

Algebra & Number Theory

Volume 12

2018

No. 5



Algebra & Number Theory

msp.org/ant

EDITORS

MANAGING EDITOR

Bjorn Poonen
Massachusetts Institute of Technology
Cambridge, USA

EDITORIAL BOARD CHAIR

David Eisenbud
University of California
Berkeley, USA

BOARD OF EDITORS

Richard E. Borcherds	University of California, Berkeley, USA	Martin Olsson	University of California, Berkeley, USA
Antoine Chambert-Loir	Université Paris-Diderot, France	Raman Parimala	Emory University, USA
J-L. Colliot-Thélène	CNRS, Université Paris-Sud, France	Jonathan Pila	University of Oxford, UK
Brian D. Conrad	Stanford University, USA	Anand Pillay	University of Notre Dame, USA
Samit Dasgupta	University of California, Santa Cruz, USA	Michael Rapoport	Universität Bonn, Germany
Hélène Esnault	Freie Universität Berlin, Germany	Victor Reiner	University of Minnesota, USA
Gavril Farkas	Humboldt Universität zu Berlin, Germany	Peter Sarnak	Princeton University, USA
Hubert Flenner	Ruhr-Universität, Germany	Joseph H. Silverman	Brown University, USA
Sergey Fomin	University of Michigan, USA	Michael Singer	North Carolina State University, USA
Edward Frenkel	University of California, Berkeley, USA	Christopher Skinner	Princeton University, USA
Andrew Granville	Université de Montréal, Canada	Vasudevan Srinivas	Tata Inst. of Fund. Research, India
Joseph Gubeladze	San Francisco State University, USA	J. Toby Stafford	University of Michigan, USA
Roger Heath-Brown	Oxford University, UK	Pham Huu Tiep	University of Arizona, USA
Craig Huneke	University of Virginia, USA	Ravi Vakil	Stanford University, USA
Kiran S. Kedlaya	Univ. of California, San Diego, USA	Michel van den Bergh	Hasselt University, Belgium
János Kollár	Princeton University, USA	Marie-France Vignéras	Université Paris VII, France
Philippe Michel	École Polytechnique Fédérale de Lausanne	Kei-Ichi Watanabe	Nihon University, Japan
Susan Montgomery	University of Southern California, USA	Shou-Wu Zhang	Princeton University, USA
Shigefumi Mori	RIMS, Kyoto University, Japan		

PRODUCTION

production@msp.org
Silvio Levy, Scientific Editor


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/>

© 2018 Mathematical Sciences Publishers

Semistable Chow–Hall algebras of quivers and quantized Donaldson–Thomas invariants

Hans Franzen and Markus Reineke

The semistable ChowHa of a quiver with stability is defined as an analog of the cohomological Hall algebra of Kontsevich and Soibelman via convolution in equivariant Chow groups of semistable loci in representation varieties of quivers. We prove several structural results on the semistable ChowHa, namely isomorphism of the cycle map, a tensor product decomposition, and a tautological presentation. For symmetric quivers, this leads to an identification of their quantized Donaldson–Thomas invariants with the Chow–Betti numbers of moduli spaces.

1. Introduction

The cohomological Hall algebra, or CoHa for short, of a quiver is defined in [Kontsevich and Soibelman 2011] as an analog of the Hall algebra construction of Ringel [1990] in equivariant cohomology of representation varieties. In [Kontsevich and Soibelman 2011] the CoHa serves as a tool for the study of quantized Donaldson–Thomas invariants of quivers, and in particular their integrality properties, since it admits a purely algebraic description as a shuffle algebra “with kernel” on spaces of symmetric polynomials. In [Efimov 2012], the CoHa of a symmetric quiver is shown to be a free super-commutative algebra, proving the positivity of quantized Donaldson–Thomas invariants in this case.

In another direction, the CoHa is used in [Franzen 2016; 2018] to determine the ring structure on the cohomology of noncommutative Hilbert schemes and more general framed moduli spaces of quiver representations, as defined in [Engel and Reineke 2009].

Already in [Franzen 2018] it turns out that a “local” version of the CoHa (the semistable CoHa), constructed via convolution on semistable loci of representation varieties with respect to a stability, is particularly useful, and that it is also convenient to replace equivariant cohomology by equivariant Chow groups.

In the present paper, we study this local version, called the semistable ChowHa, more systematically and demonstrate their utility both for understanding the structure of the CoHa and for the study of quantized Donaldson–Thomas invariants.

We prove the following structural properties of the semistable ChowHa:

The equivariant cycle map between the semistable ChowHa and the semistable CoHa is an isomorphism (Corollary 5.6), which can be viewed as a generalization of a result of [King and Walter 1995] on the cycle

MSC2010: primary 14N35; secondary 14C15, 16G20.

Keywords: cohomological Hall algebra, Donaldson–Thomas invariants, quiver moduli.

map for fine moduli spaces of quivers. We exhibit a tensor product decomposition (Theorem 6.2) of the CoHa into all semistable CoHa's for various slopes of the stability, categorifying the Harder–Narasimhan, or wall-crossing, formula of [Reineke 2003]. We give a “tautological” presentation of the semistable ChowHa (Theorem 8.1), in the spirit of [Franzen 2015], which generalizes the algebraic description of [Kontsevich and Soibelman 2011] of the CoHa. Quite surprisingly, such a tautological presentation remains valid for the equivariant Chow groups of stable loci in representation varieties (Theorem 9.1). From this, we conclude that the quantized Donaldson–Thomas invariants of a symmetric quiver are given by the Poincaré polynomials of the Chow groups of moduli spaces of stable quiver representations (Theorem 9.2). This shows that quantized Donaldson–Thomas invariants are of algebro-geometric origin; compare [Meinhardt and Reineke 2014] where quantized Donaldson–Thomas invariants are interpreted via intersection cohomology of moduli spaces of semistable quiver representations.

The proofs of these structural results basically only use the Harder–Narasimhan stratification of representation varieties of [Reineke 2003], properties of equivariant Chow groups, and the result of Efimov [2012]. All structural results are illustrated by examples in Section 10: We first give a complete description of the Hall algebra of a two-cycle quiver, which is the only symmetric quiver with known representation theory apart from the trivial and the one-loop quiver whose CoHa's are already described in [Kontsevich and Soibelman 2011]. Then we consider the only nontrivial (i.e., not isomorphic to the CoHa of a trivial quiver) semistable ChowHa for the Kronecker quiver — we observe that it is not super-commutative, but still has the same Poincaré–Hilbert series as a free super-commutative algebra; for this, the representation theory of the Kronecker quiver is used essentially. Then we illustrate the calculation of Chow–Betti numbers of moduli spaces of stable representations in the context of classical invariant theory, and finally hint at an algebraic derivation of the explicit formula of [Reineke 2012] for quantized Donaldson–Thomas invariants of multiple loop quivers.

The paper is organized as follows:

After reviewing basic facts on quiver representations (Section 2) and the definition of quantized Donaldson–Thomas invariants (Section 3), we define the semistable ChowHa in Section 4. In Section 5 we study the cycle map from ChowHa to CoHa, and prove it to be an isomorphism via induction over Harder–Narasimhan strata and framing techniques. The Harder–Narasimhan stratification is also used in Section 6 to derive the tensor product decomposition of the ChowHa. We recall the algebraic description of the CoHa of [Kontsevich and Soibelman 2011] and Efimov's theorem in Section 7. Again using the Harder–Narasimhan stratification, we obtain the algebraic description of the semistable ChowHa in Section 8. In Section 9, the previous results are combined to identify quantized Donaldson–Thomas invariants and Chow–Betti numbers. Finally, the examples mentioned above are developed in Section 10.

2. A reminder on quiver representations

Let Q be a quiver — i.e., a finite oriented graph — whose set of vertices and arrows we denote by Q_0 and Q_1 , respectively. We will often suppress the dependency on Q in the notation. The bilinear form $\chi = \chi_Q$ on \mathbb{Z}^{Q_0} defined by

$$\chi(d, e) = \sum_{i \in Q_0} d_i e_e - \sum_{\alpha: i \rightarrow j} d_i e_j = \sum_{i, j} (\delta_{i, j} - a_{i, j}) d_i e_j$$

is called the Euler form of Q . Here, $a_{i, j}$ is the number of arrows from i to j in Q . We denote the antisymmetrization $\chi(d, e) - \chi(e, d)$ of the Euler form by $\langle d, e \rangle$. Let $\Gamma = \mathbb{Z}_{\geq 0}^{Q_0}$ be the monoid of dimension vectors of Q .

Let k be a field. A representation M of Q over k is a collection of finite-dimensional vector spaces M_i with $i \in Q_0$ together with linear maps $M_\alpha : M_i \rightarrow M_j$ for every arrow $\alpha : i \rightarrow j$. See [Assem et al. 2006] for more details. The tuple $\underline{\dim} M = (\dim M_i \mid i \in Q_0) \in \Gamma$ is called the dimension vector of M . For a dimension vector $d \in \Gamma$, we define $R_d(k)$ to be the vector space

$$R_d(k) = \bigoplus_{\alpha: i \rightarrow j} \text{Hom}(k^{d_i}, k^{d_j})$$

on which we have an action of the group $G_d(k) = \prod_{i \in Q_0} \text{GL}_{d_i}(k)$ via base change. An element of $R_d(k)$ is a representation of Q on the vector spaces $(k^{d_i})_i$. Being an affine space, $R_d(k)$ admits a \mathbb{Z} -model, i.e., there exists a scheme R_d whose set of k -valued points is $R_d(k)$. Likewise, there is a group scheme G_d which is a \mathbb{Z} -model for $G_d(k)$.

We introduce a stability condition θ of Q , that is, a linear form $\mathbb{Z}^{Q_0} \rightarrow \mathbb{Z}$. For a nonzero dimension vector d , the rational number

$$\frac{\theta(d)}{\sum_i d_i}$$

is called the θ -slope of d . For a rational number μ , let $\Gamma^{\theta, \mu}$ be the submonoid of all $d \in \Gamma$ with $d = 0$ or whose θ -slope is μ . If M is a nonzero representation of Q over k , the θ -slope of M is defined as the slope of its dimension vector. A representation M of Q over k is called θ -semistable if no nonzero subrepresentation of M has larger θ -slope than M . It is called θ -stable if the θ -slope of every nonzero subrepresentation M' is strictly less than the slope of M , unless M' agrees with M . There is a Zariski-open subset $R_d^{\theta\text{-sst}}$ of the scheme R_d whose set of k -valued points is the set of θ -semistable representations of Q . There is also an open subset $R_d^{\theta\text{-st}}$ of $R_d^{\theta\text{-sst}}$ parametrizing absolutely θ -stable representations, that means $R_d^{\theta\text{-st}}(k)$ consists of those $M \in R_d(k)$ such that $M \otimes_k K$ is θ -stable for every finite extension $K \mid k$.

3. Quantum Donaldson–Thomas invariants

Fix a prime power q and let $\mathbb{F} = \mathbb{F}_q$ be the finite field with q elements. Then $R_d(\mathbb{F})$ and $G_d(\mathbb{F})$ are finite sets, and we consider the set of orbits $R_d(\mathbb{F})/G_d(\mathbb{F})$. We define the *completed Hall algebra* of Q as the vector space

$$H_q = H_q((Q)) = \left\{ f \mid f : \bigsqcup_{d \in \Gamma} R_d(\mathbb{F})/G_d(\mathbb{F}) \rightarrow \mathbb{Q} \right\}$$

equipped with the following convolution-type multiplication: for two functions f and g , we define

$$(f * g)(X) = \sum_{U \subseteq X} f(U)g(X/U),$$

the sum ranging over all subrepresentations of X . Note that this sum is finite. This multiplication turns H_q into an associative algebra. We define yet another algebra. Set $\mathbb{T}_q := \mathbb{Q}(q^{1/2})[[t_i \mid i \in \mathbb{Q}_0]]$ and let the multiplication be given by

$$t^d \circ t^e = (-q^{1/2})^{\langle d, e \rangle} t^{d+e}.$$

It is shown in [Reineke 2003] that the so-called integration map $\int : H_q \rightarrow \mathbb{T}_q$ defined by

$$\int f = \sum_{[X]} \frac{(-q^{1/2})^{\chi(\dim X, \dim X)}}{\#\text{Aut}(X)} \cdot f(X) \cdot t^{\dim X}$$

is a homomorphism of algebras. Define $\mathbf{1} \in H_q$ to be the function with $\mathbf{1}(X) = 1$ for all $[X]$. An easy computation shows that $A(q, t) := \int \mathbf{1}$ equals

$$A(q, t) = \sum_d (-q^{1/2})^{-\chi(d, d)} \prod_i \prod_{v=1}^{d_i} (1 - q^{-v})^{-1} t^d.$$

For a stability condition θ of Q and a rational number μ , we define $\mathbf{1}^{\theta, \mu} \in H_q$ as the sum of the characteristic functions on $R_d^{\theta\text{-sst}}(\mathbb{F})/G_d(\mathbb{F})$ over all $d \in \Gamma^{\theta, \mu}$. Set $A^{\theta, \mu}(q, t) = \int \mathbf{1}^{\theta, \mu}$. Using a Harder–Narasimhan type recursion, it is shown in [Reineke 2003] that:

Theorem 3.1.
$$\mathbf{1} = \prod_{\mu \in \mathbb{Q}}^{\leftarrow} \mathbf{1}^{\theta, \mu} \quad \text{in } H_q.$$

This implies that the series A and $A^{\theta, \mu}$ relate in the same way in the twisted power series ring \mathbb{T}_q .

Let R be the power series ring $\mathbb{Q}(q^{1/2})[[t_i \mid i \in \mathbb{Q}_0]]$ with the usual multiplication. Let R_+ be the set of power series without constant coefficient. There exists a unique continuous bijection $\text{Exp} : R_+ \rightarrow 1 + R_+$ such that $\text{Exp}(f + g) = \text{Exp}(f) \text{Exp}(g)$ and

$$\text{Exp}(q^{k/2} t^d) = \frac{1}{1 - q^{k/2} t^d}$$

for every $k \in \mathbb{Z}$ and $d \in \Gamma$. This function is called the plethystic exponential. We call the stability condition θ generic for the slope $\mu \in \mathbb{Q}$ (or μ -generic) if $\langle d, e \rangle = 0$ for all $d, e \in \Gamma^{\theta, \mu}$. Assuming that θ is μ -generic, the series $A^{\theta, \mu}$ can be displayed as a plethystic exponential.

Theorem 3.2 [Kontsevich and Soibelman 2011]. *For a μ -generic stability condition θ , there are polynomials $\tilde{\Omega}_d^\theta(q) = \sum_k \tilde{\Omega}_{d, 2k}^\theta q^k$ in $\mathbb{Z}[q]$ for every nonzero dimension vector d of slope μ such that*

$$A^{\theta, \mu}(q^{-1}, t) = \text{Exp} \left(\frac{1}{1-q} \sum_d (-q^{1/2})^{\chi(d, d)} \tilde{\Omega}_d^\theta(q) t^d \right).$$

Definition 3.3. If θ is a μ -generic stability condition and d is a nonzero dimension vector of slope μ then the coefficients of

$$\Omega_d^\theta(q) = \sum_{k \in \mathbb{Z}} \Omega_{d,k}^\theta q^{k/2} := q^{\chi(d,d)/2} \tilde{\Omega}_d^\theta(q) \in \mathbb{Z}[q^{\pm 1/2}]$$

are called the *quantum Donaldson–Thomas invariants* of Q with respect to θ .

When the quiver Q is symmetric, that means the Euler form of Q is a symmetric bilinear form, then every stability condition is μ -generic for every value $\mu \in \mathbb{Q}$. For example, the trivial stability condition is 0-generic and we can define the Donaldson–Thomas invariants $\Omega_{d,k}^0$ for every $d \neq 0$. Note that Theorem 3.1 implies $\Omega_{d,k}^0 = \Omega_{d,k}^\theta$ for every θ . We may therefore write $\Omega_{d,k}$ in this case.

4. The semistable ChowHa

Fix an algebraically closed field k . Abusing the notation from the first section, we will use the symbols $R_d, R_d^{\theta\text{-sst}}, R_d^{\theta\text{-st}}$, and G_d for the base extensions of the respective \mathbb{Z} -models to $\text{Spec } k$.

Let d be a dimension vector for Q . We define $\mathcal{A}_d^{\theta\text{-sst}}$ to be the G_d -equivariant Chow ring with rational coefficients of the semistable locus

$$\mathcal{A}_d^{\theta\text{-sst}}(Q) = A_{G_d}^*(R_d^{\theta\text{-sst}})_{\mathbb{Q}}.$$

For the definition of equivariant Chow groups and rings, see [Edidin and Graham 1998]. As we will always work with rational coefficients, we will often omit it in the notation. We define $\mathcal{A}^{\theta\text{-sst}, \mu}$ as the graded vector space

$$\mathcal{A}^{\theta\text{-sst}, \mu}(Q) = \bigoplus_{d \in \Gamma^{\theta, \mu}} \mathcal{A}_d^{\theta\text{-sst}}.$$

We mimic Kontsevich and Soibelman’s construction [2011] of the cohomological Hall algebra (CoHa) of a quiver with stability and trivial potential.

For two dimension vectors d and e of the same slope μ , we set $Z_{d,e}$ as the subspace of R_{d+e} of representations M which have a block upper triangular structure as indicated, i.e., for every arrow $\alpha : i \rightarrow j$, the linear map M_α sends the first d_i coordinate vectors of $k^{d_i+e_i}$ into the subspace of $k^{d_j+e_j}$ spanned by the first d_j coordinate vectors. We consider

$$R_d \times R_e \leftarrow Z_{d,e} \rightarrow R_{d+e},$$

the left-hand map sending a representation $M = \begin{pmatrix} M' & * \\ & M'' \end{pmatrix}$ to the pair (M', M'') , and the right-hand map being the inclusion. These maps are called Hecke correspondences. For a short exact sequence $0 \rightarrow M' \rightarrow M \rightarrow M'' \rightarrow 0$ of representations of the same slope, M is θ -semistable if and only if both M'

and M'' are. We thus obtain cartesian squares

$$\begin{array}{ccccc} R_d \times R_e & \longleftarrow & Z_{d,e} & \longrightarrow & R_{d+e} \\ \cup & & \cup & & \cup \\ R_d^{\text{sst}} \times R_e^{\text{sst}} & \longleftarrow & Z_{d,e} \cap R_{d+e}^{\text{sst}} & \longrightarrow & R_{d+e}^{\text{sst}}. \end{array}$$

The action of $G = G_{d+e}$ on R_{d+e} restricts to an action of the parabolic $P = \begin{pmatrix} G_d & * \\ & G_e \end{pmatrix}$ on $Z_{d,e}$, and the map $Z_{d,e} \rightarrow R_d \times R_e$ is compatible with the action of its Levi $L = G_d \times G_e$ on $R_d \times R_e$. With respect to these actions, the respective semistable loci are invariant. This gives rise to morphisms

$$(R_d^{\text{sst}} \times R_e^{\text{sst}}) \times^L G \leftarrow Z_{d,e}^{\text{sst}} \times^L G \rightarrow Z_{d,e}^{\text{sst}} \times^P G \rightarrow R_{d+e}^{\text{sst}} \times^P G \rightarrow R_{d+e}.$$

We see that:

- $(R_d^{\text{sst}} \times R_e^{\text{sst}}) \times^L G \leftarrow Z_{d,e}^{\text{sst}} \times^L G$ is a G -equivariant (trivial) vector bundle.
- $Z_{d,e}^{\text{sst}} \times^L G \rightarrow Z_{d,e}^{\text{sst}} \times^P G$ is a fibration whose fiber P/L is an affine space (as L is the Levi of P in G) and thus induces an isomorphism in G -equivariant intersection theory.
- $Z_{d,e}^{\text{sst}} \times^P G \rightarrow R_{d+e}^{\text{sst}} \times^P G$ is a G -equivariant regular embedding of relative dimension $s_1 = \sum_{\alpha:i \rightarrow j} d_i e_j$.
- $R_{d+e}^{\text{sst}} \times^P G \rightarrow R_{d+e}$ is proper as G/P is complete, and the dimension of G/P is $s_0 = \sum_i d_i e_i$.

The above morphisms give rise to maps in equivariant intersection theory

$$A_L^n(R_d^{\text{sst}} \times R_e^{\text{sst}}) \xrightarrow{\sim} A_L^n(Z_{d,e}^{\text{sst}}) \xleftarrow{\sim} A_P^n(Z_{d,e}^{\text{sst}}) \rightarrow A_P^{n+s_1}(R_{d+e}^{\text{sst}}) \rightarrow A_G^{n+s_1-s_0}(R_{d+e}^{\text{sst}}).$$

Note that $s_1 - s_0$ equals $-\chi(d, e)$, the negative of the Euler form of d and e . Composing with the equivariant exterior product map $A_{G_d}^*(R_d^{\text{sst}}) \otimes A_{G_e}^*(R_e^{\text{sst}}) \rightarrow A_L^*(R_d^{\text{sst}} \times R_e^{\text{sst}})$, we obtain a linear map

$$\mathcal{A}_d^{\text{sst}} \otimes \mathcal{A}_e^{\text{sst}} \rightarrow \mathcal{A}_{d+e}^{\text{sst}}.$$

The proof of [Kontsevich and Soibelman 2011, Theorem 1] also shows that we thus obtain an associative $\Gamma^{\theta, \mu}$ -graded algebra. In analogy to Kontsevich and Soibelman’s terminology, we define:

Definition 4.1. The algebra $\mathcal{A}^{\theta\text{-sst}, \mu}(Q)$ is called the θ -semistable *Chow–Hall algebra* (ChowHa) of slope μ of Q .

For the special case that θ is zero (i.e., $R_d^{\text{sst}} = R_d$) and $\mu = 0$, we write \mathcal{A} instead of $\mathcal{A}^{0\text{-sst}, 0}$ and call it the *ChowHa* of Q .

5. ChowHa vs. CoHa

We discuss the relation between the semistable ChowHa and Kontsevich and Soibelman’s semistable CoHa. Let k be the field of complex numbers. There is an equivariant analog (see [Edidin and Graham

1998, 2.8]) of the cycle map from [Fulton 1984, Chapter 19]. Concretely, there is a homomorphism of rings which doubles degrees

$$A_{G_d}^*(R_d^{\theta\text{-sst}}) \rightarrow H_{G_d}^*(R_d^{\theta\text{-sst}})$$

for every stability condition θ and every dimension vector d .

Theorem 5.1. *The equivariant cycle map $A_{G_d}^*(R_d^{\theta\text{-sst}}) \rightarrow H_{G_d}^*(R_d^{\theta\text{-sst}})$ is an isomorphism. In particular, $R_d^{\theta\text{-sst}}$ has no odd-dimensional G_d -equivariant cohomology.*

This theorem generalizes a result due to King and Walter [1995, Theorem 3(c)]. They show the above assertion for an acyclic quiver, an indivisible dimension vector and a stability condition for which stability and semistability coincide. In this case, there exists a geometric PG_d -quotient $R_d^{\theta\text{-(s)st}} \rightarrow M_d^\theta$, and the G_d -equivariant Chow and cohomology groups agree with the tensor product of the ordinary Chow and cohomology groups of the quotient with a polynomial ring $\mathbb{Q}[z] \cong A_{\mathbb{G}_m}^*(\text{pt}) \cong H_{\mathbb{G}_m}^*(\text{pt})$.

We need a general lemma in order to prove Theorem 5.1. Let X be a complex algebraic scheme embedded into a nonsingular variety \bar{X} of complex dimension N ; for the rest of this section, a *scheme* will be a complex algebraic scheme (see [Fulton 1984, B.1.1]) which admits such an embedding. The Borel–Moore homology $H_k(X)$ is isomorphic to the singular cohomology $H^{2N-k}(\bar{X}, \bar{X} - X)$. We consider the cycle map $\text{cl} : A_k(X) \rightarrow H_{2k}(X)$.

Lemma 5.2. *Let X be a scheme, Y a closed subscheme of X and U the open complement:*

- (1) *If both Y and U have no odd-dimensional homology then X does not have odd-dimensional homology.*
- (2) *If $A_k(Y) \rightarrow H_{2k}(Y)$ and $A_k(U) \rightarrow H_{2k}(U)$ are isomorphisms and $H_{2k+1}(U) = 0$ then $A_k(X) \rightarrow H_{2k}(X)$ is an isomorphism.*
- (3) *Suppose that $A_k(Y) \rightarrow H_{2k}(Y)$ and $A_k(X) \rightarrow H_{2k}(X)$ are isomorphisms and $H_{2k-1}(Y) = 0$. Then $A_k(U) \rightarrow H_{2k}(U)$ is also an isomorphism.*

Proof. The first assertion is clear by the long exact sequence in homology. To prove the second statement, we consider the diagram:

$$\begin{array}{ccccccc} A_k(Y) & \longrightarrow & A_k(X) & \longrightarrow & A_k(U) & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \\ \cdots & \longrightarrow & H_{2k+1}(U) & \longrightarrow & H_{2k}(Y) & \longrightarrow & H_{2k}(X) & \longrightarrow & H_{2k}(U) & \longrightarrow & H_{2k-1}(Y) & \longrightarrow & \cdots \end{array}$$

The left and right vertical maps being isomorphisms and the left-most term in the lower row being zero by assumption, the claim follows by applying the snake lemma. The third claim follows by a diagram chase in the same diagram. We give the proof for completeness. Let $u \in H_{2k}(U)$. There exists $x \in H_{2k}(X)$ with $j^*x = u$ where $j : U \rightarrow X$ is the open embedding. We find a unique $\xi \in A_k(X)$ with $\text{cl}_X \xi = x$, so $u = j^* \text{cl}_X \xi = \text{cl}_U j^* \xi$ which proves the surjectivity of cl_U . To show that cl_U is injective, let $v \in A_k(U)$ with $\text{cl}_U v = 0$. For an inverse image $\xi \in A_k(X)$ of v under j^* , we obtain $j^* \text{cl}_X \xi = 0$, whence there exists $y \in H_{2k}(Y)$ such that $i_*y = \text{cl}_X \xi$. Here $i : Y \rightarrow X$ denotes the closed immersion. Let $\eta \in A_k(Y)$

be the unique cycle with $\text{cl}_Y \eta = y$. We get $\text{cl}_X i_* \eta = i_* y = \text{cl}_X \xi$ and thus $i_* \eta = \xi$ by injectivity of cl_X . This implies $v = j^* i_* \eta = 0$. □

An immediate consequence of the above lemma is the following:

Lemma 5.3. *Suppose that a scheme X has a filtration $X = X_N \supseteq \cdots \supseteq X_1 \supseteq X_0 = \emptyset$ by closed subschemes such that the cycle map for the successive complements $S_i = X_i - X_{i-1}$ is an isomorphism for all i . Then cl_X is an isomorphism and, moreover, we have noncanonical isomorphisms*

$$A_*(X) \cong \bigoplus_{i=0}^N A_*(S_i) \quad \text{and} \quad A^*(X) \cong \bigoplus_{i=0}^N A^{*-\text{codim}_X S_i}(S_i)$$

by choosing sections of the surjections $A_*(X_i) \rightarrow A_*(S_i)$ for all i .

We now turn to an equivariant setup. Let G be a reductive linear algebraic group acting on a scheme X of complex dimension n . For an index i , we choose a representation V of G and an open subset $E \subseteq V$ such that a principal bundle quotient E/G exists and such that $\text{codim}_V(V - E) > n - i$. Then, the group

$$A_k^G(X) = A_{i+\dim V - \dim G}(X \times^G E)$$

is independent of the choice of E and V . In the same vein, for an index j with $2 \text{codim}_V(V - E) > 2n - j$, we can define equivariant Borel–Moore homology via ordinary Borel–Moore homology, namely

$$H_j^G(X) = H_{j+2 \dim V - 2 \dim G}(X \times^G E)$$

(see [Edidin and Graham 1998]). If X is smooth then $H_j^G(X)$ is dual to $H_j^{2n-j}(X \times^G E)$ which is isomorphic to $H_j^{2n-j}(X \times^G EG) = H_G^{2n-j}(X)$ (where EG is the classifying space for G).

We consider the equivariant cycle map $\text{cl} : A_k^G(X) \rightarrow H_{2k}^G(X)$ which is defined as the ordinary cycle map $\text{cl} : A_{i+\dim V - \dim G}(X \times^G E) \rightarrow H_{2k+2 \dim V - 2 \dim G}(X \times^G E)$ (again independent of $E \subseteq V$). For complementary open and closed subschemes U and Y of X which are G -invariant, we choose $E \subseteq V$ such that the principal bundle quotient E/G exists and $\text{codim}_V(V - E) > n - i$ (note that all the equivariant versions of the groups appearing in the diagram in the proof of Lemma 5.2 can be defined using E) and apply Lemma 5.2 to the complementary open/closed subschemes $U \times^G E$ and $Y \times^G E$. We thus obtain:

Corollary 5.4. *In the above equivariant situation, Lemmas 5.2 and 5.3 hold for equivariant Chow and Borel–Moore homology groups.*

Proof of Theorem 5.1. We prove Theorem 5.1 in two steps:

- (1) Prove that the odd-dimensional equivariant cohomology of the θ -semistable locus vanishes by reducing the arbitrary case to a situation where stability and semistability agree. There, the statement is known thanks to Reineke [2003].
- (2) Prove that the equivariant cycle map $A_k^{G_d}(R_d^{\theta\text{-sst}}) \rightarrow H_{2k}^{G_d}(R_d^{\theta\text{-sst}})$ is an isomorphism by induction over the Harder–Narasimhan strata.

Step 1: First, assume that d is a θ -coprime dimension vector. This means that there is no subdimension vector $0 \neq d' \leq d$ with the same slope as d apart from d itself. In this case, θ -semistability and θ -stability on R_d agree and there exists a smooth geometric PG_d -quotient $R_d^{\theta-(s)\text{st}} \rightarrow M_d^\theta$. Here, $\text{PG}_d = G_d/\mathbb{C}^\times$. In [Reineke 2003, Theorem 6.7] it is shown that the odd-dimensional cohomology of M_d^θ vanishes. But as by the existence of a geometric quotient

$$H_{G_d}^*(R_d^{\theta-\text{sst}}) \cong H^*(M_d^\theta) \otimes H_{\mathbb{C}^\times}^*(\text{pt}),$$

it follows that $R_d^{\theta-\text{sst}}$ has no odd-dimensional cohomology.

Now, let d be arbitrary. We show that for a fixed — not necessarily positive — integer k , there exist a quiver \hat{Q} , a stability condition $\hat{\theta}$, a $\hat{\theta}$ -coprime dimension vector \hat{d} , and an integer $s \geq 0$ (all depending on k) for which

$$H_k^{G_d}(R_d^{\theta-\text{sst}}) \cong H_{k+2s}^{\text{PG}_{\hat{d}}}(R_{\hat{d}}^{\hat{\theta}-(s)\text{st}}(\hat{Q})). \tag{1}$$

For a dimension vector n of Q , we consider $\hat{R}_{d,n} = R_d \times F_n$ where $F_n = \bigoplus_i \text{Hom}(\mathbb{C}^{n_i}, \mathbb{C}^{d_i})$. The space $\hat{R}_{d,n}$ is the space of representations of the framed quiver \hat{Q} of dimension vector \hat{d} which arise as follows (see [Engel and Reineke 2009, Definition 3.1]): we add an extra vertex ∞ to the vertexes of Q , i.e., $\hat{Q}_0 = Q_0 \sqcup \{\infty\}$ and, in addition to the arrows of Q , we have n_i arrows from ∞ heading to i for all $i \in Q_0$. The dimension vector \hat{d} is defined by $\hat{d}_i = d_i$ for $i \in Q_0$ and $\hat{d}_\infty = 1$ and is indivisible. The structure group $G_{\hat{d}}$ is $\mathbb{C}^\times \times G_d$, whence we can identify $\text{PG}_{\hat{d}}$ with G_d . We define $\hat{\theta}$ in the same way as in [loc. cit., Definition 3.1]. The following are equivalent for a framed representation $(M, f) \in \hat{R}_{d,n}$ (see [loc. cit., Proposition 3.3]):

- (M, f) is $\hat{\theta}$ -semistable.
- (M, f) is $\hat{\theta}$ -stable.
- M is θ -semistable and the (θ) -slope of every proper subrepresentation M' of M which contains the image of f is strictly less than the slope of M .

We denote the set of $\hat{\theta}$ -(semi)stable points of $\hat{R}_{d,n}$ with $\hat{R}_{d,n}^\theta$. It is, by the above characterization, an open subset of $R_d^{\theta-\text{sst}} \times F_n$. Let $\hat{R}_{d,n}^x$ denote the complement of $\hat{R}_{d,n}^\theta$ inside $R_d^{\theta-\text{sst}} \times F_n$. As $R_d^{\theta-\text{sst}} \times F_n$ is a G_d -equivariant vector bundle over $R_d^{\theta-\text{sst}}$, we obtain

$$H_k^{G_d}(R_d^{\theta-\text{sst}}) \cong H_{k+2d \cdot n}^{G_d}(R_d^{\theta-\text{sst}} \times F_n)$$

where $d \cdot n := \sum_i d_i n_i = \dim_{\mathbb{C}} F_n$. We thus obtain a long exact sequence

$$\dots \rightarrow H_{k+2d \cdot n}^{G_d}(\hat{R}_{d,n}^x) \rightarrow H_k^{G_d}(R_d^{\theta-\text{sst}}) \rightarrow H_{k+2d \cdot n}^{G_d}(\hat{R}_{d,n}^\theta) \rightarrow H_{k-1+2d \cdot n}^{G_d}(\hat{R}_{d,n}^x) \rightarrow \dots$$

in equivariant Borel–Moore homology. The equivariant BM homology groups $H_l^{G_d}(\hat{R}_{d,n}^x)$ vanish if l exceeds $2 \dim \hat{R}_{d,n}^x$. So in order to show that (1) is an isomorphism, it suffices to find a framing datum n such that the (complex) dimension of $\hat{R}_{d,n}^x$ is smaller than $(k - 1)/2 + d \cdot n$. As shown in the proof of

[Franzen 2018, Theorem 3.2], $\hat{R}_{d,n}^x$ is the union of Harder–Narasimhan strata

$$\hat{R}_{d,n}^x = \bigsqcup \hat{R}_{(\hat{p},q),n}^{\text{HN}}$$

over all proper subdimension vectors p of d which have the same slope (and $q = d - p$). The set $\hat{R}_{(\hat{p},q),n}^{\text{HN}}$ is defined as follows: let $L(M, f)$ be minimal among those representations of the same slope as M which contain $\text{im } f$. We set $\hat{R}_{(\hat{p},q),n}^{\text{HN}}$ as the set of all $(M, f) \in R_d^{\theta-\text{sst}} \times F_n$ with $\underline{\dim} L(M, f) = p$. As

$$\hat{R}_{(\hat{p},q),n}^{\text{HN}} \cong \left(\left(\begin{pmatrix} R_p^{\text{sst}} & * \\ & R_q^{\text{sst}} \end{pmatrix} \times \begin{pmatrix} F_p \\ 0 \end{pmatrix} \right) \times^{P_{p,q}} G_d,$$

the dimension of this stratum — if nonempty — equals

$$\sum_{\alpha:i \rightarrow j} (d_i d_j - p_i q_j) + \sum_i p_i n_i - \sum_i (d_i^2 - p_i q_i) + \sum_i d_i^2 = \dim(R_d) + d \cdot n + \chi(p, q) - q \cdot n.$$

Choosing n large enough such that

$$q \cdot n > \dim(R_d) - \frac{1}{2}(k - 1) + \chi(d - q, q)$$

for all subdimension vectors $0 \neq q \leq d$ of the same slope as d (which is possible as these are finitely many nonzero dimension vectors q), we find that the dimension of $\hat{R}_{d,n}^x$ is smaller than $(k - 1)/2 + d \cdot n$, as desired.

Similar arguments were also used by Davison and Meinhardt [2016, Lemma 4.1].

Step 2: Let Q, θ and d be arbitrary. We consider the open and closed complementary subsets R_d^{sst} and R_d^{unst} . As R_d^{sst} is smooth (of dimension $n = \sum_{\alpha:i \rightarrow j} d_i d_j$), we have $A_i^{G_d}(R_d^{\text{sst}}) \cong A_{G_d}^{n-i}(R_d^{\text{sst}})$ and $H_j^{G_d}(R_d^{\text{sst}}) \cong H_{G_d}^{2n-j}(R_d^{\text{sst}})$. By Corollary 5.4 (concretely, the equivariant analog of part (3) of Lemma 5.2), it suffices to show that

$$\text{cl} : A_*^{G_d}(R_d^{\text{unst}}) \rightarrow H_{2*}^{G_d}(R_d^{\text{unst}})$$

is an isomorphism. If $R_d^{\text{unst}} = \emptyset$, then the assertion is clear. So let us assume that R_d^{unst} is nonempty. The unstable locus admits a stratification into locally closed (irreducible) subsets $R_{d^*}^{\text{HN}}$, the Harder–Narasimhan strata, by [Reineke 2003, Proposition 3.4]. By [loc. cit., Proposition 3.7], they can be ordered in such a way that the union of the first n strata is closed for all n , thus yielding a filtration by G_d -invariant closed subsets like in Lemma 5.3. Thus it suffices to prove that

$$A_*^{G_d}(R_{d^*}^{\text{HN}}) \rightarrow H_*^{G_d}(R_{d^*}^{\text{HN}})$$

is an isomorphism for all HN types $d^* = (d^1, \dots, d^l)$ of d ; this includes showing that all HN strata have even cohomology. But by the proof of [loc. cit., Proposition 3.4], we have

$$R_{d^*}^{\text{HN}} \cong Z_{d^*}^{\text{sst}} \times^{P_{d^*}} G_d,$$

where $Z_{d^*}^{\text{sst}}$ is a (trivial) vector bundle over $R_{d^1}^{\text{sst}} \times \dots \times R_{d^l}^{\text{sst}}$, and P_{d^*} is a parabolic subgroup of G_d with Levi $G_{d^1} \times \dots \times G_{d^l}$.

In particular, $R_{d^*}^{\text{HN}}$ is smooth, therefore we can again identify Borel–Moore homology with cohomology. Moreover, $A_{G_d}^i(R_{d^*}^{\text{HN}}) \cong A_{G_{d^1} \times \dots \times G_{d^l}}^i(R_{d^1}^{\text{sst}} \times \dots \times R_{d^l}^{\text{sst}})$, and similarly for equivariant cohomology. This shows in particular that the equivariant odd-dimensional cohomology of $R_{d^*}^{\text{HN}}$ vanishes by the first step of the proof. Now we argue by induction on the dimension vector d , where the set of dimension vectors is partially ordered by $d \leq e$ if $d_i \leq e_i$ for all i ; with respect to this order, all d^ν 's are strictly smaller than d . We thus assume that the equivariant cycle map for each $R_{d^\nu}^{\text{sst}}$ is an isomorphism. Then, [Totaro 1999, Lemma 6.2], which can be generalized to equivariant Chow groups, implies that the equivariant exterior product map

$$A_{G_{d^1}}^*(R_{d^1}^{\text{sst}}) \otimes \dots \otimes A_{G_{d^l}}^*(R_{d^l}^{\text{sst}}) \rightarrow A_{G_{d^1} \times \dots \times G_{d^l}}^*(R_{d^1}^{\text{sst}} \times \dots \times R_{d^l}^{\text{sst}})$$

is an isomorphism (even with integral coefficients). As the $R_{d^\nu}^{\text{sst}}$'s have even cohomology, the Künneth map is an isomorphism. We are thus reduced to proving the assertion for minimal dimension vectors d , i.e., $d = 0$. But there the statement is obviously true. □

Remark 5.5. Theorem 5.1 is valid for integer coefficients.

Corollary 5.6. *The cycle map induces an isomorphism $\mathcal{A}^{\text{sst},\mu} \rightarrow \mathcal{H}^{\text{sst},\mu}$ of algebras.*

Proof. The multiplication both in the semistable ChowHa and in the semistable CoHa $\mathcal{H}^{\text{sst},\mu}$ are constructed by means of the same Hecke correspondences. Moreover, the cycle map is compatible with push-forward and pull-back. □

6. Tensor product decomposition

We apply Corollary 5.4 to the Harder–Narasimhan filtration, like in the proof of Theorem 5.1. There exists a filtration $R_d = X_N \supseteq \dots \supseteq X_1 \supseteq X_0 = 0$ by closed subsets such that each of the successive complements $S_i := X_i - X_{i-1}$ equals one of the Harder–Narasimhan strata $R_{d^*}^{\text{HN}}$ (and vice versa). We have argued in the proof of Theorem 5.1 that the push-forward $A_*^{G_d}(X_i) \rightarrow A_*^{G_d}(R_d)$ is injective for every i . Any choice of sections $\sigma_i : A_*^{G_d}(S_i) \rightarrow A_*^{G_d}(X_i)$ of the surjections $A_*^{G_d}(X_i) \rightarrow A_*^{G_d}(S_i)$ gives rise to injections $\tilde{\sigma}_i : A_*^{G_d}(S_i) \rightarrow A_*^{G_d}(R_d)$ and yields an isomorphism

$$\bigoplus_{d^*} A_*^{G_d}(R_{d^*}^{\text{HN}}) = \bigoplus_{i=1}^N A_*^{G_d}(S_i) \xrightarrow{\sim} A_*^{G_d}(R_d).$$

For a Harder–Narasimhan type d^* we denote the inclusion $A_*^{G_d}(R_{d^*}^{\text{HN}}) \rightarrow A_*^{G_d}(R_d)$ with $\tilde{\sigma}_{d^*}$. Using the cohomological grading of the Chow groups, we get

$$\tilde{\sigma}_{d^*} : A_{G_d}^{*-\text{codim}_{R_d}(R_{d^*}^{\text{HN}})}(R_{d^*}^{\text{HN}}) \rightarrow A_{G_d}^*(R_d).$$

The codimension of $R_{d^*}^{\text{HN}}$ in R_d can easily be computed as $\chi(d^*) := \sum_{r < s} \chi(d^r, d^s)$. Recall that the Harder–Narasimhan stratum $R_{d^*}^{\text{HN}}$ is isomorphic to $Z_{d^*}^{\text{sst}} \times^{P_{d^*}} G_d$, where $Z_{d^*}^{\text{sst}}$ is the inverse image of

$R_{d^1}^{\text{sst}} \times \cdots \times R_{d^l}^{\text{sst}}$ under the projection map of the P_{d^*} -equivariant vector bundle

$$Z_{d^*} = \begin{pmatrix} R_{d^1} & & * \\ & \ddots & \\ & & R_{d^l} \end{pmatrix} \rightarrow R_{d^1} \times \cdots \times R_{d^l}.$$

As the map $\pi : Z_{d^*} \times^{P_{d^*}} G_d \rightarrow R_d$ is proper its image is closed (and irreducible) in R_d and contains $R_{d^*}^{\text{HN}}$ as an open subset. This implies that the image of π agrees with the Zariski closure of $R_{d^*}^{\text{HN}}$. Let i be an index such that $R_{d^*}^{\text{HN}} = S_i$ in the above filtration. We summarize the situation in the following diagram:

$$\begin{array}{ccccc} R_{d^*}^{\text{HN}} & \subseteq & \overline{R_{d^*}^{\text{HN}}} & \subseteq & X_i & \subseteq & R_d \\ \cong \uparrow & & \tilde{\pi} \uparrow & & \nearrow \pi & & \\ Z_{d^*}^{\text{sst}} \times^{P_{d^*}} G_d & \subseteq & Z_{d^*} \times^{P_{d^*}} G_d & & & & \end{array}$$

The square on the left-hand side is cartesian with open inclusions. The other inclusions are closed embeddings. We pass to Chow groups. This yields the following commutative diagram

$$\begin{array}{ccccccc} A_*^{G_d}(R_{d^*}^{\text{HN}}) & \longleftarrow & A_*^{G_d}(\overline{R_{d^*}^{\text{HN}}}) & \longrightarrow & A_*^{G_d}(X_i) & \longrightarrow & A_*^{G_d}(R_d) \\ \cong \uparrow & & \tilde{\pi}_* \uparrow & & \nearrow \pi_* & & \\ A_*^{G_d}(Z_{d^*}^{\text{sst}} \times^{P_{d^*}} G_d) & \longleftarrow & A_*^{G_d}(Z_{d^*} \times^{P_{d^*}} G_d) & & & & \end{array}$$

Now we use the two natural isomorphisms $A_{G_d}^*(Z_{d^*} \times^{P_{d^*}} G_d) \cong A_{G_{d^1}}^*(R_{d^1}) \otimes \cdots \otimes A_{G_{d^l}}^*(R_{d^l})$ and $A_{G_d}^*(Z_{d^*}^{\text{sst}} \times^{P_{d^*}} G_d) \cong A_{G_{d^1}}^*(R_{d^1}^{\text{sst}}) \otimes \cdots \otimes A_{G_{d^l}}^*(R_{d^l}^{\text{sst}})$. We work with the cohomological grading here to avoid having to take degree shifts into account. Let us choose a section of the pull-back of each of the open embeddings $R_{d^v}^{\text{sst}} \subseteq R_{d^v}$. Let

$$\tau_{d^*} : A_{G_{d^1}}^*(R_{d^1}^{\text{sst}}) \otimes \cdots \otimes A_{G_{d^l}}^*(R_{d^l}^{\text{sst}}) \rightarrow A_{G_{d^1}}^*(R_{d^1}) \otimes \cdots \otimes A_{G_{d^l}}^*(R_{d^l})$$

be the resulting section. We get a section $A_{G_d}^*(Z_{d^*}^{\text{sst}} \times^{P_{d^*}} G_d) \rightarrow A_{G_d}^*(Z_{d^*} \times^{P_{d^*}} G_d)$ which we, by abuse of notation, also call τ_{d^*} . The composition

$$s_{d^*} : A_{G_d}^*(R_{d^*}^{\text{HN}}) \xrightarrow{\sim} A_{G_d}^*(Z_{d^*}^{\text{sst}} \times^{P_{d^*}} G_d) \xrightarrow{\tau_{d^*}} A_{G_d}^*(Z_{d^*} \times^{P_{d^*}} G_d) \xrightarrow{\tilde{\pi}_*} A_{G_d}^*(\overline{R_{d^*}^{\text{HN}}})$$

is then a section of the pull-back $A_{G_d}^*(\overline{R_{d^*}^{\text{HN}}}) \rightarrow A_{G_d}^*(R_{d^*}^{\text{HN}})$ because of the following:

Lemma 6.1. *Let G be an algebraic group and let*

$$\begin{array}{ccc} X' & \xrightarrow{j'} & X \\ \downarrow f' & & \downarrow f \\ Y' & \xrightarrow{j} & Y \end{array}$$

be a cartesian square of G -schemes, where j is an open embedding, f is a proper morphism and f' is an isomorphism. Let $s' : A_*^G(Y') \rightarrow A_*^G(Y)$ be a section of the open pull-back $j'^* : A_*^G(X) \rightarrow A_*^G(X')$. Then the composition $s := f_* s' f'^{-1}$ is a section of $j^* : A_*^G(Y) \rightarrow A_*^G(Y')$.

Proof. Passing to equivariant Chow groups, we obtain a commutative square

$$\begin{array}{ccc} A_*^G(X') & \xleftarrow{j'^*} & A_*^G(X) \\ \downarrow f'_* & & \downarrow f_* \\ A_*^G(Y') & \xleftarrow{j^*} & A_*^G(Y) \end{array}$$

in which f'_* is an isomorphism. Let $\alpha \in A_*^G(Y')$. We compute

$$j^* s(\alpha) = j^* f_* s' f'^{-1}(\alpha) = f'_* j'^* s' f'^{-1}(\alpha) = f'_* f'^{-1}(\alpha) = \alpha.$$

This proves the lemma. □

Now we apply the above lemma again, this time to the cartesian diagram

$$\begin{array}{ccc} R_{d^*}^{\text{HN}} & \longrightarrow & \overline{R_{d^*}^{\text{HN}}} \\ \parallel & & \downarrow \\ S_i & \longrightarrow & X_i \end{array}$$

— the vertical map being the closed immersion, and the horizontal maps the open embeddings — and to $s' = s_{d^*}$. The composition

$$\sigma_i : A_*^{G_d}(S_i) = A_*^{G_d}(R_{d^*}^{\text{HN}}) \xrightarrow{s_{d^*}} A_*^{G_d}(\overline{R_{d^*}^{\text{HN}}}) \rightarrow A_*^{G_d}(X_i)$$

is hence a section of $A_*^{G_d}(X_i) \rightarrow A_*^{G_d}(S_i)$. If we now form the inclusions $\tilde{\sigma}_{d^*} : A_*^{G_d}(R_{d^*}^{\text{HN}}) \rightarrow A_*^{G_d}(R_d)$ as described at the beginning of this section we have ensured that we obtain a commutative diagram

$$\begin{array}{ccc} A_{G_{d^1}}^*(R_{d^1}^{\text{sst}}) \otimes \cdots \otimes A_{G_{d^l}}^*(R_{d^l}^{\text{sst}}) & \xrightarrow{\cong} & A_{G_d}^*(R_d^{\text{HN}}) \\ \downarrow \tau_{d^*} & & \downarrow \tilde{\sigma}_{d^*} \\ A_{G_{d^1}}^*(R_{d^1}) \otimes \cdots \otimes A_{G_{d^l}}^*(R_{d^l}) & \longrightarrow & A_{G_d}^{*+\chi(d^*)}(R_d) \end{array}$$

in which the lower horizontal map is the ChowHa multiplication. We define the descending tensor product $\bigotimes_{\mu \in \mathbb{Q}}^{\leftarrow} \mathcal{A}^{\text{sst}, \mu}$ as the Γ -graded vector space

$$\bigoplus_d \bigoplus_{d^*} A_{G_{d^1}}^*(R_{d^1}^{\text{sst}}) \otimes \cdots \otimes A_{G_{d^l}}^*(R_{d^l}^{\text{sst}})$$

where the inner sum ranges over all Harder–Narasimhan types d^* summing to d (i.e., tuples $d^* = (d^1, \dots, d^l)$ of dimension vectors of slopes $\mu^1 > \cdots > \mu^l$ such that $d^1 + \cdots + d^l = d$). The above considerations then prove:

Theorem 6.2. *The ChowHa multiplication induces an isomorphism*

$$\bigotimes_{\mu \in \mathbb{Q}}^{\leftarrow} \mathcal{A}^{\theta - \text{sst}, \mu} \xrightarrow{\sim} \mathcal{A}$$

of Γ -graded vector spaces between the descending tensor product of the θ -semistable ChowHa's over all possible slopes and the ChowHa.

Remark 6.3. The theorem is valid with integral coefficients, for an arbitrary quiver, and does not require the stability condition to be generic. Theorem 6.2 has been proved with different methods by Rimányi [2013] for the CoHa of a Dynkin quiver which is not an orientation of E_8 .

7. Structure of the CoHa of a symmetric quiver

The CoHa and ChowHa of a quiver are described explicitly in [Kontsevich and Soibelman 2011]. Since we will make use of this description, we recall it here: The equivariant Chow ring $A_{G_d}^*(R_d) \cong A_{G_d}^*(\text{pt})$ is isomorphic to

$$\mathbb{Q}[x_{i,r} \mid i \in Q_0, 1 \leq r \leq d_i]^{W_d},$$

where $W_d = \prod_i S_{d_i}$ is the Weyl group of a maximal torus of G_d . We may regard the variables $x_{i,r}$ (located in degree 1) as a basis for the character group of this torus or as the Chern roots of the G_d -linear vector bundle $R_d \times k^{d_i} \rightarrow R_d$ with G_d acting on k^{d_i} by its i -th factor.

Theorem 7.1 [Kontsevich and Soibelman 2011, Theorem 2]. *For $f \in \mathcal{A}_d$ and $g \in \mathcal{A}_e$, the product $f * g$ equals the function*

$$\sum f(x_{i,\sigma_i(r)} \mid i, 1 \leq r \leq d_i) \cdot g(x_{i,\sigma_i(d_i+s)} \mid i, 1 \leq s \leq e_i) \cdot \prod_{i,j \in Q_0} \prod_{r=1}^{d_i} \prod_{s=1}^{e_j} (x_{j,\sigma_j(d_j+s)} - x_{i,\sigma_i(r)})^{a_{i,j} - \delta_{i,j}}.$$

The sum ranges over all (d, e) -shuffles $\sigma = (\sigma_i \mid i) \in W_{d+e}$, that means each σ_i is a (d_i, e_i) -shuffle permutation.

We assume that the stability condition θ is μ -generic. In this case, we can equip the semistable ChowHa of slope μ with a refined grading: setting

$$\mathcal{A}_{(d,n)}^{\text{sst}} = \begin{cases} A_{G_d}^{(n - \chi(d,d))/2}(R_d^{\text{sst}}) & n \equiv \chi(d, d) \pmod{2}, \\ 0 & n \not\equiv \chi(d, d) \pmod{2}, \end{cases}$$

it is easy to see that the multiplication map becomes bigraded, thus $\mathcal{A}_{(d,n)}^{\text{sst}} \otimes \mathcal{A}_{(e,m)}^{\text{sst}} \rightarrow \mathcal{A}_{(d+e,n+m)}^{\text{sst}}$.

Like in Section 3, we consider again the case of a symmetric quiver and the trivial stability condition. In this situation, it is immediate from the formula in the above theorem that $f * g = (-1)^{\chi(d,e)} g * f$ for $f \in \mathcal{A}_d$ and $g \in \mathcal{A}_e$. One can show (see [Kontsevich and Soibelman 2011, §2.6]) that there exists a bilinear form ψ on the $\mathbb{Z}/2\mathbb{Z}$ -vector space $(\mathbb{Z}/2\mathbb{Z})^{Q_0}$ such that $f \star g = (-1)^{\psi(d,e)} f * g$ is a super-commutative

multiplication, when defining the parity of an element of bidegree (d, n) to be the parity of n . We see that the generating series $P(q, t) = \sum_d \sum_k (-1)^k \dim \mathcal{A}_{(d,k)} q^{k/2} t^d$ is

$$\sum_d (-q^{1/2})^{\chi(d,d)} \prod_i \prod_{v=1}^{d_i} (1 - q^v)^{-1} t^d.$$

So, $P(q, t) = A(q^{-1}, t)$. By Theorem 3.2, the generating series has a product expansion

$$P(q, t) = \prod_d \prod_k \prod_{n \geq 0} (1 - q^{n+k/2} t^d)^{(-1)^{k-1} \Omega_{d,k}}.$$

As a free super-commutative algebra with a generator in bidegree (d, k) has the generating series $(1 - q^{k/2} t^d)^{(-1)^{k-1}} = \text{Exp}((-1)^k q^{k/2} t^d)$, Kontsevich and Soibelman [2011] made a conjecture which was eventually proved by Efimov.

Theorem 7.2 [Efimov 2012, Theorem 1.1]. *For a symmetric quiver Q , the algebra $\mathcal{A}(Q)$, equipped with the super-commutative multiplication \star , is isomorphic to a free super-commutative algebra over a $(\Gamma \times \mathbb{Z})$ -graded vector space $V = V^{\text{prim}} \otimes \mathbb{Q}[z]$, where z lives in bidegree $(0, 2)$, and $\bigoplus_k V_{d,k}^{\text{prim}}$ is finite-dimensional for every d .*

This result implies that the Donaldson–Thomas invariants $\Omega_{d,k}$ must agree with the dimension of $V_{(d,k)}^{\text{prim}}$ and must therefore be nonnegative. We will give another characterization of the primitive part of the CoHa in Theorem 9.2.

8. Tautological presentation of the semistable ChowHa

We investigate the relation between the semistable ChowHa $\mathcal{A}^{\theta-\text{sst}, \mu}$ and the ChowHa \mathcal{A} of a quiver Q . For a dimension vector d of slope μ , we consider the open embedding

$$R_d^{\text{sst}} \rightarrow R_d$$

which gives rise to a surjective map $A_{G_d}^*(R_d) \rightarrow A_{G_d}^*(R_d^{\text{sst}})$. As the Hecke correspondences for the semistable ChowHa are given by restricting the Hecke correspondences of \mathcal{A} to the semistable loci, these open pull-backs are compatible with the multiplication, i.e., they induce a surjective homomorphism of Γ -graded algebras

$$\mathcal{A} \rightarrow \mathcal{A}^{\theta-\text{sst}, \mu}.$$

Here, we regard $\mathcal{A}^{\theta-\text{sst}, \mu}$ as a Γ -graded algebra by extending it trivially to every dimension vector whose slope is not s . We can describe the kernel explicitly.

Theorem 8.1. *The kernel of the natural map $\mathcal{A}_d \rightarrow \mathcal{A}_d^{\theta-\text{sst}}$ equals the sum*

$$\sum \mathcal{A}_p * \mathcal{A}_q$$

over all pairs (p, q) of dimension vectors of Q which sum to d and such that $\mu(p) > \mu(q)$.

The key ingredient of the proof of this result is a purely intersection-theoretic lemma. Following [Fulton 1984, B.1.1], we call a k -scheme algebraic if it is separated and of finite type over $\text{Spec } k$. Thus, a variety is an algebraic scheme which is integral.

Lemma 8.2. *Let $f : X \rightarrow Y$ be a surjective, proper morphism of algebraic k -schemes. Then the push-forward $f_* : A_*(X)_{\mathbb{Q}} \rightarrow A_*(Y)_{\mathbb{Q}}$ is surjective.*

Proof. It is obviously sufficient to prove that, for every dominant morphism $f : X \rightarrow Y$ of an algebraic scheme X to a variety Y , there exists a subvariety W of X of dimension $\dim W = \dim Y$ which dominates Y . This is a local statement, so we may assume X and Y to be affine, say $X = \text{Spec } B$ and $Y = \text{Spec } A$. The morphism f corresponds to an extension $A \hookrightarrow B$ of rings. We therefore need to show that there exists a prime ideal \mathfrak{q} of B with $\mathfrak{q} \cap A = (0)$ such that the induced extension

$$Q(B/\mathfrak{q}) \mid Q(A)$$

is finite. Let $K = Q(A)$ and $R = B \otimes_A K$. By Noether normalization, there exist $b_1, \dots, b_n \in R$, algebraically independent over K , such that $K[b_1, \dots, b_n] \subseteq R$ is a finite (and hence integral) ring-extension. Without loss of generality, we may assume $b_1, \dots, b_n \in B$. Choose a set of generators c_1, \dots, c_s of R as a $K[b_1, \dots, b_n]$ -algebra and polynomials $p_i(T) \in K[b_1, \dots, b_n][T]$ such that $p_i(c_i) = 0$. We find an element $s \in A - \{0\}$ such that the coefficients of all the p_i 's lie in $A_s[b_1, \dots, b_n]$ and $p_i(c_i) = 0$ holds in B_s . This implies that B_s is an integral $A_s[b_1, \dots, b_n]$ -algebra which yields the surjectivity of the map $\text{Spec } B_s \rightarrow \text{Spec } A_s[b_1, \dots, b_n]$. We consider the prime ideal $\mathfrak{p}' = (b_1, \dots, b_n)$ of $A_s[b_1, \dots, b_n]$ and find a prime ideal \mathfrak{q}' of B_s which lies above it. Then B_s/\mathfrak{q}' is an integral extension of $A_s[b_1, \dots, b_n]/\mathfrak{p}' = A_s$ and therefore, setting $\mathfrak{q} = \mathfrak{q}' \cap B$, the extension

$$Q(B/\mathfrak{q}) = Q(B_s/\mathfrak{q}') \mid Q(A_s) = Q(A)$$

is finite. □

Proof of Theorem 8.1. Let R_d^{unst} be the complement of R_d^{sst} in R_d . Then, we have an exact sequence

$$A_m^G(R_d^{\text{unst}}) \rightarrow A_m^G(R_d) \rightarrow A_m^G(R_d^{\text{sst}}) \rightarrow 0,$$

where $G = G_d$. For a decomposition $d = p + q$, let $R_{p,q}$ be the closed subset of R_d of all representations which possess a subrepresentation of dimension vector p . It is the G -saturation of $Z_{p,q}$. The G -action gives a surjective, proper morphism

$$Z_{p,q} \times^{P_{p,q}} G \rightarrow R_{p,q},$$

where $P_{p,q}$ is the parabolic $\begin{pmatrix} G_p & * \\ & G_q \end{pmatrix}$. The unstable locus R_d^{unst} equals the union $\bigcup R_{p,q}$ over all decompositions $d = p + q$ where the slope of p is larger than the slope of q . Let us call these decompositions θ -forbidden. We obtain, using Lemma 8.2 and [Fulton 1984, Example 1.3.1(c)], that the sequence

$$\bigoplus A_m^G(Z_{p,q} \times^{P_{p,q}} G) \rightarrow A_m^G(R_d) \rightarrow A_m^G(R_d^{\text{sst}}) \rightarrow 0$$

is exact when passing to rational coefficients — the direct sum being taken over all forbidden decompositions $d = p + q$. Setting $n = \dim R_d - m$, we identify

$$A_m^G(Z_{p,q} \times^{P_{p,q}} G) \cong A_{P_{p,q}}^{n+\chi(p,q)}(Z_{p,q}) \cong A_{G_p \times G_q}^{n+\chi(p,q)}(R_p \times R_q)$$

like in Section 4. As the equivariant product map $A_{G_p}^*(R_p) \otimes A_{G_q}^*(R_q) \rightarrow A_{G_p \times G_q}^*(R_p \times R_q)$ is an isomorphism (which is clear from the explicit description given above), we have shown that

$$\bigoplus_{\substack{p+q=d \\ \text{forbidden}}} \bigoplus_{k+l=n+\chi(p,q)} A_{G_p}^k(R_p)_{\mathbb{Q}} \otimes A_{G_q}^l(R_q)_{\mathbb{Q}} \rightarrow A_{G_d}^n(R_d)_{\mathbb{Q}} \rightarrow A_{G_d}^n(R_d^{\text{sst}})_{\mathbb{Q}} \rightarrow 0$$

is an exact sequence. The first map in this sequence is precisely the ChowHa-multiplication. This proves the theorem. \square

9. The primitive part of the semistable ChowHa

As a next step, we analyze the kernel of the pull-back $A_G^*(R_d^{\text{sst}}) \rightarrow A_G^*(R_d^{\text{st}})$ induced by the open embedding of the stable locus into the semistable locus. A semistable representation $M \in R_d$ is not stable if and only if there exists a proper subrepresentation of the same slope. For a decomposition $d = p + q$ into subdimension vectors of the same slope, we define $R_{p,q}^{\text{sst}}$ as the subset of those $M \in R_d^{\text{sst}}$ which admit a subrepresentation of dimension vector p . Therefore, the set of properly semistable representations is the union

$$R_d^{\text{sst}} - R_d^{\text{st}} = \bigcup R_{p,q}^{\text{sst}}$$

over all decompositions $d = p + q$ such that p and q have the same slope and are both nonzero. We have, yet again, a surjective, proper morphism

$$\begin{pmatrix} R_p^{\text{sst}} & * \\ & R_q^{\text{sst}} \end{pmatrix} \rightarrow R_{p,q}^{\text{sst}}.$$

A result of Totaro [1999, Lemma 6.1], which can easily be transferred to equivariant Chow rings, shows that the exterior product $A_{G_p}^*(R_p^{\text{sst}}) \otimes A_{G_q}^*(R_q^{\text{sst}}) \rightarrow A_{G_p \times G_q}^*(R_p^{\text{sst}} \times R_q^{\text{sst}})$ is an isomorphism. Following the arguments of the proof of Theorem 8.1, we obtain:

Theorem 9.1. *The kernel of the surjection $\mathcal{A}_d^{\theta-\text{sst}} \rightarrow \mathcal{A}_d^{\theta-\text{st}}$ is the sum*

$$\sum \mathcal{A}_p^{\theta-\text{sst}} * \mathcal{A}_q^{\theta-\text{sst}}$$

over all decompositions $d = p + q$ into nonzero subdimension vectors of the same θ -slope.

In other words, the graded vector space $\mathcal{A}_d^{\theta-\text{st},\mu} = \bigoplus_{d \in \Gamma^{\theta,\mu}} \mathcal{A}_d^{\theta-\text{st}}$ equipped with the trivial multiplication (by which we mean that the product of two homogeneous elements of positive degree is set to be zero) is isomorphic to the quotient $\mathcal{A}_+^{\theta-\text{sst},\mu} / (\mathcal{A}_+^{\theta-\text{sst},\mu} * \mathcal{A}_+^{\theta-\text{sst},\mu})$ of the semistable ChowHa modulo the square of its augmentation ideal $\mathcal{A}_+^{\theta-\text{sst}}$.

Again, we consider the case of a symmetric quiver Q . We have deduced from Theorem 8.1 that $\mathcal{A}^{\theta\text{-sst},\mu}$ is free super-commutative over $V^{\theta,\mu} = \bigoplus_{d \in \Gamma^{\theta,\mu}} V_d$. The quotient of the augmentation ideal of a free super-commutative algebra by its square is isomorphic to the primitive part of the algebra, i.e., in our case

$$V_d \cong \mathcal{A}_d^{\theta\text{-st}} = A_{G_d}^*(R_d^{\theta\text{-st}}) \cong A_{\text{PG}_d}^*(R_d^{\theta\text{-st}}) \otimes A_{\mathbb{G}_m}^*(\text{pt})$$

for every $d \neq 0$. As $V_d = V_d^{\text{prim}} \otimes \mathbb{Q}[z]$, we deduce that

$$V_{d,k}^{\text{prim}} = \begin{cases} A_{\text{PG}_d}^{(k-\chi(d,d))/2}(R_d^{\text{st}}) & k \equiv \chi(d, d) \pmod{2}, \\ 0 & k \not\equiv \chi(d, d) \pmod{2}. \end{cases}$$

Assuming that $R_d^{\theta\text{-st}}$ is nonempty and denoting by $M_d^{\theta\text{-st}}$ the geometric quotient $R_d^{\theta\text{-st}}/\text{PG}_d$ (which we call the stable moduli space), we get

$$A_{\text{PG}_d}^j(R_d^{\theta\text{-st}}) = A_{\dim R_d - j}^{\text{PG}_d}(R_d^{\theta\text{-st}}) = A_{\dim R_d - \dim \text{PG}_d - j}(M_d^{\theta\text{-st}})$$

and $\dim M_d^{\theta\text{-st}} = \dim R_d - \dim \text{PG}_d = 1 - \chi(d, d)$. This yields that the Donaldson–Thomas invariants of Q are given by the Chow–Betti numbers of the stable moduli spaces, more precisely:

Theorem 9.2. *For a symmetric quiver Q , a stability condition θ and a dimension vector $d \neq 0$, the Donaldson–Thomas invariant $\Omega_{d,k}$ equals*

$$\Omega_{d,k} = \begin{cases} \dim A_{1-(k+\chi(d,d))/2}(M_d^{\text{st}}) & \text{if } k \equiv \chi(d, d) \pmod{2} \text{ and } M_d^{\text{st}} \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, $\Omega_{d,k}$ can only be nonzero if $\chi(d, d) \leq k \leq 2 - \chi(d, d)$.

Remark 9.3. The range for the nonvanishing of the Donaldson–Thomas invariants from the above theorem yields that the number $N_d(Q)$ in [Efimov 2012, Corollary 4.1] can be chosen as $1 - \chi(d, d)$, i.e., the dimension of $M_d^{\theta\text{-st}}$.

10. Examples

10.1. The two-cycle quiver. We start by illustrating the tensor product decomposition from Theorem 6.2. There are exactly three connected symmetric quivers which are not wild, that is, for which a classification of their finite-dimensional representations up to isomorphism is known. Namely, these are:

- The quiver L_0 of Dynkin type A_1 with a single vertex and no arrows.
- The quiver L_1 of extended Dynkin type \tilde{A}_0 consisting of a single vertex and a single loop.
- The quiver Q of extended Dynkin type \tilde{A}_1 with two vertices i and j and single arrows $i \rightarrow j$ and $j \rightarrow i$, respectively.

For the quivers L_0 and L_1 , the structure of the CoHa is determined in [Kontsevich and Soibelman 2011]. Namely, we have

$$\mathcal{A}(L_0) \cong S^*(\mathbb{Q}(1, 1)[z]) \quad \text{and} \quad \mathcal{A}(L_1) \cong S^*(\mathbb{Q}(1, 0)[z]),$$

where S^* denotes the free super-commutative algebra, $\mathbb{Q}(d, i)$ denotes a one-dimensional \mathbb{Q} -space placed in bidegree (d, i) , and z denotes the element in bidegree $(0, 2)$ as in Theorem 7.2.

The structure of the CoHa of Q is described in [Franzen 2018, Corollary 2.5]; here we give a simplified derivation of this result using the present methods. We consider the stability θ given by $\theta(d_i, d_j) = d_i$ (note that any nontrivial stability is equivalent to θ or $-\theta$ in the sense that the class of (semi)stable representation is the same). Let a representation M of Q of dimension vector d be given by vector spaces V_i and V_j and linear maps $f : V_i \rightarrow V_j$ and $g : V_j \rightarrow V_i$. We claim that this representation is θ -semistable if and only if $V_i = 0$, or $V_j = 0$, or $\dim V_i = \dim V_j$ and f is an isomorphism; moreover, it is θ -stable if it is θ -semistable and $\dim V_i, \dim V_j \leq 1$.

The case $(\dim V_i) \cdot (\dim V_j) = 0$ being trivial, we assume $\dim V_i, \dim V_j \geq 1$. Suppose M is θ -semistable. If f is not injective, we choose a vector $0 \neq v \in V_i$ in the kernel of f , yielding a subrepresentation U of dimension vector $(1, 0)$. Then we find $1 = \mu(U) \leq \mu(M) = \dim V_i / (\dim V_i + \dim V_j)$, thus $\dim V_j = 0$, a contradiction. Thus f is injective, and $(V_i, f(V_i))$ defines a subrepresentation U' of dimension vector $(\dim V_i, \dim V_i)$ of M . Then we find $\frac{1}{2} = \mu(U') \leq \mu(M)$, thus $\dim V_j \leq \dim V_i$, which already implies $\dim V_i = \dim V_j$ and shows that f is an isomorphism. Conversely every representation M consisting of vector spaces V_i and V_j of the same dimension, an isomorphism $f : V_i \rightarrow V_j$, and an arbitrary linear map $g : V_j \rightarrow V_i$ is θ -semistable: the subrepresentations of M are of the form $U = (U_i, U_j)$ for some subspaces satisfying $f(U_i) \subseteq U_j$ and $g(U_j) \subseteq U_i$. Injectivity of f implies $\dim U_i \leq \dim U_j$ and thus $\mu(U) \leq \frac{1}{2} = \mu(M)$. To show that stability forces $\dim V_i = 1 = \dim V_j$ we argue as follows: as f is an isomorphism, we may assume without loss of generality that $V_1 = V_2 = V$ and $f = \text{id}_V$. Let $v \in V$ be an eigenvector of g to some eigenvalue λ . The subspaces $U_1 = U_2 = \langle v \rangle$ then provide a subrepresentation of M of dimension vector $(1, 1)$.

This analysis provides identifications

$$\begin{aligned} A_{G_d}^*(R_d^{\theta\text{-sst}}) &\cong A_{\text{Gl}_n(k)}^*(\text{pt}) \quad \text{for } d = (n, 0) \text{ or } d = (0, n), \\ A_{G_d}^*(R_d^{\theta\text{-sst}}) &\cong A_{\text{Gl}_n(k)}^*(M_{n \times n}(k)) \quad \text{for } d = (n, n), \end{aligned}$$

which we recognize as the homogeneous parts of the CoHa of L_0 and L_1 , respectively. These identifications obviously being compatible with the respective Hecke correspondences defining the multiplications, we see that

$$\mathcal{A}^{\theta\text{-sst}, 1}(Q) \cong \mathcal{A}(L_0) \cong \mathcal{A}^{\theta\text{-sst}, 0}(Q) \quad \text{and} \quad \mathcal{A}^{\theta\text{-sst}, 1/2}(Q) \cong \mathcal{A}(L_1).$$

By Theorem 6.2, we thus arrive at an isomorphism of graded vector spaces

$$\mathcal{A}(Q) \cong S^*((\mathbb{Q}((1, 0), 1) \oplus \mathbb{Q}((0, 1), 1) \oplus \mathbb{Q}((1, 1), 0))[z]).$$

10.2. The Kronecker quiver. Now we consider the Kronecker quiver K_2 with two vertices i and j and two arrows from i to j . As we will use results from Section 5, we work over the field of complex numbers. Again we consider the stability $\theta(d_i, d_j) = d_i$. This is again a case where the representation theory of the quiver is known: up to isomorphism, there exist unique (θ -stable) indecomposable representations

P_n and I_n for each of the dimension vectors $(n, n + 1)$ and $(n + 1, n)$, respectively, for $n \geq 0$, and there exist one-parametric families $R_n(\lambda)$ of $(\theta$ -semistable) indecomposables for each of the dimension vectors (n, n) for $n \geq 0$ and $\lambda \in \mathbb{P}^1(\mathbb{C})$. Arguing as in the first example, we can conclude that

$$\mathcal{A}^{\theta\text{-sst}, \mu(d)}(K_2) \cong S^*(\mathbb{Q}(d, 1)[z]) \quad \text{for } d = (n, n + 1) \text{ or } d = (n + 1, n),$$

and

$$\mathcal{A}^{\theta\text{-sst}, \mu}(K_2) = 0 \quad \text{if } \mu \notin \left\{0, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \dots, \frac{1}{2}, \dots, \frac{4}{7}, \frac{3}{5}, \frac{2}{3}, 1\right\}.$$

It remains to consider $\mathcal{A}^{\theta\text{-sst}, 1/2}(K_2)$.

We construct a stratification of the θ -semistable locus in $R_{(n,n)}(K_2) \cong M_{n \times n}(\mathbb{C}) \times M_{n \times n}(\mathbb{C})$, on which $G = \text{Gl}_n(\mathbb{C}) \times \text{Gl}_n(\mathbb{C})$ acts via $(g, h) \cdot (A, B) = (hAg^{-1}, hBg^{-1})$. For $0 \leq r \leq n$, we define S_r as the G -saturation of the set of pairs of matrices

$$\left(\begin{pmatrix} E_r & 0 \\ 0 & N \end{pmatrix}, \begin{pmatrix} A & 0 \\ 0 & E_{n-r} \end{pmatrix} \right),$$

where E_i denotes an $i \times i$ -identity matrix, A denotes an arbitrary $r \times r$ -matrix, and N denotes a nilpotent $(n - r) \times (n - r)$ -matrix. We claim that every S_r is locally closed, their union equals the θ -semistable locus, and the closure of S_r equals the union of the $S_{r'}$ for $r' \leq r$.

The representation $R_n(\lambda)$ is given explicitly by the matrices $(E_n, \lambda E_n + J_n)$ for $\lambda \neq \infty$, and by (J_n, E_n) for $\lambda = \infty$, where J_n is the nilpotent $n \times n$ -Jordan block. As noted above, a θ -semistable representation of M of dimension vector (n, n) is of the form

$$M = R_{n_1}(\lambda_1) \oplus \dots \oplus R_{n_k}(\lambda_k)$$

for $n = n_1 + \dots + n_k$ and $\lambda_1, \dots, \lambda_k \in \mathbb{P}^1(\mathbb{C})$, uniquely defined up to reordering. Now we reorder the direct sum and assume that $\lambda_1, \dots, \lambda_j \neq \infty$ and $\lambda_{j+1} = \dots = \lambda_k = \infty$. Using the above explicit form of the representations $R_n(\lambda)$, we see that M is represented by a pair of block matrices of the form

$$\left(\begin{pmatrix} E_r & 0 \\ 0 & N \end{pmatrix}, \begin{pmatrix} A & 0 \\ 0 & E_{n-r} \end{pmatrix} \right)$$

with N nilpotent and A arbitrary. All claimed properties of the stratification follow.

Now we claim that

$$S_r \cong (\text{Gl}_n(\mathbb{C}) \times \text{Gl}_n(\mathbb{C})) \times^{\text{Gl}_r(\mathbb{C}) \times \text{Gl}_{n-r}(\mathbb{C})} (M_r(\mathbb{C}) \times N_{n-r}(\mathbb{C})),$$

where the group $\text{Gl}_r(\mathbb{C}) \times \text{Gl}_{n-r}(\mathbb{C})$ is considered as a subgroup of $\text{Gl}_n(\mathbb{C}) \times \text{Gl}_n(\mathbb{C})$ by mapping a pair (g_1, h_4) to $\left(\begin{pmatrix} g_1 & 0 \\ 0 & h_4 \end{pmatrix}, \begin{pmatrix} g_1 & 0 \\ 0 & h_4 \end{pmatrix}\right)$. We consider the stabilizer of the set of matrices in the above block form. So we take $g, h \in \text{Gl}_n(\mathbb{C})$, written as block matrices

$$g = \begin{pmatrix} g_1 & g_2 \\ g_3 & g_4 \end{pmatrix}, \quad h = \begin{pmatrix} h_1 & h_2 \\ h_3 & h_4 \end{pmatrix},$$

and assume we are given matrices $A, A' \in M_{r \times r}(\mathbb{C})$ and $N, N' \in N_{n-r}(\mathbb{C})$, the nilpotent cone of $(n-r) \times (n-r)$ matrices, such that

$$\begin{pmatrix} h_1 & h_2 \\ h_3 & h_4 \end{pmatrix} \begin{pmatrix} E_r & 0 \\ 0 & N \end{pmatrix} = \begin{pmatrix} E_r & 0 \\ 0 & N' \end{pmatrix} \begin{pmatrix} g_1 & g_2 \\ g_3 & g_4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} h_1 & h_2 \\ h_3 & h_4 \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & E_{n-r} \end{pmatrix} = \begin{pmatrix} A' & 0 \\ 0 & E_{n-r} \end{pmatrix} \begin{pmatrix} g_1 & g_2 \\ g_3 & g_4 \end{pmatrix}.$$

From these equations we first conclude $h_1 = g_1$ and $h_4 = g_4$, thus $h_2 = A'g_2$ and $g_2 = h_2N$, which yields $h_2 = A'h_2N$. By induction, this implies $h_2 = (A')^k h_2 N^k$ for all $k \geq 1$. But N is nilpotent, thus $h_2 = 0$, thus $g_2 = 0$. Similarly, we can conclude $h_3 = 0$ and $g_3 = 0$. But then g_1 and g_4 are invertible, and $A' = g_1 A g_1^{-1}$ as well as $N' = g_4 N g_4^{-1}$. This proves the claim.

To obtain information on the Chow groups from this stratification using Lemma 5.3, we first have to analyze the Chow groups of nilpotent cones.

The nilpotent cone $N_d(\mathbb{C})$ is irreducible of dimension $d^2 - d$, and the $\text{Gl}_d(\mathbb{C})$ -orbits \mathbb{O}_λ in N_d are parametrized by partitions λ in \mathcal{P}_d , the set of partitions of d (we denote by \mathcal{P} the union of all \mathcal{P}_d 's). The stabilizer G_λ of a point in \mathbb{O}_λ has dimension $\langle \lambda, \lambda \rangle = \sum_{i,j} \min(m_i, m_j) m_i m_j$, and its reductive part is isomorphic to $\prod_i \text{Gl}_{m_i}(\mathbb{C})$, where $m_i = m_i(\lambda)$ denotes the multiplicity of i as a part of λ , for $i \geq 1$. We can thus apply Lemma 5.3 and reduce the structure group — note that in characteristic zero, an orbit is isomorphic to the quotient of the group by the stabilizer of a point — to get

$$A_{\text{Gl}_d(\mathbb{C})}^*(N_d(\mathbb{C})) \cong \bigoplus_{\lambda \in \mathcal{P}_d} A_{G_\lambda}^{*+d-\langle \lambda, \lambda \rangle}(\text{pt}),$$

and the equivariant cycle map for $N_d(\mathbb{C})$ is an isomorphism.

This enables us to again apply Lemma 5.3, this time to the stratification $(S_r)_r$. We compute (using $\text{codim } S_r = n - r$):

$$\begin{aligned} A_{\text{Gl}_n(\mathbb{C}) \times \text{Gl}_n(\mathbb{C})}^*(R_{(n,n)}^{\theta\text{-sst}}(K_2)) &\cong \bigoplus_{r=0}^n A_{\text{Gl}_n(\mathbb{C}) \times \text{Gl}_n(\mathbb{C})}^{*-n+r}(S_r) \\ &\cong \bigoplus_{r=0}^n A_{\text{Gl}_r(\mathbb{C}) \times \text{Gl}_{n-r}(\mathbb{C})}^{*-n+r}(M_r(\mathbb{C}) \times N_{n-r}(\mathbb{C})) \\ &\cong \bigoplus_{r=0}^n A_{\text{Gl}_r(\mathbb{C})}^*(M_r(\mathbb{C})) \otimes A_{\text{Gl}_{n-r}(\mathbb{C})}^{*-n+r}(N_{n-r}(\mathbb{C})) \\ &\cong \bigoplus_{r=0}^n A_{\text{Gl}_r(\mathbb{C})}^*(\text{pt}) \otimes \bigoplus_{\lambda \in \mathcal{P}_{n-r}} A_{G_\lambda}^{*-\langle \lambda, \lambda \rangle}(\text{pt}). \end{aligned}$$

Summing over all n , we obtain

$$\mathcal{A}_{(*, 2*)}^{\theta\text{-sst}, 1/2}(K_2) \cong \bigoplus_{n \geq 0} A_{\text{Gl}_n(\mathbb{C}) \times \text{Gl}_n(\mathbb{C})}^*(R_{(n,n)}^{\theta\text{-sst}}(K_2)) \cong \left(\bigoplus_{r \geq 0} A_{\text{Gl}_r(\mathbb{C})}^*(\text{pt}) \right) \otimes \left(\bigoplus_{\lambda \in \mathcal{P}} A_{G_\lambda}^{*-\langle \lambda, \lambda \rangle}(\text{pt}) \right).$$

The generating function of the bigraded space $\mathcal{A}^{\theta\text{-sst},1/2}(K_2)$ therefore equals

$$\left(\sum_{n \geq 0} \frac{t^n}{(1-q) \cdots (1-q^n)}\right) \cdot \left(\sum_{\lambda} \frac{q^{-(\lambda, \lambda)} t^{|\lambda|}}{\prod_{i \geq 1} ((1-q) \cdots (1-q^{m_i}))}\right) = \prod_{i \geq 0} \frac{1}{1-q^i t} \cdot \prod_{i \geq 1} \frac{1}{1-q^i t}$$

by standard identities. We thus arrive at an isomorphism of bigraded \mathbb{Q} -spaces

$$\mathcal{A}^{\theta\text{-sst},1/2}(K_2) \cong \mathcal{S}^*(\mathbb{Q}((1, 1), 0) \oplus \mathbb{Q}((1, 1), 2))[z].$$

However, this is not an isomorphism of algebras, since we will now exhibit an example showing that the algebra $\mathcal{A}^{\theta\text{-sst},1/2}(K_2)$ is not super-commutative.

We use the algebraic description of the CoHa of Section 7 together with Theorem 8.1. We have

$$\mathcal{A}_{(m,n)}(K_2) \cong \mathbb{Q}[x_1, \dots, x_m, y_1, \dots, y_n]^{S_m \times S_n}$$

with multiplication given as in Section 7. By Theorem 8.1, $\mathcal{A}_{(1,1)}^{\theta\text{-sst},1/2}(K_2)$ is the factor of $\mathcal{A}_{(1,1)}(K_2) \cong \mathbb{Q}[x_1, y_1]$ by the image of the multiplication map $\mathcal{A}_{(1,0)}(K_2) \otimes \mathcal{A}_{(0,1)}(K_2) \rightarrow \mathcal{A}_{(1,1)}(K_2)$, thus

$$\mathcal{A}_{(1,1)}^{\theta\text{-sst},1/2}(K_2) \cong \mathbb{Q}[x, y]/(x - y)^2.$$

Again by Theorem 8.1, $\mathcal{A}_{(2,2)}^{\theta\text{-sst},1/2}(K_2)$ is the factor of $\mathcal{A}_{(2,2)}(K_2)$ by the image of the multiplication map

$$\mathcal{A}_{(2,1)}(K_2) \otimes \mathcal{A}_{(0,1)}(K_2) \oplus \mathcal{A}_{(1,0)}(K_2) \otimes \mathcal{A}_{(1,2)}(K_2) \rightarrow \mathcal{A}_{(2,2)}(K_2).$$

The degree of an element in this image is at least $-\chi((2, 1), (0, 1)) = -\chi((1, 0), (1, 2)) = 3$. A direct calculation shows that for the elements $1, x, y \in \mathcal{A}_{(1,1)}^{\theta\text{-sst},1/2}(K_2)$, we have in $\mathcal{A}_{(2,2)}^{\theta\text{-sst},1/2}(K_2)$:

$$\begin{aligned} 1 * 1 &= 2, \\ 1 * x &= y1 + y2, \\ x * 1 &= 2(x1 + x2) - (y1 + y2), \\ 1 * y &= -(x1 + x2) + 2(y1 + y2), \text{ and} \\ y * 1 &= x1 + x2. \end{aligned}$$

In particular, the (anti)commutator of 1 and x does not vanish.

10.3. Donaldson–Thomas invariants as Chow–Betti numbers. Next, we illustrate Theorem 9.2. We consider the symmetric quiver Q with two vertices i and j and $n \geq 1$ arrows from i to j and from j to i , and the dimension vector $d = (1, r)$ for $r \leq n$. To determine the quantized Donaldson–Thomas invariant $\Omega_{d,k}$, we use the stability $\theta = (r, -1)$, for which d is coprime. Therefore, $\Omega_{d,k} = \Omega_{d,k}^\theta$ equals the (suitably shifted) Poincaré polynomial of the cohomology of the moduli space $R_d^{\theta\text{-sst}}(Q)/\text{PG}_d$, which is isomorphic to a vector bundle over the Grassmannian $\text{Gr}_r(k^n)$ [Reineke 2017, §6.1]. By Theorem 9.2, we can also compute $\Omega_{d,k}$ as the (suitably shifted) Poincaré polynomial of the Chow ring of the moduli space $R_d^{0\text{-st}}(Q)/\text{PG}_d$. Again by [Reineke 2017], this moduli space is isomorphic to the space X of $n \times n$ -matrices of rank r . Mapping such a matrix to its image defines a $\text{GL}_n(k)$ -equivariant fibration

$X \rightarrow \text{Gr}_r(k^n)$, whose fiber is isomorphic to the space of $r \times n$ -matrices of highest rank. The latter being open in an affine space, its Chow ring reduces to \mathbb{Q} , thus the Chow ring of X is isomorphic to the Chow ring of $\text{Gr}_r(k^n)$ as expected.

10.4. Multiple loop quivers. Finally, we consider the quiver L_m with a single vertex and $m \geq 2$ loops. The quantized Donaldson–Thomas invariants are computed explicitly in [Reineke 2012].

All stability conditions are equivalent for this quiver. Let M_d^{simp} be the moduli space of simple (equivalently, stable) representations of L_m of dimension d . It is obtained as the geometric quotient $R_d^{\text{simp}}/\text{PGL}_d$. The Chow ring $A^*(M_d^{\text{simp}})_{\mathbb{Q}} = A_{\text{PGL}_d}^*(R_d^{\text{simp}})_{\mathbb{Q}}$ (we will always work with rational coefficients in this subsection and therefore neglect it in the notation) is a quotient of the equivariant Chow ring $A_{\text{PGL}_d}^*(R_d) = A_{\text{PGL}_d}^*(\text{pt})$. The group of characters of a maximal torus of GL_d identifies with the free abelian group in letters x_1, \dots, x_d , the natural action of the Weyl group $W = S_d$ being the permutation action. A maximal torus of PGL_d is given by the quotient of the chosen maximal torus of GL_d by the diagonally embedded multiplicative group. The corresponding Weyl group is also S_d and the character group is then the submodule

$$X_d = \text{Sp}_{(d-1,1)} = \{a_1x_1 + \dots + a_dx_d \mid a_1 + \dots + a_d = 0\}.$$

The symmetric algebra $\text{Sym}(X_d)$ over X_d is the subalgebra of $\mathbb{Q}[x_1, \dots, x_d]$ generated by $x_j - x_i$ (with $i < j$) and the equivariant Chow ring $A_{\text{PGL}_d}^*(R_d)$ is therefore $\text{Sym}(X_d)^{S_d}$ which identifies with a subalgebra of $A_{G_d}^*(\text{pt}) = \mathbb{Q}[x_1, \dots, x_d]^{S_d}$. As in the proofs of Theorems 8.1 and 9.1, the kernel of $A_{\text{PGL}_d}^*(R_d) \rightarrow A_{\text{PGL}_d}^*(R_d^{\text{simp}})$ is then given by the image of

$$\bigoplus_{\substack{p+q=d \\ p,q>0}} A_{\text{PGL}_d}^*(Z_{p,q} \times^{P_{p,q}} \text{PGL}_d) \rightarrow A_{\text{PGL}_d}^*(R_d),$$

where $P_{p,q}$ is the obvious parabolic subgroup of PGL_d —this, by the way, can be done for an arbitrary quiver and for the kernels $A_{\text{PGL}_d}^*(R_d) \rightarrow A_{\text{PGL}_d}^*(R_d^{\theta\text{-sst}})$ and $A_{\text{PGL}_d}^*(R_d^{\theta\text{-sst}}) \rightarrow A_{\text{PGL}_d}^*(R_d^{\theta\text{-st}})$. The ring $A_{\text{PGL}_d}^*(Z_{p,q})$ is isomorphic to $\text{Sym}(X_d)^{S_p \times S_q}$ and the push-forward map

$$m_{p,q} : A_{\text{PGL}_d}^*(Z_{p,q} \times^{P_{p,q}} \text{PGL}_d) \rightarrow A_{\text{PGL}_d}^*(R_d)$$

can be described algebraically and looks just like the explicit formula from [Kontsevich and Soibelman 2011, Theorem 2], i.e., given by a shuffle product with kernel $\prod_{i=1}^p \prod_{j=1}^q (x_{p+j} - x_i)^{m-1}$. The relations in $A_{\text{PGL}_d}^*(R_d)$ which present $A^*(M_d^{\text{simp}})$ thus have at least degree $(m-1)(d-1)$. In other words, for every $0 \leq i < (m-1)(d-1)$, we get

$$A^i(M_d^{\text{simp}}) \cong A_{\text{PGL}_d}^i(R_d) = \text{Sym}^i(X_d)^{S_d}.$$

The generating series of $\text{Sym}(X_d)^{S_d}$ is

$$\frac{1}{(1-q^2) \cdots (1-q^d)} = \sum_{i \geq 0} \#\{(k_2, \dots, k_d) \mid 2k_2 + \dots + dk_d\} q^i$$

and using Theorem 9.2, we obtain a description of the first few Donaldson–Thomas invariants.

Proposition 10.4.1. *For the m -loop quiver, the Donaldson–Thomas invariant $\Omega_{d,k}$ for a nonnegative integer d and an integer k of the same parity as $(1-m)d^2$ satisfying*

$$(1-m)d^2 \leq k < (1-m)(d^2 - 2d + 2)$$

computes as

$$\Omega_{d,k} = \#\{(k_2, \dots, k_d) \mid 2k_2 + \dots + dk_d = \frac{1}{2}((m-1)d^2 + k)\}.$$

We conclude the subsection with a computation of the numbers $\Omega_{2,k}$. The ring $\text{Sym}(X_2)^{S_2}$ is the subalgebra of $\mathbb{Q}[x_1, x_2]^{S_2}$ which is generated by $(x_2 - x_1)^2$. Abbreviate $\Delta = x_2 - x_1$. As a $\text{Sym}(X_2)^{S_2}$ -module, $\text{Sym}(X_2)$ is generated by 1 and Δ . The push-forward map $m_{1,1} : \text{Sym}(X_2) \rightarrow \text{Sym}(X_2)^{S_2}$ sends $f(x_1, x_2)$ to

$$(f(x_1, x_2) + (-1)^{m-1} f(x_2, x_1))\Delta^{m-1}$$

and therefore, the image of $m_{1,1}$ is the ideal of $\text{Sym}(X_2)^{S_2} = \mathbb{Q}[\Delta^2]$ which is generated by $\Delta^{2\lfloor m/2 \rfloor}$ (i.e., Δ^m if m is even and Δ^{m-1} if m is odd). We have shown that

$$\Omega_{2,k} = \begin{cases} 1 & \text{if } k \equiv 0 \pmod{4} \text{ and } 4(1-m) - 2 \leq k \leq 4(\lfloor m/2 \rfloor - m), \\ 0 & \text{otherwise.} \end{cases}$$

Acknowledgements

The authors would like to thank M. Brion, B. Davison, M. Ehrig, and M. Young for valuable discussions and remarks. We thank the referee whose comments helped to make the exposition much clearer. While doing this research, Franzen was supported by the DFG SFB Transregio 45 “Perioden, Modulräume und Arithmetik algebraischer Varietäten”. Both authors are currently supported by the DFG SFB Transregio 191 “Symplektische Strukturen in Geometrie, Algebra und Dynamik”.

