=1}^{\infty} \varphi(b_\beta^n b_\alpha^{-n})\right) = \sum_{n=1}^{\infty} b_\beta^n b_\alpha^{-n}$ would converge. In particular, we cannot write $\mathfrak{R}_0(N_0)$ as the topological direct sum $\bigoplus_{u \in N_0/\varphi(N_0)} u\varphi(\mathfrak{R}_0(N_0))$ of closed subspaces in $\mathfrak{R}_0(N_0)$, as otherwise the operator

$$\psi\colon \mathfrak{R}_0(N_0) \to \mathfrak{R}_0(N_0), \qquad \sum_{u \in J(N_0/\varphi(N_0))} u\varphi(f_u) \mapsto \varphi^{-1}(u_0) f_{u_0}$$

for the unique $u_0 \in J(N_0/\varphi(N_0)) \cap \varphi(N_0)$ would be a continuous left inverse to φ. In fact, we even have $\mathfrak{R}_0(N_0) \neq \bigoplus_{u \in N_0/\varphi(N_0)} u\varphi(\mathfrak{R}_0(N_0))$ algebraically; however, the proof of this requires the forthcoming machinery (see Remark 4.10).

In order to overcome this counterexample, we are going to consider the ring $\mathscr{R}(N_1, \ell)$ of all the skew power series of the form $f(\boldsymbol{b})$ such that $f(\varphi_t(\boldsymbol{b}))$ is convergent in $\mathfrak{R}_0$ for some $t \in T_+$. A priori it is not clear that these series form a ring, so we are going to give a more conceptual construction.

Take an arbitrary element $t \in T_+$. The conjugation by t on N_0 gives an isomorphism $\varphi_t\colon N_0 \to \varphi_t(N_0)$ of pro-p groups (since it is injective). Hence $\varphi_t(N_0)$ is also a uniform pro-p group with minimal set of generators $\{\varphi_t(n_\beta)\}_{\beta \in \Phi^+}$. So we may define the distribution algebra $D(\varphi_t(N_0)) := D(\varphi_t(N_0), K)$. The inclusion $\varphi_t(N_0) \hookrightarrow N_0$ induces an injective homomorphism of Fréchet algebras $\iota_{1,t}\colon D(\varphi_t(N_0)) \hookrightarrow D(N_0)$. It is well known [Schneider and Teitelbaum 2003] that

$$D(N_0) = \bigoplus_{n \in J(N_0/\varphi_t(N_0))} n\iota_{1,t}\big(D(\varphi_t(N_0))\big)$$

as right $D(\varphi_t(N_0))$-modules. Moreover, the direct summands are closed in $D(N_0)$. For each real number $p^{-1} < \rho < 1$, the ρ-norm on $D(N_0)$ defines a norm $r_t(\rho)$ on $D(\varphi_t(N_0))$ by restriction. This is different from the ρ-norm on $D(\varphi_t(N_0))$ (using the uniform structure on $\varphi_t(N_0)$). However, the family $(r_t(\rho))_\rho$ of norms defines the Fréchet topology on $D(\varphi_t(N_0))$. On the other hand, whenever r is a norm on $D(\varphi_t(N_0))$, we may extend r to a norm $q_t(r)$ on $D(N_0)$ by putting

$$\Bigg\| \sum_{n \in J(N_0/\varphi_t(N_0))} n\iota_{1,t}(x_n) \Bigg\|_{q_t(r)} := \max(\|x_n\|_r).$$

These norms define the Fréchet topology on $D(N_0)$. More precisely, if $\beta(t) = p^{m(\beta,t)}u(\beta,t)$ with $m(\beta,t) := \mathrm{val}_p(\beta(t)) \geq 0$ integer and $u(\beta,t) \in \mathbb{Z}_p^\times$, then:

Lemma 4.6.

$$\|x\|_\rho \leq \|x\|_{q_t(r_t(\rho))} \leq \rho^{-\sum_{\beta\in\Phi^+}(p^{m(\beta,t)}-1)}\|x\|_\rho$$

for any

$$p^{-\frac{1}{\max_{\beta\in\Phi^+} p^{m(\beta,t)}}} < \rho < 1$$

and $x \in D(N_0)$. In particular, the norms ρ and $q_t(r_t(\rho))$ define the same topology.

Proof. The inequality on the left is clear from the triangle inequality. For the other inequality, note that our assumption on ρ implies in particular that

$$\rho^{p^{m(\beta,t)}} = \rho^{p^j}\rho^{p^{m(\beta,t)}-p^j} > \rho^{p^j} p^{-\frac{p^{m(\beta,t)}-p^j}{p^{m(\beta,t)}}} > \rho^{p^j} p^{j-m(\beta,t)}$$

for all $0 \leq j < m(\beta,t)$. Hence, by Lemma 4.4, we have

$$\rho^{p^{m(\beta,t)}} = \left\| \binom{\beta(t)}{p^{m(\beta,t)}} b_\beta^{p^{m(\beta,t)}} \right\|_\rho = \|\varphi_t(b_\beta)\|_\rho.$$

Moreover, there exists an *invertible* element y in the Iwasawa algebra $\Lambda(N_{0,\beta})$ such that

$$y\varphi_t(b_\beta) \equiv \binom{\beta(t)}{p^{m(\beta,t)}} b_\beta^{p^{m(\beta,t)}} \pmod p$$

(as both sides have the same principal term). But by the choice of ρ, $|p| = 1/p < \rho^{p^{m(\beta,t)}} = \|\varphi_t(b_\beta)\|_\rho = \|\varphi_t(b_\beta)\|_{q_t(r_t(\rho))}$. Therefore we also have

$$\begin{aligned}\rho^{p^{m(\beta,t)}} &= \|b_\beta^{p^{m(\beta,t)}}\|_\rho = \left\| \binom{\beta(t)}{p^{m(\beta,t)}} b_\beta^{p^{m(\beta,t)}} \right\|_\rho = \|\varphi_t(b_\beta)\|_\rho \\ &= \|\varphi_t(b_\beta)\|_{q_t(r_t(\rho))} = \|y\varphi_t(b_\beta)\|_{q_t(r_t(\rho))} \\ &= \left\| \binom{\beta(t)}{p^{m(\beta,t)}} b_\beta^{p^{m(\beta,t)}} \right\|_{q_t(r_t(\rho))} = \|b_\beta^{p^{m(\beta,t)}}\|_{q_t(r_t(\rho))},\end{aligned}$$

whence

$$\begin{aligned}\|b_\beta^{k_\beta}\|_{q_t(r_t(\rho))} &= \|b_\beta^{p^{m(\beta,t)}}\|_{q_t(r_t(\rho))}^{k_{1,\beta}}\|b_\beta^{k_{2,\beta}}\|_{q_t(r_t(\rho))}\\ &\leq \rho^{k_{1,\beta}p^{m(\beta,t)}} \leq \rho^{-p^{m(\beta,t)}+1}\|b_\beta^{k_\beta}\|_\rho,\end{aligned}\tag{13}$$

where $k_\beta = p^{m(\beta,t)}k_{1,\beta}+k_{2,\beta}$ with $0\leq k_{2,\beta}\leq p^{m(\beta,t)}-1$ and $k_{1,\beta}$ nonnegative integers.

Now consider an element of $D(N_0)$ of the form

$$x = \sum_{\boldsymbol{k}=(k_\beta)\in\mathbb{N}^{\Phi^+}} c_{\boldsymbol{k}} \prod_{\beta\in\Phi^+} b_\beta^{k_\beta}.$$

We may assume without loss of generality that

$$J(N_0/\varphi_t(N_0)) = \left\{ \prod_{\beta\in\Phi^+} n_\beta^{j_\beta} \mid 0\leq j_\beta\leq p^{m(\beta,t)}-1 \right\},$$

where the product is taken in the reverse order. Let $\eta\in\Phi^+$ be the largest root (with respect to the ordering $<$ defined after Lemma 4.4) such that there exists a $\boldsymbol{k}\in\mathbb{N}^{\Phi^+}$ with $c_{\boldsymbol{k}}\neq 0$ and $k_\eta\neq 0$. We are going to show the estimate

$$\|x\|_{q_t(r_t(\rho))} \leq \rho^{-\sum_{\beta\leq\eta}(p^{m(\beta,t)}-1)}\|x\|_\rho$$

by induction on η. This induction has in fact finitely many steps since $|\Phi^+|<\infty$. At first we write $b_\eta^{k_\eta} = \sum_{j_\eta=0}^{p^{m(\eta,t)}-1} n_\eta^{j_\eta} f_{\boldsymbol{k},j_\eta}(\varphi_t(b_\eta))$ for each $\boldsymbol{k}\in\mathbb{N}^{\Phi^+}$. By the choice of the ordering on Φ^+, for any fixed η, the set $\prod_{\beta<\eta} N_{0,\beta}$ is a normal subgroup of N_0. Moreover, the conjugation by any element of N_0 preserves the ρ-norm on $D(N_0)$. Therefore we may write

$$\prod_{\beta\leq\eta} b_\beta^{k_\beta} = \sum_{j_\eta=0}^{p^{m(\eta,t)}-1} n_\eta^{j_\eta} x_{\boldsymbol{k},j_\eta} f_{\boldsymbol{k},j_\eta}(\varphi_t(b_\eta))$$

such that

$$x_{\boldsymbol{k},j_\eta} := n_\eta^{-j_\eta}\Bigl(\prod_{\beta<\eta} b_\beta^{k_\beta}\Bigr) n_\eta^{j_\eta} \in D\Bigl(\prod_{\beta<\eta} N_{0,\beta}\Bigr).$$

By (13), we have

$$\begin{aligned}\bigl\|f_{\boldsymbol{k},j_\eta}(\varphi_t(b_\eta))\bigr\|_{q_t(r_t(\rho))} &= \bigl\|f_{\boldsymbol{k},j_\eta}(\varphi_t(b_\eta))\bigr\|_\rho\\ &\leq \|b_\eta^{k_\eta}\|_{q_t(r_t(\rho))} \leq \rho^{-p^{m(\eta,t)}+1}\|b_\eta^{k_\eta}\|_\rho.\end{aligned}$$

Since the $r_t(\rho)$-norm is multiplicative on $D(\varphi_t(N_0))$, for any $a\in D(N_0)$ and $b\in D(\varphi_t(N_0))$ we also have $\|a\iota_{1,t}(b)\|_{q_t(r_t(\rho))} = \|a\|_{q_t(r_t(\rho))}\|b\|_{r_t(\rho)}$. Indeed, if we decompose a as $a=\sum_{n\in J(N_0/\varphi_t(N_0))} n\iota_{1,t}(a_n)$, then we have $a\iota_{1,t}(b) =$

$\sum_{n \in J(N_0/\varphi_t(N_0))} n\iota_{1,t}(a_n b)$. Now $f_{\boldsymbol{k},j_\eta}(\varphi_t(b_\eta))$ lies in $\iota_{1,t}(D(\varphi_t(N_0)))$, so we see that

$$\left\| x_{\boldsymbol{k},j_\eta} f_{\boldsymbol{k},j_\eta}(\varphi_t(b_\eta)) \right\|_{q_t(r_t(\rho))} = \| x_{\boldsymbol{k},j_\eta} \|_{q_t(r_t(\rho))} \left\| f_{\boldsymbol{k},j_\eta}(\varphi_t(b_\eta)) \right\|_{q_t(r_t(\rho))}.$$

On the other hand, the inductional hypothesis tells us that

$$\| x_{\boldsymbol{k},j_\eta} \|_{q_t(r_t(\rho))} \le \rho^{-\sum_{\beta<\eta}(p^{m(\beta,t)}-1)} \| x_{\boldsymbol{k},j_\eta} \|_\rho = \rho^{-\sum_{\beta<\eta}(p^{m(\beta,t)}-1)} \left\| \prod_{\beta<\eta} b_\beta^{k_\beta} \right\|_\rho.$$

Hence we compute

$$\begin{aligned} \|x\|_{q_t(r_t(\rho))} &= \left\| \sum_{\boldsymbol{k}} c_{\boldsymbol{k}} \sum_{j_\eta=0}^{p^{m(\eta,t)}-1} n_\eta^{j_\eta} x_{\boldsymbol{k},j_\eta} f_{\boldsymbol{k},j_\eta}(\varphi_t(b_\eta)) \right\|_{q_t(r_t(\rho))} \\ &\le \max_{\boldsymbol{k},j_\eta} \left(|c_{\boldsymbol{k}}| \left\| x_{\boldsymbol{k},j_\eta} f_{\boldsymbol{k},j_\eta}(\varphi_t(b_\eta)) \right\|_{q_t(r_t(\rho))} \right) \\ &\le \max_{\boldsymbol{k}} \left(|c_{\boldsymbol{k}}| \rho^{-\sum_{\beta<\eta}(p^{m(\beta,t)}-1)} \left\| \prod_{\beta<\eta} b_\beta^{k_\beta} \right\|_\rho \rho^{-p^{m(\eta,t)}+1} \| b_\eta^{k_\eta} \|_\rho \right) \\ &= \rho^{-\sum_{\beta\le\eta}(p^{m(\beta,t)}-1)} \|x\|_\rho. \end{aligned}$$

□

In particular, for each ρ in the range $p^{-1/\max_\beta p^{m(\beta,t)}} < \rho < 1$, the completions of $D(N_0)$ with respect to the topologies defined by $\|\cdot\|_\rho$ and by $\|\cdot\|_{q_t(r_t(\rho))}$ are the same, that is,

$$D_{[0,\rho]}(N_0) = \bigoplus_{n \in J(N_0/\varphi_t(N_0))} n\iota_{1,t}\big(D_{r_t([0,\rho])}(\varphi_t(N_0))\big), \tag{14}$$

where $D_{r_t([0,\rho])}(\varphi_t(N_0))$ denotes the completion of $D(\varphi_t(N_0))$ with respect to the norm $r_t(\rho)$.

Now we turn to the microlocalization and note first that $\varphi_t(b_\alpha) = (b_\alpha + 1)^{\alpha(t)} - 1$ is divisible by b_α. So if $\varphi_t(b_\alpha)$ is invertible in a ring, then so is b_α. On the other hand, if $p^{-1/p^{m(\alpha,t)}} < \rho < 1$, then by Lemma 4.4 we have

$$\left\| \varphi_t(b_\alpha) - \binom{\alpha(t)}{p^{m(\alpha,t)}} b_\alpha^{p^{m(\alpha,t)}} \right\|_\rho < \left\| \binom{\alpha(t)}{p^{m(\alpha,t)}} b_\alpha^{p^{m(\alpha,t)}} \right\|_\rho.$$

Hence $\varphi_t(b_\alpha)$ is invertible in the Banach algebra $D_{[\rho_1,\rho_2]}(N_0,\alpha)$ for any ρ_1, ρ_2 with $p^{-1/p^{m(\alpha,t)}} < \rho_1 < \rho_2 < 1$, since it is close to the invertible element

$$\binom{\alpha(t)}{p^{m(\alpha,t)}} b_\alpha^{p^{m(\alpha,t)}}$$

(as the binomial coefficient in this expression is not divisible by p). This shows that the microlocalization of $D_{[0,\rho_2]}(N_0)$ with respect to the multiplicative set $\varphi_t(b_\alpha)^{\mathbb{N}}$ and norm $\max(\rho_1,\rho_2)$ equals $D_{[\rho_1,\rho_2]}(N_0,\alpha)$. Therefore for each ρ_1 and ρ_2 with $p^{-1/p^{m(\alpha,t)}} < \rho_1 < \rho_2 < 1$ we obtain

$$D_{[\rho_1,\rho_2]}(N_0,\alpha) = \bigoplus_{n\in J(N_0/\varphi_t(N_0))} n\iota_{0,1}\big(D_{r_t([\rho_1,\rho_2])}(\varphi_t(N_0),\alpha)\big)$$

by microlocalizing both sides of (14). Now letting ρ_2 tend to 1 and then also $\rho_1 \to 1$, we get

$$\mathfrak{R}_0(N_0,\alpha) = \bigoplus_{n\in J(N_0/\varphi_t(N_0))} n\iota_{1,t}\big(\mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)\big) \tag{15}$$

for all $t \in T_+$. Here we define

$$\mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha) := \varinjlim_{\rho_1\to 1}\varprojlim_{\rho_2\to 1} D_{r_t([\rho_1,r(\rho_2)])}(\varphi_t(N_0),\alpha),$$

which is in general different from $\mathfrak{R}_0(\varphi_t(N_0),\alpha)$ (in which, by definition, we use norms ρ such that $\|\varphi_t(b_\beta)\|_\rho = \|\varphi_t(b_\alpha)\|_\rho$), by Example 4.5. Indeed, for $t = s$ the sum $\sum_{n=1}^{\infty}\varphi(b_\beta^n b_\alpha^{-n})$ converges in $\mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)$, but not in $\mathfrak{R}_0(\varphi_t(N_0),\alpha)$.

By entirely the same proof, we also obtain

$$\mathfrak{R}_{0,r_{t_1}(\cdot)}(\varphi_{t_1}(N_0),\alpha) = \bigoplus_{n\in J(\varphi_{t_1}(N_0)/\varphi_{t_1t_2}(N_0))} n\iota_{t_1,t_1t_2}\big(\mathfrak{R}_{0,r_{t_1t_2}(\cdot)}(\varphi_{t_1t_2}(N_0),\alpha)\big) \tag{16}$$

for each pair $t_1,t_2 \in T_+$, where ι_{t_1,t_1t_2} is the inclusion of the rings above induced by the natural inclusion $\varphi_{t_1t_2}(N_0) \hookrightarrow \varphi_{t_1}(N_0)$.

