Algebra & Number Theory

Volume 12

2018

No. 3

msp

Algebra & Number Theory

msp.org/ant

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

The subscription price for 2018 is US $340/year for the electronic version, and $535/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber 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 MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ALGEBRA AND NUMBER THEORY 12:3 (2018)
dx.doi.org/10.2140/ant.2018.12.493

Pseudo-exponential maps, variants, and quasiminimality

Martin Bays and Jonathan Kirby

We give a construction of quasiminimal fields equipped with pseudo-analytic maps, generalizing Zilber's pseudo-exponential function. In particular we construct pseudo-exponential maps of simple abelian varieties, including pseudo-$\wp$-functions for elliptic curves. We show that the complex field with the corresponding analytic function is isomorphic to the pseudo-analytic version if and only if the appropriate version of Schanuel's conjecture is true and the corresponding version of the strong exponential-algebraic closedness property holds. Moreover, we relativize the construction to build a model over a fairly arbitrary countable subfield and deduce that the complex exponential field is quasiminimal if it is exponentially-algebraically closed. This property states only that the graph of exponentiation has nonempty intersection with certain algebraic varieties but does not require genericity of any point in the intersection. Furthermore, Schanuel's conjecture is not required as a condition for quasiminimality.

1.	Introduction	493
2.	Algebraic background	498
3.	Γ-fields	500
4.	Predimension and strong extensions	507
5.	The canonical countable model	515
6.	Categoricity	523
7.	Classification of strong extensions	527
8.	Axiomatization of Γ-closed fields	529
9.	Specific applications of the general construction	531
10.	Comparison with the analytic models	537
11.	Generically Γ-closed fields	541
	Acknowledgements	547
	References	548

1. Introduction

1A. ***Exponential fields.*** The field $\mathbb{C}$ of complex numbers is well-known to be *strongly minimal*, that is, any subset of $\mathbb{C}$ definable in the ring language is either finite or cofinite. Consequently, the model theory of $\mathbb{C}$ is very tame: there is a very well-understood behaviour of the models (one model of each uncountable

Kirby was supported by EPSRC grant EP/L006375/1 .
MSC2010: primary 03C65; secondary 03C75, 12L12.
Keywords: exponential fields, predimension, categoricity, Schanuel conjecture, Ax–Schanuel, Zilber–Pink, quasiminimality, Kummer theory.

cardinality, known as uncountable categoricity) and of the definable sets (they have finite Morley rank and we can understand them geometrically in terms of algebraic varieties). The other most important mathematical field, the field $\mathbb{R}$ of real numbers, is *o-minimal*, which means that although the class of models is not well-behaved (not classifiable), there is a very good geometric understanding of the definable sets (they are the semialgebraic sets). Remarkably, Wilkie [1996] showed that when the real exponential function e^x is adjoined, the structure $\mathbb{R}_{\exp}$ is still o-minimal. Adjoining the complex exponential function e^z to $\mathbb{C}$ gives the structure $\mathbb{C}_{\exp}$, which cannot be well-behaved in terms of the class of models or the definable sets because it interprets the ring $\mathbb{Z}$. However, Zilber suggested that in the model $\mathbb{C}_{\exp}$ itself, the influence of $\mathbb{Z}$ might only extend to the countable subsets of $\mathbb{C}$. He made the following conjecture.

Conjecture 1.1 (Zilber's weak quasiminimality conjecture). The complex exponential field $\mathbb{C}_{\exp} = \langle \mathbb{C}; +, \cdot, \exp \rangle$ is *quasiminimal*: every subset of $\mathbb{C}$ definable in $\mathbb{C}_{\exp}$ is either countable or cocountable.

A slightly stronger version of the conjecture which avoids reference to definable sets is that every automorphism-invariant subset is countable or cocountable. As far as we are aware, all known approaches to the conjecture would give this stronger result anyway. If the conjecture is true then the solution sets of exponential polynomial equations, which we can call complex exponential varieties, would be expected to have good geometric properties similar to those of algebraic varieties, provided we avoid some exceptional cases like $\mathbb{Z}$. If the conjecture is false, another possibility is that $\mathbb{R}$ is definable as a subset of $\mathbb{C}$. The field $\mathbb{R}$ with $\mathbb{Z}$ as a definable subset is so-called *second-order arithmetic*, and the definable sets are extremely wild, with no geometric properties in general.

As one approach to his conjecture, Zilber [2000; 2005b] showed how to construct a quasiminimal exponential field we call $\mathbb{B}$ using a variant of Hrushovski's predimension method [1993]. He called $\mathbb{B}$ a *pseudo-exponential field* with the idea that the exponential map is a *pseudo-analytic* function.

Zilber's approach was to prove that a certain list of axioms $\mathrm{ECF}_{\mathrm{SK,CCP}}$ in the infinitary logic $L_{\omega_1,\omega}(Q)$ behaves in an analogous way to a strongly minimal first-order theory. In particular, all its models are quasiminimal and it is uncountably categorical.

Theorem 1.2. *Up to isomorphism, there is exactly one model of the axioms* $\mathrm{ECF}_{\mathrm{SK,CCP}}$ *of each uncountable cardinality, and it is quasiminimal.*

This theorem appears in [Zilber 2005b]. Some gaps in the proof were filled in the unpublished note [Bays and Kirby 2013], which this paper supersedes. In this paper we give a new construction of $\mathbb{B}$ and hence a complete proof of Theorem 1.2. The theorem suggests a stronger form of the quasiminimality conjecture which evidently implies Conjecture 1.1.

Conjecture 1.3 (Zilber's strong quasiminimality conjecture). $\mathbb{C}_{\exp}$ is isomorphic to the unique model $\mathbb{B}$ of $\mathrm{ECF}_{\mathrm{SK,CCP}}$ of cardinality continuum.

The axioms in $\mathrm{ECF}_{\mathrm{SK,CCP}}$ will be explained in Section 8, but briefly, there are two algebraic axioms which are obviously true in $\mathbb{C}_{\exp}$ and then three more axioms: Schanuel's conjecture, strong exponential-algebraic closedness, and the countable closure property. Schanuel's conjecture is a conjecture of transcendental number theory which can be seen as saying that certain systems of exponential polynomial equations do not have solutions. Strong exponential-algebraic closedness roughly says that a system of equations has solutions (even generic over any given finite set) unless that would contradict Schanuel's conjecture. The countable closure property says roughly that such systems of equations which are *balanced*, in the sense of having the same number of equations as variables, have only countably many solutions. Zilber proved the countable closure property for $\mathbb{C}_{\exp}$, so we have the following reformulation.

Theorem 1.4 [Zilber 2005b]. *Conjecture 1.3 is true if and only if Schanuel's conjecture is true and* $\mathbb{C}_{\exp}$ *is strongly exponentially-algebraically closed.*

Theorems 1.2 and 1.4 together imply that if $\mathbb{C}_{\exp}$ satisfies Schanuel's conjecture and is strongly exponentially-algebraically closed then it is quasiminimal. Schanuel's conjecture is considered out of reach, since even the very simple consequence that the numbers e and π are algebraically independent is unknown. Proving strong exponential-algebraic closedness involves finding solutions of certain systems of equations and then showing they are generic, the latter step usually done using Schanuel's conjecture. A weaker condition is *exponential-algebraic closedness* which requires the same systems of equations to have solutions, but says nothing about their genericity. We are able to remove the dependence on Schanuel's conjecture completely from Conjecture 1.1:

Theorem 1.5. *If* $\mathbb{C}_{\exp}$ *is exponentially-algebraically closed then it is quasiminimal.*

1B. *A more general construction:* Γ*-fields.* Our construction is more general and we can use it to construct also a pseudo-analytic version of the Weierstrass $\wp$-functions, the exponential maps of simple abelian varieties, and more generally other pseudo-analytic subgroups of the product of two commutative algebraic groups. For example, we prove an analogous form of Theorems 1.2 and 1.4 for $\wp$-functions. The list of axioms $\wp\mathrm{CF}_{\mathrm{SK,CCP}}(E)$ and the other notions used in the statement of the theorem will be explained in Section 9C of the paper.

Theorem 1.6. *Given an elliptic curve E over a number field $K_0 \subseteq \mathbb{C}$, the list $\wp\mathrm{CF}_{\mathrm{SK,CCP}}(E)$ of axioms is uncountably categorical and every model is quasiminimal. Furthermore, if $\wp$ is the Weierstrass function associated to $E(\mathbb{C})$, so $\exp_E = [\wp : \wp' : 1] : \mathbb{C} \to E(\mathbb{C})$ is the exponential map of $E(\mathbb{C})$, then $\mathbb{C}_\wp := \langle \mathbb{C}; +, \cdot, \exp_E \rangle \models \wp\mathrm{CF}_{\mathrm{SK,CCP}}(E)$ if and only if the analogue of Schanuel's conjecture for $\wp$ holds and $\mathbb{C}_\wp$ is strongly $\wp$-algebraically closed.*

In the most general form, we consider what we call Γ*-fields*, which are fields F of characteristic 0 equipped with a subgroup $\Gamma(F)$ of a product $G_1(F) \times G_2(F)$, where G_1 and G_2 are commutative algebraic groups. The complete definition is given in Section 3A, where we also explain how the examples we consider fit into cases (EXP), generalizing the exponential and Weierstrass $\wp$-functions above, (COR), generalizing analytic correspondences between nonisogenous elliptic curves, and (DEQ), generalizing the solution sets of certain differential equations.

Hrushovski used Fraïssé's amalgamation method, which produces countable structures. Zilber wanted uncountable structures so he instead framed his constructions in terms of existentially closed models within a certain category. He gave a framework of *quasiminimal excellent classes* [Zilber 2005a], building on Shelah's notion of an excellent $L_{\omega_1,\omega}$-sentence [1983], to prove the uniqueness of the uncountable models. The second author showed in [Kirby 2010b] that the quasiminimal excellence conditions can be checked just on the countable models, and with Hart, Hyttinen and Kesälä we proved in [Bays et al. 2014b] that the most complicated of the conditions to check — excellence — follows from the other conditions. So in this paper we recast the construction in 4 stages:

(1) We start with a suitable base Γ-field F_{base}, and describe a category $\mathcal{C}(F_{\text{base}})$ of so-called *strong extensions* of F_{base}.

(2) We apply a suitable version of *Fraïssé's amalgamation theorem* to the category to produce a countable model $M(F_{\text{base}})$.

(3) We check that $M(F_{\text{base}})$ satisfies the conditions to be part of a *quasiminimal class*, and deduce there is a unique model of cardinality continuum, which we denote by $\mathbb{M}(F_{\text{base}})$.

(4) We give the *axioms* $\Gamma\mathrm{CF}_{\mathrm{CCP}}(F_{\text{base}})$ describing the class.

As a more general form of Theorem 1.2, we prove:

Theorem 1.7. *Given an essentially finitary Γ-field F_{base} of type* (EXP), (COR), *or* (DEQ), *the list of axioms $\Gamma\mathrm{CF}_{\mathrm{CCP}}(F_{\text{base}})$ is uncountably categorical and every model is quasiminimal.*

Our notion of Γ-fields is algebraic and not every example is related to an analytic prototype. However cases (EXP) and (COR) do have many analytic examples, given in Definitions 3.1 and 3.2. We call these *analytic Γ-fields*. For these we are able to prove the countable closure property, extending Zilber's result for $\mathbb{C}_{\exp}$ and the equivalent result in [Jones et al. 2016] for $\wp$-functions.

Theorem 1.8. *Let $\mathbb{C}_\Gamma$ be an analytic Γ-field. Then $\mathbb{C}_\Gamma$ satisfies the countable closure property.*

One of the key ideas of this paper is that the amalgamation construction is done over a base Γ-field F_{base}, and that everything is done relative to that base. Pushing this idea further, we can also work over a base which is closed with respect to the quasiminimal pregeometry on the model F_{base}. This involves modifying the amalgamation construction, so we only consider extensions of F_{base} in which F_{base} remains closed with respect to the pregeometry. There is a direct analogy with extensions of differential fields. The closed base field takes the role of the field of constants, and we consider extension fields in which there are no new constants. This is useful because the differential field versions of Schanuel's conjecture then apply to say that Schanuel's conjecture is true relative to the base in the analytic Γ-fields.

In the paper [Kirby and Zilber 2014] it was shown that, assuming the conjecture on intersections with tori (CIT, also known as the multiplicative Zilber–Pink conjecture), any exponential field satisfying Schanuel's conjecture and exponential-algebraic closedness is actually strongly exponentially-algebraically closed. In this paper we are able to adapt that idea to show unconditionally that the difference between Γ-closedness and strong Γ-closedness (the analogues of exponential-algebraic closedness and strong exponential-algebraic closedness) disappears if we consider the generic version, meaning relative to a closed base. Instead of the CIT we use a theorem which we call the *horizontal semiabelian weak Zilber–Pink*, a theorem about intersections of families of algebraic varieties with cosets of algebraic subgroups of semiabelian varieties. This method allows us to prove Theorem 1.5 and a more general version:

Theorem 1.9. *Let $\mathbb{C}_\Gamma$ be an analytic Γ-field. If $\mathbb{C}_\Gamma$ is Γ-closed then it is quasiminimal.*

1C. ***An overview of the paper.*** In Section 2 we explain our conventions on viewing algebraic varieties and their profinite covers in a model-theoretic way. We also explain the relationship between subgroups and endomorphisms of the commutative algebraic groups we study.

In Section 3 we define our Γ-fields and their finitely generated extensions. We prove that finitely generated extensions of suitable (so-called *essentially finitary*) Γ-fields are determined by good bases, and that these good bases exist and are determined by finite data from a countable range of possibilities. This is the key step in proving the form of $\aleph_0$-stability which is essential for the existence of quasiminimal models. The main tool here is Kummer theory over torsion for abelian varieties.

In Section 4 we introduce the predimension notion and use it to define which extensions of Γ-fields are strong. We also use it to define a pregeometry on Γ-fields. Then we show that there is a unique *full-closure* of an essentially finitary Γ-field, and classify the strong finitely generated extensions of Γ-fields and of full Γ-fields. This completes stage (1) of the construction as described above.

Section 5 covers stage (2) of the construction. We recall a category-theoretic version of Fraïssé's amalgamation theorem which is suitably general for us. Then, starting with a suitable base Γ-field F_{base}, we consider the category $\mathcal{C}(F_{\text{base}})$ of strong extensions of F_{base} and apply the amalgamation theorem to get a countable Fraïssé limit $M(F_{\text{base}})$. We also consider a variant amalgamating only the Γ-algebraic extensions and another variant where we consider only extensions which are purely Γ-transcendental over F_{base}.

In Section 6 we show that the Fraïssé limit models we have produced are quasiminimal pregeometry structures, and hence give rise to uncountably categorical classes. In this way we get the uncountable models, in particular the model $\mathbb{M}(F_{\text{base}})$ of cardinality continuum. This is stage (3).

In Section 7 we give a classification of the finitely generated strong extensions of Γ-fields, and in Section 8 we use it to axiomatize our models and prove Theorem 1.7. This completes stage (4).

In Section 9 we consider specific instances of our Γ-fields including pseudo-exponentiation, pseudo-Weierstrass $\wp$-functions, and others, and prove Theorem 1.2 and half of Theorem 1.6.

In Section 10 we compare our models to the complex analytic prototypes. For Weierstrass $\wp$-functions we relate the Schanuel property to the André–Grothendieck conjecture on the periods of 1-motives, using work of Bertolin, finishing the proof of Theorem 1.6. We briefly discuss the literature on steps towards proving the Γ-closedness and strong Γ-closedness properties for analytic Γ-fields. Then we prove Theorem 1.8.

In Section 11 we consider Γ-fields which may not be Γ-closed but are generically so. These are the Γ-fields produced by the variant construction in which the base F_{base} remains closed with respect to the pregeometry. We state and prove the horizontal semiabelian weak Zilber–Pink, and then prove Theorems 1.5 and 1.9.

2. Algebraic background

2A. ***Algebraic varieties and groups.*** We use the standard model-theoretic foundations for algebraic varieties and algebraic groups, as described by Pillay [1998], roughly following Weil. In particular, we work in the theory ACF_0 with parameters for a field K_0. Any variety V is considered as a definable set, and using elimination of imaginaries it is in definable bijection with a constructible subset of affine space. We always assume we have chosen such a bijection, although we will not mention it explicitly. Given any field extension F of K_0, we write $V(F)$ for the points of V all of whose coordinates lie in F. In this way, V is a functor from the category of field extensions of its field of definition to the category of sets. Similarly, given any subset $A \subseteq V(F)$, we can form the subfield of F which is generated by (the coordinates of) the points in A.

In the same way, a commutative algebraic group G, defined over K_0, is considered as a functor from the category of field extensions of K_0 to the category of abelian groups. If G is an algebraic $\mathcal{O}$-module, that is, the ring $\mathcal{O}$ acts on G via regular endomorphisms, defined over K_0, we can also consider it as a functor to the category of $\mathcal{O}$-modules.

$\mathbb{G}_{\mathrm{a}}$ denotes the additive group, $\mathbb{G}_{\mathrm{a}}(F) = \langle F; + \rangle$, and $\mathbb{G}_{\mathrm{m}}$ the multiplicative group, $\mathbb{G}_{\mathrm{m}}(F) = \langle F^{\times}; \cdot \rangle$.

An algebraic group is *connected* if it has no proper finite index algebraic subgroups. The *connected component* G^o of an algebraic group G is the largest connected algebraic subgroup.

We write $G[m]$ for the m-torsion subgroup of an algebraic group G.

If G is a commutative algebraic group over a field of characteristic 0, we write LG for the (commutative) Lie algebra of G, the tangent space at the identity considered as an algebraic group. So $LG \cong \mathbb{G}_{\mathrm{a}}^{\dim(G)}$. If $\theta : G \to G'$ is an algebraic group homomorphism, then $L\theta : LG \to LG'$ is the derivative at the identity. For algebraic groups over $\mathbb{C}$, these definitions agree with the usual definitions for complex Lie groups.

2B. ***Subgroups and endomorphisms.*** Any connected algebraic subgroup H of a power $\mathbb{G}_{\mathrm{m}}^n$ of the multiplicative group can be defined by a system of monomial equations: $H = \ker(M)$ for some integer square matrix $M \in \mathrm{Mat}_n(\mathbb{Z})$ acting multiplicatively. Then $LH \leq L\mathbb{G}_{\mathrm{m}}^n = \mathbb{G}_{\mathrm{a}}^n$ is the kernel of the same matrix acting additively.

As we observe in the following lemma, the picture is almost the same when we replace $\mathbb{G}_{\mathrm{m}}$ with an abelian variety G and $\mathbb{Z}$ with its endomorphism ring $\mathrm{End}(G)$: up to finite index, subgroups are defined by $\mathcal{O}$-linear equations, namely those which define the corresponding Lie subalgebra. With a few self-contained exceptions, this lemma is essentially all we will use of the theory of abelian varieties.

Lemma 2.1. *Suppose G is $\mathbb{G}_{\mathrm{m}}$ or an abelian variety over a field of characteristic* 0, *and $\mathcal{O} = \mathrm{End}(G)$ is its endomorphism ring. Then*

(i) *any connected algebraic subgroup $H \leq G^n$ is the connected component of the kernel of an endomorphism $\eta \in \mathrm{End}(G^n) \cong \mathrm{Mat}_n(\mathcal{O})$,*

$$H = \ker(\eta)^o;$$

(ii) *$LH \leq LG^n$ is then the kernel of $L\eta \in \mathrm{End}(LG^n)$.*

Proof. (i) By Poincaré's complete reducibility theorem [Mumford 1970, p. 173], there exists an algebraic subgroup H' such that the summation map $\Sigma : H \times H' \to G^n$ is an isogeny, that is, a surjective homomorphism with finite kernel. Let m be the exponent of the kernel of Σ. Then $\theta(\Sigma(h, h')) := (mh, mh')$ defines an isogeny $\theta : G^n \to H \times H'$. Let $\pi_2 : H \times H' \to H'$ be the projection. Then $(\pi_2 \circ \theta \circ \Sigma)(h, h') = mh'$, so $\ker(\pi_2 \circ \theta)^o = (\Sigma(H \times H'[m]))^o = (H + H'[m])^o = H$.

(ii) As $H \subseteq \ker(\eta)$, the derivative $L\eta$ of η at 0 vanishes on LH, so $LH \leqslant \ker(L\eta)$. Also, $\ker(L\eta) = L\ker(\eta)$, so we have

$$\begin{aligned} \dim(\ker(L\eta)) &= \dim(L\ker(\eta)) \\ &= \dim(\ker(\eta)) && \text{since 0 is a smooth point} \\ &= \dim H && \text{since } H \text{ has finite index in } \ker(\eta) \\ &= \dim LH && \text{again since 0 is a smooth point.} \end{aligned}$$

So LH has finite index in $\ker(L\eta)$, but $\ker(L\eta) \leqslant LG^n$, which is torsion-free, so $\ker(L\eta)$ is connected, and hence $LH = \ker(L\eta)$. □

2C. *Division points and the profinite cover.*

Definition 2.2. Let G be a commutative group and let $a \in G$. A *division point* of a in G is any $b \in G$ such that, for some $m \in \mathbb{N}^+$, $mb = a$.

A *division sequence* for a in G is a sequence $(a_m)_{m \in \mathbb{N}^+}$ in G such that $a_1 = a$ and for all $m, n \in \mathbb{N}^+$ we have $na_{nm} = a_m$.

If $(a_m)_{m \in \mathbb{N}^+}$ is a division sequence for a in G we can define a group homomorphism $\theta : \mathbb{Q} \to G$ by $\theta(r/m) = ra_m$ for $r \in \mathbb{Z}$ and $m \in \mathbb{N}^+$. This gives a bijective correspondence between division sequences for a in G and group homomorphisms $\theta : \mathbb{Q} \to G$ such that $\theta(1) = a$.

Definition 2.3. The *profinite cover* $\widehat{G}$ of a commutative group G is the group of all homomorphisms $\mathbb{Q} \to G$, with the group structure defined pointwise in G. We write $\rho_G : \widehat{G} \to G$ for the evaluation homomorphism given by $\rho_G(\theta) = \theta(1)$.

Thus the set of division sequences for a in G is in bijective correspondence with $\rho_G^{-1}(a)$, and we think of elements of $\widehat{G}$ both as homomorphisms from $\mathbb{Q}$ and as division sequences.

The group $\widehat{G}$ itself is divisible and torsion-free. The image of ρ_G is the subgroup of divisible points of G, and ρ_G is injective if and only if G is torsion-free. In general, $\ker(\rho_G)$ is a profinite group built from the torsion of G (in fact, it is the product over primes l of the l-adic Tate modules of G).

For an element $a \in G$, we will often use the notation $\hat{a}$ for a chosen element of $\widehat{G}$ such that $\rho_G(\hat{a}) = a$. Of course, $\hat{a}$ is determined by a only when ρ_G is injective, that is, when G is torsion-free.

If $f : G \to H$ is a group homomorphism, we can lift it to a homomorphism $\hat{f} : \widehat{G} \to \widehat{H}$ defined by $\theta \mapsto f \circ \theta$. In particular, if $G \subseteq H$ is a subgroup then $\widehat{G}$ is naturally a subgroup of $\widehat{H}$. (In category-theoretic language, $\hat{}$ is a covariant representable functor and in fact $\rho_G : \widehat{G} \to G$ is the universal arrow from the category of divisible, torsion-free abelian groups into G.)

When G is a commutative algebraic group we think of $\widehat{G}$ also as a functor, so we write $\widehat{G}(F)$ rather than $\widehat{G(F)}$ for the group of division sequences of the group $G(F)$.

Model-theoretically we think of $\widehat{G}$ as the set of division sequences from G, which is a set of infinite tuples satisfying the divisibility conditions. It can be seen as an inverse limit of definable sets, sometimes called a pro-definable set [Kamensky 2007].

Remark 2.4. Suppose that G is a Lie group, for example the complex points of a complex algebraic group, and $\exp : LG \to G$ is the exponential map. For $a \in LG$, the sequence $(\exp(a/m))_{m \in \mathbb{N}^+}$ is a division sequence in G. In fact, the division sequences which arise this way are precisely those which converge topologically to the identity of G [Bays et al. 2014c, Remark 2.20].

3. Γ-fields

3A. Γ-*fields*. In this section we describe the analytic examples we are studying and give the definition of a Γ-field, which is intended to capture and generalize the model-theoretic algebra of the examples.

Definition 3.1 (analytic Γ-fields of type (EXP)). The graph of the usual complex exponential function is a subgroup of $\mathbb{G}_a(\mathbb{C}) \times \mathbb{G}_m(\mathbb{C})$. Similarly, if $A(\mathbb{C})$ is a complex abelian variety (or more generally any commutative complex algebraic group) of dimension d, then the graph Γ of the exponential map of A is a subgroup of $LA(\mathbb{C}) \times A(\mathbb{C})$. Here $LA(\mathbb{C})$ is the Lie algebra of A and we can identify it with the group $\mathbb{G}_a^d(\mathbb{C})$. In this paper we only consider the cases when A is $\mathbb{G}_m$ or A is a simple abelian variety of dimension d. We combine these by saying A is a *simple semiabelian variety*.

We write $\mathcal{O}$ for the ring $\mathrm{End}(A)$ of algebraic endomorphisms of A. In many cases $\mathcal{O} = \mathbb{Z}$, but sometimes, for example if A is an elliptic curve with *complex multiplication*, then $\mathcal{O}$ properly extends $\mathbb{Z}$. Any $\eta \in \mathcal{O}$ acts on LA as the derivative $d\eta$, a linear map. Thus $\mathcal{O}$ naturally acts on $LA(\mathbb{C})$ as a subring of $\mathrm{GL}_d(\mathbb{C})$, and Γ is an $\mathcal{O}$-submodule of $LA(\mathbb{C}) \times A(\mathbb{C})$.

In the case where A is an elliptic curve E, embedded in projective space $\mathbb{P}^2$ in the usual way via its Weierstrass equation, the exponential map of $E(\mathbb{C})$ is written in homogeneous coordinates as $z \mapsto (\wp(z) : \wp'(z) : 1)$, where $\wp$ is the Weierstrass $\wp$-function associated with E.

We call all of these examples *analytic* Γ-*fields of type* (EXP).

Definition 3.2 (analytic Γ-fields of type (COR)). The exponential map of a complex elliptic curve factors through $\mathbb{G}_m(\mathbb{C})$, giving an analytic map $\theta : \mathbb{G}_m(\mathbb{C}) \to E(\mathbb{C})$. More generally, there are analytic correspondences between semiabelian varieties. We take G_1 and G_2 both to be simple complex semiabelian varieties of the same dimension d, and assume G_1 and G_2 are not isogenous. Suppose $\mathrm{End}(G_1)$ and $\mathrm{End}(G_2)$ are both isomorphic to a ring $\mathcal{O}$, and furthermore there is a $\mathbb{C}$-vector space isomorphism $\psi : LG_1 \to LG_2$ which respects the actions of $\mathcal{O}$. We choose such a ψ and take Γ to be the image of the graph of ψ under $\exp_{G_1 \times G_2} : LG_1(\mathbb{C}) \times LG_2(\mathbb{C}) \to G_1(\mathbb{C}) \times G_2(\mathbb{C})$. Then Γ is an $\mathcal{O}$-submodule of $G_1(\mathbb{C}) \times G_2(\mathbb{C})$, and a complex Lie-subgroup. The graph of the map θ is an example of such a Γ, but in general Γ need not be the graph of a function.

We call these examples *analytic Γ-fields of type* (COR). By an *analytic Γ-field* we mean one of type (EXP) or type (COR).

Definition 3.3 (differential equation examples). If $f(t)$ is a holomorphic function in a neighbourhood of $0 \in \mathbb{C}$, then the pair $(x, y) = \big(f(t), \exp(f(t))\big)$ satisfies the differential equation $Dx = Dy/y$, where $D = \mathrm{d}/\mathrm{d}t$. We can consider the set Γ of solutions of the differential equation not just in a field of functions but in a differentially closed field F. Then Γ is a subgroup of $\mathbb{G}_\mathrm{a}(F) \times \mathbb{G}_\mathrm{m}(F)$. The paper [Kirby 2009] studies this situation for the differential equations satisfied by the exponential maps of semiabelian varieties S. While these S do not have to be simple, they do have to be defined over the constant field C of F. In these cases the group Γ is quite closely related to the graph of the exponential map and can be analyzed via a similar amalgamation construction.

We capture all of these three types of examples in the notion of a Γ-field. We next give the assumptions we use on the algebraic groups, and then define Γ-fields. The assumptions we make are not the most general possible, but they are what we use throughout this paper.

Definition 3.4 (conventions for K_0, G_2, $\mathcal{O}$, and $k_\mathcal{O}$). We take K_0 to be a countable field of characteristic 0, which must be a number field except in case (DEQ) below. Let G_2 be a simple semiabelian variety defined over K_0. We write $\mathcal{O}$ for the ring $\mathrm{End}(G_2)$ of algebraic (that is, regular) group endomorphisms of G_2 and assume that they are also all defined over K_0. Let $k_\mathcal{O}$ denote the ring $\mathbb{Q} \otimes_\mathbb{Z} \mathcal{O}$.

Remarks 3.5. The ring $\mathcal{O}$ has no zero divisors because G_2 is simple. So $\mathcal{O}$ embeds in $k_\mathcal{O}$. If $\mathcal{O} = \mathbb{Z}$ then $k_\mathcal{O}$ is just $\mathbb{Q}$. Every nonzero algebraic group endomorphism of a simple abelian variety is an isogeny, so becomes invertible in $k_\mathcal{O}$. Hence $k_\mathcal{O}$ is a division ring, and the $\mathcal{O}$-torsion of any $\mathcal{O}$-module is exactly the $\mathbb{Z}$-torsion.

Definition 3.6 (conventions for G_1, G, and the torsion). We consider two cases for the choice of G_1, corresponding to the above analytic examples.

- **Case** (EXP)**:** We take $G_1 = \mathbb{G}_\mathrm{a}^d$, where $d = \dim G_2$. We identify G_1 with the Lie algebra LG_2, that is, the tangent space at the identity of G_2. As in the analytic case, this identification makes G_1 into an algebraic $\mathcal{O}$-module, that is, an $\mathcal{O}$-module in which every element of $\mathcal{O}$ acts as a regular map.
- **Case** (COR)**:** G_1 is also a simple semiabelian variety defined over K_0, and with all its algebraic endomorphisms defined over K_0. We assume G_1 is not isogenous to G_2, but $\mathrm{End}(G_1) \cong \mathcal{O}$ and we choose an isomorphism, so $G_1 \times G_2$ becomes an algebraic $\mathcal{O}$-module over K_0.

Let $G = G_1 \times G_2$, and write $\pi_i : G \to G_i$ for the projection maps of the product, for $i = 1, 2$. We write the groups G_1, G_2, and G additively.

For $i = 1, 2$ the torsion of G_i is contained in $G_i(K_0^{\mathrm{alg}})$, and hence is bounded. We write Tor_i for the torsion of $G_i(F)$ for any F such that $G(F)$ contains $\mathrm{Tor}(G(K_0^{\mathrm{alg}}))$. The torsion of $G(F)$ is written $\mathrm{Tor}(G)$. It is equal to $(\mathrm{Tor}_1 \times \mathrm{Tor}_2) \cap G(F)$.

Remarks 3.7. Note that for any algebraically closed field F extending K_0 the groups $G_i(F)$ and $G(F)$ are divisible $\mathcal{O}$-modules. Furthermore, $G(F)/\mathrm{Tor}(G)$ is divisible and torsion-free, and thus a $k_\mathcal{O}$-vector space.

Definition 3.8 (Γ-fields). A Γ-*field* (with respect to the $\mathcal{O}$-module G) is a field extension A of K_0 equipped with a divisible $\mathcal{O}$-submodule $\Gamma(A)$ of $G(A)$ such that

(1) A is generated as a field by $\Gamma(A)$,

(2) the projection $\pi_i(\Gamma(A))$ in $G_i(A)$ contains Tor_i for $i = 1, 2$.

We write $\Gamma_i(A)$ for the projections $\pi_i(\Gamma(A))$. The Γ-field A is *full* if, in addition, A is algebraically closed and the projections $\Gamma_i(A)$ are equal to $G_i(A)$.

The *kernels* of a Γ-field A are defined to be

$$\ker_1(A) := \{x \in G_1(A) \mid (x, 0) \in \Gamma(A)\},$$
$$\ker_2(A) := \{y \in G_2(A) \mid (0, y) \in \Gamma(A)\}.$$

When A is full and $\ker_2(A)$ is trivial, $\Gamma(A)$ is the graph of a surjective $\mathcal{O}$-module homomorphism from $G_1(F)$ to $G_2(F)$ with kernel $\ker_1(F)$ as in the analytic examples of type (EXP). However the case (EXP) for our Γ-fields is more general.

The most difficult part of this paper uses the Kummer theory of semiabelian varieties over number fields. This is not needed for the differential equations examples, or more generally in the following variant.

Definition 3.9 (case (DEQ)). A Γ-field in case (DEQ) is the same as above except that we require the full torsion group $\mathrm{Tor}(G)$ to be contained in $\Gamma(A)$, and we relax the assumption that K_0 is a number field so it can be any countable field of characteristic 0.

Definition 3.10 (extensions of Γ-fields). An *extension* of a Γ-field A is a Γ-field B together with an inclusion of fields $A \subseteq B$ over K_0 such that $\Gamma(A) \subseteq \Gamma(B)$ and for $i = 1, 2$ we have $\ker_i(A) = \ker_i(B)$. We also say that A is a Γ-*subfield* of B.

We refer to the last condition in the definition by saying that the extension *preserves the kernels*.

One could also consider extensions of Γ-fields which do not preserve the kernels, and this would be necessary for an analysis of the first-order theory such as that done in the paper [Kirby and Zilber 2014]. However, we do not consider such extensions in this paper.

Remarks 3.11 (Γ-fields as model-theoretic structures). (1) Model-theoretically, we consider a Γ-field as a structure in the 1-sorted first-order language $L_\Gamma = \langle +, \cdot, -, \Gamma, (c_a)_{a \in K_0} \rangle$, where Γ is a relation symbol of appropriate arity to denote a subset of the group G, and we have parameters for the field K_0. Later, we will also be adding parameters for a base Γ-field F_{base}.

(2) However, our notion of Γ-field extension corresponds to an injective L_Γ-homomorphism, not necessarily an L_Γ-embedding. Specifically, it is not necessary in an extension $A \hookrightarrow B$ of Γ-fields that $\Gamma(B) \cap G(A) = \Gamma(A)$, although that will be true in most cases we will consider later, for example when the extension is *strong* (see Definition 4.3).

(3) Although we use the 1-sorted language with the sort being that of the underlying field, we also refer to elements of Γ as being from the sort Γ, rather than from the definable set Γ. Model theorists used to working with L^{eq} will see there is no important difference.

(4) By definition, the Γ-field A is determined by the submodule $\Gamma(A)$ of $G(F)$. Furthermore, an extension $A \hookrightarrow B$ is determined by the inclusion of submodules $\Gamma(A) \hookrightarrow \Gamma(B)$. Thus, if F is a monster model of ACF_0, the category of Γ-fields is equivalent to the category of divisible $\mathcal{O}$-submodules of $G(F)$ (whose projections contain Tor_1 and Tor_2), with embeddings, in a first-order language with relation symbols for all of the Zariski-closed subsets of $G(F)$ which are defined over K_0. This is more or less the setting in [Zilber 2005b].

Remark 3.12. One might also consider the case that G_1 and G_2 are equal (or isogenous, which comes to essentially the same thing). Γ can then be considered as the graph of a new (quasi)endomorphism of G_1. The situation is complicated by the need to consider the extension of the algebraic endomorphism ring generated by Γ. Analytic examples include raising to a complex power on $\mathbb{G}_\mathrm{m}$, which is analyzed with a different setup in [Zilber 2003; 2015].

In an earlier draft of this paper we tried to incorporate this into our setup, and in fact produced an example where Γ was the graph of a multivalued endomorphism θ on $\mathbb{G}_\mathrm{m}$, lifting to a generic action of the ring $\mathbb{Q}[\theta, \theta^{-1}]$ on the profinite cover $\widehat{\mathbb{G}_\mathrm{m}}$. However, this is subtly different from giving an action of the field $\mathbb{Q}(\theta)$, which is what occurs for complex powers.

While we expect that such Γ can be treated along the lines of this paper, much as we expect that the simplicity assumption on the semiabelian variety could be relaxed, these elaborations are left to future work.

3B. *Finitely generated extensions.*

Definition 3.13. Let B be a Γ-field, and let $\{A_j \mid j \in J\}$ be a set of Γ-subfields of B (each with the same kernels as B). We define $\bigwedge_{j\in J} A_j$ to be the Γ-subfield A of B such that $\Gamma(A) = \bigcap_{j\in J} \Gamma(A_j)$.

Lemma 3.14. *The Γ-field $\bigwedge_{j\in J} A_j$ is a Γ-subfield of B.*

The proof is straightforward, but we give the details because they show exactly where all the hypotheses of the definitions are used.

Proof. Let $A = \bigwedge_{j\in J} A_j$. Since $\Gamma(A)$ is defined as the intersection of a set of $\mathcal{O}$-submodules of $\Gamma(B)$, it is also an $\mathcal{O}$-submodule of $\Gamma(B)$. A is defined as the subfield of B generated by the coordinates of the points in $\Gamma(A)$, so $\Gamma(A)$ is an $\mathcal{O}$-submodule of $G(A)$.

If $a \in \ker_1(B)$, then $(a, 0) \in \Gamma(A_j)$ for all $j \in J$ because $\ker_1(A_j) = \ker_1(B)$, so $(a, 0) \in \Gamma(A)$. So $\ker_1(A) = \ker_1(B)$ and similarly $\ker_2(A) = \ker_2(B)$.

If $a \in \mathrm{Tor}_1 = \mathrm{Tor}_1(B)$ then there is $b \in G_2(B)$ such that $(a, b) \in \Gamma(B)$. Furthermore, for any $b' \in G_2(B)$ we have $(a, b') \in \Gamma(B)$ if and only if $b' - b \in \ker_2(B)$. For each $j \in J$, A_j is a Γ-subfield of B, so $\Gamma_1(A_j)$ contains $\mathrm{Tor}_1(B)$, so there is b' such that $(a, b') \in \Gamma(A_j)$. But $\ker_2(A_j) = \ker_2(B)$ by assumption, so we have $(a, b) \in \Gamma(A_j)$, and since this holds for all j we have $(a, b) \in \Gamma(A)$. Thus $\Gamma_1(A)$ contains $\mathrm{Tor}_1(B)$, and similarly $\Gamma_2(A)$ contains $\mathrm{Tor}_2(B)$.

In particular, $\Gamma(A)$ contains all the torsion from $\Gamma(B)$, so since it is the intersection of divisible $\mathcal{O}$-submodules, it is itself divisible as an $\mathcal{O}$-submodule of $G(A)$. Hence, A is a Γ-subfield of B. □

Definition 3.15. Let B be a Γ-field and $X \subseteq \Gamma(B)$ a subset. We say that

$$A = \bigwedge\{A', \text{ a } \Gamma\text{-subfield of } B \text{ (with the same kernels as } B) \mid X \subseteq \Gamma(A')\}$$

is the Γ-subfield *generated by* X, and write it as $\langle X\rangle_B$ or, more usually with the B suppressed, as $\langle X\rangle$.

We say A is a *finitely generated* Γ-field if $\Gamma(A)$ is of finite rank as an $\mathcal{O}$-module, or equivalently as a $\mathbb{Z}$-module. Equivalently, A is generated by a finite subset and $\ker_1(A)$ and $\ker_2(A)$ are of finite rank.

Note that a finitely generated Γ-field is not usually finitely generated as a field, because we insist that $\Gamma(A)$ is a divisible $\mathcal{O}$-submodule.

If Y is a subset of $\Gamma(A)$, we say that A is *finitely generated over* Y if there is a finite subset X of $\Gamma(A)$ such that A is the Γ-subfield of itself generated by $X \cup Y$. In particular, for Y a Γ-subfield of A, we have the notion of a finitely generated extension of Γ-fields. It is easy to see that an extension $A \hookrightarrow B$ of Γ-fields is finitely generated if and only if $\operatorname{ldim}_{k_{\mathcal{O}}}(\Gamma(B)/\Gamma(A))$ is finite.

Definition 3.16. The intersection of full Γ-subfields of B (with the same kernels as B) is again a full Γ-subfield. Thus we can define a full Γ-field A to be *finitely generated as a full* Γ*-field* if there is a finite subset X of A such that

$$A = \bigwedge\{A', \text{ a full } \Gamma\text{-subfield of } A \text{ (with the same kernels as } A) \mid X \subseteq \Gamma(A')\}.$$

Likewise, there is the notion of being *finitely generated as a full* Γ*-field extension*.

Except in trivial cases, a finitely generated full Γ-field is not finitely generated as a Γ-field, and a finitely generated full Γ-field extension is not finitely generated as a Γ-field extension.

Definition 3.17. Recall that an $\mathcal{O}$-submodule H of G is *pure* in G if whenever $x \in G$ and $nx \in H$ for some $n \in \mathbb{N}^+$, then $x \in H$.

Lemma 3.18. *If A is the* Γ*-subfield of B generated by X, then* $\Gamma(A)$ *is the pure* $\mathcal{O}$*-submodule of* $\Gamma(B)$ *generated by* $X \cup \pi_1^{-1}(\mathrm{Tor}_1) \cup \pi_2^{-1}(\mathrm{Tor}_2)$.

Proof. This pure $\mathcal{O}$-submodule together with the field it generates is a Γ-subfield of B with the same kernels as B, so it suffices to see that it is contained in $\Gamma(A')$ for any A' in the definition of $\langle X\rangle_B$.

$\Gamma(A')$ contains X by definition, and since $\pi_i(\Gamma(A')) = \mathrm{Tor}_i$ and A' has the same kernels as B, it also contains $\pi_i^{-1}(\mathrm{Tor}_i)$. Hence it also contains $\mathrm{Tor}(G) \cap \Gamma(B)$. Since it is divisible, it follows that it is pure in $\Gamma(B)$. $\square$

3C. *Good bases.* Let A be a Γ-field, and B a finitely generated Γ-field extension of A. So the linear dimension $\operatorname{ldim}_{k_{\mathcal{O}}}(\Gamma(B)/\Gamma(A))$ is finite. Thus we can find a *basis* for the extension, by which we mean a tuple $b = (b_1, \ldots, b_n) \in \Gamma(B)^n$ of minimal length n such that $b \cup \Gamma(A)$ generates $\Gamma(B)$, or equivalently such that $b_1 + \Gamma(A), \ldots, b_n + \Gamma(A)$ is a basis for the quotient $k_{\mathcal{O}}$-vector space $\Gamma(B)/\Gamma(A)$.

We consider the locus $\mathrm{Loc}(b/A)$ of b, that is, the smallest Zariski-closed subset of G, defined over A and containing b.

Definition 3.19. A basis $b \in \Gamma(B)^n$ for a finitely generated extension $A \hookrightarrow B$ of Γ-fields is *good* if the isomorphism type of the extension is determined up to isomorphism by the locus $\mathrm{Loc}(b/A)$. That is, whenever B' is another extension of A which is generated by a basis b' such that $\mathrm{Loc}(b'/A) = \mathrm{Loc}(b/A)$ then there is an isomorphism of Γ-fields $B \cong B'$ fixing A pointwise which takes b to b'.

Proposition 3.20. *Suppose we are in case* (DEQ), *that is*, $\mathrm{Tor}(G) \subseteq \Gamma$. *Let $A \hookrightarrow B$ be a finitely generated extension of Γ-fields. Then every basis of the extension is good.*

Proof. Suppose b is a basis of B over A, and we have another extension B' of A with basis b' such that $\mathrm{Loc}(b'/A) = \mathrm{Loc}(b/A)$. There is a (not necessarily unique) field isomorphism $\theta : B \cong B'$ over A which takes b to b'. Now, for an element $c \in G(B)$, we have $c \in \Gamma(B)$ if and only if there is $m \in \mathbb{N}$ such that mc is in the $\mathcal{O}$-linear span of $\Gamma(A)$ and b, because $\Gamma(B)$ is divisible and contains all the torsion of G. It follows that $c \in \Gamma(B)$ if and only if $\theta(c) \in \Gamma(B')$, so θ is an isomorphism of Γ-field extensions. So b is a good basis. □

In the proof it is critical that $\mathrm{Tor}(G) \subseteq \Gamma$ since otherwise some division points of the basis will be in Γ but others will not. In general we can specify an extension B of A by specifying a choice of division sequence $\hat{b}$ for a basis b such that $\hat{b} \in \hat{\Gamma}(B)$.

Definition 3.21. A Γ-field is *essentially finitary* if it is finitely generated or if it is a finitely generated extension of a countable full Γ-field.

Proposition 3.22 (existence of good bases). *Let A be an essentially finitary* Γ*-field*, *and let B be a finitely generated Γ-field extension of A* (*with the same kernels as A*). *Let b be a basis for the extension. Then there is $m \in \mathbb{N}^+$ such that any m-th division point of b in $\Gamma(B)$ is a good basis. Furthermore*, *in case* (DEQ) *we may take $m = 1$*, *so every basis is good*, *and we may even remove the assumption that A is essentially finitary.*

The bulk of the proof is contained in the following Kummer-theoretic results.

Definition 3.23. For a commutative algebraic group H, we write $\widehat{T}(H)$ for the kernel of the map $\rho_H : \widehat{H} \to H$. So $\widehat{T}(H)$ is the group of division sequences of the identity of H (which is the product over primes l of the l-adic Tate modules $T_l(H)$ of H, whence the notation).

Proposition 3.24. *Let $H = A \times \mathbb{G}_m^r$ be the product of an abelian variety and an algebraic torus. Suppose that A is defined over a number field K_0, and moreover that every endomorphism of A is also defined over K_0. Let D be either* $\mathrm{Tor}(H)$ *or $H(L)$ for an algebraically closed field extension L of K_0 and let K be a finitely generated field extension of $K_0(D)$. Let $a \in H(K)$ and suppose that a is free in H over D*, *that is*, *in no coset $H'+\gamma$ for a proper algebraic subgroup H' of H and $\gamma \in D$. Let $\hat{a} = (a_m)_{m\in\mathbb{N}^+}$ be a division sequence for a in $\widehat{H}(K^{\mathrm{alg}})$ and consider the Kummer map $\xi_a :$* $\mathrm{Gal}(K^{\mathrm{alg}}/K) \to \widehat{T}(H)$ *given by*

$$\xi_a(\sigma) = (\sigma(a_m) - a_m)_{m\in\mathbb{N}^+}.$$

Then ξ_a does not depend on the choice of division sequence $\hat{a}$, *so is well-defined*, *and the image of ξ_a is of finite index in $\widehat{T}(H)$.*

Remark 3.25. For the groups $\widehat{T}(H)$ which occur in this theorem, the finite index subgroups are precisely those which are open in the profinite topology, so the conclusion of the proposition is that $\xi_a(\mathrm{Gal}(K^{\mathrm{alg}}/K))$ is open in $\widehat{T}(H)$.

Proof of Proposition 3.24. It is straightforward that ξ_a is well-defined.

First, suppose $D = \mathrm{Tor}(H)$ and K is a *finite* extension of $K_0(D)$ — so by increasing K_0, we may assume $K = K_0(D)$. The result then follows from Kummer theory for abelian varieties. For the case $H = A$, we refer to [Bertrand 2011, Theorem 5.2], and for the generalization to $H = A \times \mathbb{G}_{\mathrm{m}}^n$ we refer to [Bays et al. 2014c, Proposition A.9].

Suppose now that $D = H(L)$, where L is an algebraically closed field. In this case, the result has a Galois-theoretic proof given as [Bays et al. 2014a, Section 3, Claim 2]. In the case that $K = K_0(D, a)$, the result follows directly from that claim; in general, it follows on noting that

$$\xi_a(\mathrm{Gal}(K^{\mathrm{alg}}/K)) \cong \mathrm{Gal}(K(\hat{a})/K) \cong \mathrm{Gal}(K_0(D, \hat{a})/K \cap K_0(D, \hat{a})),$$

and $K \cap K_0(D, \hat{a})$ is a finite extension of $K_0(D, a)$.

See also references in the introduction of [Bays et al. 2014a] for alternative proofs, and [Bertrand 2011, Theorem 5.3] for an analytic proof.

Finally, suppose $D = \mathrm{Tor}(H)$ and K is a finitely generated extension of $K_0(D)$. The result in this case follows from the first two cases. This can be seen model-theoretically in the context of [Bays et al. 2014c] as a matter of transitivity of atomicity, but we give here a direct argument.

Say B is the minimal algebraic subgroup of H such that, writing $\theta : H \to H/B$ for the quotient map, we have $\theta(a) \in (H/B)(\mathbb{Q}^{\mathrm{alg}})$. Let $K' = K \cap \mathbb{Q}^{\mathrm{alg}}$, so K is a regular extension of K' and K' is a finite extension of $K_0(D)$. Consider the diagram

$$\begin{tikzcd}
1 \arrow[d] & 0 \arrow[d] \\
\mathrm{Gal}(K^{\mathrm{alg}}/\mathbb{Q}^{\mathrm{alg}}(K)) \arrow[r, "\xi_a"] \arrow[d] & \widehat{T}(B) \arrow[d] \\
\mathrm{Gal}(K^{\mathrm{alg}}/K) \arrow[r, "\xi_a"] \arrow[d] & \widehat{T}(H) \arrow[d] \\
\mathrm{Gal}(\mathbb{Q}^{\mathrm{alg}}/K') \arrow[r, "\xi_{\theta(a)}"] \arrow[d] & \widehat{T}(H/B) \arrow[d] \\
1 & 0
\end{tikzcd}$$

where the middle horizontal map is the Kummer map for a, the top map is its restriction, and the bottom map is the Kummer map for $\theta(a)$. The vertical sequences are exact.

Say $a \in a_B + H(\mathbb{Q}^{\mathrm{alg}})$, where $a_B \in B$. By minimality of B, we have that a_B is free in B over $B(\mathbb{Q}^{\mathrm{alg}})$. Now the top map agrees with the Kummer map ξ_{a_B} in B, and so by the second case above, the map has finite index image. Now since a is free in H over $\mathrm{Tor}(H)$, we have that $\theta(a)$ is free in H/B over $\mathrm{Tor}(H/B)$, so by the first case applied to H/B, the bottom map in the above diagram also has finite index image. It follows that the central map has finite index image, as required. □

Now we prove that good bases exist.

Proof of Proposition 3.22. Let $\hat{b} \in \widehat{\Gamma}(B)^n$ be a division sequence of the basis b, and write $\hat{b} = (b_m)_{m\in\mathbb{N}^+}$. Then $\Gamma(B)$ is precisely the $\mathcal{O}$-linear span of $\Gamma(A)$ and the b_m, so to specify B up to isomorphism it is enough to specify the ACF-type of $\hat{b}$ over A. A is an essentially finitary Γ-field, so it is either finitely generated or a finitely generated extension of a countable full Γ-field A_0. In the former case, let $D = \mathrm{Tor}(G)$ and in the latter case let $D = G(A_0)$. For $i = 1, 2$, write $b_i = \pi_i(b)$ and $D_i = \pi_i(D)$, and let a_i be a $k_{\mathcal{O}}$-basis for $\pi_i(\Gamma(A))$ over D_i.

We consider the different cases in turn.

Case (EXP)**:** Since the extensions are kernel-preserving, (a_2, b_2) is $k_{\mathcal{O}}$-linearly independent over D_2, and so is free in G_2^{n+k} over D_2.

So, by Proposition 3.24, $\xi_{a_2,b_2}\big(\mathrm{Gal}(K_0(D, a, b)^{\mathrm{alg}}/K_0(D, a, b))\big)$ has finite index in $\widehat{T}(G_2^{n+k})$. In particular, its intersection with $0 \times \widehat{T}(G_2^n)$ is of finite index.

Since A is generated as a field by $K_0(D, a)$ and the division points of a_2, it follows that the image $\Xi := \xi_{b_2}\big(\mathrm{Gal}(A(b)^{\mathrm{alg}}/A(b))\big)$ has finite index in $\widehat{T}(G_2^n)$. So if m is the exponent of the finite quotient $\widehat{T}(G_2^n)/\Xi$, then $m\widehat{T}(G_2^n)$ is a subgroup of Ξ.

Hence, if b' is an m-th division point of b we have $\xi_{b'_2}\big(\mathrm{Gal}(A(b')^{\mathrm{alg}}/A(b'))\big) = \widehat{T}(G_2^n)$. So all division sequences of b' have the same ACF-type over $A(b')$, and hence b' is a good basis for B over A.

Case (COR)**:** Again, since the extensions are kernel-preserving, (a_i, b_i) is free over D_i for $i = 1, 2$. Since G_1 and G_2 are simple and nonisogenous, every algebraic subgroup of G^{k+n} is of the form $H_1 \times H_2$ for H_i a subgroup of G_i^{k+n}, so it follows that (a, b) is free in G^{k+n} over D. Since A is generated as a field by $K_0(D)$ and the division points of a_1 and of a_2, we conclude as in case (EXP).

Case (DEQ)**:** This was covered in Proposition 3.20. □

Corollary 3.26. *If A is an essentially finitary Γ-field there are*, *up to isomorphism*, *only countably many finitely generated kernel-preserving extensions of A.*

Proof. Each extension B has a good basis b, and is determined by $\mathrm{Loc}(b/A)$. Since A is countable there are only countably many algebraic varieties defined over it. □

4. Predimension and strong extensions

4A. ***Predimension.*** We define a predimension function δ as follows:

Definition 4.1. Let $A \subseteq B$ be Γ-fields. For any Γ-subfield X of B that is finitely generated over A, let

$$\delta(X/A) := \operatorname{trd}(X/A) - d \operatorname{ldim}_{k_{\mathcal{O}}}(\Gamma(X)/\Gamma(A)),$$

recalling $d = \dim G_1 = \dim G_2$.

Note that since X is assumed to be finitely generated over A, the linear dimension $\operatorname{ldim}_{k_{\mathcal{O}}}(\Gamma(X)/\Gamma(A))$ is finite, and since $\mathcal{O}$ acts by K_0-definable functions and X is the field generated by $\Gamma(X)$, the transcendence degree $\operatorname{trd}(X/A)$ is also finite. Hence the predimension is well-defined.

As a convention, for any finite $b \subset \Gamma(B)$, we set

$$\delta(b/A) := \delta(X/A),$$

where $X = \langle Ab \rangle$, the Γ-subfield of B generated by $b \cup A$.

Note that $\delta(b/A) = \operatorname{trd}(b/A) - d \operatorname{ldim}_{k_{\mathcal{O}}}(b/\Gamma(A))$.

Lemma 4.2. *Let $A \subseteq B$ be Γ-fields.*

(1) Finite character for δ:
If $b \subseteq \Gamma(B)$ is finite, there is a finitely generated Γ-subfield A_0 of A such that for any intermediate Γ-field $A_0 \subseteq A' \subseteq A$, we have $\delta(b/A) = \delta(b/A')$.

(2) Addition formula for δ:
Let X, Y be Γ-subfields of B finitely generated over A with $X \subseteq Y$. Then

$$\delta(Y/A) = \delta(Y/X) + \delta(X/A).$$

(3) Submodularity of δ:
Suppose X, Y are Γ-subfields of B with X finitely generated over $X \wedge Y$. Then, abbreviating $\langle X \cup Y \rangle$ by XY, we have

$$\delta(XY/Y) \leqslant \delta(X/X \wedge Y).$$

Proof. (1) Immediate since transcendence degree and $k_{\mathcal{O}}$-linear dimension have finite character.

(2) Note that the addition formula holds with transcendence degree or linear dimension in place of δ, so it also holds for δ by linearity.

(3) The submodularity condition is true when δ is replaced by transcendence degree. Linear dimension is modular, which means

$$\operatorname{ldim}_{k_{\mathcal{O}}}(\Gamma(XY)/\Gamma(Y)) = \operatorname{ldim}_{k_{\mathcal{O}}}(\Gamma(X)/\Gamma(X \wedge Y)),$$

so by subtracting we get the required submodularity of δ. □

4B. *Strong extensions.*

Definition 4.3. An extension $A \subseteq B$ of Γ-fields is said to be a *strong extension* if for every Γ-subfield X of B that is finitely generated over A, $\delta(X/A) \geqslant 0$. In this case, we also say that A is a *strong Γ-subfield* of B, and write $A \lhd B$.

For arbitrary Γ-fields A, B, an embedding $A \hookrightarrow B$ is said to be a *strong embedding* if the image of A is a strong Γ-subfield of B. To denote that an embedding is strong we use the notation $A \overset{\triangleleft}{\hookrightarrow} B$.

The method of predimensions and strong (also known as self-sufficient) extensions has been widely used since it was introduced by Hrushovski [1993]. We now give a few basic results which are well-known in general, but fundamental to the later development so it would be inappropriate to omit them. Some of the proofs are slightly more involved for this setting than the more well-known settings, especially those where no field is present.

Lemma 4.4. *The composition of strong embeddings is strong.*

Proof. Suppose $A \triangleleft B$ and $B \triangleleft C$. Clearly the kernels of C are the same as those of A, since both are the same as those of B. Let $X \subseteq C$ be finitely generated over A. Then $\delta(X/A) = \delta(X/X \wedge B) + \delta(X \wedge B/A)$ by the addition formula. We have $\delta(X/X \wedge B) \geqslant \delta(XB/B)$ by submodularity, and $\delta(XB/B) \geqslant 0$ because $B \triangleleft C$. Also, $\delta(X \wedge B/A) \geqslant 0$ because $A \triangleleft B$. So $\delta(X/A) \geqslant 0$. □

Given a strong extension $A \triangleleft B$ of Γ-fields, and an intermediate Γ-field X, finitely generated over A, it follows that $A \triangleleft X$ but it may not be the case that $X \triangleleft B$. However, as Y varies over finitely generated extensions of X inside B, the predimension $\delta(Y/A)$ takes integer values bounded below by 0 because $A \triangleleft B$. Thus we can replace X by a finitely generated extension X' of X, inside B, such that $\delta(X'/A)$ is minimal, and from the addition formula for δ it follows that $X' \triangleleft B$.

The next lemma shows that we can find this X' in a canonical way. It is crucial for understanding the finitely generated Γ-fields we will amalgamate, and it will allow us to understand the types in our models and prove there are only countably many of them.

Lemma 4.5. *Suppose B is a Γ-field and for each $j \in J$, A_j is a strong Γ-subfield of B. Then $\bigwedge_{j \in J} A_j$ is also strong in B.*

Proof. First we prove that if $A_1, A_2 \triangleleft B$ then $A_1 \wedge A_2 \triangleleft A_1$. So suppose X is a finitely generated Γ-field extension of $A_1 \wedge A_2$ inside A_1. Then

$$\delta(X/A_1 \wedge A_2) = \delta(X/X \wedge A_2) \geqslant \delta(XA_2/A_2) \geqslant 0$$

using submodularity and the fact that $A_2 \triangleleft B$. So $A_1 \wedge A_2 \triangleleft A_1$, but $A_1 \triangleleft B$ so, by Lemma 4.4, $A_1 \wedge A_2 \triangleleft B$. It follows by induction that if J is finite, $\bigwedge_{j \in J} A_j \triangleleft B$.

Now suppose that J is infinite and that X is a Γ-subfield of B which is finitely generated as an extension of $A = \bigwedge_{j \in J} A_j$. Then we have

$$A = A \wedge X = \bigwedge_{j \in J} (A_j \wedge X).$$

Each Γ-field $A_j \wedge X$ is in the lattice of Γ-fields intermediate between A and X. This lattice is isomorphic to the lattice of vector subspaces of the finite-dimensional vector space $\Gamma(X)/\Gamma(A)$ and so has no infinite chains. Thus there is a finite subset J_0 of J such that, writing $A_{J_0} = \bigwedge_{j \in J_0} A_j$, we have

$$A = \bigwedge_{j \in J} (A_j \wedge X) = \bigwedge_{j \in J_0} (A_j \wedge X) = A_{J_0} \wedge X.$$

Now using the result for finite intersections we have that $A_{J_0} \lhd B$, so using also submodularity we have

$$\delta(X/A) = \delta(X/A_{J_0} \wedge X) \geqslant \delta(A_{J_0}X/A_{J_0}) \geqslant 0,$$

and hence $A \lhd B$ as required. □

Consider again a strong extension $A \lhd B$ of Γ-fields, and an intermediate Γ-field X.

Definition 4.6. We define the *hull* of X in B (also known as the strong closure of X or the self-sufficient closure of X) by

$$\lceil X \rceil_B := \bigwedge \{Y \text{ a strong } \Gamma\text{-subfield of } B \mid X \subseteq Y\}.$$

The previous lemma shows that $\lceil X \rceil_B$ is indeed strong in Y, and we observe also that if X is finitely generated as an extension of A then so is $\lceil X \rceil_B$. Furthermore, if $B \lhd C$ then it is immediate that $\lceil X \rceil_C = \lceil X \rceil_B$, so we often drop the subscript B.

Lemma 4.7. *The hull operator has finite character. That is, if $A \lhd B$ and X is an intermediate Γ-field,*

$$\lceil X \rceil_B = \bigcup \{\lceil X_0 \rceil_B \mid X_0 \subseteq X \textit{ and } X_0 \textit{ is a finitely generated extension of } A\}.$$

Proof. Let U be the union in the statement of the lemma. It is immediate from the definition of the hull that if $X_0 \subseteq X$ then $\lceil X_0 \rceil_B \subseteq \lceil X \rceil_B$. It follows that $U \subseteq \lceil X \rceil_B$. Also $X \subseteq U$. Now U is a directed union of strong Γ-subfields of B, and since δ has finite character, it follows that $U \lhd B$. So $\lceil X \rceil_B \subseteq U$, as required. □

Finally in this section we give a useful lemma giving a simple sufficient condition for an extension of a strong Γ-subfield also to be strong.

Lemma 4.8. *If $A \lhd B$ and $A \subseteq A' \subseteq B$ with $\delta(A'/A) = 0$, then $A' \lhd B$.*

Proof. Let $X \subseteq B$ be a finitely generated extension of A'. Then

$$\delta(X/A') = \delta(X/A) - \delta(A'/A) = \delta(X/A) \geqslant 0. \qquad \square$$

4C. *Pregeometry.* In this section, F is any full Γ-field strongly extending a Γ-subfield F_{base}. We will use the predimension function δ to define a pregeometry on F. We could drop the assumptions that F is full and that F strongly extends some F_{base} and give a definition along the lines of that done for exponential fields in [Kirby 2010a] and for Weierstrass $\wp$-functions in [Jones et al. 2016]. However, it is sufficient for our purposes and much more straightforward to do it this way.

Definition 4.9. A Γ-subfield A of F, extending F_{base}, is Γ-*closed in* F, written $A \lhd_{\text{cl}} F$, if for any $A \subseteq B \subseteq F$ with B finitely generated over A and $\delta(B/A) \leqslant 0$ we have $B = A$.

Lemma 4.10. (1) *If $A \lhd_{\text{cl}} F$ then $A \lhd F$.*

(2) *If $A \lhd_{\text{cl}} F$ then A is a full Γ-subfield of F.*

(3) *If $A_j \lhd_{\text{cl}} F$ for $j \in J$ and $A = \bigwedge_{j \in J} A_j$ then $A \lhd_{\text{cl}} F$.*

Proof. (1) Immediate.

(2) Suppose $a \in G_1(F)$ is algebraic over A. Since F is full, there is $b \in G_2(F)$ with $(a, b) \in \Gamma(F)$. We have $\operatorname{trd}((a, b)/A) = \operatorname{trd}(b/A) \leqslant \dim G_2 = d$, so $\delta((a, b)/A) \leqslant d - d \operatorname{ldim}_{k_\mathcal{O}}((a, b)/\Gamma(A))$. If $\operatorname{ldim}_{k_\mathcal{O}}((a, b)/\Gamma(A)) = 1$ then $\delta((a, b)/A) \leqslant 0$, so since A is closed in F we have $(a, b) \in \Gamma(A)$. Otherwise $\operatorname{ldim}_{k_\mathcal{O}}((a, b)/\Gamma(A)) = 0$ so again $(a, b) \in \Gamma(A)$. A similar argument is used if $b \in G_2(F)$ is algebraic over A. Since $G_1(A)$ contains all points of G_1 that are algebraic over A, A is an algebraically closed field. Thus A is a full Γ-field.

(3) Suppose $\delta(B/A) \leqslant 0$. By submodularity and Lemma 4.5, for each j we have

$$\delta(BA_j/A_j) \leqslant \delta(B/A_j \wedge B) = \delta(B/A) - \delta(A_j \wedge B/A) \leqslant 0,$$

so $B \subseteq A_j$. Thus $B \subseteq A$. □

This notion of Γ-closedness induces a closure operator on the field F.

Definition 4.11. If $A \subseteq F$ is any subset, the Γ-*closure of A in F* is defined to be the smallest Γ-closed Γ-subfield containing A:

$$\Gamma\mathrm{cl}^F(A) := \bigwedge\{B \lhd_{\mathrm{cl}} F \mid A \subseteq B\}.$$

$\Gamma\mathrm{cl}^F(A)$ is a Γ-subfield of F, and in particular a subset of F, so $\Gamma\mathrm{cl}^F$ induces a map $\mathcal{P}F \to \mathcal{P}F$, which we also denote by $\Gamma\mathrm{cl}^F$.

Lemma 4.12. *For any Γ-subfield A of F, we have $\Gamma\mathrm{cl}^F(A) = \bigcup \mathcal{B}$, where $\mathcal{B}$ is the set of all Γ-subfields $B \subseteq F$ such that B is a finitely generated Γ-field extension of $\lceil A \rceil_F$ and $\delta(B/\lceil A \rceil_F) = 0$.*

Proof. Since $\Gamma\mathrm{cl}^F(A) \lhd F$ we have $\lceil A \rceil_F \subseteq \Gamma\mathrm{cl}^F(A)$. So $\Gamma\mathrm{cl}^F(A) = \Gamma\mathrm{cl}^F(\lceil A \rceil_F)$, and thus we may assume $A \lhd F$. Let $C = \bigcup \mathcal{B}$. Using the submodularity of δ it is easy to see that the system $\mathcal{B}$ of Γ-subfields of F is directed, so its union C is a Γ-subfield of F.

Suppose that b is a finite tuple from $\Gamma(F)$ such that $\delta(b/C) \leqslant 0$. Then by the finite character of δ and directedness of the union defining C, there is a finitely generated extension B of A inside C such that $\delta(B/A) = 0$ and $\delta(b/B) = \delta(b/C)$. Using the addition formula,

$$0 \geqslant \delta(b/B) = \delta(b/A) - \delta(B/A) = \delta(b/A) \geqslant 0.$$

So $\delta(b/A) = 0$ and hence $b \in \Gamma(C)$. Thus C is Γ-closed, so $\Gamma\mathrm{cl}^F(A) \subseteq C$.

Now suppose B is a finitely generated Γ-field extension of A with $\delta(B/A) = 0$ and $A \subseteq D \lhd_{\mathrm{cl}} F$. Then

$$\delta(BD/D) \leqslant \delta(B/B \wedge D) = \delta(B/A) - \delta(B \wedge D/A) \leqslant 0$$

because $A \lhd F$. So $\delta(B \wedge D/A) \geqslant 0$ and thus $B \subseteq D$. Hence $B \subseteq \Gamma\mathrm{cl}^F(A)$, and so $C \subseteq \Gamma\mathrm{cl}^F(A)$. □

The predimension function δ is a function depending on the sort Γ, but Γ-closure will be shown to be a pregeometry on the field sort. The next lemma allows us to move from one sort to the other.

Lemma 4.13. *If $A \lhd F$ and $a \in F \smallsetminus \Gamma\mathrm{cl}^F(A)$, there is $\alpha \in \Gamma(F)$ such that $\pi_1(\alpha) \in G_1(F)$ is interalgebraic with a over A and $\delta(\alpha/A) = 1$, and the Γ-subfield $\langle A\alpha \rangle$ of F generated by A and α satisfies $\langle A\alpha \rangle \lhd F$. We can choose α such that a is rational over $A(\pi_1(\alpha))$, and if A is essentially finitary, also such that α is* *a good basis. Furthermore, the locus* $\mathrm{Loc}(\alpha, a/A)$ *can be taken to be a particular algebraic curve defined over* K_0*, not depending on* A *or* a.

Proof. We have $\mathrm{trd}(a/A) = 1$ (since otherwise $a \in \Gamma\mathrm{cl}^F(A)$). We have fixed an identification of G_1 with a constructible subset of affine space $\mathbb{A}^N$ for some N, defined over K_0. Let f be the projection map from G_1 to the first coordinate where the projection is dominant, and choose a constructible curve $X \subseteq G_1$ by fixing the values of all the other coordinates to be values in K_0. Then X and the map f are defined over K_0. Choose $\alpha_1 \in X(F)$ with $f(\alpha_1) = a$. Then α_1 is interalgebraic with a over A, with a rational over $A(\alpha_1)$, and the locus $\mathrm{Loc}(\alpha_1, a/A)$ is defined over K_0.

Since F is full, there is $\alpha \in \Gamma(F)$ with $\pi_1(\alpha) = \alpha_1$. Then $\alpha \in \Gamma\mathrm{cl}^F(Aa)$ and $a \in \Gamma\mathrm{cl}^F(A\alpha)$. Then $\mathrm{ldim}_{k_\mathcal{O}}(\alpha/\Gamma(A)) = 1$ and so since $\delta(\alpha/A) \neq 0$ by Lemma 4.12, $\mathrm{trd}(\alpha/A) = d + 1$ and $\delta(\alpha/A) = 1$. If there were $B \supseteq \langle A\alpha\rangle$ with $\delta(B/A\alpha) < 0$ then $\delta(B/A) \leqslant 0$, which contradicts $a \notin \Gamma\mathrm{cl}^F(A)$. So $\langle A\alpha\rangle \lhd F$. If A is essentially finitary then by Proposition 3.22 we can divide α by some $m \in \mathbb{N}^+$ to ensure it is a good basis. Since α_2 is generic in $G_2(A)$ over α_1, and G_2 is defined over K_0, we deduce that $\mathrm{Loc}(\alpha, a/A)$ is defined over K_0. $\square$

Proposition 4.14. *The* Γ*-closed subsets of* F *are the closed sets of a pregeometry on* F.

Proof. It is immediate that for any subsets $A \subseteq B$ of F we have $A \subseteq \Gamma\mathrm{cl}^F(A)$, $\Gamma\mathrm{cl}^F(\Gamma\mathrm{cl}^F(A)) = \Gamma\mathrm{cl}^F(A)$, and $\Gamma\mathrm{cl}^F(A) \subseteq \Gamma\mathrm{cl}^F(B)$.

For finite character, suppose $b \in \Gamma\mathrm{cl}^F(A)$. By Lemma 4.12 there is a finitely generated extension $\lceil A\rceil_F \subseteq B$ in F such that $\delta(B/\lceil A\rceil_F) = 0$ and $b \in B$. Then there is a finite tuple $\beta \in \Gamma(F)$ with b rational over β and $\delta(\beta/A) = 0$. By finite character of δ from Lemma 4.2, there is a finitely generated Γ-subfield A_0 of $\lceil A\rceil_F$ such that for any A' with $A_0 \subseteq A' \subseteq \lceil A\rceil_F$ we have $\delta(\beta/A') = 0$. So by Lemma 4.12 again, $b \in \Gamma\mathrm{cl}^F(A_0)$.

The hull operator is a closure operator which by Lemma 4.7 has finite character. We have $A_0 \subseteq \lceil A\rceil_F$, so there is a finite subset A_{00} of A such that $\lceil A_0\rceil_F = \lceil A_{00}\rceil_F$. Hence $b \in \Gamma\mathrm{cl}^F(A_{00})$, and so $\Gamma\mathrm{cl}^F$ has finite character.

For exchange, suppose $A \lhd_{\mathrm{cl}} F$ and that $a, b \in F \smallsetminus A$ with $b \in \Gamma\mathrm{cl}^F(Aa)$. Using Lemma 4.13, we choose $\alpha, \beta \in \Gamma(F)$ corresponding to a and b, respectively.

Now $\beta \in \Gamma\mathrm{cl}^F(A\alpha)$ so there is a finitely generated Γ-field extension $A \subseteq B$ inside F with $\beta, \alpha \in \Gamma(B)$ and $\delta(B/A\alpha) = 0$. Then we have

$$\delta(B/A\beta) = \delta(B/A) - \delta(\beta/A) = \delta(B/A) - 1 = \delta(B/A) - \delta(\alpha/A) = \delta(B/A\alpha) = 0,$$

so $\alpha \in \Gamma\mathrm{cl}^F(A\beta)$, or equivalently $a \in \Gamma\mathrm{cl}^F(Ab)$. $\square$

We write $\Gamma\mathrm{dim}^F$ for the dimension with respect to the pregeometry $\Gamma\mathrm{cl}^F$. However, if F_1 and F_2 are both full Γ-fields with $F_1 \lhd_{\mathrm{cl}} F_2$ and $A \subseteq F_1$ then $\Gamma\mathrm{cl}^{F_1}(A) = \Gamma\mathrm{cl}^{F_2}(A)$. So from now on we will usually drop the superscript F and just write $\Gamma\mathrm{cl}$ and $\Gamma\mathrm{dim}$ except where it might cause confusion.

We have the usual connection between the dimension and the predimension function.

Lemma 4.15. *Suppose that $A \lhd F$ and B is a finitely generated Γ-field extension of A in F. Then*

(1) $\Gamma\dim(B/A) = \min\{\delta(C/A) \mid B \subseteq C \subseteq F\}$, *and*

(2) $B \lhd F$ *if and only if* $\Gamma\dim(B/A) = \delta(B/A)$.

Proof. Since $\lceil B \rceil_F \subseteq \Gamma\mathrm{cl}^F(B)$, we have $\Gamma\dim(B/A) = \Gamma\dim(\lceil B \rceil_F/A)$. Now it follows from the addition formula that $\delta(\lceil B \rceil_F/A) = \min\{\delta(C/A) \mid B \subseteq C \subseteq F\}$, so statement (1) reduces to the left-to-right direction of statement (2). To prove that, first assume $B \lhd F$.

Let $n = \Gamma\dim(B/A)$ and let $b_1, \dots, b_n$ be a Γcl-basis for B over A. Applying Lemma 4.13, we get $\beta_i \in \Gamma(F)$ in the closure of B with β_i corresponding to b_i. Let D be the Γ-subfield of F generated by A and $\beta_1, \dots, \beta_n$. Then $\delta(D/A) = n$ and $D \lhd F$. Furthermore, $\Gamma\mathrm{cl}^F(D) = \Gamma\mathrm{cl}^F(B)$.

Since $D \subseteq \Gamma\mathrm{cl}^F(B)$, there is $C \supseteq B \cup D$ such that $\delta(C/B) = 0$. Then

$$\delta(B/A) = \delta(C/A) - \delta(C/B) = \delta(C/A) \geqslant \delta(D/A) = n,$$

using that $D \lhd F$. Reversing the roles of B and D, the same argument shows that $\delta(B/A) \leqslant \delta(D/A)$, and so $\delta(B/A) = n = \Gamma\dim(B/A)$ as required.

The right-to-left direction of statement (2) now follows from statement (1) and the addition property. □

Remarks 4.16. (1) In the sense of the pregeometry Γcl, the set $\Gamma(F)$ is d-dimensional. Thus when $d = 1$ such as in pseudo-exponentiation and pseudo-$\wp$, we actually get a pregeometry directly on $\Gamma(F)$.

(2) In the case of pseudo-exponentiation or a pseudo-$\wp$-function, $G_1(F) = \mathbb{G}_a(F) = F$, and Γ is the graph of a function exp, so we have a bijection $\varphi : F \to \Gamma(F)$ given by $x \mapsto (x, \exp(x))$. The predimension usually considered for exponentiation, for example in [Zilber 2005b], is a function on tuples from the field sort, and in fact is just the composite $\delta \circ \varphi$ of the predimension function described here with φ.

(3) It is possible to define a predimension function directly on the field sort, even in our generality. Given any subfield A of F we write $\Gamma(A)$ for $\Gamma(F) \cap G(A)$. For any subset X of F (in the field sort) we write X^{alg} for the field-theoretic algebraic closure of $F_{\mathrm{base}}(X)$ in F.

Given subsets X, Y of F, with $\mathrm{trd}((X \cup Y)^{\mathrm{alg}}/X^{\mathrm{alg}}) < \infty$, define

$$\eta(Y/X) = \mathrm{trd}((X \cup Y)^{\mathrm{alg}}/X^{\mathrm{alg}}) - \mathrm{ldim}_{k_{\mathcal{O}}}(\Gamma((X \cup Y)^{\mathrm{alg}})/\Gamma(X^{\mathrm{alg}})),$$

which takes values in $\mathbb{Z} \cup \{-\infty\}$.

The predimension functions η and δ are closely related, and we could write

$$\eta(Y/X) = \delta((X \cup Y)^{\mathrm{alg}}/X^{\mathrm{alg}})$$

except that X^{alg} usually fails to be a Γ-subfield of F by our definition, because as a field it is not usually generated by the coordinates of the points in $\Gamma(X^{\mathrm{alg}})$. It may not even be algebraic over the field generated by those points.

It is possible to define the notion of strong embeddings of Γ-fields using this predimension function instead. Some things are easier with this approach, because the predimension is defined on a 1-dimensional sort. However we choose to work in the sort Γ because it is a vector space and hence has a modular geometry, which makes other things much easier.

4D. *Full closures.* The following theorem and proof follow [Kirby 2013, Theorem 2.18].

Theorem 4.17. *If A is a Γ-field then there is a full Γ-field extension A^{full} of A such that $A \lhd A^{\mathrm{full}}$ and A^{full} is generated as a full Γ-field by A. Furthermore, if A is essentially finitary then A^{full} is unique up to isomorphism as an extension of A.*

Proof. First we prove existence. Embed A in a large algebraically closed field F. Choose a point $a \in G_1(F)$ which is algebraic over A and not in $\pi_1(\Gamma(A))$, if such exists. Choose a division sequence $\hat{a} \in \widehat{G}_1(F)$ for a. Let $b \in G_2(F)$ be generic over A and choose a division system $\hat{b}$ for it. (Up to field isomorphism over A, $\hat{b}$ is unique.) Let A' be the field generated by A and the division sequences $\hat{a} = (a_m)_{m \in \mathbb{N}^+}$ and $\hat{b} = (b_m)_{m \in \mathbb{N}^+}$, and define $\Gamma(A')$ to be the $\mathcal{O}$-submodule of $G(A')$ generated by $\Gamma(A)$ and the points (a_m, b_m) for $m \in \mathbb{N}^+$. Since $\pi_1(\Gamma(A))$ already contains the torsion of G_1, the extension preserves the kernels. So A' is a Γ-field extension of A. We have

$$\delta(A'/A) = \operatorname{trd}(b/A) - d \operatorname{ldim}_{k_{\mathcal{O}}}(b/A) = d - d = 0,$$

so it is a strong extension. Similarly, if there is $b \in G_2(F)$ which is algebraic over A but not in $\pi_2(\Gamma(A))$, we can form a similar strong extension. Iterating these constructions, a strong full extension A^{full} of A is readily seen to exist.

Now we prove uniqueness under the additional hypothesis that A is essentially finitary. Suppose that B and B' both satisfy the conditions for A^{full}. Since A is essentially finitary it is countable, and then the construction above shows that we can take B to be countable as well. Enumerate $\Gamma(B)$ as $(s_n)_{n \in \mathbb{N}^+}$ such that for each n, either $\pi_1(s_n)$ or $\pi_2(s_n)$ is algebraic over $A \cup \{s_1, \ldots, s_{n-1}\}$. This is possible since B is generated as a full Γ-field by A.

We inductively construct a chain of strong Γ-subfields $A_n \lhd B$, each a finitely generated Γ-field extension of A such that $A_0 = A$ and $s_n \in \Gamma(A_n)$. We also construct a chain of strong embeddings $\theta_n : A_n \overset{\lhd}{\hookrightarrow} B'$. Assume we have A_n and θ_n. Let A_{n+1} be the Γ-subfield of B generated by A_n and s_{n+1}. As a field, A_{n+1} is generated by A_n and the division points of s_{n+1}. If $s_{n+1} \in \Gamma(A_n)$, then we have $A_{n+1} = A_n$ and can just take $\theta_{n+1} = \theta_n$. Otherwise, we have $\operatorname{ldim}_{k_{\mathcal{O}}}(\Gamma(A_{n+1})/\Gamma(A_n)) \geqslant 1$. By hypothesis, one of $\pi_1(s_{n+1})$ or $\pi_2(s_{n+1})$ is algebraic over A_n, say $\pi_1(s_{n+1})$. Thus $\operatorname{trd}(A_{n+1}/A_n) = \operatorname{trd}(s_{n+1}/A_n) \leqslant d$. By inductive hypothesis $A_n \lhd B$, so we have $\delta(A_{n+1}/A_n) \geqslant 0$. It follows that $\operatorname{ldim}_{k_{\mathcal{O}}}(\Gamma(A_{n+1})/\Gamma(A_n)) = 1$ and $\operatorname{trd}(A_{n+1}/A_n) = d$, so $\delta(A_{n+1}/A_n) = 0$. Thus by Lemma 4.8, $A_{n+1} \lhd B$. Also, $\pi_2(s_{n+1})$ is generic in G_2 over A_n.

Since A_n is a finitely generated Γ-field extension of A, by Proposition 3.22 there is $m \in \mathbb{N}$ such that $\{s_{n+1}/m\}$ is a good basis for the extension $A_n \lhd A_{n+1}$. Replacing s_{n+1} by s_{n+1}/m, we may assume $m = 1$. Now let W be the locus of $\pi_1(s_{n+1})$ over A_n (a variety of dimension 0, irreducible over A_n, but not necessarily absolutely irreducible), and let w be any point in W^{θ_n}, the corresponding subvariety of $G_1(B')$. Choose $v \in G_2(B')$ such that $(w, v) \in \Gamma(B')$. Since w is algebraic over $\theta_n(A_n)$, which is strong in B', the same predimension argument as above shows that v is generic in G_2 over $\theta_n(A_n)$, so

$$\mathrm{Loc}(w, v/\theta_n(A_n)) = W^{\theta_n} \times G_2 = (\mathrm{Loc}(s_{n+1}/A_n))^{\theta_n}.$$

Since s_{n+1} is a good basis over A_n, we can extend θ_n to a field embedding $\theta_{n+1} : A_{n+1} \to B'$ with $\theta(s_{n+1}) = (w, v)$, and using again that $\theta_n(A_n) \lhd B'$ we get that $\Gamma(\theta_{n+1}(A_{n+1}))$ is generated by (w, v) over $\Gamma(\theta_n(A_n))$, and hence θ_{n+1} is a Γ-field embedding.

Also, $\delta(\theta_{n+1}(A_{n+1})/\theta_n(A_n)) = 0$ so θ_{n+1} is a strong embedding by Lemma 4.8.

Now $B = \bigcup\{A_n \mid n \in \mathbb{N}\}$ and $\bigcup\{\theta_n(A_n) \mid n \in \mathbb{N}\}$ is a full Γ-subfield of B' containing A, so it must be B'. Hence $\bigcup\{\theta_n \mid n \in \mathbb{N}\}$ is an isomorphism $B \cong B'$. So A^{full} is unique, up to isomorphism as an extension of A. □

Proposition 4.18. *Let A be a countable full Γ-field. Then there are only countably many finitely generated strong full Γ-field extensions of A, up to isomorphism.*

Proof. Let $A \lhd B$ be such an extension and let b be a finite tuple generating B over A as a full Γ-field, such that $B_0 := \langle Ab \rangle \lhd B$, and of minimal length such. Then by Proposition 3.22 we may replace b by b/m for some $m \in \mathbb{N}^+$ to ensure that b is a good basis for the extension $A \lhd B_0$. Then $B = B_0^{\mathrm{full}}$, which by Theorem 4.17 is determined uniquely up to isomorphism by B_0, and by Corollary 3.26 there are only countably many choices for B_0. □

5. The canonical countable model

5A. *The amalgamation theorem.* We use the definition of *amalgamation category* from [Kirby 2009], slightly extending work of Droste and Göbel [1992], who were themselves abstracting from Fraïssé's amalgamation theorem. We restrict to the countable case. We will apply the general theory to various categories of Γ-fields with strong embeddings as morphisms. The notions of *finitely generated*, *universal*, and *saturated* all have category-theoretic translations, which we give first.

Definition 5.1. Given a category $\mathcal{K}$, an object A of $\mathcal{K}$ is said to be $\aleph_0$*-small* if and only if for every ω-chain (Z_i, γ_{ij}) in $\mathcal{K}$ with direct limit Z_ω, any arrow $A \xrightarrow{f} Z_\omega$ factors through the chain, that is, there are $i < \omega$ and $A \xrightarrow{f^*} Z_i$ such that $f = \gamma_{i\omega} \circ f^*$. We write $\mathcal{K}^{<\aleph_0}$ for the full subcategory of $\aleph_0$-small objects of $\mathcal{K}$ and $\mathcal{K}^{\leqslant\aleph_0}$ for the full subcategory of the limits of ω-chains of $\aleph_0$-small objects of $\mathcal{K}$.

Definition 5.2. Given a category $\mathcal{K}$ and a subcategory $\mathcal{K}'$, an object U of $\mathcal{K}$ is said to be $\mathcal{K}'$*-universal* if for every object A of $\mathcal{K}'$ there is an arrow $A \to U$ in $\mathcal{K}$. U is $\mathcal{K}'$*-saturated* if for every arrow $A \xrightarrow{f} B$ in $\mathcal{K}'$ and every arrow $A \xrightarrow{g} U$ in $\mathcal{K}$, there is an arrow $B \xrightarrow{h} U$ in $\mathcal{K}$ such that $g = h \circ f$. U is $\mathcal{K}'$*-homogeneous* if for every object A of $\mathcal{K}'$ and every pair of arrows $A \xrightarrow{f,g} U$ in $\mathcal{K}$, there exists an isomorphism $U \xrightarrow{h} U$ in $\mathcal{K}$ such that $g = h \circ f$.

Some authors refer to $\mathcal{K}'$-saturation as *richness* with respect to the objects and arrows from $\mathcal{K}'$.

Definition 5.3. A category $\mathcal{K}$ is an *amalgamation category* if the following hold.

(AC1) Every arrow in $\mathcal{K}$ is a monomorphism.

(AC2) $\mathcal{K}$ has direct limits (unions) of ω-chains.

(AC3) $\mathcal{K}^{<\aleph_0}$ has at most $\aleph_0$ objects up to isomorphism.

(AC4) For each object $A \in \mathcal{K}^{<\aleph_0}$ there are at most $\aleph_0$ extensions of A in $\mathcal{K}^{<\aleph_0}$, up to isomorphism.

(AC5) $\mathcal{K}^{<\aleph_0}$ has the *amalgamation property* (AP), that is, any diagram of the form

$$\begin{tikzcd} B_1 & & B_2 \\ & A \arrow[ul] \arrow[ur] & \end{tikzcd}$$

can be completed to a commuting square

$$\begin{tikzcd} & C & \\ B_1 \arrow[ur] & & B_2 \arrow[ul] \\ & A \arrow[ul] \arrow[ur] & \end{tikzcd}$$

in $\mathcal{K}^{<\aleph_0}$.

(AC6) $\mathcal{K}^{<\aleph_0}$ has the *joint embedding property* (JEP), that is, for every $B_1, B_2 \in \mathcal{K}^{<\aleph_0}$ there is $C \in \mathcal{K}^{<\aleph_0}$ and arrows

$$\begin{tikzcd} & C & \\ B_1 \arrow[ur] & & B_2 \arrow[ul] \end{tikzcd}$$

in $\mathcal{K}^{<\aleph_0}$.

The point of the definition is that the following form of Fraïssé's amalgamation theorem holds.

Theorem 5.4 [Kirby 2009, Theorem 2.18]. *If $\mathcal{K}$ is an amalgamation category then there is an object $U \in \mathcal{K}^{\leqslant\aleph_0}$, the "Fraïssé limit", which is $\mathcal{K}^{\leqslant\aleph_0}$-universal and $\mathcal{K}^{<\aleph_0}$-saturated.*

Furthermore, if $A \in \mathcal{K}^{\leqslant\aleph_0}$ is $\mathcal{K}^{<\aleph_0}$-saturated then $A \cong U$.

Remark 5.5. It follows from saturation and a back-and-forth argument that U is also $\mathcal{K}^{<\aleph_0}$-homogeneous.

5B. ***Amalgamation of Γ-fields.*** We fix a Γ-field F_{base} which is either finitely generated as a Γ-field, or is a countable full Γ-field.

The identity map on a Γ-field is obviously a strong embedding; hence, from Lemma 4.4 we have a category of strong Γ-field extensions of F_{base}, with strong embeddings as the arrows. We write $\mathcal{C}(F_{\text{base}})$ for this category, but usually abbreviate it to $\mathcal{C}$. We also consider the following full subcategories of $\mathcal{C}$.

Notation 5.6.
- $\mathcal{C}^{\text{full}}$ (or $\mathcal{C}^{\text{full}}(F_{\text{base}})$) consists of the full strong Γ-field extensions of F_{base}.
- $\mathcal{C}^{\text{fg}}$ consists of the strong Γ-field extensions of F_{base} which are finitely generated.
- $\mathcal{C}^{\text{fg-full}}$ consists of the strong Γ-field extensions of F_{base} which are full and finitely generated as full extensions.
- $\mathcal{C}^{\leqslant\aleph_0}$ consists of the strong Γ-field extensions of F_{base} which are countable.
- $\mathcal{C}^{\text{full},\leqslant\aleph_0}$ consists of the strong Γ-field extensions of F_{base} which are full and countable.

For our categories $\mathcal{C}$ and $\mathcal{C}^{\text{full}}$, it is immediate that $\aleph_0$-small just means finitely generated in the appropriate sense, and a (full) Γ-field is the union of an ω-chain of finitely generated (full) Γ-fields if and only if it is countable.

We will construct our canonical model as the Fraïssé limit of $\mathcal{C}^{\text{fg}}$. In fact it is also the Fraïssé limit of $\mathcal{C}^{\text{fg-full}}$.

In proving the amalgamation property we actually prove a stronger result, asymmetric amalgamation, which will be necessary when we come to axiomatize our models. However, the asymmetric property holds only in the case of full Γ-fields, not for $\mathcal{C}^{\text{fg}}$. We also observe that our amalgams are disjoint.

Proposition 5.7. *The categories $\mathcal{C}^{\text{full}}$ and $\mathcal{C}^{\text{full},\leqslant\aleph_0}$ have the disjoint asymmetric amalgamation property. That is, given full Γ-fields A_0, A_L, $A_R \in \mathcal{C}^{\text{full}}$, an embedding $A_0 \hookrightarrow A_L$ and a strong embedding $A_0 \overset{\lhd}{\hookrightarrow} A_R$, there exist $A \in \mathcal{C}^{\text{full}}$ and dashed arrows making the following diagram commute*:

$$\begin{tikzcd}
& A_L \arrow[dr, hook, dashed, "\lhd"] & \\
A_0 \arrow[ur, hook] \arrow[dr, hook, "\lhd"'] & & A \\
& A_R \arrow[ur, hook, dashed] &
\end{tikzcd}$$

Moreover, if the embedding $A_0 \hookrightarrow A_L$ is also strong, then so is the embedding $A_R \hookrightarrow A$; furthermore, identifying A_0, A_L, and A_R with their images in A, we have that $A_L \cap A_R = A_0$.

Proof. Since A_0 is algebraically closed as a field, we may form the free amalgam A_1 of A_L and A_R over A_0 as fields, that is, the unique (up to isomorphism) field compositum of A_L and A_R in which they are algebraically independent over A_0. We identify A_L and A_R as subfields of A_1 so, in particular, $A_L \cap A_R = A_0$. We make A_1 into a Γ-field by defining $\Gamma(A_1)$ to be the $\mathcal{O}$-submodule $\Gamma(A_L) + \Gamma(A_R)$ of $G(A_1)$.

Then $\Gamma(A_L)$ and $\Gamma(A_R)$ are $\mathcal{O}$-submodules of $\Gamma(A_1)$.

Suppose that $a \in \ker_1(A_1)$, that is, $(a, 0) \in \Gamma(A_1)$. Then there are $(a_L, b_L) \in \Gamma(A_L)$ and $(a_R, b_R) \in \Gamma(A_R)$ such that $(a, 0) = (a_L, b_L) + (a_R, b_R)$. Then $b_L = -b_R$, so

$$b_L, b_R \in \Gamma_2(A_L) \cap \Gamma_2(A_R) \subseteq G_2(A_L) \cap G_2(A_R) = G_2(A_0).$$

Since A_0 is a full Γ-field there is $a_0 \in G_1(A_0)$ such that $(a_0, b_L) \in \Gamma(A_0)$. Then $a_L - a_0 \in \ker_1(A_L) = \ker_1(A_0)$, so $a_L \in \Gamma_1(A_0)$. Similarly, $a_R \in \Gamma_1(A_0)$, so $a \in \Gamma_1(A_0)$.

Thus $\ker_1(A_1) = \ker_1(A_0)$. The same argument shows that $\ker_2(A_1) = \ker_2(A_0)$, and hence the inclusions of A_L and A_R into A_1 preserve the kernels.

Let us check that the inclusion $A_L \hookrightarrow A_1$ is strong. Let X be a Γ-subfield of A_1 which is finitely generated over A_L. Choose a basis b for the extension, say of length n. Translating by points in $\Gamma(A_L)$, we may assume that $b \in \Gamma(A_R)^n$. Now $\delta(b/A_0) \geqslant 0$ since $A_0 \lhd A_R$, so $\operatorname{trd}(b/A_0) \geqslant d \operatorname{ldim}_{k_{\mathcal{O}}}(g/\Gamma(A_0) = dn$.

Since A_R is ACF-independent from A_L over A_0, we have $\operatorname{trd}(b/A_L) = \operatorname{trd}(b/A_0)$, and we also have $\operatorname{ldim}_{k_{\mathcal{O}}}(b/\Gamma(A_L)) = n$ by assumption, so

$$\delta(X/A_L) = \operatorname{trd}(b/A_L) - d \operatorname{ldim}_{k_{\mathcal{O}}}(b/\Gamma(A_L)) = \delta(b/A_0) \geqslant 0$$

as required. Thus $A_L \lhd A_1$. The same argument shows that if the embedding $A_0 \hookrightarrow A_L$ is strong, then so is the embedding $A_R \hookrightarrow A_1$.

Now take $A = A_1^{\text{full}}$, which exists and is a strong extension of A_1, by the existence part of Theorem 4.17. Note that if A_L and A_R are countable then so is A. □

Corollary 5.8. *The category $\mathcal{C}^{\text{fg}}$ has the amalgamation property. That is, given $A_0, A_L, A_R \in \mathcal{C}^{\text{fg}}$ and strong embeddings $A_0 \overset{\lhd}{\hookrightarrow} A_L$ and $A_0 \overset{\lhd}{\hookrightarrow} A_R$ as in the following diagram, there exist $A \in \mathcal{C}^{\text{fg}}$ and dashed arrows making the diagram commute.*

$$\begin{tikzcd}
& A_L \arrow[dr, hook, dashed, "\lhd"] & \\
A_0 \arrow[ur, hook, "\lhd"] \arrow[dr, hook, "\lhd"'] & & A \\
& A_R \arrow[ur, hook, dashed, "\lhd"'] &
\end{tikzcd}$$

Proof. Let A_0, A_L, A_R be as in the statement. By the existence part of Theorem 4.17, we can extend each of the three Γ-fields to its full closure.

$$\begin{tikzcd}
& A_L \arrow[r, hook, "\lhd"] & A_L^{\text{full}} \\
A_0 \arrow[ur, hook, "\lhd"] \arrow[r, hook, "\lhd"] \arrow[dr, hook, "\lhd"'] & A_0^{\text{full}} & \\
& A_R \arrow[r, hook, "\lhd"'] & A_R^{\text{full}}
\end{tikzcd}$$

Then, because we have $A_0 \lhd A_L^{\text{full}}$ and $A_0 \lhd A_R^{\text{full}}$, by the uniqueness part of the same theorem there are embeddings as in the following diagram, which are strong by Lemma 4.8 and finite character of δ.

$$\begin{tikzcd}
 & A_L \arrow[r, hook, "\lhd"] & A_L^{\text{full}} \\
A_0 \arrow[ur, hook, "\lhd"] \arrow[r, hook, "\lhd"] \arrow[dr, hook, "\lhd"'] & A_0^{\text{full}} \arrow[ur, hook, "\lhd"] \arrow[dr, hook, "\lhd"'] & \\
 & A_R \arrow[r, hook, "\lhd"'] & A_R^{\text{full}}
\end{tikzcd}$$

By Proposition 5.7, we can complete the diagram to

$$\begin{tikzcd}
 & A_L \arrow[r, hook, "\lhd"] & A_L^{\text{full}} \arrow[dr, hook, "\lhd"] & \\
A_0 \arrow[ur, hook, "\lhd"] \arrow[r, hook, "\lhd"] \arrow[dr, hook, "\lhd"'] & A_0^{\text{full}} \arrow[ur, hook, "\lhd"] \arrow[dr, hook, "\lhd"'] & & A' \\
 & A_R \arrow[r, hook, "\lhd"'] & A_R^{\text{full}} \arrow[ur, hook, "\lhd"'] &
\end{tikzcd}$$

and then we can take A to be the Γ-subfield of A' generated by $A_L \cup A_R$, which is in $\mathcal{C}^{\text{fg}}$. $\square$

Theorem 5.9. *The categories $\mathcal{C}$ and $\mathcal{C}^{\text{full}}$ are amalgamation categories, with the same Fraïssé limit.*

Proof. Strong embeddings are injective functions, so monomorphisms. Hence (AC1) holds. It is clear that the union of a chain of (full) Γ-fields is a (full) Γ-field, so (AC2) holds. We get (AC4) from Corollary 3.26 for $\mathcal{C}$ and Proposition 4.18 for $\mathcal{C}^{\text{fg-full}}$. The amalgamation property (AC5) is proved in Proposition 5.7 and Corollary 5.8. Since every Γ-field in $\mathcal{C}$ is an extension of F_{base}, and every full Γ-field in $\mathcal{C}^{\text{full}}$ is an extension of $(F_{\text{base}})^{\text{full}}$, properties (AC3) and (AC6) follow from (AC4) and (AC5), respectively.

Thus, $\mathcal{C}$ and $\mathcal{C}^{\text{full}}$ are both amalgamation categories. Let M be the Fraïssé limit of $\mathcal{C}^{\text{full}}$. If $A \in \mathcal{C}^{\leqslant\aleph_0}$ then $A^{\text{full}} \in \mathcal{C}^{\text{full},\leqslant\aleph_0}$, so as M is $\mathcal{C}^{\text{full},\leqslant\aleph_0}$-universal there is a strong embedding $A^{\text{full}} \lhd M$, which restricts to a strong embedding $A \lhd M$. Hence M is $\mathcal{C}^{\leqslant\aleph_0}$-universal. Similarly, using Theorem 4.17 and the $\mathcal{C}^{\text{fg-full}}$-saturation of M we can see that M is also $\mathcal{C}^{\text{fg}}$-saturated. Hence M is also the Fraïssé limit of $\mathcal{C}^{\text{fg}}$. $\square$

Notation 5.10. We write $M(F_{\text{base}})$ for the Fraïssé limit in $\mathcal{C}$.

5C. Γ-*algebraic extensions.*

Definition 5.11. Let $A \lhd B$ be a strong extension of Γ-fields. The extension is Γ-*algebraic* if for all finite tuples b from $\Gamma(B)$ there is a finite tuple $c \in \Gamma(B)$ containing b such that $\delta(c/A) = 0$.

Remark 5.12. From Lemma 4.15 we see that if F is a full Γ-field such that $B \lhd F$ then the extension $A \lhd B$ is Γ-algebraic if and only if $B \subseteq \Gamma\text{cl}^F(A)$.

Let $\mathcal{C}^{\text{alg}}$ be the subcategory of $\mathcal{C}$ consisting of the Γ-algebraic extensions of F_{base}.

Proposition 5.13. *$\mathcal{C}^{\text{alg}}$ is an amalgamation category.*

Proof. The proof of Theorem 5.9 goes through, except we also have to show that the amalgam of Γ-algebraic extensions is Γ-algebraic. So suppose we have the amalgamation square

$$\begin{tikzcd} & A_L \arrow[dr, hook, dashed, "\lhd"] & \\ A_0 \arrow[ur, hook, "\lhd"] \arrow[dr, hook, "\lhd"'] & & A \\ & A_R \arrow[ur, hook, dashed, "\lhd"'] & \end{tikzcd}$$

as in Corollary 5.8 with A_L and A_R both Γ-algebraic over A_0, A' a full Γ-field, and A the Γ-subfield of A' generated by $A_L \cup A_R$. Then by Remark 5.12, we have $A_L \cup A_R \subseteq \Gamma\mathrm{cl}^{A'}(A_0)$ and so $A \subseteq \Gamma\mathrm{cl}^{A'}(A_0)$, so $A_0 \lhd A$ is Γ-algebraic. $\square$

Write M_0 (or $M_0(F_{\mathrm{base}})$) for the Fraïssé limit of $\mathcal{C}^{\mathrm{alg}}$.

Definition 5.14. A Γ-field F strongly extending F_{base} is $\aleph_0$*-saturated for* Γ*-algebraic extensions over* F_{base} if whenever $F_{\mathrm{base}} \lhd A \lhd F$ with A finitely generated over F_{base} and $A \overset{\lhd}{\to} B$ is a finitely generated Γ-algebraic extension then B embeds (necessarily strongly) into F over A.

Proposition 5.15. *$M_0(F_{\mathrm{base}})$ is the unique countable full Γ-field strongly extending F_{base} which is Γ-algebraic over F_{base} and $\aleph_0$-saturated for Γ-algebraic extensions.*

Proof. Immediate from the uniqueness part of the amalgamation theorem and Proposition 5.13. $\square$

5D. *Purely* $\boldsymbol{\Gamma}$*-transcendental extensions.* In contrast with Γ-algebraic extensions are those we call purely Γ-transcendental extensions. We discuss amalgamation of these, which gives rise to some variant constructions.

Definition 5.16. Let $A \lhd B$ be a strong extension of Γ-fields. The extension is *purely* Γ*-transcendental* if for all tuples b from $\Gamma(B)$, either $\delta(b/A) > 0$ or $b \subseteq \Gamma(A)$.

Remark 5.17. If $A \lhd B$ is an extension of full Γ-fields then it is purely Γ-transcendental if and only if A is Γ-closed in B.

Definition 5.18. When F_{base} is a full countable Γ-field, we define $\mathcal{C}_{\Gamma\text{-tr}}(F_{\mathrm{base}})$ (usually abbreviated to $\mathcal{C}_{\Gamma\text{-tr}}$) to be the full subcategory of $\mathcal{C}$ consisting of the strong purely Γ-transcendental extensions of F_{base}.

Lemma 5.19. *If $A \in \mathcal{C}_{\Gamma\text{-tr}}$ then $A^{\mathrm{full}} \in \mathcal{C}_{\Gamma\text{-tr}}$.*

Proof. Consider the case when $(a_1, a_2) \in \Gamma(A^{\mathrm{full}}) \smallsetminus \Gamma(A)$ with $a_1 \in G_1(A^{\mathrm{full}})$ algebraic over A. Since $A \lhd A^{\mathrm{full}}$ we have $\mathrm{trd}(a_2/A) = d$. If $\delta((a_1, a_2)/F_{\mathrm{base}}) \leqslant 0$ then $\mathrm{trd}(a_1, a_2/F_{\mathrm{base}}) \leqslant d$, which implies that $\mathrm{trd}(a_1/F_{\mathrm{base}}) = 0$. Since F_{base} is full, that implies $(a_1, a_2) \in \Gamma(F_{\mathrm{base}})$, a contradiction.

Replacing A by the Γ-subfield of A^{full} generated by $A \cup \{(a_1, a_2)\}$ and iterating appropriately, we see that $A^{\mathrm{full}} \in \mathcal{C}_{\Gamma\text{-tr}}$. $\square$

We will show that $\mathcal{C}_{\Gamma\text{-tr}}$ is an amalgamation category by showing that the free amalgam of purely Γ-transcendental extensions is purely Γ-transcendental, using a lemma on stable groups.

Lemma 5.20. *Let H be a commutative algebraic group defined over an algebraically closed field C. Suppose $a_1, a_2, a_3 \in H$ are pairwise algebraically independent over C and $a_1 + a_2 + a_3 = 0$. Then there is a connected algebraic subgroup U of H and cosets $c_i + U$ defined over C such that a_i is a generic point of $c_i + U$ over C, for each $i = 1, 2, 3$. In particular,* $\operatorname{trd}(a_i/C) = \dim U$ *for each i.*

Proof. This is the special case for algebraic groups of a result about stable groups due to Ziegler [2006, Theorem 1]. □

Theorem 5.21. *If F_{base} is a full countable Γ-field then $\mathcal{C}_{\Gamma\text{-tr}}$ and $\mathcal{C}^{\mathrm{full}}_{\Gamma\text{-tr}}$ are amalgamation categories.*

Notation 5.22. We write $M_{\Gamma\text{-tr}}(F_{\mathrm{base}})$ for the Fraïssé limit in $\mathcal{C}_{\Gamma\text{-tr}}(F_{\mathrm{base}})$.

Proof of Theorem 5.21. Axioms (AC1), (AC3), and (AC4) follow immediately from the fact that $\mathcal{C}_{\Gamma\text{-tr}}$ and $\mathcal{C}^{\mathrm{full}}_{\Gamma\text{-tr}}$ are full subcategories of $\mathcal{C}$. Axioms (AC2) and (AC6) are also immediate. It remains to prove (AC5), the amalgamation property.

Using Lemma 5.19, the same argument as for Corollary 5.8 allows us to reduce the amalgamation property for $\mathcal{C}_{\Gamma\text{-tr}}$ to the amalgamation property for $\mathcal{C}^{\mathrm{full}}_{\Gamma\text{-tr}}$. So suppose we have full Γ-fields

$$\begin{tikzcd}
A_L & & A_R \\
& A_0 \arrow[ul, hook, "\lhd"'] \arrow[ur, hook, "\lhd"] & \\
& F_{\mathrm{base}} \arrow[u, hook, "\lhd"] &
\end{tikzcd}$$

with A_0, A_L, and A_R all purely Γ-transcendental extensions of F_{base}. Let A_1 be the free amalgam of A_L and A_R over A_0 as in the proof of Proposition 5.7. We must show that A_1 is a purely Γ-transcendental extension of F_{base}.

So let B be a Γ-subfield of A_1 properly containing F_{base} and finitely generated over it. It remains to show that $\delta(B/F_{\mathrm{base}}) \geqslant 1$. If $B \wedge A_R \neq F_{\mathrm{base}}$ then we have

$$\begin{aligned}
\delta(B/F_{\mathrm{base}}) &= \delta(B/B \wedge A_R) + \delta(B \wedge A_R/F_{\mathrm{base}}) \\
&\geqslant \delta(BA_R/A_R) + \delta(B \wedge A_R/F_{\mathrm{base}}) \qquad \text{by submodularity} \\
&\geqslant 0 + 1 = 1,
\end{aligned}$$

the last line because $A_R \lhd A_1$ and A_R is purely Γ-transcendental over F_{base}. So in this case we are done, and similarly if $B \wedge A_L \neq F_{\mathrm{base}}$.

So we may assume that $B \wedge A_R = B \wedge A_L = F_{\mathrm{base}}$. Choose a basis $b = (b^1, \ldots, b^n)$ of B over F_{base}. For each i, choose $b^i_L \in \Gamma(A_L)$ and $b^i_R \in \Gamma(A_R)$ such that $b^i = b^i_L + b^i_R$. Let B_L and B_R be the Γ-field extensions of F_{base} generated by $b_L := (b^1_L, \ldots, b^n_L)$ and $b_R := (b^1_R, \ldots, b^n_R)$, respectively.

We claim that $B_L \wedge A_R = F_{\mathrm{base}}$ and that $A_L \wedge B_R = F_{\mathrm{base}}$. To see this, suppose that $v \in \Gamma(B_L \wedge A_R) = \Gamma(B_L) \cap \Gamma(A_R)$. Since $v \in \Gamma(B_L)$ there are $s_i \in k_{\mathcal{O}}$ and some $a \in \Gamma(F_{\mathrm{base}})$ such that $v = \sum_{i=1}^n s_i b_L^i + a$. Let $u_L = v - a = \sum_{i=1}^n s_i b_L^i$, let $u_R = \sum_{i=1}^n s_i b_R^i$, and let $u = u_L + u_R = \sum_{i=1}^n s_i b^i$. Then $v, a \in \Gamma(A_R)$, so $u_L \in \Gamma(A_R)$, and also $u_R \in \Gamma(A_R)$; hence $u \in \Gamma(A_R)$. But $u \in \Gamma(B)$ and we have $B \wedge A_R = F_{\mathrm{base}}$. So $u \in \Gamma(F_{\mathrm{base}})$ and so each $s_i = 0$, and thus $v \in \Gamma(F_{\mathrm{base}})$. So $B_L \wedge A_R = F_{\mathrm{base}}$. The same argument shows that $A_L \wedge B_R = F_{\mathrm{base}}$, and in particular $B_L \wedge B_R = F_{\mathrm{base}}$.

Let C be the Γ-subfield of A_1 generated by $B \cup B_L$, and note that it is also generated by $B_L \cup B_R$. We have $B \wedge B_L = F_{\mathrm{base}} = B_L \wedge B_R$, so applying modularity of linear dimension to the squares

$$\begin{tikzcd} & C & \\ B \arrow[ur, hook, "\triangleleft"] & & B_L \arrow[ul, hook', "\triangleleft"'] \\ & F_{\mathrm{base}} \arrow[ul, hook', "\triangleleft"] \arrow[ur, hook, "\triangleleft"'] & \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} & C & \\ B_L \arrow[ur, hook, "\triangleleft"] & & B_R \arrow[ul, hook', "\triangleleft"'] \\ & F_{\mathrm{base}} \arrow[ul, hook', "\triangleleft"] \arrow[ur, hook, "\triangleleft"'] & \end{tikzcd}$$

we get

$$\mathrm{ldim}_{k_{\mathcal{O}}}(\Gamma(B_R)/\Gamma(F_{\mathrm{base}})) = \mathrm{ldim}_{k_{\mathcal{O}}}(\Gamma(C)/\Gamma(B_L)) = \mathrm{ldim}_{k_{\mathcal{O}}}(\Gamma(B)/\Gamma(F_{\mathrm{base}})) = n,$$

and so b_R is $k_{\mathcal{O}}$-linearly independent over $\Gamma(F_{\mathrm{base}})$, and hence over $\Gamma(A_0)$, since $B_R \wedge A_0 = F_{\mathrm{base}}$. We have

$$\mathrm{trd}(b/F_{\mathrm{base}}) \geqslant \mathrm{trd}(b/A_0) \geqslant \mathrm{trd}(b/A_0 b_L) = \mathrm{trd}(b_R/A_0 b_L) = \mathrm{trd}(b_R/A_0) \geqslant dn$$

with the last three (in)equalities holding because $b = b_L + b_R$, because b_R is algebraically independent from b_L over A_0, and because $A_0 \triangleleft A_1$ and b_R is $k_{\mathcal{O}}$-linearly independent over $\Gamma(A_0)$. Similarly, $\mathrm{trd}(b/A_0) \geqslant \mathrm{trd}(b_L/A_0) \geqslant dn$.

Suppose for a contradiction that $\delta(B/F_{\mathrm{base}}) \leqslant 0$. Then we must have

$$\mathrm{trd}(b/F_{\mathrm{base}}) = \mathrm{trd}(b/A_0) = \mathrm{trd}(b_L/A_0) = \mathrm{trd}(b_R/A_0) = dn,$$

and then we also have

$$\mathrm{trd}(b, b_L/A_0) = \mathrm{trd}(b, b_R/A_0) = \mathrm{trd}(b_L, b_R/A_0) = \mathrm{trd}(b, b_L, b_R/A_0) = 2dn,$$

so b, b_L, b_R are pairwise algebraically independent over A_0.

We apply Lemma 5.20 with $H = G^n = G_1^n \times G_2^n$, $a_1 = -b$, $a_2 = b_L$, and $a_3 = b_R$ to get a connected algebraic subgroup U of G^n of dimension dn such that b is in an A_0-coset of U. Since $\mathrm{trd}(b/F_{\mathrm{base}}) = \mathrm{trd}(b/A_0) = \dim U$, and F_{base} is an algebraically closed field, the coset is actually defined over F_{base}.

G_1 and G_2 are nonisogenous and so U is of the form $U_1 \times U_2$ where each U_i is a connected subgroup of G_i^n. Since $\dim U = dn$, if $U_2 = G_2^n$ then U_1 is the trivial subgroup of G_1^n, so $\pi_1(b) \in G_1^n(A_0)$. But A_0 is a full Γ-field and so $b \in \Gamma(A_0)^n$, which contradicts $\mathrm{trd}(b/A_0) = dn$ (and $n > 0$). So U_2 must be a proper subgroup of G_2^n. Since G_2 is simple, it follows that $\pi_2(b)$ satisfies an $\mathcal{O}$-linear equation $\sum_{i=1}^n s_i \pi_2(b^i) = c$

with $c \in G_2(F_{\text{base}})$. Then, since $b \in \Gamma(B)^n$ and F_{base} is a full Γ-field, we have $\sum_{i=1}^n s_i b^i \in \Gamma(F_{\text{base}})$, which contradicts b being a basis for B over F_{base}.

So we have $\delta(B/F_{\text{base}}) \geqslant 1$, and thus A_1 is a purely Γ-transcendental extension of F_{base}, as required. □

6. Categoricity

In this section we first introduce the abstract notion of a quasiminimal pregeometry structure, which is a form of the technique of *quasiminimal excellence*, used to prove uncountable categoricity of a list of axioms. We then prove that our Γ-closed fields do satisfy the conditions to be quasiminimal pregeometry structures. Later in Section 8 we will give axioms for the Γ-closed fields, and then we can deduce categoricity of this axiomatization.

6A. ***Quasiminimal pregeometry structures.*** This definition of quasiminimal pregeometry structures comes from [Bays et al. 2014b].

Definition 6.1. Let M be an L-structure for a countable language L, equipped with a pregeometry cl (or cl_M if it is necessary to specify M). Write qftp for the quantifier-free L-type. We say that M is a *quasiminimal pregeometry structure* if the following hold:

(QM1) The pregeometry is determined by the language. That is, if a, a' are singletons, b, b' are tuples, $\mathrm{qftp}(a, b) = \mathrm{qftp}(a', b')$, and $a \in \mathrm{cl}(b)$, then $a' \in \mathrm{cl}(b')$.

(QM2) M is infinite-dimensional with respect to cl.

(QM3) *Countable closure property*:
If $A \subseteq M$ is finite then $\mathrm{cl}(A)$ is countable.

(QM4) *Uniqueness of the generic type*:
Suppose that $C, C' \subseteq M$ are countable closed subsets, enumerated such that $\mathrm{qftp}(C) = \mathrm{qftp}(C')$. If $a \in M \smallsetminus C$ and $a' \in M \smallsetminus C'$ then $\mathrm{qftp}(C, a) = \mathrm{qftp}(C', a')$ (with respect to the same enumerations for C and C').

(QM5) $\aleph_0$-*homogeneity over closed sets and the empty set*:
Let $C, C' \subseteq M$ be countable closed subsets or empty, enumerated such that $\mathrm{qftp}(C) = \mathrm{qftp}(C')$, let b, b' be finite tuples from M such that $\mathrm{qftp}(C, b) = \mathrm{qftp}(C', b')$, and let $a \in \mathrm{cl}(C, b)$. Then there is $a' \in M$ such that $\mathrm{qftp}(C, b, a) = \mathrm{qftp}(C', b', a')$.

We say M is a *weakly quasiminimal pregeometry structure* if it satisfies all the axioms except possibly (QM2).

Definition 6.2. Given M_1 and M_2 both weakly quasiminimal pregeometry L-structures, we say that an L-embedding $\theta : M_1 \hookrightarrow M_2$ is a *closed embedding* if for each $A \subseteq M_1$ we have $\theta(\mathrm{cl}_{M_1}(A)) = \mathrm{cl}_{M_2}(\theta(A))$. In particular, $\theta(M_1)$ is closed in M_2 with respect to cl_{M_2}. When θ is an inclusion map we say that M_1 is a *closed substructure* of M_2. By axiom (QM1), closed substructures are the same as closed subsets.

Definition 6.3. Given a quasiminimal pregeometry structure M, let $\mathcal{K}(M)$ be the smallest class of L-structures which contains M and all its closed substructures and is closed under isomorphism and under taking unions of directed systems of closed embeddings. We call any class of the form $\mathcal{K}(M)$ a *quasiminimal class*.

The purpose of these definitions is the categoricity theorem, Theorem 2.3 in [Bays et al. 2014b].

Fact 6.4. *If $\mathcal{K}$ is a quasiminimal class then every structure $A \in \mathcal{K}$ is a weakly quasiminimal pregeometry structure, and up to isomorphism there is exactly one structure in $\mathcal{K}$ of each cardinal dimension. In particular, $\mathcal{K}$ is uncountably categorical. Furthermore, $\mathcal{K}$ is the class of models of an $L_{\omega_1,\omega}(Q)$ sentence.*

We will verify axioms (QM1)–(QM5) for the Fraïssé limits we constructed. We first make some general observations which simplify what we have to verify.

Proposition 6.5. *Suppose that M is a countable L-structure. Then it satisfies* (QM1)–(QM5) *if and only if it satisfies the following axioms*:

(QM1′) *If a and b are finite tuples and* $\mathrm{qftp}(a) = \mathrm{qftp}(b)$ *then* $\dim(a) = \dim(b)$.

(QM2) *M is infinite-dimensional with respect to* cl.

(QM4) *Uniqueness of the generic type.*

(QM5a) $\aleph_0$-homogeneity over the empty set:
If a and b are finite tuples from M and $\mathrm{qftp}(a) = \mathrm{qftp}(b)$ *then there is $\theta \in \mathrm{Aut}(M)$ such that $\theta(a) = b$.*

(QM5b) Nonsplitting over a finite set:
If C is a closed subset of M and $b \in M$ is a finite tuple then there is a finite tuple $c \in C$ such that $\mathrm{qftp}(b/C)$ *does not split over c. That is, for all finite tuples $a, a' \in C$, if* $\mathrm{qftp}(a/c) = \mathrm{qftp}(a'/c)$ *then* $\mathrm{qftp}(a/cb) = \mathrm{qftp}(a'/cb)$.

Proof. The first axiom (QM1′) is equivalent to (QM1), because $a \in \mathrm{cl}(b)$ if and only if $\dim(a, b) = \dim(b)$, so if quantifier-free types characterize the dimension they also characterize the closure operation, and vice versa.

The countable closure property (QM3) is immediate for a countable M.

Axiom (QM5) with $C = \varnothing$ gives a back-and-forth condition which is equivalent to $\aleph_0$-homogeneity using the standard back-and-forth argument together with (QM1) and (QM4). Since M is countable, the back-and-forth construction gives (QM5a). The converse is immediate.

Finally, [Bays et al. 2014b, Corollary 5.3] shows that the case of (QM5) with C closed is equivalent to (QM5b). □

Remark 6.6. All the axioms refer to quantifier-free types with respect to a particular language, and from (QM5a) we get the conclusion that if two finite tuples from M have the same quantifier-free type then they actually have the same complete type (even the same $L_{\infty,\omega}$-type, and furthermore they lie in the same automorphism orbit, that is, they have the same Galois-type). Since M is not necessarily a saturated model of its first-order theory, it does not follow that every definable set is quantifier-free definable. Nonetheless, identifying the language which works allows us to understand the types which are realized in M.

6B. ***Verification of the quasiminimal pregeometry axioms.*** Recall that our language of Γ-fields is $L_\Gamma = \langle +, \cdot, -, \Gamma, (c_a)_{a\in K_0}\rangle$, where Γ is a relation symbol of suitable arity to denote a subset of G. We start by defining the expansion L^{QE} of L_Γ in which we have the form of quantifier-elimination described in the previous remark.

Let W be any subvariety of $G^n \times \mathbb{A}^r$ defined over F_{base}, for some $n, r \in \mathbb{N}$. (It suffices to consider those W which are the graphs of rational maps $f : W' \to \mathbb{A}^r$, with $W' \subseteq G^n$.) Let $\varphi_W(x, y)$ name the subset of $G(M)^n \times M^r$ given by

$$(x, y) \in W \quad \text{and} \quad x \in \Gamma^n \quad \text{and} \quad x \text{ is } \mathcal{O}\text{-linearly independent over } \Gamma(F_{\mathrm{base}}),$$

and let $\psi_W(y)$ be the formula $\exists x\, \varphi_W(x, y)$.

Definition 6.7. We define L^{QE} to be the expansion of L_Γ by parameters for F_{base} and relation symbols for all the formulas $\varphi_W(x, y)$ and $\psi_W(y)$.

Remark 6.8. Note that the formulas $\varphi_W(x, y)$ are always expressible in $L_{\omega_1,\omega}(L_\Gamma)$ (with parameters in F_{base}), so a priori this is an expansion of $L_\Gamma(F_{\mathrm{base}})$ by $L_{\omega_1,\omega}$-definitions. However, if the ring $\mathcal{O}$ and its action on G are definable and $\Gamma(F_{\mathrm{base}})$ is either of finite rank (which is true for example in pseudo-exponentiation) or is otherwise an L_Γ-definable set, then L^{QE} is just an expansion of $L_\Gamma(F_{\mathrm{base}})$ by first-order definitions.

For the rest of this section we use tuples both from the field sort of a model M and from $\Gamma(M)$, so to distinguish them we will use Latin letters for tuples from M and Greek letters for tuples from $\Gamma(M)$.

Theorem 6.9. *Take M to be either $M(F_{\mathrm{base}})$ or $M_{\Gamma\text{-tr}}(F_{\mathrm{base}})$, the latter only if F_{base} is a full Γ-field. Then, considered in the language L^{QE} and equipped with $\Gamma\mathrm{cl}$, M is a quasiminimal pregeometry structure.*

Proof. We verify the axioms from Proposition 6.5. The main difficulty is that the axioms refer to the field sort whereas the construction of M was done in the sort Γ, and there is no canonical way to go from one sort to the other in either direction. However, the sort Γ has rank d with respect to the pregeometry $\Gamma\mathrm{cl}$, so as we want to include the case $d > 1$ we have to verify the axioms with respect to the field sort.

First we prove (QM2). For any $n \in \mathbb{N}$, there is a strong Γ-field extension A_n of F_{base} generated by a tuple $\alpha \in \Gamma(A_n)^n$ such that α is generic in G^n over F_{base}. Then $\delta(\alpha/F_{\mathrm{base}}) = dn$. This A_n embeds strongly in M by the universality property of the Fraïssé limit, so $\Gamma\dim^M(\alpha) = dn$ by Lemma 4.15. Hence M is infinite-dimensional.

Now we prove (QM1′) and (QM5a) together. Suppose $a, b \in M^r$ with $\mathrm{qftp}_{L^{\mathrm{QE}}}(a) = \mathrm{qftp}_{L^{\mathrm{QE}}}(b)$. Choose a strong Γ-subfield $A \lhd M$ which is a finitely generated extension of F_{base} such that $a \in A^r$ and $\delta(A/F_{\mathrm{base}})$ is minimal such. Let $\alpha \in \Gamma(A)^n$ be a good basis for A over F_{base}, such that a is in the field $F_{\mathrm{base}}(\alpha)$, let $W = \mathrm{Loc}(\alpha, a/F_{\mathrm{base}})$, and let $V = \mathrm{Loc}(\alpha/F_{\mathrm{base}})$. Then $M \models \psi_W(a)$, so also $M \models \psi_W(b)$.

So there is $\beta \in \Gamma(M)^n$ such that $M \models \varphi_W(\beta, b)$. In particular, β is in $V \cap \Gamma(M)^n$ and is $k_{\mathcal{O}}$-linearly independent over $\Gamma(F_{\mathrm{base}})$. We claim that $W = \mathrm{Loc}(\beta, b/F_{\mathrm{base}})$. Suppose not, so $W' := \mathrm{Loc}(\beta, b/F_{\mathrm{base}})$ is a proper subvariety of W. We have $M \models \psi_{W'}(b)$, so since a and b have the same quantifier-free L^{QE}-type, $M \models \psi_{W'}(a)$. So there is some $\alpha' \in \Gamma(M)^n$ such that $M \models \varphi_{W'}(\alpha', a)$. W is irreducible over F_{base}, so $\dim W' < \dim W$. Since a is rational over $F_{\mathrm{base}}(\alpha)$, $\dim W = \dim V$, and so α' lies in a subvariety V' of V with $\dim V' < \dim V$. But then setting $A' = \langle F_{\mathrm{base}}, \alpha' \rangle$ we have $a \in A'^r$ and $\delta(A'/F_{\mathrm{base}}) = \delta(\alpha'/F_{\mathrm{base}}) = \dim V' - \mathrm{ldim}_{k_{\mathcal{O}}}(\alpha'/\Gamma(F_{\mathrm{base}})) < \delta(A/F_{\mathrm{base}})$, which contradicts the choice of A. Thus $\mathrm{Loc}(\beta, b/F_{\mathrm{base}}) = W$, and in particular $\mathrm{Loc}(\beta/F_{\mathrm{base}}) = V$. Let $B = \langle F_{\mathrm{base}}, \beta \rangle$. We further deduce that $B \lhd M$, since if not, the same proof would show that $\delta(A/F_{\mathrm{base}})$ would not be minimal.

By Lemma 4.15, we have $\Gamma\dim(A) = \delta(A/F_{\mathrm{base}}) = \delta(B/F_{\mathrm{base}}) = \Gamma\dim(B)$. Now $\Gamma\dim(a) = \Gamma\dim(A)$ by the minimality of $\delta(A/F_{\mathrm{base}})$, since A could be taken within $\Gamma\mathrm{cl}(a)$, and $\Gamma\dim(b) \leq \Gamma\dim(B)$, so $\Gamma\dim(b) \leq \Gamma\dim(a)$. By symmetry, $\Gamma\dim(a) = \Gamma\dim(b)$, so (QM1$'$) is proved.

Since α is a good basis, there is an isomorphism of Γ-fields $\theta_0 : A \to B$ over F_{base}, with $\theta_0(\alpha) = \beta$. Then also $\theta_0(a) = b$. Since M is $\mathcal{C}^{\mathrm{fg}}$-homogeneous (or $\mathcal{C}^{\mathrm{fg}}_{\Gamma\text{-tr}}$-homogeneous), θ_0 extends to an automorphism θ of M. That proves (QM5a).

For (QM4), suppose that $C_1, C_2 \lhd_{\mathrm{cl}} M$ with the same quantifier-free L^{QE}-type according to some enumeration, and let $\theta : C_1 \cong C_2$ be the isomorphism given by the enumeration. Suppose also that $b_1 \in M \smallsetminus C_1$ and $b_2 \in M \smallsetminus C_2$.

Using Lemma 4.13 we get $\beta_1, \beta_2 \in \Gamma(M)$ such that $b_i \in C_i(\beta_i)$ and $\mathrm{Loc}(\beta_1, b_1/C_1)$ is defined over K_0 and is equal to $\mathrm{Loc}(\beta_2, b_2/C_2)$. Also, setting $B_i := \langle C_i, \beta_i \rangle$ we have $B_i \lhd M$ and β_i is a good basis for B_i over C_i. By the definition of a good basis, the isomorphism θ extends to $\theta_1 : B_1 \cong B_2$ with $\theta_1(\beta_1) = \beta_2$ and hence $\theta_1(b_1) = b_2$.

Let $F_{\mathrm{base}} \lhd A_1 \lhd C_1$ with A_1 finitely generated over F_{base}, and let $A_2 = \theta(A_1)$. Then θ_1 restricts to an isomorphism $\theta_0 : \langle A_1 \beta_1 \rangle \cong \langle A_2 \beta_2 \rangle$. Also $\langle A_i \beta_i \rangle \lhd M$ since $\delta(\langle A_i \beta_i \rangle / A_i) = 1$ and $\beta_i \notin \Gamma\mathrm{cl}(A_i)$. Since M is $\mathcal{C}^{\mathrm{fg}}$-homogeneous (or $\mathcal{C}^{\mathrm{fg}}_{\Gamma\text{-tr}}$-homogeneous), θ_0 extends to an automorphism of M. So $\mathrm{qftp}_{L^{\mathrm{QE}}}(A_1 b_1) = \mathrm{qftp}_{L^{\mathrm{QE}}}(A_2 b_2)$ and thus, as A_1 ranges over strong Γ-subfields of C_1 finitely generated over F_{base}, we deduce that $\mathrm{qftp}_{L^{\mathrm{QE}}}(C_1 b_1) = \mathrm{qftp}_{L^{\mathrm{QE}}}(C_2 b_2)$ as required.