References

- [Assem et al. 2006] I. Assem, D. Simson, and A. Skowroński, *Elements of the representation theory of associative algebras, I*, London Math. Soc. Student Texts **65**, Cambridge Univ. Press, 2006. MR Zbl
- [Davison and Meinhardt 2016] B. Davison and S. Meinhardt, “Cohomological Donaldson–Thomas theory of a quiver with potential and quantum enveloping algebras”, preprint, 2016. arXiv
- [Edidin and Graham 1998] D. Edidin and W. Graham, “Equivariant intersection theory”, *Invent. Math.* **131**:3 (1998), 595–634. MR Zbl
- [Efimov 2012] A. I. Efimov, “Cohomological Hall algebra of a symmetric quiver”, *Compos. Math.* **148**:4 (2012), 1133–1146. MR Zbl
- [Engel and Reineke 2009] J. Engel and M. Reineke, “Smooth models of quiver moduli”, *Math. Z.* **262**:4 (2009), 817–848. MR Zbl
- [Franzen 2015] H. Franzen, “Chow rings of fine quiver moduli are tautologically presented”, *Math. Z.* **279**:3-4 (2015), 1197–1223. MR Zbl
- [Franzen 2016] H. Franzen, “On cohomology rings of non-commutative Hilbert schemes and CoHa-modules”, *Math. Res. Lett.* **23**:3 (2016), 805–840. MR Zbl
- [Franzen 2018] H. Franzen, “On the semi-stable CoHa and its modules arising from smooth models”, *J. Algebra* **503** (2018), 121–145. MR

- [Fulton 1984] W. Fulton, *Intersection theory*, Ergebnisse der Mathematik (3) **2**, Springer, 1984. MR Zbl
- [King and Walter 1995] A. D. King and C. H. Walter, “On Chow rings of fine moduli spaces of modules”, *J. Reine Angew. Math.* **461** (1995), 179–187. MR Zbl
- [Kontsevich and Soibelman 2011] M. Kontsevich and Y. Soibelman, “Cohomological Hall algebra, exponential Hodge structures and motivic Donaldson–Thomas invariants”, *Commun. Number Theory Phys.* **5**:2 (2011), 231–352. MR Zbl
- [Meinhardt and Reineke 2014] S. Meinhardt and M. Reineke, “Donaldson–Thomas invariants versus intersection cohomology of quiver moduli”, preprint, 2014. arXiv
- [Reineke 2003] M. Reineke, “The Harder–Narasimhan system in quantum groups and cohomology of quiver moduli”, *Invent. Math.* **152**:2 (2003), 349–368. MR Zbl
- [Reineke 2012] M. Reineke, “Degenerate cohomological Hall algebra and quantized Donaldson–Thomas invariants for m -loop quivers”, *Doc. Math.* **17** (2012), 1–22. MR Zbl
- [Reineke 2017] M. Reineke, “Quiver moduli and small desingularizations of some GIT quotients”, pp. 613–635 in *Representation theory: current trends and perspectives*, edited by H. Krause et al., Eur. Math. Soc., Zürich, 2017. MR Zbl
- [Rimányi 2013] R. Rimányi, “On the cohomological Hall algebra of Dynkin quivers”, preprint, 2013. arXiv
- [Ringel 1990] C. M. Ringel, “Hall algebras and quantum groups”, *Invent. Math.* **101**:3 (1990), 583–591. MR Zbl
- [Totaro 1999] B. Totaro, “The Chow ring of a classifying space”, pp. 249–281 in *Algebraic K-theory* (Seattle, 1997), edited by W. Raskind and C. Weibel, Proc. Sympos. Pure Math. **67**, Amer. Math. Soc., Providence, RI, 1999. MR Zbl

Communicated by Michel Van den Bergh

Received 2016-04-25 Revised 2018-02-13 Accepted 2018-03-31

franzen@math.uni-bonn.de

Mathematisches Institut, Universität Bonn, Germany

markus.reineke@ruhr-uni-bochum.de

Faculty of Mathematics, Ruhr-Universität Bochum, Germany

Certain abelian varieties bad at only one prime

Armand Brumer and Kenneth Kramer

An abelian surface A/\mathbb{Q} of prime conductor N is *favorable* if its 2-division field F is an S_5 -extension over \mathbb{Q} with ramification index 5 over \mathbb{Q}_2 . Let A be favorable and let B be a semistable abelian variety of dimension $2d$ and conductor N^d with $B[2]$ filtered by copies of $A[2]$. We give a sufficient class field theoretic criterion on F to guarantee that B is isogenous to A^d .

As expected from our paramodular conjecture, we conclude that there is one isogeny class of abelian surfaces for each conductor in $\{277, 349, 461, 797, 971\}$. The general applicability of our criterion is discussed in the data section.

1. Introduction	1027
2. Some review of group schemes	1031
3. The new categories	1034
4. Some Honda systems	1037
5. The local theory	1040
6. Global conclusions	1053
Appendix A. A cohomology computation in the old style	1059
Appendix B. Parabolic subgroups and an obstreperous cocycle	1061
Appendix C. Some technical lemmas on local conductors	1062
Appendix D. Some data	1065
Note added in proof	1069
Acknowledgements	1070
References	1070

1. Introduction

Let $\mathfrak{I}_d(S)$ be the set of isogeny classes of simple abelian varieties over \mathbb{Q} of dimension d with good reduction outside S , a finite set of primes. By [Faltings 1983], $\mathfrak{I}_d(S)$ is finite and it is empty when S is, by [Abrashkin 1987; Fontaine 1985]. All curves of genus 2 with good reduction outside 2 are found in [Merriman and Smart 1993; Smart 1997], yielding 165 isogeny classes of Jacobians. Factors of $J_0(2^{10})$ and Weil restrictions of elliptic curves over quadratic fields provide an additional 50 members of $\mathfrak{I}_2(\{2\})$, but the complete determination of $\mathfrak{I}_2(\{2\})$ is still open.

Research of the second author was partially supported by a PSC-CUNY Award, cycle 44, jointly funded by The Professional Staff Congress and The City University of New York.

MSC2010: primary 11G10; secondary 11R37, 11S31, 14K15.

Keywords: semistable abelian variety, group scheme, Honda system, conductor, paramodular conjecture.

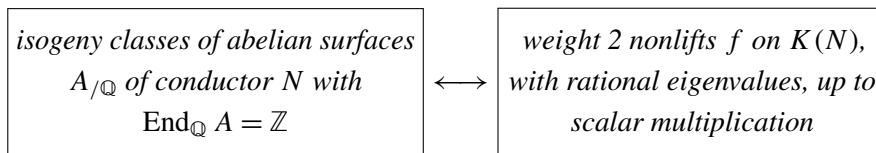
For *semistable* abelian varieties, Fontaine's nonexistence result has been slightly extended [Brumer and Kramer 2001; 2004; 2014; Calegari 2004; Schoof 2005]. It is much more challenging to find all isogeny classes when some exist.

In a beautiful sequence of papers Schoof [2005; 2012b; 2012a] shows that for $S = \{N\}$ with prime $N \leq 23$ or $S = \{3, 5\}$ the classical modular variety $J_0(N)$ or $J_0(15)$, respectively, is the only simple semistable abelian variety of arbitrary dimension, up to isogeny. To apply Faltings' isogeny theorem on abelian varieties, Schoof introduces a general result on p -divisible groups whose constituents belong to a category \underline{C} of finite flat group schemes. For the reader's convenience, the statement is included here as Theorem 3.3. For a suitable choice of category \underline{D} , depending on S , Schoof determines all simple objects and their extensions by one another. Because the Odlyzko bounds are used, the sets S to which these methods apply are severely limited.

In fact, given a finite set S of primes, it seems challenging to decide whether the dimension of the simple semistable abelian varieties good outside S is bounded.

This paper grew out of the desire to check the uniqueness of certain isogeny classes for larger conductors. Another motivation was to provide additional evidence for our conjecture. (See modification added in proof, page 1069.)

Paramodular conjecture [Brumer and Kramer 2014]. *Let $K(N)$ be the paramodular group of level N . There is a one-to-one correspondence*



in which the ℓ -adic representation of $\mathbb{T}_{\ell}(A) \otimes \mathbb{Q}_{\ell}$ and that associated to f are isomorphic for any ℓ prime to N , so that the L -series of A and f agree.

The L -series of abelian surfaces of GL_2 -type are understood via classical elliptic modular forms, while our conjecture treats all other abelian surfaces. It is verified in [Berger et al. 2015; Johnson-Leung and Roberts 2012] for the Weil restrictions of modular elliptic curves over quadratic fields, not isogenous to their conjugates. It is also compatible with twists [Johnson-Leung and Roberts 2017].

To ensure that we are not in the endoscopic case, we consider prime conductors. By [Brumer and Kramer 2014, Theorem 3.4.11], an abelian surface of prime conductor is isogenous to a Jacobian. For each N in $\{277, 349, 461, 797, 971\}$, the space of weight 2 nonlift paramodular forms on $K(N)$ is one-dimensional [Poor and Yuen 2015], so our conjecture predicts that there should be exactly one isogeny class of abelian surfaces of conductor N . In [Brumer and Kramer 2014], we proved that 277 is the smallest prime conductor. For each N listed above, there is a unique Galois module structure available for $A[2]$. For those N , $\mathbb{Q}(A[2])$ must be the Galois closure of a *favorable* quintic field as defined below.

Definition 1.1. Let N be an odd prime. A quintic extension F_0/\mathbb{Q} of discriminant $\pm 16N$ is *favorable* if the prime over 2 has ramification index 5. A *favorable polynomial* is any minimal polynomial for

a favorable quintic field. An abelian surface A of prime conductor N is *favorable* its 2-division field $\mathbb{Q}(A[2])$ is the Galois closure of a favorable quintic field.

We note some pleasant properties of favorable quintic fields.

Proposition 1.2. *Let F be the Galois closure of a favorable quintic field F_0 of discriminant $d_0 = 16N^*$ with $N^* = \pm N$. Then:*

- (i) $\text{Gal}(F/\mathbb{Q})$ is isomorphic to the symmetric group S_5 . At each prime $\mathfrak{N} \mid N$, the inertia group $\mathcal{I}_{\mathfrak{N}} = \mathcal{I}_{\mathfrak{N}}(F/\mathbb{Q})$ is generated by a transposition.
- (ii) The completion $F_{\mathfrak{P}}$ of F at each prime $\mathfrak{P} \mid 2$ is isomorphic to $\mathbb{Q}_2(\mu_5, \sqrt[5]{2})$ and the decomposition group $\mathcal{D}_{\mathfrak{P}} = \mathcal{D}_{\mathfrak{P}}(F/\mathbb{Q})$ is the Frobenius group of order 20. The sign of N^* is determined by $N^* \equiv 5(8)$.
- (iii) There is only one prime over 2 in the subfield K_{20} of F fixed by $\text{Sym}\{3, 4, 5\}$.
- (iv) If A is a favorable abelian surface, then the finite flat group scheme $A[2]_{|\mathbb{Z}_2}$ is absolutely irreducible and biconnected over \mathbb{Z}_2 .

Proof. (i) Since N exactly divides d_0 , only one prime say \mathfrak{N}_0 over N ramifies in F_0/\mathbb{Q} and the \mathcal{O}_{F_0} -ideal generated by N factors as $(N) = \mathfrak{N}_0^e \mathfrak{a}$, where \mathfrak{a} is an ideal prime to \mathfrak{N}_0 and $e > 1$. If f is the residue degree of \mathfrak{N}_0 then $N^{(e-1)f}$ divides d_0 , so $e = 2$, $f = 1$ and the other primes over N are unramified in F_0/\mathbb{Q} . Thus the completion $F_{\mathfrak{N}}$ is $\mathbb{Q}_N(\sqrt{d_0})$ and $\mathcal{I}_{\mathfrak{N}}$ has order 2. Since $\mathcal{I}_{\mathfrak{N}}$ acts nontrivially on $\sqrt{d_0}$, it is generated by a transposition. A transposition and a 5-cycle generate S_5 .

(ii) By assumption, $F_{\mathfrak{P}}/\mathbb{Q}_2$ has tame ramification of degree 5 and thus contains $\mathbb{Q}_2(\mu_5, \sqrt[5]{2})$. Since $\mathcal{D}_{\mathfrak{P}}$ is solvable, $F_{\mathfrak{P}} = \mathbb{Q}_2(\mu_5, \sqrt[5]{2})$. Any Frobenius automorphism at \mathfrak{P} is a 4-cycle, so it acts nontrivially on $\sqrt{d_0}$ and therefore $N^* \equiv 5 \pmod{8}$.

(iii) There are no transpositions in $\mathcal{D}_{\mathfrak{P}}$, so $\mathcal{D}_{\mathfrak{P}} \cap \text{Sym}\{3, 4, 5\}$ is trivial. Since $[K_{20} : \mathbb{Q}] = 20$, there is only one prime over 2 in K_{20} .

(iv) Since $\mathcal{D}_{\mathfrak{P}}$ acts on $A[2]$ via its unique 4-dimensional absolutely irreducible \mathbb{F}_2 -representation, $A[2]_{|\mathbb{Z}_2}$ has no étale or multiplicative constituents. □

A *favorable* S_5 -field is the Galois closure of a favorable quintic field. The Jacobian of a genus 2 curve C is favorable only if C has a model $y^2 = f(x)$ with f favorable, but C might have bad reduction outside N .

In general, L is a *stem field* for M if M is the Galois closure of L/\mathbb{Q} . A *pair-resolvent* for an S_5 -field F is a subfield K fixed by the centralizer of a transposition in S_5 . Then K is well-defined up to isomorphism and is a stem field for F . If r_1 and r_2 are distinct roots of a quintic polynomial f with splitting field F , we can take $K = \mathbb{Q}(r_1 + r_2)$, the fixed field of $\text{Sym}\{1, 2\} \times \text{Sym}\{3, 4, 5\}$. There is only one prime \mathfrak{p} over 2 in K by Proposition 1.2(iii). Let $\Omega_K^{(a)}$ be the maximal elementary 2-extension of K of modulus $\mathfrak{p}^a \cdot \infty$, i.e., the compositum of all quadratic extensions of K with that modulus. Write rk_a for the rank of $\text{Gal}(\Omega_K^{(a)}/K)$.

The following is a restatement of Theorem 6.1.22.

Theorem 1.3. *Let A be a favorable abelian surface of conductor N and let K be a pair-resolvent field for $F = \mathbb{Q}(A[2])$. Suppose that B is a semistable abelian variety of dimension $2d$ and conductor N^d , with $B[2]$ filtered by copies of $A[2]$. If $\text{rk}_2 = 0$ and $\text{rk}_4 \leq 1$, then B is isogenous to A^d . If B is a surface, it is isogenous to A .*

For the proof, we first construct suitable categories \underline{E} , chosen so that extensions of the simple objects \mathcal{E} in \underline{E} can be identified. Most of the paper is devoted to the study of such extension classes. A description of the extensions of \mathcal{E} by \mathcal{E} as group schemes over \mathbb{Z}_p is obtained via Honda systems. For global applications, assume that $p = 2$ and $\mathbb{Q}(\mathcal{E})$ is a favorable \mathcal{S}_5 -field. Monodromy at N restricts the extensions \mathcal{W} of \mathcal{E} by \mathcal{E} as group schemes over $\mathbb{Z}[\frac{1}{2N}]$. A comparison with local data determines when \mathcal{W} prolongs to a group scheme over $\mathbb{Z}[\frac{1}{N}]$ and leads to our class field theoretic criterion for the control of $\text{Ext}_{\underline{E}}^1(\mathcal{E}, \mathcal{E})$ required by Schoof's theorem. Ray class field information, difficult to reach over F , becomes accessible over the degree 10 field K . Moreover, we found that Theorem 1.3 and Proposition 6.1.13 have no analog for other intermediate fields of F/\mathbb{Q} . A more detailed overview of our paper follows.

The category \underline{E} of finite flat p -group schemes over $\mathbb{Z}[\frac{1}{N}]$ defined in §3 is motivated by necessary conditions for an abelian variety B to be isogenous to a product of given semistable abelian varieties A_i . It is essential to impose conductor bounds at N , without which Theorem 3.3 does not apply, as indicated in Example B.4. Thanks to Proposition A.2, we deduce in Theorem 3.7 that it suffices to study the subgroup $\text{Ext}_{[p], \underline{E}}^1(\mathcal{E}, \mathcal{E})$ consisting of classes of extensions \mathcal{W} of \mathcal{E} by \mathcal{E} such that $p\mathcal{W} = 0$.

We review group schemes and Honda systems over the ring of Witt vectors \mathbb{W} of a finite field k of characteristic p in Section 2. In Section 4, finite Honda systems are used to classify absolutely simple biconnected finite flat group schemes \mathcal{E} of rank p^4 over \mathbb{W} and describe the classes $[\mathcal{W}]$ in $\text{Ext}_{[p], \mathbb{Z}_p}^1(\mathcal{E}, \mathcal{E})$. We give the structure of the associated Galois modules E and W in Section 5 and obtain a conductor bound for the elementary abelian extension $K(W)/K(E)$ in Proposition 5.2.17. The latter improves on Fontaine's bound in our case, see Remark 5.2.19.

In Section 6, we restrict to $p = 2$ and give a class field theoretic condition equivalent to the vanishing of $\text{Ext}_{[2], \underline{E}}^1(\mathcal{E}, \mathcal{E})$ in Proposition 6.1.21. Its proof exploits the following ingredients: (i) monodromy at N , to determine the matrix groups available for $\text{Gal}(\mathbb{Q}(W)/\mathbb{Q})$ as W runs over the extensions of E by E as Galois modules, (ii) conductor bounds at $p = 2$, as described above and (iii) rigidification in Section 5.3 and (6.1.5) of the cocycles corresponding to local and global extensions of E by E , to check whether they are compatible, as needed for patching.

Appendix C contains several general facts required for the determination of abelian conductor exponents in our applications.

In Appendix D, we apply Theorem 1.3 to all the favorable quintic fields with N at most 25000 to obtain Table 1. In particular, there is a unique isogeny class of abelian surfaces for each conductor N in $\{277, 349, 461, 797, 971\}$. Curious about the wider applicability of our criterion, we studied the fields corresponding to 276109 favorable abelian surfaces of prime conductor at most 10^{10} found by an ad-hoc

search. We were surprised to discover that the uniqueness, up to isogeny, in Theorem 1.3 holds uniformly for about 11.8% of those fields. The data is summarized in Table 3.

In our companion paper [Brumer and Kramer 2018], extensions \mathcal{W} of exponent p^2 are studied and new “full image” results for certain subgroups of $\mathrm{GSp}_{2g}(\mathbb{Z}_2)$ generated by transvections are obtained. As a consequence, if A is a favorable abelian surface, then $\mathbb{Q}(A[4])$ is an elementary 2-extension of rank 11 over $\mathbb{Q}(A[2])$ with carefully controlled ramification. In Table 1, we also indicate the fields for which no favorable abelian surface can exist because there is no candidate for its 4-division field.

Write \bar{K} for the algebraic closure of K and $G_K = \mathrm{Gal}(\bar{K}/K)$. For any local or global field K , let \mathcal{O}_K be its ring of integers. If L/K is a Galois extension of number fields, let $\mathcal{D}_v(L/K)$ and $\mathcal{I}_v(L/K)$ be the decomposition and inertia subgroups of $\mathrm{Gal}(L/K)$ at a place v of L . We also use v for its restriction to each subfield of L . When the local extension L_v/K_v is abelian, $f_v(L/K)$ denotes the abelian conductor exponent of L_v/K_v . Write $f_v(V)$ for the Artin conductor exponent of a finite $\mathbb{Z}_p[\mathcal{D}_v]$ -module V .

2. Some review of group schemes

Let R be a Dedekind domain with quotient field K . Calligraphic letters are used for finite flat group schemes \mathcal{V} over R and the corresponding Roman letter for the Galois module $V = \mathcal{V}(\bar{K})$. The order of \mathcal{V} is the rank over R of its affine algebra, or equivalently the order of the finite abelian group $V = \mathcal{V}(\bar{K})$.

By the following result of Raynaud [1974], group schemes occurring as subquotients of known group schemes can be treated via their associated Galois modules. Thus, the generic fiber functor induces an isomorphism between the lattice of finite flat closed R -subgroup schemes of \mathcal{V} and that of finite flat closed K -subgroup schemes of $\mathcal{V}|_K$, where K is the field of fractions of R . The following results will be used without explicit reference.

Lemma 2.1. *Let R be a Dedekind domain with quotient field K and let \mathcal{V} be a finite flat group scheme over R with generic fiber $V = \mathcal{V}|_K$. If $W = V_2/V_1$ is a subquotient of V , for closed immersions of finite flat K -group schemes $V_1 \hookrightarrow V_2 \hookrightarrow V$, there are unique closed immersions of finite flat R -group schemes $\mathcal{V}_1 \hookrightarrow \mathcal{V}_2 \hookrightarrow \mathcal{V}$, such that $V_i = \mathcal{V}_i|_K$, and there is a unique isomorphism $\mathcal{V}_2/\mathcal{V}_1 \simeq \mathcal{W}$ compatible with $(\mathcal{V}_2/\mathcal{V}_1)|_K \simeq W$.*

Let p be a prime not dividing N , $R = \mathbb{Z}[\frac{1}{N}]$, $R' = \mathbb{Z}[\frac{1}{pN}]$ and let $\underline{\mathrm{Gr}}$ be the category of p -primary finite flat group schemes over R . Let $\underline{\mathrm{C}}$ be the category of triples $(\mathcal{V}_1, \mathcal{V}_2, \theta)$ where \mathcal{V}_1 is a finite flat \mathbb{Z}_p -group scheme, \mathcal{V}_2 a finite flat R' -group scheme and $\theta : \mathcal{V}_1 \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \rightarrow \mathcal{V}_2 \otimes_{R'} \mathbb{Q}_p$ an isomorphism of \mathbb{Q}_p -group schemes. Then Proposition 2.3 of [Schoof 2003] asserts that the functor $\underline{\mathrm{Gr}} \rightarrow \underline{\mathrm{C}}$ taking the R -group scheme \mathcal{V} to $(\mathcal{V} \otimes_R \mathbb{Z}_p, \mathcal{V} \otimes_R R', \mathrm{id} \otimes_R \mathbb{Q}_p)$ is an equivalence of categories. We can identify $\mathcal{V} \otimes_R R'$ with the Galois module V , since \mathcal{V} is étale over R' . For objects \mathcal{V}_1 and \mathcal{V}_2 of $\underline{\mathrm{Gr}}$, the Mayer–Vietoris sequence of [Schoof 2003, Corollary 2.4] specializes to:

$$\begin{array}{ccccccc} \mathrm{Hom}_{\mathbb{Q}_p}(V_1, V_2) & \leftarrow & \mathrm{Hom}_{\mathbb{Z}_p}(\mathcal{V}_1, \mathcal{V}_2) \times \mathrm{Hom}_{R'}(\mathcal{V}_1, \mathcal{V}_2) & \leftarrow & \mathrm{Hom}_R(\mathcal{V}_1, \mathcal{V}_2) & \leftarrow & 0 \\ \delta \downarrow & & & & & & (2.2) \\ \mathrm{Ext}_R^1(\mathcal{V}_1, \mathcal{V}_2) & \rightarrow & \mathrm{Ext}_{\mathbb{Z}_p}^1(\mathcal{V}_1, \mathcal{V}_2) \times \mathrm{Ext}_{R'}^1(\mathcal{V}_1, \mathcal{V}_2) & \rightarrow & \mathrm{Ext}_{\mathbb{Q}_p}^1(V_1, V_2). & & \end{array}$$

Corollary 2.3. *Let \mathcal{V}_1 and \mathcal{V}_2 be finite flat group schemes over $R = \mathbb{Z}[\frac{1}{N}]$ with \mathcal{V}_1 and \mathcal{V}_2 biconnected over \mathbb{Z}_p . The following natural maps are isomorphisms:*

$$\text{Hom}_R(\mathcal{V}_1, \mathcal{V}_2) \rightarrow \text{Hom}_{\text{Gal}}(V_1, V_2) \quad \text{and} \quad \text{Ext}_R^1(\mathcal{V}_1, \mathcal{V}_2) \rightarrow \text{Ext}_{\text{Gal}}^1(V_1, V_2).$$

If \mathcal{V} is a group scheme over R and $V_{|\mathbb{Q}_p}$ is absolutely irreducible, then

$$\text{End}_{\mathbb{Q}_p}(\mathcal{V}) = \text{End}_{R'}(\mathcal{V}) = \mathbb{F}_p, \quad \text{and} \quad \text{End}_R(\mathcal{V}) = \mathbb{F}_p.$$

In addition, $\delta = 0$ in (2.2) with $\mathcal{V}_1 = \mathcal{V}_2 = \mathcal{V}$.

Proof. The first claim follows from (2.2) and a theorem of Fontaine quoted in [Mazur 1977, Theorem 1.4]. For the second, use Schur’s lemma and a diagram chase. □

We next review some basic material on Honda systems found in [Brinon and Conrad 2009; Conrad 1999; Fontaine 1977]. Let p be a prime, k a perfect field of characteristic $p > 0$, $\mathbb{W} = \mathbb{W}(k)$ the Witt vectors and K its field of fractions. Let $\sigma : \mathbb{W} \rightarrow \mathbb{W}$ be the Frobenius automorphism characterized by $\sigma(x) \equiv x^p \pmod{p}$ for x in \mathbb{W} . The Dieudonné ring $D_k = \mathbb{W}[F, V]$ is generated by the Frobenius operator F and Verschiebung operator V . We have $FV = VF = p$, $Fa = \sigma(a)F$ and $Va = \sigma^{-1}(a)V$ for all a in \mathbb{W} .

A *Honda system* over \mathbb{W} is a pair (M, L) consisting of a finitely generated free \mathbb{W} -module M , a \mathbb{W} -submodule L and a Frobenius semilinear injective endomorphism $F : M \rightarrow M$ with $pM \subseteq F(M)$ and the induced map $L/pL \rightarrow M/FM$ an isomorphism. If F is topologically nilpotent, then (M, L) is *connected*. Since M is torsion free, M becomes a D_k -module with $V = pF^{-1}$.

A *finite Honda system* over \mathbb{W} is a pair (M, L) consisting of a left D_k -module M of finite \mathbb{W} -length and a \mathbb{W} -submodule L with $V : L \rightarrow M$ injective and the induced map $L/pL \rightarrow M/FM$ an isomorphism. If F is nilpotent on M , then (M, L) is *connected*. Morphisms are defined in the obvious manner. If (M, L) is a Honda system then $(M/p^nM, L/p^nL)$ is a finite Honda system.

Honda systems owe their importance to the following fundamental result.

Theorem 2.4 [Fontaine 1975a;1975b]. *Let k be a perfect field of characteristic $p > 0$.*

- (i) *If $p > 2$, there is a natural antiequivalence of categories $G \rightsquigarrow (\mathbf{D}(G_k), \mathbf{L}(G))$ from the category of p -divisible groups over \mathbb{W} to that of Honda systems ($\mathbf{D}(G_k)$ is the Dieudonné module of G_k). The same holds for $p = 2$ if we restrict to connected objects on both sides.*
- (ii) *If $p > 2$, there is a natural antiequivalence of categories from the category of finite flat p -primary group schemes over \mathbb{W} to that of finite Honda systems and the same holds for $p = 2$ if we restrict to connected objects on both sides.*
- (iii) *The cotangent space of G_k at the origin is $\mathbf{D}(G_k)/\mathbf{FD}(G_k)$.*
- (iv) *Both antiequivalences respect extensions of k . Moreover, if G is a p -divisible group over \mathbb{W} , then $(\mathbf{D}(G_k)/(p^n), \mathbf{L}(G)/(p^n))$ is naturally identified with the finite Honda system associated with $G[p^n]$ for all $n \geq 1$.*

Lemma 2.5. *Let (M, L) be a Honda system of exponent p . Then $M = L + FM$ is a direct sum, $\ker F = VL = VM$, $\dim \ker F = \dim L$ and $\ker V = FM$.*

Proof. Since $L/pL \rightarrow M/FM$ is an isomorphism, $M = L + FM$ is a direct sum and

$$\dim M = \dim FM + \dim L = \dim M - \dim \ker F + \dim L.$$

Hence $\dim \ker F = \dim L$ and equality holds for each inclusion in $VL \subseteq VM \subseteq \ker F$ because $V|_L$ is injective. In addition,

$$\dim L = \dim VL = \dim VM = \dim M - \dim \ker V,$$

so $M = L + \ker V$ is a direct sum and the inclusion $FM \subseteq \ker V$ is an equality. □

Let \widehat{CW}_k denote the formal k -group scheme associated to the Witt covector group functor CW_k , see [Conrad 1999; Fontaine 1977]. When k' is a finite extension of k and K' is the field of fractions of $W(k')$, we have $CW_k(k') \simeq K'/W(k')$. For any k -algebra R and $\mathbb{W} = W(k)$, let $D_k = \mathbb{W}[F, V]$ act on elements $\mathbf{a} = (\dots, a_{-n}, \dots, a_{-1}, a_0)$ of $CW_k(R)$ by $F\mathbf{a} = (\dots, a_{-n}^p, \dots, a_{-1}^p, a_0^p)$, $V\mathbf{a} = (\dots, a_{-(n+1)}, \dots, a_{-2}, a_{-1})$ and $\dot{c}\mathbf{a} = (\dots, c^{p^{-n}}a_{-n}, \dots, c^{p^{-1}}a_{-1}, ca_0)$, where \dot{c} in \mathbb{W} is the Teichmüller lift of c . Note that such lifts generate \mathbb{W} as a topological ring.

The Hasse–Witt exponential map is a homomorphism of additive groups,

$$\xi : \widehat{CW}_k(\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}}) \rightarrow \bar{K}/p\mathcal{O}_{\bar{K}} \quad \text{by} \quad (\dots, a_{-n}, \dots, a_{-1}, a_0) \mapsto \sum p^{-n} \tilde{a}_{-n}^{p^n},$$

independent of the choice of lifts \tilde{a}_{-n} in $\mathcal{O}_{\bar{K}}$. If \mathcal{U} is the group scheme of a Honda system (M, L) , the points of the Galois module U correspond to D_k -homomorphisms $\varphi : M \rightarrow \widehat{CW}_k(\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}})$ such that $\xi(\varphi(L)) = 0$ and we say that φ belongs to \mathcal{U} . The action of G_K on $U(\bar{K})$ is induced from its action on $\widehat{CW}_k(\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}})$.

We write $\dot{+}$ for the usual Witt covector addition [Conrad 1999, p. 242] and state some related elementary facts. For q a power of p and x, y in \bar{K} , the congruence $\Phi_q(x, y) \equiv ((\tilde{x} + \tilde{y})^q - \tilde{x}^q - \tilde{y}^q)/q \pmod{p\mathcal{O}_{\bar{K}}}$ defines a unique, possibly nonintegral element of $\bar{K}/p\mathcal{O}_{\bar{K}}$, independent of the choices of lifts \tilde{x} and \tilde{y} in $\mathcal{O}_{\bar{K}}$. The binomial theorem yields the following estimate.

Lemma 2.6. $\text{ord}_p((\tilde{x} + \tilde{y})^q - \tilde{x}^q - \tilde{y}^q) \geq 1 + q \min\{\text{ord}_p(\tilde{x}), \text{ord}_p(\tilde{y})\}.$

It is convenient to write $(\vec{0}, x_{-n}, \dots, x_0)$ for the element $(\dots, 0, 0, x_{-n}, \dots, x_0)$ in $\widehat{CW}_k(\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}})$. A routine calculation using the formulas in [Abrashkin 1987; Conrad 1999] gives:

Lemma 2.7. *Addition in $\widehat{CW}_k(\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}})$ specializes to*

$$(\vec{0}, u_4, u_3, u_2, u_1, u_0) \dot{+} (\vec{0}, v_2, v_1, v_0) = (\vec{0}, u_4, u_3, u_2 + v_2, w_1, w_0),$$

where $w_1 = u_1 + v_1 - \Phi_p(u_2, v_2)$ and

$$w_0 = u_0 + v_0 + \frac{1}{p}(u_1^p + v_1^p) - \Phi_{p^2}(u_2, v_2) - \frac{1}{p}(u_1 + v_1 - \Phi_p(u_2, v_2))^p.$$

3. The new categories

After a review of local conductors, we introduce the categories in which extension classes will be studied.

Fix distinct primes N and p and let K be a finite extension of \mathbb{Q}_N . If L/K is a Galois extension, let $\mathcal{D} = \mathcal{D}(L/K)$ be its Galois group and $\mathcal{I} = \mathcal{I}(L/K)$ its inertia subgroup. When \mathcal{I} acts tamely on the finite $\mathbb{Z}_p[\mathcal{D}]$ -module V , its Artin conductor exponent is given by $f_N(V) = \text{length}_{\mathbb{Z}_p} V/V^{\mathcal{I}}$. If

$$0 \rightarrow V_1 \rightarrow V \rightarrow V_2 \rightarrow 0$$

is an exact sequence of finite $\mathbb{Z}_p[\mathcal{D}]$ -modules, then $f_N(V) \geq f_N(V_1) + f_N(V_2)$.

Let A be an abelian variety over \mathbb{Q}_N with semistable bad reduction and let $\mathbb{T}_p(A)$ denote its p -adic Tate module. We freely use results of Grothendieck [Grothendieck and Raynaud 1972], reviewed in [Brumer and Kramer 2001]. The p^∞ -division field $\mathbb{Q}_N(A[p^\infty])$ depends only on the isogeny class of A , so is shared by the dual variety \hat{A} . The inertia subgroup \mathcal{I} of $\text{Gal}(\mathbb{Q}_N(A[p^\infty])/\mathbb{Q}_N)$ is pro- p cyclic and $(\sigma - 1)^2(\mathbb{T}_p(A)) = 0$ for any topological generator σ of \mathcal{I} . The fixed space $M_f(A) = \mathbb{T}_p(A)^{\mathcal{I}}$ is a \mathbb{Z}_p -direct summand $\mathbb{T}_p(A)$ and the toric space $M_t(A)$ is the \mathbb{Z}_p -submodule of $\mathbb{T}_p(A)$ orthogonal to $M_f(\hat{A})$ under the natural pairing of $\mathbb{T}_p(A)$ with $\mathbb{T}_p(\hat{A})$. Moreover, $(\sigma - 1)(\mathbb{T}_p(A))$ has finite index in $M_t(A)$. The conductor exponent of A at N , denoted $f_N(A)$, is the \mathbb{Z}_p -rank of $\mathbb{T}_p(A)/M_f(A)$. Equivalently, we have $f_N(A) = \text{rank}_{\mathbb{Z}_p} M_t(A) = \text{rank}_{\mathbb{Z}_p} (\sigma - 1)(\mathbb{T}_p(A))$.

Lemma 3.1. *Suppose that $f_N(A[p]) = f_N(A)$. Then we have $f_N(A[p^n]) = n f_N(A[p])$ for all $n \geq 1$ and $(\sigma - 1)(\mathbb{T}_p(A)) = M_t(A)$.*

Proof. In the following diagram

$$(\sigma - 1)(A[p^n]) \xleftarrow{\bar{\pi}} \frac{(\sigma - 1)(\mathbb{T}_p(A))}{(\sigma - 1)(\mathbb{T}_p(A)) \cap p^n \mathbb{T}_p(A)} \xrightarrow{\bar{j}} M_t(A)/p^n M_t(A),$$

$\bar{\pi}$ is an isomorphism induced by the natural projection $\pi : \mathbb{T}_p(A) \rightarrow A[p^n]$ and \bar{j} is an injection induced by the inclusion $j : (\sigma - 1)(\mathbb{T}_p(A)) \rightarrow M_t(A)$. Since $M_t(A)$ is a \mathbb{Z}_p -direct summand of $\mathbb{T}_p(A)$, we have $M_t(A)/p^n M_t(A) \simeq (\mathbb{Z}/p^n)^f$, where $f = f_N(A)$ and thus

$$nf = \text{length}_{\mathbb{Z}_p} M_t(A)/p^n M_t(A) \geq f_N(A[p^n]) \geq n f_N(A[p]), \tag{3.2}$$

using super-additivity of conductors for the last inequality. By assumption, the left and right sides of (3.2) are equal, so $f_N(A[p^n]) = n f_N(A[p])$. Then $\bar{j} \circ \bar{\pi}^{-1}$ is an isomorphism and $(\sigma - 1)(\mathbb{T}_p) = M_t(A)$ upon passage to the limit. \square

We recall the following elegant theorem of Schoof on p -divisible groups.

Theorem 3.3 [Schoof 2005, Theorem 8.3]. *Let $\underline{\mathcal{C}}$ be a full subcategory of the category of p -primary group schemes over $O = \mathbb{Z}[\frac{1}{N}]$, closed under taking products, closed flat subgroup schemes and quotients by closed flat subgroup schemes. Let $G = \{G_n\}$ and $H = \{H_n\}$ be p -divisible groups over O , with G_n and H_n in $\underline{\mathcal{C}}$. Suppose that:*

- (i) $R = \text{End}(G)$ is a discrete valuation ring with uniformizer π and residue field $k = R/\pi R$.

- (ii) The map $\text{Hom}_O(G[\pi], G[\pi]) \xrightarrow{\delta} \text{Ext}_{\mathbb{C}}^1(G[\pi], G[\pi])$, induced by the cohomology sequence of $0 \rightarrow G[\pi] \rightarrow G[\pi^2] \rightarrow G[\pi] \rightarrow 0$, is an isomorphism of one-dimensional k -vector spaces.
- (iii) Each H_n admits a filtration by flat closed subgroup schemes whose successive subquotients are isomorphic to $G[\pi]$.

Then H is isomorphic to G^r for some r .

For Theorem 3.3 to be applicable, the critical condition required of the category \underline{C} is that (ii) hold. Additional motivation for our choice of category is provided at the end of this section.

Definition 3.4. Let $\Sigma = \{\mathcal{E}_i \mid 1 \leq i \leq s\}$ be a collection of finite flat group schemes over $\mathbb{Z}[\frac{1}{N}]$ such that

- (i) \mathcal{E}_i is biconnected over \mathbb{Z}_p for all i and
- (ii) the Galois modules E_i are absolutely simple and pairwise nonisomorphic.

Given Σ , a category \underline{E} of finite flat group schemes \mathcal{V} over $\mathbb{Z}[\frac{1}{N}]$ is a Σ -category if the following properties are satisfied:

- E1.** Each composition factor of \mathcal{V} is isomorphic to some \mathcal{E}_i with $1 \leq i \leq s$.
- E2.** If σ_v generates inertia at $v \mid N$, then $(\sigma_v - 1)^2$ annihilates $V = \mathcal{V}(\overline{\mathbb{Q}})$.
- E3.** If n_i is the multiplicity of E_i in the semisimplification V^{ss} of V , then

$$f_N(V) = f_N(V^{ss}) = \sum n_i f_N(E_i).$$

A collection of semistable abelian varieties A_i , good outside N , is Σ -favorable if $\text{End } A_i = \mathbb{Z}$, the $\mathcal{E}_i = A_i[p]$ satisfy (i) and (ii) and $f_N(A_i) = f_N(E_i)$ for $1 \leq i \leq s$.

In particular, a favorable abelian surface A is Σ -favorable with $\Sigma = \{A[2]\}$.

Lemma 3.5. If $0 \rightarrow \mathcal{W} \rightarrow \mathcal{V} \rightarrow \overline{\mathcal{V}} \rightarrow 0$ is an exact sequence of finite flat group schemes and \mathcal{V} is in \underline{E} , then \mathcal{W} and $\overline{\mathcal{V}}$ also are in \underline{E} .

Proof. By super-additivity of conductors and **E3** for V , we have

$$f_N(V^{ss}) = f_N(W^{ss}) + f_N(\overline{V}^{ss}) \leq f_N(W) + f_N(\overline{V}) \leq f_N(V) = f_N(V^{ss}).$$

Hence **E3** is valid for both W and \overline{V} . The rest is clear. □

Lemma 3.5 implies that \underline{E} is a full subcategory of the category of p -primary group schemes over $\mathbb{Z}[\frac{1}{N}]$, closed under taking products, closed flat subgroup schemes and quotients by closed flat subgroup schemes. As in [Schoof 2005], this guarantees that $\text{Ext}_{\underline{E}}^1$ is defined.

Notation 3.6. If \mathcal{V} and \mathcal{W} in \underline{E} are annihilated by p , write $\text{Ext}_{[p], \underline{E}}^1(\mathcal{V}, \mathcal{W})$ for the subgroup of $\text{Ext}_{\underline{E}}^1(\mathcal{V}, \mathcal{W})$ whose classes are represented by extensions killed by p .

Theorem 3.7. Let $\{A_i \mid 1 \leq i \leq s\}$ be a Σ -favorable collection of abelian varieties and let \underline{E} be the Σ -category with $\Sigma = \{\mathcal{E}_i = A_i[p] \mid 1 \leq i \leq s\}$.

- (i) If B is isogenous to $\prod_i A_i^{n_i}$, then subquotients of $B[p^r]$ are in \underline{E} .
- (ii) Conversely, let B be semistable and write $B[p]^{ss} = \bigoplus n_i \mathcal{E}_i$. Suppose that $f_N(B) = \sum n_i f_N(\mathcal{E}_i)$ and

$$\mathbf{E4} : \text{Ext}_{[p], \underline{E}}^1(\mathcal{E}_i, \mathcal{E}_j) = 0, \quad \text{for all } 1 \leq i \leq j \leq s.$$

Then B is isogenous to $\prod A_i^{n_i}$.

Proof. Lemmas 3.1 and 3.5 imply the first claim. For the converse, it suffices by Lemma 3.5 to show that $B[p^r]$ belongs to \underline{E} . Property **E1** is clear and **E2** follows from semistability. By super-additivity of conductors,

$$\sum n_i f_N(\mathcal{E}_i) = f_N(B[p]^{ss}) \leq f_N(B[p]) \leq f_N(B) = \sum n_i f_N(\mathcal{E}_i).$$

Thus each weak inequality above is an equality and so

$$f_N(B[p^r]) = r f_N(B[p]) = \sum r n_i f(\mathcal{E}_i)$$

by Lemma 3.1. Hence **E3** holds and $B[p^r]$ is in \underline{E} .

Assuming **E4**, the lemma below enables us to define isotypic decompositions of the finite flat group schemes in \underline{E} . Thus the p -divisible group of B is the product of its isotypic p -divisible subgroups $H^{(i)}$. If $G^{(i)}$ is the p -divisible group of A_i , then $\text{End}(G^{(i)}) = \mathbb{Z}_p$ by the theorem of Faltings proving Tate’s conjecture. Vanishing of $\text{Ext}_{[p], \underline{E}}^1(\mathcal{E}_i, \mathcal{E}_i)$ and Proposition A.2 imply that $\text{Ext}_{\underline{E}}^1(\mathcal{E}_i, \mathcal{E}_i) = \mathbb{F}_p$ thanks to the existence of the extension $0 \rightarrow \mathcal{E}_i \rightarrow A_i[p^2] \rightarrow \mathcal{E}_i \rightarrow 0$. Theorem 3.3 now gives $H_i \simeq G_i^{n_i}$ and so the p -divisible group of B is isomorphic to that of $\prod A_i^{n_i}$. Conclude by Faltings’ theorem [1983, §5] on isogenies. □

Lemma 3.8. *Let M be a finite length module over the ring R and E_1, \dots, E_s its nonisomorphic simple constituents. Let M_i be the maximal R -submodule all of whose composition factors are isomorphic to E_i . If $\text{Ext}_R^1(E_i, E_j) = 0$ for $i \neq j$, then $M = \bigoplus M_i$, i.e., M is the sum of its isotypic components.*

Proof. If all composition factors of the R -modules N and N' are isomorphic to E_i , the same is true of $N + N'$ as a quotient of $N \oplus N'$, so the definition of M_i makes sense. The sum of the M_i is direct, since no simple module occurs in the intersection of M_j with the sum of the other isotypics. By the long exact sequence of Ext and induction, $\text{Ext}_R^1(E_i, P) = 0$ if P does not involve E_i . Let $M' = \bigoplus_{i=1}^s M_i \subsetneq M$ and let N be a minimal submodule of M containing M' . Then, after relabeling, we have $N/M' \simeq E_s$. The exact sequence $0 \rightarrow M'/M_s \rightarrow N/M_s \rightarrow E_s \rightarrow 0$ splits, so there is a submodule N' of N with $N'/M_s \simeq E_s$, contradicting maximality of M_s . □

We conclude with some comments on the definition of \underline{E} and the assumptions in Theorem 3.7. Schoof uses categories \underline{D} satisfying only **E1** and **E2**. However, as shown in Example B.4, $\text{Ext}_{[2], \underline{D}}^1(\mathcal{E}, \mathcal{E}) \neq 0$ for the cases of interest to us, violating Theorem 3.3(ii). Motivated by Theorem 3.7(i), we were led to add **E3** as a necessary condition. For the reader who might wonder why **E3** was not imposed by Schoof, we offer the following explanation.

Remark 3.9. In [Schoof 2005, §7], $p = 2$, $\dim E = 2$ and $\Delta = \text{Gal}(\mathbb{Q}(E)/\mathbb{Q}) \simeq \mathcal{S}_3$, so $H^1(\Delta, \text{Mat}_2(\mathbb{F}_2)) = 0$ and the obstruction to splitting of extensions that we encounter in Example B.4 does not arise for Schoof. Moreover, **E2** implies **E3** if $\dim E_i = 2f_N(E_i)$ for all i . Indeed, $V/V^{(\sigma_v)} \simeq (\sigma_v - 1)V \subseteq V^{(\sigma_v)}$ by **E2**. Let $V^{ss} = \bigoplus_i n_i E_i$ and write $\ell(V) = \text{length}_{\mathbb{Z}_p} V$. Then

$$2 \sum n_i f_N(E_i) = \sum n_i \dim_{\mathbb{F}_p} E_i = \ell(V) = \ell((\sigma_v - 1)V) + \ell(V^{(\sigma_v)}) \geq 2\ell((\sigma_v - 1)V) = 2f_N(V) \geq 2 \sum n_i f_N(E_i).$$

Hence $f_N(V) = \sum n_i f_N(E_i)$.

The lower bound $f_N(V) \geq \sum n_i f_N(E_i)$ holds for the conductor of V , while **E3** imposes equality. Thus, in Theorem 3.7(ii), $f_N(B)$ is as small as possible given the structure of $B[p]^{ss}$. But minimality of conductor does not guarantee that B is semistable, as the following example shows.

Example 3.10. Setzer [1981] gives an elliptic curve over $K = \mathbb{Q}(\sqrt{37})$ with everywhere good reduction:

$$C : y^2 - \epsilon y = x^3 + \frac{1}{2}(3\epsilon + 1)x^2 + \frac{1}{2}(11\epsilon + 1)x, \quad \epsilon = 6 + \sqrt{37}.$$

If B is its Weil restriction to \mathbb{Q} , then B has good reduction outside $N = 37$ and $f_N(B) = 2$ by Milne’s conductor formula [1972, Proposition 1]. Let A be any of the elliptic curves over \mathbb{Q} of conductor 37. These curves share the same group scheme $\mathcal{E} = A[2]$ and $f_N(E) = 1$. Let \underline{E} be the Σ -category with $\Sigma = \{\mathcal{E}\}$. Then $B[2]^{ss} = \mathcal{E} \oplus \mathcal{E}$ and so **E3** holds. But B has potential good reduction at N and inertia at $v \mid N$ acts on $\mathbb{T}_2(B)$ through the finite quotient $\text{Gal}(\mathbb{Q}_N(\sqrt{37})/\mathbb{Q}_N)$, so **E2** fails. Note that B was considered earlier in [Shimura 1972].

Remark 3.11. In his work on deformations, Ploner [2015] considered conditions **E1**, **E2** and **E4** for two-dimensional group schemes.

4. Some Honda systems

Recall that \mathbb{W} is the ring of Witt vectors over a finite field k of characteristic p and let K be the quotient field of \mathbb{W} . Suppose that $\mathcal{E} = A[p]$ is an absolutely simple finite flat group scheme of order p^4 where A is an abelian surface over K with biconnected good reduction. In this section, we classify the Honda systems of such \mathcal{E} ’s and those of extensions of \mathcal{E} by itself annihilated by p .

Proposition 4.1. *Let (M, L) be the Honda system for a group scheme \mathcal{E} as above. Then there is a k -basis x_1, x_2, x_3, x_4 for M such that $L = \text{span}\{x_1, x_2\}$,*

$$V = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{4.2}$$

for some λ in k^\times . Furthermore x'_1, \dots, x'_4 is another such basis if and only if $x'_1 = r^{p^2} x_1$ and $\lambda' = r^{1-p^4} \lambda$ with r in k^\times .

Proof. Let $\mathfrak{E} = (M, L)$ be the Honda system for \mathcal{E} . Refer to Lemma 2.5 as needed. Theorem 2.4, applied to the p -divisible group of A implies that $\dim L = 2$. By absolute simplicity, \mathcal{E} becomes a Raynaud \mathbb{F}_{p^4} -module scheme over the Witt vectors $\mathbb{W}(\bar{k})$ [Raynaud 1974; Tate 1997, §4]. Berthelot [1977, Lemme 2.5] shows that $M' = M \otimes_k \bar{k}$ admits a basis $\{\xi_i \mid i \in \mathbb{Z}/4\mathbb{Z}\}$ such that $F(\xi_i) = \xi_{i+1}$ or $V(\xi_{i+1}) = \xi_i$, with L' spanned by a subset of that basis.

Suppose that L' does not contain two successive basis vectors. Then we may assume that $L' = \text{span}\{\xi_1, \xi_3\}$. By injectivity of V on L , we have $V\xi_1 = \xi_0$ and $V\xi_3 = \xi_2$. Since $F(M') = \text{span}\{F(\xi_1), F(\xi_3)\}$ is 2-dimensional, $V(\xi_2) \neq \xi_1$, so $F(\xi_1) = \xi_2$ and similarly $F(\xi_3) = \xi_0$. If $\eta = \xi_1 + \xi_3$, then $F\eta = V\eta = \xi_2 + \xi_0$. Thus there is a sub-Honda system (M'', L'') of \mathfrak{E} with $M'' = \text{span}\{\eta, F\eta\}$ and $L'' = \text{span}\{\eta\}$, contradicting absolute simplicity of \mathcal{E} .

Therefore, we may assume that $L' = \text{span}\{\xi_1, \xi_2\}$. Since V is injective on L' , we cannot have $F(\xi_1) = \xi_2$, so $\xi_1 = V\xi_2 \in L' \cap VL'$ and $\dim_k(L \cap VL) = 1$ over the original ground field k . Write $x_2 = Vx_1 \neq 0$ in $L \cap VL$ with x_1 in L and so $L = \text{span}\{x_1, x_2\}$. Set $x_4 = Fx_1$ and $x_3 = F^2x_1$. Since $\dim_k \ker F = 2$ and F is nilpotent, $F^3 = 0$. By iterating F on $M = L + FM$ to find that $FM = FL + F^2L = \text{span}\{x_3, x_4\}$. Thus x_1, x_2, x_3, x_4 is a basis for M . Injectivity of V on L implies that $Vx_2 \neq 0$. But Vx_2 is in $\ker F = VL = \text{span}\{x_2, x_3\}$ and V is nilpotent. Hence $Vx_2 = \lambda x_3$ for some $\lambda \in k^\times$, resulting in matrix representations of the form (4.2).

For another such basis, x'_2 generates $L \cap VL$, so $x'_2 = r^p x_2$ with $r \in k^\times$. Then $x'_1 = r^{p^2} x_1$ and $x'_3 = F^2 x'_1 = r^{p^4} x_3$. Thus $\lambda' x'_3 = Vx'_2 = rVx_2 = r\lambda x_3 = r^{1-p^4} \lambda x'_3$ and so $\lambda' = r^{1-p^4} \lambda$ in k^\times . □

Notation 4.3. For $\lambda \in k^\times$, let $\mathfrak{E}_\lambda = (M_0, L_0)$ be the Honda system in the proposition and call x_1, x_2, x_3, x_4 a *standard basis* for \mathfrak{E}_λ . Denote the corresponding group scheme, Galois module and representation by $\mathcal{E}_\lambda, E_\lambda$ and ρ_{E_λ} respectively.

Let $\text{Ext}^1(\mathfrak{E}_\lambda, \mathfrak{E}_\lambda)$ be the group of classes of extensions of Honda systems

$$0 \rightarrow \mathfrak{E}_\lambda \xrightarrow{\iota} (M, L) \xrightarrow{\pi} \mathfrak{E}_\lambda \rightarrow 0 \tag{4.4}$$

under Baer sum [Mac Lane 1963, Chapter III, Theorem 2.1] and let $\text{Ext}^1_{[p]}(\mathfrak{E}_\lambda, \mathfrak{E}_\lambda)$ be the subgroup such that $pM = 0$.

Proposition 4.5. *If (M, L) represents a class in $\text{Ext}^1_{[p]}(\mathfrak{E}_\lambda, \mathfrak{E}_\lambda)$, there is a k -basis e_1, \dots, e_8 for M such that $\iota(x_1) = e_1, \pi(e_5) = x_1, L = \text{span}\{e_1, e_2, e_5, e_6\}$,*

$$V = \left[\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & \lambda s_2 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & \lambda s_3 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & \lambda s_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & s_1 & \lambda s_5 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \quad \text{and} \quad F = \left[\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -s_1^p & -s_5^p & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 & -s_2^p & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array} \right]$$

with s_1, s_2, s_3, s_4, s_5 in k . For $\tilde{k} = k/(\sigma^4 - 1)(k)$, the map $(M, L) \rightsquigarrow (s_1, \dots, s_5)$ induces an isomorphism of additive groups $\mathfrak{s} : \text{Ext}_{[p]}^1(\mathfrak{E}_\lambda, \mathfrak{E}_\lambda) \xrightarrow{\sim} k \oplus k \oplus k \oplus \tilde{k} \oplus k$.

Proof. Let $\{x_j \mid 1 \leq j \leq 4\}$ be a standard basis for \mathfrak{E}_λ and define $e_j = \iota(x_j)$ in (4.4). Since $0 \rightarrow L_0 \xrightarrow{\iota} L \xrightarrow{\pi} L_0 \rightarrow 0$ is exact, we can extend e_1, e_2 to a basis for L by adjoining elements \tilde{e}_5, \tilde{e}_6 of L such that $\pi(\tilde{e}_5) = x_1$ and $\pi(\tilde{e}_6) = x_2$.

From $V(\pi(\tilde{e}_5)) = \pi(\tilde{e}_6)$, we have $V\tilde{e}_5 = \tilde{e}_6 + r_1e_1 + r_2e_2 + r_3e_3 + s_1e_4$ with s_1 and all r_i in k . Replace \tilde{e}_5 by $e_5 = \tilde{e}_5 + \sigma^2(a_1)e_1 + \sigma(a_2)e_2$ and \tilde{e}_6 by $e_6 = \tilde{e}_6 + b_1e_1 + b_2e_2$ with a_i, b_i in k . Then

$$\begin{aligned} Ve_5 &= V\tilde{e}_5 + \sigma(a_1)e_2 + \lambda a_2e_3 \\ &= \tilde{e}_6 + r_1e_1 + (r_2 + \sigma(a_1))e_2 + (r_3 + \lambda a_2)e_3 + s_1e_4 \\ &= e_6 + (r_1 - b_1)e_1 + (r_2 + \sigma(a_1) - b_2)e_2 + (r_3 + \lambda a_2)e_3 + s_1e_4. \end{aligned}$$

Now choose a_i, b_i so that $V(e_5) - e_6 = s_1e_4$. Finally, let $e_8 = Fe_5$ and $e_7 = Fe_8$. Since $V(\pi(e_6)) = \lambda\pi(e_7)$, we may choose elements s_i of k such that

$$Ve_6 = \lambda(e_7 + s_2e_1 + s_3e_2 + s_4e_3 + s_5e_4). \tag{4.6}$$

This verifies the matrix representation of V . From $0 = FVe_5 = Fe_6 + \sigma(s_1)e_3$, we get $Fe_6 = -\sigma(s_1)e_3$. Apply F to (4.6) to find Fe_7 and obtain the matrix of F .

The only ambiguity left is that e_5 might be replaced by $e_5 + \sigma^2(a_1)e_1$, in which case s_4 becomes $s_4 + a_1 - \sigma^4(a_1)$ while s_1, s_2, s_3, s_5 remain unchanged.

Another extension (M', L') is equivalent to (M, L) if and only if there is an isomorphism h in the commutative diagram

$$\begin{array}{ccccccc} 0 & \longrightarrow & \mathfrak{E}_\lambda & \xrightarrow{\iota'} & (M', L') & \xrightarrow{\pi'} & \mathfrak{E}_\lambda \longrightarrow 0 \\ & & \downarrow \text{ident} & & \downarrow h & & \downarrow \text{ident} \\ 0 & \longrightarrow & \mathfrak{E}_\lambda & \xrightarrow{\iota} & (M, L) & \xrightarrow{\pi} & \mathfrak{E}_\lambda \longrightarrow 0. \end{array} \tag{4.7}$$

Let e'_1, \dots, e'_8 be a basis for (M', L') constructed as above. Since $h(e'_1), \dots, h(e'_8)$ must be another such basis, the isomorphism h exists if and only if $h(e'_1) = e_1$ and $h(e'_5) = e_5 + \sigma^2(a_1)e_1$ with a_1 in k . It follows that \mathfrak{s} is a well-defined bijection.

To verify the additivity of \mathfrak{s} , let (M, L) and (M', L') represent two classes in $\text{Ext}_{[p]}^1(\mathfrak{E}_\lambda, \mathfrak{E}_\lambda)$ and let $0 \rightarrow \mathfrak{E}_\lambda \xrightarrow{\iota''} (M'', L'') \xrightarrow{\pi''} \mathfrak{E}_\lambda \rightarrow 0$ represent their Baer sum. To obtain a k -basis for M'' let $\gamma_i = (e_i, 0)$ in $M \times M'$ for $1 \leq i \leq 4$ and $\gamma_i = (e_i, e'_i)$ for $5 \leq i \leq 8$, each of which satisfies the fiber product condition that $\pi''(\gamma_i) = \pi(e_i) = \pi'(e'_i)$. The relations are given by $\iota''(a) = (\iota(a), 0) = (0, \iota'(a))$ for all a in \mathfrak{E}_λ . We have

$$\begin{aligned} V\gamma_5 &= (Ve_5, Ve'_5) = (e_6 + s_1e_4, e'_6 + s'_1e'_4) = \gamma_6 + (s_1e_4, 0) + (0, s'_1e'_4) \\ &= \gamma_6 + (s_1e_4, 0) + (s'_1e_4, 0) = \gamma_6 + (s_1 + s'_1)\gamma_4, \end{aligned}$$

$$\begin{aligned} V\gamma_6 &= (Ve_6, Ve'_6) = \lambda(e_7, e'_7) + \sum_{1 \leq i \leq 4} \lambda(s_i e_i, s'_i e'_i) = \lambda\gamma_7 + \sum_{1 \leq i \leq 4} \lambda(s_i + s'_i)\gamma_i, \\ F\gamma_6 &= (Fe_6, Fe'_6) = -(s_1^p e_3, (s'_1)^p e'_3) = -(s_1^p e_3, 0) - (0, (s'_1)^p e'_3) \\ &= -(s_1^p e_3, 0) - ((s'_1)^p e_3, 0) = -(s_1 + s'_1)^p \gamma_3, \\ F\gamma_7 &= (Fe_7, Fe'_7) = -(s_5^p e_3, (s'_5)^p e'_3) - (s_2^p e_4, (s'_2)^p e'_4) \\ &= -(s_5 + s'_5)^p \gamma_3 - (s_2 + s'_2)^p \gamma_4. \end{aligned}$$

By completing the matrices for V and F, we find that $s''_i = s_i + s'_i$ for $1 \leq i \leq 5$. □

5. The local theory

In this section, we study the fields of points of extensions of exponent p whose Honda systems were described above. In particular, we obtain good conductor bounds. We use freely the notation of Section 2. Let K be the quotient field of \mathbb{W} and let \dot{a} be the Teichmüller lift to \mathbb{W} of a in k , with $\dot{0} = 0$. Assume that w is in $\mathcal{O}_{\bar{K}}$ and $\text{ord}_p(w) > 0$. For a in $\bar{K}/w\mathcal{O}_{\bar{K}}$, let \tilde{a} be an arbitrary lift to \bar{K} . Assertions requiring lifts are made only when the result is independent of the choices, as in the following examples. If a is not in $w\mathcal{O}_{\bar{K}}$, let $\text{ord}_p(a) = \text{ord}_p(\tilde{a})$. For w' in $\mathcal{O}_{\bar{K}}$ such that $0 < \text{ord}_p(w') \leq \text{ord}_p(w)$, let $a \equiv b \pmod{w'}$ mean that $\tilde{a} - \tilde{b}$ is in $w'\mathcal{O}_{\bar{K}}$. If $f(X)$ is in $\bar{K}[X]$, we write $f(a) \equiv 0 \pmod{w''\mathcal{O}_{\bar{K}}}$ only if $f(\tilde{a})$ is in $w''\mathcal{O}_{\bar{K}}$, for all lifts \tilde{a} of a . For this section, we write $x \sim y$ when $\text{ord}_p(x/y - 1) > 0$ and $x = y + \mathcal{O}(w)$ if $\text{ord}_p(x - y) \geq \text{ord}_p(w)$.

5.1. The irreducible case. Let \mathcal{E}_λ be the group scheme and x_1, \dots, x_4 a standard basis for the corresponding Honda system $\mathfrak{E}_\lambda = (M_0, L_0)$ from Notation 4.3. The Galois module structure of E_λ is well-known, but a description of E_λ by Witt covectors is required for our analysis of extensions of \mathcal{E}_λ by \mathcal{E}_λ . Let $F = K(E_\lambda)$, reserving Roman F and V for the Honda system Frobenius and Verschiebung operators in this section. Recall that points of the Galois module E_λ correspond to D_k -homomorphisms $\psi : M_0 \rightarrow \widehat{C\mathbb{W}}_k(\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}})$ such that $\xi(\psi(L_0)) = 0$, see Section 2.

Proposition 5.1.1. *Let $\mathfrak{X}_\lambda = \{a \in \mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}} \mid \lambda^{p^2} a^{p^4} \equiv (-p)^{p+1} a \pmod{p^{p+2}\mathcal{O}_{\bar{K}}}\}$. Given a in \mathfrak{X}_λ , define b and c in $\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}}$ by*

$$b \equiv -\frac{1}{p} \lambda^p a^{p^3} \pmod{p\mathcal{O}_{\bar{K}}} \quad \text{and} \quad c \equiv \lambda a^{p^2} \pmod{p\mathcal{O}_{\bar{K}}}.$$

- (i) *A D_k -map $\psi = \psi_a$ belongs to a point P_a of E_λ if and only if $\psi(x_1) = (\vec{0}, c, b, a)$ with a in \mathfrak{X}_λ . If so, $\psi(x_2) = (\vec{0}, c, b)$, $\psi(x_3) = (\vec{0}, \lambda^{-1}c)$ and $\psi(x_4) = (\vec{0}, a^p)$.*
- (ii) *$F = K(E_\lambda)$ is the splitting field of $f_\lambda(x) = \dot{\lambda}^{p^2} x^{p^4-1} - (-p)^{p+1}$ over K . The maximal subfield of F unramified over K is $F_0 = K(\mu_{p^4-1}, \eta)$, where η is any root of $x^{p+1} - \dot{\lambda}$. Moreover F/F_0 is tamely ramified of degree $t = (p^2 + 1)(p - 1)$. For $a \neq 0$ we have*

$$\text{ord}_p(a) = \frac{1}{t}, \quad \text{ord}_p(b) = \frac{p^2 - p + 1}{t}, \quad \text{ord}_p(c) = \frac{p^2}{t}. \tag{5.1.2}$$

(iii) \mathfrak{A}_λ is an \mathbb{F}_{p^4} -vector space under the usual operations in $\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}}$ and $a \mapsto P_a$ defines an $\mathbb{F}_p[G_K]$ -isomorphism $\mathfrak{A}_\lambda \xrightarrow{\sim} E_\lambda$.

Proof. (i) If ψ belongs to a point in E_λ , then $\psi(x_1) = (\vec{0}, c, b, a)$, since $V^3 = 0$. We obtain $\psi(x_2)$ and $\psi(x_3)$ by applying V , while $\psi(x_4) = \psi(Fx_1) = (\vec{0}, c^p, b^p, a^p)$. Use $0 = VF(x_1) = Vx_4$ to find that $c^p = b^p = 0$, so $\text{ord}_p(b), \text{ord}_p(c) \geq 1/p$. In addition, $F(x_4) = x_3$ implies that $c = \lambda a^{p^2}$. Let $\tilde{a}, \tilde{b}, \tilde{c}$ denote lifts to $\mathcal{O}_{\bar{K}}$. Vanishing of $\xi(\psi(L))$ provides the additional congruences modulo $p\mathcal{O}_{\bar{K}}$:

$$\tilde{a} + \frac{1}{p}\tilde{b}^p + \frac{1}{p^2}\tilde{c}^{p^2} \equiv 0 \quad \text{and} \quad \tilde{b} + \frac{1}{p}\tilde{c}^p \equiv 0. \tag{5.1.3}$$

Thus $p \text{ord}_p(\tilde{c}) = \text{ord}_p(p\tilde{b}) \geq 1 + \frac{1}{p}$ and so $\frac{1}{p^2}\tilde{c}^{p^2} \equiv 0$. With this simplification, the required congruences follow from (5.1.3). Furthermore, these congruences are sufficient to imply that ψ belongs to \mathcal{E}_λ when $\psi(x_1) = (\vec{0}, c, b, a)$.

(ii) If $f_\lambda(\theta) = 0$ and ζ generates μ_{p^4-1} , then the roots of f_λ have the form $\theta_j = \zeta^j\theta$ while their reductions modulo p give all nonzero elements of \mathfrak{A}_λ . For the converse, let \tilde{a} be a lift of $a \in \mathfrak{A}_\lambda$ and $g(x) = x^{p^4} - x$. Then $g(\tilde{a}/\theta) \equiv 0 \pmod{\frac{p}{\theta}\mathcal{O}_{\bar{K}}}$ and so $\tilde{a} \equiv 0$ or $\tilde{a} \equiv \theta_j \pmod{p\mathcal{O}_{\bar{K}}}$ for some j by Hensel's lemma. Hence $F = K(\mu_{p^4-1}, \theta)$ is the splitting field of f_λ . Let F_0 be the maximal subfield of F unramified over K . Since λ^{p^2} and therefore also λ is a $(p+1)$ power in $K(\theta)$, each root η of $x^{p+1} - \lambda$ is in F_0 . Furthermore, θ satisfies an Eisenstein polynomial of the form $\eta^{p^2}x^t + \omega p = 0$ over F_0 for some ω in μ_{p+1} . Hence K/F_0 is tamely ramified of degree t and we obtain the desired ordinals of a, b, c .

(iii) The embedding $\mathbb{F}_{p^4} = \mathbb{W}(\mathbb{F}_{p^4})/p\mathbb{W}(\mathbb{F}_{p^4}) \hookrightarrow \mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}}$ defines the scalar multiplication by \mathbb{F}_{p^4} . Closure of \mathfrak{A}_λ under this operation and under the usual addition in $\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}}$ is clear. The asserted Galois isomorphism follows from the correspondence between D_k -homomorphisms belonging to \mathcal{E}_λ and points of E_λ once we check that $a \mapsto P_a$ is additive. If a_1 and a_2 are in \mathfrak{A}_λ , then there is some a in \mathfrak{A}_λ such that $\psi_{a_1}(x_1) \dot{+} \psi_{a_2}(x_1) = \psi_a(x_1)$. Denote this equation of Witt covectors by $(\vec{0}, c_1, b_1, a_1) \dot{+} (\vec{0}, c_2, b_2, a_2) = (\vec{0}, c, b, a)$. Then $c = c_1 + c_2$, so $a^{p^2} = a_1^{p^2} + a_2^{p^2}$ in $\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}}$. By using lifts of a, a_1 and a_2 of the form $\omega_0\theta, \omega_1\theta$ and $\omega_2\theta$, with each ω_j in $\mu_{p^4-1} \cup \{0\}$, we find that

$$\omega_0^{p^2} \equiv \omega_1^{p^2} + \omega_2^{p^2} \equiv (\omega_1 + \omega_2)^{p^2} \pmod{\frac{p}{\theta^{p^2}}\mathcal{O}_{\bar{K}}}.$$

Since the ω 's lie in the absolutely unramified field $\mathbb{Q}_p(\mu_{p^4-1})$ and $\text{ord}_p(p/\theta^{p^2}) > 0$, we obtain $\omega_0 \equiv \omega_1 + \omega_2 \pmod{p}$ and thus $a = a_1 + a_2$ in $\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}}$. Alternatively, $\text{ord}_p(a - a_1 - a_2) \geq 1$ by the covector addition formulas in Lemma 2.7. □

Remark 5.1.4. (i) By (ii) above, the lifts of all $a \neq 0$ in \mathfrak{A}_λ to $\mathcal{O}_{\bar{K}}$ comprise the cosets $\zeta^j\theta + p\mathcal{O}_{\bar{K}}$. Thus \mathfrak{A}_λ descends to an \mathbb{F}_{p^4} -vector subspace of $\mathcal{O}_F/p\mathcal{O}_F$ and we write

$$\mathfrak{A}_\lambda(F) = \{a \in \mathcal{O}_F/p\mathcal{O}_F \mid \lambda^{p^2}a^{p^4} \equiv (-p)^{p+1}a \pmod{p^{p+2}\mathcal{O}_F}\}.$$

For α in \mathbb{F}_{p^4} and a in \mathfrak{X}_λ , we write $\alpha P_a = P_{\alpha a}$, in agreement with multiplication on Witt covectors. In fact, $\alpha \psi_a = \psi_{\alpha a}$, since evaluating on x_1 gives

$$[\alpha](\vec{0}, c_a, b_a, a) = (0, \alpha^{1/p^2} c_a, \alpha^{1/p} b_a, \alpha a) = (\vec{0}, c_{\alpha a}, b_{\alpha a}, \alpha a).$$

- (ii) If h is in the ramification subgroup of $\text{Gal}(F/K)$, then h acts on $\mathfrak{X}_\lambda(F)$ by $h(a) = \alpha a$, where $\alpha \in \mu_t$ depends on h . The structure of \mathcal{E}_λ as a Raynaud \mathbb{F}_{p^4} -module scheme is reflected by $h(P_a) = P_{h(a)} = P_{\alpha a} = \alpha P_a$. However, Frobenius in $\text{Gal}(F/K)$ acts on the scalars.
- (iii) By Proposition 5.1.1, we have $b^p \equiv -pa \pmod{p^2}$, $c^p \equiv \lambda^p a^{p^3} \equiv -pb \pmod{p^2}$ and $c^{p^2} \equiv (-p)^{p+1} a \pmod{p^{p+2}}$. These congruences are independent of the choices of lifts to $\mathcal{O}_{\bar{K}}$.

Using the local structure above, we next obtain a group scheme \mathcal{E} over $\mathbb{Z}[\frac{1}{N}]$ fulfilling the hypotheses of Definition 3.4 for a Σ -category \underline{E} with $\Sigma = \{\mathcal{E}\}$. We also determine the image of the Galois representation provided by E .

Corollary 5.1.5. *Let E be a four-dimensional symplectic module over \mathbb{F}_p and let $\rho : G_{\mathbb{Q}} \rightarrow \text{GSp}(E)$ be unramified outside $\{p, N, \infty\}$ and tamely ramified at the prime $N \neq p$. Suppose that:*

- (i) ρ restricted to a decomposition group at p is isomorphic to a local representation of the form ρ_{E_λ} as in Notation 4.3.
- (ii) Inertia at $v \mid N$ acts on E via a cyclic quotient $\langle \sigma_v \rangle$ with $(\sigma_v - 1)^2 = 0$ and $\text{rank}(\sigma_v - 1) = 1$ as a matrix.
- (iii) The fixed field of $\rho^{-1}(\text{Sp}(E))$ is $\mathbb{Q}(\mu_p)$ when p is odd.

Then there is a unique finite flat group scheme \mathcal{E} over $\mathbb{Z}[\frac{1}{N}]$ whose associated Galois representation is ρ . Moreover, the Galois image $G = \rho(G_{\mathbb{Q}})$ is $\text{GSp}_4(\mathbb{F}_p)$ for $p \geq 2$ or possibly $\text{O}_4^-(\mathbb{F}_2) \simeq S_5$ when $p = 2$.

Proof. By (i), the local representation is irreducible and so is E . We patch as described before (2.2) to get the uniqueness.

Since σ_v is a transvection by (ii), the normal subgroup P generated by transvections is nontrivial. Follow the proof of [Brumer and Kramer 2012, Proposition 2.8], using $\dim E = 4$ and the fact that N is square-free, to conclude that E is irreducible for the group P generated by transvections. If $p = 2$, we find that G is isomorphic to $\text{Sp}_4(\mathbb{F}_2) \simeq S_6$ or $\text{O}_4^-(\mathbb{F}_2) \simeq S_5$. Since 5 must divide $|G|$, we rule out $S_3 \wr S_2$. When p is odd, G contains $\text{Sp}_4(\mathbb{F}_p)$ by [Kemper and Malle 1997] and thus is isomorphic to $\text{GSp}_4(\mathbb{F}_p)$ by (iii). \square

When $p = 2$ and A is a favorable abelian surface, $\mathcal{E} = A[2]$ provides a representation ρ as in the corollary.

5.2. Extensions of exponent p . Let $0 \rightarrow \mathcal{E}_\lambda \xrightarrow{\iota} \mathcal{W} \xrightarrow{\pi} \mathcal{E}_\lambda \rightarrow 0$ be an extension of \mathcal{E}_λ by \mathcal{E}_λ killed by p with parameters $\mathbf{s}(\mathcal{W}) = [s_1 \cdots s_5]$ from Proposition 4.5. Let P_a denote the point of \mathcal{E}_λ corresponding to a in $\mathfrak{X}_\lambda(F)$, see Proposition 5.1.1(iii) and Remark 5.1.4. Then the fiber over P_a has the form $Q + \iota(E_\lambda)$ for any fixed Q in W such that $\pi(Q) = P_a$. We write $F_a = F(Q)$ for the fiber field generated over F by the coordinates of Q .

Notation 5.2.1. Write $R_u = \bar{K}/\frac{p}{u}\mathcal{O}_{\bar{K}}$, provided that u is in $\mathcal{O}_{\bar{K}}$ and $\text{ord}_p(u) < 1$.

Proposition 5.2.2. *For φ to correspond to a point of W in the fiber over $P_a \neq 0$, it is necessary and sufficient that $\varphi(e_1) = (\vec{0}, c, b, a)$ as in Proposition 5.1.1 and*

$$\varphi(e_5) = (\vec{0}, (\lambda s_2)^{1/p^2} c, (\lambda s_2)^{1/p} b + (\lambda s_3)^{1/p} c, cz, by, ax)$$

where x, y, z in \bar{K} satisfy all the following congruences:

$$\begin{aligned} \text{i)} \quad & x - y^p + p^{p-1} z^{p^2} = 0 \quad \text{in } R_a. \\ \text{ii)} \quad & y - z^p + p\lambda^{-p} \epsilon_p a^{p-p^3} = 0 \quad \text{in } R_b. \\ \text{iii)} \quad & x^{p^2} - z + wa^{-p^2} = 0 \quad \text{in } R_c. \end{aligned} \tag{5.2.3}$$

with $w = s_2 a + s_3 b + s_4 \lambda^{-1} c + s_5 a^p$, $\epsilon_p = s_1$ if $p \geq 3$ and $\epsilon_2 = s_1 - (\lambda s_2)^2$ if $p = 2$. Equivalently, z in \bar{K} satisfies $f_a(z) = 0$ in R_c , where

$$f_a(Z) = [(Z^p - p\lambda^{-p} \epsilon_p a^{p-p^3})^p - p^{p-1} Z^{p^2}]^{p^2} - Z + wa^{-p^2} \tag{5.2.4}$$

and the classes of x in R_a and y in R_b are determined by (5.2.3)(i) and (ii). When $\epsilon_p = 0$, we may instead use $f_a(Z) = Z^{p^4} - Z + wa^{-p^2}$.

Proof. Let φ in $\text{Hom}_{D_k}(M, \widehat{C\mathcal{W}}_k(\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}}))$ be an element of \mathcal{W} . Since M is generated by e_1 and e_5 as a D_k -module, φ is determined by $\varphi(e_1)$ and $\varphi(e_5)$. The injection of \mathfrak{E}_λ to M yields $\varphi(e_j) = \psi(x_j)$ for $1 \leq j \leq 4$, as in Proposition 5.1.1(i). Set $\varphi(e_5) = (\vec{0}, d_4, d_3, d_2, d_1, d_0)$, with only the five rightmost coordinates significant, since $V^5 = 0$. Applying $FV = 0$ to e_5 gives $d_4^p = d_3^p = d_2^p = d_1^p = 0$.

From the matrix representation of V , we have

$$\varphi(e_6) = V(\varphi(e_5)) \dot{+} [-s_1] \varphi(e_4) = (\vec{0}, d_4, d_3, d_2, d_1 - s_1 a^p)$$

and so $\varphi(\lambda^{-1} V e_6) = [\lambda^{-1}] (\vec{0}, d_4, d_3, d_2)$. We also have

$$\begin{aligned} \varphi(\lambda^{-1} V e_6) &= \varphi(e_7) \dot{+} \varphi(s_2 e_1) \dot{+} \varphi(s_3 e_2) \dot{+} \varphi(s_4 e_3) \dot{+} \varphi(s_5 e_4) \\ &= F^2 \varphi(e_5) \dot{+} (\vec{0}, \sigma^{-2}(s_2)c, \sigma^{-1}(s_2)b, s_2 a) \dot{+} (\vec{0}, \sigma^{-1}(s_3)c, s_3 b + s_4 \lambda^{-1} c + s_5 a^p) \\ &= (\vec{0}, d_0^{p^2}) \dot{+} (\vec{0}, s_2^{1/p^2} c, s_2^{1/p} b + s_3^{1/p} c, s_2 a + s_3 b + s_4 \lambda^{-1} c + s_5 a^p - \Phi_p(s_2^{1/p} b, s_3^{1/p} c)) \\ &= (\vec{0}, s_2^{1/p^2} c, s_2^{1/p} b + s_3^{1/p} c, s_2 a + s_3 b + s_4 \lambda^{-1} c + s_5 a^p + d_0^{p^2}). \end{aligned}$$

since $\Phi_p(s_2^{1/p} b, s_3^{1/p} c) = 0$ by (5.1.2) and Lemma 2.6. Modulo $p\mathcal{O}_{\bar{K}}$, this gives:

$$d_4 \equiv (\lambda s_2)^{1/p^2} c, \quad d_3 \equiv (\lambda s_2)^{1/p} b + (\lambda s_3)^{1/p} c, \quad d_2 \equiv \lambda(d_0^{p^2} + w). \tag{5.2.5}$$

Vanishing of the Hasse–Witt map on $\varphi(L)$ gives the following additional relations:

$$\begin{aligned} \xi(\varphi(e_5)) &= \frac{d_4^{p^4}}{p^4} + \frac{d_3^{p^3}}{p^3} + \frac{d_2^{p^2}}{p^2} + \frac{d_1^p}{p} + d_0 \equiv 0 \pmod{p\mathcal{O}_{\bar{K}}}, \\ \xi(\varphi(e_6)) &= \frac{d_4^{p^3}}{p^3} + \frac{d_3^{p^2}}{p^2} + \frac{d_1^p}{p} + d_1 - s_1 a^p \equiv 0 \pmod{p\mathcal{O}_{\bar{K}}}. \end{aligned} \tag{5.2.6}$$

Since $p^2 \text{ord}_p(d_4) \geq p^2 \text{ord}_p(c) > p + 1$, we have $p^{-3}d_4^{p^3} \equiv 0$ and $p^{-4}d_4^{p^4} \equiv 0 \pmod{p}$. Thus the d_4 -terms drop out of (5.2.6). By (5.2.5), we have

$$\begin{aligned} d_3^p &\equiv \lambda s_2 b^p + \lambda s_3 c^p \pmod{p^2}, \\ d_3^{p^2} &\equiv (\lambda s_2)^p b^{p^2} + (\lambda s_3)^p c^{p^2} \pmod{p^3}, \\ d_3^{p^3} &\equiv (\lambda s_2)^{p^2} b^{p^3} + (\lambda s_3)^{p^2} c^{p^3} \pmod{p^4}. \end{aligned}$$

In addition,

$$\text{ord}_p\left(\frac{c^{p^j}}{p^j}\right) > \text{ord}_p\left(\frac{b^{p^j}}{p^j}\right) = \frac{p^j(p^2 - p + 1)}{(p - 1)(p^2 + 1)} - j = p^{j-1}\left(1 + \frac{1}{(p - 1)(p^2 + 1)}\right) - j$$

is greater than 1 if (i) $j = 3$ and all p or (ii) $j = 2$ and $p \geq 3$. If $j = 2$ and $p = 2$, we also have $\text{ord}_2(c^4/4) > 1$ and so (5.2.6) simplifies to

$$p^{-2}d_2^{p^2} + p^{-1}d_1^p + d_0 \equiv 0 \quad \text{and} \quad p^{-1}d_2^p + d_1 - \epsilon_p a^p \equiv 0. \tag{5.2.7}$$

Let $x = d_0/a$ in R_a , $y = d_1/b$ in R_b and $z = d_2/c$ in R_c . Then (5.2.5) and (5.2.7) give (5.2.3), using the equations for a, b, c in Remark 5.1.4(iii). It follows that $f_a(z) = 0$ in R_c for f_a given by (5.2.4). When $\epsilon_p = 0$, we have

$$\text{ord}_p(z) = \frac{1}{p^4} \text{ord}_p(wa^{-p^2}) \geq -\frac{(p^2 - 1)}{p^4} \text{ord}_p(a) = -\frac{p + 1}{p^4(p^2 + 1)}$$

and thus

$$\text{ord}_p\left(\binom{p^2}{j} p^{(p-1)j} z^{p^4}\right) = (p - 1)j + 2 - \text{ord}_p(j) + p^4 \text{ord}_p(z) \geq 1,$$

i.e., the middle terms of the binomial expansion for $f_a(z)$ drop out.

Conversely, if $f_a(z) = 0$ in R_c and x and y are defined by (5.2.3)(i) and (ii), then (5.2.3)(iii) holds and we obtain a D_k -homomorphism belonging to a point of W in the fiber over P . □

Notation 5.2.8. If λ in k^\times is fixed, then a in

$$\mathfrak{R}_\lambda(F) = \{a \in \mathcal{O}_F/p\mathcal{O}_F \mid \lambda^{p^2} a^{p^4} \equiv (-p)^{p+1} a \pmod{p^{p+2}\mathcal{O}_F}\}$$

determines b and c in $\mathcal{O}_F/p\mathcal{O}_F$ by the congruences in Proposition 5.1.1. If z in R_c satisfies the resulting congruence $f_a(z) = 0$ in R_c , then z determines x in R_a and y in R_b by (5.2.3). Using the congruences in (5.2.5), set $\mathbf{d}_a(z) = (\vec{0}, d_4, d_3, cz, by, ax)$. Let φ_z be the D_k -homomorphism such that $\varphi_z(e_1) = (\vec{0}, c, b, a)$ and $\varphi_z(e_5) = \mathbf{d}_a(z)$ and let Q_z be the corresponding point in W . The fiber field generated by the point of W lying over the point P_a of E is $F_a = F(Q_z)$.

We next examine the effect of various choices of lifts on constructing a generator for the extension F_a/F . Under the assumptions of Notation 5.2.8, choose lifts to \mathcal{O}_K of λ and the entries in \mathbf{s} . By Remark 5.1.4(i), a has a lift \tilde{a} in \mathcal{O}_F . Using the congruences in Proposition 5.1.1 as equations, \tilde{a} determines lifts \tilde{b} and \tilde{c} in \mathcal{O}_F of b and c . Let \tilde{f} be the polynomial with coefficients in F obtained by using the respective lifts to replace the corresponding coefficients of $f_a(Z)$.

Corollary 5.2.9. *Construct $\tilde{f}(Z)$ in $F[Z]$ by choosing the lifts described above. If θ is any root of \tilde{f} in \bar{K} , then $F_a = F(\theta)$. If $\epsilon_p = 0$ then $h(X) = X^p - X + \tilde{w}\tilde{a}^{-p^2}$ splits completely in F_a .*

Proof. Let M be the splitting field of \tilde{f} over F . Since the p^4 solutions to the congruence $f_a(Z) = 0$ in R_c correspond to the distinct points of W in the fiber over P_a , the roots of \tilde{f} in \bar{K} remain distinct when reduced to R_c . If θ is any root of \tilde{f} , its reduction z in R_c determines the point Q_z . Thus F_a is contained in M . If γ is in $\text{Gal}(M/F_a)$, then $Q_z = \gamma(Q_z) = Q_{\gamma(z)}$, so $\gamma(z) = z$ in R_c . But then $\gamma(\theta) = \theta$, since the roots of \tilde{f} are distinct modulo $\frac{p}{c}\mathcal{O}_{\bar{K}}$. Hence $F_a = F(\theta) = M$ is independent of the various choices of lifts.

When $\epsilon_p = 0$, we have $p^4 \text{ord}_p(\theta) = \text{ord}_p(w/a^{p^2}) \geq (1-p^2) \text{ord}_p(a)$. We find that $\alpha = \theta^{p^3} + \theta^{p^2} + \theta^p + \theta$ satisfies

$$\alpha^p - \alpha \equiv \theta^{p^4} - \theta \equiv -wa^{-p^2} \pmod{\frac{p}{a^{p^2}}\mathcal{O}_L},$$

since the worst case middle term in the binomial expansion of α^p leads to

$$\text{ord}_p(p\theta^{p^3(p-1)}\theta^{p^2}) = 1 + (p^4 - p^3 + p^2) \text{ord}_p(\theta) \geq 1 - p^2 \text{ord}_p(a).$$

Hence $h(\alpha) \equiv 0 \pmod{pa^{-p^2}\mathcal{O}_L}$. Upon clearing denominators, Hensel's lemma [Lang 1994, II, §2] implies that h has a root in F_a and the other roots come by refining $\alpha + j$ with $1 \leq j \leq p - 1$. \square

A polynomial g_a of degree p^4 , analogous to f_a , but such that y in R_b satisfies $g_a(y) = 0$ in R_b , can also be derived from Proposition 5.2.2 as in the Corollary below. Then y determines x in R_a and z in R_b and thus Q_z . Choosing appropriate lifts leads to $\tilde{g}(Y)$ in $F[Y]$, such that a root of \tilde{g} also generates the extension F_a/F . Similar considerations apply to x .

Corollary 5.2.10. *Let $\mathbf{s} = [s_1 0000]$ and choose lifts $\tilde{\lambda}, \tilde{s}_1$ in \mathcal{O}_K and \tilde{a} in \mathcal{O}_F . Then $F_a = F(\vartheta)$ for any root ϑ in \bar{K} of $\tilde{g}(Y) = Y^{p^4} - Y - p\tilde{\lambda}^{-p}\tilde{s}_1\tilde{a}^{p-p^3}$. In addition, $h(X) = X^p - X - p\tilde{\lambda}^{-p}\tilde{s}_1\tilde{a}^{p-p^3}$ splits completely in F_a .*

Proof. By assumption, $w = 0$ and $\epsilon_p = s_1$. It suffices to treat $s_1 \neq 0$. In the proof of Proposition 5.2.2, we showed that $d_1^p = 0$ in $\mathcal{O}_{\bar{K}}/p\mathcal{O}_{\bar{K}}$. Hence

$$\text{ord}_p(y) = \text{ord}_p\left(\frac{d_1}{b}\right) \geq \frac{1}{p} - (p^2 - p + 1) \text{ord}_p(a) = -\frac{1}{p} \text{ord}_p(a).$$

Then $\text{ord}_p(y) > \text{ord}_p(pa^{p-p^3})$ and so $\text{ord}_p(z^p) = \text{ord}_p(pa^{p-p^3}) = (1-p)/(p^2+1)$ by (5.2.3)(ii). It follows that

$$\text{ord}_p(p^{p-1}z^{p^2}) = p - 2 + \frac{p+1}{p^2+1}. \tag{5.2.11}$$

Since (5.2.11) is positive, (5.2.3)(i) and (5.2.3)(iii) imply that

$$\text{ord}_p(y) = \frac{1}{p} \text{ord}_p(x) = \frac{1}{p^3} \text{ord}_p(z) = -\frac{1}{p^4} \left(\frac{p-1}{p^2+1} \right). \tag{5.2.12}$$

By (5.2.11), if $p \geq 3$, the term $p^{p-1}z^{p^2}$ drops out of (5.2.3)(i) and then we deduce from (5.2.3) that $g_a(y) = y^{p^4} - y + p\lambda^{-p}s_1a^{p-p^3}$ is 0 in R_b . If $p = 2$, apply Lemma 2.6 to (5.2.3)(ii) to obtain $x^4 = y^8$ in R_c . Thus $z = y^{16}$ in R_c and it again follows that $g_a(y) = 0$ in R_b . Conversely, from y satisfying $g_a(y) = 0$

in R_b , we can find x and z such that (5.2.3) holds. The concluding arguments are analogous to those in the proof of Corollary 5.2.9. □

We have focused on x, y, z in Proposition 5.2.2 because, as we show next, distinct solutions to $f_a(Z) = 0$ in R_c differ by elements of μ_{p^4-1} .

Lemma 5.2.13. *Let Q_z lie in the fiber over $P_a \neq 0$. Then every other point in the same fiber has the form $Q_{z'}$ with $z' = z + \omega$ in R_c as ω ranges over μ_{p^4-1} . If so,*

$$y' = y + \omega^p \text{ in } R_b, \quad x' = x + \omega^{p^2} \text{ in } R_a \tag{5.2.14}$$

and $Q_{z'} = Q_z + \iota(P_{a'})$ with $a' = \omega^{p^2} a$ in \mathfrak{A}_λ .

Proof. We have $f_a(z) = 0$ in R_c and we use (5.2.3)(i) and (ii) to find y and x . Putting $z' = z + \omega$ and using (5.2.14) to define y' and x' gives another solution to the congruences (5.2.3), thereby accounting for the additional $p^4 - 1$ points $Q_{z'}$ in the fiber over P_a .

Let $Q_{z'} = Q_z + \iota(P_{a'})$ and evaluate the corresponding D_k -homomorphisms at e_5 to find the equation of Witt covectors $\mathbf{d}_a(z') = \mathbf{d}_a(z) + (\vec{0}, c', b', a')$. This sum reduces to ordinary addition on coordinates in k . Indeed, apply Verschiebung twice and use Lemma 2.7 to get $cz' = cz + c'$ and so $c' = \omega c$ in k . By Remark 5.1.4(iii), c' determines b' and a' . In particular, the various lifts satisfy

$$(-p)^{p+1} a' \equiv (c')^{p^2} \equiv \omega^{p^2} c^{p^2} \equiv (-p)^{p+1} \omega^{p^2} a \pmod{p^{p+2} \mathcal{O}_{\bar{K}}}.$$

Hence $a' = \omega^{p^2} a$ in k and similarly $b' = \omega^p b$ in k . □

The next lemmas treat special cases used in the following subsection to describe Kummer generators when $p = 2$.

Lemma 5.2.15. *If $P_a \neq 0$, then the field F_a of points of the fiber over P_a equals the full field of points $K(W)$ for the Honda parameters in (5.2.16).*

Proof. If $\mathbf{s} = [s_1 s_2 000]$, use the first form of $f_a(Z)$ in Proposition 5.2.2 with $w = s_2 a$. In the remaining cases below, $\epsilon_p = 0$ and the simpler equation for $f_a(Z)$ holds. Note that $f_{\eta a}(\eta^e Z) = \eta^e f_a(Z)$ for all η in μ_{p^4-1} , with e given by

$$\frac{\mathbf{s}}{e} \begin{array}{c} \left| \begin{array}{cccc} [s_1 s_2 000] & [00 s_3 00] & [0000 s_5] & [000 s_4 0] \end{array} \right. \\ \hline \begin{array}{cccc} 1-p^2 & p^3-p^2 & p-p^2 & 0 \end{array} \end{array}. \tag{5.2.16}$$

The correspondence between the roots of $f_a(Z)$ and those of $f_{\eta a}(Z)$ induced by $z \leftrightarrow \eta^e z$ shows that $F_{\eta a} = F_a$ and so each of these fields equals $K(W)$. □

Proposition 5.2.17. *If \mathcal{W} is an extension of \mathcal{E}_λ by \mathcal{E}_λ killed by p and $L = K(W)$, then its abelian conductor exponent satisfies $\mathfrak{f}(L/F) \leq p^2$. Moreover, $\mathfrak{f}(F'/F) \leq p^2$ for every intermediate field F' of L/F .*

Proof. Let $\mathbf{s}(\mathcal{W}) = [s_1 s_2 s_3 s_4 s_5]$ and write $s_1 = \epsilon_p + \delta_p$, with $\delta_p = 0$ for odd primes p and $\delta_2 = (\lambda s_2)^2$. Then $\mathcal{W} = \mathcal{W}_1 + \cdots + \mathcal{W}_5$ is a Baer sum of group schemes corresponding to the sum of Honda parameters

$$[\epsilon_p 0000] + [\delta_p s_2 000] + [00 s_3 00] + [000 s_4 0] + [0000 s_5], \tag{5.2.18}$$

some of which may be trivial. For the fiber fields $F_a^{(j)}$ of each of these W_j , we show that $\mathfrak{f}(F_a^{(j)}/F) \leq p^2$ in the next lemmas. Since F_a is contained in the compositum of all $F_a^{(j)}$, we then have $\mathfrak{f}(F_a/F) \leq p^2$ by Lemma C.9. Furthermore, L is the compositum of all F_a as P_a varies over E_λ , so $\mathfrak{f}(L/F) \leq p^2$. Finally $\mathfrak{f}(F'/F) \leq p^2$ because the upper ramification numbering behaves well for quotients. \square

Remark 5.2.19. In contrast to the proposition, Fontaine’s higher ramification bound leads to $\mathfrak{f}(L/F) \leq p^2 + 2$ by Proposition C.12, since Proposition 5.1.1(ii) gives $e_{F/K} = e_F = (p^2 + 1)(p - 1)$. In particular, when $p = 2$, the sharper bound is essential for our applications.