Now we would like to define continuous homomorphisms

$$\varphi_{t_2t_1,t_1}\colon \mathfrak{R}_{0,r_{t_1}(\cdot)}(\varphi_{t_1}(N_0),\alpha)\to\mathfrak{R}_{0,r_{t_1t_2}(\cdot)}(\varphi_{t_1t_2}(N_0),\alpha), \quad \varphi_{t_1}(b_\beta)\mapsto\varphi_{t_1t_2}(b_\beta)$$

induced by the group isomorphism $\varphi_{t_2}\colon \varphi_{t_1}(N_0) \to \varphi_{t_1t_2}(N_0)$ so that we can take the injective limit

$$\mathfrak{R}(N_1,\ell) := \varinjlim_t \mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)$$

with respect to the maps $\varphi_{t_2t_1,t_1}$. This is not possible for all t_2 since the map φ_{t_2} will not always be norm-decreasing on monomials $\boldsymbol{b}^{\boldsymbol{k}}$ for $\boldsymbol{k} \in \mathbb{Z}^{\{\alpha\}} \times \mathbb{N}^{\Phi^+\setminus\{\alpha\}}$. To overcome this, we define the pre-ordering $\leq_\alpha$ (depending on the choice of the simple root α) on T_+ the following way: $t_1 \leq_\alpha t_2$ if and only if $|\beta(t_2t_1^{-1})| \leq |\alpha(t_2t_1^{-1})| \leq 1$ for all $\beta \in \Phi^+$. (That is, if and only if we have $m(\beta,t_2t_1^{-1}) \geq m(\alpha,t_2t_1^{-1}) \geq 0$.) In particular, $t_1 \leq_\alpha t_2$ implies $t_2t_1^{-1} \in T_+$ and it is equivalent to $1 \leq_\alpha t_2t_1^{-1}$. We also have $1 \leq_\alpha s$ for any $\alpha \in \Delta$. It is clear that $\leq_\alpha$ is transitive and reflexive. If $t_2 \leq_\alpha t_1 \leq_\alpha t_2$, then $|\beta(t_2t_1^{-1})| = 1$ for all $\beta \in \Phi^+$, whence $t_2t_1^{-1}$ lies in T_0. Therefore $\leq_\alpha$ defines a partial ordering on the quotient monoid T_+/T_0.

Lemma 4.7. *The partial ordering $\leq_\alpha$ on T_+/T_0 is right filtered, that is, any finite subset of T_+/T_0 has a common upper bound with respect to $\leq_\alpha$.*

Proof. Take any $t_1, t_2 \in T_+$ with $|\alpha(t_1)| \leq |\alpha(t_2)|$. Since the simple roots $\beta \in \Delta$ are linearly independent in $X^*(T) = \mathrm{Hom}_{\mathrm{alg}}(T, \mathbb{G}_m)$ and the pairing $X^*(T) \times X_*(T) \to \mathbb{Z}$ is perfect, we may choose $s_{\overline{\alpha}} \in T$ so that $|\beta(s_{\overline{\alpha}})| < |\alpha(s_{\overline{\alpha}})| = 1$ for all $\alpha \neq \beta \in \Delta$. Since all the positive roots are positive linear combinations of the simple roots, we see immediately that $s_{\overline{\alpha}} \in T_+$. If $\alpha \neq \gamma \in \Phi^+$, then γ is not a scalar multiple of α; hence, writing $\gamma = \sum_{\beta \in \Delta} m_{\beta,\gamma}\beta$, there is an $\alpha \neq \beta \in \Delta$ with $m_{\beta,\gamma} > 0$, whence $|\gamma(s_{\overline{\alpha}})| < 1$. So we have $t_1 \leq_\alpha t_1 s_{\overline{\alpha}}^k$ for any $k \geq 0$ and $t_2 \leq_\alpha t_1 s_{\overline{\alpha}}^k$ for k big enough. □

Fix an element $1 \leq_\alpha t \in T_+$ and let ρ_1, ρ_2 be real numbers in $p^{\mathbb{Q}}$ such that

$$p^{-1/\max\limits_{\beta \in \Phi^+} p^{m(\beta,t)+m(\alpha,t)}} < \rho_1 < \rho_2 < 1.$$

Note that $\varphi_t \colon N_0 \to \varphi_t(N_0)$ is an isomorphism of pro-p groups. Hence it induces an isometric isomorphism

$$\varphi_t \colon D_{[0,\rho_2^{p^{m(\alpha,t)}}]}(N_0) \to D_{[0,\rho_2^{p^{m(\alpha,t)}}]}(\varphi_t(N_0)),$$
$$\sum_{\boldsymbol{k}} c_{\boldsymbol{k}} \prod_{\beta} b_\beta^{k_\beta} \mapsto \sum_{\boldsymbol{k}} c_{\boldsymbol{k}} \prod_{\beta} \varphi_t(b_\beta)^{k_\beta}$$

of Banach algebras, where

$$D_{[0,\rho_2^{p^{m(\alpha,t)}}]}(\varphi_t(N_0))$$

denotes the completion of $D(\varphi_t(N_0))$ with respect to the $\rho_2^{p^{m(\alpha,t)}}$-norm defined by the set of generators $\{\varphi_t(n_\beta)\}_{\beta \in \Phi^+}$ of $\varphi_t(N_0)$. To avoid confusion, from now on we denote by the subscript ρ, N_0 the ρ-norm (as before) on $D(N_0)$ and by the subscript $\rho, \varphi_t(N_0)$ the ρ-norm on $D(\varphi_t(N_0))$. By Lemma 4.4, we have

$$\|\varphi_t(b_\beta)\|_{\rho,N_0} = \rho^{p^{m(\beta,t)}} \leq \rho^{p^{m(\alpha,t)}} = \|\varphi_t(b_\alpha)\|_{\rho,N_0}$$

for any $\beta \in \Phi^+$ and $\rho = \rho_1$ or $\rho = \rho_2$ because of our assumption $1 \leq_\alpha t$. This shows that for any monomial $\prod_{\beta \in \Phi^+} \varphi_t(b_\beta)^{k_\beta}$ (with $k_\beta \geq 0$ for all $\beta \in \Phi^+$), we have

$$\begin{aligned}\Big\| \prod_{\beta \in \Phi^+} \varphi_t(b_\beta)^{k_\beta} \Big\|_{r_t(\rho)} &= \Big\| \prod_{\beta \in \Phi^+} \varphi_t(b_\beta)^{k_\beta} \Big\|_{\rho,N_0} \\ &\leq \rho^{p^{m(\alpha,t)}\boldsymbol{k}} = \Big\| \prod_{\beta \in \Phi^+} \varphi_t(b_\beta)^{k_\beta} \Big\|_{\rho^{p^{m(\alpha,t)}},\varphi_t(N_0)},\end{aligned}$$

since both norms are multiplicative on $D(\varphi_t(N_0))$. We obtain a norm-decreasing homomorphism

$$D_{[0,\rho_2^{p^{m(\alpha,t)}}]}(N_0) \xrightarrow{\sim} D_{[0,\rho_2^{p^{m(\alpha,t)}}]}(\varphi_t(N_0)) \to D_{r_t([0,\rho_2])}(\varphi(N_0)) \\ \to D_{r_t([\rho_1,\rho_2])}(\varphi(N_0),\alpha).$$

The element $\varphi_t(b_\alpha)$ is invertible in $D_{r_t([\rho_1,\rho_2])}(\varphi_t(N_0),\alpha)$, and for each $\rho_1 \leq \rho \leq \rho_2$ and $x \in D_{[0,\rho_2]}(N_0)$, we have

$$\|\varphi_t(x)\varphi_t(b_\alpha)^{-k}\|_{r_t(\rho)} \leq \|x\|_{\rho^{p^{m(\alpha,t)}},N_0}\|b_\alpha^{-k}\|_{\rho^{p^{m(\alpha,t)}},N_0}.$$

Therefore by the universal property of microlocalization [Zábrádi 2012, Proposition A.18], we obtain a norm-decreasing homomorphism

$$\varphi_{t,1}\colon D_{[\rho_1^{p^{m(\alpha,t)}},\rho_2^{p^{m(\alpha,t)}}]}(N_0,\alpha) \to D_{r_t([\rho_1,\rho_2])}(\varphi_t(N_0),\alpha), \quad b_\beta \mapsto \varphi_t(b_\beta). \tag{17}$$

This map is not surjective in general, by Example 4.5.

Lemma 4.8. *The map* (17) *is injective.*

Proof. Take an element

$$f(\boldsymbol{b}) = \sum_{\boldsymbol{k}} d_{\boldsymbol{k}}\boldsymbol{b}^{\boldsymbol{k}} \in D_{[\rho_1^{p^{m(\alpha,t)}},\rho_2^{p^{m(\alpha,t)}}]}(N_0,\alpha)$$

and pairwise distinct $\boldsymbol{k}_1,\dots,\boldsymbol{k}_r \in \mathbb{Z}^{\{\alpha\}} \times \mathbb{N}^{\Phi^+\setminus\{\alpha\}}$. Note that

$$\|\varphi_t(b_\beta)\|_{\rho,N_0} > \left\|\varphi_t(b_\beta) - \binom{\beta(t)}{p^{m(\beta,t)}} b_\beta^{p^{m(\beta,t)}}\right\|_{\rho,N_0};$$

hence we obtain

$$\begin{aligned}\left\|\sum_{j=1}^r d_{\boldsymbol{k}_j}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}_j}\right\|_{r_t(\rho_1),r_t(\rho_2)} &= \left\|\sum_{j=1}^r d_{\boldsymbol{k}_j}\prod_{\beta\in\Phi^+}\left(\binom{\beta(t)}{p^{m(\beta,t)}} b_\beta^{p^{m(\beta,t)}}\right)^{k_{j,\beta}}\right\|_{\rho_1,\rho_2} \\ &= \max_j \left\|d_{\boldsymbol{k}_j}\prod_{\beta\in\Phi^+}\left(\binom{\beta(t)}{p^{m(\beta,t)}} b_\beta^{p^{m(\beta,t)}}\right)^{k_{j,\beta}}\right\|_{\rho_1,\rho_2} \\ &= \max_j \|d_{\boldsymbol{k}_j}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}_j}\|_{r_t(\rho_1),r_t(\rho_2)}\end{aligned}$$

using Proposition 4.3, as we have $\prod_{\beta\in\Phi^+} b_\beta^{p^{m(\beta,t)}k_{j_1,\beta}} \neq \prod_{\beta\in\Phi^+} b_\beta^{p^{m(\beta,t)}k_{j_2,\beta}}$ for $1 \leq j_1 \neq j_2 \leq r$.

Since the map $\varphi_{t,1}$ is norm-decreasing, we have $\|d_{\boldsymbol{k}}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}}\|_{r_t(\rho_1),r_t(\rho_2)} \to 0$ as $\boldsymbol{k} \to \infty$. Therefore we also have

$$\left\|\sum_{\boldsymbol{k}} d_{\boldsymbol{k}}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}}\right\|_{r_t(\rho_1),r_t(\rho_2)} = \max_{\boldsymbol{k}}\|d_{\boldsymbol{k}}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}}\|_{r_t(\rho_1),r_t(\rho_2)},$$

which is nonzero if there exists a $\boldsymbol{k}$ with $d_{\boldsymbol{k}} \neq 0$. Therefore the injectivity. □

Taking projective and injective limits, we obtain an injective ring homomorphism

$$\varphi_{t,1}\colon \mathfrak{R}_0(N_0,\alpha) \hookrightarrow \mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)$$

for any $1 \leq_\alpha t \in T_+$.

Remark 4.9. Note that $\mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)$ is a subring of $\mathfrak{R}_0(N_0,\alpha)$ via the map $\iota_{1,t}$ (for all $t \in T_+$). Hence for $1 \leq_\alpha t$ we obtain a ring homomorphism $\varphi_t = \iota_{1,t} \circ \varphi_{t,1}\colon \mathfrak{R}_0(N_0,\alpha) \to \mathfrak{R}_0(N_0,\alpha)$. However, if $1 \not\leq_\alpha t$ for some $t \in T_+$, then we in fact do not have a continuous ring homomorphism $\varphi_t\colon \mathfrak{R}_0(N_0,\alpha) \to \mathfrak{R}_0(N_0,\alpha)$. Indeed, in this case there exists a $\beta \in \Phi^+$ such that $|\beta(t)| > |\alpha(t)|$, so there exist integers $k_\beta > k_\alpha$ such that for any $p^{-|\alpha(t)|} < \rho < 1$,

$$\left\|\varphi_t\left(b_\beta^{k_\beta} b_\alpha^{-k_\alpha}\right)\right\|_\rho = \rho^{\frac{k_\beta}{|\beta(t)|} - \frac{k_\alpha}{|\alpha(t)|}} > 1;$$

therefore

$$\sum_{n=1}^{\infty} \varphi_t\left(b_\beta^{nk_\beta} b_\alpha^{-nk_\alpha}\right)$$

does not converge in $\mathfrak{R}_0(N_0,\alpha)$ even though $\sum_{n=1}^{\infty} b_\beta^{nk_\beta} b_\alpha^{-nk_\alpha}$ does.

Remark 4.10. If $\Phi^+ \neq \Delta$ (for example, if $G = \mathrm{GL}_n(\mathbb{Q}_p)$, $n > 2$), then we have

$$\begin{aligned}\mathfrak{R}_0(N_0,\alpha) &= \bigoplus_{u \in J(N_0/\varphi_t(N_0))} u\iota_{1,s}\big(\mathfrak{R}_{0,r_s(\cdot)}(\varphi(N_0),\alpha)\big) \\ &\supsetneq \bigoplus_{n \in J(N_0/\varphi(N_0))} u\varphi(\mathfrak{R}_0(N_0,\alpha)),\end{aligned}$$

by (15) (with the choice $t = s$) and Example 4.5 (which shows that $\varphi_{s,1}$ is not surjective).

In a similar fashion, we get for $t_1 \in T_+$ (and $1 \leq_\alpha t \in T_+$) an injective homomorphism

$$\varphi_{tt_1,t_1}\colon \mathfrak{R}_{0,r_{t_1}(\cdot)}(\varphi_{t_1}(N_0),\alpha) \to \mathfrak{R}_{0,r_{tt_1}(\cdot)}(\varphi_{tt_1}(N_0),\alpha).$$

In view of Lemma 4.7, we define

$$\mathscr{R}(N_1,\ell) := \varinjlim_{t \in T_+} \mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)$$

with respect to the maps φ_{t_1,t_2} for $t_2 \leq_\alpha t_1$.

Now take any $t \in T_+$ (not necessarily satisfying $1 \leq_\alpha t$). The map

$$\varphi_t := \varinjlim_{t_1} \iota_{t_1,tt_1}\colon \mathscr{R}(N_1,\ell) \to \mathscr{R}(N_1,\ell)$$

is defined as the direct limit of the inclusion maps

$$\iota_{t_1,tt_1}\colon \mathfrak{R}_{0,r_{tt_1}(\cdot)}(\varphi_{tt_1}(N_0),\alpha) \hookrightarrow \mathfrak{R}_{0,r_{t_1}(\cdot)}(\varphi_{t_1}(N_0),\alpha)$$

induced by $\varphi_{tt_1}(N_0) \subseteq \varphi_{t_1}(N_0)$. By definition, the ring $\mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)$ for any $t \in T_+$ consists of formal power series $\sum_{\boldsymbol{k}} c_{\boldsymbol{k}}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}}$ that converge in $\mathfrak{R}_0(N_0,\alpha)$. Therefore the map

$$\bigcup_{t\in T_+}\left\{\sum_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}} c_{\boldsymbol{k}}\boldsymbol{b}^{\boldsymbol{k}} \;\middle|\; \sum_{\boldsymbol{k}} c_{\boldsymbol{k}}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}} \text{ converges in } \mathfrak{R}_0(N_0,\alpha)\right\} \to \mathcal{R}(N_1,\ell),$$

$$\sum_{\boldsymbol{k}} c_{\boldsymbol{k}}\boldsymbol{b}^{\boldsymbol{k}} \mapsto \sum_{\boldsymbol{k}} c_{\boldsymbol{k}}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}} \in \mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha) \hookrightarrow \mathcal{R}(N_1,\ell)$$

is well-defined and bijective, since $\sum_{\boldsymbol{k}} c_{\boldsymbol{k}}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}}$ converges for some $t \in T_+$ and the connecting homomorphisms in the injective limit defining $\mathcal{R}(N_1,\ell)$ are injective and given by φ_{t_1,t_2} for $t_2 \leq_\alpha t_1$.