Finally, to prove (QM5b), let $C \lhd_{\mathrm{cl}} M$ and let $b \in M$ be a finite tuple. Let B be a finitely generated Γ-field extension of C such that $B \lhd M$ and $b \in B$, and let $\beta \in \Gamma(B)^n$ be a good basis for B over C with $b \in C(\beta)$. Now choose a finitely generated Γ-field extension C_0 of F_{base} in C with $C_0 \lhd C$, and a good basis γ for C_0, such that $\mathrm{Loc}(\beta, b/C)$ is defined over $F_{\mathrm{base}}(\gamma)$.

Suppose that finite tuples $a, a' \in C$ have $\mathrm{qftp}_{L^{\mathrm{QE}}}(a/\gamma) = \mathrm{qftp}_{L^{\mathrm{QE}}}(a'/\gamma)$. By (QM5a), there is a Γ-field automorphism $\theta \in \mathrm{Aut}(M/F_{\mathrm{base}}(\gamma))$ such that $\theta(a) = a'$. Let A be a strong Γ-subfield of $\Gamma\mathrm{cl}(C_0, a)$ which is finitely generated over C_0 and contains a, and let $A' = \theta(A)$. Then $A' \lhd C$.

Let $V = \mathrm{Loc}(\beta/C)$. Then $\mathrm{Loc}(\beta/A) = \mathrm{Loc}(\beta/A') = V$ because V is defined over $F_{\mathrm{base}}(\gamma)$. So, since β is a good basis, the isomorphism $\theta_0 : A \cong A'$ extends to $\theta_1 : \langle A\beta \rangle \cong \langle A'\beta \rangle$.

We claim that $\langle A\beta \rangle \lhd B$. To see this, suppose that $X \subseteq B$ is a finitely generated extension of $\langle A\beta \rangle$, and let $X_0 = X \wedge C$. Let ξ be a basis of X_0 over A. Then $\xi \cup \beta$ is a basis for X over A, since β is a basis for B over C and hence for X over $X \wedge C = X_0$. So

$$\begin{aligned}\delta(X/A\beta) &= \operatorname{trd}(\xi/A\beta) - d\,\operatorname{ldim}_{k_{\mathcal{O}}}(\xi/\Gamma(A), \beta)\\ &= \operatorname{trd}(\xi/A) - d\,\operatorname{ldim}_{k_{\mathcal{O}}}(\xi/\Gamma(A))\\ &= \delta(\xi/A) \geqslant 0\end{aligned}$$

because β is algebraically and linearly independent from C over A and $\xi \in C$. Since $B \lhd M$ we have $\langle A\beta\rangle \lhd M$. The same argument shows that $\langle A'\beta\rangle \lhd M$.

Thus, since M is $\mathcal{C}^{\mathrm{fg}}$-homogeneous (or $\mathcal{C}^{\mathrm{fg}}_{\Gamma\text{-tr}}$-homogeneous), θ_1 extends to an automorphism θ_2 of M. Now θ_2 fixes b and γ and $\theta_2(a) = a'$, so $\operatorname{qftp}_{L^{\mathrm{QE}}}(a/b\gamma) = \operatorname{qftp}_{L^{\mathrm{QE}}}(a'/b\gamma)$. Taking $c = \gamma$, considered as a tuple from the field sort of C, we see that $\operatorname{tp}(b/C)$ does not split over c, as required. □

Remark 6.10. A more complete analysis of splitting for pseudo-exponentiation was carried out in the Ph.D. thesis of Robert Henderson [2014].

We conclude this section by showing that the Γ-algebraic types over finite tuples are isolated.

Proposition 6.11. *Suppose that a, b are finite tuples in M and that $b \in \Gamma\mathrm{cl}^M(a)$. Then* $\operatorname{tp}(b/a)$ *is isolated by an L^{QE}-formula.*

Proof. Choose a finitely generated Γ-field $B \lhd M$ with $B \subseteq \Gamma\mathrm{cl}(a)$, and a good basis β for B such that $a, b \in F_{\mathrm{base}}(\beta)$. Let $W = \operatorname{Loc}(\beta, a^\frown b/F_{\mathrm{base}})$.

Then $M \models \varphi_W(\beta, a^\frown b)$ and $M \models \psi_W(a^\frown b)$. Suppose $M \models \psi_W(a^\frown c)$. Then there is a tuple γ from $\Gamma(M)$ such that $M \models \varphi_W(\gamma, a^\frown c)$. So $\operatorname{Loc}(\gamma/F_{\mathrm{base}}) \subseteq V$ but $F_{\mathrm{base}} \lhd M$ and $\operatorname{ldim}_{k_{\mathcal{O}}}(\gamma/\Gamma(F_{\mathrm{base}})) = \dim V$ by the definition of φ_W, so γ is generic in V over F_{base}. Thus $\operatorname{Loc}(\gamma/F_{\mathrm{base}}) = \operatorname{Loc}(\beta/F_{\mathrm{base}})$ so, since β is a good basis, the Γ-field C generated by γ is isomorphic to B via an isomorphism $\theta : B \to C$ such that $\theta(\beta) = \gamma$, and then necessarily $\theta(a) = a$ and $\theta(b) = c$.

Using Lemma 4.15 repeatedly,

$$\delta(C/F_{\mathrm{base}}) = \delta(B/F_{\mathrm{base}}) = \Gamma\dim(B) = \Gamma\dim(a) \leqslant \Gamma\dim(C)$$

and so $\delta(C/F_{\mathrm{base}}) = \Gamma\dim(C)$, and $C \lhd M$. Thus θ extends to an automorphism of M, so $\operatorname{tp}(c/a) = \operatorname{tp}(b/a)$, so the formula $\psi_W(a^\frown x)$ isolates $\operatorname{tp}(b/a)$. □

7. Classification of strong extensions

We next give a classification of the finitely generated strong extensions. In the next section we will use it to give axiomatizations of the classes of Γ-closed fields we have constructed, generalizing Zilber's axioms for pseudo-exponentiation.

Since G is an $\mathcal{O}$-module, each matrix $M \in \operatorname{Mat}_n(\mathcal{O})$ defines an $\mathcal{O}$-module homomorphism $G^n \xrightarrow{M} G^n$ in the usual way. If $V \subseteq G^n$, we write $M \cdot V$ for its image. Note that if V is a subvariety of G^n then $M \cdot V$ is a constructible set, and since the $\mathcal{O}$-module structure is defined over K_0, if V is defined over A then $M \cdot V$ is defined over $K_0 \cup A$. If V is irreducible then $M \cdot V$ is also irreducible.

We have $G^n = (G_1 \times G_2)^n$ and we write $x_1, \dots, x_n$ for the coordinates in G_1 and $y_1, \dots, y_n$ for the coordinates in G_2.

Definition 7.1. Let V be an irreducible subvariety of G^n. Then V is *G_1-free* if V does not lie inside any subvariety defined by an equation $\sum_{j=1}^n r_j x_j = c$ for any $r_j \in \mathcal{O}$, not all zero, and any $c \in G_1$. We define *G_2-free* the same way. We say V is *free* if it is both G_1-free and G_2-free.

V is *rotund* (for G as an $\mathcal{O}$-module) if for every matrix $M \in \mathrm{Mat}_n(\mathcal{O})$ we have

$$\dim(M \cdot V) \geqslant d \operatorname{rk} M,$$

where dim means dimension as an algebraic variety or constructible set, $\operatorname{rk} M$ is the rank of the matrix M, and $d = \dim G_1$.

V is *strongly rotund* if for every *nonzero* matrix $M \in \mathrm{Mat}_n(\mathcal{O})$ we have

$$\dim(M \cdot V) > d \operatorname{rk} M.$$

A reducible subvariety V of G^n is defined to be free, rotund, or strongly rotund if at least one of its (absolutely) irreducible components is free, rotund, or strongly rotund, respectively. If we say that such a V is *free and (strongly) rotund* then we mean that the same irreducible component is free and (strongly) rotund.

So V is free if it is "free from $\mathcal{O}$-linear dependencies", and it is rotund if all its images under suitable homomorphisms are of large dimension.

Lemma 7.2. *An irreducible subvariety $V \subseteq G^n$ is G_2-free if and only if $\pi_2(V)$ does not lie in a coset of a proper algebraic subgroup of G_2^n. If $\mathcal{O} = \mathrm{End}(G_1)$ then V is G_1-free if and only if $\pi_1(V)$ does not lie in a coset of a proper algebraic subgroup of G_1^n.*

Proof. This is an immediate consequence of Lemma 2.1(i). □

Proposition 7.3. *Suppose that A is a full Γ-field, $A \subseteq B$ is a finitely generated extension of Γ-fields, and $b \in \Gamma(B)^n$ is a basis for the extension. Let $V = \mathrm{Loc}(b/A)$.*

Then V is free. Furthermore, the extension is strong if and only if V is rotund, and it is purely Γ-transcendental if and only if V is strongly rotund.

Proof. If V is not G_1-free then, writing $b = (b_1^1, \dots, b_1^n, b_2^1, \dots, b_2^n) \in G_1^n \times G_2^n$ we have $\sum_{j=1}^n r_j b_1^j = c_1 \in G_1(A)$. Let $c_2 = \sum_{j=1}^n r_j b_2^j$. Then $(c_1, c_2) \in \Gamma(B)$ and since A is full and the extension preserves the kernels we have $(c_1, c_2) \in \Gamma(A)$. That contradicts b being a basis for the extension. So V is G_1-free and, symmetrically, G_2-free.

For $M \in \mathrm{Mat}_n(\mathcal{O})$ we have $M \cdot b \in \Gamma(B)^n$ with $\mathrm{ldim}_{k_\mathcal{O}}(M \cdot b / \Gamma(A)) = \operatorname{rk} M$. Furthermore, every finite tuple from $\Gamma(B)$ generates the same Γ-field extension of A as some tuple $M \cdot b$, because b is a basis. Thus the extension is strong if and only if for all M we have $\mathrm{trd}(M \cdot b/A) \geqslant d \operatorname{rk} M$ if and only if for all M we have $\dim(M \cdot V)^{\mathrm{Zar}} \geqslant d \operatorname{rk} M$ if and only if V is rotund.

Similarly any finite tuple from $\Gamma(B)$ which is not in $\Gamma(A)$ generates the same extension of A as a tuple $M \cdot b$ for some nonzero matrix M, and V is strongly rotund if and only if all such tuples have $\delta(M \cdot b/A) > 0$. □

Corollary 7.4. *Suppose that $A \lhd B$ is a finitely generated strong extension of essentially finitary Γ-fields, that $A^{\text{full}} \wedge B = A$, that $b \in \Gamma(B)^n$ is a basis for the extension, and that $V = \operatorname{Loc}(b/A)$. Then V is free and rotund, and it is strongly rotund if and only if B is a purely Γ-transcendental extension of A.*

Proof. First note that, since A and B are essentially finitary, by Theorem 4.17, B^{full} is uniquely determined up to isomorphism and A^{full} is uniquely determined as a Γ-subfield of B^{full}, so the condition that $A^{\text{full}} \wedge B = A$ makes unambiguous sense. Now the proof of Proposition 7.3 goes through with this weaker condition in place of $A = A^{\text{full}}$. □

8. Axiomatization of Γ-closed fields

Recall that we have fixed a ring $\mathcal{O}$ and algebraic $\mathcal{O}$-modules G_1 and G_2, both of dimension d, defined over a countable field K_0, we have the product $G = G_1 \times G_2$, and we consider structures in the language $L_\Gamma = \langle +, \cdot, -, \Gamma, (c_a)_{a \in K_0} \rangle$, where Γ is a relation symbol of appropriate arity to denote a subset of G. We are given an essentially finitary Γ-field F_{base} containing K_0, of type (EXP), (COR), or (DEQ). We add parameters for F_{base} to the language to get a language $L_{F_{\text{base}}}$. We also have an expanded language L^{QE}.

Definition 8.1. A model in the quasiminimal class $\mathcal{K}(M(F_{\text{base}}))$ is called a Γ-*closed field* (with the countable closure property, on the base F_{base}).

Theorem 8.2. *An $L_{F_{\text{base}}}$ structure F is a Γ-closed field if and only if it satisfies the following list of axioms, which we denote by $\Gamma\mathrm{CF}_{\mathrm{CCP}}(F_{\text{base}})$.*

(1) Full Γ-field: *F is an algebraically closed field containing K_0 and $\Gamma(F)$ is an $\mathcal{O}$-submodule of $G(F)$ such that the projections of $\Gamma(F)$ to $G_1(F)$ and $G_2(F)$ are surjective.*

(2) Base and kernels: *F satisfies the full atomic diagram of F_{base}.* (*In some examples we will discuss how this can be weakened.*) *Also* $\ker_i(F) = \ker_i(F_{\text{base}})$ *for* $i = 1, 2$.

(3) Predimension inequality (generalized Schanuel property): *The* predimension function

$$\delta(x/F_{\text{base}}) := \operatorname{trd}(x/F_{\text{base}}) - d \operatorname{ldim}_{k_{\mathcal{O}}}(x/\Gamma(F_{\text{base}}))$$

satisfies $\delta(x/F_{\text{base}}) \geqslant 0$ for all tuples x from $\Gamma(F)$.

(4) Strong Γ-closedness: *For every irreducible subvariety V of G^n defined over F and of dimension dn which is free and rotund for the $\mathcal{O}$-module structure on G, and every finite tuple a from $\Gamma(F)$, there is $b \in V(F) \cap \Gamma(F)^n$ such that b is $k_{\mathcal{O}}$-linearly independent over $\Gamma(F_{\text{base}}) \cup a$ (that is, no nonzero $k_{\mathcal{O}}$-linear combination of the b_i lies in the $k_{\mathcal{O}}$-linear span of $\Gamma(F_{\text{base}}) \cup a$).*

(5) Countable closure property: *For each finite subset X of F, the Γ-closure $\Gamma\mathrm{cl}^F(X)$ of X in F is countable.*

Observe that if F is a Γ-field for the $\mathcal{O}$-module G, and in particular if it is full and so satisfies axiom (1), then it satisfies axioms (2) and (3) if and only if $F_{\mathrm{base}} \lhd F$.

Lemma 8.3. *If F satisfies axioms* (1)–(3) *then it also satisfies axiom* (4) *if and only if it is $\aleph_0$-saturated for Γ-algebraic extensions* (*in the sense of Definition 5.14*).

Proof. First assume that F satisfies axioms (1)–(4). Suppose $A \lhd F$ is finitely generated over F_{base} and $A \overset{\lhd}{\hookrightarrow} B$ is a finitely generated Γ-algebraic extension. We have assumed that F_{base} is essentially finitary, so A^{full} and B^{full} are unique up to isomorphism as extensions of A and B, respectively, so A^{full} embeds (strongly) into B^{full}. Choose an embedding. Since A^{full} embeds in F we have $A^{\mathrm{full}} \wedge B$ (the intersection taken in B^{full}) embedding (strongly) into F, as summarized in the diagram below:

$$\begin{tikzcd}
& & A^{\mathrm{full}} \arrow[r] \arrow[dr, dashed] & F \\
A \arrow[r] \arrow[urr] \arrow[drr] & A^{\mathrm{full}} \wedge B \arrow[ur, dashed] \arrow[dr, dashed] & & B^{\mathrm{full}} \\
& & B \arrow[ur] &
\end{tikzcd}$$

So it remains to embed B in F over $A^{\mathrm{full}} \wedge B$, so we may assume $A = A^{\mathrm{full}} \wedge B$. Let a be a basis for A over F_{base} and let $b \in \Gamma(B)^n$ be a good basis for B over A, which exists by Proposition 3.22. Let $V = \mathrm{Loc}(b/A)$, a subvariety of G^n. Then V is free and rotund by Corollary 7.4. Since B is Γ-algebraic over A we have $\delta(b/A) = 0$, so $\dim V = dn$.

Then, by axiom (4) applied to an irreducible component of V, there is $c \in \Gamma(F)^n \cap V(F)$, $k_{\mathcal{O}}$-linearly independent over $\Gamma(F_{\mathrm{base}}) \cup \{a\}$. Since $A \lhd F$ we have $\delta(c/A) \geqslant 0$, so $\mathrm{trd}(c/A) = dn = \dim V$. Thus c is generic in V over A. Let C be the Γ-subfield of F generated by A and c. Then c is a good basis of C over A because this is a property of $\mathrm{Loc}(c/A)$, that is, of V. So, by the definition of a good basis, C is isomorphic to B over A. So F is $\aleph_0$-saturated for Γ-algebraic extensions over F_{base}.

For the converse, suppose that F is $\aleph_0$-saturated for Γ-algebraic extensions over F_{base}. Let V be a free and rotund irreducible subvariety of G^n which is defined over F and of dimension dn, and let a be a finite tuple from $\Gamma(F)$. Extending a if necessary, we may assume that $A = \langle F_{\mathrm{base}}, a\rangle \lhd F$ and that V is defined over a.

Consider a Γ-field extension B of A, generated by a tuple $b \in \Gamma(B)^n$ such that $\mathrm{Loc}(b/A) = V$. By Proposition 7.3 the extension is strong. Since V is free, $\mathrm{ldim}_{k_{\mathcal{O}}}(b/\Gamma(A)) = n$ and therefore $\delta(b/A) = \dim V - dn = 0$. So B is a Γ-algebraic extension. Thus B embeds into F over A and so we have $b \in V(F) \cap \Gamma(F)^n$, which is $k_{\mathcal{O}}$-linearly independent over $\Gamma(F_{\mathrm{base}}) \cup \{a\}$ as required. $\square$

Proof of Theorem 8.2. Suppose F is a Γ-closed field. Then, by definition, F is (isomorphic to) a closed substructure of the canonical model M or is obtained from M (and its closed substructures) as the union of a directed system of closed embeddings. If F is a closed substructure of M then certainly it is a full Γ-field strongly extending F_{base}, so it satisfies axioms (1)–(3), and it is countable so satisfies axiom (5).

For axiom (4), suppose $A \lhd F$ is finitely generated and $A \overset{\lhd}{\to} B$ is a finitely generated and Γ-algebraic extension. Since M is $\mathcal{C}^{\text{fg}}$-saturated, B embeds strongly into M over A and since F is closed in M, $B \subseteq F$. So by Lemma 8.3, F satisfies axiom (4).

So closed substructures of M satisfy axioms (1)–(5). Axioms (1)–(4) are preserved under unions of directed systems of strong embeddings, and all the axioms are preserved under unions of directed systems of closed embeddings. Hence all Γ-closed fields satisfy all five axioms of $\Gamma\text{CF}_{\text{CCP}}(F_{\text{base}})$.

Suppose now that F satisfies axioms (1)–(5). Since it satisfies axioms (1)–(3), we have the pregeometry Γcl^F on F. If F_0 is a finite-dimensional substructure of F then F_0 satisfies axioms (1)–(3) and (5) immediately and, using Lemma 8.3, also axiom (4). Let $\bar{a}$ be a Γcl^F-basis for F_0. Using Lemma 4.13, for each $a_i \in \bar{a}$, choose $\alpha_i \in \Gamma(F_0)$, interalgebraic with a_i over F_{base}. Let $C = \langle F_{\text{base}}, \alpha_1, \dots, \alpha_n \rangle$. Then F_0 is Γ-algebraic over C and is saturated for Γ-algebraic extensions so, by Proposition 5.15, $F_0 \cong M_0(C)$. Now choose an embedding of C into M. Note that $\Gamma\text{cl}^M(C)$ is also Γ-algebraic over C and is saturated for Γ-algebraic extensions so is also isomorphic to $M_0(C)$. Hence F_0 is a Γ-closed field.

Now F is the union of the directed system of all its finite-dimensional closed substructures, which by CCP are countable, and the class of Γ-closed fields is closed under such unions by definition; hence F is a Γ-closed field. □

We can now prove Theorem 1.7.

Proof of Theorem 1.7. By Theorem 6.9, $M(F_{\text{base}})$ is a quasiminimal pregeometry structure, so by Fact 6.4 the class $\mathcal{K}(M(F_{\text{base}}))$ is uncountably categorical and every model is quasiminimal. By Theorem 8.2, the list of axioms $\Gamma\text{CF}_{\text{CCP}}(F_{\text{base}})$ axiomatizes the class $\mathcal{K}(M(F_{\text{base}}))$. □

Remarks 8.4. (1) It is easy to show that axioms (1)–(4) are $L_{\omega_1,\omega}$-expressible, and axiom (5) is expressible as an $L_{\omega_1,\omega}(Q)$-sentence.

(2) If we add an ($L_{\omega_1,\omega}$-expressible) axiom stating that F is infinite dimensional to axioms (1)–(4), the only countable model is M and so we get an $\aleph_0$-categorical, and hence complete, $L_{\omega_1,\omega}$-sentence.

9. Specific applications of the general construction

We list several instances of Γ-fields that are of interest, starting with the original example.

9A. ***Pseudo-exponentiation.*** We take $K_0 = \mathbb{Q}$, $G_1 = \mathbb{G}_a$, and $G_2 = \mathbb{G}_m$. Set $\mathcal{O} = \mathbb{Z}$. Let τ be transcendental, and take F_{base} to be the field $\mathbb{Q}^{\text{ab}}(\tau)$, where $\mathbb{Q}^{\text{ab}}$ is the extension of $\mathbb{Q}$ by all roots of unity. For each $m \in \mathbb{N}^+$, choose a primitive m-th root of unity ω_m such that for all $m, n \in \mathbb{N}^+$ we have $(\omega_{mn})^n = \omega_m$. We take $\Gamma(F_{\text{base}})$ to be the graph of a homomorphism from the $\mathbb{Q}$-linear span of τ to the roots of unity such that $\tau/m \mapsto \omega_m$ for each $m \in \mathbb{N}^+$. (This F_{base} is called SK in the paper [Kirby 2013].)

Then the construction gives a class of fields F with a predicate $\Gamma(F)$ defining the graph of a surjective homomorphism from $\mathbb{G}_a(F)$ to $\mathbb{G}_m(F)$, with kernel $\tau\mathbb{Z}$, which we denote by exp. The predimension inequality is precisely Schanuel's conjecture, and the strong existential closedness axiom is known as strong exponential-algebraic closedness. Thus we obtain a proof of Theorem 1.2, which we restate in explicit form.

Theorem 9.1. *Up to isomorphism, there is exactly one model* $\langle F; +, \cdot, \exp\rangle$ *of each uncountable cardinality of the following list* $\mathrm{ECF}_{\mathrm{SK,CCP}}$ *of axioms.*

(1) ELA-field: *F is an algebraically closed field of characteristic* 0, *and* exp *is a surjective homomorphism from* $\mathbb{G}_{\mathrm{a}}(F)$ *to* $\mathbb{G}_{\mathrm{m}}(F)$.

(2) Standard kernel: *The kernel of* exp *is an infinite cyclic group generated by a transcendental element* τ.

(3) Schanuel property: *The* predimension function

$$\delta(\bar{x}) := \mathrm{trd}(\bar{x}, \exp(\bar{x})) - \mathrm{ldim}_{\mathbb{Q}}(\bar{x})$$

satisfies $\delta(\bar{x}) \geqslant 0$ *for all tuples* $\bar{x}$ *from* F.

(4) Strong exponential-algebraic closedness: *If V is a rotund, additively and multiplicatively free subvariety of* $\mathbb{G}_{\mathrm{a}}^n \times \mathbb{G}_{\mathrm{m}}^n$ *defined over F and of dimension n, and $\bar{a}$ is a finite tuple from F, then there is $\bar{x}$ in F such that* $(\bar{x}, e^{\bar{x}}) \in V$ *and $\bar{x}$ is $\mathbb{Q}$-linearly independent over $\bar{a}$ (that is, no nonzero $\mathbb{Q}$-linear combination of the x_i lies in the $\mathbb{Q}$-linear span of the a_i).*

(5) Countable closure property: *For each finite subset X of F, the exponential algebraic closure* $\mathrm{ecl}^F(X)$ *of X in F is countable.*

Proof. We apply Theorem 1.7, but note that axioms (2) and (3) are slightly different from the axioms given in the statement of Theorem 8.2. The Schanuel property holds on our choice of F_{base} because τ is transcendental, and it follows from the addition property for δ that the two versions of axiom 3 are equivalent in this case. Since τ is transcendental and the kernel is standard, it follows that F_{base} embeds strongly in F, so the two versions of axiom (2) are also equivalent. □

We denote the canonical model of cardinality continuum by $\mathbb{B}$.

9B. ***Incorporating a counterexample to Schanuel's conjecture.*** We proceed as in the previous example, except now we choose an irreducible polynomial $P(x, y) \in \mathbb{Z}[x, y]$ and take (ϵ, τ) to be a generic zero of the polynomial $P(x, y)$. (We assume that P is such that neither ϵ nor τ is zero.) Choose a division sequence (ϵ_m) for ϵ, that is, numbers such that $\epsilon_1 = \epsilon$ and $(\epsilon_{mn})^n = \epsilon_m$ for all $m, n \in \mathbb{N}^+$. Now take K to be the field $\mathbb{Q}^{\mathrm{ab}}(\tau, (\epsilon_m)_{m\in\mathbb{N}^+})$, and define $\Gamma(K)$ to be the graph of a homomorphism from the $\mathbb{Q}$-linear span of τ and 1, with $\tau/m \mapsto \omega_m$ as above and $1/m \mapsto \epsilon_m$.

Now the construction gives us a canonical model $\mathbb{B}_P$, the unique model of cardinality continuum of almost the same list of axioms as those for $\mathbb{B}$, except that Schanuel's conjecture has this exception with the formal analogues ϵ and τ of e and $2\pi i$ being algebraically dependent via the polynomial P. More precisely, the predimension axiom is replaced by an axiom scheme stating that $\exp(1)$ and τ are transcendental, that $P(\exp(1), \tau) = 0$, and the condition that for all tuples $\bar{a}$, $\mathrm{trd}(\bar{a}, \exp(\bar{a})/\tau, \exp(1)) - \mathrm{ldim}_{\mathbb{Q}}(\bar{a}/\tau, 1) \geqslant 0$.

More generally, we can take any finitely generated partial exponential field with standard kernel (that is, a finitely generated Γ-field for the appropriate groups and kernels) as F_{base} and do the same construction to build a quasiminimal exponential field $\mathbb{M}(F_{\mathrm{base}})$ of size continuum with counterexamples to the Schanuel property within a finite-dimensional $\mathbb{Q}$-vector space, but with the Schanuel property holding over that vector space. Each $\mathbb{M}(F_{\mathrm{base}})$ is unique up to isomorphism as a model of appropriate axioms, just as $\mathbb{B}$ is. One could conjecture that $\mathbb{C}_{\mathrm{exp}}$ is isomorphic to one of these. Several people have asked us if it might be possible to prove Schanuel's conjecture easily by some method showing that $\mathbb{C}_{\mathrm{exp}}$ must be isomorphic to $\mathbb{B}$, just because $\mathbb{B}$ is categorical. Examples such as these show that soft methods which ignore transcendental number theory and analytic considerations cannot hope to work.

9C. *Pseudo-Weierstrass $\wp$-functions.* Let E be an elliptic curve over a number field K_0. Choose a Weierstrass equation for E,

$$Y^2Z = 4X^3 - g_2XZ^2 - g_3Z^3$$

with $g_2, g_3 \in K_0$, which fixes an embedding of E into projective space $\mathbb{P}^2$, with homogeneous coordinates $[X : Y : Z]$. Apart from the point $O = [0 : 1 : 0]$ at infinity, we can identify E with its affine part, given by solutions in $\mathbb{A}^2$ of the equation

$$y^2 - 4x^3 - g_2x - g_3 = 0.$$

For our construction we take $G_1 = \mathbb{G}_a$ and $G_2 = E$, and we take $\mathcal{O} = \mathrm{End}(E)$, so $\mathcal{O} = \mathbb{Z}$ if E does not have complex multiplication (CM) and $\mathcal{O} = \mathbb{Z}[\tau]$ if E has CM by the imaginary quadratic τ. In the CM case, we assume that $\tau \in K_0$ (and adjoin it if not). Take ω_1 transcendental over K_0 and ω_2 transcendental over $K_0(\omega_1)$ if E does not have CM, or $\omega_2 = \tau\omega_1$ if E has CM by τ.

As a field, we define $F_{\mathrm{base}} = K_0(\mathrm{Tor}(E), \omega_1, \omega_2)$, where $\mathrm{Tor}(E)$ means the full torsion group of E, which is contained in $E(\mathbb{Q}^{\mathrm{alg}})$. We define $\Gamma(F_{\mathrm{base}})$ to be the graph of a surjective $\mathcal{O}$-module homomorphism from $\mathbb{Q}\omega_1 + \mathbb{Q}\omega_2$ to $\mathrm{Tor}(E)$, with kernel $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$. While this may not specify F_{base} up to isomorphism, we will see that Serre's open image theorem allows us to specify F_{base} with only a finite amount of extra information.

In a model M, $\Gamma(M)$ is the graph of a surjective homomorphism $\exp_{E,M} : \mathbb{G}_a(M) \to E(M)$ with kernel Λ. Using our chosen embedding of E into $\mathbb{P}^2$, we can identify the components of the function $\exp_{E,M}$ with functions $\wp, \wp' : M \to M \cup \{\infty\}$, where $\exp_{E,M}(a) = [\wp(a) : \wp'(a) : 1]$. We call the function $\wp$ a "pseudo-Weierstrass $\wp$-function".

Note that in our model M, Λ is definable by the formula $\exp_{E,M}(x) = O$. In the non-CM case, $\mathbb{Z}$ is definable by the formula $\forall y\, [y \in \Lambda \to xy \in \Lambda]$, so $\mathbb{Q}$ is also definable as the field of fractions. In the CM case, these formulas define the rings $\mathbb{Z}[\tau]$ and $\mathbb{Q}[\tau]$.

Following [Gavrilovich 2008], we will apply the following version of Serre's open image theorem to show that only a finite amount of extra information is required to specify F_{base} as a Γ-field.

Fact 9.2. *Let E be an elliptic curve defined over a number field K_0. Then there exists an $m \in \mathbb{N}$ such that every* $\mathrm{End}(E)$*-module automorphism of the torsion* $\mathrm{Tor}(E)$ *which fixes the m-torsion $E[m]$ pointwise is* *induced by a field automorphism over K_0, that is, the natural homomorphism*

$$\mathrm{Gal}(K_0^{\mathrm{alg}}/K_0(E[m])) \to \mathrm{Aut}_{\mathrm{End}(E)}(\mathrm{Tor}(E)/E[m])$$

is surjective.

Proof. When $\mathrm{End}(E) \cong \mathbb{Z}$, this is Serre's open image theorem [Serre 1972, Introduction (3)]. When E has complex multiplication, it is the analogous classical open image theorem [Serre 1972, Section 4.5, Corollaire]. □

Unfortunately the proof does not give an effective bound for m, so given an explicit K_0 and E we do not know how to compute it.

We can now prove the first half of Theorem 1.6, which we restate precisely.

Theorem 9.3. *Let E be an elliptic curve over a number field $K_0 \subseteq \mathbb{C}$. Up to isomorphism, there is exactly one model of each uncountable cardinality of the following list $\wp\mathrm{CF}_{\mathrm{SK,CCP}}(E)$ of axioms, and these models are all quasiminimal.*

(1) Full $\wp$-field: *M is an algebraically closed field of characteristic* 0, *and Γ is the graph of a surjective homomorphism from $\mathbb{G}_{\mathrm{a}}(M)$ to $E(M)$, which we denote by* $\exp_{E,M}$. *We add parameters for the number field K_0.*

(2) Kernel and base (non-CM case): *There exist $\omega_1, \omega_2 \in \mathbb{G}_{\mathrm{a}}(M)$, $\mathbb{Q}$-linearly independent, such that the kernel of $\exp_{E,M}$ is of the form $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ and, for the number m specified by Fact 9.2, the algebraic type of the pair $(\exp_{E,M}(\omega_1/m), \exp_{E,M}(\omega_2/m)) \in E[m]^2$ over the parameters K_0 is specified.*

(2) Kernel and base (CM case): *There exists a nonzero $\omega_1 \in \mathbb{G}_{\mathrm{a}}(M)$ such that the kernel of $\exp_{E,M}$ is of the form $\mathbb{Z}[\tau]\omega_1$, and for the number m specified by Fact 9.2, the algebraic type of $\exp_{E,M}(\omega_1/m) \in E[m]$ over the parameters K_0 is specified.*

(3) Predimension inequality: *The* predimension function

$$\delta(\bar{x}) := \mathrm{trd}(\bar{x}, \exp_{E,M}(\bar{x})) - \mathrm{ldim}_{k_{\mathcal{O}}}(\bar{x})$$

satisfies $\delta(\bar{x}) \geqslant 0$ for all tuples $\bar{x}$ from M where $k_{\mathcal{O}} = \mathbb{Q}$ or $\mathbb{Q}(\tau)$, as appropriate.

(4) Strong $\wp$-algebraic closedness: *M satisfies the specific case of strong Γ-closedness from Theorem 8.2 for this choice of G, $\mathcal{O}$, and F_{base}.*

(5) Countable closure property: *As in Theorem 8.2.*

Proof. Again we must show that these axioms are equivalent to those given in Theorem 8.2. As in the exponential case, we have the absolute form of the predimension inequality here, which is equivalent to the relative statement over the base together with the assertion that ω_1 is transcendental and, in the non-CM case, that ω_2 is transcendental over $K_0(\omega_1)$. It remains to show that the axioms here specify the atomic diagram of F_{base}.

So suppose that M and M' are both models of the axioms, and their bases, namely the Γ-subfields generated by the kernels, are F_{base} and $F_{\mathrm{base}}{}'$. We have K_0 as a common subfield, and the axioms give us kernel generators $(\omega_1, \omega_2) \in M^2$ and $(\omega'_1, \omega'_2) \in M'^2$ such that $(\alpha_1, \alpha_2) := (\exp_{E,M}(\omega_1/m), \exp_{E,M}(\omega_2/m))$ and $(\alpha'_1, \alpha'_2) := (\exp_{E,M'}(\omega'_1/m), \exp_{E,M'}(\omega'_2/m))$ have the same algebraic type over K_0. (In the CM case we define $\omega_2 = \tau\omega_1$ and $\omega'_2 = \tau\omega'_1$ to treat the two cases at the same time.) The points $\alpha_i \in E$ generate the m-torsion subgroup $E[m]$ of E. So we can define a field isomorphism $\sigma_1 : K_0(E[m](M)) \to K_0(E[m](M'))$ by $\alpha_i \mapsto \alpha'_i$ for $i = 1, 2$. Then we extend σ_1 arbitrarily to a field isomorphism

$$\sigma_2 : K_0(\mathrm{Tor}(E)(M)) \to K_0(\mathrm{Tor}(E)(M')).$$

Now define an $\mathrm{End}(E)$-module automorphism of $\mathrm{Tor}(E)(M)$ by

$$\exp_{E,M}\left(\frac{\omega_1}{n_1} + \frac{\omega_2}{n_2}\right) \mapsto \sigma_2^{-1}\left(\exp_{E,M'}\left(\frac{\omega'_1}{n_1} + \frac{\omega'_2}{n_2}\right)\right)$$

for all $n_1, n_2 \in \mathbb{Z}$. By construction of σ_2 this automorphism fixes $E[m]$ pointwise, so by Fact 9.2 it extends to a field automorphism σ_3 of $K_0(\mathrm{Tor}(E)(M))$. So, defining $\sigma_4 = \sigma_2 \circ \sigma_3$, we get a field isomorphism $\sigma_4 : K_0(\mathrm{Tor}(E)(M)) \to K_0(\mathrm{Tor}(E)(M'))$ such that

$$\sigma_4\left(\exp_{E,M}\left(\frac{\omega_1}{n_1} + \frac{\omega_2}{n_2}\right)\right) = \exp_{E,M'}\left(\frac{\omega'_1}{n_1} + \frac{\omega'_2}{n_2}\right)$$

for all $n_1, n_2 \in \mathbb{Z}$.

The predimension inequality implies that (ω_1, ω_2) and (ω'_1, ω'_2) have the same field-theoretic type over $\mathbb{Q}^{\mathrm{alg}}$, so we can extend σ_4 to a field isomorphism σ_5 by defining $\sigma_5(\omega_i) = \omega'_i$ for $i = 1, 2$, and this σ_5 is a Γ-field isomorphism $F_{\mathrm{base}} \to F_{\mathrm{base}}{}'$ as required. □

Later in Proposition 10.1 we will show that the predimension inequality above is the appropriate form of Schanuel's conjecture for the $\wp$-functions, thereby completing the proof of Theorem 1.6.

9D. ***Variants on $\wp$-functions.*** As in the exponential case, we can do variant constructions by changing the base field F_{base} to a different finitely generated Γ-field, to incorporate some counterexamples to the predimension inequality. We can also do constructions of "pseudo-analytic" homomorphisms $\mathbb{G}_a(M) \to E(M)$ which have no complex-analytic analogue. For example, choose an elliptic curve E without complex multiplication and take the kernel lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ for $\omega_1/\omega_2 \in \mathbb{R}$ totally real (that is, algebraic and such that it and all its conjugates are real), for example real quadratic. The construction still works perfectly well to produce a unique quasiminimal model, but no embedding of Λ into $\mathbb{C}$ can be the kernel of a meromorphic homomorphism because it is dense on the line $\mathbb{R}\omega_1$.

9E. ***Exponential maps of simple abelian varieties.*** This is the algebraic setup corresponding to the complex example described in Definition 3.1. Take $G_1 = \mathbb{G}_a^d$ and G_2 a simple abelian variety of dimension d, defined over a number field K_0. Take $\mathcal{O} = \mathrm{End}(G_2)$, and suppose these endomorphisms are also defined over K_0. Fix an embedding of K_0 into $\mathbb{C}$.

Let $\omega_1, \ldots, \omega_{2d} \in \mathbb{C}^d$ be generators of a lattice Λ such that $\mathbb{C}^d/\Lambda$ is isomorphic to $G_2(\mathbb{C})$ as a complex $\mathcal{O}$-module manifold.

We take F_{base} to be the field generated by the ω_i together with $\mathrm{Tor}(G_2)$, and $\Gamma(F_{\text{base}})$ to be the graph of an $\mathcal{O}$-module homomorphism from $\mathbb{Q}\Lambda$ onto $\mathrm{Tor}(G_2)$.

For abelian varieties of dimension greater than 1 there is no nonconjectural analogue of Serre's open image theorem, so we cannot be more specific about an axiomatization of the atomic diagram of the base. So we have no improvement on the statement of Theorem 8.2 in this case.

9F. ***Factorizations via $\mathbb{G}_m$ of elliptic exponential maps.*** The examples so far have all been of the exponential type, case (EXP). Here, we give an example in case (COR). Let $G_1 = \mathbb{G}_{\mathrm{m}}$ and $G_2 = E$, an elliptic curve without complex multiplication, defined over a number field K_0. Let $\mathcal{O} = \mathbb{Z}$.

Let ω be transcendental. As a field, $F_{\text{base}} = K_0(\omega, \mathrm{Tor}(\mathbb{G}_{\mathrm{m}}), \mathrm{Tor}(E))$, and we define $\Gamma(F_{\text{base}})$ to be the graph of a surjective homomorphism from $\mathbb{Q}\omega + \mathrm{Tor}(\mathbb{G}_{\mathrm{m}})$ onto $\mathrm{Tor}(E)$ with kernel $\mathbb{Z}\omega$. Then for $\mathbb{M} = \mathbb{M}(F_{\text{base}})$, $\Gamma(\mathbb{M})$ is the graph of a surjective homomorphism $\theta_{\mathbb{M}} : \mathbb{G}_{\mathrm{m}}(\mathbb{M}) \to E(\mathbb{M})$.

In the complex case, the exponential map of E factors through the exponential map of $\mathbb{G}_{\mathrm{m}}$ as

$$\begin{tikzcd} \mathbb{G}_{\mathrm{a}}(\mathbb{C}) \arrow[r, "{[\wp:\wp':1]}"] \arrow[d, "\exp"] & E(\mathbb{C}) \\ \mathbb{G}_{\mathrm{m}}(\mathbb{C}) \arrow[ur, "\theta"'] & \end{tikzcd}$$

and this pseudo-analytic map $\theta_{\mathbb{M}}$ is an analogue of the complex map θ. Since $E \times \mathbb{G}_{\mathrm{m}}$ is not simple, the methods of this paper do not suffice to build a field F equipped with a map θ and pseudo-exponential maps of $\mathbb{G}_{\mathrm{m}}$ and E together, in which the analogue of the above commutative diagram would hold together with a suitable predimension inequality and a categoricity theorem for a reasonable axiomatization. However, it seems likely that this is achievable by combining the methods of this paper with those of [Kirby 2009].

Question 9.4. The main obstacle to stability for the first-order theory of $\mathbb{B}$ is the kernel. In this case the kernel is just a cyclic subgroup of $\mathbb{G}_{\mathrm{m}}$, and it is known that $\mathbb{G}_{\mathrm{m}}$ equipped with such a group is superstable. So it is natural to ask whether the first-order theory of $\mathbb{M}$ in this case is actually superstable. One could even ask if any construction of type (B), say with finite rank kernels, produces a structure with a superstable first-order theory.

9G. ***Differential equations.*** We now give an example of type (DEQ). Let K_0 be a countable field of characteristic 0, let G_2 be any simple semiabelian variety of dimension d defined over K_0, and let $G_1 = \mathbb{G}_{\mathrm{a}}^d$. Let $\mathcal{O} = \mathrm{End}(G_2)$. Let C be a countable algebraically closed field extending K_0, and define $\Gamma(C) = G(C)$.

Now consider the amalgamation construction using C as the base but considering only purely Γ-transcendental extensions of C, that is, using the category $\mathcal{C}_{\Gamma\text{-tr}}(C)$ in place of $\mathcal{C}(C)$. Theorem 6.9 shows we have a quasiminimal pregeometry structure, and hence a canonical model in each uncountable cardinality. The models we obtain are quasiminimal and $\Gamma\mathrm{cl}(\varnothing) = C$.

This construction is also considered in [Kirby 2009], where it is shown that the first-order theory of these models is $\aleph_0$-stable. In that paper it is also shown that if $\langle F; +, \cdot, D\rangle$ is a differentially closed field and we define $\Gamma(F)$ to be the solution set to the exponential differential equation for G_2, then the reduct $\langle F; +, \cdot, \Gamma\rangle$ is a model of the same first-order theory, and C is the field of constants for the differential field. The paper [Kirby 2009] also considers the situation with several different groups Γ_S relating to the exponential differential equations of different semiabelian varieties S, which do not have to be simple, but they do have to be defined over the constant field C.

10. Comparison with the analytic models

Zilber conjectured that $\mathbb{C}_{\exp} \cong \mathbb{B}$ and it seems reasonable to extend the conjecture to the exponential map of any simple complex abelian variety, and indeed to other analytic functions such as the function θ from Section 9F. In each example, axioms (1) and (2) are set up to describe properties we know about these analytic functions, so verifying the conjecture amounts to verifying the other three axioms. We consider the progress towards each of the axioms in turn.

10A. ***The predimension inequalities.*** For the usual exponential function, the predimension inequality states that for all tuples a from $\mathbb{C}$, $\operatorname{trd}(a, e^a) \geqslant \operatorname{ldim}_{\mathbb{Q}}(a)$. This is precisely Schanuel's conjecture.

In the case of an elliptic curve E defined over a number field, the predimension inequality states that for all tuples a from $\mathbb{C}$, $\delta_E(a) = \operatorname{trd}(a, \exp_{E,\mathbb{C}}(a)) - \operatorname{ldim}_{k_{\mathcal{O}}}(a) \geqslant 0$.

Proposition 10.1. *The predimension inequality above for the exponential map of an elliptic curve follows from the André–Grothendieck conjecture on the periods of* 1*-motives.*

The following proof was explained to us by Juan Diego Caycedo, and follows the proof of a related statement in Section 3 of [Caycedo and Zilber 2014].

Proof. By Théorème 1.2 of [Bertolin 2002], with $s = 0$ and $n = 1$, a special case of André's conjecture (building on Grothendieck's earlier conjecture) states that if $j(E)$ is the j-invariant of E, ω_1 and ω_2 are the periods of E, η_1 and η_2 are the quasiperiods of E, $P_1, \dots, P_n \in E(\mathbb{C})$, a_i is the integral of the first kind associated with P_i, and d_i is the integral of the second kind associated with P_i, then

$$\operatorname{trd}(2\pi i, j(E), \omega_1, \omega_2, \eta_1, \eta_2, \bar{P}, a, d) \geqslant 2\operatorname{ldim}_{k_{\mathcal{O}}}(a/\omega_1, \omega_2) + 4[k_{\mathcal{O}} : \mathbb{Q}]^{-1}. \tag{1}$$

In this case, we have that $P_i = [\wp(a_i) : \wp'(a_i) : 1] = \exp_{E,\mathbb{C}}(a_i)$. Since our E is defined over a number field, $j(E)$ is algebraic. The Legendre relation states $\omega_1\eta_2 - \omega_2\eta_1 = 2\pi i$, so $j(E)$ and $2\pi i$ do not contribute to the above inequality.

If we assume that $a_1, \dots, a_n \in \mathbb{C}$ are $k_{\mathcal{O}}$-linearly independent over ω_1, ω_2, we can discard the integrals of the second kind to get the bound

$$\operatorname{trd}(\omega_1, \omega_2, \eta_1, \eta_2, a, \exp_{E,\mathbb{C}}(a)) \geqslant \operatorname{ldim}_{k_{\mathcal{O}}}(a/\omega_1, \omega_2) + 4[k_{\mathcal{O}} : \mathbb{Q}]^{-1}. \tag{2}$$

Consider the case where there is no CM, so $k_{\mathcal{O}} = \mathbb{Q}$. Throwing away η_1 and η_2 we get

$$\operatorname{trd}(\omega_1, \omega_2, a, \exp_{E,\mathbb{C}}(a)) \geqslant \operatorname{ldim}_{k_{\mathcal{O}}}(a/\omega_1, \omega_2) + 2. \tag{3}$$

From the case $n = 0$ we see that $\mathrm{trd}(\omega_1, \omega_2) = 2$ and since ω_1 and ω_2 are $\mathbb{Q}$-linearly independent we have $\delta_E(\omega_1, \omega_2) = 0$. Then (3) implies that for any a we have $\delta_E(a/\omega_1, \omega_2) \geqslant 0$, and putting these two statements together we deduce that $\delta_E(a) \geqslant 0$.

Where E does have CM (and is defined over a number field), Chudnovsky's theorem [1980, Theorem 1 and Corollary 2] gives us $\mathrm{trd}(\omega_1, \omega_2, \eta_1, \eta_2) = \mathrm{trd}(\omega_1, \pi) = 2$, so in particular $\mathrm{trd}(\omega_1) = 1$. We also have $k_{\mathcal{O}} = \mathbb{Q}(\tau)$ with $[k_{\mathcal{O}} : \mathbb{Q}] = 2$, and $\omega_2 = \omega_1\tau$, so we can discard ω_2, η_1 and η_2 from (2) to obtain

$$\mathrm{trd}(\omega_1, a, \exp_{E,\mathbb{C}}(a)) \geqslant \mathrm{ldim}_{k_{\mathcal{O}}}(a/\omega_1) + 1. \tag{4}$$

The same argument now shows that $\delta_E(a) \geqslant 0$ for any tuple a. □

Proof of Theorem 1.6. Theorem 9.3 shows that the axioms $\wp\mathrm{CF}_{\mathrm{SK,CCP}}(E)$ are uncountably categorical and that every model is quasiminimal. The analytic structure $\mathbb{C}_\wp$ is a model of the first two axioms by construction. Proposition 10.1 shows that the predimension inequality given in axiom (3) is the appropriate analogue of Schanuel's conjecture for $\wp$-functions. Axiom (5), the countable closure property, was proved in this case in [Jones et al. 2016]. □

We will give another proof of the countable closure property in Theorem 1.8 in this paper.

Our understanding of the periods conjecture uses Bertolin's translation to remove the motives, which she did only in the cases of elliptic curves and $\mathbb{G}_{\mathrm{m}}$. For abelian varieties of dimension greater than 1 we suspect that the predimension inequality axiom again follows from the André–Grothendieck periods conjecture, but there are more complications because the Mumford–Tate group plays a role and so we have not been able to verify it.

10B. ***Strong Γ-closedness.*** In the case of the usual exponentiation for $\mathbb{G}_{\mathrm{m}}$, Mantova [2016] currently has the best result towards proving the strong Γ-closedness in the complex case. They only consider the case of a variety $V \subseteq G^n$ where $n = 1$. A free and rotund $V \subseteq G^1$ is just the solution set of an irreducible polynomial $p(x, y) \in \mathbb{C}[x, y]$ which depends on both x and y, that is, the partial derivatives $\partial p/\partial x$ and $\partial p/\partial y$ are both nonzero.

Fact 10.2. *Suppose $p(x, y) \in \mathbb{C}[x, y]$ depends on both x and y. Then there are infinitely many points $x \in \mathbb{C}$ such that $p(x, e^x) = 0$. Furthermore, suppose Schanuel's conjecture is true and let a be a finite tuple from $\mathbb{C}$. Then there is $x \in \mathbb{C}$ such that (x, e^x) is a generic zero of p over a.*

The observation that there are infinitely many solutions, and the whole statement in the case that p is defined over $\mathbb{Q}^{\mathrm{alg}}$ is due to Marker [2006]. The general case stated above is due to Mantova [2016, Theorem 1.2].

10C. ***Γ-closedness.*** In Section 11 we will see that for some purposes strong Γ-closedness can be weakened to Γ-closedness.

Definition 10.3. A full Γ-field F is Γ-*closed* if for every irreducible subvariety V of G^n defined over F and of dimension dn which is free and rotund, $V(F) \cap \Gamma(F)^n$ is Zariski-dense in V.

Using the classical Rabinovich trick, one can easily show this axiom scheme is equivalent to the existence of a single point $\beta \in V(F) \cap \Gamma(F)^n$, for every such V. For the usual exponentiation, Γ-closedness is known as exponential-algebraic closedness. In this direction, Brownawell and Masser [2017, Proposition 2] have the following.

Fact 10.4. *If $V \subseteq (\mathbb{G}_a \times \mathbb{G}_m)^n(\mathbb{C})$ is an algebraic subvariety of dimension n which projects dominantly to $\mathbb{G}_a^n$ then there is $a \in \mathbb{C}^n$ such that $(a, e^a) \in V$.*

In this case V can be taken free without loss of generality, and the condition of projecting dominantly to $\mathbb{G}_a^n$ implies rotundity. However it is much stronger than rotundity. Another exposition of this theorem is given in [D'Aquino et al. 2018].

10D. ***The pregeometry and the countable closure property.*** To compare the pregeometry of our constructions such as $\mathbb{B}$ with the complex analytic models such as $\mathbb{C}_{\exp}$ we have to define the appropriate pregeometry on the complex field. Given a Γ-field F, we defined a Γ-subfield A of F to be Γ-closed in F if whenever $A \subseteq B \subseteq F$ with $\delta(B/A) \leqslant 0$ then $B \subseteq A$. One can construct Γ-fields with no proper Γ-closed subfields. Fortunately we are able to show unconditionally that there is a countable Γ-subfield of $\mathbb{C}$ which is Γ-closed in $\mathbb{C}$.

For $\mathbb{C}_{\exp}$, this was done in [Kirby 2010a] by adapting the proof of Ax's differential forms version of Schanuel's conjecture. A similar proof was given in [Jones et al. 2016] for elliptic curves. The same method ought to work for any semiabelian varieties, but here we give a different approach, applying the main result of [Ax 1972] directly to generalize a theorem of Zilber in the exponential case [Zilber 2005b, Theorem 5.12].

Let $\mathbb{C}_\Gamma$ be an analytic Γ-field, which we recall means a Γ-field from Definition 3.1 or 3.2. Then Γ is a complex Lie subgroup of $G(\mathbb{C})$, and $L\Gamma \leq LG(\mathbb{C})$ is the graph of a $\mathbb{C}$-linear isomorphism between $LG_1(\mathbb{C})$ and $LG_2(\mathbb{C})$. Thus Γ^n is a complex Lie subgroup of $G^n(\mathbb{C})$, so has a complex topology. It might not be a closed subgroup, so the topology might not be the subspace topology.

Definition 10.5. Given an algebraic subvariety $V \subseteq G^n$, write V^{isol} for the set of all isolated points of $V(\mathbb{C}) \cap \Gamma^n$ with respect to the complex topology on Γ^n. For any subset $A \subseteq \mathbb{C}$ we define $\Gamma\mathrm{cl}'(A)$ to be the subfield of $\mathbb{C}$ generated by the union of all V^{isol}, where V ranges over the algebraic subvarieties $V \subseteq G^n$ which are defined over $K_0(A)$. We consider $\Gamma\mathrm{cl}'(A)$ as a Γ-subfield of $\mathbb{C}$ by defining $\Gamma(\Gamma\mathrm{cl}'(A)) := \Gamma \cap G(\Gamma\mathrm{cl}'(A))$.

Lemma 10.6. *$\Gamma\mathrm{cl}'$ is a closure operator on $\mathbb{C}$ and for any $A \subseteq \mathbb{C}$, $\Gamma\mathrm{cl}'(A)$ is a full Γ-subfield of $\mathbb{C}$ of cardinality $|A| + \aleph_0$.*

Proof. For transitivity, suppose x is a finite tuple from $\Gamma\mathrm{cl}'(A)$ and $y \in \Gamma\mathrm{cl}'(Ax)$. We may reduce to the case that $\alpha \in W(\mathbb{C}) \cap \Gamma^n$ is isolated and the tuple x lists the coordinates of α, and $\beta \in V(\mathbb{C}) \cap \Gamma^m$ is isolated and y is a coordinate of β, with W defined over $K_0(A)$ and V over $K_0(Ax)$. Then V can be written as a fibre $V'(\alpha)$ of a subvariety $V' \subseteq G^{m+n}$ over $K_0(A)$ projecting to $W \subseteq G^n$. Then $\beta\alpha \in V' \cap \Gamma^{m+n}$ is an isolated point, so $\beta \in \Gamma\mathrm{cl}'(A)$.

Now let $A \subseteq \mathbb{C}$, and set $A' := \mathrm{acl}^{\mathbb{C}}(K_0(A))$. Let $\alpha_0 \in G_1(A')$, let $V_0 \subseteq G_1$ be the set of its conjugates, i.e., the 0-dimensional locus of α_0 over $K_0(A)$, and let $V := V_0 \times G_2 \subseteq G$. Then $V \cap \Gamma$ is nonempty, since $\pi_1 : \Gamma \twoheadrightarrow G_1(\mathbb{C})$ is surjective, and it consists of isolated points since $\ker(\pi_1)$ does. This shows that $A' \subseteq \Gamma\mathrm{cl}'(A)$, and hence in particular that $\Gamma\mathrm{cl}'$ is a closure operator. Furthermore, this argument shows that $\Gamma\mathrm{cl}'(A)$ is full.

Finally, for the cardinality calculation, note there are only $(|A| + \aleph_0)$-many algebraic varieties V defined over A and for each there can be only countably many isolated points in $V(\mathbb{C}) \cap \Gamma^n$. □

Proposition 10.7. *For an analytic* Γ*-field* $\mathbb{C}_\Gamma$, *the closure operators* $\Gamma\mathrm{cl}$ *and* $\Gamma\mathrm{cl}'$ *are the same. In particular*, $\Gamma\mathrm{cl}'$ *is a pregeometry on* $\mathbb{C}$.

To prove this we will use a lemma and Ax's theorem on the transversality of intersections between analytic subgroups and algebraic varieties.

Lemma 10.8. *If* $H \leq G^n$ *is a connected algebraic subgroup which is free then the analytic subgroup* $H(\mathbb{C}) + \Gamma^n$ *is equal to* $G^n(\mathbb{C})$.

Proof. Since G_2 is simple and not isogenous to G_1, every algebraic subgroup of G^n is of the form $H_1 \times H_2$ with H_i a subgroup of G_i^n, and since H is free it is G_2-free. Since $\mathcal{O} = \mathrm{End}(G_2)$ it follows that H is of the form $H_1 \times G_2^n$. Now since $\pi_1(\Gamma^n) = G_1^n(\mathbb{C})$ we see $H(\mathbb{C}) + \Gamma^n = G^n(\mathbb{C})$. □

Fact 10.9 [Ax 1972, Corollary 1]. *Suppose that* $\mathcal{G}$ *is a complex algebraic group*, A *is a connected analytic subgroup of* $\mathcal{G}$, U *is open in* $\mathcal{G}$, *and* X *is an irreducible analytic subvariety of* U *such that* $X \subseteq A$, X^{Zar} *is the Zariski closure of* X, *and* H *is the smallest algebraic subgroup of* $\mathcal{G}$ *containing* X. *Then*

$$\dim(H + A) \leqslant \dim X^{\mathrm{Zar}} + \dim A - \dim X.$$

Proof of Proposition 10.7. Suppose that $A \subseteq \mathbb{C}$ is $\Gamma\mathrm{cl}$-closed. Let $\alpha \in V^{\mathrm{isol}}$ for some V defined over $K_0(A)$. Replacing α with a subtuple if necessary, we may assume α is $k_{\mathcal{O}}$-linearly independent over $\Gamma(A)$. Then α is a smooth point of V and of Γ^n, so by considering tangent spaces we see that $\mathrm{trd}(\alpha/A) \leqslant \dim V \leqslant nd$, and hence $\delta(\alpha/A) \leqslant 0$. So $\alpha \in \Gamma^n(A)$ and hence A is $\Gamma\mathrm{cl}'$-closed.

Now suppose A is $\Gamma\mathrm{cl}'$-closed, and that $A \subseteq B \subseteq \mathbb{C}$ is a proper finitely generated Γ-field extension in $\mathbb{C}_\Gamma$. Let $b \in \Gamma^n(B)$ be a basis for the extension and let $V = \mathrm{Loc}(b/A)$. We will show that $\delta(b/A) > 0$.

Let X be an irreducible analytic component containing b of the analytic subset $V(\mathbb{C}) \cap \Gamma^n$ of the complex Lie group Γ^n. Since A is $\Gamma\mathrm{cl}'$-closed and $b \notin A$, $\dim X \geqslant 1$.

We claim that X has some point in A. To see this, let e be a smooth point of X, and take regular local coordinates η_i at e in G^n such that X is locally the graph of a function from the first $\dim X$ coordinates to the rest. A is algebraically closed as a field, so is topologically dense in $\mathbb{C}$. So there is a point $a \in X$ close to e such that the first $\dim X$ coordinates are in A. Let W be the intersection of V with $\eta_i = a_i$ for $i = 1, \ldots, \dim X$. Then W is defined over A and a is an isolated point of $W(\mathbb{C}) \cap \Gamma^n$; hence a is in $G^n(A)$ as required.

Suppose X^{Zar} is not G_1-free, so say $(x, y) \in X^{\mathrm{Zar}}$ implies a nontrivial $\mathcal{O}$-linear equation $\sum_{j=1}^{n} r_j x_j = c$. Then this equation holds for $\pi_1(a)$, so $c \in G_1(A)$. But then since $b \in X$, already $(x, y) \in V$ implies this equation, so V is not G_1-free, a contradiction since V is the locus of a basis over A. The same proof shows that X^{Zar} is G_2-free, so it is free.

Let H be the algebraic subgroup of G^n generated by $X^{\mathrm{Zar}}(\mathbb{C}) - b$. Then $X^{\mathrm{Zar}}(\mathbb{C}) - b$ is free, so H is free. So by Lemma 10.8, the subgroup $H + \Gamma^n$ is equal to G^n.

Applying Fact 10.9 we get

$$\dim(H + \Gamma^n) \leqslant \dim(X^{\mathrm{Zar}} - b) + \dim \Gamma^n - \dim(X - b),$$

which gives

$$2dn \leqslant \dim X^{\mathrm{Zar}} + dn - \dim X,$$

but $\dim X > 0$ and $X^{\mathrm{Zar}} \subseteq V$ so we deduce that $\dim V > nd$. Thus $\delta(b/A) > 0$. So A is Γcl-closed, as required. □

Proof of Theorem 1.8. Proposition 10.7 shows that $\Gamma\mathrm{cl} = \Gamma\mathrm{cl}'$, and so Lemma 10.6 shows that Γcl has the countable closure property. □

Remark 10.10. In [Kirby 2010a], an algebraic version of the isolated points closure $\Gamma\mathrm{cl}'$ was given, using the fact that a solution to a system of $2n$ equations of analytic functions in $2n$ variables is isolated if and only if a certain Jacobian matrix does not vanish at the point. So this gives a definition of a closure operator ecl which makes sense on any exponential field, and it was shown in [Kirby 2010a] that ecl-closed sets are strong and agree with the Γcl-closed sets as we have defined them here, and in particular that ecl is always a pregeometry. This algebraic definition of the closure operator was suggested by Macintyre [1996], although it had previously been used in the real and complex cases by Khovanskii and by Wilkie.

11. Generically Γ-closed fields

In this section we consider Γ-fields which may not be strongly Γ-closed but are generically strongly Γ-closed. Using the variant of the amalgamation construction from Section 5D, we show that such Γ-fields are also quasiminimal and that the *strong* part of strong Γ-closedness becomes redundant in this generic case.

Let K be a full Γ-field. Recall from Section 5D that an extension $K \subseteq A$ of K is *purely* Γ*-transcendental* if and only if $K \lhd_{\mathrm{cl}} A$ if and only if for all tuples b from $\Gamma(A)$, either $\delta(b/K) > 0$ or $b \subseteq \Gamma(K)$. Clearly an extension A of K is purely Γ-transcendental if and only if all of its finitely generated subextensions are.

11A. ***Generic Γ-closedness.*** Recall from Theorem 8.2 that a full Γ-field F is *strongly* Γ*-closed* if for every irreducible subvariety V of G^n defined over F and of dimension dn which is free and rotund for the $\mathcal{O}$-module structure on G, and every finite tuple a from $\Gamma(F)$, there is $b \in V(F) \cap \Gamma(F)^n$ such that b is $k_{\mathcal{O}}$-linearly independent over $\Gamma(F_{\mathrm{base}}) \cup a$.

Assuming the Schanuel property, that is, that $F_{\mathrm{base}} \lhd F$, we can extend the tuple a to a finite tuple a' such that $A := \langle F_{\mathrm{base}}, a' \rangle \lhd F$ and V is defined over A. Then taking $b \in V(F) \cap \Gamma(F)^n$ which is

$k_{\mathcal{O}}$-linearly independent over $\Gamma(F_{\text{base}}) \cup a'$, that is, over $\Gamma(A)$, we see that

$$0 \leqslant \delta(b/A) = \operatorname{trd}(b/A) - d \operatorname{ldim}_{k_{\mathcal{O}}}(b/\Gamma(A)) = \operatorname{trd}(b/A) - dn.$$

So $\operatorname{trd}(b/A) \geqslant dn$, but $b \in V$, V is defined over A, and $\dim(V) = dn$. So b is generic in V over A, and hence over $F_{\text{base}}(a)$. So it follows that in the presence of the Schanuel property, strong Γ-closedness is equivalent to the condition that for any V and a as above, there is $b \in V(F) \cap \Gamma(F)^n$ which is generic in V over $F_{\text{base}}(a)$.

Recall also the weaker property of Γ-closedness from Definition 10.3: F is Γ-*closed* if for every irreducible subvariety V of G^n defined over F and of dimension dn which is free and rotund, $V(F) \cap \Gamma(F)^n$ is Zariski-dense in V.

For the concept of generic Γ-closedness with respect to a subfield K, we need to consider extensions of the form $K \lhd_{\text{cl}} A \lhd B$ where A and B are finitely generated as extensions of the full Γ-field K. Say α is a basis of A over K and β is a basis of B over A, and that $V := \operatorname{Loc}(\beta/A)$ and $W := \operatorname{Loc}(\alpha, \beta/K)$. We also assume that $A^{\text{full}} \wedge B = A$. Then by Corollary 7.4 both V and W are free, V is rotund, and W is strongly rotund.

Definition 11.1 (generic Γ-closedness). Let F be a full Γ-field and $K \lhd_{\text{cl}} F$, $K \neq F$. Suppose $V \subseteq G^n$ is free and rotund, irreducible, and of dimension dn. Suppose also that there is $\alpha \in \Gamma(F)^r$, $k_{\mathcal{O}}$-linearly independent over $\Gamma(K)$ such that V is defined over $K(\alpha)$, and such that for $\beta \in V$, generic over $K(\alpha)$, the variety $W := \operatorname{Loc}(\alpha, \beta/K)$ is strongly rotund.

We say that F is *generically* Γ-*closed over* K (GΓC over K) if, for all such V, we have that $V(F) \cap \Gamma^n(F)$ is Zariski-dense in V.

F is *generically strongly* Γ-*closed over* K (GSΓC over K) if, whenever V and α are as above, there is $\gamma \in V(F) \cap \Gamma^n(F)$, $k_{\mathcal{O}}$-linearly independent over $\Gamma(K) \cup \alpha$.

We say F is GΓC or GSΓC without reference to K to mean G(S)ΓC over $\Gamma\mathrm{cl}^F(\varnothing)$.

Proposition 11.2. *Suppose F is a full Γ-field and $K \lhd_{\text{cl}} F$, $K \neq F$. Then F is GSΓC over K if and only if F is $\aleph_0$-saturated for Γ-algebraic extensions which are purely Γ-transcendental over K.*

Proof. This is essentially the same as the proof of Lemma 8.3. □

It is immediate from the definitions that GSΓC over K implies GΓC over K. In [Kirby and Zilber 2014] it was shown that if the conjecture on intersections with tori (CIT, now also known as the multiplicative case of the Zilber–Pink conjecture) is true, then any exponential field satisfying the Schanuel property which is exponentially-algebraically closed is also strongly exponentially algebraically closed. We use similar ideas here to prove that GΓC over K implies GSΓC over K. The Schanuel property is replaced by the assumption that K is Γ-closed in F, and instead of the Zilber–Pink conjecture it is enough to use the weak version, which is a theorem.

Given any variety S and subvarieties W, $V \subseteq S$, the typical dimension of $W \cap V$ is $\dim W + \dim V - \dim S$. If X is an irreducible component of $W \cap V$ it is said to have *atypical dimension* (for the intersection) if

$$\dim X > \dim W + \dim V - \dim S.$$

We also say that X is an *atypical component* of the intersection. For the multiplicative group, the *weak Zilber–Pink* theorem is known as *weak CIT* and is due to Zilber [2002, Corollary 3]. The semiabelian form of the statement is the following theorem [Kirby 2009, Theorem 4.6].

Fact 11.3 ("semiabelian weak Zilber–Pink", basic version). *Let S be a semiabelian variety, defined over an algebraically closed field of characteristic* 0. *Let $(W_b)_{b\in B}$ be a constructible family of constructible subsets of S. That is, B is a constructible set and W is a constructible subset of $B \times S$, with W_b the obvious projection of a fibre. Then there is a finite set $\mathcal{H}_W$ of connected proper algebraic subgroups of S such that for any $b \in B$ and any coset $c+J$ of any connected algebraic subgroup J of S, if X is an irreducible component of $W_b \cap c+J$ of atypical dimension with $c \in X$ then there is $H \in \mathcal{H}_W$ such that $X \subseteq c+H$.*

We also need a version for subvarieties not of S but of varieties of the form $U \times S$, which is sometimes called a "horizontal" family of semiabelian varieties.

Theorem 11.4 ("horizontal semiabelian weak Zilber–Pink"). *Let S be a semiabelian variety and let U be any variety. Let $(W_b)_{b\in B}$ be a constructible family of constructible subsets of $U \times S$. Then there is a finite set $\mathcal{H}_W$ of connected proper algebraic subgroups of S such that for any $b \in B$ and any coset $c+J$ of any connected algebraic subgroup J of S, if X is an irreducible component of $W_b \cap (U \times c+J)$ of atypical dimension with $c \in X$ then there is $H \in \mathcal{H}_W$ such that $X \subseteq U \times c+H$. Furthermore, H can be chosen such that we have*

$$\dim X \leqslant \dim(W_b \cap (U \times c+H)) + \dim(H \cap J) - \dim H. \tag{$*$}$$

Proof. First suppose U is a point, so $U \times S \cong S$. The main part of the statement is then Fact 11.3. For the "furthermore" part, suppose $(*)$ does not hold for the H we chose from $\mathcal{H}_W$. Then rename $\mathcal{H}_W$ as $\mathcal{H}_W^1$. We give an inductive argument to find a new $\mathcal{H}_W$ which suffices. We have the irreducible X as a component of the intersection $(W_b \cap c+H) \cap c+(H \cap J)$, and the failure of $(*)$ says that X is atypical as a component of this intersection considered as an intersection of subvarieties of $c+H$. Translating everything by c, we get that $X-c$ is an atypical component of the intersection $(W_b - c \cap H) \cap (H \cap J)$ inside H. Now apply Fact 11.3 again with H in place of S to get a proper connected algebraic subgroup H' of H from the finite set $\mathcal{H}_W^2 := \mathcal{H}_W^1 \cup \bigcup_{H\in\mathcal{H}_W^1} \mathcal{H}_{(W_b - c\cap H)_{b,c}}$ such that $X \subseteq c+H'$. If necessary we can iterate this construction and since $\dim H' < \dim H$ it stops after finitely many steps, and the $\mathcal{H}_W$ we eventually find is still finite.

Now consider arbitrary U. Suppose first that all fibres of the coordinate projection $\pi : W_b \to S$ have the same dimension k, constant with respect to b. Take $\mathcal{H}_W$ to be the finite set $\mathcal{H}_{(\pi(W_b))_{b\in B}}$ given by this theorem with U a point. We will see that this works as required.

Indeed, let $c+J$ be a coset in S, let X be an atypical component of $W_b \cap (U \times c+J)$, let Y be any irreducible component of $\pi(W_b) \cap c+J$ containing $\pi(X)$, and let $H \in \mathcal{H}_{(\pi(W_b))_{b\in B}}$ be as given by the theorem.

Then by considering dimensions of fibres, we have

$$\begin{aligned}\dim X &= \dim(\pi(X)) + k \\ &\leqslant \dim Y + k \\ &\leqslant \dim(\pi(W_b) \cap c{+}H) + \dim(H \cap J) - \dim H + k \\ &= \dim(W_b \cap (U \times c{+}H)) + \dim(H \cap J) - \dim H.\end{aligned}$$

Now for a general family $W \subseteq B \times U \times S$, write $\pi : U \times S \to S$ for the projection and define

$$W^k = \{(b, u, s) \in W \mid \dim(W_b \cap \pi^{-1}(s)) = k\}$$

for each $k = 0, \dots, \dim W$. By the definability of dimension, these W^k are all constructible subsets of W, partitioning it, and each W^k satisfies the above constancy condition on fibres. For any $c{+}J$, any component X of $W_b \cap (U \times c{+}J)$ contains a dense constructible subset which lies in some piece W^k_b. So we can take $\mathcal{H}_W$ to be $\bigcup_k \mathcal{H}_{W^k}$. □

Now we can prove that GSΓC and GΓC are equivalent.

Proposition 11.5. *Suppose F is a full Γ-field and $K \lhd_{\mathrm{cl}} F$, $K \neq F$. Then F is GSΓC over K if and only if it is GΓC over K.*

Proof. As remarked earlier, it is immediate that GSΓC over K implies GΓC over K. So assume F is GΓC over K. Let V, α, β, and W be as given in Definition 11.1. Let $V_{\alpha,\mathrm{dep}}$ be the set of points of $V(F)$ which are $k_{\mathcal{O}}$-linearly dependent over $\Gamma(K) \cup \alpha$. We shall find a proper Zariski-closed subset of V containing $V_{\alpha,\mathrm{dep}}$.

We first work in case (EXP), so G_2 is a simple semiabelian variety of dimension d and $G_1 = \mathbb{G}_\mathrm{a}^d$, which we identify with the Lie algebra LG_2 of G_2.

For a $d(r+n)$-square matrix M and a $d(r+n)$-column vector c, let $\Lambda_{M,c} \subseteq \mathbb{G}_\mathrm{a}^{d(r+n)}$ be given by $x \in \Lambda_{M,c}$ if and only if $Mx = c$. So as M and c vary, we get the family of all possible affine linear subspaces. Let $U_{M,c} = W \cap (\Lambda_{M,c} \times G_2^{r+n})$.

Now suppose $\xi \in V_{\alpha,\mathrm{dep}} \cap \Gamma(F)^n$. Let $\zeta = (\alpha, \xi) \in \Gamma(F)^{r+n}$. We also write $\zeta = (\zeta_1, \zeta_2) \in G_1^{r+n} \times G_2^{r+n}$. Let J be the smallest algebraic subgroup of G_2^{r+n} such that ζ_2 lies in a K-coset of J, say $\zeta_2 \in c_2' + J$ with $c_2' \in G_2(K)^{r+n}$. Since $\xi \in V_{\alpha,\mathrm{dep}}$, we see that J is a proper algebraic subgroup of G_2^{r+n}.

By Lemma 2.1, there is $M \in \mathrm{Mat}_{r+n}(\mathcal{O})$ such that $J = (\ker(M))^o$ and $LJ = \ker(M)$. Since K is a full Γ-field, there is $c_1' \in G_1^{r+n}(K)$ such that $c' := (c_1', c_2') \in \Gamma(K)^{r+n}$. Then $M(\zeta - c') \in \Gamma(K)^{r+n}$ since $K \leq F$ preserves the kernels. Now $\Gamma(K)$ is a $k_{\mathcal{O}}$-subspace of $\Gamma(F)$, so in particular is existentially closed as an $\mathcal{O}$-submodule, so there exists $c'' \in \Gamma(K)^{r+n}$ such that $Mc'' = M(\zeta - c')$. Set $c = (c_1, c_2) := c' + c'' \in \Gamma(K)^{r+n}$, so $M(\zeta - c) = 0$. Then $\zeta - c$ is divisible in $\ker(M)$, since $\Gamma(F)$ is divisible and torsion-free, so $\zeta - c \in LJ \times J$.

Now we have $\zeta \in U_{M,c_1} \cap (G_1^{r+n} \times c_2{+}J)$. Let X be the irreducible component of this intersection containing ζ.

We next show that X has atypical dimension for the intersection. From the definition of the predimension δ and its relationship with $\Gamma\dim$ we have

$$\begin{aligned}\dim X \geqslant \operatorname{trd}(\zeta/K) &= \delta(\zeta/K) + d\,\operatorname{ldim}_{\mathcal{O}}(\zeta/\Gamma(K))\\ &= \delta(\zeta/K) + \dim J\\ &\geqslant \Gamma\dim^F(\zeta/K) + \dim J,\\ \dim X &\geqslant \Gamma\dim^F(\alpha/K) + \dim J. \end{aligned}\tag{5}$$

Since $\alpha \in \Gamma(F)^r$ was chosen $k_{\mathcal{O}}$-linearly independent over $\Gamma(K)$ and such that $\langle K, \alpha\rangle \lhd F$, we have

$$\Gamma\dim^F(\alpha/K) = \delta(\alpha/K) = \operatorname{trd}(\alpha/K) - d\,\operatorname{ldim}_{\mathcal{O}}(\alpha/\Gamma(K)) = \operatorname{trd}(\alpha/K) - dr. \tag{6}$$

Since $W = \operatorname{Loc}(\alpha, \beta/K)$, and $V = \operatorname{Loc}(\beta/K(\alpha))$ has dimension dn, using (6) we have

$$\dim W = \dim V + \operatorname{trd}(\alpha/K) = d(r+n) + \Gamma\dim^F(\alpha/K). \tag{7}$$

From (5) and (7),

$$\begin{aligned}\dim X &\geqslant \dim W + \dim J - d(r+n)\\ &= \dim W + (\dim J + d(r+n)) - 2d(r+n)\\ &= \dim W + \dim(G_1^{r+n} \times c_2{+}J) - \dim(G^{r+n}),\end{aligned}$$

but W is free, so $\dim U_{M,c_1} < \dim W$ and so

$$\dim X > \dim U_{M,c_1} + \dim(G_1^{r+n} \times c_2{+}J) - \dim(G^{r+n}). \tag{8}$$

So X has atypical dimension.

Applying Theorem 11.4, there is a proper algebraic subgroup H of G_2^{r+n} from the finite set $\mathcal{H}_U$ such that $X \subseteq G_1^{r+n} \times c{+}H$. We have $\zeta \in X$, so $\zeta_2 \in c{+}H$. J was chosen as the smallest algebraic subgroup of S such that ζ_2 lies in a K-coset of J, so $J \subseteq H$ and hence $H \cap J = J$. So, from the "furthermore" clause of Theorem 11.4 we have

$$\dim X \leqslant \dim(U_{M,c_1} \cap (G_1^{r+n} \times c_2{+}J)) + \dim J - \dim H. \tag{9}$$

We write $TJ = LJ \times J$ and $TH = LH \times H$, thinking of them as the tangent bundles. Then we have

$$\begin{aligned}U_{M,c_1} \cap (G_1^{r+n} \times c_2{+}J) &= W \cap (c_1{+}LJ \times G_2^{r+n}) \cap (G_1^{r+n} \times c_2{+}J)\\ &= W \cap c{+}TJ\\ &= W \cap \zeta{+}TJ\\ &\subseteq W \cap \zeta{+}TH,\end{aligned}$$

so

$$\dim(U_{M,c_1} \cap (G_1^{r+n} \times c_2{+}J)) \leqslant \dim(W \cap \zeta{+}TH). \tag{10}$$

Combining (9), (10), and (5) we get

$$\Gamma\dim^F(\alpha/K) + \dim J \leqslant \dim(W \cap \zeta{+}TH) + \dim J - \dim H,$$
$$\Gamma\dim^F(\alpha/K) + \dim H \leqslant \dim(W \cap \zeta{+}TH). \tag{11}$$

By Lemma 2.1, there is $M \in \mathrm{Mat}_{r+n}(\mathcal{O})$ such that $H = \ker(M)^o$ and $LH = \ker(M)$, so $TH = \ker(M)^o$ for the action of $\mathcal{O}$ on G. So $M : G^{r+n} \to G^{r+n}$ factors as $G^{r+n} \twoheadrightarrow G^{r+n}/TH \to G^{r+n}$, where the first homomorphism is the quotient map, and the kernel of the second homomorphism is the finite group $\ker(M)/\ker(M)^o$. Now let $\theta_H : W \twoheadrightarrow W/TH$ be the restriction of the quotient map $G^{r+n} \twoheadrightarrow G^{r+n}/TH$. Then $\dim(W/TH) = \dim(M \cdot W)$. Now H is a proper subgroup of G_2^{r+n}, and $\dim(H) = \dim(\ker(M)) = d(r+n-\mathrm{rk}\, M)$, and so M is nonzero. Then since W is strongly rotund,

$$\dim(W/TH) = \dim(M \cdot W) > d\,\mathrm{rk}\, M = d(r+n) - \dim H.$$

So using the fibre dimension theorem, the dimension of a typical fibre of θ_H is

$$\begin{aligned}\dim(\text{typical fibre}) &= \dim W - \dim(W/TH)\\ &< (d(r+n) + \Gamma\dim^F(\alpha/K)) - (d(r+n) - \dim H)\\ &= \Gamma\dim^F(\alpha/K) + \dim H.\end{aligned} \tag{12}$$

The fibre of θ_H in which ζ lies is $W \cap \zeta{+}TH$, so (11) says exactly that ζ lies in a fibre of θ_H of atypical dimension. By the fibre dimension theorem, there is a proper Zariski-closed subset W_H of W, defined over K, containing all the fibres of θ_H of atypical dimension.

Since α is generic in the projection of W, and hence of W_H, the subset $V_{H,\alpha} := \{y \in V \mid (\alpha, y) \in W_H\}$ is proper Zariski-closed in V. Let $V_\alpha := \bigcup_{H \in \mathcal{H}_W} H_{H,\alpha}$. Then V_α is also a proper Zariski-closed subset of V, and we have shown that $V_{\alpha,\mathrm{dep}} \subseteq V_\alpha$.

So since F is GΓC over K, there is a point $\beta \in \Gamma(F)^n \cap V(F) \smallsetminus V_\alpha(F)$. Since $\beta \notin V_{\alpha,\mathrm{dep}}$, β is $k_{\mathcal{O}}$-linearly independent over $\Gamma(K) \cup \alpha$. Hence F is GSΓC over K as required.

The proof for case (COR) is very similar, but instead of $\zeta \in (c_1{+}LJ) \times (c_2{+}J)$ we have subgroups $J_1 \subseteq G_1^{r+n}$ and $J_2 \subseteq G_2^{r+n}$ which correspond to each other in the sense that they are connected components of solutions to the same system of $\mathcal{O}$-linear equations. So we get $\zeta \in (c_1{+}J_1) \times (c_2{+}J_2)$ with $\dim J_1 = \dim J_2 = \mathrm{ldim}_{k_{\mathcal{O}}}(\zeta / \Gamma(K)) < r+n$. Then a similar calculation shows that ζ lies in a component of the intersection $W \cap c{+}(J_1 \times J_2)$ of atypical dimension, and we apply the weak Zilber–Pink for the semiabelian variety G^{r+n} and proceed as in case (EXP). □

11B. ***Sufficient conditions for quasiminimality.***

Theorem 11.6. *Suppose F is a full Γ-field with the countable closure property which is generically Γ-closed over some countable $K \lhd_{\mathrm{cl}} F$. Then F is quasiminimal.*

Proof. We take $F_{\mathrm{base}} = K$ and consider the category $\mathcal{C}_{\Gamma\text{-tr}}(K)$. By Theorem 5.21 it is an amalgamation category so we have a Fraïssé limit $M_{\Gamma\text{-tr}}(K)$. By Theorem 6.9, $M_{\Gamma\text{-tr}}(K)$ is a quasiminimal pregeometry structure, so defines a quasiminimal class $\mathcal{K}(M_{\Gamma\text{-tr}}(K))$. Substituting Proposition 11.2 for Lemma 8.3, the proof of Theorem 8.2 shows that the models in this class are precisely the full Γ-fields which are purely Γ-transcendental extensions of K, are $\aleph_0$-saturated with respect to the Γ-algebraic extensions which are purely Γ-transcendental over K, and satisfy the countable closure property. Hence by Propositions 11.2 and 11.5, F is in $\mathcal{K}(M_{\Gamma\text{-tr}}(K))$ and hence is quasiminimal. □

If F is the complex field, in practice it might be difficult or impossible to identify a countable Γ-closed K and prove directly that F is generically Γ-closed over K. Thus the following corollaries may be more useful.

Corollary 11.7. *Suppose F is a full Γ-field with the countable closure property which is* Γ-closed. *Then F is quasiminimal.*

Proof. Clearly Γ-closedness implies generic Γ-closedness. □

Since $\mathbb{C}_{\exp}$ has the countable closure property by Theorem 1.8, this completes the proof of Theorem 1.5. We can do slightly better.

Corollary 11.8. *Suppose F is a full Γ-field with the countable closure property which is* almost Γ-closed. *That is, for all but countably many free and rotund, irreducible subvarieties $V \subseteq G^n$ of dimension dn, $\Gamma^n(F) \cap V(F)$ is Zariski-dense in V. Then F is quasiminimal.*

Proof. Suppose F is almost Γ-closed, and take K_1 to be a countable subfield of F over which all the countably many exceptional varieties V are defined. Take $K = \Gamma\mathrm{cl}^F(K_1)$. Then F is generically Γ-closed over K. □

Overall we have proved the following generalization of Theorem 1.5, which applies to the exponential function, the Weierstrass $\wp$-functions, the exponential maps of simple abelian varieties, and more.

Theorem 11.9. *Let $\mathbb{C}_\Gamma$ be an analytic Γ-field. If $\mathbb{C}_\Gamma$ is almost Γ-closed then it is quasiminimal.*

Proof. Combine Corollary 11.8 with Theorem 1.8, which gives CCP. □

Since being Γ-closed implies being almost Γ-closed, Theorem 1.9 follows. Theorem 1.5 is a special case.

Remark 11.10. We do not know if almost Γ-closedness is a necessary condition for quasiminimality. For example in the exponential case, is it possible to build an uncountable quasiminimal exponential field F with a definable family $(V_p)_{p\in P}$ of rotund and free varieties such that for only countably many p (perhaps none) there is $(\bar{x}, e^{\bar{x}}) \in V_p(F)$?

Acknowledgements

Some of this work was carried out in the Max Planck Institute for Mathematics, Bonn during the programme on Model Theory and Applications, Spring 2012. Further work was carried out at the Mathematical Sciences Research Institute, Berkeley, during the programme on Model Theory, Arithmetic Geometry and Number Theory, Spring 2014, which was supported by the NSF under Grant No. 0932078 000. Our thanks to both institutions for their support and hospitality. We would like to thank Boris Zilber and Misha Gavrilovich for useful discussions. Particular thanks are due to Juan Diego Caycedo, whose suggestions on building a pseudo-$\wp$-function were part of the initial foundations of this project, and who further contributed to the project in its earlier stages. We would also like to thank the anonymous referee for some helpful comments.

References

[Ax 1972] J. Ax, "Some topics in differential algebraic geometry, I: Analytic subgroups of algebraic groups", *Amer. J. Math.* **94** (1972), 1195–1204. MR Zbl

[Bays and Kirby 2013] M. Bays and J. Kirby, "Excellence and uncountable categoricity of Zilber's exponential fields", preprint, 2013. arXiv

[Bays et al. 2014a] M. Bays, M. Gavrilovich, and M. Hils, "Some definability results in abstract Kummer theory", *Int. Math. Res. Not.* **2014**:14 (2014), 3975–4000. MR Zbl

[Bays et al. 2014b] M. Bays, B. Hart, T. Hyttinen, M. Kesälä, and J. Kirby, "Quasiminimal structures and excellence", *Bull. Lond. Math. Soc.* **46**:1 (2014), 155–163. MR Zbl

[Bays et al. 2014c] M. Bays, B. Hart, and A. Pillay, "Universal covers of commutative finite Morley rank groups", preprint, 2014. arXiv

[Bertolin 2002] C. Bertolin, "Périodes de 1-motifs et transcendance", *J. Number Theory* **97**:2 (2002), 204–221. MR Zbl

[Bertrand 2011] D. Bertrand, "Galois descent in Galois theories", pp. 1–24 in *Arithmetic and Galois theories of differential equations*, edited by L. Di Vizio and T. Rivoal, Sémin. Congr. **23**, Soc. Math. France, Paris, 2011. MR arXiv

[Brownawell and Masser 2017] W. D. Brownawell and D. W. Masser, "Zero estimates with moving targets", *J. Lond. Math. Soc.* (2) **95**:2 (2017), 441–454. MR Zbl

[Caycedo and Zilber 2014] J. D. Caycedo and B. Zilber, "Natural models of theories of green points", preprint, 2014. arXiv

[Chudnovsky 1980] G. V. Chudnovsky, "Algebraic independence of values of exponential and elliptic functions", pp. 339–350 in *Proceedings of the International Congress of Mathematicians* (Helsinki, 1978), vol. 1, edited by O. Lehto, Acad. Sci. Fennica, Helsinki, 1980. MR Zbl

[D'Aquino et al. 2018] P. D'Aquino, A. Fornasiero, and G. Terzo, "Generic solutions of equations with iterated exponentials", *Trans. Amer. Math. Soc.* **370**:2 (2018), 1393–1407. MR Zbl

[Droste and Göbel 1992] M. Droste and R. Göbel, "A categorical theorem on universal objects and its application in abelian group theory and computer science", pp. 49–74 in *Proceedings of the International Conference on Algebra* (Novosibirsk, 1989), part III, edited by L. A. Bokut' et al., Contemp. Math. **131**, American Mathematical Society, Providence, RI, 1992. MR Zbl

[Gavrilovich 2008] M. Gavrilovich, "A remark on transitivity of Galois action on the set of uniquely divisible abelian extensions in $\mathrm{Ext}^1(E(\overline{\mathbb{Q}}), \Lambda)$", *K-Theory* **38**:2 (2008), 135–152. MR Zbl

[Henderson 2014] R. S. Henderson, *Independence in exponential fields*, Ph.D. thesis, University of East Anglia, 2014, available at https://tinyurl.com/henders14.

[Hrushovski 1993] E. Hrushovski, "A new strongly minimal set", *Ann. Pure Appl. Logic* **62**:2 (1993), 147–166. MR Zbl

[Jones et al. 2016] G. Jones, J. Kirby, and T. Servi, "Local interdefinability of Weierstrass elliptic functions", *J. Inst. Math. Jussieu* **15**:4 (2016), 673–691. MR Zbl

[Kamensky 2007] M. Kamensky, "Ind- and pro- definable sets", *Ann. Pure Appl. Logic* **147**:3 (2007), 180–186. MR Zbl

[Kirby 2009] J. Kirby, "The theory of the exponential differential equations of semiabelian varieties", *Selecta Math.* (*N.S.*) **15**:3 (2009), 445–486. MR Zbl

[Kirby 2010a] J. Kirby, "Exponential algebraicity in exponential fields", *Bull. Lond. Math. Soc.* **42**:5 (2010), 879–890. MR Zbl

[Kirby 2010b] J. Kirby, "On quasiminimal excellent classes", *J. Symbolic Logic* **75**:2 (2010), 551–564. MR Zbl

[Kirby 2013] J. Kirby, "Finitely presented exponential fields", *Algebra Number Theory* **7**:4 (2013), 943–980. MR Zbl

[Kirby and Zilber 2014] J. Kirby and B. Zilber, "Exponentially closed fields and the conjecture on intersections with tori", *Ann. Pure Appl. Logic* **165**:11 (2014), 1680–1706. MR Zbl

[Macintyre 1996] A. J. Macintyre, "Exponential algebra", pp. 191–210 in *Logic and algebra* (Pontignano, Italy, 1994), edited by A. Ursini and P. Aglianò, Lecture Notes Pure Appl. Math. **180**, Dekker, New York, 1996. MR Zbl

[Mantova 2016] V. Mantova, "Polynomial-exponential equations and Zilber's conjecture", *Bull. Lond. Math. Soc.* **48**:2 (2016), 309–320. MR Zbl

[Marker 2006] D. Marker, "A remark on Zilber's pseudoexponentiation", *J. Symbolic Logic* **71**:3 (2006), 791–798. MR Zbl

[Mumford 1970] D. Mumford, *Abelian varieties*, Tata Inst. Fundamental Res. Stud. Math. **5**, Oxford Univ. Press, 1970. MR Zbl

[Pillay 1998] A. Pillay, "Model theory of algebraically closed fields", pp. 61–84 in *Model theory and algebraic geometry*, edited by E. Bouscaren, Lecture Notes in Math. **1696**, Springer, 1998. MR Zbl

[Serre 1972] J.-P. Serre, "Propriétés galoisiennes des points d'ordre fini des courbes elliptiques", *Invent. Math.* **15**:4 (1972), 259–331. MR Zbl

[Shelah 1983] S. Shelah, "Classification theory for nonelementary classes, I: The number of uncountable models of $\psi \in L_{\omega_1,\omega}$, part A", *Israel J. Math.* **46**:3 (1983), 212–240. MR Zbl

[Wilkie 1996] A. J. Wilkie, "Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function", *J. Amer. Math. Soc.* **9**:4 (1996), 1051–1094. MR Zbl

[Ziegler 2006] M. Ziegler, "A note on generic types", preprint, 2006. arXiv

[Zilber 2000] B. Zilber, "Fields with pseudo-exponentiation", preprint, 2000. arXiv

[Zilber 2002] B. Zilber, "Exponential sums equations and the Schanuel conjecture", *J. Lond. Math. Soc.* (2) **65**:1 (2002), 27–44. MR Zbl

[Zilber 2003] B. Zilber, "Raising to powers in algebraically closed fields", *J. Math. Log.* **3**:2 (2003), 217–238. MR Zbl

[Zilber 2005a] B. Zilber, "A categoricity theorem for quasi-minimal excellent classes", pp. 297–306 in *Logic and its applications* (Ann Arbor, MI, 2003), edited by A. Blass and Y. Zhang, Contemp. Math. **380**, American Mathematical Society, Providence, RI, 2005. MR Zbl

[Zilber 2005b] B. Zilber, "Pseudo-exponentiation on algebraically closed fields of characteristic zero", *Ann. Pure Appl. Logic* **132**:1 (2005), 67–95. MR Zbl

[Zilber 2015] B. Zilber, "The theory of exponential sums", preprint, 2015. arXiv

Communicated by Jonathan Pila
Received 2017-03-17 Revised 2017-11-13 Accepted 2017-12-26

m.bays@math.uni-muenster.de *Institut für Mathematische Logik und Grundlagenforschung, Fachbereich Mathematik und Informatik, Universität Münster, Münster, Germany*

jonathan.kirby@uea.ac.uk *School of Mathematics, University of East Anglia, Norwich, United Kingdom*

ALGEBRA AND NUMBER THEORY 12:3 (2018)
dx.doi.org/10.2140/ant.2018.12.551

On faithfulness of the lifting for Hopf algebras and fusion categories

Pavel Etingof

To Alexander Kirillov, Jr. on his 50th birthday with admiration

We use a version of Haboush's theorem over complete local Noetherian rings to prove faithfulness of the lifting for semisimple cosemisimple Hopf algebras and separable (braided, symmetric) fusion categories from characteristic p to characteristic zero, showing that, moreover, any isomorphism between such structures can be reduced modulo p. This fills a gap in our earlier work. We also show that lifting of semisimple cosemisimple Hopf algebras is a fully faithful functor, and prove that lifting induces an isomorphism on Picard and Brauer–Picard groups. Finally, we show that a subcategory or quotient category of a separable multifusion category is separable (resolving an open question from our earlier work), and use this to show that certain classes of tensor functors between lifts of separable categories to characteristic zero can be reduced modulo p.

1. Introduction

Let k be an algebraically closed field of characteristic $p > 0$ and let $W(k)$ be its ring of Witt vectors, $I = (p) \subset W(k)$ the maximal ideal, K the fraction field of $W(k)$, and $\overline{K}$ its algebraic closure. In [Etingof and Gelaki 1998] it is shown that any semisimple cosemisimple Hopf algebra over k has a unique (up to an isomorphism) lift over $W(k)$. In [Etingof et al. 2005, Section 9], this result is extended to separable (braided, symmetric) fusion categories, i.e., those of nonzero global dimension.[1]

Moreover, in [Etingof et al. 2005, Section 9.3], it is claimed that lifting is faithful, i.e., if liftings of two Hopf algebras are isomorphic over $\overline{K}$ then these Hopf algebras are isomorphic, and similarly for categories (Theorem 9.6, Corollary 9.10). This is used in a number of subsequent papers.

However, it recently came to my attention that the proofs given in that paper for those faithfulness results are incomplete. Namely, the proof of Lemma 9.7 (used in the proof of Theorem 9.6) says that by Nakayama's lemma, it suffices to check the finiteness of a certain morphism ϕ of schemes over $W(k)$ modulo the maximal ideal I (i.e., over k). But it is, in fact, not clear how this follows from Nakayama's lemma. Namely, finiteness over k does imply finiteness over $W(k)/I^N$ for any $N \geq 1$, but this is not

MSC2010: 16T05.
Keywords: lifting, Hopf algebra, tensor category, separable.

[1] In [Etingof et al. 2005], separable fusion categories are called nondegenerate. But this terminology is confusing since for braided fusion categories, this term is also used in an entirely different sense: to refer to categories with trivial Müger center. For this reason, we adopt a better term "separable" introduced in [Douglas et al. 2013].

sufficient to conclude finiteness over $W(k)$.[2] In fact, the reductivity of the group of twists must be used in the proof; see Remark 2.12.

The main goal of this paper is therefore to provide complete proofs of the results on faithfulness of the lifting. This may be done by using results on geometric reductivity and power reductivity of reductive groups over rings; see [Seshadri 1977; Franjou and van der Kallen 2010]. We also prove results on integrality of stabilizers of liftings, and show that lifting is a fully faithful functor for semisimple cosemisimple Hopf algebras, and defines an isomorphism of Brauer–Picard and Picard groups of (braided) separable fusion categories. Finally, we prove that subcategories and quotient categories of separable categories are separable (resolving a question from [Etingof et al. 2005, Section 9.4]), and use this to prove that certain types of tensor functors between liftings of separable categories descend to positive characteristic.

The paper is organized as follows. In Section 2 we describe algebro-geometric preliminaries, i.e., the results on geometric reductivity and power reductivity and their applications. In Section 3 we apply these results to proving faithfulness of the lifting and integrality of stabilizers for semisimple cosemisimple Hopf algebras and prove that lifting of such Hopf algebras is a fully faithful functor. In Section 4 we generalize the results of Section 3 to tensor categories and tensor functors, thus providing complete proofs of the results of [Etingof et al. 2005, Section 9.3] and [Etingof and Gelaki 2000, Theorem 6.1]. Also, we apply these results to show that the Brauer–Picard and Picard groups of (braided) separable fusion categories are preserved by lifting. Finally, in Section 5 we show that a subcategory and quotient category of a separable category is separable, and apply it to prove a descent result for tensor functors between liftings of separable categories.

2. Auxiliary results from algebraic geometry

In this section we collect some auxiliary results from algebraic geometry that we will use below.

2A. ***Power reductivity.*** Let k be an algebraically closed field of characteristic $p > 0$ and let $W(k)$ be its ring of Witt vectors, $I = (p) \subset W(k)$ the maximal ideal, K the fraction field of $W(k)$ and $\overline{K}$ its algebraic closure.

If $\boldsymbol{X}$ is a scheme over a ring R and R' is a commutative R-algebra, then $\boldsymbol{X}_{R'}$ will denote the base change of $\boldsymbol{X}$ from R to R'.

By a *reductive group* over a commutative ring $\boldsymbol{k}$ we will mean a smooth affine group scheme with connected fibers, as in [SGA 3_{III} 1970]. Such a group $\boldsymbol{G}$ is split when it contains a split fiberwise maximal $\boldsymbol{k}$-torus as a closed $\boldsymbol{k}$-subgroup. In our applications, $\boldsymbol{G}$ will always be split, and, in fact, will be a quotient of a product of general linear groups by a central torus.

We start with restating a result from [Franjou and van der Kallen 2010], which is a combination of their Proposition 6 and Theorem 12.

[2] E.g., K is not module-finite over $W(k)$, even though it becomes module-finite (in fact, zero) upon reduction modulo I^N, and is also finite over K.

We call a ring homomorphism $\psi : B \to C$ *power-surjective* if for any element $c \in C$, some positive integer power c^N belongs to $\operatorname{Im}\psi$.

Proposition 2.1. *Let $\boldsymbol{k}$ be a commutative Noetherian ring and let $\boldsymbol{G}$ be a split reductive group over $\boldsymbol{k}$. Let A be a finitely generated commutative $\boldsymbol{k}$-algebra on which $\boldsymbol{G}$ acts rationally through algebra automorphisms. If J is a $\boldsymbol{G}$-invariant ideal in A, then the map induced by reducing mod J: $A^{\boldsymbol{G}} \to (A/J)^{\boldsymbol{G}}$ is power-surjective.*

In [Franjou and van der Kallen 2010] this property of $\boldsymbol{G}$ is called *power reductivity*. It is not assumed there that $\boldsymbol{k}$ is Noetherian, but we will use Proposition 2.1 only for Noetherian (in fact, complete local) rings $\boldsymbol{k}$.

Remark 2.2. It was explained to us by W. van der Kallen that power reductivity of $\boldsymbol{G}$ (i.e., Proposition 2.1) over complete local Noetherian rings $\boldsymbol{k}$ follows from [Thomason 1987, Theorem 3.8]. Namely, from this theorem one easily gets property (INT) of [Franjou and van der Kallen 2010] and then power reductivity. This is discussed in Section 2.4 of that paper and in Theorem 2.2 of [van der Kallen 2007]. (The latter paper assumes a base field, but this assumption is not essential.) Compare also with [Grosshans 1997, Theorem 16.9] and [Springer 1977, Lemma 2.4.7].

2B. *Faithfulness of lifting for reductive group actions.* We will need the following proposition, which is sufficient to justify the main results of [Etingof et al. 2005, Section 9.3].

Proposition 2.3. *Let $\boldsymbol{V}$ be a rational representation of a split reductive group $\boldsymbol{G}$ on an affine space defined over $W(k)$. Let $v_1, v_2 \in \boldsymbol{V}(W(k))$. Assume that the $\boldsymbol{G}$-orbits of the reductions $v_1^0, v_2^0 \in \boldsymbol{V}(k)$ are closed and disjoint. Then v_1, v_2 are not conjugate under $\boldsymbol{G}(\bar{K})$.*

Proof. This follows from [Seshadri 1977, Theorem 3, part (ii)] for $R = W(k)$, $X = \boldsymbol{V}$ and $Y = \boldsymbol{V}/\boldsymbol{G} := \operatorname{Spec} \mathcal{O}(\boldsymbol{V})^{\boldsymbol{G}}$. Namely, this theorem says that v_1^0, v_2^0 are *distinct* points of $Y(k)$. This implies that there exists a $\boldsymbol{G}$-invariant polynomial $f \in \mathcal{O}(\boldsymbol{V})$ such that $f(v_1^0) \neq f(v_2^0)$ in k. But $f(v_i^0)$ are the reductions of $f(v_i)$ mod I, so $f(v_1) \neq f(v_2)$ in $W(k)$. But then v_1, v_2 cannot be conjugate under $\boldsymbol{G}(\bar{K})$.

Here is another proof, using Proposition 2.1, for $\boldsymbol{k} = W(k)$, $A = \mathcal{O}(\boldsymbol{V})$ and $J = I\mathcal{O}(\boldsymbol{V})$. Proposition 2.1 says that for any $f \in \mathcal{O}(\boldsymbol{V}_k)^{\boldsymbol{G}(k)}$, some power f^N of f lifts to a $\boldsymbol{G}$-invariant $h \in \mathcal{O}(V)$. Now, by Haboush's theorem [1975], we can choose f so that $f(v_1^0) = 0$ and $f(v_2^0) = 1$. Then $h(v_1) \in I$ and $h(v_2) \in 1 + I$, hence $h(v_1) \neq h(v_2)$, and v_1, v_2 cannot be conjugate under $\boldsymbol{G}(\bar{K})$. □

Remark 2.4. The closedness assumption for $\boldsymbol{G}$-orbits in Proposition 2.3 cannot be removed. For instance, let $\boldsymbol{G} = \mathbb{G}_m$ and $\boldsymbol{V} = \mathbb{A}^1$ be the tautological representation of $\boldsymbol{G}$. Take $v_1 = p$ and $v_2 = 1$. Then $v_1^0 = 0$ and $v_2^0 = 1$, so their $\boldsymbol{G}$-orbits are disjoint. On the other hand, v_1 and v_2 are conjugate by the element $p \in \boldsymbol{G}$. Note that the orbit of v_2^0 is $\mathbb{A}^1 \setminus \{0\}$, hence not closed.

The reductivity assumption cannot be removed, either. Namely, let $\boldsymbol{G} = \mathbb{G}_a$ be the group of translations $x \mapsto x + b$, and $\boldsymbol{V}$ be the 2-dimensional representation of $\boldsymbol{G}$ on linear (not necessarily homogeneous) functions of x. Let $v_1 = px$ and $v_2 = 1 + px$. Then $v_1^0 = 0$, $v_2^0 = 1$. These are $\boldsymbol{G}$-invariant vectors, so their orbits are closed and disjoint. Still, v_1 is conjugate to v_2 by the transformation $x \mapsto x + 1/p$.

2C. ***Integrality of the stabilizer.*** Let $\boldsymbol{V}$ be a rational representation of a split reductive group $\boldsymbol{G}$ on an affine space defined over $W(k)$. Let $v \in \boldsymbol{V}(W(k))$ and $v^0 \in \boldsymbol{V}(k)$ be its reduction modulo p. Let $\boldsymbol{S} \subset \boldsymbol{G}$ be the scheme-theoretic stabilizer of v, and $S = \boldsymbol{S}(\bar{K})$ be the stabilizer of v in $\boldsymbol{G}(\bar{K})$. Let $\boldsymbol{S}_0 = \boldsymbol{S}_k \subset \boldsymbol{G}_k$ be the scheme-theoretic stabilizer of v^0, and $S_0 = \boldsymbol{S}_0(k) = \boldsymbol{S}(k)$ be the stabilizer of v^0 in $\boldsymbol{G}(k)$.

Proposition 2.5. *Assume that*

(i) *the $\boldsymbol{G}$-orbit of v^0 is closed*;

(ii) *$\boldsymbol{S}_0$ is finite and reduced*;

(iii) *there is a lifting map $i : S_0 \to S$ such that $i(S_0) \in \boldsymbol{G}(W(k))$ and the reduction of $i(g_0)$ modulo p equals g_0 for any $g_0 \in S_0$. In other words, the reduction map $\boldsymbol{S}(W(k)) \to \boldsymbol{S}(k) = S_0$ is surjective.*

Then i is an isomorphism. In other words, all the natural maps in the diagram

$$\boldsymbol{S}(k) \leftarrow \boldsymbol{S}(W(k)) \hookrightarrow \boldsymbol{S}(K) \hookrightarrow \boldsymbol{S}(\bar{K})$$

are isomorphisms.

Proof. Let $\boldsymbol{U}$ be a defining representation of $\boldsymbol{G}$, presenting it as a closed subgroup of $\mathrm{GL}(\boldsymbol{U})$. The main part of the proof is showing that the matrix elements of any $g \in S$ in some basis of $\boldsymbol{U}(W(k))$ are integral over $W(k)$. To this end, we want to construct a lot of S-invariants in $\mathcal{O}(\boldsymbol{U})$, so that $\mathcal{O}(\boldsymbol{U})$ is integral over the subalgebra generated by these invariants.

Let $\boldsymbol{X} = \boldsymbol{G}v^0 \subset \boldsymbol{V}_k$ be the $\boldsymbol{G}$-orbit of v^0 over k, which is closed by (i). Since $\boldsymbol{S}_0$ is reduced by (ii), the natural morphism $\boldsymbol{G}_k/\boldsymbol{S}_0 \to \boldsymbol{V}_k$ given by $g \mapsto gv^0$ defines an isomorphism $\boldsymbol{G}_k/\boldsymbol{S}_0 \cong \boldsymbol{X}$. Therefore, we have $\mathcal{O}(\boldsymbol{X} \times \boldsymbol{U}_k)^{\boldsymbol{G}(k)} = \mathcal{O}(\boldsymbol{U}_k)^{S_0}$. Thus, given a homogeneous $f \in \mathcal{O}(\boldsymbol{U}_k)^{S_0}$ of some degree ℓ, we may view f as a $\boldsymbol{G}$-invariant regular function on $\boldsymbol{X} \times \boldsymbol{U}_k$. By Proposition 2.1, some power f^N lifts to a $\boldsymbol{G}$-invariant polynomial h_f on $\boldsymbol{V} \times \boldsymbol{U}$, homogeneous of degree $N\ell$ in the second variable (as $\mathcal{O}(\boldsymbol{X} \times \boldsymbol{U}_k)$ is a quotient of $\mathcal{O}(\boldsymbol{V} \times \boldsymbol{U})$ by a $\boldsymbol{G}$-invariant ideal). Then $h_f(v, \cdot)$ is a lift over $W(k)$ of $f^N(v^0, \cdot)$, which is an S-invariant element of $\mathcal{O}(\boldsymbol{U})$, homogeneous of degree $N\ell$.

Since by (ii) S_0 is finite, Noether's theorem allows us to pick a finite collection of homogeneous generators $f_1, \dots, f_m \in \mathcal{O}(\boldsymbol{U}_k)^{S_0}$, of some degrees $\ell_1, \dots, \ell_m$. Let h_{f_j} be a lift of $f_j^{N_j}$ as above, $j = 1, \dots, m$.

Lemma 2.6. *The algebra $\mathcal{O}(\boldsymbol{U})$ is module-finite over $W(k)[h_{f_1}, \dots, h_{f_m}]$.*

Proof. First, $\mathcal{O}(\boldsymbol{U}_k)^{S_0} = k[f_1, \dots, f_m]$ is module-finite over $k[f_1^{N_1}, \dots, f_m^{N_m}]$. Also, by Noether's theorem, $\mathcal{O}(\boldsymbol{U}_k)$ is module-finite over $\mathcal{O}(\boldsymbol{U}_k)^{S_0}$. Thus, $\mathcal{O}(\boldsymbol{U}_k)$ is module-finite over $k[f_1^{N_1}, \dots, f_m^{N_m}]$.

Let $w_1^0, \dots, w_r^0$ be homogeneous module generators of $\mathcal{O}(\boldsymbol{U}_k)$ over $k[f_1^{N_1}, \dots, f_m^{N_m}]$, of degrees $p_1, \dots, p_r$. Let w_j be any homogeneous lifts of w_j^0 over $W(k)$, $j = 1, \dots, r$. We claim that $w_1, \dots, w_r$ are module generators of $\mathcal{O}(\boldsymbol{U})$ over $R := W(k)[z_1, \dots, z_m]$, where z_j acts by multiplication by h_{f_j}. Indeed, set $\deg(z_j) = N_j \ell_j$. For each degree s, we have a natural map $\psi_s : R[s - p_1] \oplus \dots \oplus R[s - p_r] \to \mathcal{O}(\boldsymbol{U})[s]$, where $[s]$ denotes the degree s part; namely, $\psi_s(b_1, \dots, b_r) = \sum_{j=1}^r b_j w_j$. The map ψ_s is a morphism of free finite-rank $W(k)$-modules, and the reduction of ψ_s mod I is surjective, since $w_1^0, \dots, w_r^0$ generate $\mathcal{O}(\boldsymbol{U}_k)$ as a module over $k[f_1^{N_1}, \dots, f_m^{N_m}]$. Hence, ψ_s is surjective as well, which implies the lemma. $\square$

By Lemma 2.6, for any linear function $F \in \boldsymbol{U}^*(W(k))$ (where $\boldsymbol{U}^*$ is the dual representation to $\boldsymbol{U}$),

$$F^n + a_1 F^{n-1} + \cdots + a_n = 0, \tag{1}$$

where $a_j \in W(k)[h_{f_1}, \ldots, h_{f_m}]$.

Let $g \in S$. Acting by g on equation (1), and using that $ga_j = a_j$ (since by construction the h_{f_j} are S-invariant), we get that gF also satisfies (1). This means that if $u \in \boldsymbol{U}(W(k))$, then $gF(u) \in \overline{K}$ is integral over $W(k)$ (as $a_j(u) \in W(k)$). Since the integral closure of $W(k)$ in $\overline{K}$ is $\overline{W(k)}$, the ring of integers in $\overline{K}$, we get that $gF(u) \in \overline{W(k)}$, hence $gF \in \boldsymbol{U}^*(\overline{W(k)})$. Thus $gF \in \boldsymbol{U}^*(Q)$, where $Q \subset \overline{W(k)}$ is a finite extension of $W(k)$. So, the matrix elements of g in some free $W(k)$-basis of $\boldsymbol{U}$ belong to Q. Thus, g is a lift of some $g_0 \in S_0$ over Q. Using (iii), consider the element $\bar{g} := gi(g_0)^{-1} \in S \cap \boldsymbol{G}(Q) = \boldsymbol{S}(Q)$, and let $\bar{g}_N \in \boldsymbol{S}(Q/(p^N))$ be the reduction of $\bar{g}$ modulo p^N. By construction, the image of $\bar{g}$ in $\boldsymbol{G}(k)$ is 1. Since $\boldsymbol{S}_0$ is finite and reduced by (ii), this implies that $\bar{g}_N = 1$ for all N. Thus, $\bar{g} = 1$ and $g = i(g_0)$, as desired. □

Remark 2.7. Condition (i) in Proposition 2.5 is essential. Indeed, take $\boldsymbol{G} = \mathrm{GL}(2)$ acting on $\boldsymbol{V} = V_2 \oplus V_1$, where V_m is the natural representation of $\boldsymbol{G}$ on homogeneous polynomials of two variables x, y of degree m. Take $v = x_1(y_1 - px_1) + y_2 \in \boldsymbol{V}(W(k))$. Then $v^0 = x_1 y_1 + y_2$. If $gv^0 = v^0$ then g preserves y_2, hence y_1, so it preserves x_1. Thus $g = 1$, hence $\boldsymbol{S}_0 = 1$. On the other hand, $S = \{1, s\}$, where $s(y) = y$, $s(x) = (1/p)y - x$.

The reductivity of $\boldsymbol{G}$ is essential as well. To see this, take $\boldsymbol{G} = \mathrm{Aff}(1)$, the group of affine linear transformations (a, b) given by $x \mapsto ax + b$, $a \neq 0$, and take $\boldsymbol{V} = \boldsymbol{Q} \oplus \boldsymbol{U}$, where $\boldsymbol{Q}$ is the space of quadratic (not necessarily homogeneous) functions of x, and $\boldsymbol{U}$ is the 1-dimensional representation with basis vector z defined by $(a, b)(z) = a^2 z$. We can then take $v = x(1 - px) + z$, so that $v^0 = x + z$. Then, as before, $\boldsymbol{S}_0 = 1$, but $S = \{1, s\}$, where $s(x) = -x + 1/p$. Note that the $\boldsymbol{G}$-orbit of v^0 is closed in this case.

2D. ***Finiteness of the orbit map.*** The results of this subsection are not needed for the proof of the main results. They are only used in the proofs of Theorems 3.5 and 4.8 and are included mainly to justify Lemma 9.7 of [Etingof et al. 2005], whose original proof is incomplete.

We keep the setting of Proposition 2.5.

Lemma 2.8. (i) *For all $r \geq 1$, $\mathcal{O}(\boldsymbol{S})/(p^r)$ is a free $W(k)/(p^r)$-module of rank $|S_0|$.*

(ii) *$\boldsymbol{S}$ is the lift of $\boldsymbol{S}_0$ to $W(k)$, i.e., $\mathcal{O}(\boldsymbol{S})$ is a free $W(k)$-module of rank $|S_0|$. Thus, the tautological morphism $\pi : \boldsymbol{G} \to \boldsymbol{G}/\boldsymbol{S}$ is finite étale.*[3]

Proof. (i) Let $\mathrm{Fun}(X, R)$ stand for the set of functions from X to R. We have a natural homomorphism $\tau : \mathcal{O}(\boldsymbol{S}) \to \mathrm{Fun}(S_0, W(k))$, given by $\tau(f)(g_0) = f(i(g_0))$. Let $\tau_r : \mathcal{O}(\boldsymbol{S})/(p^r) \to \mathrm{Fun}(S_0, W(k)/(p^r))$ be the reduction of τ modulo p^r. Then τ_r is a homomorphism between finite $W(k)/(p^r)$-modules (since $\boldsymbol{S}_0$ is finite), and it is an isomorphism modulo p. Hence τ_r is surjective. Also, the length of $\mathcal{O}(\boldsymbol{S})/(p^r)$ is at most $r|S_0|$ (as it has $|S_0|$ generators), while the length of $\mathrm{Fun}(S_0, W(k))$ equals $r|S_0|$ since it is a free $W(k)/(p^r)$-module of rank $|S_0|$. This implies that τ_r is an isomorphism, giving (i).

[3]Note that the quotient $\boldsymbol{G}/\boldsymbol{S}$ makes sense as a scheme of finite type over $W(k)$.

To prove (ii), note that τ is surjective, as it is surjective modulo p. Let $J := \operatorname{Ker}\tau$. Tensoring with K, we get a short exact sequence of K-vector spaces

$$0 \to K \otimes_{W(k)} J \to \mathcal{O}(\boldsymbol{S}_K) \to \operatorname{Fun}(S_0, K) \to 0.$$

Since $\boldsymbol{S}_K$ is automatically reduced, and $\boldsymbol{S}(K) = S$ is isomorphic to S_0 by Proposition 2.5, we obtain by counting dimensions that $K \otimes_{W(k)} J = 0$, i.e., J is the torsion ideal in $\mathcal{O}(\boldsymbol{S})$. Since $\boldsymbol{S}$ is of finite type, J is finitely generated, hence killed by p^N for some $N \geq 1$. Hence, if $J \neq 0$, then there exists $f \in J$ such that $f \notin p\mathcal{O}(S)$. Since $p^N f = 0$, this contradicts (i) for $r > N$, yielding the first statement of (ii). The second statement (the étaleness of π) follows since $\boldsymbol{S}_0$ is reduced. □

Consider the morphism $\phi : \boldsymbol{G} \to \boldsymbol{V}$ given by $\phi(g) = gv$, which we call *the orbit map*. It induces a natural morphism $\nu : \boldsymbol{G}/\boldsymbol{S} \to \boldsymbol{V}$ such that $\phi = \nu \circ \pi$. It is easy to see that every scheme-theoretic fiber of ν is either a point or empty. Hence, by [EGA IV$_4$ 1967, Proposition 17.2.6], ν is a monomorphism.

For a $W(k)$-scheme $\boldsymbol{X}$ and a closed point $x \in \boldsymbol{X}$ over k or $\overline{K}$, denote by $\widehat{\boldsymbol{X}}_x$ the formal neighborhood of $\boldsymbol{X}$. Namely, if R is a local Artinian $W(k)$-algebra with residue field k, respectively $\overline{K}$, then $\widehat{\boldsymbol{X}}_x(R)$ is the set of homomorphisms $\mathcal{O}(\boldsymbol{X}) \to R$ which lift x.

Let $\boldsymbol{Y} \subset \boldsymbol{V}$ be a closed $\boldsymbol{G}$-invariant subscheme such that ν factors through a morphism $\mu : \boldsymbol{G}/\boldsymbol{S} \to \boldsymbol{Y}$.

Proposition 2.9. *Suppose that*

(i) *for any point g of $\boldsymbol{G}/\boldsymbol{S}$ over k or $\overline{K}$, the morphism of formal neighborhoods $\mu_g : \widehat{\boldsymbol{G}/\boldsymbol{S}}_g \to \widehat{\boldsymbol{Y}}_{\mu(g)}$ induced by μ is an isomorphism*; *and*

(ii) *$\boldsymbol{Y}$ consists of finitely many closed $\boldsymbol{G}$-orbits both over k and over $\overline{K}$.*

Then μ is a closed embedding. In particular, the morphisms μ, ν, ϕ are finite.

Proof. Let us cut down the target of μ. Pick a $\boldsymbol{G}$-invariant polynomial $b_0 \in \mathcal{O}(\boldsymbol{V}_k)$ such that $b_0(v^0) = 0$ and $b_0 = 1$ on all the other orbits of $\boldsymbol{G}$ in $\boldsymbol{Y}_k$. Since by (ii), $\boldsymbol{Y}_k$ consists of finitely many closed $\boldsymbol{G}$-orbits, such b_0 exists by Haboush's theorem [1975].

Now use Proposition 2.1 to lift some power b_0^N of b_0 to a $\boldsymbol{G}$-invariant polynomial $b \in \mathcal{O}(\boldsymbol{V})$ such that $b(v) = 0$ (namely, choose any lift b of b_0^N and then replace b with $b - b(v)$). Also, consider a polynomial $c' \in \mathcal{O}(\boldsymbol{V}_K)$ such that $c'(v) = 0$ but $c' \neq 0$ on all other orbits of $\boldsymbol{G}$ on $\boldsymbol{Y}_{\overline{K}}$. Since by (ii), $\boldsymbol{Y}_{\overline{K}}$ is a union of finitely many closed $\boldsymbol{G}$-orbits, c' exists, and by setting $c' = p^M c$ for sufficiently large $M \in \mathbb{Z}_+$, we obtain a polynomial $c \in \mathcal{O}(V)$.

Now consider the closed subscheme $\boldsymbol{Z} \subset \boldsymbol{Y}$ cut out by the equations $b = 0$, $c = 0$. Then the morphism μ factors through a monomorphism $\bar{\mu} : \boldsymbol{G}/\boldsymbol{S} \to \boldsymbol{Z}$, and it suffices to show that this morphism is a closed embedding. We will do so by showing that, in fact, $\bar{\mu}$ is an isomorphism.

Lemma 2.10. *The morphism $\bar{\mu}$ is surjective.*

Proof. This follows since by construction, $\boldsymbol{Z}$ has only one $\boldsymbol{G}$-orbit both over k and over $\overline{K}$ (namely, that of v^0, respectively v). □

Lemma 2.11. *The morphism $\bar{\mu}$ is étale.*

Proof. By (i), $\bar{\mu}$ is a formally étale morphism between affine $W(k)$-schemes of finite type. This implies the statement. □

Now the proposition follows from Lemmas 2.10 and 2.11, since a surjective étale monomorphism is an isomorphism, [EGA IV$_4$ 1967, Theorem 17.9.1]. □

Remark 2.12. The reductivity of $\boldsymbol{G}$ is essential in Proposition 2.9. Namely, let $p > 2$ and consider the action of $\boldsymbol{G} = \mathbb{G}_a$ by translations $x \mapsto x + b$ on the 3-dimensional space $\boldsymbol{V}$ of quadratic polynomials in x. Let $v = x - px^2$. Then $b \circ v = x + b - p(x+b)^2 = b - pb^2 + (1 - 2pb)x - px^2$. Thus, the map $\phi : \boldsymbol{G} \to \boldsymbol{V}$, $\phi(g) = gv$, is given by

$$\phi(b) = (b - pb^2, 1 - 2pb, -p).$$

We have

$$\mathcal{O}(\phi^{-1}(0, -1, -p)) = W(k)[b]/(b - pb^2, 2 - 2pb) = W(k)[1/p] = K$$

(as $pb = 1$ in this ring). This implies that ϕ is not finite (as K is not a finitely generated $W(k)$-module), even though it is finite over $\overline{K}$ and over $W(k)/I^N$ for each N.

We note that in this example the orbits of v over $\overline{K}$ and its reduction v^0 over k are closed, and the scheme-theoretic stabilizers $\boldsymbol{S}$ of v and $\boldsymbol{S}_0$ of v^0 are both trivial. Also, one may take $\boldsymbol{Y}$ to be the curve consisting of polynomials $-px^2 + \alpha x + \beta$ such that $4p\beta = 1 - \alpha^2$. This curve is $\boldsymbol{G}$-invariant, and the map $\mu : \boldsymbol{G} = \boldsymbol{G}/\boldsymbol{S} \to \boldsymbol{Y}$ is a monomorphism which induces an isomorphism on formal neighborhoods. However, $\boldsymbol{Y}_k$ has two orbits, $\alpha = 1$ and $\alpha = -1$, and μ lands in the first one. We have a $\boldsymbol{G}$-invariant polynomial b_0 on $\boldsymbol{V}_k$ separating these orbits, namely $b_0 = \frac{1}{2}(1 - \alpha)$. But power reductivity does not apply since $\boldsymbol{G}$ is not reductive, and no power of b_0 lifts to a $\boldsymbol{G}$-invariant in $\mathcal{O}(\boldsymbol{Y})$, since $\boldsymbol{Y}_{\overline{K}}$ has only one $\boldsymbol{G}$-orbit, so the only invariant regular functions on $\boldsymbol{Y}$ are constants. As a result, we cannot define a closed subscheme $\boldsymbol{Z} \subset \boldsymbol{Y}$ such that μ factors through a *surjective* morphism $\bar{\mu} : \boldsymbol{G}/\boldsymbol{S} \to \boldsymbol{Z}$.

3. Faithfulness of the lifting for Hopf algebras

3A. ***Faithfulness of the lifting.*** If H is a semisimple cosemisimple Hopf algebra over k, let $\widetilde{H}$ denote its lift over $W(k)$ constructed in [Etingof and Gelaki 1998, Theorem 2.1], and let $\widehat{H} := \overline{K} \otimes_{W(k)} \widetilde{H}$.

Let H_1, H_2 be semisimple cosemisimple Hopf algebras over k.

Theorem 3.1 [Etingof et al. 2005, Corollary 9.10]. *If $\widehat{H_1}$ is isomorphic to $\widehat{H_2}$ then H_1 is isomorphic to H_2.*

Proof. By the assumption, $\dim H_1 = \dim H_2 = d$, a number coprime to p. Indeed, by Theorem 3.1 of [Etingof and Gelaki 1998] in a semisimple cosemisimple Hopf algebra we have $\mathbb{S}^2 = 1$, where $\mathbb{S}$ is the antipode, and by the Larson–Radford theorem [1988], we have $\mathrm{Tr}(\mathbb{S}^2) \neq 0$.

Fix identifications $\widetilde{H}_1 \cong W(k)^d$, $\widetilde{H}_2 \cong W(k)^d$ as $W(k)$-modules; they, in particular, define identifications $H_1 \cong k^d$, $H_2 \cong k^d$ as vector spaces over k.

Let $\boldsymbol{V}$ be the space of all possible pre-Hopf structures, i.e., product, coproduct, unit, counit, and antipode maps on the d-dimensional space, without any axioms, regarded as an affine scheme. In other words, if $\boldsymbol{E}=\mathbb{A}^d$ is the defining representation of $\boldsymbol{G}$, then

$$\boldsymbol{V} \;=\; \boldsymbol{E}\otimes\boldsymbol{E}\otimes\boldsymbol{E}^* \;\oplus\; \boldsymbol{E}\otimes\boldsymbol{E}^*\otimes\boldsymbol{E}^* \;\oplus\; \boldsymbol{E} \;\oplus\; \boldsymbol{E}^* \;\oplus\; \boldsymbol{E}\otimes\boldsymbol{E}^*.$$

Then $\boldsymbol{V}$ is a rational representation of $\boldsymbol{G}=\mathrm{GL}(d)$, and $\widetilde{H}_1$, $\widetilde{H}_2$ together with the above identifications give rise to two vectors $v_1, v_2 \in \boldsymbol{V}(W(k))$, while H_1, H_2 correspond to their reductions mod I, $v_1^0, v_2^0 \in \boldsymbol{V}(k)$. Moreover, by the assumption of the theorem, v_1, v_2 are conjugate under the action of $\boldsymbol{G}(\bar{K})$. We are going to show that the reductions v_1^0, v_2^0 are conjugate under the action of $\boldsymbol{G}(k)$, i.e., $H_1\cong H_2$, as claimed.

Let $H\cong k^d$ be a semisimple cosemisimple Hopf algebra over k, and u be the corresponding vector in $\boldsymbol{V}(k)$.

Lemma 3.2 [Etingof et al. 2005, Section 9]. *The $\boldsymbol{G}$-orbit $\boldsymbol{G}u$ of u is closed.*

Proof. Let $u'\in\overline{\boldsymbol{G}u}$. Then u' corresponds to a d-dimensional Hopf algebra H' such that $\mathrm{Tr}\,|_{H'}(\mathbb{S}^2)=d\neq 0$, since this is so for all points of $\boldsymbol{G}u$. Hence H' is semisimple and cosemisimple by the Larson–Radford theorem. But then the stabilizers of u, u' in $\boldsymbol{G}(k)$, which are isomorphic to $\mathrm{Aut}(H)$, $\mathrm{Aut}(H')$, are finite; see [Etingof and Gelaki 1998, Corollary 1.3]. Hence, $\dim\boldsymbol{G}u'=\dim\boldsymbol{G}u=\dim\boldsymbol{G}$. This implies that $u'\in\boldsymbol{G}u$. Hence, $\boldsymbol{G}u$ is closed. □

Thus, the orbits of v_1^0, v_2^0 are closed. Hence, Theorem 3.1 follows from Proposition 2.3. □

3B. ***Integrality of the stabilizer.***

Theorem 3.3. *Let H_1, H_2 be semisimple cosemisimple Hopf algebras over k, and $g:\widehat{H}_1\to\widehat{H}_2$ be an isomorphism. Then g maps $\widetilde{H}_1$ isomorphically to $\widetilde{H}_2$, i.e., it is a lift of an isomorphism $g_0:H_1\to H_2$ over $W(k)$. In particular, for a semisimple cosemisimple Hopf algebra H over k, the lifting map $i:\mathrm{Aut}(H)\to\mathrm{Aut}(\widehat{H})$ defined in* [Etingof and Gelaki 1998, Theorem 2.2] *is an isomorphism.*

Proof. By Theorem 3.1, we may assume that $H_1=H_2=H$. Let $v\in\boldsymbol{V}(W(k))$ be the vector corresponding to $\widetilde{H}$, and $v^0\in\boldsymbol{V}(k)$ be its reduction mod I, corresponding to H. Let $S:=\mathrm{Aut}(\widehat{H})\subset G(\bar{K})$ be the stabilizer of v, $S_0:=\mathrm{Aut}(H)\subset\boldsymbol{G}(k)$ be the stabilizer of v^0, and $\boldsymbol{S}_0$ be the scheme-theoretic stabilizer of v^0.