We next verify the lemmas needed for the proof of the Proposition. For $P_a \neq 0$ and f_a as in Proposition 5.2.2, recall that $F_a = F(Q_z)$, where $f_a(z) = 0$ in R_c . Let π_a be a uniformizer of F_a .

Lemma 5.2.20. *If $\mathbf{s} = [000s_40]$, then F_a/F is unramified of degree 1 or p .*

Proof. The claim follows from separability of $f_a(Z) = Z^{p^4} - Z + s_4$ over k . \square

Lemma 5.2.21. *For the parameters \mathbf{s} below, F_a/F is totally ramified of degree p^4 .*

- (i) *If $\mathbf{s} = [s_1 0000]$ with $s_1 \neq 0$, then $\mathfrak{f}(F_a/F) = p^2 - 2p + 2$.*
- (ii) *Let $\mathbf{s} = [s_1 s_2 s_3 s_4 s_5]$, with $s_2 \neq 0$. Set $s_1 = 0$ for odd p and $s_1 = (\lambda s_2)^2$ for $p = 2$. Then $\epsilon_p = 0$ for all p and $\mathfrak{f}(F_a/F) = p^2$.*
- (iii) *If $\mathbf{s} = [00s_3s_40]$ and $s_3 \neq 0$, then $\mathfrak{f}(F_a/F) = p$.*
- (iv) *If $\mathbf{s} = [000s_4s_5]$ and $s_5 \neq 0$, then $\mathfrak{f}(F_a/F) = p$.*

Proof. To find the conductors, we determine t in F_a to which Proposition C.5 applies. In all cases below, $g(t) - t$ is in μ_{p^4-1} for all $g \neq 1$ in $\text{Gal}(F_a/F)$ by Lemma 5.2.13 and $F_a = F(t)$.

In case (i), let $F_a = F(\vartheta)$ as in Corollary 5.2.10 and let y be the image of ϑ in R_b . Observe that by (5.2.12), F_a/F is totally ramified of degree p^4 and we have $\text{ord}_{\pi_a}(y) = \text{ord}_p(y) \text{ord}_{\pi_a}(p) = -(p - 1)^2$. Using $t = y$ gives $\mathfrak{f}(F_a/F) = p^2 - 2p + 2$.

In the remaining cases, $\epsilon_p = 0$ and $F_a = F(\theta)$ as in Corollary 5.2.9, with θ a root of $\tilde{f}(Z) = 0$ in $\mathcal{O}_{\bar{K}}$ and \tilde{f} a lift of the simpler version of f_a in Proposition 5.2.2. If z is the image of θ in R_c , then $p^4 \text{ord}_p(z) = \text{ord}_p(w) - p^2 \text{ord}_p(a)$ and so we have

case	(ii)	(iii)	(iv)
$\text{ord}_p(z)$	$-\frac{p+1}{p^4(p^2+1)}$	$-\frac{1}{p^4(p^2+1)}$	$-\frac{1}{p^3(p^2+1)}$

In cases (ii) and (iii), observe that F_a is totally ramified of degree p^4 over F , with $\text{ord}_{\pi_a}(z) = 1 - p^2$ and $1 - p$ respectively. We use $t = z$ to determine $\mathfrak{f}(F_a/F)$.

In case (iv), $w = s_5a^p + s_4a^{p^2}$. Choose $\beta \in \mathbb{W}^\times$ such that $\beta^p \equiv s_5 \pmod{p\mathbb{W}}$ and let $t = \theta^{p^3} + \beta a^{1-p}$. By Lemma 2.6, with O -notation from the start of Section 5,

$$t^p = \theta^{p^4} + s_5a^{p-p^2} + O(\pi_a) = \theta - s_4 + O(\pi_a),$$

so $\text{ord}_p(t) = 1/p \text{ord}_p(\theta) = 1/(p^4(p^2 + 1))$. Hence the ramification index of $F(t)/F$ is at least p^4 . Since $F(t) \subseteq F_a$ and $[F_a : F] \leq p^4$, we have $F_a = F(t)$, totally ramified over F . If $g(z) = z + \omega$ as in Lemma 5.2.13, then

$$g(t) - t = g(\theta)^{p^3} - \theta^{p^3} \in (\theta + \omega + \pi_a \mathcal{O}_{\bar{K}})^{p^3} - \theta^{p^3} \subseteq \omega^{p^3} + \pi_a \mathcal{O}_{\bar{K}}.$$

Proposition C.5 therefore applies with $\text{ord}_{\pi_a}(t) = 1 - p$ to give $f(F_a/F) = p$. □

5.3. Local corners. For this subsection, $p = 2$ and $K = \mathbb{Q}_2$. Let \mathcal{E} be the simple group scheme \mathcal{E}_λ of Notation 4.3, with $\lambda = 1$ necessarily. Let E be the Galois module of \mathcal{E} , $F = \mathbb{Q}_2(E)$ and $\Delta = \text{Gal}(F/\mathbb{Q}_2)$. By Proposition 5.1.1, $F = \mathbb{Q}_2(\mu_{15}, \varpi)$, with uniformizer ϖ satisfying $\varpi^5 = 2$. Fix a generator σ of the inertia subgroup of Δ and a Frobenius τ generating $\text{Gal}(F/\mathbb{Q}(\varpi))$ with $\tau\sigma\tau^{-1} = \sigma^2$. Then $\Delta = \langle \sigma, \tau \rangle$ is isomorphic to the Frobenius group of order 20 and E is the unique nontrivial irreducible module over $R = \mathbb{F}_2[\Delta]$.

Let W represent a class in $\text{Ext}_{[2], \mathbb{Q}_2}^1(E, E)$, $L = \mathbb{Q}_2(W)$ and $\mathfrak{h} = \text{Hom}_{\mathbb{F}_2}(E, E)$. Then $[W]$ corresponds to a cohomology class $[\psi]$ in $H^1(\text{Gal}(L/\mathbb{Q}_2), \mathfrak{h})$ such that

$$\rho_W(g) = \begin{bmatrix} \rho_E(g) & \psi(g)\rho_E(g) \\ 0 & \rho_E(g) \end{bmatrix} \quad \text{for all } g \in \text{Gal}(L/\mathbb{Q}_2), \tag{5.3.1}$$

as in (B.1). We introduce *corners* to rigidify ψ and facilitate comparison with the cocycles arising from global extensions.

Suppose that V is any finitely generated R -module and let $T_\sigma = \sigma^4 + \sigma^3 + \sigma^2 + \sigma + 1$ in R be the trace with respect to σ . Since σ has odd order, $V = V_0 \oplus V'$, where V_0 is the submodule on which σ acts trivially and $V' = \ker T_\sigma = (\sigma - 1)(V)$. The *corner subgroup* of V , which depends on the choice of τ , is defined as

$$\text{Cor}(V) = \{v \in V \mid \tau(v) = v \text{ and } T_\sigma(v) = 0\}.$$

If v_1, \dots, v_n is an \mathbb{F}_2 -basis for $\text{Cor}(V)$, then $Rv_i \simeq E$ and $V' = \bigoplus_{i=1}^n Rv_i$.

We consistently write P for the unique nonzero element of $\text{Cor}(E)$, so $P = P_\varpi$ as in Proposition 5.1.1(iii) and $P, \sigma(P), \sigma^2(P), \sigma^3(P)$ is an \mathbb{F}_2 -basis for E affording the matrix representations

$$s = \rho_E(\sigma) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad t = \rho_E(\tau) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{5.3.2}$$

We will also use the twisted action of \mathbb{F}_{16} on E described in Remark 5.1.4. If a primitive fifth root of unity ζ in \mathcal{O}_F is defined by $\sigma(\varpi) = \zeta\varpi$, then $\sigma(\alpha P) = \alpha\zeta P$ and $\tau(\alpha P) = \tau(\alpha)P$ for all α in \mathbb{F}_{16} .

The endomorphisms s and t belong to \mathfrak{h} , with respective minimal polynomials $s^4 + s^3 + s^2 + s + 1 = 0$ and $t^4 - 1 = 0$. We next describe \mathfrak{h} as an R -module.

Lemma 5.3.3. *An \mathbb{F}_2 -basis for $\mathfrak{h}_0 = \ker((\sigma - 1) | \mathfrak{h})$ is $1, s, s^2, s^3$, with τ acting on \mathfrak{h}_0 as one Jordan block. An \mathbb{F}_2 -basis for $\text{Cor}(\mathfrak{h})$ is t, t^2, t^3 . We have $\mathfrak{h} \simeq \mathfrak{h}_0 \oplus_{j=1}^3 \text{Rt}^j$, with each $\text{Rt}^j \simeq E$. The cohomology group $H^1(\Delta, \mathfrak{h})$ vanishes.*

Proof. The elements of \mathfrak{h}_0 are precisely the $\mathbb{F}_2[s]$ -endomorphisms of E . Since E is a cyclic $\mathbb{F}_2[s]$ -module, $\text{End}_{\mathbb{F}_2[s]}(E) = \mathbb{F}_2[s] \simeq \mathbb{F}_{16}$. The action of τ on \mathfrak{h}_0 is the action of Frobenius on \mathbb{F}_{16} and thus has one Jordan block. Similarly, the elements of $\text{Cor}(\mathfrak{h})$ are $\mathbb{F}_2[t]$ -endomorphisms of E , so contained in $\mathbb{F}_2[t]$. But only the linear combinations of t, t^2, t^3 are annihilated by the action of T_σ on \mathfrak{h} .

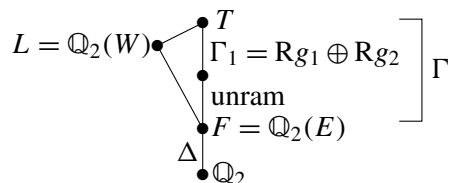
We have $H^1(\langle \tau \rangle, \mathfrak{h}_0) = H^1(\langle \tau \rangle, \mathbb{F}_{16}) = 0$ by the additive Hilbert Theorem 90 and $H^1(\langle \sigma \rangle, \mathfrak{h}) = 0$ because σ has odd order. Applying inflation-restriction with respect to the exact sequence $1 \rightarrow \langle \sigma \rangle \rightarrow \Delta \rightarrow \langle \tau \rangle \rightarrow 1$ shows that $H^1(\Delta, \mathfrak{h}) = 0$. □

Notation 5.3.4. For t as in (5.3.2), the following elements comprise $\text{Cor}(\mathfrak{h})$. Their labels are consistent with Notation 6.1.2.

$$\begin{aligned} \gamma_0 = 0, \quad \gamma_4 = t + t^2 + t^3, \quad \gamma_5 = t + t^3, \quad \gamma_9 = t^2, \\ \gamma_{11} = t + t^2, \quad \gamma'_{11} = t^2 + t^3, \quad \gamma_{15} = t^3, \quad \gamma'_{15} = t. \end{aligned} \tag{5.3.5}$$

All occur as values of extension cocycles for E by E when we range over Honda parameters, see Proposition 5.3.12 below.

Motivated by the conductor bound in Proposition 5.2.17, we assume from now on that $f_p(L/F) \leq 4$. If T is the maximal elementary 2-extension of F with ray class conductor exponent 4, then T is Galois over \mathbb{Q}_2 and we denote the action of δ in Δ on elements h of $\Gamma = \text{Gal}(T/F)$ by ${}^\delta h = \tilde{\delta} h \tilde{\delta}^{-1}$ independent of the choice of lift $\tilde{\delta}$ of δ to $\text{Gal}(T/\mathbb{Q}_2)$. We also write σ for an element of order 5 in $\text{Gal}(T/\mathbb{Q}_2)$ projecting to σ in Δ . We have the following diagram of fields and Galois groups,



where Γ_1 is the wild ramification subgroup (see Appendix C) of Γ . We next describe the complete lower ramification filtration on Γ and its structure as a module for $R = \mathbb{F}_2[\Delta]$.

Proposition 5.3.6. *Let $g_0 = \text{Artin}(\varpi, T/F)$, $g_1 = \text{Artin}(1 + \varpi + \varpi^3, T/F)$ and $g_2 = \text{Artin}(1 + \varpi^3, T/F)$. Then $\Gamma = \text{R}g_0 \oplus \text{R}g_1 \oplus \text{R}g_2 \simeq \mathbb{F}_2 \oplus E \oplus E$ and*

$$\Gamma_1 \triangleright \Gamma_2 = \Gamma_3 \triangleright \Gamma_4 = \{1\},$$

with $\Gamma_1 = \text{R}g_1 \oplus \text{R}g_2$ and $\Gamma_3 = \text{R}g_2$. There is a Frobenius Φ of order 8 in $\text{Gal}(T/\mathbb{Q}_2)$ projecting to τ in Δ and satisfying $\Phi\sigma\Phi^{-1} = \sigma^2$. In addition, $\text{Gal}(T/\mathbb{Q}_2) = \Gamma_1 \rtimes H$ with $H = \langle \sigma, \Phi \rangle$.

Proof. We use the standard filtration $U_F^{(n)}$ on local units, see (C.1). The \mathbf{R} -module structure of Γ follows from the class field theory isomorphism

$$\text{Artin}(-, T/F) : F^\times / U_F^{(4)} F^{\times 2} \xrightarrow{\sim} \Gamma.$$

In particular, \mathbf{R} acts trivially on the Frobenius g_0 of Γ , while $\mathbf{R}g_1$ and $\mathbf{R}g_2$ are isomorphic to E as \mathbf{R} -modules. Since $\Gamma_1 = \text{Artin}(U_F, T/F)$, we have $\Gamma_1 = \mathbf{R}g_1 \oplus \mathbf{R}g_2$ and similarly for Γ_2 , using $U_F^{(2)} \subset U_F^{(3)} F^{\times 2}$. Note that

$$\Gamma_1 = \ker(T_\sigma | \Gamma) = \text{Image}((\sigma - 1) | \Gamma).$$

There is a residue extension of degree 2 for T/F , so Frobenius Φ projecting to τ has order 8. Set $\Phi\sigma^3\Phi^{-1} = h\sigma$ for some h in Γ_1 . By direct computation, $T_\sigma(h) = (h\sigma)^5 = (\Phi\sigma^3\Phi^{-1})^5 = 1$. Hence $h = {}^\sigma x/x$ for some x in Γ_1 and so $(x\Phi)\sigma^3(x\Phi)^{-1} = \sigma$. Replace Φ by $x\Phi$ to guarantee that $\Phi\sigma\Phi^{-1} = \sigma^2$. Then Δ acts trivially on Φ^4 , so $\Phi^4 = g_0$. Since $H = \langle \sigma, \Phi \rangle$ is isomorphic to the Galois group of the maximal tame extension of F in T , we find that $\text{Gal}(T/\mathbb{Q}_2)$ is a semidirect product of H by the normal subgroup Γ_1 . □

Let $r_{T/L} : \text{Gal}(T/\mathbb{Q}_2) \twoheadrightarrow \text{Gal}(L/\mathbb{Q}_2)$ be the natural projection. Note that the inertia group $\text{Gal}(L/F)_1$ of $\text{Gal}(L/F)$ is the wild ramification subgroup $\text{Gal}(L/\mathbb{Q}_2)_1$ of $\text{Gal}(L/\mathbb{Q}_2)$.

Corollary 5.3.7. *The subgroup $\bar{H} = r_{T/L}(\langle \sigma, \Phi \rangle)$ of $\text{Gal}(L/\mathbb{Q}_2)$ projects onto Δ in $\text{Gal}(F/\mathbb{Q}_2)$. As \mathbf{R} -modules, $\text{Gal}(L/F)_1 \simeq E^b$, with $0 \leq b \leq 2$.*

- (i) *If L/F is totally ramified, then $\text{Gal}(L/F) = \text{Gal}(L/F)_1$ and $|\bar{H}| = 20$.*
- (ii) *Otherwise, L/F has residue degree 2, $\text{Gal}(L/F) \simeq \text{Gal}(L/F)_1 \oplus \mathbb{F}_2$ and \bar{H} has order and exponent 40.*

Proof. That \bar{H} projects onto Δ and that $\text{Gal}(L/F)_1 = r_{T/L}(\Gamma_1)$ is the direct sum of at most 2 copies of E is immediate. Moreover, L/F is totally ramified if and only if $g_0 = \Phi^4$ is in $\ker r_{T/L}$. Thus $|\bar{H}| = 20$ in case (i) and 40 in case (ii). □

Since T contains $L = \mathbb{Q}_2(W)$, the cocycle ψ in (5.3.1) inflates to $\text{Gal}(T/\mathbb{Q}_2)$. We may arrange for $\psi(\sigma) = 0$, since σ has odd order. Lemma 5.3.3 and (B.3) give injectivity of the restriction map:

$$0 \rightarrow H^1(\text{Gal}(T/\mathbb{Q}_2), \mathfrak{h}) \xrightarrow{\text{res}} H^1(\Gamma, \mathfrak{h})^\Delta = \text{Hom}_{\mathbf{R}}(\Gamma, \mathfrak{h}) \tag{5.3.8}$$

and we say that $\chi = \text{res}([\psi])$ in $\text{Hom}_{\mathbf{R}}(\Gamma, \mathfrak{h})$ belongs to W . Note that χ is determined by its values on g_0, g_1, g_2 , as defined in Proposition 5.3.6.

Lemma 5.3.9. *The field $L = \mathbb{Q}_2(W)$ is the fixed field of $\ker \chi$. Moreover:*

- (i) $\chi(g_i)$ is in $\text{Cor}(\mathfrak{h})$ for $i = 1, 2$ and $\chi(g_0)$ is in $\{0, I_4\}$.
- (ii) L/F is unramified if and only if $\chi(g_1) = \chi(g_2) = 0$.
- (iii) $\mathfrak{f}(L/F) = 4$ if and only if $\chi(g_2) \neq 0$. If $\chi(g_2) = 0$, then $\mathfrak{f}(L/F) = 0$ or 2.
- (iv) The residue degree of L/F is 1 or 2, according to whether $\chi(g_0) = 0$ or I_4 .

Proof. The matrix representation (5.3.1) shows that g in $\text{Gal}(T/\mathbb{Q}_2)$ acts trivially on W if and only if g is in $\Gamma = \text{Gal}(T/F)$ and $\chi(g) = 0$. Then items (i)–(iv) immediately follow from Proposition 5.3.6. In particular, (i) holds by considering the action of Δ on g_0, g_1 and g_2 . \square

Write \mathcal{W}_s for the extension of \mathcal{E} by \mathcal{E} of exponent 2 with Honda parameter s and W_s for its Galois module. Belonging to W_s are the cohomology class $[\psi_s]$ in $H^1(\text{Gal}(T/\mathbb{Q}_2), \mathfrak{h})$ and its restriction χ_s in $\text{Hom}_{\mathbb{F}_2[\Delta]}(\Gamma, \mathfrak{h})$, as described above. The rest of this section is devoted to evaluating χ_s as s varies.

If h is in $\Gamma = \text{Gal}(T/F)$ and Q_{z_j} is any point in the fiber over $\sigma^j(P)$, see Notation 5.2.8, any basis of the form

$$P, \sigma(P), \sigma^2(P), \sigma^3(P), Q_{z_0}, Q_{z_1}, Q_{z_2}, Q_{z_3} \tag{5.3.10}$$

yields the same matrix $\rho_{W_s}(h)$ in (5.3.1). Moreover, $h(Q_{z_j}) = Q_{z_j} + \chi_s(h)\sigma^j(P)$.

Let M/F be a finite elementary 2-extension. Define its *Kummer group* by

$$\kappa(M/F) = F^\times \cap M^{\times 2} \quad \text{and let} \quad \bar{\kappa}(M/F) = \kappa(M/F)/F^{\times 2}.$$

By definition, $F^{\times 2} \subseteq \kappa(M/F)$ and we have $M = F(\{\sqrt{\theta} \mid \theta \in \kappa(M/F)\})$. Kummer theory gives a perfect pairing

$$\text{Gal}(M/F) \times \bar{\kappa}(M/F) \rightarrow \mu_2 \quad \text{by} \quad (g, \theta) \mapsto g(\sqrt{\theta})/\theta.$$

Lemma 5.3.11. *Let $P = P_\varpi$ and let F_ϖ be the subfield of L generated by the points of W_s in the fiber over P . If $s = [10000]$, then $\kappa(F_\varpi/F)$ contains $1 + 2\varpi^4$. If $s_1 = s_2$, then $\kappa(F_\varpi/F)$ contains $1 + 2s_2\varpi^2 + 2(s_3 + s_5)\varpi^4$.*

Proof. Refer to Proposition 5.2.2. Since $p = 2$ and $\lambda = 1$, we have $\epsilon_2 = 0$ when $s_1 = s_2$. Then take the square class of the discriminant of the polynomial $h(X)$ in Corollary 5.2.9. Similarly, use Corollary 5.2.10 when $s = [10000]$. \square

We first determine χ_s when L/F is a nontrivial totally ramified extension. For compatibility with the notation for decomposition groups in Section 6, where we consider global Galois module extensions of E by E , set $\mathcal{D}_p(L/F) = \text{Gal}(L/F)$.

Proposition 5.3.12. *If L/F is totally ramified, then $\chi_s(g_0) = 0$. Depending on the conductor exponent $\mathfrak{f}(L/F)$, we have:*

- (i) $\mathfrak{f}(L/F) = 2$. Then $|\mathcal{D}_p(L/F)| = 16$, $\chi_s(g_2) = 0$ and

s	[00001]	[00100]	[10000]	[10101]	[00101]	[10001]	[10100]
$\chi_s(g_1)$	γ_{15}	γ'_{15}	γ_9	γ_4	γ_5	γ'_{11}	γ_{11}

- (ii) $\mathfrak{f}(L/F) = 4$ and $|\mathcal{D}_p(L/F)| = 16$. Then $\chi_s(g_2) = \gamma_9$ and $\chi_s(g_1) = 0$ or γ_9 according to whether $s = [11000]$ or $[01000]$.

(iii) $f(L/F) = 4$ and $|\mathcal{D}_p(L/F)| = 256$. Then $\chi_s(g_2) = \gamma_9$ and

\mathbf{s}	[11001]	[11100]	[01101]	[11101]	[01001]	[01100]
$\chi_s(g_1)$	γ_{15}	γ'_{15}	γ_4	γ_5	γ'_{11}	γ_{11}

Proof. We begin with some basic Honda parameters, from which the others can be generated by Baer sum. Recall that F_a denotes the extension of F obtained by adjoining the coordinates of the points in the fiber of W_s above one point P_a of order 2 in E .

Basic cases: (1) $\mathbf{s} = [00001]$, $[00100]$ or $[10000]$. By Lemma 5.2.21, F_a/F is totally ramified of degree 16 and $f(F_a/F) = 2$. Thus $\chi_s(g_0) = \chi_s(g_2) = 0$ by Lemma 5.3.9 and so $L = F_a$ is the subfield of T fixed by $R_{g_0} \oplus R_{g_2}$ independent of a .

(2) $\mathbf{s} = [11000]$. Lemma 5.2.15 indicates that $L = F_a$ does not depend on a . Now L/F is totally ramified of degree 16 and $f(L/F) = 4$ by Lemma 5.2.21, so $\chi_s(g_0) = 0$ but $\chi_s(g_2) \neq 0$. By Lemma 5.3.11, the Kummer group $\bar{\kappa}(L/F)$ contains the coset $\kappa = (1 + 2\varpi^2)F^{\times 2}$ and therefore equals $R\kappa$. By evaluating the pairing of Kummer theory and class field theory given by Hilbert symbols, we find that g_1 acts trivially on the square roots of elements of $\bar{\kappa}(L/F)$, so $\chi_s(g_1) = 0$.

Set $h = g_1$ in the basic case (1) and $h = g_2$ in (2). Recall that the primitive fifth root of unity ζ is defined by $\sigma(\varpi) = \zeta\varpi$. To find the matrix $\chi_s(h)$, we use a basis for W_s of the form

$$P, \sigma(P), \sigma^2(P), \sigma^3(P), Q_{z_0}, Q_{z_1}, Q_{z_2}, Q_{z_3},$$

where z_j is a root of the Honda polynomial $f_{\zeta^j\varpi}$, see Notation 5.2.8. The action of $\Delta = \text{Gal}(F/\mathbb{Q}_2)$ puts h in the corner group of $\mathcal{D}_p(L/F)$, so $\chi_s(h)$ is in $\text{Cor}(h)$ and therefore equals one of the matrices in (5.3.5). In particular, $\chi_s(h)(P) = \alpha_0 P$, with $\alpha_0 = 0$ or 1. Write $h(Q_{z_j}) = Q_{z_j} + \alpha_j P$, where $\alpha_0 = 0$ or 1 and

$$\alpha_j = c_{0j} + c_{1j}\zeta + c_{2j}\zeta^2 + c_{3j}\zeta^3 \text{ in } \mathbb{Z}[\zeta], \quad \text{for } 1 \leq j \leq 3.$$

Then the $(j+1)$ -column of the matrix $\chi_s(h)$ is $[c_{0j}, c_{1j}, c_{2j}, c_{3j}]^T \pmod 2$ by (5.3.10).

From $h(Q_{z_0}) = Q_{z_0} + \alpha_0 P$, we get $h(z_0) = z_0 + \alpha_0$ by Lemma 5.2.13. In the proof of Lemma 5.2.15, we showed that there is a correspondence between roots of f_ϖ and $f_{\zeta^j\varpi}$, allowing us to choose $z_j = \zeta^{je}z_0$, with e given by (5.2.16) and $j = 1, 2, 3$. Then $h(z_j) = z_j + \alpha_0\zeta^{je}$ in R_c . Since h is not trivial on L , we have $\alpha_0 = 1$. Further use of Lemma 5.2.13 gives

$$h(Q_{z_j}) = Q_{z_j} + \zeta^{4je} P_{\zeta^j\varpi} = Q_{z_j} + \zeta^{(1-e)j} P.$$

This determines $\chi_s(h)$ for all \mathbf{s} in the basic cases.

Remaining cases. Write $\mathbf{s} = \mathbf{t} + \mathbf{u}$, choosing Honda parameters \mathbf{t} and \mathbf{u} already treated above. Then W_s is the Baer sum of W_t and W_u and $\chi_s = \chi_t + \chi_u$.

In (ii), use $[01000] = [11000] + [10000]$. In (i), the last three entries follow by varying \mathbf{t} and \mathbf{u} among first three entries. Use $[10101] = [10000] + [00101]$ to complete (i). For (iii), let $\mathbf{t} = [11000]$ and let \mathbf{u}

run over the Honda parameters in (i), omitting [10000]. Since g_1 and g_2 are independent and nontrivial on L , we have $\text{Gal}(L/F) = Rg_1 \oplus Rg_2$ of order 256. \square

We briefly treat the remaining 16 nontrivial Honda parameters, even though Lemma 6.1.14 shows that they are not needed for our global applications.

Proposition 5.3.13. *If L/F is not totally ramified, then $\mathbf{s} = \mathbf{t} + \mathbf{u}$, where \mathbf{t} ranges over [00000] and the 15 Honda parameters in Proposition 5.3.12, while $\mathbf{u} = [00010]$. Then $\chi_{\mathbf{s}}(g_0) = I_4$, $\chi_{\mathbf{s}}(g_j) = \chi_{\mathbf{t}}(g_j)$ for $j = 1, 2$ and $\mathbb{Q}_2(W_{\mathbf{s}})$ is the compositum of $\mathbb{Q}_2(W_{\mathbf{t}})$ and the unramified quadratic extension of F .*

Proof. By Lemma 5.2.20, $F(W_{\mathbf{u}})$ is the splitting field of $Z^{16} - Z - 1$, namely the unramified quadratic extension of F . Thus $\chi_{\mathbf{u}}(g_0) = I_4$ and $\chi_{\mathbf{u}}(g_1) = \chi_{\mathbf{u}}(g_2) = 0$ by Lemma 5.3.9. The rest follows from $\chi_{\mathbf{s}} = \chi_{\mathbf{t}} + \chi_{\mathbf{u}}$. \square

6. Global conclusions

6.1. Favorable abelian surfaces. There are two irreducible \mathcal{S}_5 -representations of dimension 4 over \mathbb{F}_2 . Denote the one taking transpositions to transvections by $\iota : \mathcal{S}_5 \rightarrow \text{SL}_4(\mathbb{F}_2)$ and fix it by sending (12) $\mapsto r$ and (12345) $\mapsto s$, where

$$r = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad s = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \quad \text{Let } t = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (6.1.1)$$

The image of ι is isomorphic to the odd orthogonal group $\text{O}_4^-(\mathbb{F}_2) \subset \text{Sp}_4(\mathbb{F}_2)$. In addition, $\iota((2354)) = t$ and $\Delta = \langle s, t \rangle$ is the Frobenius group of order 20.

Fix a favorable quintic field F_0 with discriminant $d_{F_0/\mathbb{Q}} = \pm 16N$ and Galois closure F . Proposition 1.2(i) implies that the inertia group $\mathcal{I}_v(F/\mathbb{Q})$ at each place $v \mid N$ is generated by a transposition σ_v when we identify $\text{Gal}(F/\mathbb{Q})$ with \mathcal{S}_5 . In this section, E is the Galois module giving $\rho_E : \text{Gal}(F/\mathbb{Q}) = \mathcal{S}_5 \xrightarrow{\iota} \text{SL}_4(\mathbb{F}_2)$. Using the matrices r, s, t in (6.1.1), σ_v is conjugate to $\rho_E^{-1}(r)$, inertia at a some $\mathfrak{p} \mid 2$ is generated by $\sigma = \rho_E^{-1}(s)$ and $\tau = \rho_E^{-1}(t)$ is a Frobenius in the decomposition group $\mathcal{D}_{\mathfrak{p}}(F/\mathbb{Q}) = \langle \sigma, \tau \rangle$. Hence the restriction of ρ_E to $\mathcal{D}_{\mathfrak{p}}(F/\mathbb{Q})$ agrees with the representation $\rho_{E_{\lambda}}$ of Definition 4.1, as normalized in (5.3.2). By Corollary 5.1.5, E extends to a group scheme \mathcal{E} over $\mathbb{Z}[\frac{1}{N}]$. Let \underline{E} be the Σ -category introduced in Definition 3.4 with $\Sigma = \{\mathcal{E}\}$. This subsection is devoted to criteria for the validity of axiom **E4** in Theorem 3.7, needed to prove Theorem 6.1.22.

To treat extensions W of E by E of exponent 2, let $\mathcal{P} = \mathcal{P}_{E,E}$ be the parabolic group as in (B.2). We describe subgroups of \mathcal{P} in which the relevant representations ρ_W take their values.

Notation 6.1.2. Let $c : \text{Mat}_4(\mathbb{F}_2) \rightarrow \mathcal{P}$ by $c(m) = \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix}$ and $d : \mathcal{S}_5 \rightarrow \mathcal{P}$ by $d(g) = \begin{bmatrix} \iota(g) & 0 \\ 0 & \iota(g) \end{bmatrix}$. Let G_0 be the image of d . With γ_a as in 5.3.4 and $S = \mathbb{F}_2[\mathcal{S}_5]$, we define S -submodules of $\text{Mat}_4(\mathbb{F}_2) = \text{End}(E)$ with adjoint action of \mathcal{S}_5

$$\Gamma_4 = S\gamma_4, \quad \Gamma_5 = S\gamma_5, \quad \Gamma_9 = S\gamma_9, \quad \Gamma_{11} = S\gamma_{11}, \quad \Gamma_{15} = S\gamma_{15}. \quad (6.1.3)$$

Let $G_a = \langle G_0, c(\gamma_a) \rangle = c(\Gamma_a) \rtimes G_0$.

The radical of G_a equals $c(\Gamma_a)$ and has size 2^a . The abelianization of G_a is cyclic of order 2 and so defines the character $\epsilon_0 : G_a \rightarrow \mathbb{F}_2$, generalizing the additive signature on S_5 . If $a = 0$ or 4, all automorphisms of G_a are inner. The center of the other G_a 's is generated by $c(1)$ and there is an automorphism

$$\epsilon : G_a \rightarrow G_a \quad \text{by} \quad \epsilon(g) = gc(1)^{\epsilon_0(g)}. \tag{6.1.4}$$

When $a = 5$ or 9, $\text{Aut}(G_a)$ is generated by ϵ , modulo automorphisms induced from conjugation by elements of the normalizer of G_a in \mathcal{P} .

The corner group of an $\mathbb{F}_2[\Delta]$ -module consists of the elements fixed by t and annihilated by the trace T_s . Using Magma, we find the nonzero corners of Γ_a .

a	4	5	9	11	15
$\text{Cor}(\Gamma_a) - \{\gamma_0\}$	$\{\gamma_4\}$	$\{\gamma_5\}$	$\{\gamma_4, \gamma_5, \gamma_9\}$	$\{\gamma_5, \gamma_{11}, \gamma'_{11}\}$	$\{\text{all } \gamma_i\}$

(6.1.5)

Inclusions among the groups G_a follow from this table and are indicated in the Hasse diagram by ascending lines.



Moreover, G_9 is isomorphic to the fiber product of G_4 and G_5 over G_0 and similarly for the other parallelograms. When an inclusion $G_b \subset G_a$ exists, Magma extends the identity on G_0 to a surjection $f_{a,b} : G_a \twoheadrightarrow G_b$ sending γ_a to γ_b .

Definition 6.1.7. An involution g in a group H is *good* if its conjugates generate H . If g is good in $H \subseteq \mathcal{P}$ and $\text{rank}(g - 1) = 2$, then g is *very good*.

Remark 6.1.8. A Magma verification shows that each G_a has a unique conjugacy class of very good involutions, represented by $d(r)$ with r as in (6.1.1).

Proposition 6.1.9. Let L be an elementary 2-extension of $F = \mathbb{Q}(E)$, Galois over \mathbb{Q} , with L/F unramified outside $\{2, \infty\}$ and $f_p(L/F) \leq 4$ for all $p \mid 2$. Then:

- (i) The maximal subfield of L abelian over \mathbb{Q} is $\mathbb{Q}(\sqrt{N^*})$, with $N^* = \pm N \equiv 5(8)$.
- (ii) For $v \mid N$, inertia $\mathcal{I}_v(L/\mathbb{Q})$ is generated by a good involution in $\text{Gal}(L/\mathbb{Q})$.

Proof. By Proposition 1.2, F contains $\sqrt{N^*}$. For $v \mid N$, the inertia group $\mathcal{I}_v(F/\mathbb{Q})$ has order 2. Since L/F is unramified, $\mathcal{I}_v(L/\mathbb{Q})$ is generated by an involution σ_v . Intermediate fields $L \supseteq F' \supseteq F$ satisfy $f_p(F'/F) \leq f_p(L/F) \leq 4$. But Lemma C.6 implies that $f_p(F(i)/F) = 6$ and $f_p(F(\sqrt{\pm 2})/F) = 11$, so $L \cap F(i, \sqrt{2}) = F$. Since L/\mathbb{Q} is unramified outside $\{2, N, \infty\}$, item (i) follows from Kronecker–Weber. The subfield of L fixed by the normal closure of σ_v is unramified outside $\{2, \infty\}$ and is contained in $\mathbb{Q}(i)$ by [Brumer and Kramer 2001], so equals \mathbb{Q} . Thus (ii) holds. □

Corollary 6.1.10. For $[W]$ in $\text{Ext}_{[2],\mathbb{Q}}^1(E, E)$, assume that $L = \mathbb{Q}(W)$ satisfies the hypotheses in the proposition and $\text{rank } \rho_W(\sigma_v - 1) = 2$. Then $\rho_W(\text{Gal}(L/\mathbb{Q}))$ is one of the groups G_a , up to conjugation in \mathcal{P} . If $[W]$ is in $\text{Ext}_{[2],E}^1(\mathcal{E}, \mathcal{E})$, then $\text{Gal}(\mathbb{Q}(W)/\mathbb{Q})$ is conjugate to some G_a .

Proof. By the Proposition $\rho_W(\sigma_v)$ is good and so is very good by assumption. Magma verifies that the G_a represent the six conjugacy classes of subgroups of \mathcal{P} that project onto \mathcal{S}_5 and admit very good involutions. If $[W]$ is a class in $\text{Ext}_{[2],E}^1(\mathcal{E}, \mathcal{E})$, then the Proposition applies to $L = \mathbb{Q}(W)$, since $f_{\mathfrak{p}}(L/F) \leq 4$ by Proposition 5.2.17 and $\text{rank } \rho_W(\sigma_v - 1) = 2$ by **E3** of Definition 3.4. \square

Definition 6.1.11. A class $[W]$ in $\text{Ext}_{[2],\mathbb{Q}}^1(E, E)$ with $L = \mathbb{Q}(W)$ is a G_a -class if L/F is unramified outside $\{2, \infty\}$, $f_{\mathfrak{p}}(L/F) \leq 4$ for $\mathfrak{p} \mid 2$ and $\text{rank } \rho_W(\sigma_v - 1) = 2$, so that $\rho_W(\text{Gal}(L/\mathbb{Q})) = G_a$ for some a by the corollary.

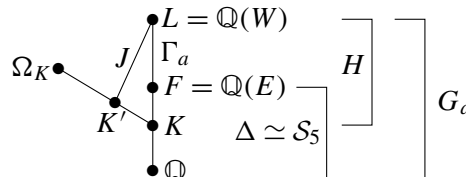
Lemma 6.1.12. Let $[W]$ be a G_a -class with $L = \mathbb{Q}(W)$.

- (i) If $[W']$ is a $G_{a'}$ -class, with $L' = \mathbb{Q}(W')$, then the Baer sum $[W''] = [W] + [W']$ is a G_b -class for some b .
- (ii) If $f_{a,b} : G_a \rightarrow G_b$ exists in (6.1.6), then the Galois module for $f_{a,b}\rho_W$ represents a G_b -class.

Proof. In (i), $[W]$ and $[W']$ correspond to classes $[\psi]$ and $[\psi']$ in $H^1(G_{\mathbb{Q}}, \mathfrak{h})$ as in (B.1) and $[W'']$ belongs to the class of $\psi'' = \psi + \psi'$. Since $L'' = \mathbb{Q}(W'')$ is a subfield of the compositum LL' , the ramification properties required of L'' in Definition 6.1.11 hold. Proposition 6.1.9 shows that $\rho(\sigma_v)$ is a good involution in G_a and so is very good, conjugate to $d(r)$ by Remark 6.1.8. Similarly for $\rho_{W'}(\sigma_v)$ in $G_{a'}$. Hence the representatives ψ and ψ' can be chosen to satisfy $\psi(\sigma_v) = \psi'(\sigma_v) = 0$. We now have $\psi''(\sigma_v) = 0$ and so $\text{rank } \rho_{W''}(\sigma_v - 1) = 2$. By Corollary 6.1.10, $[W'']$ is a G_b -class for some b .

For (ii), let L' be the subfield of L fixed by $\rho_W^{-1}(\ker f_{a,b})$. Then $f_{a,b}\rho_W$ induces an isomorphism $\rho' : \text{Gal}(L'/\mathbb{Q}) \rightarrow G_b$. The required ramification conditions hold for the subfield L' of L . As above, $\rho'(\sigma_v \mid L')$ is a good involution in G_b . Since $f_{a,b}$ is the identity on G_0 and $\rho_W(\sigma_v)$ is conjugate to $d(r)$ in G_0 so is $\rho'(\sigma_v \mid L')$. \square

Let $K = \mathbb{Q}(r_1 + r_2)$ be a pair-resolvent field for $F = \mathbb{Q}(E)$, as defined before Theorem 1.3, namely the fixed field of $\text{Sym}\{1, 2\} \times \text{Sym}\{3, 4, 5\}$. Let $\Omega_K = \Omega_K^{(4)}$ be the maximal elementary 2-extension of K of modulus $\mathfrak{p}^4\infty$, where \mathfrak{p} is the unique prime over 2 in K and ∞ allows ramification at all archimedean places. Refer to the following diagram of fields and Galois groups.



To simplify notation, also write \mathfrak{p} for a place over 2 in L and for the restrictions of \mathfrak{p} to subfields of L . Note that primes over 2 are unramified in F/K . Suppose that L is the Galois closure of K'/\mathbb{Q} . By

Lemma C.11 with M, F, K', K_1 and K there equal to the respective p -adic completions of L, F, K', K and \mathbb{Q} here, $f_p(K'/K) = f_p(L/F)$.

Proposition 6.1.13. *Let K be a pair-resolvent of F . There is a bijection*

$$\{G_a\text{-classes } [W] \text{ with } a \in \{4, 5, 9\}\} \leftrightarrow \{\text{subfields } K' \subseteq \Omega_K \text{ quadratic over } K\}$$

such that $\mathbb{Q}(W)$ is the Galois closure of K'/\mathbb{Q} .

Proof. For $v \mid N$, $\mathcal{I}_v(F/K)$ acts on the left cosets of $\text{Gal}(F/K)$ in $\text{Gal}(F/\mathbb{Q})$ with four fixed points and three orbits of size 2. Thus $(N)\mathcal{O}_K = \mathfrak{a}\mathfrak{b}^2$ where \mathfrak{a} and \mathfrak{b} are square-free, relatively prime ideals of \mathcal{O}_K of absolute norms N^4 and N^3 respectively.

Let $[W]$ be a G_a -class with a in $\{4, 5, 9\}$, $L = \mathbb{Q}(W)$ and $\rho_W : \text{Gal}(L/\mathbb{Q}) \xrightarrow{\sim} G_a$. Then $H = \text{Gal}(L/K)$ is the inverse image under $\pi : G_a \rightarrow \mathcal{S}_5$ of $\text{Gal}(F/K)$. Choose $v \mid N$ so that if σ_v generates $\mathcal{I}_v(L/\mathbb{Q})$, then $\pi(\rho_W(\sigma_v)) = (12)$. By assumption $g = \rho_W(\sigma_v)$ is very good in G_a . Magma shows that among the subgroups of index 2 in H , exactly one, say J , has the property that the action of G_a on G_a/J is faithful and g has exactly 8 fixed points in this action. Hence $K' = L^J$ is a stem field for L and in view of the factorization of $(N)\mathcal{O}_K$, no prime over N ramifies in K'/K . If $v' \mid N$ is any other choice such that $\pi(\rho_W(\sigma_{v'})) = (12)$, then $\sigma_{v'}$ is conjugate to σ_v in H and therefore gives the same J , so also the same K' . Since $f_p(K'/K) = f_p(L/F) \leq 4$ by definition of a G_a -class, K' is contained in Ω_K .

Conversely, let K' be a subfield of Ω_K quadratic over K , L the Galois closure of K'/\mathbb{Q} , $G = \text{Gal}(L/\mathbb{Q})$, $H = \text{Gal}(L/K)$ and $J = \text{Gal}(L/K')$. Then L properly contains F , since each quadratic extension of K in F ramifies at some prime over N . By Proposition 6.1.9(ii), σ_v is a good involution in G . Since no prime over N ramifies in K'/K , the action of σ_v on G/J has eight fixed points. The following group-theoretic properties of G have been established:

- (i) There is a surjection $\pi : G \rightarrow \mathcal{S}_5$ whose kernel has exponent 2 and is the radical of G .
- (ii) The abelianization of G has order 2.
- (iii) If H is the inverse image under π of the centralizer of a transposition in \mathcal{S}_5 , then there is a subgroup J of index 2 in H such that the action of G on G/J is faithful.
- (iv) There is a good involution g in G whose action on G/J has 8 fixed points.

We have (i) since the radical of $\text{Gal}(L/\mathbb{Q})$ is $\text{Gal}(L/F)$ and (ii) by Proposition 6.1.9(i).

In the Magma database of 1117 transitive groups of degree 20 only three satisfy (i)–(iv), namely G_a with a in $\{4, 5, 9\}$. Furthermore, if J is the stabilizer in \mathcal{S}_{20} of any letter, then there is a unique conjugacy class of good involutions g in G such that g acts on G/J with exactly 8 fixed points. By applying this construction to $G = \text{Gal}(L/\mathbb{Q})$, there is an isomorphism $\rho : \text{Gal}(L/\mathbb{Q}) \rightarrow G_a$ such that $\rho(\sigma_v)$ is conjugate to g and has 8 fixed points when acting on $G_a/\rho(J)$. Computation now shows the following. If $a = 4$, then $\rho(\sigma_v)$ is conjugate to $d(r)$. If a is in $\{5, 9\}$, then $\rho(\sigma_v)$ is conjugate to $d(r)$ or $d(r)\epsilon(r)$ where ϵ is the automorphism of (6.1.4). In the latter case, replace ρ by $\epsilon \circ \rho$. If W is the associated Galois module,

then its class is a G_a -class. Because any automorphism of G_a preserving the conjugacy class of $d(r)$ is conjugation by an element of \mathcal{P} , the class $[W]$ is unique. \square

Unless otherwise stated, $[W]$ now denotes a G_a -class and $L = \mathbb{Q}(W)$. Thus W represents a class in $\text{Ext}_{R'}^1(\mathcal{E}, \mathcal{E})$, where $R' = \mathbb{Z}[\frac{1}{2N}]$. By the Mayer–Vietoris sequence (2.2), W prolongs to a group scheme \mathcal{W} over $R = \mathbb{Z}[\frac{1}{N}]$ if and only if the image of $[W]$ in $\text{Ext}_{\mathbb{Q}_2}^1(\mathcal{E}, \mathcal{E})$ agrees with that of a class from $\text{Ext}_{\mathbb{Z}_2}^1(\mathcal{E}, \mathcal{E})$. If so, the other conditions in Definition 6.1.11 guarantee that $[\mathcal{W}]$ is in $\text{Ext}_{\underline{E}}^1(\mathcal{E}, \mathcal{E})$. Recall that $\mathfrak{h} = \text{Hom}_{\mathbb{F}_2}(E, E)$ and let $\psi : G_{\mathbb{Q}} \rightarrow \mathfrak{h}$ represent the class in $H^1(G_{\mathbb{Q}}, \mathfrak{h})$ associated to $[W]$, as in (B.1). Recall that at $\mathfrak{p} \mid 2$, the decomposition group $\mathcal{D}_{\mathfrak{p}}(F/\mathbb{Q})$ is isomorphic to $\Delta = \langle s, t \rangle$.

Lemma 6.1.14. *As a Δ -module, $\mathcal{D}_{\mathfrak{p}}(L/F)$ is isomorphic to E^b with $b \leq 2$.*

Proof. We may assume $\mathcal{D}_{\mathfrak{p}}(L/F) \neq 1$. Computation shows that G_a contains no subgroup of order and exponent 40 whose projection to S_5 has order 20. Conclude by using Proposition 5.3.6 and its Corollary 5.3.7. \square

Remark 6.1.15. Let $[\psi]$ in $H^1(G_{\mathbb{Q}}, \mathfrak{h})$ correspond to the G_a -class $[W]$ and write $\psi|_{\mathcal{D}_{\mathfrak{p}}}$ for the restriction to the decomposition group $\mathcal{D}_{\mathfrak{p}}$ in $G_{\mathbb{Q}}$ at a fixed place \mathfrak{p} over 2. The classes $[\mathcal{W}_{\mathfrak{s}}]$ in $\text{Ext}_{\mathbb{Z}_2}^1(\mathcal{E}, \mathcal{E})$ are classified by their Honda parameters \mathfrak{s} in $(\mathbb{F}_2)^5$. Let $[\psi_{\mathfrak{s}}]$ in $H^1(G_{\mathbb{Q}_2}, \mathfrak{h})$ correspond to $[\mathcal{W}_{\mathfrak{s}}]$. Then $[W]$ is compatible with $[\mathcal{W}_{\mathfrak{s}}]$ if and only if:

$$[\psi|_{\mathcal{D}_{\mathfrak{p}}}] = [\psi_{\mathfrak{s}}] \text{ in } H^1(G_{\mathbb{Q}_2}, \mathfrak{h}) \text{ for some Honda parameter } \mathfrak{s}. \tag{6.1.16}$$

Let $F_{\mathfrak{p}}$ be the completion of F at \mathfrak{p} and T the maximal elementary 2-extension of $F_{\mathfrak{p}}$ having conductor exponent 4. By Proposition 5.2.17, $\mathbb{Q}_2(W_{\mathfrak{s}})$ is contained in T , while the completion $L_{\mathfrak{p}}$ is contained in T by definition of a G_a -class. In the diagram below, inflation is injective and restriction is injective by (5.3.8):

$$\begin{array}{ccc} & H^1(\mathcal{D}_{\mathfrak{p}}, \mathfrak{h}) & \\ & \downarrow \text{inf} & \\ 0 & \longrightarrow H^1(\text{Gal}(T/\mathbb{Q}_2), \mathfrak{h}) \xrightarrow{\text{res}} \text{Hom}_{\mathbb{F}_2[\Delta]}(\text{Gal}(T/F_{\mathfrak{p}}), \mathfrak{h}). & \end{array} \tag{6.1.17}$$

Hence, it suffices to compare the image χ of $[\psi|_{\mathcal{D}_{\mathfrak{p}}}]$ with the image $\chi_{\mathfrak{s}}$ of $[\psi_{\mathfrak{s}}]$ in $\text{Hom}_{\mathbb{F}_2[\Delta]}(\text{Gal}(T/F_{\mathfrak{p}}), \mathfrak{h})$. Note that the values of χ and $\chi_{\mathfrak{s}}$ are corners in \mathfrak{h} . See Proposition 5.3.6 for specific generators g_0, g_1, g_2 of Γ as an $\mathbb{F}_2[\Delta]$ -module. In particular, $\chi(g_0) = 0$ by Lemmas 6.1.14 and 5.3.9(iii). Thus W prolongs to a group scheme over $R = \mathbb{Z}[\frac{1}{N}]$ exactly if there is a Honda parameter \mathfrak{s} in Proposition 5.3.12 satisfying $\chi(g_j) = \chi_{\mathfrak{s}}(g_j)$ for $j = 1, 2$.

Lemma 6.1.18. *Let $[W]$ be a G_a -class and $L = \mathbb{Q}(W)$.*

- (i) *If $\mathfrak{f}_{\mathfrak{p}}(L/F) \leq 2$ for $\mathfrak{p} \mid 2$, then W prolongs to a group scheme \mathcal{W} over R .*
- (ii) *If $a \in \{4, 5, 11\}$ and W prolongs to a group scheme over R , then $\mathfrak{f}_{\mathfrak{p}}(L/F) \leq 2$.*

Proof. Refer to Remark 6.1.15 for notation. In item (i), we have $\chi(g_2) = 0$ by Lemma 5.3.9(ii). To match χ with $\chi_{\mathfrak{s}}$ for some local Honda parameter \mathfrak{s} , we therefore consider \mathfrak{s} in Proposition 5.3.12(i), also

allowing $\mathfrak{s} = 0$. As \mathfrak{s} varies, $\chi_{\mathfrak{s}}(g_1)$ ranges over all possible corners of \mathfrak{h} and we can find a unique \mathfrak{s} such that $\chi_{\mathfrak{s}}(g_1) = \chi(g_1)$. Hence W prolongs to a group scheme \mathcal{W} over R .

In item (ii), G_a does not contain γ_9 by (6.1.5). Then $\chi(g_2) = 0$, to match $\chi_{\mathfrak{s}}(g_2)$ for some Honda parameter \mathfrak{s} in Proposition 5.3.12. Hence $f_p(L/F) \leq 2$. □

Definition 6.1.19. Let K be a pair-resolvent of F and Ω_K the maximal elementary 2-extension of K unramified outside $\{2, \infty\}$ such that $f_p(\Omega_K/K) \leq 4$ for $p \mid 2$. We say F is *amiable* if either (i) $\Omega_K = K$ or (ii) $[\Omega_K : K] = 2$ and $f_p(\Omega_K/K) = 4$.

Remark 6.1.20. For F to be amiable, all the following conditions are necessary: (i) The narrow class number of K is odd. (ii) If $a \in (1 + \mathfrak{p}^9)K_p^{\times 2}$, then $a \in K^{\times 2}$, since $f_p(K(\sqrt{a})/K) \leq 2$ by Lemma C.6. (iii) K is not totally real; otherwise $\text{rank } U_K/U_K^2 = 10$, but $\text{rank } U_p/(1 + \mathfrak{p}^9)U_p^2 = 8$.

Proposition 6.1.21. *Let \mathcal{E} be the group scheme introduced at the beginning of this section. Then $\text{Ext}_{[2],E}^1(\mathcal{E}, \mathcal{E}) = 0$ if and only if $F = \mathbb{Q}(E)$ is amiable.*

Proof. Suppose that F is amiable and let $[W]$ be a nontrivial class in $\text{Ext}_{[2],E}^1(\mathcal{E}, \mathcal{E})$. By Corollary 6.1.10, $[W]$ is G_a -class with $a \neq 0$. If $a = 11$, then $f_p(L/F) \leq 2$ by Lemma 6.1.18(ii). By diagram (6.1.6) and Lemma 6.1.12(ii), there is a G_5 -class $[W']$ with $L' = \mathbb{Q}(W')$ contained in L . Proposition 6.1.13 provides a quadratic extension K' of K contained in Ω_K with $f_p(K'/K) = f_p(L'/F) \leq f_p(L/F) \leq 2$, contradicting the amiability of F . The same argument applies when $a = 4$ or 5 . If $a = 15$ or 9 , then $[W]$ gives rise to both a G_4 -class and a G_5 -class. Then Proposition 6.1.13 provides two distinct quadratic extensions of K contained in Ω_K , again contradicting the amiability of F .

Suppose that F is not amiable. Assume first that $[\Omega_K : K] = 2$ and let $[W]$ be the G_a -class corresponding to Ω_K/K by Proposition 6.1.13. By amiability, $f_p(\Omega_K/K) \leq 2$ and so $f_p(L/F) = f_p(\Omega_K/K) \leq 2$. Then Lemma 6.1.18(i) implies that W prolongs to a nontrivial class in $\text{Ext}_{[2],E}^1(\mathcal{E}, \mathcal{E})$. Next, assume that there is a G_a -class $[W]$ with $L = \mathbb{Q}(W)$ and a $G_{a'}$ -class $[W']$ with $L' = \mathbb{Q}(W')$, coming from distinct quadratic extensions of K in Ω_K and satisfying $a, a' \in \{4, 5, 9\}$. Since a G_9 -class gives rise to a G_4 -class and a G_5 -class, we need only consider the pairs (a, a') in $\{(4, 4), (5, 5), (4, 5)\}$. In the notation of Remark 6.1.15, let χ and χ' in $\text{Hom}_{\mathbb{F}_2[\Delta]}(\text{Gal}(T/F_p), \mathfrak{h})$ belong to W and W' respectively. Then the Baer sum $W'' = W + W'$ represents a G_b -class by Lemma 6.1.12 and $\chi'' = \chi + \chi'$ belongs to W'' . By Lemma 6.1.18(i) and Lemma C.11, we may assume that $f_p(L/F) = f_p(L'/F) = 4$ and so $\chi(g_2)$ and $\chi'(g_2)$ are nontrivial, by Lemma 5.3.9(ii). In all these cases, only one nontrivial corner is available in (6.1.5), namely $\chi(g_2) = \gamma_a$ and $\chi'(g_2) = \gamma_{a'}$. If $a = a' = 4$ or 5 , then $\chi''(g_2) = 0$ and so $f(L''/F) \leq 2$. Thus W'' prolongs to a group scheme over $\mathbb{Z}[\frac{1}{N}]$. If $(a, a') = (4, 5)$, then $\chi''(g_2) = \gamma_4 + \gamma_5 = \gamma_9$, so χ'' is compatible with $\chi_{\mathfrak{s}}$ for some \mathfrak{s} in Proposition 5.3.12(i) or (ii) and the corresponding group scheme exists. □

Theorem 6.1.22. *Let A be a favorable abelian surface of prime conductor N such that $F = \mathbb{Q}(A[2])$ is amiable. If B is a semistable abelian variety of dimension $2d$ and conductor N^d , with $B[2]$ filtered by $A[2]$, then B is isogenous to A^d .*

Proof. By Proposition 1.2, $\mathcal{E} = A[2]$ satisfies the conditions in Definition 3.4 for a Σ -category \underline{E} with $\Sigma = \{\mathcal{E}\}$. Then Theorem 3.7 applies, since $\text{Ext}_{[2], \underline{E}}^1(\mathcal{E}, \mathcal{E}) = 0$ by Proposition 6.1.21 and $\text{End}(A) = \mathbb{Z}$ because A has prime conductor [Brumer and Kramer 2014]. \square

6.2. Elliptic curves of prime conductor, supersingular at 2. We briefly note how Theorem 3.7 applies to elliptic curves. Let A be an elliptic curve of prime conductor N with supersingular reduction at 2 and $\mathcal{E} = A[2]$. Then $F = \mathbb{Q}(E)$ is an \mathcal{S}_3 -extension and E is an irreducible Galois module even locally over \mathbb{Q}_2 . The only two irreducible $\mathbb{F}_2[\mathcal{S}_3]$ modules are the trivial one and E .

Proposition 6.2.1. *Let K be a cubic subfield of $F = \mathbb{Q}(E)$ and let \mathfrak{p} be the prime in K above 2. A necessary and sufficient condition for $\text{Ext}_{[2], \underline{E}}^1(\mathcal{E}, \mathcal{E}) = 0$ is that there be no quadratic extension of K of dividing conductor $\mathfrak{p}^2 \cdot \infty$.*

Proof. Only two subgroups of the parabolic group $\mathcal{P}_{E, E}$ admit good involutions. One is isomorphic to \mathcal{S}_3 and corresponds to the split extension of \mathcal{E} by itself because $H^1(\mathcal{S}_3, \text{End}(E)) = 0$ while the second is isomorphic to \mathcal{S}_4 . If M is the field of points of an extension of \mathcal{E} by \mathcal{E} annihilated by 2 and $\text{Gal}(M/\mathbb{Q}) \simeq \mathcal{S}_4$, then M is the Galois closure of a quadratic extension of K unramified at primes over p . The bound for the local conductor over 2 is given in [Schoof 2003, Proposition 6.4] and Theorem 3.7 applies. A related proof is in [Schoof 2005] for $A = J_0(N)$ with $N = 11$ and 19. \square

In the Cremona database, we find 2037 isogeny classes of elliptic curves supersingular at 2 and of prime conductor $N < 350000$. From the Brumer–McGuinness database [1990], we extract an additional 2422 isogeny classes for a total of 4459 such classes with $N \leq 10^8$. Applying the proposition above, we find 847 elliptic curves A to which Theorem 3.7 applies.

Let A_1 and A_2 be elliptic curves of prime conductor N with each $\mathcal{E}_i = A_i[2]$ biconnected over \mathbb{Z}_2 and satisfying $\text{Ext}_{[2], \underline{E}}^1(\mathcal{E}_i, \mathcal{E}_i) = 0$. Suppose that the cubic subfields K_i of $\mathbb{Q}(E_i)$ are nonisomorphic. Then $2\mathcal{O}_{K_1 K_2}$ has the prime factorization $(\mathfrak{p}_1 \mathfrak{p}_2 \mathfrak{p}_3)^3$. If $K_1 K_2$ admits no quadratic extension of conductor dividing $(\mathfrak{p}_1 \mathfrak{p}_2 \mathfrak{p}_3)^2 \infty$, then $\text{Ext}_{\underline{E}}^1(\mathcal{E}_1, \mathcal{E}_2) = 0$. We found 42 conductors N with multiple A_i to which our results apply.

As an entertaining example, Cremona’s database lists four elliptic curves of conductor 307, with $A_1 = 307A1$, $A_2 = 307C1$ and $A_3 = 307D1$ supersingular at 2. Their 2-division fields correspond to the three subfields of the ray class field of $k = \mathbb{Q}(\sqrt{-307})$ of modulus $2\mathcal{O}_k$.

Theorem 3.7 implies the following. Let B be a semistable abelian variety, good outside $N = 307$, with $B[2]^{\text{ss}} = A_1[2]^{n_1} \oplus A_2[2]^{n_2} \oplus A_3[2]^{n_3}$ for some n_i . Then B is isogenous to $A_1^{n_1} \times A_2^{n_2} \times A_3^{n_3}$. Note that we need not impose the conductor $f_N(B) = \sum n_i f_N(A_i) = \sum n_i$, thanks to Remark 3.9.

Appendix A: A cohomology computation in the old style

Let $T = \Lambda[G]$ be the group ring of a finite group G over a discrete valuation ring Λ with prime element π and finite residue field k of characteristic p . We consider a cocycle approach to $\text{Ext}_{\Lambda[G]}^1(E, E)$. Let V and W be finitely generated T -modules such that $\pi V = \pi W = 0$. A *symmetric cocycle* is a function

$f : V \times V \rightarrow W$ satisfying

$$f(v_1, v_2) = f(v_2, v_1) \quad \text{and} \quad f(v_1, v_2) + f(v_1 + v_2, v_3) = f(v_1, v_2 + v_3) + f(v_2, v_3),$$

for v 's in V , as in [Eilenberg and MacLane 1942, Theorem 7.1]. Coboundaries are symmetric cocycles such that

$$f(v_1, v_2) = g(v_1) + g(v_2) - g(v_1 + v_2),$$

for some function $g : V \rightarrow W$. The symmetric cocycle f is *enhanced* if there is a function $h : T \times V \rightarrow W$ satisfying the following for v 's in V and r, s in T :

- (i) $rf(v_1, v_2) - f(rv_1, rv_2) = h(r, v_1) + h(r, v_2) - h(r, v_1 + v_2)$.
- (ii) $h(rs, v) = rh(s, v) + h(r, sv)$.
- (iii) $f(rv, sv) = h(r + s, v) - h(r, v) - h(s, v)$.

The cohomology classes of enhanced cocycles form a k -vector space $\mathcal{D}(V, W)$.

Lemma A.1. *The functor from T -modules to abelian groups induces an exact sequence*

$$0 \rightarrow \text{Ext}_{[\pi], T}^1(V, W) \rightarrow \text{Ext}_T^1(V, W) = \mathcal{D}(V, W) \rightarrow \text{Hom}_T(V, W),$$

where $\text{Ext}_{[\pi], T}^1(V, W)$ consists of classes of extensions annihilated by π .

Proof. Let $0 \rightarrow W \xrightarrow{i} M \xrightarrow{j} V \rightarrow 0$ be an exact sequence of T -modules with $\pi V = \pi W = 0$. Let $\sigma : V \rightarrow M$ be a section of j such that $\sigma(0) = 0$. The associated cocycle is defined by $f(v_1, v_2) = \sigma(v_1) + \sigma(v_2) - \sigma(v_1 + v_2)$. If r is in T , then $h(r, v) = r\sigma(v) - \sigma(rv)$ turns f into an enhanced cocycle. For the converse, give $W \times V$ the structure of a T -module by setting

$$(w_1, v_1) + (w_2, v_2) = (w_1 + w_2 + f(v_1, v_2), v_1 + v_2), \quad r(w, v) = (rw + h(r, v), rv).$$

Hence $\text{Ext}_T^1(V, W) = \mathcal{D}(V, W)$. Given f as above, let $\iota : V \rightarrow W$ be defined by $\iota(a) = h(\pi, a)$. Since $\pi(w, v) = (\iota(v), 0)$ and π is in the center of T , we conclude that ι is a T -homomorphism and that the sequence is exact. \square

Using the lemma, we give a refined variant of [Schoof 2012b, Lemma 2.1]. Let F be a number field and R its ring of S -integers for a finite set S of primes.

Proposition A.2. *Let \mathcal{V} and \mathcal{W} be finite flat Λ -module schemes over R killed by π , with associated Galois modules V and W . Let $\text{Ext}_{[\pi], R}^1(\mathcal{V}, \mathcal{W})$ denote the subgroup of $\text{Ext}_R^1(\mathcal{V}, \mathcal{W})$ consisting of those extensions killed by π . Then there is a natural exact sequence*

$$0 \rightarrow \text{Ext}_{[\pi], R}^1(\mathcal{V}, \mathcal{W}) \rightarrow \text{Ext}_R^1(\mathcal{V}, \mathcal{W}) \rightarrow \text{Hom}_{\text{Gal}}(V, W).$$

If V is absolutely irreducible over k , then $\text{End}_{\text{Gal}}(V) = k$.

Proof. Apply Lemma A.1 with G the Galois group of a suitable finite extension of F . Then the passage from Galois modules to the associated group schemes is as in Schoof and so is left to the reader. \square

Appendix B: Parabolic subgroups and an obstreperous cocycle

For any group G , consider representations ρ_{E_i} afforded by $\mathbb{F}_p[G]$ -modules E_i for $i = 1, 2$. If g is in G and $\delta_i = \rho_{E_i}(g)$, then g acts on m in $\mathfrak{h} = \text{Hom}_{\mathbb{F}_p}(E_2, E_1)$ by $g(m) = \delta_1 m \delta_2^{-1}$. In the category of $\mathbb{F}_p[G]$ -modules, the extension classes of E_2 by E_1 under Baer sum form a group isomorphic to $H^1(G, \mathfrak{h})$. The exact sequence of $\mathbb{F}_p[G]$ -modules $0 \rightarrow E_1 \rightarrow W \rightarrow E_2 \rightarrow 0$ gives rise to a cocycle $\psi : G \rightarrow \mathfrak{h}$ such that

$$\rho_W(g) = \begin{bmatrix} \delta_1 & \psi(g)\delta_2 \\ 0 & \delta_2 \end{bmatrix} \tag{B.1}$$

and the class $[W]$ in $\text{Ext}_{\mathbb{F}_p[G]}^1(E_2, E_1)$ corresponds to that of $[\psi]$ in $H^1(G, \mathfrak{h})$. If N is a normal subgroup of G contained in $\ker \rho_W$, then $[\psi]$ comes by inflation from a unique class in $H^1(G/N, \mathfrak{h})$, also denoted by $[\psi]$.

Note that $\rho_W(G)$ lies in a *parabolic* matrix group

$$\mathcal{P} = \mathcal{P}_{E_1, E_2} = \left\{ g = \begin{bmatrix} \delta_1 & m \\ 0 & \delta_2 \end{bmatrix} \mid \delta_i = \rho_{E_i}(g), m \in \text{Mat}_{n_1, n_2}(\mathbb{F}_p) \right\} \tag{B.2}$$

with $n_i = \dim_{\mathbb{F}_p} E_i$. If $H_i = \{g \in G \mid g|_{E_i} = 1\}$ and $\Delta_i = G/H_i$, then E_i is a faithful $\mathbb{F}_p[\Delta_i]$ -module. Any normal subgroup H of G acting trivially on both E_1 and E_2 satisfies

$$\rho_W(H) \subseteq \left\{ g = \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix} \in \mathcal{P} \mid m \in \text{Mat}_{n_1, n_2}(\mathbb{F}_p) \right\}.$$

Since $H^1(H, \mathfrak{h})^{G/H} = \text{Hom}_{\mathbb{F}_p[G/H]}(H, \mathfrak{h})$, the following sequence is exact:

$$0 \rightarrow H^1(G/H, \mathfrak{h}) \xrightarrow{\text{inf}} H^1(G, \mathfrak{h}) \xrightarrow{\text{res}} \text{Hom}_{\mathbb{F}_p[G/H]}(H, \mathfrak{h}). \tag{B.3}$$

Suppose that E_1 and E_2 are $G_{\mathbb{Q}}$ -modules, $F = \mathbb{Q}(E_1, E_2)$ and $\Delta = \text{Gal}(F/\mathbb{Q})$. If the extension $W = W_{\psi}$ belongs to a cocycle $\psi : \Delta \rightarrow \mathfrak{h}$ whose class in $H^1(\Delta, \mathfrak{h})$ is not trivial, then $\mathbb{Q}(W) = F$, even though W does not split as a Δ -module.

Example B.4. Let $p = 2$ and $E = E_1 = E_2$, with $\dim_{\mathbb{F}_2}(E) = 2n$, so that \mathfrak{h} is isomorphic to $\text{Mat}_{2n}(\mathbb{F}_2)$. As in [Brumer and Kramer 2012, Remark 2.6], equip E with the irreducible symplectic representation of $\Delta \subset \text{Sp}_{2n}(\mathbb{F}_2)$ isomorphic to \mathcal{S}_m , with transvections corresponding to transpositions and $m = 2n + 1$ or $2n + 2$. When $n \geq 2$, there is a nontrivial class $[\psi]$ in $H^1(\Delta, \mathfrak{h})$ such that $\psi(g) = \epsilon(g)I_{2n}$, where $\epsilon(g) \in \mathbb{F}_2$ is the parity of the permutation g . This situation can occur when E is the kernel of multiplication by 2 on the Jacobian of a hyperelliptic curve of genus at least 2.

Suppose further that E has prime conductor N and that σ_v generates inertia in F/\mathbb{Q} at $v \mid N$. Then σ_v is a transposition in \mathcal{S}_m , so $\psi(\sigma_v) = I_{2n}$ and it follows from (B.1) that $\text{rank } \rho_W(\sigma_v - 1) = 2n$. The extension $W = W_{\psi}$ prolongs to a group scheme over $\mathbb{Z}[\frac{1}{N}]$ satisfying **E1** and **E2**, since the local cohomology group $H^1(\mathcal{D}_{\mathfrak{p}}(F/\mathbb{Q}), \mathfrak{h}) = 0$ at $\mathfrak{p} \mid 2$, as in Lemma 5.3.3. However, the minimality assumption **E3** on our category \underline{E} requires that $\text{rank } \rho_W(\sigma_v - 1) = 2$, namely the multiplicity of E in W^{ss} , so W is not in \underline{E} when $n \geq 2$.

Appendix C: Some technical lemmas on local conductors

Let K be a finite extension of \mathbb{Q}_p with uniformizer π_K , ring of integers \mathcal{O}_K and absolute ramification index $e_K = \text{ord}_{\pi_K}(p)$. Set

$$U_K^{(n)} = \{u \in \mathcal{O}_K^\times \mid \text{ord}_{\pi_K}(u - 1) \geq n\}. \tag{C.1}$$

See [Serre 1979, IV] for basic information about ramification groups and conductors. Let L/K be a finite Galois extension. The *index* of elements g in $G = \text{Gal}(L/K)$ is given by $i_{L/K}(g) = \text{ord}_{\pi_L}(g(\theta) - \theta)$ for any choice of θ in \mathcal{O}_L such that $\mathcal{O}_L = \mathcal{O}_K[\theta]$. Then $\text{ord}_{\pi_L}(g(a) - a) \geq i_{L/K}(g)$ for all a in \mathcal{O}_K . In Serre’s *lower numbering* on ramification groups, $G_j = \{g \in G \mid i_{L/K}(g) \geq j + 1\}$. Thus $G_{-1} = G$, G_0 is the inertia group, its fixed field is the maximal unramified extension of K inside L and the p -Sylow subgroup G_1 is the wild ramification subgroup of G . For g in G_0 , we have $i_{L/K}(g) = \text{ord}_{\pi_L}(g(\pi_L) - \pi_L)$. The Herbrand function is defined by

$$\varphi_{L/K}(x) = \int_0^x \frac{ds}{[G_0 : G_s]} \tag{C.2}$$

In Serre’s *upper numbering*, $G^m = G_n$ with $m = \varphi_{L/K}(n)$.

Notation C.3. Let $c_{L/K} = \max\{j \mid G_j \neq 1\}$ and let $m_{L/K} = \varphi_{L/K}(c_{L/K})$. Thus $G^{m_{L/K}} \neq 1$ but $G^{m_{L/K} + \epsilon} = 1$ for all $\epsilon > 0$. When L/K is abelian, the conductor exponent $\mathfrak{f}(L/K)$ is the smallest integer $n \geq 0$ such that $U_K^{(n)}$ is contained in the norm group $N_{L/K}(L^\times)$.

We have $\mathfrak{f}(L/K) = m_{L/K} + 1$ by [Serre 1979, XV, §2], with $c_{L/K} = m_{L/K} = -1$ and $\mathfrak{f}(L/K) = 0$ when L/K is unramified. If M/K is a Galois extension and the intermediate field L also is Galois over K , then $m_{L/K} \leq m_{M/K}$ because $\text{Gal}(M/K)^\alpha \xrightarrow{\text{res}} \text{Gal}(L/K)^\alpha$ is surjective for all α . Translation by an unramified extension of the base does not affect the conductor, as we next recall.

Lemma C.4. *If F/K is unramified, then $m_{LF/F} = m_{L/K}$. Additionally, if L/K is abelian, then $\mathfrak{f}(LF/F) = \mathfrak{f}(L/K)$.*

Proof. The restriction map $\text{Gal}(LF/F) \xrightarrow{\text{res}} \text{Gal}(L/L \cap F)$ is an isomorphism. Since F/K is unramified, π_L also is a prime element of LF . For all $s \geq 0$, it follows from the definition of the lower numbering that restriction induces an isomorphism $\text{Gal}(LF/F)_s \xrightarrow{\sim} \text{Gal}(L/L \cap F)_s = \text{Gal}(L/K)_s$. Thus the Herbrand functions of LF/F and L/K agree and the rest is clear. \square

Proposition C.5. *Let $L = K(t)$ be Galois over K , with $\text{ord}_{\pi_L}(t) = -n$ prime to p and negative. If $g(t) - t$ is a unit for all $g \neq 1$ in G_0 , then G_0 is an elementary abelian p -group and $\mathfrak{f}(L/K) = i_{L/K}(g) = n + 1$.*

Proof. By assumption, nontrivial elements g of G_0 satisfy $g(t) = t + u$ with u a unit in \mathcal{O}_L and $g(u) \equiv u \pmod{\pi_L}$. If g has order d , then

$$t = g^d(t) = t + u + g(u) + \dots + g^{d-1}(u) \equiv t + du \pmod{\pi_L},$$

so $p \mid d$. Hence $G_0 = G_1$ is a p -group and so $i = i_{L/K}(g) \geq 2$. Furthermore, $\text{ord}_\pi(g(a) - a) \geq i$ for all a in \mathcal{O}_L .

Set $\pi = \pi_L$, $\theta = 1/t = \alpha\pi^n$ and $g(\pi) - \pi = \beta\pi^i$, where α and β are units in \mathcal{O}_L . We have the following congruences modulo $\pi^{n+i}\mathcal{O}_L$:

$$\begin{aligned} g(\theta) - \theta &= (g - 1)(\alpha\pi^n) = \alpha(g - 1)(\pi^n) + g(\pi^n)(g - 1)(\alpha) \\ &\equiv \alpha(g - 1)(\pi^n) \\ &\equiv \alpha((\pi + \beta\pi^i)^n - \pi^n) \\ &\equiv \alpha\beta n\pi^{n-1+i} \end{aligned}$$

and therefore $\text{ord}_\pi(g(\theta) - \theta) = n - 1 + i$. Explicitly,

$$g(\theta) - \theta = \frac{t - g(t)}{tg(t)} = -\frac{u}{tg(t)} = -u \cdot \theta g(\theta),$$

so $\text{ord}_\pi(g(\theta) - \theta) = 2n$. Hence $i = n + 1$ and the lower ramification sequence has only one gap: $G_0 = G_n \supsetneq G_{n+1} = \{1\}$. By ramification theory, G_n is an elementary abelian p -group and we have $f(L/K) = \varphi_{L/K}(n) + 1 = n + 1$. \square

Next, we recall the conductors of Kummer extensions of degree p .

Lemma C.6. *Let K contain μ_p and $L = K(\kappa^{1/p})$ with $\kappa \in K^\times$. Then*

$$f(L/K) = \frac{pe_K}{p-1} + 1 \quad \text{if } \text{ord}_{\pi_K}(\kappa) \not\equiv 0 \pmod{p}$$

and this is maximal for cyclic extensions of K of degree p . If $\text{ord}_{\pi_K}(\kappa - 1) = n$ with $1 \leq n < pe_K/(p-1)$ and $n \not\equiv 0 \pmod{p}$, then $f(L/K) = pe_K/(p-1) - n + 1$.

Proof. In the first case, assume without loss of generality that $\text{ord}_{\pi_K}(\kappa) = 1$, so $\theta = \kappa^{1/p}$ is a prime element for L . If $g \neq 1$ in $\text{Gal}(L/K)$, then $g(\theta) - \theta = (\zeta - 1)\pi_L$ for some a p -th root of unity ζ and the conductor follows by definition.

In the second case, set $\kappa = 1 + u\pi_K^n$ with u in U_K and $\theta = \kappa^{1/p} - 1$. Then $g(\theta) = \zeta\kappa^{1/p} - 1 = \theta + (\zeta - 1)\kappa^{1/p}$, where θ satisfies $x^p + \sum_{j=1}^{p-1} \binom{p}{j} x^j = u\pi_K^n$. Let $t = \theta/(\zeta - 1)$, to find that $g(t) - t = \kappa^{1/p}$ is a unit in L and t satisfies

$$z^p + \sum_{j=1}^{p-1} a_j z^j = \frac{u\pi_K^n}{(\zeta - 1)^p} \quad \text{with} \quad a_j = \binom{p}{j} (\zeta - 1)^{j-p}. \tag{C.7}$$

For $1 \leq j \leq p - 1$, we have

$$\text{ord}_{\pi_K}(a_j) = e_K - (p - j) \frac{e_K}{p - 1} = (j - 1) \frac{e_K}{p - 1} \geq 0.$$

Put $z = t$ in (C.7) and compare ordinals on both sides, using $p \nmid n$, to see that L/K is totally ramified of degree p and

$$\text{ord}_L(t^p) = n \text{ord}_{\pi_L}(\pi_K) - p \text{ord}_{\pi_L}(\zeta - 1) = np - p \frac{pe_K}{p-1}.$$

Thus $\text{ord}_p(t) = n - pe_K/(p-1)$ and $f(L/K)$ can be found by using Proposition C.5. \square

Remark C.8. Since the choice of κ can be changed by multiplying by a suitable element of $K^{\times p}$, the only remaining cases are $n \geq pe_K/(p - 1)$. If equality holds, then (C.7) gives an integral polynomial satisfied by t whose reduction modulo π_K has the form $z^p + \bar{a}a_1z^{p-1} - \bar{b}$ with $b = u\pi^n(\zeta - 1)^{-p}$. Since a_1 and b are unit in \mathcal{O}_K , this polynomial is separable and L/K is unramified, but possibly split. If $n > pe_K/(p - 1)$, then κ is in $K^{\times p}$ and $L = K$.

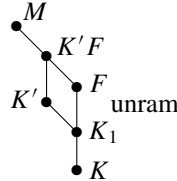
Lemma C.9. *Let L_i/K be Galois and let $m_i = m_{L_i/K}$ be the upper numbering of the last nontrivial ramification subgroup of $\text{Gal}(L_i/K)$. If $M = L_1L_2$, then $m_{M/K} = \max\{m_1, m_2\}$ and if L is a subfield of M with L/K abelian, then $\mathfrak{f}(L/K) \leq m_{M/K} + 1$.*

Proof. If $m = \max\{m_1, m_2\}$, then $m_{M/K} \geq m$. But if g is in $\text{Gal}(M/K)^\alpha$ with $\alpha > m$, then $g|_{L_i} = 1$ for $i = 1, 2$, so $g = 1$. Hence $m_{M/K} = m$. It follows that $m_{L/K} \leq m$ and therefore $\mathfrak{f}(L/K) \leq m + 1$. \square

Lemma C.10. *Assume that F/K is Galois and L/F is abelian. Let M be the Galois closure of L/K . Then M/F is abelian and $\mathfrak{f}(M/F) = \mathfrak{f}(L/F)$.*

Proof. Since $m_{L/F} \leq m_{M/F}$, we have $\mathfrak{f}(L/F) \leq \mathfrak{f}(M/F)$. If τ is in $\text{Gal}(M/K)$, then $\tau(L)/F$ is abelian and $\mathfrak{f}(\tau(L)/F) = \mathfrak{f}(L/F)$. But M is the compositum of all $\tau(L)$ as τ varies. Therefore, M/F is abelian and by Lemma C.9, $\mathfrak{f}(M/F) \leq \mathfrak{f}(L/F)$, giving equality. \square

For the next lemma, refer to the following diagram:



Lemma C.11. *Let F be the Galois closure of K_1/K and assume that F/K_1 is unramified. Let K' be an abelian extension of K_1 and let M be the Galois closure of K'/K . Then M is abelian over F and $\mathfrak{f}(M/F) = \mathfrak{f}(K'/K_1)$.*

Proof. The field M contains F because K' contains K_1 . Moreover, M is the Galois closure of $K'F/K$. Since K' is abelian over K_1 , the extension $K'F/F$ is abelian. By Lemma C.10, with L there equal to $K'F$ here, we find that M/F is abelian and $\mathfrak{f}(M/F) = \mathfrak{f}(K'F/F)$. By Lemma C.4, translation of the base via an unramified extension does not change the conductor, so $\mathfrak{f}(K'F/F) = \mathfrak{f}(K'/K_1)$. Hence $\mathfrak{f}(M/F) = \mathfrak{f}(K'/K_1)$. \square

When $L = K(V)$, where \mathcal{V} is a finite flat group scheme over \mathcal{O}_K of exponent p^n , Fontaine [1985] showed that $m_{L/K} \leq e_K(n + 1/(p - 1)) - 1$. Now consider the conductor exponent of an intermediate abelian extension.

Proposition C.12. *Let $L = K(V)$ and suppose that $K \subseteq F \subseteq F' \subseteq L$, with F'/F abelian and the relative ramification index $e_{F/K}$ equal to the tame ramification degree $[G_0 : G_1]$ of L/K . Then $\mathfrak{f}(F'/F) \leq e_F(n + 1/(p - 1)) - e_{F/K} + 1$.*

Proof. The fixed field L_1 of $H = G_1$ is the maximal subfield of L tamely ramified over K . Since $H_0 = G_1$ and $H_s = G_s$ for all $s > 0$, (C.2) gives

$$\varphi_{L/L_1}(x) = [G_0 : G_1]\varphi_{L/K}(x) = e_{F/K}\varphi_{L/K}(x), \quad \text{for all } x > 0.$$

We may assume that L properly contains L_1 . Using $c_{L/L_1} = c_{L/K}$, we have

$$m_{F'L_1/L_1} \leq m_{L/L_1} = \varphi_{L/L_1}(c_{L/L_1}) = e_{F/K}\varphi_{L/K}(c_{L/K}) = e_{F/K}m_{L/K}.$$

But F is contained in L_1 and L_1/F is unramified. Hence Lemma C.4 shows that $f(F'/F) = f(F'L_1/L_1) \leq 1 + e_{F/K}m_{L/K}$. Conclude with Fontaine's bound. \square

Appendix D: Some data

The quintic field F_0 is *amiable* if its Galois closure F is amiable as in Definition 6.1.19, so that the uniqueness in Theorem 6.1.22 applies. To check amiability, construct the pair-resolvent field K and ask Magma, under GRH, for the 2-rank of the ray class groups of K with the desired moduli, as in Theorem 1.3. A favorable abelian surface A is of *type* F_0 if $\mathbb{Q}(A[2])$ is the Galois closure of F_0 . To find representatives for isogeny classes of abelian surfaces of prime conductor N , it suffices to search for Jacobians by [Brumer and Kramer 2014, Theorem 3.4.11]. If F is amiable, then it is not totally real by Remark 6.1.20. The Magma database of quintic fields contains 1919 favorable quintic fields that are not totally real. Their absolute discriminants are at most $5 \cdot 10^6$ and 714 of them are amiable. We know Jacobians for only 82 of the latter, but expect conductors of abelian surfaces to be sparse among integers.

We tabulate explicit information for favorable fields and curves with $N < 25000$ and summarize some data for $N < 10^{10}$. In all our tables, $[a_0, a_1, a_2, \dots]$ denotes the polynomial $a_0 + a_1x + a_2x^2 + \dots$, as in Magma.

Legend for Tables 1 and 2. Table 1 on the next two pages gives a defining polynomial $f(x)$ for each of the 172 favorable quintic fields F_0 of discriminant $\pm 16N$ with $N < 25000$. Table 2 on page 1068 consists of 75 curves $y^2 = g(x)$ whose Jacobians represent distinct known isogeny classes of favorable abelian surfaces of prime conductor $N < 25000$. If C is curve number 25, 63 or 64 in that table, its leading coefficient has the form $4m^3$. These curves exhibit *mild reduction* [Brumer and Kramer 2014, p. 1162], in that C is bad at $p \mid m$ but the reduction of $J(C)$ at p is the product of two elliptic curves.

In both tables, the column marked ϵ contains an α if F_0 is amiable. For each field F_0 in Table 1, the column marked #C contains one of the following:

- The line number of a curve in Table 2 such that g has a root in F_0 .
- 0 if we can prove that no abelian surface of type F_0 exists by [Brumer and Kramer 2018].
- P if no nonlift paramodular form of that level exists, so no such surface is expected to exist.
- U if there is at most one isogeny class of that type, but it is unknown whether such an abelian surface actually exists.
- ν if F_0 is not amiable and we do not know whether or not any surface exists.

$\#F_0$	$f(x)$	N	ϵ	$\#C$	$\#F_0$	$f(x)$	N	ϵ	$\#C$
1	$[-1, -1, -2, 0, 1, 1]$	277	α	1	43	$[-2, -2, 2, 4, 2, 1]$	5309	α	U
2	$[-1, 1, 0, 0, 1, 1]$	349	α	2	44	$[-1, -3, -6, -2, 1, 1]$	5381		ν
3	$[-1, 3, 0, -2, 1, 1]$	461	α	3	45	$[3, -1, 4, 6, 1, 1]$	5437	α	0
4	$[1, 3, 2, 2, 1, 1]$	613	α	P	46	$[-2, -4, 0, 2, 2, 1]$	5651	α	26
5	$[1, 1, 2, 0, 1, 1]$	677	α	P	47	$[2, 4, -4, -4, 2, 1]$	5867		27
6	$[2, 2, 2, 2, 2, 1]$	797	α	4	48	$[2, -4, 2, -2, 2, 1]$	6277	α	U
7	$[-2, 0, 0, 0, 2, 1]$	971	α	5	49	$[-1, -1, -8, -4, 1, 1]$	6317	α	0
8	$[1, 1, 0, -2, 1, 1]$	997	α	6	50	$[-3, -5, -6, 2, 1, 1]$	6373	α	U
9	$[-1, -3, 0, 4, 1, 1]$	1051	α	7	51	$[2, 4, 0, -2, 2, 1]$	6397	α	0
10	$[2, -2, -2, 0, 2, 1]$	1061	α	U	52	$[2, -2, 0, -2, 0, 1]$	6491		28
11	$[1, -1, 2, -2, 1, 1]$	1109	α	9	53	$[2, 0, 4, 6, 0, 1]$	6701		0
12	$[-1, 3, -2, 0, 1, 1]$	1109	α	8	54	$[-2, 2, 4, -4, 0, 1]$	6763		29
13	$[-2, -4, -2, 2, 2, 1]$	1277	α	0	55	$[-1, 9, -2, -6, 1, 1]$	6907	α	U
14	$[2, -4, 4, -2, 0, 1]$	1597	α	0	56	$[2, 6, 4, 0, 0, 1]$	7013	α	U
15	$[2, -2, 2, 0, 0, 1]$	1637	α	10	57	$[-2, 0, -4, -2, 2, 1]$	7109		30
16	$[1, -3, 0, 2, 1, 1]$	1811	α	11	58	$[2, -4, -2, 4, 2, 1]$	7541	α	U
17	$[-2, 2, 2, 4, 2, 1]$	2069	α	U	59	$[2, -2, 6, 0, 0, 1]$	7549	α	U
18	$[-2, 0, 2, -2, 0, 1]$	2243	α	12	60	$[-3, 7, 2, 6, 1, 1]$	7589	α	U
19	$[3, 5, 4, 4, 1, 1]$	2269	α	U	61	$[6, 2, -8, -4, 2, 1]$	7723		ν
20	$[-3, -1, -2, 2, 1, 1]$	2341	α	13	62	$[2, 6, 0, -6, 0, 1]$	7877		31,32,33
21	$[2, 4, 2, 2, 2, 1]$	2557	α	0	63	$[11, -1, -4, -4, 1, 1]$	7963		ν
22	$[2, 4, 0, -2, 0, 1]$	2677	α	14	64	$[-2, 4, 0, -2, 2, 1]$	8243	α	34
23	$[-2, 0, 2, 0, 0, 1]$	2693		15	65	$[2, 4, 2, 2, 0, 1]$	8581		ν
24	$[2, 4, 2, 0, 0, 1]$	2909	α	U	66	$[-1, -5, -4, 6, 1, 1]$	8803		35
25	$[6, 8, 8, 6, 2, 1]$	3037	α	0	67	$[-3, 13, -4, -6, 1, 1]$	9091	α	36
26	$[2, -2, 4, 0, 0, 1]$	3109	α	U	68	$[5, 7, 0, 0, 1, 1]$	9781	α	U
27	$[-2, 4, 2, -6, 0, 1]$	3251	α	16	69	$[7, 3, -6, -4, 1, 1]$	9803		37
28	$[1, 5, 2, 4, 1, 1]$	3461	α	U	70	$[2, -2, 4, 0, 2, 1]$	9941	α	38
29	$[-1, -3, -2, -2, 1, 1]$	3499		17	71	$[7, 1, 2, -2, 1, 1]$	9949		0
30	$[2, 0, 2, 0, 0, 1]$	3557		18	72	$[2, -8, 8, 0, 0, 1]$	10037		39
31	$[2, 2, 0, 0, 0, 1]$	3637	α	19	73	$[1, -3, -4, -2, 1, 1]$	10163	α	U
32	$[2, 6, 0, -4, 0, 1]$	3701	α	20	74	$[2, 4, 0, 6, 0, 1]$	10253		0
33	$[2, 0, 0, 2, 2, 1]$	3853	α	0	75	$[-2, 2, 2, -8, 0, 1]$	10259		ν
34	$[2, 0, 0, 2, 0, 1]$	3989		21	76	$[1, 3, 6, 2, 1, 1]$	10453	α	U
35	$[-2, -2, -2, 2, 2, 1]$	3989	α	U	77	$[3, -7, 10, -6, 1, 1]$	10789		40
36	$[-1, 5, -4, -4, 1, 1]$	4003		0	78	$[2, -2, 4, -4, 0, 1]$	10837		41
37	$[2, 2, -2, -2, 2, 1]$	4157	α	22	79	$[2, 2, 6, 4, 2, 1]$	10853		42
38	$[2, -6, 4, 0, 0, 1]$	4219	α	U	80	$[6, -4, 0, -2, 0, 1]$	10949	α	43
39	$[2, 2, 0, 2, 0, 1]$	4517	α	23	81	$[1, 1, 6, -6, 1, 1]$	10957		ν
40	$[2, 0, -6, -2, 2, 1]$	5059	α	24	82	$[-3, -1, 0, 0, 1, 1]$	11117		44
41	$[-1, 1, 0, -4, 1, 1]$	5227		25	83	$[-1, -5, -6, -4, 1, 1]$	11131		ν
42	$[2, 2, 2, 0, 0, 1]$	5261	α	0	84	$[5, 11, 0, -4, 1, 1]$	11243	α	U

Table 1. Favorable quintic fields (legend on previous page; continuation on next page).