Hence we may identify

$$\mathcal{R}(N_1,\ell) = \bigcup_{t\in T_+}\left\{\sum_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}} c_{\boldsymbol{k}}\boldsymbol{b}^{\boldsymbol{k}} \;\middle|\; \sum_{\boldsymbol{k}} c_{\boldsymbol{k}}\varphi_t(\boldsymbol{b})^{\boldsymbol{k}} \text{ convergent in } \mathfrak{R}_0(N_0,\alpha)\right\} \tag{18}$$

and obtain:

Proposition 4.11. *The natural map $\varphi_t\colon \mathcal{R}(N_1,\ell) \to \mathcal{R}(N_1,\ell)$ is injective for all $t \in T_+$, and we have the decomposition*

$$\mathcal{R}(N_1,\ell) = \bigoplus_{n\in J(N_0/\varphi_t(N_0))} n\varphi_t(\mathcal{R}(N_1,\ell)).$$

In particular, $\mathcal{R}(N_1,\ell)$ is a free (right) module over itself via φ_t and it is a φ-ring over N_0 with $\varphi = \varphi_s$ in the sense of Definition 2.9.

Proof. By (16), we have

$$\mathfrak{R}_{0,r_{t_1}(\cdot)}(\varphi_{t_1}(N_0),\alpha) = \bigoplus_{n\in J(N_0/\varphi_t(N_0))} \varphi_{t_1}(n)\iota_{t_1,tt_1}\big(\mathfrak{R}_{0,r_{tt_1}(\cdot)}(\varphi_{tt_1}(N_0),\alpha)\big)$$

for any $t_1 \in T_+$. The statement follows by taking the injective limit of both sides (with respect to t_1) and noting that

$$\varphi_{t_1,1}(n) = \varphi_{t_1}(n) \in \varphi_{t_1}(N_0) \subseteq \mathfrak{R}_{0,r_{t_1}(\cdot)}(\varphi_{t_1}(N_0),\alpha)$$

for $n \in N_0 \subseteq \mathfrak{R}_0(N_0)$ and $1 \leq_\alpha t_1$; therefore n corresponds to $\varinjlim_{1\leq_\alpha t_1} (\varphi_{t_1}(n))_{t_1}$ via the identification (18). □

Remark 4.12. The ring $\mathcal{R}(N_1,\ell)$ via the description (18) consists of exactly those Laurent-series

$$x = \sum_{\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}} c_{\boldsymbol{k}}\boldsymbol{b}^{\boldsymbol{k}}$$

that converge on the open annulus of the form

$$\bigl\{\rho_2 < |z_\alpha| < 1,\ |z_\beta| \le |z_\alpha|^r \text{ for } \beta \in \Phi^+ \setminus \{\alpha\}\bigr\}, \tag{19}$$

for some $p^{-1} < \rho_2 < 1$ and $1 \le r \in \mathbb{Z}$.

Proof. If $x \in \mathcal{R}(N_1,\ell)$, then there exists a $t \in T_+$ such that $\varphi_t(x)$ converges in $\mathfrak{R}_0(N_0)$, that is, it converges in the norm $\|b_\beta\|_\rho = \rho$ for all $\beta \in \Phi^+$ for some fixed $p^{-1} < \rho_0 < 1$ and all $\rho \in (\rho_0, 1)$. By Lemma 4.7, we may assume that $|\alpha(t)| = 1$, whence $\|\varphi_t(b_\alpha)\|_\rho = \rho$ for all $\rho < 1$, as we may take $t = s_{\overline{\alpha}}^k$ for k large enough. Now let $\rho_2 := \rho_0$ and $r := \max_{\beta\in\Phi^+}\bigl([|1/\beta(t)|] + 1\bigr) \in \mathbb{Z}$. Then x converges on the annulus (19), as we have $\rho^r \le \rho^{1/|\beta(t)|} \le \|\varphi_t(b_\beta)\|_\rho$ for all $\beta \in \Phi^+ \setminus \{\alpha\}$, by Lemma 4.4.

Conversely, for any fixed $p^{-1} < \rho_2 < 1$ and integer $r \ge 1$, we need to find a $t \in T_+$ and a $\rho_0 \in (p^{-1}, 1)$ such that for all $\rho \in (\rho_0, 1)$ we have $\rho_2 < \|\varphi_t(b_\alpha)\|_\rho < 1$ and $\|\varphi_t(b_\beta)\|_\rho \le \|\varphi_t(b_\alpha)\|_\rho^r$. We take $t := s_{\overline{\alpha}}^k$ and $\rho_0 := \max\bigl(\rho_2, p^{-|\beta(t)|} \mid \beta \in \Phi^+\setminus\{\alpha\}\bigr)$, where

$$k := \max_{\beta\in\Phi^+\setminus\{\alpha\}}\left(\left[-\frac{\log r}{\log|\beta(s_{\overline{\alpha}})|}\right] + 1\right)$$

(for the definition of $s_{\overline{\alpha}}$ see the proof of Lemma 4.7). Indeed, since $|\alpha(s_{\overline{\alpha}}^k)|$ equals 1, we have

$$\rho_2 < \rho = \|\varphi_{s_{\overline{\alpha}}^k}(b_\alpha)\|_\rho < 1$$

(for any k). On the other hand, we have $|\beta(s_{\overline{\alpha}})| < 1$ for all $\alpha \ne \beta \in \Phi^+$ (whence, in particular, the definition of k makes sense), so we obtain

$$\|\varphi_t(b_\beta)\|_\rho = \max_{0\le j\le \operatorname{val}_p(\beta(t))}\bigl(\rho^{p^j} p^{j-\operatorname{val}_p(\beta(t))}\bigr) = \rho^{p^{\operatorname{val}_p(\beta(t))}} = \rho^{1/|\beta(s_{\overline{\alpha}})|^k} \le \rho^r$$

for all $\beta \in \Phi^+ \setminus \{\alpha\}$ by Lemma 4.4, with a choice of k such that $r \le 1/|\beta(s_{\overline{\alpha}})|^k$, and a choice of $p^{-|\beta(t)|} < \rho$ such that $\max_{0\le j\le \operatorname{val}_p(\beta(t))}\bigl(\rho^{p^j} p^{j-\operatorname{val}_p(\beta(t))}\bigr) = \rho^{p^{\operatorname{val}_p(\beta(t))}}$. □

Bounded rings. We write $\mathfrak{R}_0^b$ for the set of elements $x \in \mathfrak{R}_0$ such that $\lim_{\rho\to 1}\|x\|_{\rho,\rho}$ exists, and by $\mathfrak{R}_0^{\text{int}}$ the subset for which that limit is at most 1. By Proposition A.28 in [Zábrádi 2012], these are subrings of $\mathfrak{R}_0$. Moreover, since φ_t is norm-decreasing for any $1 \le_\alpha t$ (see (17)), these subrings are stable under the action of φ_t ($1 \le_\alpha t \in T_+$).

We put

$$\mathfrak{R}^b_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha) := \mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha) \cap \mathfrak{R}^b_0(N_0,\alpha),$$
$$\mathfrak{R}^{\mathrm{int}}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha) := \mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha) \cap \mathfrak{R}^{\mathrm{int}}_0(N_0,\alpha),$$

where the intersection is taken inside $\mathfrak{R}_0$ under the inclusion

$$\iota_{1,t} \colon \mathfrak{R}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha) \hookrightarrow \mathfrak{R}_0.$$

Hence

$$\mathcal{R}^b(N_1,\ell) := \varinjlim_t \mathfrak{R}^b_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)$$

and

$$\mathcal{R}^{\mathrm{int}}(N_1,\ell) := \varinjlim_t \mathfrak{R}^{\mathrm{int}}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)$$

are T_+-stable subrings of $\mathcal{R}(N_1,\ell)$ (the injective limit is taken with respect to the maps φ_{t_1,t_2} for $t_1 \leq_\alpha t_2 \in T_+$ as in the construction of $\mathcal{R}(N_1,\ell)$). Further, Lemma 4.6 shows that for any $t \in T_+$ and $x \in \mathfrak{R}_0$, we have

$$\lim_{\rho\to 1} \|x\|_\rho = \lim_{\rho\to 1} \|x\|_{q_t(r_t(\rho))}. \tag{20}$$

Indeed, we may use Lemma 4.6 in the context of $\mathfrak{R}_0$ the following way. The elements of $D_{[\rho_1,\rho_2]}(N_0,\alpha)$ are Cauchy sequences $(a_n\varphi_t(b_\alpha)^{-k_n})_{n\in\mathbb{N}}$ (in the norm $\max(\|\cdot\|_{\rho_1},\|\cdot\|_{\rho_2})$) with $a_n \in D_{[0,\rho_2]}(N_0)$ and $k_n \geq 0$. Since $\|\cdot\|_\rho$ is multiplicative for any $\rho_1 \leq \rho \leq \rho_2$ in $p^{\mathbb{Q}}$ and so is its restriction to $D(\varphi_t(N_0))$, we compute

$$\begin{aligned}
\left\|a_n\varphi_t(b_\alpha)^{-k_n}\right\|_\rho \rho^{\sum_{\beta\in\Phi^+}(p^{m(\beta,t)}-1)} &= \frac{\|a_n\|_\rho}{\|\varphi_t(b_\alpha)^{k_n}\|_\rho \rho^{-\sum_{\beta\in\Phi^+}(p^{m(\beta,t)}-1)}} \\
&\leq \frac{\|a_n\|_{q_t(r_t(\rho))}}{\|\varphi_t(b_\alpha)^{k_n}\|_{q_t(r_t(\rho))}} \\
&= \left\|a_n\varphi_t(b_\alpha)^{-k_n}\right\|_{q_t(r_t(\rho))} \\
&\leq \frac{\|a_n\|_\rho \rho^{-\sum_{\beta\in\Phi^+}(p^{m(\beta,t)}-1)}}{\|\varphi_t(b_\alpha)^{k_n}\|_\rho} \\
&\leq \left\|a_n\varphi_t(b_\alpha)^{-k_n}\right\|_\rho \rho^{-\sum_{\beta\in\Phi^+}(p^{m(\beta,t)}-1)}.
\end{aligned}$$

If $\rho \to 1$ and $n \to \infty$, we obtain (20). Combining this observation with (15), we obtain

$$\mathfrak{R}^b_0(N_0,\alpha) = \bigoplus_{n\in J(N_0/\varphi_t(N_0))} n\iota_{1,t}\big(\mathfrak{R}^b_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)\big),$$
$$\mathfrak{R}^{\mathrm{int}}_0(N_0,\alpha) = \bigoplus_{n\in J(N_0/\varphi_t(N_0))} n\iota_{1,t}\big(\mathfrak{R}^{\mathrm{int}}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha)\big).$$

So by a similar argument as for $\mathcal{R}(N_1, \ell)$, we also obtain

$$\mathcal{R}^b(N_1,\ell) = \bigoplus_{n\in J(N_0/\varphi_t(N_0))} n\varphi_t\big(\mathcal{R}^b(N_1,\ell)\big),$$
$$\mathcal{R}^{\mathrm{int}}(N_1,\ell) = \bigoplus_{n\in J(N_0/\varphi_t(N_0))} n\varphi_t\big(\mathcal{R}^{\mathrm{int}}(N_1,\ell)\big);$$

in other words, these are φ-rings over N_0 in the sense of Definition 2.9.

Remark 4.13. By [Zábrádi 2012, Lemma A.27], an element $\sum_{\boldsymbol{k}\in\mathbb{N}^{\Phi^+\setminus\{\alpha\}}\times\mathbb{Z}} c_{\boldsymbol{k}}\boldsymbol{b}^{\boldsymbol{k}}$ of $\mathcal{R}(N_1,\ell)$ (under the parametrization (18)) lies in $\mathcal{R}^b(N_1,\ell)$ (resp. in $\mathcal{R}^{\mathrm{int}}(N_1,\ell)$) if and only if $|c_{\boldsymbol{k}}|$ is bounded (resp. ≤ 1) for $\boldsymbol{k}\in\mathbb{Z}^{\{\alpha\}}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}$.

Relation with the completed Robba ring and overconvergent ring.

Lemma 4.14. *There exists a continuous (in the weak topology of $\Lambda_\ell(N_0)$) injective ring homomorphism $j_{\mathrm{int}}\colon\mathcal{R}^{\mathrm{int}}(N_1,\ell)\to\Lambda_\ell(N_0)$ respecting Laurent series expansions. The image of j_{int} is contained in $\mathbb{O}^\dagger_{\mathscr{E}}[\![N_1,\ell]\!]\subset\Lambda_\ell(N_0)$.*

Proof. We proceed in 3 steps. In Step 1 we construct a map $j_{\mathrm{int},0}=j_{\mathrm{int}|\mathfrak{R}_0^{\mathrm{int}}}\colon\mathfrak{R}_0^{\mathrm{int}}\to\Lambda_\ell(N_0)$ that is a priori continuous and o_K-linear. In Step 2 we show that $j_{\mathrm{int},0}$ is multiplicative, and hence a ring homomorphism. In Step 3 we extend it to $\mathcal{R}^{\mathrm{int}}(N_1,\ell)$ and show that the image lies in $\mathbb{O}^\dagger_{\mathscr{E}}[\![N_1,\ell]\!]\subset\mathbb{O}_{\mathscr{E}}[\![N_1,\ell]\!]=\Lambda_\ell(N_0)$.

Step 1. By Proposition 4.3 and Remark 4.13, we may write any element in $\mathfrak{R}_0^{\mathrm{int}}$ in a Laurent series expansion $\sum_{\boldsymbol{k}\in\mathbb{N}^{\Phi^+\setminus\{\alpha\}}\times\mathbb{Z}} c_{\boldsymbol{k}}\boldsymbol{b}^{\boldsymbol{k}}$ with coefficients $c_{\boldsymbol{k}}$ in o_K. So we may collect all the terms containing $b_\alpha^{k_\alpha}$ for some fixed k_α into an element of the Iwasawa algebra $\Lambda(N_1)$ to obtain an expansion $\sum_{n\in\mathbb{Z}} b_\alpha^n f_n$ with $f_n\in\Lambda(N_1)$. These power series satisfy the convergence property that there exists a real number $p^{-1}<\rho_1<1$ such that $\rho^n\|f_n\|_\rho\to 0$ as $|n|\to\infty$ for all $\rho_1<\rho<1$. In particular, if $n\to-\infty$, then $f_n\to 0$ in the compact topology of $\Lambda(N_1)$. Hence the sum $\sum_n b_\alpha^n f_n$ also converges in $\Lambda_\ell(N_0)$. In this way we have obtained a right $\Lambda(N_0)$-linear injective map $j_{\mathrm{int},0}\colon\mathfrak{R}_0^{\mathrm{int}}\to\Lambda_\ell(N_0)$.

Recall that the weak topology (see [Schneider and Venjakob 2010; Schneider and Vignéras 2011; Schneider et al. 2012] for instance) on $\Lambda_\ell(N_0)$ is defined by the open neighborhoods of 0 of the form $\mathcal{M}(r)=\mathcal{M}_\ell(N_0)^r+\mathcal{M}(N_0)^r$, where $\mathcal{M}_\ell(N_0)=\Lambda_\ell(N_0)\mathcal{M}(N_1)$ denotes the maximal ideal of $\Lambda_\ell(N_0)$ and $\mathcal{M}(N_i)$ denotes the maximal ideal of $\Lambda(N_i)\subseteq\Lambda_\ell(N_0)$ $(i=0,1)$. For any fixed $p^{-1}<\rho_1<\rho<1$, the preimage of $\mathcal{M}(r)$ in $\mathfrak{R}_0^{\mathrm{int}}\cap D_{[\rho_1,1)}(N_1,\alpha)$ contains the open ball

$$\{x\mid \|x\|_\rho<p^{-r}\}.$$

Indeed, if $x=\sum_{n\in\mathbb{Z}} b_\alpha^n f_n$, then for any $n<0$, we have $\|f_n\|_\rho<p^{-r}$, and hence $f_n\in\mathcal{M}(N_1)^r$ and $b_\alpha^n f_n\in\mathcal{M}_\ell(N_0)$. On the other hand, the positive part

$\sum_{n\geq 0} b_\alpha^n f_n$ lies in $\Lambda(N_0)$ and has ρ-norm smaller than p^{-r}, and therefore lies in $\mathcal{M}(N_0)^r$. Hence the continuity.

Step 2. Now by the continuity and linearity of $j_{\mathrm{int},0}$, it suffices to show that it is multiplicative on monomials $\boldsymbol{b^k}$. Moreover, each monomial is a linear combination of elements of the form $b_\alpha^n g$ with $g \in N_0$. In order to expand the product $(b_\alpha^{n_1} g_1)(b_\alpha^{n_2} g_2)$ into a skew Laurent series, it suffices to expand $g_1 b_\alpha^{n_2}$ with $n_2 < 0$. However, if $g_1 b_\alpha^{n_2} = \sum_n b_\alpha^n h_n$ is the expansion in $\mathcal{R}^{\mathrm{int}}(N_1,\ell)$, then $\sum_{|n|<n_0} b_\alpha^n h_n b_\alpha^{-n_2}$ tends to g_1 (as $n_0 \to +\infty$) in the topology of $\mathfrak{R}_0^{\mathrm{int}}$ (induced by the norms), and hence also in the weak topology. Therefore the expansion in $\Lambda_\ell(N_0)$ is also $g_1 b_\alpha^{n_2} = \sum_n b_\alpha^n h_n$. So the above constructed map $j_{\mathrm{int},0}$ is indeed a ring homomorphism as claimed.