By Lemma 3.2, the $\boldsymbol{G}$-orbit of v^0 is closed. Also, by [Etingof and Gelaki 1998, Theorem 1.2, Corollary 1.3], $\boldsymbol{S}_0$ is finite and reduced. Finally, by [Etingof and Gelaki 1998, Theorem 2.2], we have a lifting map $i:S_0\hookrightarrow S$ such that for any $g_0\in S_0$, $i(g_0)$ is integral, and the reduction modulo p of $i(g_0)$ equals g_0. Thus, Proposition 2.5 applies, and the result follows. □

Remark 3.4. Theorems 3.1, 3.3 also hold for quasitriangular and triangular Hopf algebras, with the same proofs.

3C. ***Finiteness of the orbit map.*** We would now like to prove an analog of Lemma 9.7 of [Etingof et al. 2005] in the Hopf algebra setting.

Theorem 3.5. *Let H be a semisimple cosemisimple Hopf algebra over k, and $v \in \boldsymbol{V}(W(k))$ be the vector corresponding to $\widetilde{H}$. Then the morphism $\phi : \boldsymbol{G} \to \boldsymbol{V}$ defined by $\phi(g) = gv$ is finite.*

Proof. Let $\boldsymbol{S}$ be the scheme-theoretic stabilizer of v. Theorem 3.3 and Lemma 2.8 imply that $\boldsymbol{S}$ is the lift of $\boldsymbol{S}_0$ over $W(k)$, and the tautological morphism $\pi : \boldsymbol{G} \to \boldsymbol{G}/\boldsymbol{S}$ is finite étale. We also have a natural morphism $\nu : \boldsymbol{G}/\boldsymbol{S} \to \boldsymbol{V}$ induced by ϕ, such that $\phi = \nu \circ \pi$.

Let $\boldsymbol{Y} \subset \boldsymbol{V}$ be the closed subscheme of Hopf algebra structures with $\mathrm{Tr}(\mathbb{S}^2) = d$. Then $\boldsymbol{Y}$ is a closed $\boldsymbol{G}$-invariant subscheme, and ν factors through a morphism $\mu : \boldsymbol{G}/\boldsymbol{S} \to \boldsymbol{Y}$. By the Larson–Radford theorem, $d \neq 0$ in k, so any Hopf algebra of dimension d over k or $\overline{K}$ is semisimple and cosemisimple. Hence, by Ştefan's theorem [1997] (restated in [Etingof and Gelaki 1998, Theorem 1.1]), $\boldsymbol{Y}$ consists of finitely many orbits both over k and over $\overline{K}$, which are closed by Lemma 3.2. Also, it follows from [Etingof and Gelaki 1998, Theorem 2.2], that μ induces an isomorphism on formal neighborhoods. Thus, Proposition 2.9 applies, and the statement follows. □

Remark 3.6. Theorems 3.1, 3.3, 3.5 are subsumed by the results of Section 4. However, we felt it was useful to give independent direct proofs of these theorems which do not use tensor categories.

3D. ***Fullness of the lifting functor.*** Finally, let us prove the following result, which appears to be new.

Theorem 3.7. *Let H_1, H_2 be semisimple cosemisimple Hopf algebras over k. Then any Hopf algebra homomorphism $\theta : \widehat{H}_1 \to \widehat{H}_2$ is a lifting of some homomorphism $\theta_0 : H_1 \to H_2$. In other words, the lifting functor defined by* [Etingof and Gelaki 1998, Corollary 2.4], *is a fully faithful embedding from the category of semisimple cosemisimple Hopf algebras over k to the category of semisimple* (*thus*, *cosemisimple*) *Hopf algebras over $\overline{K}$.*

Proof. We first prove the following lemma.

Lemma 3.8. *Let H be a semisimple cosemisimple Hopf algebra over k. Then lifting defines a bijection between Hopf ideals of H and Hopf ideals of $\widehat{H}$. The same applies to Hopf subalgebras.*

Proof. A Hopf ideal $J \subset A$ of a semisimple Hopf algebra A corresponds to a full tensor subcategory $\mathcal{C}_J \subset \mathrm{Rep}\, H$ of objects annihilated by J, and this correspondence is a bijection. Full tensor subcategories, in turn, correspond to fusion subrings of the Grothendieck ring of $\mathrm{Rep}\, A$. So the first statement follows from the fact that the Grothendieck rings of H and $\widehat{H}$ are the same. The second statement is dual to the first statement, since the orthogonal complement of a Hopf ideal in A is a Hopf subalgebra of A^*, and vice versa. □

Now let $B = \mathrm{Im}\,\theta \subset \widehat{H}_2$. Then by Lemma 3.8, B is a lifting of some Hopf subalgebra $B_0 \subset H_2$. Note that B_0 is semisimple, since its dimension divides the dimension of H_2 by the Nichols–Zoeller theorem [1989], hence is coprime to p. Thus, without loss of generality we may replace H_2 with B_0, i.e., assume that θ is surjective.

Now let $J = \mathrm{Ker}\,\theta \subset \widehat{H}_1$. Then by Lemma 3.8, J is a lift of some Hopf ideal $J_0 \subset H_1$. Let $C_0 = H_1/J_0$. Then C_0 is cosemisimple, as it is a quotient of H_1, so by the Nichols–Zoeller theorem, its dimension is coprime to p. Moreover, $\widehat{C} \cong \widehat{H}_1/J$. Thus, without loss of generality we may replace H_1 with C_0, i.e., assume that θ is an isomorphism.

But in this case the desired statement is Theorem 3.3. □

4. Faithfulness of the lifting for fusion categories

4A. ***Faithfulness of the lifting.*** Now we generalize the results of the previous section to separable fusion categories. (For basics on tensor categories we refer the reader to [Etingof et al. 2015].) Most of the proofs are parallel to the Hopf algebra case, and we will indicate the necessary modifications. We will develop the theory for ordinary fusion categories; the case of braided and symmetric categories is completely parallel.

We call a fusion category $\mathcal{C}$ *separable* if its global dimension is nonzero. This is equivalent to the definition of [Douglas et al. 2013] by Theorem 3.6.7 in that paper.

If $\mathcal{C}$ is a (braided, symmetric) separable fusion category over k, let $\widetilde{\mathcal{C}}$ be its lift to $W(k)$ constructed in [Etingof et al. 2005, Theorem 9.3, Corollary 9.4], and let $\widehat{\mathcal{C}} := \overline{K} \otimes_{W(k)} \widetilde{\mathcal{C}}$.

Let $\mathcal{C}_1, \mathcal{C}_2$ be (braided, symmetric) separable fusion categories over k. First we prove the following theorem, which is Corollary 9.9(i) of [Etingof et al. 2005].

Theorem 4.1. *If $\widehat{\mathcal{C}_1}$ is equivalent to $\widehat{\mathcal{C}_2}$ then $\mathcal{C}_1$ is equivalent to $\mathcal{C}_2$.*

Proof. We will treat the fusion case; the braided and symmetric cases are similar.

We generalize the proof of Theorem 3.1. As in [Etingof et al. 2005, Section 9.3], we may assume that $\mathcal{C}_1$ and $\mathcal{C}_2$ have the same underlying semisimple abelian category $\bar{\mathcal{C}}$ with the tensor product functor $\otimes$, a skeletal category with Grothendieck ring $\mathrm{Gr}(\bar{\mathcal{C}})$. So it has simple objects X_i, $i \in \mathbb{I}$, with $X_0 = \mathbf{1}$, and $X_i \otimes X_j = \bigoplus_m k^{N_{ij}^m} X_m$. We will also fix the unit morphism $\iota : \mathbf{1} \otimes \mathbf{1} \to \mathbf{1}$, and the coevaluation morphisms. Define a pretensor structure on $\bar{\mathcal{C}}$ to be a triple $(\Phi, \Phi', \mathrm{ev})$, where Φ is an associativity morphism, Φ' an "inverse" associativity morphism in the opposite direction, and ev is a collection of evaluation morphisms, but without any axioms. Then a tensor structure on $\bar{\mathcal{C}}$ is a pretensor structure such that $\Phi \circ \Phi' = \Phi' \circ \Phi = \mathrm{Id}$ and (Φ, ev) satisfy the axioms of a rigid tensor category (with the fixed unit and coevaluation morphisms); see [Etingof et al. 2015, Definitions 2.1.1 and 2.10.11].

Let

$$N_{ijl}^s := [X_i \otimes X_j \otimes X_l : X_s] = \sum_m N_{ij}^m N_{ml}^s = \sum_p N_{ip}^s N_{jl}^p.$$

Let $\boldsymbol{V}$ be the space of all pretensor structures on $\bar{\mathcal{C}}$, which is an affine space over $W(k)$ of dimension $2\sum_{i,j,l,s}(N_{ijl}^s)^2 + \operatorname{rank}\mathrm{Gr}(\bar{\mathcal{C}})$ (where the first summand corresponds to pairs (Φ, Φ') and the second summand to ev). Then the $\widehat{\mathcal{C}_i}$ give rise to vectors $v_i \in \boldsymbol{V}(W(k))$, and the $\mathcal{C}_i$ correspond to their reductions v_i^0 modulo I.

Now let us define the relevant affine group scheme $\boldsymbol{G}$. To this end, following [Etingof et al. 2005, Section 9.3], let $\boldsymbol{T} = \mathrm{Aut}(\otimes)$ be the group of automorphisms of the tensor product functor on $\bar{\mathcal{C}}$. Then $\boldsymbol{T}$ acts naturally on $\boldsymbol{V}$ by twisting (where after twisting we renormalize the coevaluation and unit morphisms to be the fixed ones). Let $\boldsymbol{T}_0 \subset \boldsymbol{T}$ be the subgroup of "trivial twists", i.e., ones of the form $J_{X,Y} = z_{X\otimes Y}(z_X^{-1} \otimes z_Y^{-1})$, where $z \in \mathrm{Aut}(\mathrm{Id}_{\bar{\mathcal{C}}})$ is an invertible element of the center of $\bar{\mathcal{C}}$. Then $\boldsymbol{T}_0$ is a closed central subgroup of $\boldsymbol{T}$ which acts trivially on $\boldsymbol{V}$. Let $\boldsymbol{G} := \boldsymbol{T}/\boldsymbol{T}_0$. Then $\boldsymbol{G}$ is an affine group scheme acting rationally on $\boldsymbol{V}$. Moreover, $\boldsymbol{T} = \prod_{i,j,m} \mathrm{GL}(N_{ij}^m)$, and $\boldsymbol{T}_0$ is a central torus in $\boldsymbol{T}$, hence $\boldsymbol{G}$ is a split reductive group.

Let $\mathcal{C}$ be a separable fusion category over k of some global dimension $d \neq 0$, and u be the corresponding vector in $\boldsymbol{V}(k)$. Let $\mathrm{Aut}(\mathcal{C})$ denote the group of isomorphism classes of tensor autoequivalences of $\mathcal{C}$.

Lemma 4.2 [Etingof et al. 2005, Section 9]. *The $\boldsymbol{G}$-orbit $\boldsymbol{G}u$ of u is closed.*

Proof. Let $u' \in \overline{\boldsymbol{G}u}$. Then u' corresponds to a fusion category $\mathcal{C}'$. Moreover the global dimension of $\mathcal{C}'$ is d, since this is so for all points of $\boldsymbol{G}u$, and the global dimension depends algebraically on Φ, Φ', ev. Thus, $\mathcal{C}'$ is separable. But then the stabilizers of u, u' in $\boldsymbol{G}(k)$, which are isomorphic to $\mathrm{Aut}(\mathcal{C})$, $\mathrm{Aut}(\mathcal{C}')$ are finite by [Etingof et al. 2005, Theorem 2.31] (which applies in characteristic p for separable categories; see [Etingof et al. 2005, Section 9]). Hence, $\dim \boldsymbol{G}u' = \dim \boldsymbol{G}u = \dim \boldsymbol{G}$. This implies that $u' \in \boldsymbol{G}u$. Hence, $\boldsymbol{G}u$ is closed. □

Thus, the orbits of v_1^0, v_2^0 are closed. Hence, Theorem 4.1 follows from Proposition 2.3 and the fact that the natural map $\boldsymbol{T}(W(k)) \to (\boldsymbol{T}/\boldsymbol{T}_0)(W(k))$ is surjective. □

Remark 4.3. Note that Theorem 4.1 implies [Etingof et al. 2005, Theorem 9.6]; namely, the complete local ring R in Theorem 9.6 without loss of generality may be replaced by $W(k)$.

Remark 4.4. We can now complete the proof of [Etingof and Gelaki 2000, Theorem 6.1], which states, essentially, that any semisimple cosemisimple triangular Hopf algebra over k is a twist of a group algebra. The original proof of this theorem appearing in that work is incomplete (namely, it is not clear at the end of this proof why $F \circ F' = \mathrm{Id}$). This is really a consequence of faithfulness of the lifting. Namely, if (A, R) is a semisimple cosemisimple triangular Hopf algebra over k, and $(A', R') = F \circ F'(A, R)$, then (A, R) and (A', R') have isomorphic liftings over $\bar{K}$; hence by Theorem 3.1 they are isomorphic.

Another proof of [Etingof and Gelaki 2000, Theorem 6.1] is obtained by using Theorem 4.1 for symmetric tensor categories. Namely, consider the separable symmetric fusion category $\mathcal{C} := \mathrm{Rep}(A, R)$. Then $\widehat{\mathcal{C}}$ is a symmetric fusion category over $\bar{K}$. Hence, by Deligne's theorem [1990] (see also [Etingof et al. 2015, Corollary 9.9.25]), $\widehat{\mathcal{C}} = \mathrm{Rep}_{\bar{K}}(G, z)$, where G is a finite group of order coprime to p and $z \in G$ is a central element of order ≤ 2 (the category of representations of G on superspaces with parity defined by z). Thus, $\widehat{\mathcal{C}} \cong \widehat{\mathcal{D}}$ as symmetric tensor categories, where $\mathcal{D} = \mathrm{Rep}_k(G, z)$. Hence, by Theorem 4.1, $\mathcal{C} \cong \mathcal{D}$, i.e., (A, R) is obtained by twisting of $(k[G], R_z)$, where $R_z = 1 \otimes 1$ if $z = 1$ and $R_z = \frac{1}{2}(1\otimes 1 + 1\otimes z + z\otimes 1 - z\otimes z)$ if $z \neq 1$ (note that if $z \neq 1$ then $|G|$ is necessarily even, so $p > 2$ and $\frac{1}{2} \in k$).

Remark 4.5. Let G be a finite group. Recall from [Etingof et al. 2015] that categorifications of the group ring $\mathbb{Z}G$ over a field F correspond to elements of $H^3(G, F^\times)$. Hence, the lifting map for pointed fusion categories which categorify $\mathbb{Z}G$ is the natural map $\alpha : H^3(G, k^\times) \to H^3(G, \bar{K}^\times)$ arising from the isomorphisms $H^i(G, \bar{K}^\times) \cong H^i(G, \bar{K}_f^\times)$ and $H^i(G, k^\times) \cong H^i(G, k_f^\times)$, where $\bar{K}_f^\times$ is the group of elements of finite order (i.e., roots of unity) in $\overline{K}^{\times}$, and $k_f^{\times}$ is the group of roots of unity in $k^{\times}$. The map α is then induced by the Brauer lift map $k_f^{\times} \to \overline{K}_f^{\times}$, and is clearly injective since the Brauer lift identifies $k_f^{\times}$ with a direct summand of $\overline{K}_f^{\times}$. Thus, for pointed categories faithfulness of the lifting is elementary.

Similarly, braided categorifications $\mathbb{Z}G$ over F (for abelian G) correspond to quadratic forms $G \to F^{\times}$ (see [Etingof et al. 2015]), and lifting for such categorifications is defined by the Brauer lift for quadratic forms, hence is clearly faithful.

4B. *Integrality of the stabilizer.*

Theorem 4.6. *Let $\mathcal{C}_1, \mathcal{C}_2$ be (braided, symmetric) separable fusion categories over k, and $g : \widehat{\mathcal{C}}_1 \to \widehat{\mathcal{C}}_2$ be an equivalence. Then g defines an equivalence $\widetilde{\mathcal{C}}_1 \to \widetilde{\mathcal{C}}_2$, i.e., it is isomorphic to the lift of an equivalence $g_0 : \mathcal{C}_1 \to \mathcal{C}_2$ over $W(k)$.*

Proof. As before, we treat only the fusion case; the braided and symmetric cases are similar. By Theorem 4.1, we may assume that $\mathcal{C}_1 = \mathcal{C}_2 = \mathcal{C}$. Let $v \in \boldsymbol{V}(W(k))$ be the vector corresponding to $\widetilde{\mathcal{C}}$, and $v^0 \in \boldsymbol{V}(k)$ be its reduction mod I, corresponding to $\mathcal{C}$. Let $S := \mathrm{Aut}(\widehat{\mathcal{C}})$ be the stabilizer of v, $S_0 := \mathrm{Aut}(\mathcal{C})$ be the stabilizer of v^0, and $\boldsymbol{S}_0$ the scheme-theoretic stabilizer of v^0.

By Lemma 4.2, the $\boldsymbol{G}$-orbit of v^0 is closed. By Theorem 2.27 and Theorem 2.31 of [Etingof et al. 2005] (both valid in characteristic p for separable fusion categories, see Section 9 of that paper), $\boldsymbol{S}_0$ is finite and reduced. Finally, by Theorems 9.3 and 9.4 there, we have a lifting map $i : S_0 \hookrightarrow S$, such that for all $g_0 \in \boldsymbol{G}(k)$, $i(g_0)$ is integral and the reduction of $i(g_0)$ modulo p equals g_0. Thus, Proposition 2.5 applies, and the result follows (again using that the natural map $\boldsymbol{T}(W(k)) \to (\boldsymbol{T}/\boldsymbol{T}_0)(W(k))$ is surjective). □

Remark 4.7. (1) Recall that a multifusion category is called separable if all of its component fusion categories are separable; see [Etingof et al. 2005, Section 9] and [Douglas et al. 2013]. Theorems 4.9 and 4.11 extend to separable multifusion categories with similar proofs.

(2) Theorem 4.6 is not stated explicitly in [Etingof et al. 2005], but is claimed implicitly in the (incomplete) proof of Theorem 9.6 there.

4C. *Finiteness of the orbit map.* Let us now prove Lemma 9.7 of [Etingof et al. 2005] (which completes the proofs in Section 9.3 of that paper).

Theorem 4.8. *Let $\mathcal{C}$ be a (braided, symmetric) separable fusion category over k, and $v \in \boldsymbol{V}(W(k))$ be the vector corresponding to $\widetilde{\mathcal{C}}$. Then the morphism $\phi : \boldsymbol{G} \to \boldsymbol{V}$ defined by $\phi(g) = gv$ is finite.*

Proof. We treat the case of fusion categories; the braided and symmetric cases are similar. The proof is parallel to the proof of Theorem 3.5. Namely, let $\boldsymbol{S}$ be the scheme-theoretic stabilizer of v. Then $\phi = \nu \circ \pi$, where $\pi : \boldsymbol{G} \to \boldsymbol{G}/\boldsymbol{S}$ is finite étale, and $\nu : \boldsymbol{G}/\boldsymbol{S} \to \boldsymbol{V}$. Let $\boldsymbol{Y} \subset \boldsymbol{V}$ denote the closed subscheme of vectors corresponding to fusion categories of global dimension $\tilde{d} := \dim(\widetilde{\mathcal{C}})$. Then ν factors through $\mu : \boldsymbol{G}/\boldsymbol{S} \to \boldsymbol{Y}$. By [Etingof et al. 2005, Theorem 2.27], $\boldsymbol{Y}$ consists of finitely many $\boldsymbol{G}$-orbits both over k and over $\overline{K}$. Also, these orbits are closed by Lemma 4.2. Finally, by Theorem 9.3 and Corollary 9.4 of the same paper, μ induces an isomorphism on formal neighborhoods. Thus, Proposition 2.9 applies, and the statement follows. □

4D. ***Faithfulness of the lifting and integrality of the stabilizer for tensor functors.*** Let us now prove similar results for tensor functors, i.e., Theorem 9.8 and Corollary 9.9(ii) of [Etingof et al. 2005].

Theorem 4.9. *Let $\mathcal{C}$, $\mathcal{D}$ be two (braided, symmetric) separable fusion categories over k, and $F_1, F_2 : \mathcal{C} \to \mathcal{D}$ two (braided) tensor functors. Let $g : \widehat{F_1} \to \widehat{F_2}$ be an isomorphism of lifts of F_1, F_2 over $\bar{K}$. Then g is a lift of an isomorphism $g_0 : F_1 \to F_2$. In particular, if $\widehat{F_1}$, $\widehat{F_2}$ are isomorphic then so are F_1, F_2.*

Proof. The proof is similar to the proofs of Theorems 3.1, 3.3 and Theorems 4.1, 4.6. We treat the case of tensor functors between fusion categories; the cases of braided and symmetric categories are similar.

We may assume that F_1, F_2 coincide with a given functor $\bar{F}$ as additive functors, and differ only by the tensor structures. Let $\boldsymbol{V}$ be the space of pretensor structures on $\bar{F}$, i.e., pairs (J, J') of endomorphisms of the functor $\bar{F}(\cdot \otimes \cdot)$, without any axioms. Then a tensor structure on $\bar{F}$ is such a pair (J, J') for which $J \circ J' = J' \circ J = \mathrm{Id}$, and J satisfies the tensor structure axiom, [Etingof et al. 2015, Definition 2.4.1]. Let $\boldsymbol{G}$ be the group scheme of all automorphisms of the functor $\bar{F}$. Then $\boldsymbol{G}$ is a split reductive group (a product of general linear groups) which acts on $\boldsymbol{V}$ by "gauge transformations". Moreover, the functors $\widetilde{F}_j$, $j = 1, 2$, correspond to vectors $v_j \in \boldsymbol{V}(W(k))$, and the F_j correspond to their reductions v_j^0 modulo p.

Lemma 4.10. *Let $u \in \boldsymbol{V}(k)$ be a vector representing a tensor functor F. Then the orbit $\boldsymbol{G}u$ is closed.*

Proof. Let $u' \in \overline{\boldsymbol{G}u}$. Then u' corresponds to a tensor functor F'. But the group of automorphisms of a tensor functor between separable fusion categories is finite by [Etingof et al. 2005, Theorem 2.27]. Hence the stabilizers of u, u' in $\boldsymbol{G}(k)$, which are isomorphic to $\mathrm{Aut}(F)$, $\mathrm{Aut}(F')$, are finite. Hence, $\dim \boldsymbol{G}u' = \dim \boldsymbol{G}u = \dim \boldsymbol{G}$. This implies that $u' \in \boldsymbol{G}u$. Hence, $\boldsymbol{G}u$ is closed. □

By Lemma 4.10, Proposition 2.3 applies. Thus, $F_1 \cong F_2$ as tensor functors. So we may assume without loss of generality that $F_1 = F_2 = F$ for some tensor functor F.

Let $v \in \boldsymbol{V}(W(k))$ and $v^0 \in \boldsymbol{V}(k)$ be its reduction modulo p. Let $S_0 = \mathrm{Aut}(F) \subset \boldsymbol{G}(k)$, $\boldsymbol{S}_0$ be the scheme-theoretic stabilizer of v^0, and $S = \mathrm{Aut}(\widehat{F}) \subset \boldsymbol{G}(\bar{K})$. Then $\boldsymbol{S}_0$ is finite and reduced by [Etingof et al. 2005, Theorem 2.27]. Also, by Theorem 9.3 and Corollary 9.4 of the same work, we have a lifting map $i : S_0 \hookrightarrow S$ such that for any $g_0 \in S_0$, $i(g_0)$ is integral and the reduction of $i(g_0)$ modulo p is g_0. Hence, Proposition 2.5 applies, and the result follows. □

Corollary 4.11. *For a (braided, symmetric) separable fusion category $\mathcal{C}$ over k, the lifting map i : $\mathrm{Aut}(\mathcal{C}) \to \mathrm{Aut}(\widehat{\mathcal{C}})$ defined in* [Etingof et al. 2005, Theorem 9.3], *is an isomorphism.*

Proof. Theorem 4.9 implies that i is injective, and Theorem 4.6 implies that i is surjective. □

4E. ***Application to Brauer–Picard and Picard groups of tensor categories.*** Recall from [Etingof et al. 2010] that to any fusion category $\mathcal{C}$ one can attach its Brauer–Picard groupoid $\underline{\underline{\mathrm{BrPic}}}(\mathcal{C})$. This is a 3-group, whose 1-morphisms are equivalence classes of invertible $\mathcal{C}$-bimodule categories, 2-morphisms are bimodule equivalences of such bimodule categories, and 3-morphisms are isomorphisms of such equivalences. Similarly, if $\mathcal{C}$ is braided then one can define its Picard groupoid $\underline{\underline{\mathrm{Pic}}}(\mathcal{C})$, a 3-group whose 1-morphisms are equivalence classes of invertible $\mathcal{C}$-module categories, 2-morphisms are module equivalences of such module categories, and 3-morphisms are isomorphisms of such equivalences.

Theorems 9.3 and 9.4 of [Etingof et al. 2005] then imply that if $\mathcal{C}$ is separable then we have the lifting morphism $i : \underline{\mathrm{BrPic}}(\mathcal{C}) \to \underline{\mathrm{BrPic}}(\widehat{\mathcal{C}})$ and (in the braided case) $i : \underline{\mathrm{Pic}}(\mathcal{C}) \to \underline{\mathrm{Pic}}(\widehat{\mathcal{C}})$ of the underlying 2-groups.

Corollary 4.12. *Let $\mathcal{C}$ be a separable fusion category over k.*

(i) *The lifting morphism $i : \underline{\mathrm{BrPic}}(\mathcal{C}) \to \underline{\mathrm{BrPic}}(\widehat{\mathcal{C}})$ is an isomorphism.*

(ii) *If $\mathcal{C}$ is braided then the lifting morphism $i : \underline{\mathrm{Pic}}(\mathcal{C}) \to \underline{\mathrm{Pic}}(\widehat{\mathcal{C}})$ is an isomorphism.*

Proof. (i) By [Etingof et al. 2010, Theorem 1.1], for any separable fusion category $\mathcal{D}$ one has an isomorphism $\xi : \mathrm{BrPic}(\mathcal{D}) \cong \mathrm{Aut}^{\mathrm{br}}(\mathcal{Z}(\mathcal{D}))$ of the Brauer–Picard group $\mathrm{BrPic}(\mathcal{D})$ with the group of isomorphism classes of braided autoequivalences of the Drinfeld center of $\mathcal{D}$. It is clear that this isomorphism is compatible with lifting. Therefore, the statement at the level of 1-morphisms follows from Corollary 4.11. Also, recall that $\pi_2(\underline{\mathrm{BrPic}}(\mathcal{D})) = \mathrm{Inv}(\mathcal{Z}(\mathcal{D}))$, the group of isomorphism classes of invertible objects of $\mathcal{Z}(\mathcal{D})$. Thus, at the level of 2-morphisms i comes from the obvious isomorphism $\mathrm{Inv}(\mathcal{Z}(\mathcal{C})) \cong \mathrm{Inv}(\mathcal{Z}(\widehat{\mathcal{C}}))$, which gives (i).

(ii) By [Davydov and Nikshych 2013], if $\mathcal{D}$ is braided then $\mathrm{Pic}(\mathcal{D})$ is naturally identified with the subgroup of $\mathrm{Aut}^{\mathrm{br}}(\mathcal{Z}(\mathcal{D}))$ of elements that preserve $\mathcal{D} \subset \mathcal{Z}(\mathcal{D})$ and have trivial restriction to $\mathcal{D}$. Thus, (ii) follows from (i) and Theorem 4.9. Also, recall that $\pi_2(\underline{\mathrm{Pic}}(\mathcal{D})) = \mathrm{Inv}(\mathcal{D})$, the group of isomorphism classes of invertible objects of $\mathcal{D}$. Thus, at the level of 2-morphisms i comes from the obvious isomorphism $\mathrm{Inv}(\mathcal{C}) \cong \mathrm{Inv}(\widehat{\mathcal{C}})$, which gives (ii). □

Note that in Corollary 4.12, i does not define an isomorphism of 3-groups, since π_3 of these 3-groups is the multiplicative group of the ground field, and $k^\times \ncong \overline{K}^\times$. However, we can lift i to an injection at the level of 3-morphisms. For simplicity assume that $k = \overline{\mathbb{F}}_p$ (this is not restrictive since by [Etingof et al. 2005, Theorem 2.31], any separable fusion category in characteristic p is defined over $\overline{\mathbb{F}}_p$). Then by Hensel's lemma, the surjection $W(k)^\times \to k^\times$ defined by reduction modulo p uniquely splits, since all elements of $k^\times$ are roots of unity of order coprime to p (the Brauer lift, see Remark 4.5). Then i extends to a morphism of 3-groups using the corresponding splitting $\beta : k^\times \to W(k)^\times \subset \overline{K}^\times$. In particular, using the main results of [Etingof et al. 2010], this implies the following result.

Theorem 4.13. *Let G be a finite group and $k = \overline{\mathbb{F}}_p$. Then any G-extension of $\mathcal{C}$ canonically lifts to a G-extension of $\widehat{\mathcal{C}}$, and any braided G-crossed category $\mathcal{D}$ with $\mathcal{D}_1 = \mathcal{C}$ canonically lifts to a braided G-crossed category $\widehat{\mathcal{D}}$ with $\widehat{\mathcal{D}}_1 = \widehat{\mathcal{C}}$.*

Remark 4.14. One can propose the following definition (which we are not making completely precise here). We recall (and refer, e.g., to the textbook [Yau and Johnson 2015] for details and some history) that a *linear algebraic structure* is defined by a colored PROP $\boldsymbol{P}$ (say, over $\mathbb{Z}$). Realizations of $\boldsymbol{P}$ over a commutative ring R are then $\boldsymbol{P}$-algebras over R. We call an algebraic structure $\boldsymbol{P}$ *3-separable* if every finite-dimensional realization A of $\boldsymbol{P}$ over a field k

(i) has a finite and reduced group of automorphisms (i.e., has no nontrivial derivations);

(ii) has no nontrivial first order deformations;

(iii) has a vanishing space of obstructions to deformations.

These conditions should be expressed as requiring that $H^i(A) = 0$ for $i = 1, 2, 3$ for an appropriate cohomology theory, controlling deformations of A. Then A should admit a unique lifting from a field k of characteristic p to $W(k)$, and this lifting should have the faithfulness and stabilizer integrality properties similar to Theorems 3.1, 3.3, 4.1, 4.6, 4.9: any isomorphism g of liftings of A_1, A_2 over $\bar{K}$ is defined over $W(k)$ and hence is a lifting of an isomorphism $g_0 : A_1 \to A_2$. We have seen a number of examples of 3-separable structures: semisimple cosemisimple (quasitriangular, triangular) Hopf algebras, separable (braided, symmetric) fusion categories, (braided) tensor functors between such categories.[4] It would therefore be interesting to make this notion more precise, and prove a general theorem on the existence and faithfulness of the lifting for 3-separable structures, which would unify the results of [Etingof et al. 2005, Section 9], [Etingof and Gelaki 1998], and this paper. It would also be interesting to find other examples of 3-separable structures.

5. Descent of tensor functors between separable fusion categories to characteristic p.

5A. ***Separability of subcategories and quotient categories.*** We will first prove separability of subcategories and quotients of separable categories. First we need the following result, which is a generalization of [Etingof and Ostrik 2004, Theorem 2.5].

Theorem 5.1. *Let $\mathcal{C}$ be a finite tensor category and $\mathcal{D}$ a finite indecomposable multitensor category. Let $F : \mathcal{C} \to \mathcal{D}$ be a quasitensor functor. If F is surjective* (*i.e.*, *every object of $\mathcal{D}$ is a subquotient of $F(X)$, $X \in \mathcal{C}$*) *then*

(i) *F maps projective objects to projective ones*; *and*

(ii) *$\mathcal{D}$ is an exact module category over $\mathcal{C}$.*

Proof. (i) The proof is almost identical to the proof of [Etingof and Gelaki 2017, Theorem 2.9]. We reproduce it here for the convenience of the reader.

Let P_i be the indecomposable projectives of $\mathcal{C}$. Write $F(P_i) = T_i \oplus N_i$, where T_i is projective, and N_i has no projective direct summands. Our job is to show that $N_i = 0$ for all i. So let us assume for the sake of contradiction that $N_p \neq 0$ for some p.

Let $P_i \otimes P_j = \oplus_r c_{ij}^r P_r$. Since the tensor product of a projective object with any object in $\mathcal{D}$ is projective, the objects $T_i \otimes T_j$, $T_i \otimes N_j$, and $N_i \otimes T_j$ are projective. Thus,

$$\Big(\bigoplus_i N_i\Big) \otimes N_j \supset \bigoplus_r \Big(\sum_i c_{ij}^r\Big) N_r \tag{2}$$

as a direct summand.

[4] Semisimplicity/cosemisimplicity for Hopf algebras and separability for fusion categories may be forced in the setting of linear algebraic structures by adding an auxiliary variable x and the relation $dx = 1$, where d is the global dimension.

Let $\mathbf{1} = \sum_l \mathbf{1}_l$ be the irreducible decomposition of the unit object of $\mathcal{D}$, and let s be such that $Y := \sum_r d_r N_r \mathbf{1}_s$ is nonzero (it exists since $N_p \neq 0$). Denote by X_j the simples of $\mathcal{C}$, and let d_j be their Frobenius–Perron dimensions. For any i, r, we have $\sum_j d_j c_{ij}^r = D_i d_r$, where $D_i := \mathrm{FPdim}(P_i)$. Thus, tensoring inclusion (2) on the right by $\mathbf{1}_s$, multiplying by d_j and summing over j, we get a coefficientwise inequality

$$\Big[\bigoplus_i N_i\Big][Y] \geq \Big(\sum_i D_i\Big)[Y]$$

in $\mathrm{Gr}(\mathcal{D}_s)$, where $\mathcal{D}_s := \mathcal{D} \otimes \mathbf{1}_s$. This implies that the largest eigenvalue of the matrix of $\big[\bigoplus_i N_i\big]$ on $\mathrm{Gr}(\mathcal{D}_s)$ is at least $\sum_i D_i$, which is the same as the largest eigenvalue of $\big[\bigoplus_i F(P_i)\big]$. Since F is surjective, all the entries of $\big[\bigoplus_i F(P_i)\big]$ are positive. Thus, by the Frobenius–Perron theorem (see Lemma 2.1 of [Etingof and Gelaki 2017]), $\big[\bigoplus_i N_i\big] = \big[\bigoplus_i F(P_i)\big]$. This implies that $N_i = F(P_i)$ for all i. Thus, $F(P_i)$ has no nonzero projective direct summands for all i.

However, let Q be an indecomposable projective object in $\mathcal{D}$. Then Q is injective by the quasi-Frobenius property of finite multitensor categories. Since F is surjective, Q is a subquotient, hence a direct summand of $F(P)$ for some projective $P \in \mathcal{C}$. Hence Q is a direct summand of $F(P_i)$ for some i, which gives the desired contradiction.

(ii) This follows from (i) and the fact that in a multitensor category, the tensor product of a projective object with any object is projective. □

Theorem 5.2. *Let $\mathcal{C}$, $\mathcal{D}$ be multitensor categories and let $F : \mathcal{C} \to \mathcal{D}$ be a surjective tensor functor. If $\mathcal{C}$ is separable, then so is $\mathcal{D}$. In other words, a quotient category of a separable multifusion category is separable.*

Proof. Consider first the special case when $\mathcal{C}$ is a tensor category (i.e., $\mathbf{1} \in \mathcal{C}$ is simple). Without loss of generality, we may assume that $\mathcal{D}$ is indecomposable. By Theorem 5.1, $\mathcal{D}$ is an exact $\mathcal{C}$-module category, hence semisimple (as $\mathcal{C}$ is semisimple). Moreover, we have the surjective tensor functor $F \boxtimes F : \mathcal{C} \boxtimes \mathcal{C}^{\mathrm{op}} \to \mathcal{D} \boxtimes \mathcal{D}^{\mathrm{op}}$. Let us take the dual of this functor with respect to $\mathcal{D}$. By [Etingof et al. 2005, Proposition 5.3], we get an injective tensor functor (i.e., a fully faithful embedding)

$$(F \boxtimes F)^*_{\mathcal{D}} : \mathcal{Z}(\mathcal{D}) \hookrightarrow (\mathcal{C} \boxtimes \mathcal{C}^{\mathrm{op}})^*_{\mathcal{D}}.$$

But the category $(\mathcal{C} \boxtimes \mathcal{C}^{\mathrm{op}})^*_{\mathcal{D}}$ is semisimple, since $\mathcal{C}$ is separable and $\mathcal{D}$ is semisimple. Hence, $\mathcal{Z}(\mathcal{D})$ is semisimple. Thus, by Corollary 3.5.9 of [Douglas et al. 2013], $\mathcal{D}$ is separable.

The general case now follows by applying the above special case to the surjective tensor functors $F_i : \mathcal{C}_{ii} \to F(\mathbf{1}_i) \otimes \mathcal{D} \otimes F(\mathbf{1}_i)$, where the $\mathbf{1}_i$ are the simple composition factors of $\mathbf{1}$ in $\mathcal{C}$, and where $\mathcal{C}_{ii} := \mathbf{1}_i \otimes \mathcal{C} \otimes \mathbf{1}_i$. □

The following theorem resolves an open question in [Etingof et al. 2005, Section 9.4]:

Theorem 5.3. *A (full) multitensor subcategory of a separable multifusion category is separable.*

Proof. Let $F:\mathcal{C}\hookrightarrow\mathcal{D}$ be an inclusion of $\mathcal{C}$ into a separable category $\mathcal{D}$. Then by [Etingof et al. 2005, Proposition 5.3], we have a surjective tensor functor $F^*_{\mathcal{D}}:\mathcal{D}\to\mathcal{C}^*_{\mathcal{D}}$, where $\mathcal{C}^*_{\mathcal{D}}$ is a multitensor category. By Theorem 5.2, $\mathcal{C}^*_{\mathcal{D}}$ is separable. Hence $\mathcal{D}=(\mathcal{C}^*_{\mathcal{D}})^*_{\mathcal{D}}$ is separable as well. □

5B. ***Descent of tensor functors.***

Theorem 5.4. *Let $\mathcal{C}_1,\mathcal{C}_2,\mathcal{D}$ be separable multifusion categories over k. Let $F_i:\mathcal{C}_i\to\mathcal{D}$ be tensor functors. Let $F:\widehat{\mathcal{C}_1}\to\widehat{\mathcal{C}_2}$ be a tensor functor such that $\widehat{F_2}\circ F\cong\widehat{F_1}$. Then there exists a tensor functor $F_0:\mathcal{C}_1\to\mathcal{C}_2$ such that $F_2\circ F_0\cong F_1$, and $F\cong\widehat{F_0}$. In other words, if tensor functors G and $G\circ F$ are integral then F is integral.*

Note that Theorem 3.7 is recovered from Theorem 5.4 when $\mathcal{D}=\mathrm{Vec}_k$ (namely, $H_i=\mathrm{Coend}\,F_i$). Moreover, if $\mathcal{D}=\mathrm{Rep}\,A$, where A is a semisimple k-algebra, then Theorem 5.4 gives a generalization of Theorem 3.7 to weak Hopf algebras.

Proof. The proof is similar to the proof of Theorem 3.7. If $\mathcal{C}$ is a separable multifusion category over k, there is a natural bijection between full tensor subcategories of $\mathcal{C}$ and $\widehat{\mathcal{C}}$. In particular, $\mathrm{Im}\,F$ is a lift of some multifusion subcategory $\mathcal{E}\subset\mathcal{C}_2$. By Theorem 5.3, $\mathcal{E}$ is separable. Thus, we may replace $\mathcal{C}_2$ by $\mathcal{E}$, i.e., we may assume without loss of generality that F is surjective.

Also, we may replace $\mathcal{D}$ with the image $\mathrm{Im}\,F_2$ of F_2, which is separable by Theorem 5.2 (or Theorem 5.3), i.e., we may assume without loss of generality that F_2 (hence $\widehat{F_2}$, hence $\widehat{F_1}$, hence F_1) is surjective.

Consider the dual functor $F^*_{\widehat{\mathcal{D}}}:(\widehat{\mathcal{C}_2})^*_{\widehat{\mathcal{D}}}\to(\widehat{\mathcal{C}_1})^*_{\widehat{\mathcal{D}}}=\widehat{(\mathcal{C}_1)^*_{\mathcal{D}}}$, which is an inclusion of multifusion categories by [Etingof et al. 2005, Proposition 5.3]. Thus, $(\widehat{\mathcal{C}_2})^*_{\widehat{\mathcal{D}}}$ is a lift of some multifusion subcategory $\mathcal{B}$ of $(\mathcal{C}_1)^*_{\mathcal{D}}$. Moreover, by Theorem 5.3, $\mathcal{B}$ is separable, so $(\widehat{\mathcal{C}_2})^*_{\widehat{\mathcal{D}}}=\widehat{(\mathcal{C}_2)^*_{\mathcal{D}}}\cong\widehat{\mathcal{B}}$. Let $H:\mathcal{B}\hookrightarrow(\mathcal{C}_1)^*_{\mathcal{D}}$ be the corresponding inclusion functor. Then the dual functor $H^*_{\mathcal{D}}:\mathcal{C}_1\to\mathcal{B}^*_{\mathcal{D}}$ is a surjection, and $F\cong F'\circ\widehat{H^*_{\mathcal{D}}}$, where $F':\widehat{\mathcal{B}^*_{\mathcal{D}}}\cong\widehat{\mathcal{C}_2}$ is an equivalence. By Theorem 4.6, F' is isomorphic to the lift $\widehat{F_0'}$ of an equivalence $F_0':\mathcal{B}^*_{\mathcal{D}}\cong\mathcal{C}_2$; hence F is isomorphic to the lift $\widehat{F_0}$ of a tensor functor $F_0=F_0'\circ H^*_{\mathcal{D}}:\mathcal{C}_1\to\mathcal{C}_2$. This proves the theorem. □

Theorem 5.5. *Let $\mathcal{C}_1,\mathcal{C}_2,\mathcal{D}$ be separable multifusion categories over k. Let $F_i:\mathcal{D}\to\mathcal{C}_i$ be tensor functors, such that F_1 is surjective. Let $F:\widehat{\mathcal{C}_1}\to\widehat{\mathcal{C}_2}$ be a tensor functor such that $F\circ\widehat{F_1}\cong\widehat{F_2}$. Then there exists a tensor functor $F_0:\mathcal{C}_1\to\mathcal{C}_2$ such that $F_0\circ F_1\cong F_2$, and $F\cong\widehat{F_0}$. In other words, if tensor functors G and $F\circ G$ are integral and G is surjective then F is integral.*

Proof. We may replace $\mathcal{C}_2$ with $\mathrm{Im}\,F_2$, which is separable by Theorems 5.2 or 5.3, and assume that F_2, $\widehat{F_2}$, F are surjective. Then by [Etingof et al. 2005, Proposition 5.3], we have an inclusion $(\widehat{F_1})^*_{\widehat{\mathcal{C}_2}}:(\widehat{\mathcal{C}_1})^*_{\widehat{\mathcal{C}_2}}\hookrightarrow\widehat{\mathcal{D}}^*_{\widehat{\mathcal{C}_2}}=\widehat{\mathcal{D}^*_{\mathcal{C}_2}}$. The image of $(\widehat{F_1})^*_{\widehat{\mathcal{C}_2}}$ is then a lift of some multifusion subcategory $\mathcal{B}\subset\mathcal{D}^*_{\mathcal{C}_2}$. By Theorem 5.3, $\mathcal{B}$ is separable. Let $H:\mathcal{B}\hookrightarrow\mathcal{D}^*_{\mathcal{C}_2}$ be the corresponding inclusion functor. Then $(\widehat{F_1})^*_{\widehat{\mathcal{C}_2}}$ defines an equivalence $L:(\widehat{\mathcal{C}_1})^*_{\widehat{\mathcal{C}_2}}\cong\widehat{\mathcal{B}}$ such that $(\widehat{F_1})^*_{\widehat{\mathcal{C}_2}}=\widehat{H}\circ L$. Then

$$\widehat{H}\circ L\circ F^*_{\widehat{\mathcal{C}_2}}=(\widehat{F_1})^*_{\widehat{\mathcal{C}_2}}\circ F^*_{\widehat{\mathcal{C}_2}}=(\widehat{F_2})^*_{\widehat{\mathcal{C}_2}}=\widehat{(F_2)^*_{\mathcal{C}_2}}.$$

Thus, by Theorem 5.4, $L \circ F^*_{\widehat{\mathcal{C}}_2} = \widehat{M}$ for some $M : \mathcal{C}_2 \to \mathcal{B}$. Dualizing, we get $\widehat{M^*_{\mathcal{C}_2}} = F \circ L^*_{\widehat{\mathcal{C}}_2}$, where $L^*_{\widehat{\mathcal{C}}_2} : \widehat{\mathcal{B}^*_{\mathcal{C}_2}} \to \widehat{\mathcal{C}}_1$ is an equivalence. By Theorem 4.6, $L^*_{\widehat{\mathcal{C}}_2} = \widehat{N}$ for some equivalence $N : \mathcal{B}^*_{\mathcal{C}_2} \to \mathcal{C}_1$. Thus, $F = \widehat{F}_0$, where $F_0 := M^*_{\mathcal{C}_2} \circ N^{-1}$. □

Remark 5.6. In spite of Theorems 5.4 and 5.5, in general we don't know if any tensor functor $F : \widehat{\mathcal{C}}_1 \to \widehat{\mathcal{C}}_2$ between liftings of separable (multi)fusion categories is always isomorphic to a lifting of a tensor functor $F_0 : \mathcal{C}_1 \to \mathcal{C}_2$, even in the case when $\mathcal{C}_2 = \mathrm{Vec}_k$ and $\mathcal{C}_1 = \mathrm{Rep}\, H$, where H is a semisimple cosemisimple Hopf algebra over k. In this special case, this is the question whether any Drinfeld twist J for $\widehat{H}$ is gauge equivalent to a lifting of a twist J_0 for H.

Likewise, we don't know if any fusion category or semisimple cosemisimple Hopf algebra in characteristic zero whose (global) dimension is coprime to p descends to characteristic p.

Acknowledgements

I am very grateful to Brian Conrad for useful discussions and help with writing Section 2, and to Davesh Maulik for comments on that section. I thank W. van der Kallen for Remark 2.2, D. Nikshych for Remark 4.5, and V. Ostrik for comments on a draft of this paper. I am grateful to the anonymous referee for useful comments. This work was supported by the NSF grant DMS-1502244.

References

[Davydov and Nikshych 2013] A. Davydov and D. Nikshych, "The Picard crossed module of a braided tensor category", *Algebra Number Theory* **7**:6 (2013), 1365–1403. MR Zbl

[Deligne 1990] P. Deligne, "Catégories tannakiennes", pp. 111–195 in *The Grothendieck Festschrift, II*, edited by P. Cartier et al., Progr. Math. **87**, Birkhäuser, Boston, 1990. MR Zbl

[Douglas et al. 2013] C. L. Douglas, C. Schommer-Pries, and N. Snyder, "Dualizable tensor categories", preprint, 2013. arXiv

[EGA IV$_4$ 1967] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, IV", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 5–361. MR Zbl

[Etingof and Gelaki 1998] P. Etingof and S. Gelaki, "On finite-dimensional semisimple and cosemisimple Hopf algebras in positive characteristic", *Int. Math. Res. Not.* **1998**:16 (1998), 851–864. MR Zbl

[Etingof and Gelaki 2000] P. Etingof and S. Gelaki, "The classification of triangular semisimple and cosemisimple Hopf algebras over an algebraically closed field", *Int. Math. Res. Not.* **2000**:5 (2000), 223–234. MR Zbl

[Etingof and Gelaki 2017] P. Etingof and S. Gelaki, "Exact sequences of tensor categories with respect to a module category", *Adv. Math.* **308** (2017), 1187–1208. MR Zbl

[Etingof and Ostrik 2004] P. Etingof and V. Ostrik, "Finite tensor categories", *Mosc. Math. J.* **4**:3 (2004), 627–654. MR Zbl

[Etingof et al. 2005] P. Etingof, D. Nikshych, and V. Ostrik, "On fusion categories", *Ann. of Math.* (2) **162**:2 (2005), 581–642. MR Zbl

[Etingof et al. 2010] P. Etingof, D. Nikshych, and V. Ostrik, "Fusion categories and homotopy theory", *Quantum Topol.* **1**:3 (2010), 209–273. MR Zbl

[Etingof et al. 2015] P. Etingof, S. Gelaki, D. Nikshych, and V. Ostrik, *Tensor categories*, Mathematical Surveys and Monographs **205**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Franjou and van der Kallen 2010] V. Franjou and W. van der Kallen, "Power reductivity over an arbitrary base", *Doc. Math.* Andrei A. Suslin's Sixtieth Birthday volume (2010), 171–195. MR Zbl

[Grosshans 1997] F. D. Grosshans, *Algebraic homogeneous spaces and invariant theory*, Lecture Notes in Math. **1673**, Springer, 1997. MR Zbl

[Haboush 1975] W. J. Haboush, "Reductive groups are geometrically reductive", *Ann. of Math.* (2) **102**:1 (1975), 67–83. MR Zbl

[van der Kallen 2007] W. van der Kallen, "A reductive group with finitely generated cohomology algebras", pp. 301–314 in *Algebraic groups and homogeneous spaces*, edited by V. B. Mehta, Tata Inst. Fund. Res. Stud. Math. **19**, Tata Inst. Fund. Res., Mumbai, 2007. MR Zbl

[Larson and Radford 1988] R. G. Larson and D. E. Radford, "Finite-dimensional cosemisimple Hopf algebras in characteristic 0 are semisimple", *J. Algebra* **117**:2 (1988), 267–289. MR Zbl

[Nichols and Zoeller 1989] W. D. Nichols and M. B. Zoeller, "A Hopf algebra freeness theorem", *Amer. J. Math.* **111**:2 (1989), 381–385. MR Zbl

[Seshadri 1977] C. S. Seshadri, "Geometric reductivity over arbitrary base", *Adv. Math.* **26**:3 (1977), 225–274. MR Zbl

[SGA 3_{III} 1970] M. Demazure and A. Grothendieck, *Schémas en groupes, Tome III: Structure des schémas en groupes réductifs, Exposés XIX–XXVI* (Séminaire de Géométrie Algébrique du Bois Marie 1962–1964), Lecture Notes in Math. **153**, Springer, 1970. MR Zbl

[Springer 1977] T. A. Springer, *Invariant theory*, Lecture Notes in Math. **585**, Springer, 1977. MR Zbl

[Ştefan 1997] D. Ştefan, "The set of types of n-dimensional semisimple and cosemisimple Hopf algebras is finite", *J. Algebra* **193**:2 (1997), 571–580. MR Zbl

[Thomason 1987] R. W. Thomason, "Equivariant resolution, linearization, and Hilbert's fourteenth problem over arbitrary base schemes", *Adv. Math.* **65**:1 (1987), 16–34. MR Zbl

[Yau and Johnson 2015] D. Yau and M. W. Johnson, *A foundation for PROPs, algebras, and modules*, Mathematical Surveys and Monographs **203**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

Communicated by Susan Montgomery
Received 2017-04-27 Revised 2017-10-18 Accepted 2017-12-18

etingof@math.mit.edu *Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, United States*

ALGEBRA AND NUMBER THEORY 12:3 (2018)
dx.doi.org/10.2140/ant.2018.12.571

Mean square in the prime geodesic theorem

Giacomo Cherubini and João Guerreiro

We prove upper bounds for the mean square of the remainder in the prime geodesic theorem, for every cofinite Fuchsian group, which improve on average on the best known pointwise bounds. The proof relies on the Selberg trace formula. For the modular group we prove a refined upper bound by using the Kuznetsov trace formula.

1. Introduction

There is a striking similarity between the distribution of lengths of primitive closed geodesics on the modular surface and prime numbers. If we consider the asymptotic of the associated counting function, the problem of finding optimal upper bounds on the remainder is intriguing, especially because the relevant zeta function is known to satisfy the corresponding Riemann hypothesis.

We start with reviewing briefly the framework of the problem (for a more detailed introduction we refer to [Sarnak 1980; Iwaniec 1995, §10.9]), and then state our results. Since the definitions extend to every finite volume Riemann surface, we work in this generality, and only at a later point do we specialize to the modular surface.

Every cofinite Fuchsian group Γ acts on the hyperbolic plane $\mathbb{H}$ by linear fractional transformations, and every hyperbolic element $g \in \Gamma$ is conjugated over $\mathrm{SL}_2(\mathbb{R})$ to a matrix of the form

$$\begin{pmatrix} \lambda^{1/2} & 0 \\ 0 & \lambda^{-1/2} \end{pmatrix}, \quad \lambda \in \mathbb{R}, \lambda > 1.$$

The trace of g is therefore $\mathrm{Tr}(g) = \lambda^{1/2} + \lambda^{-1/2}$, and we define its norm to be $N(g) = \lambda$. Since the trace and the norm are constant on the conjugacy class of g, if we set $P = \{\gamma g \gamma^{-1}, \gamma \in \Gamma\}$, we can define the trace and the norm of P by $\mathrm{Tr}(P) = \mathrm{Tr}(g)$ and $NP = N(g)$. Moreover, we say that g (and P) is primitive if g cannot be expressed as a positive power (greater than one) of another element $g_0 \in \Gamma$. Every class P can then be written as $P = P_0^{\nu}$ for some primitive conjugacy class P_0 and some positive integer $\nu \geq 1$.

For $X > 0$ define the counting function

$$\pi_\Gamma(X) = \sharp\{P_0 : NP_0 \leq X\}. \tag{1-1}$$

This work was supported by the Max Planck Institut für Mathematik. We would like to record our thanks to the Institute for the excellent working conditions.
MSC2010: primary 11F72; secondary 11L05, 11M36.
Keywords: prime geodesic theorem, Selberg trace formula, Kuznetsov trace formula, Kloosterman sums.

A hyperbolic conjugacy class P determines a closed geodesic on the Riemann surface $\Gamma\backslash\mathbb{H}$, and its length is given exactly by $\log NP$. For this reason we say that $\pi_\Gamma(X)$ in (1-1) counts primitive hyperbolic conjugacy classes in Γ, lengths of primitive geodesics on $\Gamma\backslash\mathbb{H}$, or, by abusing language, "prime" geodesics.

It is a result from the middle of the last century (see e.g., Huber [1961], although the result was already known to Selberg) that as $X \to \infty$ we have the asymptotic formula

$$\pi_\Gamma(X) \sim \mathrm{li}(X), \tag{1-2}$$

where $\mathrm{li}(X)$ is the logarithmic integral. The asymptotic coincides with that of the prime counting function, since we have

$$\pi(X) = \sharp\{p \le X : p \text{ prime}\} \sim \mathrm{li}(X) \quad \text{as } X \to \infty. \tag{1-3}$$

The structure of the two problems is also similar. The study of primes is strictly related to the study of the zeros of the Riemann zeta function $\zeta(s)$; on the other hand, for primitive geodesics we need to study the zeros of the Selberg zeta function $Z_\Gamma(s)$, which is defined for $\Re(s) > 1$ by

$$Z_\Gamma(s) = \prod_{P_0}\prod_{\nu=0}^{\infty}(1-(NP_0)^{-\nu-s}),$$

the outer product ranging over primitive hyperbolic classes in Γ, and extends to a meromorphic function on the complex plane.

It is an interesting question to determine the correct rate of the approximation in (1-2) and (1-3), namely to prove optimal upper bounds on the differences

$$\pi(X) - \mathrm{li}(X) \quad \text{and} \quad \pi_\Gamma(X) - \mathrm{li}(X). \tag{1-4}$$

In the case of primes, the conjectural estimate $|\pi(X) - \mathrm{li}(X)| \ll X^{1/2}\log X$ is equivalent to the Riemann hypothesis (see e.g., [Ivić 1985, Theorem 12.3]).

To understand what type of control we should expect on the difference on the right in (1-4), we introduce the weighted counting function

$$\psi_\Gamma(X) = \sum_{NP \le X} \Lambda_\Gamma(NP),$$

where $\Lambda_\Gamma(NP) = \log(NP_0)$ if P is a power of the primitive conjugacy class P_0 and $\Lambda_\Gamma(x) = 0$ otherwise. The function $\psi_\Gamma(X)$ is the analogous of the classical summatory von Mangoldt function $\psi(X)$ in the theory of primes, and studying $\pi_\Gamma(X)$ is equivalent to study $\psi_\Gamma(X)$, but it is easier to work with the latter.

The asymptotic (1-2) for $\pi_\Gamma(X)$ translates into $\psi_\Gamma(X) \sim X$ as X tends to infinity. The spectral theory of automorphic forms provides a finite number of additional secondary terms and we define the complete main term $M_\Gamma(X)$ to be the function

$$M_\Gamma(X) = \sum_{\frac{1}{2}<s_j\le 1} \frac{X^{s_j}}{s_j}, \tag{1-5}$$

where s_j is defined by $\lambda_j = s_j(1-s_j)$, λ_j are the eigenvalues of the Laplace operator Δ acting on $L^2(\Gamma\backslash\mathbb{H})$, and the sum is restricted to the *small eigenvalues*, that is, those satisfying $0 \le \lambda_j < \frac{1}{4}$. Since the eigenvalues of Δ form a discrete set with no accumulation points, the sum in (1-5) is finite. The remainder $P_\Gamma(X)$ is then defined as

$$P_\Gamma(X) = \psi_\Gamma(X) - M_\Gamma(X).$$

Various types of pointwise upper bounds have been proved in the past years on $P_\Gamma(X)$, and we review them in Remark 1.5 below. In a different direction, it is possible to study the moments of $P_\Gamma(X)$, and this has been neglected so far.

In the case of the prime number theorem, it is a classical result (assuming the Riemann hypothesis) that the normalized remainder

$$\frac{\psi(e^y) - e^y}{e^{y/2}}$$

admits moments of every order and a limiting distribution with exponentially small tails (see [Wintner 1935, p. 242] and [Rubinstein and Sarnak 1994, Theorem 1.2]). The aim of this paper is to prove the following estimate on the second moment of $P_\Gamma(X)$.

Theorem 1.1. *Let Γ be a cofinite Fuchsian group. For $A \gg 1$ we have*

$$\frac{1}{A}\int_A^{2A} |P_\Gamma(X)|^2\, dX \ll A^{4/3}.$$

Remark 1.2. By construction the zeros of the Selberg zeta function correspond to the eigenvalues of Δ. The zeros corresponding to small eigenvalues give rise to the secondary terms in the definition of $M_\Gamma(X)$, and there are no other zeros in the half plane $\Re(s) > \frac{1}{2}$. In this sense $Z_\Gamma(s)$ satisfies an analogous of the Riemann hypothesis, and the fact suggests that we might have

$$P_\Gamma(X) \ll X^{1/2+\varepsilon} \quad \forall \varepsilon > 0, \tag{1-6}$$

since this is the case for the primes under RH. However, the same method of proof fails, due to the abundance of zeros of $Z_\Gamma(s)$ (which in turn is related to the fact that $Z_\Gamma(s)$ is a function of order two, while $\zeta(s)$ is of order one), and only gives

$$P_\Gamma(X) \ll X^{3/4}. \tag{1-7}$$

For a proof we refer to [Iwaniec 1995, Theorem 10.5]. On the other hand, it is possible to prove that (1-6), if true, is optimal, since we have the Omega result by Hejhal [1983, Theorem 3.8, p. 477, and note 18, p. 503]

$$P_\Gamma(X) = \Omega_\pm(X^{1/2-\delta}) \quad \forall \delta > 0,$$

which can be strengthened to $\Omega_{\pm}(X^{1/2}(\log\log X)^{1/2})$ for cocompact groups and congruence groups.

Remark 1.3. Recently Koyama [2016] and Avdispahić [2017; 2018] have tried an approach to find upper bounds on $P_\Gamma(X)$ using a lemma of Gallagher [1970, Lemma 1]. Their results are of the type

$$P_\Gamma(X) \ll X^{\eta+\varepsilon} \quad \forall\varepsilon > 0,$$

where $\eta = \frac{7}{10}$ or $\eta = \frac{3}{4}$ for Γ cocompact or ordinary cofinite, respectively, and with X outside a set $B \subseteq [1, \infty)$ of finite logarithmic measure, that is, such that $\int_B X^{-1}\,dX < \infty$. Theorem 1.1 improves on this to the extent that η can be taken to be any number greater than $\frac{2}{3}$. Indeed, a direct consequence of Theorem 1.1 is that for every cofinite Fuchsian group and for every $\varepsilon > 0$, the set

$$B = \{X \geq 1 : |P_\Gamma(X)| \geq X^{2/3}(\log X)^{1+\varepsilon}\}$$

has finite logarithmic measure. Iwaniec [1984a, p. 187] claims without proof that "it is easy to prove that $P_\Gamma(X) \ll X^{2/3}$ for almost all X". Theorem 1.1 proves his claim (up to ε) for X outside a set of finite logarithmic measure.

In the case of the modular group we can prove a stronger bound than for the general cofinite case. In this case we can exploit the Kuznetsov trace formula, obtaining the following.

Theorem 1.4. *Let $G = \mathrm{PSL}(2, \mathbb{Z})$. For $A \gg 1$ and every $\varepsilon > 0$,*

$$\frac{1}{A}\int_A^{2A} |P_G(X)|^2\,dX \ll A^{5/4+\varepsilon}.$$

Remark 1.5. The estimate in (1-7) is currently the best known pointwise upper bound in the case of general cofinite Fuchsian groups. For the modular group and congruence groups it is possible to prove sharper estimates. Iwaniec [1984b, Theorem 2], Luo and Sarnak [1995, Theorem 1.4], Cai [2002, p. 62], and Soundararajan and Young [2013, Theorem 1.1] have worked on reducing the exponent for the modular group $G = \mathrm{PSL}(2, \mathbb{Z})$. The currently best known result is

$$P_G(X) \ll X^{\eta+\varepsilon}, \quad \text{for } \eta = \tfrac{25}{36} \text{ and every } \varepsilon > 0,$$

due to Soundararajan and Young. The exponent $\frac{7}{10}$, proved by Luo and Sarnak, holds also for congruence groups, see [Luo et al. 1995, Corollary 1.2], and for cocompact groups coming from quaternion algebras, see [Koyama 1998].

Observe that not only is the exponent $\eta = \frac{1}{2}$ out of reach, but it seems to be a hard problem reaching and, afterwards, breaking the barrier $\eta = \frac{2}{3}$. Iwaniec [1984a, p. 188; 1984b, (12)] suggested that this follows from the assumption of the Lindelöf hypothesis for Rankin L-functions. In Theorem 1.1 we prove the bound with the exponent $\frac{2}{3}$ on average, unconditionally, for every cofinite Fuchsian group. Theorem 1.4 is saying that $P_G(X) \ll X^{5/8+\varepsilon}$ on average. This goes halfway between the trivial bound (1-7) and the conjectural estimate (1-6).

In the proof of Theorem 1.4 we use a truncated formula for $\psi_\Gamma(X)$, proved by Iwaniec [1984b, Lemma 1], condensing the analysis of the Selberg trace formula, and relating $P_\Gamma(X)$ to the spectral exponential sum

$$R(X,T)=\sum_{t_j\le T} X^{it_j}.$$

The trivial bound for this sum, in view of Weyl's law (see e.g., [Iwaniec 1995, Corollary 11.2] or [Venkov 1990, Theorem 7.3]), is $|R(X,T)|\ll T^2$. Reducing the exponent of T is possible at the cost of some power of X (see [Iwaniec 1984b; Luo and Sarnak 1995]), and Petridis and Risager conjectured [2017, Conjecture 2.2] that we should have square root cancellation, that is,

$$R(X,T)\ll T^{1+\varepsilon}X^{\varepsilon}$$

for every $\varepsilon>0$. This would give the conjectural bound (1-6) for the modular group. In the appendix to [Petridis and Risager 2017] Laaksonen provides numerics that support this conjecture. In the proof of Theorem 1.4 (see Proposition 4.5) we prove that we have

$$\frac{1}{A}\int_A^{2A}|R(X,T)|^2\,dX\ll T^{2+\varepsilon}A^{1/4+\varepsilon},$$

and hence $|R(X,T)|\ll T^{1+\varepsilon}X^{1/8+\varepsilon}$ on average.

The mean square estimate in Theorem 1.4 reduces to proving a similar estimate for certain weighted sums of Kloosterman sums. We obtain the desired bounds in a relatively clean way by applying the Hardy–Littlewood–Pólya inequality [Hardy et al. 1934, Theorem 381, p. 288] in a special case: for $0<\lambda<1$, $\lambda=2(1-p^{-1})$, and $\{a_r\}$ a sequence of nonnegative numbers, then we have

$$\sum_{\substack{r,s\\ r\neq s}}\frac{a_ra_s}{|r-s|^{\lambda}}\ll_\lambda\Big(\sum_r a_r^p\Big)^{2/p}. \tag{1-8}$$

With more work it is possible to sharpen Theorem 1.4 by replacing A^ε by some power of $\log A$. To do this, one can use a version of (1-8) with explicit implied constant proved by Carneiro and Vaaler [2010, Corollary 7.2, (7.20)], or the extremal case of (1-8) with $(\lambda,p)=(1,2)$, and a logarithmic correction, proved by Li and Villavert [2011].

We also note that we simply use the Weil bound when estimating the weighted sums of Kloosterman sums and we do not exploit any cancellation amongst the Kloosterman sums. Exploring this phenomenon could lead to a power-saving improvement of Theorem 1.4.

For the proof of Theorem 1.1 we use the Selberg trace formula with a suitably chosen test function that fulfills our needs. In the general case of Γ cofinite, unlike in the modular group case, Iwaniec's truncated formula for $\psi_\Gamma(X)$ is not available. It is probably possible to prove an analogous result, but we preferred to work directly with the trace formula. In fact, in order to prove his formula, Iwaniec uses some arithmetic information on the structure of the lengths of closed geodesics on the modular surface

[Iwaniec 1984b, Lemma 4], namely their connection with primitive binary quadratic forms (proved by Sarnak [1982, §1]), and it is unclear whether a similar argument works for the general cofinite case.

Remark 1.6. We do not discuss the first moment of $P_\Gamma(X)$. Following the argument of [Phillips and Rudnick 1994, Theorems 1.1 and 1.4] it should be possible to prove that for every cofinite Fuchsian group Γ there exists a constant L_Γ such that we have

$$\lim_{A\to\infty}\frac{1}{\log A}\int_1^A\left(\frac{P_\Gamma(X)}{X^{1/2}}\right)\frac{dX}{X}=L_\Gamma.$$

2. Proof of Theorem 1.1

In this section we explain the structure of the proof of Theorem 1.1. We decided to keep this section to a more colloquial tone, and to relegate the technical computations to the next section. Hence the proof of the theorem is completed in Section 3.

The starting point of our analysis is the Selberg trace formula, that we recall from the book of Iwaniec [1995, Theorem 10.2]. An even function g is said to be an admissible test function in the trace formula if its Fourier transform $h(t)$ (see (2-4) for the convention used to define the Fourier transform) satisfies the conditions [Iwaniec 1995, (1.63)]

$$\begin{aligned}&h(t)\text{ is even,}\\&h(t)\text{ is holomorphic in the strip }|\Im(t)|\le\tfrac12+\varepsilon,\\&h(t)\ll(1+|t|)^{-2-\varepsilon}\text{ in the strip.}\end{aligned}\tag{2-1}$$

For an admissible test function g, the Selberg trace formula is the identity

$$\sum_P\frac{g(\log NP)}{2\sinh((\log NP)/2)}\Lambda_\Gamma(NP)=IE+EE+PE+DS+CS+AL,\tag{2-2}$$

where the sum on the left runs over hyperbolic conjugacy classes in Γ, and the terms appearing on the right are explained as follows.

The term IE denotes a contribution coming from the identity element, which forms a conjugacy class on its own. We have

$$IE=-\frac{\operatorname{vol}(\Gamma\backslash\mathbb{H})}{4\pi}\int_{-\infty}^{+\infty}th(t)\tanh(\pi t)\,dt.$$

The term EE denotes a contribution from the elliptic conjugacy classes in Γ (there are only finitely many such classes). Denote by $\mathcal{R}$ a primitive elliptic conjugacy class, and let $m=m_{\mathcal{R}}>1$ be the order of $\mathcal{R}$. We have

$$EE=-\sum_{\mathcal{R}}\sum_{1\le\ell<m}\left(2m\sin\left(\frac{\pi\ell}{m}\right)\right)^{-1}\int_{-\infty}^{+\infty}h(t)\frac{\cosh\left(\pi\left(1-\frac{2\ell}{m}\right)t\right)}{\cosh(\pi t)}\,dt.$$

The term PE denotes a contribution from the parabolic conjugacy classes, which are associated to the cusps of Γ. Let $\mathfrak{C}$ denote the total number of inequivalent cusps. Then we have

$$PE = \frac{\mathfrak{C}}{2\pi}\int_{-\infty}^{+\infty} h(t)\psi(1+it)\,dt + \mathfrak{C}g(0)\log 2.$$

Here $\psi(z)$ is the digamma function, i.e., the logarithmic derivative of the gamma function, $\psi(z) = (\Gamma'/\Gamma)(z)$. The term DS corresponds to the discrete spectrum, and is given by

$$DS = \sum_{t_j} h(t_j),$$

where the sum runs over the spectral parameter t_j, associated to the eigenvalue λ_j of the Laplace operator via the identity $\lambda_j = \frac{1}{4} + t_j^2$ (here and in the following we assume that $t_j \in \left[0, \frac{i}{2}\right]$ for $\lambda_j \in \left[0, \frac{1}{4}\right]$, and $t_j > 0$ for $\lambda_j > \frac{1}{4}$). The term CS comes from the continuous spectrum, and is given by

$$CS = \frac{1}{4\pi}\int_{-\infty}^{+\infty} h(t)\frac{-\varphi'}{\varphi}\left(\tfrac{1}{2}+it\right)dt,$$

where $\varphi(s)$ is the scattering determinant for the group Γ [Iwaniec 1995, p. 140]. Finally, the term AL is a single term that comes from a combination of the spectral part and the geometric part [Iwaniec 1995, (10.11) and (10.17)], and it is defined by $AL = \frac{1}{4}h(0)\operatorname{Tr}\left(\Phi\left(\frac{1}{2}\right) - \mathrm{I}\right)$, where $\Phi(s)$ is the scattering matrix associated to Γ.

A first naive choice of test function in the Selberg trace formula (2-2) is

$$g_\rho(x) = 2\sinh\left(\tfrac{1}{2}|x|\right)\mathbf{1}_{[0,\rho]}(|x|), \tag{2-3}$$

where $\rho = \log X$. If we choose $g(x)$ in this way, then the left hand side of (2-2) reduces exactly to the function $\psi_\Gamma(X)$. Unfortunately the function $g_\rho(x)$ in (2-3) is not an admissible test function, and we need therefore to take a suitable approximation of it. An analysis of the right hand side in (2-2) then leads to the desired results. To see that g_ρ is not admissible in the trace formula, consider its Fourier transform $h_\rho(t)$. We have

$$h_\rho(t) = \int_{\mathbb{R}} g_\rho(x)e^{-itx}\,dx = 2\int_0^\rho \sinh\left(\tfrac{1}{2}x\right)(e^{itx}+e^{-itx})\,dx. \tag{2-4}$$

The integral can be computed directly, which gives, for $t = \pm\frac{i}{2}$,

$$h_\rho\left(\pm\tfrac{1}{2}i\right) = 4\sinh^2\left(\tfrac{1}{2}\rho\right), \tag{2-5}$$

and, for $t \neq \pm\frac{i}{2}$,

$$h_\rho(t) = \frac{2}{\frac{1}{2}+it}\cosh\left(\rho\left(\tfrac{1}{2}+it\right)\right) + \frac{2}{\frac{1}{2}-it}\cosh\left(\rho\left(\tfrac{1}{2}-it\right)\right) - \frac{2}{\frac{1}{4}+t^2}. \tag{2-6}$$

The last integral in (2-4) shows that $h_\rho(t)$ is even and entire in t, but from (2-6) we see that we only have $h_\rho(t) \ll t^{-1}$ as t tends to infinity, and so we do not have sufficient decay as required in (2-1).

We construct two functions $g_\pm(x)$ that approximate from above and below the function g_ρ, and are admissible in the Selberg trace formula. Let $q(x)$ be an even, smooth, nonnegative function on $\mathbb{R}$ with compact support contained in $[-1, 1]$ and unit mass (i.e., $\|q\|_1 = 1$). Let $0 < \delta < \frac{1}{4}$ and define

$$q_\delta(x) = \frac{1}{\delta} q\left(\frac{x}{\delta}\right).$$

For $\rho > 1$ define $g_\pm$ to be the convolution product of the shifted function $g_{\rho\pm\delta}$ with the function q_δ, namely

$$g_\pm(x) = (g_{\rho\pm\delta} * q_\delta)(x) = \int_{\mathbb{R}} g_{\rho\pm\delta}(x-y) q_\delta(y)\, dy. \tag{2-7}$$

Taking convolution products has the advantage that the Fourier transform of the convolution is the pointwise product of the Fourier transforms of the two factors. Hence if we denote by $\hat{q}_\delta(t)$ the Fourier transform of q_δ, that is,

$$\hat{q}_\delta(t) = \int_{\mathbb{R}} q_\delta(x) e^{-itx}\, dx, \tag{2-8}$$

and by $h_\pm$ the Fourier transform of $g_\pm$, then we obtain

$$h_\pm(t) = h_{\rho\pm\delta}(t)\hat{q}_\delta(t). \tag{2-9}$$

Since the function $\hat{q}_\delta(t)$ is entire and satisfies, for $|\Im(t)| \le M < \infty$,

$$\hat{q}_\delta(t) \ll \frac{1}{1+\delta^k |t|^k} \quad \forall k \ge 0$$

(see Lemma 3.1), we conclude that $h_\pm(t)$ is an entire function and satisfies $h_\pm(t) \ll (1+|t|)^{-2-\varepsilon}$ for some $\varepsilon > 0$ in the strip $|\Im(t)| \le \frac{1}{2} + \varepsilon$, as required in (2-1). This shows that the function $g_\pm$ is an admissible test function in the trace formula. By construction, the function $g_\pm(x)$ is supported on $|x| \in [0, \rho+\delta\pm\delta]$. Moreover we have the inequalities (see Lemma 3.2), for $x \ge 0$,

$$g_-(x) + O(\delta e^{x/2} \mathbf{1}_{[0,\rho]}(x)) \le g_\rho(x) \le g_+(x) + O(\delta e^{x/2} \mathbf{1}_{[0,\rho]}(x)).$$

If we set

$$\psi_\pm(X) = \sum_P \frac{g_\pm(\log NP)}{2\sinh((\log NP)/2)} \Lambda_\Gamma(NP),$$

then using the asymptotic $\psi_\Gamma(X) \sim X$ we conclude that we have the inequalities

$$\psi_-(X) + O(\delta X) \le \psi_\Gamma(X) \le \psi_+(X) + O(\delta X).$$

From this we deduce that we have

$$\frac{1}{A}\int_A^{2A} |P(X)|^2\, dX \ll \frac{1}{A}\int_A^{2A} |\psi_\pm(X) - M_\Gamma(X)|^2\, dX + O(\delta^2 A^2), \tag{2-10}$$

and in order to prove Theorem 1.1 we give bounds on the right hand side in (2-10). We found convenient to pass to the logarithmic variable $\rho = \log X$, since this simplifies slightly the computations. Moreover, we insert a weight function in the integral to pass from the sharp mean square to a smooth one. Consider a smooth, nonnegative real function $w(\rho)$, compactly supported in $[-\varepsilon, 1+\varepsilon]$, for $0<\varepsilon<\frac{1}{4}$. In addition, assume that $0 \le w(\rho) \le 1$, and that $w(\rho)=1$ for $\rho \in [0, 1]$. Let

$$w_R(\rho) = w(\rho - R). \tag{2-11}$$

In view of the inequality

$$\frac{1}{A}\int_A^{2A} |\psi_\pm(X) - M_\Gamma(X)|^2 \, dX \ll \int_{\mathbb{R}} |\psi_\pm(e^\rho) - M_\Gamma(e^\rho)|^2 w_R(\rho)\, d\rho, \tag{2-12}$$

where $R = \log A$, we see that in order to estimate the last integral in (2-10) it suffices to upper bound the integral on the right in (2-12). In the following, abusing notation, we write $\psi_\pm(\rho)$ in place of $\psi_\pm(e^\rho)$, and $M_\Gamma(\rho)$ in place of $M_\Gamma(e^\rho)$.

At this point we can exploit the Selberg trace formula to analyze $\psi_\pm$. The terms of the discrete spectrum DS associated to the small eigenvalues $\lambda_j \in \left[0, \frac{1}{4}\right]$ (corresponding to the spectral parameter t_j in the interval $\left[0, \frac{i}{2}\right]$, and for technical convenience we include the eigenvalue $\lambda_j = \frac{1}{4}$) need particular care. We define $M_\pm(\rho)$ to be the sum of such terms, namely

$$M_\pm(\rho) := \sum_{t_j \in \left[0, \frac{i}{2}\right]} h_\pm(t_j).$$

In view of the definition of the complete main term in (1-5) it is easy to prove (see Lemma 3.3) that we have

$$M_\pm(\rho) = M_\Gamma(\rho) + O(\delta e^\rho + e^{\rho/2}).$$

Hence we have

$$\begin{aligned} \psi_\pm(\rho) - M_\Gamma(\rho) &= \psi_\pm(\rho) - M_\pm(\rho) + O(\delta e^\rho + e^{\rho/2}) \\ &= IE_\pm + EE_\pm + PE_\pm + DS'_\pm + CS_\pm + AL_\pm + O(\delta e^\rho + e^{\rho/2}), \end{aligned} \tag{2-13}$$

where the term $DS'_\pm$ denotes now the contribution from the discrete spectrum restricted to the eigenvalues $\lambda_j > \frac{1}{4}$. The first three terms and the term $AL_\pm$ in (2-13) can be bounded pointwise (see Lemma 3.5–3.7) by

$$|IE_\pm| + |EE_\pm| + |PE_\pm| + |AL_\pm| \ll e^{\rho/2} + \log(\delta^{-1}).$$

Squaring and integrating in (2-13) we obtain therefore

$$\begin{aligned} &\int_{\mathbb{R}} |\psi_\pm(\rho) - M_\Gamma(\rho)|^2 w_R(\rho)\, d\rho \\ &\qquad \ll \int_{\mathbb{R}} |DS'_\pm|^2 w_R(\rho)\, d\rho + \int_{\mathbb{R}} |CS_\pm|^2 w_R(\rho)\, d\rho + O(e^R + \delta^2 e^{2R} + \log^2(\delta^{-1})). \end{aligned} \tag{2-14}$$

In Propositions 3.10 and 3.11 we prove that the following estimate holds:

$$\int_{\mathbb{R}} |DS'_\pm|^2 w_R(\rho)\, d\rho + \int_{\mathbb{R}} |CS_\pm|^2 w_R(\rho)\, d\rho \ll \frac{e^R}{\delta} + e^{R/2} \log^2(\delta^{-1}). \tag{2-15}$$

Combining (2-14) and (2-15) we obtain

$$\int_{\mathbb{R}} |\psi_{\pm}(\rho) - M_{\Gamma}(\rho)|^2 w_R(\rho)\, d\rho \ll \frac{e^R}{\delta} + e^{R/2} \log^2(\delta^{-1}) + \delta^2 e^{2R},$$

and choosing $\delta = e^{-R/3}$ to optimize the first and the last term, we arrive at the bound

$$\int_{\mathbb{R}} |\psi_{\pm}(s) - M_{\Gamma}(s)|^2 w_R(\rho)\, d\rho \ll e^{4R/3}.$$

Recalling (2-10) and (2-12), and setting $R = \log A$, we conclude that we have

$$\frac{1}{A} \int_A^{2A} |P(X)|^2\, dX \ll \int_{\mathbb{R}} |\psi_{\pm}(\rho) - M_{\Gamma}(\rho)|^2 w_R(\rho)\, d\rho + O(\delta^2 A^2) \ll A^{4/3}.$$

This proves Theorem 1.1.

3. Technical lemmata

In this section we prove the auxiliary results needed in Section 2 to prove Theorem 1.1. We start with a simple computation to bound the function $\hat{q}_\delta(t)$ and its derivatives.

Lemma 3.1. *Let* $0 < \delta < \frac{1}{4}$, $\hat{q}_\delta(t)$ *as in* (2-8), *and let* $t \in \mathbb{R}$. *Let* $j, k \in \mathbb{N}$, $j, k \geq 0$. *We have*

$$\left| \frac{d^j \hat{q}_\delta}{dt^j}(t) \right| \ll \frac{\delta^j}{1 + |\delta t|^k},$$

and the implied constant depends on j *and* k.

Proof. From the definition of $\hat{q}_\delta(t)$ we have

$$\frac{d^j \hat{q}_\delta}{dt^j}(t) = \frac{d^j}{dt^j} \int_{\mathbb{R}} q(x) e^{-it\delta x}\, dx = (-i\delta)^j \int_{\mathbb{R}} x^j q(x) e^{-it\delta x}\, dx. \tag{3-1}$$

Bounding in absolute value we get

$$\frac{d^j \hat{q}_\delta}{dt^j}(t) \ll \delta^j. \tag{3-2}$$

Integrating by parts in the last integral of (3-1), and using that $q(x)$ is smooth and has compact support, we obtain instead

$$\frac{d^j \hat{q}_\delta}{dt^j}(t) = (-1)^k \frac{(-i\delta)^j}{(-it\delta)^k} \int_{\mathbb{R}} e^{-it\delta x} \frac{d^k}{dx^k}(x^j q(x))\, dx \ll \frac{\delta^j}{|\delta t|^k}. \tag{3-3}$$

Combining (3-2) and (3-3) we obtain

$$\frac{d^j \hat{q}_\delta}{dt^j}(t) \ll \delta^j \min(1, |\delta t|^{-k}) \ll \frac{\delta^j}{1 + |\delta t|^k}.$$

This proves the lemma. □

The function $\hat{q}_\delta(t)$ is used to construct the convolution product $g_\pm = g_{\rho\pm\delta} * q_\delta$ that approximate the function g_ρ. In the next lemma we show that g_ρ is bounded from above and below by g_+ and g_- respectively, up to a small error.

Lemma 3.2. *Let $\rho > 1$ and $0 < \delta < \frac{1}{4}$. Let $g_\rho(x)$ and $g_\pm(x)$ as in* (2-3) *and* (2-7) *respectively. For $x \geq 0$ we have*

$$g_-(x) + O(\delta e^{x/2}\mathbf{1}_{[0,\rho]}(x)) \leq g_\rho(x) \leq g_+(x) + O(\delta e^{x/2}\mathbf{1}_{[0,\rho]}(x)). \tag{3-4}$$

Proof. Observe that $g_\pm = g_\rho * q_\delta$, and since both factors in the convolution product are nonnegative, we conclude that $g_\pm(x) \geq 0$ for every $x \in \mathbb{R}$. Observe also that $g_\pm$ is supported on the compact set $\{|x| \leq \rho + \delta \pm \delta\}$. By definition of $g_\pm$ we can therefore write, for $x \geq 0$,

$$g_\pm(x) = 2\int_{\mathbb{R}} \mathbf{1}_{[0,\rho\pm\delta]}(|x-y|)\sinh\left(\tfrac{1}{2}|x-y|\right)q_\delta(y)\,dy = 2\cdot\mathbf{1}_{[0,\rho+\delta\pm\delta]}(x)\int_Q \sinh\left(\tfrac{1}{2}|x-y|\right)q_\delta(y)\,dy,$$

where $Q = [\max\{-\delta, x-(\rho\pm\delta)\}, \delta]$. For $0 \leq x \leq \delta$ we have

$$g_\rho(x) = O(\delta) \quad \text{and} \quad g_\pm(x) = O(\delta),$$

so that (3-4) holds trivially.

For $x > \delta$ we have $x - y > 0$ for every $y \in Q$, and using the addition formula for the hyperbolic sine we can write

$$\begin{aligned} g_\pm(x) &= 2\cdot\mathbf{1}_{[0,\rho+\delta\pm\delta]}(x)\int_Q \left[\cosh\left(\tfrac{1}{2}y\right)\sinh\left(\tfrac{1}{2}x\right) - \cosh\left(\tfrac{1}{2}x\right)\sinh\left(\tfrac{1}{2}y\right)\right]q_\delta(y)\,dy \\ &= 2\cdot\mathbf{1}_{[0,\rho+\delta\pm\delta]}(x)\sinh\left(\tfrac{1}{2}x\right)\int_Q q_\delta(y)\,dy + O(\delta e^{x/2}\mathbf{1}_{[0,\rho+\delta\pm\delta]}(x)). \end{aligned} \tag{3-5}$$

Now for $\delta < x < \rho \pm \delta - \delta$ we have $Q = [-\delta, \delta]$, so that (3-5) reduces to

$$g_\pm(x) = g_\rho(x) + O(\delta e^{x/2}\mathbf{1}_{[0,\rho+\delta\pm\delta]}(x)).$$

For $\rho - 2\delta \leq x \leq \rho$ we can bound by positivity in (3-5)

$$g_-(x) \leq g_\rho(x) + O(\delta e^{x/2}\mathbf{1}_{[0,\rho]}(x)),$$

so that the first inequality in (3-4) holds in this case. Finally for $\rho \leq x \leq \rho + 2\delta$ we observe that $g_\rho(x) = 0$ and $g_+(x) \geq 0$, and so we conclude that the second inequality in (3-4) holds in this case. This proves the lemma. □

The function $h_\pm(t)$ associated to $g_\pm(x)$ gives, for t corresponding to small eigenvalues, the terms appearing in the definition of the main term $M_\Gamma(X)$ in (1-5), up to small error. We prove this in the next lemma.

Lemma 3.3. *Let $\rho > 1$, $0 < \delta < \frac{1}{4}$, and let $h_\pm(t)$ as in* (2-9). *Let $t_j \in \left(0, \frac{i}{2}\right]$ be the spectral parameter associated to the eigenvalue $\lambda_j \in \left[0, \frac{1}{4}\right)$. There exists $0 < \varepsilon_\Gamma < \frac{1}{4}$ such that*

$$h_\pm(t_j) = \frac{e^{\rho(1/2+|t_j|)}}{\frac{1}{2}+|t_j|} + O(\delta e^\rho + e^{\rho(1/2-\varepsilon_\Gamma)}). \tag{3-6}$$

Proof. Recall that there are only finitely many eigenvalues λ_j in $\left(0, \frac{1}{4}\right)$, and therefore there exists $0 < \varepsilon_\Gamma < \frac{1}{4}$ such that $|t_j| \geq \varepsilon_\Gamma$ for every $\lambda_j \in \left(0, \frac{1}{4}\right)$. Observe that for $|t| \leq \frac{1}{2}$ we have

$$\hat{q}_\delta(t) = \int_{\mathbb{R}} q(x) e^{-it\delta x}\, dx = 1 + O(\delta|t|).$$

The claim then follows from (2-5) and (2-6), and from the fact that $h_\pm(t) = h_{\rho\pm\delta}(t)\hat{q}_\delta(t)$. □

Remark 3.4. Equation (3-6) also holds for $\lambda_j = \frac{1}{4}$, except that we have to multiply the first term by a factor of 2.

The next three lemmata show that in the problem of estimating $P_\Gamma(X)$ using the Selberg trace formula we can neglect the terms coming from the identity class, the elliptic classes, and the parabolic classes, as they contribute a small quantity.

Lemma 3.5. *Let $\rho > 1$, $0 < \delta < \frac{1}{4}$, and let $g_\pm$ and $h_\pm$ as in* (2-7) *and* (2-9) *respectively. Then we have*

$$\int_{-\infty}^{+\infty} t h_\pm(t) \tanh(\pi t)\, dt \ll \frac{e^{\rho/2}}{\rho^2} + \log(\delta^{-1}).$$

Proof. Recall that we have $h_\pm(t) = h_\rho(t)\hat{q}_\delta(t)$, and that by Lemma 3.1 we have $\hat{q}_\delta(t) \ll (1 + |\delta t|^k)^{-1}$ for every $k \geq 0$. Using (2-6) and the definition of the hyperbolic cosine we can write

$$h_{\rho\pm\delta}(t) = \frac{2}{1+2it}\left(e^{(\rho\pm\delta)(1/2+it)} + e^{-(\rho\pm\delta)(1/2+it)}\right) + \frac{2}{1-2it}\left(e^{(\rho\pm\delta)(1/2-it)} + e^{-(\rho\pm\delta)(1/2-it)}\right) - \frac{2}{\frac{1}{4}+t^2}. \tag{3-7}$$

Bounding in absolute value the integrand associated to the last term in (3-7), we obtain

$$\int_{\mathbb{R}} \left| \frac{t\,\hat{q}_\delta(t) \tanh(\pi t)}{\frac{1}{4}+t^2} \right| dt \ll \int_{\mathbb{R}} \frac{dt}{(1+|t|)(1+|\delta t|)} \ll \log(\delta^{-1}) + 1. \tag{3-8}$$

Now consider the integrand associated to the first term in (3-7) We integrate by parts twice and obtain

$$\begin{aligned} I &:= 2\int_{\mathbb{R}} \frac{t\hat{q}_\delta(t)\tanh(\pi t) e^{(\rho\pm\delta)(1/2+it)}}{1+2it}\, dt \\ &= 2\sum_{j=1}^{2} (-1)^{j-1} \frac{e^{(\rho\pm\delta)(1/2+it)}}{(i(\rho\pm\delta))^j} \frac{d^{j-1}}{dt^{j-1}} \left(\frac{t\hat{q}_\delta(t)\tanh(\pi t)}{(1+2it)} \right) \Bigg|_{t=-\infty}^{+\infty} \\ &\qquad + 2\int_{\mathbb{R}} \frac{e^{(\rho\pm\delta)(1/2+it)}}{(i(\rho\pm\delta))^2} \frac{d^2}{dt^2} \left(\frac{t\hat{q}_\delta(t)\tanh(\pi t)}{(1+2it)} \right) dt. \end{aligned}$$

By Lemma 3.1 we see that the boundary terms vanish, and we can bound

$$\frac{d^2}{dt^2}\left(\frac{t\hat{q}_\delta(t)\tanh(\pi t)}{(1+2it)} \right) \ll \frac{1}{(1+|\delta t|^k)} \left(\delta^2 + \frac{1}{1+|t|^2} \right),$$

for every $k \geq 0$, with implied constant depending on k. Fixing $k > 1$ we obtain

$$I \ll \frac{e^{\rho/2}}{\rho^2} \int_{\mathbb{R}} \frac{1}{(1+|\delta t|^k)} \left(\delta^2 + \frac{1}{1+|t|^2} \right) dt \ll \frac{e^{\rho/2}}{\rho^2}. \tag{3-9}$$

The other terms in (3-7) are treated similarly, and are bounded by the same quantity. Adding (3-8) and (3-9) we obtain the desired estimate. □

Lemma 3.6. *Let $\rho > 1$, $0 < \delta < \frac{1}{4}$, and $h_\pm$ as in* (2-9). *Let $m, \ell \in \mathbb{N}$, with $m \geq 2$ and $1 \leq \ell < m$. Then*

$$\int_{\mathbb{R}} h_\pm(t) \frac{\cosh(\pi(1-2\ell/m)t)}{\cosh(\pi t)}\, dt \ll \frac{e^{\rho/2}}{\rho^2},$$

with implied constant that depends on m and ℓ.

Proof. Recall that $h_\pm(t) = h_{\rho\pm\delta}(t)\hat{q}_\delta(t)$, and use (3-7) to express $h_{\rho\pm\delta}(t)$. The integrand associated to the last term in (3-7) is bounded by

$$\int_{\mathbb{R}} \frac{\cosh(\pi(1-2\ell/m)t)}{(1+t^2)\cosh(\pi t)}\, dt \ll \int_{\mathbb{R}} \frac{dt}{1+t^2} \ll 1.$$

Now consider the term $\exp\bigl((\rho\pm\delta)\bigl(\frac{1}{2}+it\bigr)\bigr)$ in (3-7). The corresponding integral contributes

$$\begin{aligned} J &:= 2\int_{\mathbb{R}} \frac{e^{(\rho\pm\delta)(1/2+it)}\hat{q}_\delta(t)\cosh(\pi(1-2\ell/m)t)}{(1+2it)\cosh(\pi t)}\, dt \\ &= 2\sum_{j=1}^{2}(-1)^{j-1}\frac{e^{(\rho\pm\delta)(1/2+it)}}{(i(\rho\pm\delta))^j}\frac{d^{j-1}}{dt^{j-1}}\biggl(\frac{\hat{q}_\delta(t)\cosh(\pi t(1-2\ell/m))}{(1+2it)\cosh(\pi t)}\biggr)\bigg|_{t=-\infty}^{+\infty} \\ &\qquad\qquad + 2\int_{\mathbb{R}} \frac{e^{(\rho\pm\delta)(1/2+it)}}{(i(\rho\pm\delta))^2}\frac{d^2}{dt^2}\biggl(\frac{\hat{q}_\delta(t)\cosh(\pi t(1-2\ell/m))}{(1+2it)\cosh(\pi t)}\biggr)\, dt. \end{aligned}$$

The boundary terms vanish, and by Lemma 3.1 we can bound

$$\frac{d^2}{dt^2}\biggl(\frac{\hat{q}_\delta(t)\cosh(\pi t(1-2\ell/m))}{(1+2it)\cosh(\pi t)}\biggr) \ll \frac{1}{(1+|t|)(1+|\delta t|^k)}\biggl(\frac{1}{1+|t|^2}+\delta^2\biggr)$$

where the implied constant depends on m and ℓ. Hence we get

$$J \ll \frac{e^{\rho/2}}{\rho^2}\int_{\mathbb{R}} \frac{1}{(1+|t|)(1+|\delta t|^k)}\biggl(\frac{1}{1+|t|^2}+\delta^2\biggr)\, dt \ll \frac{e^{\rho/2}}{\rho^2}.$$

The terms associated to the other exponentials in (3-7) are treated similarly, and are bounded by the same quantity. □

Lemma 3.7. *Let $\rho > 1$, $0 < \delta < \frac{1}{4}$, and let $g_\pm$ and $h_\pm$ as in* (2-7) *and* (2-9). *Then*

$$g_\pm(0)\log 2 + \frac{1}{2\pi}\int_{-\infty}^{+\infty} h_\pm(t)\psi(1+it)\, dt \ll \delta e^{\rho/2} + \log(\delta^{-1}).$$

Proof. We use here the formula given in [Iwaniec 1995, (10.17)] to get back to an integral involving $g_\pm$. We have

$$g_\pm(0)\log 2 + \frac{1}{2\pi}\int_{-\infty}^{+\infty} h_\pm(t)\psi(1+it)\, dt = \frac{h_\pm(0)}{4} - \gamma g_\pm(0) + \int_0^\infty \log\bigl(\sinh\bigl(\tfrac{1}{2}x\bigr)\bigr)\, dg_\pm(x). \quad (3\text{-}10)$$

By (3-7) we know that

$$\frac{h_\pm(0)}{4} = e^{\rho/2} + O(\delta e^{\rho/2} + 1), \tag{3-11}$$

and by definition of $g_\pm(x)$ we have $g_\pm(x) = O(\delta)$ for $|x| \le \delta$. The last integral in (3-10) is analyzed as follows. For $x \in [0, \delta]$ we bound

$$\frac{d}{dx} g_\pm(x) = (g_{\rho\pm\delta} * q'_\delta)(x) \ll \int_{-\delta}^{\delta} \sinh\bigl(\tfrac{1}{2}|x-y|\bigr)|q'_\delta(y)|\,dy \ll \tfrac{1}{\delta}\sinh(\delta)\|q'\|_1 \ll 1,$$

so that we have, uniformly in δ,

$$\int_0^\delta \log\bigl(\sinh\bigl(\tfrac{1}{2}x\bigr)\bigr)\,dg_\pm(x) \ll \int_0^\delta \bigl|\log\bigl(\sinh\bigl(\tfrac{1}{2}x\bigr)\bigr)\bigr|\,dx \ll 1.$$

For $x > \delta$ we integrate by parts obtaining

$$\int_\delta^\infty \log\bigl(\sinh\bigl(\tfrac{1}{2}x\bigr)\bigr)\,dg_\pm(x) = -g_\pm(\delta)\log\bigl(\sinh\bigl(\tfrac{1}{2}\delta\bigr)\bigr) - \tfrac{1}{2}\int_\delta^\infty \frac{g_\pm(x)}{\sinh\bigl(\tfrac{1}{2}x\bigr)}\cosh\bigl(\tfrac{1}{2}x\bigr)\,dx.$$

Since $g_\pm(\delta) \ll \delta$, the boundary term is bounded by $O(\delta\log(\delta^{-1}))$. If we write $\cosh\bigl(\frac{x}{2}\bigr) = \sinh\bigl(\frac{x}{2}\bigr) + e^{-x/2}$, the integral associated to $e^{-x/2}$ can be bounded by

$$\int_\delta^\infty \frac{g_\pm(x)e^{-x/2}}{\sinh\bigl(\tfrac{1}{2}x\bigr)}\,dx \ll \int_\delta^1 \frac{dx}{x} + \int_1^\infty e^{-x/2}\,dx \ll \log(\delta^{-1}) + 1.$$

Finally we have

$$-\tfrac{1}{2}\int_\delta^\infty g_\pm(x)\,dx = -2\cosh\bigl(\tfrac{1}{2}(\rho\pm\delta)\bigr) + O(1) = -e^{\rho/2} + O(\delta e^{\rho/2} + 1).$$

The exponential cancels with the first term in (3-11), and combining the other estimates we obtain the claim. □

We turn now our attention to finding upper bounds for the mean square of the spectral side in the Selberg trace formula. We start with an estimate for the integral of $h_\pm(t_1)h_\pm(t_2)$.

Lemma 3.8. *Let $0 < \delta < \frac{1}{4}$, let $h_{\rho\pm\delta}(t)$ be as in* (2-9), *and let $w_R(\rho)$ be as in* (2-11). *Let $R > 1$, and $t_1, t_2 \in \mathbb{R}$. Then we have*

$$\int_{\mathbb{R}} h_{\rho\pm\delta}(t_1)h_{\rho\pm\delta}(t_2)w_R(\rho)\,d\rho \ll \frac{e^R v(t_1)v(t_2)}{1+|t_1-t_2|^2} + \frac{e^R v(t_1)v(t_2)}{1+|t_1+t_2|^2} + e^{R/2}v(t_1^2)v(t_2^2),$$

where $v(t) = (1+|t|)^{-1}$, and the implied constant does not depend on δ.

Proof. In order to express $h_{\rho\pm\delta}$ we consider again (3-7).

Multiplying $h_{\rho\pm\delta}(t_1)$ with $h_{\rho\pm\delta}(t_2)$, the product of the first exponential in (3-7) (from the factor $h_{\rho\pm\delta}(t_1)$) and the last term from the factor $h_{\rho\pm\delta}(t_2)$ contributes

$$\int_{\mathbb{R}} \frac{4e^{(\rho\pm\delta)(1/2+it_1)}}{(1+2it_1)\big(\frac{1}{4}+t_2^2\big)} w_R(\rho)\,d\rho = \int_{\mathbb{R}} \frac{2e^{(\rho\pm\delta)(1/2+it_1)} w_R'(\rho)}{\big(\frac{1}{2}+it_1\big)^2\big(\frac{1}{4}+t_2^2\big)}\,d\rho \ll e^{R/2} v(t_1^2) v(t_2^2)\|w'\|_1 \ll e^{R/2} v(t_1^2) v(t_2^2). \tag{3-12}$$

Now consider the product of the first exponential in (3-7) for $h_{\rho\pm\delta}(t_1)$ and the same term for $h_{\rho\pm\delta}(t_2)$. This contributes

$$\int_{\mathbb{R}} \frac{4e^{(\rho\pm\delta)(1+i(t_1+t_2))}}{(1+2it_1)(1+2it_2)} w_R(\rho) d\rho \ll e^R v(t_1) v(t_2) \min(1, |t_1+t_2|^{-2}\|w''\|_1) \ll \frac{e^R v(t_1) v(t_2)}{1+|t_1+t_2|^2}. \tag{3-13}$$

The other terms in the product $h_{\rho\pm\delta}(t_1) h_{\rho\pm\delta}(t_2)$ are bounded similarly by (3-12) and (3-13), except that we need to replace $|t_1+t_2|$ by $|t_1-t_2|$ when we integrate the product $\exp((\rho\pm\delta)(\pm 1 \pm i(t_1-t_2)))$. This concludes the proof. □

In order to exploit at best the bound proved in the previous lemma, we estimate the size of the spectrum on unit intervals.

Lemma 3.9. *Let Γ be a cofinite Fuchsian group, let $\varphi(s)$ be the scattering determinant associated to Γ, and let $T > 1$. We have*

$$\sharp\{T \le t_j \le T+1\} + \int_{T\le|t|\le T+1} \left| \frac{-\varphi'}{\varphi}\big(\tfrac{1}{2}+it\big) \right| dt \ll T.$$

Proof. Recall Weyl's law in its strong form (see [Venkov 1990, Theorem 7.3])

$$\sharp\{t_j \le T\} + \frac{1}{4\pi}\int_{-T}^{T} \frac{-\varphi'}{\varphi}\big(\tfrac{1}{2}+it\big)\,dt = \frac{\operatorname{vol}(\Gamma\backslash\mathbb{H})}{4\pi} T^2 + \frac{\mathfrak{C}}{\pi} T\log T + O(T),$$

where we recall that $\mathfrak{C}$ is the number of inequivalent cusps of Γ. Consider the equation above at the point $T+1$, and subtract from it the same quantity for T. In order to shorten notation we write $f(t) = -(\varphi'/\varphi)\big(\frac{1}{2}+it\big)$. We get

$$\sharp\{T \le t_j \le T+1\} + \frac{1}{4\pi}\int_{-T-1}^{-T} f(t)\,dt + \frac{1}{4\pi}\int_{T}^{T+1} f(t)\,dt = O(T). \tag{3-14}$$

The function $f(t)$ is bounded from below by a constant k that depends on the group (this follows from the Maass–Selberg relations, see [Iwaniec 1995, (10.9)]). Hence we have

$$\int_{-T-1}^{-T} f(t)\,dt \ge k.$$

Since the number $\sharp\{T \le t_j \le T+1\}$ is nonnegative, we can write

$$\int_{T}^{T+1} f(t)\,dt \le \int_{T}^{T+1} f(t)\,dt + \int_{-T-1}^{-T} f(t)\,dt - k + 4\pi\cdot\sharp\{T \le t_j \le T+1\} = O(T) - k = O(T).$$

Again from the fact that $f(t)$ is bounded from below by a constant, we infer that in fact we have

$$\int_T^{T+1} |f(t)|\,dt \ll T.$$

Similarly we find

$$\int_{-T-1}^{-T} |f(t)|\,dt \ll T,$$

and from (3-14) we conclude that we also have $\sharp\{T \le t_j \le T+1\} \ll T$. This proves the lemma. $\square$

At this point we can prove bounds on the mean square of the discrete and continuous spectrum in the Selberg trace formula. We discuss first the discrete spectrum.

Proposition 3.10. *Let* $0 < \delta < \frac{1}{4}$, *let* $h_\pm$ *as in* (2-9), *and let* $R > 1$. *We have*

$$\int_{\mathbb{R}} \bigg|\sum_{t_j>0} h_\pm(t_j)\bigg|^2 w_R(\rho)d\rho \ll \frac{e^R}{\delta} + e^{R/2}\log^2(\delta^{-1}).$$

Proof. Recall that $h_\pm(t) = h_{\rho\pm\delta(t)}\hat{q}_\delta(t)$, and that $\hat{q}_\delta(t) \ll (1+|\delta t|^k)^{-1}$ for every $k \ge 0$. Using Lemma 3.9 this implies that the series

$$\sum_{t_j>0} h_\pm(t_j)$$

is absolutely convergent, and we can write

$$\int_{\mathbb{R}} \bigg|\sum_{t_j>0} h_\pm(t_j)\bigg|^2 w_R(\rho)\,d\rho = \sum_{t_j>0}\sum_{t_\ell>0} \hat{q}_\delta(t_j)\hat{q}_\delta(t_\ell) \int_{\mathbb{R}} h_{\rho\pm\delta}(t_j)h_{\rho\pm\delta}(t_\ell)w_R(\rho)\,d\rho.$$

By Lemma 3.8 we can estimate the integral and bound the double sum by

$$e^R \sum_{t_j,t_\ell>0} \frac{|\hat{q}_\delta(t_j)\hat{q}_\delta(t_\ell)|v(t_1)v(t_2)}{1+|t_j-t_\ell|^2} + e^{R/2} \sum_{t_j,t_\ell>0} |\hat{q}_\delta(t_j)\hat{q}_\delta(t_\ell)|v(t_1^2)v(t_2^2), \tag{3-15}$$

where $v(t) = (1+|t|)^{-1}$. Using Lemma 3.1 to bound $\hat{q}_\delta(t) \ll (1+|\delta t|^k)^{-1}$ for every $k \ge 0$, we can estimate the second sum in (3-15) by

$$e^{R/2}\bigg(\sum_{t_j\le\delta^{-1}} \frac{1}{t_j^2} + \sum_{t_j>\delta^{-1}} \frac{1}{\delta t_j^3}\bigg)^2 \ll e^{R/2}(\log^2(\delta^{-1})+1).$$

Now consider the first sum in (3-15). By symmetry and positivity, we can consider only the sum over $t_\ell \ge t_j$. Moreover we split the sum in order to optimize the bounds available. Consider a unit neighborhood of the diagonal $t_\ell = t_j$. Using Lemma 3.9 we can estimate

$$e^R \sum_{t_j>0}\sum_{t_j\le t_\ell\le t_j+1} \frac{|\hat{q}_\delta(t_j)\hat{q}_\delta(t_\ell)|}{(1+|t_j|)(1+|t_\ell|)} \ll e^R \sum_{t_j\le\delta^{-1}} \frac{1}{t_j} \sum_{t_j\le t_\ell\le t_j+1} \frac{1}{t_\ell} + \frac{e^R}{\delta^2} \sum_{t_j>\delta^{-1}} \frac{1}{t_j^2} \sum_{t_j\le t_\ell\le t_j+1} \frac{1}{t_\ell^2} \ll \frac{e^R}{\delta}.$$

The tail of the double sum, that is, the range $t_j > \delta^{-1}$ and $t_\ell > t_j + 1$, can be analyzed (we follow here the same method as in Cramér [1922]) by using a unit interval decomposition for the sum over t_ℓ, together with Lemma 3.9, to get

$$\frac{e^R}{\delta^2}\sum_{t_j>\delta^{-1}}\frac{1}{t_j^2}\sum_{t_\ell>t_j+1}\frac{1}{t_\ell^2|t_\ell-t_j|^2} \ll \frac{e^R}{\delta^2}\sum_{t_j>\delta^{-1}}\frac{1}{t_j^2}\sum_{k=1}^{\infty}\frac{1}{(t_j+k)k^2} \ll \frac{e^R}{\delta^2}\sum_{t_j>\delta^{-1}}\frac{1}{t_j^3} \ll \frac{e^R}{\delta}.$$

Finally, the range $t_j \le \delta^{-1}$ and $t_\ell > t_j + 1$ is bounded by

$$\ll e^R \sum_{t_j\le\delta^{-1}}\frac{1}{t_j}\sum_{t_\ell>t_j+1}\frac{1}{t_\ell|t_\ell-t_j|^2} \ll e^R \sum_{t_j\le\delta^{-1}}\frac{1}{t_j}\sum_{k=1}^{\infty}\frac{1}{k^2} \ll \frac{e^R}{\delta}.$$

We conclude that we have

$$\int_{\mathbb{R}}\biggl|\sum_{t_j>0} h_\pm(t_j)\biggr|^2 w_R(\rho)\,d\rho \ll \frac{e^R}{\delta} + e^{R/2}\log^2(\delta^{-1}),$$

as claimed. □

The analysis of the continuous spectrum is similar, and we obtain the same bounds. With the proposition below we conclude the list of auxiliary results needed to prove Theorem 1.1.

Proposition 3.11. *Let* $0 < \delta < \frac{1}{4}$ *and* $h_\pm$ *be as in* (2-9)*. Let* $R \ge 1$*. Then*

$$\int_{\mathbb{R}}\biggl|\int_{\mathbb{R}} h_\pm(t)\frac{-\varphi'}{\varphi}\bigl(\tfrac{1}{2}+it\bigr)\biggr|^2 w_R(\rho)\,d\rho \ll \frac{e^R}{\delta} + e^{R/2}\log^2(\delta^{-1}). \tag{3-16}$$

Proof. Recall that we have $h_\pm(t) = h_{\rho\pm\delta}(t)\hat{q}_\delta(t)$, and that $\hat{q}_\delta(t) \ll (1+|\delta t|^k)^{-1}$ for every $k \ge 0$. For simplicity we write

$$f(t) = \frac{-\varphi'}{\varphi}\bigl(\tfrac{1}{2}+it\bigr).$$

Let J denote the integral in (3-16). Due to the decay properties of $\hat{q}_\delta$ and to Lemma 3.9, J is absolutely convergent, and so we can write

$$J = \int_{\mathbb{R}}\int_{\mathbb{R}} f(t_1)\overline{f(t_2)}\hat{q}_\delta(t_1)\hat{q}_\delta(t_2)\int_{\mathbb{R}} h_{\rho\pm\delta}(t_1)h_{\rho\pm\delta}(t_2)w_R(\rho)\,d\rho dt_1 dt_2.$$

The innermost integral is bounded using Lemma 3.8. This gives

$$J \ll e^R \int_{\mathbb{R}}\int_{\mathbb{R}}\frac{|f(t_1)f(t_2)\hat{q}_\delta(t_2)\hat{q}_\delta(t_2)|}{(1+|t_1|)(1+|t_2|)(1+|t_1-t_2|^2)}\,dt_1dt_2 + e^{R/2}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{|f(t_1)f(t_2)\hat{q}_\delta(t_2)\hat{q}_\delta(t_2)|}{(1+t_1^2)(1+t_2^2)}\,dt_1dt_2. \tag{3-17}$$

The second integral in (3-17) is bounded by

$$\biggl(\int_0^{\delta^{-1}}\frac{|f(t)|}{1+t^2}\,dt + \int_{\delta^{-1}}^{\infty}\frac{|f(t)|}{\delta t^3}\,dt\biggr)^2 \ll (\log(\delta^{-1})+1)^2 \ll \log^2(\delta^{-1}),$$

where in the first inequality we have used a unit interval decomposition of the domain of integration, and Lemma 3.9 to bound the integral of $|f(t)|$ in unit intervals. Now consider the first integral in (3-17). By symmetry and positivity we can consider only the integral over $t_2 \geq t_1 \geq 0$. A unit neighborhood of the diagonal $t_1 = t_2$ gives

$$\int_0^{\delta^{-1}} \frac{|f(t_1)|}{1+t_1} \int_{t_1}^{t_1+1} \frac{|f(t_2)|}{1+t_2}\, dt_2 dt_1 + \int_{\delta^{-1}}^{\infty} \frac{|f(t_1)|}{\delta^2 t_1^3} \int_{t_1}^{t_1+1} \frac{|f(t_2)|}{\delta^2 t_2^3}\, dt_2 dt_1$$
$$\ll \int_0^{\delta^{-1}} \frac{|f(t_1)|}{1+t_1}\, dt_1 + \int_{\delta^{-1}}^{\infty} \frac{|f(t_1)|}{\delta^4 t_1^4}\, dt_1 \ll \frac{1}{\delta}.$$

The tail of the double integral, that is, the range $t_1 \geq \delta^{-1}$ and $t_2 \geq t_1 + 1$, can be bounded as follows:

$$\int_{\delta^{-1}}^{\infty} \frac{|f(t_1)|}{\delta^2 t_1^3} \int_{t_1+1}^{\infty} \frac{|f(t_2)|}{\delta^2 t_2^3 |t_2 - t_1|^2}\, dt_2 dt_1 \ll \int_{\delta^{-1}}^{\infty} \frac{|f(t_1)|}{\delta^4 t_1^5}\, dt_1 \ll 1.$$

The range $t_1 \leq \delta^{-1}$ and $t_2 \geq t_1 + 1$ contributes

$$\int_0^{\delta^{-1}} \frac{|f(t_1)|}{1+t_1} \int_{t_1+1}^{\infty} \frac{|f(t_2)|}{(1+t_2)(1+\delta^2 t_2^2)|t_2 - t_1|^2}\, dt_2 dt_1 \ll \int_0^{\delta^{-1}} \frac{|f(t_1)|}{1+t_1}\, dt_1 \ll \delta^{-1}.$$

Summarizing, we have showed that we have the bound

$$J \ll \frac{e^R}{\delta} + e^{R/2} \log^2(\delta^{-1}),$$

which is what we wanted. This proves the proposition. □

4. Modular group

This section is devoted to the proof of Theorem 1.4, concerning the case $G = \mathrm{PSL}_2(\mathbb{Z})$. Our approach starts with a lemma of Iwaniec [1984b, Lemma 1], which gives

$$\psi_G(X) = X + 2\Re\biggl(\sum_{t_j \leq T} \frac{X^{1/2+it_j}}{\frac{1}{2} + it_j}\biggr) + O\biggl(\frac{X}{T} \log^2 X\biggr),$$

where $1 \leq T \leq X^{1/2}(\log X)^{-2}$ (and it is understood that the sum runs over $t_j > 0$). Note that in this case the only small eigenvalue of Δ is $\lambda = 0$, so that $M_G(X) = X$ and $P_G(X) = \psi_G(X) - X$. From the equation above we deduce that

$$\frac{1}{A} \int_A^{2A} |P_G(X)|^2\, dX \ll \frac{1}{A} \int_A^{2A} \biggl| \sum_{t_j \leq T} \frac{X^{1/2+it_j}}{\frac{1}{2} + it_j} \biggr|^2 dX + \frac{A^2}{T^2} \log^2 A. \tag{4-1}$$

We now describe the outline of the proof of Theorem 1.4. The main idea is to use the Kuznetsov trace formula to estimate the mean square of

$$\sum_{t_j \leq T} \frac{X^{1/2+it_j}}{\frac{1}{2} + it_j}$$

or, equivalently by partial summation, the mean square of

$$R(X,T)=\sum_{t_j\le T}X^{it_j}. \tag{4-2}$$

Therefore, we start by setting up the Kuznetsov trace formula with a test function that gives us a smooth version of the sum in (4-2), and we estimate the mean square of the weighted sum of Kloosterman sums that show up in its geometric side (Lemma 4.1). This allows us to bound the spectral side of the trace formula (Lemma 4.2). In order to control the behavior of the Fourier coefficients in the spectral sums of the trace formula we use a smoothed average of these sums. This way we obtain (Lemma 4.4) a mean square estimate of a smoothed version of $R(X,T)$, from which we extract a mean square estimate for the sharp sum (Proposition 4.5).

We now start by setting up the Kuznetsov trace formula. Let $\phi(x)$ be a smooth function on $[0,\infty]$ such that

$$\begin{aligned}|\phi(x)|&\ll x, && x\to 0,\\ |\phi^{(l)}(x)|&\ll x^{-3}, && x\to\infty,\end{aligned}$$

for $l=0,1,2,3$. Define

$$\begin{aligned}\phi_0&=\frac{1}{2\pi}\int_0^\infty J_0(y)\phi(y)\,dy,\\ \phi_B(x)&=\int_0^1\int_0^\infty \xi x J_0(\xi x)J_0(\xi y)\phi(y)\,dyd\xi,\\ \phi_H(x)&=\int_1^\infty\int_0^\infty \xi x J_0(\xi x)J_0(\xi y)\phi(y)\,dyd\xi,\\ \hat{\phi}(t)&=\frac{\pi}{2i\sinh\pi t}\int_0^\infty (J_{2it}(x)-J_{-2it}(x))\phi(x)\,\frac{dx}{x},\end{aligned}$$

where J_ν is the Bessel function of the first kind and order ν. By the properties of the Hankel transform we have

$$\phi(x)=\phi_B(x)+\phi_H(x).$$

We now choose ϕ as in [Luo and Sarnak 1995]. For $X,T>1$ we set

$$\begin{aligned}\phi_{X,T}(x)&=\frac{-\sinh\beta}{\pi}x\exp(ix\cosh\beta),\\ 2\beta&=\log X+\frac{i}{T},\end{aligned}$$

and apply the Kuznetsov trace formula [1980, Theorem 1] with $\phi_{X,T}$ as the test function. Let $\{f_i\}_{i=1}^\infty$ be an orthonormal basis of Maass cusp forms for $\mathrm{SL}_2(\mathbb{Z})$, with eigenvalues $\lambda_j=\frac{1}{4}+t_j^2$. These cusp forms have Fourier expansions [Kuznetsov 1980, (2.10)]

$$f_j(z)=\sqrt{y}\sum_{n=1}^\infty \rho_j(n)K_{it_j}(2\pi ny)\cos(2\pi nx),$$

where $K_\nu(x)$ is the K-Bessel function. For $l_1, l_2 \geq 1$ the trace formula reads

$$\sum_{t_j} \hat{\phi}(t_j)\nu_j(l_1)\overline{\nu_j(l_2)} + \frac{2}{\pi}\int_0^\infty \frac{\hat{\phi}(t)}{|\zeta(1+2it)|^2}\, d_{it}(l_1)\, d_{it}(l_2)dt = \delta_{l_1,l_2}\phi_0 + \sum_{c=1}^{\infty} \frac{S(l_1,l_2,c)}{c}\phi_H\left(\frac{4\pi\sqrt{l_1 l_2}}{c}\right),$$

where $\rho_j(l) = \nu_j(l)\cosh(\pi t_j)^{1/2}$, $d_{it}(l) = \sum_{d_1 d_2 = l}(d_1/d_2)^{it}$, and $S(l_1, l_2, c)$ is the classical Kloostermann sum. By [Luo and Sarnak 1995, p. 234] we have

$$\hat{\phi}_{X,T}(t_j) = X^{it_j}e^{-t_j/T} + O(e^{-\pi t_j}), \tag{4-3}$$

$$(\phi_{X,T})_0 \ll X^{-1/2}, \tag{4-4}$$

$$\frac{2}{\pi}\int_0^\infty \frac{\hat{\phi}_{X,T}(t)}{|\zeta(1+2it)|^2}\,(d_{it}(n))^2 dt \ll T\log^2 T\, d^2(n), \tag{4-5}$$

$$S_n((\phi_{X,T})_B) \ll n^{1/2}X^{-1/2}\log^2 n, \tag{4-6}$$

where

$$S_n(\psi) = \sum_{c=1}^{\infty}\frac{S(n,n,c)}{c}\psi\left(\frac{4\pi n}{c}\right). \tag{4-7}$$

Analyzing the right hand side of the Kuznetsov trace formula, we prove a bound for the mean square of $S_n(\phi_{X,T})$ as follows.

Lemma 4.1. *Let $A, T > 2$ and let n be a positive integer. Then, for any $\varepsilon > 0$,*

$$\frac{1}{A}\int_A^{2A} |S_n(\phi_{X,T})|^2\, dX \ll_\varepsilon (nA^{1/2} + T^2)(An)^\varepsilon.$$

We postpone the proof of Lemma 4.1 to the end of the section, and we show here how to recover a similar bound on the spectral side of the Kuznetsov trace formula.

Lemma 4.2. *Let $A, T > 2$ and let n be a positive integer. Then, for any $\varepsilon > 0$,*

$$\frac{1}{A}\int_A^{2A}\left|\sum_{t_j}|\nu_j(n)|^2\hat{\phi}(t_j)\right|^2 dX \ll_\varepsilon (nA^{1/2} + T^2)(AnT)^\varepsilon.$$

Proof. By the trace formula and the bounds in equations (4-4)–(4-6) we deduce

$$\frac{1}{A}\int_A^{2A}\left|\sum_{t_j}|\nu_j(n)|^2\hat{\phi}(t_j)\right|^2 dX \ll \frac{1}{A}\int_A^{2A}|S_n(\phi_{X,T})|^2\, dX + T^2\log^4 T\, d^4(n) + \frac{n\log A\log^4 n}{A}.$$

Observing that $d^4(n) \ll n^\varepsilon$, the claim follows from Lemma 4.1. □

Once we have bounds for a fixed n, we average over $n \in [N, 2N]$. Let $h(\xi)$ be a smooth function supported in $[N, 2N]$, whose derivatives satisfy

$$|h^{(p)}(\xi)| \ll N^{-p} \quad \text{for } p = 0, 1, 2, \ldots$$

and such that

$$\int_{-\infty}^{+\infty} h(\xi)\,d\xi = N.$$

The next lemma is a result of Luo and Sarnak [1995].

Lemma 4.3. *Let h be as above. Then*

$$\sum_n h(n)|\nu_j(n)|^2 = \frac{12}{\pi^2}N + r(t_j, N),$$

$$\sum_{t_j \le T} |r(t_j, N)| \ll T^2 N^{1/2} \log^2 T.$$

Proof. See [Luo and Sarnak 1995, p. 233]. □

Lemma 4.4. *Let $A, T > 2$. Then, for any $\varepsilon > 0$,*

$$\frac{1}{A}\int_A^{2A} \Bigl|\sum_{t_j} X^{it_j} e^{-t_j/T}\Bigr|^2 dX \ll_\varepsilon A^{1/4} T^2 (AT)^\varepsilon.$$

Proof. By the previous lemma

$$\begin{aligned}\frac{1}{N}\sum_n h(n)\Bigl(\sum_{t_j}|\nu_j(n)|^2\hat{\phi}(t_j)\Bigr) &= \frac{1}{N}\sum_{t_j}\Bigl(\sum_n|\nu_j(n)|^2 h(n)\Bigr)\hat{\phi}(t_j)\\ &= \frac{12}{\pi^2}\sum_{t_j}\hat{\phi}(t_j) + \frac{1}{N}\sum_{t_j} r(t_j, N)\hat{\phi}(t_j)\\ &= \frac{12}{\pi^2}\sum_{t_j}\hat{\phi}(t_j) + O(T^2 N^{-1/2}\log^2 T), \qquad (4\text{-}8)\end{aligned}$$

where the last step can be obtained by $|\hat{\phi}(t_j)| \ll e^{-t_j/T}$ and partial summation. Note now that from (4-3) it follows that

$$\sum_{t_j}\hat{\phi}(t_j) = \sum_{t_j} X^{it_j} e^{-t_j/T} + O(1). \qquad (4\text{-}9)$$

Combining (4-8) and (4-9) we deduce that

$$\sum_{t_j} X^{it_j} e^{-t_j/T} \ll \frac{1}{N}\sum_n h(n)\Bigl(\sum_{t_j}|\nu_j(n)|^2\hat{\phi}(t_j)\Bigr) + T^2 N^{-1/2}\log^2 T.$$

Taking absolute value, squaring, and integrating over X, we infer that

$$\frac{1}{A}\int_A^{2A}\Bigl|\sum_{t_j} X^{it_j} e^{-t_j/T}\Bigr|^2 dX \ll \frac{1}{A}\int_A^{2A}\Bigl|\frac{1}{N}\sum_n h(n)\Bigl(\sum_{t_j}|\nu_j(n)|^2\hat{\phi}(t_j)\Bigr)\Bigr|^2 dX + T^4 N^{-1}\log^4 T.$$

Moreover, by Cauchy–Schwarz and the properties of h, we have

$$\left|\frac{1}{N}\sum_{n} h(n)\biggl(\sum_{t_j}|\nu_j(n)|^2\hat{\phi}(t_j)\biggr)\right|^2 \ll \frac{1}{N^2}\biggl(\sum_{n=N}^{2N}|h(n)|^2\biggr)\biggl(\sum_{n=N}^{2N}\biggl|\sum_{t_j}|\nu_j(n)|^2\hat{\phi}(t_j)\biggr|^2\biggr)$$
$$\ll \frac{1}{N}\sum_{n=N}^{2N}\biggl|\sum_{t_j}|\nu_j(n)|^2\hat{\phi}(t_j)\biggr|^2$$

Finally, we obtain

$$\frac{1}{A}\int_A^{2A}\biggl|\sum_{t_j}X^{it_j}e^{-t_j/T}\biggr|^2\,dX \ll_\varepsilon \frac{1}{N}\sum_{n=N}^{2N}\biggl(\frac{1}{A}\int_A^{2A}\biggl|\sum_{t_j}|\nu_j(n)|^2\hat{\phi}(t_j)\biggr|^2\,dX\biggr)+\frac{T^{4+\varepsilon}}{N}$$
$$\ll_\varepsilon (NA^{1/2}+T^2)(ANT)^\varepsilon+\frac{T^{4+\varepsilon}}{N},$$

where the last inequality follows from Lemma 4.2. Choose $N = A^{-1/4}T^2$ to complete the proof. □

The next step is to replace the smoothed sum $\sum X^{it_j}e^{-t_j/T}$ with the truncated one. This gives us a corresponding mean square estimate for $R(X,T)$.

Proposition 4.5. *Let $A, T > 2$, and let $R(X,T)$ be as in* (4-2). *Then, for any $\varepsilon > 0$,*

$$\frac{1}{A}\int_A^{2A}|R(X,T)|^2\,dX \ll_\varepsilon A^{1/4}T^2(AT)^\varepsilon.$$

Proof. We start by choosing a smooth function g that approximates the characteristic function of $[1,T]$. Let g be a smooth function supported on $\bigl[\frac{1}{2}, T+\frac{1}{2}\bigr]$ such that $0 \le g(\xi) \le 1$, and $g(\xi) = 1$ when $\xi \in [1,T]$. By the strong Weyl's law we know that $\bigl|\{t_j : U \le t_j \le U+1\}\bigr| \ll U$, and so we have

$$R(X,T) = \sum_{t_j} g(t_j)X^{it_j} + O(T).$$

Define $\hat{g}(\xi)$ to be the Fourier transform of $g(\xi)\exp(\xi/T)$. By [Luo and Sarnak 1995, pp. 235–236] we have that

$$\sum_{t_j} g(t_j)X^{it_j}$$
$$= \int_{-1}^{1}\hat{g}(\xi)\biggl(\sum_{t_j}(Xe^{-2\pi\xi})^{it_j}e^{-t_j/T}\biggr)d\xi + O\biggl(\sum_{t_j}\frac{e^{-t_j/T}}{t_j}+\frac{e^{-t_j/T}}{|T-t_j|+1}+\frac{\log(T+t_j)e^{-t_j/T}}{T}\biggr),$$

where the error term can be bounded by $O(T\log T)$. Also note the estimate

$$\hat{g}(\xi) \ll \min\biggl(T, \frac{1}{|x|}\biggr). \tag{4-10}$$

Defining

$$k(X,T,\xi) := \sum_{t_j}(Xe^{-2\pi\xi})^{it_j}e^{-t_j/T},$$

we deduce that

$$\begin{aligned}
&\frac{1}{A}\int_A^{2A}|R(X,T)|^2\,dX\\
&\qquad\ll \frac{1}{A}\int_A^{2A}\left|\int_{-1}^{1}\hat{g}(\xi)k(X,T,\xi)\,d\xi\right|^2dX+T^2\log^2 T\\
&\qquad\ll \frac{1}{A}\int_A^{2A}\left|\int_{|\xi|\le\delta}\hat{g}(\xi)k(X,T,\xi)\,d\xi\right|^2dX+\frac{1}{A}\int_A^{2A}\left|\int_{\delta<|\xi|\le 1}\hat{g}(\xi)k(X,T,\xi)d\xi\right|^2dX+T^2\log^2 T.
\end{aligned}$$

The first term in the sum above we bound, using Lemma 4.4 and the first bound in (4-10), by

$$\int_{|\xi|\le\delta}|\hat{g}(\xi)|^2\int_{|\xi|\le\delta}\frac{1}{A}\int_A^{2A}|k(X,T,\xi)|^2\,dXd\xi \ll_\varepsilon A^{1/4}T^4(AT)^\varepsilon\delta^2,$$

and the second term, using again Lemma 4.4 and the second bound in (4-10), by

$$\int_{\delta<|\xi|\le 1}|\hat{g}(\xi)|^2|\xi|\int_{\delta<|\xi|\le 1}\frac{1}{|\xi|}\frac{1}{A}\int_A^{2A}|k(X,T,\xi)|^2\,dXd\xi \ll_\varepsilon A^{1/4}T^2(AT)^\varepsilon\log(\delta^{-1}).$$

On taking $\delta=T^{-2}$ we obtain the claim. □

We are now able to conclude the proof of Theorem 1.4. Let $2<T\le A^{1/2}(\log A)^{-2}$. By partial summation,

$$\sum_{t_j\le T}\frac{X^{1/2+it_j}}{\frac{1}{2}+it_j}=\frac{R(X,T)X^{1/2}}{\frac{1}{2}+iT}+iX^{1/2}\int_1^T\frac{R(X,U)}{\left(\frac{1}{2}+iU\right)^2}\,dU.$$

Therefore,

$$\frac{1}{A}\int_A^{2A}\left|\sum_{t_j\le T}\frac{X^{1/2+it_j}}{\frac{1}{2}+it_j}\right|^2dX\ll\frac{1}{A}\int_A^{2A}\left|\frac{R(X,T)X^{1/2}}{\frac{1}{2}+iT}\right|^2dX+\frac{1}{A}\int_A^{2A}\left|X^{1/2}\int_1^T\frac{R(X,U)}{\left(\frac{1}{2}+iU\right)^2}\,dU\right|^2dX.$$

The first term on the right hand side is bounded, in view of Proposition 4.5, by $O(A^{5/4+\varepsilon})$. The second term can be bounded by using Cauchy–Schwarz inequality and again Proposition 4.5, giving

$$A(\log T)\int_1^T\frac{1}{U^3}\left(\frac{1}{A}\int_A^{2A}|R(X,U)|^2\,dX\right)dU\ll_\varepsilon A^{5/4+\varepsilon}\int_1^T U^{-1}\,dU\ll_\varepsilon A^{5/4+\varepsilon}.$$

Inserting these bounds in (4-1) we obtain

$$\frac{1}{A}\int_A^{2A}|\psi_G(X)-X|^2\,dX\ll_\varepsilon A^{5/4+\varepsilon}+\frac{A^2}{T^2}\log^2 A.$$

Choosing $T=A^{3/8}$ above concludes the proof of Theorem 1.4.