#F ₀	f(x)	N	ε	#C	#F ₀	f(x)	N	ε	#C
85	[-1, 5, -6, 6, 1, 1]	11261		0	129	[9, 5, -6, -4, 1, 1]	17029		ν
86	[-1, 3, 2, -4, 1, 1]	11579		45	130	[-7, 5, 4, -2, 1, 1]	17203		ν
87	[-3, 1, 0, 2, 1, 1]	11701		ν	131	[-2, 10, -12, -2, 2, 1]	17291		57
88	[2, -10, 14, -4, 0, 1]	11971		46,47	132	[-15, 13, 6, -4, 1, 1]	17317		58
89	[13, 11, -6, -6, 1, 1]	12037		ν	133	[4, -4, 8, -2, 0, 1]	17341	α	U
90	[3, -1, -2, 0, 1, 1]	12109		ν	134	[-2, 0, 4, 2, 0, 1]	17341		0
91	[3, 11, 0, -4, 1, 1]	12301		ν	135	[-4, 4, 4, 0, 0, 1]	17389	α	59
92	[2, 10, 6, -2, 0, 1]	12541	α	U	136	[3, 7, 6, 4, 1, 1]	17597	α	0
93	[10, 6, -8, -4, 2, 1]	12757		ν	137	[14, 24, 4, -6, 0, 1]	17923		ν
94	[2, 2, 4, 2, 0, 1]	12781	α	U	138	[6, -4, 6, 0, 0, 1]	18077	α	60
95	[-3, 5, -2, -4, 1, 1]	12781	α	U	139	[-1, -3, -8, -4, 1, 1]	18181	α	0
96	[-3, -5, -10, -6, 1, 1]	12907	α	U	140	[-1, -5, -4, 2, 1, 1]	18691		0
97	[-3, 1, -6, -6, 1, 1]	12923	α	48	141	[1, 7, 2, -2, 1, 1]	18757		ν
98	[-1, -1, 2, -4, 1, 1]	13003	α	U	142	[10, 4, -8, -4, 2, 1]	18869		ν
99	[-2, 2, -2, 0, 2, 1]	13037	α	0	143	[-1, 3, -8, -8, 1, 1]	19051	α	U
100	[-2, 4, -2, -4, 2, 1]	13147	α	49	144	[-2, -2, 4, 4, 2, 1]	19211		61
101	[7, -1, -2, -4, 1, 1]	13147	α	50	145	[2, 0, 4, 4, 2, 1]	19429		63
102	[2, -4, 0, 0, 0, 1]	13259		51	146	[-2, -12, -22, -8, 2, 1]	19469	α	U
103	[3, -1, 4, -4, 1, 1]	13597		0	147	[-1, -5, -14, -8, 1, 1]	19531	α	64
104	[2, 8, 8, 6, 2, 1]	13597		ν	148	[4, 0, -8, 2, 2, 1]	19597		0
105	[1, 5, 2, -12, 1, 1]	13723		52	149	[4, 4, 0, 4, 2, 1]	20389		ν
106	[6, 4, 6, 4, 2, 1]	13829	α	U	150	[1, -3, 2, 4, 1, 1]	20533	α	U
107	[1, 1, -4, -6, 1, 1]	13963		ν	151	[-2, 6, 0, 2, 2, 1]	21061	α	U
108	[-2, 6, 2, -6, 0, 1]	13997		53	152	[-2, 2, 2, -4, 2, 1]	21211	α	65
109	[4, -4, 4, 0, 2, 1]	13997		ν	153	[-5, 11, 2, -12, 1, 1]	21283		0
110	[-9, -1, 4, 0, 1, 1]	14149		ν	154	[-6, -4, 4, -4, 0, 1]	21563		66
111	[15, 13, -6, -6, 1, 1]	14197		54	155	[-14, -18, -10, -2, 2, 1]	21739	α	U
112	[2, -2, 6, -2, 2, 1]	14293		ν	156	[18, 8, -12, -6, 2, 1]	21787		67
113	[-3, -1, -2, -2, 1, 1]	14629	α	U	157	[-3, -1, 2, 2, 1, 1]	22277		68
114	[-46, 48, 6, -14, 0, 1]	14779		ν	158	[-2, 8, -8, -6, 2, 1]	22291		69
115	[2, 4, 4, 4, 0, 1]	14821	α	U	159	[-1, -3, -8, 4, 1, 1]	22637		0
116	[-2, 4, 2, -2, 0, 1]	15013		ν	160	[-3, 13, 2, 10, 1, 1]	22709		ν
117	[1, -3, 2, -4, 1, 1]	15227		ν	161	[2, 0, -6, -4, 2, 1]	22787	α	U
118	[-2, 0, 2, 0, 2, 1]	15307		55	162	[1, 9, 6, 2, 1, 1]	22861		70
119	[-2, 2, 4, 4, 0, 1]	15373	α	U	163	[-5, 13, -4, -8, 1, 1]	23003		71
120	[3, 7, 0, 0, 1, 1]	15493	α	U	164	[-3, -1, -4, -4, 1, 1]	23059	α	U
121	[-2, 4, -2, 0, 2, 1]	15581		ν	165	[1, -3, -2, 4, 1, 1]	23131		72,73
122	[5, 9, 4, 6, 1, 1]	15749		56	166	[2, -4, -2, 0, 2, 1]	23251		ν
123	[4, 0, 0, -2, 2, 1]	15749	α	U	167	[6, 4, 2, 4, 0, 1]	23669		ν
124	[2, -6, 2, 2, 2, 1]	15923	α	U	168	[-6, 2, 4, -2, 0, 1]	24109	α	0
125	[-2, 0, 10, 8, 0, 1]	16139		ν	169	[2, 8, 0, 6, 0, 1]	24469		74,75
126	[2, -2, -10, -4, 2, 1]	16451		ν	170	[2, -4, 2, 2, 0, 1]	24533		ν
127	[1, 5, 2, 0, 1, 1]	16901	α	U	171	[-6, 4, 6, -6, 0, 1]	24611	α	U
128	[-6, 4, 2, -4, 0, 1]	16981	α	U	172	[-7, -5, -2, -2, 1, 1]	24763		ν

#C	#F ₀	$g(x)$	N	ϵ	#C	#F ₀	$g(x)$	N	ϵ
1	1	[1, -4, 8, -8, 0, 4]	277	α	39	72	[1, 0, 4, 0, 0, 4]	10037	
2	2	[1, -4, 4, 4, -8, 4]	349	α	40	77	[1, 12, 44, 52, 4, 4]	10789	
3	3	[1, 8, 20, 12, -8, 4]	461	α	41	78	[13, 4, -20, -8, 8, 4]	10837	
4	6	[1, 0, 0, 4, -4, 4]	797	α	42	79	[5, 12, 0, -12, 0, 4]	10853	
5	7	[1, 4, 0, -8, 0, 4]	971	α	43	80	[-7, 12, 4, 16, 4, 4]	10949	α
6	8	[1, 0, -4, 8, -8, 4]	997	α	44	82	[1, -4, 4, -4, 8, 4]	11117	
7	9	[1, -4, 4, 0, -4, 4]	1051	α	45	86	[1, 12, 44, 44, -4, 4]	11579	
8	11	[-79, -304, -560, -200, -4, 4]	1109	α	46	88	[1, 4, 0, -4, 4, 4]	11971	
9	12	[1, 4, 4, -4, -4, 4]	1109	α	47	88	[1461041, -565424, 78052, -4092, 8, 4]	11971	
10	15	[1, 0, -4, 4, -4, 4]	1637		48	97	[1, 4, 0, -8, -4, 4]	12923	α
11	16	[5, -24, 44, -36, 8, 4]	1811	α	49	100	[1, 12, 32, 28, 8, 4]	13147	α
12	18	[1, 4, 4, 4, 8, 4]	2243	α	50	101	[1, -4, 4, -4, 4, 4]	13147	α
13	20	[-3, -4, 0, 8, 8, 4]	2341	α	51	102	[5, -28, 48, -24, -4, 4]	13259	
14	22	[5, -16, 20, -8, -4, 4]	2677	α	52	105	[1, -4, 0, 4, 8, 4]	13723	
15	23	[1, 0, 0, 4, 8, 4]	2693		53	108	[137, -356, 328, -116, 4, 4]	13997	
16	27	[1, 4, -8, -4, 4, 4]	3251	α	54	111	[9, 16, -4, -16, 0, 4]	14197	
17	29	[9, -40, 60, -32, 0, 4]	3499		55	118	[1, 4, -8, -4, 8, 4]	15307	
18	30	[1, 0, 0, 4, -8, 4]	3557		56	122	[1, 4, 4, 8, 8, 4]	15749	
19	31	[1, 0, 4, 0, 4, 4]	3637	α	57	131	[1, -4, 4, 0, -8, 4]	17291	
20	32	[161, -360, 284, -80, -4, 4]	3701	α	58	132	[-3, 8, -8, 8, -8, 4]	17317	
21	34	[1, -4, 4, 0, 0, 4]	3989		59	135	[1, 0, 0, -4, 4, 4]	17389	α
22	37	[-3, 8, -12, 12, -8, 4]	4157	α	60	138	[-3, -20, -40, -20, 4, 4]	18077	α
23	39	[1, -4, 8, -8, 4, 4]	4517	α	61	144	[-247, 552, -200, -136, 4, 4]	19211	
24	40	[-3, 8, 0, -12, 4, 4]	5059	α	62	144	[-7, 16, 4, -16, 0, 4]	19211	
25	41	[5, -20, -40, 240, -600, 500]	5227		63	145	[-3, 36, -144, 192, -108, 108]	19429	
26	46	[5185, -6384, 2664, -396, -4, 4]	5651	α	64	147	[-11, -44, 264, 440, 968, 5324]	19531	α
27	47	[73, -180, 152, -40, -8, 4]	5867		65	152	[-3, -4, 8, 4, -8, 4]	21211	α
28	52	[1, 4, 0, -8, 4, 4]	6491		66	154	[-21167, -18908, -5996, -712, 0, 4]	21563	
29	54	[-3, 4, 4, -8, 0, 4]	6763		67	156	[-3, -16, -28, -16, 4, 4]	21787	
30	57	[25, 28, -12, -16, 4, 4]	7109		68	157	[9, -32, 40, -20, 0, 4]	22277	
31	62	[41, -148, 160, -56, -4, 4]	7877		69	158	[1, -4, 8, -12, 4, 4]	22291	
32	62	[1, 8, 12, -8, -8, 4]	7877		70	162	[1, 4, 8, 4, 4, 4]	22861	
33	62	[73, -228, 232, -84, 0, 4]	7877		71	163	[5, -36, 76, -40, 4, 4]	23003	
34	64	[-591, -1160, -792, -204, -4, 4]	8243	α	72	165	[1909, -2652, 1308, -236, -4, 4]	23131	
35	66	[1, -8, 20, -12, -8, 4]	8803		73	165	[1, 8, -12, -8, 8, 4]	23131	
36	67	[1, -8, 24, -28, 4, 4]	9091	α	74	169	[1, 8, 20, 16, 0, 4]	24469	
37	69	[1, -8, 16, -8, -4, 4]	9803		75	169	[7309, -8208, 3292, -504, 4, 4]	24469	
38	70	[1, 8, 20, 16, 8, 4]	9941	α					

Table 2. Curves $y^2 = g(x)$, their 2-division fields and conductors (legend on page 1065).

Legend for Tables 3 and 4. We know 276109 curves, including 10360 mild curves with $3 \leq m \leq 53$, whose Jacobians are favorable and nonisogenous of prime conductor $N < 10^{10}$, for a total of 275494 nonisomorphic fields. Table 3 summarizes the statistics. For $0 \leq j \leq 9$, the j -th column refers to N between $j \cdot 10^9$ and $(j + 1) \cdot 10^9$. The rows A, F and α , respectively, give the number of abelian varieties, fields and amiable fields. It is remarkable that approximately 11.8% of the favorable fields are amiable, uniformly for each slice of size 10^9 . For the reader's entertainment, Table 4 lists the curves we found with largest conductors below 10^{10} and amiable Jacobians.

j	0	1	2	3	4	5	6	7	8	9	Total
A	63563	35507	29047	25450	23684	22099	20500	19505	18773	17981	276109
F	63212	35429	28998	25417	23657	22079	20479	19493	18761	17969	275494
α	7632	4290	3362	2948	2799	2606	2375	2340	2189	2127	32668

Table 3. Amiable fields among favorable fields (legend immediately above).

$P(x)$	N	$P(x)$	N
$[-90, -184, -136, -39, -1, 1]$	9882329341	$[10, 22, 7, -7, 0, 1]$	9891907261
$[11, 26, -7, -8, 0, 1]$	9893121157	$[11, 17, 3, -4, -2, 1]$	9897613669
$[-8428, -6910, -2025, -226, -1, 1]$	9898501189	$[-21, 6, 10, -1, 1, 1]$	9911121709
$[87, -106, 56, -9, -2, 1]$	9934582709	$[-61, 50, 9, -13, 0, 1]$	9982174061
$[-33, 20, -1, 10, 1, 1]$	9987633941	$[-2, -3, -15, -9, 0, 1]$	9994370909

Table 4. Curves $y^2 = 1 + 4P(x)$ of large conductor with amiable fields (legend at top of page).

Note added in proof

The paramodular conjecture should be modified to accommodate comments and examples of Frank Calegari.

Definition. An abelian fourfold B is a *fake abelian surface* if $\text{End}(B)$ is an order in a quaternion algebra over \mathbb{Q} .

Paramodular conjecture. Let \mathcal{A}_N be the set of isogeny classes of abelian surfaces A/\mathbb{Q} of conductor N with $\text{End } A = \mathbb{Z}$, let \mathcal{B}_N be the set of isogeny classes of fake abelian surfaces B/\mathbb{Q} of conductor N^2 and let \mathcal{P}_N be the set of cuspidal, nonlift Siegel paramodular newforms f of genus 2, weight 2 and level N with rational Hecke eigenvalues, up to nonzero scaling. Then there is a bijection between \mathcal{P}_N and $\mathcal{A}_N \cup \mathcal{B}_N$ such that

$$L(C, s) = \begin{cases} L(f, s, \text{spin}) & \text{if } C \in \mathcal{A}_N, \\ L(f, s, \text{spin})^2 & \text{if } C \in \mathcal{B}_N. \end{cases}$$

Acknowledgements

The authors wish to express their gratitude to the anonymous referees for their extremely careful reading of the manuscript. Their valuable suggestions helped us clarify and improve the exposition. Magma [Bosma et al. 1997], obtained with the aid of the Simons Foundation, was used in some of our computations.

References

- [Abrashkin 1987] V. A. Abrashkin, “Galois modules of group schemes of period p over the ring of Witt vectors”, *Izv. Akad. Nauk SSSR Ser. Mat.* **51**:4 (1987), 691–736, 910. In Russian; translated in *Izvestiya Math.* **31**:4 (1988), 1–46. MR Zbl
- [Berger et al. 2015] T. Berger, L. Dembélé, A. Pacetti, and M. H. Şengün, “Theta lifts of Bianchi modular forms and applications to paramodularity”, *J. Lond. Math. Soc.* (2) **92**:2 (2015), 353–370. MR Zbl
- [Berthelot 1977] P. Berthelot, “Systèmes de Honda des schémas en F_q -vectoriels”, *Bull. Soc. Math. France* **105**:3 (1977), 225–239. MR Zbl
- [Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, “The Magma algebra system, I: The user language”, *J. Symbolic Comput.* **24**:3-4 (1997), 235–265. MR Zbl
- [Brinon and Conrad 2009] O. Brinon and B. Conrad, “CMI summer school notes on p -adic Hodge Theory”, lecture notes, Stanford University, 2009, Available at <http://math.stanford.edu/~conrad/papers/notes.pdf>.
- [Brumer and Kramer 2001] A. Brumer and K. Kramer, “Non-existence of certain semistable abelian varieties”, *Manuscripta Math.* **106**:3 (2001), 291–304. MR Zbl
- [Brumer and Kramer 2004] A. Brumer and K. Kramer, “Semistable abelian varieties with small division fields”, pp. 13–37 in *Galois theory and modular forms*, edited by K.-i. Hashimoto et al., Dev. Math. **11**, Kluwer Acad. Publ., Boston, MA, 2004. MR Zbl
- [Brumer and Kramer 2012] A. Brumer and K. Kramer, “Arithmetic of division fields”, *Proc. Amer. Math. Soc.* **140**:9 (2012), 2981–2995. MR Zbl
- [Brumer and Kramer 2014] A. Brumer and K. Kramer, “Paramodular abelian varieties of odd conductor”, *Trans. Amer. Math. Soc.* **366**:5 (2014), 2463–2516. MR Zbl
- [Brumer and Kramer 2018] A. Brumer and K. Kramer, “Large 2-adic Galois image and non-existence of certain abelian surfaces over \mathbb{Q} ”, *Acta Arith.* **183**:4 (2018), 357–383. MR
- [Brumer and McGuinness 1990] A. Brumer and O. McGuinness, “310716 elliptic curves of prime conductor”, electronic reference, 1990, Available at <http://www.math.columbia.edu/~om/>.
- [Calegari 2004] F. Calegari, “Semistable abelian varieties over \mathbb{Q} ”, *Manuscripta Math.* **113**:4 (2004), 507–529. MR Zbl
- [Conrad 1999] B. Conrad, “Finite group schemes over bases with low ramification”, *Compositio Math.* **119**:3 (1999), 239–320. MR Zbl
- [Eilenberg and MacLane 1942] S. Eilenberg and S. MacLane, “Group extensions and homology”, *Ann. of Math.* (2) **43** (1942), 757–831. MR Zbl
- [Faltings 1983] G. Faltings, “Endlichkeitssätze für abelsche Varietäten über Zahlkörpern”, *Invent. Math.* **73**:3 (1983), 349–366. MR Zbl
- [Fontaine 1975a] J.-M. Fontaine, “Groupes finis commutatifs sur les vecteurs de Witt”, *C. R. Acad. Sci. Paris Sér. A-B* **280** (1975), A1423–A1425. MR Zbl
- [Fontaine 1975b] J.-M. Fontaine, “Groupes p -divisibles sur les vecteurs de Witt”, *C. R. Acad. Sci. Paris Sér. A-B* **280** (1975), A1353–A1356. MR Zbl
- [Fontaine 1977] J.-M. Fontaine, *Groupes p -divisibles sur les corps locaux*, Société Mathématique de France, Paris, 1977. Astérisque, No. 47-48. MR Zbl
- [Fontaine 1985] J.-M. Fontaine, “Il n’y a pas de variété abélienne sur \mathbf{Z} ”, *Invent. Math.* **81**:3 (1985), 515–538. MR Zbl
- [Grothendieck and Raynaud 1972] A. Grothendieck and M. Raynaud, *Modeles de neron et monodromie*, pp. 313–523, Springer, Berlin, Heidelberg, 1972. Zbl

- [Johnson-Leung and Roberts 2012] J. Johnson-Leung and B. Roberts, “Siegel modular forms of degree two attached to Hilbert modular forms”, *J. Number Theory* **132**:4 (2012), 543–564. MR Zbl
- [Johnson-Leung and Roberts 2017] J. Johnson-Leung and B. Roberts, “Twisting of Siegel paramodular forms”, *Int. J. Number Theory* **13**:7 (2017), 1755–1854. MR Zbl
- [Kemper and Malle 1997] G. Kemper and G. Malle, “The finite irreducible linear groups with polynomial ring of invariants”, *Transform. Groups* **2**:1 (1997), 57–89. MR Zbl
- [Lang 1994] S. Lang, *Algebraic number theory*, 2nd ed., Graduate Texts in Mathematics **110**, Springer, 1994. MR Zbl
- [Mac Lane 1963] S. Mac Lane, *Homology*, Grundlehren der math. Wissenschaften **114**, Springer, Berlin, 1963. MR Zbl
- [Mazur 1977] B. Mazur, “Modular curves and the Eisenstein ideal”, *Inst. Hautes Études Sci. Publ. Math.* **47** (1977), 33–186. MR Zbl
- [Merriman and Smart 1993] J. R. Merriman and N. P. Smart, “Curves of genus 2 with good reduction away from 2 with a rational Weierstrass point”, *Math. Proc. Cambridge Philos. Soc.* **114**:2 (1993), 203–214. MR Zbl
- [Milne 1972] J. S. Milne, “On the arithmetic of abelian varieties”, *Invent. Math.* **17** (1972), 177–190. MR Zbl
- [Ploner 2015] P. Ploner, “Computation of framed deformation functors”, *J. Number Theory* **156** (2015), 21–37. MR Zbl
- [Poor and Yuen 2015] C. Poor and D. S. Yuen, “Paramodular cusp forms”, *Math. Comp.* **84**:293 (2015), 1401–1438. MR Zbl
- [Raynaud 1974] M. Raynaud, “Schémas en groupes de type (p, \dots, p) ”, *Bull. Soc. Math. France* **102** (1974), 241–280. MR Zbl
- [Schoof 2003] R. Schoof, “Abelian varieties over cyclotomic fields with good reduction everywhere”, *Math. Ann.* **325**:3 (2003), 413–448. MR Zbl
- [Schoof 2005] R. Schoof, “Abelian varieties over \mathbb{Q} with bad reduction in one prime only”, *Compos. Math.* **141**:4 (2005), 847–868. MR Zbl
- [Schoof 2012a] R. Schoof, “On the modular curve $X_0(23)$ ”, pp. 317–345 in *Geometry and arithmetic*, edited by C. Faber et al., Eur. Math. Soc., Zürich, Switzerland, 2012. MR Zbl
- [Schoof 2012b] R. Schoof, “Semistable abelian varieties with good reduction outside 15”, *Manuscripta Math.* **139**:1-2 (2012), 49–70. MR Zbl
- [Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, 1979. MR Zbl
- [Setzer 1981] B. Setzer, “Elliptic curves with good reduction everywhere over quadratic fields and having rational j -invariant”, *Illinois J. Math.* **25**:2 (1981), 233–245. MR Zbl
- [Shimura 1972] G. Shimura, “Class fields over real quadratic fields and Hecke operators”, *Ann. of Math. (2)* **95** (1972), 130–190. MR Zbl
- [Smart 1997] N. P. Smart, “ S -unit equations, binary forms and curves of genus 2”, *Proc. London Math. Soc. (3)* **75**:2 (1997), 271–307. MR Zbl
- [Tate 1997] J. Tate, “Finite flat group schemes”, pp. 121–154 in *Modular forms and Fermat’s last theorem* (Boston, MA, 1995), edited by G. Cornell et al., Springer, New York, 1997. MR Zbl

Communicated by Brian Conrad

Received 2016-09-01 Revised 2017-08-20 Accepted 2017-10-23

brumer@fordham.edu

Department of Mathematics, Fordham University, Bronx, NY, United States

kkramer@qc.cuny.edu

Department of Mathematics, Queens College (CUNY), Flushing, NY, United States

Characterization of Kollár surfaces

Giancarlo Urzúa and José Ignacio Yáñez

Kollár (2008) introduced the surfaces

$$(x_1^{a_1} x_2 + x_2^{a_2} x_3 + x_3^{a_3} x_4 + x_4^{a_4} x_1 = 0) \subset \mathbb{P}(w_1, w_2, w_3, w_4)$$

where $w_i = W_i/w^*$, $W_i = a_{i+1}a_{i+2}a_{i+3} - a_{i+2}a_{i+3} + a_{i+3} - 1$, and $w^* = \gcd(W_1, \dots, W_4)$. The aim was to give many interesting examples of \mathbb{Q} -homology projective planes. They occur when $w^* = 1$. For that case, we prove that Kollár surfaces are Hwang–Keum (2012) surfaces. For $w^* > 1$, we construct a geometrically explicit birational map between Kollár surfaces and cyclic covers $z^{w^*} = l_1^{a_2 a_3 a_4} l_2^{-a_3 a_4} l_3^{a_4} l_4^{-1}$, where $\{l_1, l_2, l_3, l_4\}$ are four general lines in \mathbb{P}^2 . In addition, by using various properties on classical Dedekind sums, we prove that:

- For any $w^* > 1$, we have $p_g = 0$ if and only if the Kollár surface is rational. This happens when $a_{i+1} \equiv 1$ or $a_i a_{i+1} \equiv -1 \pmod{w^*}$ for some i .
- For any $w^* > 1$, we have $p_g = 1$ if and only if the Kollár surface is birational to a K3 surface. We classify this situation.
- For $w^* \gg 0$, we have that the smooth minimal model S of a generic Kollár surface is of general type with $K_S^2/e(S) \rightarrow 1$.

1. Introduction	1073
2. Kollár hypersurfaces	1075
3. Explicit birational map for Kollár surfaces	1077
4. Kollár surfaces are Hwang–Keum surfaces	1087
5. Kollár surfaces as branched covers of \mathbb{P}^2	1093
6. Theorems on geometric genus	1096
Acknowledgements	1104
References	1104

1. Introduction

The ground field is \mathbb{C} . Let $n \geq 3$ be an integer, and let a_1, \dots, a_n be positive integers such that there is no $(a_i, a_{i+2}, \dots, a_{i+n-2}) = (1, \dots, 1)$ when n is even. The indices are and will be taken modulo n . For every $1 \leq i \leq n$, we define the positive integers

$$W_i := \sum_{j=1}^n (-1)^{j-1} \prod_{l=i+j}^{i+n-1} a_l \quad \text{and} \quad D := \prod_{l=1}^n a_l + (-1)^{n-1}.$$

MSC2010: 14J10.

Keywords: \mathbb{Q} -homology projective planes, Dedekind sums, branched covers.

For example, for $n = 4$ we have

$$W_i = a_{i+1}a_{i+2}a_{i+3} - a_{i+2}a_{i+3} + a_{i+3} - 1 \quad \text{and} \quad D = a_1a_2a_3a_4 - 1.$$

We also define

$$w^* := \gcd(W_1, \dots, W_n).$$

Then $w^* = \gcd(W_i, W_{i+1}) = \gcd(W_i, D)$ since $a_i W_i + W_{i+1} = D$ for all i .

Set

$$w_i := \frac{W_i}{w^*} \quad \text{and} \quad d := \frac{D}{w^*}.$$

Notice that $\gcd(a_i, w^*) = 1$ for all i .

The *Kollár hypersurface* [2008] of type (a_1, \dots, a_n) is

$$X(a_1, \dots, a_n) := (x_1^{a_1} x_2 + x_2^{a_2} x_3 + \dots + x_n^{a_n} x_1 = 0) \subset \mathbb{P}(w_1, \dots, w_n).$$

Let $0 < \mu_i < w^*$ be such that $\mu_i \equiv (-1)^{i+1} \prod_{l=i+1}^{i+n-1} a_l \pmod{w^*}$. We consider the normal projective variety Y' given by the w^* -th root cover $Y' \rightarrow \mathbb{P}^{n-2} = \{y_1 + \dots + y_n = 0\} \subset \mathbb{P}^{n-1}$ branched along $\{y_1^{\mu_1} \dots y_n^{\mu_n} = 0\}$; see Section 2 for precise definitions. The map ψ associated to the linear system $|x_1^{a_1} x_2, \dots, x_n^{a_n} x_1|$ in the Kollár hypersurface shows that the varieties $X(a_1, \dots, a_n)$ and Y' are birational; this is worked out in Section 2.

In this paper we consider in detail the case $n = 4$; the surface $X = X(a_1, \dots, a_4)$ will be called *Kollár surface*. First, we note that Kollár surfaces are birational to infinitely many Kollár surfaces with $\gcd(w_i, w_{i+2}) = 1$ and $a_i > 1$ (see Theorem 5.1), and so we assume these numerical conditions to simplify the exposition. Section 3 is devoted to proving:

Theorem 1.1. *There is a configuration Γ of six rational curves in X such that, if $\widehat{X} \rightarrow X$ is a log resolution of (X, Γ) , then $\widehat{X} \rightarrow X \xrightarrow{\psi} \mathbb{P}^2$ is a morphism which factors through $Y' \rightarrow \mathbb{P}^2$ via a birational morphism $\widehat{X} \rightarrow Y'$.*

The aim of Kollár surfaces [2008] was to give examples of rational \mathbb{Q} -homology projective planes (\mathbb{Q} HPP) with ample canonical class. This occurs for $w^* = 1$ after contracting $(x_1 = x_3 = 0)$ and $(x_2 = x_4 = 0)$ in X , when these two curves have negative self-intersections (see Corollary 4.8). This contraction gives a \mathbb{Q} HPP with two cyclic quotient singularities, and when $a_i \geq 4$ for all i , the canonical class is ample. On the other hand, Hwang and Keum [2012] constructed a series of examples of \mathbb{Q} HPP with ample canonical class and same singularities as Kollár examples. In Section 4 we prove:

Theorem 1.2. *Kollár \mathbb{Q} -homology projective planes are Hwang–Keum surfaces.*

As an intriguing problem, we point out that rational \mathbb{Q} HPP with ample canonical class and cyclic quotient singularities have not yet been classified. The number of possible singularities is at most four, and examples with one, two, and three singularities have been constructed. It is conjectured that the case of four singularities is impossible [Kollár 2008; Hwang and Keum 2012].

In Section 5 we write down formulas for the invariants of Kollár surfaces via Y' when $w^* > 1$. Particularly interesting is the geometric genus, which depends on classical Dedekind sums on the exponents a_i . For example, by comparing the two models X and Y' , we write down an identity for Dedekind sums in Corollary 5.8. More importantly, in Section 6 we use new bounds on their values, essentially due to Girstmair [2017], to prove (see Theorems 6.3, 6.6, and 6.11):

Theorem 1.3. *For $w^* > 1$, we have:*

- (a) $p_g = 0$ if and only if the Kollár surface is rational. This happens when $a_i \equiv 1$ or $a_i a_{i+1} \equiv -1$ modulo w^* for some i .
- (b) $p_g = 1$ if and only if the Kollár surface is birational to a K3 surface. We classify this situation in eight cases (see Table 1).
- (c) For $w^* \gg 0$, the smooth minimal model S of a generic Kollár surface is of general type with $K_S^2/e(S) \rightarrow 1$, where K_S is the canonical class, and $e(S)$ is the topological Euler characteristic.

Moreover, we note that any p_g is realizable by some Kollár surface (Proposition 6.2), and that given $m > 0$ there exists an N such that $p_g > m$ if $w^* > N$ (Lemma 6.7). At the end, we give explicit examples of Kodaira dimension-1 elliptic fibrations (Example 6.9) and surfaces of general type (Example 6.10), arising as Kollár surfaces for w^* arbitrarily large.

2. Kollár hypersurfaces

Kollár [2008, Theorem 39] proves:

- Theorem 2.1.** (1) *The weighted projective space $\mathbb{P}(w_1, \dots, w_n)$ is well formed, and its singular set has dimension $\leq [n/2] - 1$.*
- (2) *The hypersurface $X(a_1, \dots, a_n)$ is quasismooth, and $\mathbb{P}(w_1, \dots, w_n) \setminus X(a_1, \dots, a_n)$ is smooth.*
- (3) *If $w^* = 1$, then $X(a_1, \dots, a_n)$ is birational to \mathbb{P}^{n-2} .*

To prove (3) above, Kollár uses the linear system $|x_1^{a_1} x_2, x_2^{a_2} x_3, \dots, x_n^{a_n} x_1|$. In general, this linear system defines a rational map

$$\psi : \mathbb{P}(w_1, \dots, w_n) \dashrightarrow \mathbb{P}_{y_1, \dots, y_n}^{n-1}$$

given by $y_i = x_i^{a_i} x_{i+1}$.

Proposition 2.2. *The rational map ψ defines the field extension*

$$\mathbb{C}(y_1/y_n, \dots, y_{n-1}/y_n) \subset \mathbb{C}(y_1/y_n, \dots, y_{n-1}/y_n)[z]/(z^{w^*} - f/y_n^{W_1})$$

where $z = x_1^d/y_n^{w_1}$ and $f = y_1^{a_2 a_3 \dots a_n} y_2^{-a_3 \dots a_n} y_3^{a_4 \dots a_n} \dots y_{n-1}^{(-1)^{n-2} a_n} y_n^{(-1)^{n-1}}$.

Proof. At the affine cover level, the field extension induced by ψ is

$$\mathbb{C}(y_1, \dots, y_n) \subset \mathbb{C}(y_1, \dots, y_n)[x_1]/(x_1^D - f)$$

where the other variables x_2, \dots, x_n can be written using y_1, \dots, y_n, x_1 . The action of \mathbb{C}^* compatible with the map is: given $\lambda \in \mathbb{C}^*$, $y_i \mapsto \lambda^d y_i$ and $x_i \mapsto \lambda^{w_i} x_i$. Then the rational map ψ is determined by

$$(\mathbb{C}(y_1, \dots, y_n))^{\mathbb{C}^*} \subset (\mathbb{C}(y_1, \dots, y_n)[x_1]/(x_1^D - f))^{\mathbb{C}^*}.$$

Notice that $(\mathbb{C}(y_1, \dots, y_n))^{\mathbb{C}^*} = \mathbb{C}(y_1/y_n, \dots, y_{n-1}/y_n)$, and that $z = x_1^d/y_n^{w_1}$ is a \mathbb{C}^* -invariant element such that $z^{w^*} - f/y_n^{W_1} = 0$. Since geometrically the map ψ has degree w^* , then

$$(\mathbb{C}(y_1, \dots, y_n)[x_1]/(x_1^D - f))^{\mathbb{C}^*} = \mathbb{C}(y_1/y_n, \dots, y_{n-1}/y_n)[z]/(z^{w^*} - f/y_n^{W_1}). \quad \square$$

Corollary 2.3. *The corresponding restriction map*

$$\psi|_X : X(a_1, \dots, a_n) \dashrightarrow \mathbb{P}^{n-2} = \{y_1 + \dots + y_n = 0\}$$

is cyclic of degree w^* totally branched along $(y_1 \cdots y_n = 0) \subset \mathbb{P}^{n-2}$.

In this way, we can write down another normal projective model Y' of $X(a_1, \dots, a_n)$ using a w^* -th root cover as described in [Esnault and Viehweg 1992].

As in the introduction, let $0 < \mu_i < w^*$ be such that

$$\mu_i \equiv (-1)^{i+1} \prod_{l=i+1}^{i+n-1} a_l \pmod{w^*}.$$

In $\mathbb{P}^{n-2} = \{y_1 + \dots + y_n = 0\}$, we write $L_i := \{y_i = 0\}$, and so

$$\mathcal{O}_{\mathbb{P}^{n-2}}(w_1)^{\otimes w^*} \simeq \mathcal{O}_{\mathbb{P}^{n-2}}(\mu_1 L_1 + \dots + \mu_n L_n),$$

where $w_1 w^* = W_1 = \sum_{i=1}^n \mu_i$. Then

$$Y_0 := \text{Spec}_{\mathbb{P}^{n-2}} \left(\bigoplus_{i=0}^{w^*-1} \mathcal{O}_{\mathbb{P}^{n-2}}(-w_1 i) \right) \rightarrow \mathbb{P}^{n-2}$$

is the cyclic cover given by $z^{w^*} - f/y_n^{W_1}$ above. We want to consider the normalization of Y_0 . As in [Esnault and Viehweg 1992], we define the line bundles $\mathcal{L}^{(i)}$ on \mathbb{P}^{n-2} as

$$\mathcal{L}^{(i)} := \mathcal{O}_{\mathbb{P}^{n-2}}(w_1 i) \otimes \mathcal{O}_{\mathbb{P}^{n-2}} \left(- \sum_{j=1}^n \left[\frac{\mu_j i}{w^*} \right] L_j \right)$$

for $i \in \{0, 1, \dots, w^* - 1\}$, where $[x]$ is the integer part of x . Then the normalization of Y_0 is $Y' := \text{Spec}_{\mathbb{P}^{n-2}} \left(\bigoplus_{i=0}^{w^*-1} \mathcal{L}^{(i)-1} \right)$ [Esnault and Viehweg 1992, Corollary 3.11]. Notice that $\text{gcd}(\mu_i, w^*) = 1$, and so this cyclic morphism is totally branched at the L_i .

Corollary 2.4. *There is a birational map $X(a_1, \dots, a_n) \dashrightarrow Y'$.*

In the next section we describe explicitly this birational map for $n = 4$.

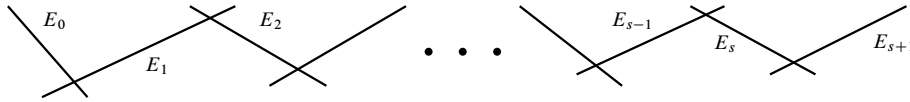


Figure 1. Exceptional divisors over $\frac{1}{m}(1, q)$, E_0 , and E_{s+1} .

3. Explicit birational map for Kollár surfaces

From now on we concentrate in the case of Kollár surfaces, where $n = 4$. We will be working with cyclic quotient surface singularities, which we now review. A cyclic quotient singularity S , denoted by $\frac{1}{m}(a, b)$, is a germ at the origin of the quotient of \mathbb{C}^2 by the action $(x, y) \mapsto (\zeta^a x, \zeta^b y)$, where ζ is a primitive m -th root of 1, and a, b are integers coprime to m [Barth et al. 2004, §III.5]. Let $0 < q < m$ be such that $aq - b \equiv 0$ modulo m . Then $\frac{1}{m}(a, b) = \frac{1}{m}(1, q)$. Let $\sigma : \tilde{S} \rightarrow S$ be the minimal resolution of S . Figure 1 shows the exceptional curves $E_i = \mathbb{P}^1$ of σ , for $1 \leq i \leq s$, and the strict transforms E_0 and E_{s+1} of $(y = 0)$ and $(x = 0)$, respectively.

The numbers $E_i^2 = -b_i$ are computed using the *Hirzebruch–Jung continued fraction*

$$\frac{m}{q} = b_1 - \frac{1}{b_2 - \frac{1}{\ddots - \frac{1}{b_s}}}$$

We denote $||[b_1, \dots, b_s]|| := m$. This continued fraction defines the sequence of integers

$$0 = \beta_{s+1} < 1 = \beta_s < \dots < q = \beta_1 < m = \beta_0$$

where $\beta_{i+1} = b_i \beta_i - \beta_{i-1}$. In this way, $\beta_{i-1}/\beta_i = [b_i, \dots, b_s]$. Partial fractions $\alpha_i/\gamma_i = [b_1, \dots, b_{i-1}]$ are computed through the sequences

$$0 = \alpha_0 < 1 = \alpha_1 < \dots < q^{-1} = \alpha_s < m = \alpha_{s+1},$$

where $\alpha_{i+1} = b_i \alpha_i - \alpha_{i-1}$ (q^{-1} is the integer such that $0 < q^{-1} < m$ and $qq^{-1} \equiv 1 \pmod{m}$), and $\gamma_0 = -1$, $\gamma_1 = 0$, and $\gamma_{i+1} = b_i \gamma_i - \gamma_{i-1}$. We have $\alpha_{i+1} \gamma_i - \alpha_i \gamma_{i+1} = -1$, $\beta_i = q \alpha_i - m \gamma_i$, and $m/q^{-1} = [b_s, \dots, b_1]$. These numbers appear in the pull-back formulas

$$\sigma^*((y = 0)) = \sum_{i=0}^{s+1} \frac{\beta_i}{m} E_i \quad \text{and} \quad \sigma^*((x = 0)) = \sum_{i=0}^{s+1} \frac{\alpha_i}{m} E_i, \tag{3-1}$$

and $K_{\tilde{S}} \equiv \sigma^*(K_S) + \sum_{i=1}^s (-1 + (\beta_i + \alpha_i)/m) E_i$.

Let $X(a_1, a_2, a_3, a_4)$ be a Kollár surface. Let

$$p_1 = (1 : 0 : 0 : 0), \quad p_2 = (0 : 1 : 0 : 0), \quad p_3 = (0 : 0 : 1 : 0), \quad p_4 = (0 : 0 : 0 : 1).$$

Proposition 3.1. *The surface $X(a_1, a_2, a_3, a_4)$ is normal and has only singularities of type $\frac{1}{w_i}(w_{i+2}, w_{i+3})$ at the points p_i when $\gcd(w_i, w_{i+2}) = 1$, and of type $\frac{1}{t_i}(t_{i+2}, w_{i+3})$ when $\gcd(w_i, w_{i+2}) = h > 1$, where $w_j = ht_j$.*

Proof. Here we follow the idea in [Iano-Fletcher 2000, §10.1]. Without loss of generality, it is enough to check the singularity at p_1 . Consider the affine cone $C_X \subset \mathbb{C}^4$ of $X(a_1, a_2, a_3, a_4)$ and the corresponding action of \mathbb{C}^* given by,

$$\lambda \in \mathbb{C}^*, \quad \lambda \cdot (x_1, x_2, x_3, x_4) = (\lambda^{w_1}x_1, \lambda^{w_2}x_2, \lambda^{w_3}x_3, \lambda^{w_4}x_4).$$

Then to study the singularities around p_1 , we check how the action behaves when we restrict to $(x_1 = 1)$. Notice that, when $x_1 \neq 0$,

$$\frac{\partial}{\partial x_2}(x_1^{a_1}x_2 + x_2^{a_2}x_3 + x_3^{a_3}x_4 + x_4^{a_4}x_1) = x_1^{a_1} + a_2x_2^{a_2-1}x_3 \neq 0,$$

so locally, by the implicit function theorem, we can write x_2 as a function of x_3 and x_4 , which become local parameters. Then the action of \mathbb{C}^* restricted to $(x_1 = 1)$ is

$$\zeta_1 \cdot (1, x_2, x_3, x_4) = (1, \zeta_1^{w_2}x_2, \zeta_1^{w_3}x_3, \zeta_1^{w_4}x_4),$$

where ζ_1 is a w_1 -th primitive root of 1. Therefore, after taking the quotient, the singularity is a cyclic singularity of type $\frac{1}{w_1}(w_3, w_4)$, if $\gcd(w_1, w_3) = 1$. If $\gcd(w_1, w_3) = h > 1$, then there are elements which fix the axis $(x_3 = 0)$, so they are quasireflections. We eliminate them by dividing $w_1 = ht_1$ and $w_3 = ht_3$ by h , obtaining that the singularity is $\frac{1}{t_1}(t_3, w_4)$. \square

Assume $a_i \geq 2$ for all i .¹ We have this key configuration of curves on $X(a_1, a_2, a_3, a_4)$ (Figure 2):

$$\begin{aligned} C_1 &:= (x_1 = x_3 = 0), \\ C_2 &:= (x_2 = x_4 = 0), \\ \Gamma_{1,2} &:= (x_3 = x_4^{a_4} + x_1^{a_1-1}x_2 = 0), \\ \Gamma_{2,3} &:= (x_4 = x_1^{a_1} + x_2^{a_2-1}x_3 = 0), \\ \Gamma_{3,4} &:= (x_1 = x_2^{a_2} + x_3^{a_3-1}x_4 = 0), \\ \Gamma_{4,1} &:= (x_2 = x_3^{a_3} + x_4^{a_4-1}x_1 = 0). \end{aligned}$$

Proposition 3.2. *The curves C_1, C_2 are smooth and rational. The curve $\Gamma_{i,j}$ is rational, and it may only have a unibranch singularity at p_j .*

Proof. The curves C_1, C_2 are obviously isomorphic to \mathbb{P}^1 . To prove the assertion about $\Gamma_{i,j}$, it is enough to do it for $\Gamma_{2,3}$. Notice that this curve lives in $(x_4 = 0) = \mathbb{P}(w_1, w_2, w_3)$, and that it is possibly singular only at $(0 : 0 : 1)$. Let us consider the $\mathbb{Z}/w_1 \oplus \mathbb{Z}/w_2 \oplus \mathbb{Z}/w_3$ quotient map

$$\mathbb{P}^2 \rightarrow \mathbb{P}(w_1, w_2, w_3)$$

¹This is to have the key configuration of curves as shown. By Theorem 5.1, Kollár surfaces with $a_i = 1$ are birationally included in our analysis. Also, check Corollary 4.8 when $w^* = 1$.

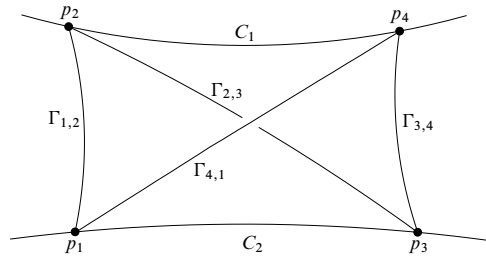


Figure 2. Key configuration of curves on a Kollár surface.

given by $(x : y : z) \mapsto (x^{w_1} : y^{w_2} : z^{w_3})$. Then the preimage of $\Gamma_{2,3}$ is

$$\Gamma'_{2,3} = (x^{w_1 a_1} + y^{w_2(a_2-1)} z^{w_3} = 0),$$

and so $\Gamma_{2,3}$ is rational since all irreducible components (branches at $(0 : 0 : 1)$) of $\Gamma'_{2,3}$ are rational curves.

To see that $\Gamma_{2,3}$ is unibranch at $(0 : 0 : 1)$, we will show that the (possible) branches of $\Gamma'_{2,3}$ form one orbit under the $\mathbb{Z}/w_1 \oplus \mathbb{Z}/w_2 \oplus \mathbb{Z}/w_3$ action. We take the canonical affine chart at $(0 : 0 : 1)$, where $\Gamma'_{2,3} = (x^{w_1 a_1} + y^{w_2(a_2-1)} = 0)$. We consider the action of \mathbb{Z}/w_3 given by $(x, y) \mapsto (\zeta_3^k x, \zeta_3^k y)$ where $k \in \mathbb{Z}$ and $\zeta_3 = e^{2\pi i/w_3}$. Notice that $\gcd(w_2, w_1) = 1$ and $\gcd(w_2, a_1) = 1$ by definition, and so we write $a_2 - 1 = rb$ and $w_1 a_1 = ra$ where $\gcd(a, b) = 1$, to factor in branches

$$x^{w_1 a_1} + y^{w_2(a_2-1)} = \prod_{c=0}^{r-1} (y^{w_2 b} - \zeta_{2r}^{2c+1} x^a)$$

where $\zeta_{2r} = e^{\pi i/r}$. Then we take $y^{w_2 b} - \zeta_{2r} x^a$ and apply $(x, y) \mapsto (\zeta_3^k x, \zeta_3^k y)$ to obtain the branch $y^{w_2 b} - \zeta_{2r} \zeta_3^{k(a-w_2 b)} x^a$, but $a - w_2 b = w_3/r$, and so it goes to $y^{w_2 b} - \zeta_{2r}^{2k+1} x^a$. Therefore, branches form one orbit, and the curve $\Gamma_{2,3}$ is unibranch at $(0 : 0 : 1)$. □

Proposition 3.3. Assume that $a_i > w^*$ for some i . Then $\Gamma_{i+2,i+3}$ is nonsingular.

Proof. We take $a_1 > w^*$ to prove that $\Gamma_{3,4}$ is nonsingular. For this we will compute the arithmetic genus of $\Gamma_{3,4}$. Let $\mathbb{P} = \mathbb{P}(w_2, w_3, w_4)$, and consider the exact sequence of sheaves $0 \rightarrow \mathcal{O}_{\mathbb{P}}(-a_2 w_2) \rightarrow \mathcal{O}_{\mathbb{P}} \rightarrow \mathcal{O}_{\Gamma_{3,4}} \rightarrow 0$. From it we have that $\chi(\mathcal{O}_{\Gamma_{3,4}}) = \chi(\mathcal{O}_{\mathbb{P}}) - \chi(\mathcal{O}_{\mathbb{P}}(-a_2 w_2))$. If $\gcd(w_2, w_4) = 1$, then by [Dolgachev 1982, §1.4] we have that $\chi(\mathcal{O}_{\mathbb{P}}) - \chi(\mathcal{O}_{\mathbb{P}}(-a_2 w_2)) = 1 - h^0(\mathbb{P}, \mathcal{O}_{\mathbb{P}}(a_2 w_2 - w_2 - w_3 - w_4))$. Then

$$p_a(\Gamma_{3,4}) = 1 - \chi(\mathcal{O}_{\Gamma_{3,4}}) = h^0(\mathbb{P}, \mathcal{O}_{\mathbb{P}}(a_2 w_2 - w_2 - w_3 - w_4)),$$

so we have to compute the number of nonnegative integer solutions of the equation $w_2 x + w_3 y + w_4 z = a_2 w_2 - w_2 - w_3 - w_4$. As $a_2 w_2 + w_3 = a_3 w_3 + w_4$, then our equation can be written as

$$w_2(x + a_2 z) + w_3(y + (1 - a_3)z) = (a_3 - 2)w_3 - w_2$$

and its solutions are

$$x = -1 - t w_3 - a_2 z, \quad y = a_3 - 2 + t w_2 + (a_3 - 1)z, \quad z = z. \tag{3-2}$$

If $x, y,$ and z are nonnegative, then $t < 0$, so we will change the sign of t and assume that $t > 0$. Then from (3-2) we obtain that

$$a_2z \leq tw_3 - 1$$

and $(a_3 - 1)z \geq tw_2 - a_3 + 2$. Hence, we have that

$$\frac{tw_3 - 1}{a_2} \geq z \geq \frac{tw_2 + 2 - a_3}{a_3 - 1}. \tag{3-3}$$

Replacing with $w_2 = \frac{1}{w^*}(a_3a_4a_1 - a_4a_1 + a_1 - 1)$ and $w_3 = \frac{1}{w^*}(a_4a_1a_2 - a_1a_2 + a_2 - 1)$ we obtain

$$ta_4a_1 - t(a_1 - 1) - \frac{t + w^*}{a_2} \geq w^*z \geq ta_4a_1 - w^* + \frac{t(a_1 - 1) + w^*}{a_3 - 1}.$$

Because $a_1 > w^*$ and $t \geq 1$, then $t(a_1 - 1) \geq w^*$, so $ta_4a_1 - w^* \geq ta_4a_1 - t(a_1 - 1)$. We have that both $(t + w^*)/a_2$ and $(t(a_1 - 1) + w^*)/(a_3 - 1)$ are positive; therefore, the right-hand side of the system (3-3) is greater than the left-hand side, so the system has no solution. Hence, the arithmetic genus of $\Gamma_{3,4}$ is zero and therefore nonsingular.

If $\gcd(w_2, w_4) = h > 1$, then $p_a(\Gamma_{3,4}) = h^1(\mathbb{P}, \mathcal{O}_{\mathbb{P}}(-a_2w_2))$. To compute it, we first have to consider the well formed weighted projective plane $\mathbb{P}' = \mathbb{P}(t_2, w_3, t_4) \simeq \mathbb{P}$, where $t_2 = w_2/h$ and $t_4 = w_4/h$, and following [Dolgachev 1982, Remarks 1.3.2], we have that $\mathcal{O}_{\mathbb{P}}(-a_2w_2) \simeq \mathcal{O}_{\mathbb{P}'}(-a_2t_2)$. Then $p_a(\Gamma_{3,4}) = h^0(\mathbb{P}', \mathcal{O}_{\mathbb{P}'}(a_2t_2 - t_2 - w_3 - t_4))$, which is equivalent to the number of nonnegative integer solutions of the equation

$$t_2x + w_3y + t_4z = a_2t_2 - t_2 - w_3 - t_4.$$

The general solution of this equation is

$$x = -1 - tw_3 - a_2z, \quad y = \frac{a_3 - 1}{h} - 1 + t_2t + \frac{a_3 - 1}{h}z, \quad z = z,$$

with $t \in \mathbb{Z}$. Then $t < 0$, and changing the sign of t as above, we have that the arithmetic genus is equal to the number of solutions of the system

$$a_1a_4t - t(a_1 - 1) - \frac{t + w^*}{a_2} \geq w^*z \geq a_1a_4t - w^* + \frac{hw^* + (a_1 - 1)t}{a_3 - 1},$$

but again, as $a_i > w^*$, then the right-hand side is greater than the left-hand side, so the arithmetic genus is 0. □

Proposition 3.4. *The map ψ is defined precisely in $X(a_1, a_2, a_3, a_4) \setminus \{p_1, p_2, p_3, p_4\}$, and it contracts*

$$\begin{aligned} \psi(C_1 \setminus \{p_2, p_4\}) &= (0 : 1 : 0 : -1), & \psi(C_2 \setminus \{p_1, p_3\}) &= (1 : 0 : -1 : 0), \\ \psi(\Gamma_{1,2} \setminus \{p_1, p_2\}) &= (-1 : 0 : 0 : 1), & \psi(\Gamma_{2,3} \setminus \{p_2, p_3\}) &= (1 : -1 : 0 : 0), \\ \psi(\Gamma_{3,4} \setminus \{p_3, p_4\}) &= (0 : 1 : -1 : 0), & \psi(\Gamma_{4,1} \setminus \{p_4, p_1\}) &= (0 : 0 : 1 : -1). \end{aligned}$$

Proof. We have that $\psi|_{\Gamma_{1,2} \setminus \{p_1, p_2\}} = (x_1^{a_1-1}x_2 : 0 : 0 : x_4^{a_4})$, and because $x_1^{a_1-1}x_2 = -x_4^{a_4}$ over $\Gamma_{1,2}$, then $\psi|_{\Gamma_{1,2} \setminus \{p_1, p_2\}} = (-1 : 0 : 0 : 1)$. This gives the result for all curves $\Gamma_{i,i+1}$.

For C_1 , let $x_4 = 1$ and $x_2 = b \neq 0$. Then the equation of the surface with these restrictions is

$$bx_1^{a_1} + b^{a_2}x_3 + x_3^{a_3} + x_1 = x_1(1 + bx_1^{a_1-1}) + x_3(b^{a_2} + x_3^{a_3-1}) = 0.$$

The map is $\psi(x_1 : b : x_3 : 1) = (bx_1^{a_1} : b^{a_2}x_3 : x_3^{a_3} : x_1)$. We multiply every coordinate by $(1 + bx_1^{a_1-1})$, and use the relation $x_1(1 + bx_1^{a_1-1}) = -x_3(b^{a_2} + x_3^{a_3-1})$, to write down $\psi(x_1 : b : x_3 : 1)$ as

$$\begin{aligned} & (bx_1^{a_1}(1 + bx_1^{a_1-1}) : b^{a_2}x_3(1 + bx_1^{a_1-1}) : x_3^{a_3}(1 + bx_1^{a_1-1}) : x_1(1 + bx_1^{a_1-1})) \\ &= (-x_3bx_1^{a_1-1}(b^{a_2} + x_3^{a_3-1}) : b^{a_2}x_3(1 + bx_1^{a_1-1}) : x_3^{a_3}(1 + bx_1^{a_1-1}) : -x_3(b^{a_2} + x_3^{a_3-1})) \\ &= (-bx_1^{a_1-1}(b^{a_2} + x_3^{a_3-1}) : b^{a_2}(1 + bx_1^{a_1-1}) : x_3^{a_3-1}(1 + bx_1^{a_1-1}) : -(b^{a_2} + x_3^{a_3-1})). \end{aligned}$$

Hence, $\psi(0 : b : 0 : 1) = (0 : b^{a_2} : 0 : -b^{a_2}) = (0 : 1 : 0 : -1)$. A similar argument works for C_2 . □

Remark 3.5. By Theorem 5.1, we know that any $X(a_1, a_2, a_3, a_4)$ has a birational model $X(a'_1, a'_2, a'_3, a'_4)$ with $\gcd(w'_i, w'_{i+2}) = 1$. From now on, we assume that $\gcd(w_1, w_3) = \gcd(w_2, w_4) = 1$.

Now we want to study the behavior of ψ on a resolution of the singularities in $X(a_1, a_2, a_3, a_4)$. To do so, we need to write this map in terms of local coordinates in the resolution, which are described in:

Theorem 3.6 [Reid 2003, Theorem 3.2]. *Let $X = \mathbb{C}^2 / (\mathbb{Z}/m)$ be a cyclic singularity of type $\frac{1}{m}(a, b)$, and let $\frac{1}{m}(a, b) = \frac{1}{m}(1, q)$ be as explained at the beginning of Section 3. Let N be the lattice $N = \mathbb{Z}^2 + \mathbb{Z} \cdot \frac{1}{m}(1, q)$, and*

$$M = \{(r, s) : r + qs \equiv 0 \pmod{m}\} \subset \mathbb{Z}^2$$

the dual lattice of invariant monomials under the action $(x, y) \mapsto (\zeta_m x, \zeta_m^q y)$ with ζ_m an m -th primitive root of unity.

Let $m/q = [b_1, \dots, b_s]$, and let z_0, z_1, \dots, z_{s+1} be vectors in N defined as

$$z_i = \frac{1}{m}(\alpha_i, \beta_i),$$

where α_i and β_i are as defined at the beginning of Section 3. Then for each $i = 0, \dots, s$, let u_i, v_i be monomials forming the dual basis of M to z_i, z_{i+1} ; that is, $u_i = (\beta_i, -\alpha_i)$ and $v_i = (-\beta_{i+1}, \alpha_{i+1})$.

Then X has a resolution of singularities $Y \rightarrow X$ constructed as

$$Y = U_0 \cup U_1 \cup \dots \cup U_s,$$

where $U_i \simeq \mathbb{C}^2$ with coordinates u_i, v_i .

The gluing $U_i \cup U_{i+1}$ and the morphism $Y \rightarrow X$ are both determined by the definition of u_i, v_i and they consist of

$$U_i \setminus (v_i = 0) \xrightarrow{\cong} U_{i+1} \setminus (u_{i+1} = 0) \quad \text{given by } u_{i+1} = v_i^{-1} \text{ and } v_{i+1} = u_i v_i^{b_i}.$$

It follows from the definition of the numbers α_i and β_i that $u_0 = x^m$ and $v_s = y^m$, and they satisfy the relations

$$x^m = u_i^{\alpha_{i+1}} v_i^{\alpha_i} \quad \text{and} \quad y^m = u_i^{\beta_{i+1}} v_i^{\beta_i}.$$

Theorem 3.7. *Let $\sigma : \widetilde{X} \rightarrow X(a_1, a_2, a_3, a_4)$ be the minimal resolution, and let*

$$\widehat{X} \xrightarrow{\psi} \widetilde{X} \xrightarrow{\sigma} X(a_1, a_2, a_3, a_4)$$

be the minimal log resolution of X together with the key configuration of curves. Then $\psi \circ \sigma \circ \varphi$ is a morphism; i.e., the indeterminacies of ψ can be resolved by $\sigma \circ \varphi$.

To prove Theorem 3.7 we have to compute the strict transform of the curves $\Gamma_{i,i+1}$ on \widetilde{X} . Let $E_{i,j}$ be the components of the exceptional divisor over the point p_i , let $\frac{1}{w_i}(w_{i+2}, w_{i+3}) = \frac{1}{w_i}(1, q_i)$, and let $\alpha_{i,j}, \beta_{i,j}$, and $\gamma_{i,j}$ be the integers defined for the continued fraction of w_i/q_i . Recall from the proof of Proposition 3.1 that x_{i+2} and x_{i+3} are toric local coordinates at p_i , so we have that $E_{i,0}$ and E_{i,s_i+1} are the strict transforms of $(x_{i+3} = 0)$ and $(x_{i+2} = 0)$ at the open set $(x_i \neq 0)$. This means that $E_{1,0} = E_{3,0}$ and $E_{2,0} = E_{4,0}$ and they correspond to the strict transforms of C_2 and C_1 , respectively. On the other hand, E_{i,s_i+1} corresponds to the strict transform of the curve $\Gamma_{i,i+1}$. (See Figure 3 to visualize the notation.) Then it remains to compute the strict transform of $\Gamma_{i,i+1}$ around the point p_{i+1} , and without loss of generality, we will compute the strict transform $\Gamma_{3,4}$ at the point p_4 . As all the results will hold locally for $\Gamma_{3,4}$, we can modify the following proofs for every $\Gamma_{i,i+1}$.

Proposition 3.8. *Let $U_{4,j}$ be the open sets of the resolution of $\frac{1}{w_4}(1, q_4)$ as defined in Theorem 3.6. Then the local equation of the strict transform of the curve $\Gamma_{3,4}$ restricted to the open set $U_{4,j}$ is*

$$\Gamma'_{34} = \begin{cases} 1 + u_j^{((a_3-1)\beta_{4,j+1}-a_2\alpha_{4,j+1})/w_4} v_j^{((a_3-1)\beta_{4,j}-a_2\alpha_{4,j})/w_4} = 0 & \text{if } a_2\alpha_{4,j+1} - (a_3 - 1)\beta_{4,j+1} \leq 0, \\ u_j^{(a_2\alpha_{4,j+1}-(a_3-1)\beta_{4,j+1})/w_4} v_j^{(a_2\alpha_{4,j}-(a_3-1)\beta_{4,j})/w_4} + 1 = 0 & \text{if } 0 \leq a_2\alpha_{4,j} - (a_3 - 1)\beta_{4,j}, \\ u_j^{(a_2\alpha_{4,j+1}-(a_3-1)\beta_{4,j+1})/w_4} + v_j^{((a_3-1)\beta_{4,j}-a_2\alpha_{4,j})/w_4} = 0 & \text{if } a_2\alpha_{4,j} - (a_3 - 1)\beta_{4,j} \leq 0 \\ & \leq a_2\alpha_{4,j+1} - (a_3 - 1)\beta_{4,j+1}. \end{cases}$$

Proof. We can assume that $x_4 = 1$ and $x_1 = 0$, so we must study the curve $(x_2^{a_2} + x_3^{a_3-1} = 0) \subset (x_4 \neq 0) \subset \mathbb{P}(w_2, w_3, w_4)$. By Theorem 3.6, to find the total transform of $\Gamma_{3,4}$ in U_i we replace x_2 and x_3 with $u_i^{\alpha_{4,i+1}/w_4} v_i^{\alpha_{4,i}/w_4}$ and $u_i^{\beta_{4,i+1}/w_4} v_i^{\beta_{4,i}/w_4}$, respectively, and so the total transform is

$$(u_i^{\alpha_{4,i+1}/w_4} v_i^{\alpha_{4,i}/w_4})^{a_2} + (u_i^{\beta_{4,i+1}/w_4} v_i^{\beta_{4,i}/w_4})^{a_3-1} = 0.$$

Recall that $\alpha_{4,i} < \alpha_{4,i+1}$ and $\beta_{4,i+1} < \beta_{4,i}$, so

$$a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i} < a_2\alpha_{4,i+1} - (a_3 - 1)\beta_{4,i+1}.$$

If both sides are negative, we factor out $(u_i^{\alpha_{4,i+1}/w_4} v_i^{\alpha_{4,i}/w_4})^{a_2}$. If both sides are positive, we factor out $(u_i^{\beta_{4,i+1}/w_4} v_i^{\beta_{4,i}/w_4})^{a_3-1}$. If $a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i} \leq 0 \leq a_2\alpha_{4,i+1} - (a_3 - 1)\beta_{4,i+1}$, we factor out $u_i^{((a_3-1)\beta_{4,i+1})/w_4}$ and $v_i^{a_2\alpha_{4,i}/w_4}$, obtaining what we wanted to prove. \square

By Proposition 3.8, the curve $\Gamma'_{3,4}$ intersects the exceptional divisor if and only if

$$a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i} \leq 0 \leq a_2\alpha_{4,i+1} - (a_3 - 1)\beta_{4,i+1}.$$

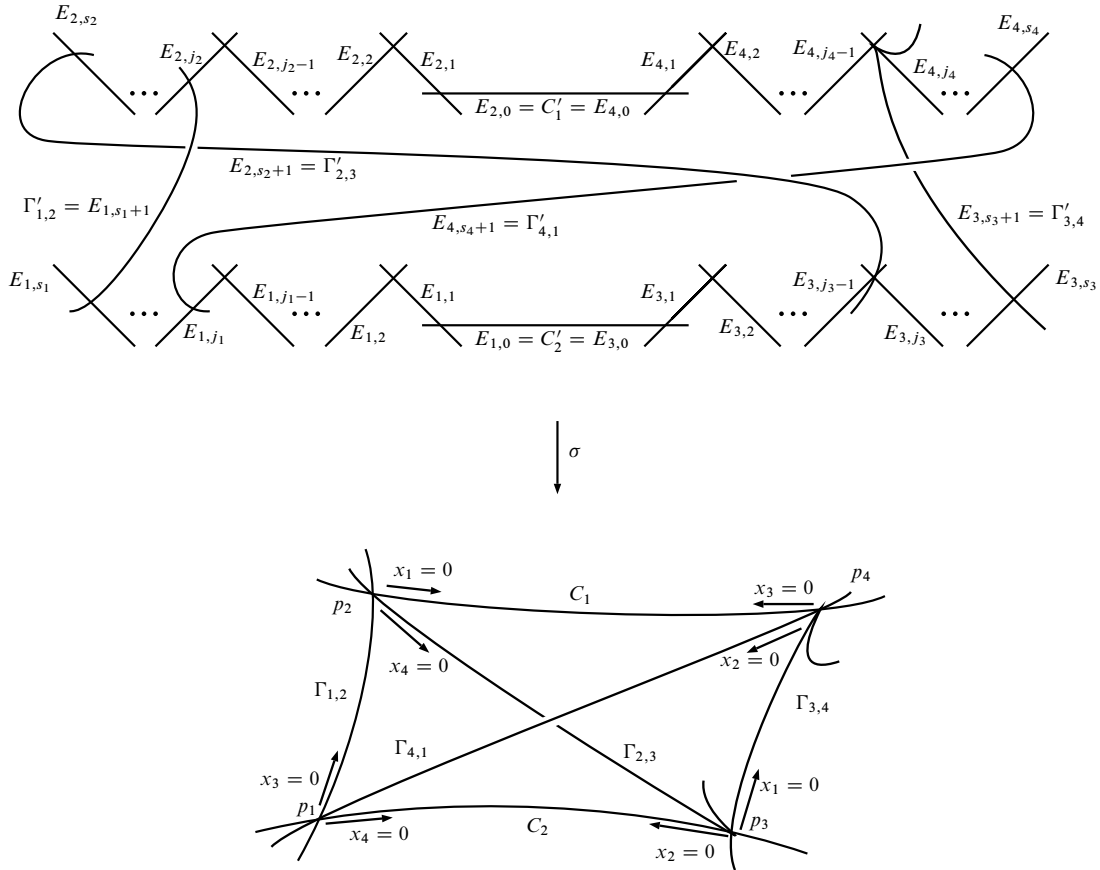


Figure 3. Key configuration of curves on $X(a_1, a_2, a_3, a_4)$ and the curve configuration of the minimal resolution \tilde{X} .

If $a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i} < 0 < a_2\alpha_{4,i+1} - (a_3 - 1)\beta_{4,i+1}$, then the curve intersects two components of the exceptional divisor, and if $a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i} = 0$ or $a_2\alpha_{4,i+1} - (a_3 - 1)\beta_{4,i+1} = 0$, then it intersects only one component.

Proposition 3.9. *Let us say that $\Gamma'_{3,4}$ intersects the exceptional divisor over p_4 at the components $E_{4,j}$ and $E_{4,j+1}$ with multiplicity m_j and m_{j+1} , respectively (possibly $m_{j+1} = 0$). Then $a_3 - 1 = \alpha_{4,j}m_j + \alpha_{4,j+1}m_{j+1}$ and $a_2 = \beta_{4,j}m_j + \beta_{4,j+1}m_{j+1}$.*

Proof. Let H be the restriction to $X(a_1, a_2, a_3, a_4)$ of a generator of the class group of $\mathbb{P}(w_1, w_2, w_3, w_4)$. We have that

$$w_1H \cdot w_2H = \frac{w_1w_2(a_3w_3 + w_4)}{w_1w_2w_3w_4} = \frac{1}{w_3} + \frac{a_3}{w_4}.$$

On the other hand, $w_1H \cdot w_2H = \sigma^*(w_1H) \cdot \sigma^*(w_2H)$, where $\sigma^*(w_1H) = \sigma^*(\Gamma_{3,4} + C_1)$, and $\sigma^*(w_2H) = \sigma^*(\Gamma_{4,1} + C_2)$. Because the pull-back of a divisor has intersection zero with any component of the

exceptional divisor, and using the pull-back formulas in (3-1), we have that

$$\begin{aligned} \sigma^*(w_1H) \cdot \sigma^*(w_2H) &= (\Gamma'_{3,4} + C'_1) \cdot \left(\sum_{i=0}^{s_3+1} \frac{\beta_{3,i}}{w_3} E_{3,i} + \sum_{i=0}^{s_4+1} \frac{\alpha_{4,i}}{w_4} E_{4,i} \right) \\ &= \Gamma'_{3,4} \cdot \sum_{i=0}^{s_3+1} \frac{\beta_{3,i}}{w_3} E_{3,i} + C'_1 \cdot \sum_{i=0}^{s_4+1} \frac{\alpha_{4,i}}{w_4} E_{4,i} + \Gamma'_{3,4} \cdot \sum_{i=0}^{s_4+1} \frac{\alpha_{4,i}}{w_4} E_{4,i} \\ &= \frac{1}{w_3} + \frac{1}{w_4} + \sum_{i=0}^{s_4+1} \frac{\alpha_{4,i}}{w_4} \Gamma'_{3,4} \cdot E_{4,i}. \end{aligned}$$

Then $a_3 - 1 = \alpha_{4,j} \Gamma'_{3,4} \cdot E_{4,j} + \alpha_{4,j+1} \Gamma'_{3,4} \cdot E_{4,j+1} = \alpha_{4,j} m_j + \alpha_{4,j+1} m_{j+1}$. To simplify the computation of the second equality, we will restrict to the plane $\mathbb{P}(w_2, w_3, w_4)$, with L a generator of the class group. We can do this because at the point p_4 the singularity is the same as the one at the point $(0 : 0 : 1) \in \mathbb{P}(w_2, w_3, w_4)$, so locally σ does not change.

Then $w_3L \cdot a_2w_2L = a_2w_2w_3/(w_2w_3w_4) = a_2/w_4$ and also

$$\sigma^*(w_3L) \cdot \sigma^*(a_2w_2L) = \Gamma'_{3,4} \cdot \sum_{i=0}^{s_4+1} \frac{\beta_{4,i}}{w_4} E_{4,i},$$

where $\sigma^*(w_3L) = \sigma^*(C_1)$ and $\sigma^*(a_2w_2L) = \sigma^*(\Gamma_{3,4})$. Then $a_2 = \beta_{4,j} m_j + \beta_{4,j+1} m_{j+1}$. □

Corollary 3.10. *If $\Gamma'_{3,4}$ intersects the exceptional divisor in one component, then it does it transversally at one point.*

Proof. Recall that in the open subset $U_{4,i}$, the exponents of the variables u_i and v_i of the strict transform of $\Gamma_{3,4}$ are $\pm(a_2\alpha_{4,i+1} - (a_3 - 1)\beta_{4,i+1})/w_4$ and $\pm(a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i})/w_4$.

Suppose that $\Gamma'_{3,4}$ intersects E_j with multiplicity m_j . Then, using Proposition 3.9, we have that $a_2 = \beta_{4,j} m_j + \beta_{4,j+1} m_{j+1}$ and $a_3 - 1 = \alpha_{4,j} m_j + \alpha_{4,j+1} m_{j+1}$, and in this case $m_{j+1} = 0$. Hence, for all i

$$\frac{a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i}}{w_4} = m_j \frac{\beta_{4,j}\alpha_{4,i} - \alpha_{4,j}\beta_{4,i}}{w_4},$$

but the singularity at p_4 was unibranch, so it is locally irreducible. Therefore, the exponents on the resolution must be relatively prime. Thus, $m_j = 1$. □

Theorem 3.11. *The curve $\Gamma'_{3,4}$ intersects the exceptional divisor in one component if and only if $\psi \circ \sigma$ is defined on the whole exceptional divisor over p_4 .*

Proof. The equation of our surface is $x_1^{a_1} x_2 + x_2^{a_2} x_3 + x_3^{a_3} x_4 + x_4^{a_4} x_1 = 0$, so locally at p_4 our surface is $(x_1^{a_1} x_2 + x_2^{a_2} x_3 + x_3^{a_3} + x_1 = 0)$. Then analytically the power series expansion of x_1 in terms of x_2 and x_3 is

$$x_1 = -x_2^{a_2} x_3 - x_3^{a_3} + (\text{higher order terms in } x_2 \text{ and } x_3).$$

Therefore, at the open set U_i

$$\sigma^*(x_1) = -(u_i^{\alpha_{4,i+1}/w_4} v_i^{\alpha_{4,i}/w_4}) a_2 (u_i^{\beta_{4,i+1}/w_4} v_i^{\beta_{4,i}/w_4}) - (u_i^{\beta_{4,i+1}/w_4} v_i^{\beta_{4,i}/w_4}) a_3 + (\text{higher order terms}),$$

and so

$$\begin{aligned} \psi \circ \sigma|_{U_i} = & \left((*) : u_i^{(a_2\alpha_{4,i+1} + \beta_{4,i+1})/w_4} v_i^{(a_2\alpha_{4,i} + \beta_{4,i})/w_4} : u_i^{a_3\beta_{4,i+1}/w_4} v_i^{a_3\beta_{4,i}/w_4} \right. \\ & \left. : -u_i^{(a_2\alpha_{4,i+1} + \beta_{4,i+1})/w_4} v_i^{(a_2\alpha_{4,i} + \beta_{4,i})/w_4} - u_i^{a_3\beta_{4,i+1}/w_4} v_i^{a_3\beta_{4,i}/w_4} + (*) \right), \end{aligned}$$

where $(*)$ are terms in u_i and v_i of degree higher than $(a_2\alpha_{4,i+1} + \beta_{4,i+1} + a_2\alpha_{4,i} + \beta_{4,i})/w_4$ and $(a_3\beta_{4,i+1} + a_3\beta_{4,i})/w_4$.

Assume now that u_i and v_i are both nonzero. If $a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i} < a_2\alpha_{4,i+1} - (a_3 - 1)\beta_{4,i+1} < 0$, then we can factor out

$$(u_i^{\alpha_{4,i+1}/w_4} v_i^{\alpha_{4,i}/w_4}) a_2 (u_i^{\beta_{4,i+1}/w_4} v_i^{\beta_{4,i}/w_4})$$

from $\psi \circ \sigma$ to obtain

$$\psi \circ \sigma|_{U_i} = ((*) : 1 : u_i^{((a_3-1)\beta_{4,i+1} - a_2\alpha_{4,i+1})/w_4} v_i^{((a_3-1)\beta_{4,i} - a_2\alpha_{4,i})/w_4} : -1 + (*)).$$

Then $(\psi \circ \sigma|_{U_i})(u_i, 0) = (\psi \circ \sigma|_{U_i})(0, v_i) = (0 : 1 : 0 : -1)$. Repeating the same procedure for $0 < a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i} < a_2\alpha_{4,i+1} - (a_3 - 1)\beta_{4,i+1}$, we obtain that, restricted to that open set U_i ,

$$(\psi \circ \sigma|_{U_i})(u_i, 0) = (\psi \circ \sigma|_{U_i})(0, v_i) = (0 : 0 : 1 : -1).$$

Now we are left with the case $a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i} \leq 0 \leq a_2\alpha_{4,i+1} - (a_3 - 1)\beta_{4,i+1}$. Suppose first that the curve $\Gamma'_{3,4}$ intersects one component of the exceptional divisor, so Proposition 3.9 implies that there is some j such that $a_2\alpha_{4,j} - (a_3 - 1)\beta_{4,j} = 0$. By Corollary 3.10, $\Gamma'_{3,4}$ intersects the exceptional divisor transversally at one point, so $(a_2\alpha_{4,j+1} - (a_3 - 1)\beta_{4,j+1})/w_4 = 1$, and $(a_2\alpha_{4,j-1} - (a_3 - 1)\beta_{4,j-1})/w_4 = -1$. Then in U_{j-1} we can still factor out

$$(u_i^{\alpha_{4,i+1}/w_4} v_i^{\alpha_{4,i}/w_4}) a_2 (u_i^{\beta_{4,i+1}/w_4} v_i^{\beta_{4,i}/w_4}),$$

so assuming that u_{j-1} and v_{j-1} are not zero, the maps looks like

$$\psi \circ \sigma|_{U_{j-1}} = ((*) : 1 : v_{j-1} : -1 - v_{j-1} + (*)).$$

Therefore, $(\psi \circ \sigma|_{U_{j-1}})(u_{j-1}, 0) = (0 : 1 : 0 : -1)$ and $(\psi \circ \sigma|_{U_{j-1}})(0, v_{j-1}) = (0 : 1 : v_{j-1} : -1 - v_{j-1})$. Doing the same for U_j we find that $(\psi \circ \sigma|_{U_j})(0, v_j) = (0 : 0 : 1 : -1)$ and $(\psi \circ \sigma|_{U_j})(u_j, 0) = (0 : u_j : 1 : -u_j - 1)$. Then we see that $\psi \circ \sigma(\bigcup_{i=0}^{j-1} E_{4,i}) = (0 : 1 : 0 : -1)$ and $\psi \circ \sigma(\bigcup_{i=j+1}^{s_4+1} E_{4,i}) = (0 : 0 : 1 : -1)$. Notice that v_{j-1} and u_j are the coordinates of the charts of $E_j \simeq \mathbb{P}^1$ and that

$$(\psi \circ \sigma|_{U_{j-1}})(0, v_{j-1}) = (0 : 1 : v_{j-1} : -1 - v_{j-1})$$

and

$$(\psi \circ \sigma|_{U_j})(u_j, 0) = (0 : u_j : 1 : -u_j - 1).$$

So $\psi \circ \sigma$ is an isomorphism from E_j onto the line $(y_1 = 0) \subset (y_1 + y_2 + y_3 + y_4 = 0) \subset \mathbb{P}^3_{y_1, y_2, y_3, y_4}$. Therefore, $\psi \circ \sigma$ is defined at the exceptional divisor over p_4 , and it is totally branched over the line $L_1 = (y_1 = 0) \subset (y_1 + y_2 + y_3 + y_4 = 0)$.

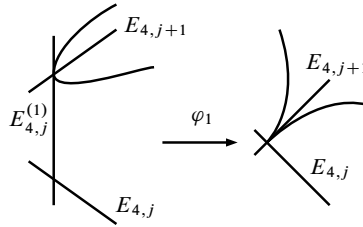


Figure 4. An example of the situation in Proposition 3.12.

Now, if $\Gamma'_{3,4}$ does not intersect transversally the exceptional divisor, then $a_2\alpha_{4,i} - (a_3 - 1)\beta_{4,i} \neq 0$ for all i , so we will have some j such that

$$a_2\alpha_{4,j} - (a_3 - 1)\beta_{4,j} < 0 < a_2\alpha_{4,j+1} - (a_3 - 1)\beta_{4,j+1},$$

and we will not be able to define the map on the open set U_j . This is because we can factor out $u_j^{a_3\beta_{4,j+1}} v_j^{a_2\alpha_{4,j} + \beta_{4,j}}$ from $\psi \circ \sigma|_{U_j}$, so the map will be

$$\begin{aligned} \psi \circ \sigma|_{U_j} = ((*) : u_j^{(a_2\alpha_{4,j+1} - (a_3 - 1)\beta_{4,j+1})/w_4} : v_j^{((a_3 - 1)\beta_{4,j} - a_2\alpha_{4,j})/w_4} \\ : -u_j^{(a_2\alpha_{4,j+1} - (a_3 - 1)\beta_{4,j+1})/w_4} - v_j^{((a_3 - 1)\beta_{4,j} - a_2\alpha_{4,j})/w_4} + (*)). \end{aligned}$$

Then if $v_j \neq 0$, $(\psi \circ \sigma|_{U_j})(0, v_j) = (0 : 0 : 1 : -1)$, and if $u_j \neq 0$, we have $(\psi \circ \sigma|_{U_j})(u_j, 0) = (0 : 1 : 0 : -1)$, and so it is not well defined when $u_j = v_j = 0$. □

Proposition 3.12. *Assume that $\Gamma'_{3,4}$ does not intersect transversally the exceptional divisor, so it intersects it at the point $(0, 0)$ of some affine open set U_j . Let $\varphi_1 : X_1 \rightarrow \tilde{X}$ be the blowup over that point, let $E_{4,j}^{(1)}$ be the new component of the exceptional divisor, and let $u_j, v'_{j,1}$ and $u'_{j,1}, v_j$ be the affine coordinates of $U_j^{(1,1)}$ and $U_j^{(1,2)}$, the two affine charts over U_j . Then they satisfy the relations $x_2^{w_4} = u_j^{\alpha_{4,j} + \alpha_{4,j+1}} v'_{j,1}{}^{\alpha_{4,j}} = u'_{j,1}{}^{\alpha_{4,j+1}} v_j^{\alpha_{4,j} + \alpha_{4,j+1}}$ and $x_3^{w_4} = u_j^{\beta_{4,j} + \beta_{4,j+1}} v'_{j,1}{}^{\beta_{4,j}} = u'_{j,1}{}^{\beta_{4,j+1}} v_j^{\beta_{4,j} + \beta_{4,j+1}}$.*

Proof. This follows from the fact that the resolution was constructed as a toric variety, and the blowup of an affine variety defined by vectors v_1 and v_2 is the variety associated to the fan generated by the vectors $v_1, v_1 + v_2$, and v_2 . (Figure 4 shows an example of the situation in the proposition.) □

Notice that, if $a_2\alpha_{4,j} - (a_3 - 1)\beta_{4,j} < 0 < a_2\alpha_{4,j+1} - (a_3 - 1)\beta_{4,j+1}$, then

$$a_2\alpha_{4,j} - (a_3 - 1)\beta_{4,j} < a_2(\alpha_{4,j} + \alpha_{4,j+1}) - (a_3 - 1)(\beta_{4,j} + \beta_{4,j+1})$$

and

$$a_2(\alpha_{4,j} + \alpha_{4,j+1}) - (a_3 - 1)(\beta_{4,j} + \beta_{4,j+1}) < a_2\alpha_{4,j+1} - (a_3 - 1)\beta_{4,j+1},$$

so we can use Proposition 3.8 to see that the strict transform of $\Gamma'_{3,4}$ in the blowup intersects at most two components of the exceptional divisor, and that the singularity of the curve is “better”. Therefore, the map $\psi \circ \sigma \circ \varphi_1$ is defined in one of the charts $U_j^{(1,i)}$, and if $a_2(\alpha_{4,j} + \alpha_{4,j+1}) - (a_3 - 1)(\beta_{4,j} + \beta_{4,j+1}) = 0$, then it is defined in all the exceptional divisor on X_1 over p_4 .

Proof of Theorem 3.7. If all the curves $\Gamma'_{i,i+1}$ intersect transversally the exceptional divisor on \tilde{X} , then the result follows from Theorem 3.11. If not, then consider the log resolution $\varphi : \hat{X} \rightarrow X$ of all the curves $\Gamma'_{i,i+1}$. Proposition 3.12 shows that the relations of the new local coordinates are compatible with the previous ones, and as the strict transform of the curves $\Gamma'_{i,i+1}$ intersect transversally the exceptional divisor, we can use the proof of Theorem 3.11 to show that the composition $\psi \circ \sigma \circ \varphi$ is defined over \hat{X} . \square

Corollary 3.13. *The morphisms $\psi \circ \sigma \circ \varphi : \hat{X} \rightarrow \mathbb{P}^2$ and $Y' \rightarrow \mathbb{P}^2$ (defined at the end of Section 2) factor through a birational morphism $\hat{X} \rightarrow Y'$ which contracts precisely six chains of smooth rational curves in*

$$(\sigma \circ \varphi)^*(C_1 + C_2 + \Gamma_{1,2} + \Gamma_{2,3} + \Gamma_{3,4} + \Gamma_{4,1}),$$

each containing one of the proper transforms of $C_1, C_2, \Gamma_{1,2}, \Gamma_{2,3}, \Gamma_{3,4}, \Gamma_{4,1}$, and each contracting to the six cyclic quotient singularities in Y' .

Proof. First, by Theorem 3.7, we note that $\psi \circ \sigma \circ \varphi : \hat{X} \rightarrow \mathbb{P}^2$ contracts precisely six chains of smooth rational curves in $(\sigma \circ \varphi)^*(C_1 + C_2 + \Gamma_{1,2} + \Gamma_{2,3} + \Gamma_{3,4} + \Gamma_{4,1})$, each containing one of the proper transforms of $C_1, C_2, \Gamma_{1,2}, \Gamma_{2,3}, \Gamma_{3,4}, \Gamma_{4,1}$. This was done locally when we proved the definition of the map in Theorem 3.11 at a certain exceptional component over the p_i . Each of these components maps to each of the four lines in \mathbb{P}^2 . Therefore, the birational map $\hat{X} \dashrightarrow Y'$ is defined over these components except possibly over the six singularities of Y' . Because there is a unique minimal resolution for normal two-dimensional singularities, the six chains of curves in \hat{X} mapping to the six nodes of the four lines in \mathbb{P}^2 must contract to the six singularities of Y' . \square

4. Kollár surfaces are Hwang–Keum surfaces

We now study the case $w^* = 1$. In this section, we allow $\gcd(w_1, w_3)$ and $\gcd(w_2, w_4)$ to be greater than 1.

In [Kollár 2008, p. 231], it is shown that the curves C_1 and C_2 are extremal rays of the $K_{X(a_1, a_2, a_3, a_4)} + (1 - \epsilon)(C_1 + C_2)$ minimal model program if $C_1^2 < 0$ and $C_2^2 < 0$. They are both contractible to quotient singularities. Hwang and Keum [2012] computed explicitly the type of these singularities.

Theorem 4.1 [Hwang and Keum 2012, Theorem 1.1]. *The contraction of the curve C_1 forms a singularity of type $\frac{1}{s_1}(w_2, w_4)$, with $s_1 = a_4 w_4 - w_3$, and the contraction of the curve C_2 forms a singularity of type $\frac{1}{s_2}(w_1, w_3)$, with $s_2 = a_3 w_3 - w_2$. If $w^* = 1$, then their Hirzebruch–Jung continued fractions are*

$$[\underbrace{2, \dots, 2}_{a_4-1}, a_3, a_1, \underbrace{2, \dots, 2}_{a_2-1}] \quad \text{and} \quad [\underbrace{2, \dots, 2}_{a_3-1}, a_2, a_4, \underbrace{2, \dots, 2}_{a_1-1}],$$

respectively.

Let $\eta : X(a_1, a_2, a_3, a_4) \rightarrow X'(a_1, a_2, a_3, a_4)$ be the contraction of C_1 and C_2 . Hwang and Keum [2012, §4] constructed several examples of rational \mathbb{Q} -homology projective planes with two cyclic singularities. In certain cases the singularities are the same as for $X'(a_1, a_2, a_3, a_4)$ when $w^* = 1$.

The construction of Hwang–Keum is as follows. Let L_1, L_2, L_3, L_4 be four general lines in \mathbb{P}^2 , and choose four points from the six intersection points, such that every L_i passes through two of them. After

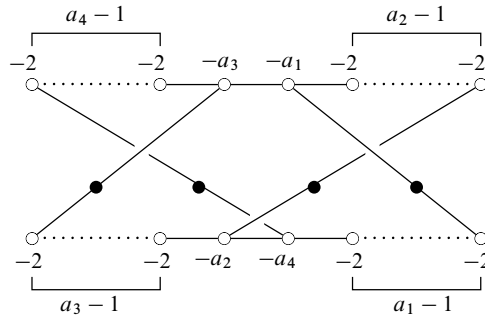
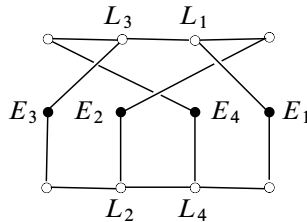


Figure 5. Curve configuration over $Z(a_1, a_2, a_3, a_4)$.

blowing up each of these four points twice, we obtain the curve configuration



where \bullet is a (-1) -curve and \circ is a (-2) -curve. We now blowup r_i times the point $E_i \cap L_i$ to obtain the surface $Z(a_1, a_2, a_3, a_4)$, where $a_i = 2 + r_i$. The curve configuration on $Z(a_1, a_2, a_3, a_4)$ is shown in Figure 5.

Let $T(a_1, a_2, a_3, a_4)$ be the surface obtained by contracting the two chains of rational curves corresponding to the white vertices. Then this surface is a rational \mathbb{Q} -homology projective plane with two cyclic singularities. By Theorem 4.1, it has the same singularities as $X'(a_1, a_2, a_3, a_4)$ when $w^* = 1$.

Theorem 4.2. *Let $X(a_1, a_2, a_3, a_4)$ be a Kollár surface with $w^* = 1$, and assume that $a_i \geq 2$ for all i . Then $X'(a_1, a_2, a_3, a_4)$ is the Hwang–Keum surface $T(a_1, a_2, a_3, a_4)$.*

To prove Theorem 4.2 we will show that we can find the same curve configuration of $Z(a_1, a_2, a_3, a_4)$ (Figure 5) in \tilde{X}' , which is the minimal resolution of $X'(a_1, a_2, a_3, a_4)$.

First of all, we prove that the rational map ψ is defined in the minimal resolution of X . For this we will use:

Proposition 4.3. *Let X be a surface with a cyclic quotient singularity at the point p , and let $C \subset X$ be a curve passing through p . Then C is nonsingular at p if and only if the strict transform of C intersects transversally at one point only one component of the exceptional divisor of the minimal resolution of X .*

Proof. The maximal cycle (which coincides with the fundamental cycle) of a cyclic quotient singularity is the (reduced) exceptional divisor. Then we can apply [Gonzalez-Sprinberg and Lejeune-Jalabert 1997, Proposition 1.1]. □

By Proposition 3.3 we have that the curves $\Gamma_{i,i+1}$ are smooth, so Proposition 4.3 says that the curves $\Gamma'_{i,i+1}$ intersect transversally the exceptional divisor over p_{i+1} . If $\gcd(w_1, w_3) = \gcd(w_2, w_4) = 1$, then we already know that the map ψ is defined on the minimal resolution of X . Therefore, we only need to check the same assertion when $\gcd(w_1, w_3) > 1$ or $\gcd(w_2, w_4) > 1$.

Proposition 4.4. *The map $\psi \circ \sigma : \tilde{X} \rightarrow \mathbb{P}^2$ is a morphism.*

Proof. We study the case over the point p_4 , with $\gcd(w_2, w_4) = h > 1$. The singularity at p_4 is $1/w_4(w_2, w_3)$ with toric coordinates x_2 and x_3 . From Proposition 3.1 we have that $1/w_4(w_2, w_3) \simeq 1/t_4(t_2, w_3)$, with toric coordinates x'_2 and x'_3 , and the relation $x'_2 = x_2$ and $x'_3 = x_3^h$. Then from Theorem 3.6 we have $Y = U_1 \cup \dots \cup U_{s_4}$ in the resolution of p_4 , with u_i, v_i the local coordinates in U_i , and the relations $x_2^{t_4} = u_i^{\alpha_{4,i}} v_i^{\alpha_{4,i+1}}$ and $x_3^{t_4} = u_i^{\beta_i} v_i^{\beta_{i+1}}$. The curve $\Gamma_{3,4} \subset \mathbb{P}(t_2, w_3, t_4)$, restricted to the open set $(x_4 = 1)$, has equation $x_2^{a_2} + x_3^{(a_3-1)/h} = 0$, and we can use Proposition 3.8 to find the equation of the curve in every U_i .

Following the proof of Proposition 3.9, by the intersection number

$$\Gamma'_{3,4} \cdot \sum_{i=0}^{s_4+1} \frac{\beta_{4,i}}{t_4} E_{4,i} = \frac{a_2}{t_4},$$

and using the fact that the curve $\Gamma'_{3,4}$ intersects transversally one component, we have that there exist $\beta_{4,j} = a_2$ and $\alpha_{4,j} = (a_3 - 1)/h$. Therefore,

$$\begin{aligned} a_2 \alpha_{4,j-1} - \frac{a_3 - 1}{h} \beta_{4,j-1} &= -1, \\ a_2 \alpha_{4,j} - \frac{a_3 - 1}{h} \beta_{4,j} &= 0, \\ a_2 \alpha_{4,j+1} - \frac{a_3 - 1}{h} \beta_{4,j+1} &= 1. \end{aligned}$$

Hence, considering the composition

$$\tilde{X} \xrightarrow{\sigma} \frac{1}{t_4}(t_2, w_3) \xrightarrow{\simeq} \frac{1}{w_4}(w_2, w_3) \xrightarrow{-\psi} X(a_1, a_2, a_3, a_4)$$

we have the hypothesis of Theorem 3.11; therefore, the map is defined on the whole exceptional divisor. \square

Proposition 4.5. *The curves C'_1 and C'_2 in \tilde{X} are (-1) -curves. To obtain the chain of curves*

$$K_1 := E_{2,s_2} \cup \dots \cup E_{2,1} \cup C'_1 \cup E_{4,1} \cup \dots \cup E_{4,s_4}$$

and

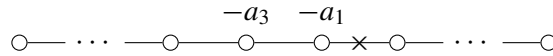
$$K_2 := E_{1,s_1} \cup \dots \cup E_{1,1} \cup C'_2 \cup E_{3,1} \cup \dots \cup E_{3,s_3}$$

we blowup \tilde{X}' on the intersection points of the curves with self-intersections $-a_3$ and $-a_1$, and $-a_2$ and $-a_4$, respectively.