Step 3. Finally, take an element $x \in \mathcal{R}^{\mathrm{int}}(N_1,\ell)$. There exists an element $1 \leq_\alpha t \in T_+$ such that $\varphi_t(x)$ lies in the image of the composite map

$$\mathfrak{R}^{\mathrm{int}}_{0,r_t(\cdot)}(\varphi_t(N_0),\alpha) \hookrightarrow \mathfrak{R}_0^{\mathrm{int}}(N_0,\alpha) \hookrightarrow \mathcal{R}^{\mathrm{int}}(N_1,\ell),$$

where the first arrow is induced by the inclusion $\varphi_t(N_0) \subseteq N_0$. Now if we reduce $j_{\mathrm{int},0}(\varphi_t(x)) \in \Lambda_\ell(N_0)$ modulo the ideal generated by $N_l - 1$ for some integer $l \geq 1$, then we obtain an element in $\varphi_t(\mathcal{O}^\dagger_{\mathcal{E}}[N_1/N_l,\ell])$. Indeed, $\varphi_t(\mathcal{O}_{\mathcal{E}}[N_1/N_l,\ell])$ is a closed subspace in $\mathcal{O}_{\mathcal{E}}[N_1/N_l,\ell]$ and all the monomials $j_{\mathrm{int},0}(\varphi_t(\boldsymbol{b^k}))$ map into this subspace under the reduction modulo $(N_l - 1)$. Hence the image lies in $\varphi_t(\mathcal{O}_{\mathcal{E}}[N_1/N_l,\ell])$. By the convergence property of elements in $\mathfrak{R}_0^{\mathrm{int}}$, we may expand

$$\varphi_t(x) = \sum_{n\in\mathbb{Z}} b_\alpha^n f_n$$

with $f_n \in \Lambda(N_1)$ and $\rho^n\|f_n\|_\rho \to 0$ as $n \to \infty$ for all $\rho_1 < \rho < 1$ and a fixed $p^{-1} < \rho_1 < 1$ depending on x. Since the reduction map $\Lambda(N_1) \to o[N_1/N_l]$ is continuous in the ρ-norm, we obtain that the reduction of $j_{\mathrm{int},0}(\varphi_t(x))$ modulo $(N_l - 1)$ also lies in $\mathcal{O}^\dagger_{\mathcal{E}}[N_1/N_l,\ell]$. Hence we have $j_{\mathrm{int},0}(\varphi_t(x)) \pmod{N_l - 1} \in \varphi_t(\mathcal{O}^\dagger_{\mathcal{E}}[N_1/N_l,\ell]) = \varphi_t(\mathcal{O}_{\mathcal{E}}[N_1/N_l,\ell]) \cap \mathcal{O}^\dagger_{\mathcal{E}}[N_1/N_l,\ell]$. Taking the limit, we see (using (7)) that $j_{\mathrm{int},0}(\varphi_t(x))$ lies in

$$\varprojlim_l \varphi_t(\mathcal{O}^\dagger_{\mathcal{E}}[N_1/N_l,\ell]) = \varphi_t(\mathcal{O}^\dagger_{\mathcal{E}}[\![N_1,\ell]\!]),$$

so we put $j_{\mathrm{int}}(x) := \varphi_t^{-1}\big(j_{\mathrm{int},0}(\varphi_t(x))\big)$. This extends the ring homomorphism $j_{\mathrm{int},0}$ to a continuous ring homomorphism $j_{\mathrm{int}}\colon \mathcal{R}^{\mathrm{int}}(N_1,\ell) \hookrightarrow \mathcal{O}^\dagger_{\mathcal{E}}[\![N_1,\ell]\!] \subset \Lambda_\ell(N_0)$, by Lemma 4.7. This map is T_+-equivariant, as it respects power series expansions. □

Now the following proposition compares $\mathcal{R}(N_1,\ell)$ with the previous construction $\mathcal{R}[\![N_1,\ell]\!]$.

Proposition 4.15. *There exists a natural T_+-equivariant ring homomorphism*

$$j\colon \mathcal{R}(N_1,\ell) \to \mathcal{R}[[N_1,\ell]]$$

with dense image.

Proof. At first we construct the map $j_0 = j_{|\mathfrak{R}_0}$ on $\mathfrak{R}_0 \subset \mathcal{R}(N_1,\ell)$ with dense image. We are going to show that for any open characteristic subgroup $H \leq N_1$, we have an isomorphism $\mathfrak{R}_0/\mathfrak{R}_0(H-1) \cong \mathcal{R}[N_1/H,\ell]$. Note that N_1 being a compact p-adic Lie group, N_1 has a system of neighborhoods of 1 consisting of open uniform characteristic subgroups (in fact N_1 is uniform — since so is N_0 by assumption — and one can take repeatedly the Frattini subgroups of N_1 that are characteristic subgroups, that is, stable under all the continuous automorphisms of N_1). So we may assume without loss of generality that H is uniform with topological generators $h_1, h_2, \dots, h_d$ with $d = \dim N_1$ as a p-adic Lie group.

Under the parametrization in Proposition 4.3, the elements of $\mathfrak{R}_0$ can be written as power series $\sum_{n\in\mathbb{Z}} b_\alpha^n f_n$ with $f_n \in D(N_1, K)$ and the convergence property that there exists a real number $\rho_1 < 1$ such that $\rho^n \| f_n \|_\rho \to 0$ (as $|n| \to \infty$) for all $\rho_1 \leq \rho < 1$. Now we have $D(N_1,K) = \bigoplus_{u\in J(N_1/H)} uD(H,K)$. Hence the right ideal $D(N_1,K)(H-1)$ in $D(N_1,K)$ is generated by the elements $h_i - 1$ for $1 \leq i \leq d$ and it is the kernel of the natural projection $\pi_H\colon D(N_1,K) \to D(N_1/H) = K[N_1/H]$. This quotient map factors through the inclusion $D(N_1,K) \hookrightarrow D_{[0,\rho]}(N_1,K)$ for any $p^{-1} < \rho < 1$. Hence $\rho^n \|\pi_H(f_n)\| \to 0$, where $\|x\| := \max_u |x_u|$ with $x = \sum_{u\in N_1/H} x_u u$ and $x_u \in K$. Therefore we obtain a map

$$\pi_H\colon \mathfrak{R}_0 \to \mathcal{R}[N_1/H,\ell] = \bigoplus_{u\in N_1/H} \mathcal{R}u, \quad \sum_{n\in\mathbb{Z}} b_\alpha^n f_n \mapsto \sum_{u\in N_1/H} \sum_{n\in\mathbb{Z}} \pi_H(f_n)_u T^n u.$$

A priori this map is only known to be K-linear, continuous, and surjective between topological K-vector spaces. So for the multiplicativity, it suffices to show that $\pi_H(\boldsymbol{b}^{\boldsymbol{k}_1}\boldsymbol{b}^{\boldsymbol{k}_2}) = \pi_H(\boldsymbol{b}^{\boldsymbol{k}_1})\pi_H(\boldsymbol{b}^{\boldsymbol{k}_2})$ for monomials $\boldsymbol{b}^{\boldsymbol{k}_i}$ with $\boldsymbol{k}_i \in \mathbb{N}\times\mathbb{Z}^d$ $(i = 1,2)$. On the other hand, these monomials are contained in the subring $\mathfrak{R}_0^{\mathrm{int}}$. By Lemma 4.14, we have a commutative diagram

$$\begin{tikzcd}
\mathfrak{R}_0^{\mathrm{int}} \arrow[r] \arrow[d, "\pi_{H,\mathcal{O}_{\mathcal{E}}^{\dagger}[[N_1,\ell]]}"'] & \mathfrak{R}_0 \arrow[d, "\pi_H"'] \\
\mathcal{O}_{\mathcal{E}}^{\dagger}[N_1/H,\ell] \arrow[r] & \mathcal{R}[N_1/H,\ell]
\end{tikzcd}$$

of o-modules such that all the maps are ring homomorphisms except possibly for π_H. However, from the commutativity of the diagram, it follows that π_H is also multiplicative on monomials and therefore a ring homomorphism. By taking the projective limit of maps π_H, we obtain a ring homomorphism $j_0\colon \mathfrak{R}_0 \to \mathcal{R}[\![N_1,\ell]\!]$ with dense image and extending $j_{\mathrm{int},0}\colon \mathfrak{R}_0^{\mathrm{int}} \hookrightarrow \mathcal{O}^{\dagger}_{\mathcal{E}}$.

Finally, the homomorphism j_0 is extended to $\mathcal{R}(N_1,\ell)$ as in the proof of Lemma 4.14. The T_+-equivariance is clear on monomials by Lemma 4.14 and follows in general from the continuity and linearity. □

Remark 4.16. The map $j\colon \mathcal{R}(N_1,\ell) \to \mathcal{R}[\![N_1,\ell]\!]$ constructed above is not injective in general. Indeed, for any root $\beta \neq \alpha$ in Φ^+, the element $\log(n_\beta) = \log(1+b_\beta)$ lies in $D(N_1) \subset \mathcal{R}(N_1,\ell)$. It is easy to see that $\log(1+b_\beta)$ is divisible by $\varphi^r(b_\beta)$ for any nonnegative integer r. Indeed, we clearly have $b_\beta \mid \log(1+b_\beta)$. Applying φ^r on the both sides of the divisibility, we obtain

$$\varphi^r(b_\beta) \mid \varphi^r(\log(1+b_\beta)) = \log(1+b_\beta)^{p^{rm_\beta}} = p^{rm_\beta}\log(1+b_\beta) \mid \log(1+b_\beta),$$

as p^{rm_β} is invertible in $\mathcal{R}$. Therefore $\log(1+b_\beta)$ lies in the kernel of π_H for all $H = N_r$, and hence also in the kernel of j.

Remark 4.17. Via the inclusion $\mathcal{O}^{\dagger}_{\mathcal{E}} \subseteq \mathcal{R}$, we also have $\mathcal{O}^{\dagger}_{\mathcal{E}}[\![N_1,\ell]\!] \subseteq \mathcal{R}[\![N_1,\ell]\!]$. However, if $N_1 \neq 1$, then we have $j_{\mathrm{int}}(\mathcal{R}^{\mathrm{int}}(N_1,\ell)) \neq j(\mathcal{R}(N_1,\ell)) \cap \mathcal{O}^{\dagger}_{\mathcal{E}}[\![N_1,\ell]\!] \subset \mathcal{R}[\![N_1,\ell]\!]$.

Proof. Assume $N_1 \neq 1$ so that we have a positive root $\beta \neq \alpha \in \Phi^+$. We proceed in three steps. In Step 1, we construct an element $x \in \mathcal{R}(N_1,\ell)$ with several properties. In Step 2, we show that $j(x)$ lies in $\mathcal{O}^{\dagger}_{\mathcal{E}}[\![N_1,\ell]\!] \subset \mathcal{R}[\![N_1,\ell]\!]$. In Step 3, we prove that $j(x)$ does not lie in $j_{\mathrm{int}}(\mathcal{R}^{\mathrm{int}}(N_1,\ell))$. The other inclusion, $j_{\mathrm{int}}(\mathcal{R}^{\mathrm{int}}(N_1,\ell)) \subset j(\mathcal{R}(N_1,\ell)) \cap \mathcal{O}^{\dagger}_{\mathcal{E}}[\![N_1,\ell]\!]$, is obvious.

Step 1. We denote by $s_n := \sum_{i=1}^{n}(-1)^{i+1}(b_\beta^i/i)$ the n-th estimating sum of $\log(1+b_\beta) \in \mathcal{R}(N_1,\ell)$. Note that $k_n := [\log_p n]$ is the smallest positive integer such that

$$p^{k_n}s_n \in \mathbb{Z}_p[N_{\beta,0}] \subseteq \mathcal{R}(N_1,\ell), \tag{21}$$

where $[\,\cdot\,]$ denotes the integer part of a real number. We further choose a sequence of real numbers $p^{-1} < \rho_1 < \cdots < \rho_n < \cdots < 1$ in $p^{\mathbb{Q}}$ such that $\lim_{n\to\infty}\rho_n = 1$. Now for any fixed positive integer n, let i_n be the smallest positive integer satisfying the properties

$$\begin{gathered}
\log_{\rho_{n-1}}(\|p^{k_{i_{n-1}}}\log(1+b_\beta)\|_{\rho_{n-1}}) + 1 < \log_{\rho_n}\big(\|p^{k_{i_n}}\log(1+b_\beta)\|_{\rho_n}\big),\\
\frac{\|\log(1+b_\beta)\|_{\rho_n}}{p^n} > \|\log(1+b_\beta) - s_{i_n}\|_{\rho_n},\\
p^{k_{i_n}/2} > \|\log(1+b_\beta)\|_{\rho_n},\\
\|\varphi^i(\log(1+b_\beta))\|_{\rho_j} > \|\varphi^i(\log(1+b_\beta) - s_{i_n})\|_{\rho_j}
\end{gathered} \tag{22}$$

for all $1 \leq i,\ j \leq n$. Such an i_n exists, as for any fixed $1 \leq i,\ j \leq n$, we have

$\lim_{k\to\infty}\|\varphi^i(\log(1+b_\beta)-s_k)\|_{\rho_j}=0$. The first condition in (22) makes the definition of i_n inductive. As a consequence, we have $\|\log(1+b_\beta)\|_{\rho_n}=\|s_{i_n}\|_{\rho_n}$ by the ultrametric inequality. Now define $j_n\in\mathbb{Z}$ such that

$$\rho_n^{j_n+1}<\frac{\|s_{i_n}\|_{\rho_n}}{p^{k_{i_n}}}=\bigl\|p^{k_{i_n}}s_{i_n}\bigr\|_{\rho_n}=\bigl\|p^{k_{i_n}}\log(1+b_\beta)\bigr\|_{\rho_n}\le\rho_n^{j_n}. \tag{23}$$

(In other words, $j_n=[\log_{\rho_n}(\|p^{k_{i_n}}\log(1+b_\beta)\|_{\rho_n})]$.) By (21), we have $j_n\ge 0$. By the first condition in (22), the sequence $(j_n)_n$ is strictly increasing: $j_{n-1}<j_n$ for all $n>1$. On the other hand,

$$(-1)^{p^{k_{i_n}}}b_\beta^{p^{k_{i_n}}}$$

is a summand in $p^{k_{i_n}}s_{i_n}$; therefore we have

$$\rho_n^{p^{k_{i_n}}}\le\|p^{k_{i_n}}s_{i_n}\|_{\rho_n}\le\rho_n^{j_n},$$

whence

$$j_n\le p^{k_{i_n}}\le i_n. \tag{24}$$

Put $x:=\sum_{n=1}^{\infty}p^{k_{i_n}}(\log(1+b_\beta)-s_{i_n})b_\alpha^{-j_n}$. Our goal in this step is to show that the sum x converges in $\mathfrak{R}_0(N_0,\alpha)\subset\mathcal{R}(N_1,\ell)$. For this it suffices to verify that for any fixed $k\ge 1$, we have $\|p^{k_{i_n}}(\log(1+b_\beta)-s_{i_n})b_\alpha^{-j_n}\|_{\rho_k}\to 0$ as $n\to\infty$. In the power series expansion of $\log(1+b_\beta)-s_{i_n}$, all the terms have degree $>i_n\ge j_n$ by (24). Therefore in the power series expansion of x all the terms have positive degree. In particular, for $k<n$ we have $\|y\|_{\rho_k}\le\|y\|_{\rho_n}$ whenever y is a monomial in the expansion of x. By (22) and (23) we obtain

$$\begin{aligned}\bigl\|p^{k_{i_n}}(\log(1+b_\beta)-s_{i_n})b_\alpha^{-j_n}\bigr\|_{\rho_k}&\le\bigl\|p^{k_{i_n}}(\log(1+b_\beta)-s_{i_n})b_\alpha^{-j_n}\bigr\|_{\rho_n}\\&<\frac{\|p^{k_{i_n}}\log(1+b_\beta)\|_{\rho_n}}{p^n}\rho_n^{-j_n}\le\frac{1}{p^n}\end{aligned}$$

for $k<n$. Hence we have $x\in\mathfrak{R}_0(N_0,\alpha)\subset\mathcal{R}(N_1,\ell)$.

Step 2. By Remark 4.16, $\log(1+b_\beta)$ lies in the kernel of π_H for all open normal subgroups $H\le N_1$. Hence, by the continuity of π_H, we obtain $\pi_H(x)=\sum_{n=1}^{\infty}\pi_H(-p^{k_{i_n}}s_{i_n}b_\alpha^{-j_n})\in\mathcal{O}_{\mathcal{E}}^{\dagger}[N_1/H,\ell]\subseteq\mathcal{R}[N_1/H,\ell]$, as we have $-p^{k_{i_n}}s_{i_n}\in\mathbb{Z}_p[N_1]$ and $\mathcal{O}_{\mathcal{E}}^{\dagger}$ is closed in $\mathcal{R}$.