Weighted sums of Kloosterman sums. In the remainder of the section we prove Lemma 4.1. Recall that we want to estimate, on average, the following sum of Kloostermann sums

$$\frac{1}{A}\int_A^{2A} |S_n(\phi_{X,T})|^2\, dX,$$

where $S_n(\psi)$ is the weighted sum of Kloosterman sums defined in (4-7), i.e.,

$$S_n(\psi) = \sum_{c=1}^{\infty} \frac{S(n,n,c)}{c}\psi\left(\frac{4\pi n}{c}\right),$$

with n a (large) positive integer. The function $\phi_{X,T}$ carries some oscillation in X, that we exploit when integrating over $X \in [A, 2A]$.

Lemma 4.6. *Let $A, T > 2$, and let z_1, z_2 be positive real numbers. Then*

$$\frac{1}{A}\int_A^{2A} \phi_{X,T}(z_1)\overline{\phi_{X,T}(z_2)}\, dX \ll z_1 z_2 A \exp\left(-\frac{A^{1/2}(z_1+z_2)}{T}\right)$$

and

$$\frac{1}{A}\int_A^{2A} \phi_{X,T}(z_1)\overline{\phi_{X,T}(z_2)}\, dX \ll \frac{z_1 z_2 A^{1/2}}{|z_1 - z_2|}.$$

For the last bound we assume $z_1 \neq z_2$.

Proof. Inserting the definition of $\phi_{X,T}$ we can write the integral as

$$\frac{z_1 z_2}{\pi^2}\int_A^{2A} |\sinh\beta|^2 \exp(i\cosh\beta z_1 + \overline{i\cosh\beta z_2})\, dX. \tag{4-11}$$

Bounding the integrand uniformly using

$$|\sinh\beta|^2 \ll X \quad \text{and} \quad \exp(i\cosh\beta z_1 + \overline{i\cosh\beta z_2}) \ll \exp\left(-\frac{A^{1/2}(z_1+z_2)}{T}\right),$$

we obtain the first bound

$$\frac{1}{A}\int_A^{2A} \phi_{X,T}(z_1)\overline{\phi_{X,T}(z_2)}\, dX \ll z_1 z_2 A \exp\left(-\frac{A^{1/2}(z_1+z_2)}{T}\right).$$

To obtain the second bound we use integration by parts in (4-11) to get

$$\frac{z_1 z_2}{\pi^2} f(X)\exp(i\cosh\beta z_1 + \overline{i\cosh\beta z_2})\Big|_A^{2A} + \frac{z_1 z_2}{\pi^2}\int_A^{2A} f'(X)\exp(i\cosh\beta z_1 + \overline{i\cosh\beta z_2})\, dX,$$

where

$$f(X) = \frac{2X|\sinh\beta|^2}{i\sinh\beta z_1 + \overline{i\sinh\beta z_2}}.$$

We can bound these terms (up to a constant) by

$$\frac{z_1 z_2 A^{3/2}}{|z_1 - z_2|},$$

which gives us the second bound

$$\frac{1}{A}\int_A^{2A} \phi_{X,T}(z_1)\overline{\phi_{X,T}(z_2)}\,dX \ll \frac{z_1 z_2 A^{1/2}}{|z_1-z_2|}.$$

This proves the lemma. □

The lemma above will provide the necessary bounds to establish Lemma 4.1.

Proof of Lemma 4.1. We expand the square in the integrand and exchange the integral with the sums, so that we can rewrite the integral as

$$\sum_{c_1,c_2=1}^{\infty} \frac{S(n,n,c_1)\overline{S(n,n,c_2)}}{c_1c_2}\frac{1}{A}\int_A^{2A} \phi_{X,T}\Big(\frac{4\pi n}{c_1}\Big)\overline{\phi_{X,T}\Big(\frac{4\pi n}{c_2}\Big)}\,dX. \tag{4-12}$$

We now split the sum in (4-12) into two sums Σ_d and Σ_{nd}, where Σ_d is the sum over the diagonal terms $c_1=c_2$, and Σ_{nd} is the sum over the terms $c_1\neq c_2$. We shall make use of the Weil bound on Kloosterman sums throughout the proof, namely

$$|S(n,n,c)| \le (n,c)^{1/2}c^{1/2}d(c). \tag{4-13}$$

Moreover, we have

$$\sum_{c\le x}(n,c)\,d^2(c) \ll x\log^3 x\,d(n).$$

We bound the diagonal terms using the first bound in Lemma 4.6 (with $z_1=z_2=(4\pi n)/c$), obtaining

$$\Sigma_d \ll n^2A\sum_c \frac{(n,c)d^2(c)}{c^3}\exp\Big(-\frac{A^{1/2}n}{Tc}\Big) \ll T^2\log^3(An)\,d(n).$$

To bound the nondiagonal terms we interpolate the two bounds in Lemma 4.6 to get, for $0<\lambda<1$,

$$\frac{1}{A}\int_A^{2A} \phi_{X,T}\Big(\frac{4\pi n}{c_1}\Big)\overline{\phi_{X,T}\Big(\frac{4\pi n}{c_2}\Big)}\,dX \ll \Big(\frac{n^2A}{c_1c_2}\Big)^{1-\lambda}\Big(\frac{nA^{1/2}}{|c_1-c_2|}\Big)^{\lambda}.$$

Therefore,

$$\begin{aligned}\Sigma_{nd} &\ll \sum_{c_1\neq c_2=1}^{\infty}\frac{|S(n,n,c_1)S(n,n,c_2)|}{c_1c_2}\Big(\frac{n^2A}{c_1c_2}\Big)^{1-\lambda}\Big(\frac{nA^{1/2}}{|c_1-c_2|}\Big)^{\lambda}\\ &\ll_\lambda (nA^{1/2})^{2-\lambda}\Big(\sum_{c=1}^{\infty}\frac{|S(n,n,c)|^{2/(2-\lambda)}}{c^2}\Big)^{2-\lambda},\end{aligned}$$

where the last inequality follows from the Hardy–Littlewood–Pólya inequality (1-8). Applying the Weil bound (4-13) we obtain

$$\begin{aligned}\Sigma_{nd} &\ll_\lambda (nA^{1/2})^{2-\lambda}\biggl(\sum_{c=1}^{\infty}\frac{((n,c)^{1/2}d(c)\sqrt{c})^{2/(2-\lambda)}}{c^2}\biggr)^{2-\lambda}\\ &\ll_\lambda (nA^{1/2})^{2-\lambda}\biggl(\sum_{c=1}^{\infty}\frac{(n,c)d^2(c)}{c^{1+(1-\lambda)/(2-\lambda)}}\biggr)^{2-\lambda}\\ &\ll_\lambda (nA^{1/2}d(n))^{2-\lambda}.\end{aligned}$$

Pick $\lambda = 1-\varepsilon$ and note that $d(n) \ll_\varepsilon n^\varepsilon$ to finish the proof. □

References

[Avdispahić 2017] M. Avdispahić, "Prime geodesic theorem of Gallagher type", preprint, 2017. arXiv

[Avdispahić 2018] M. Avdispahić, "On Koyama's refinement of the prime geodesic theorem", *Proc. Japan Acad. Ser. A Math. Sci.* **94**:3 (2018), 21–24. MR

[Cai 2002] Y. Cai, "Prime geodesic theorem", *J. Théor. Nombres Bordeaux* **14**:1 (2002), 59–72. MR Zbl

[Carneiro and Vaaler 2010] E. Carneiro and J. D. Vaaler, "Some extremal functions in Fourier analysis, II", *Trans. Amer. Math. Soc.* **362**:11 (2010), 5803–5843. MR Zbl

[Cramér 1922] H. Cramér, "Ein Mittelwertsatz in der Primzahltheorie", *Math. Z.* **12**:1 (1922), 147–153. MR JFM

[Gallagher 1970] P. X. Gallagher, "A large sieve density estimate near $\sigma = 1$", *Invent. Math.* **11**:4 (1970), 329–339. MR Zbl

[Hardy et al. 1934] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge Univ. Press, 1934. JFM

[Hejhal 1983] D. A. Hejhal, *The Selberg trace formula for* PSL(2, $\mathbb{R}$)*, II*, Lecture Notes in Math. **1001**, Springer, 1983. MR Zbl

[Huber 1961] H. Huber, "Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen, II", *Math. Ann.* **142**:4 (1961), 385–398. Appendix in **143**:5 (1961), 463–464. MR Zbl

[Ivić 1985] A. Ivić, *The Riemann zeta-function*, Wiley, New York, 1985. MR Zbl

[Iwaniec 1984a] H. Iwaniec, "Nonholomorphic modular forms and their applications", pp. 157–196 in *Modular forms* (Durham, 1983), edited by R. A. Rankin, Horwood, Chichester, UK, 1984. MR Zbl

[Iwaniec 1984b] H. Iwaniec, "Prime geodesic theorem", *J. Reine Angew. Math.* **349** (1984), 136–159. MR Zbl

[Iwaniec 1995] H. Iwaniec, *Introduction to the spectral theory of automorphic forms*, Revista Matemática Iberoamericana, Madrid, 1995. MR Zbl

[Koyama 1998] S.-y. Koyama, "Prime geodesic theorem for arithmetic compact surfaces", *Int. Math. Res. Not.* **1998**:8 (1998), 383–388. MR Zbl

[Koyama 2016] S.-y. Koyama, "Refinement of prime geodesic theorem", *Proc. Japan Acad. Ser. A Math. Sci.* **92**:7 (2016), 77–81. MR Zbl

[Kuznetsov 1980] N. V. Kuznetsov, "Petersson's conjecture for cusp forms of weight zero and Linnik's conjecture: sums of Kloosterman sums", *Mat. Sb.* (*N.S.*) **111(153)**:3 (1980), 334–383. In Russian; translation in *Math. USSR-Sb.* **39**:3 (1981), 299–342. MR Zbl

[Li and Villavert 2011] C. Li and J. Villavert, "An extension of the Hardy–Littlewood–Pólya inequality", *Acta Math. Sci. Ser. B Engl. Ed.* **31**:6 (2011), 2285–2288. MR Zbl

[Luo and Sarnak 1995] W. Z. Luo and P. Sarnak, "Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}^2$", *Inst. Hautes Études Sci. Publ. Math.* **81** (1995), 207–237. MR Zbl

[Luo et al. 1995] W. Luo, Z. Rudnick, and P. Sarnak, "On Selberg's eigenvalue conjecture", *Geom. Funct. Anal.* **5**:2 (1995), 387–401. MR Zbl

[Petridis and Risager 2017] Y. N. Petridis and M. S. Risager, "Local average in hyperbolic lattice point counting", *Math. Z.* **285**:3-4 (2017), 1319–1344. With an appendix by N. Laaksonen. MR Zbl

[Phillips and Rudnick 1994] R. Phillips and Z. Rudnick, "The circle problem in the hyperbolic plane", *J. Funct. Anal.* **121**:1 (1994), 78–116. MR Zbl

[Rubinstein and Sarnak 1994] M. Rubinstein and P. Sarnak, "Chebyshev's bias", *Experiment. Math.* **3**:3 (1994), 173–197. MR Zbl

[Sarnak 1980] P. C. Sarnak, *Prime geodesic theorems*, Ph.D. thesis, Stanford University, 1980, Available at https://tinyurl.com/sarurl.

[Sarnak 1982] P. Sarnak, "Class numbers of indefinite binary quadratic forms", *J. Number Theory* **15**:2 (1982), 229–247. Correction in **16**:2 (1983), 284. MR Zbl

[Soundararajan and Young 2013] K. Soundararajan and M. P. Young, "The prime geodesic theorem", *J. Reine Angew. Math.* **676** (2013), 105–120. MR Zbl

[Venkov 1990] A. B. Venkov, *Spectral theory of automorphic functions and its applications*, Math. and Its Appl. (Soviet Series) **51**, Kluwer, Dordrecht, 1990. MR Zbl

[Wintner 1935] A. Wintner, "On the asymptotic distribution of the remainder term of the prime-number theorem", *Amer. J. Math.* **57**:3 (1935), 534–538. MR Zbl

Communicated by Peter Sarnak
Received 2017-05-23 Revised 2017-10-26 Accepted 2017-12-30

cherubini@mpim-bonn.mpg.de *Max-Planck-Institut für Mathematik, Bonn, Germany*

guerreiro@mpim-bonn.mpg.de *Max-Planck-Institut für Mathematik, Bonn, Germany*

ALGEBRA AND NUMBER THEORY 12:3 (2018)
dx.doi.org/10.2140/ant.2018.12.599

Elliptic quantum groups and Baxter relations

Huafeng Zhang

We introduce a category $\mathcal{O}$ of modules over the elliptic quantum group of $\mathfrak{sl}_N$ with well-behaved q-character theory. We construct asymptotic modules as analytic continuation of a family of finite-dimensional modules, the Kirillov–Reshetikhin modules. In the Grothendieck ring of this category we prove two types of identities: Generalized Baxter relations in the spirit of Frenkel–Hernandez between finite-dimensional modules and asymptotic modules. Three-term Baxter TQ relations of infinite-dimensional modules.

Introduction		599
1.	Elliptic quantum groups and their representations	603
2.	Small elliptic quantum group and evaluation modules	616
3.	Kirillov–Reshetikhin modules	622
4.	Asymptotic representations	626
5.	Baxter TQ relations	640
6.	Transfer matrices and Baxter operators	642
Acknowledgments		645
References		645

Introduction

Fix $\mathfrak{sl}_N$ a special linear Lie algebra, $\mathbb{C}/(\mathbb{Z}+\mathbb{Z}\tau)$ an elliptic curve, and $\hbar$ a complex number. Associated to this triple is the *elliptic quantum group* $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$ introduced by G. Felder [1995]. It is a Hopf algebroid (neither commutative nor cocommutative) in the sense of Etingof and Varchenko [1998], so that the tensor product of two $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$-modules is naturally endowed with a module structure. In this paper we study (finite- and infinite-dimensional) representations of the elliptic quantum group.

Suppose $\hbar$ is a formal variable. $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_2)$ is an $\hbar$-deformation [Enriquez and Felder 1998] of the universal enveloping algebra of a Lie algebra $\mathfrak{sl}_2 \otimes R_\tau$, where R_τ is an algebra of meromorphic functions of $z \in \mathbb{C}$ built from the Jacobi theta function of the elliptic curve. For $\mathfrak{g}$ an arbitrary finite-dimensional simple Lie algebra, $\mathcal{E}_{\tau,\hbar}(\mathfrak{g})$ is defined [Jimbo et al. 1999] to be a quasi-Hopf algebra twist of the *affine quantum group* $U_\hbar(L\mathfrak{g})$, an $\hbar$-deformation of the loop Lie algebra $\mathfrak{g}\otimes\mathbb{C}[z,z^{-1}]$. It admits a universal

MSC2010: primary 17B37; secondary 17B10, 17B80.
Keywords: elliptic quantum groups, asymptotic representations, Yang–Baxter equation.

dynamical R-matrix in a completed tensor square, which provides solutions $R(z;\lambda) \in \mathrm{End}(V \otimes V)$, for V a suitable $\mathcal{E}_{\tau,\hbar}(\mathfrak{g})$-module, to the *quantum dynamical Yang–Baxter equation*:

$$R_{12}(z-w;\lambda+\hbar h^{(3)})R_{13}(z;\lambda)R_{23}(w;\lambda+\hbar h^{(1)}) = R_{23}(w;\lambda)R_{13}(z;\lambda+\hbar h^{(2)})R_{12}(z-w;\lambda) \in \mathrm{End}(V^{\otimes 3}).$$

Here z, w are complex spectral parameters, λ is the dynamical parameter lying in a Cartan subalgebra of $\mathfrak{g}$, the subindices of R indicate the tensor factors of $V^{\otimes 3}$ to be acted on, and the $h^{(i)}$ are grading operators arising from the weight grading on V by the Cartan subalgebra. See the comments following (1.1).

Such R-matrices $R(z;\lambda)$ appeared previously in face-type integrable models [Felder and Varchenko 1996a; Hou et al. 2003]; for instance, the R-matrix of the Andrews–Baxter–Forrester model comes from two-dimensional irreducible modules of $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_2)$, as does the 6-vertex model from the affine quantum group $U_\hbar(L\mathfrak{sl}_2)$. The definition of $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$ in [Felder 1995], by RLL exchange relations, is in the spirit of Faddeev, Reshetikhin and Takhatajan, originated from *quantum inverse scattering method*. We mention that elliptic R-matrices describe the monodromy of the quantized Knizhnik–Zamolodchikov equation associated with representations of affine quantum groups, e.g., [Frenkel and Reshetikhin 1992; Galleas and Stokman 2015; Konno 2006; Tarasov and Varchenko 1997].

Recently Aganagic and Okounkov [2016] proposed the elliptic stable envelope in equivariant elliptic cohomology, as a geometric framework to obtain elliptic R-matrices. This was made explicit [Felder et al. 2017] for cotangent bundles of Grassmannians, resulting in tensor products of two-dimensional irreducible representations of $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_2)$. The higher rank case of $\mathfrak{sl}_N$ was studied later by H. Konno [2017].

Meanwhile, Nekrasov, Pestun and Shatashvili [2018] from the 6d quiver gauge theory predicted the elliptic quantum group associated to an arbitrary Kac–Moody algebra, the precise definition of which (as an associative algebra) was proposed by Gautam and Toledano Laredo [2017a]. See also [Yang and Zhao 2017] in the context of quiver geometry.

We are interested in the representation theory of $\mathcal{E}_{\tau,\hbar}(\mathfrak{g})$ with $\hbar \in \mathbb{C}$ generic. The formal twist constructions [Enriquez and Felder 1998; Jimbo et al. 1999] from $U_\hbar(L\mathfrak{g})$ might reduce the problem to the representation theory of affine quantum groups, which is a subject developed intensively in the last three decades from algebraic, geometric and combinatorial aspects. However their work involves formal power series of $\hbar$ and infinite products in the comultiplication of $\mathcal{E}_{\tau,\hbar}(\mathfrak{g})$. Some of these divergence issues were addressed by Etingof and Moura [2002], who defined a fully faithful tenor functor between representation categories of BGG type for $U_\hbar(L\mathfrak{sl}_N)$ and $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$. Towards this functor not much is known: its image, the induced homomorphism of Grothendieck rings, etc.

In this paper we study representations of $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$ via the RLL presentation [Felder 1995] so as to bypass affine quantum groups, yet along the way we borrow ideas from the affine case. Compared to other works [Cavalli 2001; Etingof and Moura 2002; Felder and Varchenko 1996b; Gautam and Toledano Laredo 2017a; Konno 2009; 2016; Tarasov and Varchenko 2001; Yang and Zhao 2017], our approach emphasizes more on the Grothendieck ring structure of representation category. It is a higher rank extension of a recent joint work with G. Felder [2017].

The presence of the dynamical parameter λ is one of the technical difficulties of elliptic quantum groups. To resolve this, we need a commuting family of elliptic Cartan currents $\phi_j(z) \in \mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$ for $j \in J := \{1, 2, \dots, N-1\}$. They act as difference operators on an $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$-module V, and their matrix entries are meromorphic functions of $(z, \lambda) \in \mathbb{C} \times \mathfrak{h}$ where $\mathfrak{h}$ denotes the Cartan subalgebra of $\mathfrak{sl}_N$. As in [Felder and Zhang 2017], we impose the following triangularity condition:[1]

(i) There exists a basis of V, with respect to which the matrices $\phi_j(z)$ are upper triangular and their diagonal entries are independent of λ.

Our category $\mathcal{O}$ is the full subcategory of category BGG [Etingof and Moura 2002] of $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$-modules subject to condition (i), see Definition 1.7. It is abelian and monoidal. It contains most of the modules in [Cavalli 2001; Etingof and Moura 2002; Konno 2009; 2016; Tarasov and Varchenko 2001], although the proof is rather indirect. (We believe category $\mathcal{O}$ to be the image of the functor [Etingof and Moura 2002].)

We extend the *q-character* of H. Knight [1995] and Frenkel and Reshetikhin [1999] to the elliptic case. The q-character of a module V encodes the spectra of the $\phi_j(z)$, which are meromorphic functions of z thanks to condition (i). It distinguishes the isomorphism class $[V]$ in the Grothendieck ring $K_0(\mathcal{O})$, and embeds $K_0(\mathcal{O})$ in a commutative ring. Our main results are summarized as follows:

(A) Proposition 4.10 on limit construction of infinite-dimensional *asymptotic* modules $\mathscr{W}_{r,x}$, for $r \in J$ and $x \in \mathbb{C}$, from a distinguished family of finite-dimensional modules, the *Kirillov–Reshetikhin* modules.

(B) Theorem 4.15 on generalized Baxter relations à la Frenkel and Hernandez [2015], the isomorphism class of any finite-dimensional module is a polynomial of the $[\mathscr{W}_{r,x}]/[\mathscr{W}_{r,y}]$ for $r \in J$ and $x, y \in \mathbb{C}$.

(C) Corollary 5.2 relating an asymptotic module $\mathscr{W}$ to a module D and tensor products S' and S'' of asymptotic modules such that $[D][\mathscr{W}] = [S'] + [S'']$.

The above results are known in category HJ of Hernandez and Jimbo [2012] for representations over a Borel subalgebra of an affine quantum group $U_\hbar(L\mathfrak{g})$. Category HJ contains the modules $L^{\pm}_{r,a}$ for $a \in \mathbb{C}$ and r a Dynkin node of $\mathfrak{g}$. The $L^{\pm}_{r,a}$ are "prefundamental" in that their tensor products realize all irreducible objects of HJ as subquotients, and they are not modules over $U_\hbar(L\mathfrak{g})$, which makes Borel subalgebras indispensable. The Grothendieck ring of HJ is commutative.

Result (A) is the asymptotic limit construction [Hernandez and Jimbo 2012] of the $L^{-}_{r,a}$. Result (B) is the relation [Frenkel and Hernandez 2015] between finite-dimensional modules and the $L^{+}_{r,a}$. Result (C) is either QQ^*-system [Feigin et al. 2017a; Hernandez and Leclerc 2016] or $Q\tilde{Q}$-system [Frenkel and Hernandez 2016], as there are two choices of the modules D for $\mathscr{W} = L^{+}_{r,a}$.

Hernandez and Leclerc [2016] interpreted the QQ^*-system as cluster mutations of Fomin–Zelevinsky. They provided conjectural monoidal categorifications of infinite rank cluster algebras by certain subcategories of HJ.

[1]In terms of the $K_i(z)$ from (1.8), we have $\phi_j(z) = K_j(z + \ell_j\hbar)K_{j+1}(z + \ell_j\hbar)^{-1}$ where $\ell_j = (N - j - 1)/2$. These are elliptic deformations of diagonal matrices in $\mathfrak{sl}_N$.

In a quantum integrable system associated to $U_\hbar(L\mathfrak{g})$, the transfer-matrix construction defines an action of the Grothendieck ring $K_0(\mathrm{HJ})$ on the quantum space; to an isomorphism class $[V]$ is attached a transfer matrix $t_V(z)$.

Result (B) is one key step [Frenkel and Hernandez 2015] in solving the conjecture of Frenkel and Reshetikhin [1999] on the spectra of the quantum integrable system, which connects the eigenvalues of the $t_V(z)$ to the q-character of V by the so-called Baxter polynomials [Baxter 1972]. These polynomials are eigenvalues of the $t_{L^+_{r,a}}(z)$ up to an overall factor [Frenkel and Hernandez 2015]. In this sense the $L^+_{r,a}$ have simpler structures than finite-dimensional modules, and the $t_{L^+_{r,a}}(z)$ are defined as *Baxter Q operators*, as an extension of earlier works of V. Bazhanov et al. [1997; 1999; Bazhanov and Tsuboi 2008] for $\mathfrak{g}$ a special linear Lie (super)algebra. Result (C) has as consequence the Bethe Ansatz equations for the roots of Baxter polynomials [Feigin et al. 2017a; Frenkel and Hernandez 2016].

Recently category HJ was studied for quantum toroidal algebras [Feigin et al. 2017b].

For elliptic quantum groups there are no obvious Borel subalgebras. Our idea is to replace the $L^\pm_{r,a}$ over Borel subalgebras by the asymptotic modules $\mathscr{W}^{(r)}_{d,a}$ (with a new parameter $d \in \mathbb{C}$) over the *entire* quantum group, which we now explain.

Let $\theta(z) := \theta(z \mid \tau)$ be the Jacobi theta function. For $r \in J$ a Dynkin node, $a \in \mathbb{C}$ a spectral parameter, and k a positive integer, by [Cavalli 2001; Tarasov and Varchenko 2001] there exists a unique finite-dimensional irreducible module $W^{(r)}_{k,a}$ which contains a nonzero vector ω (highest weight with respect to a triangular decomposition) such that:

$$\phi_j(z)\omega = \omega, \text{ if } j \neq r \quad \text{and} \quad \phi_r(z)\omega = \frac{\theta(z + a\hbar + k\hbar)}{\theta(z + a\hbar)}\omega.$$

This is a Kirillov–Reshetikhin (KR) module, a standard terminology for affine quantum groups and Yangians once the θ symbol is removed.

The core of this paper (Section 4) is analytic continuation with respect to k. We modify the asymptotic limits $L^-_{r,a}$ of Hernandez and Jimbo [2012], as in [Felder and Zhang 2017; Zhang 2017].

Firstly the existence of the inductive system $(W^{(r)}_{k,a})_{k>0}$ in [Hernandez and Jimbo 2012] relied on a cyclicity property of M. Kashiwara, Varagnolo–Vasserot and V. Chari, which is unavailable in the elliptic case. We reduce the problem to $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_2)$ by counting "dominant weights" in q-characters (Theorem 3.4), as in the proofs of T-system of KR modules over affine quantum groups by H. Nakajima [2003] and D. Hernandez [2006].

Secondly we express the matrix coefficients of any element of $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$ acting on the $W^{(r)}_{k,a}$, viewed as functions of $k \in \mathbb{Z}_{>0}$, in products of the $\theta(k\hbar + c)$ where $c \in \mathbb{C}$ is independent of k; see Lemma 4.8. In [Hernandez and Jimbo 2012] these are polynomials in k by induction. Our proof relies on the RLL comultiplication and is explicit.

Since $\theta(k\hbar + c)$ is an entire function of k, we take k in the matrix coefficients to be a fixed complex number d. This results in the asymptotic module $\mathscr{W}^{(r)}_{d,a}$ on the inductive limit $\varinjlim W^{(r)}_{k,a}$. The module $\mathscr{W}_{r,x}$ in (A) is $\mathscr{W}^{(r)}_{x,0}$. All irreducible modules of category $\mathcal{O}$ are subquotients of tensor products of asymptotic modules.

For $\mathfrak{g}$ of general type (A)–(C) and their proofs can be adapted to affine quantum groups, whose asymptotic modules appeared in the appendix of [Zhang 2017], as well as Yangians [Gautam and Toledano Laredo 2016; 2017b]. Borel subalgebras or double Yangians are not needed.

Results (A)–(C) were established for affine quantum general linear Lie superalgebras [Zhang 2018]; their proofs require more than q-characters as counting dominant weights is inefficient [Zhang 2016]. It is interesting to consider elliptic quantum supergroups [Galleas and Stokman 2015].

For elliptic quantum groups associated with other simple Lie algebras, one possible first step would be to derive the RLL presentation; see [Guay et al. 2017; Jing et al. 2017] for Yangians.

The R-matrix of Baxter and Belavin is governed by the vertex-type elliptic quantum group [Jimbo et al. 1999]. The equivalence of representation categories between this elliptic algebra and $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$ [Etingof and Schiffmann 1998], a Vertex-IRF correspondence, might give a representation theory meaning to the original Baxter Q operator of the 8-vertex model [Baxter 1972].

The paper is structured as follows. In Section 1 we review the theory of the elliptic quantum group associated to $\mathfrak{sl}_N$ and define category $\mathcal{O}$ of representations. We show that the q-character map is an injective ring homomorphism from the Grothendieck ring $K_0(\mathcal{O})$ to a commutative ring $\mathcal{M}_{\mathrm{t}}$ of meromorphic functions. Then we present the q-character formula of finite-dimensional evaluation modules.

Section 2 is devoted to the proof of the q-character formula.

We derive in Section 3 basic facts on tensor products of KR modules (T-system, fusion) from the q-character formula. They are needed in Section 4 to construct the inductive system of KR modules and the asymptotic modules. We obtain a highest weight classification of irreducible modules in category $\mathcal{O}$. As a consequence, all standard irreducible evaluation modules of [Tarasov and Varchenko 2001] are in category $\mathcal{O}$.

In Section 5 we establish the three-term Baxter TQ relations in $K_0(\mathcal{O})$, which are infinite-dimensional analogs of the T-system. These relations are interpreted as functional relations of transfer matrices in Section 6.

1. Elliptic quantum groups and their representations

Let $N \in \mathbb{Z}_{>0}$. We introduce a category $\mathcal{O}$ (abelian and monoidal) of representations of the elliptic quantum group attached to the Lie algebra $\mathfrak{sl}_N$, and prove that its Grothendieck ring is commutative, based on q-characters.

Fix a complex number $\tau \in \mathbb{C}$ with $\mathrm{Im}(\tau) > 0$. Define the Jacobi theta function

$$\theta(z) = \theta(z \mid \tau) := -\sum_{j=-\infty}^{\infty} \exp\bigl(i\pi\bigl(j+\tfrac{1}{2}\bigr)^2\tau + 2i\pi\bigl(j+\tfrac{1}{2}\bigr)\bigl(z+\tfrac{1}{2}\bigr)\bigr), \quad i = \sqrt{-1}.$$

It is an entire function of $z \in \mathbb{C}$ with zeros lying on the lattice $\Gamma := \mathbb{Z} + \mathbb{Z}\tau$ and

$$\theta(z+1) = -\theta(z), \quad \theta(z+\tau) = -e^{-i\pi\tau - 2i\pi z}\theta(z), \quad \theta(-z) = -\theta(z).$$

Fix a complex number $\hbar \in \mathbb{C} \setminus (\mathbb{Q} + \mathbb{Q}\tau)$, which is the deformation parameter.

Let $\mathfrak{h}$ be standard Cartan subalgebra of $\mathfrak{sl}_N$; it is a complex vector space generated by the ϵ_i for $1 \le i \le N$ subject to the relation $\sum_{i=1}^N \epsilon_i = 0$. Let $\mathbb{C}^N := \oplus_{i=1}^N \mathbb{C}v_i$ and $E_{ij} \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^N)$ be the elementary matrices: $v_k \mapsto \delta_{jk} v_i$ for $1 \le i, j, k \le N$. Define the $\mathrm{End}_{\mathbb{C}}(\mathbb{C}^N \otimes \mathbb{C}^N)$-valued meromorphic functions of $(z; \lambda) \in \mathbb{C} \times \mathfrak{h}$ by

$$\boldsymbol{R}(z; \lambda) = \sum_i E_{ii}^{\otimes 2} + \sum_{i \neq j} \left(\frac{\theta(z)\theta(\lambda_{ij} - \hbar)}{\theta(z+\hbar)\theta(\lambda_{ij})} E_{ii} \otimes E_{jj} + \frac{\theta(z+\lambda_{ij})\theta(\hbar)}{\theta(z+\hbar)\theta(\lambda_{ij})} E_{ij} \otimes E_{ji} \right).$$

In the summations $1 \le i, j \le N$, and $\lambda_{ij} \in \mathfrak{h}^*$ sends $\sum_{i=1}^N c_i \epsilon_i \in \mathfrak{h}$ to $c_i - c_j \in \mathbb{C}$. By [Felder 1995], $\boldsymbol{R}(z; \lambda)$ satisfies the *quantum dynamical Yang–Baxter equation*:

$$\begin{aligned} &\boldsymbol{R}^{12}(z-w; \lambda + \hbar h^{(3)}) \boldsymbol{R}^{13}(z; \lambda) \boldsymbol{R}^{23}(w; \lambda + \hbar h^{(1)}) \\ &\qquad = \boldsymbol{R}^{23}(w; \lambda) \boldsymbol{R}^{13}(z; \lambda + \hbar h^{(2)}) \boldsymbol{R}^{12}(z-w; \lambda) \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^N)^{\otimes 3}. \end{aligned} \tag{1.1}$$

If $\boldsymbol{R}(z; \lambda) = \sum_p c_\lambda^p x_p \otimes y_p$ with $x_p, y_p \in \mathrm{End}_{\mathbb{C}}(\mathbb{C}^N)$, then

$$\boldsymbol{R}^{13}(z; \lambda + \hbar h^{(2)}) : u \otimes v_j \otimes w \mapsto \sum_p c^p_{\lambda + \hbar \epsilon_j} x_p(u) \otimes v_j \otimes y_p(w),$$

for $u, w \in \mathbb{C}^N$ and $1 \le j \le N$. The other symbols have a similar meaning.

Let $\mathbb{M} := \mathbb{M}_{\mathfrak{h}}$ be the field of meromorphic functions of $\lambda \in \mathfrak{h}$. It contains the subfield $\mathbb{C}$ of constant functions. A $\mathbb{C}$-linear map Φ of two $\mathbb{M}$-vector spaces will sometimes be denoted by $\Phi(\lambda)$ to emphasize the dependence on λ.

1A. ***Algebraic notions.*** Since the elliptic quantum groups will act on $\mathbb{M}$-vector spaces via difference operators, which are in general not $\mathbb{M}$-linear, we need to recall some basis constructions about difference operators. Our exposition follows largely [Etingof and Varchenko 1998], with minor modifications as in [Felder and Zhang 2017].

Define the category $\mathcal{V}$ as follows. An object is $X = \oplus_{\alpha \in \mathfrak{h}} X[\alpha]$ where each $X[\alpha]$ is an $\mathbb{M}$-vector space and, if nonzero, is called a weight space of weight (or $\mathfrak{h}$-weight) α. Let $\mathrm{wt}(X) \subseteq \mathfrak{h}$ be the set of weights of X. Write $\mathrm{wt}(v) = \alpha$ if $v \in X[\alpha]$.

A morphism $f : X \to Y$ in $\mathcal{V}$ is an $\mathbb{M}$-linear map which respects the weight gradings. Let $\mathcal{V}_{\mathrm{ft}}$ be the full subcategory of $\mathcal{V}$ consisting of X whose weight spaces are finite-dimensional $\mathbb{M}$-vector spaces ("ft" means finite type in [Felder and Varchenko 1996b]).

Viewed as subcategories of the category of $\mathbb{M}$-vector spaces, $\mathcal{V}$ and $\mathcal{V}_{\mathrm{ft}}$ are abelian.

Let X and Y be objects of $\mathcal{V}$. Their *dynamical tensor product* $X \widetilde{\otimes} Y$ is constructed as follows. For $\alpha, \beta \in \mathfrak{h}$, let $X[\alpha] \widetilde{\otimes} Y[\beta]$ be the quotient of the usual tensor product of $\mathbb{C}$-vector spaces $X[\alpha] \otimes_{\mathbb{C}} Y[\beta]$ by the relation

$$g(\lambda) v \otimes_{\mathbb{C}} w = v \otimes_{\mathbb{C}} g(\lambda + \hbar\beta) w, \quad \text{for } v \in X[\alpha],\, w \in Y[\beta],\, g(\lambda) \in \mathbb{M}.$$

Let $\overline{\otimes}$ denote the image of $\otimes_{\mathbb{C}}$ under the quotient. $X[\alpha]\overline{\otimes}Y[\beta]$ becomes an $\mathbb{M}$-vector space by setting $g(\lambda)(v\overline{\otimes}w) = v\overline{\otimes}g(\lambda)w$. For $\gamma \in \mathfrak{h}$, the weight space $(X\overline{\otimes}Y)[\gamma]$ is then the direct sum of the $X[\alpha]\overline{\otimes}Y[\beta]$ with $\alpha + \beta = \gamma$.

For $\alpha, \beta \in \mathfrak{h}$, a $\mathbb{C}$-linear map $\Phi : X \to Y$ is called a *difference map* of bidegree (α, β) [Etingof and Varchenko 1998, §4.2] if it sends every weight space $X[\gamma]$ to $Y[\gamma + \beta - \alpha]$, and if

$$\Phi(g(\lambda)v) = g(\lambda + \beta\hbar)\Phi(v) \quad \text{for } g(\lambda) \in \mathbb{M} \text{ and } v \in X.$$

Such a map can be recovered from its matrix as in the case of $\mathbb{M}$-linear maps. Choose $\mathbb{M}$-bases $\mathcal{B}$ and $\mathcal{B}'$ for X and Y respectively. Define the $\mathcal{B}' \times \mathcal{B}$ matrix $[\Phi]$ by taking its (b', b)-entry $[\Phi]_{b'b}(\lambda) \in \mathbb{M}$, for $b \in \mathcal{B}$ and $b' \in \mathcal{B}'$, to be the coefficient of b' in $\Phi(b)$. Then for any vector $v = \sum_{b\in\mathcal{B}} g_b(\lambda)b$ of X where $g_b(\lambda) \in \mathbb{M}$, we have[2]

$$\Phi(v) = \sum_{b'\in\mathcal{B}'} b' \sum_{b\in\mathcal{B}} [\Phi]_{b'b}(\lambda) \times g_b(\lambda + \hbar\beta).$$

When $X = Y$, a difference map is also called a difference operator. To define its matrix, we always assume $\mathcal{B}' = \mathcal{B}$.

By an algebra we mean a unital associative algebra over $\mathbb{C}$.

As in [Etingof and Varchenko 1998, Definition 4.1], an $\mathfrak{h}$-*algebra* is an algebra A, endowed with $\mathfrak{h}$-bigrading $A = \oplus_{\alpha,\beta\in\mathfrak{h}} A_{\alpha,\beta}$ which respects the algebra structure and is called the *weight decomposition*, and two algebra embeddings $\mu_{\boldsymbol{l}}, \mu_{\boldsymbol{r}} : \mathbb{M} \to A_{0,0}$ called the *left and right moment maps*, such that for $a \in A_{\alpha,\beta}$ and $g(\lambda) \in \mathbb{M}$, we have

$$\mu_{\boldsymbol{l}}(g(\lambda))a = a\mu_{\boldsymbol{l}}(g(\lambda - \hbar\alpha)) \quad \text{and} \quad \mu_{\boldsymbol{r}}(g(\lambda))a = a\mu_{\boldsymbol{r}}(g(\lambda - \hbar\beta)).$$

Call (α, β) the bidegree of elements in $A_{\alpha,\beta}$. A morphism of $\mathfrak{h}$-algebras is an algebra morphism preserving the moment maps and the weight decompositions.

From two $\mathfrak{h}$-algebras A and B we construct their tensor product $A\tilde{\otimes}B$ as follows. For $\alpha, \beta, \gamma \in \mathfrak{h}$, let $A_{\alpha,\beta}\tilde{\otimes}B_{\beta,\gamma}$ be $A_{\alpha,\beta} \otimes_{\mathbb{C}} B_{\beta,\gamma}$ modulo the relation

$$\mu_{\boldsymbol{r}}^A(g(\lambda))a \otimes_{\mathbb{C}} b = a \otimes_{\mathbb{C}} \mu_{\boldsymbol{l}}^B(g(\lambda))b \quad \text{for } a \in A_{\alpha,\beta},\ b \in B_{\beta,\gamma}, g(\lambda) \in \mathbb{M}.$$

Let $(A\tilde{\otimes}B)_{\alpha,\gamma}$ be the direct sum of the $A_{\alpha,\beta}\tilde{\otimes}B_{\beta,\gamma}$ over $\beta \in \mathfrak{h}$ ($\tilde{\otimes}$ denotes the image of $\otimes_{\mathbb{C}}$ under the quotient $\otimes_{\mathbb{C}} \to \tilde{\otimes}$). Multiplication in $A\tilde{\otimes}B$ is induced by $(a\tilde{\otimes}b)(a'\tilde{\otimes}b') = aa'\tilde{\otimes}bb'$. The moment maps are given by

$$\mu_{\boldsymbol{l}}^{A\tilde{\otimes}B} : g(\lambda) \mapsto \mu_{\boldsymbol{l}}^A(g(\lambda))\tilde{\otimes}1 \quad \text{and} \quad \mu_{\boldsymbol{r}}^{A\tilde{\otimes}B} : g(\lambda) \mapsto 1\tilde{\otimes}\mu_{\boldsymbol{r}}^B(g(\lambda)) \quad \text{for } g(\lambda) \in \mathbb{M}.$$

To an $\mathfrak{h}$-graded vector space one can attach naturally an $\mathfrak{h}$-algebra. Let X be an object of $\mathcal{V}$. Let $\mathbb{D}_{\alpha,\beta}^X$ denote the $\mathbb{C}$-vector space of difference operators $X \to X$ of bidegree (α, β). Then the direct sum

[2] Note that difference maps of bidegree (α, α) make sense for arbitrary $\mathbb{M}$-vector spaces.

$\mathbb{D}^X := \oplus_{\alpha,\beta\in\mathfrak{h}}\mathbb{D}^X_{\alpha,\beta}$ is a subalgebra of $\mathrm{End}_{\mathbb{C}}(X)$. It is an $\mathfrak{h}$-algebra structure with the moment maps

$$\mu_{\boldsymbol{r}}(g(\lambda))v = g(\lambda)v \quad \text{and} \quad \mu_{\boldsymbol{l}}(g(\lambda))v = g(\lambda+\hbar\alpha)v \quad \text{for } v \in X[\alpha],\ g(\lambda) \in \mathbb{M}.$$

Tensor products of difference operators are also difference operators. To be precise, let X and Y be two objects of $\mathcal{V}$. Let $\Phi : X \to X$ and $\Psi : Y \to Y$ be difference operators of bidegree (α, β) and (β, γ), respectively. The $\mathbb{C}$-linear map

$$X \otimes_{\mathbb{C}} Y \to X\overline{\otimes}Y, \quad v \otimes_{\mathbb{C}} w \mapsto \Phi(v)\overline{\otimes}\Psi(w)$$

is easily seen to factor through $X \otimes_{\mathbb{C}} Y \to X\overline{\otimes}Y$ and induces the $\mathbb{C}$-linear map $\Phi\overline{\otimes}\Psi : X\overline{\otimes}Y \to X\overline{\otimes}Y$, which is shown to be a difference operator of bidegree (α, γ). As in [Etingof and Varchenko 1998, Lemma 4.3], the following defines a morphism of $\mathfrak{h}$-algebras

$$\mathbb{D}^X\tilde{\otimes}\mathbb{D}^Y \to \mathbb{D}^{X\overline{\otimes}Y}, \quad \Phi\tilde{\otimes}\Psi \mapsto \Phi\overline{\otimes}\Psi.$$

1B. *Elliptic quantum groups.* For $1 \le i, j, p, q \le N$ let $R^{ij}_{pq}(z;\lambda)$ be the coefficient of $v_p \otimes v_q$ in $\boldsymbol{R}(z;\lambda)(v_i \otimes v_j)$; it can be viewed as an element of $\mathbb{M}$ after fixing $z \in \mathbb{C}$. The *elliptic quantum group* $\mathcal{E} := \mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$ is an $\mathfrak{h}$-algebra generated by[3]

$$L_{ij}(z) \in \mathcal{E}_{\epsilon_i,\epsilon_j} \quad \text{for } 1 \le i, j \le N$$

subject to the dynamical RLL relation [Etingof and Varchenko 1998, §4.4]: for $1 \le i, j, m, n \le N$,

$$\sum_{p,q=1}^{N} \mu_{\boldsymbol{l}}(R^{pq}_{mn}(z-w;\lambda))L_{pi}(z)L_{qj}(w) = \sum_{p,q=1}^{N} \mu_{\boldsymbol{r}}(R^{ij}_{pq}(z-w;\lambda))L_{nq}(w)L_{mp}(z). \tag{1.2}$$

There is an $\mathfrak{h}$-algebra morphism [Etingof and Varchenko 1998; Felder and Varchenko 1996b]

$$\Delta : \mathcal{E} \to \mathcal{E}\tilde{\otimes}\mathcal{E}, \quad L_{ij}(z) \mapsto \sum_{k=1}^{N} L_{ik}(z)\tilde{\otimes}L_{kj}(z), \quad \text{for } 1 \le i, j \le N \tag{1.3}$$

which is coassociative $(1\tilde{\otimes}\Delta)\Delta = (\Delta\tilde{\otimes}1)\Delta$ and is called the coproduct. For $u \in \mathbb{C}$,

$$\Phi_u : \mathcal{E} \to \mathcal{E}, \quad L_{ij}(z) \mapsto L_{ij}(z+u\hbar) \quad \text{for } 1 \le i, j \le N \tag{1.4}$$

extends uniquely to an $\mathfrak{h}$-algebra automorphism (spectral parameter shift).

Strictly speaking, $\mathcal{E}$ is not well-defined as an $\mathfrak{h}$-algebra because of the additional parameter z; this is resolved in [Konno 2016] by viewing $z, \hbar$ as formal variables. In this paper we are mainly concerned with representations in which (1.2)–(1.4) make sense as identities of difference operators depending analytically on z.

[3] We use $\mathfrak{sl}_N$, as in [Felder 1995; Felder and Varchenko 1996b], to emphasize that $\mathfrak{h}$ is the Cartan subalgebra of $\mathfrak{sl}_N$. Other works [Cavalli 2001; Konno 2016] use $\mathfrak{gl}_N$ for the reason that the elliptic quantum determinant is not fixed to be 1.

Let $\mathfrak{S}_N$ be the group of permutations of $\{1, 2, \ldots, N\}$. For $1 \le k \le N$, let $\mathfrak{S}^k$ be the subgroup of permutations which fix the last k letters. The k-th *fundamental weight* ϖ_k and *elliptic quantum minor* $\mathcal{D}_k(z)$ are defined by [Tarasov and Varchenko 2001, (2.5)]:

$$\varpi_k := \sum_{i=1}^{k} \epsilon_i \in \mathfrak{h}, \tag{1.5}$$

$$\Theta_k(\lambda) := \prod_{N-k+1 \le i < j \le N} \theta(\lambda_{ij}) \in \mathbb{M}^\times,$$

$$\mathcal{D}_k(z) := \frac{\mu_{\boldsymbol{r}}(\Theta_k(\lambda))}{\mu_{\boldsymbol{l}}(\Theta_k(\lambda))} \sum_{\sigma \in \mathfrak{S}^k} \operatorname{sign}(\sigma) \prod_{i=N}^{N-k+1} L_{\sigma(i),i}(z + (N-i)\hbar) \in \mathcal{E}. \tag{1.6}$$

Here $\operatorname{sign}(\sigma) \in \{\pm 1\}$ denotes the signature of the permutation σ. We take the descending product over $N \ge i \ge N-k+1$ in (1.6). Set $\varpi_0 := 0$.

We shall need the following elements $\hat{L}_k(z)$ of $\mathcal{E}_{\epsilon_k,\epsilon_k}$ as in [Tarasov and Varchenko 2001, (4.1)]:

$$\hat{L}_N(z) := L_{NN}(z) \quad \text{and} \quad \hat{L}_k(z) = L_{kk}(z) \prod_{j=k+1}^{N} \frac{\mu_{\boldsymbol{r}}(\theta(\lambda_{kj}))}{\mu_{\boldsymbol{l}}(\theta(\lambda_{kj}))}. \tag{1.7}$$

Theorem 1.1 [Tarasov and Varchenko 2001, Proposition 2.1] and [Konno 2016, (E.18)]. *$\mathcal{D}_N(z)$ is central in $\mathcal{E}$ and grouplike*: $\Delta(\mathcal{D}_N(z)) = \mathcal{D}_N(z) \tilde{\otimes} \mathcal{D}_N(z)$.

The simple roots $\alpha_i := \epsilon_i - \epsilon_{i+1}$ for $1 \le i < N$ generate a free abelian subgroup $\boldsymbol{Q}$ of $\mathfrak{h}$, called the root lattice. Let $\boldsymbol{Q}_+$ and $\boldsymbol{Q}_-$ be submonoids of $\boldsymbol{Q}$ generated by the α_i and $-\alpha_i$ respectively. Define the lexicographic partial ordering $\prec$ on $\mathfrak{h}$ as follows:

$$\alpha \prec \beta \quad \text{if } \beta - \alpha = n_l \alpha_l + \sum_{i=l+1}^{N-1} n_i \alpha_i \in \boldsymbol{Q} \text{ with } n_l \in \mathbb{Z}_{>0}.$$

This is weaker than the standard ordering $\alpha \le \beta$ if $\beta - \alpha \in \boldsymbol{Q}_+$.

Corollary 1.2. *$\mathcal{D}_k(z)$ commutes with the $L_{ij}(w)$ for $N-k < i, j \le N$ and $\Delta(\mathcal{D}_k(z)) - \mathcal{D}_k(z) \tilde{\otimes} \mathcal{D}_k(z)$ is a finite sum $\sum_\alpha x_\alpha \tilde{\otimes} y_\alpha$ over $\{\alpha \in \mathfrak{h} \mid -\varpi_{N-k} \prec \alpha\}$ where x_α and y_α are of bidegree $(-\varpi_{N-k}, \alpha)$ and $(\alpha, -\varpi_{N-k})$ respectively.*

The proof of the corollary is postponed to Section 2A.

1C. ***Categories.*** *From now on unless otherwise stated vector spaces, linear maps and bases are defined over* $\mathbb{M}$. Let X be an object of $\mathcal{V}_{\mathrm{ft}}$. A representation of $\mathcal{E}$ on X consists of difference operators $L_{ij}^X(z) : X \to X$ of bidegree (ϵ_i, ϵ_j) for $1 \le i, j \le N$ depending on $z \in \mathbb{C}$ with the following properties:[4]

[4]This is called a representation of *finite type* in [Felder and Varchenko 1996b]. From condition (M1) it follows that the coefficients of the $L_{ij}^X(z)$ are meromorphic functions with respect to *any* basis of X.

(M1) There exists a basis of X with respect to which all the matrix entries of the difference operators $L_{ij}^X(z)$ are meromorphic functions of $(z,\lambda)\in\mathbb{C}\times\mathfrak{h}$.

(M2) Equation (1.2) holds in $\mathbb{D}^X$ with μ_l and μ_r being moment maps in $\mathbb{D}^X$.

Call X an $\mathcal{E}$-module. Property (M2) can be interpreted as an $\mathfrak{h}$-algebra morphism $\mathcal{E}\to\mathbb{D}^X$ sending $L_{ij}(z)\in\mathcal{E}$ to the difference operator $L_{ij}^X(z)$ on X. Applying ρ to the elements of (1.6)–(1.7), one gets difference operators $\mathcal{D}_k^X(z)$ and $\hat{L}_k^X(z)$ acting on X with bidegree $(-\varpi_{N-k},-\varpi_{N-k})$ and (ϵ_k,ϵ_k), respectively. When no confusion arises, we shall drop the superscript X from L^X, $\mathcal{D}^X$ and $\hat{L}^X$ to simplify notations.

A morphism $\Phi: X\to Y$ of $\mathcal{E}$-modules is a linear map which respects the $\mathfrak{h}$-gradings (so that Φ is a morphism in category $\mathcal{V}$) and satisfies $\Phi L_{ij}^X(z)=L_{ij}^Y(z)\Phi$ for $1\le i,j\le N$. The category of $\mathcal{E}$-modules is denoted by Rep. It is a subcategory of $\mathcal{V}_{\mathrm{ft}}$ and is abelian since the kernel and cokernel of a morphism of $\mathcal{E}$-modules, as $\mathfrak{h}$-graded $\mathbb{M}$-vector spaces, are naturally $\mathcal{E}$-modules.[5]

Definition 1.3 [Etingof and Moura 2002, §4]. $\tilde{\mathcal{O}}$ is the full subcategory of Rep whose objects X are such that $\mathrm{wt}(X)$ is contained in a *finite* union of cones $\mu+\boldsymbol{Q}_-$ with $\mu\in\mathfrak{h}$.

For X and Y objects in category $\tilde{\mathcal{O}}$, the $L_{ij}^{X\overline{\otimes}Y}(z):=\sum_{k=1}^N L_{ik}^X(z)\overline{\otimes}L_{kj}^Y(z)$ define a representation of $\mathcal{E}$ on $X\overline{\otimes}Y$ which is easily seen to be in category $\tilde{\mathcal{O}}$. So $\tilde{\mathcal{O}}$ is a monoidal subcategory of $\mathcal{V}$. Similarly, $\tilde{\mathcal{O}}$ is an abelian subcategory of Rep.

Definition 1.4 [Felder and Zhang 2017, §2]. An object in $\mathcal{F}_{\mathrm{mer}}$ consists of a finite-dimensional vector space V equipped with difference operators $D_l(z): V\to V$ of bidegree $(-\varpi_{N-l},-\varpi_{N-l})$ (see footnote 2) for $1\le l\le N$ depending on $z\in\mathbb{C}$ such that:

(M3) There exists an ordered basis of V with respect to which the matrices of the difference operators $D_l(z)$ are upper triangular, the diagonal entries are nonzero meromorphic functions of $z\in\mathbb{C}$, and the off-diagonal entries are meromorphic functions of $(z,\lambda)\in\mathbb{C}\times\mathfrak{h}$.

A morphism $\Phi: V\to W$ in $\mathcal{F}_{\mathrm{mer}}$ is a linear map commuting with the $D_l(z)$. (Namely, $\Phi D_l^V(z)=D_l^W(z)\Phi: V\to W$ for $1\le l\le N$. Here we add the superscripts V and W in the $D_l(z)$ to indicate the space on which they act.)

For V an object of $\mathcal{F}_{\mathrm{mer}}$, the operators $D_l(z)$ being invertible because of the triangularity, one has a unique factorization of operators for $1\le l\le N$:

$$D_l(z)=K_N(z)K_{N-1}(z+\hbar)K_{N-2}(z+2\hbar)\cdots K_{N-l+1}(z+(l-1)\hbar). \tag{1.8}$$

Notably $K_l(z): X\to X$ is a difference operator of bidegree (ϵ_l,ϵ_l). Property (M3) still holds if the $D_l(z)$ are replaced by the $K_l(z)$.

[5]In other works [Cavalli 2001; Etingof and Moura 2002; Felder and Varchenko 1996b; Gautam and Toledano Laredo 2017a; Konno 2009; 2016; Tarasov and Varchenko 2001; Yang and Zhao 2017] a module V is an $\mathfrak{h}$-graded $\mathbb{C}$-vector space; morphisms of modules depend on the dynamical parameter λ, so do their kernel and cokernel; the abelian category structure is nontrivial. The scalar extension gives a module $V\otimes_{\mathbb{C}}\mathbb{M}$ in the present situation. Since our modules and morphisms are $\mathbb{M}$-linear, the dependence of kernels and images on the dynamical parameter does not matter.

The forgetful functor from $\mathcal{F}_{\mathrm{mer}}$ to the category of finite-dimensional vector spaces equips $\mathcal{F}_{\mathrm{mer}}$ with an abelian category structure. (For a proof, we refer to [Felder and Zhang 2017, §2.1] where another characterization of category $\mathcal{F}_{\mathrm{mer}}$ in terms of Jordan–Hölder series is given.) Let us describe its Grothendieck group $K_0(\mathcal{F}_{\mathrm{mer}})$.

The multiplicative group $\mathbb{M}_{\mathbb{C}}^{\times}$ of nonzero meromorphic functions of $z \in \mathbb{C}$ contains a subgroup $\mathbb{C}^{\times}$ of nonzero constant functions. Let $\mathcal{M}$ be the quotient group of $(\mathbb{M}_{\mathbb{C}}^{\times})^N$ by its subgroup formed of $(c_1, c_2, \dots, c_N) \in (\mathbb{C}^{\times})^N$ such that $c_1 c_2 \cdots c_N = 1$. We show that $K_0(\mathcal{F}_{\mathrm{mer}})$ has a $\mathbb{Z}$-basis indexed by $\mathcal{M}$.

For $\boldsymbol{f} = (f_1(z), f_2(z), \dots, f_N(z)) \in (\mathbb{M}_{\mathbb{C}}^{\times})^N$, the vector space $\mathbb{M}$ with the following difference operators $D_l(z)$ is an object in category $\mathcal{F}_{\mathrm{mer}}$ denoted by $\mathbb{M}_{\boldsymbol{f}}$:

$$g(\lambda) \mapsto g(\lambda - \hbar\varpi_{N-l}) f_N(z) f_{N-1}(z+\hbar) f_{N-2}(z+2\hbar) \cdots f_{N-l+1}(z+(l-1)\hbar).$$

We have $K_l(z)g(\lambda) = g(\lambda + \hbar\epsilon_l) f_l(z)$. As a consequence of (M3) in Definition 1.4, all irreducible objects of category $\mathcal{F}_{\mathrm{mer}}$ are of this form.

Lemma 1.5. *Let $\boldsymbol{e}, \boldsymbol{f} \in (\mathbb{M}_{\mathbb{C}}^{\times})^N$. The objects $\mathbb{M}_{\boldsymbol{e}}$ and $\mathbb{M}_{\boldsymbol{f}}$ are isomorphic in category $\mathcal{F}_{\mathrm{mer}}$ if and only if $\boldsymbol{e}, \boldsymbol{f}$ have the same image under the quotient $(\mathbb{M}_{\mathbb{C}}^{\times})^N \twoheadrightarrow \mathcal{M}$.*

Proof. Write $\boldsymbol{e} = (e_1(z), e_2(z), \dots, e_N(z))$ and $\boldsymbol{f} = (f_1(z), f_2(z), \dots, f_N(z))$.

Sufficiency: assume $e_l(z) = f_l(z)c_l$ with $c_l \in \mathbb{C}^{\times}$ and $c_1 c_2 \cdots c_N = 1$. For $1 \le l < N$, choose b_l such that $c_l = e^{b_l \hbar}$. Set $b_N := -b_1 - b_2 - \cdots - b_{N-1}$. Then $e^{b_N \hbar} = c_1^{-1} c_2^{-1} \cdots c_{N-1}^{-1} = c_N$ and the following is a well-defined element of $\mathbb{M}^{\times}$:

$$\varphi(x_1\epsilon_1 + x_2\epsilon_2 + \cdots + x_N\epsilon_N) = e^{b_1x_1 + b_2x_2 + \cdots + b_Nx_N} \quad \text{for } x_1, x_2, \dots, x_N \in \mathbb{C}.$$

Indeed $\varphi(\alpha+\beta) = \varphi(\alpha)\varphi(\beta)$ and $\varphi(x\epsilon_1 + x\epsilon_2 + \cdots + x\epsilon_N) = 1$ for $x \in \mathbb{C}$. Notably,

$$\varphi(\lambda + \hbar\epsilon_l) = \varphi(\lambda)\varphi(\hbar\epsilon_l) = e^{\hbar b_l}\varphi(\lambda) = c_l\varphi(\lambda).$$

So $\mathbb{M}_{\boldsymbol{e}} \to \mathbb{M}_{\boldsymbol{f}}$, $g(\lambda) \mapsto g(\lambda)\varphi(\lambda)$ is an isomorphism in category $\mathcal{F}_{\mathrm{mer}}$.

Let $\Phi : \mathbb{M}_{\boldsymbol{e}} \to \mathbb{M}_{\boldsymbol{f}}$ be an isomorphism in category $\mathcal{F}_{\mathrm{mer}}$. Set $\varphi(\lambda) := \Phi(1)$. Then $\varphi(\lambda) \in \mathbb{M}^{\times}$. Applying $\Phi K_l(z) = K_l(z)\Phi$ to 1 we get

$$\varphi(\lambda + \hbar\epsilon_l) f_l(z) = \varphi(\lambda) e_l(z).$$

So $e_l(z)/f_l(z) = \varphi(\lambda + \hbar\epsilon_l)/\varphi(\lambda)$, being independent of z, is a constant function $c_l \in \mathbb{C}^{\times}$. We have $e_l(z) = f_l(z)c_l$ and $\varphi(\lambda + \hbar\epsilon_l) = c_l\varphi(\lambda)$. It follows that

$$\varphi(\lambda) = \varphi(\lambda + \hbar\epsilon_1 + \hbar\epsilon_2 + \cdots + \hbar\epsilon_N) = c_1 c_2 \cdots c_N \varphi(\lambda),$$

which implies $c_1 c_2 \cdots c_N = 1$. So $\boldsymbol{e}$ and $\boldsymbol{f}$ have the same image in $\mathcal{M}$. □

For each $\boldsymbol{f} \in \mathcal{M}$, let us fix a preimage $\boldsymbol{f}'$ in $(\mathbb{M}_{\mathbb{C}}^{\times})^N$ and set $\mathbb{M}(\boldsymbol{f}) := \mathbb{M}_{\boldsymbol{f}'}$. Then the isomorphism classes $[\mathbb{M}(\boldsymbol{f})]$ for $\boldsymbol{f} \in \mathcal{M}$ form a $\mathbb{Z}$-basis of $K_0(\mathcal{F}_{\mathrm{mer}})$. When no confusion arises, we identify an element of $(\mathbb{M}_{\mathbb{C}}^{\times})^N$ with its image in $\mathcal{M}$.

Lemma 1.6. *Let V be in category $\mathcal{F}_{\text{mer}}$. Assume B is an ordered basis of V with respect to which the matrices of the difference operators $K_l(z)$ are upper triangular. Then for $b \in B$ and $1 \leq l \leq N$ there exist $\varphi_b(\lambda) \in \mathbb{M}^\times$ and $f_{b,l}(z) \in \mathbb{M}_{\mathbb{C}}^\times$ such that*

$$[K_l]_{bb}(z;\lambda) = f_{b,l}(z)\frac{\varphi_b(\lambda)}{\varphi_b(\lambda+\hbar\epsilon_l)}.$$

Recall that $[K_l]_{bb}(z;\lambda)$ is the coefficient of b in $K_l(z)b$. This lemma says that if the matrices of the $K_l(z)$ are upper triangular, then their diagonal entries must be of the form $f(z)h(\lambda)$, and the $h(\lambda)$ can be gauged away uniformly.

More precisely, the new basis $\{\varphi_b(\lambda)b \mid b \in B\}$ with the ordering induced from B satisfies (M3) in Definition 1.4; the diagonal entry of $K_l(z)$ associated to $\varphi_b(\lambda)b$ is $f_{b,l}(z)$. This yields the following identity in the Grothendieck group $K_0(\mathcal{F}_{\text{mer}})$:

$$[V] = \sum_{b\in B}[\mathbb{M}(f_{b,1}(z), f_{b,2}(z), \dots, f_{b,N}(z))].$$

Proof. Write $B = \{b_1 < b_2 < \dots < b_m\}$. We proceed by induction on the dimension $m = \dim(V)$. If $m = 1$, then there exist $\boldsymbol{f} = (f_1(z), f_2(z), \dots, f_N(z)) \in (\mathbb{M}_{\mathbb{C}}^\times)^N$ and an isomorphism $\Phi : \mathbb{M}_{\boldsymbol{f}} \to V$ in category $\mathcal{F}_{\text{mer}}$. Let $\Phi(1) = \varphi(\lambda)b_1$. Then applying $\Phi K_l(z) = K_l(z)\Phi$ to 1 we obtain the desired identity

$$f_l(z)\varphi(\lambda) = [K_l]_{b_1b_1}(z;\lambda)\varphi(\lambda+\hbar\epsilon_l).$$

If $m > 1$, then the subspace V' of V spanned by $(b_1, b_2, \dots, b_{m-1})$ is stable by the $K_l(z)$ and $D_l(z)$ by the triangularity assumption. So V' is an object of category $\mathcal{F}_{\text{mer}}$ and we obtain a short exact sequence $0 \to V' \to V \to V/V' \to 0$. The rest is clear by applying the induction hypothesis to V' and V/V', which have ordered bases $\{b_1 < b_2 < \dots < b_{m-1}\}$ and $\{b_m + V'\}$ respectively. □

Definition 1.7. $\mathcal{O}$ is the full subcategory of $\tilde{\mathcal{O}}$ consisting of $\mathcal{E}$-modules X such that $X[\mu]$ endowed with the action of the $\mathcal{D}_l(z)$ belongs to $\mathcal{F}_{\text{mer}}$ for all $\mu \in \text{wt}(X)$.

The definition of $\tilde{\mathcal{O}}$ is standard as in the cases of Kac–Moody algebras [Kac 1990] and quantum affinizations [Hernandez 2005]. Definition 1.7 is a special feature of elliptic quantum groups. It is meant to loosen the dependence on the dynamical parameter λ.[6]

$\mathcal{O}$ is an abelian subcategory of $\tilde{\mathcal{O}}$. For X in category $\mathcal{O}$, (1.8) defines difference operators $K_l(z) : X \to X$ of bidegree (ϵ_l, ϵ_l) for $1 \leq l \leq N$.[7]

Following [Cavalli 2001, Definition 2.1], a nonzero weight vector of a module X in category $\tilde{\mathcal{O}}$ is called *singular* if it is annihilated by the $L_{ij}(z)$ for $1 \leq j < i \leq N$.

Lemma 1.8. *Let X be in category $\mathcal{O}$. If $v \in X$ is singular, then $K_i(z)v = \hat{L}_i(z)v$ for all $1 \leq i \leq N$.*

[6] For the elliptic quantum group associated to an arbitrary finite-dimensional simple Lie algebra, Gautam and Toledano Laredo [2017a, §2.3] defined a category of *integrable* modules on which the action of the elliptic Cartan currents, analogs of $\mathcal{D}_k(z)$, is independent of λ. The asymptotic modules that we will construct in Section 4 are not integrable.

[7] The $K_l(z)$ do not come from the elliptic quantum group, yet formally they are elliptic Cartan currents $K_l^+(z)$ in [Konno 2016, Corollary E.24], arising from a Gauss decomposition of an $\hat{L}$-matrix [Ding and Frenkel 1993].

Proof. Descending induction on i: for $i = N$ we have $K_N(z) = L_{NN}(z) = \hat{L}_N(z)$. Assume the statement for $i > N - t$ where $1 \leq t < N$. We need to prove the case $i = N - t$. Let α be the weight of v and let Y be the submodule of X generated by v. By [Cavalli 2001, Lemma 2.3], Y is linearly spanned by vectors of the form

$$L_{p_1q_1}(z_1)L_{p_2q_2}(z_2)\cdots L_{p_nq_n}(z_n)v,$$

where $1 \leq p_l \leq q_l \leq N$ and $z_l \in \mathbb{C}$ for $1 \leq l \leq n$. So $\alpha + \epsilon_p - \epsilon_q \notin \mathrm{wt}(Y)$ for $1 \leq p < q \leq N$, and any nonzero vector $\omega \in Y[\alpha]$ is singular. Apply $\mathcal{D}_k(z)$ to ω. At the right-hand side of (1.6) only the term $\sigma = \mathrm{Id}$ is nonzero and equal to $\hat{L}_N(z)\hat{L}_{N-1}(z+\hbar)\cdots\hat{L}_{N-k+1}(z+(k-1)\hbar)\omega$ by (1.7). It follows that

$$\begin{aligned}
\mathcal{D}_{t+1}(z)v &= \underline{\hat{L}_N(z)}\hat{L}_{N-1}(z+\hbar)\cdots\hat{L}_{N-t+1}(z+(t-1)\hbar)\hat{L}_{N-t}(z+t\hbar)v \\
&= K_N(z)\underline{\hat{L}_{N-1}(z+\hbar)}\cdots\hat{L}_{N-t+1}(z+(t-1)\hbar)\hat{L}_{N-t}(z+t\hbar)v \\
&\ \vdots \\
&= K_N(z)K_{N-1}(z+\hbar)\cdots K_{N-t+1}(z+(t-1)\hbar)\hat{L}_{N-t}(z+t\hbar)v.
\end{aligned}$$

Here we applied the induction hypothesis to $N, N-1, \ldots, N-t+1$ successively to singular vectors to the right of the underlines. Since the $K_l(z)$ are invertible, in view of (1.8) we must have $\hat{L}_{N-t}(z+t\hbar)v = K_{N-t}(z+t\hbar)v$. $\square$

We extend the q-character theory of H. Knight and Frenkel and Reshetikhin to category $\mathcal{O}$, as in [Felder and Zhang 2017, §3]. Take the product group $\mathcal{M}_{\mathrm{w}} := \mathcal{M} \times \mathfrak{h}$, by viewing $\mathfrak{h}$ as an additive group. Let $\varpi : \mathcal{M}_{\mathrm{w}} \twoheadrightarrow \mathfrak{h}$ be the projection to the second component.

As in [Hernandez and Leclerc 2016, §3.2], let $\mathcal{M}_{\mathrm{t}}$ be the set of formal sums $\sum_{f \in \mathcal{M}_{\mathrm{w}}} c_f \boldsymbol{f}$ with integer coefficients $c_f \in \mathbb{Z}$ such that for $\mu \in \mathfrak{h}$, all but finitely many c_f with $\varpi(\boldsymbol{f}) = \mu$ is zero; the set $\{\varpi(\boldsymbol{f}) : c_f \neq 0\}$ is contained in a finite union of cones $\nu + \boldsymbol{Q}_-$ with $\nu \in \mathfrak{h}$. Make $\mathcal{M}_{\mathrm{t}}$ into a ring; addition is the usual one of formal sums and multiplication is induced from that of $\mathcal{M}_{\mathrm{w}}$.

Definition 1.9. Let X be in category $\mathcal{O}$. For $\mu \in \mathrm{wt}(X)$, since $X[\mu]$ equipped with the difference operators $\mathcal{D}_k(z)$ is in category $\mathcal{F}_{\mathrm{mer}}$, in the Grothendieck group of which we have $[X[\mu]] = \sum_{i=1}^{\dim X[\mu]} [\mathbb{M}(\boldsymbol{f}^{\mu,i})]$ where $\boldsymbol{f}^{\mu,i} \in \mathcal{M}$ for $1 \leq i \leq \dim X[\mu]$. Each of the $(\boldsymbol{f}^{\mu,i};\mu) \in \mathcal{M}_{\mathrm{w}}$ is called an *e-weigth* of X. Let $\mathrm{wt}_e(X)$ be the set of e-weigths of X. The *q-character* of X is defined to be

$$\chi_q(X) := \sum_{\mu \in \mathrm{wt}(X)} \sum_{i=1}^{\dim X[\mu]} (\boldsymbol{f}^{\mu,i};\mu) \in \mathcal{M}_{\mathrm{t}}.$$

Proposition 1.10. *Let X and Y be in category $\mathcal{O}$. The $\mathcal{E}$-module $X \overline{\otimes} Y$ is also in category $\mathcal{O}$ and $\chi_q(X \overline{\otimes} Y) = \chi_q(X)\chi_q(Y)$.*

Proof. Clearly $X \overline{\otimes} Y$ is in category $\tilde{\mathcal{O}}$. Let us verify property (M3) of Definition 1.4. The idea is almost the same as that of [Felder and Zhang 2017, Proposition 3.9], which in turn followed [Frenkel and Reshetikhin 1999, §2.4]. For $\alpha, \beta \in \mathfrak{h}$, let us choose ordered bases $(v_i^{\alpha})_{1 \leq i \leq p_\alpha}$ and $(w_j^{\beta})_{1 \leq j \leq q_\beta}$ for $X[\alpha]$

and $Y[\beta]$, respectively, satisfying (M3). Note that $(v_i^\alpha \overline{\otimes} w_j^\beta)_{\alpha,\beta,i,j}$ forms a basis $\mathcal{B}$ of $X\overline{\otimes}Y$. Choose a partial order $\trianglelefteq$ on $\mathcal{B}$ with the properties

$$v_i^\alpha \overline{\otimes} w_j^\beta \trianglelefteq v_r^\alpha \overline{\otimes} w_s^\beta \quad \text{if } i \le r \text{ and } j \le s,$$
$$v_i^\alpha \overline{\otimes} w_j^\beta \vartriangleleft v_r^\gamma \overline{\otimes} w_s^\delta \quad \text{if } \gamma \prec \alpha \text{ and } \beta \prec \delta.$$

For $1 \le k \le N$, by Corollary 1.2, $\mathcal{D}_k^{X\overline{\otimes}Y}(z)(v_r^\gamma \overline{\otimes} w_s^\delta) = \mathcal{D}_k^X(z)v_r^\gamma \overline{\otimes} \mathcal{D}_k^Y(z) w_s^\delta + Z$ where Z is a finite sum of vectors in $X[\gamma + \varpi_{N-k} + \eta]\overline{\otimes} Y[\delta - \varpi_{N-k} - \eta]$ for $\eta \in \mathfrak{h}$ such that $-\varpi_{N-k} \prec \eta$. So the ordered basis $\mathcal{B}$ induces an upper triangular matrix for $\mathcal{D}_k^{X\overline{\otimes}Y}(z)$ whose diagonal entry associated to $v_r^\gamma \overline{\otimes} w_s^\delta$ is the product of those associated to v_r^γ and w_s^δ. This implies (M3) for the weight spaces $(X\overline{\otimes}Y)[\alpha]$ with bases $\mathcal{B} \cap (X\overline{\otimes}Y)[\alpha]$ and the multiplicative formula of q-characters as well. □

For $f(z) \in \mathbb{M}_{\mathbb{C}}^\times$ and $\alpha \in \mathfrak{h}$ we make the simplifications

$$f(z) := (f(z), \dots, f(z); 0) \quad \text{and} \quad e^\alpha := (1, \dots, 1; \alpha) \in \mathcal{M}_{\mathrm{w}}.$$

Definition 1.11. Let $1 \le i, k \le N$ such that $i \neq N$. Set $\ell_k := (N-k-1)/2$. For $a \in \mathbb{C}$, define the following elements of $\mathcal{M}_{\mathrm{w}}$:

$$A_{i,a} := \Big(\underbrace{1, \dots, 1}_{i-1}, \frac{\theta(z+(a-\ell_i)\hbar)}{\theta(z+(a-\ell_i-1)\hbar)}, \frac{\theta(z+(a-\ell_i)\hbar)}{\theta(z+(a-\ell_i+1)\hbar)}, \underbrace{1, \dots, 1}_{N-i-1}; \alpha_i\Big).$$

$$\Psi_{k,a} := \big(\underbrace{\theta(z+(a-\ell_k)\hbar), \dots, \theta(z+(a-\ell_k)\hbar)}_{k}, \underbrace{1, \dots, 1}_{N-k}; a\varpi_k\big).$$

$$Y_{k,a} := \Bigg(\underbrace{\frac{\theta\big(z+\big(a-\ell_k+\frac{1}{2}\big)\hbar\big)}{\theta\big(z+\big(a-\ell_k-\frac{1}{2}\big)\hbar\big)}, \dots, \frac{\theta\big(z+\big(a-\ell_k+\frac{1}{2}\big)\hbar\big)}{\theta\big(z+\big(a-\ell_k-\frac{1}{2}\big)\hbar\big)}}_{k}, \underbrace{1, \dots, 1}_{N-k}; \varpi_k\Bigg).$$

$$\boxed{k}_a := \Bigg(\underbrace{\frac{\theta(u+\hbar)\theta(u-\hbar)}{\theta(u)^2}, \dots, \frac{\theta(u+\hbar)\theta(u-\hbar)}{\theta(u)^2}}_{k-1}, \frac{\theta(u+\hbar)}{\theta(u)}, \underbrace{1, \dots, 1}_{N-k}; \epsilon_k\Bigg)\Bigg|_{u=z+a\hbar}.$$

$A_{i,a}$, $Y_{k,a}$ and $\Psi_{k,a}$ are elliptic analogs of generalized simple roots, fundamental ℓ-weight [Frenkel and Reshetikhin 1999] and prefundamental weight [Hernandez and Jimbo 2012]. Set $c_{ij} := 2\delta_{ij} - \delta_{i,j\pm 1}$ and $Y_{0,a} = \Psi_{0,a} := 1$. Then (in the products $1 \le j \le N$)

$$Y_{k,a} = \frac{\Psi_{k,a+\frac{1}{2}}}{\Psi_{k,a-\frac{1}{2}}} \quad \text{and} \quad A_{i,a} = \prod_j \frac{\Psi_{j,a+\frac{c_{ij}}{2}}}{\Psi_{j,a-\frac{c_{ij}}{2}}} = Y_{i,a-\frac{1}{2}} Y_{i,a+\frac{1}{2}} \prod_{j=i\pm 1} Y_{j,a}^{-1}. \tag{1.9}$$

The interplay of A, Ψ is the source of the three-term Baxter's Relation (5.32) in category $\mathcal{O}$. Note that A and Y can also be written in terms of $\boxed{k}$:

$$A_{i,a} = \boxed{i}_{a-\ell_i} \boxed{i+1}^{-1}_{a-\ell_i}, \qquad Y_{k,a} = \prod_{j=1}^{k} \boxed{j}_{a-\ell_k-\frac{1}{2}+j-k}, \tag{1.10}$$

$$\Psi_{N,a} = \theta\big(z + \big(a + \tfrac{1}{2}\big)\hbar\big), \qquad Y_{N,a} = \frac{\theta(z+(a+1)\hbar)}{\theta(z+a\hbar)}. \tag{1.11}$$

1D. ***Vector representations.*** Let $\boldsymbol{V} := \oplus_{i=1}^{N} \mathbb{M} v_i$ with $\mathfrak{h}$-grading $\boldsymbol{V}[\epsilon_i] = \mathbb{M} v_i$. Rewriting (1.1) in the form of (1.2), we obtain an $\mathcal{E}$-module structure on $\boldsymbol{V}$:

$$L_{ij}(z) v_k = \sum_{l=1}^{N} \frac{\theta(z+\hbar)}{\theta(z)} R_{il}^{jk}(z;\lambda) v_l.$$

The factor $\theta(z+\hbar)/\theta(z)$ is used to simplify the q-character, see (1.12).

If $i \le N-k+1$, since $L_{pq}(z)v_i = 0$ for all $N \ge p > q > N-k$, only the term $\sigma = \mathrm{Id}$ in (1.6) survives and

$$\begin{aligned}
\mathcal{D}_k(z) v_i &= \frac{\mu_{\boldsymbol{r}}(\Theta_k(\lambda))}{\mu_{\boldsymbol{l}}(\Theta_k(\lambda))} \prod_{j=N}^{N-k+1} L_{jj}(z+(N-j)\hbar) v_i = g_k^i(z;\lambda) v_i, \\
g_k^i(z;\lambda) &= \prod_{j>N-k} \frac{\theta(\lambda_{ij}+\hbar)}{\theta(\lambda_{ij})} \quad \text{for } i \le N-k, \\
g_k^{N-k+1}(z;\lambda) &= \frac{\theta(z+k\hbar)}{\theta(z+(k-1)\hbar)}.
\end{aligned}$$

If $i > N-k+1$, then $L_{N-k+1,i}(z) v_{N-k+1} = \theta(\hbar)\theta(z+\lambda_{N-k+1,i}) v_i / (\theta(z)\theta(\lambda_{N-k+1,i}))$. By Corollary 1.2, $\mathcal{D}_k(z) v_i = g_k^{N-k+1}(z;\lambda) v_i$. Let us perform a change of basis (see [Konno 2016, (E.2)])

$$\tilde{v}_i := v_i \prod_{l>i} \theta(\lambda_{il}+\hbar) \in \boldsymbol{V}[\epsilon_i].$$

After a direct computation, we obtain

$$\mathcal{D}_k(z)\tilde{v}_i = \tilde{v}_i \times \begin{cases} 1 & \text{for } i \le N-k, \\ \frac{\theta(z+k\hbar)}{\theta(z+(k-1)\hbar)} & \text{for } i > N-k. \end{cases} \tag{1.12}$$

The basis $\{\tilde{v}_1 < \tilde{v}_2 < \cdots < \tilde{v}_N\}$ of $\boldsymbol{V}$ satisfies property (M3) of Definition 1.4, so $\boldsymbol{V}$ is in category $\mathcal{O}$. For $a \in \mathbb{C}$, let $\boldsymbol{V}(a)$ be the pullback of $\boldsymbol{V}$ by the spectral parameter shift Φ_a in (1.4). Naturally $\boldsymbol{V}(a)$ is in category $\mathcal{O}$; it is called a *vector representation.* Combining with (1.8) we have

$$\chi_q(\boldsymbol{V}(a)) = \boxed{1}_a + \boxed{2}_a + \cdots + \boxed{N}_a.$$

1E. ***Highest weight modules.*** Let X be in category $\tilde{\mathcal{O}}$. A nonzero weight vector $v \in X[\alpha]$ is called a *highest weight vector* if it is singular and $\hat{L}_k(z)v = f_k(z)v$ for $1 \le k \le N$; here the $f_k(z) \in \mathbb{M}_{\mathbb{C}}^{\times}$. Call $(f_1(z), f_2(z), \ldots, f_N(z); \alpha) \in \mathcal{M}_{\mathrm{w}}$ the highest weight of v; by Lemma 1.8 it belongs to $\mathrm{wt}_e(X)$ if X is in category $\mathcal{O}$.

If there is a highest weight vector $v \in X[\alpha]$ of X which also generates the whole module, then X is called a *highest weight module*, see [Cavalli 2001, Definition 2.1]. In this case, by [Cavalli 2001, Lemma 2.3], $X[\alpha] = \mathbb{M}v$ and $\mathrm{wt}(X) \subseteq \alpha + \boldsymbol{Q}_-$, so the highest weight vector is unique up to scalar product. This implies that X admits a unique irreducible quotient. The highest weight of v is also called the highest weight of X; it is of multiplicity one in $\chi_q(X)$ if X is in category $\mathcal{O}$.

All irreducible modules in category $\mathcal{O}$ are of highest weight.

By [Cavalli 2001, Theorem 2.8] two irreducible highest weight modules in category $\tilde{\mathcal{O}}$ are isomorphic if and only if their highest weights are identical in $\mathcal{M}_{\mathrm{w}}$; all singular vectors of an irreducible highest weight module in category $\tilde{\mathcal{O}}$ are proportional. It follows that the q-characters distinguish irreducible modules in category $\mathcal{O}$.

Let $\mathcal{R}$ be the set of $\boldsymbol{d} \in \mathcal{M}_{\mathrm{w}}$ which appears as the highest weight of an irreducible module in category $\mathcal{O}$. For $\boldsymbol{d} \in \mathcal{R}$, let us fix an irreducible module $S(\boldsymbol{d})$ in category $\mathcal{O}$ of highest weight $\boldsymbol{d}$. Let $\mathcal{R}_0$ and $\mathcal{R}_{\mathrm{fd}}$ be the set of $\boldsymbol{d} \in \mathcal{R}$ such that $S(\boldsymbol{d})$ is one-dimensional and finite-dimensional, respectively.

We shall need the *completed* Grothendieck group $K_0(\mathcal{O})$. Its definition is the same as that in [Hernandez and Leclerc 2016, §3.2]; elements are formal sums $\sum_{\boldsymbol{d}\in\mathcal{R}} c_{\boldsymbol{d}}[S(\boldsymbol{d})]$ with integer coefficients $c_{\boldsymbol{d}} \in \mathbb{Z}$ such that $\oplus_{\boldsymbol{d}} S(\boldsymbol{d})^{\oplus|c_{\boldsymbol{d}}|}$ is in category $\mathcal{O}$ and addition is the usual one of formal sums. As in the case of Kac–Moody algebras [Kac 1990, §9.6], for $\boldsymbol{d} \in \mathcal{R}$ the multiplicity $m_{\boldsymbol{d},X}$ of $S(\boldsymbol{d})$ in any object X of category $\mathcal{O}$ is well-defined due to Definition 1.7, and $[X] := \sum_{\boldsymbol{d}} m_{\boldsymbol{d},X}[S(\boldsymbol{d})]$ belongs to $K_0(\mathcal{O})$. In the case $X = S(\boldsymbol{d})$ the right-hand side is simply $[S(\boldsymbol{d})]$ as $m_{\boldsymbol{e},S(\boldsymbol{d})} = \delta_{\boldsymbol{d},\boldsymbol{e}}$ for $\boldsymbol{e} \in \mathcal{R}$.

By Proposition 1.10, $K_0(\mathcal{O})$ is endowed with a ring structure with multiplication $[X][Y] = [X \overline{\otimes} Y]$ for X and Y in category $\mathcal{O}$. Together with Definition 1.9, we obtain

Corollary 1.12. *The assignment* $[X] \mapsto \chi_q(X)$ *defines an injective morphism of rings* $\chi_q : K_0(\mathcal{O}) \to \mathcal{M}_{\mathrm{t}}$. *In particular*, $K_0(\mathcal{O})$ *is commutative.*

Let $\mathcal{O}_{\mathrm{fd}}$ be the full subcategory of $\mathcal{O}$ consisting of finite-dimensional modules. It is abelian and monoidal. Its Grothendieck ring $K_0(\mathcal{O}_{\mathrm{fd}})$ admits a $\mathbb{Z}$-basis $[S(\boldsymbol{d})]$ for $\boldsymbol{d} \in \mathcal{R}_{\mathrm{fd}}$, and is commutative as a subring of $K_0(\mathcal{O})$.

By Proposition 1.10, $S(\boldsymbol{d})\overline{\otimes}S(\boldsymbol{e})$ admits an irreducible subquotient $S(\boldsymbol{de})$, so the three sets $\mathcal{R} \supset \mathcal{R}_{\mathrm{fd}} \supset \mathcal{R}_0$ are submonoids of $\mathcal{M}_{\mathrm{w}}$.

Lemma 1.13. *Let* $\boldsymbol{d} = ((f_k(z))_{1\le k\le N}; \mu) \in \mathcal{M}_{\mathrm{w}}$.

(i) *Suppose* $\boldsymbol{d} \in \mathcal{R}$. *Then for* $1 \le k < N$ *we have*

$$\frac{f_k(z)}{f_{k+1}(z)} = c \prod_{l=1}^{n} \frac{\theta(z + a_l\hbar)}{\theta(z + b_l\hbar)} \quad \textit{and} \quad \mu_{k,k+1} = \sum_{l=1}^{n} (a_l - b_l)$$

for certain $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n \in \mathbb{C}$ *and* $c \in \mathbb{C}^\times$.

(ii) *If* $\boldsymbol{d} \in \mathcal{R}_{\mathrm{fd}}$, *then* (*i*) *holds and after a rearrangement of the* a_l, b_l *we have* $a_l - b_l \in \mathbb{Z}_{\ge 0} + \hbar^{-1}\Gamma$ *for all* l.

(iii) $\boldsymbol{d} \in \mathcal{R}_0$ *if and only if* (*ii*) *holds with* $a_l - b_l \in \hbar^{-1}\Gamma$ *for all* l.

Proof. Results (i) and (iii) are essentially [Felder and Varchenko 1996b, Theorems 6 and 9], which can be proved as in [Felder and Zhang 2017, Theorem 4.1] by replacing L_{+-}, L_{-+} therein with $L_{k,k+1}$, $L_{k+1,k}$. Result (ii) comes from either [Cavalli 2001, Theorem 5.1] or [Felder and Zhang 2017, Corollary 4.6]. □

As examples $Y_{N,a}, \Psi_{N,a} \in \mathcal{R}_0$. Call an e-weigth $\boldsymbol{e} \in \mathcal{M}_{\mathrm{w}}$ *dominant* or *rational* if $\boldsymbol{e} = \boldsymbol{dm}$ where $\boldsymbol{d} \in \mathcal{R}_0$ and $\boldsymbol{m}$ is a product of the $Y_{i,a}$ or the $\Psi_{i,a}\Psi_{i,b}^{-1}$, respectively with $a, b \in \mathbb{C}$ and $1 \le i \le N$. Lemma 1.13 implies that all elements of $\mathcal{R}_{\mathrm{fd}}$ or $\mathcal{R}$ are dominant or rational, respectively.

Theorem 1.14 [Cavalli 2001]. $\mathcal{R}_{\mathrm{fd}}$ *is the set of dominant e-weigths.*

Proof. It suffices to prove $Y_{n,a} \in \mathcal{R}_{\mathrm{fd}}$ for $1 \le n < N$. Note that $V(w)$ and γ from [Cavalli 2001, (1.19)] correspond to our $\boldsymbol{V}(-w/\hbar)\overline{\otimes} S(\theta(z-w)/\theta(z-w-\hbar))$ and $-\hbar$. Let us rephrase [Cavalli 2001, Theorem 4.4] in terms of the $\boldsymbol{V}$ by replacing z and w in [loc. cit.] with $-a\hbar$ and z.

The $\mathcal{E}$-module $\boldsymbol{V}(a)\overline{\otimes}\boldsymbol{V}(a+1)\overline{\otimes}\cdots\overline{\otimes}\boldsymbol{V}(a+n-1)$ admits an irreducible quotient S which contains a singular vector ω of weight ϖ_n such that $\hat{L}_k(z)\omega = \Lambda_k(z)g_k(\lambda)\omega$ where for $1 \le k \le N$ (set $\delta_{k\le n} = 1$ if $1 \le k \le n$ and $\delta_{k \le n} = 0$ if $n < k \le N$)

$$\Lambda_k(z) = \frac{\theta(z+(a+1)\hbar)}{\theta(z+a\hbar)}\frac{\theta(z+(a+n)\hbar)}{\theta(z+(a+n-\delta_{k\le n})\hbar)} \quad \text{for } g_k(\lambda) \in \mathbb{M}^{\times}.$$

As a subquotient of tensor products of vector representations, S belongs to category $\mathcal{O}$. By Lemma 1.6, the $g_k(\lambda)$ can be gauged away, and the highest weight of S is $\Lambda_N(z)Y_{n,a-1+(N+n)/2} \in \mathcal{R}_{\mathrm{fd}}$. This implies $Y_{n,a-1+(N+n)/2} \in \mathcal{R}_{\mathrm{fd}}$. □

A sharp difference from the affine case [Hernandez and Jimbo 2012, Theorem 3.11] is that category $\mathcal{O}$ does not admit prefundamental modules, i.e., $\Psi_{r,a} \notin \mathcal{R}$ if $r < N$. One might want to introduce a larger category with well-behaved q-character theory, so that modules of highest weight $\Psi_{r,a}$ exist. For this purpose, the finite-dimensionality of weight spaces should be dropped because of [Felder and Varchenko 1996b, Theorem 9]. The recent work [Bittmann 2017] on representations of affine quantum groups is in this direction.

1F. ***Young tableaux and q-character formula.*** Let $\mathcal{P}$ be the set partitions with at most N parts, i.e., N-tuples of nonnegative integers $(\mu_1 \ge \mu_2 \ge \cdots \ge \mu_N)$. To such a partition we associate a Young diagram

$$Y_\mu := \{(i, j) \in \mathbb{Z}^2 \mid 1 \le i \le N, 1 \le j \le \mu_i\},$$

and the set $\mathcal{B}_\mu$ of Young tableaux of shape Y_μ. We put the Young diagram at the northwest position so that $(i, j) \in Y_\mu$ corresponds to the box at the i-th row (from bottom to top) and j-th column (from right to left). By a tableau we mean a function $T : Y_\mu \to \{1 < 2 < \cdots < N\}$ weakly increasing at each row (from left to right) and strictly increasing at each column (from top to bottom).

For $\mu = (\mu_1 \ge \mu_2 \ge \cdots \ge \mu_N) \in \mathcal{P}$ and $a \in \mathbb{C}$, we have the dominant e-weigth

$$\theta_{\mu,a} := \left(\frac{\theta(z+(a+\mu_1)\hbar)}{\theta(z+a\hbar)}, \frac{\theta(z+(a+\mu_2)\hbar)}{\theta(z+a\hbar)}, \ldots, \frac{\theta(z+(a+\mu_N)\hbar)}{\theta(z+a\hbar)}; \sum_{j=1}^{N} \mu_j \epsilon_j\right).$$

The associated irreducible module in category $\mathcal{O}_{\mathrm{fd}}$ is denoted by $S_{\mu,a}$.

Theorem 1.15. *Let $\mu \in \mathcal{P}$ and $a \in \mathbb{C}$. For the $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$-module $S_{\mu,a}$ we have*

$$\chi_q(S_{\mu,a}) = \sum_{T\in\mathcal{B}_\mu} \prod_{(i,j)\in Y_\mu} \boxed{T(i,j)}_{a+j-i} \in \mathcal{M}_{\mathrm{t}}. \tag{1.13}$$

For $\nu = (1 \geq 0 \geq 0 \geq \cdots \geq 0)$, we have $S_{\nu,a} \cong \boldsymbol{V}(a)$, and (1.13) specializes to the q-character formula in Section 1D. As an illustration of the theorem, let $N = 3$ and $\mu = (2 \geq 1 \geq 0)$. Pictorially $\mathcal{B}_\mu$ consists of

$$\begin{array}{cc} & \boxed{1} \\ \boxed{1} & \boxed{2} \end{array}, \quad \begin{array}{cc} & \boxed{1} \\ \boxed{1} & \boxed{3} \end{array}, \quad \begin{array}{cc} & \boxed{1} \\ \boxed{2} & \boxed{2} \end{array}, \quad \begin{array}{cc} & \boxed{1} \\ \boxed{2} & \boxed{3} \end{array}, \quad \begin{array}{cc} & \boxed{1} \\ \boxed{3} & \boxed{3} \end{array}, \quad \begin{array}{cc} & \boxed{2} \\ \boxed{1} & \boxed{3} \end{array}, \quad \begin{array}{cc} & \boxed{2} \\ \boxed{2} & \boxed{3} \end{array}, \quad \begin{array}{cc} & \boxed{2} \\ \boxed{3} & \boxed{3} \end{array}.$$

The fourth tableau gives rise to the term $\boxed{2}_{a+1}\boxed{3}_{a}\boxed{1}_{a-1}$ in $\chi_q(S_{\mu,a})$.

Remark 1.16. Theorem 1.15 is an elliptic analog of the q-character formula for affine quantum groups [Frenkel and Mukhin 2002, Lemma 4.7]. In principle it can be deduced from the functor of Gautam and Toledano Laredo [2017a, § 6]. This is a functor from finite-dimensional representations of affine quantum groups to those of elliptic quantum groups (including our $S_{\mu,a}$), and it respects affine and elliptic q-characters.

The proof of Theorem 1.15 will be given in Section 2D. It is in the spirit of [Frenkel and Mukhin 2002], based on small elliptic quantum groups of Tarasov and Varchenko [2001].

2. Small elliptic quantum group and evaluation modules

The aim of this section is to prove Corollary 1.2 and Theorem 1.15.

Recall that $\mathfrak{h}$ is the $\mathbb{C}$-vector space generated by the ϵ_i for $1 \leq i \leq N$ subject to the relation

$$\epsilon_1 + \epsilon_2 + \cdots + \epsilon_N = 0.$$

For $1 \leq k \leq N$, define the $\mathbb{C}$-vector space $\mathfrak{h}_k$ to be the quotient of $\mathfrak{h}$ by $\epsilon_1 = \epsilon_2 = \cdots = \epsilon_{N-k} = 0$. (By convention $\mathfrak{h}_N = \mathfrak{h}$.) The quotient $\mathfrak{h} \twoheadrightarrow \mathfrak{h}_k$ induces an embedding $\mathbb{M}_{\mathfrak{h}_k} \hookrightarrow \mathbb{M}$.

Let $\mathcal{E}_k^{\mathfrak{h}}$ and $\mathcal{E}_k$ be the $\mathfrak{h}$-algebra and $\mathfrak{h}_k$-algebra, respectively, generated by the $L_{ij}(z)$ for $N-k < i, j \leq N$ subject to relation (1.2) with summations $N-k < p, q \leq N$. (This makes sense because the $R_{ij}^{pq}(z;\lambda)$ for $N-k < i, j, p, q \leq N$ belong to $\mathbb{M}_{\mathfrak{h}_k}$.) The following defines an $\mathfrak{h}_k$-algebra morphism

$$\Delta_k : \mathcal{E}_k \to \mathcal{E}_k \tilde{\otimes} \mathcal{E}_k \quad \text{and} \quad L_{ij}(z) \mapsto \sum_{p=N-k+1}^{N} L_{ip}(z) \tilde{\otimes} L_{pj}(z).$$

One has natural algebra morphisms $\mathcal{E}_k \to \mathcal{E}_k^{\mathfrak{h}} \to \mathcal{E}$ sending $L_{ij}(z)$ to itself; the second is an $\mathfrak{h}$-algebra morphism. $\mathcal{D}_1(z), \mathcal{D}_2(z), \ldots, \mathcal{D}_k(z)$ from (1.6) are well-defined in $\mathcal{E}_k^{\mathfrak{h}}$ and $\mathcal{E}_k$. Their images in $\mathcal{E}$ are the first k elliptic quantum minors.

2A. ***Proof of Corollary 1.2.*** The $\mathfrak{h}_k$-algebra with coproduct $(\mathcal{E}_k, \Delta_k)$ is isomorphic to the usual elliptic quantum group $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_k)$; here we view $\mathfrak{h}_k$ as a Cartan subalgebra of $\mathfrak{sl}_k$ so that $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_k)$ is an $\mathfrak{h}_k$-algebra. Under this isomorphism, by (1.6), $\mathcal{D}_k(z) \in \mathcal{E}_k$ corresponds to the k-th elliptic quantum minor of $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_k)$. So Theorem 1.1 can be applied to $(\mathcal{E}_k, \mathcal{D}_k(z), \Delta_k)$ and then to the algebra morphism $\mathcal{E}_k \to \mathcal{E}$. The first statement of the corollary is obvious, and the second is based on the fact that for $i, j > N-k$ the difference $\Delta - \Delta_k$ at $L_{ij}(z)$ is a finite sum over $\alpha \in \mathfrak{h}$ of elements in $\mathcal{E}_{\epsilon_i,\alpha} \widetilde{\otimes} \mathcal{E}_{\alpha,\epsilon_j}$ with $\epsilon_{N-k+1} \prec \alpha$ and so $\epsilon_i, \epsilon_j \prec \alpha$. $\square$

We believe $0 \neq \alpha + \varpi_{N-k} \in \boldsymbol{Q}_+$ in Corollary 1.2, as in [Damiani 1998, §7] and [Zhang 2016, §3].

2B. ***Small elliptic quantum group of Tarasov–Varchenko.*** Let us define the linear form $\lambda_i \in \mathfrak{h}^*$ by taking i-th component for $1 \leq i \leq N$,

$$x_1\epsilon_1 + x_2\epsilon_2 + \cdots + x_N\epsilon_N \mapsto x_i - \frac{1}{N}(x_1 + x_2 + \cdots + x_N).$$

The linear form λ_{ij} of Section 1 is $\lambda_i - \lambda_j$. For $\gamma \in \mathfrak{h}$ and $1 \leq i, j \leq N$, set $\gamma_i := \lambda_i(\gamma)$ and $\gamma_{ij} := \gamma_i - \gamma_j$ as *complex numbers*. We hope this is not to be confused with the previously defined *vectors* $\lambda_i \in \mathfrak{h}^*$ and $\epsilon_i, \alpha_i, \varpi_i \in \mathfrak{h}$.

Following [Tarasov and Varchenko 2001, §3], let $\mathbb{M}_2$ be the ring of meromorphic functions $f(\lambda^{\{1\}}, \lambda^{\{2\}})$ of $(\lambda^{\{1\}}, \lambda^{\{2\}}) \in \mathfrak{h} \oplus \mathfrak{h}$ whose location of singularities in $\lambda^{\{1\}}$ does not depend on $\lambda^{\{2\}}$ and vice versa. For brevity, we write $f(\lambda^{\{1\}})$ or $f(\lambda^{\{2\}})$ instead of $f(\lambda^{\{1\}}, \lambda^{\{2\}})$ if the function does not depend on the other variable.

Definition 2.1 [Tarasov and Varchenko 2001]. The *small elliptic quantum group* $\mathfrak{e} := \mathfrak{e}_{\tau,\hbar}(\mathfrak{sl}_N)$ is the algebra with generators $\mathbb{M}_2$ and t_{ij} for $1 \leq i, j \leq N$ and subject to relations: $\mathbb{M}_2$ is a subalgebra. For $f(\lambda^{\{1\}}, \lambda^{\{2\}}) \in \mathbb{M}_2$ and $1 \leq i, j, k, l \leq N$,

$$\begin{aligned} t_{ij} f(\lambda^{\{1\}}, \lambda^{\{2\}}) &= f(\lambda^{\{1\}} + \hbar\epsilon_i, \lambda^{\{2\}} + \hbar\epsilon_j) t_{ij}, \\ t_{ij} t_{ik} &= t_{ik} t_{ij}, \\ t_{ik} t_{jk} &= \frac{\theta(\lambda^{\{1\}}_{ij} - \hbar)}{\theta(\lambda^{\{1\}}_{ij} + \hbar)} t_{jk} t_{ik} && \text{for } i \neq j, \\ \frac{\theta(\lambda^{\{2\}}_{jl} - \hbar)}{\theta(\lambda^{\{2\}}_{jl})} t_{ij} t_{kl} - \frac{\theta(\lambda^{\{1\}}_{ik} - \hbar)}{\theta(\lambda^{\{1\}}_{ik})} t_{kl} t_{ij} &= \frac{\theta(\lambda^{\{1\}}_{ik} + \lambda^{\{2\}}_{jl})\theta(-\hbar)}{\theta(\lambda^{\{1\}}_{ik})\theta(\lambda^{\{2\}}_{jl})} t_{il} t_{kj} && \text{for } i \neq k \text{ and } j \neq l. \end{aligned}$$

Here $\lambda^{\{1\}}_{ij} = \lambda^{\{1\}}_i - \lambda^{\{1\}}_j$ and $\lambda^{\{2\}}_{ij} = \lambda^{\{2\}}_i - \lambda^{\{2\}}_j$.

The small elliptic quantum group $\mathfrak{e}$ is equipped with an $\mathfrak{h}$-algebra structure: Elements of $\mathbb{M}_2$ are of bidegree $(0, 0)$. t_{ij} is of bidegree (ϵ_j, ϵ_i). The moment maps are given by

$$\mu_{\boldsymbol{l}}(g(\lambda)) = g(\lambda^{\{2\}}) \quad \text{and} \quad \mu_{\boldsymbol{r}}(g(\lambda)) = g(\lambda^{\{1\}}).$$

Let X be an object of $\mathcal{V}_{\mathrm{ft}}$. A representation ρ of $\mathfrak{e}$ on X is a morphism of $\mathfrak{h}$-algebras $\rho : \mathfrak{e} \to \mathbb{D}^X$ such that for $f(\lambda^{\{1\}}, \lambda^{\{2\}}) \in \mathbb{M}_2$ and $v \in X[\gamma]$,

$$\rho(f(\lambda^{\{1\}}, \lambda^{\{2\}})) : v \mapsto f(\lambda, \lambda + \hbar\gamma)v.$$

A morphism of two representations (ρ, X) and (σ, Y) is a morphism $\Phi : X \to Y$ in $\mathcal{V}_{\mathrm{ft}}$ such that $\Phi\rho(t_{ij}) = \sigma(t_{ij})\Phi$ for $1 \le i, j \le N$. Let rep be the category of $\mathfrak{e}$-modules. The following result is [Tarasov and Varchenko 2001, Corollary 3.4].

Corollary 2.2. *Let (ρ, X) be a representation of $\mathfrak{e}$ on X. Then, for $a \in \mathbb{C}$,*

$$L_{ij}(z) \mapsto \frac{\theta(z + a\hbar + \lambda_i^{\{2\}} - \lambda_j^{\{1\}})}{\theta(z + a\hbar)}\rho(t_{ji})$$

defines a representation of $\mathcal{E}$ on X, called the evaluation module $X(a)$.

There is a flip of the subscripts i and j because the bidegrees of L_{ij} and t_{ij} are flips of each other. See also [Tarasov and Varchenko 2001, (3.6)] where $\mathcal{T}_{ij}(u)$ comes from t_{ji}.

$X \mapsto X(a)$ defines a functor $\mathrm{ev}_a : \mathrm{rep} \to \mathrm{Rep}$. Let $\mathcal{F}$ be the full subcategory of rep whose objects are finite-dimensional $\mathfrak{e}$-modules X with $X(x)$ being in category $\mathcal{O}$. Then ev_a restricts to a functor of abelian categories $\mathcal{F} \to \mathcal{O}_{\mathrm{fd}}$, and induces an injective morphism of Grothendieck groups $K_0(\mathcal{F}) \hookrightarrow K_0(\mathcal{O}_{\mathrm{fd}})$.

For $1 \le k \le N$, define $\hat{t}_k \in \mathfrak{e}$ in the same way as (1.7):[8]

$$\hat{t}_N(z) := t_{NN} \quad \text{and} \quad \hat{t}_k(z) = t_{kk} \prod_{j=k+1}^{N} \frac{\mu_{\boldsymbol{r}}(\theta(\lambda_{kj}))}{\mu_{\boldsymbol{l}}(\theta(\lambda_{kj}))}.$$

Let $\mu \in \mathfrak{h}$. There exists a unique (up to isomorphism) irreducible $\mathfrak{e}$-module V_μ with the property V_μ admits a nonzero vector v of weight μ such that $\hat{t}_k v = v$, $t_{ij} v = 0$ for $1 \le i, j, k \le N$ and $j < k$; it is called *standard* in [Tarasov and Varchenko 2001, §4]. Let L_μ denote the complex irreducible module over the simple Lie algebra $\mathfrak{sl}_N$ of highest weight μ. For $\nu \in \mathfrak{h}$, let $d_\mu[\nu] = \dim_{\mathbb{C}} L_\mu[\nu]$ where $L_\mu[\nu]$ is the weight space of weight ν.

Theorem 2.3 [Tarasov and Varchenko 2001, Theorem 5.9]. *The $\mathfrak{e}$-module V_μ is finite-dimensional if and only if $\mu_{ij} \in \mathbb{Z}_{\ge 0} + \hbar^{-1}\Gamma$ for $1 \le i < j \le N$. If $\tilde{\mu} \in \mathfrak{h}$ is such that $\mu_{ij} - \tilde{\mu}_{ij} \in \hbar^{-1}\Gamma$ and $\tilde{\mu}_{ij} \in \mathbb{Z}_{\ge 0}$ for $i < j$, then* $\dim V_\mu[\mu + \gamma] = d_{\tilde{\mu}}[\tilde{\mu} + \gamma]$ *for $\gamma \in \boldsymbol{Q}_-$.*

In the theorem $\tilde{\mu}$ is uniquely determined by μ since $\mathbb{Z} \cap h^{-1}\Gamma = \{0\}$. Such an $\mathfrak{e}$-module V_μ is in category $\mathcal{F}$. Indeed, the evaluation module $V_\mu(a)$ is irreducible in category $\tilde{\mathcal{O}}$ of highest weight

$$\left(\frac{\theta(z + (\mu_1 + a)\hbar)}{\theta(z + a\hbar)}, \frac{\theta(z + (\mu_2 + a)\hbar)}{\theta(z + a\hbar)}, \ldots, \frac{\theta(z + (\mu_N + a)\hbar)}{\theta(z + a\hbar)}; \mu\right).$$

One checks that such an e-weigth is dominant. So $V_\mu(a)$ is in category $\mathcal{O}_{\mathrm{fd}}$ by Theorem 1.14. The *character* $\chi(V_\mu)$ of V_μ is $\sum_\gamma d_{\tilde{\mu}}[\tilde{\mu} + \gamma]e^{\mu+\gamma} \in \mathcal{M}_{\mathrm{t}}$.

[8] The $\hat{t}_a$ are slightly different from the $\hat{t}_{aa}$ in [Tarasov and Varchenko 2001, (4.1)]. Yet they play the same role.

The isomorphism classes $[V_\mu]$ where $\mu \in \mathfrak{h}$ and $\mu_{ij} \in \mathbb{Z}_{\geq 0} + \hbar^{-1}\Gamma$ for $i < j$ form a $\mathbb{Z}$-basis of $K_0(\mathcal{F})$, and $[V_\mu] \mapsto \chi(V_\mu)$ extends uniquely to a morphism of abelian groups $\chi : K_0(\mathcal{F}) \to \mathcal{M}_{\mathrm{t}}$, which is injective thanks to the linear independence of characters of irreducible representations of the simple Lie algebra $\mathfrak{sl}_N$.

2C. *Category $\mathcal{O}'_{\mathrm{fd}}$*. We are going to prove Theorem 1.15 by induction on N. The idea is to view the irreducible $\mathcal{E}$-module $S_{\mu,a}$ as an $\mathcal{E}^{\mathfrak{h}}_{N-1}$-module and to apply the induction hypothesis. For this purpose, we need to adapt carefully the definitions of finite-dimensional module category $\mathcal{O}_{\mathrm{fd}}$ and its q-characters in Section 1C to $\mathcal{E}^{\mathfrak{h}}_{N-1}$. To distinguish with $\mathcal{E}$ and to simplify notations, we shall add a prime (instead of the index $N-1$) to objects related to $\mathcal{E}^{\mathfrak{h}}_{N-1}$. Notably $\mathfrak{h}' := \mathfrak{h}_{N-1}$.

We define category $\mathcal{O}'_{\mathrm{fd}}$. An object is a *finite-dimensional* $\mathfrak{h}$-graded vector space X (viewed as an object of category $\mathcal{V}_{\mathrm{ft}}$) endowed with difference operators $L^X_{ij}(z) : X \to X$ of bidegree (ϵ_i, ϵ_j) for $2 \leq i, j \leq N$ depending on $z \in \mathbb{C}$ such that:

(M1') There exists a basis of X with respect to which the matrix entries of the difference operators $L^X_{ij}(z)$ are meromorphic functions of $(z, \lambda) \in \mathbb{C} \times \mathfrak{h}$.

(M2') $L_{ij}(z) \mapsto L^X_{ij}(z)$ defines an $\mathfrak{h}$-algebra morphism $\mathcal{E}^{\mathfrak{h}}_{N-1} \to \mathbb{D}^X$.

(M3') X admits an ordered weight basis with respect to which the matrices of the difference operators $\mathcal{D}^X_l(z)$ for $1 \leq l < N$ are upper triangular and their diagonal entries are nonzero meromorphic functions of $z \in \mathbb{C}$.

A morphism in category $\mathcal{O}'_{\mathrm{fd}}$ a linear map $\Phi : X \to Y$ such that $\Phi L^X_{ij}(z) = L^Y_{ij}(z)\Phi$ for $2 \leq i, j \leq N$. Category $\mathcal{O}'_{\mathrm{fd}}$ is an abelian subcategory of $\mathcal{V}_{\mathrm{ft}}$.

The $\mathfrak{h}$-algebra morphism $\mathcal{E}^{\mathfrak{h}}_{N-1} \to \mathcal{E}$ induces restriction functor $\mathcal{O}_{\mathrm{fd}} \to \mathcal{O}'_{\mathrm{fd}}$.

Let X be in category $\mathcal{O}'_{\mathrm{fd}}$. Equation (1.8) defines difference operators $K^X_l(z) : X \to X$ of bidegree (ϵ_l, ϵ_l) for $2 \leq l \leq N$. Condition (M3') implies that for each weight α, the weight space $X[\alpha]$ admits an ordered basis B_α with respect to which the matrix of $K^X_l(z)$ is upper triangular and has as diagonal entries $f_{b,l}(z) \in \mathbb{M}^\times_{\mathbb{C}}$ for $b \in B_\alpha$. Following Definition 1.9, we define the q-character of X to be

$$\chi'_q(X) = \sum_{\alpha \in \mathrm{wt}(X)} \sum_{b \in B_\alpha} (1, f_{b,2}(z), f_{b,3}(z), \ldots, f_{b,N}(z); \alpha) \in \mathcal{M}_{\mathrm{t}}.$$

It is independent of the choice of the bases B_α, as one can use category $\mathcal{F}_{\mathrm{mer}}$ to characterize the $f_{b,l}(z)$, see the comments after Lemma 1.6.

Remark 2.4. Let X be in category $\mathcal{O}_{\mathrm{fd}}$, viewed as an object of $\mathcal{O}'_{\mathrm{fd}}$. Then $\chi'_q(X)$ is obtained from $\chi_q(X)$ by replacing each e-weigth $\boldsymbol{g}$ of the $\mathcal{E}$-module X with $\boldsymbol{g}'$; here for $\boldsymbol{g} = (g_1(z), g_2(z), \ldots, g_N(z); \alpha) \in \mathcal{M}_{\mathrm{t}}$ we define

$$\boldsymbol{g}' := (1, g_2(z), g_3(z), \ldots, g_N(z); \alpha) \in \mathcal{M}_{\mathrm{t}}.$$

Reciprocally, if X is an irreducible $\mathcal{E}$-module in $\mathcal{O}_{\mathrm{fd}}$ of highest weight $(e_1(z), e_2(z), \ldots, e_N(z); \alpha) \in \mathcal{R}_{\mathrm{fd}}$, then $\chi_q(X)$ can be recovered from $\chi'_q(X)$. Indeed, since the N-th elliptic quantum minor is central, by

Schur's lemma, it acts on X as a scalar. Each e-weigth $(f_1(z), f_2(z), \dots, f_N(z); \beta)$ of the $\mathcal{E}$-module X is determined by the its last N components in $\chi_q'(X)$ as follows:

$$e_1(z+(N-1)\hbar)e_2(z+(N-2)\hbar)\cdots e_N(z) = f_1(z+(N-1)\hbar)f_2(z+(N-2)\hbar)\cdots f_N(z).$$

The highest weight theory in Section 1E carries over to category $\mathcal{O}'_{\mathrm{fd}}$ since $\hat{L}_k(z) \in \mathcal{E}^{\mathfrak{h}}_{N-1}$ for $2 \le k \le N$. Irreducible objects in $\mathcal{O}'_{\mathrm{fd}}$ are classified by their highest weight, and the q-character map is an injective morphism from the Grothendieck group $K_0(\mathcal{O}'_{\mathrm{fd}})$ to the additive group $\mathcal{M}_{\mathrm{t}}$. Let $\mathcal{P}'$ be the set of partitions with at most $N-1$ parts $(\nu_2 \ge \nu_3 \ge \cdots \ge \nu_N)$. For such a partition and for $c, a \in \mathbb{C}$,

$$\left(1, \frac{\theta(z+(a+\nu_2)\hbar)}{\theta(z+a\hbar)}, \frac{\theta(z+(a+\nu_3)\hbar)}{\theta(z+a\hbar)}, \dots, \frac{\theta(z+(a+\nu_N)\hbar)}{\theta(z+a\hbar)}; c\epsilon_1 + \sum_{j=2}^{N} \nu_j \epsilon_j\right)$$

is the highest weight of an irreducible $\mathcal{E}^{\mathfrak{h}}_{N-1}$-module in category $\mathcal{O}'_{\mathrm{fd}}$, which is denoted by $S'_{\nu,c,a}$. As in Section 1F, ν is identified with its Young diagram Y_ν. Let $\mathcal{B}'_\nu$ be the set of Young tableaux $Y_\nu \to \{2 < 3 < \cdots < N\}$ of shape ν.

Lemma 2.5. *Assume that Theorem 1.15 is true for $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_{N-1})$-modules. Then for $\nu \in \mathcal{P}'$ and $c, a \in \mathbb{C}$, the q-character of the $\mathcal{E}^{\mathfrak{h}}_{N-1}$-module $S'_{\nu,c,a}$ is*

$$\chi_q'(S'_{\nu,c,a}) = e^{c\epsilon_1} \sum_{T \in \mathcal{B}'_\nu} \prod_{(i,j)\in Y_\nu} \boxed{T(i,j)}'_{a+j-i} \in \mathcal{M}_{\mathrm{t}}.$$

Proof. We shall need $\mathcal{E}_{N-1}$-modules which are $\mathfrak{h}'$-graded $\mathbb{M}_{\mathfrak{h}'}$-vector spaces; similar category of finite-dimensional modules and q-characters are defined, based on the $\mathfrak{h}'$-algebra isomorphism $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_{N-1}) \cong \mathcal{E}_{N-1}$ in Section 2A.

For $\nu := (\nu_2 \ge \nu_3 \ge \cdots \ge \nu_N) \in \mathcal{P}'$ and $a \in \mathbb{C}$ there exists a unique (up to isomorphism) irreducible $\mathcal{E}_{N-1}$-module, denoted by $S'_{\nu,a}$, which contains a nonzero vector ω of $\mathfrak{h}'$-weight $\nu_2\epsilon_2 + \nu_3\epsilon_3 + \cdots + \nu_N\epsilon_N$ such that

$$L_{ij}(z)\omega = 0 \quad \text{and} \quad \hat{L}_k(z)\omega = \frac{\theta(z+(a+\nu_k)\hbar)}{\theta(z+a\hbar)}\omega$$

for $2 \le i, j, k \le N$ with $j < i$. We endow the $\mathbb{M}$-vector space $X := \mathbb{M} \otimes_{\mathbb{M}_{\mathfrak{h}'}} S'_{\nu,a}$ with an $\mathcal{E}^{\mathfrak{h}}_{N-1}$-module structure in category $\mathcal{O}'_{\mathrm{fd}}$.

Let w be a nonzero weight vector in $S'_{\nu,a}$. Its $\mathfrak{h}'$-weight is written uniquely in the form

$$(\nu_2\epsilon_2 + \nu_3\epsilon_3 + \cdots + \nu_N\epsilon_N) + (x_2\alpha_2 + x_3\alpha_3 + \cdots + x_{N-1}\alpha_{N-1}) \in \mathfrak{h}',$$

where $x_j \in \mathbb{Z}_{\le 0}$. Define the $\mathfrak{h}$-weight of $g(\lambda) \otimes_{\mathbb{M}_{\mathfrak{h}'}} w$, for $g(\lambda) \in \mathbb{M}^\times$, to be

$$(c\epsilon_1 + \nu_2\epsilon_2 + \nu_3\epsilon_3 + \cdots + \nu_N\epsilon_N) + (x_2\alpha_2 + x_3\alpha_3 + \cdots + x_{N-1}\alpha_{N-1}) \in \mathfrak{h},$$

and define the action of $L_{ij}(z)$ for $2 \le i, j \le N$ by the formula

$$L_{ij}(z)(g(\lambda) \otimes_{\mathbb{M}_{\mathfrak{h}'}} w) = g(\lambda + \hbar\epsilon_j) \otimes_{\mathbb{M}_{\mathfrak{h}'}} L_{ij}(z)w.$$

(M1')–(M2') are clear from the $\mathcal{E}_{N-1}$-module structure on $S'_{\nu,a}$. Choose an ordered weight basis B of $S'_{\nu,a}$ over $\mathbb{M}_{\mathfrak{h}'}$ such that the matrices of $\mathcal{D}_k(z)$ for $1 \leq k < N$ are upper triangular and their diagonal entries belong to $\mathbb{M}_{\mathbb{C}}^{\times}$. Then the ordered basis $\{1 \otimes_{\mathbb{M}_{\mathfrak{h}'}} b \mid b \in B\} =: B'$ of X satisfies (M3'). So X is in category $\mathcal{O}'_{\mathrm{fd}}$.

The matrices of $\mathcal{D}_k(z)$ with respect to the basis B' of X and the basis B of $S'_{\nu,a}$ are the same. So $\chi'_q(X)$ up to a normalization factor $e^{c\epsilon_1}$, is equal to the q-character of the $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_{N-1})$-module $S_{\nu,a}$. The latter is given by (1.13).

X has a unique (up to scalar) singular vector and is of highest weight, so it is irreducible. A comparison of highest weights shows that $X \cong S'_{\nu,c,a}$. $\square$

Fix $\mu \in \mathcal{P}$ a partition with at most N parts. Given a tableau $T \in \mathcal{B}_\mu$, by deleting the boxes $\boxed{1}$ in T, we obtain a Young diagram $T^{-1}(\{2, 3, \dots, N\})$ with at most $N-1$ rows, which corresponds to a partition in $\mathcal{P}'$, denoted by ν_T. Let $\mathcal{W}_\mu$ be the set of all such ν_T with $T \in \mathcal{B}_\mu$. For $\nu \in \mathcal{W}_\mu$, define c_ν to be the cardinal of the finite subset $Y_\mu \setminus Y_\nu$ of $\mathbb{Z}^2$.

Again take the example $N = 3$ and $\mu = (2 \geq 1)$ after Theorem 1.15. The eight tableaux in $\mathcal{B}_\mu$ with $\boxed{1}$ deleted give four Young diagrams and partitions

$$\ydiagram{1} = (1), \quad \ydiagram{2} = (2), \quad \ydiagram{1,1} = (1 \geq 1), \quad \ydiagram{1+1,2} = (2 \geq 1).$$

The corresponding integers c_ν are 2, 1, 1, 0.

Lemma 2.6. *Let $\mu \in \mathcal{P}$ and $a \in \mathbb{C}$. In the Grothendieck group $K_0(\mathcal{O}'_{\mathrm{fd}})$*

$$[S_{\mu,a}] = \sum_{\nu \in \mathcal{W}_\mu} [S'_{\nu,c_\nu,a}].$$

Proof. Let $\mathfrak{e}'$ be the subalgebra of $\mathfrak{e}$ generated by $\mathbb{M}_2$ and the t_{ij} for $2 \leq i, j \leq N$. One can define similar abelian category $\mathcal{F}'$ of $\mathfrak{e}'$-modules (which are $\mathfrak{h}$-graded $\mathbb{M}$-vector spaces) equipped with:

(a) The evaluation functor $\mathrm{ev}'_a : \mathcal{F}' \to \mathcal{O}'_{\mathrm{fd}}$ from $\mathfrak{e}'$-modules to $\mathcal{E}^{\mathfrak{h}}_{N-1}$-modules.

(b) The injective character map $\chi : K_0(\mathcal{F}') \to \mathcal{M}_{\mathfrak{t}}$ from the $\mathfrak{h}$-grading.

Theorem 2.3 applied to the $\mathfrak{h}'$-algebra $\mathfrak{e}_{\tau,\hbar}(\mathfrak{sl}_{N-1})$, from the scalar extension in the proof of Lemma 2.5, one obtains an irreducible object $V'_{\nu,c}$ in category $\mathcal{F}'$ for $\nu = (\nu_2 \geq \nu_3 \geq \cdots \geq \nu_N) \in \mathcal{P}'$ and $c \in \mathbb{C}$ with the following properties:

(c) $V'_{\nu,c}$ admits a nonzero vector v of weight $c\epsilon_1 + \nu_2\epsilon_2 + \nu_3\epsilon_3 + \cdots + \nu_N\epsilon_N$ and $\hat{t}_k v = v$, $t_{ij} v = 0$ for $2 \leq i, j, k \leq N$ and $i < j$.

(d) $\chi(V'_{\nu,c})$ is equal to the character of the irreducible $\mathfrak{sl}'_{N-1}$-module of highest weight $c\epsilon_1 + \nu_2\epsilon_2 + \nu_3\epsilon_3 + \cdots + \nu_N\epsilon_N$; here $\mathfrak{sl}'_{N-1}$ is the parabolic Lie subalgebra of $\mathfrak{sl}_N$ (with the same Cartan algebra $\mathfrak{h}$) associated to the simple roots $\alpha_2, \alpha_3, \dots, \alpha_{N-1}$.

By comparing highest weight we observe that $\mathrm{ev}'_a(V'_{\nu,c}) \cong S'_{\nu,c,a}$ in category $\mathcal{O}'_{\mathrm{fd}}$.

Let $\mu = (\mu_1 \geq \mu_2 \geq \cdots \geq \mu_N) \in \mathcal{P}$. Set $\bar{\mu} := \mu_1\epsilon_1 + \mu_2\epsilon_2 + \cdots + \mu_N\epsilon_N$. Then $S_{\mu,a} \cong \mathrm{ev}_a(V_{\bar{\mu}})$ in category $\mathcal{O}_{\mathrm{fd}}$. By diagram chasing

$$\begin{tikzcd}
K_0(\mathcal{O}_{\mathrm{fd}}) \arrow[d, "\mathrm{Res}"'] & K_0(\mathcal{F}) \arrow[l, "\mathrm{ev}_a"'] \arrow[r, "\chi"] \arrow[d, "\mathrm{Res}"'] & \mathcal{M}_{\mathrm{t}} \\
K_0(\mathcal{O}'_{\mathrm{fd}}) & K_0(\mathcal{F}') \arrow[l, "\mathrm{ev}'_a"'] \arrow[ur, "\chi"'] &
\end{tikzcd}$$

Lemma 2.6 is equivalent to the character identity $\chi(V_{\bar{\mu}}) = \sum_{\nu \in \mathcal{W}_\mu} \chi(V'_{\nu,c_\nu})$. Since the left- and right-hand sides are the character of a representations of $\mathfrak{sl}_N$ by Theorem 2.3 and $\mathfrak{sl}'_{N-1}$ by (d), respectively, this identity is a consequence of the branching rule for representations of the reductive Lie algebras $\mathfrak{sl}_N \supset \mathfrak{sl}'_{N-1}$. □

2D. ***Proof of Theorem 1.15.*** We proceed by induction on N. For $N = 1$ and $\mu = (n)$, since $S_{\mu,a}$ is one-dimensional, its q-character is equal to its highest weight

$$\left(\frac{\theta(z+(a+n)\hbar)}{\theta(z+a\hbar)}; n\epsilon_1\right) = \prod_{j=1}^{n} \boxed{1}_{a+j-1}.$$

Suppose $N > 1$. By Lemma 2.5, the induction hypothesis in the case of $N-1$ gives the q-character formula for all the $\mathcal{E}^{\mathfrak{h}}_{N-1}$-modules $S'_{\nu,c,a}$ where $\nu \in \mathcal{P}'$ and $c \in \mathbb{C}$. So the q-character $\chi'_q(S_{\mu,a})$ of the $\mathcal{E}^{\mathfrak{h}}_{N-1}$-module $S_{\mu,a}$ is known by Lemma 2.6.

Since $S_{\mu,a}$ is an irreducible $\mathcal{E}$-module in category $\mathcal{O}_{\mathrm{fd}}$, by Remark 2.4, $\chi_q(S_{\mu,a})$ can be recovered from $\chi'_q(S_{\mu,a})$. Since $\mathcal{B}_\mu$ is the disjoint union of the $\mathcal{B}'_\nu$ for $\nu \in \mathcal{W}_\mu$, it suffices to check that for each e-weigth $(m_1^T(z), m_2^T(z), \ldots, m_N^T(z); \alpha)$ at the right-hand side of (1.13), where $T \in \mathcal{B}_\mu$, the following product

$$m^T(z) := m_1^T(z+(N-1)\hbar)m_2^T(z+(N-2)\hbar)\cdots m_N^T(z)$$

is the eigenvalue of scalar action of $\mathcal{D}_N(z)$ on $S_{\mu,a}$. Notice first that

$$\prod_{p=1}^{N} \frac{\theta(z+(a+N-p+\mu_p)\hbar)}{\theta(z+(a+N-p)\hbar)} = \prod_{(i,j)\in Y_\mu} \frac{\theta(z+(a+j-i+N)\hbar)}{\theta(z+(a+j-i+N-1)\hbar)}.$$

By (1.12), each box $\boxed{i}_x$ contributes to $\theta(z+(x+N)\hbar)/\theta(z+(x+N-1)\hbar)$, so the right-hand side of the identity is exactly $m^T(z)$. By Remark 2.4, the left-hand side is the scalar of $\mathcal{D}_N(z)$ acting on $S_{\mu,a}$. This completes the proof of Theorem 1.15. □

3. Kirillov–Reshetikhin modules

We study certain irreducible $\mathcal{E}$-modules via q-characters.

Fix $a \in \mathbb{C}$. For $k \in \mathbb{C}$ and $1 \leq r \leq N$, define the *asymptotic e-weigth*

$$\boldsymbol{w}_{k,a}^{(r)} := \Psi_{r,a+k}\Psi_{r,a}^{-1} \in \mathcal{M}_{\mathrm{w}}. \tag{3.14}$$

Assume $k \in \mathbb{Z}_{\geq 0}$. We identify $k\varpi_r$ with the partition $(k \geq k \geq \cdots \geq k)$ where k appears r times. Then $\boldsymbol{w}_{k,a}^{(r)} = Y_{r,a+\frac{1}{2}} Y_{r,a+\frac{3}{2}} \cdots Y_{r,a+k-\frac{1}{2}} = \theta_{k\varpi_r,a}$ by (1.9)–(1.10), and the finite-dimensional irreducible $\mathcal{E}$-module $S(\boldsymbol{w}_{k,a}^{(r)})$ in category $\mathcal{O}_{\mathrm{fd}}$ is denoted by $W_{k,a}^{(r)}$ and called *Kirillov–Reshetikhin module* (KR module).