Proof. We have the commutative diagram

$$\begin{array}{ccc} \tilde{X} & \xrightarrow{\sigma} & X(a_1, a_2, a_3, a_4) \\ \downarrow & & \downarrow \eta \\ \tilde{X}' & \xrightarrow{\sigma'} & X'(a_1, a_2, a_3, a_4) \end{array}$$

Then, to obtain the chain of curves K_1 we have to blowup on the exceptional divisor over the singularity $\frac{1}{s_1}(w_2, w_4)$. This is because, if no blowup were needed, then C'_1 would be some of the curves in the exceptional divisor over the singularity $\frac{1}{s_1}(w_2, w_4)$, so we would have that $w_2 \leq a_4 - 1$ or $w_4 \leq a_2 - 1$, which can happen only if one of the a_i is 1. Recall from Theorem 4.1 that the Hirzebruch–Jung continued fraction of the singularity $\frac{1}{s_1}(w_2, w_4)$ is $[2, \dots, 2, a_3, a_1, 2, \dots, 2]$. Then we want to show that the blowups needed must be done between the curves with self-intersection $-a_3$ and $-a_1$. For this, we will rule out every other possibility. Suppose first that the blowups are done on the point

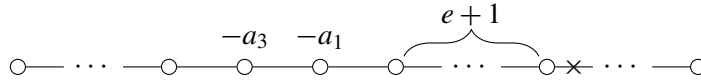


Then we would obtain that the continued fraction associated to the singularity at p_2 would have an β_i such that

$$\beta_i \geq |[2, \dots, 2, a_3, a_1 + 1]|,$$

$\underbrace{\hspace{10em}}_{a_4-1}$

but $|\overbrace{[2, \dots, 2]}^{a_4-1}, a_3, a_1 + 1| = w_2 + 2 + a_3a_4 - 2a_4 > w_2$, which is a contradiction. If the blowups are done on the point



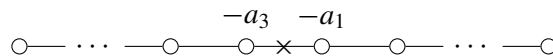
with $e \geq 0$, we would have

$$\beta_i \geq |[2, \dots, 2, a_3, a_1, \underbrace{2, \dots, 2}_e, 3]|,$$

$\underbrace{\hspace{10em}}_{a_4-1}$

but $|\overbrace{[2, \dots, 2]}^{a_4-1}, a_3, a_1, \overbrace{2, \dots, 2}^e, 3| = (2e + 3)w_2 - (2e + 1)a_3a_4 - 2a_4 + 1 > w_2$.

Therefore, the blowups to obtain the chain of curves K_1 desired have to be done at the point



□

From the proof of Proposition 4.5, we have that the singularity at p_i of the Kollár surface has Hirzebruch–Jung continued fraction

$$[\dots, c_i, \underbrace{2, \dots, 2}_{a_{i+2}-1}]$$

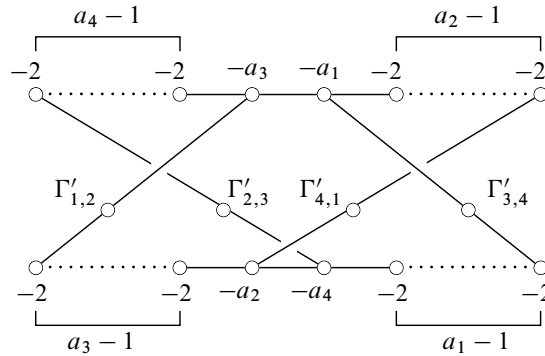


Figure 6. Curve configuration on \tilde{X}' .

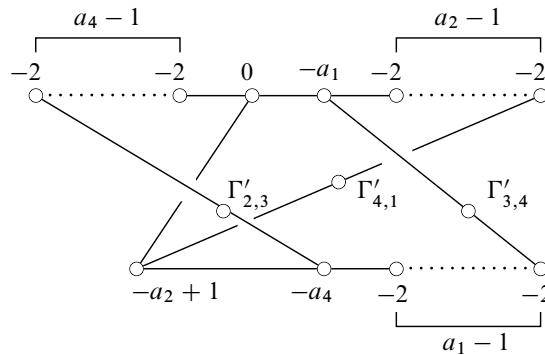


Figure 7. Contraction of $\Gamma'_{1,2}$ and the chain of (-2) -curves.

with $c_i > 2$. The intersection of $\Gamma'_{i-1,i}$ with the exceptional divisor over p_i is $\beta_{i,j}/w_i = a_{i+2}/w_i$, so the curve $\Gamma'_{i-1,i}$ intersects the exceptional divisor over p_i at the mentioned component with self-intersection $-c_i$. This is because $\beta_{i,s_i+1} = 0$ and $\beta_{i,s_i} = 1$, and $\beta_{i,k-1} = b_k \beta_{i,k} - \beta_{i,k+1}$. This implies that $\beta_{i,s_i-(a_2-1)} = a_2 = \beta_j$. Therefore, we have the curve configuration shown in Figure 6.

Proposition 4.6. *The curves $\Gamma'_{i,i+1}$ are (-1) -curves.*

Proof. We have a birational morphism $\psi \circ \sigma : \tilde{X} \rightarrow \mathbb{P}^2$, so it is a composition of blowups, which contracts (-1) -curves to reach \mathbb{P}^2 . We start by contracting the curves from the proof of Proposition 4.5 to obtain \tilde{X}' with the curve configuration in Figure 6. Recall from Theorem 3.11 that the image of the curves with self-intersection $-a_i$ are the four lines in general position in \mathbb{P}^2 , so they cannot be contracted. In addition, by Corollary 3.13 the birational morphism $\tilde{X}' \rightarrow \mathbb{P}^2$ is an isomorphism outside of the configuration in Figure 6. Then, one of the $\Gamma'_{i,i+1}$ is a (-1) -curve; say that it is $\Gamma'_{1,2}$. We contract $\Gamma'_{1,2}$ and the chain of (-2) -curves connected to it, to obtain the diagram in Figure 7.

By repeating the procedure, we obtain that all curves $\Gamma'_{i,i+1}$ are (-1) -curves. □

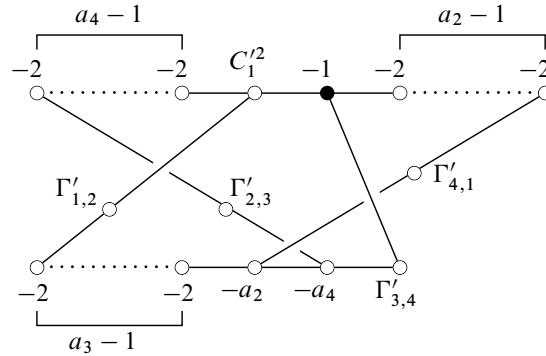


Figure 8. Curve configuration on \widehat{X}' .

Proof of Theorem 4.2. From Propositions 4.5 and 4.6, we conclude that \widetilde{X}' and $Z(a_1, a_2, a_3, a_4)$ are obtained from the same sequence of blowups of \mathbb{P}^2 . Therefore, $\widetilde{X}' \simeq Z(a_1, a_2, a_3, a_4)$ and so $X'(a_1, a_2, a_3, a_4) \simeq T(a_1, a_2, a_3, a_4)$. \square

Remark 4.7. We note that, if $w^* \neq 1$, then the surface $T(a_1, a_2, a_3, a_4)$ does not correspond to a Kollár surface, so Kollár surfaces with $w^* = 1$ and $a_i \geq 2$ are strictly contained in Hwang–Keum surfaces.

Finally, we check what happens when some $a_i = 1$, say $a_1 = 1$.

Corollary 4.8. *Let $a_1 = 1$. Then the point p_4 is smooth, and the map ψ is defined in the log resolution \widehat{X} of the key curves. The curve $\Gamma_{3,4}$ is smooth, and ψ does not contract C_1 . The surface \widehat{X} is obtained by doing blowups from $Z(1, a_2, a_3, a_4)$. The curve $C_1 \subset X(1, a_2, a_3, a_4)$ is contractible if and only if $a_3 > a_2$.*

Proof. If $a_1 = 1$, then $w_2 = a_4(a_3 - 1)$ and $w_4 = a_3 - 1$. Then by Proposition 3.1 we have that the point p_4 is smooth, and at the point p_2 the singularity is of type $\frac{1}{a_4}(1, a_2a_3a_4 - a_3a_4 + a_4 - 1) = \frac{1}{a_4}(1, a_4 - 1)$. The curve $\Gamma_{1,2}$ intersects transversally the curve C_1 at the point $(0 : -1 : 0 : 1)$, and following the proof of Proposition 3.4 we have that $\psi(0 : 1 : 0 : b) = (b : -1 - b : 0 : 1)$, so the curve ψ does not contract C_1 . The curve $\Gamma_{3,4}$ restricted to the weighted projective plane $(x_1 = 0)$ and to the open set $(x_4 \neq 1)$ is $(x_2^{a_2} + x_3 = 0) \subset \mathbb{A}^2$, so it is smooth and to obtain the log resolution \widehat{X} it is necessary to do a_2 blowups.

Now assume that all the other $a_i \geq 2$. Therefore, C_2 is contractible, and by contracting it and all the other (-1) -curves in \widehat{X} we obtain the surface \widehat{X}' with the curve configuration shown in Figure 8. If also $a_2 = 1$, then all the points are smooth but point p_2 with a singularity of type $\frac{1}{a_4}(1, a_4 - 1)$, and we obtain the curve configuration on \widehat{X} shown in Figure 9.

Following the proof of Proposition 4.6 we have that the curves $\Gamma'_{i,i+1}$ are (-1) -curves, $C_1^{l_2} = -a_3$, and $C_2^{l_2} = -a_4$. Therefore, $\widehat{X}' \simeq Z(1, a_2, a_3, a_4)$, and by contracting the (-1) -curve in the top chain along with the (-2) -curves to the right, we obtain that $C_1^{l_2} = -a_3 + a_2$. Therefore, C_2 is contractible if and only if $C_1^{l_2} < 0$, and this is equivalent to $a_3 > a_2$. \square

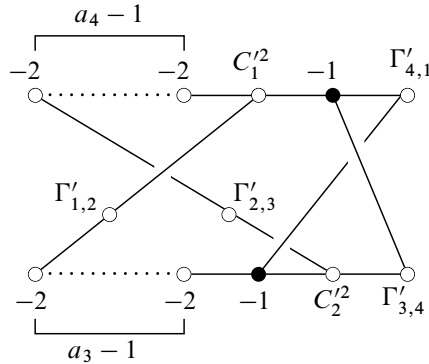


Figure 9. Curve configuration on X'_n when $a_2 = 1$.

5. Kollár surfaces as branched covers of \mathbb{P}^2

We now consider the birational model $Y' := \text{Spec}_{\mathbb{P}^2}(\bigoplus_{i=0}^{w^*-1} \mathcal{L}^{(i)-1})$ of $X(a_1, a_2, a_3, a_4)$, which was defined at the end of Section 2 as the w^* -th root cover of $(L_1^{\mu_1} L_2^{\mu_2} L_3^{\mu_3} L_4^{\mu_4} = 0) \subset \mathbb{P}^2$. We recall that $0 < \mu_i < w^*$ are

$$\mu_1 \equiv a_2 a_3 a_4, \quad \mu_2 \equiv -a_3 a_4, \quad \mu_3 \equiv a_4, \quad \mu_4 \equiv -1$$

modulo w^* , and that by definition $\gcd(\mu_i, w^*) = 1$. The lines L_1, L_2, L_3, L_4 form a plane curve with six nodes. We also recall that

$$\mathcal{L}^{(i)} := \mathcal{O}_{\mathbb{P}^2}(w_1 i) \otimes \mathcal{O}_{\mathbb{P}^2}\left(-\sum_{j=1}^4 \left\lfloor \frac{\mu_j i}{w^*} \right\rfloor L_j\right)$$

for $i \in \{0, 1, \dots, w^* - 1\}$, where $[x]$ is the integer part of x , and $w_1 w^* = \sum_{i=1}^4 \mu_i$. Let Y be the minimal resolution of all singularities in Y' .

Theorem 5.1. *The Kollár surface $X(a_1, a_2, a_3, a_4)$ is birational to*

$$X(a'_1, a'_2, a'_3, a'_4) \subset \mathbb{P}(w'_1, w'_2, w'_3, w'_4)$$

with $\gcd(w'_1, w'_3) = \gcd(w'_2, w'_4) = 1$, for infinitely many 4-tuples (a'_1, a'_2, a'_3, a'_4) .

Proof. By Corollary 2.4, the surface $X(a_1, a_2, a_3, a_4)$ is birational to Y' , and so for any $t_i \in \mathbb{Z}_{\geq 0}$ we have that $X(a_1, a_2, a_3, a_4)$ is birational to

$$X(a_1 + t_1 w^*, a_2 + t_2 w^*, a_3 + t_3 w^*, a_4 + t_4 w^*),$$

as soon as $w^* = \gcd(W'_1, \dots, W'_4)$ for the corresponding W'_i . This is because, for a fixed w^* , the isomorphism type of Y' depends only on the multiplicities μ_i modulo w^* . In this way, we must find $t_i \in \mathbb{Z}_{\geq 0}$ such that $\gcd(w'_1, w'_3) = \gcd(w'_2, w'_4) = 1$, and $w^* = \gcd(W'_1, \dots, W'_4)$.

First, choose t_3 such that $\gcd(a_3 + t_3 w^*, 6(a_4 - 1)) = 1$, and let $a'_3 := a_3 + t_3 w^*$ and $W'_1 := a_2 a'_3 a_4 - a'_3 a_4 + a_4 - 1 = w'_1 w^*$. Next take t_2 such that $\gcd(w'_1 + t_2 a'_3 a_4, 6(a_4 - 1)) = 1$, and then

define $a'_2 := a_2 + t_2 w^*$. Now we will choose t_1 such that the final weights $(w''_1, w''_2, w''_3, w''_4)$ satisfy $\gcd(w''_1, w''_3) = \gcd(w''_2, w''_4) = 1$, and $w^* = \gcd(W''_1, \dots, W''_4)$.

Let $W''_2 := a'_3 a_4 a_1 - a_4 a_1 + a_1 - 1 = w'_2 w^*$, $W''_3 := a_4 a_1 a'_2 - a_1 a'_2 + a'_2 - 1 = w'_3 w^*$, and $W''_4 := a_1 a'_2 a'_3 - a'_2 a'_3 + a'_3 - 1 = w'_4 w^*$, and define

$$\begin{aligned} W''_1 &:= w'_1 w^* = w'_1 w^*, & W''_2 &:= w''_2 w^* = (w'_2 + t(a'_3 a_4 - a_4 + 1)) w^*, \\ W''_3 &:= w'_3 w^* = (w'_3 + t(a_4 a'_2 - a'_2)) w^*, & W''_4 &:= w'_4 w^* = (w'_4 + t a'_2 a'_3) w^*, \end{aligned}$$

where t will be found.

Let $w''_1 = \prod q_{1,j}^{\lambda_{1,j}}$ be its prime factorization. Then we have to find a solution t for $w'_4 + t a'_2 a'_3 \not\equiv 0 \pmod{q_{1,j}}$, $w'_3 + t a'_2 (a_4 - 1) \not\equiv 0 \pmod{q_{1,j}}$, and $t \not\equiv 0 \pmod{q_{1,j}}$, for all j . This t will exist because we have that $\gcd(a_4 - 1, w''_1) = 1$, and that all $p_{1,j}$ are greater than 3, by the previous choice of t_2 and t_3 .

By the Chinese remainder theorem, we know that the solutions are of the form $t_1 + r \cdot \prod q_{1,j}$, $r \in \mathbb{Z}$. Hence, we have that $\gcd(w''_1, w''_3) = \gcd(w''_1, w''_4) = 1$, for any choice of r . Therefore, considering

$$w''_2 = w'_2 + t_1(a'_3 a_4 - a_4 + 1) + r \cdot (a'_3 a_4 - a_4 + 1) \cdot \prod q_{1,j}$$

and $w''_4 = w'_4 + t_1 a'_2 a'_3 + r \cdot a'_2 a'_3 \cdot \prod q_{1,j}$, it is enough to find an $r \in \mathbb{Z}_{\geq 0}$ such that $\gcd(w''_2, w''_4) = 1$. Let

$$\begin{aligned} A &:= w'_2 + t_1(a'_3 a_4 - a_4 + 1), & B &:= (a'_3 a_4 - a_4 + 1) \cdot \prod q_{1,j}, \\ C &:= w'_4 + t_1 a'_2 a'_3, & D &:= a'_2 a'_3 \cdot \prod q_{1,j}. \end{aligned}$$

Notice that $\gcd(A, B) = 1$ by the definition of w'_2 and the way t_1 was obtained. Let $AD - BC = q_{2,1}^{\lambda_{2,1}} q_{2,2}^{\lambda_{2,2}} \cdots q_{2,l}^{\lambda_{2,l}}$ with $q_{2,j}$ a prime number, and let r_1 be a solution of

$$A + Br \not\equiv 0 \pmod{q_{2,j}}. \tag{5-1}$$

Now assume that $\gcd(w''_2, w''_4) = \gcd(A + Br_1, C + Dr_1) > 1$. This means that there is a prime $p \neq q_{2,j}$ for all j , such that it divides both $A + Br$ and $C + Dr$. Then consider the linear transformation $T : (\mathbb{Z}/p\mathbb{Z})^2 \rightarrow (\mathbb{Z}/p\mathbb{Z})^2$ associated to the matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$. This matrix maps the vector $(1, r_1)$ to $(0, 0)$, so the matrix is singular. But the determinant $AD - BC \not\equiv 0 \pmod{p}$, which is a contradiction. Therefore, $\gcd(A + Br_1, C + Dr_1) = 1$. Let $a'_1 := a_1 + (t_1 + r_1 \cdot \prod p_{1,j}) w^*$. This gives us that $X(a'_1, a'_2, a'_3, a_4) \subset \mathbb{P}(w''_1, w''_2, w''_3, w''_4)$ is birational to $X(a_1, a_2, a_3, a_4)$, with $\gcd(w''_1, w''_3) = \gcd(w''_2, w''_4)$, and because $\gcd(w''_1, w''_4) = 1$, then $w^* = \gcd(W''_1, \dots, W''_4)$. Because (5-1) has infinite solutions, then we have infinite 4-tuples $(a''_1, a''_2, a''_3, a''_4)$ that satisfy the result. \square

Corollary 5.2. *Let Y' be an n -th root cover of $(L_1^{\mu_1} L_2^{\mu_2} L_3^{\mu_3} L_4^{\mu_4} = 0) \subset \mathbb{P}^2$, with $\gcd(\mu_i, n) = 1$ for all i . Then Y' is birational to a Kollár surface.*

Proof. If we multiply the μ_i by a unit ξ of $\mathbb{Z}/n\mathbb{Z}$, then the n -th root cover does not change. So we take ξ such that $\xi \mu_4 = -1$. In this way, we have to solve the system $a_2 a_3 a_4 \equiv \xi \mu_1$, $-a_3 a_4 \equiv \xi \mu_2$, $a_4 \equiv \xi \mu_3$, and $a_1 a_2 a_3 a_4 \equiv 1$ modulo n , which has a solution because ξ and the μ_i are units in $\mathbb{Z}/n\mathbb{Z}$. Then, with

those a_i we can use Theorem 5.1 to find numbers a'_i such that $X(a'_1, a'_2, a'_3, a'_4)$ is a Kollár surface with $w^* = n$, and birational to Y' . □

We want to compute the main numerical invariants of Y . For that we first define the following numbers.

Definition 5.3. Let $n > 1$ be an integer, and let a, b be integers coprime to n .

(1) We define the generalized Dedekind sum [Hirzebruch and Zagier 1974, p. 94] as

$$s(a, b; n) = \sum_{i=1}^{n-1} \left(\left(\frac{ia}{n} \right) \right) \left(\left(\frac{ib}{n} \right) \right)$$

where $((x)) = x - [x] - \frac{1}{2}$ for any rational number x .

(2) Let $0 < q < n$ be such that $aq \equiv b$ modulo n . We define the HJ length $l = l(a, b; n)$ as the length of the Hirzebruch–Jung continued fraction

$$\frac{n}{q} = [b_1, \dots, b_l].$$

Dedekind sums and Hirzebruch–Jung continued fractions relate as (see, e.g., [Barkan 1977; Urzúa 2010, Example 3.5])

$$12s(a, b; n) = \frac{q + q^{-1}}{n} + \sum_{i=1}^{l(a,b;n)} (b_i - 3),$$

where $0 < q^{-1} < n$ and $qq^{-1} \equiv 1$ modulo n . We recall that Y is the minimal resolution of Y' .

Proposition 5.4. We have that $\pi_1(Y) = 0$, and

$$p_g(Y) = 2s(1, 1; w^*) + \sum_{i < j} s(\mu_i, \mu_j; w^*)$$

where $s(1, 1; w^*) = w^*/12 + 1/(6w^*) - \frac{1}{4}$.

Proof. The finite morphism $Y' \rightarrow \mathbb{P}^2$ is completely ramified at four lines. By pulling back to Y a trivial pencil through one point in one of these lines, one can compute $\pi_1(Y) = 0$; for details see the proof of [Urzúa 2010, Theorem 8.5]. This also shows that $\chi(\mathbb{C}_Y) = 1 + p_g(Y)$. Then we use [Urzúa 2010, Proposition 3.2] to find the formula for $p_g(Y)$. The term $2s(1, 1; w^*)$ turns out to be exactly the expression not involving Dedekind sums in [Urzúa 2010, Proposition 3.2]. □

Remark 5.5. Since the geometric genus $p_g(Y)$ is a nonnegative number, we have $2s(1, 1; w^*) + \sum_{i < j} s(\mu_i, \mu_j; w^*) \geq 0$, which can be rewritten using basic properties of Dedekind sums as

$$p_g(Y) = 2s(1, 1; w^*) - \sum_{i=1}^4 s(1, a_i; w^*) + s(1, a_1a_4; w^*) + s(1, a_1a_2; w^*) \geq 0.$$

We will tell more on this expression in the next section.

Proposition 5.6. *We have that $e(Y) = w^* + 2 + \sum_{i < j} l(\mu_i, \mu_j; w^*)$, and*

$$K_Y^2 = w^* + \frac{4}{w^*} + 4 + \sum_{i < j} (12s(\mu_i, \mu_j; w^*) - l(\mu_i, \mu_j; w^*)).$$

Proof. See [Urzúa 2010, Proposition 3.6], and use Noether’s formula. □

Corollary 5.7. *For $X = X(a_1, a_2, a_3, a_4)$ we have $e(X) = w^* + 4$, $\pi_1(X) = 0$, and $p_g(X) = 2s(1, 1; w^*) - \sum_{i=1}^4 s(1, a_i; w^*) + s(1, a_1a_4; w^*) + s(1, a_1a_2; w^*)$.*

Corollary 5.8. *Let $\gcd(w_i, w_{i+2}) = 1$ for all i . Then*

$$12 \left(\sum_{i < j} s(\mu_i, \mu_j; w^*) + \sum_i s(w_{i+2}, w_{i+3}; w_i) \right) = \frac{d(d - \sum_i w_i)^2}{\prod_i w_i} - \sum_i \frac{2}{w_i} - \frac{w^{*2} - 6w^* + 4}{w^*}.$$

Proof. Let $X = X(a_1, a_2, a_3, a_4)$. We are going to compute $p_g(X)$ from its minimal resolution, and then the equality follows from $p_g(X) = p_g(Y)$. Let $\tilde{X} \rightarrow X$ be the minimal resolution of singularities, so $p_g(\tilde{X}) = p_g(X)$. As in the proof of Proposition 3.4 in [Urzúa 2010] and the formula right before Proposition 5.4, we have

$$K_{\tilde{X}}^2 - K_X^2 = -12 \sum_i s(w_{i+2}, w_{i+3}; w_i) - \sum_i l(w_{i+2}, w_{i+3}; w_i) + \sum_i \frac{2(w_i - 1)}{w_i},$$

and $e(\tilde{X}) - e(X) = \sum_i l(w_{i+2}, w_{i+3}; w_i)$. Since $K_X^2 = d(d - \sum_i w_i)^2 / \prod_i w_i$ and $e(X) = w^* + 4$, then

$$p_g(\tilde{X}) = \frac{d(d - \sum_{i=1}^4 w_i)^2}{12w_1w_2w_3w_4} - \sum_i s(w_{i+2}, w_{i+3}; w_i) - \frac{1}{6} \sum_i \frac{1}{w_i} + \frac{w^*}{12}$$

is a consequence of the Noether’s equality $12\chi(\mathcal{O}_{\tilde{X}}) = K_X^2 + e(\tilde{X})$. □

6. Theorems on geometric genus

In this section we prove results related to the geometric genus of Kollár surfaces. All our computations will be done in terms of generalized Dedekind sums. We note that the (classical) Dedekind sum $s(q, n)$ is equal to $s(1, q; n)$ and $s(a, b; n) = s(1, a^{-1}b; n)$, and so all properties of $s(q, n)$ are properties of $s(a, b; n)$ [Hirzebruch and Zagier 1974, Chapter II]. For example, we have the reciprocity law:

Theorem 6.1 (see, e.g., [Hirzebruch and Zagier 1974, p. 93]). *If n and k are relatively prime, then*

$$s(1, k; n) + s(1, n; k) = \frac{1}{12} \left(\frac{n}{k} + \frac{1}{nk} + \frac{k}{n} \right) - \frac{1}{4}. \tag{6-1}$$

Throughout this section, w^* will be greater than 1. All equalities involving \equiv will be modulo w^* , unless otherwise stated. The symbol q^{-1} will denote the inverse of q modulo w^* . To avoid confusion, we will write $\frac{1}{q}$ when it corresponds to a number in \mathbb{Q} .

Proposition 6.2. *Any $n \geq 0$ is realizable as the geometric genus of a Kollár surface.*

Proof. We know that $w^* = 1$ implies rationality, and so $p_g = 0$. Assume that $n > 0$, and let $w^* = 3n + 1$ and $a_1 \equiv 3^{-1}$, $a_2 \equiv 3$, and $a_3 \equiv a_4 \equiv w^* - 1$. This gives the w^* -th root cover Y with $\mu_1 = 3$ and $\mu_2 = \mu_3 = \mu_4 = w^* - 1$. The geometric genus of Y is

$$\begin{aligned} p_g(Y) &= 5s(1, 1; w^*) - 3s(1, 3; w^*) \\ &= 5\left(\frac{w^*}{12} + \frac{1}{6w^*} - \frac{1}{4}\right) - 3\left(\frac{w^*}{36} + \frac{1}{4w^*} + \frac{1}{36w^*} - \frac{1}{18} - \frac{1}{4}\right) \\ &= n. \end{aligned} \quad \square$$

6.1. $p_g = 0$ surfaces are rational.

Theorem 6.3. *Let $X = X(a_1, a_2, a_3, a_4)$ be a Kollár surface with $w^* > 1$. Then the following are equivalent:*

- (a) $p_g(X) = 0$.
- (b) $a_i \equiv 1$ or $a_i a_{i+1} \equiv -1$ modulo w^* for some i .
- (c) X is rational.

Lemma 6.4. *Let $0 < a < n$ be relatively prime. Then:*

- (1) $s(1, 1; n) > 2s(1, a; n)$ if $a \not\equiv 1$.
- (2) $s(1, 1; n) > 3s(1, a; n)$ if $a \not\equiv 1, 2, 2^{-1}$.
- (3) $s(1, 1; n) > 4s(1, a; n)$ if $a \not\equiv 1, 2, 2^{-1}, 3, 3^{-1}$.

Proof. First of all, using the reciprocity law we have

$$\begin{aligned} 2s(1, 2; n) &= \frac{n^2 - 6n + 5}{12n} < s(1, 1; n), \\ 3s(1, 3; n) &\leq \frac{n^2 - 7n + 10}{12n} < s(1, 1; n), \\ 4s(1, 4; n) &\leq \frac{n^2 - 6n + 17}{12n} < s(1, 1; n) \end{aligned}$$

with $\gcd(n, 2) = 1$, $\gcd(n, 3) = 1$, and $\gcd(n, 4) = 1$, respectively, and $n \geq 6$. Notice that $s(1, 1; n) = (n - 1)(n - 2)/12n$. Girstmair [2017, Theorem 1] describes how Dedekind sums $s(1, m; n)$ grow for a fixed m , given a positive integer k . To do so, Girstmair divides the numbers $1 \leq m \leq n - 1$ as ordinary and not ordinary, and proves that, if m is ordinary, then $s(1, m; n) \leq n/(12(k + 1)) + O(1)$ and, if m is not ordinary, then there exist $d \in \{1, \dots, 2k + 1\}$ and $c \in \{0, 1, \dots, d\}$, $\gcd(c, d) = 1$, such that $s(1, m; n) = n/(12dq) + O(1)$, where $q = md - nc$.

First assume that $k = 2$. Notice that $s(1, 1; n)/2 = n/24 + O(1)$; also if m is ordinary, then $s(1, m; n) \leq n/36 + O(1)$, and if m is not ordinary and $dq \geq 3$, then $s(1, m; n) \leq n/36 + O(1)$. Therefore, we have to find a bound for the three $O(1)$ involved, and find an N such that, if $n > N$, then $s(1, 1; n)/2 > s(m, n)$ for ordinary numbers and nonordinary numbers with $qd \geq 3$. The procedure to do so is shown by

Girstmair [2017, Theorem 2], and for the case $k = 2$ such N is 132. The nonordinary numbers with $qd \leq 2$ correspond to $m \equiv 1, 2, 2^{-1}$, but the first case was ruled out in the proposition, and the inequality for 2 and 2^{-1} was shown at the beginning of the proof. Therefore, we have (1) for $n > 132$, and using a computer we can check that it holds true for every n .

For $k = 3$ and $k = 4$ we obtain similar results, with $N = 320$ and $N = 630$, respectively. The cases with $qd \leq 3$ and $qd \leq 4$ are the ones ruled out in the proposition, and using a computer we can check that (2) and (3) are true for $n \leq 320$ and $n \leq 630$. □

Corollary 6.5. (1) $2s(1, 1; n) - 2s(1, 2; n) + s(1, 4; n) - s(1, 3; n) + s(1, 2 \cdot 3^{-1}; n) - s(1, 4 \cdot 3^{-1}; n) > 0$
for all $n > 5$.

(2) $2s(1, 1; n) - s(1, 2; n) - s(1, 3; n) - s(1, 4; n) + s(1, 6; n) - s(1, 2 \cdot 3^{-1}; n) + s(1, 4 \cdot 3^{-1}; n) > 0$
for all $n > 7$.

(3) $2s(1, 1; n) - s(1, 2; n) - s(1, 3; n) - s(1, 5; n) + s(1, 6; n) + s(1, 2 \cdot 5^{-1}; n) - s(1, 6 \cdot 5^{-1}; n) > 0$
for all $n > 7$.

Proof. Using the inequalities from Lemma 6.4 we see that to prove (1) it is enough to prove that $\frac{2}{3}s(1, 1; n) + s(1, 4; n) + s(1, 2 \cdot 3^{-1}; n) - s(1, 4 \cdot 3^{-1}; n) > 0$. On the other hand, we have that $s(1, 4; n) > 0$ if $n \notin \{7, 13, 19, 25, 31\}$, $s(1, -2 \cdot 3^{-1}; n) < s(1, 1; n)/3$ if $n \notin \{5, 7\}$, and $s(1, 4 \cdot 3^{-1}; n) < s(1, 1; n)/3$ if $n \neq 5$. Therefore, if n is not one of those cases, then the inequality holds. We check the remaining cases and find that (1) is false only if $n = 5$. We repeat the same argument and prove that we have to check the cases when $n \in \{7, 11, 13, 19, 25, 31\}$ for (2), and when $n \in \{7, 13, 19, 31\}$ for (3). Both cases give us that (2) or (3) are false only if $n = 7$. □

Proof of Theorem 6.3. By Corollary 5.7, we have that the geometric genus of $X(a_1, a_2, a_3, a_4)$ is

$$p_g(X) = 2s(1, 1; w^*) - \sum_{i=1}^4 s(1, a_i; w^*) + s(1, a_1 a_4; w^*) + s(1, a_1 a_2; w^*).$$

(c) \implies (a). This is trivial.

(a) \implies (b). Assume that $a_i \not\equiv 1$ and $a_i a_{i+1} \not\equiv -1$ for all i . First, if $a_i \not\equiv 2, 2^{-1}$ and $a_i a_{i+1} \not\equiv -2, -2^{-1}$ for all i , then by Lemma 6.4(2) we have that $p_g > 2s(1, 1; w^*) - \frac{6}{3}s(1, 1; w^*) > 0$. Therefore, it is enough to rule out the cases when $a_1 \equiv 2$ or $a_1 a_2 \equiv -2^{-1}$. First suppose that $a_1 \equiv 2$, so

$$p_g = 2s(1, 1; w^*) + s(1, 2a_2; w^*) + s(1, 2a_4; w^*) - s(1, 2; w^*) - \sum_{i=2}^4 s(1, a_i; w^*),$$

and we have to check the cases when we cannot use Lemma 6.4(3).

If $a_3 \equiv 2$ or $a_3 \equiv 2^{-1}$, then $a_1 a_2 \equiv -1$ or $a_4 \equiv 1$, respectively, so they satisfy the hypothesis for $p_g = 0$.

If $a_2 \equiv 2^{-1}$, $2a_2 \equiv -2$, $2a_4 \equiv -2$, $a_4 \equiv 3^{-1}$, or $2a_2 \equiv -3$, then one of the terms is equal to $s(1, 1; w^*)$ or two of the terms cancel, so by Lemma 6.4(1) we have that $p_g > 0$.

If $a_2 \equiv 2$, $2a_2 \equiv -2^{-1}$, or $2a_4 \equiv -2^{-1}$, then

$$p_g = 2s(1, 1; w^*) - 2s(1, 2; w^*) + s(1, 4; w^*) - s(1, 3; w^*) + s(1, 2 \cdot 3^{-1}; w^*) - s(1, 4 \cdot 3^{-1}; w^*)$$

and by Corollary 6.5(1) $p_g > 0$ when $w^* > 5$. If $w^* = 5$, then it satisfies the conditions for $p_g = 0$.

If $a_2 \equiv 3$ or $2a_4 \equiv -3^{-1}$, then

$$p_g = 2s(1, 1; w^*) - s(1, 2; w^*) - s(1, 3; w^*) - s(1, 4; w^*) + s(1, 6; w^*) - s(1, 2 \cdot 3^{-1}; w^*) + s(1, 4 \cdot 3^{-1}; w^*)$$

and by Corollary 6.5(2) $p_g > 0$ when $w^* > 7$. If $w^* = 7$, then it satisfies the conditions for $p_g = 0$.

If $a_4 \equiv 3$ or $2a_2 \equiv -3^{-1}$, then

$$p_g = 2s(1, 1; w^*) - s(1, 2; w^*) - s(1, 3; w^*) - s(1, 5; w^*) + s(1, 6; w^*) + s(1, 2 \cdot 5^{-1}; w^*) - s(1, 6 \cdot 5^{-1}; w^*)$$

and by Corollary 6.5(3) $p_g > 0$ when $w^* > 7$. If $w^* = 7$, then it satisfies the conditions for $p_g = 0$.

These cover all the cases for $a_1 \equiv 2$. Now assume that $a_1 a_2 \equiv -2^{-1}$, so

$$p_g = 2s(1, 1; w^*) - s(1, 2; w^*) + s(1, a_1 a_4; w^*) + s(1, 2a_2; w^*) - \sum_{i=2}^4 s(1, a_i; w^*),$$

and we proceed as in the previous case.

If $a_1 a_4 \equiv -2$ or $a_1 a_4 \equiv -2^{-1}$, then $a_1 \equiv 1$ or $a_4 \equiv 1$, respectively, so they satisfy the hypothesis for $p_g = 0$.

If $a_2 \equiv 3^{-1}$ or $a_3 \equiv 3$, then two of the terms in the sum cancel, so by Lemma 6.4(1) we have that $p_g > 0$.

If $a_4 \equiv 3^{-1}$ or $2a_2 \equiv -3^{-1}$, then

$$p_g = 2s(1, 1; w^*) - s(1, 2; w^*) - s(1, 3; w^*) - s(1, 4; w^*) + s(1, 6; w^*) - s(1, 2 \cdot 3^{-1}; w^*) + s(1, 4 \cdot 3^{-1}; w^*)$$

and by Corollary 6.5(2) $p_g > 0$ when $w^* > 7$. If $w^* = 7$, then it satisfies the conditions for $p_g = 0$.

If $a_2 \equiv 3$ or $a_3 \equiv 3^{-1}$, then

$$p_g = 2s(1, 1; w^*) - s(1, 2; w^*) - s(1, 3; w^*) - s(1, 5; w^*) + s(1, 6; w^*) + s(1, 2 \cdot 5^{-1}; w^*) - s(1, 6 \cdot 5^{-1}; w^*)$$

and by Corollary 6.5(3) $p_g > 0$ when $w^* > 7$. If $w^* = 7$, then it satisfies the conditions for $p_g = 0$.

These cover all the cases for $a_1 a_2 \equiv -2^{-1}$.

(b) \implies (c). Notice that (b) implies the existence of μ_i and μ_j such that $\mu_i + \mu_j \equiv 0 \pmod{w^*}$. Consider the trivial pencil of lines through $L_i \cap L_j$. Since $\mu_i + \mu_j \equiv 0 \pmod{w^*}$, this pencil defines a pencil of smooth rational curves in Y via pull-back. Therefore, Y is rational, and so is X . \square

6.2. $p_g = 1$ surfaces are K3. In Table 1, we show the total transform of the key configuration of curves after successively blowing down several (-1) -curves from the minimal resolution of the indicated surfaces $X(a_1, a_2, a_3, a_4)$.

$X(a_1, a_2, a_3, a_4)$	w^*	total transform of key configuration
$X(7, 7, 15, 15)$	4	
$X(8, 9, 14, 22)$	5	
$X(11, 27, 10, 18)$	7	
$X(17, 14, 42, 18)$	11	
$X(20, 21, 43, 22)$	13	
$X(26, 56, 39, 64)$	17	
$X(29, 30, 42, 32)$	19	
$X(47, 51, 63, 91)$	20	

Table 1. List for $p_g = 1$.

Theorem 6.6. *Let $X = X(a_1, a_2, a_3, a_4)$ be a Kollár surface with $w^* > 1$. Then the following are equivalent:*

- (a) $p_g(X) = 1$.
- (b) X is birational to one of the eight surfaces in Table 1.
- (c) X is birational to a K3 surface.

Proof. (c) \implies (a). It is trivial.

(a) \implies (b). First we prove:

Lemma 6.7. *Let m be a positive integer. Then there is a positive integer N such that, if $w^* > N$ and $p_g \neq 0$, then $p_g > m$.*

Proof. If all a_i and $-a_1a_2$ and $-a_1a_4$ are not equivalent to $2, 2^{-1}, 3, 3^{-1}$, then by Lemma 6.4(3)

$$p_g > 2s(1, 1; w^*) - \frac{6}{4}s(1, 1; w^*) = \frac{1}{2}s(1, 1; w^*).$$

Also we note that, if we fix two of these values, say for example $a_1 \equiv 2$ and $a_1a_2 \equiv -3$, then the rest of the a_i are completely determined, and they are equivalent to $2, 2^{-1}, 3, 3^{-1}$ only for finitely many w^* . Therefore, if we set two of the $a_i, -a_1a_2$, or $-a_1a_4$ to be equivalent to 3 or 3^{-1} , then for $w^* \gg 0$

$$p_g > 2s(1, 1; w^*) - \frac{2}{3}s(1, 1; w^*) - s(1, 1; w^*) = \frac{1}{3}s(1, 1; w^*).$$

If one of the values is 2 or 2^{-1} and the other is 3 or 3^{-1} , then for $w^* \gg 0$

$$p_g > 2s(1, 1; w^*) - \frac{1}{2}s(1, 1; w^*) - \frac{1}{3}s(1, 1; w^*) - s(1, 1; w^*) = \frac{1}{6}s(1, 1; w^*).$$

Both of these cases happen when $w^* > 28$; hence, we have to check the case when two of the values are 2 or 2^{-1} . This was done in the proof of Theorem 6.3, and the only relevant case is when p_g is $2s(1, 1; w^*) - 2s(1, 2; w^*) + s(1, 4; w^*) - s(1, 3; w^*) + s(1, 2 \cdot 3^{-1}; w^*) - s(1, 4 \cdot 3^{-1}; w^*)$. For $w^* \gg 0$

$$p_g > 2s(1, 1; w^*) - s(1, 1; w^*) - \frac{1}{3}s(1, 1; w^*) - \frac{1}{2}s(1, 1; w^*) + s(1, 4; w^*),$$

and because $s(1, 4; w^*) \geq 0$ for $w^* \geq 15$, we have that $p_g > s(1, 1; w^*)/6$.

Therefore, N is the first integer such that $s(1, 1; N) > 6m$. □

To prove this implication, we first use Lemma 6.7 for $m = 1$, which gives us that $N = 75$. We check using a computer all the possible w^* -th root covers for $w^* \leq 75$, and find that there are eight cases with $p_g = 1$, which are represented by a Kollár surface in Table 1.

(b) \implies (c). We prove this implication by means of the following simple lemma:

Lemma 6.8. *Let S be a smooth projective surface with $p_g = 1$ and $q = 0$. Assume it has an effective connected divisor F with $F^2 = 0$ and $p_a(F) = 1$, and a (-2) -curve C such that $F \cdot C = 1$. Then S is birational to a K3 surface, and F is a fiber of an elliptic fibration $S \rightarrow \mathbb{P}^1$, where C is a section.*

Proof. Notice that F has the type of a nonmultiple fiber of an elliptic fibration. We want to get such a fibration on S . By the Riemann–Roch inequality and $F \cdot (F - K_S) = 0$, we have $h^0(F) + h^2(F) \geq \chi(\mathcal{O}_S) = 2$. Since in addition $h^2(F) = h^0(K_S - F)$ and $C \cdot (K_S - F) = -1$, we have $h^2(F) = 0$. Therefore, there is a fibration $S \rightarrow \mathbb{P}^1$ with general fiber of genus 1 and F is a fiber. Let S' be the relative minimal model of this fibration. By the canonical class formula, $K_S \sim (-2 + \chi(\mathcal{O}_S))F + \sum_i (m_i - 1)G_i + E$ where G_i are the multiple fibers, and E is the exceptional divisor from $S \rightarrow S'$. But there is a section C , and so $G_i = 0$ for all i . Then S' has trivial canonical class, and so it is a K3 surface. \square

We now go case by case, showing what the support $\text{supp}(F)$ of F is and its type (using Kodaira's notation), and showing C . Here we are choosing F and C ; there are other choices in general:

$$(4) \text{supp}(F) = \sum_{i=1}^6 F_i + L_1 + L_2 + L_4 + F_{16} + F_{17} + F_{18}, \text{ type I}_{12}, \text{ and } C = F_7.$$

$$(5) \text{supp}(F) = F_1 + F_{16} + F_{17} + L_4, \text{ type IV, and } C = F_2.$$

$$(7) \text{supp}(F) = F_1 + F_{16} + F_{17} + L_4, \text{ type III, and } C = F_{15}.$$

$$(11) \text{supp}(F) = F_6 + L_2 + F_{17} + F_7, \text{ type II, and } C = F_5.$$

$$(13) \text{supp}(F) = F_1 + F_2 + L_4 + L_3 + F_8 + \sum_{i=10}^{15} F_i, \text{ type III}^*, \text{ and } C = F_3.$$

$$(17) \text{supp}(F) = L_2 + \sum_{i=7}^9 F_i + F_{12} + L_3 + F_{13} + F_{16}, \text{ type IV, and } C = F_{11}.$$

$$(19) \text{supp}(F) = F_4 + L_1 + F_5 + F_6 + F_7 + L_2 + F_{15}, \text{ type II, and } C = F_3.$$

$$(20) \text{supp}(F) = F_3 + L_1 + F_4 + F_5 + F_6 + L_2 + F_{14}, \text{ type II, and } C = F_2. \quad \square$$

6.3. $p_g \geq 2$ generic surfaces are of general type. In this subsection, we assume that $p_g \geq 2$. We recall that Kollár surfaces are simply connected. By classification of algebraic surfaces, the Kodaira dimension of the associate surface Y is either 1 or 2. We first present families of explicit examples for each of the two possible Kodaira dimensions, and then we show the general picture for $w^* \gg 0$.

Let $g : Y' \rightarrow \mathbb{P}^2$ be the normal w^* -th root cover branched on

$$(L_1^{\mu_1} L_2^{\mu_2} L_3^{\mu_3} L_4^{\mu_4} = 0),$$

and let $f : Y \rightarrow \mathbb{P}^2$ be g composed with the minimal resolution of singularities of Y' . Let $p_{i,j} = L_i \cap L_j$ for $i < j$. Let $E_{i,j,k}$ be the k -th exceptional curve over $p_{i,j}$. Then

$$K_Y \equiv f^* \left(-3H + \frac{w^* - 1}{w^*} (L_1 + L_2 + L_3 + L_4) \right) - \sum_{i < j} \sum_k \left(1 - \frac{\alpha_{i,j,k} + \beta_{i,j,k}}{w^*} \right) E_{i,j,k}$$

where H is a line in \mathbb{P}^2 . We have

$$-3H + \frac{w^* - 1}{w^*} (L_1 + L_2 + L_3 + L_4) \equiv \frac{w^* - 4}{4w^*} (L_1 + L_2 + L_3 + L_4)$$

and

$$f^*(L_i + L_j) \equiv w^* L'_i + w^* L'_j + \sum_k (\alpha_{i,j,k} + \beta_{i,j,k}) E_{i,j,k}$$

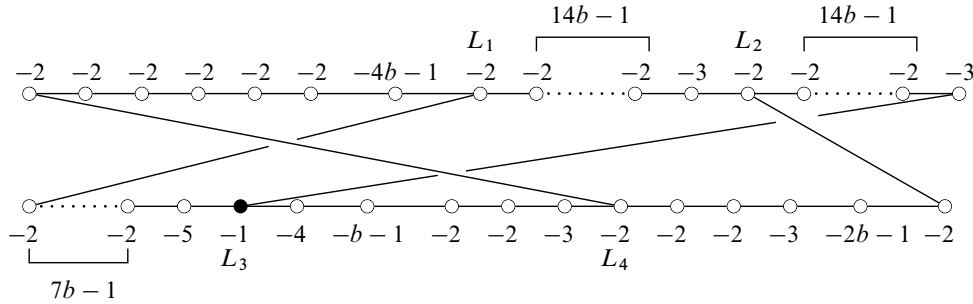


Figure 10. Curve configuration of a general type example.

for $i \neq j$, and so

$$K_Y \equiv \frac{w^* - 4}{4} (L'_1 + L'_2 + L'_3 + L'_4) + \sum_{i < j} \sum_k \left(\frac{\alpha_{i,j,k} + \beta_{i,j,k} - 4}{4} \right) E_{i,j,k},$$

where we are using notation and facts from the beginning of Section 3, and $L'_i \simeq \mathbb{P}^1$ is the (reduced, irreducible) preimage of L_i .

Example 6.9. Let $b \geq 2$. Consider $w^* = 4(b - 1)$, $\mu_1 = \mu_2 = 1$, and $\mu_3 = \mu_4 = 2b - 3$. Then, over $p_{1,2}$ and $p_{3,4}$ we have A_{w^*-1} singularities in Y' , and over the rest of the $p_{i,j}$ we have $\frac{1}{w^*}(1, 2b - 1)$. Notice that $w^*/(2b - 1) = [2, b, 2]$. We have that $L_i'^2 = -2$, and

$$K_Y \equiv \frac{b - 2}{2} \left(2 \sum_i L'_i + \sum_k 2(E_{1,2,k} + E_{3,4,k}) + (E_{1,3,k} + E_{1,4,k} + E_{2,3,k} + E_{2,4,k}) \right).$$

Therefore, Y is a minimal surface with $K_Y^2 = 0$ and $e(Y) = 3w^* + 12$, and so $p_g(Y) = b - 1$. The surface Y is K3 when $b = 2$, and Kodaira dimension 1 when $b > 2$. In fact, one can show that $E_{1,3,2}, E_{1,4,2}, E_{2,3,2}, E_{2,4,2}$ are sections (and $(-b)$ -curves) for an elliptic fibration $Y \rightarrow \mathbb{P}^1$, and the complement of them in the support above of K_Y give two $I_{w^*}^*$ singular fibers (using Kodaira notation).

Example 6.10. Let $b \geq 1$. Consider $w^* = 28b + 1$, $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 4$, and $\mu_4 = 28b - 6$. Then over $p_{i,j}$ we have:

- $(p_{1,2})$ $\frac{1}{w^*}(1, w^* - 2)$ and $[2, \dots, 2, 3]$ with $(14b - 1)$ 2s.
- $(p_{1,3})$ $\frac{1}{w^*}(1, 7b)$ and $[5, 2, \dots, 2]$ with $(7b - 1)$ 2s.
- $(p_{1,4})$ $\frac{1}{w^*}(1, 7)$ and $[4b + 1, 2, 2, 2, 2, 2, 2]$.
- $(p_{2,3})$ $\frac{1}{w^*}(1, w^* - 2)$ and $[2, \dots, 2, 3]$ with $(14b - 1)$ 2s.
- $(p_{2,4})$ $\frac{1}{w^*}(1, 14b + 4)$ and $[2, 2b + 1, 3, 2, 2]$.
- $(p_{3,4})$ $\frac{1}{w^*}(1, 7b + 2)$ and $[4, b + 1, 2, 2, 3]$.

One can also compute that $L_1'^2 = L_2'^2 = L_4'^2 = -2$ and $L_3'^2 = -1$. The configuration of all these curves is shown in Figure 10.

One can verify that $\alpha_{i,j,k} + \beta_{i,j,k} > 4$ for all i, j, k . Therefore, by the formula above, K_Y can be written with positive coefficients supported in the configuration of curves, so that to obtain the minimal model Y'' of Y we only need to contract L'_3 since $(w^* - 4)/4 > 1$ (and see the figure). We compute using the formulas above $K_{Y''}^2 = 7(3b - 1)$, $e(Y'') = 63b + 19$, and $p_g(Y'') = 7b$. In this way, Y'' is of general type for any b .

We now consider prime numbers $w^* \gg 0$ and partitions

$$\mu_1 + \mu_2 + \mu_3 + \mu_4 = w^*$$

with $0 < \mu_i < w^*$. Let \mathcal{S} be the set of all partitions. Then, as we did before, there are smooth projective surfaces Y constructed as w^* -th root covers $Y \rightarrow Y' \rightarrow \mathbb{P}^2$, and there are infinitely many Kollár surfaces $X(a_1, a_2, a_3, a_4)$ birational to each Y . Let X_{\min} be a minimal (smooth) model for Y (and so for all $X(a_1, a_2, a_3, a_4)$). The following is based on [Urzúa 2010; 2017]:

Theorem 6.11. *There is $\mathcal{S}' \subset \mathcal{S}$ with $|\mathcal{S}'|/w^* \rightarrow 0$ as $w^* \gg 0$ such that, if $\{\mu_1, \mu_2, \mu_3, \mu_4\} \in \mathcal{S} \setminus \mathcal{S}'$, then X_{\min} is a simply connected surface of general type with $K_{X_{\min}}^2/e(X_{\min}) \rightarrow 1$ as $w^* \gg 0$.*

Proof. By Proposition 5.6, we have $e(Y) = w^* + 2 + \sum_{i < j} l(\mu_i, \mu_j; w^*)$, and

$$K_Y^2 = w^* + \frac{4}{w^*} + 4 + \sum_{i < j} 12s(\mu_i, \mu_j; w^*) - l(\mu_i, \mu_j; w^*).$$

Notice that by Theorem 4.1 in [Urzúa 2017], both $e(Y) \gg 0$ and $K_Y^2 \gg 0$. In particular Y is of general type by classification of algebraic surfaces. We also note that $K_{Y'}$ is ample since it is numerically $(1 - 4/w^*)$ times the pull-back of the class of a line. Thus, by Theorem 4.3 in [Urzúa 2017], the number of potential (-1) -curves to be contracted over w^* tends to zero as w^* approaches infinity, and so X_{\min} satisfies $K_{X_{\min}}^2/e(X_{\min}) \rightarrow 1$ as $w^* \gg 0$. \square

Acknowledgements

We are grateful to DongSeon Hwang for numerous discussions about \mathbb{Q} HPPs, to Kurt Girstmair for useful discussions on Dedekind sums, and to the anonymous referees for their many helpful comments. This is part of the master's thesis of the second author at the Pontificia Universidad Católica de Chile. A large part of this paper was written while the authors were visiting the Department of Mathematics of the University of Massachusetts Amherst. We are thankful for the hospitality. The authors were supported by the FONDECYT regular grant 1150068.

References

- [Barkan 1977] P. Barkan, "Sur les sommes de Dedekind et les fractions continues finies", *C. R. Acad. Sci. Paris Sér. A-B* **284**:16 (1977), A923–A926. MR Zbl
- [Barth et al. 2004] W. P. Barth, K. Hulek, C. A. M. Peters, and A. Van de Ven, *Compact complex surfaces*, 2nd ed., *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)* **4**, Springer, 2004. MR Zbl

- [Dolgachev 1982] I. Dolgachev, “Weighted projective varieties”, pp. 34–71 in *Group actions and vector fields* (Vancouver, 1981), edited by J. B. Carrell, Lecture Notes in Mathematics **956**, Springer, 1982. MR Zbl
- [Esnault and Viehweg 1992] H. Esnault and E. Viehweg, *Lectures on vanishing theorems*, DMV Seminar **20**, Birkhäuser, 1992. MR Zbl
- [Girstmair 2017] K. Girstmair, “The largest values of Dedekind sums”, *Int. J. Number Theory* **13**:6 (2017), 1579–1583. MR Zbl
- [Gonzalez-Sprinberg and Lejeune-Jalabert 1997] G. Gonzalez-Sprinberg and M. Lejeune-Jalabert, “Families of smooth curves on surface singularities and wedges”, *Ann. Polon. Math.* **67**:2 (1997), 179–190. MR Zbl
- [Hirzebruch and Zagier 1974] F. Hirzebruch and D. Zagier, *The Atiyah–Singer theorem and elementary number theory*, Mathematics Lecture Series **3**, Publish or Perish, 1974. MR Zbl
- [Hwang and Keum 2012] D. Hwang and J. Keum, “Construction of singular rational surfaces of Picard number one with ample canonical divisor”, *Proc. Amer. Math. Soc.* **140**:6 (2012), 1865–1879. MR Zbl
- [Iano-Fletcher 2000] A. R. Iano-Fletcher, “Working with weighted complete intersections”, pp. 101–173 in *Explicit birational geometry of 3-folds*, edited by A. Corti and M. Reid, London Mathematical Society Lecture Note Series **281**, Cambridge University, 2000. MR Zbl
- [Kollár 2008] J. Kollár, “Is there a topological Bogomolov–Miyaoaka–Yau inequality?”, *Pure Appl. Math. Q.* **4**:2 (2008), 203–236. MR Zbl
- [Reid 2003] M. Reid, “Surface cyclic quotient singularities and Hirzebruch–Jung resolutions”, preprint, 2003, Available at <http://homepages.warwick.ac.uk/~masda/surf/more/cyclic.pdf>.
- [Urzúa 2010] G. Urzúa, “Arrangements of curves and algebraic surfaces”, *J. Algebraic Geom.* **19**:2 (2010), 335–365. MR Zbl
- [Urzúa 2017] G. Urzúa, “Chern slopes of surfaces of general type in positive characteristic”, *Duke Math. J.* **166**:5 (2017), 975–1004. MR Zbl

Communicated by Gavril Farkas

Received 2016-12-23 Revised 2018-01-29 Accepted 2018-03-17

urzua@mat.puc.cl

*Facultad de Matemáticas, Pontificia Universidad Católica de Chile,
Campus San Joaquín, Santiago, Chile*

yanez@math.utah.edu

*Department of Mathematics, University of Utah, Salt Lake City, UT,
United States*

Représentations de réduction unipotente pour $SO(2n+1)$ III: Exemples de fronts d'onde

Jean-Loup Waldspurger

Soit G un groupe $SO(2n+1)$ défini sur un corps p -adique. Nous calculons le front d'onde des représentations irréductibles anti-tempérées de $G(F)$ qui sont de réduction unipotente. Le front d'onde d'une telle représentation est l'orbite orthogonale duale à l'orbite symplectique qui intervient dans le paramètre d'Arthur de cette représentation.

Let G be a group $SO(2n+1)$ defined over a p -adic field. We compute the wave front set of the antitempered irreducible representations of $G(F)$ which are of unipotent reduction. The wave front set of such representations is the orthogonal orbit dual to the symplectic orbit appearing in the Arthur's parametrization of the representation.

Introduction	1107
1. Combinatoire	1110
2. Calcul de caractères	1138
3. Fronts d'onde	1150
Index des notations	1169
Index des notations de [Waldspurger 2018]	1169
Remerciement	1170
Bibliographie	1170

Introduction

Cet article est la suite de [Waldspurger 2018; 2016b]. Le corps de base F est local, non-archimédien et de caractéristique nulle. On note p sa caractéristique résiduelle. Un entier $n \geq 1$ est fixé pour tout l'article. On suppose $p > 6n + 4$. On introduit les groupes G_{iso} et G_{an} suivants. Le groupe G_{iso} est le groupe spécial orthogonal d'un espace V_{iso} de dimension $2n + 1$ sur F muni d'une forme quadratique Q_{iso} et G_{an} est le groupe spécial orthogonal d'un espace V_{an} de dimension $2n + 1$ sur F muni d'une forme quadratique Q_{an} . Le groupe G_{iso} est déployé et G_{an} en est la forme intérieure non déployée. Pour un indice $\sharp = \text{iso}$ ou an , on note $\text{Irr}_{\text{unip}, \sharp}$ l'ensemble des classes d'isomorphismes de représentations admissibles irréductibles de $G_{\sharp}(F)$ qui sont tempérées et de réduction unipotente, cf. [Waldspurger 2018, §1.3]

MSC2010: 22E50.

Mots-clefs: representation of unipotent reduction, dual orbit, wave front set, unipotent orbit.

pour la définition de cette propriété. On note $\text{Irr}_{\text{tunip}}$ la réunion disjointe de $\text{Irr}_{\text{tunip,iso}}$ et $\text{Irr}_{\text{tunip,an}}$. Pour une partition symplectique λ de $2n$, fixons un homomorphisme algébrique $\rho_\lambda : \text{SL}(2; \mathbb{C}) \rightarrow \text{Sp}(2n; \mathbb{C})$ paramétré par λ , cf. [Waldspurger 2018, §1.3]. On note $Z(\lambda)$ le commutant dans $\text{Sp}(2n; \mathbb{C})$ de l'image de ρ_λ . Soit $s \in Z(\lambda)$ un élément semi-simple dont toutes les valeurs propres sont de module 1. On note $Z(s, \lambda)$ le commutant de s dans $Z(\lambda)$, $\mathbf{Z}(\lambda, s)$ son groupe de composantes connexes et $\mathbf{Z}(\lambda, s)^\vee$ le groupe des caractères de $\mathbf{Z}(\lambda, s)$. La paramétrisation de Langlands prend la forme suivante, cf. [Waldspurger 2018, §1.3] : $\text{Irr}_{\text{tunip}}$ est paramétré par l'ensemble des classes de conjugaison (en un sens facile à préciser) de triplets (λ, s, ϵ) , où λ et s sont comme ci-dessus et $\epsilon \in \mathbf{Z}(\lambda, s)^\vee$. On note $\mathfrak{Irr}_{\text{tunip}}$ cet ensemble de triplets. Ce paramétrage a été obtenu par différents auteurs : Lusztig [1995], Mœglin [1996b, théorème 5.2] et Arthur [2013, théorème 2.2.1] dans le cas du groupe G_{iso} .

Dans [Mœglin et Waldspurger 2003; Waldspurger 2016b], on a montré que les représentations construites par Lusztig vérifiaient les propriétés de compatibilité à l'endoscopie qui les caractérisent. En particulier, dans le cas du groupe G_{iso} , ces représentations sont les mêmes que celles d'Arthur. Pour $(\lambda, s, \epsilon) \in \mathfrak{Irr}_{\text{tunip}}$, on note $\pi(\lambda, s, \epsilon)$ la représentation tempérée qui lui est associée par Lusztig. L'involution introduite par Zelevinsky dans le cas du groupe $\text{GL}(n)$ a été généralisée par Aubert et par Schneider et Stuhler aux groupes réductifs quelconques. On la note D et on pose $\delta(\lambda, s, \epsilon) = D(\pi(\lambda, s, \epsilon))$.

Soit $\mathfrak{t} = \text{iso}$ ou an et soit π une représentation admissible irréductible de $G_{\mathfrak{t}}(F)$. Notons $\mathfrak{g}_{\mathfrak{t}}$ l'algèbre de Lie de $G_{\mathfrak{t}}$. Harish-Chandra a prouvé que, dans un voisinage de l'origine, le caractère de π , descendu par l'exponentielle à $\mathfrak{g}_{\mathfrak{t}}(F)$, était combinaison linéaire de transformées de Fourier d'intégrales orbitales nilpotentes. Fixons une clôture algébrique \bar{F} de F et notons $\bar{\mathcal{N}}(\pi)$ l'ensemble des orbites nilpotentes \mathcal{O} dans $\mathfrak{g}_{\mathfrak{t}}(\bar{F})$ vérifiant la condition suivante : il existe une orbite nilpotente \mathcal{O} dans $\mathfrak{g}_{\mathfrak{t}}(F)$, qui est incluse dans \mathcal{O} et qui intervient avec un coefficient non nul dans le développement ci-dessus du caractère de π . On dit que π admet un front d'onde si $\bar{\mathcal{N}}(\pi)$ admet un unique élément maximal. Dans ce cas, on dit que cet élément maximal est le front d'onde de π . Les orbites nilpotentes dans $\mathfrak{g}_{\mathfrak{t}}(\bar{F})$ sont paramétrées par les partitions orthogonales de $2n + 1$ et nous identifions ces deux ensembles. On conjecture (ce qui est peut-être hasardeux) que toute représentation admissible irréductible admet un front d'onde. Signalons que, dans le cas où le corps de base est non pas p -adique, mais réel, la notion de front d'onde est également définie et se révèle importante, cf. par exemple [Barbasch et Vogan 1985].

En modifiant quelque peu une construction de Spaltenstein, on définit une "dualité" qui envoie une partition symplectique λ de $2n$ sur une partition orthogonale $d(\lambda)$ de $2n + 1$. La partition $d(\lambda)$ est toujours spéciale et la dualité d n'est pas bijective (par contre, sa restriction au sous-ensemble des partitions symplectiques spéciales de $2n$ est une bijection entre cet ensemble et celui des partitions orthogonales spéciales de $2n + 1$). On démontre dans cet article le résultat suivant.

Théorème. *Soit $(\lambda, s, \epsilon) \in \mathfrak{Irr}_{\text{tunip}}$. Alors la représentation $\delta(\lambda, s, \epsilon)$ admet un front d'onde et celui-ci est la partition $d(\lambda)$.*

Remarquons que l'on retrouve dans notre cas particulier le théorème 1.4 de [Mœglin 1996a] : ce front d'onde est une partition spéciale. Notre théorème n'est pas très nouveau. Mœglin [1996b, théorème 3.3.5]

a démontré un résultat similaire. Ses hypothèses étaient plus générales que les nôtres. D'une part, elle considérait tous les groupes classiques et pas seulement les groupes spéciaux orthogonaux. Surtout, elle considérait les représentations dont le paramètre de Langlands, sous sa forme habituelle, se restreint au groupe de Weil en une somme de caractères d'ordre au plus 2, éventuellement ramifiés. Nous nous limitons au cas de réduction unipotente, ce qui exclut les caractères ramifiés. Toutefois, notre résultat n'est pas inclus dans celui de [Mœglin 1996b] : avec nos notations, celui-ci suppose que les termes de λ sont tous distincts. La démonstration est aussi entièrement différente.

Soit $(\lambda, s, \epsilon) \in \mathfrak{Irr}_{\text{unip}}$, et posons $\delta = \delta(\lambda, s, \epsilon)$. Notons \sharp l'indice iso ou an tel que δ soit une représentation de $G_{\sharp}(F)$. Dans [Waldspurger 2016a], on a donné une formule qui calcule la restriction du caractère de δ aux éléments compacts de $G_{\sharp}(F)$ (ceux qui sont contenus dans un sous-groupe compact). A fortiori cette formule calcule la restriction du caractère à un voisinage de l'origine. Cette restriction est somme de distributions que l'on peut calculer si l'on connaît les restrictions de δ aux différents sous-groupes compacts maximaux de $G_{\sharp}(F)$, ou plus exactement les représentations des groupes "résiduels" qui s'en déduisent. La construction de Lusztig donne les renseignements voulus. A partir de là, en utilisant de nombreux travaux de Lusztig (faisceaux-caractères, correspondance de Springer généralisée, etc.), on traduit l'assertion à démontrer en termes de représentations de groupes de Weyl. Il s'agit en gros de savoir quelles sont les représentations qui peuvent intervenir dans certaines restrictions d'une représentation d'un produit de groupes de Weyl déterminée par δ . C'est un problème combinatoire que nous avons longuement étudié dans [Waldspurger 2001] et les résultats de cette référence permettent de conclure.

Remarque. Dans [Waldspurger 2001], le groupe était supposé non ramifié, ce qui est le cas de G_{iso} mais pas de G_{an} . En fait, cette hypothèse ne servait qu'à utiliser des résultats d'homogénéité qui n'étaient alors connus que sous cette hypothèse restrictive. Ils sont maintenant connus sans cette hypothèse, cf. [DeBacker 2002], et la plupart des résultats de [Waldspurger 2001], en particulier ceux que l'on utilisera, s'étendent au cas général.

Évidemment, il serait tentant d'appliquer la même méthode non pas à la bête représentation $\delta(\lambda, s, \epsilon)$, mais à la représentation tempérée $\pi(\lambda, s, \epsilon)$. Indiquons où est le problème. Les représentations des groupes "résiduels" associés à $\delta(\lambda, s, \epsilon)$ sont bien calculées par Lusztig, mais en termes de représentations de groupes de Weyl peu explicites. Plus précisément, il apparaît des représentations non irréductibles dont la décomposition en composantes irréductibles est dictée par des variantes de polynômes de Kazhdan-Lusztig. Ces représentations sont notées $\rho_{\lambda, \epsilon}$ dans notre article, mais il ne s'agit plus du même couple λ, ϵ , notons-les ici $\rho_{\nu, \tau}$. Il y a un ordre (partiel) naturel sur l'ensemble des représentations irréductibles et on contrôle très bien le terme minimal du développement de $\rho_{\nu, \tau}$ en composantes irréductibles. Il s'avère que cela nous suffit pour conclure. Si l'on remplace $\delta(\lambda, s, \epsilon)$ par $\pi(\lambda, s, \epsilon)$, les représentations $\rho_{\nu, \tau}$ sont remplacées par leur produit tensoriel avec sgn , le caractère signe du groupe de Weyl sous-jacent. Comme on peut s'y attendre, cela inverse l'ordre : on connaît le terme maximal du développement de $\text{sgn} \otimes \rho_{\nu, \tau}$. Mais maintenant, l'ordre va dans le mauvais sens et connaître le terme maximal ne permet plus de conclure.

1. Combinatoire

1.1. Partitions et représentations des groupes de Weyl. On appelle partition une classe d'équivalence de suites décroissantes finies de nombres entiers positifs ou nuls, deux suites étant équivalentes si elles ne diffèrent que par des termes nuls. Pour une telle partition $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r)$, on pose $S(\lambda) = \sum_{j=1, \dots, r} \lambda_j$ et on note $l(\lambda)$ le plus grand entier j tel que $\lambda_j \neq 0$. Cas particulier : on note \emptyset la partition $(0, \dots, 0)$ et on pose $l(\emptyset) = 0$. On note mult_λ la fonction sur $\mathbb{N} - \{0\}$ telle que, pour tout i dans cet ensemble, $\text{mult}_\lambda(i)$ est le nombre d'entiers j tels que $\lambda_j = i$. On pose aussi $\text{mult}_\lambda(\geq i) = \sum_{i' \geq i} \text{mult}_\lambda(i')$. On note $\text{Jord}(\lambda)$ l'ensemble des $i \geq 1$ tels que $\text{mult}_\lambda(i) \geq 1$. Soit $k \in \mathbb{N}$. A équivalence près, on peut supposer $r \geq k$ et on pose $S_k(\lambda) = \sum_{j=1, \dots, k} \lambda_j$. Pour $N \in \mathbb{N}$, on note $\mathcal{P}(N)$ l'ensemble des partitions λ telles que $S(\lambda) = N$. Plus généralement, pour un entier $k \geq 1$, on note $\mathcal{P}_k(N)$ l'ensemble des k -uples de partitions $(\lambda_1, \dots, \lambda_k)$ tels que $S(\lambda_1) + \dots + S(\lambda_k) = N$. On utilisera plus loin des variantes de cette notation, par exemple $\mathcal{P}_k^{\text{symp}}(2N)$ etc. On définit de la façon usuelle la transposition $\lambda \mapsto {}^t\lambda$ dans $\mathcal{P}(N)$ et les applications

$$(\lambda_1, \lambda_2) \mapsto \lambda_1 + \lambda_2 \quad \text{et} \quad (\lambda_1, \lambda_2) \mapsto \lambda_1 \cup \lambda_2$$

qui envoient $\mathcal{P}_2(N)$ dans $\mathcal{P}(N)$. On définit un ordre partiel sur $\mathcal{P}(N)$: pour deux partitions $\lambda_1, \lambda_2 \in \mathcal{P}(N)$, $\lambda_1 \leq \lambda_2$ si et seulement si $S_k(\lambda_1) \leq S_k(\lambda_2)$ pour tout $k \in \mathbb{N}$.

Plusieurs notations ci-dessus se généralisent aux suites finies $\alpha = (\alpha_1, \dots, \alpha_r)$ de nombres réels pas forcément décroissantes. Par exemple, si k est un entier tel que $0 \leq k \leq r$, on pose $S_k(\alpha) = \sum_{j=1, \dots, k} \alpha_j$. On utilisera aussi la notation $\alpha_{\leq k} = S_k(\alpha)$. Si α et β sont deux suites de même longueur, on note $\alpha + \beta$ la suite $(\alpha_1 + \beta_1, \dots, \alpha_r + \beta_r)$.

Pour tout ensemble X , on note $\mathbb{C}[X]$ l'espace vectoriel complexe de base X . Pour tout groupe fini W , on note \widehat{W} l'ensemble des classes de représentations irréductibles de W . En identifiant une telle représentation à son caractère, l'espace $\mathbb{C}[X]$ s'identifie à celui des fonctions de W dans \mathbb{C} qui sont invariantes par conjugaison.

Soit $N \in \mathbb{N}$. On note \mathfrak{S}_N le groupe des permutations de l'ensemble $\{1, \dots, N\}$. On sait paramétrer $\widehat{\mathfrak{S}}_N$ par $\mathcal{P}(N)$, on note $\rho(\lambda)$ la représentation irréductible correspondant à une partition λ (en particulier la représentation triviale de \mathfrak{S}_N est paramétrée par la partition $\lambda = (n)$). On note sgn le caractère signe usuel de \mathfrak{S}_N . Si une représentation irréductible ρ est paramétrée par la partition λ , $\rho \otimes \text{sgn}$ est paramétrée par ${}^t\lambda$.

On note W_N le groupe de Weyl d'un système de racines de type B_N ou C_N (avec la convention $W_0 = \{1\}$). On sait paramétrer \widehat{W}_N par $\mathcal{P}_2(N)$, on note $\rho(\alpha, \beta)$ la représentation irréductible correspondant à un couple de partitions (α, β) (en particulier, la représentation triviale est paramétrée par $((N), \emptyset)$). On note sgn le caractère signe usuel de W_N et sgn_{CD} le caractère dont le noyau est le sous-groupe W_N^D d'un système de racines de type D_N . Si une représentation irréductible ρ est paramétrée par le couple de partitions (α, β) , $\rho \otimes \text{sgn}$ est paramétrée par $({}^t\beta, {}^t\alpha)$ et $\rho \otimes \text{sgn}_{CD}$ est paramétrée par (β, α) .

Supposons $N \geq 1$. Pour $(\alpha, \beta) \in \mathcal{P}_2(N)$, les restrictions à W_N^D de $\rho(\alpha, \beta)$ et $\rho(\beta, \alpha)$ sont équivalentes. Si $\alpha \neq \beta$, ces restrictions sont irréductibles, on les note $\rho^D(\alpha, \beta)$ ou $\rho^D(\beta, \alpha)$. Si $\alpha = \beta$, la restriction de $\rho(\alpha, \alpha)$ à W_N^D se décompose en deux représentations irréductibles, que l'on note $\rho^D(\alpha, \alpha, +)$ et

$\rho^D(\alpha, \alpha, -)$. Elles sont conjuguées par un élément de $W_N - W_N^D$ et on n'aura pas besoin de les distinguer. Toutes les représentations irréductibles de W_N^D sont ainsi obtenues.

1.2. Symboles. Pour tout ensemble fini X , on note $|X|$ le nombre d'éléments de X . Si X est un ensemble de nombres, on note $S(X)$ la somme des éléments de X . Pour tout nombre réel x , on note $[x]$ sa partie entière.

Soit $N \in \mathbb{N}$. Un symbole de rang N est une classe d'équivalence de couples (X, Y) de sous-ensembles finis de \mathbb{N} , vérifiant la condition

$$S(X) + S(Y) - \left[\left(\frac{1}{2}(|X| + |Y| - 1) \right)^2 \right] = N.$$

La relation d'équivalence est engendrée par les deux relations (qui préservent l'égalité précédente) :

$$(X, Y) \sim (X', Y') \quad \text{où} \quad X' = \{x + 1; x \in X\} \cup \{0\}, \quad Y' = \{y + 1; y \in Y\} \cup \{0\};$$

$$(X, Y) \sim (Y, X).$$

Remarque. Par abus de terminologie, on appellera plutôt symbole un couple (X, Y) représentant une classe d'équivalence.

Le défaut d'un symbole (X, Y) est la valeur absolue de $|X| - |Y|$ (il ne dépend que de la classe d'équivalence de (X, Y)). Pour $D \in \mathbb{N}$, on note $\mathcal{S}_{N,D}$ l'ensemble des symboles de rang N et de défaut D .

On regroupe les symboles en familles : deux symboles sont dans la même famille si et seulement si on peut les représenter par des couples (X, Y) et (X', Y') tels que $X \cup Y = X' \cup Y'$ et $X \cap Y = X' \cap Y'$. La parité du défaut est constante sur chaque famille. Toute famille de symboles de défaut impair contient un unique symbole spécial, c'est-à-dire représenté par un couple (X, Y) de la forme $X = (x_1 \geq \dots \geq x_{r+1})$, $Y = (y_1 \geq \dots \geq y_r)$ et tel que

$$x_1 \geq y_1 \geq x_2 \geq y_2 \geq \dots \geq y_r \geq x_{r+1}.$$

Toute famille de symboles de défaut pair contient un unique symbole spécial, c'est-à-dire représenté par un couple (X, Y) de la forme $X = (x_1 \geq \dots \geq x_r)$, $Y = (y_1 \geq \dots \geq y_r)$ et tel que

$$x_1 \geq y_1 \geq x_2 \geq y_2 \geq \dots \geq y_r.$$

Soit (X, Y) un symbole de rang N . Fixons un entier d majorant les éléments de $X \cup Y$. Posons

$$X' = \{d, \dots, 0\} - \{d - y; y \in Y\}, \quad Y' = \{d, \dots, 0\} - \{d - x; x \in X\}.$$

On vérifie que (X', Y') est un symbole de rang N . A équivalence près, il ne dépend pas du choix de d et ne dépend que de la classe d'équivalence de (X, Y) . Cette construction définit une "dualité" $(X, Y) \mapsto d(X, Y) = (X', Y')$ dans l'ensemble des symboles de rang N . Cette dualité conserve le défaut et est involutive : $d \circ d$ est l'identité. Elle se restreint en une involution du sous-ensemble des symboles spéciaux. Enfin, deux symboles sont dans une même famille si et seulement si leurs images par dualité le

sont. Autrement dit, si (X, Y) est dans la famille du symbole spécial $(X^{\text{sp}}, Y^{\text{sp}})$, alors $d(X, Y)$ est dans la famille du symbole spécial $d(X^{\text{sp}}, Y^{\text{sp}})$.

Soit $\rho \in \widehat{W}_N$. Comme en 1.1, on lui associe un couple de partitions $(\alpha, \beta) \in \mathcal{P}_2(N)$. On choisit un entier $r \geq l(\alpha), l(\beta)$. On pose $X = \alpha + \{r, \dots, 0\}$, $Y = \beta + \{r-1, \dots, 0\}$. Alors (X, Y) est un symbole de rang N et de défaut 1 dont la classe ne dépend pas de r . On pose $\text{symb}(\rho) = (X, Y)$. L'application $\text{symb} : \widehat{W}_N \rightarrow \mathcal{S}_{N,1}$ ainsi définie est bijective. On a $\text{symb}(\rho \otimes \text{sgn}) = d \circ \text{symb}(\rho)$.

Soit $\rho \in \widehat{W}_N^D$. Comme en 1.1, on lui associe un couple de partitions $(\alpha, \beta) \in \mathcal{P}_2(N)$. On choisit un entier $r \geq l(\alpha), l(\beta)$. On pose $X = \alpha + \{r-1, \dots, 0\}$, $Y = \beta + \{r-1, \dots, 0\}$. Alors (X, Y) est un symbole de rang N et de défaut 0 dont la classe ne dépend pas de r . On pose $\text{symb}(\rho) = (X, Y)$. L'application $\text{symb} : \widehat{W}_N^D \rightarrow \mathcal{S}_{N,0}$ ainsi définie est surjective. Ses fibres ont un ou deux éléments, celles à deux éléments étant formées des couples de la forme $\rho(\alpha, \alpha, +)$, $\rho(\alpha, \alpha, -)$. On a $\text{symb}(\rho \otimes \text{sgn}) = d \circ \text{symb}(\rho)$.

1.3. Correspondance de Springer, cas symplectique. Soit $n \in \mathbb{N}$. On note $\mathcal{P}^{\text{symp}}(2n)$ l'ensemble des partitions symplectiques de $2n$, c'est-à-dire les $\lambda \in \mathcal{P}(2n)$ telles que $\text{mult}_\lambda(i)$ est pair pour tout entier i impair. Pour une telle partition, on note $\text{Jord}_{\text{bp}}(\lambda)$ l'ensemble des entiers $i \geq 2$ pairs tels que $\text{mult}_\lambda(i) \geq 1$. Plus précisément, pour un entier $k \geq 1$, on note $\text{Jord}_{\text{bp}}^k(\lambda)$ l'ensemble des $i \geq 2$ pairs tels que $\text{mult}_\lambda(i) = k$.

On note $\mathcal{P}^{\text{symp}}(2n)$ l'ensemble des couples (λ, ϵ) où $\lambda \in \mathcal{P}^{\text{symp}}(2n)$ et $\epsilon \in \{\pm 1\}^{\text{Jord}_{\text{bp}}(\lambda)}$. La correspondance de Springer généralisée établit une bijection entre $\mathcal{P}^{\text{symp}}(2n)$ et l'ensemble des couples (ρ, k) tels que

$$k \in \mathbb{N} \quad \text{et} \quad k(k+1) \leq 2n; \quad \rho \in \widehat{W}_{n-k(k+1)/2}.$$

On note $(\rho_{\lambda, \epsilon}, k_{\lambda, \epsilon})$ le couple associé à un élément $(\lambda, \epsilon) \in \mathcal{P}^{\text{symp}}(2n)$. L'entier $k_{\lambda, \epsilon}$ se calcule de la façon suivante. Notons $i_1 > \dots > i_m > 0$ les entiers pairs i tels que $\text{mult}_\lambda(i)$ soit impair. Posons

$$h = \sum_{j=1, \dots, m} (-1)^j (1 - \epsilon(i_j)).$$

Alors $k_{\rho, \epsilon} = \sup(h, -h-1)$. En particulier, si $\epsilon = 1$, c'est-à-dire $\epsilon(i) = 1$ pour tout $i \in \text{Jord}_{\text{bp}}(\lambda)$, on a $k_{\lambda, 1} = 0$ et $\rho_{\lambda, 1} \in \widehat{W}_n$.

Une partition $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots) \in \mathcal{P}^{\text{symp}}(2n)$ est spéciale si et seulement si λ_{2j-1} et λ_{2j} sont de même parité pour tout $j \geq 1$. Cela équivaut à ce que ${}^t\lambda$ soit symplectique. En notant $\mathcal{P}^{\text{symp}, \text{sp}}(2n)$ l'ensemble des partitions symplectiques spéciales de $2n$, l'application $\lambda \mapsto {}^t\lambda$ est une involution de $\mathcal{P}^{\text{symp}, \text{sp}}(2n)$. Considérons une partition $\lambda \in \mathcal{P}^{\text{symp}, \text{sp}}(2n)$ et définissons $i_1 > \dots > i_m$ comme ci-dessus. Si m est pair, on pose $m' = m$; si m est impair, on pose $m' = m+1$ et $i_{m'} = 0$. On appelle intervalle de λ un ensemble Δ de l'une des formes suivantes :

- pour un entier $h = 1, \dots, m'/2$, Δ est l'ensemble des i tels que $i = 0$ ou $i \geq 1$ et $\text{mult}_\lambda(i) \geq 1$ et tels que $i_{2h-1} \geq i \geq i_{2h}$;
- $\Delta = \{i\}$ où i est un entier pair tel que $i = 0$ ou $i \geq 2$ et $\text{mult}_\lambda(i) \geq 1$ et tel qu'il n'existe pas d'entier $h = 1, \dots, m'/2$ de sorte que $i_{2h-1} \geq i \geq i_{2h}$.

Parce que λ est spéciale, on vérifie que les intervalles sont formés d'entiers pairs (c'est évident dans le deuxième cas ci-dessus, un peu moins dans le premier). Ils forment une partition de l'ensemble $\text{Jord}_{\text{bp}}(\lambda) \cup \{0\}$. On ordonne les intervalles : $\Delta > \Delta'$ si $i > i'$ pour tous $i \in \Delta, i' \in \Delta'$. On note Δ_{\min} le plus petit intervalle (c'est celui qui contient 0). On note $\text{Int}(\lambda)$ l'ensemble des intervalles de λ .

L'application $\lambda \mapsto \text{symb}(\rho_{\lambda,1})$ est une bijection de $\mathcal{P}^{\text{symp,sp}}(2n)$ sur l'ensemble des symboles spéciaux de rang n et de défaut 1. Pour $\lambda \in \mathcal{P}^{\text{symp,sp}}(2n)$, on a défini en [Waldspurger 2001, VIII.17] une bijection fam entre la famille du symbole $\text{symb}(\rho_{\lambda,1})$ et l'ensemble des

$$(\tau, \delta) \in (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \times (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \quad \text{tels que } \tau(\Delta_{\min}) = \delta(\Delta_{\min}) = 0.$$

Soit $\lambda \in \mathcal{P}^{\text{symp}}(2n)$. Il existe une unique partition spéciale $\text{sp}(\lambda) \in \mathcal{P}^{\text{symp,sp}}(2n)$ telle que $\text{symb}(\rho_{\lambda,1})$ et $\text{symb}(\rho_{\text{sp}(\lambda),1})$ soient dans la même famille. Il est connu que $\lambda \leq \text{sp}(\lambda)$ et que $\text{sp}(\lambda)$ est la plus petite partition symplectique spéciale λ' telle que $\lambda \leq \lambda'$. Plus généralement, on a le lemme suivant.

Lemme. (i) Soit $(\lambda, \epsilon) \in \mathcal{P}^{\text{symp}}(2n)$, supposons $k_{\lambda,\epsilon} = 0$. Notons $\text{sp}(\lambda, \epsilon)$ l'unique partition spéciale telle que $\text{symb}(\rho_{\lambda,\epsilon})$ et $\text{symb}(\rho_{\text{sp}(\lambda,\epsilon),1})$ soient dans la même famille. Alors $\lambda \leq \text{sp}(\lambda, \epsilon)$.

(ii) Soit $\lambda \in \mathcal{P}^{\text{symp,sp}}(2n)$. Pour $\epsilon \in \{\pm 1\}^{\text{Jord}_{\text{bp}}(\lambda)}$, les conditions suivantes sont équivalentes :

(a) $k_{\lambda,\epsilon} = 0$ et $\text{sp}(\lambda, \epsilon) = \lambda$;

(b) ϵ est constant sur tout $\Delta \in \text{Int}(\lambda)$ et, dans le cas où $\Delta_{\min} \neq \{0\}$, $\epsilon(i) = 1$ pour tout $i \in \Delta_{\min} - \{0\}$.

(iii) Soit $\lambda \in \mathcal{P}^{\text{symp,sp}}(2n)$. L'application $\epsilon \mapsto \text{fam} \circ \text{symb}(\rho_{\lambda,\epsilon})$ est une bijection entre l'ensemble des ϵ décrits au (ii) et le sous-ensemble des

$$(\tau, \delta) \in (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \times (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \quad \text{tels que } \delta = 0 \text{ et } \tau(\Delta_{\min}) = 0.$$

La preuve est similaire à celle du lemme 1.4 ci-dessous.

1.4. Correspondance de Springer, cas orthogonal impair. Soit $n \in \mathbb{N}$. On note $\mathcal{P}^{\text{orth}}(2n+1)$ l'ensemble des partitions orthogonales de $2n+1$, c'est-à-dire les $\lambda \in \mathcal{P}(2n+1)$ telles que $\text{mult}_{\lambda}(i)$ est pair pour tout entier $i > 0$ pair. Pour une telle partition, on note $\text{Jord}_{\text{bp}}(\lambda)$ l'ensemble des entiers $i \geq 1$ impairs tels que $\text{mult}_{\lambda}(i) \geq 1$. Plus précisément, pour un entier $k \geq 1$, on note $\text{Jord}_{\text{bp}}^k(\lambda)$ l'ensemble des $i \geq 1$ impairs tels que $\text{mult}_{\lambda}(i) = k$.

On note $\mathcal{P}^{\text{orth}}(2n+1)$ l'ensemble des couples (λ, ϵ) où $\lambda \in \mathcal{P}^{\text{orth}}(2n+1)$ et $\epsilon \in (\{\pm 1\}^{\text{Jord}_{\text{bp}}(\lambda)})/\{\pm 1\}$, le groupe $\{\pm 1\}$ s'envoyant diagonalement dans $\{\pm 1\}^{\text{Jord}_{\text{bp}}(\lambda)}$. En pratique, on relèvera ϵ en un élément de $\{\pm 1\}^{\text{Jord}_{\text{bp}}(\lambda)}$. Sauf indication contraire, les formules que nous écrirons ne dépendront pas du choix de ce relèvement. La correspondance de Springer généralisée établit une bijection entre $\mathcal{P}^{\text{orth}}(2n+1)$ et l'ensemble des couples (ρ, k) tels que

$$k \in \mathbb{N}, \quad k \text{ est impair et } k^2 \leq 2n+1; \quad \rho \in \widehat{W}_{n-(k^2-1)/2}.$$

On note $(\rho_{\lambda,\epsilon}, k_{\lambda,\epsilon})$ le couple associé à un élément $(\lambda, \epsilon) \in \mathcal{P}^{\text{orth}}(2n+1)$. L'entier $k_{\lambda,\epsilon}$ se calcule de la façon suivante. Notons $i_1 > \dots > i_m$ les entiers impairs i tels que $\text{mult}_\lambda(i)$ soit impair. Posons

$$h = \sum_{j=1,\dots,m} (-1)^j (1 - \epsilon(i_j)).$$

Alors $k_{\lambda,\epsilon} = |h+1|$. En particulier, si $\epsilon = 1$, c'est-à-dire $\epsilon(i) = 1$ pour tout $i \in \text{Jord}_{\text{bp}}(\lambda)$, on a $k_{\lambda,1} = 1$ et $\rho_{\lambda,1} \in \widehat{W}_n$.

Une partition $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots) \in \mathcal{P}^{\text{orth}}(2n+1)$ est spéciale si et seulement si λ_1 est impair et λ_{2j} et λ_{2j+1} sont de même parité pour tout $j \geq 1$. Cela équivaut à ce que ${}^t\lambda$ soit orthogonale. En notant $\mathcal{P}^{\text{orth,sp}}(2n+1)$ l'ensemble des partitions orthogonales spéciales de $2n+1$, l'application $\lambda \mapsto {}^t\lambda$ est une involution de $\mathcal{P}^{\text{orth,sp}}(2n+1)$. Considérons une partition $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n+1)$ et définissons $i_1 > \dots > i_m$ comme ci-dessus. L'entier m est forcément impair. On appelle intervalle de λ un ensemble Δ de l'une des formes suivantes :

- pour un entier $h = 0, \dots, (m-1)/2$, Δ est l'ensemble des $i \geq 1$ tels que $\text{mult}_\lambda(i) \geq 1$ et tels que $i_{2h} \geq i \geq i_{2h+1}$, avec la convention $i_0 = \infty$;
- $\Delta = \{i\}$ où i est un entier impair tel que $\text{mult}_\lambda(i) \geq 1$ et tel qu'il n'existe pas d'entier $h = 0, \dots, (m-1)/2$ de sorte que $i_{2h} \geq i \geq i_{2h+1}$.

Parce que λ est spéciale, on vérifie que les intervalles sont formés d'entiers impairs. Ils forment une partition de l'ensemble $\text{Jord}_{\text{bp}}(\lambda)$. On ordonne les intervalles comme dans le cas symplectique. On note Δ_{\min} le plus petit intervalle et Δ_{\max} le plus grand. On note $\text{Int}(\lambda)$ l'ensemble des intervalles.

L'application $\lambda \mapsto \text{symb}(\rho_{\lambda,1})$ est une bijection de $\mathcal{P}^{\text{orth,sp}}(2n+1)$ sur l'ensemble des symboles spéciaux de rang n et de défaut 1. Pour $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n+1)$, on a défini en [Waldspurger 2001, VIII.19] une bijection fam entre la famille du symbole $\text{symb}(\rho_{\lambda,1})$ et l'ensemble des $(\tau, \delta) \in (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \times (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)}$ tels que $\tau(\Delta_{\max}) = \delta(\Delta_{\min}) = 0$.

Soit $\lambda \in \mathcal{P}^{\text{orth}}(2n+1)$. Il existe une unique partition spéciale $\text{sp}(\lambda) \in \mathcal{P}^{\text{orth,sp}}(2n+1)$ telle que $\text{symb}(\rho_{\lambda,1})$ et $\text{symb}(\rho_{\text{sp}(\lambda),1})$ soient dans la même famille. Il est connu que $\lambda \leq \text{sp}(\lambda)$ et que $\text{sp}(\lambda)$ est la plus petite partition orthogonale spéciale λ' telle que $\lambda \leq \lambda'$. Plus généralement, on a le lemme suivant.

Lemme. (i) Soit $(\lambda, \epsilon) \in \mathcal{P}^{\text{orth}}(2n+1)$, supposons $k_{\lambda,\epsilon} = 1$. Notons $\text{sp}(\lambda, \epsilon)$ l'unique partition spéciale telle que $\text{symb}(\rho_{\lambda,\epsilon})$ et $\text{symb}(\rho_{\text{sp}(\lambda,\epsilon),1})$ soient dans la même famille. Alors $\lambda \leq \text{sp}(\lambda, \epsilon)$.

(ii) Soit $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n+1)$. Pour $\epsilon \in \{\pm 1\}^{\text{Jord}_{\text{bp}}(\lambda)}$, les conditions suivantes sont équivalentes :

- (a) $k_{\lambda,\epsilon} = 1$ et $\text{sp}(\lambda, \epsilon) = \lambda$;
- (b) ϵ est constant sur tout $\Delta \in \text{Int}(\lambda)$.

(iii) Soit $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n+1)$. L'application $\epsilon \mapsto \text{fam} \circ \text{symb}(\rho_{\lambda,\epsilon})$ est une bijection entre l'ensemble des ϵ décrits au (ii), modulo le groupe diagonal $\{\pm 1\}$, et le sous-ensemble des

$$(\tau, \delta) \in (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \times (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \quad \text{tels que } \delta = 0 \text{ et } \tau(\Delta_{\max}) = 0.$$

Preuve. Soit (λ, ϵ) comme en (i). On peut supposer $\lambda = (\lambda_1, \dots, \lambda_{2n+1})$. Dans la suite $\lambda + \{2n, \dots, 0\}$, il y a n nombres pairs notés $2z_1 > \dots > 2z_n$ et $n + 1$ nombres impairs notés $2z'_1 + 1 > \dots > 2z'_{n+1} + 1$. On note $z = (z_1, \dots, z_n)$ et $z' = (z'_1, \dots, z'_{n+1})$ puis

$$A^\sharp = z' + \{n, \dots, 0\} = (a_1^\sharp, \dots, a_{n+1}^\sharp), \quad B^\sharp = z + \{n-1, \dots, 0\} = (b_1^\sharp, \dots, b_n^\sharp).$$

On vérifie que

$$a_1^\sharp \geq b_1^\sharp \geq a_2^\sharp \geq \dots \geq b_n^\sharp \geq a_{n+1}^\sharp. \quad (1)$$

On voit aussi qu'il y a une unique bijection croissante $i \mapsto \Sigma_i$ entre l'ensemble $\text{Jord}_{\text{bp}}(\lambda)$ et celui des sous-ensembles non vides de $(A^\sharp \cup B^\sharp) - (A^\sharp \cap B^\sharp)$ formés d'entiers consécutifs, et maximaux pour cette propriété. Posons

$$A = \left(A^\sharp - \bigcup_{i \in \text{Jord}_{\text{bp}}(\lambda), \epsilon_i = -1} (\Sigma_i \cap A^\sharp) \right) \cup \left(\bigcup_{i \in \text{Jord}_{\text{bp}}(\lambda), \epsilon_i = -1} (\Sigma_i \cap B^\sharp) \right),$$

$$B = \left(B^\sharp - \bigcup_{i \in \text{Jord}_{\text{bp}}(\lambda), \epsilon_i = -1} (\Sigma_i \cap B^\sharp) \right) \cup \left(\bigcup_{i \in \text{Jord}_{\text{bp}}(\lambda), \epsilon_i = -1} (\Sigma_i \cap A^\sharp) \right).$$

Si on multiplie ϵ par l'élément diagonal -1 , on échange A et B . On peut donc choisir le relèvement ϵ de sorte que $|A| \geq |B|$. L'hypothèse $k_{\lambda, \epsilon} = 1$ entraîne alors que $|A| = n + 1$ et $|B| = n$. On définit les suites $X = (x_1, \dots, x_{n+1})$ et $Y = (y_1, \dots, y_n)$ par $X + \{n, \dots, 0\} = A$ et $Y + \{n-1, \dots, 0\} = B$. Alors $\text{symb}(\rho_{\lambda, \epsilon}) = (X, Y)$.