Step 3. Assume finally that $j_{\mathrm{int}}(z)=j(x)$ for some $z\in\mathcal{R}^{\mathrm{int}}(N_1,\ell)$. Both z and $j(x)\in\mathcal{O}_{\mathcal{E}}^{\dagger}[\![N_1,\ell]\!]\subset\mathcal{O}_{\mathcal{E}}[\![N_1,\ell]\!]$ have a power series expansion. By the injectivity of j_{int}, these expansions are equal. Hence, put $z=\sum_{\boldsymbol{k}\in\mathbb{Z}\times\mathbb{N}^{\Phi^+\setminus\{\alpha\}}}d_{\boldsymbol{k}}\boldsymbol{b}^{\boldsymbol{k}}$ with $d_{\boldsymbol{k}}\in\mathbb{Z}_p$. By the definition of $\mathcal{R}^{\mathrm{int}}(N_1,\ell)$, there exists an element $t\in T_+$ such that $\varphi_t(z)$ lies in $\mathfrak{R}_0^{\mathrm{int}}$. This means that there exists a positive integer K_0 such that for all fixed $k \geq K_0$ and $\varepsilon > 0$ we have $\|\varphi_t(d_{\boldsymbol{k}}\boldsymbol{b}^{\boldsymbol{k}})\|_{\rho_k} < \varepsilon$ for all but finitely many $\boldsymbol{k} \in \mathbb{Z} \times \mathbb{N}^{\Phi^+ \setminus \{\alpha\}}$. In particular, for any fixed $k \geq K_0$ we have

$$\left\|\varphi_t(-p^{k_{i_n}} s_{i_n} b_\alpha^{-j_n})\right\|_{\rho_k} < \varepsilon$$

for all but finitely many positive integers n since the sequence j_n is strictly increasing by construction; therefore the terms in $x = \sum_{n=1}^{\infty} p^{k_{i_n}}(\log(1+b_\beta) - s_{i_n})b_\alpha^{-j_n}$ cannot cancel each other. Now we clearly have $\|\varphi_t(b_\alpha)\|_{\rho_k} \leq \rho_k$. On the other hand, we compute (for $n > \max(k, m(\beta, t))$ large enough)

$$\begin{aligned}\|\varphi_t(-p^{k_{i_n}} s_{i_n})\|_{\rho_k} &= \frac{\|\varphi^{m(\beta,t)}(s_{i_n})\|_{\rho_k}}{p^{k_{i_n}}} \\ &= \frac{\|\varphi^{m(\beta,t)}(\log(1+b_\beta))\|_{\rho_k}}{p^{k_{i_n}}} \\ &= \frac{\|\log(1+b_\beta)\|_{\rho_k}}{p^{m(\beta,t)+k_{i_n}}}.\end{aligned}$$

Hence we obtain

$$\begin{aligned}\varepsilon &> \|\varphi_t(-p^{k_{i_n}} s_{i_n} b_\alpha^{-j_n})\|_{\rho_k} \\ &\geq \frac{\|\log(1+b_\beta)\|_{\rho_k}}{p^{m(\beta,t)+k_{i_n}} \rho_k^{j_n}} \\ &> \frac{\rho_k \|\log(1+b_\beta)\|_{\rho_k}}{p^{m(\beta,t)+k_{i_n}} \|p^{k_{i_n}} \log(1+b_\beta)\|_{\rho_n}^{\log_{\rho_n} \rho_k}} \\ &= \frac{\rho_k \|\log(1+b_\beta)\|_{\rho_k}}{p^{m(\beta,t)}} \frac{p^{k_{i_n}(\log_{\rho_n} \rho_k - 1)}}{\|\log(1+b_\beta)\|_{\rho_n}^{\log_{\rho_n} \rho_k}} \\ &> \frac{\rho_k \|\log(1+b_\beta)\|_{\rho_k}}{p^{m(\beta,t)}} p^{k_{i_n}(\frac{1}{2}\log_{\rho_n} \rho_k - 1)},\end{aligned}$$

using (22) and (23). This is a contradiction, as the right side tends to ∞ as $n \to \infty$. Therefore $j(x)$ is not in the image of j_{int} as claimed. □

Remark 4.18. The elements of $\mathcal{R}[\![N_1, \ell]\!]$ cannot be expanded as a skew Laurent series of the form $\sum_{\boldsymbol{k} \in \mathbb{Z}^{\Phi^+}} d_{\boldsymbol{k}} \boldsymbol{b}^{\boldsymbol{k}}$ in general. Indeed, the sum

$$\sum_{n=1}^{\infty} \varphi^n(b_\beta)/p^{2^n} = \sum_{n=1}^{\infty}((b_\beta + 1)^{p^n} - 1)/p^{2^n}$$

converges in $\mathcal{R}[\![N_1, \ell]\!]$ for any simple root $\beta \neq \alpha$ but does not have a skew Laurent-series expansion, as the coefficient of b_β in its expansion would be the nonconvergent sum $\sum_{n=1}^{\infty} p^{n-2^n}$.

We end this section with a diagram showing all the rings constructed.

$$\begin{tikzcd}
\mathcal{O}_{\mathcal{E}} \arrow[rrr, hook] & & & \mathcal{O}_{\mathcal{E}}[\![N_1,\ell]\!] = \Lambda_\ell(N_0) \\
\mathcal{O}_{\mathcal{E}}^\dagger \arrow[u, hook] \arrow[r, hook] \arrow[d, hook] & \mathfrak{R}_0^{\mathrm{int}}(N_0,\alpha) \arrow[r, hook] \arrow[d, hook] & \mathcal{R}^{\mathrm{int}}(N_1,\ell) \arrow[r, hook, "j_{\mathrm{int}}"] \arrow[d, hook] & \mathcal{O}_{\mathcal{E}}^\dagger[\![N_1,\ell]\!] \arrow[u, hook] \arrow[d, hook] \\
\mathcal{E}^\dagger \arrow[r, hook] \arrow[d, hook] & \mathfrak{R}_0^{bd}(N_0,\alpha) \arrow[r, hook] \arrow[d, hook] & \mathcal{R}^{bd}(N_1,\ell) \arrow[r, hook, "j_{\mathrm{int}}\otimes_{\mathbb{Z}_p}\mathbb{Q}_p"] \arrow[d, hook] & \mathcal{E}^\dagger[\![N_1,\ell]\!] \arrow[d, hook] \\
\mathcal{R} \arrow[r, hook] & \mathfrak{R}_0(N_0,\alpha) \arrow[r, hook] & \mathcal{R}(N_1,\ell) \arrow[r, "j"] & \mathcal{R}[\![N_1,\ell]\!]
\end{tikzcd} \qquad (25)$$

Here $\mathcal{R}(N_1,\ell)$ consists of Laurent series $\sum_{\boldsymbol{k}} c_{\boldsymbol{k}} \boldsymbol{b}^{\boldsymbol{k}}$ with $c_{\boldsymbol{k}} \in K$ that converge on the open annulus of the form

$$\bigl\{\rho_2 < |z_\alpha| < 1,\ |z_\beta| \le |z_\alpha|^r \text{ for } \beta \in \Phi^+ \setminus \{\alpha\}\bigr\}$$

for some $0 < \rho_2 < 1$ and $1 \le r \in \mathbb{Z}$. The elements of $\mathfrak{R}_0(N_0,\alpha)$ are exactly those for which we can take $r = 1$. Their analogous integral (resp. bounded) versions consist of those Laurent series having the same convergence condition for which $c_{\boldsymbol{k}} \in o_K$ for all $\boldsymbol{k} \in \mathbb{Z}^{\{\alpha\}} \times \mathbb{N}^{\Phi^+\setminus\{\alpha\}}$ (resp. for which $\{c_{\boldsymbol{k}} \mid \boldsymbol{k} \in \mathbb{Z}^{\{\alpha\}} \times \mathbb{N}^{\Phi^+\setminus\{\alpha\}}\} \subset K$ bounded).

4.1. ***Towards an equivalence of categories for overconvergent and Robba rings.*** Propositions 3.1 and 3.7 apply in both the cases $R = \mathcal{O}_{\mathcal{E}}$ and $R = \mathcal{O}_{\mathcal{E}}^\dagger$. In both cases the category $\mathfrak{M}(R,\varphi)$ is the category of *étale* φ-modules over R. By the main result of [Cherbonnier and Colmez 1998] (see also [Kedlaya 2012]), we also have an equivalence of categories between finite free étale (φ,Γ)-modules over $\mathcal{O}_{\mathcal{E}}^\dagger$ and finite free étale (φ,Γ)-modules over $\mathcal{O}_{\mathcal{E}}$ given by the base change $\mathcal{O}_{\mathcal{E}} \otimes_{\mathcal{O}_{\mathcal{E}}^\dagger} \cdot$. On the other hand, T_ℓ acts by automorphisms on an object D in $\mathfrak{M}(\mathcal{O}_{\mathcal{E}}, T_+)$ and also on an object $D^\dagger$ in $\mathfrak{M}(\mathcal{O}_{\mathcal{E}}^\dagger, T_+)$. Since automorphisms correspond to automorphisms in an equivalence of categories, we obtain:

Proposition 4.19. *The functors*

$$\mathcal{O}_{\mathcal{E}} \otimes_{\mathcal{O}_{\mathcal{E}}^\dagger} \cdot : \mathfrak{M}(\mathcal{O}_{\mathcal{E}}^\dagger, T_+) \to \mathfrak{M}(\mathcal{O}_{\mathcal{E}}, T_+), \quad \cdot^\dagger : \mathfrak{M}(\mathcal{O}_{\mathcal{E}}, T_+) \to \mathfrak{M}(\mathcal{O}_{\mathcal{E}}^\dagger, T_+)$$

are quasi-inverse equivalences of categories.

For the Robba ring $\mathcal{R}$, étaleness is stronger than what we assumed for a module $D_{\mathrm{rig}}^\dagger$ to belong to $\mathfrak{M}(\mathcal{R},\varphi)$. The category $\mathfrak{M}(\mathcal{R},\varphi)$ is just the category of φ-modules over the Robba ring. Recall that an object $D_{\mathrm{rig}}^\dagger$ in $\mathfrak{M}(\mathcal{R},\varphi)$ is *étale* (or unit-root, or pure of slope zero) whenever it comes from an overconvergent étale φ-module $D^\dagger$ over the ring of "overconvergent" power series $\mathcal{O}_{\mathcal{E}}^\dagger$ by base extension. We denote by $\mathfrak{M}^0(\mathcal{R}, \varphi)$ the category of étale φ-modules over the Robba ring $\mathcal{R}$. We consequently define the categories $\mathfrak{M}^0(\mathcal{R}, T_+)$, $\mathfrak{M}^0(\mathcal{R}[\![N_1, \ell]\!], \varphi)$, and $\mathfrak{M}^0(\mathcal{R}[\![N_1, \ell]\!], T_+)$ as the full subcategory of étale objects in the corresponding categories without superscript 0. Via the equivalence of categories 3.7, étale objects correspond to each other. Combining this observation with the main result of [Berger 2002] leads to:

Corollary 4.20. *We have a commutative diagram of equivalences of categories*

$$\begin{tikzcd}
\mathfrak{M}^0(\mathcal{R}, T_+) \arrow[d] & \mathfrak{M}(\mathcal{E}^\dagger, T_+) \arrow[l] \arrow[r] \arrow[d] & \mathfrak{M}(\mathcal{E}, T_+) \arrow[d] \\
\mathfrak{M}^0(\mathcal{R}[\![N_1, \ell]\!], T_+) & \mathfrak{M}(\mathcal{E}^\dagger[\![N_1, \ell]\!], T_+) \arrow[l] \arrow[r] & \mathfrak{M}(\mathcal{E}[\![N_1, \ell]\!], T_+).
\end{tikzcd}$$

Proof. The left horizontal arrows are also equivalences of categories by [Berger 2002], noting that T_ℓ acts via automorphisms on both types of objects in the upper row. □

Remark 4.21. The category $\mathfrak{M}^0(\mathcal{R}[\![N_1, \ell]\!], T_+)$ of étale T_+-modules is embedded into the bigger category $\mathfrak{M}(\mathcal{R}[\![N_1, \ell]\!], T_+)$. So we may speak of *trianguline* objects in $\mathfrak{M}^0(\mathcal{R}[\![N_1, \ell]\!], T_+)$ as in the classical case (see for instance [Berger 2011]). Indeed, we call an object $M_{\mathrm{rig}}^\dagger$ in $\mathfrak{M}^0(\mathcal{R}[\![N_1, \ell]\!], T_+)$ trianguline if it becomes a successive extension of objects in $\mathfrak{M}(\mathcal{R}[\![N_1, \ell]\!], T_+)$ of rank 1 after a finite base extension $L \otimes_K \cdot$. It is clear that trianguline objects correspond to trianguline objects via the first vertical arrow in Corollary 4.20.

Remark 4.22. It would be interesting to construct a noncommutative version of the "big" rings $\tilde{A}_{\mathbb{Q}_p}$ and $\tilde{A}_{\mathbb{Q}_p}^\dagger$ in [Kedlaya 2012] and generalize (the proofs of) Theorems 2.3.5, 2.4.5, and 2.6.2 to this noncommutative setting. For this, one would need a generalization for results in the present paper to base fields other than $\mathbb{Q}_p$.

Remark 4.23. Since we have the natural inclusions $\mathcal{O}_{\mathcal{E}}^\dagger \hookrightarrow \mathcal{R}^{\mathrm{int}}(N_1, \ell) \hookrightarrow \mathcal{O}_{\mathcal{E}}^\dagger[\![N_1, \ell]\!]$ we have a fully faithful functor

$$\Theta := \bigl(\mathcal{R}^{\mathrm{int}}(N_1, \ell) \otimes_{\mathcal{O}_{\mathcal{E}}^\dagger} \cdot\bigr) \circ \bigl(\mathcal{O}_{\mathcal{E}}^\dagger \otimes_{\ell, \mathcal{O}_{\mathcal{E}}^\dagger[[N_1, \ell]]} \cdot\bigr) \circ \bigl(\mathcal{O}_{\mathcal{E}}^\dagger[\![N_1, \ell]\!] \otimes_{\mathcal{R}^{\mathrm{int}}(N_1, \ell)} \cdot\bigr)$$

from the category $\mathfrak{M}(\mathcal{R}^{\mathrm{int}}(N_1, \ell), T_+)$ to itself. Whether or not it is essentially surjective (or equivalently, naturally isomorphic to the identity functor) is not clear. However, we have $\Theta \cong \Theta \circ \Theta$ naturally.

Proof. The faithfulness is clear since the objects in the category $\mathfrak{M}(\mathcal{R}^{\mathrm{int}}(N_1, \ell), T_+)$ are free modules, the maps $\mathcal{O}_{\mathcal{E}}^\dagger \hookrightarrow \mathcal{R}^{\mathrm{int}}(N_1, \ell) \hookrightarrow \mathcal{O}_{\mathcal{E}}^\dagger[\![N_1, \ell]\!]$ are injective, and the functor

$$\mathcal{O}^\dagger_{\mathcal{E}} \otimes_{\ell,\mathcal{O}^\dagger_{\mathcal{E}}[[N_1,\ell]]} \bullet$$

in the middle is an equivalence of categories by Proposition 3.7. The assertion $\Theta \cong \Theta \circ \Theta$ is also clear by Proposition 3.7. For the fullyness let $f \colon \Theta(\mathcal{M}_1) \to \Theta(\mathcal{M}_2)$ be a morphism in $\mathfrak{M}(\mathcal{R}^{\mathrm{int}}(N_1, \ell))$. Then we have $\Theta(f - \Theta(f)) = 0$, and by the faithfulness of Θ, we obtain $f = \Theta(f)$. □

Acknowledgements

This research was supported by the Hungarian OTKA grant K-100291. The proof of Proposition 3.1 is a direct generalization of Theorem 8.20 in [Schneider et al. 2012] and grew out of related discussions with Marie-France Vigneras and Peter Schneider. I express my gratitude to both of them for allowing me to include this proof here and for other valuable discussions. I am especially indebted to Peter Schneider, from whom I learned most of what I know about p-adic representation theory. I would like to thank Torsten Schoeneberg for reading the manuscript and for his comments. I also benefited from enlightening conversations with Kiran Kedlaya. Finally, I thank the referee for a careful reading of the manuscript.

References

[Berger 2002] L. Berger, "Représentations p-adiques et équations différentielles", *Invent. Math.* **148**:2 (2002), 219–284. MR 2004a:14022 Zbl 1113.14016

[Berger 2011] L. Berger, "Trianguline representations", *Bull. Lond. Math. Soc.* **43**:4 (2011), 619–635. MR 2012h:11079 Zbl 1236.11053

[Berger 2013] L. Berger, "Multivariable Lubin–Tate (φ, Γ)-modules and filtered φ-modules", preprint, 2013, Available at http://perso.ens-lyon.fr/laurent.berger/articles/article23.pdf.