The $Y_{i,a+m}$ and $A_{i,a+m}$ for $1 \leq i \leq N$ and $m \in \frac{1}{2}\mathbb{Z}$ are linearly independent in the abelian group $\mathcal{M}_{\mathrm{w}}$, and generate the subgroups $\mathcal{P}_a$ and $\mathcal{Q}_a$ and the submonoids $\mathcal{P}_a^+$ and $\mathcal{Q}_a^+$, respectively. The inverses of these submonoids are denoted by $\mathcal{P}_a^-$ and $\mathcal{Q}_a^-$ respectively. By (1.13) and (1.10),

$$\mathrm{wt}_e(S_{\mu,a}) \subset \theta_{\mu,a}\mathcal{Q}_a^- \subset \mathcal{P}_a \quad \text{for } \mu \in \mathscr{P}.$$

Indeed, let $T_\mu \in \mathscr{B}_\mu$ be such that the associated monomial in (1.13) is $\theta_{\mu,a}$. Then for $S \in \mathscr{B}_\mu$, we must have $S(i,j) \geq T_\mu(i,j)$ for all $(i,j) \in Y_\mu$.

Following [Frenkel and Mukhin 2001, § 6], we call $\boldsymbol{f} \in \mathcal{P}_a$ *right-negative* if the factors $Y_{i,a+m}$ with $1 \leq i < N$ appearing in $\boldsymbol{f}$, for which $m \in \frac{1}{2}\mathbb{Z}$ is minimal, have negative powers.

Lemma 3.1 [Frenkel and Mukhin 2001]. *Let $\boldsymbol{e}, \boldsymbol{f} \in \mathcal{P}_a$. If $\boldsymbol{e}$ and $\boldsymbol{f}$ are right-negative, then so is $\boldsymbol{ef}$.*

All elements in $\mathcal{Q}_a^-$ different from 1 are right-negative by (1.9).

Lemma 3.2. *Let $k \in \mathbb{Z}_{>0}$ and $1 \leq r < N$.*

(1) *For $1 \leq l \leq k$, $\boldsymbol{w}_{k,a}^{(r)} A_{r,a}^{-1} A_{r,a+1}^{-1} \cdots A_{r,a+l-1}^{-1}$ is an e-weigth of $W_{k,a}^{(r)}$ of multiplicity one in $\chi_q(W_{k,a}^{(r)})$.*

(2) *An e-weigth of $W_{k,a}^{(r)}$ different from those in* (1) *and from $\boldsymbol{w}_{k,a}^{(r)}$ must belong to $\boldsymbol{w}_{k,a}^{(r)} A_{r,a}^{-1} A_{s,a-\frac{1}{2}}^{-1} \mathcal{Q}_a^-$ for certain $1 \leq s < N$ with $s = r \pm 1$.*

(3) *Any e-weigth of $W_{k,a}^{(r)}$ is either $\boldsymbol{w}_{k,a}^{(r)}$ or right-negative.*

Proof. The Young diagram $Y_{k\varpi_r}$ is a rectangle of r rows and k columns. For (1)–(2) the proof of [Zhang 2018, Lemma 3.4] works by applying Theorem 1.15 to $W_{k,a}^{(r)} \cong S_{k\varpi_r, a-\ell_r}$. For (3), $\boldsymbol{w}_{k,a}^{(r)} A_{r,a}^{-1}$ is right-negative, and so is any element of $\boldsymbol{w}_{k,a}^{(r)} A_{r,a}^{-1} \mathcal{Q}_a^-$. □

For $1 \leq r < N$ and $k, t, a \in \mathbb{C}$, define as in [Frenkel and Hernandez 2016, §4.3] and [Zhang 2018, Remark 3.2]

$$\boldsymbol{d}_{k,a}^{(r,t)} := \frac{\Psi_{r,a+t}}{\Psi_{r,a}} \prod_{s=r\pm 1} \frac{\Psi_{s,a-\frac{1}{2}}}{\Psi_{s,a-\frac{1}{2}-k}} = \boldsymbol{w}_{t,a}^{(r)} \prod_{s=r\pm 1} \boldsymbol{w}_{k,a-\frac{1}{2}-k}^{(s)} \in \mathcal{M}_{\mathrm{w}}. \tag{3.15}$$

If $k, t \in \mathbb{Z}_{\geq 0}$, then $\boldsymbol{d}_{k,a}^{(r,t)} \in \mathcal{R}_{\mathrm{fd}}$ and set $D_{k,a}^{(r,t)} := S(\boldsymbol{d}_{k,a}^{(r,t)})$.

Lemma 3.3. *Let $1 \leq r < N$ and $m, k \in \mathbb{Z}_{>0}$.*

(1) *The dominant e-weigths of $W_{k+m-1,1}^{(r)} \overline{\otimes} W_{k,0}^{(r)}$ and $W_{k-1,1}^{(r)} \overline{\otimes} W_{k+m,0}^{(r)}$ are*

$$\boldsymbol{w}_{k+m-1,1}^{(r)} \boldsymbol{w}_{k,0}^{(r)} \quad \textit{and} \quad \boldsymbol{w}_{k+m-1,1}^{(r)} \boldsymbol{w}_{k,0}^{(r)} A_{r,1}^{-1} A_{r,2}^{-1} \cdots A_{r,l}^{-1} \quad \textit{for } 1 \leq l \leq k,$$
$$\boldsymbol{w}_{k-1,1}^{(r)} \boldsymbol{w}_{k+m,0}^{(r)} \quad \textit{and} \quad \boldsymbol{w}_{k-1,1}^{(r)} \boldsymbol{w}_{k+m,0}^{(r)} A_{r,1}^{-1} A_{r,2}^{-1} \cdots A_{r,l}^{-1} \quad \textit{for } 1 \leq l < k,$$

respectively. All such e-weigths are of multiplicity one.

(2) *The module $W_{k-1,1}^{(r)} \overline{\otimes} W_{k+m,0}^{(r)}$ is irreducible.*

Proof. For (1), one can copy the last two paragraphs of the proof of [Fourier and Hernandez 2014, Theorem 4.1], since the right-negativity property of KR modules in the elliptic case (Lemma 3.2) is the same as in the affine case. Let T be the tensor product module of (2). Suppose T is not irreducible. Then there exists $1 \le l \le k-1$ such that T admits an irreducible subquotient $S \cong S(\boldsymbol{d}_l)$ where by (1.9)

$$\boldsymbol{d}_l := \boldsymbol{w}_{k-1,1}^{(r)} \boldsymbol{w}_{k+m,0}^{(r)} \prod_{j=1}^{l} A_{r,j}^{-1} = \frac{\Psi_{r,k}\Psi_{r,k+m}}{\Psi_{r,l+1}\Psi_{r,l}} \prod_{s=r\pm1} \frac{\Psi_{s,l+\frac{1}{2}}}{\Psi_{s,\frac{1}{2}}}.$$

Set $\mu := \varpi(\boldsymbol{d}_l)$. The weight space $S[\mu-\alpha_r]$ is nonzero since the Ψ_r do not cancel in $\boldsymbol{d}_l$, and its possible e-weigths are $\boldsymbol{d}_l A_{r,l}^{-1}$, $\boldsymbol{d}_l A_{r,l+1}^{-1}$ since S is a subquotient of $W_{k-l-1,l+1}^{(r)} \overline{\otimes} W_{k-l+m,l}^{(r)} \overline{\otimes} (\overline{\otimes}_{s=r\pm1} W_{l,\frac{1}{2}}^{(s)})$. If $\boldsymbol{d}_l A_{r,l}^{-1}$ is an e-weigth of S, then

$$\boldsymbol{w}_{k-1,1}^{(r)} \boldsymbol{w}_{k+m,0}^{(r)} A_{r,1}^{-1} A_{r,2}^{-1} \cdots A_{r,l-1}^{-1} A_{r,l}^{-2} \in \mathrm{wt}_e(T) = \mathrm{wt}_e(W_{k-1,1}^{(r)})\, \mathrm{wt}_e(W_{k+m,0}^{(r)})$$

which contradicts with the q-characters of KR modules in Lemma 3.2. So $k > l+1$ and $S[\mu-\alpha_r] = \mathbb{M}v \neq 0$. Let ω be a highest weight vector of S. Then

$$p := \mu_{r,r+1} = 2k-2l+m-1, \quad L_{r,r+1}(z)\omega = A(z;\lambda)v, \quad L_{r+1,r}(z)v = B(z;\lambda)\omega$$

for some meromorphic functions A and B of $(z,\lambda) \in \mathbb{C} \times \mathfrak{h}$. For $1 \le i \le N$, let $g_i(z) \in \mathbb{M}_{\mathbb{C}}^{\times}$ be the i-th component of $\boldsymbol{d}_l \in \mathcal{M}_{\mathrm{w}}$. Then $L_{ii}(z)\omega = g_i(z)\varphi_i(\lambda)\omega$ for certain $\varphi_i(\lambda) \in \mathbb{M}^{\times}$ by (1.7). Set $h(z) := g_r(z)/g_{r+1}(z)$. We have

$$h(z) = \frac{\theta(z+(k-\ell_r)\hbar)\theta(z+(k+m-\ell_r)\hbar)}{\theta(z+(l+1-\ell_r)\hbar)\theta(z+(l-\ell_r)\hbar)} = \frac{\theta(z-w_1)\theta(z-w_2)}{\theta(z-w_3)\theta(z-w_4)},$$

where $w_1 := (\ell_r-k)\hbar$, $w_2 := (\ell_r-k-m)\hbar$ and so on. Applying (1.2) with $(i,j) = (r+1,r) = (n,m)$ to ω, as in the proof of [Felder and Zhang 2017, Theorem 4.1], we obtain

$$\left(\frac{\theta(z-w+\lambda_{r,r+1}+p\hbar)\theta(\hbar)}{\theta(z-w+\hbar)\theta(\lambda_{r,r+1}+p\hbar)} g_{r+1}(z)g_r(w) - \frac{\theta(z-w+\lambda_{r,r+1})\theta(\hbar)}{\theta(z-w+\hbar)\theta(\lambda_{r,r+1})} g_{r+1}(w)g_r(z)\right)\varphi_r(\lambda+\hbar\epsilon_{r+1})\varphi_{r+1}(\lambda)$$
$$= \frac{\theta(z-w)\theta(\lambda_{r,r+1}+\hbar)}{\theta(z-w+\hbar)\theta(\lambda_{r,r+1})} B(w;\lambda)A(z;\lambda+\hbar\epsilon_r).$$

Multiplying both sides by $\theta(z-w+\hbar)/(g_{r+1}(z)g_{r+1}(w))$ and noticing $g_{r+1}(z) = \theta(z-w_3)/\theta(z-w_3-l\hbar)$, one can evaluate w at w_1 and w_2 to obtain identities of meromorphic functions of (z,λ):

$$\tilde{A}(z;\lambda)x_i(\lambda) = \frac{\theta(z-w_i+\lambda_{r,r+1})}{\theta(z-w_i)} f(\lambda)h(z) \quad \text{for } i = 1,2.$$

Here we set $\varphi(\lambda) := \varphi_r(\lambda+\hbar\epsilon_{r+1})\varphi_{r+1}(\lambda)$ and

$$\tilde{A}(z;\lambda) := \frac{A(z;\lambda+\hbar\epsilon_r)}{g_{r+1}(z)}, \quad x_i(\lambda) := \frac{B(w_i;\lambda)}{g_{r+1}(w_i)}, \quad f(\lambda) := -\frac{\theta(\hbar)\varphi(\lambda)}{\theta(\lambda_{r,r+1}+\hbar)}.$$

Since $f(\lambda)h(z) \neq 0$, we have $x_i(\lambda) \neq 0$ and so

$$\frac{\theta(z-w_1+\lambda_{r,r+1})\theta(z-w_2)}{x_1(\lambda)\theta(z-w_3)\theta(z-w_4)} = \frac{\theta(z-w_2+\lambda_{r,r+1})\theta(z-w_1)}{x_2(\lambda)\theta(z-w_3)\theta(z-w_4)}$$

as nonzero meromorphic functions of (z, λ). This forces $w_1 - w_2 = m\hbar \in \mathbb{Z} + \mathbb{Z}\tau$, which certainly does not hold. This proves (3). $\square$

Theorem 3.4. *For $1 \le r < N$, $t \in \mathbb{Z}_{\ge 0}$ and $k > 0$, we have the following identities in the Grothendieck ring of category $\mathcal{O}_{\mathrm{fd}}$:*

$$[D^{(r,t)}_{k,k+1}] + [W^{(r)}_{k-1,1}][W^{(r)}_{k+t+1,0}] = [W^{(r)}_{k+t,1}][W^{(r)}_{k,0}], \tag{3.16}$$

$$[D^{(r,t+1)}_{k,k}][W^{(r)}_{k+t,0}] = [D^{(r,0)}_{k+t+1,k+t+1}][W^{(r)}_{k-1,0}] + [D^{(r,t)}_{k,k}][W^{(r)}_{k+t+1,0}]. \tag{3.17}$$

Proof. Set $T := W^{(r)}_{k+t,1} \overline{\otimes} W^{(r)}_{k,0}$ and $\boldsymbol{d} := \boldsymbol{w}^{(r)}_{k+t,1}\boldsymbol{w}^{(r)}_{k,0}$. Then $S := S(\boldsymbol{d})$ is an irreducible subquotient of T and by (3.14)–(3.15)

$$\boldsymbol{d} = \boldsymbol{w}^{(r)}_{k-1,1}\boldsymbol{w}^{(r)}_{k+t+1,0} \quad \text{and} \quad \boldsymbol{d}^{(r,t)}_{k,k+1} = A^{-1}_{r,1}A^{-1}_{r,2}\cdots A^{-1}_{r,k}\boldsymbol{d}.$$

Set $m = t+1$ in Lemma 3.3. Then $S \cong W^{(r)}_{k-1,1} \overline{\otimes} W^{(r)}_{k+t+1,0}$, and there is exactly one dominant e-weigth (counted with multiplicity) in $\mathrm{wt}_e(T) \setminus \mathrm{wt}_e(S)$, namely $\boldsymbol{d}^{(r,t)}_{k,k+1}$. This proves (3.16), which implies after taking spectral parameter shifts

$$[D^{(r,t+1)}_{k,k}] = [W^{(r)}_{k+t+1,0}][W^{(r)}_{k,-1}] - [W^{(r)}_{k-1,0}][W^{(r)}_{k+t+2,-1}],$$

$$[D^{(r,0)}_{k+t+1,k+t+1}] = [W^{(r)}_{k+t+1,0}][W^{(r)}_{k+t+1,-1}] - [W^{(r)}_{k+t,0}][W^{(r)}_{k+t+2,-1}],$$

$$[D^{(r,t)}_{k,k}] = [W^{(r)}_{k+t,0}][W^{(r)}_{k,-1}] - [W^{(r)}_{k-1,0}][W^{(r)}_{k+t+1,-1}].$$

Equation (3.17) becomes a trivial identity involving only KR modules. $\square$

$D^{(r,t)}_{k,k+1}$ is *special* in the sense of [Nakajima 2003] as it contains only one dominant e-weigth. For $t = 0$, $D^{(r,0)}_{k,k+1} \cong W^{(r-1)}_{k,\frac{1}{2}} \overline{\otimes} W^{(r+1)}_{k,\frac{1}{2}}$ by showing that the tensor product is special as in [Nakajima 2003], and (3.16) is the T-system of KR modules.

Corollary 3.5. *Let $1 \le r < N$, $a \in \mathbb{C}$ and $k, t \in \mathbb{Z}_{>0}$.*

(1) $\boldsymbol{d}^{(r,t)}_{k,a}A^{-1}_{r,a}A^{-1}_{r,a+1}\cdots A^{-1}_{r,a+l-1} \in \mathrm{wt}_e(D^{(r,t)}_{k,a})$ *for $1 \le l \le t$.*

(2) *Any e-weigth of $D^{(r,t)}_{k,a}$ different from those in* (1) *and $\boldsymbol{d}^{(r,d)}_{k,a}$ belongs to $\boldsymbol{d}^{(r,t)}_{k,a}\{A^{-1}_{r,a-k-1}, A^{-1}_{s,a-k-\frac{1}{2}}\}\mathcal{Q}^-_a$, for certain $1 \le s < N$ with $s = r \pm 1$.*

Proof. This comes from Lemma 3.2 and Theorem 3.4. $\square$

Lemma 3.6. *Let $1 \le r < N$ and $t \in \mathbb{Z}_{\ge 0}$. There is a short exact sequence*

$$0 \to D^{(r,t)}_{1,a} \to W^{(r)}_{t+1,a-1} \overline{\otimes} W^{(r)}_{1,a-2} \to W^{(r)}_{t+2,a-2} \to 0$$

of $\mathcal{E}$-modules in category $\mathcal{O}_{\mathrm{fd}}$.

Proof. Let T and S be the second and third terms above (zero excluded). Let ω_1 and ω_2 be highest weight vectors of $W^{(r)}_{t+1,a-1}$ and $W^{(r)}_{1,a-2}$ respectively. Then $\omega_1 \overline{\otimes} \omega_2$ is a highest weight vector of T and generates a submodule T'. Suppose $T' = T$. Then T is a highest weight module whose highest weight is equal to that of the irreducible module S. There is a surjective morphism of modules $T \to S$, the kernel of which is $D_{1,a}^{(r,t)}$ by (3.16) (one applies a spectral parameter shift Φ_{a-2} to the equation with $k=1$). This is the desired short exact sequence.

Suppose $T \neq T'$. Then $[T'] = [S]$ or $[T'] = [D_{k,a}^{(r,t)}]$. By comparing highest weights, we have $[T'] = [S]$. So the weight space $T'[(t+2)\varpi_r - \alpha_r]$ is one-dimensional. Corollary 2.2 applied to $W_{t+1,a-1}^{(r)} \cong S_{(t+1)\varpi_r, a-\ell_r-1}$, one finds $g(\lambda) \in \mathbb{M}^\times$ such that $L_{r+1,r+1}(z)\omega_1 = \omega_1$ and (set $b := a - \ell_r - 1$)

$$L_{r,r+1}(z)\omega_1 = \frac{\theta(z+(b+t)\hbar+\lambda_{r,r+1})}{\theta(z+b\hbar)}\omega_1' \quad \text{and} \quad L_{rr}(z)\omega_1 = \frac{\theta(z+(b+t+1)\hbar)}{\theta(z+b\hbar)}g(\lambda)\omega_1,$$

where $0 \neq \omega_1'$ is of weight $(t+1)\varpi_r - \alpha_r$. Similarly $L_{r+1,r+1}(z)\omega_2 = \omega_2$ and

$$L_{r,r+1}(z)\omega_2 = \frac{\theta(z+(b-1)\hbar+\lambda_{r,r+1})}{\theta(z+(b-1)\hbar)}\omega_2'$$

with $\omega_2' \neq 0$ of weight $\varpi_r - \alpha_r$. Since ω_1, ω_2 are highest weight vectors, we have

$$\begin{aligned}
&L_{r,r+1}(z)(\omega_1 \overline{\otimes} \omega_2) \\
&= L_{r,r+1}(z)\omega_1 \overline{\otimes} L_{r+1,r+1}(z)\omega_2 + L_{rr}(z)\omega_1 \overline{\otimes} L_{r,r+1}(z)\omega_2 \\
&= \left(\frac{\theta(z+(b+t)\hbar+\lambda_{r,r+1})}{\theta(z+b\hbar)}\omega_1'\right)\overline{\otimes}\omega_2 + \frac{\theta(z+(b+t+1)\hbar)}{\theta(z+b\hbar)}g_1(\lambda)\omega_1 \overline{\otimes}\left(\frac{\theta(z+(b-1)\hbar+\lambda_{r,r+1})}{\theta(z+(b-1)\hbar)}\omega_2'\right).
\end{aligned}$$

Setting $z = -(b+t+1)\hbar$ we obtain $\omega_1' \overline{\otimes} \omega_2 \in T'$, and so $\omega_1 \otimes \omega_2' \in T'$. The weight space $T'[(t+2)\varpi_r - \alpha_r]$ is at least two-dimensional, a contradiction. □

Lemma 3.6 is inspired by [Moura and Pereira 2017, § 5.3], to transform identities in the Grothendieck group into exact sequences by restriction to $\mathfrak{sl}_2$, see [Chari 2002]. More generally, we have the short exact sequences in category $\mathcal{O}_{\mathrm{fd}}$ by [Felder and Zhang 2017, Proposition 4.3 and Corollary 4.5]:[9]

$$0 \to D_{k,k+1}^{(r,t)} \to W_{k+t,1}^{(r)} \overline{\otimes} W_{k,0}^{(r)} \to W_{k-1,1}^{(r)} \overline{\otimes} W_{k+t+1,0}^{(r)} \to 0,$$

$$0 \to D_{k+t+1,k+t+1}^{(r,0)} \overline{\otimes} W_{k-1,0}^{(r)} \to D_{k,k}^{(r,t+1)} \overline{\otimes} W_{k+t,0}^{(r)} \to D_{k,k}^{(r,t)} \overline{\otimes} W_{k+t+1,0}^{(r)} \to 0.$$

These exact sequences hold for affine quantum (super)groups [Fourier and Hernandez 2014; Zhang 2018]. In the super case the proof is more delicate since Lemma 3.2(3) fails.

4. Asymptotic representations

We construct infinite-dimensional modules in category $\mathcal{O}$ as inductive limits ($k \to \infty$) of the KR modules $W_{k,a}^{(r)}$ for fixed $1 \leq r < N$ and $a := \ell_r$.

The general strategy follows that of Hernandez and Jimbo [2012]:

(i) Produce an inductive system of vector spaces $W_{0,a}^{(r)} \subseteq W_{1,a}^{(r)} \subseteq W_{2,a}^{(r)} \subseteq \cdots$.

(ii) Prove that the matrix entries of the $L_{ij}(z)$ are *good* functions of $k \in \mathbb{Z}_{\geq 0}$.

[9] The elliptic quantum group of [Felder and Zhang 2017] is slightly different as it is defined by another R-matrix, which is gauge equivalent to the present $\boldsymbol{R}$ by [Enriquez and Felder 1998].

(iii) Define the module structure on the inductive limit of (i).

Step (i) is done in Lemma 4.2, step (ii) in Lemma 4.8, and step (iii) in Proposition 4.10. We shall see that the proofs in each step are different from [Hernandez and Jimbo 2012].

In what follows, by $k > l$ we implicitly assume that $k, l \in \mathbb{Z}_{\geq 0}$. For $k > l$, set $Z_{kl} := W^{(r)}_{k-l,a+l} \cong S_{(k-l)\varpi_r,l}$ and fix a highest weight vector $\omega_{kl} \in Z_{kl}$. By (1.7), we have for $1 \leq i \leq r < j \leq N$,

$$L_{ii}(z)\omega_{kl} = \omega_{kl}\frac{\theta(z+k\hbar)}{\theta(z+l\hbar)}\prod_{q=r+1}^{N}\frac{\theta(\lambda_{iq}+(k-l+1)\hbar)}{\theta(\lambda_{iq}+\hbar)} \quad \text{for } L_{jj}(z)\omega_{kl} = \omega_{kl}.$$

Note that $Z_{k0} = W^{(r)}_{k,a}$, and we simply write $\omega_{k0} =: \omega_k$.

Lemma 4.1. *Let $t > k > l > m$. There exists a unique morphism of $\mathcal{E}$-modules*

$$\mathcal{G}^l_{k,m} : Z_{kl}\overline{\otimes}Z_{lm} \to Z_{km}$$

such that $\mathcal{G}^l_{k,m}(\omega_{kl}\overline{\otimes}\omega_{lm}) = \omega_{km}$. Moreover the following diagram commutes:

$$\begin{tikzcd} Z_{tk}\overline{\otimes}Z_{kl}\overline{\otimes}Z_{lm} \arrow[r, "\mathcal{G}^k_{t,l}\overline{\otimes}\mathrm{Id}"] \arrow[d, "\mathrm{Id}\overline{\otimes}\mathcal{G}^l_{k,m}"'] & Z_{tl}\overline{\otimes}Z_{lm} \arrow[d, "\mathcal{G}^l_{t,m}"] \\ Z_{tk}\overline{\otimes}Z_{km} \arrow[r, "\mathcal{G}^k_{t,m}"] & Z_{tm} \end{tikzcd} \tag{4.18}$$

Proof. (Uniqueness) Let F and G be two such morphisms and let X be the image of $F - G$. Then $\omega_{km} \notin X$. If $X \neq 0$, then X has a highest weight vector $v \neq 0$, which is proportional to ω_{km} by the irreducibility of Z_{km}, a contradiction. So $X = 0$ and $F = G$. The commutativity of (4.18) is proved in the same way.

(Existence) Let $b \in \mathbb{C}$ and $n \in \mathbb{Z}_{>0}$. By Lemma 3.6 there exists a surjective $\mathcal{E}$-linear map

$$W^{(r)}_{n-1,b+1}\overline{\otimes}W^{(r)}_{1,b} \to W^{(r)}_{n,b}.$$

An induction on n shows that the $\mathcal{E}$-module $W^{(r)}_{1,b+n-1}\overline{\otimes}W^{(r)}_{1,b+n-2}\overline{\otimes}\cdots\overline{\otimes}W^{(r)}_{1,b+1}\overline{\otimes}W^{(r)}_{1,b}$ can be projected onto $W^{(r)}_{n,b}$. Setting $(n, b) = (k - m, a + m)$ we obtain a surjective $\mathcal{E}$-linear map

$$g : Z_{k,k-1}\overline{\otimes}Z_{k-1,k-2}\overline{\otimes}\cdots\overline{\otimes}Z_{m+2,m+1}\overline{\otimes}Z_{m+1,m} =: T \twoheadrightarrow Z_{km}.$$

Taking (n, b) to be $(k-l, a+l)$ and $(l-m, a+m)$, we project the first $k-l$ and the last $l-m$ tensor factors of T onto Z_{kl} and Z_{lm} respectively. The tensor product of these projections gives $f : T \twoheadrightarrow Z_{kl}\overline{\otimes}Z_{lm}$. Since $\omega_{kl}\overline{\otimes}\omega_{lm}$, ω_{km} and $\omega := \omega_{k,k-1}\overline{\otimes}\omega_{k-1,k-2}\overline{\otimes}\cdots\overline{\otimes}\omega_{m+2,m+1}\overline{\otimes}\omega_{m+1,m} \in T$ are highest weight vectors of the same e-weigth, by surjectivity one can assume $f(\omega) = \omega_{kl}\overline{\otimes}\omega_{lm}$ and $g(\omega) = \omega_{km}$. It suffices to prove that g factorizes through f, and so $g = \mathcal{G}^l_{k,m} f$. Set $Y := \ker(f)$ and $Z := \ker(g)$. The image of g being irreducible, Z is a maximal submodule of T. Since $\omega \notin Y + Z$, we have $Y + Z = Z$ and $Y \subseteq Z$. $\square$

We need two special cases of the $\mathcal{G}$; for $k > l$ and $t-1 > l$,

$$\mathcal{F}_{k,l} = \mathcal{G}^{l}_{k,0} : Z_{kl}\overline{\otimes} W^{(r)}_{l,a} \to W^{(r)}_{k,a} \quad \text{and} \quad \mathcal{G}_{t,l} = \mathcal{G}^{l+1}_{t,l} : Z_{t,l+1}\overline{\otimes} Z_{l+1,l} \to Z_{tl}.$$

As in [Hernandez and Jimbo 2012, §4.2], for $k > l$ define the restriction map

$$F_{k,l} : W^{(r)}_{l,a} \to W^{(r)}_{k,a}, \quad v \mapsto \mathcal{F}_{k,l}(\omega_{kl}\overline{\otimes} v).$$

It is a difference map of bidegree $((l-k)\varpi_r, 0)$.

Applying (4.18) with $t > k > l > 0$ to $\omega_{tk}\overline{\otimes}\omega_{kl}\overline{\otimes} W^{(r)}_{l,a}$ gives $F_{t,k}F_{k,l} = F_{t,l}$. So $(W^{(r)}_{l,a}, F_{k,l})$ is an inductive system of vector spaces.[10]

Applying (4.18) with $k > l+1 > l > 0$ to $\omega_{k,l+1}\overline{\otimes} Z_{l+1,l}\overline{\otimes} W^{(r)}_{l,a}$, we obtain

$$\mathcal{F}_{k,l}(\mathcal{G}_{k,l}(\omega_{k,l+1}\overline{\otimes} v)\overline{\otimes} w) = F_{k,l}\mathcal{F}_{l+1,l}(v\overline{\otimes} w) \quad \text{for } v\overline{\otimes} w \in Z_{l+1,l}\overline{\otimes} W^{(r)}_{l,a}. \tag{4.19}$$

Lemma 4.2. *The linear maps $F_{k,l}$ are injective.*

Proof. Assume $K := \ker(F_{k,l}) \neq 0$; it is a graded subspace of $W^{(r)}_{l,a}$. Choose $\mu \in \mathrm{wt}(K)$ such that $\mu+\alpha_i \notin \mathrm{wt}(K)$ for all $1 \le i < N$ and fix $0 \neq w \in K[\mu]$. We show that w is a singular vector, so $w \in \mathbb{M}\omega_l$ and $\omega_l \in K$, a contradiction. It suffices to prove that $L_{ji}(z)w \in K$ for all $1 \le i < j \le N$; this implies $L_{ji}(z)w = 0$ because by assumption on μ the weight space $K[\mu+\epsilon_i-\epsilon_j]$ vanishes.

Suppose $j > r$. If $1 \le p \le N$ and $p \neq j$, then $(k-l)\varpi_r + \epsilon_p - \epsilon_j \notin \mathrm{wt}(Z_{kl})$ by Theorem 1.15. It follows that for $v \in W^{(r)}_{l,a}$ we have in $Z_{kl}\overline{\otimes} W^{(r)}_{l,a}$,

$$L_{ji}(z)(\omega_{kl}\overline{\otimes} v) = L_{jj}(z)\omega_{kl}\overline{\otimes} L_{ji}(z)v = \omega_{kl}\overline{\otimes} L_{ji}(z)v.$$

It follows that $L_{ji}(z)K \subseteq K$ because of the commutativity:

$$L_{ji}(z)F_{k,l} = F_{k,l}L_{ji}(z) \quad \text{for } 1 \le i,\ j \le N \text{with } j > r. \tag{4.20}$$

Suppose $j \le r$. For $p > r$ since $r \ge j > r$ we have $L_{pi}(z)w \in K$ and so $L_{pi}(z)w = 0$. For $p \le r$, by Theorem 1.15, $L_{jp}(z)\omega_{kl} = 0$ if $p \neq j$. This implies

$$L_{ji}(z)(\omega_{kl}\overline{\otimes} w) = L_{jj}(z)\omega_{kl}\overline{\otimes} L_{ji}(z)w = \frac{\theta(z+k\hbar)}{\theta(z+l\hbar)} g(\lambda)(\omega_{kl}\overline{\otimes} L_{ji}(z)w)$$

for certain $g(\lambda) \in \mathbb{M}^\times$. Applying $\mathcal{F}_{k,l}$ we obtain $F_{k,l}L_{ji}(z)w = 0$, as desired. ☐

In what follows k and l denote positive integers, while i, j, m, n, p, q, s, t, u and v denote the integers between 1 and N related to the Lie algebra $\mathfrak{sl}_N$.

Lemma 4.3. *For $k > l$ and $1 \le i \le N$ we have*

$$K_i(z)F_{k,l} = \left(\frac{\theta(z+k\hbar)}{\theta(z+l\hbar)}\right)^{\delta_{i\le r}} F_{k,l}K_i(z). \tag{4.21}$$

[10]In the affine case [Hernandez and Jimbo 2012, (4.26)] the structure map comes from the stronger fact that $Z_{kl}\overline{\otimes} Z_{lm}$ is of highest weight with Z_{km} being the irreducible quotient.

Proof. We compute $\mathcal{D}_i(z)(\omega_{kl}\overline{\otimes}v)$ for $v\in W^{(r)}_{l,a}$ based on the coproduct of Corollary 1.2. If $-\varpi_{N-k}\prec\alpha$ then $\alpha+\varpi_{N-k}\notin\boldsymbol{Q}_-$ and $(k-l)\omega_{kl}+\alpha+\varpi_{N-k}\notin\mathrm{wt}(Z_{kl})$. The extra terms $x_\alpha\tilde{\otimes}y_\alpha$ in the coproduct do not contribute, and so $\mathcal{D}_i(z)(\omega_{kl}\overline{\otimes}v)=\mathcal{D}_i(z)\omega_{kl}\overline{\otimes}\mathcal{D}_i(z)v$. By (1.8) a similar identity holds when $\mathcal{D}_i(z)$ is replaced by $K_i(z)$, because $K_i(z)\omega_{kl}=(\theta(z+k\hbar)/\theta(z+l\hbar))^{\delta_{i\le r}}\omega_{kl}$ is independent of λ. By applying $\mathcal{F}_{k,l}$ to the new identity involving $K_i(z)$, we obtain (4.21). □

From now until Corollary 4.7, we shall fix integers j and p with the condition $1\le j\le r<p\le N$. By Corollary 2.2, there are elements $\omega^{jp}_{kl}\in Z_{kl}$ for $k>l$ such that

$$L_{jp}(z)\omega_{kl}=\frac{\theta(z+(k-1)\hbar+\lambda_{jp})}{\theta(z+l\hbar)}\omega^{jp}_{kl}.$$

Indeed $\omega^{jp}_{kl}=t_{pj}\omega_{kl}$ in the evaluation module $Z_{kl}\cong V_{(k-l)\varpi_r}(l)$. Since $Y_{(k-l)\varpi_r}$ is a rectangle, $\mathbb{M}\omega^{jp}_{kl}$ is the weight space of weight $(k-l)\varpi_r+\epsilon_p-\epsilon_j$.

Lemma 4.4. *In the $\mathcal{E}$-module Z_{kl} we have $\omega^{jp}_{kl}\neq 0$ and*

$$L_{pj}(z)\omega^{jp}_{kl}=-\omega_{kl}\frac{\theta(z+l\hbar-\lambda_{jp})\theta((k-l)\hbar)\theta(\hbar)}{\theta(z+l\hbar)\theta(\lambda_{jp})\theta(\lambda_{jp}+\hbar)}\prod_{r<q\neq p}\frac{\theta(\lambda_{jq}+(k-l+1)\hbar)}{\theta(\lambda_{jq}+\hbar)}.$$

The product is taken over integers q such that $r+1\le q\le N$ and $q\neq p$.

Proof. The weight grading on $Z_{kl}=S_{(k-l)\varpi_r,l}$ indicates $t_{jp}\omega^{jp}_{kl}=g(\lambda)\omega_{kl}$ for certain $g(\lambda)\in\mathbb{M}$. The last relation of Definition 2.1 with $a=d=j$ and $c=b=p$ is applied to the highest weight vector ω_{kl}, the second term vanishes and

$$\frac{\theta(\lambda_{jp}+(k-l+1)\hbar)}{\theta(\lambda_{jp}+(k-l)\hbar)}g(\lambda)=\frac{\theta((k-l)\hbar)\theta(-\hbar)}{\theta(\lambda_{jp})\theta(\lambda_{jp}+(k-l)\hbar)}\prod_{q>r}\frac{\theta(\lambda_{jq}+(k-l+1)\hbar)}{\theta(\lambda_{jq}+\hbar)}.$$

This implies $\omega^{jp}_{kl}\neq 0$. We conclude that $L_{pj}(z)\omega^{jp}_{kl}=\theta(z+l\hbar-\lambda_{jp})g(\lambda)\omega_{kl}/\theta(z+l\hbar)$. □

Lemma 4.5. *Let $k-1>l$. In the $\mathcal{E}$-module $Z_{k,l+1}\overline{\otimes}Z_{l+1,l}$ we have*

$$L_{pj}(z)(\boldsymbol{a}^{(l)}_{jp}(k;\lambda)(\omega_{k,l+1}\overline{\otimes}\omega^{jp}_{l+1,l})-\omega^{jp}_{k,l+1}\overline{\otimes}\omega_{l+1,l})=0,$$

where

$$\boldsymbol{a}^{(l)}_{jp}(k;\lambda):=\frac{\theta((k-l-1)\hbar)\theta(\lambda_{jp}-\hbar)}{\theta(\hbar)\theta(\lambda_{jp})}\prod_{r<q\neq p}\frac{\theta(\lambda_{jq}+(k-l)\hbar)}{\theta(\lambda_{jq}+\hbar)}.$$

Furthermore $\mathcal{G}_{k,l}(\boldsymbol{a}^{(l)}_{jp}(k;\lambda)(\omega_{k,l+1}\overline{\otimes}\omega^{jp}_{l+1,l})-\omega^{jp}_{k,l+1}\overline{\otimes}\omega_{l+1,l})=0.$

Proof. We compute $L_{pj}(z)(\omega^{jp}_{k,l+1}\overline{\otimes}\omega_{l+1,l})=\sum_{q=1}^N L_{pq}(z)\omega^{jp}_{k,l+1}\overline{\otimes}L_{qj}(z)\omega_{l+1,l}$. Since $\omega_{l+1,l}$ is a highest weight vector, the terms with $q>j$ vanish. The weight of $L_{qj}(z)\omega^{jp}_{k,l+1}$ is $(k-l-1)\varpi_r+\epsilon_q-\epsilon_j$, which does not belong to $\mathrm{wt}(Z_{k,l+1})$ for $q<j$. So only the term $q=j$ survives. By Lemma 4.4

$$
\begin{aligned}
&L_{pj}(z)(\omega^{jp}_{k,l+1}\bar{\otimes}\omega_{l+1,l})\\
&\quad= L_{pj}(z)\omega^{jp}_{k,l+1}\bar{\otimes}L_{jj}(z)\omega_{l+1,l}\\
&\quad= -\frac{\theta(z+(l+1)\hbar-\lambda_{jp})\theta((k-l-1)\hbar)\theta(\hbar)}{\theta(z+(l+1)\hbar)\theta(\lambda_{jp})\theta(\lambda_{jp}+\hbar)}\prod_{r<q\neq p}\frac{\theta(\lambda_{jq}+(k-l)\hbar)}{\theta(\lambda_{jq}+\hbar)}\omega_{k,l+1}\\
&\qquad\qquad\bar{\otimes}\frac{\theta(z+(l+1)\hbar)}{\theta(z+l\hbar)}\prod_{q>r}\frac{\theta(\lambda_{jq}+2\hbar)}{\theta(\lambda_{jq}+\hbar)}\omega_{l+1,l}\\
&\quad= -\frac{\theta(z+l\hbar-\lambda_{jp})\theta((k-l-1)\hbar)\theta(\hbar)}{\theta(z+l\hbar)\theta(\lambda_{jp}+\hbar)^2}\prod_{r<q\neq p}\frac{\theta(\lambda_{jq}+(k-l+1)\hbar)}{\theta(\lambda_{jq}+\hbar)}(\omega_{k,l+1}\bar{\otimes}\omega_{l+1,l}).
\end{aligned}
$$

Similar arguments lead to

$$
\begin{aligned}
L_{pj}(z)(\omega_{k,l+1}\bar{\otimes}\omega^{jp}_{l+1,l}) &= L_{pp}(z)\omega_{k,l+1}\bar{\otimes}L_{pj}(z)\omega^{jp}_{l+1,l}\\
&= -\frac{\theta(z+l\hbar-\lambda_{jp})\theta(\hbar)^2}{\theta(z+l\hbar)\theta(\lambda_{jp})\theta(\lambda_{jp}+\hbar)}\prod_{r<q\neq p}\frac{\theta(\lambda_{jq}+2\hbar)}{\theta(\lambda_{jq}+\hbar)}(\omega_{k,l+1}\bar{\otimes}\omega_{l+1,l}).
\end{aligned}
$$

The ratio of the two coefficients of $\omega_{kl}\bar{\otimes}\omega_{l+1,l}$ above is $\boldsymbol{a}^{(l)}_{jp}(k;\lambda+\hbar\epsilon_j)$, which is easily seen to be independent of z. For the last identity, let x be the vector in the argument of $\mathscr{G}_{k,l}$. Then both $\mathscr{G}_{k,l}(x)$ and ω^{jp}_{kl} belong to the one-dimensional weight space of weight $(k-l)\varpi_r+\epsilon_j-\epsilon_p$. These two vectors are proportional, the first is annihilated by $L_{pj}(z)$, while the second is not. So $\mathscr{G}_{k,l}(x)=0$. □

Corollary 4.6. *Let $k-1>l$. In the $\mathcal{E}$-module Z_{kl} we have*

$$L_{jp}(z)\omega_{kl}=\mathscr{G}_{k,l}(\omega_{k,l+1}\bar{\otimes}\omega^{jp}_{l+1,l})\times\boldsymbol{b}^{(l)}_{jp}(k,z;\lambda),$$

where

$$\boldsymbol{b}^{(l)}_{jp}(k,z;\lambda):=\frac{\theta(z+(k-1)\hbar+\lambda_{jp})\theta((k-l)\hbar)}{\theta(z+l\hbar)\theta(\hbar)}\prod_{r<q\neq p}\frac{\theta(\lambda_{jq}+(k-l)\hbar)}{\theta(\lambda_{jq}+\hbar)}.$$

Proof. The idea is similar to [Zhang 2018, Lemma 7.6]. We compute $L_{jp}(z)(\omega_{k,l+1}\bar{\otimes}\omega_{l+1,l})$. As in the proof of Lemma 4.5, only two terms survive:

$$
\begin{aligned}
L_{jp}(z)(\omega_{k,l+1}\bar{\otimes}\omega_{l+1,l}) &= L_{jj}(z)\omega_{k,l+1}\bar{\otimes}L_{jp}(z)\omega_{l+1,l}+L_{jp}(z)\omega_{k,l+1}\bar{\otimes}L_{pp}(z)\omega_{l+1,l}\\
&= \frac{\theta(z+k\hbar)}{\theta(z+(l+1)\hbar)}\prod_{q>r}\frac{\theta(\lambda_{jq}+(k-l)\hbar)}{\theta(\lambda_{jq}+\hbar)}\omega_{k,l+1}\bar{\otimes}\frac{\theta(z+l\hbar+\lambda_{jp})}{\theta(z+l\hbar)}\omega^{jp}_{l+1,l}\\
&\qquad+\frac{\theta(z+(k-1)\hbar+\lambda_{jp})}{\theta(z+(l+1)\hbar)}\omega^{jp}_{k,l+1}\bar{\otimes}\omega_{l+1,l}\\
&= \boldsymbol{e}^{(l)}_{jp}(k,z;\lambda)(\omega_{k,l+1}\bar{\otimes}\omega^{jp}_{l+1,l})+\frac{\theta(z+k\hbar+\lambda_{jp})}{\theta(z+(l+1)\hbar)}(\omega^{jp}_{k,l+1}\bar{\otimes}\omega_{l+1,l}).
\end{aligned}
$$

Here $\boldsymbol{e}^{(l)}_{jp}(k,z;\lambda)$ is the following meromorphic function of $(k,z,\lambda)\in\mathbb{C}\times\mathbb{C}\times\mathfrak{h}$:

$$\frac{\theta(z+k\hbar)\theta(z+l\hbar+\lambda_{jp})}{\theta(z+(l+1)\hbar)\theta(z+l\hbar)}\frac{\theta(\lambda_{jp}+(k-l-1)\hbar)}{\theta(\lambda_{jp})}\prod_{r<q\neq p}\frac{\theta(\lambda_{jq}+(k-l)\hbar)}{\theta(\lambda_{jq}+\hbar)}.$$

Set $x:=\boldsymbol{a}^{(l)}_{jp}(k;\lambda)(\omega_{k,l+1}\overline{\otimes}\omega^{jp}_{l+1,l})-\omega^{jp}_{k,l+1}\overline{\otimes}\omega_{l+1,l}$, which is in the kernel of $\mathcal{G}_{k,l}$ by Lemma 4.5. It follows that for any $g(\lambda)\in\mathbb{M}$ we have

$$L_{jp}(z)\omega_{kl}=L_{jp}(z)\mathcal{G}_{k,l}(\omega_{k,l+1}\overline{\otimes}\omega_{l+1,l})=\mathcal{G}_{k,l}(L_{jp}(z)(\omega_{k,l+1}\overline{\otimes}\omega_{l+1,l})+g(\lambda)x).$$

Let us fix $g(z;\lambda):=\theta(z+k\hbar+\lambda_{jp})/\theta(z+(l+1)\hbar)$. Then $L_{jp}(z)(\omega_{k,l+1}\overline{\otimes}\omega_{l+1,l})+g(z;\lambda)x$ is proportional to $\omega_{k,l+1}\overline{\otimes}\omega^{jp}_{l+1,l}$ and $L_{jp}(z)\omega_{kl}=\mathcal{G}_{k,l}(\omega_{k,l+1}\overline{\otimes}\omega^{jp}_{l+1,l})\times\boldsymbol{b}^{(l)}_{jp}(k,z;\lambda)$ where

$$\begin{aligned}\boldsymbol{b}^{(l)}_{jp}(k,z;\lambda)&=\boldsymbol{e}^{(l)}_{jp}(k,z;\lambda)+g(z;\lambda)\boldsymbol{a}^{(l)}_{jp}(k;\lambda)\\&=\boldsymbol{b}(k,z;\lambda)\times\prod_{r<q\neq p}\frac{\theta(\lambda_{jq}+(k-l)\hbar)}{\theta(\lambda_{jq}+\hbar)},\\ \boldsymbol{b}(k,z;\lambda)&:=\frac{\theta(z+k\hbar)\theta(z+l\hbar+\lambda_{jp})}{\theta(z+(l+1)\hbar)\theta(z+l\hbar)}\frac{\theta(\lambda_{jp}+(k-l-1)\hbar)}{\theta(\lambda_{jp})}+\frac{\theta(z+k\hbar+\lambda_{jp})\theta((k-l-1)\hbar)\theta(\lambda_{jp}-\hbar)}{\theta(z+(l+1)\hbar)\theta(\lambda_{jp})\theta(\hbar)}.\end{aligned}$$

The function $\boldsymbol{b}(k,z;\lambda)$, viewed as an entire function of k, satisfies the same double periodicity as $\theta(k\hbar)\theta(k\hbar+z+\lambda_{jp}-(l+1)\hbar)$. One checks that $\boldsymbol{b}(l,z;\lambda)=0$. This implies

$$\boldsymbol{b}(k,z;\lambda)=\theta(k\hbar+z-\hbar+\lambda_{jp})\theta(k\hbar-l\hbar)f(z;\lambda),$$

where $f(z;\lambda)$ is a meromorphic function of $(z;\lambda)\in\mathbb{C}\times\mathfrak{h}$ independent of k. Now setting $k\hbar=-z$, we obtain $f(z;\lambda)=1/(\theta(z+l\hbar)\theta(\hbar))$. □

Corollary 4.7. *Let $1\le i,j\le N$ with $j\le r$. For $k-1>l$ and $x\in W^{(r)}_{l,a}$*

$$\begin{aligned}L_{ji}(z)F_{k,l}(x)=F_{k,l}\frac{\theta(z+k\hbar)}{\theta(z+l\hbar)}\mu_{\boldsymbol{l}}\Big(\prod_{q=r+1}^{N}\frac{\theta(\lambda_{jq}+(k-l+1)\hbar)}{\theta(\lambda_{jq}+\hbar)}\Big)L_{ji}(z)x\\+F_{k,l+1}\mathcal{F}_{l+1,l}\Big(\sum_{p=r+1}^{N}\omega^{jp}_{l+1,l}\overline{\otimes}\mu_{\boldsymbol{l}}(\boldsymbol{b}^{(l)}_{jp}(k,z;\lambda))L_{pi}(z)x\Big).\end{aligned}\tag{4.22}$$

Proof. Consider $L_{ji}(z)F_{k,l}(x)=\mathcal{F}_{k,l}\big(\sum_{p=1}^{N}L_{jp}(z)\omega_{kl}\overline{\otimes}L_{pi}(z)x\big)$. As in the proof of Lemma 4.5, $L_{jp}(z)\omega_{kl}=0$ if $p\notin\{j,r+1,r+2,\dots,N\}$. For $p=j$, we obtain the first row of (4.22), while for $r<p\le N$, Corollary 4.6 and (4.19) with $v=\omega^{jp}_{l+1,l}$ give the second row. □

Fix weight bases $\mathcal{B}_l$ of $W^{(r)}_{l,a}$ for $l>0$ uniformly so that $F_{k,l}(\mathcal{B}_l)\subseteq\mathcal{B}_k$.

We view $\boldsymbol{b}_{jp}^{(l)}(c, z; \lambda)$ in Corollary 4.6 as a meromorphic function of $(c, z, \lambda) \in \mathbb{C}^2 \times \mathfrak{h}$. For $1 \le i, j \le N$, $l > 0$ and $c, z \in \mathbb{C}$, define $\mathscr{L}_{ji}^{(l)}(c, z) : W_{l,a}^{(r)} \to W_{l+1,a}^{(r)}$ by

$$\mathscr{L}_{ji}^{(l)}(c, z)x = F_{l+1,l} L_{ji}(z)x = \frac{\theta(z + (\gamma_j + \delta_{ij} - 1)\hbar + \lambda_{ji})}{\theta(z)} F_{l+1,l} t_{ij} x \qquad \text{for } j > r,$$

$$\begin{aligned}\mathscr{L}_{ji}^{(l)}(c, z)x = \frac{\theta(z + c\hbar)}{\theta(z + l\hbar)} \prod_{q=r+1}^{N} \frac{\theta(\lambda_{jq} + (c + \gamma_{jq} + \delta_{ij} - \delta_{iq})\hbar)}{\theta(\lambda_{jq} + (l + \gamma_{jq} + \delta_{ij} - \delta_{iq})\hbar)} F_{l+1,l} L_{ji}(z)x \\ + \sum_{p=r+1}^{N} \boldsymbol{b}_{jp}^{(l)}(c, z; \lambda + (\gamma + l\varpi_r + \epsilon_i - \epsilon_p)\hbar) \mathscr{F}_{l+1,l}(\omega_{l+1,l}^{jp} \overline{\otimes} L_{pi}(z)x) \quad \text{for } j \le r.\end{aligned}$$

Here $x \in W_{l,a}^{(r)}[\gamma + l\varpi_r]$ and δ_{ij} is the usual Kronecker symbol. Corollary 2.2 applied to the evaluation module $W_{l,a}^{(r)} \cong V_{l\varpi_r}(0)$ indicates that for $b' \in \mathcal{B}_{l+1}$ and $b \in \mathcal{B}_l$

> $\mathscr{L}_{ji}^{(l)}(c, z)$ is a difference map of bidegree $(\epsilon_j - \varpi_r, \epsilon_i)$. Its matrix entry $[\mathscr{L}_{ji}^{(l)}]_{b'b}(c, z; \lambda)$ is a meromorphic function of $(c, z, \lambda) \in \mathbb{C}^2 \times \mathfrak{h}$. Moreover, $\theta(z)\theta(z + l\hbar)[\mathscr{L}_{ji}^{(l)}]_{b'b}(c, z; \lambda)$ is entire on (c, z) for generic λ.

As a unification of (4.20) and (4.22), we have

$$L_{ji}(z) F_{k,l} = F_{k,l+1} \mathscr{L}_{ji}^{(l)}(k, z) \quad \text{for } k > l + 1. \tag{4.23}$$

For $k \in \mathbb{Z}_{>0}$ and $z \in \mathbb{C}$ let $\Xi(c; k, z)$ be the set of entire functions $F(c)$ of $c \in \mathbb{C}$ with the following double periodicity:

$$F(c + \hbar^{-1}) = (-1)^k F(c) \quad \text{and} \quad F(c + \tau\hbar^{-1}) = (-1)^k e^{-ki\pi\tau - 2ki\pi c\hbar - 2i\pi z} F(c).$$

A typical example is $\theta(c\hbar)^{k-1}\theta(c\hbar + z)$. Such a function is called *homogeneous*. If $f(c), g(c) \in \Xi(c; k, z)$, then we write $f(c) \approx g(c)$.

Note that $\Xi(c; k, z)\Xi(c; k', z') \subseteq \Xi(c; k + k', z + z')$.

Lemma 4.8. *Let $b \in \mathcal{B}_l$ be of weight $\gamma + l\varpi_r$ and $b' \in \mathcal{B}_{l+1}$. For $j > r$ the matrix entry $[\mathscr{L}_{ji}^{(l)}]_{b'b}(c, z; \lambda)$ is independent of c. For $j \le r$ as entire functions of c*

$$[\mathscr{L}_{ji}^{(l)}]_{b'b}(c, z; \lambda) \approx \theta(z + c\hbar) \prod_{q=r+1}^{N} \theta(\lambda_{jq} + (c + \gamma_{jq} + \delta_{ij} - \delta_{iq})\hbar).$$

Moreover, $\theta(z)[\mathscr{L}_{ji}^{(l)}]_{b'b}(c, z; \lambda)$ is an entire function of (c, z) for generic λ.

Proof. In the case $j > r$, Corollary 2.2 is applied to $W_{l,a}^{(r)} \cong S_{l\varpi_r,0}$, the matrix entry is of the form $\theta(z + (\gamma_j + \delta_{ij} - 1)\hbar + \lambda_{ji})g_1(\lambda)/\theta(z)$ for $g_1(\lambda) \in \mathbb{M}$. Assume $j \le r$. By Corollary 4.6 the matrix entry is of the form $E(c, z; \lambda)g_2(\lambda)/(\theta(z)\theta(z + l\hbar))$, where $g_2(\lambda) \in \mathbb{M}$ and $E(c, z; \lambda)$ is an entire function of

$(c, z, \lambda) \in \mathbb{C} \times \mathbb{C} \times \mathfrak{h}$. As functions of z and c, respectively, we have

$$E(c, z\hbar; \lambda) \in \Xi(z; 2, (c + l + \gamma_j + \delta_{ij} - 1)\hbar + \lambda_{ji}),$$

$$E(c, z; \lambda) \in \Xi(c; N - r + 1, \sum_{q=r+1}^{N} (\lambda_{jq} + (\gamma_{jq} + \delta_{ij} - \delta_{iq})\hbar) + z).$$

On the other hand, for $k > l + 1$ we have by Corollary 2.2 and (4.23),

$$F_{k,l+1}\mathscr{L}_{ji}^{(l)}(k, z)b = L_{ji}(z)F_{k,l}v = \frac{\theta(z + \lambda_{ji} + (k + \gamma_j + \delta_{ij} - 1)\hbar)}{\theta(z)} t_{ij} F_{k,l} b.$$

The right-hand side as a function of z is regular at $z = -l\hbar$, so are any of the coefficients of the left-hand side $E(k, z; \lambda)g_2(\lambda)/(\theta(z)\theta(z + l\hbar))$. This forces $E(k, -l\hbar; \lambda) = 0$ and

$$E(c, z; \lambda) = \theta(z + l\hbar)\theta(z + (c + \gamma_j + \delta_{ij} - 1)\hbar + \lambda_{ji})D(c; \lambda)g_3(\lambda),$$

where $g_3(\lambda) \in \mathbb{M}$ and $D(c; \lambda)$ is an entire function of (c, λ). Applying the double periodicity with respect to c once more, we obtain the desired result. $\square$

Lemma 4.9. *Let $f(c)$ be a homogeneous entire function. If $f(k) = 0$ for infinitely many integers k, then $f(c)$ is identically zero.*

Proof. By definition the homogeneous entire function $f(c)$, if nonzero, can be written as a product of theta functions $\theta(c\hbar + z)$. Since $\hbar \notin \mathbb{Q} + \mathbb{Q}\tau$, each of these theta functions of c can not have zeroes at infinitely many integers. $\square$

Let W_∞ be the inductive limit of the inductive system $(W_{l,a}^{(r)}, F_{k,l})$ of vector spaces (over $\mathbb{M}$), with the $F_l : W_{l,a}^{(r)} \to W_\infty$ for $l > 0$ being the structural maps.

From now on fix $d \in \mathbb{C}$. A vector $0 \neq w \in W_\infty$ is of weight $d\varpi_r + \gamma$ if there exist $l > 0$ and $w' \in W_{l,a}^{(r)}[l\varpi_r + \gamma]$ such that $w = F_l(w')$. The weight grading is independent of the choice of l because $F_{k,l}$ sends $W_{l,a}^{(r)}[l\varpi_r + \gamma]$ to $W_{k,a}^{(r)}[k\varpi_r + \gamma]$. Let W_∞^d denote the resulting object of $\mathcal{V}$. By construction $\mathrm{wt}(W_\infty^d) \subseteq d\varpi_r + \boldsymbol{Q}_-$, and $F_l : W_{l,a}^{(r)} \to W_\infty^d$ is a difference map of bidegree $((l - d)\varpi_r, 0)$.

Let $\gamma \in \boldsymbol{Q}_-$. The injective maps $F_{k,l}$ together with Theorems 1.15 and 2.3 imply that

$$\dim(W_{k,a}^{(r)}[k\varpi_r + \gamma]) = d_{k\varpi_r}[k\varpi_r + \gamma],$$

as $k \to \infty$, converges to an integer which is exactly $\dim(W_\infty^d[d\varpi_r + \gamma])$. So W_∞^d is an object of $\mathcal{V}_{\mathrm{ft}}$. Our goal is to make W_∞^d into an $\mathcal{E}$-module in category $\mathcal{O}$ with favorable q-character.[11]

[11] In the affine case, the matrix entries of analogs of $\mathscr{L}_{ji}^{(l)}(k, z)$ are Laurent polynomials of $e^{k\hbar}$. Hernandez and Jimbo [2012] proved this by using elimination theorems of q-characters, and then took the limit $e^{k\hbar} \to 0$ as $k \to \infty$ to obtain modules over Borel subalgebras of affine quantum groups. Later in [Zhang 2017; 2018] an elementary proof of polynomiality was given based on $\mathfrak{sl}_2$-representation theory, which by taking limit $e^{k\hbar} \to e^{d\hbar}$ as $k \to \infty$ (with $d \in \mathbb{C}$ a new parameter) resulted in modules over affine quantum groups. Here we adapt the second approach to the elliptic case.

For $1 \le i, j \le N$ and $z \in \mathbb{C}$ with $\theta(z) \neq 0$, the $\mathscr{L}_{ji}^{(l)}(d, z)$ constitute a morphism of inductive system of $\mathbb{C}$-vector spaces:

$$\begin{tikzcd} W_{l,a}^{(r)} \arrow[r, "\mathscr{L}_{ji}^{(l)}(d,z)"] \arrow[d, "F_{l',l}"'] & W_{l+1,a}^{(r)} \arrow[d, "F_{l'+1,l+1}"] \\ W_{l',a}^{(r)} \arrow[r, "\mathscr{L}_{ji}^{(l')}(d,z)"] & W_{l'+1,a}^{(r)} \end{tikzcd} \quad \text{for } l' > l.$$

Indeed, the matrix entries of $F_{l'+1,l+1}\mathscr{L}_{ji}^{(l)}(c, z)$ and $\mathscr{L}_{ji}^{(l')}(c, z)F_{l',l}$, as difference maps $W_{l,a}^{(r)} \to W_{l'+1,a}^{(r)}$, are homogeneous entire functions of c with the same double periodicity by Lemma 4.8 and are equal at all integers c larger than $l' + 1$ by (4.23). By Lemma 4.9 these two maps coincide for all $c \in \mathbb{C}$. Define

$$\mathscr{L}_{ji}^{d}(z) := \varinjlim \mathscr{L}_{ji}^{(l)}(d, z) \in \mathrm{Hom}_{\mathbb{C}}(W_{\infty}^{d}, W_{\infty}^{d}).$$

For $x \in W_{\infty}^{d}[d\varpi_r + \gamma]$ with $x = F_l(x')$ and $x' \in W_{l,a}^{(r)}[l\varpi_r + \gamma]$, we have

$$\mathscr{L}_{ji}^{d}(z)x = F_{l+1}\mathscr{L}_{ji}^{(l)}(d, z)x'. \tag{4.24}$$

The difference maps $\mathscr{L}_{ji}^{(l)}(d, z)$ and F_{l+1} are of bidegree $(\epsilon_j - \varpi_r, \epsilon_i)$ and $((l + 1 - d)\varpi_r, 0)$ respectively. So $\mathscr{L}_{ji}^{d}(z)$ is a difference operator of bidegree (ϵ_j, ϵ_i).

Proposition 4.10. *$(W_{\infty}^{d}, \mathscr{L}_{ji}^{d}(z))$ is an $\mathcal{E}$-module in category $\mathcal{O}$. Moreover,*

$$\chi_q(W_{\infty}^{d}, \mathscr{L}_{ji}^{d}(z)) = \boldsymbol{w}_{d,a}^{(r)} \times \lim_{k\to\infty} (\boldsymbol{w}_{k,a}^{(r)})^{-1} \chi_q(W_{k,a}^{(r)}). \tag{4.25}$$

Proof. We need to prove conditions (M1)–(M3) of Section 1C. First (M1) follows from (4.24) and from the comments before (4.23). To prove (M2), let $x \in W_{\infty}^{d}[d\varpi_r + \gamma]$ and $x' \in W_{l,a}^{(r)}[l\varpi_r + \gamma]$ such that $x = F_l(x')$. We assume l so large that $W_{\infty}^{d}[d\varpi_r + \gamma]$ and $W_{l,a}^{(r)}[l\varpi_r + \gamma]$ have the same dimension.

Step I: Proof of (M2) We need to show that for $1 \le i, j, m, n \le N$

$$\sum_{p,q} R_{mn}^{pq}(z - w; \lambda + (\epsilon_i + \epsilon_j + d\varpi_r + \gamma)\hbar)\mathscr{L}_{pi}^{d}(z)\mathscr{L}_{qj}^{d}(w)x = \sum_{s,t} R_{st}^{ij}(z - w; \lambda)\mathscr{L}_{nt}^{d}(w)\mathscr{L}_{ms}^{d}(z)x \in W_{\infty}^{d}.$$

Here at the right-hand side we have used $R_{mn}^{pq}(z; \lambda) = R_{mn}^{pq}(z; \lambda + \hbar\epsilon_p + \hbar\epsilon_q)$ to move R to the left. By (4.24) it is enough to prove the equation

$$\begin{aligned}\sum_{p,q} R_{mn}^{pq}(z - w; \lambda + (\epsilon_i + \epsilon_j + c\varpi_r + \gamma)\hbar)\mathscr{L}_{pi}^{(l+1)}(c, z)\mathscr{L}_{qj}^{(l)}(c, w)x' \\ = \sum_{s,t} R_{st}^{ij}(z - w; \lambda)\mathscr{L}_{nt}^{(l+1)}(c, w)\mathscr{L}_{ms}^{(l)}(c, z)x' \in W_{l+2,a}^{(r)}.\end{aligned} \tag{4.26}$$

Let $A_1(c, z, w)$ and $A_2(c, z, w)$ denote the left-hand side and the right-hand side of this equation without x'. These are difference maps $W_{l,a}^{(r)} \to W_{l+2,a}^{(r)}$ of bidegree $(\epsilon_m + \epsilon_n - 2\varpi_r, \epsilon_i + \epsilon_j)$, as $R_{mn}^{pq} \neq 0$ implies $\epsilon_m + \epsilon_n = \epsilon_p + \epsilon_q$.

Claim 1. *For $b \in \mathcal{B}_l$ of weight $l\varpi_r + \gamma$ and $b' \in \mathcal{B}_{l+2}$, as entire functions of c,*

$$[A_1]_{b'b}(c, z, w; \lambda) \approx [A_2]_{b'b}(c, z, w; \lambda).$$

This is divided into four cases. For simplicity let us drop b', b, z, w and λ from A_1 and A_2.

Case 1.1: $m, n > r$. $A_1(c)$ and $A_2(c)$ are independent of c by Lemma 4.8.

Case 1.2: $m, n \le r$. At the left-hand side of (4.26) we have $\{p, q\} = \{m, n\}$ and so R^{pq}_{mn} is independent of c. At the right-hand side $\{s, t\} = \{i, j\}$. Therefore

$$\begin{aligned} A_1(c) \approx \theta(c\hbar + z) \prod_{u>r} \theta(\lambda_{pu} + (c + \gamma_{pu} + \delta_{ip} - \delta_{iu} + \delta_{jp} - \delta_{ju} - \delta_{qp} + \delta_{qu})\hbar) \\ \times \theta(c\hbar + w) \prod_{v>r} \theta(\lambda_{qv} + (c + \gamma_{qv} + \delta_{jq} - \delta_{jv} + \delta_{iq} - \delta_{iv})\hbar), \\ A_2(c) \approx \theta(c\hbar + w) \prod_{u>r} \theta(\lambda_{nu} + (c + \gamma_{nu} + \delta_{tn} - \delta_{tu} + \delta_{sn} - \delta_{su} - \delta_{mn} + \delta_{mu})\hbar) \\ \times \theta(c\hbar + z) \prod_{v>r} \theta(\lambda_{mv} + (c + \gamma_{mv} + \delta_{sm} - \delta_{sv} + \delta_{tm} - \delta_{tv})\hbar). \end{aligned}$$

These formulas are deduced from Lemma 4.8. One needs to take into account the shifts of γ and λ. For example at the left-hand side of (4.26), the terms $\mathcal{L}_{qj}$ and $\mathcal{L}_{pi}$ shift γ and λ by $\epsilon_j - \epsilon_q$ and $\hbar\epsilon_i$, respectively. The right-hand sides of these two formulas lie in $\Xi(c; 2+2N-2r, e)$ with $e \in \mathbb{C}$ independent of the choices of p, q, s and t.

Case 1.3: $m \le r < n$. At the right-hand side $\{s, t\} = \{i, j\}$ and

$$\begin{aligned} A_2(c) &\approx \theta(c\hbar + z) \prod_{v>r} \theta(\lambda_{mv} + (c + \gamma_{mv} + \delta_{sm} - \delta_{sv} + \delta_{tm} - \delta_{tv})\hbar) \\ &\approx \theta(c\hbar + z) \prod_{v>r} \theta(\lambda_{mv} + (c + \gamma_{mv} + \delta_{im} - \delta_{iv} + \delta_{jm} - \delta_{jv})\hbar). \end{aligned}$$

The last term is independent of s and t. On the other hand $A_1(c) = E(c) + F(c)$ where E and F correspond to $(p, q) = (m, n)$ and $(p, q) = (n, m)$ respectively and so

$$\begin{aligned} E(c) &\approx \frac{\theta(f - \hbar)}{\theta(f)} \theta(c\hbar + z) \prod_{u>r} \theta(\lambda_{mu} + (c + \gamma_{mu} + \delta_{im} - \delta_{iu} + \delta_{jm} - \delta_{ju} + \delta_{nu})\hbar), \\ F(c) &\approx \frac{\theta(f + z - w)}{\theta(f)} \theta(c\hbar + w) \prod_{v>r} \theta(\lambda_{mv} + (c + \gamma_{mv} + \delta_{jm} - \delta_{jv} + \delta_{im} - \delta_{iv})\hbar). \end{aligned}$$

Here $f := c\hbar + \lambda_{mn} + (\gamma_{mn} + \delta_{im} - \delta_{in} + \delta_{jm} - \delta_{jn})\hbar$. We observe easily that $A_2(c) \approx E(c) \approx F(c)$ and so $A_1(c) \approx A_2(c)$ are homogeneous.

Case 1.4: $n \le r < m$. This is parallel to the third case.

Claim 2. *In Claim 1 equality holds for $c = k \in \mathbb{Z}_{>l+2}$.*

Let us apply $F_{k,l+2}$ to (4.26) with $c = k$ and $x' = b$. By (4.23)

$$F_{k,l+2}\mathcal{L}^{(l+1)}_{pi}(k, z)\mathcal{L}^{(l)}_{qj}(k, w) = L_{pi}(z) F_{k,l+1}\mathcal{L}^{(l)}_{qj}(k, w)x' = L_{pi}(z)L_{qj}(w)F_{k,l},$$

and similarly $F_{k,l+2}\mathscr{L}_{nt}^{(l+1)}(d,w)\mathscr{L}_{ms}^{(l)}(d,z)b = L_{nt}(w)L_{ms}(z)F_{k,l}b$. We obtain the defining relation $RLL = LLR$ of the $\mathcal{E}$-module $W_{k,a}^{(r)}$ applied to the vector $F_{k,l}(b)$. Since $F_{k,l+2}$ is injective, (4.26) holds for $c = k$ and $x' = b$. This proves Claim 2.

Together with Lemma 4.9, we obtain equality in Claim 1 for all $c \in \mathbb{C}$. This proves (4.26).

Step II: Let $1 \leq i \leq N$. We have by (1.6) and (4.26)

$$\mathcal{D}_i(z)x = \frac{\Theta_i(\lambda)}{\Theta_i(\lambda+(d\varpi_r+\gamma)\hbar)} F_{l+i}\mathscr{D}_i^{(l)}(d,z)x'. \tag{4.27}$$

Here $\mathscr{D}_i^{(l)}(c,z) = \sum_{\sigma\in\mathfrak{S}^i} T_\sigma(c,z)$ and $T_\sigma(c,z) : W_{l,a}^{(i)} \to W_{l+i,a}^{(i)}$, for $\sigma \in \mathfrak{S}^i$, is given by

$$\mathrm{sign}(\sigma)\mathscr{L}_{\sigma(N),N}^{(l+i-1)}(c,z)\mathscr{L}_{\sigma(N-1),N-1}^{(l+i-2)}(c,z+\hbar)\cdots\mathscr{L}_{\sigma(N-i+1),N-i+1}^{(l)}(c,z+(i-1)\hbar).$$

Each $T_\sigma(c,z)$ is a difference map of bidegree $(-\varpi_{N-i}-i\varpi_r, -\varpi_{N-i})$. Define the meromorphic function of $(c,z) \in \mathbb{C}^2$ (note that l is fixed)

$$g(c,z) = 1 \ \text{ if } i < N+1-r \quad \text{and} \quad g(c,z) = \prod_{p=N-i+1}^{r} \frac{\theta(z+(N-p+c)\hbar)}{\theta(z+(N-p+l)\hbar)} \ \text{ otherwise.}$$

Claim 3. *For $b \in \mathcal{B}_l$ of weight $l\varpi_r+\gamma$ and $b' \in \mathcal{B}_{l+i}$, as entire functions of $c \in \mathbb{C}$,*

$$[T_\sigma]_{b'b}(c,z;\lambda) \approx g_i(c,z)\Theta_i(\lambda+(c\varpi_r+\gamma)\hbar).$$

The idea is the same as Claim 1, based on Lemma 4.8. If $N-i+1 > r$, then $T_\sigma^b(c,z;\lambda)$ and $\Theta_i(\lambda+(c\varpi_r+\gamma)\hbar)$ are independent of c, and we are done.

Assume $N-i+1 \leq r$. By (1.5) and Lemma 4.8,

$$\Theta_i(\lambda+(c\varpi_r+\gamma)\hbar) \approx \prod_{p=N-i+1}^{r}\prod_{u=r+1}^{N}\theta(\lambda_{pu}+(c+\gamma_{pu})\hbar),$$

$$[T_{\mathrm{Id}}]_{b,b'}(c,z;\lambda) \approx \prod_{p=N-i+1}^{r}\theta(z+(c+N-p)\hbar)\prod_{u=r+1}^{N}\theta(\lambda_{pu}^{(p)}+(c+\gamma_{pu}+1)\hbar).$$

Here $\lambda^{(p)} = \lambda + \hbar\sum_{v=p+1}^{N}\epsilon_v$ and so $\lambda_{pu}^{(p)} = \lambda_{pu} - \hbar$ for $p \leq r < u$. The case $\sigma = \mathrm{Id}$ in Claim 3 is now obvious. It remains to show $[T_\sigma]_{b'b}(c,z;\lambda) \approx [T_{\sigma'}]_{b'b}(c,z;\lambda)$ for all $\sigma,\sigma' \in \mathfrak{S}^i$. One can assume $\sigma' = \sigma s_j$ where $s_j = (j, j+1)$ is a simple transposition with $N-i+1 \leq j < N-1$. Let us define

$$p := \sigma(j+1), q := \sigma(j), \quad l' := l+i+j-1-N, \quad w := z+(N-j)\hbar.$$

Then we have the decomposition of difference maps

$$T_\sigma(c,z) = \mathrm{sign}(\sigma)A(c,z)U_{pq}(c,w)B(c,z) \quad \text{and} \quad T_{\sigma'}(c,z) = \mathrm{sign}(\sigma)A(c,z)U_{qp}(c,w)B(c,z).$$

The difference maps A, B, U are defined by (descending order in the products)

$$A(c,z)=\prod_{u=N}^{j+2}\mathscr{L}^{(l+i-1-N+u)}_{\sigma(u),u}(c,z+(N-u)\hbar): W^{(r)}_{l'+2,a}\to W^{(r)}_{l+i-1,a},$$
$$B(c,z)=\prod_{u=j-1}^{N-i+1}\mathscr{L}^{(l+i-1-N+u)}_{\sigma(u),u}(c,z+(N-u)\hbar): W^{(r)}_{l,a}\to W^{(r)}_{j-1,a},$$
$$U_{pq}(c,w)=\mathscr{L}^{(l'+1)}_{p,j+1}(c,w-\hbar)\mathscr{L}^{(l')}_{qj}(c,w): W^{(r)}_{l',a}\to W^{(r)}_{l'+2,a}.$$

Flipping p and q one gets U_{qp}. Now $[T_\sigma]_{b'b}(c,z;\lambda)\approx[T_{\sigma'}]_{b'b}(c,z;\lambda)$ is a consequence of the following claim.

Claim 4. *For $y\in\mathcal{B}_{l'}$ of weight $l'\varpi_r+\eta$ and $y'\in\mathcal{B}_{l'+2}$, as entire functions of c,*

$$[U_{pq}]_{y'y}(c,w;\lambda)\approx[U_{qp}]_{y'y}(c,w;\lambda).$$

If $p,q\le r$, then by Lemma 4.8 (setting $\eta'=\eta+\epsilon_j-\epsilon_q$ and $\lambda'=\lambda+\hbar\epsilon_{j+1}$)

$$[U_{pq}]_{y'y}(c,w;\lambda)\approx\theta(w+(c-1)\hbar)\prod_{u=r+1}^{N}\theta(\lambda_{pu}+(c+\eta'_{pu}+\delta_{p,j+1}-\delta_{u,j+1})\hbar)$$
$$\times\,\theta(w+c\hbar)\prod_{v=r+1}^{N}\theta(\lambda'_{qv}+(c+\eta_{qv}+\delta_{jq}-\delta_{jv})\hbar).$$

We have $U^{b'}_{pq}(c,w;\lambda)\in\Xi(c;2N-2r+2,e)$ with $e=e(p,q)$ symmetric on p,q. So $[U_{pq}]_{y'y}(c,w;\lambda)\approx[U_{qp}]_{y'y}(c,w;\lambda)$.

The other cases of p and q are proved in the same way as in Claim 1.

Step III: Proof of (M3) Let $k>l+i$. Notice that $\mathcal{D}_i(z)\omega_{kl}=g_i(k,z)\omega_{kl}$. From the proof of Lemma 4.3 and from (4.23) and (4.27) we get

$$g_i(k,z)F_{k,l}\mathcal{D}_i(z)x'=\mathcal{D}_i(z)F_{k,l}(x')=F_{k,l+i}\frac{\Theta_i(\lambda)}{\Theta_i(\lambda+(k\varpi_r+\gamma)\hbar)}\mathscr{D}^{(l)}_i(k,z)x'.$$

Applying F_k to this identity and multiplying by $\Theta_i(\lambda+(k\varpi_r+\gamma)\hbar)$ we have

$$\Theta_i(\lambda+(k\varpi_r+\gamma)\hbar)g_i(k,z)F_l\mathcal{D}_i(z)x'=\Theta_i(\lambda)F_{l+i}\mathscr{D}^{(l)}_i(k,z)x'\quad\text{for } k>l+i.$$

Both sides after taking coefficients with respect to a basis of $W^d_\infty[d\varpi_r+\gamma]$ can be viewed as entire functions of $k\in\mathbb{C}$, and they satisfy the same double periodicity by Claim 3. By Lemma 4.9, the above identity holds for all $k\in\mathbb{C}$. Taking $k=d$, by (4.27), we obtain $\mathcal{D}_i(z)x=g_i(d,z)F_l\mathcal{D}_i(z)x'$.

Let B be a basis of $W^{(r)}_{l,a}[l\varpi_r+\gamma]$ satisfying the upper triangular property of (M3). Then the basis $F_l(B)$ of $W^d_\infty[d\varpi_r+\gamma]$ satisfies the same property. The $\mathcal{E}$-module W^d_∞ is in category $\mathcal{O}$. The diagonal entry of $\mathcal{D}_i(z)$ associated to $F_l(x')\in W^d_\infty$ for $x'\in B$ is equal to that of $\mathcal{D}_i(z)$ associated to $x'\in W^{(r)}_{l,a}$ multiplied by $g_i(d,z)$. The q-character formula in (4.25) follows from the explicit formula of $g_i(d,z)$. □

Question 4.11. Let $F(c)$ be a finite sum of homogeneous entire functions. If $F(k) = 0$ for infinitely many integers k, then is $F(c)$ identically zero?

If the answer to this question is affirmative, then the proof of Proposition 4.10 can be largely simplified, Claims 1, 3 and 4 are not necessary.[12]

Remark 4.12. By Lemma 4.8, $W_\infty^d \cong \mathscr{W}(0)$ with $\mathscr{W}$ an $\mathfrak{e}$-module of character $\lim_{k\to\infty} e^{(d-k)\varpi_r}\chi(W_{k,a}^{(r)})$, so it is in the image of the functor [Etingof and Moura 2002, Proposition 4.1]. By Lemma 4.2 and its proof, $\mathscr{W}$ contains a unique highest weight vector up to scalar. Let Q be the quotient of standard Verma module $M_{d\varpi_r,1}$ in [Tarasov and Varchenko 2001, Proposition 4.7] by $t_{a+1,a}v_{d\varpi_r,1}$ for $a \neq r$. Then $\mathscr{W}$ is the contragradient module to Q in [Tarasov and Varchenko 2001, §6]. It is interesting to have a direct proof of $\mathscr{W}(0)$ being in category $\mathcal{O}$.

For $x \in \mathbb{C}$ let $\mathscr{W}_{d,x}^{(r)}$ be the pullback of W_∞^d by Φ_{x-a} in (1.4); it is called *an asymptotic module*. Set $\mathscr{W}_{d,x}^{(N)} := S(\boldsymbol{w}_{d,x}^{(N)})$ and $\mathscr{W}_{s,x} := \mathscr{W}_{x,0}^{(s)}$ for $1 \leq s \leq N$.

Corollary 4.13. (i) *$\mathcal{R}$ is the set of rational e-weigths.*

(ii) *For any $\mathcal{E}$-module M in category $\mathcal{O}$, we have* $\mathrm{wt}_e(M) \subset \mathcal{R}$.

(iii) *For $d, x \in \mathbb{C}$ and $1 \leq r \leq N$ we have in $\mathcal{M}_\mathfrak{t}$ and $K_0(\mathcal{O})$ respectively*

$$\chi_q(\mathscr{W}_{d,x}^{(r)}) = \boldsymbol{w}_{d,x}^{(r)} \times \chi_q(\mathscr{W}_{0,x}^{(r)}) \quad \textit{and} \quad [\mathscr{W}_{d,0}^{(r)}][\mathscr{W}_{0,x}^{(r)}] = [\mathscr{W}_{d-x,x}^{(r)}][\mathscr{W}_{x,0}^{(r)}]. \tag{4.28}$$

Proof. Conclusion (iii) comes from (4.25), as in the proof of [Felder and Zhang 2017, Theorem 3.11]. As a highest weight of $\mathscr{W}_{d,x}^{(r)}$, $\boldsymbol{w}_{d,x}^{(r)}$ belongs to $\mathcal{R}$. Together with Lemma 1.13 we obtain (i). In (ii) one may assume M irreducible. Then M is a subquotient of a tensor product of asymptotic modules. Since e-weigths of an asymptotic module are rational, we conclude from the multiplicative structure of q-characters in Proposition 1.10. □

In Section 2B the evaluation module $V_\mu(x)$ is an irreducible highest weight module in category $\tilde{\mathcal{O}}$. Its highest weight is easily shown to be rational.

Corollary 4.14. *$V_\mu(x)$ is in category $\mathcal{O}$ for $\mu \in \mathfrak{h}$ and $x \in \mathbb{C}$.*

Finite-dimensional modules in category $\mathcal{O}$ are related to the asymptotic modules by *generalized Baxter relations* in the sense of Frenkel and Hernandez [2015, Theorem 4.8]; see [Felder and Zhang 2017, Corollary 4.7] and [Zhang 2017, Theorem 5.11] for a closer situation.

Theorem 4.15. *Let V be a finite-dimensional $\mathcal{E}$-module in category $\mathcal{O}$. Then*

$$[V] = \sum_{j=1}^{\dim V} [S(\boldsymbol{d}_j)] \times \boldsymbol{m}_j \tag{4.29}$$

[12]In the affine case, by footnote 11 the situation is much easier; a Laurent polynomial vanishing at infinitely many integers must be zero, see [Zhang 2017, §2].

in a fraction ring of the Grothendieck ring of $\mathcal{O}$. Here $\boldsymbol{d}_j \in \mathcal{R}_0$ and $\boldsymbol{m}_j$ is a product of the $[\mathcal{W}_{r,x}]/[\mathcal{W}_{r,y}]$ with $x, y \in \mathbb{C}$ and $1 \le r < N$.

Proof. The idea is the same as [Frenkel and Hernandez 2015]. Since the q-character map is injective, one can replace isomorphism classes with q-characters. $\chi_q(V)$ is the sum of its e-weigths, the number of which is $\dim V$. By Corollary 4.13, any e-weigth $\boldsymbol{e}$ is of the form $\boldsymbol{d}\prod \Psi_{r,x}/\Psi_{r,y} = \chi_q(S(\boldsymbol{d}))\prod \chi_q(\mathcal{W}_{r,x})/\chi_q(\mathcal{W}_{r,y})$, where $\boldsymbol{d} \in \mathcal{R}_0$ and the product is over $1 \le r < N$ and $x, y \in \mathbb{C}$. This proves (4.29) in terms of q-characters. $\square$

To compare with [Frenkel and Hernandez 2015, Theorem 4.8], one imagines that for $1 \le r < N$ and $x \in \mathbb{C}$ there existed a *positive prefundamental module* $L^+_{r,x}$ in category $\mathcal{O}$ with q-character $\chi_q(L^+_{r,x}) = \Psi_{r,x} \times \chi(L^+_{r,0})$ as in [Frenkel and Hernandez 2015, Theorem 4.1]. Then $[\mathcal{W}_{r,x}]/[\mathcal{W}_{r,y}] = [L^+_{r,x}]/[L^+_{r,y}]$. Note that the q-character of $\mathcal{W}^{(r)}_{0,x}$ in (4.28) is different from its character.

Example 4.16. Let $N = 3$. Consider the vector representation $\boldsymbol{V}$ of Section 1D:

$$\boxed{1}_0 = \frac{\Psi_{1,\frac{3}{2}}}{\Psi_{1,\frac{1}{2}}}, \quad \boxed{2}_0 = \frac{\Psi_{1,-\frac{1}{2}}}{\Psi_{1,\frac{1}{2}}}\frac{\Psi_{2,1}}{\Psi_{2,0}}, \quad \boxed{3}_0 = \frac{\theta(z+\hbar)}{\theta(z)}\frac{\Psi_{2,-1}}{\Psi_{2,0}},$$

$$[\boldsymbol{V}] = \frac{[\mathcal{W}_{1,\frac{3}{2}}]}{[\mathcal{W}_{1,\frac{1}{2}}]} + \frac{[\mathcal{W}_{1,-\frac{1}{2}}]}{[\mathcal{W}_{1,\frac{1}{2}}]}\frac{[\mathcal{W}_{2,1}]}{[\mathcal{W}_{2,0}]} + \frac{[\mathcal{W}_{3,\frac{1}{2}}]}{[\mathcal{W}_{3,-\frac{1}{2}}]}\frac{[\mathcal{W}_{2,-1}]}{[\mathcal{W}_{2,0}]}.$$

Example 4.17. Let us construct the $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_2)$-module $\mathcal{W}_{1,\Lambda}$ from [Felder and Varchenko 1996b, Theorem 3]. As in [loc. cit.] set $\eta = -\frac{1}{2}\hbar$, $\lambda = \lambda_{12}$ and $(a, b, c, d) = (L_{11}, L_{12}, L_{21}, L_{22})$. For $\Lambda \in \mathbb{Z}_{>0}$, consider the evaluation module $L_\Lambda((\Lambda-1)\eta)$ with basis $(e_k)_{0 \le k \le \Lambda}$. Note that k indicates the basis vectors, while Λ the integer parameter of a KR module. Let us make a change of basis (the second product is empty if $k = 0$)

$$v_k := e_k \prod_{i=1}^{\Lambda} \frac{\theta(\lambda+(i-k)\hbar)}{\theta(\hbar)} \times \prod_{j=1}^{k} \frac{\theta(\lambda-j\hbar)}{\theta((\Lambda-k+j)\hbar)} \quad \text{for } 0 \le k \le \Lambda.$$

Tensoring $L_\Lambda((\Lambda-1)\eta)$ with the one-dimensional module of highest weight $\theta(w+\Lambda\hbar)/\theta(w)$, we obtain another irreducible module V_Λ with basis $(v_k = v_k \overline{\otimes} 1)_{0 \le k \le \Lambda}$; here to follow [loc. cit.] w denotes z. We have $\mathrm{wt}(v_k) = (\Lambda-k)\epsilon_1 + k\epsilon_2$ and

$$\begin{aligned}
a(w)v_k &= \frac{\theta(w+(\Lambda-k)\hbar)}{\theta(w)}\frac{\theta(\lambda+(\Lambda-k+1)\hbar)}{\theta(\lambda+(1-k)\hbar)}v_k,\\
b(w)v_k &= \frac{\theta(w+\lambda+(\Lambda-k-1)\hbar)}{\theta(w)}\frac{\theta((\Lambda-k)\hbar)\theta(\hbar)}{\theta(\lambda-\hbar)\theta(\lambda)}v_{k+1},\\
c(w)v_k &= -\frac{\theta(w-\lambda+(k-1)\hbar)}{\theta(w)}\frac{\theta(k\hbar)}{\theta(\hbar)}v_{k-1},\\
d(w)v_k &= \frac{\theta(w+k\hbar)}{\theta(w)}\frac{\theta(\lambda-(k+1)\hbar)\theta(\lambda-k\hbar)}{\theta(\lambda-\hbar)\theta(\lambda)}v_k.
\end{aligned}$$

We have $t_{12}v_k = -v_{k-1}\theta(k\hbar)/\theta(\hbar)$ and v_0 is of highest weight $\boldsymbol{w}^{(1)}_{\Lambda,0}$, so $V_\Lambda \cong W^{(1)}_{\Lambda,0}$. The bases (v_k) trivialize the inductive system (V_Λ) because the inductive maps commute with t_{12} by (4.20). For $\Lambda \in \mathbb{C}$, the above formulas define an $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_2)$-module structure on $\oplus_{k=0}^{\infty}\mathbb{M}v_k$, with $\mathrm{wt}(v_k) = (\Lambda - k)\epsilon_1 + k\epsilon_2$. This is the desired $\mathscr{W}_{1,\Lambda}$.

General formulas for the $\mathcal{E}_{\tau,\hbar}(\mathfrak{sl}_N)$-module $\mathscr{W}_{1,\Lambda}$ can be found in [Cavalli 2001, §3.4].

5. Baxter TQ relations

We derive three-term relations in the Grothendieck ring $K_0(\mathcal{O})$ for the asymptotic modules. For $1 \le r < N$ and $k, x, t \in \mathbb{C}$, by Corollary 4.13, $\boldsymbol{d}^{(r,t)}_{k,x} \in \mathcal{R}$ and is the highest weight of an irreducible module $D^{(r,t)}_{k,x}$ in category $\mathcal{O}$.

Call a complex number $c \in \mathbb{C}$ *generic* if $c \notin \frac{1}{2}\mathbb{Z} + \frac{1}{\hbar}(\mathbb{Z} + \mathbb{Z}\tau)$. This condition is equivalent to $\mathcal{Q}_a \cap \mathcal{Q}_{a+c} = \{1\}$ for all $a \in \mathbb{C}$.

Theorem 5.1. *Let $1 \le r < N$, $t \in \mathbb{Z}_{>0}$ and $k, a, b \in \mathbb{C}$ with k generic. Then*

$$\chi_q(D^{(r,t)}_{k,a}) = \boldsymbol{d}^{(r,t)}_{k,a}\left(1 + \sum_{l=1}^{t} A^{-1}_{r,a}A^{-1}_{r,a+1}\cdots A^{-1}_{r,a+l-1}\right)\prod_{s=r\pm1}\chi_q(\mathscr{W}^{(s)}_{0,a-k-\frac{1}{2}}) \tag{5.30}$$

and $D^{(r,0)}_{k,a} \cong \overline{\otimes}_{s=r\pm1}\mathscr{W}^{(s)}_{k,a-k-\frac{1}{2}}$.

Proof. Set $x := a - k - \frac{1}{2}$. Define $\boldsymbol{d} := \boldsymbol{d}^{(r,t)}_{k,a}$ and for $1 \le l \le t$,

$$\boldsymbol{m}_l := \boldsymbol{m}_0 A^{-1}_{r,a}A^{-1}_{r,a+1}\cdots A^{-1}_{r+a+l-1} \quad \text{and} \quad \boldsymbol{m}_0 := \boldsymbol{d}\prod_{j=1}^{t}\boldsymbol{w}^{(r)}_{k+j-\frac{1}{2},x}.$$

By (1.9) and (3.14)–(3.15), we have for $0 \le l \le t$

$$\boldsymbol{m}_l = \prod_{j=l+1}^{t}\frac{\Psi_{r,a+j}}{\Psi_{r,x}} \times \prod_{j=1}^{l}\frac{\Psi_{r,a+j-2}}{\Psi_{r,x}} \times \prod_{s=r\pm1}\frac{\Psi_{s,a+l-\frac{1}{2}}}{\Psi_{s,x}}.$$

Let us introduce the tensor products for $0 \le l \le t$,

$$\begin{aligned} S^l &:= (\overline{\otimes}^t_{j=l+1}\mathscr{W}^{(r)}_{k+j+\frac{1}{2},x})\overline{\otimes}(\overline{\otimes}^l_{j=1}\mathscr{W}^{(r)}_{k+j-\frac{3}{2},x})\overline{\otimes}(\overline{\otimes}_{s=r\pm1}\mathscr{W}^{(s)}_{k+l,x}),\\ T &:= D^{(r,t)}_{k,a}\overline{\otimes}(\overline{\otimes}^t_{j=1}\mathscr{W}^{(r)}_{k+j-\frac{1}{2},x}). \end{aligned}$$

Equation (5.30) is equivalent to $\chi_q(T) = \sum_{l=0}^{t}\chi_q(S^l)$ in view of (4.28).

Given two elements $\chi = \sum_{\boldsymbol{f}} c_{\boldsymbol{f}}\boldsymbol{f}$ and $\chi' := \sum_{\boldsymbol{f}} c'_{\boldsymbol{f}}\boldsymbol{f}$ of $\mathcal{M}_\mathrm{t}$, we say that χ is bounded above by χ' if $c_{\boldsymbol{f}} \le c'_{\boldsymbol{f}}$ for all $\boldsymbol{f} \in \mathcal{M}_\mathrm{w}$. When this is the case, χ' is bounded below by χ. If χ is bounded below and above by χ', then $\chi = \chi'$.

Claim 1. *The S^l are irreducible. In particular,* $D^{(r,0)}_{k,a} \cong \overline{\otimes}_{s=r\pm1}\mathscr{W}^{(s)}_{k,x}$.

Fix $0 \leq l \leq t$. Let $S' := S(\boldsymbol{m}_l)$. For $n \in \mathbb{Z}_{>0}$, set

$$S'_n := (W^{(r)}_{n,x})^{\overline{\otimes} t} \overline{\otimes} (\overline{\otimes}_{s=r\pm 1} W^{(s)}_{n,x}) \quad \text{and} \quad \boldsymbol{s}'_n := (\boldsymbol{w}^{(r)}_{n,x})^t \prod_{s=r\pm 1} \boldsymbol{w}^{(s)}_{n,x}.$$

By Lemma 3.2, any e-weigth $\boldsymbol{s}'_n \boldsymbol{e} \in \mathcal{P}_x$ of S'_n different from $\boldsymbol{s}'_n$ is right-negative. So S'_n is irreducible. Viewing S'_n as an irreducible subquotient of

$$S' \overline{\otimes} (\overline{\otimes}^t_{j=l+1} \mathscr{W}^{(r)}_{n-k-j-\frac{1}{2}, a+j}) \overline{\otimes} (\overline{\otimes}^l_{j=1} \mathscr{W}^{(r)}_{n-k-j+\frac{3}{2}, a+j-2}) \overline{\otimes} (\overline{\otimes}_{s=r\pm 1} \mathscr{W}^{(s)}_{n-k-l, a+l-\frac{1}{2}}),$$

we have $\boldsymbol{e} = \boldsymbol{e}' \prod_{j=1}^t \boldsymbol{e}_j \prod_{s=r\pm 1} \boldsymbol{e}^{(s)}$ where $\boldsymbol{m}_l \boldsymbol{e}'$, $\boldsymbol{w}^{(r)}_{n-k-j-\frac{1}{2}, a+j} \boldsymbol{e}_j$ for $l < j \leq t$, $\boldsymbol{w}^{(r)}_{n-k-j+\frac{3}{2}, a+j-2} \boldsymbol{e}_j$ for $1 \leq j \leq l$, and $\boldsymbol{w}^{(s)}_{n-k-l, a+l-\frac{1}{2}} \boldsymbol{e}^{(s)}$ are e-weigths of the corresponding tensor factors. By Lemma 3.2 and Proposition 4.10,

$$\boldsymbol{e}, \boldsymbol{e}' \in \mathcal{Q}_x^- \quad \text{and} \quad \boldsymbol{e}_j, \boldsymbol{e}^{(s)} \in \mathcal{Q}_a^-.$$

Since $a - x = k + \frac{1}{2}$ is generic, $\mathcal{Q}_a^- \cap \mathcal{Q}_x^- = \{1\}$ and so $\boldsymbol{e} = \boldsymbol{e}'$. The normalized q-character of S' is bounded below by that of S'_n for all $n \in \mathbb{Z}_{>0}$. On the other hand, viewing S' as an irreducible subquotient of S^l and applying (4.25) to S^l, we see that the normalized q-character of S' is bounded above by the limit of that of S'_n as $n \to \infty$. Therefore $S^l \cong S'$ is irreducible.

Claim 2. *For $1 \leq l \leq t$, we have $\boldsymbol{d} A^{-1}_{r,a} A^{-1}_{r,a+1} \cdots A^{-1}_{r,a+l-1} \in \mathrm{wt}_e(D^{(r,t)}_{k,a})$. It follows that $\boldsymbol{m}_l \in \mathrm{wt}_e(T)$.*

Let us view the KR module $W^{(r)}_{t,a}$ as an irreducible subquotient of

$$D^{(r,t)}_{k,a} \overline{\otimes} (\overline{\otimes}_{s=r\pm 1} \mathscr{W}^{(s)}_{-k, a-\frac{1}{2}}).$$

By Lemma 3.2, $\boldsymbol{w}^{(r)}_{t,a} A^{-1}_{r,a} A^{-1}_{r,a+1} \cdots A^{-1}_{r,a+l-1} \in \mathrm{wt}_e(W^{(r)}_{t,a})$. The $A^{-1}_{r,a+j}$ must arise from $\mathrm{wt}_e(D^{(r,t)}_{k,a})$ instead of any of the $\mathrm{wt}_e(W^{(s)}_{-k,a-\frac{1}{2}})$ with $s \neq r$.

For $0 \leq j, l \leq t$, since $\mathrm{wt}_e(S^l) \subset \boldsymbol{m}_l \mathcal{Q}_x^-$ and $\boldsymbol{m}_j \in \boldsymbol{m}_l \mathcal{Q}_a$, we have $\boldsymbol{m}_j \in \mathrm{wt}_e(S^l)$ if and only if $l = j$. Therefore, all the S^l appear as irreducible subquotients of T, and they are mutually nonisomorphic. So $\chi_q(T)$ is bounded below by $\sum_{l=0}^t \chi_q(S^l)$.

Claim 3. *$\chi_q(D^{(r,t)}_{k,a})$ is bounded above by*

$$\boldsymbol{d}\left(1 + \sum_{l=1}^t A^{-1}_{r,a} A^{-1}_{r,a+1} \cdots A^{-1}_{r+a+l-1}\right) \prod_{s=r\pm 1} \chi_q(\mathscr{W}^{(s)}_{0,x}).$$

Fix $\boldsymbol{d}\boldsymbol{f} \in \mathrm{wt}_e(D^{(r,t)}_{k,a})$. For $n \in \mathbb{Z}_{>0}$, viewing $D^{(r,t)}_{k,a}$ as a subquotient of

$$D^{(r,t)}_{n,a} \overline{\otimes} (\overline{\otimes}_{s=r\pm 1} \mathscr{W}^{(s)}_{k-n,x})$$

gives $\boldsymbol{f} = \boldsymbol{f}_n \prod_{s=r\pm 1} \boldsymbol{f}^{(s)}$ where by Lemma 3.2 and Corollary 3.5

$$\boldsymbol{f}_n \boldsymbol{d}^{(r,t)}_{n,a} \in \mathrm{wt}_e(D^{(r,t)}_{n,a}) \quad \text{and} \quad \boldsymbol{f}^{(s)} \boldsymbol{w}^{(s)}_{k-n,x} \in \mathrm{wt}_e(\mathscr{W}^{(s)}_{k-n,x}) = \boldsymbol{w}^{(s)}_{k-n,x} \, \mathrm{wt}_e(\mathscr{W}^{(s)}_{0,x}).$$

It follows that $\boldsymbol{f}_n \in \mathcal{Q}_a^-$, $\boldsymbol{f}^s \in \mathcal{Q}_x^-$ and $\boldsymbol{f} \in \mathcal{Q}_a^- \mathcal{Q}_x^-$.

Let $n \in \mathbb{Z}_{>0}$ be large enough so that $\boldsymbol{f} \in \mathcal{Q}_{a;n}^{-}\mathcal{Q}_{x}^{-}$ where $\mathcal{Q}_{a;n}^{-}$ is the submonoid of $\mathcal{Q}_{a}^{-}$ generated by the $A_{i,a+m}^{-1}$ for $1 \le i < N$ and $m \in \frac{1}{2}\mathbb{Z}$ with $m > -n$. Since $a - x = k + \frac{1}{2}$ is generic, Corollary 3.5 implies that

$$\boldsymbol{f}_n \in \{1,\, A_{r,a}^{-1},\, A_{r,a}^{-1}A_{r,a+1}^{-1},\, \dots,\, A_{r,a}^{-1}A_{r,a+1}^{-1}\cdots A_{r+a+t-1}^{-1}\}$$

is uniquely determined by $\boldsymbol{f}$. The coefficient of $\boldsymbol{df}$ in $\chi_q(M_{k,a}^{(r)})$ is bounded above by that of $\prod_{s=r\pm 1}\boldsymbol{f}^{(s)}$ in $\prod_{s=r\pm 1}\chi_q(\mathcal{W}_{0,x}^{(s)})$. This proves the claim.

It follows from Claim 3 that $\chi_q(T)$ is bounded above by

$$\boldsymbol{m}_0\Bigg(1 + \sum_{l=1}^{t} A_{r,a}^{-1}A_{r,a+1}^{-1}\cdots A_{r+a+l-1}^{-1}\Bigg)\prod_{s=r\pm 1}\chi_q(\mathcal{W}_{0,x}^{(s)}) \times \prod_{j=1}^{t}\chi_q(\mathcal{W}_{0,x}^{(r)}) = \sum_{l=0}^{t}\chi_q(S^l).$$

Since "bounded below" also holds, we obtain the exact formula for $\chi_q(T)$, which implies (5.30). This completes the proof of the theorem. □

Claim 1 is in the spirit of [Frenkel and Hernandez 2015, Theorem 4.11], and Claim 3 [Hernandez and Leclerc 2016, (6.14)], [Frenkel and Hernandez 2016, §4.3] and [Zhang 2018, Theorem 3.3], the main difference being the nonexistence of prefundamental modules. If both k and t are generic, then $\chi_q(D_{k,a}^{(r,t)})$ is obtained from the right-hand side of (5.30) by replacing $\sum_{l=1}^{t}$ therein with $\sum_{l=1}^{\infty}$.

Corollary 5.2. *Let $k \in \mathbb{C}$ be generic and $1 \le r < N$. In $K_0(\mathcal{O})$ holds*

$$[D_{k,k+\frac{1}{2}}^{(r,1)}][\mathcal{W}_{r,k+\frac{1}{2}}] = [\mathcal{W}_{r,k-\frac{1}{2}}]\prod_{s=r\pm 1}[\mathcal{W}_{s,k+1}] + [\mathcal{W}_{r,k+\frac{3}{2}}]\prod_{s=r\pm 1}[\mathcal{W}_{s,k}]. \tag{5.31}$$

Proof. From (5.30) and the injectivity of the q-character map we obtain

$$[D_{k,a}^{(r,t)}][\mathcal{W}_{a-b+t-1,b}^{(r)}] = [D_{k+t,a+t}^{(r,0)}][\mathcal{W}_{a-b-1,b}^{(r)}] + [D_{k,a}^{(r,t-1)}][\mathcal{W}_{a-b+t,b}^{(r)}] \tag{5.32}$$

for $a, b \in \mathbb{C}$ and $t \in \mathbb{Z}_{>0}$. (5.31) is the special case $(t, a, b) = \big(1, k + \frac{1}{2}, 0\big)$ of this identity in view of the tensor product decomposition of $D_{k,a}^{(r,0)}$ in Theorem 5.1. □

Equation (5.32) can be viewed as a generic version of (3.17).

6. Transfer matrices and Baxter operators

We have obtained three types of identities (4.28), (4.29), and (5.31) in the Grothendieck ring $K_0(\mathcal{O})$. These are viewed as universal functional relations [Bazhanov et al. 1997; 1999; Bazhanov and Tsuboi 2008] in the sense that when specialized to quantum integrable systems they imply functional relations of transfer matrices. In this section, we study one such example, with the quantum space being a tensor product of vector representations [Hou et al. 2003].

Fix $\ell := N\kappa$ with $\kappa \in \mathbb{Z}_{>0}$ and $a_1, a_2, \dots, a_\ell \in \mathbb{C} \setminus \Gamma$. Set $I := \{1, 2, \dots, N\}$. Let I_0^ℓ be the subset of I^ℓ formed of $\underline{i}$ such that $\epsilon_{i_1} + \epsilon_{i_2} + \cdots + \epsilon_{i_\ell} = 0 \in \mathfrak{h}$. Upon identification $\underline{i} := v_{i_1}\overline{\otimes}v_{i_2}\overline{\otimes}\cdots\overline{\otimes}v_{i_\ell}$, the weight space $\boldsymbol{V}^{\overline{\otimes}\ell}[0]$ has basis I_0^ℓ.

Let $\mathbb{D}_p$ be the set of formal sums $\sum_{\alpha\in\mathfrak{h}} p^\alpha T_\alpha f_\alpha(z;\lambda)$ such that: The $f_\alpha(z;\lambda)$ are meromorphic functions of $(z,\lambda)\in\mathbb{C}\times\mathfrak{h}$. The set $\{\alpha : f_\alpha\neq 0\}$ is contained in a finite union of cones $\nu+\boldsymbol{Q}_-$ with $\nu\in\mathfrak{h}$. Make $\mathbb{D}_p$ into a ring: Addition is the usual one of formal sums. Multiplication is induced from

$$p^\alpha T_\alpha f(z;\lambda)\times p^\beta T_\beta g(z;\lambda) = p^{\alpha+\beta} T_{\alpha+\beta} f(z;\lambda+\hbar\beta) g(z;\lambda). \tag{6.33}$$

As in [Felder and Varchenko 1996a; Felder and Zhang 2017], we construct a ring morphism $[X]\mapsto t_X(z)$ from $K_0(\mathcal{O})$ to the ring $\mathrm{M}(I_0^\ell;\mathbb{D}_p)$ of $I_0^\ell\times I_0^\ell$ matrices with coefficients in $\mathbb{D}_p$. (We think of $\mathrm{M}(I_0^\ell;\mathbb{D}_p)$ as a ring of formal difference operators on $\boldsymbol{V}^{\tilde{\otimes}\ell}[0]$.)

Let X be an object of category $\mathcal{O}$. To $\underline{i},\underline{j}\in I_0^\ell$ we associate

$$L^X_{\underline{i}\,\underline{j}}(z) := L^X_{i_1 j_1}(z+a_1) L^X_{i_2 j_2}(z+a_2)\cdots L^X_{i_\ell j_\ell}(z+a_\ell)\in (D_X)_{0,0}.$$

Since $(D_X)_{0,0}\subseteq \mathrm{End}_{\mathbb{M}}(X)$, one can take trace of $L^X_{\underline{i}\,\underline{j}}(z)$ over weight spaces of X.

Definition 6.1. The *transfer matrix* associated to an object X in category $\mathcal{O}$ is the matrix $t_X(z)\in\mathrm{M}(I_0^\ell;\mathbb{D}_p)$ whose $(\underline{i},\underline{j})$-th entry for $\underline{i},\underline{j}\in I_0^\ell$ is

$$\sum_{\alpha\in\mathrm{wt}(X)} p^\alpha T_\alpha\times \mathrm{Tr}_{X[\alpha]}(L^X_{\underline{i}\,\underline{j}}(z)|_{X[\alpha]})\in\mathbb{D}_p.$$

Almost all of the results and comments in [Felder and Zhang 2017, §5] hold true after slight modification in our present situation. In the following, we focus on the modification of these results, referring to [Felder and Zhang 2017] for their proofs.

We remark that $t_V(z)|_{p=1}$ can be identified with the transfer matrix $T(z)$ in [Hou et al. 2003, (2.22)] where the $E_{\tau,\eta}(sl_n)$-module W is $V_{\Lambda_1}(a_1)\otimes V_{\Lambda_1}(a_2)\otimes\cdots\otimes V_{\Lambda_1}(a_\ell)$.

The transfer matrix associated to the one-dimensional module of highest weight $g(z)\in\mathbb{M}_{\mathbb{C}}^\times$ is the scalar matrix $\prod_{i=1}^\ell g(z+a_i)$.

For $1\le r\le N$ and $x\in\mathbb{C}$, consider the $\mathcal{E}$-module $\mathscr{W}'_{r,x} := \mathscr{W}_{r,x}\tilde{\otimes} S(\theta(z-\ell_r\hbar))$ in category $\mathcal{O}$. By Lemma 4.8, the matrix entries of the difference operators $L_{ij}(z)$ for $1\le i,j\le N$, with respect to any basis of $\mathscr{W}'_{r,x}$, are entire functions of $z\in\mathbb{C}$.

Definition 6.2. The r-th *Baxter Q-operator* for $1\le r\le N$ is defined to be

$$Q_r(u) := t_{\mathscr{W}'_{r,u\hbar^{-1}}}(z)|_{z=0}\quad\text{for } u\in\mathbb{C}. \tag{6.34}$$

Since $\mathscr{W}'_{N,x} = S\big(\theta\big(z+\big(x+\tfrac12\big)\hbar\big)\big)$ is one-dimensional, $Q_N(z)=\prod_{i=1}^\ell\theta\big(z+a_i+\tfrac12\hbar\big)$.

Let $1\le r<N$. Then $Q_r(z)=p^{z\hbar^{-1}\varpi_r}T_{z\hbar^{-1}\varpi_r}\tilde{Q}(z)$ and $\tilde{Q}(z)$ is a power series in the $p^{-\alpha_i}T_{-\alpha_i}$ for $1\le i<N$. The leading term $\tilde{Q}_0(z)$ of $\tilde{Q}(z)$ is invertible. Indeed $\tilde{Q}_0(0)$ is the scalar matrix $\prod_{j=1}^\ell\theta(a_j)\in\mathrm{M}(I_\ell;\mathbb{C})$, which is invertible because $\theta(a_j)\neq 0$ by assumption. (One can prove furthermore that with respect to certain order on I_0^ℓ, the matrix $\tilde{Q}_0(z)$ is upper triangular, whose entries are meromorphic functions of $(z;\lambda)\in\mathbb{C}\times\mathfrak{h}$ and entire on z.) Therefore $Q_r(z)\in\mathrm{GL}(I_0^\ell;\mathbb{D}_p)$.

Similarly one can show that $t_{\mathscr{W}'_{r,x}}(z)$ is invertible for $x\in\mathbb{C}$.

Proposition 6.3. *Let X and Y be in category $\mathcal{O}$ and let $x, u \in \mathbb{C}$.*

(i) $t_{\Psi_u^* X}(z) = t_X(z + u\hbar)$.

(ii) $t_X(z)t_Y(z) = t_{X\overline{\otimes}Y}(z)$.

(iii) $t_{\mathcal{W}_{r,x}}(z)t_{\mathcal{W}_{r,0}}(z + u\hbar) = t_{\mathcal{W}_{r,x-u}}(z + u\hbar)t_{\mathcal{W}_{r,u}}(z)$.

(iv) $t_X(z)t_Y(w) = t_Y(w)t_X(z)$.

In (iv), we replace one of the z in (6.33) with w to define the multiplication. It is proved as in [Frenkel and Hernandez 2015, Theorem 5.3]; the commutativity of transfer matrices is a consequence of the commutativity of the Grothendieck ring $K_0(\mathcal{O})$. The standard proof by using the Yang–Baxter equation [Baxter 1972] would require braiding in category $\mathcal{O}$, whose existence is not clear.

Conclusion (ii) and the fact that $t_X(z)$ only depends on the isomorphism class $[X]$ of X imply that $[X] \mapsto t_X(z)$ is a ring homomorphism $\mathrm{tr}_p : K_0(\mathcal{O}) \to \mathrm{M}(I_0^\ell; \mathbb{D}_p)$. Applying tr_p to (4.28) we obtain (iii). Replace $(\mathcal{W}, x, u, z)$ with $(\mathcal{W}', z\hbar^{-1} + x, z\hbar^{-1}, 0)$ in (iii) and take the inverse of $Q_r(z)$ and $t_{\mathcal{W}'_{r,0}}(z)$. We have

$$\frac{t_{\mathcal{W}_{r,x}}(z)}{t_{\mathcal{W}_{r,0}}(z)} = \frac{t_{\mathcal{W}'_{r,x}}(z)}{t_{\mathcal{W}'_{r,0}}(z)} = \frac{Q_r(z + x\hbar)}{Q_r(z)}, \tag{6.35}$$

as in [Felder and Zhang 2017, Theorem 5.6(i)]. Now applying tr_p to (4.29), we obtain

Corollary 6.4. *Let V be a finite-dimensional $\mathcal{E}$-module in category $\mathcal{O}$. Then in* (4.29) *replacing V, $S(\boldsymbol{d}_j)$ and the $[\mathcal{W}_{r,a}]/[\mathcal{W}_{r,b}]$ with $t_V(z)$, $t_{S(\boldsymbol{d}_j)}(z)$ and $Q_r(z + a\hbar)/Q_r(z + b\hbar)$ respectively, we obtain an identity in* $\mathrm{M}(I_0^\ell; \mathbb{D}_p)$.

This forms the generalized Baxter relations for transfer matrices. If the prefundamental modules $L_{r,a}^+$ before Example 4.16 existed, then we would have defined alternatively the r-th Baxter operator $Q_r^{\mathrm{FH}}(z) = t_{L_{r,0}^+}(z)$ as a *real* transfer matrix [Frenkel and Hernandez 2015, §5.5] and so $Q_r(z + a\hbar)/Q_r(z + b\hbar) = Q_r^{\mathrm{FH}}(z + a\hbar)/Q_r^{\mathrm{FH}}(z + b\hbar)$ based on $[\mathcal{W}_{r,a}]/[\mathcal{W}_{r,b}] = [L_{r,a}^+]/[L_{r,b}^+]$.

As an illustration of the corollary, let us be in the situation of Example 4.16:

$$t_V(z) = \frac{Q_1\big(z + \frac{3}{2}\hbar\big)}{Q_1\big(z + \frac{1}{2}\hbar\big)} + \frac{Q_1\big(z - \frac{1}{2}\hbar\big)}{Q_1\big(z + \frac{1}{2}\hbar\big)} \frac{Q_2(z + \hbar)}{Q_2(z)} + \frac{Q_2(z - \hbar)}{Q_2(z)} \prod_{j=1}^{\ell} \frac{\theta(z + a_j + \hbar)}{\theta(z + a_j)}.$$

Apply tr_p to (5.31), divide both sides by the second term, and then perform the change of variable $z + \big(k + \frac{1}{2}\big)\hbar \mapsto w$. By (6.35) and Proposition 6.3(i)

$$X_k^{(r)}(w)\frac{Q_r(w)}{Q_r(w - \hbar)} = 1 + \frac{Q_r(w + \hbar)}{Q_r(w - \hbar)} \prod_{s=r\pm 1} \frac{Q_s\big(w - \frac{1}{2}\hbar\big)}{Q_s\big(w + \frac{1}{2}\hbar\big)}. \tag{6.36}$$

This forms three-term Baxter TQ relations for transfer matrices, where

$$X_k^{(r)}(w) = t_{D_{k,0}^{(r,1)}}(w) \prod_{s=r\pm 1} t_{\mathcal{W}_{s,k+1}}^{-1}\big(w - \big(k + \tfrac{1}{2}\big)\hbar\big).$$

By (6.36), $X_k^{(r)}(z) \in \mathrm{M}(I_0^\ell; \mathbb{D}_p)$ is independent of the choice of generic $k \in \mathbb{C}$.

In the homogeneous case $a_1 = a_2 = \cdots = a_\ell = a$, the entries of the matrix $\tilde{Q}_r(z)$, as entire functions of z, in general do not satisfy the uniform double periodicity of [Felder and Zhang 2017, Theorem 5.6(ii)]. By "uniform" we mean the multipliers with respect to $z+1$ and $z+\tau$ only depend on (a, z, ℓ). This is because the transfer matrix construction in [Felder and Zhang 2017] is based on a slightly different elliptic quantum group; see footnote 9.

We follow [Frenkel and Hernandez 2016, §5] to derive the Bethe Ansatz equations from (6.36). Let u be a zero of $Q_r(z)$. Suppose $X_k^{(r)}(z)$, $Q_r(z-\hbar)$ and $Q_s\big(z+\frac{1}{2}\hbar\big)$ for $s \neq r \pm 1$ have no poles at $z = u$. (This is a genericity condition.) Then as in [Frenkel and Hernandez 2016, (5.16)]

$$\frac{Q_r(u+\hbar)}{Q_r(u-\hbar)} \prod_{s=r\pm 1} \frac{Q_s\big(u-\frac{1}{2}\hbar\big)}{Q_s\big(u+\frac{1}{2}\hbar\big)} = -1. \tag{6.37}$$

To compare with [Frenkel and Hernandez 2016], we can assume furthermore that eigenvalues of $Q_r(z)$ are of the form $p^{z\hbar^{-1}\varpi_r} \prod_{i=1}^{d_r} \theta(z-u_{r;i})$ based on [Felder and Zhang 2017, Remark 5.8]. Then

$$p^{\alpha_r} \prod_{i=1}^{d_r} \frac{\theta(u_{r;k}+\hbar-u_{r;i})}{\theta(u_{r;k}-\hbar-u_{r;i})} \prod_{s=r\pm 1} \prod_{j=1}^{d_s} \frac{\theta\big(u_{r;k}+\frac{1}{2}\hbar-u_{s;j}\big)}{\theta\big(u_{r;k}+\frac{1}{2}\hbar-u_{s;j}\big)} = -1 \quad \text{for } 1 \le k \le d_r.$$

We remark that similar Bethe Ansatz equations for $\mathcal{E}$ appeared in [Hou et al. 2003, (3.45)].

For affine quantum groups and toroidal $\mathfrak{gl}_1$, the genericity condition of Bethe Ansatz equations has been dropped in [Feigin et al. 2017a; 2017b].

Acknowledgments

The author thanks Giovanni Felder, David Hernandez, Bernard Leclerc, Marc Rosso and Vitaly Tarasov for fruitful discussions, and the anonymous referees for their valuable comments and suggestions. This work was supported by the National Center of Competence in Research SwissMAP — The Mathematics of Physics of the Swiss National Science Foundation, during the author's postdoctoral stay at ETH Zürich.

References

[Aganagic and Okounkov 2016] M. Aganagic and A. Okounkov, "Elliptic stable envelope", preprint, 2016. arXiv

[Baxter 1972] R. J. Baxter, "Partition function of the eight-vertex lattice model", *Ann. Physics* **70** (1972), 193–228. MR Zbl

[Bazhanov and Tsuboi 2008] V. V. Bazhanov and Z. Tsuboi, "Baxter's Q-operators for supersymmetric spin chains", *Nuclear Phys. B* **805**:3 (2008), 451–516. MR Zbl

[Bazhanov et al. 1997] V. V. Bazhanov, S. L. Lukyanov, and A. B. Zamolodchikov, "Integrable structure of conformal field theory, II: Q-operator and DDV equation", *Comm. Math. Phys.* **190**:2 (1997), 247–278. MR Zbl

[Bazhanov et al. 1999] V. V. Bazhanov, S. L. Lukyanov, and A. B. Zamolodchikov, "Integrable structure of conformal field theory, III: The Yang–Baxter relation", *Comm. Math. Phys.* **200**:2 (1999), 297–324. MR Zbl

[Bittmann 2017] L. Bittmann, "Asymptotics of standard modules of quantum affine algebras", preprint, 2017. arXiv

[Cavalli 2001] A. M. Cavalli, *On representations of the elliptic quantum group $E_{\gamma,\tau}(\mathfrak{gl}_N)$*, Ph.D. thesis, Swiss Federal Institute of Technology, 2001.

[Chari 2002] V. Chari, "Braid group actions and tensor products", *Int. Math. Res. Not.* **2002**:7 (2002), 357–382. MR Zbl

[Damiani 1998] I. Damiani, "La R-matrice pour les algèbres quantiques de type affine non tordu", *Ann. Sci. École Norm. Sup.* (*4*) **31**:4 (1998), 493–523. MR Zbl

[Ding and Frenkel 1993] J. T. Ding and I. B. Frenkel, "Isomorphism of two realizations of quantum affine algebra $U_q(\widehat{\mathfrak{gl}(n)})$", *Comm. Math. Phys.* **156**:2 (1993), 277–300. MR Zbl

[Enriquez and Felder 1998] B. Enriquez and G. Felder, "Elliptic quantum groups $E_{\tau,\eta}(\mathfrak{sl}_2)$ and quasi-Hopf algebras", *Comm. Math. Phys.* **195**:3 (1998), 651–689. MR Zbl

[Etingof and Moura 2002] P. I. Etingof and A. A. Moura, "On the quantum Kazhdan–Lusztig functor", *Math. Res. Lett.* **9**:4 (2002), 449–463. MR Zbl

[Etingof and Schiffmann 1998] P. Etingof and O. Schiffmann, "A link between two elliptic quantum groups", *Asian J. Math.* **2**:2 (1998), 345–354. MR Zbl

[Etingof and Varchenko 1998] P. Etingof and A. Varchenko, "Solutions of the quantum dynamical Yang–Baxter equation and dynamical quantum groups", *Comm. Math. Phys.* **196**:3 (1998), 591–640. MR Zbl

[Feigin et al. 2017a] B. Feigin, M. Jimbo, T. Miwa, and E. Mukhin, "Finite type modules and Bethe ansatz equations", *Ann. Henri Poincaré* **18**:8 (2017), 2543–2579. MR Zbl

[Feigin et al. 2017b] B. Feigin, M. Jimbo, T. Miwa, and E. Mukhin, "Finite type modules and Bethe ansatz for quantum toroidal $\mathfrak{gl}_1$", *Comm. Math. Phys.* **356**:1 (2017), 285–327. MR Zbl

[Felder 1995] G. Felder, "Elliptic quantum groups", pp. 211–218 in *XIth International Congress of Mathematical Physics* (Paris, 1994), edited by D. Iagolnitzer, Int. Press, Cambridge, 1995. MR Zbl arXiv

[Felder and Varchenko 1996a] G. Felder and A. Varchenko, "Algebraic Bethe ansatz for the elliptic quantum group $E_{\tau,\eta}(\mathfrak{sl}_2)$", *Nuclear Phys. B* **480**:1-2 (1996), 485–503. MR Zbl

[Felder and Varchenko 1996b] G. Felder and A. Varchenko, "On representations of the elliptic quantum group $E_{\tau,\eta}(\mathfrak{sl}_2)$", *Comm. Math. Phys.* **181**:3 (1996), 741–761. MR Zbl

[Felder and Zhang 2017] G. Felder and H. Zhang, "Baxter operators and asymptotic representations", *Selecta Math.* (*N.S.*) **23**:4 (2017), 2947–2975. MR Zbl

[Felder et al. 2017] G. Felder, R. Rimányi, and A. Varchenko, "Elliptic dynamical quantum groups and equivariant elliptic cohomology", preprint, 2017. arXiv

[Fourier and Hernandez 2014] G. Fourier and D. Hernandez, "Schur positivity and Kirillov–Reshetikhin modules", *SIGMA* **10** (2014), art. id. 058. MR Zbl

[Frenkel and Hernandez 2015] E. Frenkel and D. Hernandez, "Baxter's relations and spectra of quantum integrable models", *Duke Math. J.* **164**:12 (2015), 2407–2460. MR Zbl

[Frenkel and Hernandez 2016] E. Frenkel and D. Hernandez, "Spectra of quantum KdV Hamiltonians, Langlands duality, and affine opers", 2016. To appear in *Comm. Math. Phys.* arXiv

[Frenkel and Mukhin 2001] E. Frenkel and E. Mukhin, "Combinatorics of q-characters of finite-dimensional representations of quantum affine algebras", *Comm. Math. Phys.* **216**:1 (2001), 23–57. MR Zbl

[Frenkel and Mukhin 2002] E. Frenkel and E. Mukhin, "The Hopf algebra Rep $U_q\widehat{\mathfrak{gl}}_\infty$", *Selecta Math.* (*N.S.*) **8**:4 (2002), 537–635. MR Zbl

[Frenkel and Reshetikhin 1992] I. B. Frenkel and N. Y. Reshetikhin, "Quantum affine algebras and holonomic difference equations", *Comm. Math. Phys.* **146**:1 (1992), 1–60. MR Zbl

[Frenkel and Reshetikhin 1999] E. Frenkel and N. Reshetikhin, "The q-characters of representations of quantum affine algebras and deformations of $\mathcal{W}$-algebras", pp. 163–205 in *Recent developments in quantum affine algebras and related topics* (Raleigh, NC, 1998), edited by N. Jing and K. C. Misra, Contemp. Math. **248**, Amer. Math. Soc., Providence, RI, 1999. MR Zbl

[Galleas and Stokman 2015] W. Galleas and J. V. Stokman, "On connection matrices of quantum Knizhnik–Zamolodchikov equations based on Lie super algebras", 2015. To appear in *Adv. Stud. Pure Math.* arXiv

[Gautam and Toledano Laredo 2016] S. Gautam and V. Toledano Laredo, "Yangians, quantum loop algebras, and abelian difference equations", *J. Amer. Math. Soc.* **29**:3 (2016), 775–824. MR Zbl

[Gautam and Toledano Laredo 2017a] S. Gautam and V. Toledano Laredo, "Elliptic quantum groups and their finite-dimensional representations", preprint, 2017. arXiv

[Gautam and Toledano Laredo 2017b] S. Gautam and V. Toledano Laredo, “Meromorphic tensor equivalence for Yangians and quantum loop algebras”, *Publ. Math. Inst. Hautes Études Sci.* **125** (2017), 267–337. MR Zbl

[Guay et al. 2017] N. Guay, V. Regelskis, and C. Wendlandt, “Equivalences between three presentations of orthogonal and symplectic Yangians”, preprint, 2017. arXiv

[Hernandez 2005] D. Hernandez, “Representations of quantum affinizations and fusion product”, *Transform. Groups* **10**:2 (2005), 163–200. MR Zbl

[Hernandez 2006] D. Hernandez, “The Kirillov–Reshetikhin conjecture and solutions of T-systems”, *J. Reine Angew. Math.* **596** (2006), 63–87. MR Zbl

[Hernandez and Jimbo 2012] D. Hernandez and M. Jimbo, “Asymptotic representations and Drinfeld rational fractions”, *Compos. Math.* **148**:5 (2012), 1593–1623. MR Zbl

[Hernandez and Leclerc 2016] D. Hernandez and B. Leclerc, “Cluster algebras and category $\mathcal{O}$ for representations of Borel subalgebras of quantum affine algebras”, *Algebra Number Theory* **10**:9 (2016), 2015–2052. MR Zbl

[Hou et al. 2003] B. Y. Hou, R. Sasaki, and W.-L. Yang, “Algebraic Bethe ansatz for the elliptic quantum group $E_{\tau,\eta}(\mathfrak{sl}_n)$ and its applications”, *Nuclear Phys. B* **663**:3 (2003), 467–486. MR Zbl

[Jimbo et al. 1999] M. Jimbo, H. Konno, S. Odake, and J. Shiraishi, “Quasi-Hopf twistors for elliptic quantum groups”, *Transform. Groups* **4**:4 (1999), 303–327. MR Zbl

[Jing et al. 2017] N. Jing, M. Liu, and A. Molev, “Isomorphism between the R-matrix and Drinfeld presentations of Yangian in types B, C and D”, preprint, 2017. arXiv

[Kac 1990] V. G. Kac, *Infinite-dimensional Lie algebras*, 3rd ed., Cambridge Univ. Press, 1990. MR Zbl

[Knight 1995] H. Knight, “Spectra of tensor products of finite-dimensional representations of Yangians”, *J. Algebra* **174**:1 (1995), 187–196. MR Zbl

[Konno 2006] H. Konno, “Dynamical R matrices of elliptic quantum groups and connection matrices for the q-KZ equations”, *SIGMA* **2** (2006), art. id. 091. MR Zbl

[Konno 2009] H. Konno, “Elliptic quantum group $U_{q,p}(\widehat{\mathfrak{sl}}_2)$, Hopf algebroid structure and elliptic hypergeometric series”, *J. Geom. Phys.* **59**:11 (2009), 1485–1511. MR Zbl

[Konno 2016] H. Konno, “Elliptic quantum groups $U_{q,p}(\widehat{\mathfrak{gl}}_N)$ and $E_{q,p}(\widehat{\mathfrak{gl}}_N)$”, 2016. To appear in *Adv. Stud. Pure Math.* arXiv

[Konno 2017] H. Konno, “Elliptic weight functions and elliptic q-KZ equation”, *J. Integrable Syst.* **2**:1 (2017), art. id. xyx011.

[Moura and Pereira 2017] A. Moura and F. Pereira, “On tensor products of a minimal affinization with an extreme Kirillov–Reshetikhin module for type A”, *Transform. Groups* (online publication November 2017).