Soit $k \in \{1, \dots, 2n+1\}$. On va majorer $S_k(\lambda)$ en fonction de (X, Y) . Tout d'abord

$$S_k(\lambda) = S_k(\lambda + \{2n, \dots, 0\}) - \frac{1}{2}k(4n+1-k). \quad (2)$$

Notons $j_1 < \dots < j_s$ les indices j tels que λ_j soit impair et $h_1 < \dots < h_t$ les indices h pour lesquels λ_h est pair. Puisque la somme des λ_j vaut $2n+1$ qui est impair, s est impair et $t = 2n+1-s$ est pair. Puisque tout nombre pair non nul intervient avec multiplicité paire, la parité de t entraîne que 0 intervient aussi avec multiplicité paire. Il en résulte aussi que $h_{2r} = h_{2r-1} + 1$ pour tout $r = 1, \dots, t/2$. Pour $r = 1, \dots, s$, il y a $j_r - r$ termes λ_j qui sont pairs et strictement supérieurs à λ_{j_r} . Puisque ces termes interviennent avec multiplicité paire, j_r est de même parité que r . Soient $s_k \in \{1, \dots, s\}$ et $t_k \in \{1, \dots, t\}$ les plus grands entiers tels que $j_{s_k} \leq k$ et $h_{t_k} \leq k$. On a $s_k + t_k = k$. Les k premiers termes de $\lambda + \{2n, \dots, 0\}$ sont les

$$\lambda_{j_u} + 2n + 1 - j_u \quad \text{pour } u = 1, \dots, s_k, \quad (3)$$

$$\lambda_{h_v} + 2n + 1 - h_v \quad \text{pour } v = 1, \dots, t_k. \quad (4)$$

D'après les propriétés de nos suites, il y a $[(s_k + 1)/2]$ éléments impairs et $[s_k/2]$ éléments pairs parmi les éléments (3). Si t_k est pair, il y a $t_k/2$ éléments impairs et $t_k/2$ éléments pairs parmi les éléments (4). Si t_k est impair et h_{t_k} est pair, il y a $(t_k + 1)/2$ éléments impairs et $(t_k - 1)/2$ éléments pairs parmi les éléments (4). Si t_k est impair et h_{t_k} est impair, il y a $(t_k - 1)/2$ éléments impairs et $(t_k + 1)/2$ éléments

pairs parmi les éléments (4). En réunissant les deux types d'éléments, on voit que, parmi les k premiers termes de $\lambda + \{2n, \dots, 0\}$, il y a $[(k+1)/2] + \eta$ termes impairs et $[k/2] - \eta$ termes pairs, où

- $\eta = 1$ si t_k et s_k sont impairs et h_{t_k} est pair,
- $\eta = -1$ si t_k et h_{t_k} sont impairs et s_k est pair,
- $\eta = 0$ dans les autres cas.

Les k premiers termes de $\lambda + \{2n, \dots, 0\}$ sont donc $2z'_1 + 1, \dots, 2z'_{[(k+1)/2] + \eta} + 1$ et $2z_1, \dots, 2z_{[k/2] - \eta}$.

Supposons λ_k impair. Alors $j_{s_k} = k$. On a dit que s_k est de la même parité que j_{s_k} , donc que k , donc $t_k = k - s_k$ est pair. Donc $\eta = 0$ et il résulte de la description ci-dessus que

$$S_k(\lambda + \{2n, \dots, 0\}) = 2S_{[(k+1)/2]}(z') + [(k+1)/2] + 2S_{[k/2]}(z). \quad (5)$$

Supposons λ_k pair. Alors $h_{t_k} = k$. Si $\eta = 0$, le calcul est le même que ci-dessus et on a (5). Supposons $\eta = 1$. Alors $k = h_{t_k}$ est pair et les k premiers termes de $\lambda + \{2n, \dots, 0\}$ sont $2z'_1 + 1, \dots, 2z'_{k/2+1} + 1$ et $2z_1, \dots, 2z_{k/2-1}$. Le dernier de ces termes est $\lambda_k + 2n + 1 - k$ qui est impair, donc c'est $2z'_{k/2+1} + 1$. Mais, t_k étant impair, on a $h_{t_k+1} = h_{t_k} + 1 = k + 1$ et $\lambda_k = \lambda_{k+1}$. Le $k+1$ -ième terme de $\lambda + \{2n, \dots, 0\}$ est $\lambda_k + 2n - k$ qui est pair, c'est donc le premier terme pair strictement inférieur à $2z_{k/2-1}$, autrement dit, c'est $2z_{k/2}$. Les égalités $\lambda_k + 2n + 1 - k = 2z'_{k/2+1} + 1$ et $\lambda_k + 2n - k = 2z_{k/2}$ entraînent $z'_{k/2+1} = z_{k/2}$. Les k premiers termes de $\lambda + \{2n, \dots, 0\}$ sont donc aussi bien $2z'_1 + 1, \dots, 2z'_{k/2} + 1$ et $2z_1, \dots, 2z_{k/2-1}, 2z_{k/2} + 1$. On obtient alors

$$S_k(\lambda + \{2n, \dots, 0\}) = 2S_{[(k+1)/2]}(z') + [(k+1)/2] + 1 + 2S_{[k/2]}(z). \quad (6)$$

Supposons maintenant $\eta = -1$. Alors $k = h_{t_k}$ est impair et les k premiers termes de $\lambda + \{2n, \dots, 0\}$ sont $2z'_1 + 1, \dots, 2z'_{(k-1)/2} + 1$ et $2z_1, \dots, 2z_{(k+1)/2}$. Le dernier de ces termes est $\lambda_k + 2n + 1 - k$ qui est pair, donc c'est $2z_{(k+1)/2}$. Comme ci-dessus, on a $\lambda_k = \lambda_{k+1}$. Le $k+1$ -ième terme de $\lambda + \{2n, \dots, 0\}$ est $\lambda_k + 2n - k$ qui est impair, c'est donc le premier terme impair strictement inférieur à $2z'_{(k-1)/2} + 1$, autrement dit, c'est $2z'_{(k+1)/2} + 1$. Les égalités $\lambda_k + 2n + 1 - k = 2z_{(k+1)/2}$ et $\lambda_k + 2n - k = 2z'_{(k+1)/2} + 1$ entraînent $z'_{(k+1)/2} + 1 = z_{(k+1)/2}$. Les k premiers termes de $\lambda + \{2n, \dots, 0\}$ sont donc aussi bien $2z'_1 + 1, \dots, 2z'_{(k-1)/2} + 1, 2z'_{(k+1)/2} + 2$ et $2z_1, \dots, 2z_{(k-1)/2}$. On obtient encore (6).

Supposons encore λ_k pair. Puisque $h_{t_k} = k = s_k + t_k$, les conditions de parité sur t_k sont redondantes dans la définition de η . On voit que $\eta = \pm 1$ si et seulement si $k + s_k$ est impair. On voit aussi que s_k est de la même parité que $S_k(\lambda)$. On a obtenu que, si $S_k(\lambda) + k$ est pair, on a la formule (5) tandis que, si $S_k(\lambda) + k$ est impair, on a la formule (6). Posons alors, pour $k = 1, \dots, 2n + 1$,

$$v_k(\lambda) = 1 \quad \text{si } \lambda_k \text{ est pair et } S_k(\lambda) + k \text{ est impair,} \quad v_k(\lambda) = 0 \quad \text{sinon.}$$

On a la formule générale :

$$S_k(\lambda + \{2n, \dots, 0\}) = 2S_{[(k+1)/2]}(z') + [(k+1)/2] + v_k(\lambda) + 2S_{[k/2]}(z). \quad (7)$$

Remarquons que, d'après le calcul ci-dessus, on a

$$\text{si } \nu_k(\lambda) = 1, \quad \text{alors } k \leq 2n \text{ et } \lambda_{k+1} = \lambda_k. \quad (8)$$

D'après la définition des termes A^\sharp et B^\sharp , on a les égalités

$$\begin{aligned} S_{[(k+1)/2]}(z') &= S_{[(k+1)/2]}(A^\sharp) - [(k+1)/2](2n+1 - [(k+1)/2])/2, \\ S_{[k/2]}(z) &= S_{[k/2]}(B^\sharp) - [k/2](2n-1 - [k/2])/2. \end{aligned}$$

Avec les formules (2) et (7), on obtient

$$S_k(\lambda) = 2S_{[(k+1)/2]}(A^\sharp) + 2S_{[k/2]}(B^\sharp) + \nu_k(\lambda) + c_k,$$

où c_k est un nombre qui ne dépend pas de λ . Notons $A^\sharp \sqcup B^\sharp$ la réunion des suites A^\sharp et B^\sharp , les termes étant rangés en ordre décroissant mais comptés avec leur multiplicité (c'est-à-dire qu'un terme intervenant dans les deux suites intervient avec multiplicité 2). La propriété (1) entraîne que la réunion (en ce sens) des $[(k+1)/2]$ plus grands termes de A^\sharp et des $[(k-1)/2]$ plus grands termes de B^\sharp n'est autre que la suite des k plus grands termes de $A^\sharp \sqcup B^\sharp$. La formule précédente se récrit

$$S_k(\lambda) = 2S_k(A^\sharp \sqcup B^\sharp) + \nu_k(\lambda) + c_k.$$

Avec une définition similaire, on a $A^\sharp \sqcup B^\sharp = A \sqcup B$, donc aussi $S_k(A^\sharp \sqcup B^\sharp) = S_k(A \sqcup B)$. Il existe deux entiers $e, f \in \mathbb{N}$ tels que $e + f = k$ et que la famille des k plus grands éléments de $A \sqcup B$ soit la réunion des familles des e plus grands éléments de A et des f plus grands éléments de B . Alors

$$S_k(A \sqcup B) = S_e(A) + S_f(B). \quad (9)$$

Par définition de X et Y , on a

$$S_e(A) = S_e(X) + e(2n+1-e)/2, \quad S_f(B) = S_f(Y) + f(2n-1-f)/2.$$

D'où

$$S_k(\lambda) = 2S_e(X) + 2S_f(Y) + \nu_k(\lambda) - e(2-e) - f^2 + c'_k,$$

où $c'_k = c_k + (2n-1)k$ est indépendant de λ . Le terme $S_e(X) + S_f(Y)$ est la somme de k termes de la famille $X \sqcup Y$, donc il est majoré par la somme des k plus grands termes de cette famille :

$$S_e(X) + S_f(Y) \leq S_k(X \sqcup Y). \quad (10)$$

D'où

$$S_k(\lambda) \leq 2S_k(X \sqcup Y) + \nu_k(\lambda) - e(e-2) - f^2 + c'_k. \quad (11)$$

Posons $\underline{\lambda} = \text{sp}(\lambda, \epsilon)$. Reprenons le calcul en remplaçant (λ, ϵ) par $(\underline{\lambda}, 1)$. On souligne les objets associés à cette paire. Parce que le caractère ϵ est remplacé par 1, on a les égalités $\underline{A} = \underline{A}^\sharp$, $\underline{B} = \underline{B}^\sharp$ et l'on voit que $\underline{e} = [(k+1)/2]$ et $\underline{f} = [k/2]$. Parce que le symbole $(\underline{X}, \underline{Y})$ est spécial, la réunion des

$[(k+1)/2]$ plus grands termes de \underline{X} et des $[k/2]$ plus grands termes de \underline{Y} n'est autre que la famille des k plus grands termes de $\underline{X} \sqcup \underline{Y}$. L'analogue de l'inégalité (10) est donc une égalité et on obtient

$$S_k(\underline{\lambda}) = 2S_k(\underline{X} \sqcup \underline{Y}) + \nu_k(\underline{\lambda}) - [(k+1)/2]([(k+1)/2] - 2) - [k/2]^2 + c'_k.$$

Par définition de $\text{sp}(\lambda, \epsilon)$, les symboles (X, Y) et $(\underline{X}, \underline{Y})$ sont dans la même famille, d'où $X \sqcup Y = \underline{X} \sqcup \underline{Y}$. En comparant (11) avec l'égalité ci-dessus, on obtient

$$S_k(\lambda) \leq S_k(\underline{\lambda}) + \nu_k(\lambda) - e(e-2) - f^2 - \nu_k(\underline{\lambda}) + [(k+1)/2]([(k+1)/2] - 2) + [k/2]^2. \quad (12)$$

On vérifie que, pour deux entiers e, f tels que $e + f = k$, on a

$$e(e-2) + f^2 \geq [(k+1)/2]([(k+1)/2] - 2) + [k/2]^2,$$

l'égalité n'étant vérifiée que pour les couples $(e, f) = ([k/2], [k/2])$ ou, si k est pair, $(e, f) = (k/2 + 1, k/2 - 1)$. On obtient

$$S_k(\lambda) \leq S_k(\underline{\lambda}) + \nu_k(\lambda) - \nu_k(\underline{\lambda}). \quad (13)$$

Supposons $S_k(\lambda) > S_k(\underline{\lambda})$. L'inégalité précédente force $\nu_k(\lambda) = 1$. Donc λ_k est pair, $k + S_k(\lambda)$ est impair et, d'après (8), $\lambda_{k+1} = \lambda_k$. Les entiers $k-1 + S_{k-1}(\lambda)$ et $k+1 + S_{k+1}(\lambda)$ sont pairs, donc $\nu_{k-1}(\lambda) = \nu_{k+1}(\lambda) = 0$. L'inégalité (13) entraîne donc $S_{k-1}(\lambda) \leq S_{k-1}(\underline{\lambda})$ et $S_{k+1}(\lambda) \leq S_{k+1}(\underline{\lambda})$ (pour être précis, si $k = 1$, notre calcul ne s'applique pas à $k-1$ mais, dans ce cas, l'inégalité $S_0(\lambda) \leq S_0(\underline{\lambda})$ est triviale). Les deux inégalités $S_{k-1}(\lambda) \leq S_{k-1}(\underline{\lambda})$ et $S_k(\lambda) > S_k(\underline{\lambda})$ entraînent $\lambda_k > \underline{\lambda}_k$. Donc $\lambda_{k+1} = \lambda_k > \underline{\lambda}_k \geq \underline{\lambda}_{k+1}$. Alors l'inégalité $S_k(\lambda) > S_k(\underline{\lambda})$ entraîne $S_{k+1}(\lambda) > S_{k+1}(\underline{\lambda})$, contrairement à ce que l'on a vu ci-dessus. Cette contradiction prouve l'inégalité $S_k(\lambda) \leq S_k(\underline{\lambda})$. Cela étant vrai pour tout k , on conclut $\lambda \leq \underline{\lambda}$, ce qui démontre le (i) de l'énoncé.

Supposons maintenant λ spéciale et $\lambda = \underline{\lambda}$. Considérons l'inégalité (12) pour k impair. Les termes relatifs à λ et $\underline{\lambda}$ s'annulent et il reste

$$e(e-2) + f^2 \leq ((k+1)/2)((k+1)/2 - 2) + ((k-1)/2)^2.$$

Comme on l'a dit, cela entraîne que $e = (k+1)/2$ et $f = (k-1)/2$. Parce que λ est spéciale, on calcule facilement les termes A^\sharp et B^\sharp . Puisque λ_1 est impair, le terme $\lambda_1 + 2n$ l'est aussi, donc c'est $2z'_1 + 1$. Pour $h = 1, \dots, n$, les termes λ_{2h} et λ_{2h+1} sont de même parité donc les termes $\lambda_{2h} + 2n + 1 - 2h$ et $\lambda_{2h+1} + 2n - 2h$ sont de parité opposée. Par récurrence, ce sont les termes $2z_h$ et $2z'_{h+1} + 1$. Cela permet le calcul des termes z_h et z'_{h+1} , puis des termes a_{h+1}^\sharp et b_h^\sharp . On obtient

$$\begin{aligned} a_1^\sharp &= (\lambda_1 - 1)/2 + 2n, \\ b_h^\sharp &= a_{h+1}^\sharp = \lambda_{2h}/2 + 2n - 2h && \text{pour } h = 1, \dots, n, \text{ si } \lambda_{2h} = \lambda_{2h+1} \text{ est pair,} \\ b_h^\sharp &= (\lambda_{2h} + 1)/2 + 2n - 2h && \text{si } \lambda_{2h} \text{ et } \lambda_{2h+1} \text{ sont impairs,} \\ a_{h+1}^\sharp &= (\lambda_{2h+1} - 1)/2 + 2n - 2h && \text{si } \lambda_{2h} \text{ et } \lambda_{2h+1} \text{ sont impairs.} \end{aligned}$$

Considérons l'intervalle maximal Δ_{\max} de λ . Notons ses éléments $i_1 > \dots > i_t$. Ils sont impairs. Les multiplicités de i_1, \dots, i_{t-1} sont paires et celle de i_t est impaire. On note ces multiplicités

$$2m_1, \dots, 2m_{t-1}, 2m_t + 1.$$

L'intervalle correspond aux éléments suivants de $A^\sharp \sqcup B^\sharp$:

$$a_1^\sharp > b_1^\sharp > \dots > a_{m_1}^\sharp > b_{m_1}^\sharp > a_{m_1+1}^\sharp > b_{m_1+1}^\sharp > \dots > a_{m_{\leq 2}}^\sharp > b_{m_{\leq 2}}^\sharp > \dots > a_{m_{\leq t-1}+1}^\sharp > b_{m_{\leq t-1}+1}^\sharp > \dots > b_{m_{\leq t}}^\sharp > a_{m_{\leq t}+1}^\sharp$$

(on rappelle que $m_{\leq i} = m_1 + \dots + m_i$). Supposons que ϵ ne vaut pas 1 sur Δ_{\max} . Soit s le plus petit élément de $\{1, \dots, t\}$ tel que $\epsilon(i_s) = -1$. Appliquons l'égalité (9) à $k = 2m_{\leq s-1} + 1$. Comme on l'a dit plus haut, on a $e = m_{\leq s-1} + 1$, $f = m_{\leq s-1}$. On obtient

$$a_1^\sharp + \dots + a_{m_{\leq s-1}+1}^\sharp + b_1^\sharp + \dots + b_{m_{\leq s-1}}^\sharp = a_1 + \dots + a_{m_{\leq s-1}+1} + b_1 + \dots + b_{m_{\leq s-1}}.$$

Par construction de A et B et par définition de s , les termes de ces ensembles sont égaux à ceux de A^\sharp et B^\sharp jusqu'à l'indice $m_{\leq s-1}$ et l'égalité précédente devient

$$a_{m_{\leq s-1}+1}^\sharp = a_{m_{\leq s-1}+1}.$$

Par contre, passer de (A^\sharp, B^\sharp) à (A, B) échange les termes correspondant à l'entier i_s . Hormis le cas $s = t$ et $m_t = 0$, on a donc $a_{m_{\leq s-1}+1} = b_{m_{\leq s-1}+1}^\sharp$. Quand $s = t$ et $m_t = 0$, $a_{m_{\leq s-1}+1}$ est un terme de la famille $A^\sharp \sqcup B^\sharp$ qui n'est pas dans l'ensemble écrit ci-dessus. Dans tous les cas, on obtient $a_{m_{\leq s-1}+1} < a_{m_{\leq s-1}+1}^\sharp$ ce qui contredit l'égalité de ces termes prouvée ci-dessus. Cette contradiction conclut : ϵ vaut 1 sur Δ_{\max} .

Considérons maintenant un intervalle $\Delta \neq \Delta_{\max}$. On note ses éléments $i_1 > \dots > i_t$. Le premier indice j tel que $\lambda_j = i_1$ est forcément pair. Notons le $2u$. Si $t = 1$, la multiplicité de i_1 est paire. On la note $2m$. Alors l'intervalle correspond aux éléments suivants de $A^\sharp \sqcup B^\sharp$:

$$b_u^\sharp < a_{u+1}^\sharp < \dots < b_{u+m-1}^\sharp < a_{u+m}^\sharp.$$

Si $t > 1$, les multiplicités de i_1 et i_t sont impaires et, pour $1 < s < t$, celle de i_s est paire. On les note respectivement $2m_1 + 1, 2m_t + 1, 2m_s$. Alors l'intervalle correspond aux éléments suivants de $A^\sharp \sqcup B^\sharp$:

$$b_u^\sharp < a_{u+1}^\sharp < \dots < b_{u+m_1}^\sharp < a_{u+m_1+1}^\sharp < \dots < b_{u+m_{\leq 2}}^\sharp < a_{u+m_{\leq 2}+1}^\sharp < \dots < b_{u+m_{\leq t-1}}^\sharp < a_{u+m_{\leq t-1}+1}^\sharp < \dots < a_{u+m_{\leq t}+1}^\sharp.$$

Supposons par récurrence que ϵ est constant sur tout intervalle strictement supérieur à Δ . Pour ces intervalles, ou bien on ne change pas les termes de A^\sharp et B^\sharp leur correspondant, ou bien on les échange. Mais on voit ci-dessus que chaque intervalle contribue autant à A_\sharp qu'à B_\sharp . Cela ne perturbe pas les numérotations des termes postérieurs, on veut dire par là que la contribution à A (resp. B) de l'intervalle Δ commence par a_{u+1} (resp. b_u). Si Δ est réduit à i_1 , on n'a rien à démontrer : ϵ est forcément constant sur Δ . Supposons $t > 1$ et que ϵ ne soit pas constant sur Δ . Notons s le plus petit élément de $\{2, \dots, t\}$ tel que $\epsilon_{i_{s-1}} \neq \epsilon_{i_s}$. Appliquons l'égalité (9) à $k = 2u - 1$ (on sait qu'alors $e = u$ et $f = u - 1$) et à $k = 2u + 2m_{\leq s-1} + 1$ (on sait qu'alors $e = u + m_{\leq s-1} + 1$ et $f = u + m_{\leq s-1}$). Par différence, on obtient

$$b_u^\sharp + a_{u+1}^\sharp + \dots + b_{u+m_{\leq s-1}}^\sharp + a_{u+m_{\leq s-1}+1}^\sharp = b_u + a_{u+1} + \dots + b_{u+m_{\leq s-1}} + a_{u+m_{\leq s-1}+1}. \quad (14)$$

Supposons d'abord $\epsilon(i_{s-1}) = 1$ et $\epsilon(i_s) = -1$. Les termes de l'ensemble A sont égaux à ceux de A^\sharp entre les indices $u + 1$ et $u + m_{\leq s-1}$. Les termes de l'ensemble B sont égaux à ceux de B^\sharp entre les indices u et $u + m_{\leq s-1}$. L'égalité (14) devient

$$a_{u+m_{\leq s-1}+1}^\sharp = a_{u+m_{\leq s-1}+1}.$$

Parce que $\epsilon_{i_s} = -1$, on échange les termes correspondant à l'entier i_s . Hormis le cas $s = t$ et $m_t = 0$, on a donc

$$a_{u+m_{\leq s-1}+1} = b_{u+m_{\leq s-1}+1}^\sharp.$$

Si $s = t$ et $m_t = 0$, $a_{u+m_{\leq s-1}+1}$ est un terme de la famille $A^\sharp \sqcup B^\sharp$ qui est au-delà de ceux écrits ci-dessus. Dans tous les cas, on obtient $a_{u+m_{\leq s-1}+1} < a_{u+m_{\leq s-1}+1}^\sharp$ ce qui contredit l'égalité de ces termes prouvée ci-dessus.

Supposons maintenant $\epsilon(i_{s-1}) = -1$ et $\epsilon(i_s) = 1$. Les entiers i_1, \dots, i_{s-1} contribuent à A et B en échangeant leur contribution à A^\sharp et B^\sharp . D'où

$$a_{u+1} = b_u^\sharp, \quad \dots, \quad a_{u+m_{\leq s-1}+1} = b_{u+m_{\leq s-1}}^\sharp, \quad b_u = a_{u+1}^\sharp, \quad \dots, \quad b_{u+m_{\leq s-1}-1} = a_{u+m_{\leq s-1}}^\sharp.$$

L'égalité (14) devient

$$b_{u+m_{\leq s-1}} = a_{u+m_{\leq s-1}+1}^\sharp.$$

Par contre, l'entier i_s contribue par les mêmes termes à A et A^\sharp comme à B et B^\sharp . Mais les indices sont décalés et on a

$$a_{u+m_{\leq s-1}+2} = a_{u+m_{\leq s-1}+1}^\sharp$$

et, hormis le cas $s = t$ et $m_t = 0$, $b_{u+m_{\leq s-1}} = b_{u+m_{\leq s-1}}^\sharp$. Si $s = t$ et $m_t = 0$, $b_{u+m_{\leq s-1}}$ est un terme de la famille $A^\sharp \sqcup B^\sharp$ qui est au-delà de ceux écrits ci-dessus. Dans tous les cas, on obtient $b_{u+m_{\leq s-1}} < a_{u+m_{\leq s-1}+1}^\sharp$ ce qui contredit l'égalité de ces termes prouvée ci-dessus.

Ces contradictions prouvent que ϵ est constant sur Δ . Cela prouve que, sous les hypothèses du (ii) de l'énoncé, la condition (a) implique (b).

Soit maintenant $(\lambda, \epsilon) \in \mathcal{P}^{\text{orth}}(2n+1)$, supposons λ spéciale et ϵ constant sur les intervalles de λ . En notant $i_1 > \dots > i_m$ les éléments de $\text{Jord}_{\text{bp}}(\lambda)$ intervenant avec multiplicité impaire, on a $\epsilon(i_{2h}) = \epsilon(i_{2h+1})$ pour tout $h = 1, \dots, (m-1)/2$. La recette indiquée plus haut pour calculer $k_{\lambda, \epsilon}$ montre que cet entier vaut 1. Relevons ϵ en l'élément de $\{\pm 1\}^{\text{Jord}_{\text{bp}}(\lambda)}$ qui vaut 1 sur le plus grand intervalle. On a calculé ci-dessus les termes A^\sharp, B^\sharp, A, B . Remarquons que les deux premiers sont aussi les termes \underline{A} et \underline{B} associés à $(\lambda, 1)$. On en déduit facilement les termes $\underline{X}, \underline{Y}, X$ et Y . On voit que les ensembles suivants contribuent de la même façon à \underline{X} et X comme à \underline{Y} et Y :

— l'intervalle maximal ; sa contribution est de la forme

$$x_1 = y_1 > x_2 = y_2 > \dots > x_u = y_u > x_{u+1};$$

— tout couple $\lambda_{2h}, \lambda_{2h+1}$ d'éléments pairs donc égaux ; sa contribution est de la forme $y_h = x_{h+1}$;

— tout intervalle non maximal sur lequel ϵ vaut 1 ; sa contribution est de la forme

$$y_u > x_{u+1} = y_{u+1} > \cdots > x_v = y_v > x_{v+1}.$$

Par contre, la contribution à \underline{X} et \underline{Y} d'un intervalle sur lequel ϵ vaut -1 est de la forme

$$\underline{y}_u > \underline{x}_{u+1} = \underline{y}_{u+1} > \cdots > \underline{x}_v = \underline{y}_v > \underline{x}_{v+1},$$

tandis que sa contribution à X et Y est

$$x_{u+1} = \underline{y}_u > y_u = \underline{x}_{u+1} = x_{u+2} = \underline{y}_{u+1} > \cdots > y_{v-1} = \underline{x}_v = x_{v+1} = \underline{y}_v > y_v = \underline{x}_{v+1}.$$

Il est immédiat que $X \sqcup Y = \underline{X} \sqcup \underline{Y}$, autrement dit les symboles (X, Y) et $(\underline{X}, \underline{Y})$ sont dans la même famille. Cela prouve que $\lambda = \mathrm{sp}(\lambda, \epsilon)$, donc que la relation (b) du (ii) de l'énoncé entraîne la relation (a).

Conservons les hypothèses sur (λ, ϵ) . On relève ϵ comme ci-dessus. Posons $(\tau, \delta) = \mathrm{fam} \circ \mathrm{symb}(\rho_{\lambda, \epsilon})$. En utilisant la description du symbole (X, Y) faite ci-dessus et la définition de l'application fam de [Waldspurger 2001, VIII.19], on calcule, pour tout intervalle Δ :

- $\delta(\Delta) = 0$;
- $\tau(\Delta) = 0$ si ϵ vaut 1 sur Δ et $\tau(\Delta) = 1$ si ϵ vaut -1 .

Le (iii) de l'énoncé s'en déduit. □

1.5. Correspondance de Springer, cas orthogonal pair. Soit $n \in \mathbb{N}$. On note $\mathcal{P}^{\mathrm{orth}}(2n)$ l'ensemble des partitions orthogonales de $2n$, c'est-à-dire les $\lambda \in \mathcal{P}(2n)$ telles que $\mathrm{mult}_\lambda(i)$ est pair pour tout entier i pair. Pour une telle partition, on note $\mathrm{Jord}_{\mathrm{bp}}(\lambda)$ l'ensemble des entiers $i \geq 1$ impairs tels que $\mathrm{mult}_\lambda(i) \geq 1$. Plus précisément, pour un entier $k \geq 1$, on note $\mathrm{Jord}_{\mathrm{bp}}^k(\lambda)$ l'ensemble des $i \geq 1$ impairs tels que $\mathrm{mult}_\lambda(i) = k$.

Pour $\lambda \in \mathcal{P}^{\mathrm{orth}}(2n)$, disons que λ est exceptionnelle si $\mathrm{Jord}_{\mathrm{bp}}(\lambda) = \emptyset$. Si $n > 0$, on introduit l'ensemble $\underline{\mathcal{P}}^{\mathrm{orth}}(2n)$ formé des partitions $\lambda \in \mathcal{P}^{\mathrm{orth}}(2n)$ non exceptionnelles et des paires $(\lambda, +)$ et $(\lambda, -)$ pour les partitions $\lambda \in \mathcal{P}^{\mathrm{orth}}(2n)$ exceptionnelles.

Justifions cette définition. Notons $\overline{\mathbb{F}}_q$ une clôture algébrique de \mathbb{F}_q et $O(2n)$ le groupe orthogonal évident sur $\overline{\mathbb{F}}_q$. L'ensemble $\mathcal{P}^{\mathrm{orth}}(2n)$ paramètre les classes de conjugaison unipotentes par $O(2n)(\overline{\mathbb{F}}_q)$ dans $\mathrm{SO}(2n)(\overline{\mathbb{F}}_q)$. Mais il arrive que de telles classes se coupent en deux classes de conjugaison par $\mathrm{SO}(2n)(\overline{\mathbb{F}}_q)$. Cela arrive précisément quand la classe est paramétrée par une partition λ exceptionnelle. Alors l'ensemble $\underline{\mathcal{P}}^{\mathrm{orth}}(2n)$ paramètre les classes de conjugaison unipotentes par $\mathrm{SO}(2n)(\overline{\mathbb{F}}_q)$ dans $\mathrm{SO}(2n)(\overline{\mathbb{F}}_q)$. Si $n = 0$, on pose $\underline{\mathcal{P}}^{\mathrm{orth}}(0) = \mathcal{P}^{\mathrm{orth}}(0) = \{\emptyset\}$. Il y a en tout cas une application évidente de $\underline{\mathcal{P}}^{\mathrm{orth}}(2n)$ dans $\mathcal{P}^{\mathrm{orth}}(2n)$. Si $\underline{\lambda}$ est un élément de $\underline{\mathcal{P}}^{\mathrm{orth}}(2n)$, on note sans plus de commentaire $\lambda \in \mathcal{P}^{\mathrm{orth}}(2n)$ son image.

On note $\mathcal{P}^{\mathrm{orth}}(2n)$ l'ensemble des couples (λ, ϵ) où $\lambda \in \mathcal{P}^{\mathrm{orth}}(2n)$ et $\epsilon \in (\{\pm 1\}^{\mathrm{Jord}_{\mathrm{bp}}(\lambda)})/\{\pm 1\}$, le groupe $\{\pm 1\}$ s'envoyant diagonalement dans $\{\pm 1\}^{\mathrm{Jord}_{\mathrm{bp}}(\lambda)}$. On note $\underline{\mathcal{P}}^{\mathrm{orth}}(2n)$ l'ensemble des couples $(\underline{\lambda}, \epsilon)$ où $\underline{\lambda} \in \underline{\mathcal{P}}^{\mathrm{orth}}(2n)$ et $\epsilon \in (\{\pm 1\}^{\mathrm{Jord}_{\mathrm{bp}}(\lambda)})/\{\pm 1\}$. La correspondance de Springer généralisée établit une bijection entre $\underline{\mathcal{P}}^{\mathrm{orth}}(2n)$ et l'ensemble des couples (ρ, k) tels que

- $k \in \mathbb{N}$, k est pair et $k^2 \leq 2n$;
- si $k > 0$, $\rho \in \widehat{W}_{n-k^2/2}$; si $k = 0$, $\rho \in \widehat{W}_n^D$.

On note $(\rho_{\lambda,\epsilon}, k_{\lambda,\epsilon})$ le couple associé à un élément $(\lambda, \epsilon) \in \mathcal{P}^{\text{orth}}(2n)$. L'entier $k_{\lambda,\epsilon}$ ne dépend que de l'image (λ, ϵ) de (λ, ϵ) dans $\mathcal{P}^{\text{orth}}(2n)$. Il se calcule de la façon suivante. Notons $i_1 > \dots > i_m$ les entiers impairs i tels que $\text{mult}_\lambda(i)$ soit impair. Posons

$$h = \sum_{j=1, \dots, m} (-1)^j (1 - \epsilon(i_j)).$$

Alors $k_{\lambda,\epsilon} = |h|$. En particulier, si $\epsilon = 1$, c'est-à-dire $\epsilon(i) = 1$ pour tout $i \in \text{Jord}_{\text{bp}}(\lambda)$, on a $k_{\lambda,1} = 0$ et $\rho_{\lambda,1} \in \widehat{W}_n^D$.

Une partition

$$\lambda = (\lambda_1 \geq \lambda_2 \geq \dots) \in \mathcal{P}^{\text{orth}}(2n)$$

est spéciale si et seulement si λ_{2j-1} et λ_{2j} sont de même parité pour tout $j \geq 1$. Cela équivaut à ce que ${}^t\lambda$ soit symplectique. Notons $\mathcal{P}^{\text{orth,sp}}(2n)$ l'ensemble des partitions orthogonales spéciales de $2n$. Considérons une partition $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n)$ et définissons $i_1 > \dots > i_m$ comme ci-dessus. L'entier m est forcément pair. On appelle intervalle de λ un ensemble Δ de l'une des formes suivantes :

- pour un entier $h = 1, \dots, m/2$, Δ est l'ensemble des $i \geq 1$ tels que $\text{mult}_\lambda(i) \geq 1$ et tels que $i_{2h-1} \geq i \geq i_{2h}$;
- $\Delta = \{i\}$ où i est un entier impair tel que $\text{mult}_\lambda(i) \geq 1$ et tel qu'il n'existe pas d'entier $h = 1, \dots, m/2$ de sorte que $i_{2h-1} \geq i \geq i_{2h}$.

Parce que λ est spéciale, on vérifie que les intervalles sont formés d'entiers impairs. Ils forment une partition de $\text{Jord}_{\text{bp}}(\lambda)$. On ordonne les intervalles comme dans le cas symplectique. On note Δ_{\min} (resp. Δ_{\max}) le plus petit (resp. grand) intervalle. On note $\text{Int}(\lambda)$ l'ensemble des intervalles.

Pour $(\lambda, \epsilon) \in \mathcal{P}^{\text{orth}}(2n)$, le symbole $\text{symb}(\rho_{\lambda,\epsilon})$ ne dépend que de λ , on le note abusivement $\text{symb}(\rho_{\lambda,\epsilon})$. L'application $\lambda \mapsto \text{symb}(\rho_{\lambda,1})$ est une bijection de $\mathcal{P}^{\text{orth,sp}}(2n)$ sur l'ensemble des symboles spéciaux de rang n et de défaut 0. Pour $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n)$, on a défini en [Waldspurger 2001, VIII.19] une bijection fam entre la famille du symbole $\text{symb}(\rho_{\lambda,1})$ et un certain sous-ensemble de $(\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \times (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)}$.

Soit $\lambda \in \mathcal{P}^{\text{orth}}(2n)$. Il existe une unique partition spéciale $\text{sp}(\lambda) \in \mathcal{P}^{\text{orth,sp}}(2n)$ telle que $\text{symb}(\rho_{\lambda,1})$ et $\text{symb}(\rho_{\text{sp}(\lambda),1})$ soient dans la même famille. Il est connu que $\lambda \leq \text{sp}(\lambda)$ et que $\text{sp}(\lambda)$ est la plus petite partition orthogonale spéciale λ' telle que $\lambda \leq \lambda'$. Plus généralement, on a le lemme suivant.

Lemme. (i) Soit $(\lambda, \epsilon) \in \mathcal{P}^{\text{orth}}(2n)$, supposons $k_{\lambda,\epsilon} = 0$. Notons $\text{sp}(\lambda, \epsilon)$ l'unique partition spéciale telle que $\text{symb}(\rho_{\lambda,\epsilon})$ et $\text{symb}(\rho_{\text{sp}(\lambda,\epsilon),1})$ soient dans la même famille. Alors $\lambda \leq \text{sp}(\lambda, \epsilon)$.

(ii) Soit $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n)$. Pour $\epsilon \in \{\pm 1\}^{\text{Jord}_{\text{bp}}(\lambda)}$, les conditions suivantes sont équivalentes :

- (a) $k_{\lambda,\epsilon} = 0$ et $\text{sp}(\lambda, \epsilon) = \lambda$;
- (b) ϵ est constant sur tout $\Delta \in \text{Int}(\lambda)$.

(iii) Soit $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n)$. L'application $\epsilon \mapsto \text{fam} \circ \text{symb}(\rho_{\lambda,\epsilon})$ est une bijection entre l'ensemble des ϵ décrits au (ii), modulo le groupe diagonal $\{\pm 1\}$, et le sous-ensemble des

$$(\tau, \delta) \in (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \times (\mathbb{Z}/2\mathbb{Z})^{\text{Int}(\lambda)} \quad \text{tels que } \delta = 0 \text{ et } \tau(\Delta_{\max}) = 0.$$

La preuve est similaire à celle du lemme précédent.

1.6. Dualité, cas symplectique-orthogonal impair. Soit $n \in \mathbb{N}$. Notons $\mathcal{S}_{n,1}^{\mathrm{sp}}$ l'ensemble des symboles spéciaux de rang n et de défaut impair (ce défaut est alors 1). On dispose de bijections

$$\mathcal{P}^{\mathrm{symp},\mathrm{sp}}(2n) \rightarrow \mathcal{S}_{n,1}^{\mathrm{sp}}, \quad \lambda \mapsto \mathrm{symb}(\rho_{\lambda,1}); \quad \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n+1) \rightarrow \mathcal{S}_{n,1}^{\mathrm{sp}}, \quad \lambda \mapsto \mathrm{symb}(\rho_{\lambda,1})$$

et d'une involution d de $\mathcal{S}_{n,1}^{\mathrm{sp}}$. On en déduit des bijections

$$d : \mathcal{P}^{\mathrm{symp},\mathrm{sp}}(2n) \rightarrow \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n+1) \quad \text{et} \quad d : \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n+1) \rightarrow \mathcal{P}^{\mathrm{symp},\mathrm{sp}}(2n)$$

inverses l'une de l'autre définies par la formule commune $\mathrm{symb}(\rho_{d(\lambda),1}) = d \circ \mathrm{symb}(\rho_{\lambda,1})$.

Soit $\lambda \in \mathcal{P}^{\mathrm{symp},\mathrm{sp}}(2n)$. On vérifie qu'il y a une unique bijection décroissante de $\mathrm{Int}(\lambda)$ sur $\mathrm{Int}(d(\lambda))$. Notons $\Delta_1 > \dots > \Delta_r$ les intervalles de λ et $\Delta'_1 > \dots > \Delta'_r$ ceux de $d(\lambda)$. On a dit que l'involution d des symboles échangeait les familles de $\mathrm{symb}(\rho_{\lambda,1})$ et de $\mathrm{symb}(\rho_{d(\lambda),1})$. D'autre part, ces familles sont paramétrées par des sous-ensembles de $(\mathbb{Z}/2\mathbb{Z})^{\mathrm{Int}(\lambda)} \times (\mathbb{Z}/2\mathbb{Z})^{\mathrm{Int}(\lambda)}$ et de $(\mathbb{Z}/2\mathbb{Z})^{\mathrm{Int}(d(\lambda))} \times (\mathbb{Z}/2\mathbb{Z})^{\mathrm{Int}(d(\lambda))}$, respectivement. Soit (X, Y) un symbole dans la famille de $\mathrm{symb}(\rho_{\lambda,1})$, notons $(\tau, \delta) = \mathrm{fam}(X, Y)$ et $(\tau', \delta') = \mathrm{fam} \circ d(X, Y)$. On vérifie les égalités

$$\tau'(\Delta'_h) = \tau(\Delta_{r+1-h}), \quad \delta'(\Delta'_h) = \delta(\Delta_{r-h})$$

pour tout $h = 1, \dots, r$, avec la convention $\delta(\Delta_0) = 0$. En particulier cette application échange l'ensemble des (τ, δ) tels que $\delta = 0$ et celui des (τ', δ') tels que $\delta' = 0$.

On a défini des applications $\mathrm{sp} : \mathcal{P}^{\mathrm{symp}}(2n) \rightarrow \mathcal{P}^{\mathrm{symp},\mathrm{sp}}(2n)$ et $\mathrm{sp} : \mathcal{P}^{\mathrm{orth}}(2n+1) \rightarrow \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n+1)$. On étend les bijections d en des applications encore notées $d : \mathcal{P}^{\mathrm{symp}}(2n) \rightarrow \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n+1)$ et $d : \mathcal{P}^{\mathrm{orth}}(2n+1) \rightarrow \mathcal{P}^{\mathrm{symp},\mathrm{sp}}(2n)$ par la formule commune $d(\lambda) = d \circ \mathrm{sp}(\lambda)$. Il est connu que ces applications sont décroissantes : pour $\lambda, \lambda' \in \mathcal{P}^{\mathrm{symp}}(2n)$ (resp. $\lambda, \lambda' \in \mathcal{P}^{\mathrm{orth}}(2n+1)$), $\lambda \leq \lambda'$ entraîne $d(\lambda') \leq d(\lambda)$.

Soit $\lambda \in \mathcal{P}^{\mathrm{symp}}(2n)$. On vérifie que $d(\lambda)$ est la plus grande partition $\mu \in \mathcal{P}^{\mathrm{orth}}(2n+1)$ telle que $\mu \leq {}^t(\lambda \cup \{1\})$, cf. [Mœglin et Renard 2017, paragraphe 7].

Soit $\mu \in \mathcal{P}^{\mathrm{orth}}(2n+1)$. Écrivons $\mu = (\mu_1 = \dots = \mu_s > \mu_{s+1} \geq \dots)$. Posons

$$\mu' = (\mu_1 = \dots = \mu_{s-1} \geq \mu_s - 1 \geq \mu_{s+1} \geq \dots).$$

On vérifie que $d(\mu)$ est la plus grande partition $\lambda \in \mathcal{P}^{\mathrm{symp}}(2n)$ telle que $\lambda \leq {}^t\mu'$, cf. [Mœglin et Renard 2017, paragraphe 7].

Soit $\lambda \in \mathcal{P}^{\mathrm{symp},\mathrm{sp}}(2n)$ (resp. $\lambda \in \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n+1)$). On a défini l'ensemble $\mathrm{Int}(\lambda)$. Soit $\Delta \in \mathrm{Int}(\lambda)$. On note $J(\Delta)$ l'ensemble des indices $j \geq 1$ tels que $\lambda_j \in \Delta$. Hormis les cas particuliers ci-dessous, on note $j_{\min}(\Delta)$ (resp. $j_{\max}(\Delta)$) le plus petit (resp. grand) élément de $J(\Delta)$. Les cas particuliers sont : λ symplectique et $\Delta = \Delta_{\min}$, auquel cas on pose $j_{\max}(\Delta) = \infty$; λ orthogonal et $\Delta = \Delta_{\max}$, auquel cas $j_{\min}(\Delta)$ n'est pas défini (plus exactement, on peut le définir en appliquant la définition ci-dessus, on obtient $j_{\min}(\Delta_{\max}) = 1$, mais cette valeur perturberait nos calculs et on considère que $j_{\min}(\Delta_{\max})$ n'est pas défini). Remarquons que, si $\lambda \in \mathcal{P}^{\mathrm{symp},\mathrm{sp}}(2n)$, les $j_{\min}(\Delta)$ sont impairs et les $j_{\max}(\Delta)$ sont pairs (ou ∞); si $\lambda \in \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n+1)$, les $j_{\min}(\Delta)$ sont pairs et les $j_{\max}(\Delta)$ sont impairs.

On définit une suite de nombres $\zeta(\lambda) = (\zeta(\lambda)_1, \zeta(\lambda)_2, \dots)$ par

$$\zeta(\lambda)_j = \begin{cases} 1 & \text{s'il existe } \Delta \in \text{Int}(\lambda) \text{ tel que } j = j_{\min}(\Delta), \\ -1 & \text{s'il existe } \Delta \in \text{Int}(\lambda) \text{ tel que } j = j_{\max}(\Delta), \\ 0 & \text{dans les autres cas.} \end{cases}$$

Lemme. Soit $\lambda \in \mathcal{P}^{\text{symp,sp}}(2n)$ (resp. $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n+1)$). On a l'égalité

$${}^t d(\lambda) = \lambda + \zeta(\lambda).$$

Preuve. On suppose $\lambda \in \mathcal{P}^{\text{symp,sp}}(2n)$, la preuve étant similaire dans le cas orthogonal. Montrons d'abord

(15) ${}^t d(\lambda)$ est la plus petite partition orthogonale spéciale ν de $2n+1$ telle que $\nu \geq \lambda \cup \{1\}$.

Puisque $d(\lambda)$ est orthogonale et spéciale, sa transposée l'est également. L'inégalité $d(\lambda) \leq {}^t(\lambda \cup \{1\})$ entraîne ${}^t d(\lambda) \geq \lambda \cup \{1\}$. Inversement, soit ν une partition orthogonale spéciale de $2n+1$ telle que $\nu \geq \lambda \cup \{1\}$. Alors ${}^t \nu$ est encore orthogonale et vérifie ${}^t \nu \leq {}^t(\lambda \cup \{1\})$. Donc ${}^t \nu \leq d(\lambda)$ puis $\nu \geq {}^t d(\lambda)$. Cela démontre (15).

On vérifie facilement que $\lambda + \zeta(\lambda)$ est une partition, c'est-à-dire $\lambda_j + \zeta(\lambda)_j \geq \lambda_{j+1} + \zeta(\lambda)_{j+1}$ pour tout $j \geq 1$. Puisque tout intervalle $\Delta \neq \Delta_{\min}$ crée un terme $j_{\min}(\Delta)$ pour lequel $\zeta(\lambda)_{j_{\min}(\Delta)} = 1$ et un terme $j_{\max}(\Delta)$ pour lequel $\zeta(\lambda)_{j_{\max}(\Delta)} = -1$ et puisque le dernier intervalle Δ_{\min} crée seulement un $j_{\min}(\Delta_{\min})$ la somme totale des $\zeta(\lambda)_j$ vaut 1 et $\lambda + \zeta(\lambda) \in \mathcal{P}(2n+1)$. Si λ_1 est impair, λ_1 n'est pas dans un intervalle et $\zeta(\lambda)_1 = 0$ donc $\lambda_1 + \zeta(\lambda)_1$ est impair. Si λ_1 est pair, il appartient à un intervalle Δ (le plus grand intervalle). On a $j_{\min}(\Delta) = 1$, d'où $\zeta(\lambda)_1 = 1$ et $\lambda_1 + \zeta(\lambda)_1$ est encore impair. Considérons un entier $h \geq 1$ et distinguons les cas :

- λ_{2h} et λ_{2h+1} sont impairs. Comme ci-dessus, on a alors $\zeta(\lambda)_{2h} = \zeta(\lambda)_{2h+1} = 0$ et les termes $\lambda_{2h} + \zeta(\lambda)_{2h}$ et $\lambda_{2h+1} + \zeta(\lambda)_{2h+1}$ sont impairs.
- λ_{2h} est impair et λ_{2h+1} est pair. Dans ce cas $\zeta(\lambda)_{2h} = 0$ mais λ_{2h+1} appartient à un intervalle Δ tel que $j_{\min}(\Delta) = 2h+1$, donc $\zeta(\lambda)_{2h+1} = 1$; les termes $\lambda_{2h} + \zeta(\lambda)_{2h}$ et $\lambda_{2h+1} + \zeta(\lambda)_{2h+1}$ sont impairs.
- λ_{2h} est pair et λ_{2h+1} est impair. Dans ce cas $\zeta(\lambda)_{2h+1} = 0$ mais λ_{2h} appartient à un intervalle Δ tel que $j_{\max}(\Delta) = 2h$, donc $\zeta(\lambda)_{2h} = -1$; les termes $\lambda_{2h} + \zeta(\lambda)_{2h}$ et $\lambda_{2h+1} + \zeta(\lambda)_{2h+1}$ sont impairs.
- λ_{2h} et λ_{2h+1} sont pairs et distincts. Dans ce cas, λ_{2h} appartient à un intervalle Δ tel que $j_{\max}(\Delta) = 2h$ et λ_{2h+1} appartient à l'intervalle suivant Δ' tel que $j_{\min}(\Delta') = 2h+1$; on a $\zeta(\lambda)_{2h} = -1$ et $\zeta(\lambda)_{2h+1} = 1$; les termes $\lambda_{2h} + \zeta(\lambda)_{2h}$ et $\lambda_{2h+1} + \zeta(\lambda)_{2h+1}$ sont impairs.
- λ_{2h} et λ_{2h+1} sont pairs et égaux. Dans ce cas, $\lambda_{2h} = \lambda_{2h+1}$ appartient à un intervalle Δ tel que $j_{\min}(\Delta) < 2h < 2h+1 < j_{\max}(\Delta)$ et $\zeta(\lambda)_{2h} = \zeta(\lambda)_{2h+1} = 0$; les termes $\lambda_{2h} + \zeta(\lambda)_{2h}$ et $\lambda_{2h+1} + \zeta(\lambda)_{2h+1}$ sont pairs et égaux.

Cela montre d'abord que les termes pairs de la partition $\lambda + \zeta(\lambda)$ interviennent par paires, donc sont de multiplicité paire, c'est-à-dire que $\lambda + \zeta(\lambda)$ est orthogonale. Cela montre ensuite que deux termes $\lambda_{2h} + \zeta(\lambda)_{2h}$ et $\lambda_{2h+1} + \zeta(\lambda)_{2h+1}$ sont de la même parité. Donc $\lambda + \zeta(\lambda)$ est spéciale.

Pour $k \geq 1$, on voit que $S_k(\zeta(\lambda))$ vaut 1 s'il existe $\Delta \in \mathrm{Int}(\lambda)$ tel que $j_{\min}(\Delta) \leq k < j_{\max}(\Delta)$ et vaut 0 sinon. Écrivons $\lambda = (\lambda_1 \geq \dots \geq \lambda_l > 0)$. Alors $\lambda \cup \{1\} = (\lambda_1, \dots, \lambda_l, 1, 0)$. Si $k \leq l$, on a

$$S_k(\lambda + \zeta(\lambda)) = S_k(\lambda) + S_k(\zeta(\lambda)) \geq S_k(\lambda) = S_k(\lambda \cup \{1\}).$$

Si $k \geq l + 1$, on a $k \geq j_{\min}(\Delta_{\min})$ et $S_k(\zeta(\lambda)) = 1$. Le même calcul conduit à l'égalité

$$S_k(\lambda + \zeta(\lambda)) = S_k(\lambda \cup \{1\}).$$

Donc $\lambda + \zeta(\lambda) \geq \lambda \cup \{1\}$.

Soit maintenant ν une partition orthogonale spéciale de $2n + 1$ telle que $\nu \geq \lambda \cup \{1\}$. Soit $k \geq 1$. On a $S_k(\nu) \geq S_k(\lambda \cup \{1\}) \geq S_k(\lambda)$. Supposons que $S_k(\nu) < S_k(\lambda + \zeta(\lambda))$. Alors $S_k(\nu) = S_k(\lambda)$ et $S_k(\zeta(\lambda)) = 1$. Donc il existe un intervalle Δ tel que $j_{\min}(\Delta) \leq k < j_{\max}(\Delta)$. On vérifie alors que $S_k(\lambda)$ est pair. Supposons de plus k impair. Alors $S_k(\nu)$ est impair parce que ν est spéciale. L'égalité $S_k(\nu) = S_k(\lambda)$ est contradictoire. Cela démontre que, pour k impair, $S_k(\nu) \geq S_k(\lambda + \zeta(\lambda))$. Supposons maintenant que k est pair. Les inégalités $j_{\min}(\Delta) \leq k < j_{\max}(\Delta)$ et le fait que $j_{\min}(\Delta)$ est impair tandis que $j_{\max}(\Delta)$ est pair ou infini entraînent que $j_{\min}(\Delta) \leq k - 1 < j_{\max}(\Delta)$ et $j_{\min}(\Delta) < k + 1 < j_{\max}(\Delta)$. D'après ce que l'on vient de démontrer, on a $S_{k-1}(\nu) \geq S_{k-1}(\lambda + \zeta(\lambda)) = S_{k-1}(\lambda) + 1$. Avec l'égalité $S_k(\nu) = S_k(\lambda)$, cela entraîne $\nu_k < \lambda_k$. D'autre part, λ_k et λ_{k+1} sont dans un même intervalle. L'entier k étant pair, cela entraîne $\lambda_{k+1} = \lambda_k$, donc $\nu_{k+1} \leq \nu_k < \lambda_k = \lambda_{k+1}$. Avec l'égalité $S_k(\nu) = S_k(\lambda)$, cela entraîne $S_{k+1}(\nu) < S_{k+1}(\lambda)$, ce qui contredit l'hypothèse $\nu \geq \lambda \cup \{1\}$. Cette contradiction démontre encore l'inégalité $S_k(\nu) \geq S_k(\lambda + \zeta(\lambda))$. Celle-ci est donc vraie pour tout k , d'où $\nu \geq \lambda + \zeta(\lambda)$.

On a donc prouvé que $\lambda + \zeta(\lambda)$ était la plus petite partition orthogonale spéciale ν de $2n + 1$ telle que $\nu \geq \lambda \cup \{1\}$. Le lemme résulte alors de (15). \square

1.7. Dualité, cas orthogonal pair. Soit $n \in \mathbb{N}$. Notons $\mathcal{S}_{n,0}^{\mathrm{sp}}$ l'ensemble des symboles spéciaux de rang n et de défaut pair (ce défaut est alors 0). On dispose d'une bijection

$$\mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n) \rightarrow \mathcal{S}_{n,0}^{\mathrm{sp}}, \quad \lambda \mapsto \mathrm{symb}(\rho_{\lambda,1})$$

et d'une involution d de $\mathcal{S}_{n,0}^{\mathrm{sp}}$. On en déduit une involution $d : \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n) \rightarrow \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n)$ définie par la formule $\mathrm{symb}(\rho_{d(\lambda),1}) = d \circ \mathrm{symb}(\rho_{\lambda,1})$.

Soit $\lambda \in \mathcal{P}^{\mathrm{orth},\mathrm{sp}}(2n)$. On vérifie qu'il y a une unique bijection décroissante de $\mathrm{Int}(\lambda)$ sur $\mathrm{Int}(d(\lambda))$. Notons $\Delta_1 > \dots > \Delta_r$ les intervalles de λ et $\Delta'_1 > \dots > \Delta'_r$ ceux de $d(\lambda)$. On a dit que l'involution d des symboles échangeait les familles de $\mathrm{symb}(\rho_{\lambda,1})$ et de $\mathrm{symb}(\rho_{d(\lambda),1})$. D'autre part, ces familles sont paramétrées par des sous-ensembles de $(\mathbb{Z}/2\mathbb{Z})^{\mathrm{Int}(\lambda)} \times (\mathbb{Z}/2\mathbb{Z})^{\mathrm{Int}(\lambda)}$ et de $(\mathbb{Z}/2\mathbb{Z})^{\mathrm{Int}(d(\lambda))} \times (\mathbb{Z}/2\mathbb{Z})^{\mathrm{Int}(d(\lambda))}$, respectivement. Soit (X, Y) un symbole dans la famille de $\mathrm{symb}(\rho_{\lambda,1})$, notons $(\tau, \delta) = \mathrm{fam}(X, Y)$ et $(\tau', \delta') = \mathrm{fam} \circ d(X, Y)$. On vérifie les égalités suivantes, pour tout $h = 1, \dots, r$:

$$\delta'(\Delta'_h) = \delta(\Delta_{r-h}),$$

avec la convention $\delta(\Delta_0) = 0$;

— si le défaut de (X, Y) est strictement positif,

$$\tau'(\Delta'_h) = \tau(\Delta_{r+1-h});$$

— si ce défaut est nul,

$$\tau'(\Delta'_h) = \tau(\Delta_{r+1-h}) - \tau(\Delta_1).$$

En particulier cette application échange l'ensemble des (τ, δ) tels que $\delta = 0$ et celui des (τ', δ') tels que $\delta' = 0$.

On a défini l'application $\text{sp} : \mathcal{P}^{\text{orth}}(2n) \rightarrow \mathcal{P}^{\text{orth,sp}}(2n)$. On étend l'involution d en une application encore notée $d : \mathcal{P}^{\text{orth}}(2n) \rightarrow \mathcal{P}^{\text{orth,sp}}(2n)$ par la formule $d(\lambda) = d \circ \text{sp}(\lambda)$. Il est connu que cette application est décroissante : pour $\lambda, \lambda' \in \mathcal{P}^{\text{orth}}(2n)$, $\lambda \leq \lambda'$ entraîne $d(\lambda') \leq d(\lambda)$.

Soit $\lambda \in \mathcal{P}^{\text{orth}}(2n)$. On vérifie que $d(\lambda)$ est la plus grande partition $\mu \in \mathcal{P}^{\text{orth}}(2n)$ telle que $\mu \leq {}^t\lambda$, cf. [Mœglin et Renard 2017, paragraphe 7].

Soit $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n)$. On a défini l'ensemble $\text{Int}(\lambda)$. Soit $\Delta \in \text{Int}(\lambda)$. On note $J(\Delta)$ l'ensemble des indices $j \geq 1$ tels que $\lambda_j \in \Delta$. On note $j_{\min}(\Delta)$ (resp. $j_{\max}(\Delta)$) le plus petit (resp. grand) élément de $J(\Delta)$. Le nombre $j_{\min}(\Delta)$ (resp. $j_{\max}(\Delta)$) est impair (resp. pair). On définit une suite de nombres $\zeta(\lambda) = (\zeta(\lambda)_1, \zeta(\lambda)_2, \dots)$ par

$$\zeta(\lambda)_j = \begin{cases} 1 & \text{si il existe } \Delta \in \text{Int}(\lambda) \text{ tel que } j = j_{\min}(\Delta), \\ -1 & \text{si il existe } \Delta \in \text{Int}(\lambda) \text{ tel que } j = j_{\max}(\Delta), \\ 0 & \text{dans les autres cas.} \end{cases}$$

Lemme. Soit $\lambda \in \mathcal{P}^{\text{orth,sp}}(2n)$. On a l'égalité ${}^t d(\lambda) = \lambda + \zeta(\lambda)$.

La preuve est similaire à celle du lemme précédent.

1.8. Dualité et induction. Considérons une famille $\mathbf{n} = (n_1, \dots, n_t, n_0)$ d'entiers positifs ou nuls. Posons $n = \sum_{j=0, \dots, t} n_j$. Posons

$$\mathcal{P}^{\text{orth}}(\mathbf{n}) = \mathcal{P}(n_1) \times \dots \times \mathcal{P}(n_t) \times \mathcal{P}^{\text{orth}}(2n_0 + 1), \quad \mathcal{P}^{\text{symp}}(\mathbf{n}) = \mathcal{P}(n_1) \times \dots \times \mathcal{P}(n_t) \times \mathcal{P}^{\text{symp}}(2n_0).$$

On définit une opération d'induction

$$\mathcal{P}^{\text{orth}}(\mathbf{n}) \rightarrow \mathcal{P}^{\text{orth}}(2n + 1), \quad \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_t, \lambda_0) \mapsto \text{ind}(\boldsymbol{\lambda})$$

de la façon suivante : $\text{ind}(\boldsymbol{\lambda})$ est la plus grande partition orthogonale λ telle que

$$\lambda \leq (\lambda_1 + \lambda_1) + \dots + (\lambda_t + \lambda_t) + \lambda_0.$$

L'ensemble $\mathcal{P}^{\text{orth}}(\mathbf{n})$ étant le produit d'ensembles ordonnés, il l'est aussi par l'ordre produit. On vérifie que l'application d'induction est strictement croissante.

On définit l'application

$$\text{cup} : \mathcal{P}^{\text{symp}}(\mathbf{n}) \rightarrow \mathcal{P}^{\text{symp}}(2n)$$

par la formule

$$\text{cup}(\lambda_1, \dots, \lambda_t, \lambda_0) = (\lambda_1 \cup \lambda_1) \cup \dots \cup (\lambda_t \cup \lambda_t) \cup \lambda_0.$$

On définit enfin une dualité $d : \mathcal{P}^{\text{symp}}(\mathbf{n}) \rightarrow \mathcal{P}^{\text{orth}}(\mathbf{n})$. C'est le produit des applications $\lambda \mapsto {}^t\lambda$ sur chaque facteur $\mathcal{P}(n_i)$ pour $i = 1, \dots, t$ et de la dualité $d : \mathcal{P}^{\text{symp}}(2n_0) \rightarrow \mathcal{P}^{\text{orth,sp}}(2n_0+1) \subset \mathcal{P}^{\text{orth}}(2n_0+1)$. On a alors (Cf. [Barbasch et Vogan 1985, corollaire A.4]) :

Lemme. *Pour $\lambda = (\lambda_1, \dots, \lambda_t, \lambda_0) \in \mathcal{P}^{\text{symp}}(\mathbf{n})$, on a l'égalité $d \circ \text{cup}(\lambda) = \text{ind} \circ d(\lambda)$.*

1.9. Induction endoscopique. Soient $n_1, n_2 \in \mathbb{N}$, posons $n = n_1 + n_2$. Soient $\lambda_1 \in \mathcal{P}^{\text{symp,sp}}(2n_1)$ et $\lambda_2 \in \mathcal{P}^{\text{orth,sp}}(2n_2)$. Rappelons la définition de l'induite endoscopique $\text{ind}(\lambda_1, \lambda_2) \in \mathcal{P}^{\text{symp}}(2n)$, cf. [Waldspurger 2001, XI.6]. On note J^+ l'ensemble des entiers $j \geq 1$ tels que

$\lambda_{1,j}$ est pair, $\lambda_{2,j}$ est impair et il existe $\Delta \in \text{Int}(\lambda_1) \cup \text{Int}(\lambda_2)$ de sorte que $j = j_{\min}(\Delta)$ (cela entraîne que j est impair).

On note J^- l'ensemble des entiers $j \geq 1$ tels que

$\lambda_{1,j}$ est pair, $\lambda_{2,j}$ est impair et il existe $\Delta \in \text{Int}(\lambda_1) \cup \text{Int}(\lambda_2)$ de sorte que $j = j_{\max}(\Delta)$ (cela entraîne que j est pair).

On vérifie que J^+ et J^- ont même nombre d'éléments et que, si on note leurs éléments $j_1^+ < \dots < j_r^+$ et $j_1^- < \dots < j_r^-$, on a

$$j_1^+ < j_1^- < j_2^+ < j_2^- < \dots < j_r^+ < j_r^-.$$

On note $\xi = (\xi_1, \xi_2, \dots)$ la famille définie par $\xi_j = 1$ si $j \in J^+$, $\xi_j = -1$ si $j \in J^-$ et $\xi_j = 0$ pour $j \geq 1$ tel que $j \notin J^+ \cup J^-$. Alors $\text{ind}(\lambda_1, \lambda_2) = \lambda_1 + \lambda_2 + \xi$.

Proposition. $d(\lambda_1) \cup d(\lambda_2) \leq d(\text{ind}(\lambda_1, \lambda_2))$.

Preuve. Les deux membres de l'inégalité à prouver sont des partitions de $2n+1$. Les partitions $d(\lambda_1)$ et $d(\lambda_2)$ sont orthogonales, leur réunion l'est aussi. D'après la caractérisation de $d(\text{ind}(\lambda_1, \lambda_2))$ donnée en 1.6, il suffit de prouver l'inégalité

$$d(\lambda_1) \cup d(\lambda_2) \leq {}^t(\text{ind}(\lambda_1, \lambda_2) \cup \{1\}), \quad \text{ou encore} \quad {}^t d(\lambda_1) + {}^t d(\lambda_2) \geq \text{ind}(\lambda_1, \lambda_2) \cup \{1\},$$

ou encore, d'après les lemmes 1.6 et 1.7 et la définition ci-dessus,

$$\lambda_1 + \zeta(\lambda_1) + \lambda_2 + \zeta(\lambda_2) \geq (\lambda_1 + \lambda_2 + \xi) \cup \{1\}.$$

Cette inégalité se traduit par les inégalités suivantes, pour tout $k \geq 1$:

$$S_k(\zeta(\lambda_1)) + S_k(\zeta(\lambda_2)) \geq S_k(\xi), \quad \text{si } k \leq l(\text{ind}(\lambda_1, \lambda_2)); \quad (16)$$

$$S_k(\zeta(\lambda_1)) + S_k(\zeta(\lambda_2)) \geq S_k(\xi) + 1, \quad \text{si } k > l(\text{ind}(\lambda_1, \lambda_2)). \quad (17)$$

Les entiers $S_k(\zeta(\lambda_1))$, $S_k(\zeta(\lambda_2))$ et $S_k(\xi)$ valent toujours 0 ou 1. L'inégalité (16) est donc vérifiée si $S_k(\xi) = 0$. Supposons $S_k(\xi) = 1$. Avec les notations introduites plus haut, il existe alors $s \in \{1, \dots, r\}$

tel que $j_s^+ \leq k < j_s^-$. L'entier λ_{1,j_s^-} est pair et l'entier λ_{2,j_s^-} est impair. Il existe donc $\Delta_1 \in \text{Int}(\lambda_1)$ et $\Delta_2 \in \text{Int}(\lambda_2)$ tels que $\lambda_{1,j_s^-} \in \Delta_1$ et $\lambda_{2,j_s^-} \in \Delta_2$. Posons $u = \max(j_{\min}(\Delta_1), j_{\min}(\Delta_2))$. Alors $\lambda_{1,u} \in \Delta_1$ est pair et $\lambda_{2,u} \in \Delta_2$ est impair. En appliquant la définition de J^+ , on voit que $u \in J^+$. On a aussi $u \leq j_s^-$, donc $u \leq j_s^+$. On a alors $j_{\min}(\Delta_2) \leq u \leq j_s^+ \leq k < j_s^- \leq j_{\max}(\Delta_2)$ et $j_{\min}(\Delta_1) \leq u \leq j_s^+ \leq k < j_s^- \leq j_{\max}(\Delta_1)$. Ces inégalités entraînent $S_k(\zeta(\lambda_1)) = S_k(\zeta(\lambda_2)) = 1$. On a alors l'égalité

$$S_k(\zeta(\lambda_1)) + S_k(\zeta(\lambda_2)) = 2 = S_k(\xi) + 1,$$

qui est plus forte que (16). Cela prouve cette inégalité (16).

Supposons $k > l(\text{ind}(\lambda_1, \lambda_2))$. Si $S_k(\xi) = 1$, on vient de voir que l'inégalité (17) est vérifiée (et que c'est une égalité). On peut donc supposer $S_k(\xi) = 0$ et il suffit de montrer que $S_k(\zeta(\lambda_1)) = 1$. Puisque $k > l(\text{ind}(\lambda_1, \lambda_2))$, on a $\lambda_{1,k} + \lambda_{2,k} + \xi_k = 0$. Si $\xi_k \neq -1$, cela force $\lambda_{1,k} = 0$. Si $\xi_k = -1$, alors d'une part $\lambda_{1,k} \leq 1$, d'autre part $k \in J^-$. Donc $\lambda_{1,k}$ est pair et on a encore $\lambda_{1,k} = 0$. Donc $k \in J(\Delta_{1,\min})$, où $\Delta_{1,\min}$ est le plus petit élément de $\text{Int}(\lambda_1)$. Cela entraîne $S_k(\zeta(\lambda_1)) = 1$, ce qui achève la démonstration. \square

1.10. Intervalles relatifs, induction endoscopique régulière. On conserve les données n_1, n_2, λ_1 et λ_2 . On pose $\lambda = \text{ind}(\lambda_1, \lambda_2)$.

On a défini en [Waldspurger 2001, XI.11] un ensemble d'intervalles de λ . La terminologie est mal choisie car il se peut que λ soit spéciale et que cet ensemble ne soit pas celui défini en 1.3 ci-dessus. Nous appellerons ici intervalles relatifs (à λ_1 et λ_2) ces nouveaux intervalles. Rappelons leur définition. On pose

$$\begin{aligned} \mathcal{J} &= \{j_{\min}(\Delta); \Delta \in \text{Int}(\lambda_1) \cup \text{Int}(\lambda_2)\} \cup \{j_{\max}(\Delta); \Delta \in \text{Int}(\lambda_1) \cup \text{Int}(\lambda_2)\}, \\ \mathcal{J}^+ &= \{j_{\min}(\Delta); \Delta \in \text{Int}(\lambda_1)\} \cap \{j_{\min}(\Delta); \Delta \in \text{Int}(\lambda_2)\}, \\ \mathcal{J}^- &= \{j_{\max}(\Delta); \Delta \in \text{Int}(\lambda_1)\} \cap \{j_{\max}(\Delta); \Delta \in \text{Int}(\lambda_2)\}. \end{aligned}$$

Remarquons que \mathcal{J} contient ∞ qui est $j_{\max}(\Delta)$ pour le plus petit $\Delta \in \text{Int}(\lambda_1)$. Appelons intervalle relatif d'indices tout intervalle d'entiers $\{j, \dots, j'\}$ (avec éventuellement $j' = \infty$) vérifiant l'une des conditions suivantes :

$$(18) \quad j = j' \in \mathcal{J}^+ \cup \mathcal{J}^-;$$

$$(19) \quad j < j', \quad j \text{ et } j' \text{ sont deux termes consécutifs de } \mathcal{J} \text{ et il existe un unique } d \in \{1, 2\} \text{ et un unique } \Delta_d \in \text{Int}(\lambda_d) \text{ de sorte que } j_{\min}(\Delta_d) \leq j < j' \leq j_{\max}(\Delta_d).$$

Pour tout tel intervalle relatif d'indices $J = \{j, \dots, j'\}$, on pose $D(J) = \{\lambda_{j''}; j \leq j'' \leq j'\}$. On appelle intervalle relatif un tel ensemble $D(J)$. Inversement, pour un intervalle relatif D , on note $J(D)$ l'intervalle relatif d'indices J dont il provient et on note $j_{\min}(D)$ (resp. $j_{\max}(D)$) le plus petit (resp. grand) terme de $J(D)$. On note $\text{Int}_{\lambda_1, \lambda_2}(\lambda)$ l'ensemble de ces intervalles relatifs. On montre que cet ensemble d'intervalles relatifs forme une partition de $\text{Jord}_{\text{bp}}(\lambda) \cup \{0\}$.

Remarque. Comme on l'a dit ci-dessus, la définition des intervalles relatifs n'est pas la même que celle des intervalles d'une partition spéciale donnée en 1.3. Ces deux types d'intervalles n'ont pas les mêmes propriétés. En particulier, si Δ est un intervalle d'une partition symplectique spéciale, $j_{\min}(\Delta)$ est impair

et $j_{\max}(\Delta)$ est pair (ou ∞). Tandis que, pour un intervalle relatif D comme ci-dessus, $j_{\min}(D)$ et $j_{\max}(D)$ sont de parité quelconque.

On dit que λ_1 et λ_2 induisent régulièrement λ si et seulement si tout intervalle relatif est réduit à un élément. Autrement dit, $\text{Int}_{\lambda_1, \lambda_2}(\lambda)$ est la partition maximale de $\text{Jord}_{\text{bp}}(\lambda) \cup \{0\}$.

Supposons que λ_1 et λ_2 induisent régulièrement λ . On définit alors une fonction $\tau_{\lambda_1, \lambda_2} : \text{Jord}_{\text{bp}}(\lambda) \rightarrow \mathbb{Z}/2\mathbb{Z}$ de la façon suivante. Soit $i \in \text{Jord}_{\text{bp}}(\lambda)$. L'ensemble $\{i\}$ est un intervalle relatif. Remarquons que $\text{mult}_{\lambda}(i) = 1$ si et seulement si $J(\{i\})$ n'a qu'un élément, autrement dit $J(\{i\})$ est du type (18). Si $\text{mult}_{\lambda}(i) = 1$, on pose $\tau_{\lambda_1, \lambda_2}(i) = 0$. Si $\text{mult}_{\lambda}(i) \geq 2$, $J(\{i\})$ est du type (19) et on note $d(i) \in \{1, 2\}$ l'indice tel qu'il existe $\Delta_{d(i)} \in \text{Int}(\lambda_{d(i)})$ de sorte que $J(\{i\}) \subset \{j_{\min}(\Delta_{d(i)}), \dots, j_{\max}(\Delta_{d(i)})\}$. On pose $\tau_{\lambda_1, \lambda_2}(i) = d(i) + 1 \pmod{2\mathbb{Z}}$.

1.11. Une proposition d'existence. Soient $n \in \mathbb{N}$ et $\lambda \in \mathcal{P}^{\text{symp}}(2n)$. On se limite ici au cas où tous les termes de λ sont pairs. En particulier, λ est spéciale. Fixons une fonction $\tau : \text{Jord}_{\text{bp}}(\lambda) \rightarrow \mathbb{Z}/2\mathbb{Z}$ telle que $\tau(i) = 0$ pour tout $i \in \text{Jord}_{\text{bp}}(\lambda)$ tel que $\text{mult}_{\lambda}(i) = 1$.

Proposition. Soient λ et τ comme ci-dessus. Il existe $n_1, n_2 \in \mathbb{N}$ tels que $n_1 + n_2 = n$ et il existe $\lambda_1 \in \mathcal{P}^{\text{symp}, \text{sp}}(2n_1)$ et $\lambda_2 \in \mathcal{P}^{\text{orth}, \text{sp}}(2n_2)$ tels que

- (a) λ_1 et λ_2 induisent régulièrement λ ;
- (b) $d(\lambda_1) \cup d(\lambda_2) = d(\lambda)$;
- (c) $\tau_{\lambda_1, \lambda_2} = \tau$.

Preuve. Notons \mathfrak{J}^+ l'ensemble des $j \geq 1$ tels que j soit impair et $\lambda_j > \lambda_{j+1}$. Notons \mathfrak{J}^- l'ensemble des $j \geq 2$ tels que j soit pair et $\lambda_{j-1} > \lambda_j$. Les ensembles \mathfrak{J}^+ et \mathfrak{J}^- sont disjoints et leur réunion est égale à la réunion des couples $\{2k-1, 2k\}$, pour $k \geq 1$, tels que $\lambda_{2k-1} > \lambda_{2k}$. On note $\mathfrak{x} = (\mathfrak{x}_1, \mathfrak{x}_2, \dots)$ la suite telle que $\mathfrak{x}_j = 1$ si $j \in \mathfrak{J}^+$, $\mathfrak{x}_j = -1$ si $j \in \mathfrak{J}^-$ et $\mathfrak{x}_j = 0$ si $j \notin \mathfrak{J}^+ \cup \mathfrak{J}^-$.

On prolonge la fonction τ à $\text{Jord}_{\text{bp}}(\lambda) \cup \{0\}$ en posant $\tau(0) = 0$. Soit $d \in \{1, 2\}$. Pour $j \geq 1$, disons que j et $j+1$ sont d -liés si et seulement s'ils vérifient l'une des conditions suivantes :

- (20) $\lambda_j = \lambda_{j+1}$ et $\tau(\lambda_j) = d + 1$ (on veut dire par là $\tau(\lambda_j) \equiv d + 1 \pmod{2\mathbb{Z}}$) ;
- (21) j est impair et $\lambda_j > \lambda_{j+1}$.

Pour deux entiers $1 \leq j \leq j'$, disons qu'ils sont d -liés si et seulement si k et $k+1$ sont d -liés pour tout $k = j, \dots, j'-1$. C'est une relation d'équivalence. On note \mathfrak{Int}_d l'ensemble des classes d'équivalence dont le nombre d'éléments est au moins 2. Pour $\mathfrak{J} \in \mathfrak{Int}_d$, on note $j_{\min}(\mathfrak{J})$ (resp. $j_{\max}(\mathfrak{J})$) le plus petit (resp. plus grand) élément de \mathfrak{J} (éventuellement $j_{\max}(\mathfrak{J}) = \infty$). Montrons que :

- (22) L'ensemble \mathfrak{Int}_d est fini ; il contient un élément infini si et seulement si $d = 1$.
- (23) Pour $\mathfrak{J} \in \mathfrak{Int}_d$, $j_{\min}(\mathfrak{J})$ est impair et $j_{\max}(\mathfrak{J})$ est pair ou infini.
- (24) Pour tout $k \geq 1$, il existe au moins un $d \in \{1, 2\}$ et un $\mathfrak{J} \in \mathfrak{Int}_d$ tel que $\{2k-1, 2k\} \subset \mathfrak{J}_d$; les deux éléments de $\{1, 2\}$ vérifient cette condition si et seulement si $2k-1 \in \mathfrak{J}^+$ (ce qui équivaut à $2k \in \mathfrak{J}^-$).

- (25) Pour tout $k \geq 1$, il existe au plus un $d \in \{1, 2\}$ et un $\mathfrak{J} \in \mathfrak{Int}_d$ tel que $\{2k, 2k+1\} \subset \mathfrak{J}_d$.
- (26) Pour tout $j \geq 1$, il existe au moins un $d \in \{1, 2\}$ et un $\mathfrak{J} \in \mathfrak{Int}_d$ tel que $j \in \mathfrak{J}$; les deux éléments de $\{1, 2\}$ vérifient cette condition si et seulement si $j \in \mathfrak{J}^+ \cup \mathfrak{J}^-$.

Pour $j \geq l(\lambda) + 1$, on a $\lambda_j = \lambda_{j+1} = 0$ et, puisque $\tau(0) = 0$, j et $j+1$ sont 1-liés mais pas 2-liés. Donc $\{l(\lambda) + 1, \dots, \infty\}$ est contenu dans une classe infinie $\mathfrak{J}_{1, \min} \in \mathfrak{Int}_1$ tandis que, pour $j \geq l(\lambda) + 2$, $\{j\}$ forme une classe pour la 2-équivalence donc n'est pas contenu dans un élément de \mathfrak{Int}_2 . Cela prouve (22).

Soit $\mathfrak{J} \in \mathfrak{Int}_d$, posons $j = j_{\min}(\mathfrak{J})$. Montrons que j est impair. Puisque \mathfrak{J} a au moins deux éléments, j et $j+1$ sont d -liés. Si la condition (21) est vérifiée, j est impair et on a terminé. Si (20) est vérifiée, on a $\tau(\lambda_j) = d+1$. Si $j = 1$, j est impair et on a terminé. Sinon, puisque j est l'élément minimal de \mathfrak{J} , $j-1$ et j ne sont pas d -liés. Alors le couple $(j-1, j)$ ne vérifie pas (21). Donc $j-1$ est pair ou $\lambda_{j-1} = \lambda_j$. Dans le premier cas, j est impair et on a terminé. Dans le deuxième cas, on a $\tau(\lambda_{j-1}) = \tau(\lambda_j) = d+1$ mais alors $(j-1, j)$ vérifie (20) et $j-1$ et j sont d -liés, ce qui n'est pas le cas. Cela démontre l'assertion. Un raisonnement similaire prouve que $j_{\max}(\mathfrak{J})$ est pair s'il n'est pas infini. D'où (23).

Soit $k \geq 1$. Pour $d = 1, 2$, dire qu'il existe $\mathfrak{J} \in \mathfrak{Int}_d$ tel que $\{2k-1, 2k\} \subset \mathfrak{J}$ équivaut à ce que $2k-1$ et $2k$ soient d -liés. Si $\lambda_{2k-1} > \lambda_{2k}$, $2k-1$ et $2k$ vérifient (21) et sont d -liés pour les deux éléments $d = 1, 2$. Mais on a aussi $2k-1 \in \mathfrak{J}^+$ et l'assertion (24) est vérifiée dans ce cas. Si $\lambda_{2k-1} = \lambda_{2k}$, (21) n'est pas vérifiée. Alors $2k-1$ et $2k$ sont d -liés pour l'unique élément $d = \tau(\lambda_{2k-1}) + 1$. On a aussi $2k-1 \notin \mathfrak{J}^+$ et (24) est encore vérifiée.

Soient $k \geq 1$ et $d = 1, 2$. Le couple $(2k, 2k+1)$ ne vérifie pas la condition (21). Si $2k, 2k+1$ sont d -liés, la condition (20) est satisfaite. Donc $d = \tau(\lambda_{2k}) + 1$ est uniquement déterminé. D'où (25).

Soit $j \geq 1$. Posons $k = [(j+1)/2]$. On a $j \in \{2k-1, 2k\}$. Soit $d = 1, 2$ et $\mathfrak{J} \in \mathfrak{Int}_d$. D'après (23), les conditions $j \in \mathfrak{J}$ et $\{2k-1, 2k\} \subset \mathfrak{J}$ sont équivalentes. Alors (26) résulte de (24).

Pour $d = 1, 2$, définissons une fonction $p_d : \mathbb{N} - \{0\} \rightarrow \mathbb{Z}/2\mathbb{Z} : p_d(j) = 1$ s'il existe $\mathfrak{J} \in \mathfrak{Int}_d$ tel que $j \in \mathfrak{J}$, $p_d(j) = 0$ sinon. La relation (23) entraîne

$$p_d(j) = p_d(j+1) \quad \text{si } j \text{ est impair.}$$

La définition de \mathfrak{x} et l'assertion (26) entraînent l'égalité

$$\mathfrak{x}_j \equiv p_1(j) + p_2(j) + 1 \pmod{2\mathbb{Z}}. \quad (27)$$

On va montrer qu'il existe des suites d'entiers positifs ou nuls λ_1 et λ_2 vérifiant les conditions suivantes, pour $j \geq 1$:

$$(28) \quad \lambda_{1,j} + \lambda_{2,j} + \mathfrak{x}_j = \lambda_j;$$

$$(29) \quad \text{pour } d = 1, 2, \lambda_{d,j} \equiv d + p_d(j) \pmod{2\mathbb{Z}};$$

(30) pour $d = 1, 2$, on a

- (a) $\lambda_{d,j} = \lambda_{d,j+1}$ si j est pair, $p_d(j) = 1$ et il n'existe pas de $\mathfrak{J} \in \mathfrak{Int}_d$ tel que $j = j_{\max}(\mathfrak{J})$ ou si j est impair et $p_d(j) = 0$;

- (b) $\lambda_{d,j} > \lambda_{d,j+1}$ si j est pair et il existe $\mathfrak{J} \in \mathfrak{Int}_d$ tel que $j = j_{\max}(\mathfrak{J})$;
(c) $\lambda_{d,j} \geq \lambda_{d,j+1}$ si j est impair et $p_d(j) = 1$ ou si j est pair et $p_d(j) = 0$.

On raisonne par récurrence descendante sur j . Pour $j \geq l(\lambda) + 2$, on pose $\lambda_{1,j} = \lambda_{2,j} = 0$. On a vu dans la preuve de (22) que j était contenu dans $\mathfrak{J}_{1,\min}$ et n'était contenu dans aucun élément de \mathfrak{Int}_2 . On a aussi $j \notin \mathfrak{J}^+ \cup \mathfrak{J}^-$ donc $\mathfrak{x}_j = 0$. On voit que toutes nos conditions sont vérifiées.

On fixe j et on suppose que l'on a fixé des termes $\lambda_{1,j'}$ et $\lambda_{2,j'}$ pour $j' > j$ de sorte que les conditions soient vérifiées pour ces j' . Pour $d = 1, 2$, soit $e_d \in \mathbb{Z}$, posons $\lambda_{d,j} = \lambda_{d,j+1} + e_d$. Traduisons les conditions ci-dessus en termes des entiers e_1 et e_2 . La condition (28) étant vérifiée pour $j + 1$, on voit que cette condition pour j équivaut à

$$e_1 + e_2 = \lambda_j - \lambda_{j+1} + \mathfrak{x}_{j+1} - \mathfrak{x}_j. \quad (31)$$

La condition (29) étant vérifiée pour $j + 1$, cette condition pour j équivaut à

$$e_d \equiv p_d(j) + p_d(j + 1) \pmod{2\mathbb{Z}}. \quad (32)$$

Remarquons que, si (31) est vérifiée et si (32) l'est pour un $d \in \{1, 2\}$, cette condition (32) est aussi vérifiée pour l'autre élément de $\{1, 2\}$: cela résulte de la parité de λ_j et de λ_{j+1} et de la relation (27). La condition (30) se traduit par les conditions $e_d = 0$ dans le cas (a), $e_d > 0$ dans le cas (b) et $e_d \geq 0$ dans le cas (c). Remarquons que, dans le cas (a), la condition $e_d = 0$ est compatible avec (32), autrement dit on a $p_d(j) + p_d(j + 1) \equiv 0 \pmod{2\mathbb{Z}}$. En effet, si j est impair, on a toujours $p_d(j) = p_d(j + 1)$. Si j est pair, la condition de (30)(a) est d'une part que $p_d(j) = 1$ donc il existe $\mathfrak{J} \in \mathfrak{Int}_d$ tel que $j \in \mathfrak{J}$, d'autre part que j n'est pas l'élément maximal de \mathfrak{J} . Donc $j + 1 \in \mathfrak{J}$ et $p_d(j + 1) = p_d(j)$.

Supposons la condition (30)(a) vérifiée pour au moins un $d = 1, 2$, disons pour $d = 1$ pour fixer la notation. On n'a pas le choix pour e_1 : on pose $e_1 = 0$. Comme on vient de le dire, la condition (32) est vérifiée pour $d = 1$. La condition (31) ne laisse plus le choix pour e_2 : on pose $e_2 = \lambda_j - \lambda_{j+1} + \mathfrak{x}_{j+1} - \mathfrak{x}_j$. Puisque (31) est vérifiée et aussi (32) pour $d = 1$, (32) est aussi vérifiée pour $d = 2$. Il reste à vérifier que e_2 vérifie les conditions résultant de (30). Supposons d'abord j pair. L'hypothèse que (30)(a) est vérifiée pour $d = 1$ signifie, comme on l'a vu ci-dessus, qu'il existe $\mathfrak{J}_1 \in \mathfrak{Int}_1$ tel que $\{j, j + 1\} \subset \mathfrak{J}_1$. D'après (25), cette condition ne peut pas être réalisée pour $d = 2$. Donc (30)(a) n'est pas vérifiée pour $d = 2$. Si (30)(c) est vérifiée pour $d = 2$, on doit seulement voir que $e_2 \geq 0$. Or, puisque j est pair, on a $-\mathfrak{x}_j \geq 0$ et $\mathfrak{x}_{j+1} \geq 0$, donc $\lambda_j - \lambda_{j+1} + \mathfrak{x}_{j+1} - \mathfrak{x}_j \geq 0$ comme on le voulait. Si (30)(b) est vérifiée pour $d = 2$, on doit montrer que $e_2 > 0$. On a $p_1(j) = 1$ d'après (30)(a) pour $d = 1$ et $p_2(j) = 1$ d'après (30)(b) pour $d = 2$. Alors $j \in \mathfrak{J}^-$ d'après (26) et $-\mathfrak{x}_j = 1$. Donc $\lambda_j - \lambda_{j+1} + \mathfrak{x}_{j+1} - \mathfrak{x}_j > 0$ comme on le voulait. Supposons maintenant j impair. L'hypothèse que (30)(a) est vérifiée pour $d = 1$ signifie que $p_1(j) = 0$. D'après (26), on a $p_2(j) = 1$ et $j \notin \mathfrak{J}^+$, donc aussi $j + 1 \notin \mathfrak{J}^-$. Ces deux dernières relations entraînent $\mathfrak{x}_j = \mathfrak{x}_{j+1} = 0$ et $e_2 = \lambda_j - \lambda_{j+1}$. La relation $p_2(j) = 1$ entraîne que (30)(c) est vérifiée pour $d = 2$ et que l'on doit seulement prouver que $e_2 \geq 0$, ce qui est clair d'après la formule précédente.

Supposons maintenant que (30)(a) n'est vérifiée ni pour $d = 1$, ni pour $d = 2$. Supposons la condition (30)(b) vérifiée pour au moins un $d = 1, 2$, disons pour $d = 1$. Cela entraîne que j est pair. Choisissons

pour e_1 le plus petit entier strictement positif vérifiant la condition (32). On a $e_1 = 1$ ou 2 . Posons $e_2 = \lambda_j - \lambda_{j+1} + \varepsilon_{j+1} - \varepsilon_j - e_1$. Comme ci-dessus, on doit montrer que e_2 vérifie les conditions résultant de (30). On a supposé que (30)(a) n'était pas vérifiée pour $d = 2$. Supposons que (30)(c) soit vérifiée pour $d = 2$. Il faut voir que $e_2 \geq 0$. D'après (30)(b) pour $d = 1$, il existe $\mathfrak{J}_1 \in \mathfrak{Int}_1$ tel que $j = j_{\max}(\mathfrak{J}_1)$. Donc j et $j + 1$ ne sont pas 1-liés. D'après (30)(c) pour $d = 2$ et parce que j est pair, on a $p_2(j) = 0$ donc j et $j + 1$ ne sont pas 2 liés. Si $\lambda_j = \lambda_{j+1}$ la condition (20) est vérifiée pour un d donc j et $j + 1$ sont d -liés pour ce d . Puisque ce n'est pas le cas, on a $\lambda_j \neq \lambda_{j+1}$, donc $\lambda_j \geq \lambda_{j+1} + 2$, puisque les termes de λ sont pairs. Le même calcul que plus haut conduit à l'inégalité cherchée $e_2 \geq 0$. Supposons maintenant (30)(b) vérifiée pour $d = 2$. On doit prouver $e_2 > 0$. On vient de montrer que j et $j + 1$ n'étaient pas 1-liés. Pour la même raison, ils ne sont pas 2-liés et cela entraîne encore $\lambda_j \geq \lambda_{j+1} + 2$. Les conditions (30)(b) pour $d = 1, 2$ entraînent que $p_1(j) = p_2(j) = 1$, donc $j \in \mathfrak{J}^-$ d'après (26). Alors $-\varepsilon_j = 1$ et on voit que $e_2 > 0$.

Il reste le cas où (30)(c) est vérifiée pour $d = 1, 2$. Puisque $p_d(j) = 1$ pour au moins un d , cette hypothèse entraîne que j est impair et $p_1(j) = p_2(j) = 1$. Donc $j \in \mathfrak{J}^+$, puis $j + 1 \in \mathfrak{J}^-$. Ces relations entraînent que $\varepsilon_j = 1$ et $\varepsilon_{j+1} = -1$ et aussi que $\lambda_j > \lambda_{j+1}$, donc $\lambda_j \geq \lambda_{j+1} + 2$. Puisque j est impair, on a $p_1(j + 1) = p_1(j) = 1$. La condition (32) pour $d = 1$ signifie que e_1 doit être pair. Choisissons $e_1 = 0$, qui vérifie la condition résultant de (30)(c) pour $d = 1$. Posons $e_2 = \lambda_j - \lambda_{j+1} + \varepsilon_{j+1} - \varepsilon_j - e_1 = \lambda_j - \lambda_{j+1} - 2$. On a $e_2 \geq 0$, ce qui vérifie la condition résultant de (30)(c) pour $d = 2$. Cela démontre l'existence de nos suites λ_1 et λ_2 .