[Breuil 2004] C. Breuil, "Invariant L et série spéciale p-adique", *Ann. Sci. École Norm. Sup.* (4) **37**:4 (2004), 559–610. MR 2005j:11039 Zbl 1113.14016

[Cherbonnier and Colmez 1998] F. Cherbonnier and P. Colmez, "Représentations p-adiques surconvergentes", *Invent. Math.* **133**:3 (1998), 581–611. MR 2000d:11146 Zbl 0928.11051

[Coates et al. 2005] J. Coates, T. Fukaya, K. Kato, R. Sujatha, and O. Venjakob, "The GL_2 main conjecture for elliptic curves without complex multiplication", *Publ. Math. Inst. Hautes Études Sci.* 101 (2005), 163–208. MR 2007b:11172 Zbl 1108.11081

[Colmez 2010a] P. Colmez, "La série principale unitaire de $\mathrm{GL}_2(\mathbb{Q}_p)$", pp. 213–262 in *Représentations p-adiques de groupes p-adiques, II: Représentations de* $\mathrm{GL}_2(\mathbb{Q}_p)$ *et* (ϕ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Société Mathématique de France, Paris, 2010. MR 2011g:22026 Zbl 1242.11095

[Colmez 2010b] P. Colmez, "(ϕ, Γ)-modules et représentations du mirabolique de $\mathrm{GL}_2(\mathbb{Q}_p)$", pp. 61–153 in *Représentations p-adiques de groupes p-adiques, II: Représentations de* $\mathrm{GL}_2(\mathbb{Q}_p)$ *et* (ϕ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Société Mathématique de France, Paris, 2010. MR 2011i:11170 Zbl 1235.11107

[Colmez 2010c] P. Colmez, “Représentations de $\mathrm{GL}_2(\mathbb{Q}_p)$ et (ϕ, Γ)-modules”, pp. 281–509 in *Représentations p-adiques de groupes p-adiques, II: Représentations de* $\mathrm{GL}_2(\mathbb{Q}_p)$ *et* (ϕ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Société Mathématique de France, Paris, 2010. MR 2011j:11224 Zbl 1218.11107

[Dixon et al. 1999] J. D. Dixon, M. P. F. du Sautoy, A. Mann, and D. Segal, *Analytic pro–p groups*, 2nd ed., Cambridge Studies in Advanced Mathematics **61**, Cambridge University Press, 1999. MR 2000m:20039 Zbl 0934.20001

[Fontaine 1990] J.-M. Fontaine, “Représentations p-adiques des corps locaux, I”, pp. 249–309 in *The Grothendieck Festschrift*, vol. 2, edited by P. Cartier et al., Progr. Math. **87**, Birkhäuser, Boston, 1990. MR 92i:11125 Zbl 0743.11066

[Fresnel and van der Put 2004] J. Fresnel and M. van der Put, *Rigid analytic geometry and its applications*, Progress in Mathematics **218**, Birkhäuser, Boston, 2004. MR 2004i:14023 Zbl 1096.14014

[Kedlaya 2012] K. S. Kedlaya, “New methods for (φ, Γ)-modules”, preprint, 2012, Available at http://math.mit.edu/~kedlaya/papers/new-phigamma.pdf.

[Kisin 2010] M. Kisin, “Deformations of $G_{\mathbb{Q}_p}$ and $\mathrm{GL}_2(\mathbb{Q}_p)$ representations”, pp. 511–528 in *Représentations p-adiques de groupes p-adiques, II: Représentations de* $\mathrm{GL}_2(\mathbb{Q}_p)$ *et* (ϕ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Société Mathématique de France, Paris, 2010. MR 2011e:11185 Zbl 1233.11126

[Paškūnas 2013] V. Paškūnas, “The image of Colmez’s Montreal functor”, preprint, 2013. To appear in *Publ. Math. IHES*. arXiv 1005.2008

[Schneider and Teitelbaum 2003] P. Schneider and J. Teitelbaum, “Algebras of p-adic distributions and admissible representations”, *Invent. Math.* **153**:1 (2003), 145–196. MR 2004g:22015 Zbl 1028.11070

[Schneider and Venjakob 2010] P. Schneider and O. Venjakob, “Localizations and completions of skew power series rings”, *Amer. J. Math.* **132**:1 (2010), 1–36. MR 2011e:16081 Zbl 1191.16041

[Schneider and Vignéras 2011] P. Schneider and M.-F. Vignéras, “A functor from smooth o-torsion representations to (ϕ, Γ)-modules”, pp. 525–601 in *On certain L-functions*, edited by J. Arthur et al., Clay Math. Proc. **13**, Amer. Math. Soc., Providence, RI, 2011. MR 2012e:11194 Zbl 05932925

[Schneider et al. 2012] P. Schneider, M.-F. Vignéras, and G. Zábrádi, “From étale P_+-representations to G-equivariant sheaves on G/P”, preprint, 2012, Available at http://www.math.uni-muenster.de/u/pschnei/publ/pre/reverse-final-revised.pdf. To appear in *Automorphic forms and Galois representations*, LMS Lecture Notes Series, Cambridge Univ. Press.

[Zábrádi 2011] G. Zábrádi, “Exactness of the reduction on étale modules”, *J. Algebra* **331** (2011), 400–415. MR 2012c:11253 Zbl 1227.22022

[Zábrádi 2012] G. Zábrádi, “Generalized Robba rings”, *Israel J. Math.* **191**:2 (2012), 817–887. MR 3011498 Zbl 1277.22008

Communicated by Marie-France Vignéras
Received 2012-08-18 Revised 2013-04-05 Accepted 2013-05-07

zger@cs.elte.hu
Department of Algebra and Number Theory,
Institute of Mathematics, Eötvös Loránd University,
Pázmány Péter sétány 1/C, Budapest H-1117, Hungary

ALGEBRA AND NUMBER THEORY 8:1 (2014)
dx.doi.org/10.2140/ant.2014.8.243

The Tannakian formalism and the Langlands conjectures

David Kazhdan, Michael Larsen and Yakov Varshavsky

Let H be a connected reductive group over an algebraically closed field of characteristic zero, and let Γ be an abstract group. In this note, we show that every homomorphism of Grothendieck semirings $\phi : K_0^+[H] \to K_0^+[\Gamma]$, which maps irreducible representations to irreducible, comes from a group homomorphism $\rho : \Gamma \to H(K)$. We also connect this result with the Langlands conjectures.

Introduction

Let F be a global function field, Γ_F the absolute Galois group of F, G a split connected reductive group over F, ℓ a prime number different from the characteristic of F, and $\hat{G} = {}^L G^0$ the connected Langlands dual group over $\overline{\mathbb{Q}}_\ell$.

Recall that a weak Langlands conjecture asserts that for every pair (π, ω), where π is an automorphic representation of G, whose central character is of finite order, and ω is a representation of $\hat{G}$, there exists a unique semisimple ℓ-adic representation $\rho_{\pi,\omega}$ of Γ_F, whose L^S-function is equal to the L^S-function of (π, ω).

Moreover, a strong Langlands conjecture asserts that there exists a $\hat{G}$-valued ℓ-adic representation $\rho_\pi : \Gamma_F \to \hat{G}(\overline{\mathbb{Q}}_\ell)$ (not unique in general) such that the composition $\omega \circ \rho_\pi$ is isomorphic to $\rho_{\pi,\omega}$ for each representation ω.

The main result of this note implies that in some cases the strong Langlands conjecture follows from the weak one. More specifically, we show the existence of ρ_π in the case when $\rho_{\pi,\omega}$ is irreducible for each irreducible representation ω. In this case, ρ_π is unique up to conjugation, and the Zariski closure of its image contains the derived group of G.

Our result is a corollary of the following variant of the Tannakian formalism. Let H be a connected reductive group over an algebraically closed field K of characteristic zero, and let Γ be an abstract group. Then every homomorphism of groups $\rho : \Gamma \to H(K)$ induces a homomorphism of Grothendieck semirings

Kazhdan was partially supported by ISF grant 1438/06. Larsen was partially supported by NSF grant DMS-0800705. Varshavsky was partially supported by ISF grant 598/09.
MSC2010: primary 11R39; secondary 11F80, 17B10, 18D10.
Keywords: Tannaka duality, Langlands conjectures.

$\rho^* : K_0^+[H] \to K_0^+[\Gamma]$. In this note, we show a partial converse of this assertion. Namely, we show that every homomorphism of Grothendieck semirings $\phi : K_0^+[H] \to K_0^+[\Gamma]$, which maps irreducible representations to irreducibles, comes from a group homomorphism $\rho : \Gamma \to H(K)$. In particular, we show that a connected reductive group is determined by its Grothendieck semiring.

This note was inspired by a combination of a work in progress [Kazhdan and Varshavsky $\geq$ 2014], where it is shown that the weak Langlands conjecture holds in some cases, and a work [Larsen and Pink 1990], which indicates that one does not need the full Tannakian structure in order to reconstruct a connected reductive group.

1. Main results

Let K be an algebraically closed field of characteristic zero.

1.1. (a) For every algebraic group G over K, we denote by $K_0^+[G]$ the Grothendieck semiring of the category of rational representations of G.

In other words, $K_0^+[G]$ is the set of equivalence classes of finite dimensional semisimple representations of G. For every representation ω of G, we denote by $[\omega]$ its class (or more precisely, the class of its semisimplification) in $K_0^+[G]$. For every pair of semisimple representations ω_1 and ω_2 of G, we have $[\omega_1]+[\omega_2]=[\omega_1 \oplus \omega_2]$ and $[\omega_1]\cdot[\omega_2]=[\omega_1 \otimes \omega_2]$.

(b) Note that a representation ω of G is irreducible if and only if its class $[\omega] \in K_0^+[G]$ is irreducible, that is, it cannot be realized as a nontrivial sum $[\omega_1]+[\omega_2]$ of elements of $K_0^+[G]$.

(c) Every homomorphism $\rho : G \to H$ of algebraic groups over K gives rise to the homomorphism $\rho^* : K_0^+[H] \to K_0^+[G]$ of semirings, where $\rho^*([\omega]) := [\omega \circ \rho]$.

The following result asserts that each connected reductive group is determined by its Grothendieck semiring:

Theorem 1.2. *Let G and H be two connected reductive groups over K, and let $\phi : K_0^+[G] \xrightarrow{\sim} K_0^+[H]$ be an isomorphism of semirings.*

Then there exists an isomorphism $\rho : H \xrightarrow{\sim} G$ such that $\rho^* = \phi$. Moreover, ρ is unique up to conjugation.

Remark 1.3. Note, by comparison, that, if G is connected, semisimple, and simply connected, then the Grothendieck *ring* $K_0[G]$ is isomorphic to $\mathbb{Z}[x_1, \dots, x_r]$, where r is the rank of G. Thus, for such groups, $K_0[G]$ encodes only the rank.

1.4. Now let Γ be an abstract group, and let $K_0^+[\Gamma]$ be the Grothendieck semiring of the category of finite-dimensional representations of Γ over K. Every group homomorphism $\rho : \Gamma \to G(K)$ gives rise to the homomorphism $\rho^* : K_0^+[G] \to K_0^+[\Gamma]$ of Grothendieck semirings.

We have the following version of the Tannakian formalism:

Theorem 1.5. *Let $\phi : K_0^+[G] \to K_0^+[\Gamma]$ be a homomorphism of semirings that maps irreducible elements to irreducibles.*

Then there exists a homomorphism $\rho : \Gamma \to G(K)$ such that $\rho^* = \phi$. Moreover, the Zariski closure of the image of each such ρ contains G^{der}, and ρ is unique up to conjugation.

Remarks 1.6. (a) Conversely, let $\rho : \Gamma \to G(K)$ be a homomorphism such that the Zariski closure of $\rho(\Gamma)$ contains G^{der}. Then the homomorphism $\rho^* : K_0^+[G] \to K_0^+[\Gamma]$ maps irreducible elements to irreducibles.

(b) The result fails completely if one does not assume that ϕ maps irreducible elements to irreducible.

Indeed, let G be SL_2, and let Γ be the group with one element. In this case, for each integer $k \geq 2$, there exists a (unique) homomorphism of semirings

$$\phi_k : K_0^+[\mathrm{SL}_2] \to K_0^+[\Gamma] = \mathbb{Z}_{\geq 0},$$

which maps the standard representation of SL_2 to $k \in \mathbb{Z}_{\geq 0}$. Only ϕ_2 corresponds to a (unique) homomorphism $\Gamma \to \mathrm{SL}_2(K)$.

1.7 (Chevalley space). (a) Let $c_G := \operatorname{Spec} K[G]^G$ be the Chevalley space of G, where the action of G on $K[G]$ is induced by the adjoint action of G on itself. For every representation ω of G, its trace $\mathrm{Tr}_\omega \in K[G]^G \subset K[G]$ is a regular function on c_G.

(b) Let $\chi_G : G \to c_G$ be the canonical projection, induced by the embedding $K[c_G] = K[G]^G \hookrightarrow K[G]$. Then for each $g \in G$ and each representation ω of G, we have an equality $\mathrm{Tr}_\omega(\chi_G(g)) = \mathrm{Tr}_\omega(g)$.

The following result is a more explicit formulation of Theorem 1.5:

Corollary 1.8. *Let $f : \Gamma \to c_G(K)$ be a map of sets.*

Suppose that, for every irreducible algebraic representation ω of G, there exists an irreducible finite-dimensional representation ρ_ω of Γ over K such that

$$\mathrm{Tr}_{\rho_\omega}(\gamma) = \mathrm{Tr}_\omega(f(\gamma)) \quad \textit{for all } \gamma \in \Gamma. \tag{1-1}$$

Then there exists a homomorphism $\rho : \Gamma \to G(K)$ such that

$$\chi_G(\rho(\gamma)) = f(\gamma) \quad \textit{for all } \gamma \in \Gamma. \tag{1-2}$$

The Zariski closure of $\rho(\Gamma)$ contains G^{der}, and ρ is unique up to conjugation.

Remark 1.9. Conversely, assume that there exists a homomorphism $\rho : \Gamma \to G(K)$ satisfying (1-2) and such that the Zariski closure of $\rho(\Gamma)$ contains G^{der}. Then for every irreducible representation $\omega : G \to \mathrm{GL}_n$ the representation

$$\rho_\omega := \omega \circ \rho : \Gamma \to \mathrm{GL}_n(K)$$

is irreducible and satisfies (1-1) (use 1.7(b)).

1.10 (Application to the Langlands conjectures). Let F be a global field, F^{sep} a separable closure, $\Gamma_F = \mathrm{Gal}(F^{\mathrm{sep}}/F)$ the absolute Galois group, and ℓ a prime number different from the characteristic of F. Let $\hat{G}$ be a connected reductive group over $\overline{\mathbb{Q}}_\ell$.

By *ℓ-adic* and *$\hat{G}$-valued ℓ-adic* representations of Γ_F, we mean continuous homomorphisms $\rho : \Gamma_F \to \mathrm{GL}_n(\overline{\mathbb{Q}}_\ell)$ and $\rho : \Gamma_F \to \hat{G}(\overline{\mathbb{Q}}_\ell)$, respectively, which are unramified for almost all places of F.

There are well-defined traces $\mathrm{Tr}_\rho(\mathrm{Frob}_v)$ and $\mathrm{Tr}_{\omega\circ\rho}(\mathrm{Frob}_v)$ for almost all places v of F and all representations ω of $\hat{G}$, respectively.

The following analogue of Corollary 1.8 has applications to Langlands conjectures:

Corollary 1.11. *Let $\hat{G}$ be a reductive group over $\overline{\mathbb{Q}}_\ell$, Σ a cofinite subset of the set of places of F, and $f : \Sigma \to c_{\hat{G}}(\overline{\mathbb{Q}}_\ell)$ any map of sets.*

Assume that, for every irreducible algebraic representation ω of G, there exists an irreducible ℓ-adic representation ρ_ω of Γ_F such that

$$\mathrm{Tr}_{\rho_\omega}(\mathrm{Frob}_v) = \mathrm{Tr}_\omega(f(v)) \quad \textit{for almost all } v \in \Sigma. \tag{1-3}$$

Then there exists a $\hat{G}$-valued ℓ-adic representation $\rho : \Gamma_K \to \hat{G}(\overline{\mathbb{Q}}_\ell)$ such that

$$\chi_{\hat{G}}(\rho(\mathrm{Frob}_v)) = f(v) \quad \textit{for almost all } v \in \Sigma. \tag{1-4}$$

The Zariski closure of $\rho(\Gamma_F)$ contains $\hat{G}^{\mathrm{der}}$, and ρ is unique up to conjugation.

2. Determining a connected reductive group from its Grothendieck semiring

In this section, we are going to prove Theorem 1.2. Michael Mueger called our attention to the fact that at least two proofs of this theorem already exist in the literature: [McMullen 1984] and [Handelman 1993]. Nevertheless, we feel that this new proof has merits (including brevity) that justify presenting it.

Let G be a connected reductive group. We will fix a Borel subgroup $B \subset G$ and a maximal torus $T \subset B$. Let $\alpha_1, \dots, \alpha_r$ be the simple roots of G with respect to (B, T), and let W be the Weyl group of (G, T).

2.1. (a) We set $U := X^*(T) \otimes \mathbb{R}$. For each subset $X \subset U$, we denote by $\mathrm{Conv}(X) \subset U$ the convex hull of X.

(b) For each dominant weight ν of G, we denote by V_ν the irreducible representation of G with highest weight ν.

(c) We define a partial order on $X^*(T)$ by the rule

$$\mu \leq \lambda \quad \text{if and only if} \quad \lambda = \mu + \sum_{i=1}^{r} x_i\alpha_i \text{ and } x_i \geq 0 \text{ for all } i.$$

Proposition 2.2. *Let μ and λ be two dominant weights of G. The following conditions are equivalent*:

(a) $\mu \leq \lambda$.

(b) $\mathrm{Conv}(W\mu) \subset \mathrm{Conv}(W\lambda)$.

(c) *There exists a finite-dimensional representation V' of G such that, for every n, every irreducible factor of $V_\mu^{\otimes n}$ is a factor of $V_\lambda^{\otimes n} \otimes V'$.*

Proof. (a) $\Longrightarrow$ (b). Notice that, since μ is dominant, we have $w\mu \leq \mu$ for all $w \in W$. Therefore, our assumption $\mu \leq \lambda$ implies that $w\mu \leq \lambda$ for all $w \in W$. Thus, our assertion follows from the following lemma:

Lemma 2.3. *Let μ and λ be two weights of G such that $w\mu \leq \lambda$ for all $w \in W$. Then* $\mathrm{Conv}(W\mu) \subset \mathrm{Conv}(W\lambda)$.