[Nakajima 2003] H. Nakajima, “t-analogs of q-characters of Kirillov–Reshetikhin modules of quantum affine algebras”, *Represent. Theory* **7** (2003), 259–274. MR Zbl

[Nekrasov et al. 2018] N. Nekrasov, V. Pestun, and S. Shatashvili, “Quantum geometry and quiver gauge theories”, *Comm. Math. Phys.* **357**:2 (2018), 519–567. MR Zbl

[Tarasov and Varchenko 1997] V. Tarasov and A. Varchenko, *Geometry of q-hypergeometric functions, quantum affine algebras and elliptic quantum groups*, Astérisque **246**, Soc. Math. France, Paris, 1997. MR Zbl

[Tarasov and Varchenko 2001] V. Tarasov and A. Varchenko, “Small elliptic quantum group $e_{\tau,\gamma}(\mathfrak{sl}_N)$”, *Mosc. Math. J.* **1**:2 (2001), 243–286. MR Zbl

[Yang and Zhao 2017] Y. Yang and G. Zhao, “Quiver varieties and elliptic quantum groups”, preprint, 2017. arXiv

[Zhang 2016] H. Zhang, “RTT realization of quantum affine superalgebras and tensor products”, *Int. Math. Res. Not.* **2016**:4 (2016), 1126–1157. MR Zbl

[Zhang 2017] H. Zhang, “Asymptotic representations of quantum affine superalgebras”, *SIGMA* **13** (2017), art. id. 066. MR Zbl

[Zhang 2018] H. Zhang, “Length-two representations of quantum affine superalgebras and Baxter operators”, *Comm. Math. Phys.* **358**:2 (2018), 815–862. MR

Communicated by Edward Frenkel
Received 2017-06-29 Revised 2017-12-17 Accepted 2018-03-10

huafeng.zhang@math.univ-lille1.fr *Laboratoire Paul Painlevé, Université de Lille, Villeneuve d'Ascq, France*

mathematical sciences publishers msp

ALGEBRA AND NUMBER THEORY 12:3 (2018)
dx.doi.org/10.2140/ant.2018.12.649

Differential forms in positive characteristic, II: cdh-descent via functorial Riemann–Zariski spaces

Annette Huber and Shane Kelly

This paper continues our study of the sheaf associated to Kähler differentials in the cdh-topology and its cousins, in positive characteristic, without assuming resolution of singularities. The picture for the sheaves themselves is now fairly complete. We give a calculation $\mathcal{O}_{\mathrm{cdh}}(X) \cong \mathcal{O}(X^{\mathrm{sn}})$ in terms of the seminormalisation. We observe that the category of representable cdh-sheaves is equivalent to the category of seminormal varieties. We conclude by proposing some possible connections to Berkovich spaces and F-singularities in the last section. The tools developed for the case of differential forms also apply in other contexts and should be of independent interest.

1. Introduction

This paper continues the programme started in [Huber and Jörder 2014] (characteristic 0) and [Huber, Kebekus and Kelly 2017] (positive characteristic). For a survey see also [Huber 2016].

Programme. Let us quickly summarise the main idea. Sheaves of differential forms are very rich sources of invariants in the study of algebraic invariants of smooth algebraic varieties. However, they are much less well-behaved for singular varieties. In characteristic 0, the use of the h-topology — replacing Kähler differentials with their sheafification in this Grothendieck topology — is very successful. It unifies several ad hoc notions and simplifies arguments. In positive characteristic, resolution of singularities would imply that the cdh-sheafification could be used in a very similar way. Together with the results of [Huber, Kebekus and Kelly 2017], we now have a fairly complete unconditional picture, at least for the sheaves themselves. We refer to the follow-up [Huber and Kelly $\geq$ 2018] for results on cohomological descent, where, however, many questions remain open.

Results. There are a number of weaker cousins of the h-topology in the literature. They exclude the Frobenius morphism but still allow abstract blowups. The cdh-topology (see Section 2B) is the most well-established, appearing prominently in work on motives and K-theory, and having connections to rigid geometry; see [Friedlander and Voevodsky 2000; Voevodsky 2000; Suslin and Voevodsky 2000a; Cisinski and Déglise 2015; Cortiñas et al. 2008; Geisser and Hesselholt 2010; Cisinski 2013; Kerz et al. 2018; Morrow 2016], for example. We write Ω^n_τ for the sheafification of the presheaf $X \mapsto \Gamma(X, \Omega^n_{X/k})$ with respect to the topology τ.

MSC2010: primary 14G17; secondary 14F20.
Keywords: differential forms, cdh-topology, valuation rings, seminormalization, singularities.

Theorem 1.1. *Suppose k is a perfect field.*

(1) ([Huber, Kebekus and Kelly 2017], also Theorem 4.12) *For a smooth k-scheme X and $n \geq 0$,*

$$\Gamma(X, \Omega^n_{X/k}) = \Omega^n_{\mathrm{cdh}}(X).$$

(2) (Theorem 5.4, Proposition 5.9) *For any finite type separated k-scheme X, the restrictions $\Omega^n_{\mathrm{cdh}}|_{X_{\mathrm{Zar}}}$ and $\Omega^n_{\mathrm{cdh}}|_{X_{\mathrm{et}}}$ to the small Zariski and étale sites are coherent $\mathcal{O}_X$-modules.*

(3) (Proposition 6.2) *For functions, we have isomorphisms, functorial in X,*

$$\mathcal{O}_{\mathrm{cdh}}(X) = \mathcal{O}(X^{\mathrm{sn}}),$$

where X^{sn} is the seminormalisation of the variety X; see Section 2C.

(4) (Proposition 6.9) *For top degree differentials, we have*

$$\Omega^d_{\mathrm{cdh}}(X) \cong \varinjlim_{\substack{X' \to X \\ \text{proper birational}}} \Gamma(X', \Omega^d_{X'/k}), \qquad \dim X = d.$$

By combining (1) and (4) we deduce a corollary that involves only Kähler differentials. To our knowledge the formula is new:

Corollary 1.2 (Corollary 6.11). *Let k be a perfect field and X a smooth k-scheme. We have*

$$\Gamma(X, \Omega^d_{X/k}) \cong \varinjlim_{\substack{X' \to X \\ \text{proper birational}}} \Gamma(X', \Omega^d_{X'/k}), \qquad \dim X = d.$$

Note that it could also be deduced directly from [Huber, Kebekus and Kelly 2017, Theorem 5.8].

In contrast to the case of characteristic 0, the sheaves Ω^n_{cdh} are not torsion free (this was shown in [Huber, Kebekus and Kelly 2017, Example 3.6] by "pinching" along the Frobenius of a closed subscheme). So not only does lacking access to resolution of singularities cause proofs to become harder, the existence of inseparable field extensions actually changes some of the results.

Main tool. Our main tool, taking the role of a desingularisation of a variety X, is the category

$$\mathrm{val}(X),$$

a functorial variant of the Riemann–Zariski space, which we now discuss. Recall that the Riemann–Zariski space of an integral variety X is (as a set) given by the set of (all, not necessarily discrete) X-valuation rings of $k(X)$. The Riemann–Zariski space is only functorial for dominant morphisms of integral varieties. We replace it by the category $\mathrm{val}(X)$ (see Definition 2.21) of X-schemes of the form $\mathrm{Spec}(R)$ with R either a field of finite transcendence degree over k or a valuation ring of such a field. In [Huber, Kebekus and Kelly 2017], we were focussing on the discrete valuation rings — this turned out to be useful, but allowing nonnoetherian valuation rings yields a much better tool.

We define $\Omega^n_{\mathrm{val}}(X)$ as the set of global sections of the presheaf $\mathrm{Spec}(R) \mapsto \Omega^n_{R/k}$ on $\mathrm{val}(X)$; that is,

$$\Omega^n_{\mathrm{val}}(X) = \varprojlim_{R \in \mathrm{val}(X)} \Omega^n_{R/k}.$$

A global section admits the following very explicit description (Lemma 3.7). It is uniquely determined by specifying an element $\omega_x \in \Omega^n_{\kappa(x)/k}$ for every point $x \in X$ subject to two compatibility conditions: If R is an X-valuation ring of a residue field $\kappa(x)$, then ω_x has to be integral, i.e., contained in $\Omega^n_{R/k} \subseteq \Omega^n_{\kappa(x)/k}$. If ξ is the image of its special point in X, then $\omega_R|_{\mathrm{Spec}(R/\mathfrak{m})}$ has to agree with $\omega_\xi|_{\mathrm{Spec}(R/\mathfrak{m})}$ in $\Omega^n_{(R/\mathfrak{m})/k}$. The above mentioned results are deduced by establishing eh-descent (and therefore cdh- and rh-descent) for Ω^n_{val}. This is used to show the following:

Theorem 1.3 (Theorem 4.12). $\Omega^n_{\mathrm{rh}} = \Omega^n_{\mathrm{cdh}} = \Omega^n_{\mathrm{eh}} = \Omega^n_{\mathrm{val}}$.

This gives a very useful characterisation of Ω^n_{cdh}, and answers the open question [Huber, Kebekus and Kelly 2017, Proposition 5.13]. We also show that using other classes of valuation rings (e.g., rank one or strictly henselian or removing the transcendence degree bound) produces the same sheaf (see Proposition 3.11).

Special cases and the* sdh*-topology. There are two special cases, both of particular importance and with better properties: the case of 0-forms (i.e., functions) and the case of the “canonical sheaf” (i.e., d-forms on the category of k-schemes of dimension at most d). In both cases, the resulting sheaves even have descent for the sdh-topology introduced in [Huber, Kebekus and Kelly 2017]:

$$\mathcal{O}_{\mathrm{cdh}} = \mathcal{O}_{\mathrm{sdh}}, \qquad \Omega^d_{\mathrm{cdh}} = \Omega^d_{\mathrm{sdh}}$$

(see Remark 6.7 and Proposition 6.12). Recall that every variety is locally smooth in the sdh-topology by de Jong’s theorem on alterations, and the hope was that requiring morphisms to be separably decomposed would prohibit pathologies caused by purely inseparable extensions. Unfortunately, $\Omega^n(X) \neq \Omega^n_{\mathrm{sdh}}(X)$ for general n and X smooth; see [Huber, Kebekus and Kelly 2017].

Seminormalisation. For functions, we have the explicit computation

$$\mathcal{O}_{\mathrm{cdh}}(X) = \mathcal{O}(X^{\mathrm{sn}}), \tag{1}$$

where X^{sn} is the seminormalisation of the variety X (see Proposition 6.2). Here, the seminormalisation $X^{\mathrm{sn}} \to X$ is the universal morphism, which induces isomorphisms of topological spaces and residue fields; see Section 2C. In fact, we have this result for all representable presheaves.

Theorem 1.4 (Proposition 6.14). *Suppose S is a noetherian scheme and the normalisation of every finite type S-scheme is also finite type* (*i.e., that S is Nagata, as defined in Remark 2.7*)*. For instance, S might be the spectrum of a perfect field. Then for all separated finite type S-schemes X and Y, the canonical morphisms*

$$h^Y_{\mathrm{rh}}(X) = h^Y(X^{\mathrm{sn}}) \tag{2}$$

are isomorphisms, *where* $h^Y(-) = \hom_S(-, Y)$. *The natural maps*

$$h^Y_{\mathrm{rh}} \to h^Y_{\mathrm{cdh}} \to h^Y_{\mathrm{eh}} \to h^Y_{\mathrm{sdh}} \to h^Y_{\mathrm{val}} \tag{3}$$

are isomorphisms of presheaves on $\mathsf{Sch}^{\mathrm{ft}}_S$.

In characteristic 0, equation (2) is already formulated for the h-topology by Voevodsky [1996, Section 3.2], and generalised to algebraic spaces by Rydh [2010]. The theorem confirms that we have identified the correct analogy in positive characteristic.

As (2) confirms, the cdh-topology is not subcanonical. In fact, the full subcategory spanned by those cdh-sheaves which are sheafifications of representable sheaves is equivalent to the category of seminormal schemes (Corollary 6.17). In particular, the subcategory of smooth or even normal varieties remains unchanged; however, we lose information about certain singularities, e.g., cuspidal singularities are smoothened out. Depending on the question, this might be considered an advantage or a disadvantage. We strongly argue that it is an advantage for the natural questions of birational algebraic geometry, where differential forms are a main tool.

Recall that if X is integral, the ring of global sections of the structure sheaf of the normalisation $\widetilde{X}$ is the intersection of all X-valuation rings of the function field:

$$\Gamma(\widetilde{X}, \mathcal{O}_{\widetilde{X}}) = \bigcap_{\text{valuation rings } R \text{ of } k(X)} R.$$

In light of $\mathcal{O}_{\mathrm{cdh}} \cong \mathcal{O}_{\mathrm{val}}$, equation (1) has the following neat interpretation:

Scholium 1.5. *If X is reduced*, *the ring of global sections of the structure sheaf of the seminormalisation X^{sn} is the "intersection" of all X-valuation rings*:

$$\Gamma(X^{\mathrm{sn}}, \mathcal{O}_{X^{\mathrm{sn}}}) = \varprojlim_{\substack{\mathrm{Spec}(R) \to X \\ \text{with } R \text{ a valuation ring}}} R.$$

The seminormalisation was introduced and studied quite some time ago; see for example [Traverso 1970; Swan 1980], and for a historical survey see [Vitulli 2011]. The original motivation for considering the seminormalisation (or rather, the closely related and equivalent in characteristic 0 concept of weak normalisation) was to make the moduli space of positive analytic d-cycles on a projective variety "more normal" without changing its topology, i.e., without damaging too much the way that it solved its moduli problem [Andreotti and Norguet 1967]. Clearly, the cdh- and eh-topology are relevant to these moduli questions. Indeed, the cdh-topology already appears in the study of moduli of cycles in [Suslin and Voevodsky 2000a, Theorem 4.2.11]. In relation to this, let us point out that basically all of the present paper works for arbitrary unramified étale sheaves commuting with filtered limits of schemes, and in particular applying $(-)_{\mathrm{val}}$ to various Hilbert presheaves provides alternative constructions for Suslin and Voevodsky's relative cycle presheaves $z(X/S, r), z^{\mathrm{eff}}(X/S, r), c(X/S, r), c^{\mathrm{eff}}(X/S, r)$ introduced in [Suslin and Voevodsky 2000a], and heavily used in their work on motivic cohomology. Such matters, however, go beyond the scope of this current paper.

Outline of the paper. We start in Section 2 by collecting basic notation and facts on the Grothendieck topologies that we use, seminormality and valuation rings. In Section 2E we introduce our main tool, different categories of local rings above a given X, and discuss the relation to the Riemann–Zariski space.

In Section 3 we discuss and compare the presheaves on schemes of finite type over the base induced from presheaves on our categories of local rings. Everything is then applied to the case of differential forms.

In Section 4 we verify sheaf conditions on these presheaves. In Theorem 4.12, this culminates in our main comparison theorem on differential forms.

Section 5 establishes coherence of Ω^n_{rh} locally on the Zariski or even small étale site of any X. In Section 6 we turn to the special examples of $\mathcal{O}$, the canonical sheaf, and more generally the category of representable sheaves. Finally, in Section 7 we outline interesting open connections to the theory of Berkovich spaces and to F-singularities.

2. Commutative algebra and general definitions

2A. ***Notation.*** Throughout, k is assumed to be a perfect field. The case of interest is the case of positive characteristic. Sometimes we use a separated noetherian base scheme S. This includes the case $S = \mathrm{Spec}(k)$ of course.

The valuation rings we use are *not* assumed to be noetherian!

We denote by $\mathsf{Sch}^{\mathrm{ft}}_S$ the category of separated schemes of finite type over S, and write $\mathsf{Sch}^{\mathrm{ft}}_k$ when considering the case $S = \mathrm{Spec}(k)$.

We write $\Omega^n(X)$ for the vector space of k-linear n-differential forms on a k-scheme X, often denoted elsewhere by $\Gamma(X, \Omega^n_{X/S})$. Note the assignment $X \mapsto \Omega^n(X)$ is functorial in X.

Following [Stacks project, Tag 01RN], we call a morphism of schemes with finitely many irreducible components $f : X \to Y$ *birational* if it induces a bijection between the sets of irreducible components and an isomorphism on the residue fields of generic points. In the case of varieties (or more generally, if f is of finite presentation and X, Y are generically reduced [EGA IV$_3$ 1966, Théorème 8.10.5(i)]) this is equivalent to the existence of dense open subsets $U \subseteq X$ and $V \subset Y$ such that f induces an isomorphism $U \to V$.

2B. ***Topologies.***

Definition 2.1 (cdp-morphism). A morphism $f : Y \to X$ is called a cdp-*morphism* if it is proper and completely decomposed, where by "completely decomposed" we mean that for every (not necessarily closed) point $x \in X$ there is a point $y \in Y$ with $f(y) = x$ and $[k(y) : k(x)] = 1$.

These morphisms are also referred to as *proper* cdh-*covers* (e.g., by Suslin and Voevodsky [2000a]), or *envelopes* (e.g., by Fulton [1998]).

Remark 2.2 (rh, cdh and eh-topologies).

(1) Recall that the rh-topology on $\mathsf{Sch}^{\mathrm{ft}}_S$ is generated by the Zariski topology and cdp-morphisms [Goodwillie and Lichtenbaum 2001]. In a similar vein, the cdh-topology is generated by the

Nisnevich topology and cdp-morphisms [Suslin and Voevodsky 2000a, §5]. The eh-topology is generated by the étale topology and cdp-morphisms [Geisser 2006].

(2) We have an even stronger statement: every rh-, cdh- or eh- covering $U \to X$ in $\mathsf{Sch}_S^{\mathsf{ft}}$ admits a refinement of the form $W \to V \to X$, where $V \to X$ is a cdp-morphism and $W \to V$ is a Zariski, Nisnevich or étale covering, respectively [Suslin and Voevodsky 2000a, proof of Proposition 5.9].

2C. ***Seminormality.*** A special case of a cdp-morphism is the seminormalisation.

Definition 2.3 [Swan 1980]. A reduced ring A is *seminormal* if for all $b, c \in A$ satisfying $b^3 = c^2$, there is an $a \in A$ with $a^2 = b, a^3 = c$. Equivalently, every morphism $\operatorname{Spec}(A) \to \operatorname{Spec}(\mathbb{Z}[t^2, t^3])$ factors through $\operatorname{Spec}(\mathbb{Z}[t]) \to \operatorname{Spec}(\mathbb{Z}[t^2, t^3])$.

Definition 2.4. Recall that an inclusion of rings $A \subseteq B$ is called *subintegral* if $\operatorname{Spec}(B) \to \operatorname{Spec}(A)$ is a completely decomposed homeomorphism [Swan 1980, §2]. In other words, an inclusion is subintegral if it induces an isomorphism on topological spaces, and residue fields.

Remark 2.5. Any (not necessarily injective) ring map $\phi : A \to B$ inducing a completely decomposed homeomorphism is *integral* in the sense that for every $b \in B$ there is a monic $f(x) \in A[x]$ such that b is a solution to the image of $f(x)$ in $B[x]$ [Stacks project, Tag 04DF]. Consequently, a posteriori, subintegral inclusions are contained in the normalisation.

We have the following nice properties.

Lemma 2.6. *Let A be a reduced ring and $(A_i)_I$ a (not necessarily filtered) diagram of reduced rings.*

(1) *If A is normal, it is seminormal.*

(2) [Swan 1980, Corollary 3.3] *If all A_i are seminormal, then so is $\varprojlim_{i \in I} A_i$.*

(3) [Swan 1980, Corollary 3.4] *If the total ring of fractions $Q(A)$ is a product of fields, then A is seminormal if and only if for every subintegral extension $A \subseteq B \subseteq Q(A)$ we have $A = B$. In particular, this holds if A is noetherian or A is a valuation ring.*

(4) [Swan 1980, §2, Theorem 4.1] *There exists a universal morphism $A \to A^{\mathrm{sn}}$ with target a seminormal reduced ring. The morphism* $\operatorname{Spec}(A^{\mathrm{sn}}) \to \operatorname{Spec}(A)$ *is subintegral.*

(5) [Swan 1980, Corollary 4.6] *For any multiplicative set $S \subseteq A$ we have $A^{\mathrm{sn}}[S^{-1}] = A[S^{-1}]^{\mathrm{sn}}$; in particular, if A is seminormal, so is $A[S^{-1}]$.*

Remark 2.7. Recall that a scheme S is called *Nagata* if it is locally noetherian, and if for every $X \in \mathsf{Sch}_S^{\mathsf{ft}}$, the normalisation $X^n \to X$ [Stacks project, Tag 035E] is finite. Well-known examples are fields or the ring of integers $\mathbb{Z}$; more generally, quasi-excellent rings are Nagata [Stacks project, Tag 07QV].

Actually, the definition of Nagata [Stacks project, Tag 033S] is: for every point $x \in X$, there is an open $U \ni x$ such that $R = \mathcal{O}_X(U)$ is a Nagata ring [Stacks project, Tag 032R], i.e., noetherian and for every prime $\mathfrak{p}$ of R the quotient $R/\mathfrak{p}$ is N-2 [Stacks project, Tag 032F], i.e., for every finite field extension $L/\mathsf{Frac}(R/\mathfrak{p})$, the integral closure of $R/\mathfrak{p}$ in L is finite over $R/\mathfrak{p}$. However, Nagata proved [Stacks project, Tag 0334] that Nagata rings are characterised by being noetherian and universally Japanese [Stacks project, Tag 032R], but being universally Japanese is the same as having the property that every finite type ring morphism $R \to R'$ with R' a domain, the integral closure of R' in its fraction field is finite over R' [Stacks project, Tag 032F, Tag 0351]. Since we can assume the U from above is affine, and finiteness of the normalisation is detected locally, one sees that our definition above is equivalent to the standard one.

Proposition 2.8. *Let X be a scheme. There exists a universal morphism*

$$X^{\mathrm{sn}} \to X$$

from a scheme whose structure sheaf is a sheaf of seminormal reduced rings, called the seminormalisation *of X.*

(1) *It is a homeomorphism of topological spaces.*

(2) *If $X = \mathrm{Spec}(A)$ then $X^{\mathrm{sn}} = \mathrm{Spec}(A^{\mathrm{sn}})$.*

(3) *If the normalisation $X^n \to X$ is finite (e.g., if $X \in \mathsf{Sch}_k^{\mathrm{ft}}$, or more generally, if $X \in \mathsf{Sch}_S^{\mathrm{ft}}$ with S a Nagata scheme), then $X^{\mathrm{sn}} \to X$ is a finite* cdp*-morphism.*

Proof. Replacing X with its associated reduced scheme X_{red} we can assume that X is reduced. Define X^{sn} to be the ringed space with the same underlying topological space as X, and structure sheaf the sheaf obtained from $U \mapsto \mathcal{O}(U)^{\mathrm{sn}}$. It satisfies the appropriate universality by the universal property of $(-)^{\mathrm{sn}}$ for rings.

Since $(-)^{\mathrm{sn}}$ commutes with inverse limits and localisation by Lemma 2.6 parts (2) and (5), for any reduced ring A the structure sheaf of $\mathrm{Spec}(A^{\mathrm{sn}})$ is a sheaf of seminormal rings, and we obtain a canonical morphism $\mathrm{Spec}(A^{\mathrm{sn}}) \to \mathrm{Spec}(A)^{\mathrm{sn}}$ of ringed spaces. By the universal properties, $(-)^{\mathrm{sn}}$ commutes with colimits, so for any point $x \in X$ we have $\mathcal{O}_{X^{\mathrm{sn}},x} = \mathcal{O}_{X,x}^{\mathrm{sn}}$, and consequently, $\mathrm{Spec}(A^{\mathrm{sn}}) \to \mathrm{Spec}(A)^{\mathrm{sn}}$ is an isomorphism of ringed spaces. From this we deduce that in general X^{sn} is a scheme, and $X^{\mathrm{sn}} \to X$ is a completely decomposed morphism of schemes (see Lemma 2.6(4)).

Finally, by the universal property, there is a factorisation $X^n \to X^{\mathrm{sn}} \to X$. So if the normalisation is finite, then so is the seminormalisation. $\square$

The following well-known property explains the significance of the seminormalisation in our context.

Lemma 2.9. *Let $\mathcal{F}$ be an* rh*-sheaf on $\mathsf{Sch}_S^{\mathrm{ft}}$. Take $X \in \mathsf{Sch}_S^{\mathrm{ft}}$ and $X' \to X$ a completely decomposed homeomorphism. Then*

$$\mathcal{F}(X) \to \mathcal{F}(X^{\mathrm{red}}) \to \mathcal{F}(X')$$

are isomorphisms. In particular,

$$\mathcal{F}(X) \cong \mathcal{F}(X^{\mathrm{sn}})$$

if $X^{\mathrm{sn}} \in \mathsf{Sch}_S^{\mathrm{ft}}$ (for example, if S is Nagata).

Proof. The map $X^{\mathrm{red}} \to X$ is cdp and we have $(X^{\mathrm{red}} \times_X X^{\mathrm{red}}) = X^{\mathrm{red}}$. The isomorphism for X^{red} follows from the sheaf sequence. The same argument applies to X': Since $\mathcal{F}$ is an rh-sheaf we can assume X' is reduced. Then since $X' \to X$ is a finite [Stacks project, Tag 04DF] completely decomposed homeomorphism, so are the projections $X' \times_X X' \to X'$, and it follows that the diagonal induces an isomorphism $X' \xrightarrow{\sim} (X' \times_X X')^{\mathrm{red}}$. □

Lemma 2.10. *Suppose that*

$$\begin{tikzcd} E \arrow[r, "j"] \arrow[d, "q"'] & X' \arrow[d, "p"] \\ Z \arrow[r, "i"] & X \end{tikzcd}$$

is a commutative square in $\mathrm{Sch}^{\mathrm{ft}}_S$, *with* i, j *closed immersions*, p, q *finite surjective*, *and* p *an isomorphism outside* $i(Z)$.

Then the pushout $Z \sqcup_E X'$ *exists in* $\mathrm{Sch}^{\mathrm{ft}}_S$, *is reduced if both* Z *and* X' *are reduced*, *and the canonical morphism* $Z \sqcup_E X' \to X$ *is a finite completely decomposed homeomorphism. In particular*,

$$Z \sqcup_E X' \cong X$$

if Z, X' *are reduced and* X *is seminormal.*

Proof. First consider the case where X is affine (so all four schemes are affine). The pushout exists by [Ferrand 2003, Scolie 4.3, Théorème 5.1] and is given explicitly by the spectrum of the pullback of the underlying rings. By construction it is reduced as Z and X' are. The underlying set of the pushout is the pushout of the underlying sets [Ferrand 2003, Scolie 4.3]. As p is an isomorphism outside $i(Z)$ and q is surjective, it follows that $Z \sqcup_E X' \to X$ is a bijection on the underlying sets. The existence of the (continuous) map (of topological spaces) $Z \sqcup_E X' \to X$ shows that every open set of X is an open set of $Z \sqcup_E X'$. Conversely, if W is a closed set of $Z \sqcup_E X'$ then its preimage in X' is closed. As p is proper, this implies that W is also closed in X. In other words, every open set of $Z \sqcup_E X'$ is an open set of X. Hence $Z \sqcup_E X' \to X$ is a homeomorphism with reduced source. Now $Z \sqcup X' \to X$ is finite and completely decomposed, so $Z \sqcup_E X' \to X$ is also finite and completely decomposed. To summarise, $Z \sqcup_E X' \to X$ is a finite completely decomposed homeomorphism. That is, if both are reduced it comes from a subintegral extension of rings. If in addition X is seminormal, this implies that it is an isomorphism (Lemma 2.6(3)).

For a general $X \in \mathrm{Sch}^{\mathrm{ft}}_S$, the pushout exists by [Ferrand 2003, Théorème 7.1] (one checks the condition (ii) easily by pulling back an open affine of X to X'). Then the properties of $Z \sqcup_E X' \to X$ claimed in the statement can be verified on an open affine cover of X. As long as

$$U \times_X (Z \sqcup_E X') \cong (U \times_X Z) \sqcup_{U \times_X E} (U \times_X X')$$

for every open affine $U \subseteq X$, it follows from the case where X is affine. This latter isomorphism can be checked in the category of locally ringed spaces using the explicit description of [Ferrand 2003, Scolie 4.3]; it is also a special case of [Ferrand 2003, Lemme 4.4]. □

2D. ***Valuation rings.*** Recall that an integral domain R is called a *valuation ring* if for all $x \in K = \mathsf{Frac}(R)$, at least one of x or x^{-1} is in $R \subseteq K$. If R contains a field k, we say that R is a *k-valuation ring*. We say R is a *valuation ring of K* to emphasise that $K = \mathsf{Frac}(R)$.

The name valuation ring comes from the fact that the abelian group $\Gamma_R = K^*/R^*$ equipped with the relation "$a \leq b$ if and only if $b/a \in (R-\{0\})/R^*$" is a totally ordered group, and the canonical homomorphism $v : K^* \to K^*/R^*$ is a valuation, in the sense that $v(a+b) \geq \min(v(a), v(b))$ for all $a, b \in K^*$. Conversely, for any valuation on a field, the set of elements with nonnegative value are a valuation ring in the above sense. If $\Gamma_R = K^*/R^*$ is isomorphic to $\mathbb{Z}$ we say that R is a *discrete valuation ring*. Every noetherian valuation ring is either a discrete valuation ring or a field.

One of the many, varied characterisations of valuation rings is the following.

Proposition 2.11 [Bourbaki 1964, Chapitre VI, §1.2, Théorème 1]. *Let $R \subseteq K$ be a subring of a field. Then R is a valuation ring if and only if its set of ideals is totally ordered.*

In particular, this implies that the maximal ideal is unique, that is, every valuation ring is a local ring. The cardinality of the set of nonzero prime ideals of a valuation ring is called its *rank*. As the set of primes is totally ordered, the rank agrees with the Krull dimension.

Corollary 2.12. *Let R be a valuation ring. If $\mathfrak{p} \subseteq R$ is a prime ideal, then both the quotient $R/\mathfrak{p}$ and the localisation $R_{\mathfrak{p}}$ are again valuation rings.*

Corollary 2.13. *Let R be a valuation ring and $S \subseteq R-\{0\}$ a multiplicative set. Then $R[S^{-1}]$ is a valuation ring. In fact, $\mathfrak{p} = \bigcup_{S \cap \mathfrak{q} = \varnothing} \mathfrak{q}$ is a prime and $R[S^{-1}] = R_{\mathfrak{p}}$.*

Proof. By Corollary 2.12 it suffices to show the second claim. Clearly $\mathfrak{p}$ is prime. Recall that the canonical inclusion $R \to R[S^{-1}]$ induces an isomorphism of *locally ringed spaces* between $\mathrm{Spec}(R[S^{-1}])$ and $\{\mathfrak{q} : \mathfrak{q} \cap S = \varnothing\} \subseteq \mathrm{Spec}(R)$. Since this set has a maximal element, namely $\mathfrak{p}$, the morphism $\mathrm{Spec}(R_{\mathfrak{p}}) \to \mathrm{Spec}(R[S^{-1}])$ is an isomorphism of *locally ringed spaces*. Applying $\Gamma(-, \mathcal{O}_-)$ gives the desired ring isomorphism. ☐

Given a prime $\mathfrak{p}$ of a valuation ring R, we can of course reconstruct R from the valuation rings $R_{\mathfrak{p}}$ and $R/\mathfrak{p}$ and their canonical maps to the residue field $\kappa(\mathfrak{p})$:

Lemma 2.14. *Let R be a valuation ring and $\mathfrak{p}$ a prime ideal. Then the diagram*

$$\begin{tikzcd} R \arrow[r, hook] \arrow[d] & R_{\mathfrak{p}} \arrow[d] \\ R/\mathfrak{p} \arrow[r, hook] & R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}} \cong \mathsf{Frac}(R/\mathfrak{p}) \end{tikzcd} \tag{4}$$

is cartesian and the canonical R-module morphism $\mathfrak{p} \to \mathfrak{p}R_{\mathfrak{p}}$ is an isomorphism.

Proof. We have to check that an element a/s of $R_{\mathfrak{p}}$ is in R if its reduction modulo $\mathfrak{p}R_{\mathfrak{p}}$ is in $R/\mathfrak{p}$. This amounts to showing that $\mathfrak{p} = \mathfrak{p}R_{\mathfrak{p}}$: if there is $b \in R$ which agrees with a/s mod $\mathfrak{p}R_{\mathfrak{p}}$, then $b - a/s \in \mathfrak{p}R_{\mathfrak{p}}$. But if $\mathfrak{p} = \mathfrak{p}R_{\mathfrak{p}}$, then $b - a/s \in \mathfrak{p} \subset R$, i.e., $a/s \in R \subseteq R_{\mathfrak{p}}$.

Let $x = a/s \in \mathfrak{p}R_{\mathfrak{p}}$ with $a \in \mathfrak{p}$ and $s \in R - \mathfrak{p}$. Let v be the valuation of R. We compare $v(a)$ and $v(s)$. If we had $v(s) \geq v(a)$, then $x^{-1} \in R$ and hence $s = x^{-1}a \in \mathfrak{p}$ because $\mathfrak{p}$ is an ideal. This is a contradiction. Hence $v(s) < v(a)$ and $x \in R$. Now we have the equation $sx = a$ in R; hence $sx \in \mathfrak{p}$. As $\mathfrak{p}$ is a prime ideal and $s \notin \mathfrak{p}$, this implies $x \in \mathfrak{p}$. □

Another one of the many characterisations of valuation rings is the following.

Proposition 2.15 [Stacks project, Tag 092S; Olivier 1983]. *A local ring R is a valuation ring if and only if every submodule of every flat R-module is again a flat R-module.*

From this we immediately deduce the following, which is crucial for Corollary 2.18.

Corollary 2.16. *Let R be a valuation ring. If $R \to A$ is a flat R-algebra with $A \otimes_R A \to A$ also flat, then for every prime ideal $\mathfrak{p} \subset A$ the localisation $A_{\mathfrak{p}}$ is again a valuation ring.*

Proof. If A satisfies the hypotheses, then so does $A_{\mathfrak{p}}$, so we can assume A is local, and it suffices to show that every sub-A-module of a flat A-module is a flat A-module. Recall that flatness of $R \to A$ implies the forgetful functor $U : A\text{-mod} \to R\text{-mod}$ preserves flatness, because $- \otimes_A A \otimes_R - \cong (U-) \otimes_R -$. Recall also that flatness of $A \otimes_R A \to A$ implies that U *detects* flatness, because we have isomorphisms

$$- \otimes_A - \cong (- \otimes_A (A \otimes_R A) \otimes_A -) \otimes_{(A \otimes_R A)} A \quad \text{and} \quad (- \otimes_A (A \otimes_R A) \otimes_A -) \cong (U-) \otimes_R (U-).$$

Since U preserves and detects flatness, and preserves monomorphisms, the claim now follows from Proposition 2.15. □

As one might expect, the rank is bounded by the transcendence degree.

Proposition 2.17 [Bourbaki 1964, Chapitre VI, §10.3, Corollaire 1; Gabber and Ramero 2003, 6.1.24; Engler and Prestel 2005, Corollary 3.4.2]. *Suppose that R' is a valuation ring and $K' = \mathrm{Frac}(R')$. Let K'/K be a field extension. Note that $R = K \cap R'$ is again a valuation ring, and the inclusion $R \subseteq R'$ induces a field extension of residue fields $\kappa \subseteq \kappa'$, and a morphism $\Gamma_R \to \Gamma_{R'}$ of totally ordered groups. With this notation, we have*

$$\mathrm{tr.d}(K'/K) + \mathrm{rk}\, R \geq \mathrm{tr.d}(\kappa'/\kappa) + \mathrm{rk}\, R'.$$

In particular, if R' is a k-valuation ring for some field k, we have

$$\mathrm{tr.d}(K'/k) \geq \mathrm{tr.d}(\kappa'/k) + \mathrm{rk}\, R',$$

and finiteness of $\mathrm{tr.d}(K'/k)$ *implies finiteness of* $\mathrm{rk}\, R'$.

Recall that a local ring R is called *strictly henselian* if every faithfully flat étale morphism $R \to A$ admits a retraction. Every local ring R admits a "smallest" local morphism towards a strictly henselian local ring, which is unique up to nonunique isomorphism. The target of any such morphism is called *the strict henselisation* and is denoted by R^{sh}. There are various ways to construct this. One standard construction is to choose a separable closure κ^s of the residue field κ and take the colimit

$$R^{\mathrm{sh}} = \varinjlim_{R \to A \to \kappa^s} A \tag{5}$$

over factorisations such that $R \to A$ is étale.

From Corollary 2.16 we deduce the following.

Corollary 2.18. *For any valuation ring R, the strict henselisation is again a valuation ring. If R is of finite rank, then the following hold*:

(1) *If $R \to A$ is an étale algebra, $R \to A_{\mathfrak{p}}$ is also an étale algebra for all $\mathfrak{p} \subset A$.*

(2) *In the colimit* (5), *it suffices to consider those étale R-algebras A which are valuation rings.*

(3) *Every étale covering $U \to$* Spec(R) *admits a Zariski covering $V \to U$ such that V is a disjoint union of spectra of valuation rings.*

This corollary actually holds whenever the primes of R are well-ordered by [Bourbaki 1964, Chapitre VI, §8.3, Théorème 1], and might very well be true in general, but finite rank suffices for our purposes.

Proof. Recall that the diagonal is open immersion in the case of an unramified morphism. Hence the assumptions of Corollary 2.16 are satisfied for all étale R-algebras and their cofiltered limits. In particular, R^{sh} is a valuation ring.

(1) It suffices to show that $A \to A_{\mathfrak{p}}$ is of finite type. First note that since $\mathrm{Spec}(A) \to \mathrm{Spec}(R)$ is étale, it is quasifinite, and so since R has finite rank, A has finitely many primes. For every prime $\mathfrak{q} \subset A$ such that $\mathfrak{q} \not\subseteq \mathfrak{p}$, choose one $\pi_{\mathfrak{q}} \in \mathfrak{q} \backslash (\mathfrak{q} \cap \mathfrak{p})$. Then $A_{\mathfrak{p}} = A[\pi_{\mathfrak{q}_1}^{-1}, \dots, \pi_{\mathfrak{q}_n}^{-1}]$ (see Corollary 2.13).

(2) We can replace each A with $A_{\mathfrak{p}}$ without affecting the colimit (see part (1)).

(3) Part (1) also implies that $\mathrm{Spec}(A_{\mathfrak{p}}) \to \mathrm{Spec}(A)$ is an open immersion. Each of the finitely many $A_{\mathfrak{p}}$ is a valuation ring by Corollary 2.16 □

Just as strictly henselian rings are "local rings" for the étale topology, strictly henselian valuation rings are the "local rings" for the eh-topology.

Proposition 2.19 [Gabber and Kelly 2015, Theorem 2.6]. *A k-ring R is a valuation ring* (*resp. henselian valuation ring*, *strictly henselian valuation ring*) *if and only if for every* rh-*covering* (*resp.* cdh-*covering*, eh-*covering*) $\{U_i \to X\}$ *in* $\mathsf{Sch}_k^{\mathrm{ft}}$, *the morphism of sets of k-scheme morphisms*

$$\coprod_{i \in I} \mathrm{hom}(\mathrm{Spec}(R), U_i) \to \mathrm{hom}(\mathrm{Spec}(R), X)$$

is surjective.

The key input into the present paper is the same as for [Huber, Kebekus and Kelly 2017].

Theorem 2.20. *For every finitely generated extension K/k and every k-valuation ring R of K the map*

$$\Omega^n(R) \to \Omega^n(K)$$

is injective for all $n \geq 0$, i.e., Ω^n is torsion free on val(k).

This is due to Gabber and Ramero [2003, Corollary 6.5.21] for $n = 1$. The general case is deduced in [Huber, Kebekus and Kelly 2017, Lemma A.4].

2E. *Presheaves on categories of local rings.* We fix a base scheme S. The case of main interest for the present paper is $S = \mathrm{Spec}(k)$ with k a field.

Definition 2.21. Let X/S be a scheme of finite type. We use the following notation:

- $\mathrm{val}^{\mathrm{big}}(X)$ is the category of X-schemes $\mathrm{Spec}(R)$ with R either a field extension of k or a k-valuation ring of such a field.
- $\mathrm{val}(X)$ is the full subcategory of $\mathrm{val}^{\mathrm{big}}(X)$ of those $\pi : \mathrm{Spec}(R) \to X$ such that the transcendence degree $\mathrm{tr.d}(k(\eta)/k(\pi(\eta)))$ is finite (but we do not demand the field extension to be finitely generated), where η is the generic point of $\mathrm{Spec}(R)$.
- $\mathrm{val}^{\le r}(X)$ is the full subcategory of $\mathrm{val}(X)$ of those $\mathrm{Spec}(R) \to X$ such that rank of the valuation ring R is $\le r$.
- $\mathrm{dvr}(X)$ is the full subcategory of $\mathrm{val}(X)$ of those $\mathrm{Spec}(R) \to X$ such that $R = \mathcal{O}_{Y,y}$ for some $Y \in \mathsf{Sch}^{\mathsf{ft}}_X$ and some point $y \in Y$ of codimension ≤ 1 which is regular.
- $\mathrm{shval}(X)$ is the full subcategory of $\mathrm{val}(X)$ of those $\mathrm{Spec}(R) \to X$ such that R is strictly henselian.
- $\mathrm{rval}(X)$ is the full subcategory of $\mathrm{val}(X)$ of those $\mathrm{Spec}(R) \to X$ such that R is a valuation ring of a residue field of X.

We generically denote one of the above categories of local rings by $\mathrm{loc}(X)$. We also write loc for $\mathrm{loc}(S)$.

Note that the morphisms in the above categories are not required to be induced by local homomorphisms of local rings. All X-morphisms are allowed.

Definition 2.22. Let $\mathcal{F}$ be a presheaf on loc. We say that $\mathcal{F}$ is *torsion free* if

$$\mathcal{F}(R) \to \mathcal{F}(\mathsf{Frac}(R))$$

is injective for all valuation rings R in loc and their field of fractions $\mathsf{Frac}(R)$.

We could have also called this property *separated*, since when $\mathcal{F}$ is representable, it is the valuative criterion for separatedness.

Lemma 2.23. *For any $X \in \mathsf{Sch}^{\mathsf{ft}}_S$, let $X' \to X$ be any completely decomposed homeomorphism* (*in particular*, *if $X^{\mathrm{sn}} \in \mathsf{Sch}^{\mathsf{ft}}_S$, for example if S is Nagata, we can take $X' = X^{\mathrm{sn}}$*). *Then*

$$\mathrm{loc}(X) = \mathrm{loc}(X')$$

for all of the above categories of local rings.

Proof. This follows from the universal property of $(-)^{\mathrm{sn}}$ since valuation rings are normal. □

The category val should be seen as a fully functorial version of the Riemann–Zariski space.

Definition 2.24. Let X be an integral S-scheme of finite type with generic point η. As a set, the Riemann–Zariski space $\mathsf{RZ}(X)$, called the "Riemann surface" in [Zariski and Samuel 1960, §17, p. 110], is the set of (not necessarily discrete) valuation rings over X of the function field $k(X)$; see also [Temkin 2011, before Remark 2.1.1, after Remark 2.1.2, before Proposition 2.2.1, Proposition 2.2.1, Corollary 3.4.7]. We turn it into a topological space by using as a basis the sets of the form

$$E(A') = \{R \in \mathsf{RZ}(X) | A' \subset R\},$$

where $\mathrm{Spec}(A)$ is an affine open of X, and A' is a finitely generated sub-A-algebra of $k(X)$; see [Temkin 2011, before Lemma 3.1.1, before Lemma 3.1.8]. It has a canonical structure of locally ringed space induced by the assignment $U \mapsto \bigcap_{R \in U} R$ for open subsets $U \subseteq \mathsf{RZ}(X)$. One can equivalently define $\mathsf{RZ}(X)$ as the inverse limit of all proper birational morphisms $Y \to X$, taking the inverse limit in the category of locally ringed spaces. In particular, as a set it is the inverse limit of the underlying sets of the Y, equipped with the coarsest topology making the projections $\varprojlim Y \to Y$ continuous, and the structure sheaf is the colimit of the inverse images of the $\mathcal{O}_Y$ along the projections $\varprojlim Y \to Y$.

This topological space is quasicompact, in the sense that every open cover admits a finite subcover; see [Zariski and Samuel 1960, Theorem 40] for the case $S = \mathrm{Spec}(k)$, and [Temkin 2011, Proposition 3.1.10] for general S.

Note that the Riemann–Zariski space is functorial only for dominant morphisms. Our category $\mathrm{loc}(X)$ above is the functorial version: it is the union of the Riemann–Zariski-spaces of all integral X-schemes of finite type.

3. Presheaves on categories of valuation rings

3A. ***Generalities.*** We now introduce our main player. We fix a base scheme S.

Definition 3.1. Let X/S be of finite type. Let $\mathcal{F}$ be a presheaf on one of the categories of local rings $\mathrm{loc}(X)$ of Definition 2.21 over X.

We define $\mathcal{F}_{\mathrm{loc}}(X)$ as a global section of the presheaf $\mathrm{Spec}(R) \mapsto \mathcal{F}(\mathrm{Spec}(R))$ on $\mathrm{loc}(X)$, i.e., as the projective limit

$$\mathcal{F}_{\mathrm{loc}}(X) = \varprojlim_{\mathrm{loc}(X)} \mathcal{F}(\mathrm{Spec}(R))$$

over the respective categories.

Remark 3.2. This means that an element of $\mathcal{F}_{\mathrm{loc}}(X)$ is defined as a system of elements $s_R \in \mathcal{F}(\mathrm{Spec}(R))$ indexed by objects $\mathrm{Spec}(R) \to X \in \mathrm{loc}(X)$ which are compatible in the sense that for every morphism $\mathrm{Spec}(R) \to \mathrm{Spec}(R')$ in $\mathrm{loc}(X)$, we have $s_{R'}|_R = s_R$. Note that this is an abuse of notation, since the element s_R does *not only* depend on R but also on the structure map $\mathrm{Spec}(R) \to X$. Most of the time the structure map will be clear from the context.

Remark 3.3. We have the following equivalent definitions of $\mathcal{F}_{\mathrm{loc}}(X)$.

(1) $\mathcal{F}_{\mathrm{loc}}(X)$ is the equaliser of the canonical maps

$$\mathcal{F}(X) = \mathrm{eq}\left(\prod_{\mathrm{Spec}(R)\to X\in\mathrm{loc}(X)} \mathcal{F}(R) \overset{d^0}{\underset{d^1}{\rightrightarrows}} \prod_{\mathrm{Spec}(R)\to\mathrm{Spec}(R')\to X\in\mathrm{loc}(X)} \mathcal{F}(R)\right).$$

In particular, since valuation rings are the "local rings" for the rh-site [Gabber and Kelly 2015], the construction $\mathcal{F}_{\mathrm{val}}$ can be thought of as a naïve Godement sheafification (it differs in general from the Godement sheafification because colimits do not commute with infinite products).

(2) $\mathcal{F}_{\mathrm{loc}}$ is the (restriction to $\mathsf{Sch}^{\mathrm{ft}}_S$ of the) right Kan extension along the inclusion $\iota : \mathrm{loc} \subseteq \mathsf{Sch}_S$:

$$\mathcal{F}_{\mathrm{val}} = (\iota^! \mathcal{F})|_{\mathsf{Sch}^{\mathrm{ft}}_S}.$$

(3) $\mathcal{F}_{\mathrm{loc}}(X)$ is the set of natural transformations

$$\mathcal{F}_{\mathrm{loc}}(X) = \mathrm{hom}_{\mathsf{PreShv(loc)}}(h^X, \mathcal{F}),$$

where $h^X = \mathrm{hom}_{\mathsf{Sch}_S}(-, X)$.

Lemma 3.4. *Let $\mathcal{F}$ be a presheaf on* loc. *Then the assignment $X \mapsto \mathcal{F}_{\mathrm{loc}}(X)$ defines a presheaf $\mathcal{F}_{\mathrm{loc}}$ on $\mathsf{Sch}^{\mathrm{ft}}_S$.*

Proof. Composition of $\mathrm{Spec}(R) \to X$ with a morphism $f : X \to Y$ of schemes of finite type over k defines a functor $\mathrm{loc}(X) \to \mathrm{loc}(Y)$ and hence a homomorphism of limits $f^* : \mathcal{F}_{\mathrm{loc}}(Y) \to \mathcal{F}_{\mathrm{loc}}(X)$. □

Remark 3.5. We show in Proposition 3.11 that for $S = \mathrm{Spec}(k)$ with k a perfect field, we have

$$\Omega^n_{\mathrm{val}} = \Omega^n_{\mathrm{val}^{\mathrm{big}}} = \Omega^n_{\mathrm{val}^{\leq 1}} = \Omega^n_{\mathrm{shval}}$$

as presheaves on $\mathsf{Sch}^{\mathrm{ft}}_k$. In [Huber, Kebekus and Kelly 2017], we systematically studied the case of the category dvr. If every $X \in \mathsf{Sch}^{\mathrm{ft}}_k$ admits a proper birational morphism from a smooth k-scheme, we also have

$$\Omega^n_{\mathrm{val}} = \Omega^n_{\mathrm{dvr}}$$

because both are equal to Ω^n_{cdh} in this case. In positive characteristic, the only cases that we know $\Omega^n_{\mathrm{val}}(X) = \Omega^n_{\mathrm{dvr}}(X)$ unconditionally are if either $n = 0$ (see Remark 6.7), $n = \dim X$ (see Proposition 6.12), $\dim X < n$ (in which case both are zero) or $\dim X \leq 3$.

Lemma 3.6. *Let $\mathcal{F}$ be a presheaf on* $\mathrm{val}^{\mathrm{big}}$. *Then $\mathcal{F}_{\mathrm{val}} = \mathcal{F}_{\mathrm{val}^{\mathrm{big}}}$ as presheaves on $\mathsf{Sch}^{\mathrm{ft}}_S$. A section over X is uniquely determined by the value on the residue fields of X and their valuation rings, that is, the maps $\mathcal{F}_{\mathrm{loc}}(X) \to \prod_{\mathrm{rval}(X)} \mathcal{F}(R)$ are injective for* $\mathrm{loc} = \mathrm{val}, \mathrm{val}^{\mathrm{big}}$ *and all $X \in \mathsf{Sch}^{\mathrm{ft}}_S$.*

Proof. We begin with the second statement. Suppose $s, t \in \mathcal{F}_{\mathrm{loc}}(X)$ are sections such that $s|_R = t|_R$ for all $R \in \mathrm{rval}(X)$. If $R \subseteq K$ is a valuation ring, then by Proposition 2.17, the intersection $R' = R \cap \kappa(x) \subseteq K$ is a valuation ring. Then the sections s_R, t_R for $\mathrm{Spec}(R) \to X$ (see Remark 3.2) agree by the compatibility condition $s_R = s_{R'}|_R = t_{R'}|_R = t_R$.

Now the first statement. Since we have a factorisation

$$\mathcal{F}_{\mathrm{val}^{\mathrm{big}}}(X) \to \mathcal{F}_{\mathrm{val}}(X) \to \prod_{\mathrm{rval}(X)} \mathcal{F}(R),$$

it follows that the first map is injective, and we only need to show it is surjective.

Let $s \in \mathcal{F}_{\mathrm{val}}(X)$ be a global section. Defining $t_R := s_{R'}|_R$, with R and R' as above, we get a candidate element $t = (t_R) \in \prod_{\mathrm{val}^{\mathrm{big}}(X)} \mathcal{F}(R)$ which is potentially in $\mathcal{F}_{\mathrm{val}^{\mathrm{big}}}(X)$. Let $\mathrm{Spec}(R_1) \to \mathrm{Spec}(R_2)$ be a morphism in $\mathrm{val}^{\mathrm{big}}(X)$. Let R_1', R_2' be the valuation rings of residue fields of X corresponding to R_1, R_2 as above, but since there is not necessarily a morphism $\mathrm{Spec}(R_1') \to \mathrm{Spec}(R_2')$, we also set $\mathfrak{p}_2' = \ker(R_2' \to R_1)$ and $S = \kappa(\mathfrak{p}_2') \cap R_1$, to obtain the commutative diagram

$$\begin{tikzcd}
 & & \mathrm{Spec}(R_1') \\
 & \mathrm{Spec}(S) \arrow[ur] \arrow[dr] & \\
\mathrm{Spec}(R_1) \arrow[ur] \arrow[dr] & & \mathrm{Spec}(R_2') \\
 & \mathrm{Spec}(R_2) \arrow[ur] &
\end{tikzcd} \tag{6}$$

in $\mathrm{val}^{\mathrm{big}}(X)$, with $R_1', R_2', S \in \mathrm{val}(X)$ by Proposition 2.17. Now the result follows from a diagram chase: we have

$$t_{R_1} = s_{R_1'}|_{R_1} = s_{R_1'}|_S|_{R_1} = s_S|_{R_1} = s_{R_2'}|_S|_{R_1} = s_{R_2'}|_{R_2}|_{R_1} = t_{R_2}|_{R_1}. \qquad \square$$

Recall that a presheaf on loc is torsion free if it sends dominant morphisms to monomorphisms (see Definition 2.22).

Lemma 3.7. *Let* $\mathrm{loc} \in \{\mathrm{val}, \mathrm{val}^{\mathrm{big}}, \mathrm{val}^{\leq 1}, \mathrm{rval}\}$. *Let $\mathcal{F}$ be a torsion free presheaf on* $\mathrm{loc}(X)$. *Then* $\mathcal{F}_{\mathrm{loc}}(X)$ *is canonically isomorphic to*

$$\left\{ (s_x) \in \prod_{x \in X} \mathcal{F}(x) \;\middle|\; \begin{array}{l} \textit{for every } T \to X \in \mathrm{loc}(X) \textit{ there exists} \\ \qquad s_T \in \mathcal{F}(T) \textit{ such that } (s_x)|_{\prod_{t \in T} F(t)} = s_R|_{\prod_{t \in T} F(t)} \end{array} \right\}. \tag{7}$$

It is perhaps worth noting that the description in (7) is basically the presheaf rs F from the proof of [Kelly 2012, Proposition 3.6.12].

Proof. By torsion freeness, the projection $\mathcal{F}_{\mathrm{loc}}(X) \to \mathcal{F}_{\mathrm{val}^{\leq 0}}(X)$ is injective. By functoriality the map $\mathcal{F}_{\mathrm{val}^{\leq 0}}(X) \to \prod_{x \in X} \mathcal{F}(x)$ is injective.

Assume conversely we are given a system of s_x as in (7). As in the proof of the previous lemma, this gives us a candidate section $t_R \in \mathcal{F}(\mathrm{Spec}(R))$ for all $\mathrm{Spec}(R) \in \mathrm{loc}(X)$. It remains to check compatibility of these sections. Let $\mathrm{Spec}(R_1) \to \mathrm{Spec}(R_2)$ be a morphism in $\mathrm{loc}(X)$. The generic point of $\mathrm{Spec}(R_1)$ maps to a point of $\mathrm{Spec}(R_2)$ corresponding to a prime ideal $\mathfrak{p}_2 \subset R_2$. By torsion freeness, we may replace R_1 by its field of fractions and R_2 by $(R_2)_{\mathfrak{p}_2}$. In other words, R_1 is a field containing the residue field of R_2. Now the same diagram chase as for (6) works. Since $S = R_2'/\mathfrak{m}$ and $x = \mathrm{Spec}(R_1')$ is the image of $\mathrm{Spec}(R_1)$ in X, keeping in mind the condition of (7) above, we have

$$t_{R_1} = s_x|_{R_1} = s_x|_{R_2'/\mathfrak{m}}|_{R_1} = s_{\mathrm{Frac}(R_2')}|_{R_2'/\mathfrak{m}}|_{R_1} = s_{\mathrm{Frac}(R_2')}|_{R_2}|_{R_1} = t_{R_2}|_{R_1}. \qquad \square$$

3B. ***Reduction to strictly henselian valuation rings.*** The aim of this section is to establish that using strictly henselian local rings gives the same result.

Proposition 3.8. *Let $\mathcal{F}$ be a presheaf on* val *that commutes with filtered colimits and satisfies the sheaf condition for the étale topology. Then the canonical projection morphism*

$$\mathcal{F}_{\mathrm{val}}(X) \to \mathcal{F}_{\mathrm{shval}}(X)$$

is an isomorphism.

Proof. Let $s, t \in \mathcal{F}_{\mathrm{val}}(X)$ be sections such that the induced elements of $\mathcal{F}_{\mathrm{shval}}(X)$ agree. Let $x \in X$ be a point with residue field κ. The separable closure κ^s of κ is in the category shval. By assumption,

$$\mathcal{F}(\kappa^s) = \varinjlim_{\lambda/\kappa} \mathcal{F}(\kappa),$$

where λ runs through the finite extensions of κ contained in κ^s. The vanishing of s_{κ^s} implies that there is one such λ with $s_\lambda = 0$. As $\lambda \subseteq \kappa^s$, the morphism $\mathrm{Spec}(\lambda) \to \mathrm{Spec}(\kappa)$ is étale. As a consequence of étale descent, we know that the map $\mathcal{F}(\kappa) \to \mathcal{F}(\lambda)$ is injective, hence $s_\kappa = t_\kappa$. The same argument also applies to a valuation ring R of κ and its strict henselisation, viewed as the colimit of (5); see Corollary 2.18(2). Hence the morphism in the statement is injective.

Now let $t \in \mathcal{F}_{\mathrm{shval}}(X)$. For any morphism $\mathrm{Spec}(R) \to X$ in $\mathrm{val}(X)$ with strict henselisation R^{sh}, let $G = \mathsf{Gal}(\mathsf{Frac}(R^{\mathrm{sh}})/\mathsf{Frac}(R))$. Since $\mathsf{Frac}(R) = \mathsf{Frac}(R^{\mathrm{sh}})^G$, and $R = R^{\mathrm{sh}} \cap \mathsf{Frac}(R)$, we have $R = (R^{\mathrm{sh}})^G$. In particular, the element $t_{R^{\mathrm{sh}}}$ lifts to $\mathcal{F}(R) \subseteq \mathcal{F}(R^{\mathrm{sh}})$, as $t_{R^{\mathrm{sh}}}$ must be compatible with every morphism $\mathrm{Spec}(R^{\mathrm{sh}}) \to \mathrm{Spec}(R^{\mathrm{sh}})$ in $\mathrm{shval}(X)$. In this way, we obtain an element t_R for every $\mathrm{Spec}(R) \to X$ in $\mathrm{val}(X)$. We want to know that these form a section of $\mathcal{F}_{\mathrm{val}}(X)$. But compatibility with morphisms $\mathrm{Spec}(R) \to \mathrm{Spec}(R')$ of $\mathrm{val}(X)$ follows from the definition of the t_R, the fact that strict henselisations are functorial [Stacks project, Tag 08HR], and the morphisms $\mathcal{F}(R) \to \mathcal{F}(R^{\mathrm{sh}})$ being injective, which we just proved. □

3C. ***Reduction to rank one.*** The aim of this section is to establish that using rank one valuation rings gives the same result:

Proposition 3.9. *Let $\mathcal{F}$ be a torsion free presheaf on* val. *Assume that for every valuation ring R in* val *and prime ideal $\mathfrak{p}$ the diagram*

$$\begin{tikzcd} \mathcal{F}(R) \arrow[r, hook] \arrow[d] & \mathcal{F}(R_{\mathfrak{p}}) \arrow[d] \\ \mathcal{F}(R/\mathfrak{p}) \arrow[r, hook] & \mathcal{F}(\mathsf{Frac}(R/\mathfrak{p})) \end{tikzcd}$$

is cartesian. Then the natural restriction

$$\mathcal{F}_{\mathrm{val}} \to \mathcal{F}_{\mathrm{val}^{\leq 1}}$$

is an isomorphism.

Proof. Recall that $\mathrm{val}^{\leq r}$ is defined to be the subcategory of val using only valuations of rank at most r, and $\mathcal{F}_{\mathrm{val}^{\leq r}}$ denotes the presheaf obtained using only $\mathrm{val}^{\leq r}$. There are canonical morphisms

$$\mathcal{F}_{\mathrm{val}} \to \mathcal{F}_{\mathrm{val}^{\leq r}} \to \mathcal{F}_{\mathrm{val}^{\leq r-1}}$$

for all $r \geq 1$. By torsion freeness these are all subpresheafs of $\mathcal{F}_{\mathrm{val}^{\leq 0}}$, and so the two morphisms above are monomorphisms. Moreover, $\mathcal{F}_{\mathrm{val}} = \bigcap_r \mathcal{F}_{\mathrm{val}^{\leq r}}$ because by definition, all fields in val are of finite transcendence degree over k and hence all valuation rings have finite rank (see Proposition 2.17). Hence it suffices to show that $\mathcal{F}_{\leq r} \to \mathcal{F}_{\leq r-1}$ is an epimorphism for all $r \geq 2$.

Let $\mathfrak{p}$ be a prime ideal of a valuation ring R. By Corollary 2.12 both $R_{\mathfrak{p}}$ and $R/\mathfrak{p}$ are valuation rings and if $\mathfrak{p}$ is not maximal or zero, then $R_{\mathfrak{p}}$ and $R/\mathfrak{p}$ are of rank smaller than R. Indeed, $\operatorname{rk} R = \operatorname{rk} R_{\mathfrak{p}} + \operatorname{rk} R/\mathfrak{p}$ since the rank is equal to the Krull dimension and the set of ideals, and in particular prime ideals, of a valuation ring is totally ordered (see Proposition 2.11).

Let $t \in \mathcal{F}_{\mathrm{val}^{\leq r-1}}(X)$ be a section. If $r \geq 2$, we choose a canonical candidate $s \in \prod_{R \in \mathrm{val}^{\leq r}} \mathcal{F}(\operatorname{Spec}(R))$ for an element of $\mathcal{F}_{\mathrm{val}^{\leq r}}(X)$ in the preimage of t: for every valuation ring of rank r, take $\mathfrak{p}$ to be any prime ideal with $R_{\mathfrak{p}}, R/\mathfrak{p} \in \mathrm{val}^{\leq r-1}$ and construct a section over R using the cartesian square of the assumption. Since the morphisms $\mathcal{F}(\operatorname{Spec}(R)) \to \mathcal{F}(\operatorname{Spec}(R_{\mathfrak{p}}))$ are monomorphisms, the choice of $\mathfrak{p}$ does not matter.

It remains to check that this candidate section $s \in \prod_{\mathrm{val}^{\leq r}} \mathcal{F}(R)$ is actually a section of $\mathcal{F}_{\mathrm{val}^{\leq r}}(X)$, i.e., for any X-morphism of valuation rings $\operatorname{Spec}(R') \to \operatorname{Spec}(R)$ with R or R' (or both) of rank r, we want to know that the element s_R restricts to $s_{R'}$. The morphism $\mathcal{F}(R') \to \mathcal{F}(\mathsf{Frac}(R'))$ is injective, so it suffices to consider the case when R' is some field L. Then $R \to L$ factors as

$$R \to R/\mathfrak{p} \to \mathsf{Frac}(R/\mathfrak{p}) \to L,$$

where $\mathfrak{p}$ is the prime $\mathfrak{p} = \ker(R \to L)$. That s_R is sent to $s_{R/\mathfrak{p}}$ comes from the independence of the choice of $\mathfrak{p}$ that we used to construct s. For the same reason, $s_{R/\mathfrak{p}}$ is sent to $s_{\mathsf{Frac}(R/\mathfrak{p})}$. Finally, both $\mathsf{Frac}(R/\mathfrak{p})$ and L, being fields, are of rank zero, and so $s_{\mathsf{Frac}(R/\mathfrak{p})} = t_{\mathsf{Frac}(R/\mathfrak{p})}$ is sent to $s_L = t_L$ since t is already a section of $\mathcal{F}_{\mathrm{val}^{\leq r-1}}$. □

Corollary 3.10. *For any scheme Y, writing h^Y for the presheaf* $\hom_{\mathsf{Sch}_k}(-, Y)$, *the canonical maps are isomorphisms*

$$h^Y_{\mathrm{val}^{\mathrm{big}}} \cong h^Y_{\mathrm{val}} \cong h^Y_{\mathrm{val}^{\leq 1}} \cong h^Y_{\mathrm{shval}}.$$

Proof. The first isomorphism is Lemma 3.6. The second one follows from Proposition 3.9 — note that

$$\operatorname{Spec}(R) = \operatorname{Spec}(R/\mathfrak{p}) \amalg_{\operatorname{Spec}(\mathsf{Frac}(R/\mathfrak{p}))} \operatorname{Spec}(R_{\mathfrak{p}})$$

by [Ferrand 2003, Théorème 5.1]. The last one follows from Proposition 3.8. □

3D. ***The case of differential forms.*** Now we show that the previous material applies to the case of differential forms:

Proposition 3.11. *Let $S = \mathrm{Spec}(k)$ with k a perfect field. The canonical morphisms*

$$\begin{tikzcd}
 & & \Omega^n_{\mathrm{shval}} \\
\Omega^n_{\mathrm{val}^{\mathrm{big}}} \arrow[rru, bend left] \arrow[r, "\text{Lemma 3.6}", "\cong"'] \arrow[rrd, bend right] & \Omega^n_{\mathrm{val}} \arrow[ru, "\text{Proposition 3.8}"', "\cong"] \arrow[rd, "\text{Proposition 3.9}", "\cong"'] & \\
 & & \Omega^n_{\mathrm{val}^{\le 1}}
\end{tikzcd}$$

are isomorphisms of presheaves on $\mathsf{Sch}^{\mathrm{ft}}_k$.

Proof. The presheaf Ω^n on $\mathrm{val}^{\mathrm{big}}$ is an étale sheaf and commutes with direct limits. Hence we may apply Proposition 3.8 in order to show the comparison to $\Omega^n_{\mathrm{shval}}$. For the final isomorphism we want to apply Proposition 3.9. The rest of this section is devoted to checking the necessary cartesian diagram. □

Lemma 3.12. *Let R be a k-valuation ring and $\mathfrak{p}$ a prime ideal. Then the diagram*

$$\begin{tikzcd}
\Omega^1(R) \arrow[r, hook] \arrow[d, two heads] & \Omega^1(R_{\mathfrak{p}}) \arrow[d, two heads] \\
\Omega^1(R/\mathfrak{p}) \arrow[r, hook] & \Omega^1(\mathsf{Frac}(R/\mathfrak{p}))
\end{tikzcd}$$

is cartesian.

Proof. The R-module $\Omega^1(R)$ is flat because it is torsion free over a valuation ring. We tensor the diagram (4) of Lemma 2.14 with the flat R-module $\Omega^1(R)$ and obtain the cartesian diagram

$$\begin{tikzcd}
\Omega^1(R) \arrow[r, hook] \arrow[d] & \Omega^1(R_{\mathfrak{p}}) \arrow[d] \\
\Omega^1(R) \otimes_R R/\mathfrak{p} \arrow[r, hook] & \Omega^1(R_{\mathfrak{p}}) \otimes_{R_{\mathfrak{p}}} R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}
\end{tikzcd}$$

In the next step we use the fundamental exact sequence for differentials of a quotient [Matsumura 1970, Theorem 58, p. 187] and obtain the following diagram with exact columns:

$$\begin{tikzcd}
\mathfrak{p}/\mathfrak{p}^2 \arrow[r, "\text{epi}"] \arrow[d] & \mathfrak{p}R_{\mathfrak{p}}/\mathfrak{p}^2R_{\mathfrak{p}} \arrow[d] \\
\Omega^1(R) \otimes_R R/\mathfrak{p} \arrow[r, hook, "\text{monic}"] \arrow[d, two heads, "\text{epi}"'] & \Omega^1(R_{\mathfrak{p}}) \otimes_{R_{\mathfrak{p}}} R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}} \arrow[d] \\
\Omega^1(R/\mathfrak{p}) \arrow[r, hook] \arrow[d] & \Omega^1(R_{\mathfrak{p}}/\mathfrak{p}R_{\mathfrak{p}}) \arrow[d] \\
0 & 0
\end{tikzcd} \tag{8}$$

The top horizontal arrow is an isomorphism, and in particular a surjection, because $\mathfrak{p} = \mathfrak{p}R_{\mathfrak{p}}$. A small diagram chase now shows that the second square is cartesian. Putting the two diagrams together, we get the claim. $\square$

Lemma 3.13. *Let R be a k-valuation ring and $\mathfrak{p}$ a prime ideal. Then the diagram*

$$\begin{tikzcd} \Omega^n(R) \arrow[r] \arrow[d] & \Omega^n(R_{\mathfrak{p}}) \arrow[d] \\ \Omega^n(R/\mathfrak{p}) \arrow[r] & \Omega^n(\mathsf{Frac}(R/\mathfrak{p})) \end{tikzcd}$$

is cartesian for all $n \geq 0$.

Proof. We have already done the cases $n = 0, 1$. We want to go from $n = 1$ to general n by taking exterior powers. We write $\Omega^1(R)$ as the union of its finitely generated sub-R-modules

$$\Omega^1(R) = \bigcup N.$$

We write $\overline{N}$ for the image of N in $\Omega^1(R/\mathfrak{p})$. Note that

$$\begin{aligned} \Omega^1(R/\mathfrak{p}) &= \bigcup \overline{N}, \\ \Omega^1(R_{\mathfrak{p}}) = \Omega^1(R)_{\mathfrak{p}} &= \bigcup N_{\mathfrak{p}}, \\ \Omega^1(R_{\mathfrak{p}}/\mathfrak{p}) = \Omega^1(R/\mathfrak{p})_{\mathfrak{p}} &= \bigcup \overline{N}_{\mathfrak{p}}. \end{aligned}$$

The module N is a torsion free finitely generated module over the valuation ring R, hence free. The $R_{\mathfrak{p}}$, $R/\mathfrak{p}$, $R_{\mathfrak{p}}/\mathfrak{p}$ modules $N_{\mathfrak{p}}$, $N/\mathfrak{p}$, $N_{\mathfrak{p}}/\mathfrak{p}$ are therefore also free, and the same is true for all the exterior powers. So for all $n \geq 0$ we get the following cartesian diagram:

$$\begin{tikzcd} \bigwedge^n N \arrow[r] \arrow[d] & \bigwedge^n N_{\mathfrak{p}} \arrow[d] \\ \bigwedge^n (N/\mathfrak{p}) \arrow[r] & \bigwedge^n (N_{\mathfrak{p}}/\mathfrak{p}) \end{tikzcd}$$

Note that if the rank of N is one, and $n = 0$, this is just the cartesian diagram from Lemma 2.14. Note also that at this stage we are working with $N/\mathfrak{p}$ and $N_{\mathfrak{p}}/\mathfrak{p}$, instead of $\overline{N}$ and $\overline{N}_{\mathfrak{p}}$.

Passing to the direct limit, we have established as a first step that the diagram

$$\begin{tikzcd} \Omega^n(R) \arrow[r] \arrow[d] & \Omega^n(R_{\mathfrak{p}}) \arrow[d] \\ \Omega^n(R)/\mathfrak{p} \arrow[r] & \Omega^n(R_{\mathfrak{p}})/\mathfrak{p} \end{tikzcd}$$

is cartesian.

Let $\pi : \Omega^n(R)/\mathfrak{p} \to \Omega^n(R/\mathfrak{p})$ and $\theta : \Omega^n(R_\mathfrak{p})/\mathfrak{p} \to \Omega^n(R_\mathfrak{p}/\mathfrak{p}R_\mathfrak{p})$ be the natural maps. We want to show that

$$\begin{tikzcd}
\Omega^n(R)/\mathfrak{p} \arrow[r] \arrow[d] & \Omega^n(R_\mathfrak{p})/\mathfrak{p} \arrow[d] \\
\Omega^n(R/\mathfrak{p}) \arrow[r] & \Omega^n(R_\mathfrak{p}/\mathfrak{p})
\end{tikzcd}$$

is also cartesian. Let $\pi^{-1}\overline{N} \subseteq \Omega^1(R)/\mathfrak{p}$ be the preimage of $\overline{N}$. Similarly, let $\theta^{-1}\overline{N}_\mathfrak{p} \subseteq \Omega^1(R_\mathfrak{p})/\mathfrak{p}$ be the preimage of $\overline{N}_\mathfrak{p}$ by $\theta : \Omega^1(R_\mathfrak{p})/\mathfrak{p} \to \Omega^1(R_\mathfrak{p}/\mathfrak{p})$. In particular, we have the following cube, for which the two side squares are cartesian by definition, the front square is the cartesian square from diagram (8) on page 666, and a diagram chase then shows that the back square is also cartesian. Note that the lower and upper faces are probably not cartesian, but this does not affect the argument.

$$\begin{tikzcd}
\pi^{-1}\overline{N} \arrow[rr, hook] \arrow[dr, hook] \arrow[dd, two heads, "\pi"'] & & \theta^{-1}\overline{N}_\mathfrak{p} \arrow[dr, hook] \arrow[dd, two heads, "\theta" near start] & \\
& \Omega^1(R)/\mathfrak{p} \arrow[rr, hook] \arrow[dd, two heads] & & \Omega^1(R_\mathfrak{p})/\mathfrak{p} \arrow[dd, two heads] \\
\overline{N} \arrow[rr, hook] \arrow[dr, hook] & & \overline{N}_\mathfrak{p} \arrow[dr, hook] & \\
& \Omega^1(R/\mathfrak{p}) \arrow[rr, hook] & & \Omega^1(R_\mathfrak{p}/\mathfrak{p})
\end{tikzcd}$$

We claim that the back square stays cartesian when passing to higher exterior powers. We show this by comparing the kernels of $\bigwedge^n \pi$ and $\bigwedge^n \theta$, cf. the diagram chase of diagram (8) on page 666. More precisely, to show that the higher exterior powers of the back square are cartesian, it suffices to show that $\ker \bigwedge^n \pi \to \ker \bigwedge^n \theta$ is a surjection. We will show that it is an isomorphism.

Note that since the back face is cartesian, and the horizontal morphism is a monomorphism, we have $\ker(\pi) \cong \ker(\theta)$. Let X be this common kernel. The module $\overline{N} \subset \Omega^1(R/\mathfrak{p})$ is torsion free and finitely generated over the valuation ring $R/\mathfrak{p}$, and therefore it is free, and in particular, projective. Hence π admits a splitting σ.

The map $\iota = \sigma \otimes_{R/\mathfrak{p}} R_\mathfrak{p}/\mathfrak{p}$ is then a splitting of θ compatible with σ. In particular, we have compatible decompositions

$$\pi^{-1}\overline{N} = X \oplus \overline{N}, \qquad \theta^{-1}\overline{N}_\mathfrak{p} = X \oplus \overline{N}_\mathfrak{p}.$$

The X's are the same due to the square being cartesian. Hence

$$\bigwedge^n_{R/\mathfrak{p}} \pi^{-1}\overline{N} = \bigoplus_{i=0}^n \bigwedge^i_{R/\mathfrak{p}} X \underset{R/\mathfrak{p}}{\otimes} \bigwedge^{n-i}_{R/\mathfrak{p}} \overline{N},$$

$$\bigwedge^n_{R_\mathfrak{p}/\mathfrak{p}} \theta^{-1}\overline{N}_\mathfrak{p} = \bigoplus_{i=0}^n \bigwedge^i_{R_\mathfrak{p}/\mathfrak{p}} X \underset{R_\mathfrak{p}/\mathfrak{p}}{\otimes} \bigwedge^{n-i}_{R_\mathfrak{p}/\mathfrak{p}} \overline{N}_\mathfrak{p}.$$

The kernels of $\bigwedge^n \pi$ and $\bigwedge^n \pi_F$ are given by the analogous sums but indexed by $i = 1, \dots, n$. Note that X is an $R_\mathfrak{p}/\mathfrak{p} = \mathsf{Frac}(R/\mathfrak{p})$-vector space. Since $R_\mathfrak{p}/\mathfrak{p} = \mathsf{Frac}(R/\mathfrak{p})$, if A, B are $R_\mathfrak{p}/\mathfrak{p}$-vector spaces, then

$$A \otimes_{R/\mathfrak{p}} B = A \otimes_{R_\mathfrak{p}/\mathfrak{p}} B.$$

Hence

$$\bigwedge^i_{R/\mathfrak{p}} X = \bigwedge^i_{R_\mathfrak{p}/\mathfrak{p}} X$$

and finally

$$\Big(\bigwedge^i_S X\Big) \otimes_S \overline{N} = \Big(\bigwedge^i_S X\Big) \otimes_F F \otimes_S \overline{N} = \Big(\bigwedge^i_F X\Big) \otimes_F \overline{N}_\mathfrak{p},$$

where $S = R/\mathfrak{p}$ and $F = \mathsf{Frac}(R/\mathfrak{p}) = R_\mathfrak{p}/\mathfrak{p}$. There are similar formulas for higher exterior powers of $\overline{N}_\mathfrak{p}$ and $\overline{N}$. So we have shown that $\ker \bigwedge^n \pi \to \ker \bigwedge^n \theta$ is an isomorphism as claimed. □

Remark 3.14. This finishes the proof of Proposition 3.11.

4. Descent properties of Ω^n_{val}

Our presheaf of interest, the presheaf Ω^n, is a sheaf for the étale topology. This has far reaching consequences.

Remark 4.1. Let us point out that we have written Ω^n everywhere because this is our main object of study, but everything in this section is valid for any presheaf $\mathcal{F}$ on $\langle \mathrm{val}(S), \mathsf{Sch}^{\mathrm{ft}}_S \rangle$, the full subcategory of S-schemes whose objects are those of $\mathrm{val}(S)$ and $\mathsf{Sch}^{\mathrm{ft}}_S$, satisfying:

(Co) $\mathcal{F}$ commutes with filtered colimits,

(Et) $\mathcal{F}$ satisfies the sheaf condition for the étale topology,

(TF) $\mathcal{F}$ is torsion free in the sense that $\mathcal{F}(R) \to \mathcal{F}(\mathsf{Frac}(R))$ is injective for every valuation ring R.

For example, if Y is any scheme, then $\mathcal{F}(-) = h^Y(-) = \hom(-, Y)$ satisfies these conditions.

Proposition 4.2. *Suppose that $\mathcal{F}$ is a presheaf on* $\mathrm{shval}(X)$. *Then $\mathcal{F}_{\mathrm{shval}}$ is an* eh-*sheaf. Similarly, $\mathcal{F}_{\mathrm{val}}$ is an* rh-*sheaf for any presheaf $\mathcal{F}$ on* $\mathrm{val}(X)$.

Remark 4.3. If we had defined a category hval of henselian valuation rings, we could also have said that $\mathcal{F}_{\mathrm{hval}}$ is a cdh-sheaf for any presheaf $\mathcal{F}$ on $\mathrm{hval}(X)$.

Proof. We only give the proof for the $\mathrm{shval}(X)$, $\mathcal{F}_{\mathrm{shval}}$, eh case, as the same proof works for $\mathrm{val}(X)$, $\mathcal{F}_{\mathrm{val}}$, rh. Let $U \to X$ be an eh-cover. The map $\mathcal{F}_{\mathrm{shval}}(X) \to \mathcal{F}_{\mathrm{shval}}(U)$ is injective because by Proposition 2.19 every morphism $\mathrm{Spec}(R) \to X$ from a strictly henselian valuation ring factors through U.

Now suppose $s \in \mathcal{F}_{\mathrm{shval}}(U)$ satisfies the sheaf condition for the cover $U \to X$. Let $f : \mathrm{Spec}(R) \to X$ be in $\mathrm{shval}(X)$. By choosing a lift $g : \mathrm{Spec}(R) \to U \to X$, we obtain a candidate section $s_g \in \mathcal{F}(\mathrm{Spec}(R))$. We claim that it is independent of the choice of lift. Let g' be a second lift. The pair (g, g') defines a morphism $\mathrm{Spec}(R) \to U \times_X U$. By assumption, s is in the kernel of $\mathcal{F}_{\mathrm{shval}}(U) \xrightarrow{pr_1 - pr_2} \mathcal{F}_{\mathrm{shval}}(U \times_X U)$. In particular,

$s_g = s_{g'} \in \mathcal{F}(\mathrm{Spec}(R))$ in the (g, g')-component. Let $f_1 : \mathrm{Spec}(R_1) \to X$ and $f_2 : \mathrm{Spec}(R_2) \to X$ be in $\mathrm{loc}(X)$ and $\mathrm{Spec}(R_1) \to \mathrm{Spec}(R_2)$ an X-morphism. The choice of a lift $g_2 : \mathrm{Spec}(R_2) \to U$ also induces a lift g_1. The section s_{g_2} restricts to s_{g_1}, hence s_{f_2} restricts to s_{f_1}. Our candidate components define an element of $\mathcal{F}_{\mathrm{val}}(X)$. □

Corollary 4.4. *Let k be a perfect field. The presheaf Ω^n_{val} is an* eh-*sheaf on* $\mathsf{Sch}^{\mathrm{ft}}_k$. *More generally*, $\mathcal{F}_{\mathrm{val}}$ *is an* eh-*sheaf on* $\mathsf{Sch}^{\mathrm{ft}}_S$ *for any presheaf $\mathcal{F}$ on* $\mathrm{val}(S)$ *which commutes with filtered colimits and satisfies the sheaf condition for the étale topology.*

Proof. The proof follows from $\Omega^n_{\mathrm{val}} = \Omega^n_{\mathrm{shval}}$ (see Proposition 3.8). □

This immediately implies the following:

Corollary 4.5. *The map of presheaves $\Omega^n \to \Omega^n_{\mathrm{val}}$ on $\mathsf{Sch}^{\mathrm{ft}}_k$ induces maps of presheaves*

$$\Omega^n_{\mathrm{rh}} \to \Omega^n_{\mathrm{cdh}} \to \Omega^n_{\mathrm{eh}} \to \Omega^n_{\mathrm{val}}. \tag{9}$$

More generally, this is true for any presheaf $\mathcal{F}$ satisfying (Co) *and* (Et) *from Remark 4.1.*

We find it worthwhile to restate the following theorem from [Huber, Kebekus and Kelly 2017]. The original statement is for Ω^n and $S = \mathrm{Spec}(k)$, but one may check directly that the proof works for any Zariski sheaf and the relative Riemann–Zariski space (Definition 2.24) of [Temkin 2011].

Theorem 4.6 [Huber, Kebekus and Kelly 2017, Theorem A.3]. *For any presheaf $\mathcal{F}$ on $\mathsf{Sch}^{\mathrm{ft}}_S$ the following are equivalent*:

(1) cf. [Huber, Kebekus and Kelly 2017, Hypothesis H] *For every integral $X \in \mathsf{Sch}^{\mathrm{ft}}_S$ and $s, t \in \mathcal{F}(X)$ such that $s|_U = t|_U$ for some dense open $U \subseteq X$, there exists a proper birational morphism $X' \to X$ with $s|_{X'} = t|_{X'}$.*

(2) cf. [Huber, Kebekus and Kelly 2017, Hypothesis V] *For any $R \in \mathrm{val}(S)$ with fraction field K the morphism $\mathcal{F}'(R) \to \mathcal{F}'(K)$ is injective. That is, $\mathcal{F}'$ is torsion free* (*Definition 2.22*).

Here $\mathcal{F}'$ is the presheaf sending $R \in \mathrm{val}(S)$ to the colimit $\mathcal{F}'(R) = \varinjlim_{(R \downarrow \mathsf{Sch}^{\mathrm{ft}}_S)} \mathcal{F}(Y)$ over factorisations $\mathrm{Spec}(R) \to Y \to S$ *through $Y \in \mathsf{Sch}^{\mathrm{ft}}_S$. Clearly, $(\Omega^n)' = \Omega^n$.*

More generally, for any presheaf $\mathcal{F}$ satisfying (Co) *and* (Et) *from Remark 4.1, one has $\mathcal{F}' = \mathcal{F}$, and* (2) *is exactly the condition* (TF).

Corollary 4.7. *The maps $\Omega^n_{\mathrm{rh}} \to \Omega^n_{\mathrm{cdh}} \to \Omega^n_{\mathrm{eh}} \to \Omega^n_{\mathrm{val}}$ are injective. More generally, the same is true for any presheaf $\mathcal{F}$ satisfying* (Co), (Et) *and* (TF) *from Remark 4.1.*

Proof. See [Huber, Kebekus and Kelly 2017, Corollary 5.10, Proposition 5.12]. It suffices to show that for any $X \in \mathsf{Sch}^{\mathrm{ft}}_S$ and any $\omega, \omega' \in \Omega^n(X)$ with $\omega|_{\prod_{x \in X} \Omega^n(x)} = \omega'|_{\prod_{x \in X} \Omega^n(x)}$ there is a cdp-morphism $X' \to X$ with $\omega|_{X'} = \omega'|_{X'}$. By noetherian induction, it suffices to show that Theorem 4.6(1) is satisfied. But by Theorem 4.6, this is equivalent to being torsion free. □

In order to prove surjectivity, we need a strong compactness property.

Lemma 4.8. *Let $(Y_i)_{i\in I}$ be a filtered system of nonempty noetherian topological spaces, with transition morphisms $\pi_{ij} : Y_i \to Y_j$. Then there is a system of nonempty irreducible closed subsets $Z_i \subseteq Y_i$ such that $\overline{\pi_{ij}(Z_i)} = Z_j$.*

Proof. To every i we attach a finite set F_i as follows: Let $V_1, \dots, V_n$ be the irreducible components of Y_i. Let F_i be the set of the irreducible components of all multiple intersections $V_{m_1} \cap \dots \cap V_{m_k}$ for all $1 \le k \le n$ and all choices of m_j. We define a transition map $F_i \to F_j$ by mapping an element $W \in F_i$ to the smallest element of F_j containing $\pi_{ij}(W)$. This defines a filtered system of nonempty finite sets $(F_i)_{i\in I}$. Its projective limit is nonempty by [Stacks project, Tag 086J]. Let $(W_i)_{i\in I}$ be an element of the limit.

Now for each $j \in I$, consider the partially ordered set of closures of images $\{\overline{\pi_{ij}(W_i)} \subseteq Y_j : i \le j\}$. We claim that there is an i_j such that $\overline{\pi_{i'j}(W_{i'})} = \overline{\pi_{i_j j}(W_{i_j})}$ for all $i' \le i_j$. Indeed, if not, then we can construct a strictly decreasing sequence of closed subsets of Y_j, contradicting the fact that it is a noetherian topological space. Define $Z_j = \overline{\pi_{i_j j}(W_{i_j})} \subseteq Y_j$ for such an i_j. It follows from our definitions of the Z_j that for every $i, j \in I$ with $i \le j$, we have $\overline{\pi_{ij}(Z_i)} = Z_j$. □

Let X be integral and $Y \to X$ proper birational. Fix $\omega, \omega' \in \Omega^n_{\mathrm{val}}(X)$ (or in $\mathcal{F}_{\mathrm{val}}(X)$ if we are using an $\mathcal{F}$ as in Remark 4.1) and define $Y^{\omega\neq\omega'} \subset Y$ to be the subset of points $y \in Y$ for which $\omega_y \neq \omega'_y$. We view it as a topological space with the induced topology.

Lemma 4.9. *The topological space $Y^{\omega\neq\omega'}$ is noetherian. In addition, every irreducible closed subset of $Y^{\omega\neq\omega'}$ has a unique generic point.*

Proof. The topological space of Y is noetherian. Subspaces of noetherian topological spaces with the induced topology are noetherian (easy exercise), and hence $Y^{\omega\neq\omega'}$ is noetherian. Let $Z^{\omega\neq\omega'} \subseteq Y^{\omega\neq\omega'}$ be irreducible. Note its closure $Z \subseteq Y$ is also irreducible. Let η be the generic point of Z in the scheme Y. Assume $\eta \notin Z^{\omega\neq\omega'}$. By definition this means $\omega_\eta = \omega'_\eta$. Hence, since Ω^n commutes with filtered colimits, $\omega = \omega'$ also on some open dense $U \subseteq Z$ of the reduction of Z. As $Z^{\omega\neq\omega'}$ is dense in Z, the intersection $U \cap Z^{\omega\neq\omega'}$ is nonempty, implying the existence of a point $y \in Y^{\omega\neq\omega'}$ for which $\omega_y = \omega'_y$, and contradicting the definition of $Y^{\omega\neq\omega'}$. Hence $\eta \in Z^{\omega\neq\omega'}$ and we have found a generic point. Uniqueness follows easily from uniqueness of generic points in Y. □

We are implicitly using the Riemann–Zariski space of X (see Definition 2.24) in the following proof.

Lemma 4.10. *Let X be integral and $\omega, \omega' \in \Omega^n_{\mathrm{val}}(X)$ such that $\omega_{k(X)} = \omega'_{k(X)}$. Then there exists a proper birational $Y \to X$ such that $\omega|_Y = \omega'|_Y \in \Omega^n_{\mathrm{val}}(Y)$. More generally, this is true for a presheaf $\mathcal{F}$ satisfying* (Co), (Et) *and* (TF) *from Remark 4.1.*

Proof. Suppose the contrary — that $Y^{\omega\neq\omega'}$ is nonempty for all proper birational maps $Y \to X$. The Y's and hence also their subspaces $Y^{\omega\neq\omega'}$ form a filtered system of noetherian topological spaces. By Lemma 4.8, if all $Y^{\omega\neq\omega'}$ are nonempty, then there exists a system $Z_Y \subseteq Y^{\omega\neq\omega'}$ of irreducible subspaces of the $Y^{\omega\neq\omega'}$, such that the map $Z_{Y'} \to Z_Y$ induced by any birational $Y' \to Y \to X$ is dominant. Their generic points (which exist and are unique by Lemma 4.9) define a point $y = (y_i) \in \varprojlim Y$ for which

$\omega_{k(y_i)} \neq \omega'_{k(y_i)}$ for all y_i. But Ω^n commutes with cofiltered limits of rings, so $\omega|_{k(y)} \neq \omega'|_{k(y)}$, where $k(y)$ is the residue field of the valuation ring $R = \varinjlim \mathcal{O}_{Y_i,y_i}$ in the locally ringed space $\varprojlim Y$. But R is an X-valuation ring of $k(X)$, so by torsion freeness we must also have $\omega|_{k(X)} \neq \omega'|_{k(X)}$, contradicting our assumption that $\omega|_{k(X)} = \omega'|_{k(X)}$. □

Proposition 4.11. *The map $\Omega^n_{\mathrm{rh}} \to \Omega^n_{\mathrm{val}}$ is surjective. More generally, $\mathcal{F}_{\mathrm{rh}} \to \mathcal{F}_{\mathrm{val}}$ is surjective for any presheaf $\mathcal{F}$ satisfying* (Co), (Et) *and* (TF) *from Remark 4.1.*

Proof. As both are rh-sheaves, we may work rh-locally. In particular, without loss of generality X is integral with function field K. Choose $\omega = (\omega_R)_R \in \Omega^n_{\mathrm{val}}(X)$.

We start with an X-valuation ring $R \subset K$, i.e., $R \in \mathsf{RZ}(X)$. The form ω_R is already defined on some ring A_R of finite type over X. Let $\tilde{\omega}_R$ be this class. For all R' in the Zariski-open subset of $\mathsf{RZ}(X)$ defined by A_R (i.e., $A_R \subset R'$), we have $\omega_{R'} = \tilde{\omega}_R$ because $\Omega^n(R')$ is torsion free and the equality holds in $\Omega^n(K)$.

As $\mathsf{RZ}(X)$ is quasicompact, we may cover it by finitely many of these and obtain a finite set of rings $A_i \subset K$ which are of finite type over X. On each of these we are given a differential form $\omega_{A_i} \in \Omega^n(A_i)$ inducing all ω_R for $R \in \mathsf{RZ}(X)$ containing A_i. The forms $\omega|_{A_i}$ and ω_{A_i} in $\Omega^n_{\mathrm{val}}(\mathrm{Spec}(A_i))$ agree in $\Omega^n_{\mathrm{val}}(K)$. By Lemma 4.10 there exists a blowup $W_i \to \mathrm{Spec}(A_i)$ such that $\omega_{A_i}|_{W_i} = \omega|_{W_i}$. Hence, $\omega|_{W_i}$ is represented by a class ω_{W_i} in $\Omega^n_{\mathrm{rh}}(W_i) \subset \Omega^n_{\mathrm{val}}(W_i)$.

By Nagata compactification, there is a factorisation

$$W_i \to \overline{W}_i \to X,$$

where the first map is a dense open immersion and the second is proper and birational. Let V be the closure of $\mathrm{Spec}(K)$ in $\overline{W}_1 \times_X \cdots \times_X \overline{W}_n$. The canonical morphism $V \to X$ is proper and birational. Every point of V is dominated by a valuation ring of K [EGA II 1961, Proposition 7.1.7]. There is necessarily at least one of the A_i contained in it, and so the base changes $V_i = W_i \times_{\overline{W}_i} V$ form an open cover of V. Let $\omega_{V_i} \in \Omega^n_{\mathrm{rh}}(V_i)$ be the restriction of ω_{W_i}.

We have $\omega_{V_i} = \omega_{V_j}$ in $\Omega^n_{\mathrm{val}}(V_i \cap V_j)$ (and hence $\Omega^n_{\mathrm{rh}}(V_i \cap V_j)$ by Corollary 4.7) because both agree with the restriction of ω. By Zariski-descent, the ω_{V_i} glue to a global differential form $\omega_V \in \Omega^n_{\mathrm{rh}}(V)$ representing $\omega|_V$ in $\Omega^n_{\mathrm{rh}}(V) \subset \Omega^n_{\mathrm{val}}(V)$.

Now let $Z \subset X$ be the exceptional locus of $V \to X$ and let $E \subset V$ be its preimage. By induction on the dimension there is a class ω_Z in $\Omega^n_{\mathrm{rh}}(Z)$ mapping to $\omega|_Z \in \Omega^n_{\mathrm{val}}(Z)$. We know that ω_Z and ω_Y agree on E because $\Omega^n_{\mathrm{rh}}(E) \to \Omega^n_{\mathrm{val}}(E)$ is injective and both represent $\omega|_E$. By rh-descent this gives a class $\omega_X \in \Omega^n_{\mathrm{val}}(X)$ mapping to ω. □

Theorem 4.12. *The canonical morphisms of presheaves on* $\mathsf{Sch}^{\mathrm{ft}}_k$

$$\Omega^n_{\mathrm{rh}} \to \Omega^n_{\mathrm{cdh}} \to \Omega^n_{\mathrm{eh}} \to \Omega^n_{\mathrm{val}}$$

are isomorphisms. More generally, these are isomorphisms for any presheaf $\mathcal{F}$ satisfying (Co), (Et) *and*

(TF) *from Remark 4.1. Moreover*,

$$\Omega^n(X) = \Omega^n_{\mathrm{val}}(X)$$

when $S = \operatorname{Spec}(k)$ *for* X *smooth.*

Proof. Corollary 4.7 says that the maps are injective. As the composition is surjective, they are all isomorphisms. The smooth case is then a consequence of [Huber, Kebekus and Kelly 2017, Theorem 5.11], which says that $\Omega(X) \cong \Omega^n_{\mathrm{rh}}(X)$ for smooth X. □

Remark 4.13. This still leaves open whether $\Omega^n_{\mathrm{eh}} \to \Omega^n_{\mathrm{dvr}}$ is an isomorphism. Weak resolution of singularities would imply that this is an isomorphism in general; see [Huber, Kebekus and Kelly 2017, Proposition 5.13].

5. Quasicoherence

The sheaves $\Omega^n_{\mathrm{eh}}|_X$ are obviously sheaves of $\mathcal{O}_X$-modules. We want to show that they are coherent. The main step is actually quasicoherence.

5A. *Quasicoherence.*

Lemma 5.1. *Let A be a finite type k-algebra, $f \in A$ not nilpotent, and $\omega \in \Omega^n_{\mathrm{val}}(A)$ be such that $\omega|_{A_f} = 0$. Then $f\omega = 0$. Moreover, the map*

$$\Omega^n_{\mathrm{val}}(A)_f \to \Omega^n_{\mathrm{val}}(A_f)$$

is injective.

Note this would follow directly from quasicoherence of $\Omega^n_{\mathrm{val}}|_{\operatorname{Spec}(A)}$. We want to prove it directly in order to show quasicoherence down the line.

Proof. By torsion freeness it suffices to show that $f\omega_x = 0$ for all residue fields $\kappa(x)$ of points $x \in \operatorname{Spec}(A)$ (Lemma 3.7). It suffices to consider the case $f|_{\kappa(x)} \neq 0$. But then $A \to \kappa(x)$ factors through A_f, and the claim follows from the assumption $\omega|_{A_f} = 0$.

We now turn to injectivity. Let ω/f^N be in the kernel of

$$\Omega^n_{\mathrm{val}}(A)_f \to \Omega^n_{\mathrm{val}}(\operatorname{Spec} A_f).$$

This means that $\omega_R/f^N = 0$ for every $R \in \mathrm{val}(A_f)$. As f is invertible in this ring, this implies $\omega_R = 0$. That is, ω satisfies the assumption of the first assertion. Hence $f\omega = 0$ in $\Omega^n_{\mathrm{val}}(A)$ and this implies $\omega/f^N = 0$ in the localization. □

In order to proceed, we need a lemma from algebraic geometry.

Lemma 5.2. *Let $U \subset X$ be an open immersion of integral schemes, and let $V \to U$ be a* cdp-*morphism with integral connected components. Then there is a cartesian diagram*

$$\begin{array}{ccc} V & \xrightarrow{j} & Y \\ \downarrow & & \downarrow{\scriptstyle p} \\ U & \longrightarrow & X \end{array}$$

with p a cdp-*morphism and j an open immersion.*

Proof. Let V' be an irreducible component of V. We factor $V' \to U \to X$ as a (dense) open immersion $V' \to Y'$ followed by a proper map. It is easy to see that $V' = Y' \times_X U$ because the map $V' \to U \times_X Y'$ is both proper and a (dense) open immersion. We define Y as the disjoint union of all Y' and $X \setminus U$. □

Proposition 5.3. *Let X be an integral k-scheme of finite type. Then the restriction to the small Zariski site $\Omega^n_{\mathrm{rh}}|_{X_{\mathrm{Zar}}}$ is quasicoherent.*

Proof. It is sufficient to show that for every integral ring A and $f \in A$, the canonical morphism $\Omega^n_{\mathrm{rh}}(A)_f \to \Omega^n_{\mathrm{rh}}(A_f)$ is an isomorphism. By the last lemma and because $\Omega^n_{\mathrm{rh}} = \Omega^n_{\mathrm{val}}$, the map is injective.

To show that it is surjective, it suffices to check that for every

$$\omega \in \Omega^n_{\mathrm{rh}}(A_f)$$

there is N such that $f^N\omega$ lifts to $\Omega^n_{\mathrm{rh}}(A)$. We put $X = \mathrm{Spec}(A)$, $U = \mathrm{Spec}(A_f)$. There is an rh-cover $\{V_i \to U\}_{i=1}^m$ such that ω is represented by an algebraic differential form on $\coprod V_i$. The strategy is to show that, up to multiplication by f, the form ω is actually representable on a cdp-morphism of U, and then descend it to U. Lemma 5.2 is key.

We can choose the cover in the form

$$V_i \to V \to U \tag{10}$$

with $\pi : V \to U$ a cdp-morphism, V reduced and $V_i \to V$ open immersions. Moreover, we may assume that V is the disjoint union of its irreducible components and that they are birational over their image in U (because we will want to apply Lemma 5.2).

We now construct the following cartesian squares whose vertical morphisms are proper envelopes, and horizontal ones are open immersions:

$$\begin{tikzcd}
\widetilde{W}_i \cap \widetilde{W}_j \arrow[r] \arrow[d] & \widetilde{W}_i \arrow[r] \arrow[d] & \widetilde{W} \arrow[d] \\
\underset{\omega_i - \omega_j = 0 \in \Omega^n(W_{ij})}{W_{ij}} \arrow[r] \arrow[d] & {} \arrow[r] \arrow[d] & \widetilde{W}_{ij} \arrow[d, "p_{ij}"] \\
V_i \cap V_j \arrow[r] & \underset{\omega_i \in \Omega^n(V_i)}{V_i} \arrow[r] & V
\end{tikzcd}$$

Let $\omega_i \in \Omega^n(V_i)$ be the representing form. Since ω_i came from an element $\omega \in \Omega^n_{\mathrm{rh}}(U)$, the differences $\omega_i - \omega_j$ vanish in $\Omega^n_{\mathrm{rh}}(V_i \times_U V_j)$ by the exact sequence

$$0 \to \Omega^n_{\mathrm{rh}}(U) \to \bigoplus \Omega^n_{\mathrm{rh}}(V_i) \to \bigoplus \Omega^n_{\mathrm{rh}}(V_i \times_U V_j),$$

and hence vanish in $\Omega^n_{\mathrm{rh}}(V_i \times_V V_j) = \Omega^n_{\mathrm{rh}}(V_i \cap V_j)$. Consequently, there is an rh-cover of $V_i \cap V_j$ on which $\omega_i - \omega_j$ vanishes as a section of the presheaf Ω^n. We can assume that this cover is again of the form (10). Since Ω^n is Zariski separated, we find that there is a cdp-morphism $W_{ij} \to V_i \cap V_j$ such that

$\omega_i - \omega_j$ vanishes on W_{ij}. By Lemma 5.2 there is a commutative diagram

$$\begin{tikzcd} W_{ij} \arrow[r, "j"] \arrow[d] & \widetilde{W}_{ij} \arrow[d, "p_{ij}"] \\ V_i \cap V_j \arrow[r] & V \end{tikzcd}$$

with an open immersion j and a proper envelope p_{ij}. Let $\widetilde{W} = \prod_V \widetilde{W}_{ij}$ be the fibre product of all $\widetilde{W}_{ij}$ over V. Hence $\widetilde{W} \to V$ is a proper envelope which factors through all p_{ij}. Let $\widetilde{W}_i$ be the preimage of V_i in $\widetilde{W}$. Now consider the above big diagram. The differences of the restrictions $\omega_i - \omega_j$ vanish in $\Omega^n(\widetilde{W}_i \cap \widetilde{W}_j)$, and $\{\widetilde{W}_i \to \widetilde{W}\}$, being the pullback of the Zariski cover $\{V_i \to V\}$ is also a Zariski cover. Hence, we can lift the restrictions $\tilde{\omega}_i \in \Omega^n(\widetilde{W}_i)$ to a section

$$\omega_{\widetilde{W}} \in \Omega^n(\widetilde{W}).$$

We find that $\omega \in \Omega^n_{\mathrm{rh}}(U)$ and $\omega_{\widetilde{W}} \in \Omega^n(\widetilde{W})$ agree in $\Omega^n_{\mathrm{rh}}(\widetilde{W})$ since $\{\widetilde{W}_i \to \widetilde{W}\}$ is a Zariski cover, and they agree on $\widetilde{W}_i$. In other words, at this point, we have shown that $\omega|_{\widetilde{W}}$ is in the image of $\Omega^n \to \Omega^n_{\mathrm{rh}}$.

Again by Lemma 5.2 there is a cartesian diagram

$$\begin{tikzcd} \widetilde{W} \arrow[r, "j"] \arrow[d] & Y \arrow[d, "p"] \\ U \arrow[r] & X \end{tikzcd}$$

with j an open immersion and p a proper envelope. The open subset U in X is by definition the complement of $V(f)$, hence the same is true for $\widetilde{W}$ in Y. As Ω^n_Y is coherent, this implies that there is N such that $f^N \omega_{\widetilde{W}}$ extends to Y.

Let ω_Y be an extension of $f^N \omega_{\widetilde{W}}$ to $\Omega^n(Y)$. Let $Z = V(f) \subset X$ and E its preimage in Y. Consider the exact sequence

$$0 \to \Omega^n_{\mathrm{rh}}(X) \to \Omega^n_{\mathrm{rh}}(Y) \oplus \Omega^n_{\mathrm{rh}}(Z) \to \Omega^n_{\mathrm{rh}}(E).$$

The class $(f\omega_Y, 0)$ maps to $f\omega_Y|_E$. As $f = 0$ in all of E, this equals zero. Hence there is a class $\omega'_X \in \Omega^n_{\mathrm{rh}}(X)$ such that $\omega'_X|_{\widetilde{W}} = f^{N+1}\omega_{\widetilde{W}}$ in $\Omega^n_{\mathrm{rh}}(\widetilde{W})$. Recall that two paragraphs ago we mentioned that $\omega|_{\widetilde{W}} = \omega_{\widetilde{W}}$ in $\Omega^n_{\mathrm{rh}}(\widetilde{W})$. So in fact, we know that $\omega'_X|_{\widetilde{W}} = f^{N+1}\omega|_{\widetilde{W}}$ in $\Omega^n_{\mathrm{rh}}(\widetilde{W})$. But $\widetilde{W} \to U$ is a cdp-morphism, so it follows that $\omega'_X|_U = f^{N+1}\omega$ in $\Omega^n_{\mathrm{rh}}(U)$. □

5B. *Coherence.*

Theorem 5.4. *Let X be of finite type over k. Then $\Omega^n_{\mathrm{rh}}|_{X_{\mathrm{Zar}}}$ is coherent.*

Proof. We write $\Omega^n_{\mathrm{rh}}|_X = \Omega^n_{\mathrm{rh}}|_{X_{\mathrm{Zar}}}$ for briefness. Let $i : X_{\mathrm{red}} \to X$ be the reduction. We have $i_*(\Omega^n_{\mathrm{rh}}|_{X_{\mathrm{red}}}) = \Omega^n_{\mathrm{rh}}|_X$. Hence we may assume that X is reduced. Let $X = X_1 \cup \cdots \cup X_N$ be the decomposition into reduced irreducible components. Let $i_j : X_j \to X$ and $i_{jl} : X_j \cap X_l \to X$ be the closed immersions. By rh-descent, we have an exact sequence of sheaves of $\mathcal{O}_X$-modules

$$0 \to \Omega^n_{\mathrm{rh}}|_X \to \bigoplus_{j=1}^N i_{j*}(\Omega^n_{\mathrm{rh}}|_{X_j}) \to \bigoplus_{j,l=1}^N i_{jl*}(\Omega^n_{\mathrm{rh}}|_{X_j\cap X_l}).$$

By induction on the dimension it suffices to consider the irreducible case.

Let X be integral with function field K. Let $\widetilde{X}$ be its normalisation. The map $\pi : \widetilde{X} \to X$ is an isomorphism outside some closed proper subset $Z \subset X$. Let E be its preimage in $\widetilde{X}$. From the blowup sequence we obtain

$$0 \to \Omega^n_{\mathrm{rh}}|_X \to \pi_*(\Omega^n_{\mathrm{rh}}|_{\widetilde{X}}) \oplus i_{Z*}(\Omega^n_{\mathrm{rh}}|_Z) \to \pi_* i_{E*}(\Omega^n_{\mathrm{rh}}|_E).$$

Hence by induction on the dimension, we may assume that X is normal. We have shown in Proposition 5.3 that the sheaf $\Omega^n_{\mathrm{rh}}|_X$ is quasicoherent in the integral case. Let $j : X^{\mathrm{sm}} \to X$ be the inclusion of the smooth locus with closed complement Z. It is of codimension at least 2. Hence $j_*\Omega^n_{X^{\mathrm{sm}}}$ is coherent. By Theorem 4.12, $\Omega^n_{X^{\mathrm{sm}}} \cong \Omega^n_{\mathrm{rh}}|_{X^{\mathrm{sm}}}$ so we have a map $\Omega^n_{\mathrm{rh}}|_X \to j_*\Omega^n_{X^{\mathrm{sm}}}$. Its image is coherent because it is a quasicoherent subsheaf of a coherent sheaf. Its kernel $\mathcal{K}$ is also quasicoherent. We claim that it is a subsheaf of $i_{Z*}(\Omega^n_{\mathrm{rh}}|_Z)$. It suffices to prove that the canonical composition $\mathcal{K}(U) \to \Omega^n_{\mathrm{rh}}|_X(U) \to i_{Z*}(\Omega^n_{\mathrm{rh}}|_Z)(U)$ is a monomorphism for all open $U \subseteq X$. Replacing X with U, it suffices to consider the case $U = X$. Then this morphism is canonically identified with the morphism $\ker(\Omega^n_{\mathrm{val}}(X) \to \Omega^n_{\mathrm{val}}(X^{\mathrm{sm}})) \to \Omega^n_{\mathrm{val}}(Z)$, which is injective because $\Omega^n_{\mathrm{val}}(X) \to \Omega^n_{\mathrm{val}}(X^{\mathrm{sm}}) \times \Omega^n_{\mathrm{val}}(Z)$ is injective (Lemma 3.7). By induction on the dimension we can assume that $i_{Z*}\Omega^n_{\mathrm{rh}}|_Z$ is also coherent, so the kernel $\mathcal{K}$ is coherent as well. As a quasicoherent extension of two coherent sheaves the sheaf $\Omega^n_{\mathrm{rh}}|_X$ is coherent. ☐

5C. *Torsion.* We return to the question of torsion forms. As in [Huber, Kebekus and Kelly 2017], we denote by $\operatorname{tor}\Omega^n_{\mathrm{eh}}(X)$ the submodule of torsion sections, i.e., those vanishing on some dense open subset. There is an obvious source of torsion classes: Let $f : Y \to X$ be proper birational with centre $Z \subset X$ and preimage $E \subset Y$. Any $\omega \in \ker(\Omega^n_{\mathrm{eh}}(Z) \to \Omega^n_{\mathrm{eh}}(E))$ gives rise to a torsion class on X by the blowup sequence. By [Huber, Kebekus and Kelly 2017, Example 5.15] this kernel can indeed be nonzero. We have established in Lemma 4.10 that all torsion classes arise in this way.

Proposition 5.5. *Let X be of finite type over k.*

(1) *The presheaf $\mathcal{T}_X : U \mapsto \operatorname{tor}\Omega^n_{\mathrm{rh}}(U)$ is a coherent sheaf of $\mathcal{O}_X$-modules on X_{Zar}.*

(2) *There is a proper birational morphism $f : Y \to X$ such that*

$$\mathcal{T}_X = \ker(\Omega^n_{\mathrm{rh}}|_{X_{\mathrm{Zar}}} \to f_*\Omega^n_{\mathrm{rh}}|_{Y_{\mathrm{Zar}}}).$$

Proof. By reductions similar to those in the first paragraph of the proof of Theorem 5.4 it suffices to show coherence for X integral, and since quasicoherent subsheaves of coherent sheaves are coherent, it suffices to show quasicoherence for X integral.

Let $X = \operatorname{Spec}(A)$, $U = \operatorname{Spec}(A_f)$. We have to show that $\mathcal{T}_X(U) = \mathcal{T}_X(X)_f$. Let

$$\omega \in \mathcal{T}_X(U) \subset \Omega^n_{\mathrm{rh}}|_{X_{\mathrm{Zar}}}(U) = (\Omega^n_{\mathrm{rh}}(X))_f.$$

Then ω is of the form $\tilde{\omega}/f^N$ with $\tilde{\omega} \in \Omega^n_{\mathrm{rh}}(X)$. By assumption, ω vanishes at the generic point of U, which is equal to the generic point of X. Hence the same is true for $\tilde{\omega}$. This finishes the proof of coherence.

Any form vanishing on a blowup is torsion, i.e., $\ker(\Omega^n_{\mathrm{rh}}|_{X_{\mathrm{Zar}}} \to f_*\Omega^n_{\mathrm{rh}}|_{Y_{\mathrm{Zar}}}) \subseteq \mathcal{T}_X$ for any proper birational $f : Y \to X$, and so our job is to find a Y for which this inclusion is surjective. By Theorem 4.12, $\Omega^n_{\mathrm{rh}} = \Omega^n_{\mathrm{val}}$, and so by Lemma 4.10, for any $\omega \in \mathcal{T}_X(X)$, there is a proper birational morphism $Y_\omega \to X$ for which $\omega \in \ker(\Omega^n_{\mathrm{rh}}(X) \to \Omega^n_{\mathrm{rh}}(Y))$.

If X is affine, the $\mathcal{O}(X)$-module $\mathcal{T}_X(X)$ is finitely generated. We can find a proper birational $Y \to X$ killing the generators and hence all of $\mathcal{T}_X(X)$, and by coherence even all of $\mathcal{T}_X$. If X is not affine, let $X = U_1 \cup \cdots \cup U_m$ be an affine cover, and $Y_i \to U_i$ proper birational morphisms killing all of $\mathcal{T}_{U_i}$. By Nagata compactification, there is a proper birational morphism $\overline{Y_i} \to X$ such that $Y_i = U_i \times_X \overline{Y_i}$. Let V be the closure of $\operatorname{Spec}(k(X))$ in $\overline{Y_1} \times_X \cdots \times_X \overline{Y_m}$. It is equipped with the open cover $\{U_i \times_X V \to V\}$. Moreover, each $U_i \times_X V \to V \to X$ factors through Y_i, so $\mathcal{T}_X = \bigcup_{i=1}^m \ker(\Omega^n_{\mathrm{rh}}|_{X_{\mathrm{Zar}}} \to f_*\Omega^n_{\mathrm{rh}}|_{(U_i \times_X V)_{\mathrm{Zar}}})$. Then by Zariski descent we deduce that $\mathcal{T}_X = \ker(\Omega^n_{\mathrm{rh}}|_{X_{\mathrm{Zar}}} \to f_*\Omega^n_{\mathrm{rh}}|_{(V)_{\mathrm{Zar}}})$. ☐

It now becomes an interesting question to understand whether a given X admits such a blowup $Y \to X$ such that there is a point ξ in the centre over which all residue fields of Y_ξ are inseparable over $\kappa(\xi)$. This does not happen in the smooth case: *any* blowup of a regular scheme is completely decomposed (unconditionally) [Huber, Kebekus and Kelly 2017, Proposition 2.12]. In the example in [Huber, Kebekus and Kelly 2017, Example 3.6] the point ξ had codimension 1 and Y was the normalisation. It induced a purely inseparable field extension of $k(\xi)$.

One might wonder if assuming X is normal is enough to avoid this pathology. Let us show that it is not.

Example 5.6. Here we give an example of a normal variety X over a perfect field k of positive characteristic p, a point $\xi \in X$ and a blowup $Y \to X$ such that for every point in the fibre $Y_\xi \to \xi$ the residue field extension is inseparable.

In particular, for every X-valuation ring R of $k(X)$ sending its special point to ξ, the field extension $k(\xi) \subseteq R/m$ is inseparable.

Our variety is

$$X = \operatorname{Spec}\left(\frac{k[s,t,x,y,z]}{z^p - sx^p - ty^p}\right),$$

from [Suslin and Voevodsky 2000b, Example 3.5.10]. Let Y be the blowup of this variety at the ideal (x, y, z). The blowup Y admits an open affine covering by affine schemes with rings

$$\frac{k\left[s,t,\frac{x}{z},\frac{y}{z},z\right]}{1 - s\left(\frac{x}{z}\right)^p - t\left(\frac{y}{z}\right)^p}, \qquad \frac{k\left[s,t,\frac{x}{y},y,\frac{z}{y}\right]}{\left(\frac{z}{y}\right)^p - s\left(\frac{x}{y}\right)^p - t}, \qquad \frac{k\left[s,t,x,\frac{y}{x},\frac{z}{x}\right]}{\left(\frac{z}{x}\right)^p - s - t\left(\frac{y}{x}\right)^p},$$

and the intersections of these open affines with the exceptional fibre E are

$$\frac{k\big[s,t,\frac{x}{z},\frac{y}{z}\big]}{1-s\big(\frac{x}{z}\big)^p-t\big(\frac{y}{z}\big)^p},\qquad \frac{k\big[s,t,\frac{x}{y},\frac{z}{y}\big]}{\big(\frac{z}{y}\big)^p-s\big(\frac{x}{y}\big)^p-t},\qquad \frac{k\big[s,t,\frac{y}{x},\frac{z}{x}\big]}{\big(\frac{z}{x}\big)^p-s-t\big(\frac{y}{x}\big)^p},$$

respectively, all lying over the singular locus $V(x,y,z)=\operatorname{Spec}(k[s,t])\subseteq X$. Our point ξ is the generic point $\xi=\operatorname{Spec}(k(s,t))$. Every point of the fibre E_ξ has residue field an inseparable extension of $k(\xi)$: Consider for example the fibre of the rightmost affine for concreteness (all three fibres are isomorphic up to an automorphism of $k(s,t)$). It factors as

$$\operatorname{Spec}\left(\frac{k(s,t)\big[\frac{y}{x},\frac{z}{x}\big]}{\big(\frac{z}{x}\big)^p-s-t\big(\frac{y}{x}\big)^p}\right)\to\operatorname{Spec}\big(k(s,t)\big[\tfrac{y}{x}\big]\big)\to\operatorname{Spec}(k(s,t)).$$

Any residue field $k(\zeta)$ of a point ζ of the leftmost affine scheme is a finite field extension of the residue field $k(\theta)$ of a point $\theta\in\mathbb{A}^1_{k(s,t)}$ in the middle. This latter $k(\theta)$ is generated over $k(s,t)$ by the image of $\frac{y}{x}$. This finite field extension $k(\zeta)/k(\theta)$ is purely inseparable so long as $s+t\big(\frac{y}{x}\big)^p$ is nonzero in $k(\theta)$. If $k(\theta)=k\big(s,t,\frac{y}{x}\big)$, then clearly this is the case. If $k(\theta)=k(s,t)\big[\frac{y}{x}\big]/f\big(\frac{y}{x}\big)$ for some irreducible polynomial $f\big(\frac{y}{x}\big)\in k(s,t)\big[\frac{y}{x}\big]$, then $s+t\big(\frac{y}{x}\big)^p$ is nonzero if and only if $\big(\frac{y}{x}\big)^p=-\frac{s}{t}\mod f\big(\frac{y}{x}\big)$. But since $\big(\frac{y}{x}\big)^p+\frac{s}{t}$ is irreducible in $k(s,t)\big[\frac{y}{x}\big]$, this would imply $f\big(\frac{y}{x}\big)=\big(\frac{y}{x}\big)^p+\frac{s}{t}$, in which case the subextension $k(\zeta)/k(\theta)/k(s,t)$ is already purely inseparable.

Now the blowup $Y\to X$ is birational and proper, and so any X-valuation ring R of $k(X)$ is uniquely a Y-valuation ring of $k(Y)$, and if the special point of R is sent to ξ, then the lift sends this special point to some point $\eta\in E_\xi$. But any field extension which contains an inseparable field extension is inseparable, and $k(\xi)\to R/\mathfrak{m}$ contains $k(\xi)\to k(\eta)$, hence, $k(\xi)\to R/\mathfrak{m}$ is inseparable.

Remark 5.7. Let us also observe that the rightmost open affine contains the point $\xi'=V\big(\frac{y}{x}\big)$ of E_ξ with residue field $k(s,t)\big[\frac{z}{x}\big]/\big(\big(\frac{z}{x}\big)^p-s\big)\cong k(s^{1/p},t)$. In particular, we have produced a blowup $Y\to X$, a point $\xi\in X$, and a point $\xi'\in Y$ over it for which $\Omega^1(k(\xi))\to\Omega^1(k(\xi'))$ is neither injective nor zero.

5D. *The étale case.* We recall:

Definition 5.8. Let $X\in\mathsf{Sch}^{\mathrm{ft}}_k$. A presheaf $\mathcal{F}$ of $\mathcal{O}$-modules on the small étale site X_{et} is called *coherent* if for all X' étale over X, the sheaf $\mathcal{F}|_{X'_{\mathrm{Zar}}}$ is a coherent sheaf for the Zariski-topology and, in addition, for all $\pi:X_1\to X_2$ étale, the natural map

$$\big(\pi^*\mathcal{F}|_{X_{2,\mathrm{Zar}}}\big)\otimes_{\pi^*\mathcal{O}_{X_2}}\mathcal{O}_{X_1}\to\mathcal{F}|_{X_{1,\mathrm{Zar}}}$$

is an isomorphism.

The left-hand side is nothing but the pullback in the category of quasicoherent sheaves. We reserve the notation π^* for the pullback of abelian sheaves.

Proposition 5.9. *For all $X\in\mathsf{Sch}^{\mathrm{ft}}_k$, the sheaf $\Omega^n_{\mathrm{eh}}|_{X_{\mathrm{et}}}$ is coherent.*

Proof. We already know coherence for the Zariski-topology. By replacing X_2 by X, it suffices to check the condition for all étale morphisms $X'\to X$. Both sides are sheaves for the Zariski-topology; hence it suffices to consider the case where X and X' are both affine and $\pi : X' \to X$ is surjective. (Indeed, first assume X and X' affine, then cover the image of X' by affines U_i and replace X by U_i and X' by the preimage of U_i.) We fix such a π.

We need to show that

$$\Omega^n_{\mathrm{eh}}(X) \otimes_{\mathcal{O}(X)} \mathcal{O}(X') \to \Omega^n_{\mathrm{eh}}(X') \tag{11}$$

is an isomorphism. Note that the Ω^n analogue of this assertion holds.

Step 1: Consider the morphism of presheaves on the category of separated schemes of finite type over X

$$Y \mapsto \big(\Omega^n(Y) \otimes_{\mathcal{O}(X)} \mathcal{O}(X') \to \Omega^n(Y \times_X X')\big). \tag{12}$$

We claim that it is an isomorphism. Indeed, both sides are sheaves for the Zariski-topology (the left-hand side because $\mathcal{O}(X) \to \mathcal{O}(X')$ is flat). Hence it suffices to consider the case where Y is affine. Let $Y' = Y \times_X X'$. It is also affine. Then the map (12) identifies as

$$\Omega^n(Y) \otimes_{\mathcal{O}(X)} \mathcal{O}(X') = \Omega^n(Y) \otimes_{\mathcal{O}(Y)} \mathcal{O}(Y') \to \Omega^n(Y'),$$

and hence it is an isomorphism.

Step 2: We now sheafify the morphism of presheaves with respect to the eh-topology. As $\mathcal{O}(X) \to \mathcal{O}(X')$ is flat, it commutes with sheafification, and we get an isomorphism of eh-sheaves

$$\Omega^n_{\mathrm{eh}} \otimes_{\mathcal{O}(X)} \mathcal{O}(X') \to (\Omega^n(- \times_X X'))_{\mathrm{eh}}. \tag{13}$$

Step 3: We claim that

$$(\Omega^n(- \times_X X'))_{\mathrm{eh}}(Y) = \Omega^n_{\mathrm{eh}}(Y \times_X X'). \tag{14}$$

It suffices to show that every eh-cover of $Y' = Y \times_X X'$ can be refined by the pullback of an eh-cover of Y. Let $Z \to Y'$ be an eh-cover. The composition $Z \to Y' \to Y$ is also an eh-cover. The pullback $Z \times_Y Z \to Z \to Y'$ is the required refinement.

Combining the isomorphisms (13) and (14) in the case $Y = X$ gives the desired isomorphism (11). $\square$

6. Special cases

The cases of forms of degree zero or top degree are easier to handle than the general case. In this section we study these special cases.

6A. 0-*differentials.*

This section is about the sheafification of the presheaf $\Omega^0 = \mathcal{O}$. For expositional reasons we work over a field. For more general bases, and more general representable presheaves, see Section 6C.

In contrast to the general case, $\mathcal{O}_{\mathrm{val}}$ is torsion free.

Lemma 6.1. *For every $X \in \mathrm{Sch}^{\mathrm{ft}}_k$, every dense open immersion $U \subseteq X$ in $\mathrm{Sch}^{\mathrm{ft}}_k$ induces an inclusion $\mathcal{O}_{\mathrm{val}}(X) \subseteq \mathcal{O}_{\mathrm{val}}(U)$.*

Proof. First suppose it is true for irreducible schemes, and let $X_1, \dots, X_n$ be the irreducible components of X. Let $U_1, \dots, U_n$ be their intersections with U. Each U_i is dense in X_i. Let $f, g \in \mathcal{O}_{\mathrm{val}}(X)$ such that $f|_U = g|_U$. Then $f|_{U_i} = g|_{U_i}$. By the irreducible case, $f|_{X_i} = g|_{X_i}$. By cdh-descent

$$\mathcal{O}_{\mathrm{val}}(X) \subseteq \mathcal{O}_{\mathrm{val}}(X_i)$$

and hence $f = g$.

Now consider X irreducible. Since $\mathcal{O}_{\mathrm{val}}(X) = \mathcal{O}_{\mathrm{val}}(X_{\mathrm{red}})$ we can assume X is integral. Consider the description of Lemma 3.7. If two sections $(s_x), (t_x) \in \mathcal{O}_{\mathrm{val}}(X)$ are equal on a dense open, then $s_\eta = t_\eta$ where η is the generic point of X. Consequently, for any valuation ring R of the form $\eta \to \mathrm{Spec}(R) \to X$, the lifts $s_{\mathrm{Spec}(R)}, t_{\mathrm{Spec}(R)}$ are also equal, as well as their images in R/m, and from there we deduce that $s_x = t_x$ where x is the image of $\mathrm{Spec}(R/\mathfrak{m}) \to X$. But for every point $x \in X$ there is a valuation ring $R_x \subset k(X)$ such that the special point of $\mathrm{Spec}(R)$ maps to x [EGA II 1961, Proposition 7.1.7]. □

Recall the notion of the seminormalisation of a variety (Definition 2.3).

Proposition 6.2. *Let $Y \in \mathsf{Sch}_k^{\mathrm{ft}}$. Then the canonical morphism*

$$\mathcal{O}_{\mathrm{val}}(Y) \cong \mathcal{O}(Y^{\mathrm{sn}})$$

is an isomorphism. The same is true for $\mathcal{O}_{\mathrm{rh}}$, $\mathcal{O}_{\mathrm{cdh}}$ and $\mathcal{O}_{\mathrm{eh}}$ in place of $\mathcal{O}_{\mathrm{val}}$.

Remark 6.3. The proof below works for a general noetherian base scheme S. We (the authors) do not know if the seminormalisation A^{sn} is always noetherian in this setting, but the definition of $\mathcal{O}_{\mathrm{val}}(A^{\mathrm{sn}})$ is, clearly, still valid, and the careful reader will note that the proof below still works, regardless.

Proof. Both $\mathcal{O}_{\mathrm{val}}(-)$ and $\mathcal{O}((-)^{\mathrm{sn}})$ are invariant under $(-)_{\mathrm{red}}$, and are Zariski sheaves, so it suffices to consider the case where Y is reduced and affine. Let $Y = \mathrm{Spec}(A)$. As $\mathrm{Spec}(A^{\mathrm{sn}}) \to \mathrm{Spec}(A)$ is a completely decomposed homeomorphism, $\mathrm{val}(A^{\mathrm{sn}}) \to \mathrm{val}(A)$ is an equivalence of categories, so $\mathcal{O}_{\mathrm{val}}(A) \to \mathcal{O}_{\mathrm{val}}(A^{\mathrm{sn}})$ is an isomorphism. Hence, it suffices to show that if A is a seminormal ring, then $\mathcal{O}(A) = \mathcal{O}_{\mathrm{val}}(A)$. Define $A_{\mathrm{val}} = \mathcal{O}_{\mathrm{val}}(\mathrm{Spec}(A))$. By Lemma 2.6(3) it suffices to show that $A \to A_{\mathrm{val}}$ is subintegral.

First we show that it is integral. By the argument in Lemma 6.1, or by its statement and the comparison of Theorem 4.12, both of the canonical morphisms $A \to A_{\mathrm{val}} \to \prod_{\mathfrak{p}\text{ minimal}} A_{\mathfrak{p}}$ is an embedding into a product of fields [Stacks project, Tag 00EW]. Since $\mathrm{Spec}(A)$ is a noetherian topological space, there are finitely many of them. Now by the definition of A_{val}, the image of A_{val} in each $A_{\mathfrak{p}}$ is contained in any valuation ring of $A_{\mathfrak{p}}$ containing the image of A. Since the normalisation is the intersection of these valuation rings [Stacks project, Tag 090P], it follows that the extension $A \subseteq A_{\mathrm{val}}$ is integral.

To show that $\mathrm{Spec}(A_{\mathrm{val}}) \to \mathrm{Spec}(A)$ is a completely decomposed homeomorphism, it suffices to show that for all fields κ which are residue fields of A or A_{val},

$$\hom_k(A_{\mathrm{val}}, \kappa) \to \hom_k(A, \kappa) \tag{15}$$

is an isomorphism.

Surjectivity: We construct a section of the map of sets (15). For every morphism $\phi : A \to \kappa$ in $\mathrm{val}(A)$ to a residue field κ of A, there is a canonical extension $\phi : A \xrightarrow{\iota} A_{\mathrm{val}} \xrightarrow{\pi_\phi} \kappa$ making the triangle commute:

just take $\pi_\phi : \varprojlim_{\phi' : A \to R' \in \mathrm{val}(A)} R' \to \kappa$ to be the projection to the ϕ-th component of the limit. For κ a general field, take π_ϕ to be the map associated to the residue field corresponding to ϕ.

Injectivity: We show that the section $\phi \mapsto \pi_\phi$ we have just constructed is surjective. That is, for an arbitrary residue field $\psi : A_{\mathrm{val}} \to \kappa$, we claim that $\psi = \pi_{A \to A_{\mathrm{val}} \to \kappa}$. If $\psi \circ \iota$ is the canonical map $A \to A_{\mathfrak{p}}$ to one of the fractions fields of an irreducible component of $\mathrm{Spec}(A)$, then there is a unique lift because $A \subseteq A_{\mathrm{val}} \subseteq \prod_{\mathfrak{p} \text{ minimal}} A_{\mathfrak{p}}$, as we observed above. For a general residue field of A_{val}, there is an A_{val}-valuation ring $R \subseteq A_{\mathfrak{p}}$ of the fraction field $A_{\mathfrak{p}}$ of any irreducible component containing $\mathrm{Spec}(\kappa) \in \mathrm{Spec}(A_{\mathrm{val}})$ whose special point maps to $\mathrm{Spec}(\kappa)$ [EGA II 1961, Proposition 7.1.7] (for the nonnoetherian version, cf. [Stacks project, Tag 00IA]). So we have the commutative diagram

$$\begin{array}{ccccc} R/\mathfrak{m} & \longleftarrow & R & \longrightarrow & A_{\mathfrak{p}} \\ \uparrow & & \uparrow & & \\ \kappa & \xleftarrow{\psi} & A_{\mathrm{val}} & & \end{array}$$

As we have just discussed, the map $A_{\mathrm{val}} \to A_{\mathfrak{p}}$ must be the unique extension $\pi_{A \to A_{\mathrm{val}} \to A_{\mathfrak{p}}}$ of the canonical $A \to A_{\mathfrak{p}}$. As $R \to A_{\mathfrak{p}}$ is injective, the map $A_{\mathrm{val}} \to R$ must also be the canonical $\pi_{A \to A_{\mathrm{val}} \to R}$, and therefore $A_{\mathrm{val}} \to R/\mathfrak{m}$ is $\pi_{A \to A_{\mathrm{val}} \to R \to R/\mathfrak{m}}$, and by injectivity of $\kappa \to R/\mathfrak{m}$, we conclude $\psi = \pi_{A \to A_{\mathrm{val}} \to \kappa}$.