Fixons donc de telles suites λ_1 et λ_2 . La condition (30) entraîne que ce sont des partitions, c'est-à-dire qu'elles sont décroissantes. Montrons que :

(33) Il existe des entiers n_1, n_2 tels que $n_1 + n_2 = n$, que λ_1 appartienne à $\mathcal{P}^{\text{symp}, \text{sp}}(2n_1)$ et que λ_2 appartienne à $\mathcal{P}^{\text{orth}, \text{sp}}(2n_2)$.

On voit qu'il s'agit de prouver que, pour $d = 1, 2$ et $k \geq 1$, les termes $\lambda_{d, 2k-1}$ et $\lambda_{d, 2k}$ sont de même parité et que, quand cette parité est celle de d , on a $\lambda_{d, 2k-1} = \lambda_{d, 2k}$. La première propriété résulte de (29) et de l'égalité $p_d(2k - 1) = p_d(2k)$. Si la parité de $\lambda_{d, 2k-1}$ est celle de d , cette même relation (29) entraîne $p_d(2k - 1) = 0$. Mais alors (30)(a) est vérifiée pour $2k - 1$, d'où $\lambda_{d, 2k-1} = \lambda_{d, 2k}$. D'où (33).

Grâce à cette relation, on peut définir les ensembles d'intervalles $\text{Int}(\lambda_1)$ et $\text{Int}(\lambda_2)$ et, comme en 1.9, les ensembles J^+ et J^- et la fonction ξ . Montrons que

$$\{J(\Delta); \Delta \in \text{Int}(\lambda_1)\} = \mathfrak{Int}_1, \quad \{J(\Delta); \Delta \in \text{Int}(\lambda_2)\} = \mathfrak{Int}_2, \quad J^+ = \mathfrak{J}^+, \quad J^- = \mathfrak{J}^-, \quad \xi = \varepsilon. \quad (34)$$

Soit $d = 1, 2$. La réunion des $J(\Delta)$ quand Δ parcourt $\text{Int}(\lambda_d)$ est l'ensemble des $j \geq 1$ tels que $\lambda_{d, j} \equiv d + 1 \pmod{2\mathbb{Z}}$. En vertu de (29), c'est l'ensemble des $j \geq 1$ tels que $p_d(j) = 1$, autrement dit c'est la réunion des éléments de \mathfrak{Int}_d . On a donc un même ensemble d'indices découpé de deux façons en intervalles disjoints : les $J(\Delta)$ pour $\Delta \in \text{Int}(\lambda_d)$ ou les $\mathfrak{J} \in \mathfrak{Int}_d$. Pour prouver que ces découpages sont les mêmes, il suffit de prouver que les éléments maximaux de ces intervalles sont les mêmes, c'est-à-dire

$$\{j_{\max}(\Delta); \Delta \in \text{Int}(\lambda_d)\} = \{j_{\max}(\mathfrak{J}); \mathfrak{J} \in \mathfrak{Int}_d\}.$$

Comme on l'a vu en (22), l'infini intervient dans les deux ensembles si $d = 1$ et n'y intervient pas si $d = 2$. Soit $j \geq 1$. Par définition de $\text{Int}(\lambda_d)$, j appartient à l'ensemble de gauche ci-dessus si et seulement si j est pair, $\lambda_{d,j} \equiv d + 1 \pmod{2\mathbb{Z}}$ et $\lambda_{d,j} > \lambda_{d,j+1}$. On vient de voir que la congruence est équivalente à $p_d(j) = 1$. Les relations (30) entraînent alors que ces conditions équivalent à ce que j soit de la forme $j_{\max}(\mathfrak{J})$ pour un $\mathfrak{J} \in \mathfrak{Int}_d$. Cela démontre les deux premières égalités de (34). Soit $j \in J^+$. Alors j est impair $\lambda_{1,j}$ et $\lambda_{2,j}$ sont "de bonne parité", d'où, comme on l'a vu, $p_d(j) = 1$ pour $d = 1, 2$. Alors $j \in \mathfrak{J}^+$ d'après (26) et l'imparité de j . Inversement, soit $j \in \mathfrak{J}^+$. Alors j est impair et, en inversant le raisonnement précédent, $\lambda_{1,j}$ et $\lambda_{2,j}$ sont de bonne parité. Il existe $\Delta_1 \in \text{Int}(\lambda_1)$ et $\Delta_2 \in \text{Int}(\lambda_2)$ tels que $j \in J(\Delta_1) \cap J(\Delta_2)$. Si $j = 1$, on a évidemment $j = j_{\min}(\Delta_1) = j_{\min}(\Delta_2)$ et $j \in J^+$. Si $j > 1$, l'assertion (25) implique qu'il existe d tel que $j - 1$ n'appartienne pas à \mathfrak{J}_d , où $\mathfrak{J}_d = J(\Delta_d)$. Alors $j = j_{\min}(\mathfrak{J}_d)$ pour ce d , ou encore $j = j_{\min}(\Delta_d)$. Par définition de l'ensemble J^+ , on a alors $j \in J^+$. Cela prouve l'égalité $J^+ = \mathfrak{J}^+$ et l'égalité $J^- = \mathfrak{J}^-$ se démontre de même. Ces égalités et les définitions de ξ et \mathfrak{r} entraînent la dernière égalité de (34).

L'égalité $\xi = \mathfrak{r}$ et la relation (28) entraînent l'égalité $\text{Ind}(\lambda_1, \lambda_2) = \lambda$. Montrons que

$$\lambda_1 \text{ et } \lambda_2 \text{ induisent régulièrement } \lambda. \quad (35)$$

Cela signifie que tout intervalle relatif est réduit à un seul élément. Soit D un tel intervalle relatif. Évidemment, si $J(D)$ est réduit à un seul élément, D aussi. Supposons que $J(D)$ a au moins deux éléments. Il vérifie la relation (19). Pour fixer la notation, supposons que l'entier d qui figure dans cette relation soit 1. Il existe donc $\mathfrak{J}_1 \in \mathfrak{Int}_1$ tel que $J(D) \subset \mathfrak{J}_1$. Considérons deux éléments consécutifs $j, j+1 \in J(D)$. Supposons qu'il existe $\mathfrak{J}_2 \in \mathfrak{Int}_2$ tel que $\{j, j+1\} \subset \mathfrak{J}_2$. On a $j_{\min}(\mathfrak{J}_2) < j+1 \leq j_{\max}(D)$. Puisque les termes $j_{\min}(D)$ et $j_{\max}(D)$ sont par définition des éléments consécutifs de \mathcal{J} , cela entraîne $j_{\min}(\mathfrak{J}_2) \leq j_{\min}(D)$. De même $j_{\max}(D) \leq j_{\max}(\mathfrak{J}_2)$. Alors $J(D) \subset \mathfrak{J}_2$ ce qui est exclu par (19). Cela démontre que, pour deux éléments $j, j+1 \in J(D)$, il n'existe pas de $\mathfrak{J}_2 \in \mathfrak{Int}_2$ tel que $\{j, j+1\} \subset \mathfrak{J}_2$. Donc j et $j+1$ sont 1-liés mais pas 2-liés. En se reportant aux relations (20) et (21) qui définissent la liaison, on voit que, si j est impair, le fait que j et $j+1$ ne sont pas 2-liés entraîne que $\lambda_j = \lambda_{j+1}$, tandis que, si j est pair, le fait que j et $j+1$ sont 1-liés entraîne la même égalité. Cette égalité pour tout couple $j, j+1 \in J(D)$ entraîne que λ_j est constant pour $j \in J(D)$, ce que l'on voulait démontrer.

Montrons que

$$\tau_{\lambda_1, \lambda_2} = \tau. \quad (36)$$

Soit $i \in \text{Jord}_{\text{bp}}(\lambda)$. Si $\text{mult}_\lambda(i) = 1$, on a $\tau_{\lambda_1, \lambda_2}(i) = \tau(i) = 0$ par définition. Supposons $\text{mult}_\lambda(i) \geq 2$. Comme ci-dessus, il existe un unique $d = 1, 2$ et un unique $\mathfrak{J}_d \in \mathfrak{Int}_d$ tel que $J(\{i\}) \subset \mathfrak{J}_d$. On a alors $\tau_{\lambda_1, \lambda_2}(i) = d + 1$. Considérons un couple $j, j+1 \in J(\{i\})$. Ils sont d -liés et on a $\lambda_j = \lambda_{j+1} = i$. L'une des relations (20) ou (21) est vérifiée pour d et ce ne peut être que (20). Donc $\tau(i) = d + 1$, d'où l'égalité cherchée $\tau_{\lambda_1, \lambda_2}(i) = \tau(i)$.

Montrons qu'on a l'égalité

$$\zeta(\lambda_1) + \zeta(\lambda_2) = \zeta(\lambda) + \xi. \quad (37)$$

Soit $j \geq 1$. Supposons j impair. Chacune des quatre fonctions vaut 0 ou 1 en j . Supposons d'abord $\zeta(\lambda_1)_j = \zeta(\lambda_2)_j = 1$. Alors il existe $\mathfrak{J}_1 \in \mathfrak{Int}_1$ et $\mathfrak{J}_2 \in \mathfrak{Int}_2$ de sorte que $j = j_{\min}(\mathfrak{J}_1) = j_{\min}(\mathfrak{J}_2)$. D'après (26) et (34), on a $j \in J^+$, d'où $\xi_j = 1$. Si $j = 1$, j est le plus petit indice tel que λ_j appartienne au plus grand intervalle de λ (il s'agit ici des intervalles au sens des partitions spéciales) donc $\zeta(\lambda)_j = 1$. Si $j > 1$, l'hypothèse sur j implique que $j - 1$ et j ne sont ni 1-liés, ni 2-liés. Si $\lambda_{j-1} = \lambda_j$, $j - 1$ et j sont d -liés pour le d tel que $\tau(\lambda_j) = d + 1$, cf. (20). C'est impossible donc $\lambda_{j-1} > \lambda_j$. Puisque j est impair, c'est la condition pour que j soit de la forme $j = j_{\min}(\Delta)$ pour un $\Delta \in \text{Int}(\lambda)$. Donc $\zeta(\lambda)_j = 1$. L'égalité (37) est vérifiée en j . Supposons maintenant $\zeta(\lambda_1)_j = 1$ et $\zeta(\lambda_2)_j = 0$ (un raisonnement analogue vaut si on échange les indices 1 et 2). Il existe $\mathfrak{J}_1 \in \mathfrak{Int}_1$ tel que $j = j_{\min}(\mathfrak{J}_1)$ mais il n'y a pas de \mathfrak{J}_2 vérifiant la même égalité. Supposons d'abord $p_2(j) = 1$. De nouveau, $j \in J^+$ et $\xi_j = 1$. Puisque $p_2(j) = 1$, le fait que j ne soit pas le plus petit élément d'un élément de \mathfrak{Int}_2 entraîne que $j \geq 2$ et que $j - 1$ et j sont 2-liés. Puisque $j - 1$ est pair, cette condition implique $\lambda_{j-1} = \lambda_j$. Donc $\zeta(\lambda)_j = 0$ et on obtient l'égalité cherchée. Supposons au contraire $p_2(j) = 0$. Alors $j \notin J^+$ et $\xi_j = 0$. Si $j = 1$, on a $\zeta(\lambda)_j = 1$ comme ci-dessus. Sinon, $j - 1$ et j ne sont pas 1-liés (car $j = j_{\min}(\mathfrak{J}_1)$) et ne sont pas 2-liés (car $p_2(j) = 0$). Comme ci-dessus, cela entraîne $\lambda_{j-1} > \lambda_j$ et $\zeta(\lambda)_j = 1$. D'où l'égalité cherchée. Supposons enfin $\zeta(\lambda_1)_j = \zeta(\lambda_2)_j = 0$. D'après (26), on peut supposer par exemple $p_1(j) = 1$. Comme ci-dessus, l'hypothèse $\zeta(\lambda_1)_j = 0$ implique alors $j \geq 2$ et $j - 1$ et j sont 1-liés. D'où $\lambda_{j-1} = \lambda_j$ et $\tau(\lambda_j) = 0$. La première relation entraîne $\zeta(\lambda)_j = 0$. La seconde entraîne que $j - 1$ et j ne sont pas 2-liés. Si $p_2(j) = 1$, j est de la forme $j_{\min}(\mathfrak{J}_2)$ et alors $\zeta(\lambda)_j = 1$ contrairement à l'hypothèse. Donc $p_2(j) = 0$ et $j \notin J^+$. Donc $\xi_j = 0$ et on obtient l'égalité cherchée. Des calculs similaires valent dans le cas j pair. Cela prouve (37).

Cette égalité entraîne

$$\lambda_1 + \zeta(\lambda_1) + \lambda_2 + \zeta(\lambda_2) = \lambda_1 + \lambda_2 + \xi + \zeta(\lambda) = \lambda + \zeta(\lambda).$$

En utilisant les lemmes 1.6 et 1.7, cette égalité se transforme en

$${}^t d(\lambda_1) + {}^t d(\lambda_2) = {}^t d(\lambda), \quad \text{qui équivaut à } d(\lambda_1) \cup d(\lambda_2) = d(\lambda). \quad \square$$

1.12. Multiplicités. Soient $n, n', n'', n_1, n_2 \in \mathbb{N}$ tels que $n = n' + n'' = n_1 + n_2$. Soient $\rho_1 \in \widehat{W}_{n_1}$ et $\rho_2 \in \widehat{W}_{n_2}^D$. A ρ_1 est associé un symbole (X_1, Y_1) de rang n_1 et de défaut 1. On note λ_1 la partition symplectique spéciale de $2n_1$ associée à la famille de (X_1, Y_1) et on pose $(\tau_1, \delta_1) = \text{fam}(X_1, Y_1)$. A ρ_2 est associé un symbole (X_2, Y_2) de rang n_2 et de défaut 0. On note λ_2 la partition orthogonale spéciale de $2n_2$ associée à la famille de (X_2, Y_2) et on pose $(\tau_2, \delta_2) = \text{fam}(X_2, Y_2)$. On définit des représentations ρ_2^+ et ρ_2^- de W_{n_2} de la façon suivante. Introduisons le couple $(\alpha_2, \beta_2) \in \mathcal{P}_2(n_2)$ qui paramètre ρ_2 . C'est-à-dire que, si $\alpha_2 \neq \beta_2$, $\rho_2 = \rho^D(\alpha_2, \beta_2)$; si $\alpha_2 = \beta_2$, il existe un signe $\eta = \pm$ tel que $\rho_2 = \rho^D(\alpha_2, \alpha_2, \eta)$. Dans ce dernier cas, on pose $\rho_2^+ = \rho_2^- = \rho(\alpha_2, \alpha_2)$. Si $\alpha_2 \neq \beta_2$, on sait que l'on peut permuter α_2 et β_2 . Supposons α_2 plus grand que β_2 pour l'ordre lexicographique (pour le plus petit indice j tel que $\alpha_{2,j} \neq \beta_{2,j}$, on a $\alpha_{2,j} > \beta_{2,j}$). On pose

$$\rho_2^+ = \rho(\alpha_2, \beta_2) \quad \text{et} \quad \rho_2^- = \rho(\beta_2, \alpha_2).$$

Soient $(\lambda', \epsilon') \in \mathcal{P}^{\mathrm{symp}}(2n')$, $(\lambda'', \epsilon'') \in \mathcal{P}^{\mathrm{symp}}(2n'')$. Considérons l'hypothèse

$$k_{\lambda', \epsilon'} = k_{\lambda'', \epsilon''} = 0. \quad (\mathrm{Hyp})$$

Supposons-la vérifiée. Dans [Waldspurger 2018, §1.8 et §1.10], on a défini des espaces $\mathcal{R} = \bigoplus_{\gamma \in \Gamma} \mathcal{R}(\gamma)$, $\mathcal{R}^{\mathrm{glob}} \subset \mathcal{R}$ et une application linéaire $\rho_{\iota} : \mathcal{R} \rightarrow \mathcal{R}^{\mathrm{glob}}$ (ces objets sont relatifs à l'entier n). Posons $\gamma = (0, 0, n', n'')$. C'est un élément de Γ et $\rho_{\lambda', \epsilon'} \otimes \rho_{\lambda'', \epsilon''}$ s'identifie à un élément de $\mathcal{R}(\gamma)$. On dispose donc de l'élément $\rho_{\iota}(\rho_{\lambda', \epsilon'} \otimes \rho_{\lambda'', \epsilon''}) \in \mathcal{R}$. Remarquons en passant que l'élément \mathbf{a} de [Waldspurger 2018, §1.10] vaut $(0, 0, 0, 1)$. Posons $\theta = (0, 0, n_1, n_2)$. C'est aussi un élément de Γ et, pour $\zeta = \pm$, $\rho_1 \otimes \rho_2^{\zeta}$ s'identifie à un élément de $\mathcal{R}(\theta)$. On peut définir la multiplicité $m(\rho_1, \rho_2^{\zeta}; \rho_{\lambda', \epsilon'}, \rho_{\lambda'', \epsilon''})$ de $\rho_1 \otimes \rho_2^{\zeta}$ dans $\rho_{\iota}(\rho_{\lambda', \epsilon'} \otimes \rho_{\lambda'', \epsilon''})$ par la formule usuelle

$$m(\rho_1, \rho_2^{\zeta}; \rho_{\lambda', \epsilon'}, \rho_{\lambda'', \epsilon''}) = |W_{n_1}|^{-1} |W_{n_2}|^{-1} \sum_{w_1 \in W_{n_1}, w_2 \in W_{n_2}} \rho_1(w_1) \rho_2^{\zeta}(w_2) \rho_{\iota}(\rho_{\lambda', \epsilon'} \otimes \rho_{\lambda'', \epsilon''})(w_1 \times w_2).$$

On n'a pas besoin d'introduire des conjugaisons complexes dans cette formule puisqu'on sait que les représentations irréductibles des groupes de type W_n ont des caractères réels. En réfléchissant à la définition de $\rho_{\iota}(\rho_{\lambda', \epsilon'} \otimes \rho_{\lambda'', \epsilon''})$, on voit que sa restriction à $\mathcal{R}(\theta)$ est une "vraie" représentation, ce qui entraîne que la multiplicité ci-dessus est un entier naturel.

On a défini en 1.9 l'induite endoscopique $\mathrm{ind}(\lambda_1, \lambda_2) \in \mathcal{P}^{\mathrm{symp}}(2n)$.

Proposition. *On suppose vérifiée l'hypothèse (Hyp). Soit $\zeta = \pm$. Si $m(\rho_1, \rho_2^{\zeta}; \rho_{\lambda', \epsilon'}, \rho_{\lambda'', \epsilon''}) \neq 0$, alors $\lambda' \cup \lambda'' \leq \mathrm{ind}(\lambda_1, \lambda_2)$.*

Cette proposition, comme la suivante, se déduit des résultats de [Waldspurger 2001]. Nous donnerons la preuve dans le paragraphe suivant.

1.13. Multiplicités, cas particulier. On conserve les données du paragraphe précédent. Posons $\lambda = \mathrm{ind}(\lambda_1, \lambda_2)$. On suppose de plus :

λ est à termes pairs; λ_1 et λ_2 induisent régulièrement λ ; $\delta_1 = \delta_2 = 0$.

On définit des fonctions $\delta^+, \delta^-, \tau^+, \tau^- : \mathrm{Jord}_{\mathrm{bp}}(\lambda) \rightarrow \mathbb{Z}/2\mathbb{Z}$ de la façon suivante, où on utilise les notations des paragraphes 1.9 et 1.10. Soit $i \in \mathrm{Jord}_{\mathrm{bp}}(\lambda)$. On a $\{i\} \in \mathrm{Int}_{\lambda_1, \lambda_2}(\lambda)$ puisque λ_1 et λ_2 induisent régulièrement λ . On pose $\delta^+(i) = \delta^-(i) = 0$ sauf dans le cas où $j_{\max}(\{i\}) \in J^+$. Dans ce cas, il existe d'uniques $\Delta_1 \in \mathrm{Int}(\lambda_1)$ et $\Delta_2 \in \mathrm{Int}(\lambda_2)$ tels que $j_{\max}(\{i\}) \in J(\Delta_1) \cap J(\Delta_2)$ et on pose

$$\delta^+(i) = \tau_1(\Delta_1) + \tau_2(\Delta_2) + 1, \quad \delta^-(i) = \tau_1(\Delta_1) + \tau_2(\Delta_2).$$

Si $\mathrm{mult}_{\lambda}(i) = 1$, il existe comme ci-dessus d'uniques $\Delta_1 \in \mathrm{Int}(\lambda_1)$ et $\Delta_2 \in \mathrm{Int}(\lambda_2)$ tels que $j_{\max}(\{i\}) \in J(\Delta_1) \cap J(\Delta_2)$ et on pose $\tau^+(i) = \tau^-(i) = \tau_1(\Delta_1)$. Supposons $\mathrm{mult}_{\lambda}(i) \geq 2$. Alors il existe un unique $d = 1, 2$ et un unique $\Delta_d \in \mathrm{Int}(\lambda_d)$ tels que $J(\{i\}) \subset J(\Delta_d)$. Si $d = 1$, on pose $\tau^+(i) = \tau^-(i) = \tau_1(\Delta_1)$. Si $d = 2$, on pose

$$\tau^+(i) = \tau_2(\Delta_2), \quad \tau^-(i) = \tau_2(\Delta_2) + 1.$$

Si i n'est pas l'élément maximal de $\text{Jord}_{\text{bp}}(\lambda)$, on note i^+ le plus petit élément de $\text{Jord}_{\text{bp}}(\lambda)$ strictement supérieur à i . Si i est l'élément maximal, on pose par convention $\delta^+(i^+) = \delta^-(i^+) = 1$.

Remarque. (1) On vérifie sur ces formules que $\tau^+ + \tau^- = \tau_{\lambda_1, \lambda_2}$, cf. 1.10.

(2) On a montré en [Waldspurger 2001, XI.29 remarque] que, pour tout $i \in \text{Jord}_{\text{bp}}(\lambda)$, on a la congruence $\delta^+(i) + \delta^-(i) \equiv \text{mult}_\lambda(\geq i) \pmod{2\mathbb{Z}}$. Cela équivaut à $\text{mult}_\lambda(i) \equiv \delta^+(i) + \delta^-(i) - \delta^+(i^+) - \delta^-(i^+) \pmod{2\mathbb{Z}}$.

Soit $\zeta = \pm$. On introduit les deux conditions suivantes :

(A) $^\zeta$ (i) $\lambda' \cup \lambda'' = \lambda$;

(ii) pour tout $i \in \text{Jord}_{\text{bp}}(\lambda)$, $\text{mult}_{\lambda''}(i) \equiv \delta^{-\zeta}(i) - \delta^{-\zeta}(i^+) \pmod{2\mathbb{Z}}$;

(iii) pour tout $i \in \text{Jord}_{\text{bp}}(\lambda')$, $\epsilon'(i) = (-1)^{\tau^\zeta(i)}$; pour tout $i \in \text{Jord}_{\text{bp}}(\lambda'')$, $\epsilon''(i) = (-1)^{\tau^{-\zeta}(i)}$.

(B) $^\zeta$ (i) $\lambda' \cup \lambda'' = \lambda$;

(ii) l'hypothèse (Hyp) de 1.12 est vérifiée et $m(\rho_1, \rho_2^\zeta; \rho_{\lambda', \epsilon'}, \rho_{\lambda'', \epsilon''}) \neq 0$.

Proposition. Pour $\zeta = \pm$, les conditions (A) $^\zeta$ et (B) $^\zeta$ sont équivalentes. Si elles sont vérifiées, on a $m(\rho_1, \rho_2^\zeta; \rho_{\lambda', \epsilon'}, \rho_{\lambda'', \epsilon''}) = 1$.

Preuve de la proposition et de la précédente. On devra utiliser la propriété générale suivante. Soient m, m', m'' trois éléments de \mathbb{N} tels que $m' + m'' = m$, soient $(\alpha, \beta) \in \mathcal{P}_2(m)$, $(\alpha', \beta') \in \mathcal{P}_2(m')$ et $(\alpha'', \beta'') \in \mathcal{P}_2(m'')$. Le groupe $W_{m'} \times W_{m''}$ se plonge naturellement dans W_m et ce plongement est bien défini à conjugaison près. D'où un foncteur de restriction $\text{res}_{W_{m'} \times W_{m''}}^{W_m}$. On a :

(38) Supposons que $\rho(\alpha', \beta') \otimes \rho(\alpha'', \beta'')$ intervienne dans $\text{res}_{W_{m'} \times W_{m''}}^{W_m}(\rho(\alpha, \beta))$; alors

(a) $S(\alpha') + S(\alpha'') = S(\alpha)$, $S(\beta') + S(\beta'') = S(\beta)$;

(b) $\alpha' \cup \alpha'' \leq \alpha$, $\beta' \cup \beta'' \leq \beta$;

(c) $\alpha \leq \alpha' + \alpha''$, $\beta \leq \beta' + \beta''$.

(39) Supposons que (a) soit vérifiée ainsi que l'une des conditions suivantes :

(b') $\alpha' \cup \alpha'' = \alpha$, $\beta' \cup \beta'' = \beta$;

(c') $\alpha = \alpha' + \alpha''$, $\beta = \beta' + \beta''$;

alors $\rho(\alpha', \beta') \otimes \rho(\alpha'', \beta'')$ intervient dans $\text{res}_{W_{m'} \times W_{m''}}^{W_m}(\rho(\alpha, \beta))$ avec multiplicité 1.

Si on oublie les conditions (b) et (b'), cela résulte de [Geck et Pfeiffer 2000, Lemmas 6.1.2, 6.1.3]. En remarquant que la multiplicité de $\rho(\alpha', \beta') \otimes \rho(\alpha'', \beta'')$ dans $\text{res}_{W_{m'} \times W_{m''}}^{W_m}(\rho(\alpha, \beta))$ est égale à celle de $(\rho(\alpha', \beta') \otimes \text{sgn}) \otimes (\rho(\alpha'', \beta'') \otimes \text{sgn})$ dans $\text{res}_{W_{m'} \times W_{m''}}^{W_m}(\rho(\alpha, \beta) \otimes \text{sgn})$, c'est à dire celle de $\rho({}^t\beta', {}^t\alpha') \otimes \rho({}^t\beta'', {}^t\alpha'')$ dans $\text{res}_{W_{m'} \times W_{m''}}^{W_m}(\rho({}^t\beta, {}^t\alpha))$, on récupère ces assertions (b) et (b').

Plaçons-nous dans la situation du paragraphe précédent (c'est-à-dire qu'on lève provisoirement l'hypothèse que λ_1 et λ_2 induisent régulièrement λ) et posons l'hypothèse (Hyp), c'est-à-dire

$$k_{\lambda', \epsilon'} = k_{\lambda'', \epsilon''} = 0. \quad (40)$$

Notons Π la composante de $\rho\iota(\rho_{\lambda',\epsilon'} \otimes \rho_{\lambda'',\epsilon''})$ dans $\mathcal{R}(\theta)$, où $\theta = (0, 0, n_1, n_2)$. On note \mathcal{N} l'ensemble des quadruplets $\mathbf{n} = (n'_1, n'_2, n''_1, n''_2)$ tels que

$$n' = n'_1 + n'_2, \quad n'' = n''_1 + n''_2, \quad n_1 = n'_1 + n''_1, \quad n_2 = n'_2 + n''_2.$$

Pour un tel quadruplet, posons $W_{\mathbf{n}} = W_{n'_1} \times W_{n'_2} \times W_{n''_1} \times W_{n''_2}$. Ce groupe se plonge dans $W_{n_1} \times W_{n_2}$ et dans $W_{n'} \times W_{n''}$ de façon évidente, les plongements étant bien définis à conjugaison près. D'où des foncteurs d'induction $\text{ind}_{W_{\mathbf{n}}}^{W_{n_1} \times W_{n_2}}$ et de restriction $\text{res}_{W_{\mathbf{n}}}^{W_{n'} \times W_{n''}}$. Notons $\text{sgn}_{CD,\mathbf{n}}$ le caractère de $W_{\mathbf{n}}$ qui est le produit tensoriel du caractère sgn_{CD} de W_{n_2} et des caractères triviaux des autres facteurs. Alors, par définition de $\rho\iota$, on a l'égalité

$$\Pi = \bigoplus_{\mathbf{n} \in \mathcal{N}} \text{ind}_{W_{\mathbf{n}}}^{W_{n_1} \times W_{n_2}} (\text{sgn}_{CD,\mathbf{n}} \otimes \text{res}_{W_{\mathbf{n}}}^{W_{n'} \times W_{n''}} (\rho_{\lambda',\epsilon'} \otimes \rho_{\lambda'',\epsilon''})).$$

Le terme $m(\rho_1, \rho_2^\zeta; \rho_{\lambda',\epsilon'}, \rho_{\lambda'',\epsilon''})$ est la multiplicité de $\rho_1 \otimes \rho_2^\zeta$ dans Π . Supposons que cette multiplicité soit non nulle. On peut alors fixer $\mathbf{n} \in \mathcal{N}$ et une représentation irréductible $\rho_{\mathbf{n}} = \rho'_1 \otimes \rho'_2 \otimes \rho''_1 \otimes \rho''_2$ de $W_{\mathbf{n}}$ telle que

$$\rho_{\mathbf{n}} \text{ intervient dans } \text{res}_{W_{\mathbf{n}}}^{W_{n'} \times W_{n''}} (\rho_{\lambda',\epsilon'} \otimes \rho_{\lambda'',\epsilon''}), \quad (41)$$

et telle que $\rho_1 \otimes \rho_2^\zeta$ intervient dans $\text{ind}_{W_{\mathbf{n}}}^{W_{n_1} \times W_{n_2}} (\text{sgn}_{CD,\mathbf{n}} \otimes \rho_{\mathbf{n}})$. Cette dernière condition équivaut à

$$\text{sgn}_{CD,\mathbf{n}} \otimes \rho_{\mathbf{n}} \text{ intervient dans } \text{res}_{W_{\mathbf{n}}}^{W_{n_1} \times W_{n_2}} (\rho_1 \otimes \rho_2^\zeta). \quad (42)$$

Introduisons les couples de partitions qui paramètrent les différentes représentations intervenant, que l'on note avec les mêmes indices et exposants que celles-ci : par exemple (α'_1, β'_1) paramètre ρ'_1 . On fait une exception, dont la raison sera expliquée plus loin : (α_2, β_2) paramètre ρ_2^+ . Remarquons que

$$\text{sgn}_{CD,\mathbf{n}} \otimes \rho_{\mathbf{n}} = \rho'_1 \otimes \rho'_2 \otimes \rho''_1 \otimes (\text{sgn}_{CD} \otimes \rho''_2),$$

et que $\text{sgn}_{CD} \otimes \rho''_2$ est paramétrée par (β''_2, α''_2) . Supposons d'abord $\zeta = +$ (on identifie ci-dessous les signes \pm à ± 1). En appliquant (38), les relations (25) et (26) entraînent :

$$\alpha' \leq \alpha'_1 + \alpha'_2, \quad \beta' \leq \beta'_1 + \beta'_2, \quad \alpha'' \leq \alpha''_1 + \alpha''_2, \quad \beta'' \leq \beta''_1 + \beta''_2; \quad (43)$$

$$\alpha'_1 \cup \alpha''_1 \leq \alpha_1, \quad \beta'_1 \cup \beta''_1 \leq \beta_1, \quad \alpha'_2 \cup \beta''_2 \leq \alpha_2, \quad \beta'_2 \cup \alpha''_2 \leq \beta_2. \quad (44)$$

Ce sont exactement les relations de [Waldspurger 2001, p. 377]. Plus précisément, dans cette référence, on considère le cas où V est symplectique. Les données ι_1 et ι_2 sont $(\lambda_1, \tau_1, \delta_1)$ et $(\lambda_2, \tau_2, \delta_2)$. Le terme ι_0 est $(0, 0, 1)$. Si maintenant $\zeta = -$, la représentation ρ_2^- est paramétrée par (β_2, α_2) et on obtient des relations comme ci-dessus, où on permute α_2 et β_2 . On voit que seules les dernières relations sont modifiées. La condition (44) devient :

$$\alpha'_1 \cup \alpha''_1 \leq \alpha_1, \quad \beta'_1 \cup \beta''_1 \leq \beta_1, \quad \beta'_2 \cup \alpha''_2 \leq \alpha_2, \quad \alpha'_2 \cup \beta''_2 \leq \beta_2. \quad (45)$$

Ce sont de nouveau les relations de [Waldspurger 2001, p. 377], où maintenant le terme ι_0 est $(0, 0, -1)$. Remarquons qu'en [Waldspurger 2001], on a supposé $\alpha_2 \geq \beta_2$ pour l'ordre lexicographique, ce qui explique que l'on a défini ci-dessus le couple (α_2, β_2) comme celui qui paramètre ρ_2^+ .

La condition (40) et l'existence de couples de partitions (α'_1, β'_1) , etc., vérifiant les conditions (43) et (44) ou (45) équivalent à l'appartenance du quadruplet $(\lambda', \epsilon'; \lambda'', \epsilon'')$ à l'ensemble $\mathcal{I}_L(\iota_1, \iota_2; \iota_0)$ défini en [Waldspurger 2001, p. 377] (le L de cette référence est un réseau autodual dont la seule utilité, dans le chapitre en question, est de déterminer les entiers n' et n''). La proposition XI.28 de [Waldspurger 2001] affirme qu'on a alors l'inégalité $\lambda' \cup \lambda'' \leq \lambda$. Cela prouve la proposition 1.12.

On revient aux hypothèses du présent paragraphe, c'est-à-dire que λ_1 et λ_2 induisent régulièrement λ . Supposons satisfaite la condition (B) $^\zeta$. Les hypothèses ci-dessus sont aussi satisfaites, donc $(\lambda', \epsilon'; \lambda'', \epsilon'')$ appartient à l'ensemble $\mathcal{I}_L(\iota_1, \iota_2; \iota_0)$, où $\iota_0 = (0, 0, \zeta)$. De plus, on a par hypothèse l'égalité $\lambda' \cup \lambda'' = \lambda$. Le (i) de la proposition XI.29 de [Waldspurger 2001] affirme alors que $(\lambda', \epsilon'; \lambda'', \epsilon'')$ appartient à l'ensemble $\mathcal{I}_L^{\max}(\iota_1, \iota_2; \iota_0)$ défini page 380 de [Waldspurger 2001]. Compte tenu de l'hypothèse que l'induction est régulière, cette appartenance signifie précisément que (A) $^\zeta$ est satisfaite. Inversement, supposons vérifiée cette condition (A) $^\zeta$. Comme on vient de le dire, cela signifie que $(\lambda', \epsilon'; \lambda'', \epsilon'')$ appartient à l'ensemble $\mathcal{I}_L^{\max}(\iota_1, \iota_2; \iota_0)$, a fortiori à l'ensemble $\mathcal{I}_L(\iota_1, \iota_2; \iota_0)$. Par définition de celui-ci, cela entraîne que (40) est vérifié. Le (ii) de la proposition XI.29 de [Waldspurger 2001] affirme qu'il y a un unique quadruplet de partitions (α'_1, β'_1) , etc., vérifiant les relations (43) et (44) ou (45) et que, pour ce quadruplet, les inégalités figurant dans ces relations sont des égalités. La multiplicité $m(\rho_1, \rho_2^\zeta; \rho_{\lambda', \epsilon'}, \rho_{\lambda'', \epsilon''})$ est la somme sur les $\mathbf{n} \in \mathcal{N}$ et les représentations irréductibles $\rho_{\mathbf{n}}$ du produit de la multiplicité de $\rho_{\mathbf{n}}$ dans $\text{res}_{W_{\mathbf{n}}}^{W_{n'} \times W_{n''}}(\rho_{\lambda', \epsilon'} \otimes \rho_{\lambda'', \epsilon''})$ et de la multiplicité de $\text{sgn}_{CD, \mathbf{n}} \otimes \rho_{\mathbf{n}}$ dans $\text{res}_{W_{\mathbf{n}}}^{W_{n_1} \times W_{n_2}}(\rho_1 \otimes \rho_2^\zeta)$. On vient de voir qu'il y a un unique \mathbf{n} et une unique $\rho_{\mathbf{n}}$ pour lesquels ces multiplicités ne sont pas nulles. Pour ce couple, les inégalités (43) et (44) ou (45) sont des égalités. Grâce à (39), on en déduit que les multiplicités en question valent 1. Alors

$$m(\rho_1, \rho_2^\zeta; \rho_{\lambda', \epsilon'}, \rho_{\lambda'', \epsilon''}) = 1.$$

Donc (B) $^\zeta$ est vérifiée ainsi que la dernière assertion de la proposition 1.13. □

2. Calcul de caractères

2.1. Caractères de représentations. Dans cette deuxième section, on reprend les données et notations de [Waldspurger 2018 ; 2016b]. Rappelons les principales. Le corps de base F est local non-archimédien et de caractéristique nulle. On note p sa caractéristique résiduelle. Un entier $n \geq 1$ est fixé. On suppose

$$p > 6n + 4.$$

On considère deux espaces vectoriels sur F de dimension $2n + 1$, notés V_{iso} et V_{an} , munis de formes quadratiques non dégénérées Q_{iso} et Q_{an} . On note G_{iso} et G_{an} les groupes spéciaux orthogonaux de $(V_{\text{iso}}, Q_{\text{iso}})$ et $(V_{\text{an}}, Q_{\text{an}})$. On suppose G_{iso} déployé et G_{an} non quasi-déployé. Pour un indice $\sharp = \text{iso}$ ou an , on fixe une mesure de Haar sur $G_\sharp(F)$ comme en [Waldspurger 2016b, 1.1]. On note $\text{Irr}_{\text{unip}, \sharp}$

l'ensemble des classes d'isomorphismes de représentations admissibles irréductibles de $G_{\sharp}(F)$ qui sont de réduction unipotente, cf. [Waldspurger 2018, §1.3].

Soit $\pi \in \text{Irr}_{\text{unip}, \sharp}$. A π est associé son caractère-distribution, c'est-à-dire la forme linéaire Θ_{π} sur $C_c^{\infty}(G_{\sharp}(F))$ définie par $\Theta_{\pi}(f) = \text{trace } \pi(f)$. Restreignons-nous aux fonctions f dont le support est formé d'éléments compacts de $G_{\sharp}(F)$, c'est-à-dire d'éléments dont les valeurs propres dans une clôture algébrique de F sont de valuation nulle. La représentation π étant de niveau 0, on a donné dans [Waldspurger 2016a] une formule pour $\Theta_{\pi}(f)$, que nous allons expliciter.

Dans [Waldspurger 2016a, paragraphe 10], on a introduit un ensemble $\text{Fac}_{\text{max}}^*(G_{\sharp})$. A tout $(\mathcal{F}, \nu) \in \text{Fac}_{\text{max}}^*(G_{\sharp})$ sont associés un sous-groupe compact $K_{\mathcal{F}}^{\dagger}$ de $G_{\sharp}(F)$ et un sous-ensemble $K_{\mathcal{F}}^{\nu} \subset K_{\mathcal{F}}^{\dagger}$. Le groupe $G_{\sharp}(F)$ agit naturellement sur $\text{Fac}_{\text{max}}^*(G_{\sharp})$. Il résulte facilement des définitions que l'ensemble des orbites pour cette action est en bijection avec l'ensemble des triplets (n', n'', ζ) , où $(n', n'') \in D(n)$ (c'est-à-dire $n', n'' \in \mathbb{N}$ et $n' + n'' = n$) et $\zeta = \pm$, soumis aux restrictions suivantes :

- dans le cas où $\sharp = \text{iso}$, on a $\zeta = +$ si $n'' = 0$ et $\zeta = -$ si $n'' = 1$;
- dans le cas où $\sharp = \text{an}$, on a $n'' \geq 1$.

On peut choisir un ensemble de représentants des orbites dans $\text{Fac}_{\text{max}}^*(G_{\sharp})$ de sorte que, si un élément (\mathcal{F}, ν) de cet ensemble correspond à un triplet (n', n'', ζ) , le groupe $K_{\mathcal{F}}^{\dagger}$ soit égal au groupe $K_{n', n''}^{\pm}$ de [Waldspurger 2018, §1.2] et l'ensemble $K_{\mathcal{F}}^{\nu}$ soit égal à $K_{n', n''}^{\zeta}$.

Considérons un triplet (n', n'', ζ) comme ci-dessus. On dispose de la fonction $\text{proj}_{\text{cusp}}(\text{Res}_{n', n''}^{\zeta}(\pi)) \in \mathcal{R}^{\text{par, glob}}$, cf. [Waldspurger 2018, §1.5]. On peut considérer que c'est une fonction sur $K_{n', n''}^{\zeta}$, invariante par $K_{n', n''}^u$. On pose

$$\Theta_{\pi, \text{cusp}}(f) = \sum_{(n', n'', \zeta)} \text{mes}(K_{n', n''}^{\pm})^{-1} \int_{G_{\sharp}(F)} \int_{G_{\sharp}(F)} f(g^{-1}hg) \text{proj}_{\text{cusp}}(\text{Res}_{n', n''}^{\zeta}(\pi))(h) dh dg.$$

Cette intégrale est convergente dans cet ordre. Les (n', n'', ζ) sont soumis aux restrictions ci-dessus. Mais on peut en fait lever celles-ci parce la fonction $\text{proj}_{\text{cusp}}(\text{Res}_{n', n''}^{\zeta}(\pi))$ est nulle si elles ne sont pas vérifiées.

Considérons maintenant une partition $\mathbf{m} = (m_1 \geq \dots \geq m_t > 0) \in \mathcal{P}(\leq n)$ (c'est-à-dire $S(\mathbf{m}) := m_1 + \dots + m_t \leq n$), posons $n_0 = n - S(\mathbf{m})$. On suppose $n_0 \geq 1$ si $\sharp = \text{an}$. On associe à \mathbf{m} un sous-groupe de Levi $M \subset G_{\sharp}$. Avec les notations de [Waldspurger 2018, §1.1], c'est l'ensemble des éléments qui, pour tout $j = 1, \dots, t$, stabilisent les deux sous-espaces de V_{\sharp} engendrés respectivement par $v_{n_0+m_1+\dots+m_{j-1}+1}, \dots, v_{n_0+m_1+\dots+m_j}$ et par $v_{2n+2-n_0-m_1-\dots-m_j}, \dots, v_{2n+2-n_0-m_1-\dots-m_{j-1}-1}$. On a

$$M \simeq \text{GL}(m_1) \times \dots \times \text{GL}(m_t) \times G_{n_0, \sharp},$$

où $G_{n_0, \sharp}$ est l'analogue de G_{\sharp} quand n est remplacé par n_0 (ce groupe est trivial si $\sharp = \text{iso}$ et $n_0 = 0$). Pour tout $j = 1, \dots, t$, fixons un sous-groupe compact maximal $K_{m_j} \subset \text{GL}(m_j; F)$ et notons $K_{m_j}^u$ son radical pro- p -unipotent. On note $K_{\mathbf{m}}$ (resp. $K_{\mathbf{m}}^u$) le produit de ces groupes. On note aussi A_M le plus grand tore déployé central dans M , c'est-à-dire le produit des centres des groupes $\text{GL}(m_j)$. On a défini en [Waldspurger 2016a, paragraphe 11] un ensemble $\text{Fac}_{\text{max}}^*(M)_{G_{\sharp}\text{-comp}}$. À tout élément (\mathcal{F}_M, ν) de cet ensemble est associé un sous-groupe $K_{\mathcal{F}_M}^{\dagger}$ de $M(F)$. Le groupe $M(F)$ agit naturellement sur

$\text{Fac}_{\max}^*(M)_{G_{\sharp}\text{-comp}}$. On voit que l'ensemble des orbites est en bijection avec l'ensemble des triplets (n', n'', ζ) tels que $(n', n'') \in D(n_0)$ et $\zeta = \pm$, soumis aux restrictions similaires à celles ci-dessus. On peut choisir un ensemble de représentants des orbites de sorte que, si un élément (\mathcal{F}_M, ν) de cet ensemble correspond à un triplet (n', n'', ζ) , le groupe $K_{\mathcal{F}_M}^{\dagger}$ soit égal à $A_M(F)K_m \times K_{n', n''}^{\pm}$.

L'analogie pour ce groupe M de l'espace $\mathcal{R}^{\text{par, glob}}$ est l'espace

$$\mathcal{R}_m^{\text{par, glob}} = C^{\text{GL}(m_1)} \otimes \dots \otimes C^{\text{GL}(m_r)} \otimes \mathcal{R}_{n_0}^{\text{par, glob}},$$

cf. [Waldspurger 2018, §1.5]. On introduit l'application linéaire $\text{Res}^M : \mathbb{C}[\text{Irr}_{\text{unip}, M}] \rightarrow \mathcal{R}_m^{\text{par, glob}}$ analogue à Res . Soit P un sous-groupe parabolique de G_{\sharp} de composante de Levi M . Le semi-simplifié du module de Jacquet π_P s'identifie à un élément de $\mathbb{C}[\text{Irr}_{\text{unip}, M}]$. D'après [Waldspurger 2018, §1.5(1)] (qui résulte directement de [Moy et Prasad 1996, Proposition 6.7]) le terme $\text{Res}(\pi_P)$ ne dépend pas du choix de P et on a l'égalité

$$\text{Res}^M(\pi_P) = \text{res}_m \circ \text{Res}(\pi).$$

Notons ce terme $\text{Res}_m(\pi)$ et notons ses diverses composantes $\text{Res}_{m, n', n''}^{\zeta}(\pi)$. On dispose de la projection cuspidale $\text{proj}_{\text{cusp}}(\text{Res}_{m, n', n''}^{\zeta}(\pi))$. On peut considérer que c'est une fonction sur $K_m \times K_{n', n''}^{\zeta}$, invariante par $K_m^u \times K_{n', n''}^u$. Pour une fonction $\phi \in C_c^{\infty}(M(F))$, posons

$$\begin{aligned} \Theta_{\pi, \text{cusp}}^M(\phi) &= \sum_{(n', n'', \zeta)} \text{mes}((A_M(F)K_m \times K_{n', n''}^{\pm})/A_M(F))^{-1} \\ &\quad \times \int_{M(F)/A_M(F)} \int_{M(F)} \phi(m^{-1}ym) \text{proj}_{\text{cusp}}(\text{Res}_{m, n', n''}^{\zeta}(\pi))(y) dy dm. \end{aligned}$$

Cette intégrale converge dans cet ordre. Fixons un groupe P comme ci-dessus, notons U son radical unipotent. Fixons une mesure de Haar sur $U(F)$. De la mesure sur $M(F)$ (fixée comme en [Waldspurger 2016b, 1.1]) et de celle sur $U(F)$ se déduit une mesure invariante à gauche sur $P(F)$, puis une pseudo-mesure sur $P(F) \backslash G_{\sharp}(F)$ (pseudo parce qu'elle s'applique à des fonctions qui ne sont pas invariantes à gauche par $P(F)$ mais qui se transforment selon le module usuel δ_P). Définissons une fonction f_U sur $M(F)$ par

$$f_U(m) = \delta_P(m)^{1/2} \int_{U(F)} f(mu) du.$$

En vertu de notre hypothèse sur le support de f , on peut aussi bien supprimer le facteur $\delta_P(m)^{1/2}$, il vaut 1 si l'intégrale est non nulle. D'autre part, pour $g \in G_{\sharp}(F)$, on définit la fonction ${}^g f$ sur $G_{\sharp}(F)$ par ${}^g f(h) = f(g^{-1}hg)$. On pose

$$\Theta_{\pi, m, \text{cusp}}(f) = \int_{P(F) \backslash G_{\sharp}(F)} \Theta_{\pi, \text{cusp}}^M(({}^g f)_U) dg.$$

Ce terme ne dépend pas du choix de P . Remarquons que le terme $\Theta_{\pi, \text{cusp}}(f)$ introduit plus haut est égal à $\Theta_{\pi, \emptyset, \text{cusp}}(f)$, où on a noté \emptyset l'unique partition de 0.

Rappelons que l'on suppose que le support de f est formé d'éléments compacts de $G_{\sharp}(F)$. Le théorème 12 de [Waldspurger 2016a] affirme l'égalité

$$\Theta_{\pi}(f) = \sum_m 2^{-l(m)} \mathrm{mult}_m^{-1} \Theta_{\pi, m, \mathrm{cusp}}(f), \tag{*}$$

où on a posé

$$\mathrm{mult}_m = \prod_{i \geq 1} \mathrm{mult}_m(i)!$$

et noté $l(m)$ le nombre de termes non nuls de m (qui est noté t plus haut). La somme porte sur les partitions indiquées plus haut, c'est-à-dire $m \in \mathcal{P}(\leq n)$ si $\sharp = \mathrm{iso}$ et $m \in \mathcal{P}(\leq n-1)$ si $\sharp = \mathrm{an}$.

Remarque. Le théorème 12 de [Waldspurger 2016a] n'est pas tout-à-fait énoncé comme ci-dessus mais on voit facilement que les deux énoncés sont équivalents.

2.2. Un lemme élémentaire. Soit $\sharp = \mathrm{iso}$ ou an . Pour $g \in G_{\sharp}(F)$, on dit que g est topologiquement unipotent si et seulement si $\lim_{m \rightarrow \infty} g^{p^m} = 1$. Pour $X \in \mathfrak{g}_{\sharp}(F)$, on dit que X est topologiquement nilpotent si et seulement si $\lim_{m \rightarrow \infty} X^m = 0$. Sous certaines hypothèses sur p (du type $p > A + B \mathrm{val}_F(p)$, où val_F est la valuation usuelle de F), l'exponentielle est définie sur l'ensemble des éléments topologiquement nilpotents de $\mathfrak{g}_{\sharp}(F)$ et est une bijection de cet ensemble sur celui des éléments topologiquement unipotents de $G_{\sharp}(F)$. Pour simplifier les hypothèses sur p , on remplace l'exponentielle par l'application E définie par $E(X) = (1 + X/2)/(1 - X/2)$. Pour $p > 2$, c'est une bijection de l'ensemble des éléments topologiquement nilpotents de $\mathfrak{g}_{\sharp}(F)$ sur celui des éléments topologiquement unipotents de $G_{\sharp}(F)$. Rappelons que l'on a supposé $p > 6n + 4$, a fortiori $p > 2$.

Soit $(n', n'') \in D(n)$. On suppose $n'' \geq 1$ si $\sharp = \mathrm{an}$. On a défini en [Waldspurger 2018, §1.2] le réseau $L_{n', n''} \subset V_{\sharp}$, le sous-groupe compact $K_{n', n''}^+$ de $G_{\sharp}(F)$ et son radical pro- p -unipotent $K_{n', n''}^u$. On définit deux réseaux $\mathfrak{k}_{n', n''}$ et $\mathfrak{k}_{n', n''}^u$ de $\mathfrak{g}_{\sharp}(F)$: ce sont les sous-ensembles des éléments $X \in \mathfrak{g}_{\sharp}(F)$ tels que $X(L_{n', n''}) \subset L_{n', n''}$ (ce qui entraîne aussi $X(L_{n', n''}^*) \subset L_{n', n''}^*$) (resp. $X(L_{n', n''}) \subset \varpi L_{n', n''}^*$ et $X(L_{n', n''}^*) \subset L_{n', n''}$). On vérifie que, pour $X \in \mathfrak{g}_{\sharp}(F)$ topologiquement nilpotent, on a

$$X \in \mathfrak{k}_{n', n''} \iff E(X) \in K_{n', n''}^+, \quad X \in \mathfrak{k}_{n', n''}^u \iff E(X) \in K_{n', n''}^u.$$

Posons $\mathbf{G} = \mathrm{SO}(2n' + 1) \times \mathrm{SO}(2n'')_{\sharp}$, avec les notations de [Waldspurger 2018, §1.1]. On sait que $K_{n', n''}^+ / K_{n', n''}^u \simeq \mathbf{G}(\mathbb{F}_q)$. Notons \mathfrak{g} l'algèbre de Lie de \mathbf{G} . On vérifie que $\mathfrak{k}_{n', n''} / \mathfrak{k}_{n', n''}^u \simeq \mathfrak{g}(\mathbb{F}_q)$. On note encore E l'application définie par $E(X) = (1 + X/2)/(1 - X/2)$ sur l'ensemble des éléments nilpotents de $\mathfrak{g}(\mathbb{F}_q)$. C'est une bijection de cet ensemble sur celui des éléments unipotents de $\mathbf{G}(\mathbb{F}_q)$.

Soit $f \in C_c^{\infty}(G_{\sharp}(F))$. Supposons que le support de f est formé d'éléments topologiquement unipotents. On déduit de f une fonction $f_{\mathrm{Lie}} \in C_c^{\infty}(\mathfrak{g}_{\sharp}(F))$. Son support est formé d'éléments topologiquement nilpotents. Pour un tel élément X , on a $f_{\mathrm{Lie}}(X) = f(E(X))$. On déduit aussi de f une fonction f_{red} sur $K_{n', n''}^+$ telle que, pour tout $g \in K_{n', n''}^+$,

$$f_{\mathrm{red}}(g) = \int_{K_{n', n''}^u} f(gh) dh.$$

Cette fonction est invariante par $K_{n',n''}^u$, on peut considérer que c'est une fonction sur $\mathbf{G}(\mathbb{F}_q)$. Elle est alors à support unipotent. On en déduit une fonction $f_{\text{red,Lie}}$ sur $\mathfrak{g}(\mathbb{F}_q)$: celle-ci est à support nilpotent et, pour un élément nilpotent $X \in \mathfrak{g}(\mathbb{F}_q)$, on a l'égalité $f_{\text{red,Lie}}(X) = f_{\text{red}}(E(X))$. Enfin, on déduit de f_{Lie} une fonction $f_{\text{Lie,red}}$ sur $\mathfrak{k}_{n',n''}$: pour X dans cet ensemble,

$$f_{\text{Lie,red}}(X) = \int_{\mathfrak{k}_{n',n''}^u} f_{\text{Lie}}(X + Y) dY.$$

Cette fonction est invariante par translations par $\mathfrak{k}_{n',n''}^u$. On peut la considérer comme une fonction sur $\mathfrak{g}(\mathbb{F}_q)$. Elle est alors à support nilpotent.

Lemme. $f_{\text{red,Lie}} = f_{\text{Lie,red}}$.

Preuve. En détaillant les définitions, on voit qu'il s'agit de démontrer l'assertion suivante :

- (46) Soit $X \in \mathfrak{k}_{n',n''}$ un élément topologiquement nilpotent ; alors l'application $Y \mapsto E(X)^{-1}E(X + Y)$ envoie bijectivement $\mathfrak{k}_{n',n''}^u$ sur $K_{n',n''}^u$ et préserve les mesures.

La démonstration est élémentaire ; on la laisse au lecteur. □

On a effectué les constructions ci-dessus pour le groupe $G_{\mathfrak{k}}$ afin de ne pas introduire de notations supplémentaires. Mais il est clair que les mêmes constructions et le même lemme valent pour les groupes de Levi de $G_{\mathfrak{k}}$ et nous les utiliserons pour ceux-ci.

2.3. Calcul du caractère sur les éléments topologiquement unipotents. Pour $\mathfrak{k} = \text{iso}$ ou an , soit $\pi \in \text{Irr}_{\text{unip},\mathfrak{k}}$. On a défini l'élément $\text{Res}(\pi) \in \mathcal{R}^{\text{par, glob}}$ et l'isomorphisme $k : \mathcal{R}^{\text{glob}} \rightarrow \mathcal{R}^{\text{par, glob}}$ en [Waldspurger 2018, §1.5 et §1.9]. On note κ_π l'élément de $\mathcal{R}^{\text{glob}}$ tel que $\text{Res}(\pi) = k(\kappa_\pi)$. Soit $f \in C_c^\infty(G_{\mathfrak{k}}(F))$. On suppose que tout élément du support de f est topologiquement unipotent. Un tel élément est compact, donc $\Theta_\pi(f)$ est donné par la formule (*) de 2.1. Nous allons expliciter cette formule à l'aide de l'élément $\kappa_\pi \in \mathcal{R}^{\text{glob}}$.

Considérons un entier $n_0 \in \{0, \dots, n\}$, une décomposition $n_0 = n' + n''$ et une partition $\mathbf{m} = (m_1, \dots, m_t > 0) \in \mathcal{P}(n - n_0)$. Ces données sont soumises aux mêmes restrictions qu' en 2.1 : si $\mathfrak{k} = \text{an}$, on a $n_0 \geq 1$ et $n'' \geq 1$. On a associé à ces données un groupe de Levi M de $G_{\mathfrak{k}}$. Soit $\phi \in C_c^\infty(M(F))$, supposons que le support de ϕ est formé d'éléments topologiquement unipotents. On va d'abord calculer

$$I = \int_{M(F)} \phi(y) \sum_{\zeta} \text{proj}_{\text{cusp}}(\text{Res}_{\mathbf{m},n',n''}^\zeta(\pi))(y) dy.$$

La somme porte sur les signes $\zeta = \pm$, soumis aux conditions : si $\mathfrak{k} = \text{iso}$, on a $\zeta = +$ si $n'' = 0$ et $\zeta = -$ si $n'' = 1$. La deuxième fonction dans l'intégrale est à support dans le groupe compact $K_{\mathbf{m}} \times K_{n',n''}^\pm$ et est invariante par $K_{\mathbf{m}}^u \times K_{n',n''}^u$. On peut l'identifier à une fonction sur le groupe $\mathbf{M}^\pm(\mathbb{F}_q)$, où

$$\mathbf{M}^\pm = \text{GL}(m_1) \times \dots \times \text{GL}(m_t) \times \text{SO}(2n' + 1) \times O(2n'')_{\mathfrak{k}}.$$

Définissons une fonction ϕ_{res} sur $K_{\mathbf{m}} \times K_{n',n''}^{\pm}$ par

$$\phi_{\mathrm{res}}(y) = \int_{K_{\mathbf{m}}^u \times K_{n',n''}^u} \phi(yh) dh.$$

On peut la considérer elle-aussi comme une fonction sur $\mathbf{M}^{\pm}(\mathbb{F}_q)$. On a l'égalité

$$I = \sum_{y \in \mathbf{M}^{\pm}(\mathbb{F}_q)} \phi_{\mathrm{res}}(y) \sum_{\zeta} \mathrm{proj}_{\mathrm{cusp}}(\mathrm{Res}_{\mathbf{m},n',n''}^{\zeta}(\pi))(y).$$

On dispose de l'application

$$\mathrm{res}_{\mathbf{m}} : \mathcal{R}^{\mathrm{glob}} \rightarrow \mathbb{C}[\widehat{\mathfrak{S}}_{m_1}] \times \cdots \times \mathbb{C}[\widehat{\mathfrak{S}}_{m_t}] \otimes \mathcal{R}_{n_0}^{\mathrm{glob}}$$

obtenue en itérant la construction de [Waldspurger 2018, §1.8]. Notons $\Gamma_{n',n''}$ l'ensemble des $\gamma = (r', r'', N', N'') \in \Gamma_{n_0}$ tels que $r'^2 + r' + N' = n'$, $r''^2 + N'' = n''$. Pour un tel γ , notons $\mathrm{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma}$ la composante dans

$$\mathbb{C}[\widehat{\mathfrak{S}}_{m_1}] \times \cdots \times \mathbb{C}[\widehat{\mathfrak{S}}_{m_t}] \otimes \mathcal{R}_{\gamma}$$

de $\mathrm{res}_{\mathbf{m}}(\kappa_{\pi})$. Excluons d'abord le cas où $\sharp = \mathrm{iso}$ et $n'' = 1$. On voit que

$$\sum_{\zeta} \mathrm{proj}_{\mathrm{cusp}}(\mathrm{Res}_{\mathbf{m},n',n''}^{\zeta}(\pi)) = \sum_{\gamma \in \Gamma_{n',n''}} \mathrm{proj}_{\mathrm{cusp}} \circ k^M(\mathrm{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma}).$$

Fixons $\gamma = (r', r'', N', N'') \in \Gamma_{n',n''}$. Dans [Mœglin et Waldspurger 2003, 2.12 et 2.13] on a introduit des fonctions $k(r', w')$ sur $\mathrm{SO}(2n'+1)(\mathbb{F}_q)$ pour $w' \in W_{N'}$ et $k(r'', w)$ sur $O(2n'')_{\sharp}(\mathbb{F}_q)$ pour $w'' \in W_{N''}$. Une construction analogue vaut pour les groupes $\mathrm{GL}(m_j)$: pour $w \in \mathfrak{S}_{m_j}$, on définit une fonction $k(w)$ sur $\mathrm{GL}(m_j; \mathbb{F}_q)$. Posons

$$W(\mathbf{m}, N', N'') = \mathfrak{S}_{m_1} \times \cdots \times \mathfrak{S}_{m_t} \times W_{N'} \times W_{N''}.$$

La fonction $\mathrm{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma}$ est définie sur ce groupe. Pour $w = (w_1, \dots, w_t, w', w'') \in W(\mathbf{m}, N', N'')$, on pose $k(r', r''; w) = k(w_1) \otimes \cdots \otimes k(w_t) \otimes k(r', w') \otimes k(r'', w'')$. Il résulte des définitions que

$$k^M(\mathrm{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma}) = |W(\mathbf{m}, N', N'')|^{-1} \sum_{w \in W(\mathbf{m}, N', N'')} \mathrm{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma}(w) k(r', r''; w).$$

Dans chacun des groupes \mathfrak{S}_{m_j} , $W_{N'}$ et $W_{N''}$, on définit usuellement la notion d'élément elliptique. L'application k entrelace la projection $\mathrm{proj}_{\mathrm{cusp}}$ et la projection sur les éléments elliptiques. Donc

$$\mathrm{proj}_{\mathrm{cusp}} \circ k^M(\mathrm{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma}) = |W(\mathbf{m}, N', N'')|^{-1} \sum_{w \in W(\mathbf{m}, N', N'')_{\mathrm{ell}}} \mathrm{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma}(w) k(r', r''; w),$$

où $W(\mathbf{m}, N', N'')_{\mathrm{ell}}$ est le sous-ensemble des éléments elliptiques de $W(\mathbf{m}, N', N'')$. On obtient

$$I = \sum_{\gamma = (r', r'', N', N'') \in \Gamma_{n',n''}} |W(\mathbf{m}, N', N'')|^{-1} \sum_{w \in W(\mathbf{m}, N', N'')_{\mathrm{ell}}} \mathrm{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma}(w) \sum_{y \in \mathbf{M}^{\pm}(\mathbb{F}_q)} \phi_{\mathrm{res}}(y) k(r', r''; w)(y).$$

Les hypothèses sur le support de ϕ entraînent que ϕ_{res} est à support unipotent. D’après la proposition 2.16 de [Mœglin et Waldspurger 2003], $k(r', r''; w)$ est nulle sur les unipotents sauf si $r' = r'' = 0$. Il ne reste qu’un seul γ qui contribue, à savoir l’élément $\gamma_{n', n''} = (0, 0, n', n'')$. D’où

$$I = |W(\mathfrak{m}, n', n'')|^{-1} \sum_{w \in W(\mathfrak{m}, n', n'')_{\text{ell}}} \text{res}_{\mathfrak{m}}(\kappa_{\pi})_{\gamma_{n', n''}}(w) \sum_{y \in \mathbf{M}^{\pm}(\mathbb{F}_q)} \phi_{\text{res}}(y)k(w)(y),$$

où on a posé simplement $k(w) = k(0, 0; w)$. A ce point, on peut supprimer l’hypothèse restrictive faite plus haut. Si $\sharp = \text{iso}$ et $n'' = 1$, on a $I = 0$ car on se limite à $\zeta = -$ et $K_{n'', \text{iso}}^-$ ne contient pas d’élément topologiquement unipotent. Mais la formule ci-dessus donne le même résultat, car pour l’unique élément elliptique $w'' \in W_{1, \text{ell}}$, on a $k(0, w'')_{\text{iso}} = 0$, cf. [Mœglin et Waldspurger 2003, 2.13]. Notons \mathbf{M} la composante neutre de \mathbf{M}^{\pm} et \mathfrak{m} son algèbre de Lie. On dispose de l’application E de 2.2, qui est une bijection de l’ensemble $\mathfrak{m}_{\text{nil}}(\mathbb{F}_q)$ des éléments nilpotents de $\mathfrak{m}(\mathbb{F}_q)$ sur l’ensemble $\mathbf{M}_{\text{unip}}(\mathbb{F}_q)$ des éléments unipotents de $\mathbf{M}(\mathbb{F}_q)$. Notons $\phi_{\text{red, Lie}}$ la fonction sur $\mathfrak{m}(\mathbb{F}_q)$ qui est nulle hors des éléments nilpotents et qui vérifie $\phi_{\text{red, Lie}}(X) = \phi_{\text{red}}(E(X))$ pour tout X nilpotent. Pour $w \in W(\mathfrak{m}, n', n'')_{\text{ell}}$, définissons de même une fonction $k(w)_{\text{Lie}}$. On obtient

$$I = |W(\mathfrak{m}, n', n'')|^{-1} \sum_{w \in W(\mathfrak{m}, n', n'')_{\text{ell}}} \text{res}_{\mathfrak{m}}(\kappa_{\pi})_{\gamma_{n', n''}}(w) I_w, \tag{47}$$

où

$$I_w = \sum_{X \in \mathfrak{m}(\mathbb{F}_q)} \phi_{\text{red, Lie}}(X)k(w)_{\text{Lie}}(X).$$

Fixons $w = (w_1, \dots, w_t, w', w'') \in W(\mathfrak{m}, n', n'')_{\text{ell}}$. On peut supposer $\text{sgn}_{CD}(w'') = 1$ si $\sharp = \text{iso}$, $\text{sgn}_{CD}(w'') = -1$ si $\sharp = \text{an}$, sinon la fonction $k(0, w'')$ est nulle sur $\text{SO}_{\sharp}(\mathbb{F}_q)$, cf. [Mœglin et Waldspurger 2003, 2.13]. A tout w_j est associée une classe de conjugaison de sous-tore maximal elliptique dans $\text{GL}(m_j)$ (qui est d’ailleurs l’unique telle classe). A w' (resp. w'') est associée une classe de conjugaison de sous-tore maximal elliptique dans $\text{SO}(2n' + 1)$ (resp. $\text{SO}(2n'')_{\sharp}$). On fixe des tores dans ces classes de conjugaison et on note T_w leur produit qui est donc un sous-tore maximal elliptique dans \mathbf{M} . On dispose de l’induction de Deligne–Lusztig de T_w à \mathbf{M} . Ce foncteur vaut aussi pour les algèbres de Lie. Notons \mathfrak{t}_w l’algèbre de Lie de T_w et considérons la fonction caractéristique de $\{0\}$ dans $\mathfrak{t}_w(\mathbb{F}_q)$. On note Q_w son image par induction de Deligne–Lusztig, qui est une fonction sur $\mathfrak{m}(\mathbb{F}_q)$, à support nilpotent. On a l’égalité

$$k(w)_{\text{Lie}} = 2^{\beta} Q_w, \quad \text{où } \beta = 0 \text{ si } n'' = 0, \quad \beta = 1 \text{ si } n'' > 0. \tag{48}$$

En effet, d’après nos définitions de [Mœglin et Waldspurger 2003, 2.12 et 2.13], $k(w)$ est égal à $(-1)^n 2^{\beta}$ fois la trace d’un Frobenius sur un faisceau-caractère. D’après [Lusztig 1990, Theorem 1.14], cette trace est égale, sur les unipotents, $(-1)^n$ fois l’image par induction de Deligne–Lusztig de la fonction caractéristique de 1 dans $T_w(\mathbb{F}_q)$. En descendant par l’application E à l’algèbre de Lie, on obtient (48).

En [Waldspurger 2016b, 1.1], on a fixé un caractère ψ_F de F de conducteur $\varpi \mathfrak{o}$. Il lui est associé un caractère de \mathbb{F}_q grâce auquel on définit, comme en [Waldspurger 2016b, 1.1], une transformation de

Fourier $\varphi \mapsto \hat{\varphi}$ dans $C_c^\infty(\mathfrak{m}(\mathbb{F}_q))$. On la normalise de sorte que $\hat{\hat{\varphi}}(X) = \varphi(-X)$. D'après la proposition 7.2 et l'égalité 6.15(a) de [Lusztig 1992], on a l'égalité

$$\hat{Q}_w(X) = \text{sgn}(w)q^{-n/2}Q_w(X) \text{ pour tout élément nilpotent } X \in \mathfrak{m}_{\text{nil}}(\mathbb{F}_q). \tag{49}$$

Fixons un point $X_w \in \mathfrak{t}_w(\mathbb{F}_q)$ en position générale. Notons $\varphi[X_w]$ la fonction caractéristique de la classe de conjugaison par $M(\mathbb{F}_q)$ de X_w . D'après [Waldspurger 2001, proposition II.8], on a l'égalité

$$\hat{\varphi}[X_w](X) = \hat{Q}_w(X) \text{ pour tout élément nilpotent } X \in \mathfrak{m}_{\text{nil}}(\mathbb{F}_q). \tag{50}$$

En rassemblant (48), (49) et (50), on obtient l'égalité $k(w)_{\text{Lie}}(X) = \text{sgn}(w)q^{n/2}2^\beta \hat{\varphi}[X_w](X)$, pour $X \in \mathfrak{m}_{\text{nil}}(\mathbb{F}_q)$. D'où

$$I_w = \text{sgn}(w)q^{n/2}2^\beta \sum_{X \in \mathfrak{m}(\mathbb{F}_q)} \phi_{\text{red,Lie}}(X) \hat{\varphi}[X_w](X),$$

puis, par la formule de Parseval,

$$I_w = \text{sgn}(w)q^{n/2}2^\beta \sum_{X \in \mathfrak{m}(\mathbb{F}_q)} \hat{\phi}_{\text{red,Lie}}(X) \varphi[X_w](X).$$

Ou encore, en explicitant la fonction $\varphi[X_w]$,

$$I_w = \text{sgn}(w)q^{n/2}2^\beta |\mathbf{T}_w(\mathbb{F}_q)|^{-1} \sum_{x \in M(\mathbb{F}_q)} \hat{\phi}_{\text{red,Lie}}(x^{-1}X_w x).$$

La conjugaison se fait ici par le groupe $M(\mathbb{F}_q)$ et on rappelle que M est la composante neutre de M^\pm . Mais, dans la formule ci-dessus, on peut remplacer X_w par un conjugué quelconque par un élément de $M^\pm(\mathbb{F}_q)$. Un tel conjugué vérifie les mêmes propriétés que X_w . On peut donc remplacer la conjugaison par $M(\mathbb{F}_q)$ par la conjugaison par $M^\pm(\mathbb{F}_q)$ tout entier, à condition de diviser par $[M^\pm(\mathbb{F}_q) : M(\mathbb{F}_q)]$, qui vaut précisément 2^β . D'où

$$I_w = \text{sgn}(w)q^{n/2}|\mathbf{T}_w(\mathbb{F}_q)|^{-1} \sum_{x \in M^\pm(\mathbb{F}_q)} \hat{\phi}_{\text{red,Lie}}(x^{-1}X_w x). \tag{51}$$

Notons \mathfrak{m} l'algèbre de Lie de M . Comme en 2.2, on définit une fonction ϕ_{Lie} sur $\mathfrak{m}(F)$: elle est à support topologiquement nilpotent ; pour $X \in \mathfrak{m}(F)$ topologiquement nilpotent, on a $\phi_{\text{Lie}}(X) = \phi(E(X))$. On en déduit une fonction $\phi_{\text{Lie,red}}$ sur $\mathfrak{k}_m \oplus \mathfrak{k}_{n',n''}$ (avec une définition évidente de \mathfrak{k}_m et, ci-dessous, de $\mathfrak{k}_{n',n''}^u$) par

$$\phi_{\text{Lie,red}}(X) = \int_{\mathfrak{k}_m^u \oplus \mathfrak{k}_{n',n''}^u} \phi_{\text{Lie}}(X + Y) dY$$

pour tout $X \in \mathfrak{k}_m \oplus \mathfrak{k}_{n',n''}$. On peut considérer que c'est une fonction sur $\mathfrak{m}(\mathbb{F}_q)$. Le lemme 2.2 dit que $\phi_{\text{red,Lie}} = \phi_{\text{Lie,red}}$. On dispose de la fonction $\hat{\phi}_{\text{Lie}}$ (la transformée de Fourier de ϕ_{Lie}) dont on déduit comme ci-dessus une fonction $(\hat{\phi}_{\text{Lie}})_{\text{red}}$ sur $\mathfrak{k}_m \oplus \mathfrak{k}_{n',n''}$, que l'on peut considérer comme une fonction sur $\mathfrak{m}(\mathbb{F}_q)$. On vérifie l'égalité

$$(\hat{\phi}_{\text{Lie}})_{\text{red}} = \hat{\phi}_{\text{Lie,red}}.$$

Dans la formule (51), remplaçons $\hat{\phi}_{\text{red,Lie}}$ par $(\hat{\phi}_{\text{Lie}})_{\text{red}}$. Les termes de la formule vivent dans $\mathfrak{m}(\mathbb{F}_q)$ mais on peut les relever dans $\mathfrak{k}_m \oplus \mathfrak{k}_{n',n''}$. On relève ainsi X_w en un élément de ce réseau que l'on note X_w . La somme en $x \in M^\pm(\mathbb{F}_q)$ devient une intégrale sur $K_m \times K_{n',n''}^\pm$, divisée par la mesure de $K_m \times K_{n',n''}$. On obtient

$$I_w = \text{sgn}(w)q^{n/2}|T_w(\mathbb{F}_q)|^{-1} \text{mes}(K_m \times K_{n',n''})^{-1} \int_{K_m \times K_{n',n''}^\pm} (\hat{\phi}_{\text{Lie}})_{\text{red}}(x^{-1}X_w x) dx. \tag{52}$$

Notons T_w le centralisateur de X_w et \mathfrak{t}_w son algèbre de Lie. Le tore T_w est non ramifié sur F et possède une structure naturelle sur \mathfrak{o} . On a $\mathfrak{t}_w(\mathfrak{o}) = \mathfrak{t}_w(F) \cap (\mathfrak{k}_m \oplus \mathfrak{k}_{n',n''})$ et $\varpi \mathfrak{t}_w(\mathfrak{o}) = \mathfrak{t}_w(F) \cap (\mathfrak{k}_m^u \oplus \mathfrak{k}_{n',n''}^u)$. Posons $\mathcal{X}_w = X_w + \varpi \mathfrak{t}_w(\mathfrak{o})$. Montrons que

pour tout $x \in K_m \times K_{n',n''}^\pm$,

$$(\hat{\phi}_{\text{Lie}})_{\text{red}}(x^{-1}X_w x) = \text{mes}(\mathcal{X}_w)^{-1} \int_{K_m \times K_{n',n''}^u} \int_{\mathcal{X}_w} \hat{\phi}_{\text{Lie}}(x^{-1}y^{-1}Y_w y x) dY_w dy. \tag{53}$$

On se ramène immédiatement au cas $x = 1$ en conjuguant par x la fonction $\hat{\phi}_{\text{Lie}}$. Supposons donc $x = 1$. Posons $T_w(F)^u = T_w(F) \cap (K_m^u \times K_{n',n''}^u)$. C'est l'image par E de $\varpi \mathfrak{t}_w(\mathfrak{o})$, on a donc $\text{mes}(T_w(F)^u) = \text{mes}(\mathcal{X}_w)$. Les éléments Y_w appartiennent à $\mathfrak{t}_w(F)$ donc $T_w(F)$ commute à ces éléments. On peut remplacer l'intégrale en $y \in K_m^u \times K_{n',n''}^u$ du membre de droite ci-dessus par une intégrale en $y \in T_w(F)^u \setminus (K_m^u \times K_{n',n''}^u)$, multipliée par $\text{mes}(\mathcal{X}_w)$. Ce facteur fait disparaître son inverse qui figure dans ce membre de droite. Considérons l'application

$$\iota : T_w(F)^u \setminus (K_m^u \times K_{n',n''}^u) \times \varpi \mathfrak{t}_w(\mathfrak{o}) \rightarrow \mathfrak{m}(F), \quad (y, Z) \mapsto y^{-1}(X_w + Z)y - X_w.$$

Il est clair que son image est contenue dans $\mathfrak{k}_m^u \oplus \mathfrak{k}_{n',n''}^u$. Montrons qu'elle est injective. Si (y, Z) et (y', Z') ont même image, on a $y^{-1}(X_w + Z)y = y'^{-1}(X_w + Z')y'$. Le point X_w est en position générale et ses valeurs propres (dans une clôture algébrique $\overline{\mathbb{F}}_q$ de \mathbb{F}_q) sont distinctes. Les valeurs propres de $X_w + Z$ et $X_w + Z'$ sont entières (dans une clôture algébrique de F) et leurs réductions dans $\overline{\mathbb{F}}_q$ sont les mêmes que celles de X_w . On en déduit aisément que les points $X_w + Z$ et $X_w + Z'$ ne peuvent être conjugués que s'ils sont égaux. Donc $Z = Z'$. Alors $y'y^{-1}$ commute à $X_w + Z'$ et appartient donc à $T_w(F)$. Cela prouve l'injectivité de ι . L'application ι est différentiable. Sa dérivée en un point (y, Z) est l'application

$$\mathfrak{t}_w(F) \setminus \mathfrak{m}(F) \times \mathfrak{t}_w(F) \rightarrow \mathfrak{m}(F), \quad (\mathfrak{Y}, \mathfrak{Z}) \mapsto y^{-1}([X_w + Z, \mathfrak{Y}] + \mathfrak{Z})y.$$

Celle-ci est bijective et, parce que les valeurs propres de $X_w + Z$ sont entières et de réductions toutes distinctes, on vérifie qu'elle préserve les mesures. Donc ι est un isomorphisme local, de jacobien constant de valeur 1. On en déduit que l'image de ι est ouverte dans $\mathfrak{k}_m^u \oplus \mathfrak{k}_{n',n''}^u$ et que cette image a même mesure que l'espace de départ. D'autre part, l'image de ι est clairement compacte et l'espace de départ a même mesure que $\mathfrak{k}_m^u \oplus \mathfrak{k}_{n',n''}^u$. Cela entraîne que ι est un isomorphisme préservant les mesures de $T_w(F)^u \setminus (K_m^u \times K_{n',n''}^u) \times \varpi \mathfrak{t}_w(\mathfrak{o})$ sur $\mathfrak{k}_m^u \oplus \mathfrak{k}_{n',n''}^u$. Le membre de droite de (53) (en $x = 1$) s'écrit

$$\int_{T_w(F)^u \setminus (K_m^u \times K_{n',n''}^u)} \int_{\varpi \mathfrak{t}_w(\mathfrak{o})} \hat{\phi}_{\text{Lie}}(\iota(y, Z) + X_w) dZ dy.$$

D'après les propriétés de ι , c'est aussi

$$\int_{\mathfrak{e}_m^u \oplus \mathfrak{e}_{n',n''}^u} \hat{\phi}_{\mathrm{Lie}}(X_w + Y) dY.$$

Mais ceci est la définition de $(\hat{\phi}_{\mathrm{Lie}})_{\mathrm{red}}(X_w)$. Cela démontre (53).

Utilisons (53) pour transformer (52). L'intégrale en $K_m^u \times K_{n',n''}^u$ est absorbée par celle en $K_m \times K_{n',n''}$ mais introduit un facteur $\mathrm{mes}(K_m^u \times K_{n',n''}^u)$ qui compense l'inverse de cette mesure intervenant dans (52). On obtient

$$I_w = \mathrm{sgn}(w)q^{n/2}|\mathbf{T}_w(\mathbb{F}_q)|^{-1} \mathrm{mes}(\mathcal{X}_w)^{-1} \int_{K_m \times K_{n',n''}^\pm} \int_{\mathcal{X}_w} \hat{\phi}_{\mathrm{Lie}}(x^{-1}Y_w x) dY_w dx.$$

Rappelons que l'on a supposé $\mathrm{sgn}_{CD}(w'') = 1$ si $\sharp = \mathrm{iso}$ et $\mathrm{sgn}_{CD}(w'') = -1$ si $\sharp = \mathrm{an}$. Notons $W(\mathbf{m}, n', n'')_{\mathrm{ell}, \sharp}$ le sous-ensemble des éléments de $W(\mathbf{m}, n', n'')_{\mathrm{ell}}$ dont la composante w'' vérifie cette condition. En revenant à (47), on obtient

$$I = |W(\mathbf{m}, n', n'')|^{-1} \sum_{w \in W(\mathbf{m}, n', n'')_{\mathrm{ell}, \sharp}} \mathrm{res}_m(\kappa_\pi)_{\gamma_{n',n''}}(w) \mathrm{sgn}(w)q^{n/2}|\mathbf{T}_w(\mathbb{F}_q)|^{-1} \\ \times \mathrm{mes}(\mathcal{X}_w)^{-1} \int_{K_m \times K_{n',n''}^\pm} \int_{\mathcal{X}_w} \hat{\phi}_{\mathrm{Lie}}(x^{-1}Y_w x) dY_w dx. \quad (54)$$

Notons plus précisément $I_{n',n''}(\phi)$ cette expression. En 2.1, on a défini un terme $\Theta_{\pi, \mathrm{cusp}}^M(\phi)$. On a l'égalité

$$\Theta_{\pi, \mathrm{cusp}}^M(\phi) = \sum_{n', n''} \mathrm{mes}(A_M(F) \backslash (A_M(F)K_m \times K_{n',n''}^\pm))^{-1} \int_{A_M(F) \backslash M(F)} I_{n',n''}({}^m\phi) dm,$$

où (n', n'') parcourt $D(n)$ avec la restriction $n'' \geq 1$ si $\sharp = \mathrm{an}$ et où on a noté ${}^m\phi$ la fonction $x \mapsto \phi(m^{-1}xm)$. On peut oublier la restriction sur n'' : si $\sharp = \mathrm{an}$ et $n'' = 0$, la formule (54) vaut 0 car l'ensemble $W(\mathbf{m}, n', n'')_{\mathrm{ell}, \sharp}$ est vide. On voit que l'image par transformation de Fourier de $({}^m\phi)_{\mathrm{Lie}}$ est ${}^m(\hat{\phi}_{\mathrm{Lie}})$. Les intégrales sur $K_m \times K_{n',n''}^\pm$ de la formule (54) sont absorbées par l'intégrale sur $A_M(F) \backslash M(F)$, mais introduisent des facteurs $\mathrm{mes}(K_m \times K_{n',n''}^\pm)$. Notons $A_M(F)^c$ le plus grand sous-groupe compact de $A_M(F)$. C'est aussi l'intersection de $A_M(F)$ et de K_m . Donc

$$\mathrm{mes}(A_M(F) \backslash (A_M(F)K_m \times K_{n',n''}^\pm)) = \mathrm{mes}(K_m \times K_{n',n''}^\pm) \mathrm{mes}(A_M(F)^c)^{-1}.$$

D'où

$$\Theta_{\pi, \mathrm{cusp}}^M(\phi) = \sum_{(n', n'') \in D(n)} \mathrm{mes}(A_M(F)^c) |W(\mathbf{m}, n', n'')|^{-1} \\ \times \sum_{w \in W(\mathbf{m}, n', n'')_{\mathrm{ell}, \sharp}} \mathrm{res}_m(\kappa_\pi)_{\gamma_{n',n''}}(w) \mathrm{sgn}(w)q^{n/2}|\mathbf{T}_w(\mathbb{F}_q)|^{-1} \\ \times \mathrm{mes}(\mathcal{X}_w)^{-1} \int_{A_M(F) \backslash M(F)} \int_{\mathcal{X}_w} \hat{\phi}_{\mathrm{Lie}}(m^{-1}Y_w m) dY_w dm. \quad (55)$$

On peut encore remplacer l'intégrale en $A_M(F) \backslash M(F)$ par une intégrale sur $T_w(F) \backslash M(F)$, à condition de multiplier par $\mathrm{mes}(A_M(F) \backslash T_w(F))$. Notons $T_w(F)^c$ le plus grand sous-groupe compact de $T_w(F)$.

Parce que T_w est non ramifié, on a $T_w(F) = A_M(F)T_w(F)^c$, d'où

$$\text{mes}(A_M(F)\backslash T_w(F)) = \text{mes}(A_M(F)^c)^{-1} \text{mes}(T_w(F)^c).$$

Le premier facteur compense son inverse qui figure dans la formule ci-dessus. On a introduit plus haut le sous-groupe $T_w(F)^u$ de $T_w(F)$ et on a $T_w(F)^c/T_w(F)^u \simeq \mathbf{T}_w(\mathbb{F}_q)$. De plus, $\text{mes}(T_w(F)^u) = \text{mes}(\varpi \mathfrak{t}_w(\mathfrak{o}))$. On a fixé sur $\mathfrak{t}_w(F)$ la mesure autoduale. Puisque $\mathfrak{t}_w(\mathfrak{o})$ et $\varpi \mathfrak{t}_w(\mathfrak{o})$ sont duaux pour le bicaractère $(X, Y) \mapsto \psi_F(\text{trace}(XY))$, on calcule

$$\text{mes}(\varpi \mathfrak{t}_w(\mathfrak{o})) = [\mathfrak{t}_w(\mathfrak{o}) : \varpi \mathfrak{t}_w(\mathfrak{o})]^{-\frac{1}{2}} = q^{-n/2}.$$

D'où

$$\text{mes}(T_w(F)^c) = q^{-n/2} |\mathbf{T}_w(\mathbb{F}_q)|.$$

Ces termes compensent leurs inverses figurant dans la formule (55). Finalement

$$\begin{aligned} \Theta_{\pi, \text{cusp}}^M(\phi) &= \sum_{(n', n'') \in D(n_0)} |W(\mathbf{m}, n', n'')|^{-1} \sum_{w \in W(\mathbf{m}, n', n'')_{\text{ell}, \sharp}} \text{res}_m(\kappa_\pi)_{\gamma_{n', n''}}(w) \text{sgn}(w) \text{mes}(\mathcal{X}_w)^{-1} \\ &\quad \times \int_{T_w(F)\backslash M(F)} \int_{\mathcal{X}_w} \hat{\phi}_{\text{Lie}}(m^{-1}Y_w m) dY_w dm. \end{aligned} \quad (56)$$

Soit maintenant $f \in C_c^\infty(G_\sharp(F))$. On suppose que le support de f est formé d'éléments topologiquement unipotents. En 2.1, on a défini le terme

$$\Theta_{\pi, m, \text{cusp}}(f) = \int_{P(F)\backslash G_\sharp(F)} \Theta_{\pi, \text{cusp}}^M(({}^g f)_U) dg.$$

On définit la fonction f_{Lie} , cf. 2.2. Posons $\phi = f_U$. On voit que

$$\phi_{\text{Lie}}(X) = \int_{\mathfrak{u}(F)} f_{\text{Lie}}(X + Y) dY,$$

où dY est la mesure de Haar sur $\mathfrak{u}(F)$ telle que l'exponentielle de $\mathfrak{u}(F)$ sur $U(F)$ préserve les mesures. D'où aussi

$$\hat{\phi}_{\text{Lie}}(X) = \int_{\mathfrak{u}(F)} \hat{f}_{\text{Lie}}(X + Y) dY.$$

Ou encore

$$\hat{\phi}_{\text{Lie}}(X) = |\det(\text{ad}(X)|_{\mathfrak{u}(F)})|_F \int_{U(F)} \hat{f}_{\text{Lie}}(u^{-1}Xu) du,$$

où $|\cdot|_F$ est la valeur absolue usuelle de F . On peut remplacer f par ${}^g f$. En posant $\phi = ({}^g f)_U$, on obtient

$$\hat{\phi}_{\text{Lie}}(X) = |\det(\text{ad}(X)|_{\mathfrak{u}(F)})|_F \int_{U(F)} \hat{f}_{\text{Lie}}(g^{-1}u^{-1}Xug) du.$$

Les éléments Y_w intervenant dans (56) vérifient $|\det(\text{ad}(Y_w)|_{\mathfrak{u}(F)})|_F = 1$ car les valeurs propres des réductions X_w sont toutes distinctes. On en déduit l'égalité

$$\int_{P(F)\backslash G_\sharp(F)} \int_{T_w(F)\backslash M(F)} \hat{\phi}_{\text{Lie}}(m^{-1}Y_w m) dm dg = \int_{T_w(F)\backslash G_\sharp(F)} \hat{f}_{\text{Lie}}(g^{-1}Y_w g) dg.$$

On obtient

$$\Theta_{\pi, \mathbf{m}, \text{cusp}}(f) = \sum_{(n', n'') \in D(n_0)} |W(\mathbf{m}, n', n'')|^{-1} \sum_{w \in W(\mathbf{m}, n', n'')_{\text{ell}, \sharp}} \text{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma_{n', n''}}(w) \text{sgn}(w) \text{mes}(\mathcal{X}_w)^{-1} \\ \times \int_{T_w(F) \backslash G(F)} \int_{\mathcal{X}_w} \hat{f}_{\text{Lie}}(g^{-1} Y_w g) dY_w dg.$$

Dans [Waldspurger 2001, p. 53], on a introduit la distribution

$$\varphi \mapsto \int_{T_w(F) \backslash G_{\sharp}(F)} \varphi(g^{-1} Y_w g) dg$$

sur $C_c^{\infty}(\mathfrak{g}_{\sharp}(F))$ (elle y est notée $\phi_{\theta}(X_T, \varphi)$). On a montré en [Waldspurger 2001, corollaire III.5] que sa restriction à un certain sous-espace $\mathcal{H} \subset C_c^{\infty}(\mathfrak{g}_{\sharp}(F))$ ne dépendait pas de l'élément $Y_w \in \mathcal{X}_w$ (elle ne dépend d'ailleurs pas non plus du choix de X_w mais cela résulte déjà de nos calculs ci-dessus). L'espace \mathcal{H} est défini ainsi. Soit B un sous-groupe d'Iwahori de $G_{\sharp}(F)$. Il lui correspond un sous- \mathfrak{o} -réseau \mathfrak{b} de $\mathfrak{g}_{\sharp}(F)$. Notons $C_c^{\infty}(\mathfrak{g}_{\sharp}(F)/\mathfrak{b})$ le sous-espace des fonctions invariantes par \mathfrak{b} . Alors \mathcal{H} est la somme de ces espaces $C_c^{\infty}(\mathfrak{g}_{\sharp}(F)/\mathfrak{b})$ quand \mathfrak{b} décrit tous les sous-groupes d'Iwahori de G_{\sharp} . On voit facilement que \mathcal{H} est exactement le sous-espace des fonctions φ telles que $\hat{\varphi}$ soit à support topologiquement nilpotent. En particulier, \hat{f}_{Lie} appartient à \mathcal{H} . Donc l'intégrale

$$\int_{T_w(F) \backslash G(F)} \hat{f}_{\text{Lie}}(g^{-1} Y_w g) dg$$

ne dépend pas du point $Y_w \in \mathcal{X}_w$. Intégrer cette formule en Y_w revient à la multiplier par $\text{mes}(\mathcal{X}_w)$ et ce facteur compense son inverse figurant dans la formule plus haut. On obtient simplement

$$\Theta_{\pi, \mathbf{m}, \text{cusp}}(f) = \sum_{(n', n'') \in D(n_0)} |W(\mathbf{m}, n', n'')|^{-1} \sum_{w \in W(\mathbf{m}, n', n'')_{\text{ell}, \sharp}} \text{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma_{n', n''}}(w) \text{sgn}(w) \\ \times \int_{T_w(F) \backslash G(F)} \hat{f}_{\text{Lie}}(g^{-1} X_w g) dg. \quad (57)$$

Pour expliciter davantage la formule obtenue, introduisons l'élément $\gamma_0 = (0, 0, n)$ de Γ (cf. [Waldspurger 2018, §1.8]) et la composante $\kappa_{\pi, 0}$ de κ_{π} dans la composante $\mathcal{R}(\gamma_0)$ de \mathcal{R} . Soit $(\alpha, \beta', \beta'') \in \mathcal{P}_3(n)$. On a défini en [Waldspurger 2018, §1.8] la valeur $\kappa_{\pi, 0}(w_{\alpha, \beta', \beta''})$. Associons à notre triplet de partitions la partition $\mathbf{m} = \alpha$ et les entiers $n' = S(\beta')$, $n'' = S(\beta'')$, $n_0 = n' + n''$. Soit $w = (w_1, \dots, w_t, w', w'')$ un élément de $W(\mathbf{m}, n', n'')_{\text{ell}}$ tel que w' et w'' soient paramétrés par les partitions (\emptyset, β') et (\emptyset, β'') . Les définitions entraînent que

$$\text{res}_{\mathbf{m}}(\kappa_{\pi})_{\gamma_{n', n''}}(w) = \kappa_{\pi, 0}(w_{\alpha, \beta', \beta''}).$$

Posons $\text{sgn}(w_{\alpha, \beta', \beta''}) = \text{sgn}(w)$ et définissons une distribution $\phi_{\alpha, \beta', \beta''}$ sur $C_c^{\infty}(\mathfrak{g}_{\sharp}(F))$ par :

- si $\sharp = \text{iso}$ et $l(\beta'')$ est impair ou si $\sharp = \text{an}$ et $l(\beta'')$ est pair, $\phi_{\alpha, \beta', \beta''} = 0$;
- si $\sharp = \text{iso}$ et $l(\beta'')$ est pair ou si $\sharp = \text{an}$ et $l(\beta'')$ est impair,

$$\phi_{\alpha, \beta', \beta''}(\varphi) = \int_{T_w(F) \backslash G(F)} \varphi(g^{-1} X_w g) dg$$

pour tout $\varphi \in C_c^\infty(\mathfrak{g}_\#(F))$.