Proof. Suppose $\mathrm{Conv}(W\mu)$ is not contained in $\mathrm{Conv}(W\lambda)$. Then there exists $w \in W$ such that $w\mu \notin \mathrm{Conv}(W\lambda)$. As $\mathrm{Conv}(W\lambda)$ is W-stable, it follows that $\mu \notin \mathrm{Conv}(W\lambda)$ and hence also $w\mu \notin \mathrm{Conv}(W\lambda)$ for all $w \in W$.

By the separation lemma, there exists $\theta \in U^*$ such that $\theta(\mu) > \theta(w\lambda)$ for all $w \in W$. This is an open condition, so we may choose θ such that $\theta(\alpha_i) \neq 0$ for each $i = 1, \ldots, r$. Replacing θ by $w\theta$ and μ by $w\mu$ for some $w \in W$, we may assume in addition that $\theta(\alpha_i) > 0$ for each $i = 1, \ldots, r$.

By our assumption, $\mu = \lambda - \sum_{i=1}^{r} x_i\alpha_i$ with each $x_i \geq 0$. Therefore,

$$\theta(\mu) = \theta(\lambda) - \sum_{i=1}^{r} x_i\theta(\alpha_i) \leq \theta(\lambda),$$

contradicting our assumption $\theta(\mu) > \theta(\lambda)$. □

(b) $\Longrightarrow$ (c). We start with the following lemma:

Lemma 2.4. *Let X be a finite subset of a finite-dimensional Euclidean space E. Then there exists a compact subset Y of E such that*

$$\mathrm{Conv}(nX) \subset Y + \underbrace{X + X + \cdots + X}_{n} \tag{2-1}$$

for all positive integers n.

Proof. Let $m := |X|$, and let Y denote the ball of vectors of norm at most $R := 2m \max_{x\in X} \|x\|$. We claim that inclusion (2-1) holds for this Y.

Let X be the set $\{x_1, \dots, x_m\}$. Then every vector in $\mathrm{Conv}(nX)$ is of the form

$$v := a_1 n x_1 + \cdots + a_m n x_m,$$

where the a_i are nonnegative and sum to 1. Let $b_i := \lfloor n a_i \rfloor$ for $i \geq 2$ and $b_1 = n - (b_2 + \cdots + b_m)$. As $|b_i - a_i n| < 1$ for $i > 1$, we have

$$|b_1 - a_1 n| = |n - (b_2 + \cdots + b_m) + (a_2 n + \cdots + a_m n - n)| < m - 1.$$

Thus,

$$\|(b_1 x_1 + \cdots + b_m x_m) - v\| \leq \sum_{i=1}^{m} |b_i - a_i n| \|x_i\| < R,$$

and of course, $b_1 x_1 + \cdots + b_m x_m$ belongs to the n-fold iterated sum of X. □

Now we return to the proof of the proposition. We assume that $\mathrm{Conv}(W\mu) \subset \mathrm{Conv}(W\lambda)$, let $X = W\lambda$, and fix a compact set Y satisfying (2-1). Denote by V' the direct sum of all representations V_ν where ν ranges over the dominant weights in WY.

If n is a positive integer, the highest weight χ of any irreducible factor of $V_\mu^{\otimes n}$ is a weight of $V_\mu^{\otimes n}$. Therefore, $\chi \leq n\mu$; hence, by the implication (a) $\Longrightarrow$ (b) shown above, χ is an element of

$$\mathrm{Conv}(Wn\mu) = n\,\mathrm{Conv}(W\mu) \subset n\,\mathrm{Conv}(W\lambda).$$

By (2-1), χ can be written as a sum of n elements of $W\lambda$ and an element of WY, which is necessarily in the weight group. Thus, χ has the form $\sum_{i=1}^{n} w_i \lambda + w'\nu$ for some $w_1, \dots, w_n, w' \in W$ and some highest weight ν of V'.

Using the conjecture of Parthasarathy, Ranga Rao, and Varadarajan, proven in [Kumar 1988], we conclude that V_χ is an irreducible factor of $V_\lambda^{\otimes n} \otimes V_\nu$ and hence also an irreducible factor of $V_\lambda^{\otimes n} \otimes V'$.

(c) $\Longrightarrow$ (a). Now suppose that there exists a finite-dimensional representation V' of G such that, for every n, every irreducible factor of $V_\mu^{\otimes n}$ must be a factor of $V_\lambda^{\otimes n} \otimes V'$ as well. Then every weight of $V_\mu^{\otimes n}$ must be a weight of $V_\lambda^{\otimes n} \otimes V'$, and in particular, this is true for the weight $n\mu$. Thus, $n\mu = \lambda_n + \nu_n$ for some weights λ_n of $V_\lambda^{\otimes n}$ and $\nu_n \in V'$.

Note that $\lambda_n = n\lambda - \sum_{i=1}^{r} n_i \alpha_i$ for some $n_i \in \mathbb{Z}_{\geq 0}$. Therefore, $n\mu$ is equal to $n\lambda - \sum_{i=1}^{r} n_i \alpha_i + \nu_n$; hence, for each $n \in \mathbb{N}$, we have an equality

$$\lambda - \mu = \sum_{i=1}^{r} \frac{n_i}{n} \alpha_i - \frac{1}{n} \nu_n.$$

Next we recall that the set of weights of V' is finite, so the expression $(1/n)\nu_n \in U$ tends to zero when n tends to infinity. Hence, the difference $\lambda - \mu$ equals $\sum_{i=1}^{r} x_i\alpha_i$, where each $x_i = \lim_{n\to\infty} n_i/n$ is nonnegative. This shows that $\mu \le \lambda$. □

Corollary 2.5. *The root datum of G can be reconstructed from the semiring $K_0^+[G]$.*

Proof. We divide our construction into steps as follows.

Step 1. First we claim that the partially ordered set of dominant weights of G can be reconstructed from the semiring $K_0^+[G]$.

For this, we note that the map $\mu \mapsto [V_\mu]$ gives a bijection between the set of dominant weights of G and the set of irreducible objects of $K_0^+[G]$.

Proposition 2.2 implies that for two dominant weights μ and λ of G we have $\mu \le \lambda$ if and only if there exists $\theta \in K_0^+[G]$ such that, for all $n \in \mathbb{N}$ and all irreducible elements $[V_\nu] \in K_0^+[G]$, we have

$$[V_\mu]^n - [V_\nu] \in K_0^+[G] \implies [V_\lambda]^n\theta - [V_\nu] \in K_0^+[G].$$

Step 2. For every triple λ, μ, ν of dominant weights of G, we have $\lambda = \mu + \nu$ if and only if λ is the largest dominant weight such that V_λ is an irreducible factor of $V_\mu \otimes V_\nu$. Therefore, Proposition 2.2 implies that the semigroup structure on the set of dominant weights of G can be reconstructed from the semiring $K_0^+[G]$.

Step 3. The group of weights $X^*(T)$ of G is the group completion of the semigroup of dominant weights. The group of coweights of G, $X_*(T)$, is given as the group of homomorphisms

$$X_*(T) = \mathrm{Hom}(X^*(T), \mathbb{Z}).$$

Note that there is a canonical isomorphism between $\mathrm{Aut}(X^*(T))$ and $\mathrm{Aut}(X_*(T))$.

Step 4. We claim that $\alpha \in X^*(T)$ is a simple root if and only if it is a minimal nonzero weight of T for which there exists a dominant weight $\lambda \in X^*(T)$ such that $V_{2\lambda-\alpha}$ is an irreducible factor of $V_\lambda^{\otimes 2}$.

More precisely, we claim that, for every dominant weight λ, the maximal weights $\mu \ne 2\lambda$ such that V_μ is an irreducible factor of $V_\lambda^{\otimes 2}$ are precisely weights of the form $2\lambda - \alpha$, where α is a simple root satisfying $\langle \check{\alpha}, \lambda\rangle > 0$.

To show this, we observe that every maximal weight $\mu \ne 2\lambda$ in $V_\lambda^{\otimes 2}$ is of the form $2\lambda - \alpha$, where α is a simple root satisfying $\langle \check{\alpha}, \lambda\rangle > 0$. Now the assertion follows from the fact that, for such an α, the weight $2\lambda - \alpha$ has multiplicity one in $V_{2\lambda}$ and multiplicity two in $V_\lambda^{\otimes 2}$.

By Step 1, the set of simple roots can therefore be reconstructed from the semiring $K_0^+[G]$.

Step 5. For each simple root α of G, the corresponding simple coroot $\check{\alpha} \in X_*$ can be characterized by the following condition: for every dominant weight μ, the pairing $\langle\check{\alpha}, \mu\rangle$ is the unique element $m \in \mathbb{Z}_{\geq 0}$ such that $2\mu - m\alpha$ is dominant but $2\mu - (m+1)\alpha$ is not dominant. Indeed,

$$\langle\check{\alpha}, 2\mu - m\alpha\rangle = 2\langle\check{\alpha}, \mu\rangle - m\langle\check{\alpha}, \alpha\rangle = 2\langle\check{\alpha}, \mu\rangle - 2m$$

is nonnegative if and only if $m \leq \langle\check{\alpha}, \mu\rangle$ while, for every other simple root $\alpha' \neq \alpha$ of G with a corresponding simple coroot $\check{\alpha}'$, we have

$$\langle\check{\alpha}', 2\mu - m\alpha\rangle = 2\langle\check{\alpha}', \mu\rangle - m\langle\check{\alpha}', \alpha\rangle \geq 2\langle\check{\alpha}', \mu\rangle \geq 0$$

for all $m \geq 0$. Thus, the set of simple coroots can also be reconstructed from $K_0^+[G]$.

Step 6. After having reconstructed all simple coroots $\check{\alpha}$, we reconstruct all simple reflections $s_\alpha \in \mathrm{Aut}(X_*(T))$, hence the Weyl group $W \subset \mathrm{Aut}(X_*(T))$, as the subgroup generated by simple reflections. Next we reconstruct the set of all roots of G, as images of the simple roots under W, and likewise for the coroots of G. This completes the reconstruction of the whole root datum of G. □

2.6. *Proof of Theorem 1.2.* An isomorphism of semirings $\phi : K_0^+[G] \xrightarrow{\sim} K_0^+[H]$ induces a bijection between irreducible objects and hence a bijection between dominant weights of G and H, which we denote by $\tilde{\phi}$.

The proof of Corollary 2.5 shows that $\tilde{\phi}$ extends to an isomorphism between the root data of G and H. It therefore comes from an isomorphism of algebraic groups $\rho : H \xrightarrow{\sim} G$.

We claim that $\rho^* : K_0^+[G] \xrightarrow{\sim} K_0^+[H]$ is equal to ϕ. It is enough to show that, for each dominant weight λ of G, we have $\phi([V_\lambda]) = \rho^*([V_\lambda])$. Both expressions, however, are equal to $[V_{\tilde{\phi}(\lambda)}]$.

Conversely, if $\rho : H \to G$ is an isomorphism such that $\rho^* = \phi$, then for each dominant weight λ of G we have $\rho^*([V_\lambda]) = \phi([V_\lambda]) = [V_{\tilde{\phi}(\lambda)}]$, so ρ induces the isomorphism $\tilde{\phi}$ between the root data; hence, ρ is unique up to conjugation. □

3. The Tannakian formalism

In this section, we are going to prove Theorem 1.5. Throughout the section, we will assume that the hypotheses of Theorem 1.5 hold. For each irreducible representation ω of G, we choose an irreducible representation ρ_ω of Γ such that $[\rho_\omega] = \phi([\omega])$.

Lemma 3.1. (a) *Let ω' and ω'' be two irreducible representations of G, and let $\omega' \otimes \omega'' \cong \bigoplus \omega_i$ be a decomposition of their tensor product into irreducibles. Then $\rho_{\omega'} \otimes \rho_{\omega''} \cong \bigoplus \rho_{\omega_i}$.*

(b) *If ω is a trivial (one-dimensional) representation* **1** *of G, then ρ_ω is a trivial representation of Γ.*

(c) *The representation ω is one-dimensional if and only if ρ_ω is one-dimensional.*

(d) *For each irreducible representation ω of G, we have $\rho_{\omega^*} \cong (\rho_\omega)^*$.*

(e) *Let ω' and ω'' be two irreducible representations of G such that $\rho_{\omega'} \cong \rho_{\omega''}$. Then restrictions $\omega'|_{G^{\mathrm{der}}}$ and $\omega''|_{G^{\mathrm{der}}}$ are isomorphic.*

Proof. (a) By hypothesis, we have $[\omega'] \cdot [\omega''] = \sum_i [\omega_i]$. Since ϕ is a homomorphism of semirings, we conclude that

$$[\rho_{\omega'} \otimes \rho_{\omega''}] = \phi([\omega']) \cdot \phi([\omega'']) = \sum_i \phi([\omega_i]) = \big[\bigoplus \rho_{\omega_i}\big].$$

Since $\rho_{\omega'}$ and $\rho_{\omega''}$ are irreducible, their tensor product $\rho_{\omega'} \otimes \rho_{\omega''}$ is semisimple (see [Chevalley 1955, p. 88]). Therefore, $\rho_{\omega'} \otimes \rho_{\omega''} \cong \bigoplus \rho_{\omega_i}$.

(b) This follows from the observation that $\omega = \mathbf{1}$ if and only if $\omega \otimes \omega \cong \omega$.

(c) This follows from the observation that ω is one-dimensional if and only if $\omega \otimes \omega$ is irreducible.

(d) Note that the representation $\omega \otimes \omega^*$ has a trivial subrepresentation $\mathbf{1}$. Therefore, by (a) and (b), the representation $\rho_{\omega^*} \otimes \rho_{\omega}$ has a subrepresentation $\rho_{\mathbf{1}} \cong \mathbf{1}$. Since ρ_{ω} and ρ_{ω^*} are irreducible, this implies that $\rho_{\omega^*} \cong (\rho_{\omega})^*$.

(e) If $\rho_{\omega'} \cong \rho_{\omega''}$, then the tensor product $\rho_{\omega'} \otimes (\rho_{\omega''})^* \cong \rho_{\omega'} \otimes \rho_{\omega''^*}$ contains a subrepresentation $\mathbf{1}$. Using (a) and (c), we conclude that the tensor product $\omega' \otimes \omega''^*$ has a one-dimensional subrepresentation ξ. Since ω' and ω'' are irreducible, we conclude that $\omega' \cong \omega'' \otimes \xi$; thus, the restrictions $\omega'|_{G^{\mathrm{der}}}$ and $\omega''|_{G^{\mathrm{der}}}$ are isomorphic. □

3.2. For every irreducible representation ω of G, we denote by z_ω its central character. Let Z be the center of G, and denote by ι the embedding $\Gamma \xrightarrow{\sim} \Gamma \times \{1\} \hookrightarrow \Gamma \times Z(K)$.

Lemma 3.3. (a) *There exists a unique homomorphism of semirings $\tilde{\phi} : K_0^+[G] \to K_0^+[\Gamma \times Z(K)]$ such that*

$$\tilde{\phi}([\omega]) = [\rho_\omega \boxtimes z_\omega] \quad \textit{for each irreducible } \omega. \tag{3-1}$$

(b) *The map $\tilde{\phi}$ is injective, maps irreducibles to irreducibles, and satisfies $\iota^* \circ \tilde{\phi} = \phi$.*

(c) *Assume that there exists a homomorphism $\rho : \Gamma \to G(K)$ such that $\rho^* = \phi$, and let $\tilde{\rho} : \Gamma \times Z(K) \to G(K)$ be a homomorphism defined by $\tilde{\rho}(\gamma, z) := \rho(\gamma) \cdot z$. Then $\tilde{\rho}^* = \tilde{\phi}$.*

Proof. (a) Since the additive Grothendieck semigroup $K_0^+[G]$ is freely generated by irreducible elements $[\omega]$, there exists a unique homomorphism of semigroups $\tilde{\phi} : K_0^+[G] \to K_0^+[\Gamma \times Z(K)]$ that satisfies (3-1). It remains to show that for every two representations ω' and ω'' of G we have an equality

$$\tilde{\phi}([\omega'] \cdot [\omega'']) = \tilde{\phi}([\omega']) \cdot \tilde{\phi}([\omega'']). \tag{3-2}$$

By the additivity of $\tilde{\phi}$, we may assume that ω' and ω'' are irreducible. Let $\omega' \otimes \omega'' \cong \bigoplus \omega_i$ be a decomposition of their tensor product into irreducibles. Then

$[\omega'] \cdot [\omega''] = \sum_i [\omega_i]$; hence, the left-hand side of (3-2) is equal to

$$\tilde{\phi}\Bigl(\sum_i [\omega_i]\Bigr) = \sum_i \tilde{\phi}([\omega_i]) = \sum_i [\rho_{\omega_i} \boxtimes z_{\omega_i}]$$

while the right-hand side of (3-2) is equal to

$$[\rho_{\omega'} \boxtimes z_{\omega'}] \cdot [\rho_{\omega''} \boxtimes z_{\omega''}] = [(\rho_{\omega} \otimes \rho_{\omega''}) \boxtimes z_{\omega'} z_{\omega''}].$$

Since the central character of each ω_i is equal to $z_{\omega'} z_{\omega''}$, equality (3-2) follows from Lemma 3.1(a).