The claim about $\mathcal{O}_{\mathrm{rh}}, \mathcal{O}_{\mathrm{cdh}}, \mathcal{O}_{\mathrm{eh}}$ follows from Theorem 4.12. □

Recall the sdh-topology introduced in [Huber, Kebekus and Kelly 2017, Section 6.2]. It is generated by étale covers and those proper surjective maps that are separably decomposed, i.e., any point has a preimage such that the residue field extension is finite and separable. By de Jong's theorem on alterations [1996], every $X \in \mathrm{Sch}_k^{\mathrm{ft}}$ is sdh-locally smooth. However, sdh-descent fails for differential forms [Huber, Kebekus and Kelly 2017, Proposition 6.6]. The situation is better in degree zero.

Proposition 6.4. *Let k be perfect. Then $\mathcal{O}_{\mathrm{val}} = \mathcal{O}_{\mathrm{sdh}}$.*

Proof. The sdh-topology is stronger than the eh-topology, and we know $\mathcal{O}_{\mathrm{val}} = \mathcal{O}_{\mathrm{eh}}$ (Theorem 4.12). Hence, we have a canonical morphism $\mathcal{O}_{\mathrm{val}} \to \mathcal{O}_{\mathrm{sdh}}$, and an isomorphism $(\mathcal{O}_{\mathrm{val}})_{\mathrm{sdh}} \cong \mathcal{O}_{\mathrm{sdh}}$. If we can show that $\mathcal{O}_{\mathrm{val}}$ is already an sdh sheaf, we are done. The topology is generated by proper separably decomposed morphisms and étale covers. We already know that $\mathcal{O}_{\mathrm{val}}$ is an étale sheaf. Hence it suffices to show that if $Y \to X$ is a proper sdh-cover, which generically is finite and separable, then

$$0 \to \mathcal{O}_{\mathrm{val}}(X) \to \mathcal{O}_{\mathrm{val}}(Y) \to \mathcal{O}_{\mathrm{val}}(Y \times_X Y)$$

is exact.

Recall that since $\mathcal{O}$ is torsion free on valuation rings, $\mathcal{O}_{\mathrm{val}}(X) \to \prod_{x \in X} \kappa(x)$ is injective (Lemma 3.7). For $y \in Y$ with image $x \in X$, the induced map $\kappa(x) \to \kappa(y)$ is injective as a map of fields. As $Y \to X$ is surjective, this implies that $\mathcal{O}_{\mathrm{val}}(X) \to \mathcal{O}_{\mathrm{val}}(Y)$ is injective.

Let $f \in \mathcal{O}_{\mathrm{val}}(Y)$ be in the kernel of the second map. We want to define an element $g \in \mathcal{O}_{\mathrm{val}}(X)$ and start with the component $g_x \in \kappa(x)$ for $x \in X$. Let $y \in Y$ be a preimage of X such that the residue field extension $\kappa(y)/\kappa(x)$ is finite and separable. Let $\lambda/\kappa(x)$ be a finite Galois extension containing $\kappa(y)$. We have a canonical map $l : \mathrm{Spec}(\lambda) \to Y$, and any $\kappa(x)$-automorphism σ of λ gives us a second map $l \circ \sigma$, and this pair of maps define some $(l, l \circ \sigma) : \mathrm{Spec}(\lambda) \to Y \times_X Y$. By the assumption that f is a cocycle, $f_l = (\pi_1^* f)_{(l,l\sigma)} = (\pi_2^* f)_{(l,l\sigma)} = f_{l\sigma}$ in λ. That is, $f_l \in \lambda$ is $\mathsf{Gal}(\lambda/\kappa(x))$-invariant, and therefore, actually lies in $\kappa(x) \subseteq \lambda$. We define $g_{\kappa(x)} := f_\lambda$.

Note that another consequence of the cocycle condition is that $g_{\kappa(x)}$ is independent of the choice of y, even without assuming separability or finiteness. For any other y' over x, we can choose an extension $K/\kappa(x)$ containing both $\kappa(y)$ and $\kappa(y')$, leading to a map $\mathrm{Spec}(K) \to Y \times_X Y$, to which we can apply the cocycle condition to find that $f_y = f_{y'}$ in K via the chosen embeddings, and therefore they also agree in $\kappa(x)$.

It remains to show that the tuple $(g_{\kappa(x)})_{x \in X}$ defines a section of $\mathcal{O}_{\mathrm{val}}(X)$. We continue with the criterion of Lemma 3.7. Let $R \subset \kappa(x)$ be a valuation ring over X. Let $y \in Y$ again be a preimage of x. There is a valuation ring $S \subset \kappa(y)$ such that $R = S \cap \kappa(x)$ [Bourbaki 1964, Chapitre VI, §3.3, Proposition 5]. By the valuative criterion for properness, S is a Y-valuation on $\kappa(y)$. As $f \in \mathcal{O}_{\mathrm{val}}(Y)$, the element $g_{\kappa(x)} = f_{\kappa(y)}$ is in $S \subseteq \kappa(y)$, but it is also in $\kappa(x)$, so $g_{\kappa(x)}$ is in $R \subseteq \kappa(x)$. Therefore, let us write it as g_R. Let x_0 be the image of $\mathrm{Spec}(R/\mathfrak{m}_R) \to X$, and y_0 the image of $\mathrm{Spec}(S/\mathfrak{m}_S) \to Y$). To finish, we must show that g_R agrees with g_{x_0} in $R/\mathfrak{m}$. But $g_R|_{R/\mathfrak{m}_R}|_{S/\mathfrak{m}_S} = g_R|_S|_{S/\mathfrak{m}_S} = f_S|_{S/\mathfrak{m}_S} = f_{S/\mathfrak{m}_S} = f_{y_0}|_{S/\mathfrak{m}_S} = g_{x_0}|_{y_0}|_{S/\mathfrak{m}_S} = g_{x_0}|_{R/\mathfrak{m}_R}|_{S/\mathfrak{m}_S}$ and $R/\mathfrak{m}_R \to S/\mathfrak{m}_S$ is injective, so $g_R|_{R/\mathfrak{m}_R} = g_{x_0}|_{R/\mathfrak{m}_R}$. □

Remark 6.5. Let us point out where the above proof breaks for Ω^n. The argument for injectivity is actually valid because we can choose the preimage y of x to be separable. The construction of each $g_{\kappa(x)}$ is fine, as well as independence of the choice of y used in the construction. However, for y over x which are not separable, we cannot necessarily check that $g_x|_y = f_y$. Choosing y/x separable in the last paragraph, we can show that each $g_{\kappa(x)}$ lifts to any X-valuation ring of $\kappa(x)$, but we cannot ensure that $R/\mathfrak{m}_R \to S/\mathfrak{m}_S$ is separable, nor its image $x_0 \to y_0$, so we cannot check that we have a well-defined section.

In fact, not being able to control this kind of ramification is precisely why the sdh-topology is not suitable for working with differential forms; cf. [Huber, Kebekus and Kelly 2017, Example 6.5].

On the other hand, Proposition 6.4 is valid for any representable presheaf h^Y for any scheme Y. Moreover, using the same proof, we can show that $\Omega^n_{\mathrm{val}}(X) = \Omega^n_{\mathrm{sdh}}(X)$ whenever $\dim X \leq n$.

Proposition 6.6.
$$\mathcal{O}_{\mathrm{sdh}} = \mathcal{O}_{\mathrm{dvr}}.$$

Proof. The same arguments as in the last proof show that $\mathcal{O}_{\mathrm{dvr}}$ has sdh-descent. (In the above notation: if R is a discrete valuation ring, then S can also be chosen as a discrete valuation ring.) As pointed out before, any $X \in \mathsf{Sch}_k^{\mathrm{ft}}$ is smooth locally for the sdh-topology. Hence it suffices to compare the values on smooth varieties. In this case we have on the one hand $\mathcal{O}_{\mathrm{sdh}}(X) = \mathcal{O}_{\mathrm{val}}(X) = \mathcal{O}(X)$, on the other hand $\mathcal{O}_{\mathrm{dvr}}(X) = \mathcal{O}(X)$ by [Huber, Kebekus and Kelly 2017, Remark 4.3.3]. □

Remark 6.7. Hence we have

$$\mathcal{O}_{\mathrm{rh}} = \mathcal{O}_{\mathrm{cdh}} = \mathcal{O}_{\mathrm{eh}} = \mathcal{O}_{\mathrm{val}} = \mathcal{O}_{\mathrm{dvr}} = \mathcal{O}_{\mathrm{sdh}}.$$

However, in positive characteristic $\mathcal{O}_{\mathrm{h}} \neq \mathcal{O}_{\mathrm{val}}$ because $\mathcal{O}_{\mathrm{val}}$ does not have descent for Frobenius covers. See Proposition 6.2 and Remark 6.15(1).

The following property is well-known for the ordinary structure sheaf under the assumption that Y is normal. It will be useful in connection with cohomological descent questions [Huber and Kelly $\geq$ 2018].

Proposition 6.8. *Let $Y' \to Y$ be a* cdp*-morphism in $\mathrm{Sch}_k^{\mathrm{ft}}$ with geometrically connected fibres. Then*

$$\mathcal{O}_{\mathrm{eh}}(Y) \to \mathcal{O}_{\mathrm{eh}}(Y')$$

is an isomorphism.

Proof. We use induction on the dimension of Y. Note the hypotheses are preserved by all base changes along Y. Without loss of generality, all schemes are assumed to be reduced. If $\dim Y = -1$ (i.e., $Y = \varnothing$), then $Y' = \varnothing$ and the proposition follows from $\mathcal{O}_{\mathrm{eh}}(\varnothing) = 0$. In dimension ≥ 0, let $\pi : \widetilde{Y} \to Y$ be the normalisation of Y. Let $Z \subset Y$ be the locus where π fails to be an isomorphism and $E = \pi^{-1}Z$ the preimage. Let $\widetilde{Y}'$, Z' and E' be the base changes to Y'. We have a commutative diagram of blowup sequences

$$\begin{array}{ccccccc} 0 & \longrightarrow & \mathcal{O}_{\mathrm{eh}}(Y') & \longrightarrow & \mathcal{O}_{\mathrm{eh}}(\widetilde{Y}') \oplus \mathcal{O}_{\mathrm{eh}}(Z') & \longrightarrow & \mathcal{O}_{\mathrm{eh}}(E') \\ & & \uparrow & & \uparrow & & \uparrow \\ 0 & \longrightarrow & \mathcal{O}_{\mathrm{eh}}(Y) & \longrightarrow & \mathcal{O}_{\mathrm{eh}}(\widetilde{Y}) \oplus \mathcal{O}_{\mathrm{eh}}(Z) & \longrightarrow & \mathcal{O}_{\mathrm{eh}}(E) \end{array}$$

By the induction hypothesis, it now suffices to prove $\mathcal{O}_{\mathrm{eh}}(\widetilde{Y}) \cong \mathcal{O}_{\mathrm{eh}}(\widetilde{Y}')$. Since $\mathcal{O}_{\mathrm{eh}}((-)^{\mathrm{sn}}) = \mathcal{O}((-)^{\mathrm{sn}})$ by Proposition 6.2, so it suffices, in fact, to show that $\mathcal{O}(\widetilde{Y}'^{\mathrm{sn}}) \to \mathcal{O}(\widetilde{Y})$ is an isomorphism.

Since $\widetilde{Y}'^{\mathrm{sn}} \to \widetilde{Y}'$ is a proper homeomorphism, we are dealing with a proper surjective morphism $\widetilde{Y}'^{\mathrm{sn}} \to \widetilde{Y}$ to a normal scheme. In this situation, Stein factorisation [EGA III$_1$ 1961, Théorème 4.3.1] gives us a factorisation $\widetilde{Y}'^{\mathrm{sn}} \to W \to \widetilde{Y}$ such that $\mathcal{O}(\widetilde{Y}'^{\mathrm{sn}}) \cong \mathcal{O}(W)$ and such that $W \to \widetilde{Y}$ is finite. Since $\widetilde{Y}'^{\mathrm{sn}} \to \widetilde{Y}$ has connected fibres, so does $W \to \widetilde{Y}$ [EGA III$_1$ 1961, Corollaire 4.3.3]. We also deduce that because $\widetilde{Y}'^{\mathrm{sn}}$ is reduced so is W, and because $\widetilde{Y}'$ is completely decomposed, so is $\widetilde{Y}'^{\mathrm{sn}} \to \widetilde{Y}$ and therefore $W \to \widetilde{Y}$ also. In particular, since $W \to \widetilde{Y}$ is completely decomposed and W reduced, the fibre over the generic point of $\widetilde{Y}$ must be an isomorphism. Replacing W with its normalisation, $\widetilde{W} \to \widetilde{Y}$, we have a finite birational morphism between normal schemes. This can only be an isomorphism, so $W \to \widetilde{Y}$ was an isomorphism, and $\mathcal{O}(W) = \mathcal{O}(\widetilde{Y})$.

To summarise, $\mathcal{O}_{\mathrm{eh}}(\widetilde{Y}') \cong \mathcal{O}_{\mathrm{eh}}(\widetilde{Y}'^{\mathrm{sn}}) \cong \mathcal{O}(\widetilde{Y}'^{\mathrm{sn}}) \cong \mathcal{O}(W) \cong \mathcal{O}(\widetilde{Y})$. □

6B. ***Top degree differentials.*** Recall the notion of a birational morphism of schemes in the nonreduced case from Section 2A.

Proposition 6.9. *Let $X/k \in \mathrm{Sch}_k^{\mathrm{ft}}$ be of dimension at most d.*

(1) *$\Omega^d_{\mathrm{cdh}}(X)$ is a birational invariant, i.e., it remains unchanged under proper surjective birational morphisms.*

(2) *We have*
$$\Omega^d_{\mathrm{cdh}}(X) = \varinjlim_{X'\to X} \Omega^d(X'),$$
where the colimit is over proper surjective birational morphisms $X' \to X$.

(3) *Elements of $\Omega^d_{\mathrm{cdh}}(X) = \Omega^d_{\mathrm{val}}(X)$ are determined by their value on the total ring of fractions $Q(X)$, and the integrality condition only needs to be tested on valuations of the function fields. In particular, it is torsion free.*

(4) *More precisely, if X is irreducible of dimension d, then by Lemma 3.7,*
$$\Omega^d_{\mathrm{val}}(X) = \bigcap_{\mathrm{Spec}(k(X))\to\mathrm{Spec}(R)\to X\in\mathrm{rval}(X)} \Omega^d(R). \tag{16}$$
In general, if $X = X_1 \cup \cdots \cup X_N$ is the decomposition into irreducible component, then
$$\Omega^d_{\mathrm{val}}(X) = \bigoplus_{i=1}^{N} \Omega^d_{\mathrm{val}}(X_i) \quad \textit{and} \quad \Omega^d_{\mathrm{val}}(X) = \Omega^d_{\mathrm{val}}(\tilde{X}).$$

Proof. Note that $\Omega^d_{\mathrm{cdh}} = \Omega^d_{\mathrm{val}}$ vanishes on schemes of dimension less that d. Hence the first statement is immediate from the sequence for abstract blowup squares.

The third statement follows from Lemma 3.7, the fact that $\Omega^n(K) = 0$ for $n > \mathrm{trdeg}(K/k)$ and the valuative criterion for properness. The explicit formula is immediate from this.

For the second statement consider
$$X \mapsto \widetilde{\Omega}^d(X) := \varinjlim_{X'\to X} \Omega^d(X').$$
By definition this is a birational invariant. We claim that $\widetilde{\Omega}^d$ is torsion free. Note that X' can always be refined by the disjoint union of its irreducible components with their reduced structure. Let ω be a torsion element of $\widetilde{\Omega}^d(X)$. It is represented by a differential form on some $X_1 \to X$. After restriction to some further $X_2 \to X_1$ it vanishes on a dense open subset. Then there is a proper birational morphism $X_3 \to X_2$ such that $\omega|_{X_3} = 0$. This was shown in [Huber, Kebekus and Kelly 2017, Theorem A.3] (for a recap see Theorem 4.6 combined with Theorem 2.20). Hence $\omega = 0$ in the direct limit.

By torsion freeness, we have $\widetilde{\Omega}^d(X) = 0$ if the dimension of X is less than d. Hence $\widetilde{\Omega}^d$ is a presheaf on the category of k-schemes of dimension at most d. It is a Zariski sheaf because Ω^d is. It has descent for abstract blowup squares by birational invariance and vanishing in smaller dimensions. Hence it is an rh-sheaf. By the universal property, there is a natural map
$$\Omega^d_{\mathrm{rh}} \to \widetilde{\Omega}^d.$$
The map
$$\widetilde{\Omega}^d(X) = \varinjlim_{X'\to X} \Omega^d(X') \to \varinjlim_{X'\to X} \Omega^d_{\mathrm{rh}}(X') = \Omega^d_{\mathrm{rh}}(X)$$
induces a natural map in the other direction. We check that they are inverse to each other. Both sheaves are torsion free; hence it suffices to consider generic points where it is true. ☐

Remark 6.10. Note that the description in (16) can also be interpreted as the global sections on the Riemann–Zariski space $\Gamma(\mathsf{RZ}(X), \Omega^d_{RZ(X)/k})$.

In the smooth case, this gives a formula involving only ordinary differential forms.

Corollary 6.11. *Let X be a smooth k-scheme of dimension d. Then*

$$\Omega^d(X) = \varinjlim_{X' \to X} \Omega^d(X'),$$

where the colimit is over proper birational morphisms $X' \to X$.

Proposition 6.12. *On the category of k-schemes of dimension at most d, we have*

$$\Omega^d_{\mathrm{val}} = \Omega^d_{\mathrm{eh}} = \Omega^d_{\mathrm{sdh}} = \Omega^d_{\mathrm{dvr}}.$$

Proof. The first isomorphism is Theorem 4.12. For the other two, the same proofs as for Proposition 6.4 and Proposition 6.6 work. □

6C. *Representable sheaves.* Note that $\Omega^0 = \mathcal{O} = \hom(-, \mathbb{A}^1)$. In this section we extend our results to all representable sheaves over a general noetherian base S. We use the following notation for representable presheaves on $\mathsf{Sch}^{\mathrm{ft}}_S$:

$$h^Y(-) = \hom_{\mathsf{Sch}_S}(-, Y), \qquad Y \in \mathsf{Sch}^{\mathrm{ft}}_S.$$

Note that this presheaf satisfies the properties of Remark 4.1. Notice also that the h^Y are torsion free in the sense of Definition 2.22 — this is exactly the valuative criterion for separatedness.

Lemma 6.13. *Suppose that the noetherian base scheme S is Nagata. Let $X' \to X$ be a finite completely decomposed surjective morphism in Sch_S, and suppose that X' is seminormal. Then the coequalisers*

$$\mathrm{coeq}(X' \times_X X' \rightrightarrows X') = C, \qquad \mathrm{coeq}((X' \times_X X')^{\mathrm{sn}} \rightrightarrows X') = D$$

exist in Sch_S, we have $D = X^{\mathrm{sn}}$ and the canonical morphisms $D \to C \to X$ are finite completely decomposed homeomorphisms.

Proof. Using the description [Ferrand 2003, Scolie 4.3], one easily constructs the coequaliser in the category of locally ringed spaces by taking the coequaliser in the category of sets, equipping it with the quotient topology, and the equaliser of the direct images of the structure sheaves. Using this description, one readily deduces from $X' \to X$ being finite that $D \to C \to X$ are homeomorphisms. Note that $X^{\mathrm{sn}} \to X$ is also a homeomorphism. Now since $X' \to X$ is an rh-cover, it follows from Proposition 6.2 and Remark 6.3 that $\mathcal{O}(X^{\mathrm{sn}}) = \mathrm{eq}(\mathcal{O}(X') \rightrightarrows \mathcal{O}((X' \times_X X')^{\mathrm{sn}})$. The same holds for any open $U \subseteq X$. That is, the canonical morphism $D \to X^{\mathrm{sn}}$ of locally ringed spaces is an isomorphism on topological spaces *and* structure sheaves. In other words, it is an isomorphism. Finally, note that we have a canonical inclusion of sheaves $\mathcal{O}_X \subseteq \mathcal{O}_C \subseteq \mathcal{O}_{X^{\mathrm{sn}}}$. For any open affine $U \subseteq X$ of X, it follows that $\mathrm{Spec}(\mathcal{O}_{X^{\mathrm{sn}}}(U)) \to \mathrm{Spec}(\mathcal{O}_C(U)) \to \mathrm{Spec}(\mathcal{O}_X(U))$ are homeomorphisms on topological spaces. Hence, C is a scheme. □

Proposition 6.14. *Suppose that the noetherian base scheme S is Nagata. Then for every $X, Y \in \mathsf{Sch}^{\mathrm{ft}}_S$ the canonical morphisms*

$$h^Y_{\mathrm{rh}}(X) = h^Y(X^{\mathrm{sn}}) \tag{17}$$

are isomorphisms. The natural maps

$$h^Y_{\mathrm{rh}} \to h^Y_{\mathrm{cdh}} \to h^Y_{\mathrm{eh}} \to h^Y_{\mathrm{sdh}} \to h^Y_{\mathrm{val}} \tag{18}$$

are isomorphisms of presheaves on $\mathsf{Sch}^{\mathrm{ft}}_S$.

Proof. We claim that h^Y_{red} is rh-separated, where

$$h^Y_{\mathrm{red}}(-) = h^Y((-)_{\mathrm{red}}).$$

Let $f, g \in h^Y_{\mathrm{red}}(X)$ with $f = g$ in $h^Y_{\mathrm{rh}}(X)$. In particular, $f_\eta = g_\eta$ at all generic points of X. But as X_{red} is reduced and $Y \to S$ is separated, this implies $f = g$. Since h^Y_{red} is a Zariski sheaf, in light of the factorisation of Remark 2.2(2), we have

$$h^Y_{\mathrm{rh}}(X) = h^Y_{\mathrm{cdp}}(X)$$

when X is reduced (see [Kelly 2017, Proposition 2.1.16]), and hence, in general, as both h^Y_{rh} and h^Y_{cdp} are unchanged by reducing the structure sheaf.

By cdp-separatedness of h^Y_{red} we have

$$h^Y_{\mathrm{cdp}}(X) = \varinjlim_{\mathrm{cdp}} \check{H}^0(X'/X, h^Y_{\mathrm{red}}), \tag{19}$$

where the colimit is over all cdp-covers $p : X' \to X$. We claim that

$$\varinjlim_{\text{comp. dec. homeo.}} \check{H}^0(X'/X, h^Y_{\mathrm{red}}) \to \varinjlim_{\mathrm{cdp}} \check{H}^0(X'/X, h^Y_{\mathrm{red}}) \tag{20}$$

is an isomorphism, where the first colimit is over completely decomposed homeomorphisms. Let $p : X' \to X$ be a cdp-cover. For such a cover, define $X'' = \underline{\mathrm{Spec}}\, p_*\mathcal{O}_{X'}$. Since $q : X' \to X''$ is proper, the topological space of X'' is the quotient of the topological space X' via this morphism. Hence, any morphism $f' \in \mathrm{eq}(h^Y_{\mathrm{red}}(X') \rightrightarrows h^Y_{\mathrm{red}}(X' \times_Y X'))$ factors through X'' as a morphism of *topological spaces*. But we have $q_*\mathcal{O}_{X'} = \mathcal{O}_{X''}$ by construction, and therefore f comes from some $f'' \in h^Y_{\mathrm{red}}(X'')$. Then, since $(X' \times_X X')_{\mathrm{red}} \to (X'' \times_X X'')_{\mathrm{red}}$ is dominant with reduced source, we actually have $f'' \in \check{H}^0(X''/X, h^Y_{\mathrm{red}})$. Replacing X'' by $(X'')^{\mathrm{sn}}$ (this is where we use the assumption that S is Nagata), we are in the situation of Lemma 6.13, and find that $f''|_{(X'')^{\mathrm{sn}}}$ comes from some $g \in h^Y_{\mathrm{red}}(D)$ for some completely decomposed homeomorphism $D \to X$. As $(D \times_X D)_{\mathrm{red}} = D_{\mathrm{red}}$, we have $g \in \check{H}^0(D/X, h^Y_{\mathrm{red}})$, so we have shown that (20) is surjective. It is clearly injective, as any refinement $X'' \to X' \to X$ of a completely decomposed homeomorphism by a cdp-morphism is dominant. Hence, (20) is an isomorphism.

Finally, it follows from Lemma 2.6(3) that $X^{\mathrm{sn}} \to X$ is an initial object in the category of completely decomposed homeomorphisms to X. So

$$h^Y_{\mathrm{cdp}}(X) = \varinjlim_{\text{comp. dec. homeo.}} \check{H}^0(X'/X, h^Y_{\mathrm{red}}) = \check{H}^0(X^{\mathrm{sn}}/X, h^Y_{\mathrm{red}}) = h^Y(X^{\mathrm{sn}}).$$

The isomorphisms (18), except for h^Y_{sdh}, are Theorem 4.12. Injectivity of $h^Y_{\mathrm{sdh}} \to h^Y_{\mathrm{val}}$ has the same proof as injectivity of $h^Y_{\mathrm{eh}} \to h^Y_{\mathrm{val}}$; see the proof of Corollary 4.7. □

Remark 6.15. (1) This is analogous to the comparison $h^Y_{\mathrm{h}}(X) = h^Y(X^{\mathrm{sn}})$ in characteristic zero; see [Huber and Jörder 2014, Proposition 4.5; Voevodsky 1996, Section 3.2].

(2) In fact, in general we have $h^Y_{\mathrm{h}}(X) = h^Y(X^{\mathrm{wn}})$, where Y^{wn} is the absolute weak normalisation [Rydh 2010, Definition B.1]. For noetherian reduced schemes in pure positive characteristic, Y^{wn} is the perfect closure of Y in $\prod_{i=1}^n K_i^a$, the product of the algebraic closures of the function fields of its irreducible components. This holds much more generally: it is true for any algebraic space Y locally of finite presentation [Rydh 2010, Theorem 8.16].

In particular, the categories of representable rh-, cdh-, and eh-sheaves on $\mathsf{Sch}^{\mathrm{ft}}_S$ agree.

Corollary 6.16 (cf. [Voevodsky 1996, Theorem 3.2.9]). *Suppose the noetherian scheme S is Nagata. The category of representable* rh*-sheaves on $\mathsf{Sch}^{\mathrm{ft}}_S$ is a localisation of the category $\mathsf{Sch}^{\mathrm{ft}}_S$ with respect to completely decomposed homeomorphisms. In other words, it is obtained by formally inverting morphisms of the form $X_{\mathrm{red}} \to X$, then formally inverting subintegral extensions.*

Proof. Certainly, the functor $X \mapsto h^X_{\mathrm{rh}}$ factors through the localisation functor $(-)_{\mathrm{red}} : \mathsf{Sch}^{\mathrm{ft}}_S \to (\mathsf{Sch}^{\mathrm{ft}}_S)_{\mathrm{red}}$. Now, it is straightforward to check that the class $\mathcal{S}$ of subintegral extensions of reduced schemes are a multiplicative system:

(1) $\mathcal{S}$ is closed under composition.

(2) For every $t : Z \to Y$ in $\mathcal{S}$ and $g : X \to Y$ in $(\mathsf{Sch}^{\mathrm{ft}}_S)_{\mathrm{red}}$ there is an $s : W \to X$ in $\mathcal{S}$ and $f : W \to Z$ in $(\mathsf{Sch}^{\mathrm{ft}}_S)_{\mathrm{red}}$ with $gs = tf$.

(3) If $f, g : X \rightrightarrows Y$ are parallel morphisms in $(\mathsf{Sch}^{\mathrm{ft}}_S)_{\mathrm{red}}$, then the following are equivalent:
   (a) $sf = sg$ for some $s \in \mathcal{S}$ with source Y.
   (b) $ft = gt$ for some $t \in \mathcal{S}$ with target X.

(In fact, the latter two conditions are equivalent to $f = g$ in this case.)

Since $\mathcal{S}$ is a multiplicative system, the hom sets in the localisation $\mathcal{S}^{-1}(\mathsf{Sch}^{\mathrm{ft}}_S)_{\mathrm{red}}$ are calculated by the formula

$$\hom_{\mathcal{S}^{-1}(\mathsf{Sch}^{\mathrm{ft}}_S)_{\mathrm{red}}}(X, Y) = \varinjlim_{X' \to X \in \mathcal{S}} \hom(X', Y) = \hom(X^{\mathrm{sn}}, Y);$$

see [Gabriel and Zisman 1967; Weibel 1994, Theorem 10.3.7]. But, by Proposition 6.14, this is equal to $\hom_{\mathsf{Shv}_{\mathrm{rh}}(\mathsf{Sch}^{\mathrm{ft}}_S)}(h^X_{\mathrm{rh}}, h^Y_{\mathrm{rh}})$. □

Corollary 6.17 (cf. [Voevodsky 1996, Theorem 3.2.10]). *Suppose our noetherian base scheme S is Nagata. Let $\mathsf{Shv}^{\mathrm{rep}}_{\mathrm{rh}}(\mathsf{Sch}^{\mathrm{ft}}_S) \subseteq \mathsf{Shv}_{\mathrm{rh}}(\mathsf{Sch}^{\mathrm{ft}}_S)$ denote the full subcategory of representable* rh*-sheaves. The Yoneda functor $h^-_{\mathrm{rh}} : \mathsf{Sch}^{\mathrm{ft}}_S \to \mathsf{Shv}^{\mathrm{rep}}_{\mathrm{rh}}(\mathsf{Sch}^{\mathrm{ft}}_S)$ admits a left adjoint. The counit of the adjunction is the seminormalisation $X^{\mathrm{sn}} \to X$. In particular, for any schemes $X, Y \in \mathsf{Sch}^{\mathrm{ft}}_S$ with X seminormal, one has*

$$\hom(h^X_{\mathrm{rh}}, h^Y_{\mathrm{rh}}) = \hom_{\mathsf{Sch}^{\mathrm{ft}}_S}(X, Y).$$

Proof. We have $\hom_{\mathsf{Shv}_{\mathrm{rh}}}(h^X_{\mathrm{rh}}, h^Y_{\mathrm{rh}}) = h^Y_{\mathrm{rh}}(X) = h^Y(X^{\mathrm{sn}}) = \hom(X^{\mathrm{sn}}, Y)$. □

7. Future directions

7A. *Relation to Berkovich spaces.* Berkovich [1990] introduced a generalisation of rigid geometry in terms of seminorms. The sheaves $\Omega^n_{\mathrm{val}\le 1}$ seem to be connected to Berkovich spaces. We review a very small part of the theory.

Recall that a *multiplicative nonarchimedean norm* on a field K is a group homomorphism $|\cdot| : K^* \to \mathbb{R}_{>0}$ (the latter equipped with multiplication) such that $|f+g| \le \max(|f|, |g|)$. This is usually extended to a map of sets $|\cdot| : K \to \mathbb{R}_{\ge 0}$ by setting $|0| = 0$.

Key Lemma 7.1. *The set of multiplicative nonarchimedean norms on a field K is the same as the set of pairs $(R, |\cdot|)$ where R is a valuation ring of K, and $|\cdot| : K^*/R^* \to \mathbb{R}_{>0}$ is an injective group homomorphism. Under this bijection, the ring R corresponds to $|\cdot|^{-1}[0,1]$. Since $\mathbb{R}_{>0}$ has no convex subgroups, such valuation rings R necessarily have rank* 1 (*or rank* 0 *if $R = K$*).

Proof. Obvious. □

Let K be a field equipped with a multiplicative nonarchimedean norm $\|\cdot\| : K^* \to \mathbb{R}_{>0}$. If X is a K-variety, the *Berkovich space* X^{an} of X, as a set, consists of pairs $(x, |\cdot|)$ where x is a point of X, and $|\cdot| : \kappa(x) \to \mathbb{R}_{>0}$ is a multiplicative nonarchimedean norm extending $\|\cdot\|$. This set is equipped with a structure of locally ringed space such that the projection $\pi : X^{\mathrm{an}} \to X; (x, |\cdot|) \mapsto x$ is a morphism of locally ringed spaces.

On the other hand, recall that Lemma 3.7 described $\Omega^n_{\mathrm{val}}(X)$ as

$$\left\{ (s_x) \in \prod_{x \in X} \Omega^n(x) \;\middle|\; \begin{array}{r} \text{for every } k\text{-valuation ring } R \subseteq k(x) \text{ of rank} \le 1 \text{ we have} \\ s_x \in \Omega^n(R) \text{ and } s_x|_{R/\mathfrak{m}} = s_y|_{R/\mathfrak{m}} \text{ where } y = \mathrm{Image}(\mathrm{Spec}(R/\mathfrak{m})) \end{array} \right\}. \tag{21}$$

In particular, every section $s \in \Omega^n_{\mathrm{val}}(X)$ gives a function $X^{\mathrm{an}} \to \coprod_{(x,|\cdot|)} \Omega^n(\mathscr{H}(x))$ such that the image of $(x, |\cdot|)$ lands in the corresponding component. Here, $\mathscr{H}(x)$ is the completion $\widehat{\kappa(x)}$ of the normed field $\kappa(x)$. Similarly, we could apply Ω^n to the structure sheaf of the locally ringed space X^{an}, and obtain a ring morphism $\Omega^n_{X^{\mathrm{an}}/K}(X^{\mathrm{an}}) \to \coprod_{(x,|\cdot|)} \Omega^n(\mathscr{H}(x))$.

Question 7.2. Is the image of one of

$$\Omega^n_{X^{\mathrm{an}}/K}(X^{\mathrm{an}}) \to \coprod_{(x,|\cdot|)} \Omega^n(\mathscr{H}(x)) \quad \text{and} \quad \Omega^n_{\mathrm{val}}(X) \to \coprod_{(x,|\cdot|)} \Omega^n(\mathscr{H}(x))$$

contained in the image of the other? Does $\Omega^n_{\mathrm{val}}(X) = \Omega^n_{\mathrm{cdh}}(X)$ have an intrinsic description in terms of analytic spaces?

7B. *F-singularities and reflexive differentials.* Recall that reflexive differentials are defined as the double dual $\Omega^{[n]}_X = (\Omega^n_X)^{**}$. One of the results of [Huber and Jörder 2014] is that on a klt base space X, cdh-differentials recover reflexive differential forms:

$$\Omega_X^{[n]}(X) = \Omega^n_{\mathrm{cdh}}(X).$$

One can ask when such a formula is possible in positive characteristic. For example, what is an appropriate replacement for the klt base space hypothesis?

There is an active area of research in positive characteristic birational geometry studying singularities defined via the Frobenius which are analogues of singularities arising in the minimal model program. These former are called F-singularities; log terminal, log canonical, rational and du Bois correspond to F-regular, F-pure/F-split, F-rational and F-injective, respectively; see [Schwede 2010, Remark 17.11]. Under this dictionary, Kawamata log terminal (i.e., klt) corresponds to strongly F-regular [Schwede 2010, Corollary 17.10].

Consequently, we arrive at the following question.

Question 7.3 (Blickle). If a normal scheme X is strongly F-regular, do we have $\Omega_X^{[n]}(X) \cong \Omega^n_{\mathrm{cdh}}(X)$?

In the special case $n = d_X = \dim X$, we have

$$\Omega^{d_X}_{\mathrm{cdh}}(X) = \Omega^{d_X}_{\mathrm{val}}(X) = \varprojlim_{R\in \mathrm{val}(X)} \Omega^{d_X}(R) = \varprojlim_{R\in \mathrm{val}(X), R\subseteq k(X)} \Omega^{d_X}(R) = \Omega^{d_X}_{RZ(X)}(RZ(X)),$$

and so the question becomes:

Question 7.4. If a normal scheme X of dimension d_X is strongly F-regular, do we have $\Omega_X^{[d_X]}(X) \cong \varinjlim_{X'\to X} \Omega^{d_X}_{X'}(X')$? Here the colimit is over proper birational morphisms $X' \to X$.

Remark 7.5. Under the assumption of resolution of singularities, Question 7.4 is true: being strongly F-regular implies being pseudorational, which means (by definition) that for any proper birational morphism $\pi : X' \to X$ the direct image $\pi_*\omega_{X'}$ of the canonical dualizing sheaf $\omega_{X'}$ of X' is ω_X, the canonical dualizing sheaf of X. If X' is smooth over the base, we have $\omega_{X'} = \Omega^{d_X}_{X/k}$. On the other hand, we also have $\Omega_X^{[d_X]} = \omega_X$. So if we restrict the colimit to those X' which are smooth, we have

$$\Omega_X^{[d_X]} = \omega_X = \varinjlim_{\substack{X'\to X\\ X'\text{smooth}}} \omega_{X'}(X') = \varinjlim_{\substack{X'\to X\\ X'\text{smooth}}} \Omega^{d_X}_{X'}(X').$$

Under the assumption of resolution of singularities, the colimit over X' smooth is the same as the colimit over all X'.

We remark that an alternative description of reflexive differentials when X is klt of characteristic zero is given in [Greb et al. 2011] by $\Omega_X^{[p]} \cong \pi_*\Omega^p_{X'}$, where $\pi : X' \to X$ is a log resolution, and that this implies an isomorphism $\Omega_X^{[d_X]}(X) \cong \Omega^{d_X}_{\mathrm{val}}(X)$ (in characteristic zero), where $d_X = \dim X$. That is, the characteristic zero version of Question 7.3 is true.

Let us list some facts about strongly F-regular schemes that allow us to replace the F-regular hypothesis in Question 7.3 with an equivalent more explicit hypothesis, which may help develop a strategy for a proof.

To begin with, a ring R is *strongly F-regular* if and only if its test ideal is equal to R [Schwede 2010, Proposition 16.9]. This implies that R is Cohen–Macaulay, and in particular, that the sheaf ω_R is a dualising object in the derived category.

As for test ideals, it is shown in [Blickle et al. 2015] that in positive characteristic, the test ideal of a normal variety X over a perfect field can be defined as the intersection of the images of certain trace maps, with the intersection taken over all generically finite proper separable maps $\pi : Y \to X$ with Y regular. Here, the trace map comes from the trace morphism $\pi_! \pi^! \omega_X^\bullet \to \omega_X^\bullet$. Moreover, there exists a $Y \to X$ in the indexing set whose image agrees with this intersection.

So in the affine case, with our replacement hypothesis which is equivalent to "strongly F-regular", Question 7.3 becomes:

Question 7.6. Let X be a normal affine variety over a perfect field of positive characteristic, with $\mathbb{Q}$-Cartier canonical divisor. Suppose that for every generically finite proper separable morphism $\pi : Y \to X$ with regular source, the trace morphism

$$\pi_* \mathcal{O}_Y(\lceil K_Y - \pi^* K_X \rceil) \xrightarrow{\text{trace}} \mathcal{O}_X.$$

is surjective; cf. [Blickle et al. 2015, Main Theorem]. Let $\iota : X_{\mathrm{reg}} \to X$ be the inclusion of the regular locus. Is the canonical morphism

$$\Omega^n_{\mathrm{val}}(X) \to \Omega^n_{\mathrm{val}}(X_{\mathrm{reg}})$$

an isomorphism?

In this formulation, we have used that for any normal scheme X with regular locus $\iota : X_{\mathrm{reg}} \to X$ one has $\Omega_X^{[n]} = \iota_* \Omega^n_{X_{\mathrm{reg}}}$. Since we know that $\Omega^n = \Omega^n_{\mathrm{val}}$ on regular schemes, we obtain the description

$$\Omega_X^{[n]}(X) = (\iota_* \Omega^n_{X_{\mathrm{reg}}})(X) = \Omega^n_{X_{\mathrm{reg}}}(X_{\mathrm{reg}}) = \Omega^n_{\mathrm{val}}(X_{\mathrm{reg}}).$$

Acknowledgements

It is a pleasure to thank Manuel Blickle, Fumiharu Kato, Stefan Kebekus, Simon Pepin-Lehalleur, Kay Rülling, David Rydh and Vasudevan Srinivas for inspiring discussions and answers to our mathematical questions. We also greatly appreciate the referee's careful reading.

References

[Andreotti and Norguet 1967] A. Andreotti and F. Norguet, "La convexité holomorphe dans l'espace analytique des cycles d'une variété algébrique", *Ann. Scuola Norm. Sup. Pisa* (3) **21** (1967), 31–82. MR Zbl

[Berkovich 1990] V. G. Berkovich, *Spectral theory and analytic geometry over non-Archimedean fields*, Mathematical Surveys and Monographs **33**, American Mathematical Society, Providence, RI, 1990. MR Zbl

[Blickle et al. 2015] M. Blickle, K. Schwede, and K. Tucker, "F-singularities via alterations", *Amer. J. Math.* **137**:1 (2015), 61–109. MR Zbl

[Bourbaki 1964] N. Bourbaki, *Éléments de mathématique: Algèbre commutative, Chapitres 5–6*, Actualités Scientifiques et Industrielles **1308**, Hermann, Paris, 1964. MR Zbl

[Cisinski 2013] D.-C. Cisinski, "Descente par éclatements en K-théorie invariante par homotopie", *Ann. of Math.* (2) **177**:2 (2013), 425–448. MR Zbl

[Cisinski and Déglise 2015] D.-C. Cisinski and F. Déglise, "Integral mixed motives in equal characteristic", *Doc. Math.* Extra volume: Alexander S. Merkurjev's sixtieth birthday (2015), 145–194. MR Zbl

[Cortiñas et al. 2008] G. Cortiñas, C. Haesemeyer, M. Schlichting, and C. Weibel, "Cyclic homology, cdh-cohomology and negative K-theory", *Ann. of Math.* (2) **167**:2 (2008), 549–573. MR Zbl

[EGA II 1961] A. Grothendieck, "Eléments de géométrie algébrique, II: Étude globale élémentaire de quelques classes de morphismes", *Inst. Hautes Études Sci. Publ. Math.* **8** (1961), 5–222. MR Zbl

[EGA III_1 1961] A. Grothendieck, "Eléments de géométrie algébrique, III: Étude cohomologique des faisceaux cohérents, I", *Inst. Hautes Études Sci. Publ. Math.* **11** (1961), 5–167. MR Zbl

[EGA IV_3 1966] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* **28** (1966), 5–255. MR Zbl

[Engler and Prestel 2005] A. J. Engler and A. Prestel, *Valued fields*, Springer, 2005. MR Zbl

[Ferrand 2003] D. Ferrand, "Conducteur, descente et pincement", *Bull. Soc. Math. France* **131**:4 (2003), 553–585. MR Zbl

[Friedlander and Voevodsky 2000] E. M. Friedlander and V. Voevodsky, "Bivariant cycle cohomology", pp. 138–187 in *Cycles, transfers, and motivic homology theories*, Ann. of Math. Stud. **143**, Princeton University Press, 2000. MR Zbl

[Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Mathematik (3) **2**, Springer, 1998. MR Zbl

[Gabber and Kelly 2015] O. Gabber and S. Kelly, "Points in algebraic geometry", *J. Pure Appl. Algebra* **219**:10 (2015), 4667–4680. MR Zbl

[Gabber and Ramero 2003] O. Gabber and L. Ramero, *Almost ring theory*, Lecture Notes in Math. **1800**, Springer, 2003. MR Zbl

[Gabriel and Zisman 1967] P. Gabriel and M. Zisman, *Calculus of fractions and homotopy theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete **35**, Springer, 1967. MR Zbl

[Geisser 2006] T. Geisser, "Arithmetic cohomology over finite fields and special values of ζ-functions", *Duke Math. J.* **133**:1 (2006), 27–57. MR Zbl

[Geisser and Hesselholt 2010] T. Geisser and L. Hesselholt, "On the vanishing of negative K-groups", *Math. Ann.* **348**:3 (2010), 707–736. MR Zbl

[Goodwillie and Lichtenbaum 2001] T. G. Goodwillie and S. Lichtenbaum, "A cohomological bound for the h-topology", *Amer. J. Math.* **123**:3 (2001), 425–443. MR Zbl

[Greb et al. 2011] D. Greb, S. Kebekus, S. J. Kovács, and T. Peternell, "Differential forms on log canonical spaces", *Publ. Math. Inst. Hautes Études Sci.* **114** (2011), 87–169. MR Zbl

[Huber 2016] A. Huber, "Differential forms in algebraic geometry — a new perspective in the singular case", *Port. Math.* **73**:4 (2016), 337–367. MR Zbl

[Huber and Jörder 2014] A. Huber and C. Jörder, "Differential forms in the h-topology", *Algebr. Geom.* **1**:4 (2014), 449–478. MR Zbl

[Huber and Kelly $\geq$ 2018] A. Huber and S. Kelly, "Differential forms in positive characteristic, III: Towards cdh cohomological descent", work in progress.

[Huber, Kebekus and Kelly 2017] A. Huber, S. Kebekus, and S. Kelly, "Differential forms in positive characteristic avoiding resolution of singularities", *Bull. Soc. Math. France* **145**:2 (2017), 305–343. MR Zbl

[de Jong 1996] A. J. de Jong, "Smoothness, semi-stability and alterations", *Inst. Hautes Études Sci. Publ. Math.* **83** (1996), 51–93. MR Zbl

[Kelly 2012] S. Kelly, *Triangulated categories of motives in positive characteristic*, Ph.D. thesis, Université Paris 13, 2012. arXiv

[Kelly 2017] S. Kelly, *Voevodsky motives and $l$dh-descent*, Astérisque **391**, Société Mathématique de France, Paris, 2017. MR Zbl

[Kerz et al. 2018] M. Kerz, F. Strunk, and G. Tamme, "Algebraic K-theory and descent for blow-ups", *Invent. Math.* **211**:2 (2018), 523–577. MR Zbl

[Matsumura 1970] H. Matsumura, *Commutative algebra*, W. A. Benjamin, New York, 1970. MR Zbl

[Morrow 2016] M. Morrow, "A historical overview of pro cdh descent in algebraic K-theory and its relation to rigid analytic varieties", preprint, 2016. arXiv

[Olivier 1983] J.-P. Olivier, "Going up along absolutely flat morphisms", *J. Pure Appl. Algebra* **30**:1 (1983), 47–59. MR Zbl

[Rydh 2010] D. Rydh, "Submersions and effective descent of étale morphisms", *Bull. Soc. Math. France* **138**:2 (2010), 181–230. MR Zbl

[Schwede 2010] K. Schwede, "F-singularities and Frobenius splitting notes", lecture notes, University of Utah, 2010, available at https://www.math.utah.edu/~schwede/frob/RunningTotal.pdf.

[Stacks project] P. Belmans, A. J. de Jong, et al., "The Stacks project", electronic reference, 2005–, available at http://stacks.math.columbia.edu.

[Suslin and Voevodsky 2000a] A. Suslin and V. Voevodsky, "Bloch–Kato conjecture and motivic cohomology with finite coefficients", pp. 117–189 in *The arithmetic and geometry of algebraic cycles* (Banff, AB, 1998), edited by B. B. Gordon et al., NATO Sci. Ser. C Math. Phys. Sci. **548**, Kluwer Academic Publishers, Dordrecht, Netherlands, 2000. MR Zbl

[Suslin and Voevodsky 2000b] A. Suslin and V. Voevodsky, "Relative cycles and Chow sheaves", pp. 10–86 in *Cycles, transfers, and motivic homology theories*, Ann. of Math. Stud. **143**, Princeton University Press, 2000. MR Zbl

[Swan 1980] R. G. Swan, "On seminormality", *J. Algebra* **67**:1 (1980), 210–229. MR Zbl

[Temkin 2011] M. Temkin, "Relative Riemann–Zariski spaces", *Israel J. Math.* **185** (2011), 1–42. MR Zbl

[Traverso 1970] C. Traverso, "Seminormality and Picard group", *Ann. Scuola Norm. Sup. Pisa* (3) **24** (1970), 585–595. MR Zbl

[Vitulli 2011] M. A. Vitulli, "Weak normality and seminormality", pp. 441–480 in *Commutative algebra — Noetherian and non-Noetherian perspectives*, edited by M. Fontana et al., Springer, 2011. MR Zbl

[Voevodsky 1996] V. Voevodsky, "Homology of schemes", *Selecta Math.* (*N.S.*) **2**:1 (1996), 111–153. MR Zbl

[Voevodsky 2000] V. Voevodsky, "Triangulated categories of motives over a field", pp. 188–238 in *Cycles, transfers, and motivic homology theories*, Ann. of Math. Stud. **143**, Princeton University Press, 2000. MR Zbl

[Weibel 1994] C. A. Weibel, *An introduction to homological algebra*, Cambridge Studies in Advanced Mathematics **38**, Cambridge University Press, 1994. MR Zbl

[Zariski and Samuel 1960] O. Zariski and P. Samuel, *Commutative algebra*, vol. II, D. Van Nostrand Co., Princeton, NJ, 1960. MR Zbl

Communicated by János Kollár
Received 2017-07-05 Revised 2018-01-10 Accepted 2018-03-10

annette.huber@math.uni-freiburg.de *Mathematisches Institut, Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany*

shanekelly@math.titech.ac.jp *Department of Mathematics, Tokyo Institute of Technology, Tokyo, Japan*

ALGEBRA AND NUMBER THEORY 12:3 (2018)
dx.doi.org/10.2140/ant.2018.12.693

Nilpotence order growth of recursion operators in characteristic p

Anna Medvedovsky

We prove that the killing rate of certain degree-lowering "recursion operators" on a polynomial algebra over a finite field grows slower than linearly in the degree of the polynomial attacked. We also explain the motivating application: obtaining a lower bound for the Krull dimension of a local component of a big mod p Hecke algebra in the genus-zero case. We sketch the application for $p=2$ and $p=3$ in level one. The case $p=2$ was first established in by Nicolas and Serre in 2012 using different methods.

1. Introduction	693
2. Preliminaries	695
3. The nilpotence growth theorem (NGT)	697
4. A toy case of the NGT	699
5. Applications to mod p Hecke algebras	701
6. The proof of the NGT begins	703
7. The content function and its properties	707
8. Content of some proper fractions	711
9. The nilgrowth witness: finishing the proof	716
10. Complements	718
Acknowledgements	720
References	721

1. Introduction

The main goal of this document is to prove the following *nilpotence growth theorem*, about the killing rate of a recursion operator on a polynomial algebra over a finite field under repeated application:

Theorem A (nilpotence growth theorem; see also Theorem 1). *Let $\mathbb{F}$ be a finite field of characteristic p, and suppose that $T:\mathbb{F}[y]\to\mathbb{F}[y]$ is a degree-lowering $\mathbb{F}$-linear operator satisfying the following condition*:

> *The sequence $\{T(y^n)\}_n$ of polynomials in $\mathbb{F}[y]$ satisfies a linear recursion over $\mathbb{F}[y]$ whose companion polynomial $X^d+a_1X^{d-1}+\cdots+a_d\in\mathbb{F}[y][X]$ has both total degree d and y-degree d.*[(i)]

Then there exists a constant $\alpha<1$ so that the minimum power of T that kills y^n is $O(n^\alpha)$.

MSC2010: primary 11T55; secondary 11B85, 11F03, 11F33.
Keywords: linear recurrences in characteristic p, modular forms modulo p, congruences between modular forms, mod p Hecke algebras, p-regular sequences, base representation of numbers.

[(i)] The *companion polynomial* of a linear recurrence $s_n=a_1s_{n-1}+\cdots+a_ds_{n-d}$ satisfied by a sequence $\{s_n\}_n$ for all $n\geq d$ is $X^d-a_1X^{d-1}-\cdots-a_d$. See Section 2D.

To prove this theorem, we reduce to the case where the companion polynomial of the recursion has an "empty middle" in its degree-d homogeneous part: that is, when for some $a \in \mathbb{F}$ it has the form $X^d + ay^d +$ (terms of total degree $< d$). Then we prove this empty-middle case (see Theorem 4 below) by constructing a function $c : \mathbb{F}[y] \to \mathbb{N} \cup \{-\infty\}$ that grows like $(\deg f)^\alpha$ and whose value is lowered by every application of T. In the special case where d is a power of p, the function c takes y^n to the integer obtained by writing n in base d and then reading the expansion in some smaller base, so that the sequence $\{c(y^n)\}_n$ is p-regular in the sense of Allouche and Shallit [1992]. The proof that $c(T(y^n)) < c(y^n)$, by strong induction, uses higher-order recurrences depending on n, so that n is compared to numbers whose base-d expansion is not too different.

It is the author's hope that ideas from p-automata theory can eventually be used to sharpen and generalize the nilpotence growth theorem.

Motivating application of the nilpotence growth theorem. The motivating application for the nilpotence growth theorem (Theorem A above) is the *nilpotence method* for establishing lower bounds on dimensions of local components of Hecke algebras acting on mod p modular forms of tame level N. These Hecke algebra components were first studied by Jochnowitz [1982] in the 1970s, but the first full structure theorem did not appear until over thirty years later. In 2012 Nicolas and Serre used recurrences satisfied by Hecke operators (see (5-1)) to describe the Hecke action on modular forms modulo 2 completely explicitly [Nicolas and Serre 2012a], leading to a Hecke algebra structure result for $p = 2$ and $N = 1$ [Nicolas and Serre 2012b]. Unfortunately their explicit formulas appear not to generalize beyond $p = 2$. The structure of mod p Hecke algebras for $p \geq 5$ was subsequently established by very different techniques by Bellaïche and Khare [2015] for $N = 1$ and later generalized by Deo [2017] to all N. The Bellaïche–Khare method deduces information about mod p Hecke algebra components from corresponding characteristic-zero Hecke algebra components, which are known to be big by the Gouvêa–Mazur "infinite fern" construction ([Gouvêa and Mazur 1998]; see also [Emerton 2011, Corollary 2.28]). The nilpotence method is yet a third technique, coming out of an idea of Bellaïche for tackling the case $p = 3$ and $N = 1$ as outlined in [Bellaïche and Khare 2015, Appendix], and implemented and developed in level one for $p = 2, 3, 5, 7, 13$ in the present author's Ph.D. dissertation [Medvedovsky 2015]. Like the Nicolas–Serre approach, the nilpotence method stays entirely in characteristic p and makes use of Hecke recurrences; but instead of explicit Hecke action formulas, the nilpotence growth theorem now plays the crucial dimension-bounding role. See Section 5 below for a taste of this method for $p = 2, 3$, which completes the determination of the structure of the Hecke algebra for $p = 3$ begun in [Bellaïche and Khare 2015, Appendix] and recovers the Nicolas–Serre result for $p = 2$. In fact, the nilpotence method via the nilpotence growth theorem in its current form gives lower bounds on dimensions of mod p Hecke algebras of level N so long as the genus of the modular curve $X_0(Np)$ is zero; see [Medvedovsky 2015] and the forthcoming [Medvedovsky $\geq$ 2018] for details.

Structure of this document. After a few preliminary definitions in Section 2, we state and discuss a more general version of the nilpotence growth theorem (NGT); see Theorem 1 in Section 3. In Section 4, we prove a toy version of the NGT (Theorem 2). In Section 5, we use the toy version of NGT to prove that the mod p level-one Hecke algebra for $p = 2, 3$ has the form $\mathbb{F}_p[\![x, y]\!]$. This section illustrates the motivating application of the nilpotence growth theorem and is not required for the rest of the document. This is a reasonable stopping point for a first reading.

In Section 6 the proof of the NGT begins in earnest. There is a short overview of the structure of the proof in Section 6A. In Section 6B, we reduce to working over a finite field. In Section 6C, we reduce to the empty-middle NGT (Theorem 4). In Section 6D, we give the inductive argument that reduces the proof of the empty-middle NGT to finding a *nilgrowth witness* function satisfying certain properties. The next three sections are combinatorial in nature, as we construct a nilgrowth witness. In Section 7, we discuss base-b representation of numbers and introduce the *content* function. In Section 8, we prove a number of technical inequalities about the content function. In Section 9 we finally construct a nilgrowth witness out of the content function, finishing the proof of the empty-middle NGT, and hence of the NGT in full. Finally in Section 10, we state a more precise version of the toy NGT and speculate on the optimality of some bounds.

2. Preliminaries

This section contains a brief review of a few unconnected algebraic notions. All rings and algebras are assumed to be commutative, with unity. We use the convention that the set of natural numbers starts with zero: $\mathbb{N} = \{0, 1, 2, \ldots\}$. Below, R is always a ring.

2A. ***Structure of finite rings.*** If R is finite, then R is artinian, hence a finite product of finite local rings. If R is a finite local ring with maximal ideal $\mathfrak{m}$, then the residue field $R/\mathfrak{m}$ is a finite field of characteristic p. Moreover, the graded pieces $\mathfrak{m}^n/\mathfrak{m}^{n+1}$ are finite $R/\mathfrak{m}$-vector spaces, so that R has cardinality a power of p. Basic examples of finite local rings include $\mathbb{F}_p[t]/(t^k)$ and $\mathbb{Z}/p^k\mathbb{Z}$.

2B. ***Degree filtration on a polynomial algebra.*** If $0 \neq f = \sum_{n\geq 0} c_n y^n$ is a polynomial in $R[y]$, then its *y-degree*, or just *degree*, is as usual defined to be $\deg f := \max\{n : c_n \neq 0\}$. For $f = 0$, set $\deg f := -\infty$.

The degree function gives $R[y]$ the structure of a *filtered algebra*. Let $R[y]_n := \{f \in R[y] : \deg f \leq n\}$, and then $R[y] = \bigcup_{n\geq 0} R[y]_n$ and multiplication preserves the filtration as required.

2C. ***Local nilpotence and the nilpotence index.*** Let M be any R-module and $T \in \operatorname{End}_R(M)$ an R-linear endomorphism. (In applications to Hecke algebras, R will be a finite field, M an infinite-dimensional R-algebra of modular forms, and T a Hecke operator.)

The operator $T : M \to M$ is *locally nilpotent* on M if every element of M is annihilated by some power of T. If T is locally nilpotent and f in M is nonzero, we define the *nilpotence index of f with respect to T* as

$$N_T(f) := \max\{k \geq 0 : T^k f \neq 0\}.$$

Also set $N_T(0) := -\infty$.

Suppose $R = K$ is a field, $M = K[y]$, and $T : M \to M$ preserves the degree filtration; that is, $T(K[y]_n) \subset K[y]_n$. Then T is locally nilpotent if and only if T strictly lowers degrees, in which case we also have $N_T(f) \leq \deg f$.

For example, $T = \frac{d}{dy}$ is locally nilpotent on $K[y]$. If K has characteristic zero, then $N_T(f) = \deg f$; otherwise $N_T(f) \leq \operatorname{char} K - 1$.

2D. *Linear recurrences and companion polynomials.* Now suppose that M is an R-algebra, and M' is an M-module (we will usually take $M' = M$). A sequence $s = \{s_n\} \in M'^{\mathbb{N}}$ satisfies an *M-linear recurrence of order d* if there exist elements $a_0, a_1, \ldots, a_d \in M$ so that

$$a_0 s_n = a_1 s_{n-1} + \cdots + a_d s_{n-d} \quad \text{for all } n \geq d. \tag{2-1}$$

Unlike some authors, we do not assume that a_d is nonzero or not a zero divisor, but we do insist that the recursion already hold for $n = d$. The *companion polynomial* of this linear recurrence is $P(X) = a_0 X^d - a_1 X^{d-1} - \cdots - a_d \in M[X]$. If $a_0 = 1$, then the recurrence is said to be *monic*; we will always assume below that our linear recurrences are monic unless stated otherwise.

Example. The sequence $s = \{0, 1, y, y^2, y^3, y^4, \ldots\} \in R[y]^{\mathbb{N}}$ satisfies an $R[y]$-linear recursion of minimal order 2; we have $s_n = y s_{n-1}$ for all $n \geq 2$, but not for $n = 1$. The companion polynomial of the recurrence is therefore $X^2 - yX$.

Given any sequence s in $M'^{\mathbb{N}}$, the set of companion polynomials of (not necessarily monic) M-linear recurrences satisfied by s forms an ideal of $M[X]$. We record this observation in the following form:

Fact. If a sequence $s \in M'^{\mathbb{N}}$ satisfies the recurrence defined by some monic $P \in M[X]$, then it also satisfies the recurrence defined by PQ for any other monic $Q \in M[X]$.

In characteristic p we get the following corollary, of which we will make crucial use:

Corollary 2.1. *If R has characteristic p and $s \in M'^{\mathbb{N}}$ satisfies the order-d recurrence*

$$s_n = a_1 s_{n-1} + a_2 s_{n-2} + \cdots + a_d s_{n-d} \quad \textit{for all } n \geq d,$$

then for every $k \geq 0$ the sequence s also satisfies the order-dp^k deeper *recurrence*

$$s_n = a_1^{p^k} s_{n-p^k} + a_2^{p^k} s_{n-2p^k} + \cdots + a_d^{p^k} s_{n-dp^k} \quad \textit{for all } n \geq dp^k. \tag{2-2}$$

Proof. Let $P = X^d - a_1 X^{d-1} - \cdots - a_d$ be the companion polynomial of a recursion satisfied by s. By the fact above, the sequence s also satisfies the recurrence whose companion polynomial is

$$P^{p^k} = X^{dp^k} - a_1^{p^k} X^{dp^k - p^k} - a_2^{p^k} X^{dp^k - 2p^k} - \cdots - a_d^{p^k},$$

which is exactly what is expressed in (2-2). □

If M is a domain embedded into a field K, we have the following well-known characterization of power sequences in $\overline{K}^{\mathbb{N}}$ satisfying a fixed M-linear recurrence:

Fact. An element α in $\overline{K}$ is a root of monic $P \in M[X]$ if and only if the sequence $\{\alpha^n\}_n = \{1, \alpha, \alpha^2, \dots\}$ satisfies the linear recurrence with companion polynomial P.

If the companion polynomial of such an M-linear recurrence has no repeated roots in $\overline{K}$, it follows from the proposition that *every* solution to the recurrence is a linear combination of such power sequences on the roots of the companion polynomial. One can further describe all $\overline{K}$-sequences satisfying a general M-linear recursion — see, for example, [Conrad 2016], particularly the historical references on page 2 — but we will not need this below.

3. The nilpotence growth theorem (NGT)

3A. ***Statement of the NGT.*** We are now ready to state the most general version of the nilpotence growth theorem (NGT). From now on, we will assume R to be a *finite* ring, and $M = R[y]$.

Theorem 1 (nilpotence growth theorem). *Let R be a finite ring, and suppose that $T : R[y] \to R[y]$ is an R-linear operator satisfying the following two conditions*:

(1) *T lowers degrees*: $\deg T(f) < \deg f$ *for every nonzero f in* $R[y]$.

(2) *The sequence $\{T(y^n)\}_n$ satisfies a* filtered *linear recursion over* $R[y]$: *there exist $a_1, \dots, a_d \in R[y]$, with* $\deg a_i \le i$ *for each i, so that for all $n \ge d$,*

$$T(y^n) = a_1 T(y^{n-1}) + \cdots + a_d T(y^{n-d}).$$

Suppose further that

(3) *the coefficient of y^d in a_d is invertible in R.*

Then there exists a constant $\alpha < 1$ so that $N_T(y^n) \ll n^\alpha$.

In other words, Theorem 1 implies that, under a mild technical assumption (condition (3)), the nilpotence index of a degree-lowering operator defined by a filtered linear recursion grows *slower than linearly* in the degree. The mild technical assumption is necessary in the theorem as stated; see the discussion in (4) in Section 3B below.

3B. ***Discussion of the NGT.*** (1) *Connection with* Theorem A: If $T : R[y] \to R[y]$ satisfies the conditions of Theorem 1, then the companion polynomial of the recursion satisfied by the sequence $\{T(y^n)\}_n$ is

$$P_T = X^d - a_1 X^{d-1} - \cdots - a_d \in R[y, X].$$

The condition $\deg a_i \le i$ from (2) guarantees that the total degree of P_T is exactly d. In particular, in the case where $R = \mathbb{F}$ is a finite field, condition (3) implies that $\deg_y P_T = \deg a_d = d$. In other words, Theorem 1 over a finite field reduces to Theorem A.

(2) *Condition* (1) *guarantees that T is locally nilpotent*: Moreover, $N_T(y^n) \le n$, so that the function $n \mapsto N_T(y^n)$ a priori grows no faster than linearly.

(3) *Condition* (2) *and connection to recursion operators*: The condition that the sequence $\{T(y^n)\}_n$ satisfies a linear recurrence is the definition of a *recursion operator*, a notion that will be explored in a future paper. A natural source of *filtered* recursion operators (that is, satisfying additional degree bounds as in condition (2) above) comes from the action of Hecke operators on algebras of modular forms of a fixed level. Namely, *if f is a modular form of weight k and level N and T is a Hecke operator acting on the algebra M of forms of level N, then the sequence $\{T(f^n)\}_n$ satisfies an M-linear recursion with companion polynomial $X^d + a_1 X^{d-1} + \cdots + a_d$, where a_i comes from weight ki.*

See (5-1) and (5-2) below for examples over $\mathbb{F}_p$, [Medvedovsky 2015, chapter 6] for a proof in level one when T is a prime Hecke operator, or [Medvedovsky $\geq$ 2018] for more details.

The history of Hecke recurrences appears to be relatively brief. Hecke recurrences over $\mathbb{F}_2$ were crucially used by Nicolas and Serre to obtain the structure of the mod 2 Hecke algebra in level one [Nicolas and Serre 2012a; 2012b]. Earlier, Al Hajj Shehadeh, Jaafar, and Khuri-Makdisi [2009] had investigated two-dimensional Hecke recurrences over $\mathbb{Q}$ satisfied by the array $\{T_\ell(E_4^n E_6^m)\}_{n,m}$; Buzzard and Calegari [2005, p. 594] had used Hecke recurrences for U_2 acting on a power basis of overconvergent 2-adic modular functions in their study of slopes.

(4) *Condition* (3) *is necessary as stated*: Consider the operator $T : R[y] \to R[y]$ defined by $T(y^n) = s_n$, where $\{s_n\}$ is the sequence $\{0, 1, y, y^2, \ldots\}$ with companion polynomial $X^2 - yX$ from the example on page 696. All conditions except (3) are satisfied, and it is easy to see that $N_T(y^n) = n$ in this case.

For an example with $a_d \neq 0$, consider the operator T with the defining companion polynomial $P_T = X^2 + yX + y$ and initial values $[T(1), T(y)] = [0, 1]$. By induction, $\deg T(y^n) = n - 1$. Therefore $N_T(y^n) = n$.

Computationally, it appears that if $R = \mathbb{F}_p$ and $\deg a_d < d$ but there exists an i with $0 < i < d$ so that $\deg a_i = i$, then either N_T grows logarithmically or else it grows linearly. In that sense, it appears that "fullness" of degree at the end of P_T (that is, the presence of a y^d term) appears to be, at least generically, necessary to compensate for "fullness" of degree in the middle (that is, the presence of a $y^i X^{d-i}$ term for some $0 < i < d$), if one wants the growth of N_T to be sublinear but not degenerate. But the phenomenon is not well understood.

(5) *The constant* α: The power α depends on R and d only, and tends to 1 as $d \to \infty$. More precisely, the dependence on R is only through its maximal residue characteristic; the length of R as a module over itself affects only the implicit constant of the growth condition $N_T(y^n) \ll n^\alpha$. If the inequality $\deg a_i < i$ is strict for every $i < d$ ("empty middle" case) we can take α to be $\log_{p^k}(p^k - 1)$ for k satisfying $d \leq p^k$. See Theorems 2 or 4 below.

(6) *Finite characteristic is necessary — a counterexample in characteristic zero*: Consider the operator T on $\mathbb{Q}[y]$ with $P_T = X^2 - yX - y^2$ and degree-lowering initial terms $[T(1), T(y)] = [0, 1]$. This satisfies all three conditions of the NGT. It is easy to see that $T(y^n) = F_n y^{n-1}$, where F_n is the n-th Fibonacci number; the recursion is $s_n = y s_{n-1} + y^2 s_{n-2}$. Therefore

$$T^k(y^n) = F_n F_{n-1} \cdots F_{n-k+1} y^{n-k},$$

so that $N_T(y^n) = n$. (Compare to characteristic p, where the operator defined by $T(y^n) = F_n y^{n-1}$ on $\mathbb{F}_p[y]$ satisfies $T^{p+1} \equiv 0$.) See also Proposition 10.1 for a family of examples in any degree.

Computationally, it appears that generic characteristic-zero examples that do not degenerate (to $\log n$ growth) all exhibit linear growth. In contrast, over a finite field, one observes $O(n^\alpha)$ growth for various $\alpha < 1$.

(7) *Finiteness of R is necessary as stated — a counterexample over $\mathbb{F}_p(t)$, due to Paul Monsky*: Let $P_T = X^2 - tyX - y^2$ and start with $[0, 1]$ again. Then $T(y^n) = F_n(t)y^{n-1}$ with $F_n(t) \in \mathbb{F}_p[t]$ monic of degree $n-1$, so that $N_T(y^n) = n$ again. However, see the empty-middle case (Theorem 4) for a special case that does hold for infinite rings of characteristic p.

4. A toy case of the NGT

Fix a prime p and take $R = \mathbb{F}_p$ for simplicity (in fact any ring of characteristic p works here). We prove the following special case of the NGT for recurrences with empty middle (i.e., whose companion polynomials have no maximal-degree cross terms) and whose order is a power of p.

Theorem 2 (toy case of NGT). *Let $q = p^k$ for some $k \geq 1$. Suppose $T : \mathbb{F}_p[y] \to \mathbb{F}_p[y]$ is a degree-lowering linear operator so that the sequence $\{T(y^n)\}_n$ satisfies an $\mathbb{F}_p[y]$-linear recursion with companion polynomial*

$$P = X^q + (\textit{terms of total degree} < q) + ay^q \in \mathbb{F}_p[y][X]$$

for some $a \in \mathbb{F}_p$. Then $N_T(y^n) \ll n^{\log(q-1)/\log q}$.

Most of the main features of the proof of the NGT (Theorem 1) are already present in the proof of this toy case. We include the toy case here because the proof is technically much simpler; understanding it may suffice for all but the most curious readers.

4A. *The content function.* For $q = 3$, following Bellaïche (see the appendix of [Bellaïche and Khare 2015]), we define a function $c : \mathbb{N} \to \mathbb{N}$ depending on q as follows. Given an integer n, we write it in base q as $n = \sum_i n_i q^i$ with $0 \leq n_i < q$, only finitely many of which are nonzero, and define the *q-content of n* as $c(n) := c_q(n) := \sum_i n_i (q-1)^i$. For example, since $71 = [2\,4\,1]_5$ in base 5, the 5-content of 71 is $2 \cdot 4^2 + 4 \cdot 4 + 1 = 49$.

The following properties of the content function are easy to check. See also Section 7A, where the content function and variations are discussed in detail.

Proposition 4.1. (1) $c(n) \ll n^{\log_q(q-1)}$.

(2) *$c(q^k n) = (q-1)^k c(n)$ for all $k \geq 0$.*

(3) *If $0 \leq n < q$, then $c(n) = n$.*

(4) *If i is a digit base q and $n \geq i$ has no more than 2 digits base q, then $c(n) - c(n-i)$ is either i or $i-1$.*

(5) *If $q \leq n < q^2$, then $c(n-q) = c(n) - q + 1$.*

4B. ***Setup of the proof.*** We now define the q-content of a polynomial $f \in \mathbb{F}_p[y]$ through the q-content of its degree. More precisely, if $0 \neq f = \sum a_n y^n$, let $\tilde{c}(f) := \max\{c(n) : a_n \neq 0\}$. Also set $\tilde{c}(0) := -\infty$. For example, the 3-content of $2y^9 + y^7 + y^2$ is $\max\{c(9), c(7), c(2)\} = 5$.

Now let $T : \mathbb{F}_p[y] \to \mathbb{F}_p[y]$ be a degree-lowering recursion operator whose companion polynomial

$$P = X^q + a_1(y)X^{q-1} + \cdots + a_q(y) \in \mathbb{F}_p[y][X]$$

satisfies $\deg a_i(y) < i$ for $1 \leq i < q$ and $\deg a_q(y) \leq q$.

To prove Theorem 2, we will show that T lowers the q-content of any $f \in \mathbb{F}_p[y]$; that is, that $\tilde{c}(Tf) < \tilde{c}(f)$. Since $\tilde{c}(f) < 0$ only if $f = 0$, the fact that T lowers q-content implies that $N_T(f) \leq c(f)$. Proposition 4.1(1) will then imply $N_T(f) \ll (\deg f)^{\log(q-1)/\log q}$, as desired.

It suffices to prove that $\tilde{c}(Tf) < \tilde{c}(f)$ for $f = y^n$. We will proceed by strong induction on n, each time using a deeper recursion of order q^{k+1} corresponding to P^{q^k}, with k chosen so that $q^{k+1} \leq n < q^{k+2}$. We learned this technique from Gerbelli-Gauthier's proof [2016] of the key technical lemmas of Nicolas and Serre [2012a]. Using deeper recurrences with induction allows us to compare n to $n - iq^k$, which has the same last k digits base q, rather than to $n - i$, whose base-q expansion may look very different.

4C. ***The induction.*** The base case is $n < q$, in which case being q-content-lowering is the same thing as being degree-lowering (Proposition 4.1(3)).

For $n \geq q$, we must show that $\tilde{c}(T(y^n)) < c(n)$ assuming that $\tilde{c}(T(y^m)) < c(m)$ for all $m < n$. As above, choose $k \geq 0$ with $q^{k+1} \leq n < q^{k+2}$. By Corollary 2.1, the sequence $\{T(y^n)\}$ satisfies the order-q^{k+1} recurrence

$$T(y^n) = a_1(y)^{q^k} T(y^{n-q^k}) + a_2(y)^{q^k} T(y^{n-2q^k}) + \cdots + a_q(y)^{q^k} T(y^{n-q^{k+1}}).$$

Pick a term y^m appearing in $T(y^n)$ with nonzero coefficient; we want to show that $c(m) < c(n)$. From the recursion, y^m appears with nonzero coefficient in $a_i(y)^{q^k} T(y^{n-iq^k})$ for some i. More precisely, y^m appears in $y^{jq^k} T(y^{n-iq^k})$ for some y^j appearing in $a_i(y)$, so that either $j < i$ or $i = j = q$. Then y^{m-jq^k} appears in $T(y^{n-iq^k})$, and by induction we know that $c(m - jq^k) < c(n - iq^k)$. To conclude that $c(m) < c(n)$, it would suffice to show that

$$c(n) - c(m) \geq c(n - iq^k) - c(m - jq^k),$$

or, equivalently, that

$$c(n) - c(n - iq^k) \geq c(m) - c(m - jq^k).$$

Since subtracting multiples of q^k leaves the last k digits of n base q untouched, we may replace n and m by $q^k \lfloor \frac{n}{q^k} \rfloor$ and $q^k \lfloor \frac{m}{q^k} \rfloor$, respectively, and then use Proposition 4.1(2) to cancel out a factor of $(q-1)^k$. In other words, we must show that

$$c(n) - c(n - i) \geq c(m) - c(m - j)$$

for n, m, i and j satisfying $i \le n < q^2$ and $j \le m < n$ and either $j < i$ or $i = j = q$. But this is an easy consequence of Proposition 4.1(4)–(5): For $j < i$, we know that $c(n) - c(n-i)$ is at least $i - 1$ and $c(m) - c(m-j)$ is at most $j \le i - 1$. And for $i = j = q$ both sides equal $q - 1$.

This completes the proof of Theorem 2.

4D. ***Toy case versus general case.*** The proof of the full NGT (Theorem 1) proceeds by first reducing to the empty-middle case over a finite field (Theorem 4 below), of which Theorem 2 is a special case where the order of the recursion is a power of p. Apart from the reduction step, most of the difficulty in generalizing from Theorem 2 to Theorem 4 comes from working with more general versions of the content function to accommodate any recursion order. Namely, we will extend the content function to rational numbers and prove sufficiently strong analogues of Proposition 4.1 for the induction to proceed, see Sections 7–9.

5. Applications to mod p Hecke algebras

This section gives an indication of how the NGT can give information about lower bounds of mod p Hecke algebras, the author's main motivation for proving the theorem. More precisely, in this section we will use Theorem 2 to complete the proof of Theorem 24 of [Bellaïche and Khare 2015, Appendix], which establishes the structure of the mod 3 Hecke algebra of level one.(ii) Simultaneously and using the same methods, we will give an alternate proof of the main result of Nicolas and Serre [2012b], the structure of the mod 2 Hecke algebra in level one. See Theorem 3 below.

More generally, the NGT can be used to obtain lower bounds on Krull dimensions of local components of big mod p Hecke algebras acting on forms of level N in the case where $X_0(Np)$ has genus zero, for this is precisely the condition for the algebra of modular forms of level N mod p to be a polynomial algebra over $\mathbb{F}_p$. For more details, see [Medvedovsky 2015] (for level one) or [Medvedovsky $\ge$ 2018]. To generalize the nilpotence method to all (p, N), one must generalize the NGT to all rings of S-integers in characteristic-p global fields, with the max-pole-order filtration generalizing degree.

We work in level one with $p \in \{2, 3\}$. Let $M = M(1, \mathbb{F}_p) \subset \mathbb{F}_p[\![q]\!]$ be the space of modular forms of level one modulo p in the sense of Swinnerton-Dyer and Serre (that is, reductions of integral q-expansions). For $p = 2, 3$ Swinnerton-Dyer observes [1973] that $M = \mathbb{F}_p[\Delta]$, where $\Delta = \prod_{i=1}^{n}(1-q^n)^{24} \in \mathbb{F}_p[\![q]\!]$. Standard dimension formulas show that $M_k := \mathbb{F}_p[\Delta]_k$, the polynomials in Δ of degree bounded by k, coincides with the space of mod p reductions of q-expansions of forms of weight $12k$, and hence is Hecke invariant. Further, one can show that $K := \langle \Delta^n : p \nmid n \rangle_{\mathbb{F}_p} \subset M$ is the kernel of the operator U_p, which implies that K and the finite-dimensional subspaces $K_k := M_k \cap K$ are all Hecke invariant.

Let $A_k \subset \mathrm{End}_{\mathbb{F}_p}(K_k)$ be the algebra generated by the action of the Hecke operators T_ℓ with ℓ prime and $\ell \neq p$. Since $K_k \hookrightarrow K_{k+1}$, we have $A_{k+1} \twoheadrightarrow A_k$. Let $A := \varprojlim_k A_k$. Then A is a profinite ring

(ii) More precisely, we prove a weaker version of [Bellaïche and Khare 2015, Proposition 35]. In the notation of Section 7B here, we show for $f \in \mathbb{F}_3[\Delta]$ that $c_{3,2}(T_2 f) \le c_{3,2}(f) - 1$ and $c_{9,6}(T_7' f) \le c_{9,6}(f) - 3$. This suffices to complete the proof of [loc. cit., Theorem 24], but we do not prove the stronger claim that $c_{3,2}(T_7' f) \le c_{3,2}(f) - 2$.

embedding into $\mathrm{End}(K)$; it is the shallow Hecke algebra acting on forms of level one mod p. The standard pairing $A \times K \to \mathbb{F}_p$ given by $\langle T, f\rangle \mapsto a_1(Tf)$ is nondegenerate on both sides and continuous in the profinite topology on A. Therefore A is in continuous duality with K. By work of Tate [1994] and Serre [1986, p. 710, Note 229.2] we know that Δ is the only Hecke eigenform in $K \otimes \overline{\mathbb{F}}_p$.(iii) This implies that A is a local $\mathbb{F}_p$-algebra with maximal ideal $\mathfrak{m}$ and residue field $\mathbb{F}_p$ generated by the modified Hecke operators $T'_\ell := T_\ell - a_\ell(\Delta)$, acting locally nilpotently on $M = \mathbb{F}_p[\Delta]$. Using deformation theory of Chenevier pseudorepresentations, one can deduce:

Proposition 5.1. *There is a surjection* $\mathbb{F}_p[\![x, y]\!] \twoheadrightarrow A$ *given by* $\begin{cases} x \mapsto T_3,\ y \mapsto T_5 & \text{if } p = 2, \\ x \mapsto T_2,\ y \mapsto T'_7 & \text{if } p = 3. \end{cases}$

For $p = 2$, the fact that A is generated by T_3 and T_5 was first proved without deformation theory in [Nicolas and Serre 2012a]. For $p = 3$, Proposition 5.1 is stated [Bellaïche and Khare 2015, Appendix], using deformation theory of reducible Rouquier pseudocharacters developed in [Bellaïche 2012]. See [Medvedovsky 2015, Chapter 7] for detailed computations of tangent spaces to reducible local components of mod p Hecke algebras in level one.

The main result of this section is the following:

Theorem 3. *The surjection* $\mathbb{F}_p[\![x, y]\!] \twoheadrightarrow A$ *of Proposition 5.1 is an isomorphism.*

The key input will be Theorem 2, as well as the following observation: if T is any Hecke operator and f is a modular form in a Hecke invariant algebra M, then the sequence $\{T(f^n)\}_n$ satisfies an M-linear recursion. For more details on the Hecke recursion, see [Medvedovsky 2015, Chapter 6] or [Medvedovsky $\geq$ 2018], but here we will only need some special cases for $f = \Delta$ already given in [Nicolas and Serre 2012a; Bellaïche and Khare 2015, Appendix]. For $p = 2$, we have, as in [Nicolas and Serre 2012a, (13)–(14)],

$$\begin{aligned} T_3(\Delta^n) &= \Delta\, T_3(\Delta^{n-3}) + \Delta^4\, T_3(\Delta^{n-4}), \quad n \geq 3, \\ T_5(\Delta^n) &= \Delta^2\, T_5(\Delta^{n-2}) + \Delta^4\, T_5(\Delta^{n-4}) + \Delta\, T_5(\Delta^{n-5}) + \Delta^6\, T_5(\Delta^{n-6}), \quad n \geq 6, \end{aligned} \tag{5-1}$$

with companion polynomials $P_3 = X^4 + \Delta X + \Delta^4$ and $P_5 = X^6 + \Delta^2 X^4 + \Delta^4 X^2 + \Delta X + \Delta^6$. Note that $\{T_5(\Delta^n)\}_n$ also satisfies the recursion defined by $P_5^* = P_5(X^2 + \Delta^2) = X^8 + \Delta X^3 + \Delta^3 X + \Delta^8$. And for $p = 3$, the recursions satisfied by $\{T_2(\Delta^n)\}$ and $\{T'_7(\Delta^n)\}$ have companion polynomials

$$P_2 = X^3 - \Delta X + \Delta^3 \quad \text{and} \quad P'_7 = X^9 - \Delta X^5 - \Delta^2 X^4 + (\Delta^4 - \Delta)X^2 + (\Delta^5 + \Delta^2)X - \Delta^9. \tag{5-2}$$

See Lemma 33 in [Bellaïche and Khare 2015, Appendix].(iv)

Proof of Theorem 3. Let T and S be the generators of A from Proposition 5.1. Then T, S are filtered and degree-lowering recursion operators on $\mathbb{F}_p[\Delta]$, each satisfying the conditions of Theorem 2. In other words, there exists an $\alpha < 1$ so that $N(\Delta^n) := N_T(\Delta^n) + N_S(\Delta^n) \ll n^\alpha$.

(iii) Alternatively, one can use an observation of Serre to conclude that any Hecke eigenform in $K \otimes \overline{\mathbb{F}}_p$ is in fact defined over $\mathbb{F}_p$, reducing the eigenform search to a finite computation. See [Bellaïche and Khare 2015, Section 1.2 footnote].

(iv) Lemma 33 in [Bellaïche and Khare 2015, Appendix] gives the degree-8 recurrence satisfied by $\{T_7(\Delta^n)\}_n$; the recurrence satisfied by $\{\Delta^n + T_7(\Delta^n)\}_n$ has an extra factor of $X - \Delta$.

We now claim that the Hilbert–Samuel function of A grows *faster than linearly*, so that the Krull dimension of A is at least 2. Indeed, the Hilbert–Samuel function of A sends a positive integer k to

$$\dim_{\mathbb{F}_p} A/\mathfrak{m}^k = \dim_{\mathbb{F}_p} K[\mathfrak{m}^k] \geq \#\{n : \Delta^n \in K,\ \mathfrak{m}^k \Delta^n = 0\} = \#\{n \text{ prime to } p : N(\Delta^n) \geq k\} \gg k^{1/\alpha},$$

which is certainly faster than linear, since $\frac{1}{\alpha} > 1$. Therefore it grows at least quadratically, and the Krull dimension of A is at least 2. By the Hauptidealsatz, the kernel of the surjection $\mathbb{F}_p[\![x, y]\!] \twoheadrightarrow A$ from Proposition 5.1 is trivial. □

Using the more precise bounds on α from Theorem 4, we can conclude that, for $p = 2$ we have $\alpha = \max\{\log_4 2, \log_8 4\} = \frac{2}{3}$ and for $p = 3$ we have $\alpha = \max\{\log_3 2, \log_9 6\} \approx 0.815$. Compare to $\alpha = \frac{1}{2}$ obtained for $p = 2$ by Nicolas and Serre [2012a, §4.1]. Computations suggest that $\alpha = \frac{1}{2}$ also holds for $p = 3$, but we have not been able to prove this.

6. The proof of the NGT begins

We now begin the proof of Theorem 1.

6A. ***Overview of the proof.*** The proof proceeds as follows.

(1) *Reduce the NGT to the case where R is a finite field*: See Section 6B below.

(2) *Reduce to the empty-middle case*: The NGT over a finite field is implied by a special empty-middle case (Theorem 4), where the companion polynomial has no terms of maximal total degree except for X^d and y^d (i.e., the highest-degree homogeneous part has an empty middle). Note that Theorem 4 holds over any ring of characteristic p. See Section 6C below for the statement of Theorem 4 and the reduction step.

(3) *Prove Theorem 4*: The main idea of the proof is as follows. Given an operator T satisfying the conditions of Theorem 4, we define a function $c_T : \mathbb{N} \to \mathbb{N}$ that grows like n^α for some $\alpha < 1$. We extend this function to polynomials in $R[y]$ via the degree. Finally, we use strong induction to prove that applying T strictly lowers the c_T-value of any polynomial in $R[y]$. Therefore, $N_T(y^n)$ is bounded by $c_T(n) \asymp n^\alpha$.

The key features of this kind of proof are already present in the proof of Theorem 2 in Section 4.

6B. ***Reduction to the case where R is a finite field.***

Proposition 6.1. *If Theorem 1 is true whenever R is a finite field*, *then Theorem 1 is true.*

Proof. First, suppose R is a finite artinian local ring with maximal ideal $\mathfrak{m}$ and finite residue field $\mathbb{F}$. Let ℓ be the least positive integer so that $\mathfrak{m}^\ell = 0$.

Let $T : R[y] \to R[y]$ be the operator in the statement of the theorem, and write $\overline{T} : \mathbb{F}[y] \to \mathbb{F}[y]$ for the operator obtained by tensoring with the quotient map $R \twoheadrightarrow \mathbb{F}$. Theorem 1 for $\mathbb{F}$ guarantees that $N_{\overline{T}}(y^n) \ll n^\alpha$ for some $\alpha < 1$. Let

$$g(n) := \max_{n' \leq n}\{N_{\overline{T}}(y^{n'}) + 1\} \ll n^\alpha,$$

so that g is nondecreasing, integer-valued, and satisfies $\overline{T}^{g(\deg f)} f = 0$ for every f in $\mathbb{F}[y]$. Lifting back to R, we get that $T^{g(\deg f)} f$ is in $\mathfrak{m}[y]$ for all $f \in R[y]$.[(V)] More generally, if f is in $\mathfrak{m}^i[y]$, then $T^{g(\deg f)}$ sends f to $\mathfrak{m}^{i+1}[y]$. Since $\mathfrak{m}^\ell = 0$, we have $T^{\ell g(\deg f)} f = 0$ for every $f \in R[y]$, so that $N_T(y^n) \leq \ell g(n) - 1 \ll n^\alpha$.

In the general case, R is a finite product of finite artinian local rings R_i, and an R-linear operator $T : R[y] \to R[y]$ decomposes as $\sum T_i$ where $T_i : R_i[y] \to R_i[y]$ is the R_i-linear restriction of T to R_i. From the paragraph above, we can choose $\alpha_i < 1$ so that $N_{T_i}(y^n) \ll n^{\alpha_i}$ for all n. Then $\alpha = \max_i\{\alpha_i\}$ works for T. □

6C. ***Reduction to the empty-middle NGT.*** From now on we fix a prime p. Theorem 1 over a finite field of characteristic p is implied by the following special case in which the shape of the recursion satisfied by T is restricted. Note that the statement below has no finiteness restrictions on the base ring, and no restriction on the coefficient of y^d.

Theorem 4 (empty-middle NGT). *Let R be a ring of characteristic p, and suppose that T is a degree-lowering linear operator on $R[y]$ so that the sequence $\{T(y^n)\}_n$ satisfies a linear recursion whose companion polynomial has the shape*

$$X^d + ay^d + (\textit{terms of total degree} \leq d - D)$$

for some $D \geq 1$ and some constant $a \in R$. Let $b \geq d$ be a power of p, and suppose that either $b - d \leq 1$ or that $D \leq \frac{b}{2}$. Then

$$N_T(y^n) \ll n^\alpha \quad \textit{for } \alpha = \frac{\log(b-D)}{\log b}.$$

The case $d = b$ and $D = 1$ has already been established in Theorem 2; an analogous argument extends to $d = b$ and any D with $1 \leq D \leq b - 1$.

Proposition 6.2 (empty-middle NGT implies NGT). *Theorem 4 implies Theorem 1.*

Proof. By Proposition 6.1, we may assume that we are working over a finite field $\mathbb{F}$. Let

$$P = X^d + a_1 X^{d-1} + \cdots + a_d \in \mathbb{F}[y][X]$$

be the filtered recursion satisfied by the sequence $\{T(y^n)\}_n$ as in the setup of Theorem 1; recall that we insist that $\deg a_d = d$. We will show that P divides a polynomial of the form

$$X^e - y^e + (\text{terms of total degree} < e)$$

for $e = q^m(q-1)$, where q is a power of p and $m \geq 0$. Then the sequence $\{T(y^n)\}_n$ will also satisfy the recursion associated to a polynomial whose shape fits the requirements of Theorem 4.

Let H be the degree-d homogeneous part of P, so that $P = H + (\text{terms of total degree} < d)$. We claim that there exists a homogeneous polynomial $S \in \mathbb{F}[y, X]$ so that $H \cdot S = X^e - y^e$ for some positive

[(V)] If $\mathfrak{a} \subset R$ is an ideal, write $\mathfrak{a}[y] \subset R[y]$ for the ideal of polynomials all of whose coefficients are in $\mathfrak{a}$.

integer e of required form. Once we find such an S, we know that $P \cdot S$ will have the desired shape $X^e - y^e +$ (terms of total degree $< e$).

To find S, we dehomogenize the problem by setting $y = 1$. Let $h(X) := H(1, X) \in \mathbb{F}[X]$, a monic polynomial of degree d and nonzero constant coefficient. Let $\mathbb{F}'$ be the splitting field of $h(X)$; under our assumptions on a_d, all the roots of $h(X)$ are nonzero. Let q be the cardinality of $\mathbb{F}'$. Every nonzero element $\alpha \in \mathbb{F}'$, and hence every root of $h(X)$, satisfies $\alpha^{q-1} = 1$.

Finally, let q^m be a power of q not less than any multiplicity of any root of $h(X)$. Since every root of h satisfies the polynomial $X^{q-1} - 1$, we know that $h(X)$ divides the polynomial $(X^{q-1} - 1)^{q^m} = X^{q^m(q-1)} - 1$. Set $e = q^m(q-1)$, and let $s(X)$ be the polynomial in $\mathbb{F}[X]$ satisfying $h(X)s(X) = X^e - 1$.

Now we finally "rehomogenize" again; if $S \in \mathbb{F}[y, X]$ is the homogenization of $s(X)$, then $Q \cdot S = X^e - y^e$, so that S is the homogeneous scaling factor for P that we seek. □

6D. ***The main induction for the proof the empty-middle NGT.*** From now on, having already fixed p, we will always assume that R is a ring of characteristic p, not necessarily finite.

Definition. If $T : R[y] \to R[y]$ an R-linear operator, we will call T a (d, D)-*NRO*, for *nilpotent recursion operator*, if T satisfies the conditions of Theorem 4; that is, T lowers degrees, and $\{T(y^n)\}$ satisfies an $R[y]$-linear recursion with companion polynomial

$$X^d + ay^d + (\text{terms of total degree} \le d - D)$$

for some $d \ge 1$ and some $D \ge 1$. Note that any (d, D)-NRO is a (d, D')-NRO for any $1 \le D' \le D$.

The proof of Proposition 6.2 shows that, if R is a finite field, then any T satisfying the conditions of Theorem 1 is in fact a (d, D)-NRO for some d and D.

On the other hand, we make the following definition, for any triple (b, d, D) with $b \ge d \ge D \ge 1$:

Definition. A function $c : \mathbb{Q}_{\ge 0} \to \mathbb{Q}_{\ge 0}$ is a (b, d, D)-*nilgrowth witness* if it satisfies the following properties:

(1) *Discreteness*: $c(\mathbb{N})$ is contained in a lattice of $\mathbb{Q}$ (that is, $\exists M \in \mathbb{N}$ with $Mc(\mathbb{N}) \subset \mathbb{N}$).

(2) *Growth property*: $c(n) \asymp n^{\log(b-D)/\log b}$ as $n \to \infty$.

(3) *Base property*: $0 = c(0) < c(1) < \cdots < c(d-1)$ and $c(d-D) < c(d)$.

(4) *Step property*: For any $k \ge 0$, and any pair $(i, j) \in \{0, 1, \ldots, d\}^2$ with either $(i, j) = (d, d)$ or $i - j \ge D$, and any integers n, m satisfying $db^k \le n < db^{k+1}$ and $jb^k \le m$ we have

$$c(n) - c(n - ib^k) \ge c(m) - c(m - jb^k).$$

In this section, we prove, using strong induction, that if b is a power of p, then the growth of the nilpotence index of a (d, D)-NRO is bounded by the growth of a (b, d, D)-nilgrowth witness.

Proposition 6.3. *Let T be a (d, D)-NRO, and $b \ge d$ a power of p. If c is a (b, d, D)-witness, then $N_T(y^n) \le c(n)$.*

Before proving Proposition 6.3, we record a corollary using the growth property above.

Corollary 6.4. *Suppose that T is a (d, D)-NRO, and let $b = p^{\lceil \log_p d \rceil}$. If there exists a (b, d, D)-witness, then $N_T(y^n) \ll n^{\log(b-D)/\log b}$.*

In other words, in this section we reduce the proof of Theorem 4 to establishing the existence of a $(p^{\lceil \log_p d \rceil}, d, D)$-witness for the (d, D)-NRO T, provided that D is not too big.

Proof of Proposition 6.3. Given a (b, d, D)-witness c, we define a new function $\tilde{c} : R[y] \to \mathbb{N} \cup \{-\infty\}$ via

$$\tilde{c}\Big(\sum a_n y^n\Big) := \max\{c(n) : a_n \neq 0\} \quad \text{and} \quad \tilde{c}(0) := -\infty.$$

We will show that T lowers the $\tilde{c}$-value of polynomials in $R[y]$: that is, that for any nonzero $f \in R[y]$, we have $\tilde{c}(T(f)) < \tilde{c}(f)$.

It suffices to show this for $f = y^n$.

Write x_n for $T(y^n)$. We will use strong induction to show that $\tilde{c}(x_n) < c(n)$.

The base case is all n with $0 \le n < d$. Since $\deg x_n < n$, the statement $\tilde{c}(x_n) < c(n)$ for $n < d$ is implied by the statement that c is strictly increasing on $\{0, 1, \dots, d-1\}$. This is the base property above.

For $n > d$, let $k \ge 0$ be the integer so that $d \cdot b^k \le n < d \cdot b^{k+1}$. Let $P(X) \in \mathbb{F}[y][X]$ be the companion polynomial of the given recursion satisfied by the sequence $\{x_n\}$. Let

$$\mathcal{I} := \{(i, j) : 0 \le j < j + D \le i \le d\} \cup \{(d, d)\}.$$

By assumption, P has the form

$$P = X^d + \sum_{(i,j)\in\mathcal{I}} a_{i,j}\, y^j X^{d-i}$$

for some $a_{i,j} \in R$. By Corollary 2.1, the sequence $\{x_n\}$ also satisfies the order-db^k recursion corresponding to P^{b^k}: namely, for all $n \ge db^k$, we have

$$x_n = -\sum_{(i,j)\in\mathcal{I}} a_{i,j}\, y^{jb^k} x_{n-ib^k}.$$

We will show that, if y^m appears with nonzero coefficient in one of the terms on the right-hand side above, then $c(m) < c(n)$. Since $\tilde{c}(x_n)$ is equal to one of these $c(m)$s, this will imply our claim. So suppose that y^m appears with nonzero coefficient in the (i, j)-term on the right-hand side. That is, y^m appears in $y^{jb^k} x_{n-ib^k}$ for some $(i, j) \in \mathcal{I}$. That means that y^{m-jb^k} appears in x_{n-ib^k}. Note that $i \ge D$, so that $n - ib^k < n$, and the induction assumption applies; since y^{m-jb^k} appears in x_{n-ib^k}, we can assume that $c(m - jb^k) < c(n - ib^k)$.

To show that $c(m) < c(n)$, it therefore suffices to show that

$$c(n) - c(m) \ge c(n - ib^k) - c(m - jb^k),$$

since the latter is assumed to be strictly positive. But this, slightly rearranged, is just the step property from the definition of a (b, d, D)-witness above. $\square$

We now aim to construct a (b, d, D)-witness if $b - d \le 1$ or if $D \le \frac{b}{2}$. This will occupy the next three sections. In Section 7 we investigate the properties of a *content* function, which writes numbers in one base and reads them in another. In Section 8, we establish some inequalities about the content of rational numbers. In Section 9, we use the content function to construct a (b, d, D)-nilgrowth witness, completing the proof of Theorem 4.

7. The content function and its properties

In this section we will introduce a function $c : \mathbb{Q}_{\ge 0} \to \mathbb{Q}_{\ge 0}$ that will serve as a nilgrowth witness in the proof of Theorem 4. This type of function was first introduced in the appendix to [Bellaïche and Khare 2015], loosely inspired by the "code" of [Nicolas and Serre 2012a].

7A. ***Base-b representation of numbers.*** We fix an integer $b \ge 2$ to be the *base*.

Let $\mathcal{D}(b) = \{0, \dots, b-1\}$ be the alphabet of digits base b, and $\mathcal{D}(b)^*$ the set of finite words on $\mathcal{D}(b)$, including the empty word ϵ. The number-of-letters function for a word $x \in \mathcal{D}(b)^*$ will be denoted by $\ell(x)$, for *length*.

Let $\mathcal{R}(b)$ be the set of all bi-infinite pointed words

$$\underline{x} = \dots x_2 x_1 x_0 . x_{-1} x_{-2} \dots$$

on $\mathcal{D}(b)$ that start with ${}^{\infty}0$ (the digit 0 repeated infinitely to the left). The set of finite words $\mathcal{D}(b)^*$ naturally embeds into $\mathcal{R}(b)$ via $x \mapsto ({}^{\infty}0)x.(0^{\infty})$, where 0^{∞} is the digit 0 repeated infinitely to the right. More generally, any pointed right-infinite word will be viewed as an element of $\mathcal{R}(b)$ by appending ${}^{\infty}0$ on the left. For $\underline{x} \in \mathcal{R}(b)$, we can define the real number $\pi_b(\underline{x}) \in \mathbb{R}_{\ge 0}$ by reading it as a sequence of digits base b, via $\pi_b(\underline{x}) := \sum_i x_i b^i$. Since $x_i = 0$ for $i \gg 0$, this sum converges. The map π_b is not injective; indeed, $\sum_{i<k} (b-1) b^i = b^k$, so that for any finite word w and digit $x \ne b-1$, and any radix point placement, we have $\pi_b(wx(b-1)^{\infty}) = \pi_b(w(x+1)0^{\infty})$. But we can choose a section of π_b by restricting the domain: Let $\mathcal{R}'(b) \subset \mathcal{R}(b)$ the subset of those that do not end with $(b-1)^{\infty}$. Then the reading-base-b function $\pi_b : \mathcal{R}'(b) \to \mathbb{R}_{\ge 0}$ is a bijection, and the inverse map $\rho_b : \mathbb{R}_{\ge 0} \to \mathcal{R}'(b)$ takes a nonnegative real number q to its *normal* (that is, not ending in $(b-1)^{\infty}$) base-b representation $\underline{x} = \rho_b(q)$ satisfying $\pi_b(\underline{x}) = q$.

The base-b representation $\rho_b(q)$ is *eventually periodic* (that is, ends with z^{∞} for some finite word z) if and only if $q \in \mathbb{Q}_{\ge 0}$. For $q \in \mathbb{Q}_{\ge 0}$, then, we know that

$$\rho_b(q) = x.yz^{\infty},$$

where x, y, z are in $\mathcal{D}(b)$. If we insist that x does not start with 0 and that first y and then z have minimal length among such representations, then x, y, and z are defined uniquely. We will assume this minimality from now on. Note that by construction x and y may be empty words, but z has length at least 1.

We define, then, three constants associated to $q \in \mathbb{Q}_{\geq 0}$:

$$\begin{aligned}
\ell(q) = \ell_b(q) &:= \ell(x) = \max\{0, \lfloor \log_b q \rfloor + 1\}, \\
s(q) = s_b(q) &:= \ell(y) = \min\{k \geq 0 : \text{denominator of } b^k q \text{ is prime to } b\}, \\
t(q) = t_b(q) &:= \ell(z) = \min\{k \geq 1 : \text{denominator of } b^{s(q)} q \text{ divides } b^k - 1\}.
\end{aligned}$$

In particular, we know that, for $q \in \mathbb{Q}_{\geq 0}$, we have

$$q = n + \frac{u}{b^{s(q)}} + \frac{m}{b^{s(q)}(b^{t(q)} - 1)}, \tag{7-1}$$

where n, u, and m are all integers with $n = \lfloor q \rfloor = \pi_b(x)$, $u = \pi_b(y)$, and $m = \pi_b(z)$.

We will need the following very simple lemma.

Lemma 7.1. *Given $q \in \mathbb{Q}_{\geq 0}$ and a base b, if $q' \in q\mathbb{N}$, then*

(1) $s_b(q') \leq s_b(q)$,

(2) $t_b(q')$ *divides* $t_b(q)$.

The proof follows from the fact that the denominator of q' is a divisor of the denominator of q. Alternatively, one can consider the effect of multiplication by integers on base-b expansions.

7B. *The content function.* Now let $b, \beta \geq 2$ be bases. Define the (b, β)-*content* of any q in $\mathbb{R}_{\geq 0}$ to be the result of reading the normal base-b representation of q in base β,

$$c_{b,\beta}(q) := \pi_\beta(\rho_b(q)).$$

Note that π_β makes sense as a function $\mathcal{R}(b) \to \mathbb{R}_{\geq 0}$; the series $\sum_{i \leq k} x_i b^i$ always converges if the x_i are bounded.

Examples. (1) Since $\rho_5(196) = 1241$, we have $c_{5,3}(196) = 1 \cdot 3^3 + 2 \cdot 3^2 + 4 \cdot 3 + 1 = 58$.

(2) We have $\rho_7\big(\frac{1}{3}\big) = 0.(2)^\infty$. Therefore $c_{7,5}(n) = 2 \sum_{i \geq 1} 5^{-i} = \frac{1}{2}$.

(3) $c_{8,3}\big(\frac{1}{6}\big) = \pi_3(0.1(25)^\infty) = \frac{1}{3} + \big(\frac{2}{3^2} + \frac{5}{3^3}\big) \sum_{i \geq 0} 3^{-2i} = \frac{1}{3} + \frac{11}{27} \cdot \frac{9}{8} = \frac{19}{24}$.

The following lemma, which will be used frequently, is an easy computation.

Lemma 7.2. *For $q \in \mathbb{Q}_{\geq 0}$, let $s = s_b(q)$ and $t = t_b(q)$. Then if $q = n + \dfrac{u}{b^s} + \dfrac{m}{b^s(b^t - 1)}$ as in* (7-1) *above, we have*

$$c_{b,\beta}(q) = c_{b,\beta}(n) + \frac{c_{b,\beta}(u)}{\beta^s} + \frac{c_{b,\beta}(m)}{\beta^s(\beta^t - 1)}.$$

We will also use the following growth estimate:

Lemma 7.3. *We have $c_{b,\beta}(n) \asymp n^{\log_b \beta}$. More precisely, for $n \geq 1$, we have*

$$\beta^{-1} n^{\log_b \beta} < c_{b,\beta}(n) < \frac{\beta(b-1)}{\beta - 1} n^{\log_b \beta}.$$

Proof. Let $\ell = \ell_b(n)$, so that for $n \geq 1$ we have $\ell = 1 + \lfloor \log_b n \rfloor$, or $\log_b n < \ell \leq 1 + \log_b n$. Then, on one hand, the (b, β)-content of n is bounded above by π_β of the infinite pointed word $(b-1)^\ell.(b-1)^\infty$:

$$c_{b,\beta}(n) < \sum_{k<\ell} (b-1)\beta^k = \frac{b-1}{\beta-1}\beta^\ell \leq \frac{\beta(b-1)}{\beta-1}\beta^{\log_b n} = \frac{\beta(b-1)}{\beta-1} n^{\log_b \beta}.$$

On the other hand, the (b, β)-content of n is at least π_β of the pointed word $1(0^{\ell-1}).0^\infty$:

$$c_{b,\beta}(n) \geq \beta^{\ell-1} > \beta^{-1} n^{\log_b \beta}. \qquad \square$$

In particular, if $\beta < b$, then $c_{b,\beta}$ grows slower than linearly in n.

Remarks. (1) If $\beta < b$, then $c_{b,\beta}$ is never monotonically increasing, even on the integers. Indeed,

$$c_{b,\beta}(b^k - 1) = \frac{b-1}{\beta-1}(\beta^k - 1),$$

which is greater than $c_{b,\beta}(b^k) = \beta^k$ as soon as $\beta^k > \frac{\beta-1}{b-2}$.

(2) The sequence $\{c_{b,\beta}(n)\}_{n\in\mathbb{N}}$ is *b-regular* in the sense of Allouche and Shallit [1992]. A sequence $\{s_n\} \in \mathbb{Q}^\mathbb{N}$ is said to be b-regular if there exists a $\mathbb{Q}$-vector space V, endomorphisms $M_0, \ldots, M_{b-1}$ of V, a vector $v \in V$, and a functional $\lambda : V \to \mathbb{Q}$ so that $s_{\pi_b(n_\ell \ldots n_0)} = \lambda(M_{n_\ell} \cdots M_{n_0} v)$. For the sequence $\{c_{b,\beta}(n)\}$, we can take $V = \mathbb{Q}^2$, $M_i = \left[\begin{smallmatrix} 1 & i \\ 0 & \beta \end{smallmatrix}\right]$, $v = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ and $\lambda = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. Indeed, b-regularity appears to be lurking in many places in this theory, but we have not yet been able to make use of it.[vi]

Finally, we record a trivial scaling property of $c_{b,\beta}$:

Lemma 7.4. *For $n \in \mathbb{R}_{\geq 0}$, and any $k \in \mathbb{Z}$, we have $c_{b,\beta}(b^k n) = \beta^k c_{b,\beta}(n)$.*

7C. ***Content and the carry-digit word.*** Even though the content function is not monotonic, it behaves well under addition. For $m, n \in \mathbb{R}_{\geq 0}$, we define $r_b(m, n) \in \mathcal{R}(2)$ to be the word of carry digits when the sum $s = m + n$ is computed in base b. More precisely, let $\rho_b(m) = \underline{m}$, $\rho_b(n) = \underline{n}$, and $\rho_b(s) = \underline{s}$, and let $r_b(m, n) := \underline{r}$ satisfying

$$m_i + n_i + r_{i-1} = s_i + b r_i \quad \text{for all } i \text{ in } \mathbb{Z}. \tag{7-2}$$

Since m_i, n_i, and s_i are all 0 for $i > \ell_b(s)$, and since $r_i \in \{0, 1\}$, the set of equations above defines r_i uniquely, inductively down from $i = \ell_b(s)$.

Examples. (1) We compute $r_3(77, 11)$ starting with $\rho_3(77) = 2212$ and $\rho_3(11) = 102$. When the addends have finite base-b expansions, the carry digits appear as a byproduct of the base-b addition algorithm. Below, we read off that $r_3(77, 5) = 1101$:

$$\begin{array}{r} {\scriptstyle 1101} \\ 2212 \\ +102 \\ \hline 10021. \end{array}$$

[vi] For example, the Nicolas–Serre code sequence $h(n)$ defined in [Nicolas and Serre 2012a, §4.1] is 2-regular, as are its constituent parts $n_3(n)$ and $n_5(n)$.

Note the shift one space to the right in our indexing the conventional carry digit notation.

(2) We compute $r_5\left(\frac{53}{60}, \frac{23}{100}\right)$. We have $\rho_5\left(\frac{53}{60}\right) = 0.4(20)^\infty$ and $\rho_5\left(\frac{23}{100}\right) = 0.10(3)^\infty$. Their sum is $\frac{167}{150} = \pi_5(1.02(40)^\infty)$. Comparing the base-5 expansions of the two addends with the expansion of the sum allows us to compute the carry digits left to right.

$$\begin{array}{r} {\scriptstyle 1\ 00101010\ldots} \\ 0.42020202\ldots \\ +0.10333333\ldots \\ \hline 1.02404040\ldots \end{array}$$

Therefore $r_5\left(\frac{53}{60}, \frac{23}{100}\right) = 0.10(01)^\infty$. In this case, we will get the same infinite carry-digit word if we take the "limit" of the finite carry-digit words obtained by truncating the expansions of the two addends.

(3) Note that $r_{10}\left(\frac{1}{3}, \frac{2}{3}\right) = 0.1^\infty$, even though any finite truncation of the decimal expansions of $\frac{1}{3}$ and $\frac{2}{3}$ would yield no carry digits in the sum. If the addends are not in $\mathbb{Z}\left[\frac{1}{b}\right]$ but the sum is, one computes the expansion of the sum before computing the carry-digit word.

The carry digit word exactly keeps track of the difference between values of $c_{b,\beta}$:

Lemma 7.5. *For m, n in $\mathbb{R}_{\geq 0}$, we have*

$$c_{b,\beta}(m) + c_{b,\beta}(n) = c_{b,\beta}(m+n) + (b-\beta)\pi_\beta(r_b(m,n)).$$

Proof. Let $s = m + n$, and let $\underline{m}$, $\underline{n}$, $\underline{s}$ be the corresponding base-b expansions and $\underline{r}$ the carry-digit word. Scaling (7-2) by β^i and summing up over all i gives us

$$\sum m_i \beta^i + \sum n_i \beta^i + \sum r_{i-1} \beta^i = \sum s_i \beta^i + b \sum r_i \beta^i$$

or, equivalently,

$$c_{b,\beta}(m) + c_{b,\beta}(n) + \beta\pi_\beta(\underline{r}) = c_{b,\beta}(m+n) + b\pi_\beta(\underline{r}). \qquad \square$$

We will typically use Lemma 7.5 when comparing $c_{b,\beta}(m)$ and $c_{b,\beta}(n)$ by analyzing $c_{b,\beta}(m-n)$ and $r_b(m-n, n)$.

Examples. (1) We have $c_{3,2}(77) = \pi_2(2212) = 28$, and $c_{3,2}(88) = \pi_2(10021) = 21$. The difference is accounted for by

$$c_{3,2}(88-77) = \pi_2(102) = 6 \quad \text{and} \quad \pi_2(r_3(77, 11)) = \pi_2(1101) = 13.$$

Since $28 + 6 = 21 + 13$, we are consistent with Lemma 7.5.

(2) Let $(b,\beta)=(5,3)$, and consider the $\frac{53}{60}+\frac{23}{100}$ example from above. Using Lemma 7.2, we find that

$$\begin{aligned}
c_{5,3}\left(\tfrac{53}{60}\right) &= \pi_3(0.4(20)^\infty) = \frac{\pi_3(4)}{3}+\frac{\pi_3(20)}{3(3^2-1)} = \frac{19}{12}\\
c_{5,3}\left(\tfrac{23}{100}\right) &= \pi_3(0.10(3)^\infty) = \frac{\pi_3(10)}{3^2}+\frac{\pi_3(3)}{3^2(3-1)} = \frac{1}{2}\\
c_{5,3}\left(\tfrac{167}{150}\right) &= \pi_3(1.02(40)^\infty) = \pi_3(1)+\frac{\pi_3(02)}{3^2}+\frac{\pi_3(40)}{3^2(3^2-1)} = \frac{25}{18}\\
\pi_3\left(r_5\left(\tfrac{53}{60},\tfrac{23}{100}\right)\right) &= \pi_3(0.10(01)^\infty) = \frac{\pi_3(10)}{3^2}+\frac{\pi_3(01)}{3^2(3^2-1)} = \frac{25}{72}.
\end{aligned}$$

As expected from Lemma 7.5, we have $\frac{19}{12}+\frac{1}{2}=\frac{25}{12}=\frac{25}{18}+2\cdot\frac{25}{72}$.

(3) Finally, let $b=10$ and return to the addition equation $\frac{1}{3}+\frac{2}{3}=1$. For $a\in\mathcal{D}(9)$, we have $\pi_\beta(0.a^\infty)=\frac{a}{\beta-1}$. Therefore the two sides of the Lemma 7.5 equation agree:

$$\begin{aligned}
&\text{(LHS)} & c_{10,\beta}\left(\tfrac{1}{3}\right)+c_{10,\beta}\left(\tfrac{2}{3}\right) &= \pi_\beta(0.3^\infty)+\pi_\beta(0.6^\infty) = \tfrac{9}{\beta-1}\\
&\text{(RHS)} & c_{10,\beta}(1)+(10-\beta)\pi_\beta\left(r_{10}\left(\tfrac{1}{3},\tfrac{2}{3}\right)\right) &= 1+(10-\beta)\pi_\beta(0.1^\infty) = \tfrac{9}{\beta-1}.
\end{aligned}$$

8. Content of some proper fractions

In this section, we will prove inequalities about $(b,b-D)$-content of some proper fractions that we will use in Section 9 to produce a (b,d,D)-nilgrowth witness.

8A. ***Unit fractions in base b.*** Let $b\geq 2$ be a base, and fix a denominator d with $1<d\leq b$. To motivate the discussion, we note that

$$\frac{1}{d}=\frac{\frac{1}{b}}{1-\frac{b-d}{b}}=\sum_{k\geq 1}(b-d)^{k-1}b^{-k}.$$

Therefore, the base-b expansion of $\frac{1}{d}$ "wants" to be $0.1(b-d)(b-d)^2(b-d)^3\ldots$. Of course, unless $b-d\leq 1$, this is not possible; $(b-d)^k$ is not a digit base b for k large enough. However, letting $\rho_b\left(\frac{1}{d}\right)=0.a_1a_2a_3\ldots$, we can say the following:

Lemma 8.1. *For $k\geq 1$, we have $a_i=(b-d)^{i-1}$ for $i=1,\ldots,k$ if and only if $(b-d)^k<d$.*

Proof. We will establish this claim by induction on k. The a_k can be defined recursively via

$$a_k=\left\lfloor\frac{b^k}{d}-\sum_{i=1}^{k-1}a_ib^{k-i}\right\rfloor=\left\lfloor\frac{b^k}{d}\right\rfloor-\sum_{i=1}^{k-1}a_ib^{k-i}. \tag{8-1}$$

For $k=1$, we have $a_1=\left\lfloor\frac{b}{d}\right\rfloor\geq 1$. Therefore $a_1=1$ if and only if $\frac{b}{d}<2$, which is equivalent to $(b-d)^1<d$. So the claim for $k=1$ is true.

Now suppose we already know that $a_i = (b-d)^{i-1}$ for $i < k$. Then $a_k = (b-d)^{k-1}$ if and only if

$$\begin{aligned}1 &> \left(\frac{b^k}{d} - \sum_{i=1}^{k-1} a_i b^{k-i}\right) - (b-d)^{k-1} \\ &= \frac{b^k}{d} - (b^{k-1} + (b-d)b^{k-2} + \cdots + (b-d)^{k-2}b + (b-d)^{k-1}) \\ &= \frac{b^k}{d} - \frac{b^k - (b-d)^k}{b-(b-d)} = \frac{(b-d)^k}{d},\end{aligned}$$

as desired. □

The same argument also implies the immediate:

Corollary 8.2. *If $a_i = (b-d)^{i-1}$ for $i < k$, then $a_k = (b-d)^{k-1} + \lfloor \frac{1}{d}(b-d)^k \rfloor$.*

We can now delineate what $\frac{1}{d}$ must look like in base b.

Lemma 8.3. *For $d \le b$, the base-b expansion of $\frac{1}{d}$ falls into one of five mutually exclusive cases.*

(1) $\rho_b\left(\frac{1}{d}\right) = 0.2^+\ldots$ *(in other words, $a_1 \ge 2$) if and only if $d \le \frac{b}{2}$.*

(2) $\rho_b\left(\frac{1}{d}\right) = 0.13^+\ldots$ *(i.e., $a_1 = 1$ and $a_2 \ge 3$) if and only if $\frac{b}{2} < d \le \frac{b^2}{b+3} = b - 3 + \frac{9}{b+3}$.*

(3) $\rho_b\left(\frac{1}{d}\right) = 0.124^+\ldots$ *(i.e., $a_1 = 1$, $a_2 = 2$, and $a_3 \ge 4$) if and only if $b > 6$ and $d = b-2$.*

(4) $\rho_b\left(\frac{1}{d}\right) = 0.1^\infty$ *if and only if $d = b-1$.*

(5) $\rho_b\left(\frac{1}{d}\right) = 0.10^\infty$ *if and only if $d = b$.*

Proof. If $b-d = 0$ or $b-d = 1$, then $a_i = (b-d)^{i-1}$ for all i (Lemma 8.1). Assume $b-d \ge 2$. If $d \le \frac{b}{2}$, then $\frac{1}{d} \ge \frac{2}{b}$, so that $a_1 \ge 2$, as claimed. Otherwise, we must have $(b-d)^1 < d$, so that $a_1 = 1$ and $a_2 \ge b-d$. This means that $a_2 \ge 3$ unless *both* $b-d = 2$ and $(b-d)^2 < d$, in which case we have $d > 4$ (and hence $b > 6$), and $a_2 \ge (b-d)^2 = 4$. □

8B. *The carry-digit word for a proper fraction in base b.* Keeping the notation b and d, we additionally fix a D with $1 \le D < d$ and investigate $\rho_b\left(\frac{D}{d}\right) =: 0.e_1e_2e_3\ldots$. The e_k satisfy the same type of recursion as the a_k, namely,

$$e_k = \left\lfloor \frac{b^k D}{d} \right\rfloor - \sum_{i=1}^{k-1} e_i b^{k-i}.$$

In particular, $e_1 = \lfloor \frac{Db}{d} \rfloor$, and the following generalization of Lemma 8.1 and Corollary 8.2 holds:

Lemma 8.4. *For $k \ge 0$, we have $e_i = D(b-d)^{i-1}$ for $i = 1, \ldots, k$ if and only if $D(b-d)^k < d$. Moreover, if $D(b-d)^k < d$, then $e_{k+1} = D(b-d)^k + \lfloor \frac{1}{d} D(b-d)^{k+1} \rfloor$.*

In order to understand the relationship between the a_k and the e_k we define the carry-digit word $\underline{r} = 0.r_1r_2r_3\ldots$ for the addition problem $\sum_{i=1}^{D} \frac{1}{d} = \frac{D}{d}$. (See Section 7C for definitions.) In other words, set $\underline{r}^{(i)} := 0.r_1^{(i)} r_2^{(i)} r_3^{(i)} \cdots := r_b\left(\frac{i}{d}, \frac{1}{d}\right)$ for $1 \le i \le D-1$, and then set $r_j := \sum_{i=1}^{d-1} r_j^{(i)}$.

Then $0 \le r_j \le D-1$ for every i; since $\frac{D}{d} < 1$, we have $r_1 = 0$. Putting together equations (7-2) applied to $\frac{i}{d} + \frac{1}{d}$ for $1 \le i < D$ gives us the precise relationship between the a_k, e_k, and r_k:

$$e_k = Da_k + r_{k+1} - br_k. \tag{8-2}$$

In fact, we have a closed formula for r_k:

Lemma 8.5. *For $k \ge 1$ we have*

$$r_k = \left\lfloor \frac{Db^{k-1}}{d} \right\rfloor - D\left\lfloor \frac{b^{k-1}}{d} \right\rfloor = \left\lfloor \frac{D(b-d)^{k-1}}{d} \right\rfloor - D\left\lfloor \frac{(b-d)^{k-1}}{d} \right\rfloor.$$

Proof. The formula is true for $k = 1$ (in our case, all quantities are 0). For $k \ge 1$, we will establish the formula for $k+1$, starting with (8-2):

$$\begin{aligned} r_{k+1} = e_k - Da_k + br_k &= \left\lfloor \frac{Db^k}{d} \right\rfloor - \sum_{i=1}^{k-1} e_i b^{k-i} - D\left\lfloor \frac{b^k}{d} \right\rfloor + D\sum_{i=1}^{k-1} a_i b^{k-i} + br_k \\ &= \left\lfloor \frac{Db^k}{d} \right\rfloor - D\left\lfloor \frac{b^k}{d} \right\rfloor + \sum_{i=1}^{k-1} b^{k-i}(br_i - r_{i+1}) + br_k = \left\lfloor \frac{Db^k}{d} \right\rfloor - D\left\lfloor \frac{b^k}{d} \right\rfloor. \end{aligned}$$

Here we used (8-2) in the form $-e_i + Da_i = br_i - r_{i+1}$ for $1 \le i < k$ to pass from the first line to the second, and cancellation of a telescoping sum for the final equality. Finally, we note that

$$\left\lfloor \frac{Db^k}{d} \right\rfloor - D\left\lfloor \frac{b^k}{d} \right\rfloor = \left\lfloor \frac{D(b-d)^k}{d} \right\rfloor - D\left\lfloor \frac{(b-d)^k}{d} \right\rfloor$$

because the intervening terms $\frac{1}{d}\bigl(D\sum_{i=1}^{k}\binom{k}{i}b^{k-i}(-d)^i\bigr)$ are integers, and hence can pass through the greatest-integer function to cancel. □

Corollary 8.6. (1) *If $(b-d)^k < d$ for some $k \ge 0$, then for every $i \le k+1$, we have $r_i = \left\lfloor \frac{1}{d} D(b-d)^{i-1} \right\rfloor$.*
(2) *If $D(b-d)^k < d$ for some $k \ge 0$, then for every $i \le k+1$, we have $r_i = 0$.*

8C. *(b, β)-Content of proper fractions.* We fix a triple (b, d, D) with $1 \le D \le d \le b$ subject to the conditions that $b \ge 2$, and further impose the condition that $\beta := b - D \ge 2$. Recall the content function $c_{b,\beta}$ from Section 7B, and let $c^d_{b,\beta} : \mathbb{Q}_{\ge 0} \to \mathbb{Q}_{\ge 0}$ be the function defined by

$$c^d_{b,\beta}(n) := c_{b,\beta}\bigl(\tfrac{n}{d}\bigr).$$

Whenever the triple (b, d, D) is understood, we write $c = c^d_{b,\beta}$, and let $\underline{r} = 0.r_1r_2\ldots$ be the carry-digit word for $\frac{1}{d} + \cdots + \frac{1}{d} = \frac{D}{d}$ as in Section 8B. Lemma 7.5 implies that

$$c(D) = Dc(1) - D\pi_\beta(\underline{r}). \tag{8-3}$$

In this section, we will establish some lower bounds on $c(1)$ and $c(D)$. First, we dispatch the cases $d = b$ and $d = b-1$, which yield easy explicit formulas.

Lemma 8.7. *Suppose that $d = b$ or $d = b-1$, and $0 \le i < d$. Then $c(i) = \frac{i}{d-D}$.*

Proof. Computation, see Lemma 8.3. Note that $D \le \min\{d, b-2\}$ precludes the possibility that $d = D$. $\square$

Proposition 8.8. *If $d \le b-2$ and $b > 6$, then $c(1) \ge \frac{\beta+1}{\beta(\beta-1)}$.*

Remark. It is a simple exercise to check that the only exceptions for $b \le 6$ are in fact $(b, d, D) = (4, 2, 2)$, $(5, 3, 3)$, or $(6, 4, 4)$ by exhausting all cases.

Proof. We go through the first three cases of Lemma 8.3. Note that $\beta = 2$ implies that all of the inequalities in $2 \le b-d \le b-D = \beta$ are equalities, so that $d = D$ and $d = b-2$. In particular, if $\beta = 2$ and $b > 6$, then we must be in the third case.

(1) If $\rho_b\big(\frac{1}{d}\big)$ starts with 0.2^+, then $c(1) \ge \pi_\beta(0.2) = \frac{2}{\beta}$. Since $\beta \ge 3$, we have $2 \ge \frac{\beta+1}{\beta-1}$, so that $c(1) \ge \frac{\beta+1}{\beta(\beta-1)}$, as desired.

(2) If $\rho_b\big(\frac{1}{d}\big)$ starts with 0.13^+, then we have $c(1) \ge \rho_\beta(0.13) = \frac{\beta+3}{\beta^2}$. This last is no less than $\frac{\beta+1}{\beta(\beta-1)}$ if and only if $\beta^2 + 2\beta - 3 = (\beta+3)(\beta-1) \ge (\beta+1)\beta = \beta^2 + \beta$. Therefore $\beta \ge 3$ again implies $c(1) \ge \frac{\beta+1}{\beta(\beta-1)}$, as desired.

(3) If $\rho_b\big(\frac{1}{d}\big)$ starts with 0.124^+, then

$$c(1) \ge \pi_\beta(0.124) = \frac{\beta^2 + 2\beta + 4}{\beta^3} = \frac{\beta+1}{\beta(\beta-1)} + \frac{2\beta - 4}{\beta^4 - \beta^3}.$$

Since $\beta \ge 2$, our claim is established. $\square$

Corollary 8.9. *If $d \le b-2$ and $D \le \frac{b}{2}$, then $c(1) \ge \frac{D}{\beta(\beta-1)}$.*

Proof. For $D \le \frac{b}{2}$, we have $\beta = b - D \ge b - \frac{b}{2} = \frac{b}{2} \ge D$. Therefore Proposition 8.8 establishes the desired inequality, the exceptional cases $(b, d, D) = (4, 2, 2)$, $(5, 3, 3)$, or $(6, 4, 4)$ being easy to check explicitly. $\square$

In the next proposition we will show that $c(D)$ is not too small, provided that d is not too big relative to b, or, failing that, that D is not too big relative to b and d.

Proposition 8.10. *Suppose $D < d \le b-2$ and* at least one *of the following conditions is satisfied*:

(1) $d \le \frac{b}{2}$.

(2) $D < d\big(1 - \frac{1}{b-d}\big)$.

Then $c(D) \ge \frac{D(\beta+1)}{\beta(\beta-1)}$.

Remark. Computationally, it appears that the optimal statement is as follows. If $D < d \le b-2$, then $c(D) \ge \frac{D(\beta+1)}{\beta(\beta-1)}$ if and only if at least one of the following is true: $(1')$ $(b-d)^2 > b-1$ or (2) $D < d\big(1 - \frac{1}{b-d}\big)$. Note that Condition (1) above implies condition $(1')$, but this latter is strictly weaker. Here we only prove Proposition 8.10 as stated.

Before proving Proposition 8.10, some preparatory lemmas.