(b) By construction, for each irreducible element $[\omega]$, the element $\tilde{\phi}([\omega]) = [\rho_\omega \boxtimes z_\omega]$ is irreducible, and

$$\iota^* \tilde{\phi}([\omega]) = \iota^*([\rho_\omega \boxtimes z_\omega]) = [\rho_\omega] = \phi([\omega]).$$

This implies that $\tilde{\phi}$ maps irreducibles to irreducibles and satisfies $\iota^* \circ \tilde{\phi} = \phi$.

Finally, since as additive semigroups $K_0^+[G]$ and $K_0^+[\Gamma]$ are freely generated by irreducibles, in order to show that $\tilde{\phi}$ is injective, it is enough to show that it is injective on irreducibles.

Let ω' and ω'' be two irreducible representations of G such that $\tilde{\phi}([\omega']) = \tilde{\phi}([\omega''])$. Then $\rho_{\omega'} \cong \rho_{\omega''}$ and $z_{\omega'} = z_{\omega''}$. Using Lemma 3.1(e), we conclude that $\omega'|_{G^{\mathrm{der}}} \cong \omega''|_{G^{\mathrm{der}}}$ and $\omega'|_Z = \omega''|_Z$. Hence, $\omega' \cong \omega''$, implying the injectivity.

(c) It is enough to show that $\tilde{\rho}^*([\omega]) = \tilde{\phi}([\omega])$ when $[\omega]$ is irreducible. Both expressions, however, are equal to $[\rho_\omega \boxtimes z_\omega]$. □

3.4. *Proof of Theorem 1.5.* First we will show the existence of ρ under the assumption that $\phi : K_0^+[G] \to K_0^+[\Gamma]$ is injective.

Let $\mathscr{C}$ be the full subcategory of $\operatorname{Rep}\Gamma$ consisting of semisimple representations $\tau \in \operatorname{Rep}\Gamma$ such that $[\tau] = \phi([\omega])$ for some $[\omega] \in K_0^+[G]$. Since $\phi([\omega])$ is irreducible for each irreducible $[\omega]$, $\mathscr{C}$ is a semisimple abelian subcategory. Since ϕ is a homomorphism of semirings, $\mathscr{C}$ is a rigid tensor subcategory of $\operatorname{Rep}\Gamma$ (use Lemma 3.1(a)–(d)) and hence a Tannakian category. Let $f : \mathscr{C} \to \mathrm{Vec}_{\mathrm{K}}$ be the forgetful functor, and let $H := \operatorname{Aut}^{\otimes}(f)$ be the group of tensor automorphisms of f.

By the Tannakian formalism (see, for example, [Deligne and Milne 1982, Theorem 2.11]), H is an affine group scheme, and f induces an equivalence of tensor categories $\mathscr{C} \xrightarrow{\sim} \operatorname{Rep} H$. Since G is an algebraic group, the category $\operatorname{Rep} G$ has a tensor generator ω. Then an element $\rho_\omega \in \operatorname{Rep}\Gamma$ such that $[\rho_\omega] = \phi([\omega])$ must be a tensor generator of $\mathscr{C} \cong \operatorname{Rep} H$. This implies that H is an algebraic group (see [Deligne and Milne 1982, Proposition 2.20]). Moreover, since $\mathscr{C} \cong \operatorname{Rep} H$ is semisimple, the group H is reductive (see [Deligne and Milne 1982, Proposition 2.23]).

Every element of $\gamma \in \Gamma$ defines a tensor automorphism of f over K. Hence we get a group homomorphism $\pi : \Gamma \to H(K)$ such that $\pi^* : \operatorname{Rep} H \to \operatorname{Rep}\Gamma$ is the inverse of the equivalence $f : \mathscr{C} \xrightarrow{\sim} \operatorname{Rep} H$.

By construction, the homomorphism $\phi : K_0^+[G] \to K_0^+[\Gamma]$ decomposes as $K_0^+[G] \xrightarrow{\phi'} K_0^+[H] \xrightarrow{\pi^*} K_0^+[\Gamma]$, and the homomorphism ϕ' is surjective. By our assumption, ϕ' is also injective; hence, it is an isomorphism. Since G is connected, we conclude that H is connected as well (use, for example, [Deligne and Milne 1982, Corollary 2.22]). Therefore, by Theorem 1.2, there exists an isomorphism $\rho' : H \xrightarrow{\sim} G$ such that $\phi' = \rho'^*$. Then the composition $\rho := \rho' \circ \pi : \Gamma \to G(K)$ satisfies $\rho^* = \pi^* \circ \phi' = \phi$.

To show the existence of ρ in general, we consider the homomorphism of Grothendieck semirings $\tilde{\phi} : K_0^+[G] \to K_0^+[\Gamma \times Z(K)]$, considered in Lemma 3.3(a).

Then $\tilde{\phi}$ is injective, so by the particular case shown above, there exists a homomorphism $\tilde{\rho} : \Gamma \times Z(K) \to G(K)$ such that $\tilde{\rho}^* = \tilde{\phi}$. Then the composition $\rho := \tilde{\rho} \circ \iota : \Gamma \to G(K)$ satisfies $\rho^* = \iota^* \circ \tilde{\phi} = \phi$.

Conversely, let $\rho : \Gamma \to G(K)$ be a homomorphism such that $\rho^* = \phi$. To show that the Zariski closure of $\rho(\Gamma)$ contains G^{der}, it suffices to show that the homomorphism $\tilde{\rho} : \Gamma \times Z(K) \to G(K)$ from Lemma 3.3(c) has a Zariski closed image.

Let $H \subset G$ be the Zariski closure of the image of $\tilde{\rho}$, and denote by i the inclusion $H \hookrightarrow G$. Then $\tilde{\rho}^* = \tilde{\phi} : K_0^+[G] \to K_0^+[\Gamma \times Z(K)]$ factors through $i^* : K_0^+[G] \to K_0^+[H]$. In particular, i^* is injective and maps irreducibles to irreducibles. Then using Chevalley's theorem ([Borel 1991, Theorem 5.1] or [Deligne and Milne 1982, Proposition 2.21]), i has to be an isomorphism.

Finally, to show that ρ is unique up to conjugation, it suffice to show that $\tilde{\rho} : \Gamma \times Z(K) \to G(K)$ is unique up to conjugation. Thus, we can replace ρ by $\tilde{\rho}$ and ϕ by $\tilde{\phi}$, thereby assuming that ϕ is injective.

Then, using the notation of the existence part, the tensor functor $\rho^* : \operatorname{Rep} G \to \operatorname{Rep} \Gamma$ decomposes as a composition $\operatorname{Rep} G \xrightarrow{\psi} \operatorname{Rep} H \xrightarrow{\pi^*} \operatorname{Rep} \Gamma$ of tensor functors. By the Tannakian formalism, there exists a homomorphism $\rho' : H \to G$ such that $\rho'^* = \psi$. Then ρ is conjugate to the composition $\rho' \circ \pi$, so it remains to show that the conjugacy class of ρ' is uniquely defined.

We have seen that ϕ decomposes as $K_0^+[G] \xrightarrow{\phi'} K_0^+[H] \xrightarrow{\pi^*} K_0^+[\Gamma]$; therefore, $\rho'^* : K_0^+[G] \xrightarrow{\sim} K_0^+[H]$ coincides with ϕ'. Hence, the uniqueness assertion for ρ' follows from Theorem 1.2. □

4. Two corollaries

In this section, we are going to prove Corollaries 1.8 and 1.11.

Lemma 4.1. *Assume that the hypotheses of Corollary 1.8 hold.*

(a) *There exists a unique homomorphism of semirings $\phi : K_0^+[G] \to K_0^+[\Gamma]$ such that $\phi([\omega]) = [\rho_\omega]$ for each irreducible ω.*

(b) *Let $\rho : \Gamma \to G(K)$ be a group homomorphism. Then equality* (1-2) *holds for ρ if and only if $\omega \circ \rho \cong \rho_\omega$ for all irreducible ω.*

Proof. (a) Since the semigroup $K_0^+[G]$ is freely generated by irreducible elements, there exists a unique homomorphism of semigroups $\phi : K_0^+[G] \to K_0^+[\Gamma]$ such that $\phi([\omega]) = [\rho_\omega]$ for each irreducible ω.

It remains to show that for every two representations ω' and ω'' of G we have an equality

$$\phi([\omega'] \cdot [\omega'']) = \phi([\omega']) \cdot \phi([\omega'']).$$

Since a semisimple representation is determined by its trace, it is enough to show that

$$\mathrm{Tr}_{\phi([\omega']\cdot[\omega''])}(\gamma) = \mathrm{Tr}_{\phi([\omega'])}(\gamma) \cdot \mathrm{Tr}_{\phi([\omega''])}(\gamma) \tag{4-1}$$

for all $\gamma \in \Gamma$. First we observe that for all $\gamma \in \Gamma$ and all $[\omega] \in K_0^+[G]$ we have an equality

$$\mathrm{Tr}_{\phi([\omega])}(\gamma) = \mathrm{Tr}_{[\omega]}(f(\gamma)). \tag{4-2}$$

Indeed, by additivity, it is enough to show (4-2) for ω irreducible. In this case, the assertion follows from equalities $\phi([\omega]) = [\rho_\omega]$ and (1-1).

Using (4-2), our desired equality (4-1) can be written in the form

$$\mathrm{Tr}_{[\omega']\cdot[\omega'']}(f(\gamma)) = \mathrm{Tr}_{[\omega']}(f(\gamma)) \cdot \mathrm{Tr}_{[\omega'']}(f(\gamma)).$$

Therefore, it follows from the multiplicativity of the trace map $\mathrm{Tr} : K_0^+[G] \to K[G]$.

(b) Since functions Tr_ω with ω irreducible generate $K[c_G]$ as a K-vector space (see [Steinberg 1965, Theorem 6.1(a)]), the equality (1-2) is equivalent to the equality

$$\mathrm{Tr}_\omega(\chi_G(\rho(\gamma))) = \mathrm{Tr}_\omega(f(\gamma)) \tag{4-3}$$

for all $\gamma \in \Gamma$ and all irreducible ω. Since the left side of (4-3) equals $\mathrm{Tr}_{\omega\circ\rho}(\gamma)$ (see 1.7(b)) while the right-hand side of (4-3) equals $\mathrm{Tr}_{\rho_\omega}(\gamma)$ by (1-1), equality (4-3) is equivalent to the equality $\mathrm{Tr}_{\omega\circ\rho} = \mathrm{Tr}_{\rho_\omega}$ for all irreducible ω. But this is equivalent to the desired isomorphism $\omega \circ \rho \cong \rho_\omega$. □

4.2. *Proof of Corollary 1.8.* By Lemma 4.1(a), there exists a unique homomorphism of semirings $\phi : K_0^+[G] \to K_0^+[\Gamma]$ such that $\phi([\omega]) = [\rho_\omega]$ for each irreducible ω. Then by Theorem 1.5, there exists a homomorphism $\rho : \Gamma \to G(K)$ such that $\rho^* = \phi$. In particular, we have that $[\omega \circ \rho]$ is equal to $\phi([\omega]) = [\rho_\omega]$ for each irreducible ω. Then by Lemma 4.1(b), the equality (1-2) holds for ρ.

Conversely, let $\rho : \Gamma \to G(K)$ be a homomorphism, satisfying (1-2). Then by Lemma 4.1(b), $\rho^*([\omega]) = [\omega \circ \rho]$ is equal to $\phi([\omega]) = [\rho_\omega]$ for each irreducible ω. Thus $\rho^* : K_0^+[G] \to K_0^+[\Gamma]$ is equal to ϕ. It then follows from Theorem 1.5 that ρ is unique up to conjugation, and that the Zariski closure of $\rho(\Gamma)$ contains G^{der}. □

4.3. *Proof of Corollary 1.11.* The argument is very similar to that of Corollary 1.8.

As in Lemma 4.1(a), there exists a unique homomorphism of semirings

$$\phi : K_0^+[\hat{G}] \to K_0^+[\Gamma_F]$$

such that $\phi([\omega]) = [\rho_\omega]$ for each irreducible ω. Indeed, arguing as in Lemma 4.1(a) word for word, we reduce ourselves to the equality (4-1). Moreover, by the Chebotarev density theorem, it is enough to show equality (4-1) when $\gamma = \mathrm{Frob}_v$ for almost all $v \in \Sigma$.

Then we reduce the problem to showing that

$$\mathrm{Tr}_{\phi([\omega])}(\mathrm{Frob}_v) = \mathrm{Tr}_{[\omega]}(f(v))$$

for all irreducible $[\omega]$ and almost all $v \in \Sigma$. But the latter equality follows from equalities $\phi([\omega]) = [\rho_\omega]$ and (1-3).

By Theorem 1.5, there now exists a homomorphism $\rho : \Gamma_F \to \hat{G}(\overline{\mathbb{Q}}_\ell)$ such that $\rho^* = \phi$.

We claim that, for every representation ω of $\hat{G}$, the composition $\omega \circ \rho$ is a semisimple ℓ-adic representation. By additivity, it is enough to show in the case when ω is irreducible. However, in this case,

$$[\omega \circ \rho] = \rho^*([\omega]) = [\rho_\omega]$$

is irreducible; hence, $\omega \circ \rho \cong \rho_\omega$ is an irreducible ℓ-adic representation.

Choosing ω to be a faithful representation of $\hat{G}$, we conclude that ρ is continuous and unramified almost everywhere.

Finally, arguing exactly as in Lemma 4.1(b) (and using the isomorphisms $\omega \circ \rho \cong \rho_\omega$), we conclude that ρ satisfies the equality (1-4).

Conversely, let $\rho : \Gamma_F \to \hat{G}(\overline{\mathbb{Q}}_\ell)$ be a $\hat{G}$-valued ℓ-adic representation satisfying (1-4). Again arguing exactly as in Lemma 4.1(b) and using the Chebotarev density theorem, we conclude that $\rho^*([\omega]) = [\omega \circ \rho]$ is equal to $\phi([\omega]) = [\rho_\omega]$ for each irreducible ω. Thus, $\rho^* : K_0^+[\hat{G}] \to K_0^+[\Gamma_F]$ is equal to ϕ.

Therefore, it follows from Theorem 1.5 that ρ is unique up to conjugation and that the Zariski closure of $\rho(\Gamma_F)$ contains $\hat{G}^{\mathrm{der}}$. □

Acknowledgments

We would like to acknowledge the contribution of Pavel Etingof, whose suggestions led to substantial simplification and conceptual clarification of this paper.

Part of the work was done while Kazhdan and Varshavsky visited the University of Chicago. The rest of the work was done while Varshavsky visited Indiana University. We thank both these institutions for stimulating atmosphere and financial support. We would also like to thank Michael Mueger for calling our attention to two previous proofs of Theorem 1.2.

References

[Borel 1991] A. Borel, *Linear algebraic groups*, 2nd ed., Graduate Texts in Mathematics **126**, Springer, New York, 1991. MR 92d:20001 Zbl 0726.20030

[Chevalley 1955] C. Chevalley, *Théorie des groupes de Lie, III: Théorèmes généraux sur les algèbres de Lie*, Actualités Sci. Ind. **1226**, Hermann & Cie, Paris, 1955. MR 16,901a

[Deligne and Milne 1982] P. Deligne and J. S. Milne, "Tannakian categories", pp. 101–228 in *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Mathematics **900**, Springer, Berlin, 1982. MR 84m:14046 Zbl 0477.14004

[Handelman 1993] D. Handelman, "Representation rings as invariants for compact groups and limit ratio theorems for them", *Internat. J. Math.* **4**:1 (1993), 59–88. MR 94c:22005 Zbl 0788.22005

[Kazhdan and Varshavsky $\geq$ 2014] D. Kazhdan and Y. Varshavsky, "On the cohomology of the moduli spaces of F-bundles: stable cuspidal Deligne–Lusztig part", In preparation.

[Kumar 1988] S. Kumar, "Proof of the Parthasarathy–Ranga Rao–Varadarajan conjecture", *Invent. Math.* **93**:1 (1988), 117–130. MR 89j:17009 Zbl 0668.17008

[Larsen and Pink 1990] M. Larsen and R. Pink, "Determining representations from invariant dimensions", *Invent. Math.* **102**:2 (1990), 377–398. MR 92c:22026 Zbl 0687.22004

[McMullen 1984] J. R. McMullen, "On the dual object of a compact connected group", *Math. Z.* **185**:4 (1984), 539–552. MR 85e:22010 Zbl 0513.43007

[Steinberg 1965] R. Steinberg, "Regular elements of semisimple algebraic groups", *Inst. Hautes Études Sci. Publ. Math.* 25 (1965), 49–80. MR 31 #4788 Zbl 0136.30002

Communicated by Richard Taylor
Received 2012-09-02 Revised 2013-08-20 Accepted 2013-09-19

kazhdan@math.huji.ac.il *Einstein Institute of Mathematics, Hebrew University,*
Givat Ram, 91904 Jerusalem, Israel

mjlarsen@indiana.edu *Department of Mathematics, Indiana University, Rawles Hall,*
Bloomington, IN 47405-5701, United States

vyakov@math.huji.ac.il *Einstein Institute of Mathematics, Hebrew University,*
Givat Ram, 91904 Jerusalem, Israel

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *ANT* are usually in English, but articles written in other languages are welcome.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 8 No. 1 2014

1937-0652(2014)8:1;1-A