Lemma 8.11. *Under the assumption $d > \frac{b}{2}$, condition (2) from Proposition 8.10 is equivalent to the inequality $r_2 < b-d-1$.*

Proof. Apply Corollary 8.6 for $k = 0$ to deduce that $r_2 = \lfloor \frac{1}{d} D(b-d) \rfloor$. Then $r_2 < b - d - 1$ if and only if $\frac{D(b-d)}{d} < b - d - 1$ if and only if $D < \frac{d(b-d-1)}{b-d}$, which is condition (2), as desired. □

As before, let $\rho_b\left(\frac{1}{d}\right) = 0.a_1 a_2 \ldots$ and $\rho_b\left(\frac{D}{d}\right) = 0.e_1 e_2 \ldots$, and let $\underline{r} = 0.r_1 r_2 \ldots$ be the carry digits as in Section 8B. Using (8-2) and the partial-sum cutoffs $c(D) \geq \pi_\beta(0.e_1 \ldots e_k) = \sum_{i=1}^{k} e_i \beta^{-i}$, we get the following partial-sum versions of (8-3):

Lemma 8.12. *For any $k \geq 1$, the quantity $c(D)$ satisfies the following inequality*:

$$c(D) \geq \frac{D \sum_{i=1}^{k} (a_i - r_i) \beta^{k-i} + r_{k+1}}{\beta^k}.$$

Corollary 8.13. *Any of the following conditions are sufficient to guarantee $c(D) \geq \frac{D(\beta+1)}{\beta(\beta-1)}$*:

(1) $\beta \geq 3$ *and* $a_1 \geq 2$;

(2) $r_2 \geq \frac{2D}{\beta-1}$;

(3) $\beta \geq 3$ *and* $a_2 - r_2 \geq 3$;

(4) $a_2 - r_2 = 2$ *and* $r_3 \geq \frac{2D}{\beta-1}$;

(5) $\beta \geq 3$ *and* $a_2 - r_2 = 2$ *and* $a_3 - r_3 \geq 3$.

Proof. We use Lemma 8.12 for each specified k. Recall that $r_1 = 0$.

(1) $k = 1$, use estimate $r_2 \geq 0$. We have $c(D) \geq \frac{2D}{\beta} \geq \frac{D(\beta+1)}{\beta(\beta-1)}$, since $2 \geq \frac{\beta+1}{\beta-1}$ for $\beta \geq 3$.

(2) $k = 1$, use estimate $a_1 \geq 1$:

$$c(D) \geq \frac{D + r_2}{\beta} \geq \frac{D + \frac{2D}{\beta-1}}{\beta} = \frac{D(\beta+1)}{\beta(\beta-1)}.$$

(3) $k = 2$, use estimate $a_1 \geq 1$ and $r_3 \geq 0$:

$$c(D) \geq \frac{\beta(D + r_2) + (Da_2 - br_2)}{\beta^2} = \frac{D(\beta + a_2 - r_2)}{\beta^2}.$$

This last being greater than $\frac{D(\beta+1)}{\beta(\beta-1)}$ is equivalent to $(\beta + a_2 - r_2)(\beta - 1) \geq \beta(\beta+1)$, or $a_2 - r_2 \geq \frac{2\beta}{\beta-1}$. For $\beta \geq 3$, this is guaranteed by $a_2 - r_2 \geq 3$.

(4) $k = 2$, use estimate $a_1 \geq 1$:

$$\frac{c(D)}{D} \geq \frac{\beta + (a_2 - r_2) + \frac{r_3}{D}}{\beta^2} \geq \frac{\beta + 2 + \frac{2}{\beta-1}}{\beta^2} = \frac{\beta+1}{\beta(\beta-1)}.$$

(5) $k = 3$, use estimate $a_1 \geq 1$ and $r_4 \geq 0$:

$$\frac{c(D)}{D} \geq \frac{\beta^2 + (a_2 - r_2)\beta + (a_3 - r_3)}{\beta^3} \geq \frac{\beta^2 + 2\beta + 3}{\beta^3} = \frac{\beta+1}{\beta(\beta-1)} + \frac{\beta-3}{\beta^3(\beta-1)}.$$ □

Proof of Proposition 8.10. Note that the assumptions $D < d \leq b - 2$ guarantee that $\beta \geq 3$.

If condition (1) holds, then $a_1 \geq 2$ (Lemma 8.3(1)) so that Corollary 8.13(1) gives what we want. If condition (1) fails, but condition (2) holds, then by Lemma 8.11, we have $r_2 \leq b - d - 2$. Moreover, from

Lemma 8.4 for $D = 1$ and $k = 1$, we know that $a_2 \geq b - d$. If either inequality is strict, Corollary 8.13(3) gives us the desired inequality. Therefore it remains to consider the case $r_2 = b - d - 2$ $\bigl($so that $\frac{d(b-d-2)}{b-d} \leq D < \frac{d(b-d-1)}{b-d}\bigr)$ and $a_2 = b - d$ (so that $(b-d)^2 < d$[vii]).

We now estimate the third digits. By Corollary 8.6 and Lemma 8.4, we have $r_3 = \bigl\lfloor \frac{1}{d} D(b-d)^2 \bigr\rfloor$ and $a_3 \geq (b-d)^2$. Condition (2) implies that $r_3 \leq \frac{1}{d} D(b-d)^2 < (b-d)(b-d-1)$, so that

$$a_3 - r_3 > (b-d)^2 - (b-d)(b-d-1) = b - d \geq 3,$$

so that the desired inequality holds by Corollary 8.13(5). □

9. The nilgrowth witness: finishing the proof

Let $b \geq d \geq D \geq 1$ be integers subject to the conditions $b \geq 2$ and $\beta := b - D \geq 2$, as before. In this section we exhibit a (b, d, D)-nilgrowth witness and complete the proof of Theorem 4.

Recall from Section 8C that $c^d_{b,\beta} : \mathbb{Q}_{\geq 0} \to \mathbb{Q}_{\geq 0}$ is the function defined by

$$c^d_{b,\beta}(n) := c_{b,\beta}\bigl(\tfrac{n}{d}\bigr).$$

Also define the integer constant $M^d_{b,\beta} := \beta^{s_b(1/d)}(\beta^{t_b(1/d)} - 1)$. Here $c_{b,\beta}$ is the (b, β)-content function, first defined in Section 7B, and s_b and t_b count the number of digits after the decimal point of the preperiod and the period, respectively, of base-b expansions; see definition before (7-1).

The following theorem, combined with Corollary 6.4, will prove Theorem 4, completing in turn the proof of Theorem 1.

Theorem 5. *If $b - d \leq 1$, or if $D \leq \frac{b}{2}$, then the function $c^d_{b,b-D}$ is a (b, d, D)-nilgrowth witness.*

We begin the proof of Theorem 5. Recall from Section 6D that a (b, d, D)-nilgrowth witness must satisfy four properties: discreteness, growth, base, and step. We establish the first two immediately.

Lemma 9.1 (discreteness property). *For any $n \in \mathbb{N}$, we have $M^d_{b,\beta} c^d_{b,\beta}(n) \in \mathbb{N}$.*

Proof. It suffices to see that $\beta^{s_b(1/d)}(\beta^{t_b(1/d)} - 1)c_{b,\beta}(n)$ is an integer for $n \in \frac{1}{d}\mathbb{N}$. For $n = \frac{1}{d}$ this follows from Lemma 7.2, and for general $n \in \frac{1}{d}\mathbb{N}$ from Lemmas 7.2 and 7.1. □

Lemma 9.2 (growth property). *We have $c^d_{b,\beta}(n) \asymp n^{\log_b \beta}$.*

Proof. Lemma 7.3. □

It remains to establish the base property and the step property.

For $m, n \in \mathbb{Q}_{\geq 0}$ with $m \geq n$, set

$$R(m, n) := R^d_{b,\beta}(m, n) := D\pi_\beta r_b\bigl(\tfrac{m-n}{d}, \tfrac{n}{d}\bigr).$$

[vii] Incidentally this implies the failure of condition (1′), which should conjecturally replace condition (1) as noted in the remark after the statement of Proposition 8.10.

Here r_b is the carry-digit word, as in Section 7C. We then have, for m, n as above

$$c(n) - c(n-m) = c(m) - R(n, m). \tag{9-1}$$

This is just a restatement of Lemma 7.5, in the form in which we will use it below.

We now use the technical results of Section 8 to prove that our candidate nilgrowth witness satisfies the base property and the step property.

Lemma 9.3 (base property). *Suppose that $d = b$ or $d = b-1$ or $D \le \frac{b}{2}$. Then we have*:

(1) $c(d-D) \le c(d)$;

(2) $0 = c(0) < c(1) < \cdots < c(d-1)$.

Remark. Part (1) is in fact true without the assumption $D \le \frac{b}{2}$, but we do not need this greater generality. Part (2) above is not generally true if $D > \frac{b}{2}$. For example, for $(b, d, D) = (7, 5, 5)$, we have $c(2) = c(3) = 3$; and for $(b, d, D) = (11, 9, 7)$ we have $c(4) = \frac{334}{195} > \frac{316}{195} = c(5)$. The condition delineated here is certainly not optimal, however.

Proof. If $d = b$ or $d = b-1$, then both statements are immediate from the formula in Lemma 8.7. (Note that $c(d)$ is always 1.) Assume therefore that $d \le b-2$.

(1) If $d = D$, then the inequality is trivial; so assume $D < d$. By (9-1), we have

$$c(d) - c(d-D) = c(D) - R(d, D).$$

Certainly $r_b\left(\frac{D}{d}, \frac{d-D}{d}\right)$ can be no greater than 0.1^∞. Therefore

$$R(d, D) = D\pi_\beta r_b\left(\tfrac{D}{d}, \tfrac{d-D}{d}\right) \le D\pi_\beta(0.1^\infty) = \tfrac{D}{\beta-1}.$$

On the other hand, by Proposition 8.10, we know that $c(D) \ge \frac{D(\beta+1)}{\beta(\beta-1)} > \frac{D}{\beta-1}$. Therefore $c(d) > c(d-D)$ (and in fact the inequality is strict).

(2) It suffices to show that, for $0 < i < d$, we have $c(i) > c(i-1)$. By (9-1) this is equivalent to the inequality $c(1) > R(i, 1)$. Since $i < d$, we know that

$$R(i, 1) = D\pi_\beta r_b\left(\tfrac{i-1}{d}, \tfrac{1}{d}\right) \le D\pi_\beta(0.01^\infty) = \tfrac{D}{\beta(\beta-1)}.$$

Now Corollary 8.9 completes the claim. □

Lemma 9.4 (step property). *Suppose $d = b$ or $d = b-1$ or $D \le \frac{b}{2}$. If $(i, j) \in \mathcal{I}$, and n, m are integers with $db^k \le n < db^{k+1}$ and $jb^k \le m$, then*

$$c(n) - c(n - ib^k) \ge c(m) - c(m - jb^k).$$

Here as before $\mathcal{I} = \{(i, j) : 0 \le j < j + D \le i \le d\} \cup \{(d, d)\}$ is the set of pairs (i, j) so that $y^j X^{d-i}$ can appear in the companion polynomial of the recursion in question; see proof of Proposition 6.3.

Proof. We use Lemma 7.4 to divide each equation by β^k, and replace n and m by $\frac{n}{b^k}$ and $\frac{m}{b^k}$, respectively. We therefore seek to show that for $n, m \in \mathbb{Z}\big[\frac{1}{b}\big]_{\geq 0}$ satisfying $d \leq n < db$ and $m \geq j$, we have

$$c(n) - c(n-i) \geq c(m) - c(m-j).$$

Using (9-1) twice and rearranging terms, the desired statement is equivalent to

$$c(i) - c(j) \geq R(n,i) - R(m,j).$$

Since $R(m,j)$ is nonnegative, it suffices to show that

$$c(i) - c(j) \geq R(n,i).$$

We now take two cases. If $i = d$, then $R(n,i) = 0$, so that it suffices to show that $c(d) \geq c(j)$ for $(d, j) \in \mathcal{I}$. For $j = d$, this is clear; and for $j \leq i - D$, this follows from both parts of Lemma 9.3 above.

If, on the other hand, $i < d$, then (9-1) gives us $c(i) - c(j) = c(i-j) - R(i,j)$, which reduces the desired statement to

$$c(i-j) \geq R(n,i) + R(i,j). \tag{9-2}$$

If $d = b$ or $d = b-1$, then $R(i,j) = 0$; since $c(i-j) = \frac{i-j}{d-D} \geq \frac{D}{d-D}$ (Lemma 8.7), it remains to show that $R(n,i) \leq \frac{D}{d-D}$. If $d = b$, then at most one digit is carried, so that $R(n,i) \leq D\pi_\beta(0.1) = \frac{D}{\beta}$. And if $d = b-1$, then every digit may be carried, so that $R(n,i) \leq D\pi_\beta(0.1^\infty) = \frac{D}{\beta-1}$. In both cases, the desired inequality holds.

On the other hand if $d \leq b-2$ $\big($and so $D \leq \frac{b}{2}\big)$, then we reason as follows. The left-hand side of desired inequality (9-2) is bounded below by $c(D)$, and the right-hand side is bounded above by

$$D\pi_\beta(0.1^\infty) + D\pi_\beta(0.01^\infty) = \frac{D(\beta+1)}{\beta(\beta-1)}.$$

Therefore it suffices to show that $c(D) \geq \frac{D(\beta+1)}{\beta(\beta-1)}$ which is established in Proposition 8.10. □

Lemmas 9.3 and 9.4 complete the proof of Theorem 5, which in turn completes the proof of Theorem 4, and hence of Theorem 1.

10. Complements

10A. ***Refinement of Theorem 2.*** We state a refinement of the toy version of the NGT. One can also obtain similar refinements of Theorem 4.

Theorem 6 (refined toy NGT). *Let $\mathbb{F}$ be a field of characteristic p and let $q = p^k$. Suppose that $T : \mathbb{F}[y] \to \mathbb{F}[y]$ is an $\mathbb{F}$-linear operator satisfying the following two conditions*:

(1) *For $f \in \mathbb{F}[y]$, we have* $\deg T(f) \leq \deg f - E$ *for some $E \geq 1$.*

(2) *The sequence $\{T(y^n)\}_n$ satisfies a linear recursion whose companion polynomial has the shape*

$$P = (X + cy)^d + (\textit{terms of total degree} \leq d - D) \in \mathbb{F}[y][X]$$

for some $d \leq q$, $D \geq 1$, and $c \in \mathbb{F}$.

Then

$$N_T(f) \leq \frac{(q-D)(q-1)}{E(q-D-1)}(\deg f)^{\log(q-D)/\log q}.$$

Remark. Since the total degree of the companion polynomial of the recursion is the same as the order, it suffices to check the condition that $\deg T(f) \leq \deg f - E$ on $f = 1, y, \ldots, y^{d-1}$ only.

Proof. The sequence $\{T(y^n)\}_n$ also satisfies the linear recursion with companion polynomial

$$P' = (X + cy)^{q-d} P = X^q + cy^q + (\text{terms of total degree} \leq q - D).$$

Let $c = c_{q,q-D}$, define $\tilde{c} : \mathbb{F}[y] \to \mathbb{N} \cup \{-\infty\}$ from c as in the proof of Theorem 2, and follow the same inductive argument mutatis mutandis to show that $\tilde{c}(Tf) \leq \tilde{c}(f) - E$. (The main adjustment is in Proposition 4.1(4); if i is a digit base q and $n \geq i$ has no more than 2 digits base q, then $c(n) - c(n-i)$ is either i or $i - D$; see Lemma 7.5 for a conceptual explanation.)

We have therefore shown that $N_T(y^n) \leq \frac{c(n)}{E}$. Lemma 7.3 completes the proof. □

10B. *Comments on α in Theorem 4.* How optimal is the order of growth of the nilpotence index α from the empty-middle NGT?

To this end, if K is a field, and $T : K[y] \to K[y]$ a degree-lowering linear operator, let

$$\alpha(T) := \limsup_{n\to\infty} \frac{\log N_T(y^n)}{\log n},$$

and let

$$\alpha_K(d, D) = \sup_{T \in \mathcal{L}_K(d,D)} \{\alpha(T)\},$$

where $\mathcal{L}_K(d, D)$ is the set of degree-lowering operators $T : K[y] \to K[y]$ with $\{T(y^n)\}_n$ satisfying a recurrence with companion polynomial $X^d + cy^d + (\text{terms of total degree} \leq d - D)$ for some $c \in K$. Since $N_T(y^n) \leq n$, we know that $\alpha_K(d, D) \leq 1$. The following proposition clarifies that studying $\alpha_K(d, D)$ is only interesting in characteristic p.

Proposition 10.1. *If K has characteristic zero and $D < d$, then $\alpha_K(d, D) = 1$.*

Proof. Fix d, and consider the recursion operator $T : K[y] \to K[y]$ defined by the companion polynomial $P = X^d - y^d - y$, corresponding to the recurrence $T(y^n) = (y^d + y)T(y^{n-d})$, and initial values $\{T(y^n)\}_{n=0}^{d-1} = \{0, \ldots, 0, 1\}$. We will show that $N_T(y^{kd+d-1}) = \lfloor \frac{k}{d-1} \rfloor + 1$, which will establish that $\alpha_T = 1$.

Indeed, from the recurrence, we have $T(y^n) = 0$ if $n \not\equiv -1 \bmod d$, and $T(y^{kd+d-1}) = (y^d + y)^k$. For $f = \sum a_n y^n \in K[y]$, write $e(f)$ for the set $\{n : a_n \neq 0\}$ of exponents appearing in f. From above, we see that

$$e(T(y^{kd+d-1})) = \{k, k + (d-1), k + 2(d-1), \ldots, kd\}.$$

More generally, we can show by induction that the set $S_{m,k} := e(T^m(y^{kd+d-1}))$ is an arithmetic progression of common difference $d-1$, greatest term $d(k-(m-1)(d-1))$, and length $k-(m-1)(d-1)+1$, so long as $k \geq (m-1)(d-1)$; otherwise the set is empty and $T^m(y^{kd+d-1}) = 0$. Indeed, from the explicit formulation of $T(y^n)$, we see that if $S_{m,k}$ is as claimed, then the greatest element of $S_{m+1,k} = N-(d-1)$, where N is the greatest element of $S_{m,k}$ congruent to $d-1$ modulo d. Since the maximum element of $S_{m,k}$ is congruent to 0 modulo d, and every successive smaller element is $d-1$ less, we see that N is the d-th greatest element of $S_{m,k}$. In other words,

$$N = d(k-(m-1)(d-1)) - (d-1)^2,$$

so that the greatest element of $S_{m+1,k}$ is $N-(d-1) = d(k-m(d-1))$, as desired. Since the relevant coefficients are positive and we are in infinite characteristic, no cancellation of intermediate terms is possible. Finally, since $T^m(y^{kd+d-1}) \neq 0$ if and only if $k \geq (m-1)(d-1)$, we have $N_T(y^{kd+d-1}) = \lfloor \frac{k}{d-1} \rfloor + 1$, as claimed. □

For $\mathbb{F}$ a field of characteristic p, let us confine our inquiry to the case where d can be taken to be a power of p, as in Theorem 6 above. Theorem 6 tells us that $\alpha_{\mathbb{F}}(p^k, D) \leq \log(p^k - D)/\log p^k$. How optimal is this estimate? Computationally, it appears that for $k = 1$ this inequality is optimal. A few examples for $D = 1$:

Examples. (1) $p = 3$: The recursion operator T with companion polynomial $X^3 + yX - y^3$ and initial values $\{0, 1, y\}$ appears to achieve $N_T(y^n) = c_{3,2}(n)$ infinitely often.

(2) $p = 5$: The recursion operator T with companion polynomial

$$X^5 + 3yX^3 + y^2X^2 + 3y^3X + 4y^5$$

and initial values $[0, 1, y, y^2, y^3]$ appears to achieve $N_T(y^n) = c_{5,4}(n)$ for "most" n; every counterexample n has 0s in its base-5 expansion.

(3) $p = 7$: The recursion operator T with companion polynomial

$$X^7 + 3y^2X^4 + 6y^3X^3 + 5y^4X^2 + 3y^5X + 6y^7$$

appears to achieve $N_T(y^n) = c_{7,6}(n)$ for most n. For $n < 1000$, there are only 36 counterexamples, and $c_{7,6}(n) - N_T(y^n) \leq 3$ for each one.

(4) $p = 11$. The recursion operator T with companion polynomial

$$P_T = X^{11} + 6yX^9 + 2y^2X^8 + 3y^3X^7 + 6y^4X^6 + 8y^6X^4 + y^8X^2 + 9y^9X + 10y^{11}$$

appears to achieve $N_T(y^n) = c_{11,10}(n)$ for most n. For $n < 1000$, there are only 8 counterexamples, and $N_T(y^n) = c_{11,10}(n) - 1$ for each one.

The estimate appears not to be optimal as soon as $k \geq 2$.

Acknowledgements

I am greatly indebted to my Ph.D. advisor Joël Bellaïche for sparking and supporting this investigation into modular forms modulo p. The initial idea for the nilpotence method is his — a most generous gift.

I would also like to thank Paul Monsky for many illuminating discussions on the topic of mod p modular forms and Hecke algebras, and Kiran Kedlaya for helpful comments on an earlier version of Theorem 1. Part of the writing of this document was carried out at the Max Planck Institute for Mathematics in Bonn, and I am grateful for their hospitality. Extensive computations related to this project were coded and carried out in SAGE [$\geq$ 2018].

References

[Al Hajj Shehadeh et al. 2009] H. Al Hajj Shehadeh, S. Jaafar, and K. Khuri-Makdisi, "Generating functions for Hecke operators", *Int. J. Number Theory* **5**:1 (2009), 125–140. MR Zbl

[Allouche and Shallit 1992] J.-P. Allouche and J. Shallit, "The ring of k-regular sequences", *Theoret. Comput. Sci.* **98**:2 (1992), 163–197. MR Zbl

[Bellaïche 2012] J. Bellaïche, "Pseudodeformations", *Math. Z.* **270**:3-4 (2012), 1163–1180. MR Zbl

[Bellaïche and Khare 2015] J. Bellaïche and C. Khare, "Level 1 Hecke algebras of modular forms modulo p", *Compos. Math.* **151**:3 (2015), 397–415. MR Zbl

[Buzzard and Calegari 2005] K. Buzzard and F. Calegari, "Slopes of overconvergent 2-adic modular forms", *Compos. Math.* **141**:3 (2005), 591–604. MR Zbl

[Conrad 2016] K. Conrad, "Solving linear recursions over all fields", preprint, 2016, Available at https://tinyurl.com/kconrecu.

[Deo 2017] S. V. Deo, "Structure of Hecke algebras of modular forms modulo p", *Algebra Number Theory* **11**:1 (2017), 1–38. MR Zbl

[Emerton 2011] M. Emerton, "p-adic families of modular forms (after Hida, Coleman, and Mazur)", exposé 1013, pp. 31–61 in *Séminaire Bourbaki,* 2009/2010, Astérisque **339**, Soc. Math. France, Paris, 2011. MR Zbl

[Gerbelli-Gauthier 2016] M. Gerbelli-Gauthier, "The order of nilpotence of Hecke operators mod 2: a new proof", *Res. Number Theory* **2** (2016), art. id. 7. MR Zbl

[Gouvêa and Mazur 1998] F. Q. Gouvêa and B. Mazur, "On the density of modular representations", pp. 127–142 in *Computational perspectives on number theory* (Chicago, 1995), edited by D. A. Buell and J. T. Teitelbaum, AMS/IP Stud. Adv. Math. **7**, Amer. Math. Soc., Providence, RI, 1998. MR Zbl

[Jochnowitz 1982] N. Jochnowitz, "A study of the local components of the Hecke algebra mod l", *Trans. Amer. Math. Soc.* **270**:1 (1982), 253–267. MR Zbl

[Medvedovsky 2015] A. Medvedovsky, *Lower bounds on dimensions of mod-p Hecke algebras: the nilpotence method*, Ph.D. thesis, Brandeis University, 2015, Available at https://search.proquest.com/docview/1687761247.

[Medvedovsky $\geq$ 2018] A. Medvedovsky, "Lower bounds on dimensions of mod-p Hecke algebras in the genus-zero case", in preparation.

[Nicolas and Serre 2012a] J.-L. Nicolas and J.-P. Serre, "Formes modulaires modulo 2: l'ordre de nilpotence des opérateurs de Hecke", *C. R. Math. Acad. Sci. Paris* **350**:7-8 (2012), 343–348. MR Zbl

[Nicolas and Serre 2012b] J.-L. Nicolas and J.-P. Serre, "Formes modulaires modulo 2: structure de l'algèbre de Hecke", *C. R. Math. Acad. Sci. Paris* **350**:9-10 (2012), 449–454. MR Zbl

[SAGE $\geq$ 2018] SAGE, "SAGE software", Available at http://www.sagemath.org.

[Serre 1986] J.-P. Serre, *Œuvres, III: 1972-1984*, Springer, 1986. MR Zbl

[Swinnerton-Dyer 1973] H. P. F. Swinnerton-Dyer, "On l-adic representations and congruences for coefficients of modular forms", pp. 1–55 in *Modular functions of one variable, III* (Antwerp, 1972), edited by W. Kuyk and J.-P. Serre, Lecture Notes in Math. **350**, Springer, 1973. Correction in *Modular functions of one variable, IV*, Lecture Notes in Math. **476**, Springer, 1975, pp. 149. MR Zbl

[Tate 1994] J. Tate, "The non-existence of certain Galois extensions of $\mathbb{Q}$ unramified outside 2", pp. 153–156 in *Arithmetic geometry* (Tempe, AZ, 1993), edited by N. Childress and J. W. Jones, Contemp. Math. **174**, Amer. Math. Soc., Providence, RI, 1994. MR Zbl

Communicated by Kiran S. Kedlaya
Received 2017-07-27 Revised 2018-01-12 Accepted 2018-02-23

medved@mpim-bonn.mpg.de *Max-Planck-Institut für Mathematik, Bonn, Germany*

ALGEBRA AND NUMBER THEORY 12:3 (2018)
dx.doi.org/10.2140/ant.2018.12.723

Algebraic de Rham theory for weakly holomorphic modular forms of level one

Francis Brown and Richard Hain

We establish an Eichler–Shimura isomorphism for weakly modular forms of level one. We do this by relating weakly modular forms with rational Fourier coefficients to the algebraic de Rham cohomology of the modular curve with twisted coefficients. This leads to formulae for the periods and quasiperiods of modular forms.

1. Introduction

Let M_n denote the $\mathbb{Q}$-vector space of modular forms of weight n and level one with rational Fourier coefficients. Let $S_n \subset M_n$ denote the subspace of cusp forms. The Eichler–Shimura isomorphism [Eichler 1957; Shimura 1959] is usually expressed as a pair of isomorphisms

$$\begin{aligned} M_{n+2} \otimes_{\mathbb{Q}} \mathbb{C} &\xrightarrow{\sim} H^1(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}})^+ \otimes_{\mathbb{Q}} \mathbb{C}, \\ S_{n+2} \otimes_{\mathbb{Q}} \mathbb{C} &\xrightarrow{\sim} H^1(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}})^- \otimes_{\mathbb{Q}} \mathbb{C}, \end{aligned} \tag{1-1}$$

where the right-hand side denotes group cohomology with coefficients in $V_n^{\mathrm{B}} = \mathrm{Sym}^n V^{\mathrm{B}}$, where V^{B} denotes the standard two-dimensional representation of SL_2 over $\mathbb{Q}$ with basis $\boldsymbol{a}, \boldsymbol{b}$, and $\pm$ denote eigenspaces with respect to the real Frobenius (complex conjugation). One wants to think of this theorem as a special case of the comparison isomorphism between algebraic de Rham cohomology and Betti cohomology, each of which has a natural $\mathbb{Q}$-structure. The rational structures on these groups then enables one to define periods. Each cuspidal Hecke eigenspace, like an elliptic curve, should have four periods (two periods and two quasiperiods) corresponding to the entries of a 2×2 period matrix. However, the isomorphisms (1-1) do not generate enough periods since each one only produces a single period for every modular form. To obtain a full set of periods, one needs to consider "modular forms of the second kind."

In this note, we compute the algebraic de Rham cohomology of the moduli stack $\mathcal{M}_{1,1}$ of elliptic curves and relate it to weakly holomorphic modular forms (modular forms which are holomorphic on the upper half plane but with poles at the cusp). From this, we deduce a $\mathbb{Q}$-de Rham–Eichler–Shimura isomorphism, and a definition of the period matrix of a Hecke eigenspace.

Brown was partially supported by ERC grant 724638. Hain was partially supported by National Science Foundation through grant DMS-1406420. He was also supported by ERC grant 724638 during a visit to Oxford during which this paper was written.
MSC2010: primary 11F11; secondary 11F23, 11F25, 11F67.
Keywords: weakly holomorphic modular form, algebraic de Rham cohomology.

Before stating the main results, it may be instructive to review the familiar case of an elliptic curve E over $\mathbb{Q}$ with equation $y^2 = 4x^3 - ux - v$. The de Rham cohomology $H^1_{\mathrm{dR}}(E, \mathbb{Q})$ is a two-dimensional vector space over $\mathbb{Q}$, as is the Betti (singular) cohomology $H^1(E(\mathbb{C}); \mathbb{Q})$. The comparison isomorphism is a canonical isomorphism

$$H^1_{\mathrm{dR}}(E; \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C} \xrightarrow{\sim} H^1(E(\mathbb{C}); \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C}. \tag{1-2}$$

On the other hand, the space $F^1 H^1_{\mathrm{dR}}(E; \mathbb{Q}) := H^0(E; \Omega^1_{E/\mathbb{Q}})$ is one-dimensional and spanned by the holomorphic differential dx/y. The Betti cohomology $H^1(E(\mathbb{C}); \mathbb{Q})$ splits into two eigenspaces under the action of complex conjugation, with eigenvalues ± 1. The analogue of the Eichler–Shimura isomorphisms, in this setting, are exactly the formulae

$$H^0(E; \Omega^1_{E/\mathbb{Q}}) \otimes_{\mathbb{Q}} \mathbb{C} \xrightarrow{\sim} H^1(E(\mathbb{C}); \mathbb{Q})^+ \otimes_{\mathbb{Q}} \mathbb{C},$$
$$H^0(E; \Omega^1_{E/\mathbb{Q}}) \otimes_{\mathbb{Q}} \mathbb{C} \xrightarrow{\sim} H^1(E(\mathbb{C}); \mathbb{Q})^- \otimes_{\mathbb{Q}} \mathbb{C},$$

given by integrating the form dx/y over invariant and antiinvariant cycles in $E(\mathbb{C})$ with respect to complex conjugation, respectively. This is clearly a weaker statement than the comparison isomorphism (1-2). To obtain all periods of the elliptic curve, one needs to consider, in addition, period integrals of the differential $x\,dx/y$ of the *second kind*, which provides an isomorphism

$$H^1_{\mathrm{dR}}(E; \mathbb{Q}) \cong \mathbb{Q}\frac{dx}{y} \oplus \mathbb{Q}\frac{x\,dx}{y}. \tag{1-3}$$

Remark 1.1. There is also an isomorphism of

$$H^1(E(\mathbb{C}); \mathbb{C})/H^0(E; \Omega^1_{E/\mathbb{Q}}) \otimes_{\mathbb{Q}} \mathbb{C}$$

with the space of antiholomorphic differentials on $E(\mathbb{C})$. Since this isomorphism is only defined over $\mathbb{C}$, one loses the rational structure, and cannot define periods in this manner. The analogue for modular forms is the isomorphism

$$(H^1(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}}) \otimes_{\mathbb{Q}} \mathbb{C})/(M_{n+2} \otimes_{\mathbb{Q}} \mathbb{C})$$

with the space of antiholomorphic cusp forms of weight $n+2$.

1A. ***Statement of the theorem.*** Let $\mathcal{E}$ denote the universal elliptic curve over the moduli stack $\mathcal{M}_{1,1}$ (over $\mathbb{Q}$) of elliptic curves. It defines a rank two algebraic vector bundle $\mathscr{V}$, equipped with the Gauss–Manin connection ∇. For all $n \geq 1$, set

$$\mathscr{V}_n = \mathrm{Sym}^n \mathscr{V}$$

and denote the induced connection by ∇ also. Grothendieck defined algebraic de Rham cohomology, which is a finite-dimensional $\mathbb{Q}$-vector space:

$$H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathscr{V}_n).$$

In order to describe this space in terms of modular forms, for each $n \in \mathbb{Z}$, let $M_n^!$ denote the $\mathbb{Q}$-vector space of weakly holomorphic modular forms of weight n that have a Fourier expansion

$$\sum_{k \geq -N} a_k q^k$$

with $a_k \in \mathbb{Q}$. They can have negative weights. Such a form is called a cusp form if $a_0 = 0$. Let $S_n^! \subset M_n^!$ denote the subspace of cusp forms. Now consider the differential operator

$$\mathcal{D} = q \frac{d}{dq}. \tag{1-4}$$

It does not in general preserve modularity, but an identity due to Bol [1949] implies that its powers induce a linear map

$$\mathcal{D}^{n+1} : M_{-n}^! \to S_{n+2}^!$$

for every $n \geq 0$. Our main theorem was inspired by the recent paper [Guerzhoy 2008]. After writing this note, we learnt that similar results for modular curves of higher level were implicitly obtained by Coleman [1996] in the p-adic setting, and independently by Scholl [1985]. As pointed out in the very recent paper [Kazalicki and Scholl 2016], a description of algebraic de Rham cohomology in terms of modular forms of the second kind seems not to have been stated explicitly anywhere in the literature up until that point. An approach using the Cousin resolution was subsequently given in [Candelori 2014]. The following theorem can be indirectly deduced from the results of these papers by viewing a modular form of level one as an invariant form of levels 3 and 4.

Theorem 1.2. *For each $n \geq 0$, there is a canonical isomorphism of $\mathbb{Q}$-vector spaces*

$$\varpi : M_{n+2}^! / \mathcal{D}^{n+1} M_{-n}^! \xrightarrow{\sim} H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathcal{V}_n).$$

The space on the left contains the space of holomorphic modular forms as a subspace:

$$M_{n+2} \subset M_{n+2}^! / \mathcal{D}^{n+1} M_{-n}^!.$$

More precisely, the group $H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathcal{V}_n)$ carries a natural Hodge filtration

$$H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathcal{V}_n) = F^0 \supset F^1 = \cdots = F^{n+1} \supset F^{n+2} = 0$$

and F^{n+1} is the image of M_{n+2} under ϖ. That is,

$$\varpi : M_{n+2} \xrightarrow{\sim} F^{n+1} H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathcal{V}_n).$$

A splitting of the Hodge filtration is discussed in Section 6.

1B. ***Comparison isomorphism.*** Grothendieck's algebraic de Rham theorem implies (see Section 3) that there is a canonical isomorphism of complex vector spaces

$$H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathcal{V}_n) \otimes_{\mathbb{Q}} \mathbb{C} \xrightarrow{\sim} H^1(\mathcal{M}_{1,1}(\mathbb{C}); \mathbb{V}_n^{\mathrm{B}}),$$

where the right-hand space is the Betti (singular) cohomology of $\mathcal{M}_{1,1}(\mathbb{C})$ with coefficients in the complex local system $\mathbb{V}_n^{\mathrm{B}} = \mathrm{Sym}^{n-1} R^1\pi_*\mathbb{C}$, where $\pi : \mathcal{E} \to \mathcal{M}_{1,1}$ is the universal elliptic curve. Let V_n^{B} denote its fibre $H^0(\mathfrak{H}, \mathbb{V}_n^{\mathrm{B}})$ at the tangent vector $\partial/\partial q$. Since $\mathcal{M}_{1,1}(\mathbb{C})$ is the orbifold quotient of the upper half plane by $\mathrm{SL}_2(\mathbb{Z})$, its cohomology is computed by group cohomology of $\mathrm{SL}_2(\mathbb{Z})$ and we immediately deduce the following consequence of Grothendieck's theorem:

Corollary 1.3. *There is a canonical isomorphism*

$$\mathrm{comp}_{\mathrm{B,dR}} : H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathscr{V}_n) \otimes_{\mathbb{Q}} \mathbb{C} \xrightarrow{\sim} H^1(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}} \otimes_{\mathbb{Q}} \mathbb{C}).$$

Combined with the previous theorem, we deduce an algebraic de Rham version of the Eichler–Shimura isomorphism. It is the analogue for modular forms of the isomorphism (1-2).

Corollary 1.4. *There is a canonical isomorphism*

$$M^!_{n+2}/\mathcal{D}^{n+1}M^!_{-n} \otimes_{\mathbb{Q}} \mathbb{C} \xrightarrow{\sim} H^1(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}} \otimes_{\mathbb{Q}} \mathbb{C}). \tag{1-5}$$

The dimension of the space on the left-hand side was computed in [Guerzhoy 2008], the dimension of the right-hand space by Eichler–Shimura: both are $1 + 2\dim S_n$. Restricting the previous isomorphism to the subspace M_{n+2} of holomorphic modular forms, and projecting onto the positive or negative eigenspaces with respect to complex conjugation on the right-hand space gives back the two isomorphisms (1-1).

For any weakly holomorphic modular form $f \in M^!_{n+2}$, its image under the comparison isomorphism is given explicitly by the cohomology class of the cocycle:

$$\gamma \mapsto (2\pi i)^{n+1} \int_{\gamma^{-1}z_0}^{z_0} f(z)(z\boldsymbol{a} - \boldsymbol{b})^n \, dz, \tag{1-6}$$

where $\boldsymbol{a}, \boldsymbol{b}$ is a basis of V^{B}, which we think of as the first rational Betti cohomology group of the elliptic curve $\mathbb{C}/(\mathbb{Z} \oplus z\mathbb{Z})$. Its cohomology class does not depend on the choice of basepoint $z_0 \in \mathfrak{H}$. A different version of this map (and without the rational structures) was described in [Bringmann et al. 2013]. See also [Bruggeman et al. 2014, Theorem A].

1C. ***Periods.*** The isomorphism (1-5) is compatible with the action of Hecke operators. The action of Hecke operators on the left-hand side was defined in [Guerzhoy 2008]. The eigenspace of an Eisenstein series is one-dimensional, that corresponding to a cusp form is two-dimensional. Let f be a cusp Hecke eigenform and $K_f \subset \mathbb{R}$ the field generated by its Fourier coefficients. Let

$$V_f^{\mathrm{dR}} \subset \big(M^!_{n+2}/D^{n+1}M^!_{-n}\big) \otimes_{\mathbb{Q}} K_f$$

denote the Hecke eigenspace generated by f. It is a two-dimensional K_f-vector space. Let

$$V_f^{\mathrm{B}} \subset H^1\big(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}} \otimes_{\mathbb{Q}} K_f\big)$$

denote the corresponding Betti eigenspace. It is also a two-dimensional K_f-vector space, and decomposes into invariant and antiinvariant eigenspaces with respect to the real Frobenius. We deduce from (1-5) a

canonical isomorphism
$$\mathrm{comp}_{\mathrm{B,dR}} : V_f^{\mathrm{dR}} \otimes_{K_f} \mathbb{C} \xrightarrow{\sim} V_f^{\mathrm{B}} \otimes_{K_f} \mathbb{C}.$$

Definition 1.5. Define a *period matrix* P_f of f to be the matrix of $\mathrm{comp}_{\mathrm{B,dR}}$ written in a K_f-basis of V_f^{dR} and V_f^{B}. We can assume that the basis of V_f^{B} is compatible with decomposition into eigenspaces for the action of the real Frobenius. It is of the form
$$P_f = \begin{pmatrix} \eta_f^+ & \omega_f^+ \\ i\eta_f^- & i\omega_f^- \end{pmatrix},$$
where $\omega_f^{\pm}, \eta_f^{\pm} \in \mathbb{R}$. It is well-defined up to right multiplication by a lower-triangular matrix with entries in K_f, and entries in $K_f^{\times}$ on the diagonal. The $\omega_f^+, i\omega_f^-$ are the holomorphic periods [Manin 1973].

Remark 1.6. Only the holomorphic periods $\omega_f^+, i\omega_f^-$ can be obtained from the classical Eichler–Shimura isomorphisms (1-1).

Theorem 1.7. *If f has weight $2n$, then* $\det(P_f) \in (2\pi i)^{2n-1} K_f^{\times}$.

2. Definitions and background

2A. *Weakly holomorphic modular forms.*

Definition 2.1. For every $n \in \mathbb{Z}$, let $M_n^!$ denote the $\mathbb{Q}$-algebra of weakly holomorphic modular forms of level 1 and weight n with rational Fourier coefficients. It is the $\mathbb{Q}$-vector space of holomorphic functions $f : \mathfrak{H} \to \mathbb{C}$ on the upper half plane $\mathfrak{H}$ such that
$$f(\gamma z) = (cz+d)^n f(z) \quad \text{for all} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \tag{2-1}$$
which admit a Fourier expansion of the form
$$f = \sum_{k \geq -N} a_k q^k, \quad a_k \in \mathbb{Q}.$$
The space $S_n^! \subset M_n^!$ of cusp forms is the subspace of functions satisfying $a_0 = 0$.

Proposition 2.2 (Bol's identity). *For all $n \geq 0$, there is a linear map*
$$\mathcal{D}^{n+1} : M_{-n}^! \to M_{n+2}^!.$$

Proof. This result follows automatically from our proof of Theorem 1.2. We provide a more direct proof for completeness.

For $f : \mathfrak{H} \to \mathbb{C}$ a real analytic modular form of weight $n \in \mathbb{Z}$ define
$$d_n f = \frac{1}{2\pi i}\left(\frac{\partial f}{\partial z} + \frac{nf}{z - \bar{z}}\right).$$
It is well-known by Maass, and easily verified, that this operator respects the transformation property (2-1), and therefore $d_n f$ is a real analytic modular form of weight $n+2$.

The proposition follows from the following identity, for all $n \geq 0$:

$$\mathcal{D}^{n+1} = d_n d_{n-2} \cdots d_{2-n} d_{-n}.$$

To verify this, write a real analytic function on $\mathfrak{H}$ as a formal power series in $(z - \bar{z})$ and $\bar{z}$. It suffices to verify the formula for

$$f_{a,b} = (2\pi i(z - \bar{z}))^a (-2\pi i \bar{z})^b,$$

where $a, b \geq 0$. We check that $d_m f_{a,b} = (a+m) f_{a-1,b}$, and hence

$$d_n d_{n-2} \cdots d_{2-n} d_{-n}\big(f_{a,b}\big) = a(a-1) \cdots (a-n) f_{a-n-1,b}.$$

On the other hand, $\log q = 2\pi i z$ and $\log \bar{q} = -2\pi i \bar{z}$ and hence

$$f_{a,b} = (\log q + \log \bar{q})^a (\log \bar{q})^b.$$

Since $\mathcal{D} = q\partial/\partial q = \partial/\partial(\log q)$, the operator $\mathcal{D}^{n+1}$ acts on $f_{a,b}$ in an identical manner. $\square$

2B. *Moduli of elliptic curves.* Let $\Bbbk$ be a commutative ring with $6 \in \Bbbk^\times$. In much of this paper, $\Bbbk$ will be either the universal such ring, $\Bbbk = \mathbb{Z}\big[\frac{1}{6}\big]$, or $\Bbbk = \mathbb{Q}$. The formula

$$\lambda \cdot u = \lambda^4 u, \quad \lambda \cdot v = \lambda^6 v$$

defines a left action

$$\mu : \mathbb{G}_m \times \mathbb{A}^2 \to \mathbb{A}^2 \tag{2-2}$$

of the multiplicative group $\mathbb{G}_m$ on the affine plane $\mathbb{A}^2 := \operatorname{Spec} \Bbbk[u, v]$, and defines a grading on $\Bbbk[u, v]$ called the weight. The discriminant function

$$\Delta = u^3 - 27v^2$$

has weight 12 under this action, so that $\mathbb{G}_m$ also acts on the graded ring $\Bbbk[u, v][\Delta^{-1}]$. Let D be the vanishing locus of the discriminant Δ. Set

$$X = \mathbb{A}^2 - D := \operatorname{Spec} \Bbbk[u, v][\Delta^{-1}].$$

Then X is the moduli scheme of elliptic curves over $\Bbbk$ together with a nonzero abelian differential, [Katz and Mazur 1985, §(2.2.6)]. Its coordinate ring

$$\mathcal{O}(X) = \Bbbk[u, v][\Delta^{-1}] = \bigoplus_{m \text{ even}} \operatorname{gr}_m \mathcal{O}(X)$$

is graded by the $\mathbb{G}_m$ action. The sheaf of regular functions on X will be denoted by $\mathcal{O}_X$.

The universal elliptic curve $\mathcal{E}$ over X is the subvariety of $\mathbb{P}^2 \times_\Bbbk X$ which is the Zariski closure of the affine scheme defined by the equation

$$y^2 = 4x^3 - ux - v \in \mathcal{O}(X)[x, y].$$

The universal abelian differential on $\mathcal{E}$ is dx/y. The multiplicative group $\mathbb{G}_m$ acts on (x, y) by $\lambda \cdot (x, y) = (\lambda^2 x, \lambda^3 y)$ and on the abelian differential by $\lambda \cdot dx/y = \lambda^{-1} dx/y$.

The moduli stack of elliptic curves over $\Bbbk$ is the stack quotient

$$\mathcal{M}_{1,1}/\Bbbk = \mathbb{G}_m \backslash\backslash X$$

of X by $\mathbb{G}_m$. Its Deligne–Mumford compactification is

$$\overline{\mathcal{M}}_{1,1}/\Bbbk = \mathbb{G}_m \backslash\backslash Y,$$

where $Y = \mathbb{A}^2 - \{0\}$. In down to earth terms, to work on $\mathcal{M}_{1,1}/\Bbbk$ is to work $\mathbb{G}_m$-equivariantly on X, and to work on $\overline{\mathcal{M}}_{1,1}/\Bbbk$ is to work $\mathbb{G}_m$-equivariantly on Y.

2C. *Upper half plane description.* When $\Bbbk = \mathbb{C}$, this description of $\mathcal{M}_{1,1}$ relates to the traditional upper half plane model via Eisenstein series. Denote by G_{2n} the normalised Eisenstein series

$$\mathsf{G}_{2n}(q) = -\frac{B_{2n}}{4n} + \sum_{m=1}^{\infty} \sigma_{2n-1}(m) q^m$$

of weight $2n$ where B_k is the k-th Bernoulli number, $q = e^{2\pi i z}$, and $\sigma_k(m) = \sum_{d|m} d^k$ the divisor function.

Define a map $\rho : \mathfrak{H} \to X(\mathbb{C})$ by $\rho(z) = (u(z), v(z))$, where

$$u = 20\mathsf{G}_4(z) \quad \text{and} \quad v = \tfrac{7}{3}\mathsf{G}_6(z). \tag{2-3}$$

The map ρ factors through the punctured q-disk and induces a graded ring isomorphism to the space of holomorphic modular forms

$$\rho^* : \mathbb{Q}[u, v] \xrightarrow{\sim} \bigoplus_{n \text{ even}} M_n.$$

The pull-back of the discriminant Δ is the Ramanujan τ-function $\Delta(z)$, which vanishes nowhere on $\mathfrak{H}$. It follows that ρ induces a graded ring isomorphism

$$\rho^* : \mathcal{O}(X) \xrightarrow{\sim} \bigoplus_{n \text{ even}} M_n^!.$$

and that the image of ρ is indeed contained in $X(\mathbb{C})$. The ring of weakly modular forms is therefore nothing other than the affine ring of functions on X.

The elliptic curve $\mathbb{C}/(\mathbb{Z} \oplus z\mathbb{Z})$ is mapped isomorphically to the elliptic curve

$$y^2 = 4x^3 - 20\mathsf{G}_4(z) - \tfrac{7}{3}\mathsf{G}_6(z)$$

by $w \mapsto (\wp_z(w)/(2\pi i)^2, \wp_z'(w)/(2\pi i)^3)$, where $\wp_z(w)$ denotes the Weierstrass $\wp$-function. The abelian differential dx/y pulls back to $2\pi i \, dw$. This implies that the map (2-3) induces an isomorphism

$$\mathrm{SL}_2(\mathbb{Z}) \backslash\backslash \mathfrak{H} \xrightarrow{\sim} (\mathbb{G}_m \backslash\backslash X)(\mathbb{C}) = \mathcal{M}_{1,1}(\mathbb{C})$$

of analytic stacks, since $\rho(\gamma z) = ((cz+d)^4 u(z), (cz+d)^6 v(z))$ for $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ of the form (2-1), by modularity of G_4 and G_6.

The following lemma motivates the choice of the basis of one-forms considered later.

Lemma 2.3. *Pulling back along* (2-3) *we have*

$$\frac{2u\,dv - 3v\,du}{\Delta} = \frac{2}{3}\frac{dq}{q}.$$

Proof. Formulae due to Ramanujan imply that

$$\mathcal{D}E_2 = (E_2^2 - E_4)/12, \quad \mathcal{D}E_4 = (E_2E_4 - E_6)/3, \quad \mathcal{D}E_6 = (E_2E_6 - E_4^2)/2,$$

where $E_{2n} = -(4n/B_{2n})\mathsf{G}_{2n}$ are the Eisenstein series normalised such that their constant Fourier coefficient is 1. These are easily verified by computing the first few terms in their Fourier expansion [Zagier 2008, Proposition 15]. It follows that

$$3E_6\,\mathcal{D}E_4 - 2E_4\,\mathcal{D}E_6 = E_4^3 - E_6^2 = 1728\Delta.$$

Now substitute $E_4 = 240\mathsf{G}_4 = 12u$ and $E_6 = -504\mathsf{G}_6 = -216v$. $\square$

Remark 2.4. The functions u, v have the following q-expansions:

$$u = 20\,\mathsf{G}_4 = \tfrac{1}{12} + 20q + 180q^2 + 560q^3 + 1460q^4 + \cdots$$
$$v = \tfrac{7}{3}\,\mathsf{G}_6 = -\tfrac{1}{216} + \tfrac{7}{3}q + 77q^2 + \tfrac{1708}{3}q^3 + \tfrac{7399}{3}q^4 + \cdots$$

They have coefficients in $\mathbb{Z}\left[\frac{1}{6}\right]$. Furthermore,

$$\Delta^{-1} = \frac{1}{q} + 24 + 324q + 3200q^2 + 25650q^3 + 176256q^4 + \cdots$$

has integer coefficients. It follows that

$$\frac{M^!_{n+2}}{\mathcal{D}^{n+1}M^!_{-n}}$$

has a natural $\mathbb{Z}\left[\frac{1}{6}\right]$-structure, given by series whose Fourier coefficients lie in $\mathbb{Z}\left[\frac{1}{6}\right]$. This is because the operator $\mathcal{D}$ acts on the ring $\mathbb{Z}[\![q]\!]$.

The $\mathbb{Z}\left[\frac{1}{6}\right]$ structure on the affine ring of $X/\mathbb{Z}\left[\frac{1}{6}\right]$ coincides, by Remark 2.4, with the $\mathbb{Z}\left[\frac{1}{6}\right]$ structure on Fourier expansions.

2D. *Differential forms.* The following one-forms play a special role:

$$\psi = \frac{1}{12}\frac{d\Delta}{\Delta}, \quad \omega = \frac{3}{2}\left(\frac{2u\,dv - 3v\,du}{\Delta}\right) \in \Omega^1(X). \tag{2-4}$$

They have weights 0 and -2, respectively. By Lemma 2.3, we have

$$\rho^*\omega = \frac{dq}{q}.$$

Both ω and ψ have logarithmic singularities along the discriminant locus D.

Lemma 2.5. *If $h \in \mathrm{gr}_m \mathcal{O}(X)$, then $\Delta\, \psi \wedge \omega = \frac{3}{4}\, du \wedge dv$ and*

$$\frac{dh}{h} \wedge \omega = m\psi \wedge \omega, \quad d(h\omega) = (m-2)h\psi \wedge \omega. \tag{2-5}$$

Proof. Since $d\log(h_1 h_2) = d\log(h_1) + d\log(h_2)$, it suffices to verify the first equation for $h = u$ and $h = \Delta$. For the second equation, use $d\omega = -2\psi \wedge \omega$ to write

$$d(h\omega) = h\frac{dh}{h} \wedge \omega - 2h\psi \wedge \omega$$

and conclude using the first equation. □

In particular, since ψ and ω are pointwise linearly independent, we have

$$\Omega^1(X) = \mathcal{O}(X)du \oplus \mathcal{O}(X)dv \cong \mathcal{O}(X)\psi \oplus \mathcal{O}(X)\omega.$$

Corollary 2.6. *For every $f \in \mathrm{gr}_m \mathcal{O}(X)$ we have*

$$df = \vartheta(f)\omega + mf\psi,$$

where $\vartheta : \mathcal{O}(X) \to \mathcal{O}(X)$ is the derivation of weight two defined by

$$\vartheta = 6v\frac{\partial}{\partial u} + \frac{u^2}{3}\frac{\partial}{\partial v}. \tag{2-6}$$

Proof. There exist unique $f_0, f_1 \in \mathcal{O}(X)$ such that

$$df = f_0\omega + f_1\psi.$$

Use (2-5) to deduce that $df \wedge \omega = mf\psi \wedge \omega$, which yields $f_1 = mf$. The linear map $\vartheta : f \mapsto f_0$ is a derivation which necessarily satisfies $\vartheta(\Delta) = 0$. This is certainly true of the formula (2-6). It therefore suffices to verify that $\vartheta(u) = 6v$. For this, use the fact that $d\Delta = 3u^2\, du - 54v\, dv$ to deduce that $du \wedge \psi = \frac{-9v}{2}(du \wedge dv)/\Delta$. Comparing with $\omega \wedge \psi = -\frac{3}{4}(du \wedge dv)/\Delta$ implies that $\vartheta(u) = 6v$ as required. □

Remark 2.7. The derivation ϑ is closely related to the Serre derivative [Zagier 2008, (53)].

Consider the pull-back along (2-2) followed by the natural map

$$\Omega^1_X \xrightarrow{\mu^*} \Omega^1_{\mathbb{G}_m \times X} \to \Omega^1_{(\mathbb{G}_m \times X)/X}$$

to relative Kähler differentials. Taking global sections gives a natural $\mathcal{O}(X)$-linear map

$$\pi^* : \Omega^1(X) \to \mathcal{O}(X) \otimes_{\mathbb{k}} \Omega^1(\mathbb{G}_m) = \mathcal{O}(X)[\lambda^{\pm}]\, d\log\lambda,$$

where λ is the coordinate on $\mathbb{G}_m$.

Say that an element of $\Omega^1(X)$ is *proportional* to ω if it lies in the subspace $\mathcal{O}(X)\omega$ of $\Omega^1(X)$ spanned by ω.

Lemma 2.8. *A form $\eta \in \Omega^1(X)$ is proportional to ω if and only if $\pi^*\eta = 0$. Furthermore, $\pi^*\psi = d\log\lambda$.*

Proof. Via the natural isomorphism of $\mathcal{O}(X)[\lambda^{\pm}]$-modules,

$$\big(\Omega^1(\mathbb{G}_m)\otimes_{\Bbbk}\mathcal{O}(X)\big)\oplus\big(\Omega^1(X)\otimes_{\Bbbk}\Bbbk[\lambda^{\pm}]\big)\to\Omega^1(\mathbb{G}_m\times X),$$

we can compute

$$\mu^*(\omega)=(0,\lambda^{-2}\omega)\quad\text{and}\quad\mu^*(\psi)=(d\log\lambda,\psi).\tag{2-7}$$

This follows from calculating

$$\mu^*(2u\,dv-3v\,du)=2\lambda^4u\,d(\lambda^6v)-3\lambda^6v\,d(\lambda^4u)=\lambda^{10}\omega$$

and noting that the terms involving $d\lambda$ cancel. More generally, we verify that for any $h\in\operatorname{gr}_m\mathcal{O}(X)$, we have

$$\mu^*\,dh=d(\lambda^m h)=\lambda^m\,dh+m\lambda^{m-1}h\,d\lambda$$

and therefore $\mu^*\,d\log h=(m\,d\log\lambda,d\log h)$. Setting $h=\Delta$ proves (2-7), from which the lemma immediately follows. □

2E. ***Alternative description of $\mathcal{M}_{1,1/\mathbb{Q}}$.*** A complementary approach to constructing $\mathcal{M}_{1,1}$ as a stack over $\Bbbk$ is as a quotient of the affine subscheme Z of X defined by $\Delta=1$. Its affine ring $\mathcal{O}(Z)$ is $\Bbbk[u,v]/I$, where I is the graded ideal of $\Bbbk[u,v]$ generated by $\Delta-1$, where $\Bbbk$ is any commutative ring with $6\in\Bbbk^\times$. Since Δ has weight 12, the affine group scheme

$$\boldsymbol{\mu}_{12}=\operatorname{Spec}\Bbbk[\lambda]/(\lambda^{12}-1),$$

with λ group-like, acts on Z.

Remark 2.9. The affine scheme Z is isomorphic, over $\Bbbk$, to the Fermat cubic minus its identity element. In fact, the closure of the locus $u^3-27v^2=1/4$ is isomorphic to the Fermat cubic $x^3+y^3=1$, where $u=1/(x+y)$ and $v=(x-y)/6(x+y)$.

The inclusion $\boldsymbol{\mu}_{12}\hookrightarrow\mathbb{G}_m$ induces an isomorphism of (the constant group scheme) $\mathbb{Z}/12\mathbb{Z}$ with the character group of $\boldsymbol{\mu}_{12}$. Denote the congruence class of n mod 12 by $[n]$. The $\boldsymbol{\mu}_{12}$ action on Z gives a $\mathbb{Z}/12\mathbb{Z}$-grading of its coordinate ring:

$$\mathcal{O}(Z)=\bigoplus_{n\text{ mod }12}\mathcal{O}(Z)_{[n]}.$$

If n is odd then $\mathcal{O}(Z)_{[n]}=0$. Since the inclusion $j:Z\hookrightarrow X$ is $\boldsymbol{\mu}_{12}$ equivariant, the restriction homomorphism induces a ring homomorphism

$$j^*:\mathcal{O}(X)^{\mathbb{G}_m}\to\mathcal{O}(Z)^{\boldsymbol{\mu}_{12}}\tag{2-8}$$

and, for each $m\in\mathbb{Z}$, a homomorphism

$$j^*:\operatorname{gr}_m\mathcal{O}(X)\to\mathcal{O}(Z)_{[m]}\tag{2-9}$$

of modules over (2-8).

Lemma 2.10. *The homomorphism* (2-8) *is an isomorphism, so that*

$$j : \boldsymbol{\mu}_{12}\backslash\backslash Z \xrightarrow{\sim} \mathbb{G}_m\backslash\backslash X$$

is an isomorphism of stacks. Moreover, for each $m \in \mathbb{Z}$, (2-9) *is an isomorphism.*

Proof. Since $\Delta - 1$ is $\boldsymbol{\mu}_{12}$-invariant, it suffices for the first part to show that

$$\mathcal{O}(X)^{\mathbb{G}_m} = \Bbbk[u, v, \Delta^{-1}]^{\mathbb{G}_m} \longrightarrow \Bbbk[\bar{u}, \bar{v}]^{\boldsymbol{\mu}_{12}}/I^{\boldsymbol{\mu}_{12}} = \mathcal{O}(Z)^{\boldsymbol{\mu}_{12}}$$

is an isomorphism, where $\bar{u} = j^*(u)$ and $\bar{v} = j^*(v)$ are the images of u, v. Since $\bar{u}$ has weight 4 and $\bar{v}$ has weight 6, it follows that $\Bbbk[\bar{u}, \bar{v}]^{\boldsymbol{\mu}_{12}} = \Bbbk[\bar{u}^3, \bar{v}^2]$. The inverse is induced by the map

$$\Bbbk[\bar{u}^3, \bar{v}^2] \to \Bbbk[u, v, \Delta^{-1}]^{\mathbb{G}_m}$$

which sends $\bar{u}^3$ to $u^3\Delta^{-1}$ and $\bar{v}^2$ to $v^2\Delta^{-1}$. It vanishes on $I^{\boldsymbol{\mu}_{12}}$. This proves that (2-8) is an isomorphism.

For all integers k, multiplication by Δ^k gives an isomorphism

$$\mathrm{gr}_m\mathcal{O}(X) \xrightarrow{\sim} \mathrm{gr}_{m+12k}\mathcal{O}(X)$$

of $\mathrm{gr}_0\mathcal{O}(X) = \mathcal{O}(X)^{\mathbb{G}_m}$-modules. It therefore suffices to prove that (2-9) is an isomorphism for $m = 0, 4, 6, 8, 10, 14$, which form a complete set of representatives for even numbers modulo 12. But $\mathcal{O}(Z)$ is isomorphic to the free $\mathbb{Z}/12\mathbb{Z}$-graded $\mathcal{O}(Z)^{\boldsymbol{\mu}_{12}} \cong \Bbbk[\bar{u}^3, \bar{v}^2]$-module generated by monomials in $\bar{u}$ and $\bar{v}$ that are of degree < 3 in $\bar{u}$ and degree < 2 in $\bar{v}$, where $\bar{u}^3 - 27\bar{v}^2 = 1$. These are

$$1, \bar{u}, \bar{v}, \bar{u}^2, \bar{u}\bar{v}, \bar{u}^2\bar{v},$$

and have weights 0, 4, 6, 8, 10, 14, respectively. Similarly, $\mathrm{gr}_m\mathcal{O}(X)$ is isomorphic to the free graded $\mathcal{O}(X)^{\mathbb{G}_m} \cong \Bbbk[u^3\Delta^{-1}, v^2\Delta^{-1}]$-module generated by the monomials $1, u, v, u^2, uv, u^2v$. $\square$

2F. *Gauss–Manin connection* [Katz 1973, §A1]**.** The vector bundle $\mathscr{V}$ over X is defined to be the restriction of the trivial rank 2 vector bundle

$$\overline{\mathscr{V}} := \mathcal{O}_Y\mathsf{S} \oplus \mathcal{O}_Y\mathsf{T}$$

on $Y := \mathbb{A}^2 - \{0\}$ to X. The multiplicative group acts on it by

$$\lambda \cdot \mathsf{S} = \lambda\mathsf{S}, \quad \lambda \cdot \mathsf{T} = \lambda^{-1}\mathsf{T}.$$

So $\overline{\mathscr{V}}$ can be regarded as a vector bundle on the moduli stack $\overline{\mathcal{M}}_{1,1}$ and $\mathscr{V}$ as a vector bundle on $\mathcal{M}_{1,1}$. Set $\mathscr{V}_n = \mathrm{Sym}^n\,\mathscr{V}$ for all $n \geq 1$.

The connection on $\overline{\mathscr{V}}$, and its symmetric powers

$$\overline{\mathscr{V}}_n := \mathrm{Sym}^n\,\overline{\mathscr{V}} = \bigoplus_{j+k=n} \mathcal{O}_Y\mathsf{S}^j\mathsf{T}^k$$

is defined by

$$\nabla = d + \begin{pmatrix} \mathsf{S} & \mathsf{T} \end{pmatrix} \begin{pmatrix} \psi & \omega \\ -\frac{u}{12}\omega & -\psi \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \mathsf{S}} \\ \frac{\partial}{\partial \mathsf{T}} \end{pmatrix}. \tag{2-10}$$

It is $\mathbb{G}_m$-invariant, and thus defines a rational connection on $\overline{\mathscr{V}}_n$ with regular singularities and nilpotent residue along the discriminant divisor D.

Set $\mathbb{V} = R^1\pi_*\Bbbk_X$ where $\pi : \mathcal{E} \to X$ is the universal elliptic curve. It is proved in [Hain 2013, Proposition 19.6] that when $\Bbbk \subset \mathbb{C}$ there is a natural isomorphism

$$\mathbb{V}^{\mathrm{B}} \otimes_{\Bbbk} \mathcal{O}_X^{\mathrm{an}} \cong \mathscr{V}^{\mathrm{an}} \otimes_{\Bbbk} \mathbb{C}$$

of bundles with connection over $X(\mathbb{C})$, where the left hand bundle is endowed with the Gauss–Manin connection. Under this isomorphism T corresponds to the section dx/y of $(R^1\pi_*\mathbb{C}_X) \otimes_{\mathbb{C}} \mathcal{O}_X$ and S to the section $x\,dx/y$. For later use, we set

$$\mathbb{V}_n^{\mathrm{B}} = \operatorname{Sym}^n \mathbb{V}^{\mathrm{B}}. \tag{2-11}$$

When pulled back to the upper half plane (and q-disk), the connection can also be written down in terms of the frame A and T, where A is the section corresponding to the Poincaré dual of the element $\boldsymbol{a}$ of $H_1(\mathbb{Z}/(\mathbb{Z}+z\mathbb{Z}))$ corresponding to the curve from 0 to 1. The two framings are related by[1]

$$\mathsf{S} = \mathsf{A} + 2\,\mathsf{G}_2(q)\mathsf{T}. \tag{2-12}$$

In this frame, the Gauss–Manin connection is given by

$$\nabla = d + \mathsf{A}\frac{\partial}{\partial \mathsf{T}} \otimes \frac{dq}{q} = 2\pi i\Big(\mathcal{D} + \mathsf{A}\frac{\partial}{\partial \mathsf{T}}\Big) \otimes dz, \tag{2-13}$$

where, as above, $\mathcal{D}$ is the differential operator $q\partial/\partial q = (2\pi i)^{-1}\partial/\partial z$.

Lemma 2.11. *If $f = \sum_{j+k=n} f^{j,k}\mathsf{A}^j\mathsf{T}^k$ is a real analytic section of $\mathscr{V}_n$, then*

$$\nabla f = 2\pi i \sum_{j+k=n} \big(\mathcal{D} f^{j,k} + (k+1) f^{j-1,k+1}\big)\, dz \otimes \mathsf{A}^j\mathsf{T}^k,$$

where we define $f^{n+1,-1} = f^{-1,n+1} = 0$.

Proof. Apply the connection (2-13) to the section f. □

Pulling back via the map π^* defines a relative connection on $\mathbb{G}_m$ over $\mathcal{O}_X$, which by (2-7) is of the form

$$\pi^*(\nabla) = d + \begin{pmatrix} \mathsf{S} & \mathsf{T} \end{pmatrix} \begin{pmatrix} \frac{d\lambda}{\lambda} & 0 \\ 0 & -\frac{d\lambda}{\lambda} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \mathsf{S}} \\ \frac{\partial}{\partial \mathsf{T}} \end{pmatrix} \tag{2-14}$$

A section of this relative connection is flat if and only if it is $\mathbb{G}_m$-equivariant.

[1] This follows from the formulas in [Hain 2013, §19.3]: T and S are obtained from $\hat{T}$ and $\hat{S}$ there by setting $\xi = (2\pi i)^{-1}$ and taking A to be $\boldsymbol{a}$.

3. De Rham cohomology

In this section, we take $\Bbbk = \mathbb{Q}$. Since X is affine, it follows from the version of Grothendieck's algebraic de Rham theorem with coefficients in a connection [Deligne 1970, Corollary 6.3] that for each $n \geq 0$, $H^1_{\mathrm{dR}}(X, \mathscr{V}_n)$ is computed by the complex

$$\Omega^\bullet(X, \mathscr{V}_n) := \bigl[\Gamma(X, \mathscr{V}_n) \xrightarrow{\nabla} \Gamma(X, \Omega^1_X \otimes_{\mathcal{O}_X} \mathscr{V}_n) \xrightarrow{\nabla} \Gamma(X, \Omega^2_X \otimes_{\mathcal{O}_X} \mathscr{V}_n)\bigr].$$

Since ∇ is equivariant with respect to the action of $\mathbb{G}_m$, we obtain the subcomplex

$$\Omega^\bullet(X, \mathscr{V}_n)^{\mathbb{G}_m} = \bigl[\Gamma(X, \mathscr{V}_n)^{\mathbb{G}_m} \xrightarrow{\nabla} \Gamma(X, \Omega^1_X \otimes_{\mathcal{O}_X} \mathscr{V}_n)^{\mathbb{G}_m} \xrightarrow{\nabla} \Gamma(X, \Omega^2_X \otimes_{\mathcal{O}_X} \mathscr{V}_n)^{\mathbb{G}_m}\bigr] \tag{3-1}$$

of invariant forms. Since $\mathbb{G}_m$ is connected, this also computes the de Rham cohomology of X with coefficients in $\mathscr{V}_n$. The Leray spectral sequence for the $\mathbb{G}_m$-bundle $p : X \to \mathcal{M}_{1,1}$

$$H^j(\mathcal{M}_{1,1}, R^k p_* \mathscr{V}_n) \to H^{j+k}(X, \mathscr{V}_n)$$

has only two nonvanishing rows, which implies that there is an exact sequence

$$0 \to H^j_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n) \xrightarrow{p^*} H^j_{\mathrm{dR}}(X, \mathscr{V}_n) \to H^{j-1}_{\mathrm{dR}}(\mathcal{M}_{1,1}, R^1 p_* \mathscr{V}_n) \to 0 \tag{3-2}$$

for all $j \geq 0$, where we recall that $\mathscr{V}_n$ denotes both the bundle $p_* \mathscr{V}_n$ on $\mathcal{M}_{1,1}$ and $\mathscr{V}_n$ on X. Since $\mathscr{V}_n$ is trivial on the $\mathbb{G}_m$ orbits on X,

$$H^{j-1}_{\mathrm{dR}}(\mathcal{M}_{1,1}, R^1 p_* \mathscr{V}_n) \cong H^1_{\mathrm{dR}}(\mathbb{G}_m; \mathbb{Q}) \otimes_{\mathbb{Q}} H^{j-1}_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n).$$

If $n > 0$, then $H^j(\mathcal{M}_{1,1}, \mathscr{V}_n)$ vanishes when $j \neq 1$. We deduce natural isomorphisms

$$p^* : H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n) \xrightarrow{\sim} H^1_{\mathrm{dR}}(X, \mathscr{V}_n)$$

for all $n > 0$ and

$$\pi^* : H^1_{\mathrm{dR}}(X; \mathscr{V}_0) \xrightarrow{\sim} H^1_{\mathrm{dR}}(\mathbb{G}_m; \mathbb{Q}).$$

By computations in Section 2D, the left-hand side is generated by $[\psi]$.

Remark 3.1. Since $\mathcal{M}_{1,1} = \boldsymbol{\mu}_{12} \backslash\backslash Z$, the homology of the complex $\Omega^\bullet(Z, j^* \mathscr{V}_n)^{\boldsymbol{\mu}_{12}}$ is $H^\bullet_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n)$. Below (see Proposition 3.8) we show that there is a canonical isomorphism

$$\Omega^\bullet(X, \mathscr{V}_n)^{\mathbb{G}_m} \cong \Omega^\bullet(Z, j^* \mathscr{V}_n)^{\boldsymbol{\mu}_{12}} \oplus \Omega^\bullet(Z, j^* \mathscr{V}_n)^{\boldsymbol{\mu}_{12}}[-1],$$

where the shift is given by multiplication by ψ. This gives a natural splitting of (3-2), where the shift $[-1]$ is given by cup product with $[\psi]$:

$$H^\bullet_{\mathrm{dR}}(X, \mathscr{V}_n) \cong H^\bullet_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n) \oplus H^\bullet_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n)[-1]. \tag{3-3}$$

3A. ***The subcomplex*** $\Omega^\bullet(X,\mathscr{V}_n)^\omega$**.** Recall that we can identify $\mathrm{gr}_n\mathcal{O}(X)$ with $M_n^!$ via the isomorphism

$$\rho^*:\mathrm{gr}_n\mathcal{O}(X)=\mathrm{gr}_n\mathbb{Q}[u,v][\Delta^{-1}]\xrightarrow{\sim} M_n^!$$

since every holomorphic modular form can be uniquely written as a polynomial in the Eisenstein series u and v.

Definition 3.2. For each $n\geq 0$, define

$$\varpi: M_{n+2}^!\to\Gamma(X,\Omega_X^1\otimes_{\mathcal{O}_X}\mathscr{V}_n)^{\mathbb{G}_m} \tag{3-4}$$

to be the $\mathbb{Q}$-linear map that takes f to $\omega_f := f\omega\,\mathsf{T}^n$, where ω and T were defined in (2-4) and Section 2F.

The first task in proving Theorem 1.2 is to show that ϖ induces a linear map

$$\varpi: M_{n+2}^!/D^{n+1}M_{-n}^!\to H^1_{\mathrm{dR}}(\mathcal{M}_{1,1};\mathscr{V}_n). \tag{3-5}$$

Here we take the first steps in this direction. The following lemma implies that ω_f is ∇-closed.

Lemma 3.3. *For all $j,k\geq 0$ and $f\in M_{k-j+2}^!$, the one-form*

$$\eta=f\omega\,\mathsf{S}^j\mathsf{T}^k\in\Gamma(X,\omega\otimes\mathscr{V}_{j+k})$$

is $\mathbb{G}_m$-invariant and satisfies $\nabla\eta=0$. Consequently,

$$\Gamma(X,\omega\otimes\mathscr{V}_{j+k})^{\mathbb{G}_m}\subset\ker\nabla. \tag{3-6}$$

Proof. We have $f\in\mathrm{gr}_{k-j+2}\mathcal{O}(X)$. By the Leibniz rule

$$\nabla\eta=d(f\omega)\mathsf{S}^j\mathsf{T}^k-f\omega\wedge\nabla\mathsf{S}^j\mathsf{T}^k. \tag{3-7}$$

From the connection formula (2-10), we have

$$\nabla\mathsf{S}^j\mathsf{T}^k\equiv(j-k)\,\psi\,\mathsf{S}^j\mathsf{T}^k\pmod{\omega\mathcal{O}(X)}$$

and therefore

$$f\omega\wedge\nabla\mathsf{S}^j\mathsf{T}^k=(k-j)f\psi\wedge\omega\mathsf{S}^j\mathsf{T}^k.$$

By the second equation of (2-5), we find that

$$d(f\omega)=(k-j+2-2)f\,\psi\wedge\omega=(k-j)f\,\psi\wedge\omega.$$

Substituting the two previous expressions into (3-7) implies that $\nabla\eta=0$. □

Lemma 3.4. *The image of $\nabla:\Gamma(X,\mathscr{V}_n)^{\mathbb{G}_m}\to\Gamma(X,\Omega_X^1\otimes\mathscr{V}_n)^{\mathbb{G}_m}$ lies in the subspace $\Gamma(X,\omega\otimes\mathscr{V}_n)^{\mathbb{G}_m}$ of forms proportional to ω.*

Proof. It suffices, by Lemma 2.8, to show that if

$$f=\sum_{j+k=n}f^{j,k}\mathsf{S}^j\mathsf{T}^k$$

is $\mathbb{G}_m$-equivariant, then $\pi^*(\nabla f) = 0$. It follows from (2-14) that

$$\pi^*(\nabla f) = \sum_{j+k=n} \Big(\pi^* df^{j,k} + (j-k)\lambda^{k-j} f^{j,k} \frac{d\lambda}{\lambda}\Big) \mathsf{S}^j \mathsf{T}^k.$$

Each term in brackets vanishes, since it expresses the fact that $\lambda \cdot f^{j,k} = \lambda^{k-j} f^{j,k}$, i.e., $f^{j,k} \in \mathrm{gr}_{k-j}\mathcal{O}$, which is equivalent to the $\mathbb{G}_m$-equivariance of f since S has weight $+1$ and T has weight -1. □

This lemma implies that ∇ acts on $\Gamma(X, \mathscr{V}_n)^{\mathbb{G}_m}$ via the connection

$$\nabla^\omega = \omega\vartheta + \begin{pmatrix} \mathsf{S} & \mathsf{T} \end{pmatrix} \begin{pmatrix} 0 & \omega \\ -\frac{u}{12}\omega & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \mathsf{S}} \\ \frac{\partial}{\partial \mathsf{T}} \end{pmatrix}, \tag{3-8}$$

where ϑ is the operator (2-6).

The following is an immediate consequence of Lemmas 3.3 and 3.4.

Corollary 3.5. *For all $n \geq 0$,*

$$\Omega^\bullet(X, \mathscr{V}_n)^\omega := \big[\Gamma(X, \mathscr{V}_n)^{\mathbb{G}_m} \xrightarrow{\nabla^\omega} \Gamma(X, \omega \otimes \mathscr{V}_n)^{\mathbb{G}_m} \longrightarrow 0\big] \tag{3-9}$$

is a subcomplex of $\Omega^\bullet(X, \mathscr{V}_n)^{\mathbb{G}_m}$.

The kernel of the restriction mapping

$$j^* : \Omega^\bullet(X, \mathscr{V}_n)^{\mathbb{G}_m} \to \Omega^\bullet(Z, j^*\mathscr{V}_n)^{\boldsymbol{\mu}_{12}}$$

contains the ideal generated by ψ. It therefore induces a homomorphism

$$j^* : \Omega^\bullet(X, \mathscr{V}_n)^\omega \to \Omega^\bullet(Z, j^*\mathscr{V}_n)^{\boldsymbol{\mu}_{12}}. \tag{3-10}$$

Lemma 3.6. *For all $n \geq 0$, the restriction map* (3-10) *is an isomorphism.*

Proof. Both complexes have length 2. That j^* is an isomorphism in degree 0 follows directly from Lemma 2.10. To prove the assertion in degree 1, note that elements of $\Gamma(X, \omega \otimes \mathscr{V}_n)^{\mathbb{G}_m}$ are of the form

$$\sum_{k+l=n} f^{k,l} \omega \mathsf{S}^k \mathsf{T}^l \quad \text{where } f^{k,l} \in \mathrm{gr}_{l-k+2}\mathcal{O}(X)$$

and elements of $\Gamma(Z, j^*\mathscr{V}_n)^{\boldsymbol{\mu}_{12}}$ are of the form

$$\sum_{k+l=n} g^{k,l} j^*(\omega) \mathsf{S}^k \mathsf{T}^l \quad \text{where } g^{k,l} \in \mathrm{gr}_{[l-k+2]}\mathcal{O}(Z),$$

since $j^*(\omega)$ generates $\Omega^1(Z)$, which follows from the fact that $\omega \wedge \psi \neq 0$ and therefore $j^*(\omega) \neq 0$. Injectivity follows from Lemma 2.10: if $j^*(f^{k,l})$ vanishes in $\mathrm{gr}_{[l-k+2]}\mathcal{O}(Z)$, it follows that $f^{k,l} = 0$. For the surjectivity, observe that

$$j^*\Big(\sum_{k+l=n} G^{k,l} \omega \mathsf{S}^k \mathsf{T}^l\Big) = \sum_{k+l=n} g^{k,l} j^*(\omega) \mathsf{S}^k \mathsf{T}^l,$$

where $G^{k,l} \in \mathrm{gr}_{l-k+2}\mathcal{O}(X)$ is the unique preimage of $g^{k,l} \in \mathrm{gr}_{[l-k+2]}\mathcal{O}(Z)$. □

Since the map $Z \to \boldsymbol{\mu}_{12}\backslash\backslash Z = \mathcal{M}_{1,1}$ is an étale map of stacks over $\mathbb{Q}$, the complex

$$\Gamma(Z, \Omega^\bullet_Z \otimes j^*\mathcal{V}_n)^{\boldsymbol{\mu}_{12}} \tag{3-11}$$

computes $H^\bullet_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n)$ for all $n \geq 0$. Thus we have:

Corollary 3.7. *The complex $\big(\Omega^\bullet(X, \mathcal{V}_n)^\omega, \nabla^\omega\big)$ computes $H^\bullet_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n)$ for all $n \geq 0$.*

We conclude this section by showing that the isomorphism (3-3) lifts to the level of de Rham complexes.

Proposition 3.8. *There is a canonical isomorphism*

$$\Omega^\bullet(X, \mathcal{V}_n)^{\mathbb{G}_m} \cong \Omega^\bullet(Z, j^*\mathcal{V}_n)^{\boldsymbol{\mu}_{12}} \oplus \Omega^\bullet(Z, j^*\mathcal{V}_n)^{\boldsymbol{\mu}_{12}}[-1]$$

of complexes.

Proof. The quotient of $\Omega^\bullet(X, \mathcal{V}_n)^{\mathbb{G}_m}$ by $\Omega^\bullet(X, \mathcal{V}_n)^\omega$ is $\psi \otimes \Omega^\bullet(X, \mathcal{V}_n)^\omega$:

$$\begin{tikzcd}
0 \arrow[r] & \Gamma(X, \mathcal{V}_n)^{\mathbb{G}_m} \arrow[r, "\nabla^\omega"] \arrow[d, equal] & \Gamma(X, \omega \otimes \mathcal{V}_n)^{\mathbb{G}_m} \arrow[rr] \arrow[d] & & 0 \arrow[d] & \\
0 \arrow[r] & \Gamma(X, \mathcal{V}_n)^{\mathbb{G}_m} \arrow[r, "\nabla"] \arrow[d] & \Gamma(X, \Omega^1_X \otimes \mathcal{V}_n)^{\mathbb{G}_m} \arrow[r, "\nabla"] \arrow[d] & \Gamma(X, \Omega^2_X \otimes \mathcal{V}_n)^{\mathbb{G}_m} \arrow[rr] \arrow[d, equal] & & 0 \\
0 \arrow[r] & 0 \arrow[r] & \psi \otimes \Gamma(X, \mathcal{V}_n)^{\mathbb{G}_m} \arrow[r, "\nabla^\omega"] & \psi \otimes \Gamma(X, \omega \otimes \mathcal{V}_n)^{\mathbb{G}_m} \arrow[rr] & & 0
\end{tikzcd}$$

This exact sequence of complexes is naturally split since the third row defines a subcomplex of the second. The result follows from Lemma 3.6 since $\psi \otimes \Omega^\bullet(X, \mathcal{V}_n)^\omega \cong \Omega^\bullet(X, \mathcal{V}_n)^\omega[-1]$. $\square$

4. Proof of Theorem 1.2

In this section, we work over $\Bbbk = \mathbb{Q}$.

4A. *Heads and tails.* We show that the "head" of an element f of $\Gamma(X, \omega \otimes \mathcal{V}_n)^{\mathbb{G}_m}$ is related to the "tail" of ∇f by the Bol operator.

Lemma 4.1. *For all $n \geq 0$ there exists a unique $\mathbb{Q}$-linear map*

$$\phi : M^!_{-n} \to \Gamma(X, \mathcal{V}_n)^{\mathbb{G}_m}$$

such that if we write

$$\phi = \sum_{j+k=n} \phi^{j,k} \mathsf{S}^j \mathsf{T}^k,$$

where $\phi^{j,k} \in \mathrm{Hom}(M^!_{-n}, \mathrm{gr}_{k-j}\mathcal{O}(X))$, then we have

$$\phi^{n,0}(f) = f \quad \text{and} \quad \nabla\phi(f) \in \omega\mathcal{O}(X)\mathsf{T}^n. \tag{4-1}$$

In other words, given a weakly holomorphic modular form f of weight $-n$, there is a unique section of $\mathcal{V}_n$ which coincides with f in the first component, and whose image under ∇ vanishes in all components save the last.

Proof. Suppose that $f \in M^!_{-n}$. We shall construct the components $f^{j,k} := \phi^{j,k}(f)$ inductively in k. For $k = 0$, we have $f^{n,0} = f$. The connection acts via (3-8) which we write

$$\nabla^\omega = \omega\Big(\vartheta + \mathsf{S}\frac{\partial}{\partial \mathsf{T}} - \frac{u}{12}\,\mathsf{T}\frac{\partial}{\partial \mathsf{S}}\Big).$$

Suppose that $f^{a,b}$ is defined for $b \le k < n$. The coefficient of $\mathsf{S}^j\mathsf{T}^k$ in $\nabla^\omega\phi(f)$ is

$$\omega\Big(\vartheta(f^{j,k}) + (k+1)\, f^{j-1,k+1} - (j+1)\frac{u}{12} f^{j+1,k-1}\Big).$$

There is a unique $f^{j-1,k+1} \in \mathrm{gr}_{k-j+2}\mathcal{O}(X)$ that makes this vanish; namely

$$f^{j-1,k+1} = \frac{1}{k+1}\Big(\frac{j+1}{12} u f^{j+1,k-1} - \vartheta(f^{j,k})\Big).$$

By induction, these equations determine $\phi(f)$ uniquely. □

Note that the inductive definition of ϕ involves dividing by $k+1$ for $1 \le k < n$.

Lemma 4.2 (heads and tails). *The diagram*

$$\begin{tikzcd} M^!_{-n} \arrow[r, "\phi"] \arrow[d, "\mathcal{D}^{n+1}/n!"'] & \Gamma(X, \mathcal{V}_n)^{\mathbb{G}_m} \arrow[d, "\nabla"] \\ M^!_{n+2} \arrow[r, "\varpi"] & \Gamma(X, \omega \otimes \mathcal{V}_n)^{\mathbb{G}_m} \end{tikzcd}$$

commutes for all $n \ge 0$, where $\mathcal{D}^{n+1}$ is the Bol operator.

Proof. Let $f \in M^!_{-n}$, and let $\phi(f)$ be the unique section constructed in Lemma 4.1, whose coefficient of S^n is f. Perform the change of gauge (2-12). In this gauge, the coefficient of A^n in $\phi(f)$ is f:

$$\phi(f) = \sum_{j+k=n} F^{j,k}\mathsf{A}^j\mathsf{T}^k, \quad \text{where } F^{n,0} = f.$$

The defining property of ϕ is that $\nabla\phi(f)$ is a multiple of $\omega\mathsf{T}^n$. Let $r(f) \in \mathcal{O}(X)$ be the coefficient. That is,

$$\nabla\phi(f) = r(f)\omega\,\mathsf{T}^n.$$

This condition is preserved under the change of gauge (2-12), so that

$$\nabla\phi(f) = r(f)\omega\,\mathsf{T}^n = r(f)\,\frac{dq}{q}\mathsf{T}^n = 2\pi i r(f)\,dz\mathsf{T}^n.$$

By Lemma 2.11, we obtain the system of equations

$$F^{n,0} = f, \quad \mathcal{D}F^{j,k} + (k+1)F^{j-1,k+1} = 0, \quad \mathcal{D}F^{0,n} = r(f).$$

It follows that $r(f) = (-1)^n\mathcal{D}^{n+1} f/n!$. This proves the result since if f is nonzero n must be even and $(-1)^n = 1$. □

Remark 4.3. The previous two lemmas imply a relation between the Bol operator, multiplication by u, and the Serre derivative ϑ. Compare [Swinnerton-Dyer 1973, (25)].

We have therefore proved the existence of (3-5). It remains to prove that it is an isomorphism.

4B. *Proof of injectivity of* (3-5). Suppose that $\omega_g \in \Gamma(X, \omega \otimes \mathcal{V}_n)^{\mathbb{G}_m}$ is exact. Write $\omega_g = \nabla f$, where $f \in \Gamma(X, \mathcal{V}_n)^{\mathbb{G}_m}$. Consider the linear map

$$\Gamma(X, \mathcal{V}_n)^{\mathbb{G}_m} \to M^!_{-n}, \qquad \sum_{j+k=n} f^{j,k} \mathsf{S}^j \mathsf{T}^k \mapsto f^{n,0}.$$

Since ω_g has the property that all coefficients $\mathsf{S}^j \mathsf{T}^k$ except for $j = 0,\ k = n$ vanish, it follows from the uniqueness in Lemma 4.1 that $f = \phi(f^{n,0})$. By Lemma 4.2, it follows that ω_g is in the image of $D^{n+1} M^!_{-n}$.

4C. *Proof of surjectivity* (3-5). To complete the proof, we must show that the map (3-5) is surjective. By the algebraic de Rham Theorem (Section 3), every class in $H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathcal{V}_n)$ is represented by a section of the form

$$\eta = \sum_{j+k=n} f^{j,k} \omega\, \mathsf{S}^j \mathsf{T}^k,$$

where $f^{j,k} \in \mathrm{gr}_{k-j+2} \mathcal{O}(X)$. We show by induction that such a form is equivalent, modulo the image of ∇, to one in which $f^{j,k}$ vanishes for all $j > 0$. Suppose that $0 \le j \le n$ is largest such that $f^{j,k}$ is nonzero. If j is zero then there is nothing to prove, so assume $j > 0$. It follows from the connection formula (2-10) that

$$\nabla(g\, \mathsf{S}^{j-1} \mathsf{T}^{k+1}) \equiv (k+1) g\, \omega\, \mathsf{S}^j \mathsf{T}^k \mod \mathsf{T}^{k+1}$$

for any $g \in \mathcal{O}(X)$ of weight $k - j + 2$. Therefore, by replacing η by

$$\eta - \frac{1}{(k+1)} \nabla(f^{j,k}\, \mathsf{S}^{j-1} \mathsf{T}^{k+1}),$$

we can assume that $f^{j,k}$ vanishes, since the second term lies in $\Gamma(X, \omega \otimes \mathcal{V}_n)^{\mathbb{G}_m}$ by Lemma 3.4. Proceeding in this manner, we deduce that every cohomology class in $H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathcal{V}_n)$ is represented by a form in the image of (3-4).

Remark 4.4. The above argument works in characteristic 0 or $> \max\{3, n\}$.

5. Periods of Cusp Forms

Let $f \in S_n$ and $g \in S^!_n$. Write $f = \sum_{m>0} a_m q^m$ and $g = \sum_m b_m q^m$. In [Guerzhoy 2008, Theorem 1], a Hecke-equivariant pairing is defined, up to a sign, by

$$\{f, g\} = \sum_{m \in \mathbb{Z}} \frac{a_m b_{-m}}{m^{n-1}}. \tag{5-1}$$

In this section we show it extends to all $f \in S^!_n$ and that it corresponds to the image of $\omega_f \otimes \omega_g$ under the de Rham incarnation of the cup product

$$H^1_{\mathrm{cusp}}(\mathcal{M}_{1,1}, \mathbb{V}^{\mathrm{B}}_n) \otimes H^1_{\mathrm{cusp}}(\mathcal{M}_{1,1}, \mathbb{V}^{\mathrm{B}}_n) \to H^2_{\mathrm{cusp}}(\mathcal{M}_{1,1}, \mathbb{Q}) \xrightarrow{\sim} \mathbb{Q}$$

induced by the natural pairing $\mathbb{V}_n^{\mathrm{B}} \otimes \mathbb{V}_n^{\mathrm{B}} \to \mathbb{Q}$, where $\mathbb{V}_n^{\mathrm{B}}$ is the local system (2-11).

5A. *Inner products.* Let $V^{\mathrm{dR}} = \mathbb{Q}\mathsf{A} \oplus \mathbb{Q}\mathsf{T}$ be the fibre of the vector bundle $\overline{\mathcal{V}}$ over the cusp. At $q = 0$,

$$\mathsf{S} = \mathsf{A} + 2\mathsf{G}_2(0)\mathsf{T} = \mathsf{A} - \tfrac{1}{12}\mathsf{T}$$

so V^{dR} is also spanned by S and T. Define a skew symmetric inner product

$$\langle\ ,\ \rangle_{\mathrm{dR}} : V^{\mathrm{dR}} \otimes V^{\mathrm{dR}} \to \mathbb{Q}$$

by declaring that $\langle \mathsf{T}, \mathsf{A}\rangle = 1$. This is the natural inner product on V^{dR} and corresponds to the cup product pairing on the first de Rham cohomology group of an elliptic curve (see [Hain 2013, Proposition 19.1]). It extends to a $(-1)^n$ symmetric inner product on $V_n^{\mathrm{dR}} := \mathrm{Sym}^n\, V^{\mathrm{dR}}$ by

$$\langle v_1 v_2 \cdots v_n, w_1 w_2 \cdots w_n\rangle = \sum_{\sigma\in\Sigma_n} \prod_{j=1}^{n} \langle v_j, w_{\sigma(j)}\rangle \qquad v_j, w_k \in V^{\mathrm{dR}}.$$

In particular,

$$\langle \mathsf{A}^n, \mathsf{T}^n\rangle = \langle \mathsf{S}^n, \mathsf{T}^n\rangle = (-1)^n n! \tag{5-2}$$

This inner product induces an inner product $\langle\ ,\ \rangle_{\mathrm{dR}} : \overline{\mathcal{V}}_n \otimes \overline{\mathcal{V}}_n \to \mathcal{O}_X$ which is flat with respect to the connection ∇.

There is also a Betti version. As in the introduction, we set

$$V^{\mathrm{B}} = \mathbb{Q}\boldsymbol{a} \oplus \mathbb{Q}\boldsymbol{b}$$

and $V_n^{\mathrm{B}} = \mathrm{Sym}^n\, V_{\mathrm{B}}$. The unique skew symmetric inner product $\langle\ ,\ \rangle_{\mathrm{B}}$ on V_{B} satisfying $\langle \boldsymbol{a}, \boldsymbol{b}\rangle_{\mathrm{B}} = 1$ induces a $(-1)^n$ symmetric inner product

$$\langle\ ,\ \rangle : V_n^{\mathrm{B}} \otimes V_n^{\mathrm{B}} \to \mathbb{Q}$$

satisfying $\langle \boldsymbol{a}^n, \boldsymbol{b}^n\rangle = n!$ as above.

For each tangent vector $\vec{v} = e^{\lambda}\partial/\partial q$ of the origin of the q-disk there is a comparison isomorphism

$$\mathrm{comp}_{\mathrm{B,dR}} : V^{\mathrm{dR}} \otimes \mathbb{C} \to V^{\mathrm{B}} \otimes \mathbb{C},$$

which is defined by

$$\mathrm{comp}_{\mathrm{B,dR}}(\mathsf{A}) = \boldsymbol{a}, \quad \mathrm{comp}_{\mathrm{B,dR}}(\mathsf{T}) = -2\pi i \boldsymbol{b} + \lambda \boldsymbol{a}.$$

It corresponds to the limit mixed Hodge structure on the first cohomology of the first order smoothing of the nodal cubic in the direction of $\vec{v}$. Observe that, for all λ, two inner products are related by

$$\mathrm{comp}_{\mathrm{B,dR}}^{*}\langle\ ,\ \rangle_{\mathrm{dR}} = (2\pi i)^{-n}\langle\ ,\ \rangle_{\mathrm{B}} \tag{5-3}$$

via the comparison map $\mathrm{comp}_{\mathrm{B,dR}} : V_n^{\mathrm{dR}} \otimes \mathbb{C} \to V_n^{\mathrm{B}} \otimes \mathbb{C}$.

5B. ***Residue maps.*** The pullback of $(\mathcal{V}_n, \nabla)$ to a formal neighbourhood of the origin of the q-disk $\mathbb{D}^*$ along the map

$$\rho : \mathbb{D}^* \to X(\mathbb{C})$$

defined in Section 2C is the free $\mathbb{Q}[\![q]\!]$-module with basis $\{\mathsf{A}^n, \mathsf{A}^{n-1}\mathsf{T}, \dots, \mathsf{T}^n\}$ endowed with the connection (2-13). It can also be expressed in the frame $\{\mathsf{S}^j\mathsf{T}^k\}$ using the formal change of gauge (2-12).

The fraction field of $\mathbb{Q}[\![q]\!]$ is the ring $\mathbb{Q}((q)) := \mathbb{Q}[\![q]\!][q^{-1}]$ of Laurent series. Set

$$\Omega^1(\mathbb{D}^*, \mathcal{V}_n) := \mathbb{Q}((q))dq \otimes V_n^{\mathrm{dR}}.$$

Define the local residue map $\Omega^1(\mathbb{D}^*, \mathcal{V}_n) \to \mathbb{Q}$ to be the composite

$$\mathrm{Res} : \Omega^1(\mathbb{D}^*, \mathcal{V}_n) \to V_n^{\mathrm{dR}} \to V_n^{\mathrm{dR}}/\mathsf{A}V_{n-1}^{\mathrm{dR}} \xrightarrow{\sim} \mathbb{Q} \tag{5-4}$$

of the usual residue map with the map that sends A to 0 and T to 1. That is,

$$\mathrm{Res}\biggl(\sum_{j+k=n}\sum_{m\in\mathbb{Z}} a_m^{j,k} q^m \frac{dq}{q}\mathsf{A}^j\mathsf{T}^k\biggr) = a_0^{0,n}.$$

Observe that if $f \in \mathbb{Q}((q)) \otimes V_n$, then $\mathrm{Res}(\nabla f) = 0$, since by Equation (2-13), an exact section satisfies $\nabla f \equiv 0 \bmod \mathsf{A}$.

The residue map

$$\mathrm{Res} : \Gamma(X, \omega \otimes \mathcal{V}_n)^{\mathbb{G}_m} \to \mathbb{Q}$$

is defined to be the composite of the restriction map

$$\rho^* : \Gamma(X, \omega \otimes \mathcal{V}_n)^{\mathbb{G}_m} \to \Omega^1(\mathbb{D}^*, \mathcal{V}_n).$$

followed by the residue map (5-4).

5C. ***Cuspidal de Rham cohomology.*** Define the local cuspidal de Rham complex $\Omega^\bullet(\mathbb{D}^*, \mathcal{V}_n)$ by

$$\Omega^0_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n) := \biggl\{\sum_{j+k=n} a_m^{j,k} q^m \mathsf{A}^j\mathsf{T}^k : a_0^{n,0} = 0\biggr\}$$

$$\Omega^1_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n) := \ker\biggl(\mathrm{Res} : \mathbb{Q}((q))\frac{dq}{q} \otimes V_n^{\mathrm{dR}} \to \mathbb{Q}\biggr).$$

It is closed under the differential ∇. Since $H^0(\Omega^\bullet(\mathbb{D}^*, \mathcal{V}_n))$ is spanned by A^n, we see that $H^0(\Omega^\bullet_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n))$ vanishes. The following lemma implies that this complex is acyclic. It is a local version of Lemma 4.2. Its proof is similar.

Lemma 5.1 (local heads and tails). *For all $n \geq 0$, every element of $\Omega^1_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n)$ is exact in $\Omega^\bullet_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n)$. If*

$$f = \sum_{j+k=n} f^{j,k}\mathsf{A}^j\mathsf{T}^k \in \mathbb{Q}((q))V_n^{\mathrm{dR}} \quad \textit{and} \quad \nabla(f) = \sum_{k\neq 0} a_k q^k \frac{dq}{q}\mathsf{T}^n,$$

then

$$(-1)^n n!\left(q\frac{d}{dq}\right)^{n+1} f^{n,0} = \sum_{k\neq 0} a_k q^k.$$

Consequently, the head $f^{n,0}$ of f is

$$f^{n,0} = \frac{(-1)^n}{n!}\sum_{k\neq 0}\frac{a_k q^k}{k^{n+1}}. \qquad \square$$

The restriction mapping $\rho : \Omega^\bullet(X, \mathcal{V}_n)^\omega \to \Omega^\bullet(\mathbb{D}^*, \mathcal{V}_n)$ commutes with ∇. Consequently,

$$\Omega^\bullet_{\mathrm{cusp}}(X, \mathcal{V}_n)^\omega := \rho^{-1}\big(\Omega^\bullet_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n)\big)$$

is a subcomplex of $\Omega^\bullet(X, \mathcal{V}_n)^\omega$.

Lemma 5.2. *The complex $\Omega^\bullet_{\mathrm{cusp}}(X, \mathcal{V}_n)^\omega$ computes the cuspidal de Rham cohomology of $\mathcal{M}_{1,1}/\mathbb{Q}$. That is, the comparison isomorphism*

$$\mathrm{comp}_{\mathrm{B,dR}} : H^\bullet_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n)\otimes\mathbb{C} \xrightarrow{\sim} H^\bullet(\mathrm{SL}_2(\mathbb{Z}), V_n^{\mathrm{B}})\otimes\mathbb{C}$$

restricts to an isomorphism

$$\mathrm{comp}_{\mathrm{B,dR}} : H^\bullet_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n)\otimes\mathbb{C} \xrightarrow{\sim} H^\bullet_{\mathrm{cusp}}(\mathrm{SL}_2(\mathbb{Z}), V_n^{\mathrm{B}})\otimes\mathbb{C}.$$

Proof. This follows directly from the exactness of the sequence

$$0 \to \Omega^\bullet_{\mathrm{cusp}}(X, \mathcal{V}_n)^\omega \to \Omega^\bullet(X, \mathcal{V}_n)^\omega \to \left[\mathbb{Q}\mathsf{A}^n \xrightarrow{0} \mathbb{Q}\frac{dq}{q}\mathsf{T}^n\right] \to 0.$$

Alternatively, it can be deduced directly from Theorem 1.2. $\square$

5D. *The residue pairing.* By Lemma 5.1, the first homology of $\Omega^\bullet_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n)$ vanishes. Define the local residue pairing

$$\{\ ,\ \} : \Omega^1_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n)\otimes\Omega^1(\mathbb{D}^*, \mathcal{V}_n) \to \mathbb{Q}$$

as follows. Lemma 5.1 implies that each $\xi \in \Omega^1_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n)$ can be uniquely written $\xi = \nabla F$, where $F \in \Omega^0_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n)$. We set

$$\{\xi, \eta\} := \{\nabla F, \eta\} = \mathrm{Res}_{q=0}\langle F, \eta\rangle.$$

Lemma 5.3. *This pairing is well-defined. We have*

$$\left\{\sum_{j\neq 0} a_j q^j \frac{dq}{q}\mathsf{T}^n, \sum_{k\in\mathbb{Z}} b_k q^k \frac{dq}{q}\mathsf{T}^n\right\} = \sum_{k+l=0}\frac{a_k b_l}{k^{n+1}}.$$

It is $(-1)^{n+1}$ symmetric when restricted to $\Omega^1_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n)^{\otimes 2}$.

Proof. That it is well-defined follows from the uniqueness of F. To see that it is $(-1)^{n+1}$ symmetric on $\Omega^1_{\mathrm{cusp}}(\mathbb{D}^*, \mathscr{V}_n)$, write $\xi = \nabla F$ and $\eta = \nabla G$, where $F, G \in \Omega^0_{\mathrm{cusp}}(\mathbb{D}^*, \mathscr{V}_n)$. By flatness of the inner product,

$$d\langle F, G\rangle = \langle \nabla F, G\rangle - (-1)^n \langle \nabla G, F\rangle.$$

The symmetry property follows as the left-hand side has vanishing residue.

The formula is an immediate consequence of Lemma 5.1 as

$$\begin{aligned}\left\{\nabla\biggl(\sum_{j+k=n} f^{j,k}\mathsf{A}^j\mathsf{T}^k\biggr), g\,dq\,\mathsf{T}^n\right\} &= \sum_{j+k=n} \operatorname{Res}_{q=0} f^{j,k} g\langle \mathsf{A}^j\mathsf{T}^k, \mathsf{T}^n\rangle \\ &= (-1)^n n!\operatorname{Res}_{q=0} f^{n,0} g.\end{aligned}$$ □

By composing with the restriction maps

$$\Omega^1_{\mathrm{cusp}}(X, \mathscr{V}_n)^\omega \to \Omega^1_{\mathrm{cusp}}(\mathbb{D}^*, \mathscr{V}_n) \text{ and } \Omega^1(\mathcal{M}_{1,1}, \mathscr{V}_n) \to \Omega^1(\mathbb{D}^*, \mathscr{V}_n)$$

we obtain a well defined pairing

$$\{\ ,\ \} : \Omega^1_{\mathrm{cusp}}(X, \mathscr{V}_n)^\omega \otimes \Omega^1(\mathcal{M}_{1,1}, \mathscr{V}_n) \to \mathbb{Q}. \tag{5-5}$$

Lemma 5.4. *The pairing* (5-5) *has the property that* $\{\nabla f, \xi\} = 0$ *for all* $f \in \Omega^0(X, \mathscr{V}_n)^\omega$ *and* $\xi \in \Omega^1(X, \mathscr{V}_n)^\omega$. *Consequently, it induces a well-defined pairing*

$$\textstyle\int^{\mathrm{dR}} : H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n) \otimes H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n) \to \mathbb{Q}$$

such that the diagram

$$\begin{tikzcd} S^!_n \otimes M^!_n \arrow[r, "\{\ ,\ \}"] \arrow[d, "\varpi\otimes\varpi"'] & \mathbb{Q} \\ H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n) \otimes H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathscr{V}_n) \arrow[ur, "\int^{\mathrm{dR}}"'] & \end{tikzcd}$$

commutes.

Proof. The one-form $\langle f, \xi\rangle$ is an element of $\Gamma(X, \omega \otimes \mathscr{V}_0)^{\mathbb{G}_m}$, and therefore defines a class in the cohomology of the complex (3-9). We showed that the latter computes $H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}; \mathscr{V}_0) = 0$, and so $\langle f, \xi\rangle$ is exact. Its residue therefore vanishes.

Since the pairing (5-5) is $(-1)^{n+1}$ symmetric on $\Omega^1_{\mathrm{cusp}}(\mathcal{M}_{1,1}, \mathscr{V}_n)$, and since $\nabla f \in \Omega^1_{\mathrm{cusp}}(\mathcal{M}_{1,1}, \mathscr{V}_n)$ for all $f \in \Omega^0(X, \mathscr{V}_n)$, this implies that $\{\xi, \nabla f\} = 0$ for all $\xi \in \Omega^1_{\mathrm{cusp}}(\mathcal{M}_{1,1}, \mathscr{V}_n)$ and that the pairing is well-defined on cohomology.

The formula for the pairing is a direct consequence of Lemma 5.3. □

5E. ***Relation to cup product.*** Our final task is to determine how $\int^{\mathrm{dR}}$ is related to the cup product. For this we need to discuss relative cohomology and its relation to cuspidal cohomology. Let Γ_∞ be the subgroup of $\mathrm{SL}_2(\mathbb{Z})$ generated by $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$. Throughout this section, we assume that $n > 0$.

We have the exact sequence

$$0 \to H^0(\Gamma_\infty; V_n^{\mathrm{B}}) \to H^1(\mathrm{SL}_2(\mathbb{Z}), \Gamma_\infty; V_n^{\mathrm{B}}) \to H^1_{\mathrm{cusp}}(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}}) \to 0, \tag{5-6}$$

which is dual to the exact sequence

$$0 \to H^1_{\mathrm{cusp}}(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}}) \to H^1(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}}) \to H^1(\Gamma_\infty; V_n^{\mathrm{B}}) \to 0$$

under the cup product pairing

$$\int^{\mathrm{B}} : H^1(\mathrm{SL}_2(\mathbb{Z}), \Gamma_\infty; V_n^{\mathrm{B}}) \otimes H^1(\mathrm{SL}_2(\mathbb{Z}); V_n^{\mathrm{B}}) \to H^2(\mathrm{SL}_2(\mathbb{Z}), \Gamma_\infty; \mathbb{Q}) \cong \mathbb{Q}$$

induced by $\langle\ ,\ \rangle_{\mathrm{B}}$.

Remark 5.5. It is helpful to note that $H^0(\Gamma_\infty, V_n^{\mathrm{B}}) = \mathbb{Q}\boldsymbol{a}^n$ and $H^1(\Gamma_\infty, V_n^{\mathrm{B}}) = V_n^{\mathrm{B}}/\boldsymbol{a}V_{n-1}^{\mathrm{B}} \cong \mathbb{Q}\boldsymbol{b}^n$.

The algebraic analogue of the relative cohomology group above is the de Rham cohomology

$$H^\bullet_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathbb{D}^*; \mathcal{V}_n),$$

which is defined to be the cohomology of the complex

$$\Omega^\bullet(X, \mathbb{D}^*; \mathcal{V}_n)^\omega := \mathrm{cone}\big[\Omega^\bullet(X, \mathcal{V}_n)^\omega \to \Omega^\bullet(\mathbb{D}^*, \mathcal{V}_n)\big][-1].$$

There is a comparison isomorphism

$$\mathrm{comp}_{\mathrm{B,dR}} : H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathbb{D}^*; \mathcal{V}_n) \otimes \mathbb{C} \to H^1(\mathrm{SL}_2(\mathbb{Z}), \Gamma_\infty; V_n^{\mathrm{B}}) \otimes \mathbb{C}.$$

There is a short exact sequence

$$0 \to \mathbb{Q}\mathsf{A}^n \to H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathbb{D}^*; \mathcal{V}_n) \to H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n) \to 0 \tag{5-7}$$

which maps to (5-6) after tensoring both sequences with $\mathbb{C}$.

In order to use the cup product to construct a pairing between $H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n)$ and $H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n)$, we need to choose a splitting of (5-7). Here we use the splitting

$$s : H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n) \to H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathbb{D}^*; \mathcal{V}_n)$$

that is defined by taking the class of $\xi \in \Omega^1_{\mathrm{cusp}}(X, \mathcal{V}_n)^\omega$ to the class of (ξ, F), where F is the unique element of $\Omega^0_{\mathrm{cusp}}(\mathbb{D}^*, \mathcal{V}_n)$ whose derivative is the restriction of ξ to $\mathbb{D}^*$.

Proposition 5.6. *The diagram*

$$\begin{tikzcd}
H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n) \otimes H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathcal{V}_n) \otimes \mathbb{C} \arrow[r, "\int^{\mathrm{dR}}"] \arrow[d, "\mathrm{comp}^{\otimes 2}_{\mathrm{B,dR}} \circ (s \otimes 1)"'] & \mathbb{C}\,\boldsymbol{e}^{\mathrm{dR}} \arrow[d] \\
H^1(\mathrm{SL}_2(\mathbb{Z}), \Gamma_\infty; V_n) \otimes H^1(\mathcal{M}^{\mathrm{an}}_{1,1}; \mathbb{V}_n) \otimes \mathbb{C} \arrow[r, "\int^{\mathrm{B}}"] & \mathbb{C}\,\boldsymbol{e}^{\mathrm{B}}
\end{tikzcd}$$

commutes, where $\boldsymbol{e}^{\mathrm{dR}} = (2\pi i)^{n+1}\boldsymbol{e}^{\mathrm{B}}$ *and where* $\textstyle\int^{\mathrm{B}}$ *denotes the cup product induced by* $\langle\ ,\ \rangle_{\mathrm{B}}$ *evaluated on the fundamental class of* $\overline{\mathcal{M}}_{1,1}(\mathbb{C})$.

Proof. Suppose that $\xi \in \Omega^1_{\mathrm{cusp}}(X; \mathscr{V}_n)$ is cuspidal and $\eta \in \Omega^1(X, \mathscr{V}_n)^{\omega}$. They represent cohomology classes. Let U be the analytic q-disk $\{q \in \mathbb{C} : |q| < e^{-2\pi}\}$ and $U' = U - \{0\}$. Since ξ is cuspidal, its restriction to U' is exact. Choose a meromorphic section $\widetilde{F} \in \Gamma(U', \mathscr{V}_n^{\mathrm{an}})$ such that $\nabla\widetilde{F} = \xi$ on U' and the image F of $\widetilde{F}$ in $\Omega^0(\mathbb{D}^*, \mathscr{V}_n)$ is the unique element of $F \in \Omega^0_{\mathrm{cusp}}(\mathbb{D}^*, \mathscr{V}_n)$ such that ∇F is the q-expansion of ξ.

Choose $r, R \in \mathbb{R}$ such that $0 < r < R < e^{-2\pi}$. Choose a smooth function $\rho : U \to \mathbb{R}_{\geq 0}$ which vanishes outside the annulus

$$A := \{q \in U : r \leq |q| \leq R\}$$

and is equal to 1 identically when $|q| \leq r$. Then $\tilde{\xi} := \xi - \nabla(\rho\widetilde{F})$ extends by 0 to a smooth 1-form on the orbifold $\overline{\mathcal{M}}^{\mathrm{an}}_{1,1}$ with values in $\mathscr{V}_n \otimes \mathbb{C}$, which vanishes in a neighbourhood of the cusp and equals ξ on $|q| > R$.

Since $\rho\widetilde{F}$ is a smooth section of $\mathscr{V}_n \otimes \mathbb{C}$ over $\mathcal{M}^{\mathrm{an}}_{1,1}$, $\tilde{\xi}$ is a smooth form that represents the same class in $H^1_{\mathrm{cusp}}(\mathcal{M}_{1,1}, \mathscr{V}_n)$ as ξ. Since $\nabla(\rho\widetilde{F})$ is supported in the annulus A and since $\xi \wedge \eta = 0$, $\tilde{\xi} \wedge \eta$ is supported in A. Using (5-3), we have

$$\begin{aligned}
\textstyle\int^{\mathrm{B}}(\xi \otimes \eta) = \textstyle\int_{\overline{\mathcal{M}}^{\mathrm{an}}_{1,1}} \langle \tilde{\xi}, \eta\rangle_{\mathrm{B}} &= (2\pi i)^n \int_{\overline{\mathcal{M}}^{\mathrm{an}}_{1,1}} \langle \tilde{\xi}, \eta\rangle_{\mathrm{dR}} = -(2\pi i)^n \int_A d\langle \rho\widetilde{F}, \eta\rangle_{\mathrm{dR}} \\
&\overset{\text{(Stokes)}}{=} (2\pi i)^n \int_{|q|=r} \langle \widetilde{F}, \eta\rangle_{\mathrm{dR}} = (2\pi i)^{n+1} \operatorname{Res}_{q=0} \widetilde{F}\eta \\
&= (2\pi i)^{n+1}\{\xi, \eta\} = (2\pi i)^{n+1} \textstyle\int^{\mathrm{dR}}(\xi \otimes \eta). \qquad \square
\end{aligned}$$

Remark 5.7. The Hecke invariance of Guerzhoy's inner product (5-1) on cusp forms follows directly from this description of the inner product using the projection formula.[2]

Remark 5.8. The section s defined above cannot be Hecke invariant. If it were, the Eisenstein series G_{n+2} would be orthogonal to $S^!_{n+2}$ under the inner product $\{\ ,\ \}$. However, this is not the case, since it would contradict the discussion in Section 6: orthogonality with respect to G_{n+2} and cuspidality are two distinct linear conditions on the space of weakly holomorphic modular forms.

Consider, by way of example, the case $n = 10$. There is a unique $\mathbb{Q}$-linear combination $f_{-1} \in S^!_{12}$ of the weakly holomorphic modular forms $\mathsf{G}_{24}\Delta^{-1}$, G_{12} and Δ such that $f_{-1} = q^{-1} + O(q^2)$. It is given explicitly by

$$f_{-1} = q^{-1} + 47709536\,q^2 + 39862705122\,q^3 + 7552626810624\,q^4 + \cdots.$$

Its class in $S^!_{12}/\mathcal{D}^{11}M^!_{-10}$ is a Hecke eigenform, with the same eigenvalues as Δ. Since the leading coefficient of Δ is q, we have

$$\{\Delta, f_{-1}\} = 1.$$

[2]This states that if $f : X \to Y$ is a smooth proper morphism, then $(f_*a) \cdot b = f_*(a \cdot f^*b)$, where $a \in H^\bullet(X; f^*\mathbb{V})$ and $b \in H^\bullet(Y, \mathbb{V})$.

On the other hand, from the Fourier expansion

$$\mathsf{G}_{12} = \tfrac{691}{65520} + q + 2049q^2 + 177148q^3 + \cdots ,$$

we find that a naive application of the formula (5-1) to f_{-1} and G_{12} would give $\{\mathsf{G}_{12}, f_{-1}\} = 1$, and hence $S^!_{12}$ is not orthogonal to $\mathsf{G}_{12}\mathbb{Q}$.

5F. ***Proof of Theorem 1.7.*** Suppose that f is a Hecke eigen cusp form of weight $2n$. Denote the associated 2-dimensional subspace of $H^1(\mathcal{M}_{1,1}, \mathcal{V}_{2n-2}) \otimes K_f$ by V_f. It has Betti and de Rham realisations V_f^{B} and V_f^{dR} related by the comparison isomorphism $\mathrm{comp}_{\mathrm{B,dR}} : V_f^{\mathrm{dR}} \otimes_{K_f} \mathbb{C} \to V_f^{\mathrm{B}} \otimes_{K_f} \mathbb{C}$. The cup product induces a nondegenerate, skew-symmetric pairing

$$\langle\ ,\ \rangle : V_f \otimes V_f \to K_f(-2n+1).$$

It has Betti and de Rham realisations

$$\langle\ ,\ \rangle_{\mathrm{B}} = \textstyle\int^{\mathrm{B}} \otimes \boldsymbol{e}^{\mathrm{B}} \quad \text{and} \quad \langle\ ,\ \rangle_{\mathrm{dR}} = \textstyle\int^{\mathrm{dR}} \otimes \boldsymbol{e}^{\mathrm{dR}}.$$

Let α_f, β_f be a K_f-de Rham basis of V_f^{dR}. There is a K_f basis r_f^+, r_f^- of V_f^{B} with $\langle r_f^+, r_f^- \rangle_{\mathrm{B}} = \boldsymbol{e}_{\mathrm{B}}$. Then

$$\begin{pmatrix}\alpha_f & \beta_f\end{pmatrix} = \begin{pmatrix} r_f^+ & r_f^- \end{pmatrix} \begin{pmatrix} \eta_f^+ & \omega_f^+ \\ i\eta_f^- & i\omega_f^- \end{pmatrix} = \begin{pmatrix} r_f^+ & r_f^- \end{pmatrix} P_f,$$

where $\eta_f^\pm$ and $\omega_f^\pm$ are real numbers. By Proposition 5.6,

$$\begin{aligned}\big\langle \mathrm{comp}_{\mathrm{B,dR}}(\omega_f), \mathrm{comp}_{\mathrm{B,dR}}(\eta_f)\big\rangle_{\mathrm{B}} &= \langle \eta_f^+ r_f^+ + i\eta_f^- r_f^-,\ \omega_f^+ r_f^+ + i\omega_f^- r_f^- \rangle_{\mathrm{B}} \\ &= \det(P_f)\,\boldsymbol{e}_{\mathrm{B}} \\ &= (2\pi i)^{-2n+1} \det(P_f)\,\boldsymbol{e}_{\mathrm{dR}}.\end{aligned}$$

Since $\langle \omega_f, \eta_f \rangle_{\mathrm{dR}} \in K_f^\times\, \boldsymbol{e}_{\mathrm{dR}}$, this implies that $\det(P_f) \in (2\pi i)^{2n-1} K_f^\times$.

6. A $\mathbb{Q}$-de Rham splitting of the Hodge filtration

Let $\ell = \dim S_n$, so that

$$\dim M_n^! / \mathcal{D}^{n-1} M_{2-n}^! = \dim H^1_{\mathrm{dR}}(\mathcal{M}_{1,1}, \mathcal{V}_{n-2}) = 2\ell + 1.$$

Let ord_∞ denote the order of vanishing at the cusp. It follows from the Riemann–Roch formula, as noted in [Duke and Jenkins 2008], that

$$\mathrm{ord}_\infty f \le \ell$$

for all $f \in M_n^!$. Furthermore, it was shown in [Guerzhoy 2008] that for any $f \in M_n^!$, there exists a unique representative of f modulo $\mathcal{D}^{n-1} M_{2-n}^!$ such that

$$\mathrm{ord}_\infty f \ge -\ell.$$

Since the dimension of $M_n^!/\mathcal{D}^{n-1}M_{2-n}^!$ is exactly $2\ell+1$, it follows that such an f is uniquely determined by its $2\ell+1$ Fourier coefficients $(a_{-\ell},\ldots,a_\ell)\in\mathbb{Q}^{2\ell+1}$, where

$$f=\sum_{n\geq -\ell} a_n q^n,$$

and, conversely, any vector $(a_{-\ell},\ldots,a_\ell)\in\mathbb{Q}^{2\ell+1}$ uniquely determines an element in $M_n^!/D^{n-1}M_{2-n}^!$. It follows that the functions $f\in M_n^!$ of the form

$$f_m=q^m+O(q^{\ell+1})$$

for every $-\ell\leq m\leq \ell$, are a $\mathbb{Q}$-basis for $H^1_{\mathrm{dR}}(\mathcal{M}_{1,1};\mathscr{V}_n)$, by Theorem 1.2. These functions satisfy some remarkable properties, studied in [Duke and Jenkins 2008].

This basis simultaneously gives a $\mathbb{Q}$-de Rham splitting of the Hodge filtration

$$H^1_{\mathrm{dR}}(\mathcal{M}_{1,1},\mathscr{V}_{n-2})=F^0\supset F^1=\cdots=F^{n-2}\supset F^{n-1}=M_n\supset F^n=0$$

and of the weight filtration

$$0=W_{n-2}\subset H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1},\mathscr{V}_{n-2})=W_{n-1}\subset W_n=\cdots=W_{2n-3}\subset W_{2n-2}=H^1_{\mathrm{dR}}(\mathcal{M}_{1,1},\mathscr{V}_{n-2}).$$

The splitting of

$$0\to F^{n-1}H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1},\mathscr{V}_{n-2})\to H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1},\mathscr{V}_{n-2})\to \mathrm{gr}_F^{n-1}H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1},\mathscr{V}_{n-2})\to 0$$

is given by lifting $\mathrm{gr}^0_F H^1_{\mathrm{cusp,dR}}(\mathcal{M}_{1,1},\mathscr{V}_{n-2})$ to the subspace of $M_n^!/\mathcal{D}^{n-1}M_{2-n}^!$ consisting of those f whose Fourier coefficients a_j vanish when $0\leq j\leq\ell$. The splitting of the weight filtration is given by the Eisenstein series.

Appendix: Framings

The aim of this appendix is to explain the choice of the power of $2\pi i$ in the cocycle formula (1-6). Here we take $\Bbbk=\mathbb{Q}$.

Recall that $\pi:\mathcal{E}\to X$ denotes the universal elliptic curve and that $\mathbb{V}_n^{\mathrm{B}}$ denotes the n-th symmetric power of $R^1\pi_*\mathbb{Q}$. It underlies a variation of Hodge structure of weight n. To give a framing of $\mathbb{V}_n^{\mathrm{B}}$, it suffices to give a framing of $\mathbb{V}^{\mathrm{B}}:=\mathbb{V}_1^{\mathrm{B}}$.

The pullback of $\mathbb{V}^{\mathrm{B}}$ to $\mathfrak{H}$ along $\rho:\mathfrak{H}\to X$ is the trivial local system whose fibre over $z\in\mathfrak{H}$ is $H^1(E_z)$, where $E_z:=\mathbb{C}/(\mathbb{Z}\oplus z\mathbb{Z})$. We identify $H^1(E_z)$ with its dual $H_1(E_z)\cong\mathrm{Hom}(H_1(E_z),\mathbb{Z})$ by Poincaré duality

$$\mathrm{P}:H_1(E_z)\to H^1(E_z),\quad \mathrm{P}(c):=\langle c,\ \rangle,$$

where $\langle\ ,\ \rangle$ denotes the intersection pairing. On the level of Hodge structures, Poincaré duality is an isomorphism

$$\mathrm{P}:H_1(E_z,\mathbb{Q})\to H^1(E_z,\mathbb{Q}(1)).$$

Denote the standard basis of $H_1(E_z;\mathbb{Z})$ by $\boldsymbol{a}, \boldsymbol{b}$. These classes correspond to the lattice points 1 and z, respectively. Denote the dual basis of $H^1(E_z) \cong \operatorname{Hom}(H_1(E_z), \mathbb{Z})$ by $\check{\boldsymbol{a}}, \check{\boldsymbol{b}}$. Then $\check{\boldsymbol{b}} = \mathrm{P}(\boldsymbol{a})$ and $\check{\boldsymbol{a}} = \mathrm{P}(\boldsymbol{b})$. We identify the (Betti) components of $H_1(E)$ and $H^1(E)$ via P. With this identification

$$dw = -\boldsymbol{b} + z\boldsymbol{a},$$

where w is the coordinate in the universal covering $\mathbb{C}$ of E_z. The abelian differential dx/y on the elliptic curve corresponding to $\rho(z)$ is $2\pi i\, dw$. This is the section T of $\mathscr{V}_n$. So T corresponds to the section

$$\mathsf{T} = 2\pi i\, dw = 2\pi i(z\boldsymbol{a} - \boldsymbol{b})$$

of $\mathscr{V}^{\mathrm{an}} := \mathbb{V} \otimes \mathcal{O}(\mathfrak{H})$.

Each $f \in M^!_{n+2}$ corresponds to an element $h(u, v) \in \mathcal{O}(X)$. The corresponding 1-form ω_f is $h\mathsf{T}^n\omega$. So

$$\rho^*\omega_f = (2\pi i)^n f(z)(z\boldsymbol{a} - \boldsymbol{b})^n \frac{dq}{q} = (2\pi i)^{n+1} f(z)(z\boldsymbol{a} - \boldsymbol{b})^n\, dz.$$

Acknowledgments

This paper is an outgrowth of a collaboration which originated in 2014–2015 during a stay at the Institute for Advanced Study. Both authors wish to thank the IAS for its hospitality and support. We would also like to thank Ma Luo for providing the isomorphism of Z (defined in Section 2E) with the Fermat cubic, and Jesse Silliman for bringing Coleman's work [1996] to our attention. Francis Brown also thanks the IHES for hospitality.

References

[Bol 1949] G. Bol, "Invarianten linearer differentialgleichungen", *Abh. Math. Sem. Univ. Hamb.* **16**:3-4 (1949), 1–28. MR Zbl

[Bringmann et al. 2013] K. Bringmann, P. Guerzhoy, Z. Kent, and K. Ono, "Eichler–Shimura theory for mock modular forms", *Math. Ann.* **355**:3 (2013), 1085–1121. MR Zbl

[Bruggeman et al. 2014] R. Bruggeman, Y. Choie, and N. Diamantis, "Holomorphic automorphic forms and cohomology", preprint, 2014. arXiv

[Candelori 2014] L. Candelori, "Harmonic weak Maass forms of integral weight: a geometric approach", *Math. Ann.* **360**:1-2 (2014), 489–517. MR Zbl

[Coleman 1996] R. F. Coleman, "Classical and overconvergent modular forms", *Invent. Math.* **124**:1-3 (1996), 215–241. MR Zbl

[Deligne 1970] P. Deligne, *Équations différentielles à points singuliers réguliers*, Lecture Notes in Math. **163**, Springer, 1970. MR Zbl

[Duke and Jenkins 2008] W. Duke and P. Jenkins, "On the zeros and coefficients of certain weakly holomorphic modular forms", *Pure Appl. Math. Q.* **4**:4 (2008), 1327–1340. MR Zbl

[Eichler 1957] M. Eichler, "Eine Verallgemeinerung der Abelschen Integrale", *Math. Z.* **67** (1957), 267–298. MR Zbl

[Guerzhoy 2008] P. Guerzhoy, "Hecke operators for weakly holomorphic modular forms and supersingular congruences", *Proc. Amer. Math. Soc.* **136**:9 (2008), 3051–3059. MR Zbl

[Hain 2013] R. Hain, "Notes on the universal elliptic KZB connection", 2013. To appear in *Pure Appl. Math. Q.* arXiv

[Katz 1973] N. M. Katz, "p-adic properties of modular schemes and modular forms", pp. 69–190 in *Modular functions of one variable, III* (Antwerp, 1972), edited by W. Kuyk and J.-P. Serre, Lecture Notes in Math. **350**, Springer, 1973. MR Zbl

[Katz and Mazur 1985] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Annals of Math. Studies **108**, Princeton Univ. Press, 1985. MR Zbl

[Kazalicki and Scholl 2016] M. Kazalicki and A. J. Scholl, "Modular forms, de Rham cohomology and congruences", *Trans. Amer. Math. Soc.* **368**:10 (2016), 7097–7117. MR Zbl

[Manin 1973] Y. I. Manin, "Periods of parabolic forms and p-adic Hecke series", *Mat. Sb.* (*N.S.*) **92(134)**:3 (1973), 378–401. In Russian; translated in *Math. USSR-Sb.* **21**:3 (1973), 371–393. MR Zbl

[Scholl 1985] A. J. Scholl, "Modular forms and de Rham cohomology; Atkin–Swinnerton–Dyer congruences", *Invent. Math.* **79**:1 (1985), 49–77. MR Zbl

[Shimura 1959] G. Shimura, "Sur les intégrales attachées aux formes automorphes", *J. Math. Soc. Japan* **11** (1959), 291–311. MR Zbl

[Swinnerton-Dyer 1973] H. P. F. Swinnerton-Dyer, "On l-adic representations and congruences for coefficients of modular forms", pp. 1–55 in *Modular functions of one variable, III* (Antwerp, 1972), edited by W. Kuyk and J.-P. Serre, Lecture Notes in Math. **350**, Springer, 1973. Correction in *Modular functions of one variable, IV*, Lecture Notes in Math. **476** (1975), 149. MR Zbl

[Zagier 2008] D. Zagier, "Elliptic modular forms and their applications", pp. 1–103 in *The 1-2-3 of modular forms*, edited by K. Ranestad, Springer, 2008. MR Zbl

Communicated by Yuri Manin
Received 2017-08-03 Revised 2017-12-22 Accepted 2018-01-22

francis.brown@all-souls.ox.ac.uk , *All Souls College, Oxford, United Kingdom*

hain@math.duke.edu *Department of Mathematics, Duke University, Durham, NC, United States*

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 12 No. 3 2018

Pseudo-exponential maps, variants, and quasiminimality 493
MARTIN BAYS and JONATHAN KIRBY

On faithfulness of the lifting for Hopf algebras and fusion categories 551
PAVEL ETINGOF

Mean square in the prime geodesic theorem 571
GIACOMO CHERUBINI and JOÃO GUERREIRO

Elliptic quantum groups and Baxter relations 599
HUAFENG ZHANG

Differential forms in positive characteristic, II: cdh-descent via functorial Riemann–Zariski spaces 649
ANNETTE HUBER and SHANE KELLY

Nilpotence order growth of recursion operators in characteristic p 693
ANNA MEDVEDOVSKY

Algebraic de Rham theory for weakly holomorphic modular forms of level one 723
FRANCIS BROWN and RICHARD HAIN

1937-0652(2018)12:3;1-E