La distinction entre les deux cas provient de ce que X_w n'existe que si $w \in W(\mathbf{m}, n', n'')_{\text{ell}, \#}$, c'est-à-dire si $\text{sgn}(w'')$ vaut 1 si $\# = \text{iso}$, -1 si $\# = \text{an}$, ce qui se traduit par les conditions indiquées. Cette définition dépend des choix de w dans sa classe de conjugaison et de l'élément X_w . Mais nous n'appliquerons cette distribution qu'à des éléments de l'espace \mathcal{H} . Comme on l'a dit ci-dessus, cette restriction ne dépend pas de ces choix. Dans la formule (57), l'intégrale devient $\phi_{\alpha, \beta', \beta''}(\hat{f}_{\text{Lie}})$. Cette formule devient une somme, indexée par les triplets $(\alpha, \beta', \beta'')$ tels que $\alpha = \mathbf{m}$, de termes ne dépendant que de ces triplets. Chaque triplet $(\alpha, \beta', \beta'')$ intervient avec une certaine multiplicité. Celle-ci est le produit de $|W(\mathbf{m}, n', n'')|^{-1}$ et du nombre d'éléments $w = (w_1, \dots, w_t, w', w'') \in W(\mathbf{m}, n', n'')_{\text{ell}}$ tels que w' et w'' soient paramétrés par (\emptyset, β') et (\emptyset, β'') . Pour toute partition λ , posons

$$z(\lambda) = \left(\prod_{j=1, \dots, l(\lambda)} 2\lambda_j \right) \prod_{i \geq 1} \text{mult}_\lambda(i)!,$$

et posons

$$z(\alpha, \beta', \beta'') = z(\alpha)z(\beta')z(\beta'').$$

On voit que la multiplicité précédente est égale à

$$2^{l(\alpha)} \text{mult!}_\alpha z(\alpha, \beta', \beta'')^{-1}.$$

Alors (57) se récrit

$$\Theta_{\pi, \mathbf{m}, \text{cusp}}(f) = \sum_{(\alpha, \beta', \beta'') \in \mathcal{P}_3(n); \alpha = \mathbf{m}} 2^{l(\alpha)} \text{mult!}_\alpha z(\alpha, \beta', \beta'')^{-1} \text{sgn}(w_{\alpha, \beta', \beta''}) \kappa_{\pi, 0}(w_{\alpha, \beta', \beta''}) \phi_{\alpha, \beta', \beta''}(\hat{f}_{\text{Lie}}).$$

Le résultat de 2.1 est que $\Theta_\pi(f)$ est la somme sur \mathbf{m} des expressions ci-dessus, multipliées par $2^{-l(\mathbf{m})} \text{mult!}_\mathbf{m}^{-1}$. D'où

$$\Theta_\pi(f) = \sum_{(\alpha, \beta', \beta'') \in \mathcal{P}_3(n)} z(\alpha, \beta', \beta'')^{-1} \text{sgn}(w_{\mu, \beta', \beta''}) \kappa_{\pi, 0}(w_{\alpha, \beta', \beta''}) \phi_{\alpha, \beta', \beta''}(\hat{f}_{\text{Lie}}). \tag{58}$$

3. Fronts d'onde

3.1. Rappel sur les orbites unipotentes. Soit $\# = \text{iso}$ ou an . On appelle orbite nilpotente une classe de conjugaison par $G_\#(F)$ d'éléments nilpotents dans $\mathfrak{g}_\#(F)$. On note $\text{Nil}_\#$ l'ensemble des orbites nilpotentes. Les orbites nilpotentes sont classifiées par des données $(\mu, (q_i)_{i \in \text{Jord}_{\text{bp}}(\mu)})$, où $\mu \in \mathcal{P}^{\text{orth}}(2n+1)$; pour tout $i \in \text{Jord}_{\text{bp}}(\mu)$, q_i est une classe d'équivalence d'une forme quadratique non dégénérée sur un espace vectoriel sur F de dimension $\text{mult}_\mu(i)$; le noyau anisotrope de la forme quadratique $\bigoplus_{i \in \text{Jord}_{\text{bp}}(\mu)} q_i$ est équivalent à celui de $Q_\#$.

Pour une orbite nilpotente \mathcal{O} , on note $\mu(\mathcal{O})$ la partition associée à \mathcal{O} .

Une classification analogue vaut pour les groupes $\text{SO}(2n+1)$ et $\text{SO}(2n)_\#$ définis sur \mathbb{F}_q . Il y a une petite perturbation dans le cas du groupe $\text{SO}(2n)_{\text{iso}}$. La classification ci-dessus vaut pour les classes de conjugaison par $O(2n)_{\text{iso}}(\mathbb{F}_q)$ et non pas par $\text{SO}(2n)_{\text{iso}}(\mathbb{F}_q)$. Il peut y avoir des classes de conjugaison par

$O(2n)_{\mathrm{iso}}(\mathbb{F}_q)$ qui se coupent en deux classes de conjugaison par $\mathrm{SO}(2n)_{\mathrm{iso}}(\mathbb{F}_q)$. A ces deux classes sont associées les mêmes données $(\mu, (q_i)_{i \in \mathrm{Jord}_{\mathrm{bp}}(\mu)})$.

La définition suivante va nous être utile. Considérons deux espaces vectoriels l_1 et l_2 sur \mathbb{F}_q de dimensions d_1 et d_2 , respectivement. Soient q_1 et q_2 , des formes quadratiques non dégénérées sur ces espaces. À isomorphisme près, il existe un unique triplet (V, Q, L) vérifiant les conditions suivantes :

- V est un espace vectoriel sur F de dimension $d_1 + d_2$;
- Q est une forme quadratique non dégénérée sur V ;
- $L \subset V$ est un réseau presque autodual, c'est-à-dire $\varpi L^* \subseteq L \subseteq L^*$;
- (l_1, q_1) est isomorphe à (l', Q') et (l_2, q_2) est isomorphe à (l'', Q'') (rappelons que $l' = L/\varpi L^*$, $l'' = L^*/L$ et que Q' et Q'' sont les formes sur ces espaces qui se déduisent naturellement de Q , cf. [Waldspurger 2018, §1.1]).

On note Q_{q_1, q_2} cette forme quadratique Q dont la classe d'équivalence est bien déterminée.

Considérons l'ensemble \mathbf{Nil}_{\sharp} des paires $(\mathcal{O}_1, \mathcal{O}_2)$ telles que

- il existe $n_1, n_2 \in \mathbb{N}$, avec $n_1 + n_2 = n$ et $n_2 \geq 1$ si $\sharp = \mathrm{an}$, de sorte que \mathcal{O}_1 soit une orbite nilpotente dans $\mathfrak{so}(2n_1 + 1)(\mathbb{F}_q)$ et \mathcal{O}_2 est une orbite nilpotente dans $\mathfrak{so}(2n_2)_{\sharp}(\mathbb{F}_q)$.

À une telle paire, on va associer une orbite nilpotente $\mathcal{O}_{\mathcal{O}_1, \mathcal{O}_2}$. Notons

$$(\mu_1, (q_{1,i})_{i \in \mathrm{Jord}_{\mathrm{bp}}(\mu_1)}) \quad \text{et} \quad (\mu_2, (q_{2,i})_{i \in \mathrm{Jord}_{\mathrm{bp}}(\mu_2)})$$

les paramètres de \mathcal{O}_1 et \mathcal{O}_2 . On pose $\mu = \mu_1 \cup \mu_2$ et, pour tout $i \in \mathrm{Jord}_{\mathrm{bp}}(\mu)$, $q_i = Q_{q_{1,i}, q_{2,i}}$ (avec $q_{1,i}$ ou $q_{2,i} = 0$ si $i \notin \mathrm{Jord}_{\mathrm{bp}}(\mu_1)$ ou $i \notin \mathrm{Jord}_{\mathrm{bp}}(\mu_2)$). On vérifie que $(\mu, (q_i)_{i \in \mathrm{Jord}_{\mathrm{bp}}(\mu)})$ classe une orbite nilpotente dans $\mathfrak{g}_{\sharp}(F)$. Alors $\mathcal{O}_{\mathcal{O}_1, \mathcal{O}_2}$ est cette orbite unipotente. L'application

$$\mathbf{Nil}_{\sharp} \rightarrow \mathbf{Nil}_{\sharp}, \quad (\mathcal{O}_1, \mathcal{O}_2) \mapsto \mathcal{O}_{\mathcal{O}_1, \mathcal{O}_2}$$

est surjective.

Pour $\mathcal{O} \in \mathbf{Nil}_{\sharp}$, on note $I_{\mathcal{O}}$ l'intégrale orbitale associée à \mathcal{O} . Pour la définir, il faut bien sûr fixer une mesure sur \mathcal{O} invariante par conjugaison. La définition de cette mesure n'aura pas d'importance pour nous.

Soit $(\mathcal{O}_1, \mathcal{O}_2) \in \mathbf{Nil}_{\sharp}$. En [Waldspurger 2001, IX.2], on a défini une fonction $h_{\mathcal{O}_1, \mathcal{O}_2} \in C_c^{\infty}(\mathfrak{g}_{\sharp}(F))$ (dans cette référence, les éléments de \mathbf{Nil}_{\sharp} étaient notés \check{N}). Elle vérifie les propriétés suivantes :

- $h_{\mathcal{O}_1, \mathcal{O}_2} \in \mathcal{H}$; l'espace \mathcal{H} a été défini en 2.3 ; c'est celui des fonctions φ dont la transformée de Fourier est à support topologiquement nilpotent ;

$$(59) \quad \text{pour } \mathcal{O} \in \mathbf{Nil}_{\sharp} \text{ dont l'adhérence } \bar{\mathcal{O}} \text{ ne contient pas } \mathcal{O}_{\mathcal{O}_1, \mathcal{O}_2}, \quad I_{\mathcal{O}}(h_{\mathcal{O}_1, \mathcal{O}_2}) = 0 ;$$

$$(60) \quad \text{pour } \mathcal{O} = \mathcal{O}_{\mathcal{O}_1, \mathcal{O}_2}, \quad I_{\mathcal{O}}(h_{\mathcal{O}_1, \mathcal{O}_2}) \neq 0.$$

On définit une fonction $f_{\mathcal{O}_1, \mathcal{O}_2} \in C_c^{\infty}(G_{\sharp}(F))$ comme suit (cf. [Waldspurger 2001, lemme IX.4]) : c'est la fonction à support topologiquement unipotent telle que $(f_{\mathcal{O}_1, \mathcal{O}_2})_{\mathrm{Lie}} = \hat{h}_{\mathcal{O}_1, \mathcal{O}_2}$.

3.2. Développement des caractères à l'origine. Soit π une représentation lisse et irréductible de $G_{\sharp}(F)$. D'après Harish-Chandra, on sait qu'il existe une unique famille de nombres complexes $(c_{\mathcal{O}}(\pi))_{\mathcal{O} \in \mathbf{Nil}_{\sharp}}$

et un voisinage $V(\pi)$ de 1 dans $G_{\sharp}(F)$ de sorte que les propriétés suivantes soient vérifiées. Le voisinage $V(\pi)$ est invariant par conjugaison par $G_{\sharp}(F)$ et est formé d'éléments topologiquement unipotents. Soit $f \in C_c^\infty(G_{\sharp}(F))$. On suppose que le support de f est contenu dans $V(\pi)$. En particulier, on peut associer à f une fonction f_{Lie} sur $\mathfrak{g}_{\sharp}(F)$, à support topologiquement nilpotent. Alors on a l'égalité

$$\Theta_{\pi}(f) = \sum_{\mathcal{O} \in \text{Nil}_{\sharp}} c_{\mathcal{O}}(\pi) I_{\mathcal{O}}(\hat{f}_{\text{Lie}}). \quad (61)$$

Remarquons que les coefficients $c_{\mathcal{O}}(\pi)$ ne sont pas tous nuls. En effet, si f est la fonction caractéristique d'un sous-groupe ouvert compact H contenu dans $V(\pi)$, $\Theta_{\pi}(f)$ est égal au produit de la mesure de H et de la dimension du sous-espace des invariants par H dans l'espace de π . Ce terme est non nul si H est assez petit. On dit que π admet un front d'onde s'il existe $\mu(\pi) \in \mathcal{P}^{\text{orth}}(2n+1)$ de sorte que

- pour tout $\mathcal{O} \in \text{Nil}_{\sharp}$ tel que $c_{\mathcal{O}}(\pi) \neq 0$, on a $\mu(\mathcal{O}) \leq \mu(\pi)$;
- il existe $\mathcal{O} \in \text{Nil}_{\sharp}$ tel que $c_{\mathcal{O}}(\pi) \neq 0$ et $\mu(\mathcal{O}) = \mu(\pi)$.

Évidemment, $\mu(\pi)$ est unique si elle existe. On conjecture que toute représentation lisse irréductible admet un front d'onde. Supposons que π admette un front d'onde. On montre que

- $\mu(\pi)$ est une partition spéciale, cf. [Mœglin 1996a, théorème 1.4];
- pour tout $\mathcal{O} \in \text{Nil}_{\sharp}$ tel que $\mu(\mathcal{O}) = \mu(\pi)$, on a $c_{\mathcal{O}}(\pi) \geq 0$, cf. [Mœglin et Waldspurger 1987, corollaire 1.17].

Remarque. La construction d'Harish-Chandra utilise l'exponentielle et non pas notre exponentielle tronquée E . Mais le résultat est le même, avec les mêmes coefficients, que l'on utilise l'une ou l'autre de ces applications.

Dans le cas où $\pi \in \text{Irr}_{\text{unip}, \sharp}$, on peut prendre pour voisinage $V(\pi)$ l'ensemble tout entier des éléments topologiquement unipotents de $G_{\sharp}(F)$. En effet, pour f à support topologiquement unipotent, $\Theta_{\pi}(f)$ est calculé par la formule (58). Or il résulte de [DeBacker 2002, Theorem 2.1.5] que le membre de droite de cette formule est de la même forme que celui de (61) ci-dessus. Ces deux expressions doivent coïncider si le support de f est dans un voisinage assez petit de l'origine. Les distributions $f \mapsto I_{\mathcal{O}}(\hat{f}_{\text{Lie}})$ sont linéairement indépendantes, même si on les restreint aux fonctions vérifiant cette condition de support. Cela implique que les coefficients sont les mêmes dans les deux expressions. Donc (61) est valable pour toute f à support topologiquement unipotent.

3.3. Le théorème. Pour $\sharp = \text{iso}$ ou an , notons $\text{Irr}_{\text{tunip}, \sharp}$ le sous-ensemble des représentations admissibles irréductibles de $G_{\sharp}(F)$ qui sont tempérées et de réduction unipotente. Notons $\text{Irr}_{\text{tunip}}$ la réunion disjointe de $\text{Irr}_{\text{tunip}, \text{iso}}$ et $\text{Irr}_{\text{tunip}, \text{an}}$. Dans [Waldspurger 2018, §1.3], on a adapté l'habituelle classification de Langlands : l'ensemble $\text{Irr}_{\text{tunip}}$ est paramétré par un ensemble $\mathfrak{Irr}_{\text{tunip}}$ de triplets (λ, s, ϵ) . En particulier, le terme λ est un élément de $\mathcal{P}^{\text{symp}}(2n)$. Pour un tel triplet, on a noté $\pi(\lambda, s, \epsilon)$ la représentation qui lui est associée par Lusztig (elle est tempérée). On a introduit l'involution D de Zelevinsky–Aubert–Schneider–Stuhler en [Waldspurger 2018, §1.7] et une dualité d entre partitions en 1.6 et 1.7 ci-dessus. On pose $\delta(\lambda, s, \epsilon) = D(\pi(\lambda, s, \epsilon))$.

Théorème. Soit $(\lambda, s, \epsilon) \in \mathcal{Irr}_{\mathrm{unip}}$. Alors $\delta(\lambda, s, \epsilon)$ admet un front d'onde et on a l'égalité

$$\mu(\delta(\lambda, s, \epsilon)) = d(\lambda).$$

La fin de l'article est consacrée à la démonstration du théorème.

3.4. Une première réduction. En [Waldspurger 2018, §1.3], on a introduit le sous-ensemble $\mathcal{Irr}_{\mathrm{unip-quad}}$ des triplets $(\lambda, s, \epsilon) \in \mathcal{Irr}_{\mathrm{unip}}$ tels que $s^2 = 1$. Supposons que le théorème soit prouvé pour les triplets $(\lambda, s, \epsilon) \in \mathcal{Irr}_{\mathrm{unip-quad}}$ tels que λ n'ait que des termes pairs. Montrons que le théorème résulte de ce cas particulier.

Soit $\sharp = \mathrm{iso}$ ou an et soit P un sous-groupe parabolique de G_{\sharp} , de composante de Levi M . Soit π^M une représentation lisse irréductible de $M(F)$, notons $\pi = \mathrm{Ind}_P^{G_{\sharp}}(\pi^M)$ son induite et supposons π irréductible. On a défini en 3.2 la notion de front d'onde pour le groupe G_{\sharp} mais on sait bien que la définition est générale et s'applique en particulier au Levi M . Supposons que π^M admette un front d'onde. Si

$$M = \mathrm{GL}(m_1) \times \cdots \times \mathrm{GL}(m_t) \times G_{\sharp, n_0},$$

$\mu(\pi^M)$ est alors une famille $(\mu_1, \dots, \mu_t, \mu_0)$ où, pour $i = 1, \dots, t$, $\mu_i \in \mathcal{P}(m_i)$ et $\mu_0 \in \mathcal{P}^{\mathrm{orth}}(2n_0 + 1)$. On a défini en 1.8 une opération d'induction qui envoie $\mu(\pi^M)$ sur une partition $\mathrm{ind}(\mu(\pi^M)) \in \mathcal{P}^{\mathrm{orth}}(2n + 1)$.

Sous ces hypothèses, on a

$$\pi \text{ admet un front d'onde et on a } \mu(\pi) = \mathrm{ind}(\mu(\pi^M)). \quad (62)$$

Preuve. Pour $f \in C_c^\infty(G_{\sharp}(F))$, on a l'égalité $\Theta_{\pi}(f) = \Theta_{\pi^M}(f_P)$, où f_P est l'habituel "terme constant" de f . On définit facilement un voisinage $V(\pi)$ de 1 dans $G_{\sharp}(F)$ invariant par conjugaison et formé d'éléments topologiquement unipotents, de sorte que, si f est à support dans $V(\pi)$, f_P soit à support dans $V(\pi^M)$. Pour une telle fonction f , on a alors

$$\Theta_{\pi}(f) = \sum_{\mathcal{O}^M \in \mathrm{Nil}^M} c_{\mathcal{O}^M}(\pi^M) I_{\mathcal{O}^M}(f_P),$$

où Nil^M est l'analogie de Nil_{\sharp} pour le groupe M . Notons \mathfrak{u} le radical nilpotent de l'algèbre de Lie \mathfrak{p} . Pour $\mathcal{O}^M \in \mathrm{Nil}^M$ et $\mathcal{O} \in \mathrm{Nil}_{\sharp}$, on dit que \mathcal{O} est induite de \mathcal{O}^M si \mathcal{O} coupe $\mathcal{O}^M + \mathfrak{u}(F)$ selon un ouvert non vide. On note cette relation $\mathcal{O} \subset \mathrm{ind}(\mathcal{O}^M)$. Si on se plaçait sur la clôture algébrique \bar{F} , il y aurait une et une seule orbite induite mais, parce que l'on travaille sur F , il y en a plusieurs en général. Cette opération d'induction d'orbites est reliée à l'induction des partitions par la relation suivante :

$$\text{si } \mathcal{O} \subset \mathrm{ind}(\mathcal{O}^M), \quad \text{alors } \mu(\mathcal{O}) = \mathrm{ind}(\mu(\mathcal{O}^M)).$$

On a une égalité

$$I_{\mathcal{O}^M}(f_P) = \sum_{\mathcal{O} \subset \mathrm{ind}(\mathcal{O}^M)} c_{\mathcal{O}^M, \mathcal{O}} I_{\mathcal{O}}(f),$$

avec des coefficients $c_{\mathcal{O}^M, \mathcal{O}} > 0$. D'où

$$\Theta_\pi(f) = \sum_{\mathcal{O} \in \text{Nil}_\sharp} c_{\mathcal{O}}(\pi) I_{\mathcal{O}}(f),$$

où, pour tout $\mathcal{O} \in \text{Nil}_\sharp$, on a

$$c_{\mathcal{O}}(\pi) = \sum_{\substack{\mathcal{O}^M \in \text{Nil}^M \\ \mathcal{O} \subset \text{ind}(\mathcal{O}^M)}} c_{\mathcal{O}^M}(\pi^M) c_{\mathcal{O}^M, \mathcal{O}}. \tag{63}$$

Si $c_{\mathcal{O}}(\pi) \neq 0$, il existe \mathcal{O}^M tel que $\mathcal{O} \subset \text{ind}(\mathcal{O}^M)$ et $c_{\mathcal{O}^M}(\pi^M) \neq 0$. On a alors $\mu(\mathcal{O}) = \text{ind}(\mu(\mathcal{O}^M))$ et $\mathcal{O}^M \leq \mu(\pi^M)$. D'où $\mu(\mathcal{O}) \leq \text{ind}(\mu(\pi^M))$ car l'opération d'induction est croissante. Inversement, soit $\mathcal{O}_0^M \in \text{Nil}^M$ tel que $c_{\mathcal{O}_0^M}(\pi^M) \neq 0$ et $\mu(\mathcal{O}_0^M) = \mu(\pi^M)$. Soit $\mathcal{O} \subset \text{ind}(\mathcal{O}_0^M)$. On a $\mu(\mathcal{O}) = \text{ind}(\mu(\pi^M))$. Montrons que $c_{\mathcal{O}}(\pi) \neq 0$. Soit \mathcal{O}^M intervenant de façon non nulle dans la formule (63). On a $c_{\mathcal{O}^M}(\pi^M) \neq 0$ donc $\mu(\mathcal{O}^M) \leq \mu(\pi^M)$. Si cette relation n'est pas une égalité, on a $\text{ind}(\mu(\mathcal{O}^M)) < \text{ind}(\mu(\pi^M))$ car l'opération d'induction est strictement croissante. Cela contredit la relation $\text{ind}(\mu(\mathcal{O}^M)) = \mu(\mathcal{O}) = \text{ind}(\mu(\pi^M))$. Donc $\mu(\mathcal{O}^M) = \mu(\pi^M)$. Alors le coefficient $c_{\mathcal{O}^M}(\pi^M) c_{\mathcal{O}^M, \mathcal{O}}$ est strictement positif. Par construction, il y a au moins une telle orbite \mathcal{O}^M , à savoir \mathcal{O}_0^M . Le coefficient $c_{\mathcal{O}}(\pi)$ est une somme non vide de termes strictement positifs, donc $c_{\mathcal{O}}(\pi) > 0$, ce qui achève la démonstration de (62) \square

Soit $(\lambda, s, \epsilon) \in \mathfrak{Irr}_{\text{unip}, \sharp}$. En reprenant les considérations de [Waldspurger 2018, §1.3], on voit qu'il existe

- un sous-groupe parabolique P de G_\sharp de composante de Levi

$$M = \text{GL}(m_1) \times \cdots \times \text{GL}(m_t) \times G_{\sharp, n_0};$$

- pour tout $j = 1, \dots, t$, un caractère non ramifié χ_j de F^\times tel que $\chi_j^2 \neq 1$ si m_j est pair;
- un élément $(\lambda_0, s_0, \epsilon_0) \in \mathfrak{Irr}_{\text{unip-quad}, n_0, \sharp}$ tel que λ_0 n'ait que des termes pairs;

de sorte que les propriétés suivantes soient vérifiées :

$$\pi(\lambda, s, \epsilon) = \text{Ind}_P^{G_\sharp} (\text{st}_{m_1}(\chi_1 \circ \det) \otimes \cdots \otimes \text{st}_{m_t}(\chi_t \circ \det) \otimes \pi(\lambda_0, s_0, \epsilon_0)), \tag{64}$$

où st_{m_j} est la représentation de Steinberg de $\text{GL}(m_j; F)$:

$$\lambda = (m_1, m_1) \cup \cdots \cup (m_t, m_t) \cup \lambda_0. \tag{65}$$

Remarque. Le couple (λ_0, s_0) se déduit de (λ, s) en éliminant les termes impairs de λ ainsi que les termes pairs pour lesquels la valeur propre correspondante de s est différente de ± 1 . On voit que $\mathbf{Z}(\lambda, s) = \mathbf{Z}(\lambda_0, s_0)$ (avec les notations de l'introduction) et ϵ s'identifie à un caractère ϵ_0 de $\mathbf{Z}(\lambda_0, s_0)$. L'égalité (64) exprime la compatibilité du paramétrage de Langlands avec l'induction. Cette compatibilité est bien vérifiée par les représentations construites par Lusztig, cf., par exemple, [Waldspurger 2004, lemme 3.8].

En appliquant l'involution D , on déduit de (64) l'égalité

$$\delta(\lambda, s, \epsilon) = \text{Ind}_P^{G_\sharp} (\delta^M),$$

où

$$\delta^M = (\chi_1 \circ \det) \otimes \cdots \otimes (\chi_t \circ \det) \otimes \delta(\lambda_0, s_0, \epsilon_0).$$

Puisqu'on suppose connu le théorème pour $\delta(\lambda_0, s_0, \epsilon_0)$ (et que les fronts d'onde des représentations des groupes $GL(m_j)$ sont bien connus), il résulte de (62) que $\delta(\lambda, s, \epsilon)$ admet un front d'onde et que

$$\mu(\delta(\lambda, s, \epsilon)) = \text{ind}(\mu(\delta^M)).$$

Le front d'onde d'un caractère de $GL(m_j; F)$ est ${}^t(m_j)$, c'est-à-dire la partition composée de m_j fois le nombre 1. On a donc

$$\mu(\delta^M) = ({}^t(m_1), \dots, {}^t(m_t), d(\lambda_0)).$$

Posons $\lambda = ((m_1), \dots, (m_t), \lambda_0)$. Avec les définitions de 1.8, cette dernière relation s'écrit $\mu(\delta^M) = d(\lambda)$ tandis que l'égalité (65) s'écrit $\lambda = \text{cup}(\lambda)$. En appliquant le lemme 1.8, on obtient $\text{ind}(\mu(\delta^M)) = d(\lambda)$. D'où $\mu(\delta(\lambda, s, \epsilon)) = d(\lambda)$, ce qui démontre le théorème.

3.5. Traduction de ce que l'on veut démontrer. On fixe désormais un élément $(\lambda, s, \epsilon) \in \mathfrak{Irr}_{\text{unip}}$ et on pose $\delta = \delta(\lambda, s, \epsilon)$. On note \sharp l'indice tel que δ soit une représentation de $G_{\sharp}(F)$. On veut prouver que δ admet un front d'onde et que $\mu(\delta) = d(\lambda)$. Montrons qu'il suffit de prouver :

(66) Pour tout $(\mathcal{O}_1, \mathcal{O}_2) \in \mathbf{Nil}_{\sharp}$, la relation $\Theta_{\delta}(f_{\mathcal{O}_1, \mathcal{O}_2}) \neq 0$ entraîne $\mu(\mathcal{O}_1) \cup \mu(\mathcal{O}_2) \leq d(\lambda)$.

(67) Il existe $(\mathcal{O}_1, \mathcal{O}_2) \in \mathbf{Nil}_{\sharp}$ tel que $\Theta_{\delta}(f_{\mathcal{O}_1, \mathcal{O}_2}) \neq 0$ et $\mu(\mathcal{O}_1) \cup \mu(\mathcal{O}_2) = d(\lambda)$.

Comme on l'a dit en 3.2, on peut prendre pour voisinage $V(\delta)$ l'ensemble tout entier des éléments topologiquement unipotents. En particulier, le développement de 3.2 vaut pour toute fonction $f_{\mathcal{O}_1, \mathcal{O}_2}$. D'après la définition de cette fonction, on a

$$\Theta_{\delta}(f_{\mathcal{O}_1, \mathcal{O}_2}) = \sum_{\mathcal{O} \in \mathbf{Nil}_{\sharp}} c_{\mathcal{O}}(\delta) I_{\mathcal{O}}(h_{\mathcal{O}_1, \mathcal{O}_2}). \quad (68)$$

Soit \mathcal{O}_0 un élément maximal dans l'ensemble des $\mathcal{O} \in \mathbf{Nil}_{\sharp}$ pour lesquels $c_{\mathcal{O}}(\delta) \neq 0$. Appliquons l'égalité (68) à une paire $(\mathcal{O}_1, \mathcal{O}_2)$ telle que $\mathcal{O}_{\mathcal{O}_1, \mathcal{O}_2} = \mathcal{O}_0$. En vertu de (59), il ne reste dans la somme que des \mathcal{O} pour lesquels $\bar{\mathcal{O}}$ contient \mathcal{O}_0 . Par maximalité de \mathcal{O}_0 , il ne reste donc que \mathcal{O}_0 . Le coefficient $c_{\mathcal{O}_0}(\delta)$ est non nul par hypothèse et l'intégrale orbitale $I_{\mathcal{O}_0}(h_{\mathcal{O}_1, \mathcal{O}_2})$ ne l'est pas par (60). Donc $\Theta_{\delta}(f_{\mathcal{O}_1, \mathcal{O}_2}) \neq 0$. D'où $\mu(\mathcal{O}_0) \leq d(\lambda)$ d'après (66). Ceci étant vrai pour tout élément maximal \mathcal{O}_0 , c'est vrai pour tout élément : pour tout \mathcal{O} tel que $c_{\mathcal{O}}(\delta) \neq 0$, on a $\mu(\mathcal{O}) \leq d(\lambda)$. En appliquant maintenant (68) pour une paire $(\mathcal{O}_1, \mathcal{O}_2)$ vérifiant (67), le même calcul montre que $c_{\mathcal{O}}(\delta) \neq 0$ pour l'orbite $\mathcal{O} = \mathcal{O}_{\mathcal{O}_1, \mathcal{O}_2}$. Pour cette orbite, on a $\mu(\mathcal{O}) = d(\lambda)$. Cela vérifie les propriétés requises pour que δ admette un front d'onde et que l'on ait $\mu(\delta) = d(\lambda)$.

3.6. Début du calcul. Fixons un couple $(\mathcal{O}_1, \mathcal{O}_2) \in \mathbf{Nil}_{\sharp}$, posons

$$\mu_1 = \mu(\mathcal{O}_1), \quad \mu_2 = \mu(\mathcal{O}_2), \quad S(\mu_1) = 2n_1 + 1, \quad S(\mu_2) = 2n_2.$$

Si $\sharp = \text{iso}$, on pose $W_{n_2, \text{iso}} = W_{n_2}^D$. Si $\sharp = \text{an}$, auquel cas $n_2 > 0$, on note $W_{n_2, \text{an}} = \{w \in W_{n_2}; \text{sgn}_{CD}(w) = -1\}$. On fixe des éléments nilpotents $Y_1 \in \mathcal{O}_1$ et $Y_2 \in \mathcal{O}_2$.

La formule (58) calcule $\Theta_\delta(f_{\mathcal{O}_1, \mathcal{O}_2})$ en fonction de termes $\phi_{\alpha, \beta_1, \beta_2}(h_{\mathcal{O}_1, \mathcal{O}_2})$ pour $(\alpha, \beta_1, \beta_2) \in \mathcal{P}_3(n)$. On a calculé ce terme en [Waldspurger 2001, proposition 3.5]. On va rappeler ce résultat en modifiant quelque peu ses notations. Pour $w_1 \in W_{n_1}$, on définit une certaine fonction $Q_{w_1}^\natural$ sur l'ensemble des éléments nilpotents de $\text{SO}(2n_1 + 1)(\mathbb{F}_q)$, cf. [Waldspurger 2001, VIII.13]. Elle est invariante par conjugaison par $\text{SO}(2n_1 + 1)(\mathbb{F}_q)$ et ne dépend que de la classe de conjugaison de w_1 . Pour $w_2 \in W_{n_2, \sharp}$, on définit de même une fonction $Q_{w_2}^\natural$ sur l'ensemble des éléments nilpotents de $\text{SO}(2n_2)_\sharp(\mathbb{F}_q)$, cf. [Waldspurger 2001, VIII.13]. Elle est invariante par conjugaison par $\text{SO}(2n_2)_\sharp(\mathbb{F}_q)$ et ne dépend que de la classe de conjugaison par $W_{n_2}^D$ de w_2 .

Remarque. Dans le cas où $\sharp = \text{an}$, la construction de [Waldspurger 2001] était un peu différente. On y avait fixé une certaine symétrie élémentaire $w_\phi \in W_{n_2, \text{an}}$ et défini une fonction $Q_{w_2}^\natural$ indexée non pas par un élément $w_2 \in W_{n_2, \text{an}}$, mais par un élément $w_2 \in W_{n_2}^D$. Cette fonction ne dépendait que de la classe de w_ϕ -conjugaison de w_2 . Notre présente fonction $Q_{w_2}^\natural$ est la fonction $Q_{w_\phi w_2}^\natural$ de [Waldspurger 2001].

Notons $W(\alpha, \beta_1, \beta_2)$ l'ensemble des paires $(w_1, w_2) \in W_{n_1} \times W_{n_2, \sharp}$ vérifiant la condition suivante. Notons (α_1, β'_1) la paire de partitions paramétrant la classe de conjugaison de w_1 et (α_2, β'_2) celle qui paramètre la classe de conjugaison par W_{n_2} de w_2 . Alors

$$\alpha = \alpha_1 \cup \alpha_2, \quad \beta'_1 = \beta_1, \quad \beta'_2 = \beta_2.$$

Pour une telle paire (w_1, w_2) , posons

$$[w_1, w_2] = \frac{z(\alpha)}{z(\alpha_1)z(\alpha_2)},$$

cf. 2.3 pour la définition de ces termes ;

- $\eta(w_1, w_2) = 2$ si $n_2 \geq 1$ et la classe de conjugaison de w_2 par W_{n_2} coïncide avec sa classe de conjugaison par $W_{n_2}^D$;
- $\eta(w_1, w_2) = 1$ si $n_2 = 0$ ou si $n_2 \geq 1$ et la classe de conjugaison de w_2 par W_{n_2} se coupe en deux classes de conjugaison par $W_{n_2}^D$.

Fixons un ensemble de représentants $\mathcal{W}(\alpha, \beta_1, \beta_2)$ des classes de conjugaison par $W_{n_1} \times W_{n_2}^D$ dans $W(\alpha, \beta_1, \beta_2)$.

La proposition IX.5 de [Waldspurger 2001] affirme alors l'existence d'un demi-entier $d_{\mathcal{O}_1, \mathcal{O}_2}$ (ne dépendant que de $(\mathcal{O}_1, \mathcal{O}_2)$) de sorte que

$$\phi_{\alpha, \beta_1, \beta_2}(h_{\mathcal{O}_1, \mathcal{O}_2}) = q^{d_{\mathcal{O}_1, \mathcal{O}_2}} \sum_{(w_1, w_2) \in \mathcal{W}(\alpha, \beta_1, \beta_2)} \eta(w_1, w_2) [w_1, w_2] Q_{w_1}^\natural(Y_1) Q_{w_2}^\natural(Y_2).$$

Cette formule peut se simplifier. Pour $w_1 \in W_{n_1}$, notons $Z(w_1)$ son centralisateur dans W_{n_1} . Pour $w_2 \in W_{n_2}$, notons $Z^D(w_2)$ son centralisateur dans $W_{n_2}^D$. Le terme $Q_{w_1}^\natural(Y_1) Q_{w_2}^\natural(Y_2)$ ne dépendant que de la classe de conjugaison de (w_1, w_2) par $W_{n_1} \times W_{n_2}^D$, on peut remplacer la somme sur le système de représentants

$\mathcal{W}(\alpha, \beta_1, \beta_2)$ par une somme sur $W(\alpha, \beta_1, \beta_2)$, à condition de multiplier chaque terme indexé par (w_1, w_2) par l'inverse du nombre d'éléments de sa classe de conjugaison par $W_{n_1} \times W_{n_2}^D$. Cet inverse est égal à

$$|Z(w_1)| |Z^D(w_2)| |W_{n_1}|^{-1} |W_{n_2}^D|^{-1}.$$

D'autre part, soit $(w_1, w_2) \in W(\alpha, \beta_1, \beta_2)$. On vérifie l'égalité

$$z(\alpha, \beta_1, \beta_2)^{-1} \eta(w_1, w_2)[w_1, w_2] = |Z(w_1)|^{-1} |Z^D(w_2)|^{-1}.$$

La formule ci-dessus se réécrit

$$z(\alpha, \beta_1, \beta_2)^{-1} \phi_{\alpha, \beta_1, \beta_2}(h_{\mathcal{O}_1, \mathcal{O}_2}) = q^{d_{\mathcal{O}_1, \mathcal{O}_2}} |W_{n_1}|^{-1} |W_{n_2}^D|^{-1} \sum_{(w_1, w_2) \in W(\alpha, \beta_1, \beta_2)} Q_{w_1}^{\natural}(Y_1) Q_{w_2}^{\natural}(Y_2).$$

Reportons cette égalité dans la formule (58). On obtient

$$\begin{aligned} & \Theta_{\delta}(f_{\mathcal{O}_1, \mathcal{O}_2}) \\ &= q^{d_{\mathcal{O}_1, \mathcal{O}_2}} |W_{n_1}|^{-1} |W_{n_2}^D|^{-1} \sum_{(\alpha, \beta_1, \beta_2) \in \mathcal{P}_3(n)} \mathrm{sgn}(w_{\alpha, \beta_1, \beta_2}) \kappa_{\delta, 0}(w_{\alpha, \beta_1, \beta_2}) \sum_{(w_1, w_2) \in W(\alpha, \beta_1, \beta_2)} Q_{w_1}^{\natural}(Y_1) Q_{w_2}^{\natural}(Y_2). \end{aligned}$$

Sommer en $(\alpha, \beta_1, \beta_2)$ puis en $(w_1, w_2) \in W(\alpha, \beta_1, \beta_2)$ revient à sommer sur tout $(w_1, w_2) \in W_{n_1} \times W_{n_2, \sharp}$.

D'où

$$\Theta_{\delta}(f_{\mathcal{O}_1, \mathcal{O}_2}) = q^{d_{\mathcal{O}_1, \mathcal{O}_2}} |W_{n_1}|^{-1} |W_{n_2}^D|^{-1} \sum_{(w_1, w_2) \in W_{n_1} \times W_{n_2, \sharp}} \mathrm{sgn}(w_1) \mathrm{sgn}(w_2) \kappa_{\delta, 0}(w_1 \times w_2) Q_{w_1}^{\natural}(Y_1) Q_{w_2}^{\natural}(Y_2).$$

Pour toute paire $(\rho_1, \rho_2) \in \widehat{W}_{n_1} \times \widehat{W}_{n_2}$, posons

$$m_{\delta}(\rho_1, \rho_2) = |W_{n_1}|^{-1} |W_{n_2}|^{-1} \sum_{(w_1, w_2) \in W_{n_1} \times W_{n_2}} \kappa_{\delta, 0}(w_1, w_2) \rho_1(w_1) \rho_2(w_2).$$

Posons aussi

$$\chi_{\rho_1}^{\natural} = |W_{n_1}|^{-1} \sum_{w_1 \in W_{n_1}} \rho_1(w_1) Q_{w_1}^{\natural} \quad \text{et} \quad \chi_{\rho_2, \sharp}^{\natural} = |W_{n_2}^D|^{-1} \sum_{w_2 \in W_{n_2, \sharp}} \rho_2(w_2) Q_{w_2}^{\natural}.$$

Pour $(w_1, w_2) \in W_{n_1} \times W_{n_2}$, on a l'égalité

$$\mathrm{sgn}(w_1) \mathrm{sgn}(w_2) \kappa_{\delta, 0}(w_1 \times w_2) = \sum_{(\rho_1, \rho_2) \in \widehat{W}_{n_1} \times \widehat{W}_{n_2}} m_{\delta}(\rho_1 \otimes \mathrm{sgn}, \rho_2 \otimes \mathrm{sgn}) \rho_1(w_1) \rho_2(w_2).$$

On obtient alors

$$\Theta_{\delta}(f_{\mathcal{O}_1, \mathcal{O}_2}) = q^{d_{\mathcal{O}_1, \mathcal{O}_2}} \sum_{(\rho_1, \rho_2) \in \widehat{W}_{n_1} \times \widehat{W}_{n_2}} m_{\delta}(\rho_1 \otimes \mathrm{sgn}, \rho_2 \otimes \mathrm{sgn}) \chi_{\rho_1}^{\natural}(Y_1) \chi_{\rho_2, \sharp}^{\natural}(Y_2). \quad (69)$$

Soit $(\mu_1, \eta_1) \in \mathcal{P}^{\mathrm{orth}}(2n_1 + 1)$. Supposons $k_{\mu_1, \eta_1} = 1$. On a défini en [Waldspurger 2001, VIII.13] une fonction $\chi_{\mu_1, \eta_1}^{\natural}$ sur l'ensemble des éléments nilpotents de $\mathfrak{so}(2n_1 + 1)(\mathbb{F}_q)$. Il existe un demi-entier d_{μ_1, η_1}

tel que $\chi_{\mu_1, \eta_1}^{\natural} = q^{d_{\mu_1, \eta_1}} \chi_{\rho_{\mu_1, \eta_1}}^{\natural}$. On note $\mathcal{P}^{\text{orth}}(2n_1 + 1; k = 1)$ le sous-ensemble des $(\mu_1, \eta_1) \in \mathcal{P}^{\text{orth}}(2n_1 + 1)$ tels que $k_{\mu_1, \eta_1} = 1$. Rappelons que l'application

$$\mathcal{P}^{\text{orth}}(2n_1 + 1; k = 1) \rightarrow \widehat{W}_{n_1}, \quad (\mu_1, \eta_1) \mapsto \rho_{\mu_1, \eta_1}$$

est bijective.

Soit $(\mu_2, \eta_2) \in \underline{\mathcal{P}}^{\text{orth}}(2n_2)$, cf. 1.5. Supposons $k_{\mu_2, \eta_2} = 0$. On a défini en [Waldspurger 2001, VIII.13] une fonction $\chi_{\mu_2, \eta_2, \sharp}^{\natural}$ sur l'ensemble des éléments nilpotents de $\mathfrak{so}(2n_2)_{\sharp}(\mathbb{F}_q)$ (on a ajouté un indice \sharp à la notation de [Waldspurger 2001]). Soit $(\mu_2, \eta_2) \in \mathcal{P}^{\text{orth}}(2n_2)$ tel que $k_{\mu_2, \eta_2} = 0$. Si μ_2 n'est pas exceptionnel, (μ_2, η_2) est aussi un élément de $\underline{\mathcal{P}}^{\text{orth}}(2n_2)$; la représentation ρ_{μ_2, ϵ_2} est définie ainsi que ses prolongements $\rho_{\mu_2, \epsilon_2}^+$ et $\rho_{\mu_2, \epsilon_2}^-$, cf. 1.12. La fonction $\chi_{\mu_2, \eta_2, \sharp}^{\natural}$ est aussi définie. Si μ_2 est exceptionnel, auquel cas $\eta_2 = 1$, μ_2 se relève en les deux éléments $(\mu_2, +)$ et $(\mu_2, -)$ de $\underline{\mathcal{P}}^{\text{orth}}(2n_2)$. Les représentations $\rho_{\mu_2, +, \eta_2}$ et $\rho_{\mu_2, -, \eta_2}$ sont définies. Les représentations $\rho_{\mu_2, +, \eta_2}^+$, $\rho_{\mu_2, +, \eta_2}^-$, $\rho_{\mu_2, -, \eta_2}^+$ et $\rho_{\mu_2, -, \eta_2}^-$ sont toutes égales. On note $\rho_{\mu_2, \eta_2}^+ = \rho_{\mu_2, \eta_2}^-$ ce prolongement. On pose

$$\chi_{\mu_2, \eta_2, \sharp}^{\natural} = \chi_{\mu_2, +, \eta_2, \sharp}^{\natural} + \chi_{\mu_2, -, \eta_2, \sharp}^{\natural}.$$

En tout cas, il existe un demi-entier d_{μ_2, η_2} tel que les égalités suivantes soient vérifiées :

$$\begin{aligned} \text{si } \sharp = \text{iso}, \quad \chi_{\mu_2, \eta_2, \text{iso}}^{\natural} &= q^{d_{\mu_2, \eta_2}} \chi_{\rho_{\mu_2, \eta_2, \text{iso}}^+}^{\natural} = q^{d_{\mu_2, \eta_2}} \chi_{\rho_{\mu_2, \eta_2, \text{iso}}^-}^{\natural}; \\ \text{si } \sharp = \text{an}, \quad \chi_{\mu_2, \eta_2, \text{an}}^{\natural} &= q^{d_{\mu_2, \eta_2}} \chi_{\rho_{\mu_2, \eta_2, \text{an}}^+}^{\natural} = -q^{d_{\mu_2, \eta_2}} \chi_{\rho_{\mu_2, \eta_2, \text{an}}^-}^{\natural}. \end{aligned}$$

Remarquons que, quand μ_2 est exceptionnel, on a $\chi_{\mu_2, \eta_2, \text{an}}^{\natural} = 0$.

On note $\mathcal{P}^{\text{orth}}(2n_2; k = 0)$ l'ensemble des $(\mu_2, \eta_2) \in \mathcal{P}^{\text{orth}}(2n_2)$ tels que $k_{\mu_2, \eta_2} = 0$. Rappelons que \widehat{W}_{n_2} est réunion disjointe des ensembles $\{\rho_{\mu_2, \eta_2}^+, \rho_{\mu_2, \eta_2}^-\}$ pour $(\mu_2, \eta_2) \in \mathcal{P}^{\text{orth}}(2n_2; k = 0)$ avec μ_2 non exceptionnel et des ensembles $\{\rho_{\mu_2, +, \eta_2}^+\} = \{\rho_{\mu_2, +, \eta_2}^-\} = \{\rho_{\mu_2, -, \eta_2}^+\} = \{\rho_{\mu_2, -, \eta_2}^-\}$ pour $(\mu_2, \eta_2) \in \mathcal{P}^{\text{orth}}(2n_2; k = 0)$ avec μ_2 exceptionnel.

On pose

$$\mathcal{PP}^{\text{orth}}(n_1, n_2) = \mathcal{P}^{\text{orth}}(2n_1 + 1; k = 1) \times \mathcal{P}^{\text{orth}}(2n_2; k = 0)$$

pour abrégier la notation. Pour $(\mu_1, \eta_1; \mu_2, \eta_2) \in \mathcal{PP}^{\text{orth}}(n_1, n_2)$, posons

— si $\sharp = \text{iso}$, $n_2 \neq 0$ et μ_2 n'est pas exceptionnel,

$$m_{\delta, \text{iso}}(\mu_1, \eta_1; \mu_2, \eta_2) = m_{\delta}(\rho_{\mu_1, \eta_1} \otimes \text{sgn}, \rho_{\mu_2, \eta_2}^+ \otimes \text{sgn}) + m_{\delta}(\rho_{\mu_1, \eta_1} \otimes \text{sgn}, \rho_{\mu_2, \eta_2}^- \otimes \text{sgn});$$

— si $\sharp = \text{an}$, $n_2 \neq 0$ et μ_2 n'est pas exceptionnel,

$$m_{\delta, \text{an}}(\mu_1, \eta_1; \mu_2, \eta_2) = m_{\delta}(\rho_{\mu_1, \eta_1} \otimes \text{sgn}, \rho_{\mu_2, \eta_2}^+ \otimes \text{sgn}) - m_{\delta}(\rho_{\mu_1, \eta_1} \otimes \text{sgn}, \rho_{\mu_2, \eta_2}^- \otimes \text{sgn});$$

— si $\sharp = \text{iso}$ et si $n_2 = 0$ ou μ_2 est exceptionnel

$$m_{\delta, \text{iso}}(\mu_1, \eta_1; \mu_2, \eta_2) = \frac{1}{2}(m_{\delta}(\rho_{\mu_1, \eta_1} \otimes \text{sgn}, \rho_{\mu_2, \eta_2}^+ \otimes \text{sgn}) + m_{\delta}(\rho_{\mu_1, \eta_1} \otimes \text{sgn}, \rho_{\mu_2, \eta_2}^- \otimes \text{sgn}));$$

— si $\sharp = \text{an}$ et μ_2 est exceptionnel

$$m_{\delta, \text{an}}(\mu_1, \eta_1; \mu_2, \eta_2) = 0.$$

On voit alors que la formule (69) se réécrit

$$\begin{aligned} \Theta_\delta(f_{\mathcal{O}_1, \mathcal{O}_2}) &= q^{d_{\mathcal{O}_1, \mathcal{O}_2}} \sum_{(\mu_1, \eta_1; \mu_2, \eta_2) \in \mathcal{PP}^{\text{orth}}(n_1, n_2)} q^{-d_{\mu_1, \eta_1} - d_{\mu_2, \eta_2}} m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2) \chi_{\mu_1, \eta_1}^\sharp(Y_1) \chi_{\mu_2, \eta_2, \sharp}^\sharp(Y_2). \end{aligned} \quad (70)$$

3.7. Traduction des conditions en termes de représentations de groupes de Weyl. Montrons qu'il nous suffit de prouver les deux assertions suivantes :

- (71) Soient $n_1, n_2 \in \mathbb{N}$ avec $n_1 + n_2 = n$ et $n_2 \geq 1$ si $\sharp = \text{an}$; soient $(\mu_1, \eta_1; \mu_2, \eta_2) \in \mathcal{PP}^{\text{orth}}(n_1, n_2)$; supposons $m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2) \neq 0$; alors $\mu_1 \cup \mu_2 \leq d(\lambda)$.
- (72) Il existe $n_1, n_2 \in \mathbb{N}$ avec $n_1 + n_2 = n$ et $n_2 \geq 1$ si $\sharp = \text{an}$ et il existe $(\mu_1, \eta_1; \mu_2, \eta_2) \in \mathcal{PP}^{\text{orth}}(n_1, n_2)$ tels que $m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2) \neq 0$ et $\mu_1 \cup \mu_2 = d(\lambda)$.

Soit $(\mathcal{O}_1, \mathcal{O}_2) \in \mathbf{Nil}_\sharp$. On en déduit des entiers n_1, n_2 comme en 3.6. Supposons $\Theta_\delta(f_{\mathcal{O}_1, \mathcal{O}_2}) \neq 0$. La formule (70) implique qu'il existe $(\mu_1, \eta_1; \mu_2, \eta_2) \in \mathcal{PP}^{\text{orth}}(n_1, n_2)$ tel que

$$m_{\delta, \text{iso}}(\mu_1, \eta_1; \mu_2, \eta_2) \neq 0 \quad \text{et} \quad \chi_{\mu_1, \eta_1}^\sharp(Y_1) \chi_{\mu_2, \eta_2, \sharp}^\sharp(Y_2) \neq 0.$$

Or la fonction $\chi_{\mu_1, \eta_1}^\sharp$ n'est non nulle que sur les orbites nilpotentes \mathcal{O}'_1 vérifiant $\mu(\mathcal{O}'_1) \leq \mu_1$, cf. [WalDSPURGER 2001, VIII.13]. Puisque $Y_1 \in \mathcal{O}_1$, cela entraîne $\mu(\mathcal{O}_1) \leq \mu_1$. La fonction $\chi_{\mu_2, \eta_2, \sharp}^\sharp$ vérifie une propriété analogue. Donc $\mu(\mathcal{O}_2) \leq \mu_2$. Grâce à (71), on a aussi $\mu_1 \cup \mu_2 \leq d(\lambda)$. Donc $\mu(\mathcal{O}_1) \cup \mu(\mathcal{O}_2) \leq d(\lambda)$, ce qui vérifie la propriété (66).

Fixons des données vérifiant (72). Considérons la somme

$$\Psi = \sum_{\eta'_1, \eta'_2} q^{-d_{\mu_1, \eta'_1} - d_{\mu_2, \eta'_2}} m_{\delta, \sharp}(\mu_1, \eta'_1; \mu_2, \eta'_2) \chi_{\mu_1, \eta'_1}^\sharp \chi_{\mu_2, \eta'_2, \sharp}^\sharp,$$

où η'_1 parcourt les éléments de $\{\pm 1\}^{\text{Jord}_{\text{bp}}(\mu_1)} / \{\pm 1\}$ tels que $k(\mu_1, \eta'_1) = 1$ et η'_2 parcourt les éléments de $\{\pm 1\}^{\text{Jord}_{\text{bp}}(\mu_2)} / \{\pm 1\}$ tels que $k(\mu_2, \eta'_2) = 1$. C'est une fonction invariante par conjugaison sur le produit des ensembles d'éléments nilpotents de $\mathfrak{so}(2n_1 + 1)(\mathbb{F}_q)$ et de $\mathfrak{so}(2n_2)_\sharp(\mathbb{F}_q)$. Notons $\mathcal{U}(\mu_1)$ la réunion des orbites nilpotentes \mathcal{O}_1 dans $\mathfrak{so}(2n_1 + 1)(\mathbb{F}_q)$ telles que $\mu(\mathcal{O}_1) = \mu_1$. Notons $\mathcal{U}(\mu_2)$ la réunion des orbites nilpotentes \mathcal{O}_2 dans $\mathfrak{so}(2n_2)_\sharp(\mathbb{F}_q)$ telles que $\mu(\mathcal{O}_2) = \mu_2$. Quand η'_1 décrit les éléments ci-dessus, les restrictions à $\mathcal{U}(\mu_1)$ des fonctions $\chi_{\mu_1, \eta'_1}^\sharp$ sont linéairement indépendantes, cf. [WalDSPURGER 2001, VIII.13]. Quand η'_2 décrit les éléments ci-dessus, les restrictions à $\mathcal{U}(\mu_2)$ des fonctions $\chi_{\mu_2, \eta'_2, \sharp}^\sharp$ sont elles aussi linéairement indépendantes (on doit remarquer que, si $\sharp = \text{an}$, l'hypothèse $m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2) \neq 0$ implique que μ_2 n'est pas exceptionnel). L'hypothèse $m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2) \neq 0$ implique donc que la restriction de Ψ à $\mathcal{U}(\mu_1) \times \mathcal{U}(\mu_2)$ est non nulle. Fixons donc des orbites $\mathcal{O}_1 \subset \mathcal{U}_1$ et $\mathcal{O}_2 \subset \mathcal{U}_2$ telles que Ψ soit non nulle sur $\mathcal{O}_1 \times \mathcal{O}_2$. On a $\mu(\mathcal{O}_1) \cup \mu(\mathcal{O}_2) = \mu_1 \cup \mu_2 = d(\lambda)$. Appliquons la formule (70) au

couple $(\mathcal{O}_1, \mathcal{O}_2)$. Notons plutôt $(\mu'_1, \eta'_1; \mu'_2, \eta'_2)$ les termes indexant la somme de cette formule. Je dis que, si $(\mu'_1, \mu'_2) \neq (\mu_1, \mu_2)$, le terme

$$m_{\delta, \#}(\mu'_1, \eta'_1; \mu'_2, \eta'_2) \chi_{\mu'_1, \eta'_1}^{\natural}(Y_1) \chi_{\mu'_2, \eta'_2, \#}^{\natural}(Y_2)$$

est nul. En effet, la non-nullité des deux derniers termes entraîne comme plus haut les inégalités $\mu(\mathcal{O}_1) \leq \mu'_1$ et $\mu(\mathcal{O}_2) \leq \mu'_2$, c'est-à-dire $\mu_1 \leq \mu'_1$ et $\mu_2 \leq \mu'_2$. La non-nullité du premier terme entraîne $\mu'_1 \cup \mu'_2 \leq d(\lambda)$ d'après (71). Puisque $\mu_1 \cup \mu_2 = d(\lambda)$, les inégalités précédentes sont forcément des égalités. Donc $\mu'_1 = \mu_1$ et $\mu'_2 = \mu_2$, contrairement à l'hypothèse. Dans la somme de (70) ne restent donc que les $(\mu'_1, \eta'_1; \mu'_2, \eta'_2)$ pour lesquels $\mu'_1 = \mu_1$ et $\mu'_2 = \mu_2$. C'est-à-dire que l'on obtient

$$\Theta_{\delta}(f_{\mathcal{O}_1, \mathcal{O}_2}) = q^{d_{\mathcal{O}_1, \mathcal{O}_2}} \Psi(Y_1, Y_2),$$

où $(Y_1, Y_2) \in \mathcal{O}_1 \times \mathcal{O}_2$. D'où $\Theta_{\delta}(f_{\mathcal{O}_1, \mathcal{O}_2}) \neq 0$. Alors $(\mathcal{O}_1, \mathcal{O}_2)$ vérifie (67).

3.8. Une description de $\text{Res}(\delta)$. On a fixé $(\lambda, s, \epsilon) \in \mathfrak{Irr}_{\text{unip}}$ en 3.5. Nous supposons désormais que c'est un élément de $\mathfrak{Irr}_{\text{unip-quad}}$, ce qui nous suffit d'après 3.4. On a montré en [Waldspurger 2018, §2.2] que cet ensemble s'identifiait à $\mathcal{P}_2^{\text{symp}}(2n)$. On note $(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-)$ l'élément de cet ensemble auquel s'identifie (λ, s, ϵ) . On a $\lambda = \lambda^+ \cup \lambda^-$ et cette décomposition est déterminée par l'élément s .

Notons $\mathfrak{n}(\lambda^+, \lambda^-)$ l'ensemble des couples de partitions symplectiques (ν^+, ν^-) vérifiant les conditions suivantes :

- (73) (a) $\nu^+ \cup \nu^- = \lambda$;
 (b) pour tout entier $i \in \mathbb{N}$ impair, $\text{mult}_{\nu^-}(i) = 0$;
 (c) pour tout $i \in \text{Jord}_{\text{bp}}(\lambda^+) \cap \text{Jord}_{\text{bp}}(\lambda^-)$, $\text{mult}_{\nu^-}(i) \leq 2$;
 (d) pour tout $i \in \text{Jord}_{\text{bp}}(\lambda)$ tel que $\text{mult}_{\lambda^+}(i) = 0$ ou $\text{mult}_{\lambda^-}(i) = 0$, $\text{mult}_{\nu^-}(i) \leq 1$.

Notons $\mathfrak{N}(\lambda_1, \lambda_2)$ l'ensemble des quadruplets $(\nu^+, \xi^+, \nu^-, \xi^-) \in \mathfrak{Irr}_{\text{unip-quad}}$ tels que $(\nu^+, \nu^-) \in \mathfrak{n}(\lambda^+, \lambda^-)$. Fixons un tel quadruplet $(\nu^+, \xi^+, \nu^-, \xi^-)$. Soit $i \in \text{Jord}_{\text{bp}}(\lambda)$. On définit un nombre complexe $e(i)$ par les formules suivantes, où on pose par convention $\xi^+(i) = 1$ si $i \notin \text{Jord}_{\text{bp}}(\nu^+)$ et $\xi^-(i) = 1$ si $i \notin \text{Jord}_{\text{bp}}(\nu^-)$:

- (74) (a) si $\text{mult}_{\nu^-}(i) = 0$, $e(i) = \xi^+(i)^{\text{mult}_{\lambda^-}(i)}$;
 (b) si $\text{mult}_{\nu^-}(i) = 1$, $\text{mult}_{\lambda^+}(i) \geq 1$ et $\text{mult}_{\lambda^-}(i) \geq 1$,

$$e(i) = \epsilon^-(i) \xi^-(i) \xi^+(i)^{\text{mult}_{\lambda^-}(i)-1} + \epsilon^+(i) \xi^+(i)^{\text{mult}_{\lambda^-}(i)}$$
 ;
 (c) si $\text{mult}_{\nu^-}(i) = 1$ et $\text{mult}_{\lambda^+}(i) = 0$, $e(i) = \epsilon^-(i) \xi^-(i) \xi^+(i)^{\text{mult}_{\lambda^-}(i)-1}$;
 (d) si $\text{mult}_{\nu^-}(i) = 1$ et $\text{mult}_{\lambda^-}(i) = 0$, $e(i) = \epsilon^+(i)$;
 (e) si $\text{mult}_{\nu^-}(i) = 2$, $e(i) = \epsilon^+(i) \epsilon^-(i) \xi^-(i) \xi^+(i)^{\text{mult}_{\lambda^-}(i)-1}$.

On pose

$$e(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-; \nu^+, \xi^+, \nu^-, \xi^-) = 2^{-|\text{Jord}_{\text{bp}}(\lambda^+)| - |\text{Jord}_{\text{bp}}(\lambda^-)|} \prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} e(i).$$

D'autre part, en [Waldspurger 2018, §1.11], on a associé à (v^+, ξ^+, v^-, ξ^-) un élément de \mathcal{R} que l'on a noté $j(\rho_{v^+, \xi^+} \otimes \rho_{v^-, \xi^-})$. Le terme j était un isomorphisme entre \mathcal{R} et un autre espace. Distinguer ces deux espaces nous était alors utile. Ce ne l'est plus, on identifie l'espace en question à \mathcal{R} grâce à l'isomorphisme j et on fait disparaître ce j de la notation.

Lemme. *On a l'égalité*

$$\kappa_\delta = \sum_{(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}(\lambda^+, \lambda^-)} e(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-; v^+, \xi^+, v^-, \xi^-) \rho \iota(\rho_{v^+, \xi^+} \otimes \rho_{v^-, \xi^-}).$$

Preuve. Rappelons quelques notations de [Waldspurger 2018, §1.3]. On fixe un homomorphisme $\rho_\lambda : \mathrm{SL}(2; \mathbb{C}) \rightarrow \mathrm{Sp}(2n; \mathbb{C})$ paramétré par λ et on note $Z(\lambda)$ le commutant dans $\mathrm{Sp}(2n; \mathbb{C})$ de son image. Le terme s appartient à $Z(\lambda)$ et vérifie $s^2 = 1$. On note $Z(\lambda, s)$ le commutant de s dans $Z(\lambda)$, $\mathbf{Z}(\lambda, s)$ son groupe des composantes connexes et $\mathbf{Z}(\lambda, s)^\vee$ le groupe des caractères de $\mathbf{Z}(\lambda, s)$. On a $\epsilon \in \mathbf{Z}(\lambda, s)^\vee$. Considérons un sous-ensemble $H \subset Z(\lambda, s)$ qui s'envoie bijectivement sur $\mathbf{Z}(\lambda, s)$ et est formé d'éléments h vérifiant $h^2 = 1$. Pour $h \in H$, le triplet (λ, s, h) appartient à l'ensemble $\mathfrak{Cn}\delta_{\mathrm{unip}\text{-}quad}$ de [Waldspurger 2018, §2.2]. Il lui est associé une représentation virtuelle

$$\Pi(\lambda, s, h) = \sum_{\epsilon' \in \mathbf{Z}(\lambda, s)^\vee} \pi(\lambda, s, \epsilon') \epsilon'(h). \quad (75)$$

Par inversion de Fourier dans le groupe $\mathbf{Z}(\lambda, s)$, on a

$$\pi(\lambda, s, \epsilon) = |\mathbf{Z}(\lambda, s)|^{-1} \sum_{h \in H} \Pi(\lambda, s, h) \epsilon(h).$$

On applique $\mathrm{Res} \circ D$ à cette égalité :

$$\mathrm{Res}(\delta) = |\mathbf{Z}(\lambda, s)|^{-1} \sum_{h \in H} \mathrm{Res} \circ D \circ \Pi(\lambda, s, h) \epsilon(h).$$

Utilisons l'involution $\mathcal{F}^{\mathrm{par}}$ de [Waldspurger 2018, §1.9]. Puisque, justement, c'est une involution, on peut composer à gauche le membre de droite ci-dessus par $\mathcal{F}^{\mathrm{par}} \circ \mathcal{F}^{\mathrm{par}}$. Le théorème 2.7 de [Waldspurger 2018] n'est plus conditionnel puisqu'on a démontré en [Waldspurger 2016b] le théorème 2.1 de [Waldspurger 2018]. Il nous dit que $\mathcal{F}^{\mathrm{par}} \circ \mathrm{Res} \circ D \circ \Pi(\lambda, s, h) = \mathrm{Res} \circ D \circ \Pi(\lambda, h, s)$. D'où

$$\mathrm{Res}(\delta) = |\mathbf{Z}(\lambda, s)|^{-1} \sum_{h \in H} \mathcal{F}^{\mathrm{par}} \circ \mathrm{Res} \circ D \circ \Pi(\lambda, h, s) \epsilon(h).$$

On peut encore développer le membre de droite en utilisant (75) où l'on échange s et h :

$$\mathrm{Res}(\delta) = |\mathbf{Z}(\lambda, s)|^{-1} \sum_{h \in H} \sum_{\xi \in \mathbf{Z}(\lambda, h)^\vee} \mathcal{F}^{\mathrm{par}} \circ \mathrm{Res} \circ D(\pi(\lambda, h, \xi)) \xi(s) \epsilon(h). \quad (76)$$

L'ensemble $\mathfrak{Cn}\delta_{\mathrm{unip}\text{-}quad}$ s'identifie à $\mathcal{P}_4^{\mathrm{symp}}(2n)$. Plus précisément, les éléments de $\mathfrak{Cn}\delta_{\mathrm{unip}\text{-}quad}$ de la forme (λ, s, h) (c'est-à-dire dont les deux premiers termes sont nos éléments fixés λ et s) s'identifient aux quadruplets $(\lambda^{++}, \lambda^{-+}, \lambda^{+-}, \lambda^{--}) \in \mathcal{P}_4^{\mathrm{symp}}(2n)$ tels que $\lambda^{++} \cup \lambda^{+-} = \lambda^+$ et $\lambda^{-+} \cup \lambda^{--} = \lambda^-$. Si

(λ, s, h) correspond ainsi à $(\lambda^{++}, \lambda^{-+}, \lambda^{+-}, \lambda^{--})$, l'image de h dans

$$\mathbf{Z}(\lambda, s) \simeq (\mathbb{Z}/2\mathbb{Z})^{\text{Jord}_{\text{bp}}(\lambda^+)} \times (\mathbb{Z}/2\mathbb{Z})^{\text{Jord}_{\text{bp}}(\lambda^-)}$$

est

$$(\text{mult}_{\lambda^{+-}}(i))_{i \in \text{Jord}_{\text{bp}}(\lambda^+)} \times (\text{mult}_{\lambda^{--}}(i))_{i \in \text{Jord}_{\text{bp}}(\lambda^-)}, \quad (77)$$

où il s'agit en fait des images des multiplicités dans $\mathbb{Z}/2\mathbb{Z}$. On voit que l'on peut choisir H de sorte que l'ensemble des (λ, s, h) pour $h \in H$ s'identifie à l'ensemble des quadruplets satisfaisant les conditions ci-dessus et de plus : $\text{mult}_{\lambda^{+-}}(i) \leq 1$ et $\text{mult}_{\lambda^{--}}(i) \leq 1$ pour tout i . On peut évidemment renforcer des inégalités en $\text{mult}_{\lambda^{+-}}(i) \leq \inf(1, \text{mult}_{\lambda^+}(i))$ et $\text{mult}_{\lambda^{--}}(i) \leq \inf(1, \text{mult}_{\lambda^-}(i))$ (par exemple, $\text{mult}_{\lambda^{+-}}(i) \leq \text{mult}_{\lambda^+}(i)$ puisque $\lambda^+ = \lambda^{++} \cup \lambda^{+-}$). On choisit ainsi l'ensemble H . Pour tout $h \in H$, continuons à noter $(\lambda^{++}, \lambda^{-+}, \lambda^{+-}, \lambda^{--})$ le quadruplet associé à (λ, s, h) et posons $v^+ = \lambda^{++} \cup \lambda^{-+}$, $v^- = \lambda^{+-} \cup \lambda^{--}$. Le couple (v^+, v^-) appartient à notre ensemble $\mathfrak{n}(\lambda^+, \lambda^-)$. Le groupe $\mathbf{Z}(\lambda, h)$ s'identifie à

$$(\mathbb{Z}/2\mathbb{Z})^{\text{Jord}_{\text{bp}}(v^+)} \times (\mathbb{Z}/2\mathbb{Z})^{\text{Jord}_{\text{bp}}(v^-)}$$

et un élément $\xi \in \mathbf{Z}(\lambda, h)^\vee$ s'identifie à un couple (ξ^+, ξ^-) . Le quadruplet (v^+, ξ^+, v^-, ξ^-) appartient à $\mathfrak{N}(\lambda^+, \lambda^-)$. Mais l'application $h \mapsto (v^+, v^-)$ n'est pas injective (h est une classe de conjugaison par $\mathbf{Z}(\lambda, s)$ et (v^+, v^-) paramètre sa classe de conjugaison par $\mathbf{Z}(\lambda)$). L'égalité (76) se récrit

$$\text{Res}(\delta) = \sum_{(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}(\lambda^+, \lambda^-)} f(v^+, \xi^+, v^-, \xi^-) \mathcal{F}^{\text{par}} \circ \text{Res} \circ D(\pi(v^+, \xi^+, v^-, \xi^-)),$$

où :

- pour (λ, h, ξ) correspondant à (v^+, ξ^+, v^-, ξ^-) , on a noté $\pi(v^+, \xi^+, v^-, \xi^-) = \pi(\lambda, h, \xi)$;
- $f(v^+, \xi^+, v^-, \xi^-)$ est la somme des $|\mathbf{Z}(\lambda, s)|^{-1} \epsilon(h) \xi(s)$ sur les $h \in H$ d'image (v^+, v^-) .

Pour $(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}(\lambda^+, \lambda^-)$, on a

$$\text{Res} \circ D(\pi(v^+, \xi^+, v^-, \xi^-)) = \text{Rep} \circ \rho \iota(\rho_{v^+, \xi^+} \otimes \rho_{v^-, \xi^-})$$

d'après la proposition 1.11 de [Waldspurger 2018]. On a aussi $\mathcal{F}^{\text{par}} \circ \text{Rep} = k$, cf. [Waldspurger 2018, §1.9]. Donc

$$\text{Res}(\delta) = \sum_{(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}(\lambda^+, \lambda^-)} f(v^+, \xi^+, v^-, \xi^-) k \circ \rho \iota(\rho_{v^+, \xi^+} \otimes \rho_{v^-, \xi^-}).$$

Puisque κ_δ est l'élément de \mathcal{R} tel que $\text{Res}(\delta) = k(\kappa_\delta)$, on obtient la formule de l'énoncé, à condition de prouver l'égalité

$$f(v^+, \xi^+, v^-, \xi^-) = e(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-; v^+, \xi^+, v^-, \xi^-) \quad \text{pour tout } (v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}(\lambda^+, \lambda^-).$$

Fixons donc $(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}(\lambda^+, \lambda^-)$. On compare tout de suite le facteur $|\mathbf{Z}(\lambda, s)|^{-1}$ figurant dans la définition de $f(v^+, \xi^+, v^-, \xi^-)$ avec le facteur $2^{-|\text{Jord}_{\text{bp}}(\lambda^+)| - |\text{Jord}_{\text{bp}}(\lambda^-)|}$ figurant dans celle de $e(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-; v^+, \xi^+, v^-, \xi^-)$: ils sont égaux. On note f le terme $f(v^+, \xi^+, v^-, \xi^-)$ privé de ce facteur et on doit prouver que $f = \prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} e(i)$, avec les notations précédant l'énoncé. Soit $h \in H$

d'image (ν^+, ν^-) . Notons encore $(\lambda^{++}, \lambda^{-+}, \lambda^{+-}, \lambda^{--})$ le quadruplet associé à (λ, s, h) . D'après (77), on a

$$\epsilon(h) = \left(\prod_{i \in \text{Jord}_{\text{bp}}(\lambda^+)} \epsilon^+(i)^{\text{mult}_{\lambda^{+-}}(i)} \right) \left(\prod_{i \in \text{Jord}_{\text{bp}}(\lambda^-)} \epsilon^-(i)^{\text{mult}_{\lambda^{--}}(i)} \right).$$

On simplifie cette égalité en

$$\epsilon(h) = \prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} \epsilon^+(i)^{\text{mult}_{\lambda^{+-}}(i)} \epsilon^-(i)^{\text{mult}_{\lambda^{--}}(i)},$$

avec la convention $\epsilon^+(i) = 1$ si $i \notin \text{Jord}_{\text{bp}}(\lambda^+)$ et $\epsilon^-(i) = 1$ pour $i \notin \text{Jord}_{\text{bp}}(\lambda^-)$. Le quadruplet associé à (λ, h, s) est $(\lambda^{++}, \lambda^{+-}, \lambda^{-+}, \lambda^{--})$. Le couple (ξ^+, ξ^-) s'identifie à un élément de $\mathbf{Z}(\lambda, h)^\vee$ et on a la formule similaire (avec une convention analogue) :

$$\xi(s) = \prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} \xi^+(i)^{\text{mult}_{\lambda^{-+}}(i)} \xi^-(i)^{\text{mult}_{\lambda^{--}}(i)}.$$

Posons simplement $m^+(i) = \text{mult}_{\lambda^{+-}}(i)$ et $m^-(i) = \text{mult}_{\lambda^{--}}(i)$. On a $\text{mult}_{\lambda^{-+}}(i) = \text{mult}_{\lambda^-}(i) - m^-(i)$ et on obtient

$$\epsilon(h)\xi(s) = \prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} \epsilon^+(i)^{m^+(i)} \epsilon^-(i)^{m^-(i)} \xi^+(i)^{\text{mult}_{\lambda^-}(i) - m^-(i)} \xi^-(i)^{m^-(i)}.$$

Pour $i \in \text{Jord}_{\text{bp}}(\lambda)$, notons $E(i)$ l'ensemble des couples $(m^+, m^-) \in \{0, 1\}^2$ tels que

$$m^+ \leq \inf(1, \text{mult}_{\lambda^+}(i)) \quad \text{et} \quad m^- \leq \inf(1, \text{mult}_{\lambda^-}(i)); \quad m^+ + m^- = \text{mult}_{\nu^-}(i).$$

L'application

$$h \mapsto (m^+(i), m^-(i))_{i \in \text{Jord}_{\text{bp}}(\lambda)}$$

identifie l'ensemble des $h \in H$ d'image (ν^+, ν^-) avec $\prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} E(i)$. On voit alors que $f = \prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} f_i$, où, pour $i \in \text{Jord}_{\text{bp}}(\lambda)$, on a posé

$$f_i = \sum_{(m^+, m^-) \in E(i)} \epsilon^+(i)^{m^+} \epsilon^-(i)^{m^-} \xi^+(i)^{\text{mult}_{\lambda^-}(i) - m^-} \xi^-(i)^{m^-}.$$

Il reste à démontrer l'égalité $f_i = e_i$ pour tout $i \in \text{Jord}_{\text{bp}}(\lambda)$. C'est un calcul élémentaire que l'on effectue en distinguant chacun des cas (74)(a) à (74)(e). On le laisse au lecteur. \square

3.9. Preuve de (71). On fixe $n_1, n_2 \in \mathbb{N}$ avec $n_1 + n_2 = n$ et $n_2 \geq 1$ si $\sharp = \text{an}$. On fixe $(\mu_1, \eta_1; \mu_2, \eta_2) \in \mathcal{PP}^{\text{orth}}(n_1, n_2)$ et on suppose $m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2) \neq 0$. D'après la définition de $m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2) \neq 0$, on peut fixer $\zeta' = \pm$ tel que $m_{\delta}(\rho_{\mu_1, \eta_1} \otimes \text{sgn}, \rho_{\mu_2, \eta_2}^{\zeta'} \otimes \text{sgn}) \neq 0$.

On a défini en les paragraphes 1.4 et 1.5 les partitions spéciales $\text{sp}(\mu_1, \eta_1)$ et $\text{sp}(\mu_2, \eta_2)$. D'après les résultats de ces paragraphes, on a

$$\mu_1 \leq \text{sp}(\mu_1, \eta_1), \quad \mu_2 \leq \text{sp}(\mu_2, \eta_2). \quad (78)$$

Le symbole associé à ρ_{μ_1, η_1} appartient à la famille de $\text{sp}(\mu_1, \eta_1)$. Posons $\lambda_1 = d(\text{sp}(\mu_1, \eta_1))$ et $\rho_1 = \rho_{\mu_1, \eta_1} \otimes \text{sgn}$. D'après 1.6, le symbole associé à ρ_1 appartient à la famille de λ_1 . Posons $\lambda_2 = d(\text{sp}(\mu_2, \eta_2))$

et $\rho_2 = \rho_{\mu_2, \eta_2} \otimes \text{sgn}$ si μ_2 n'est pas exceptionnel. Si μ_2 est exceptionnel, on relève μ_2 en un élément $\underline{\mu}_2$ de $\mathcal{P}^{\text{orth}}(2n_2)$ et on pose $\rho_2 = \rho_{\underline{\mu}_2, \eta_2} \otimes \text{sgn}$. On a de même : le symbole associé à ρ_2 appartient à la famille de λ_2 . La représentation $\rho_{\underline{\mu}_2, \eta_2}^{\zeta'} \otimes \text{sgn}$ est l'un des prolongements de ρ_2 à W_{n_2} donc est de la forme ρ_2^ζ pour un $\zeta = \pm$ (on n'a pas en général $\zeta = \zeta'$ mais peu importe). On a donc $m_\delta(\rho_1, \rho_2^\zeta) \neq 0$. Rappelons que ce terme est la "multiplicité" de $\rho_1 \otimes \rho_2^\zeta$ dans $\kappa_{\delta, 0}$. La fonction κ_δ est calculée par le lemme 3.8 (rappelons que l'on suppose $(\lambda, s, \epsilon) \in \mathfrak{Irr}_{\text{unip-quad}}$). La fonction $\kappa_{\delta, 0}$ est calculée par une formule analogue, où l'on se restreint aux $(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}(\lambda^+, \lambda^-)$ tels que $k_{v^+, \xi^+} = k_{v^-, \xi^-} = 0$. Notons ce sous-ensemble $\mathfrak{N}_0(\lambda^+, \lambda^-)$. D'après ce lemme, on peut fixer un élément $(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}_0(\lambda^+, \lambda^-)$ tel que la multiplicité de $\rho_1 \otimes \rho_2^\zeta$ dans $\rho_\iota(\rho_{v^+, \xi^+} \otimes \rho_{v^-, \xi^-})$ est non nulle. On sait que la représentation ρ_{v^+, ξ^+} est de la forme

$$\rho_{v^+, \xi^+} = \sum_{\dot{v}^+, \dot{\xi}^+} c(v^+, \xi^+; \dot{v}^+, \dot{\xi}^+) \rho_{\dot{v}^+, \dot{\xi}^+},$$

où $(\dot{v}^+, \dot{\xi}^+)$ parcourt les éléments de $\mathcal{P}^{\text{symp}}(S(v^+))$ tels que $k_{\dot{v}^+, \dot{\xi}^+} = 0$. On note $\mathcal{P}^{\text{symp}}(S(v^+); k=0)$ l'ensemble de ces éléments. Le coefficient $c(v^+, \xi^+; \dot{v}^+, \dot{\xi}^+)$ n'est non nul que si $v^+ \leq \dot{v}^+$. De mêmes propriétés valent pour ρ_{v^-, ξ^-} . On peut donc fixer $(\dot{v}^+, \dot{\xi}^+) \in \mathcal{P}^{\text{symp}}(S(v^+); k=0)$ et $(\dot{v}^-, \dot{\xi}^-) \in \mathcal{P}^{\text{symp}}(S(v^-); k=0)$ tels que

$$v^+ \leq \dot{v}^+, \quad v^- \leq \dot{v}^-, \quad (79)$$

et la multiplicité de $\rho_1 \otimes \rho_2^\zeta$ dans $\rho_\iota(\rho_{\dot{v}^+, \dot{\xi}^+} \otimes \rho_{\dot{v}^-, \dot{\xi}^-})$ soit non nulle. Cette multiplicité est exactement le terme $m(\rho_1, \rho_2^\zeta; \rho_{\dot{v}^+, \dot{\xi}^+}, \rho_{\dot{v}^-, \dot{\xi}^-})$ défini en 1.12. D'après la proposition de ce paragraphe, la non-nullité de cette multiplicité entraîne

$$\dot{v}^+ \cup \dot{v}^- \leq \text{ind}(\lambda_1, \lambda_2). \quad (80)$$

Par définition de $\mathfrak{N}(\lambda^+, \lambda^-)$, on a $v^+ \cup v^- = \lambda$. Les inégalités (79) et (80) entraînent $\lambda \leq \text{ind}(\lambda_1, \lambda_2)$ d'où aussi $d(\text{ind}(\lambda_1, \lambda_2)) \leq d(\lambda)$ puisque la dualité est décroissante. On applique la proposition 1.9 : $d(\lambda_1) \cup d(\lambda_2) \leq d(\text{ind}(\lambda_1, \lambda_2))$, d'où aussi $d(\lambda_1) \cup d(\lambda_2) \leq d(\lambda)$. Or $d(\lambda_1) = \text{sp}(\mu_1, \eta_1)$ et $d(\lambda_2) = \text{sp}(\mu_2, \eta_2)$ par définition. En utilisant les inégalités (78), on en déduit $\mu_1 \cup \mu_2 \leq d(\lambda)$, ce qui démontre (71).

3.10. Preuve de (72). On a supposé $(\lambda, s, \epsilon) \in \mathfrak{Irr}_{\text{unip-quad}}$. Maintenant, on suppose de plus que les termes de λ sont tous pairs. C'est loisible d'après 3.4.

On fixe une fonction $\tau : \text{Jord}_{\text{bp}}(\lambda) \rightarrow \mathbb{Z}/2\mathbb{Z}$ vérifiant les conditions suivantes :

(81) (a) Pour $i \in \text{Jord}_{\text{bp}}(\lambda)$ tel que $\text{mult}_\lambda(i) = 1$, $\tau(i) = 0$.

(b) Pour $i \in \text{Jord}_{\text{bp}}(\lambda^+) \cap \text{Jord}_{\text{bp}}(\lambda^-)$, $(-1)^{\tau(i)} = \epsilon^+(i)\epsilon^-(i)$.

Remarquons que les deux cas sont exclusifs : dans le cas (b), on a $\text{mult}_\lambda(i) \geq 2$.

Appliquant la proposition 1.11, on introduit des entiers $n_1, n_2 \in \mathbb{N}$ tels que $n_1 + n_2 = n$ et des partitions $\lambda_1 \in \mathcal{P}^{\text{symp, sp}}(2n_1)$, $\lambda_2 \in \mathcal{P}^{\text{orth, sp}}(2n_2)$ telles que λ_1 et λ_2 induisent régulièrement λ , $d(\lambda_1) \cup d(\lambda_2) = d(\lambda)$ et $\tau_{\lambda_1, \lambda_2} = \tau$. On fixe des couples (τ_1, δ_1) et (τ_2, δ_2) paramétrant des symboles (X_1, Y_1) dans la famille de λ_1 et (X_2, Y_2) dans la famille de λ_2 , avec $\delta_1 = 0$ et $\delta_2 = 0$. On pose $\mu_1 = d(\lambda_1)$, $\mu_2 = d(\lambda_2)$. Ce sont des

partitions spéciales. Le symbole $d(X_1, Y_1)$ appartient à la famille de μ_1 et est paramétré par un couple (τ'_1, δ'_1) tel que $\delta'_1 = 0$. D'après le lemme 1.4, c'est le symbole de la représentation ρ_{μ_1, η_1} pour un couple $(\mu_1, \eta_1) \in \mathcal{P}^{\text{orth}}(2n_1 + 1; k = 1)$. Le symbole $d(X_2, Y_2)$ appartient à la famille de μ_2 et est paramétré par un couple (τ'_2, δ'_2) tel que $\delta'_2 = 0$. Supposons μ_2 non exceptionnel. D'après le lemme 1.5, $d(X_2, Y_2)$ est le symbole de la représentation ρ_{μ_2, η_2} pour un couple $(\mu_2, \eta_2) \in \mathcal{P}^{\text{orth}}(2n_2; k = 0)$. Dans le cas où μ_2 est exceptionnel, on a de même un couple (μ_2, η_2) mais la représentation ρ_{μ_2, η_2} doit être remplacée par $\rho_{\underline{\mu}_2, \eta_2}$, où $\underline{\mu}_2$ est l'un des relèvements de μ_2 dans $\underline{\mathcal{P}}^{\text{orth}}(2n_2)$. On va montrer que le quadruplet $(\mu_1, \eta_1; \mu_2, \eta_2)$ vérifie la condition (72).

Tout d'abord, on a $\mu_1 \cup \mu_2 = d(\lambda_1) \cup d(\lambda_2) = d(\lambda)$.

On veut prouver que $m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2) \neq 0$. Posons $\text{sgn}_{\sharp} = 1$ si $\sharp = \text{iso}$, $\text{sgn}_{\sharp} = -1$ si $\sharp = \text{an}$. Le terme $m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2)$ est égal à

$$m_{\delta}(\rho_{\mu_1, \eta_1} \otimes \text{sgn}, \rho_{\mu_2, \eta_2}^+ \otimes \text{sgn}) + \text{sgn}_{\sharp} m_{\delta}(\rho_{\mu_1, \eta_1} \otimes \text{sgn}, \rho_{\mu_2, \eta_2}^- \otimes \text{sgn}), \tag{82}$$

éventuellement divisé par $\frac{1}{2}$. La représentation $\rho_{\mu_1, \eta_1} \otimes \text{sgn}$ n'est autre que la représentation ρ_1 de W_{n_1} dont le symbole est (X_1, Y_1) . Les représentations $\rho_{\mu_2, \eta_2}^+ \otimes \text{sgn}$ et $\rho_{\mu_2, \eta_2}^- \otimes \text{sgn}$ sont les prolongements (éventuellement égaux) à W_{n_2} d'une représentation $\rho_2 \in W_{n_2}^D$ dont le symbole est (X_2, Y_2) . Le terme (82) est égal, au signe près, à

$$m_{\delta}(\rho_1, \rho_2^+) + \text{sgn}_{\sharp} m_{\delta}(\rho_1, \rho_2^-). \tag{83}$$

Soit $\zeta = \pm$. En reprenant la preuve du paragraphe précédent, on calcule

$$\begin{aligned} & m_{\delta}(\rho_1, \rho_2^{\zeta}) \\ &= \sum_{(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}_0(\lambda^+, \lambda^-)} e(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-; v^+, \xi^+, v^-, \xi^-) \\ & \quad \times \sum_{(\dot{v}^+, \dot{\xi}^+) \in \mathcal{P}^{\text{symp}}(S(v^+); k=0)} \sum_{(\dot{v}^-, \dot{\xi}^-) \in \mathcal{P}^{\text{symp}}(S(v^-); k=0)} c(v^+, \xi^+; \dot{v}^+, \dot{\xi}^+) \\ & \quad \times c(v^-, \xi^-; \dot{v}^-, \dot{\xi}^-) m(\rho_1, \rho_2^{\zeta}; \rho_{v^+, \xi^+}, \rho_{v^-, \xi^-}). \end{aligned}$$

Comme dans le paragraphe précédent, la non-nullité du terme que l'on somme entraîne les inégalités (79) et (80) de ce paragraphe. On a $v^+ \cup v^- = \lambda$ et, ici, on sait par hypothèse que $\text{ind}(\lambda_1, \lambda_2) = \lambda$. Ces inégalités (79) et (80) sont donc des égalités. Maintenant que $v^+ = \dot{v}^+$, on sait que la relation $c(v^+, \xi^+; \dot{v}^+, \dot{\xi}^+) \neq 0$ équivaut à $\dot{\xi}^+ = \xi^+$ et que, si elle est vérifiée, on a $c(v^+, \xi^+; \dot{v}^+, \dot{\xi}^+) = 1$. De même bien sûr pour les objets associés à v^- . Cela nous débarrasse des sommes en $(\dot{v}^+, \dot{\xi}^+)$ et $(\dot{v}^-, \dot{\xi}^-)$ et des coefficients $c(v^+, \xi^+; \dot{v}^+, \dot{\xi}^+)$ et $c(v^-, \xi^-; \dot{v}^-, \dot{\xi}^-)$. On a simplement

$$m_{\delta}(\rho_1, \rho_2^{\zeta}) = \sum_{(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}_0(\lambda^+, \lambda^-)} e(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-; v^+, \xi^+, v^-, \xi^-) m(\rho_1, \rho_2^{\zeta}; \rho_{v^+, \xi^+}, \rho_{v^-, \xi^-}).$$

Les éléments $(v^+, \xi^+, v^-, \xi^-) \in \mathfrak{N}_0(\lambda^+, \lambda^-)$ pour lesquels $m(\rho_1, \rho_2^{\zeta}; \rho_{v^+, \xi^+}, \rho_{v^-, \xi^-}) \neq 0$ sont exactement les éléments de $\mathfrak{N}(\lambda^+, \lambda^-)$ qui vérifient l'hypothèse (B)⁵ de 1.13. D'après la proposition de

ce paragraphe, ce sont aussi les éléments de $\mathfrak{N}(\lambda^+, \lambda^-)$ qui vérifient (A) $^\zeta$ et, pour ces éléments, on a $m(\rho_1, \rho_2^\zeta; \rho_{v^+, \xi^+}, \rho_{v^-, \xi^-}) = 1$. La condition (A) $^\zeta$ se décompose en deux :

$$\text{pour } i \in \text{Jord}_{\text{bp}}(\lambda), \quad \text{mult}_{v^-}(i) \equiv c^\zeta(i), \quad (84)$$

où on a posé $c^\zeta(i) = \delta^{-\zeta}(i) - \delta^{-\zeta}(i^+)$;

$$\text{pour } i \in \text{Jord}_{\text{bp}}(v^+), \quad \xi^+(i) = (-1)^{\tau^\zeta(i)}; \quad \text{pour } i \in \text{Jord}_{\text{bp}}(v^-), \quad \xi^-(i) = (-1)^{\tau^{-\zeta}(i)}. \quad (85)$$

Notons $n^\zeta(\lambda_1, \lambda_2)$ l'ensemble des $(v^+, v^-) \in n(\lambda_1, \lambda_2)$ vérifiant la condition (84). Dans la suite du calcul, pour $(v^+, v^-) \in n^\zeta(\lambda_1, \lambda_2)$, notons (ξ^+, ξ^-) le couple déterminé par la condition (85). On obtient

$$m_\delta(\rho_1, \rho_2^\zeta) = \sum_{(v^+, v^-) \in n^\zeta(\lambda^+, \lambda^-)} e(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-; v^+, \xi^+, v^-, \xi^-). \quad (86)$$

Soit $i \in \text{Jord}_{\text{bp}}(\lambda)$. Définissons un ensemble $M^\zeta(i)$ par les égalités suivantes :

- si $c^\zeta(i) = 1$, $M^\zeta(i) = \{1\}$;
- si $c^\zeta(i) = 0$ et $\text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) = 0$, $M^\zeta(i) = \{0\}$;
- si $c^\zeta(i) = 0$ et $\text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) \neq 0$, $M^\zeta(i) = \{0, 2\}$.

On vérifie à l'aide de la définition de 3.8 et de la relation (84) ci-dessus que l'application $(v^+, v^-) \mapsto (\text{mult}_{v^-}(i))_{i \in \text{Jord}_{\text{bp}}(\lambda)}$ est une bijection de $n^\zeta(\lambda^+, \lambda^-)$ sur $\prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} M^\zeta(i)$.

Soit $i \in \text{Jord}_{\text{bp}}(\lambda)$ et $m \in M^\zeta(i)$. Définissons un nombre $e^\zeta(i, m)$ par les égalités suivantes :

$$(87) \text{ (a) } e^\zeta(i, 0) = (-1)^{\tau^\zeta(i) \text{mult}_{\lambda^-}(i)};$$

$$\text{(b) si } \text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) \neq 0,$$

$$e^\zeta(i, 1) = \epsilon^+(i)(-1)^{\tau^\zeta(i) \text{mult}_{\lambda^-}(i)} + \epsilon^-(i)(-1)^{\tau^{-\zeta}(i) + \tau^\zeta(i)(\text{mult}_{\lambda^-}(i) - 1)};$$

$$\text{(c) si } \text{mult}_{\lambda^+}(i) = 0, \quad e^\zeta(i, 1) = \epsilon^-(i)(-1)^{\tau^{-\zeta}(i) + \tau^\zeta(i)(\text{mult}_{\lambda^-}(i) - 1)};$$

$$\text{(d) si } \text{mult}_{\lambda^-}(i) = 0, \quad e^\zeta(i, 1) = \epsilon^+(i);$$

$$\text{(e) } e^\zeta(i, 2) = \epsilon^+(i)\epsilon^-(i)(-1)^{\tau^{-\zeta}(i) + \tau^\zeta(i)(\text{mult}_{\lambda^-}(i) - 1)}.$$

Pour $(v^+, v^-) \in n^\zeta(\lambda^+, \lambda^-)$, on vérifie à l'aide de la définition de 3.8 et de la relation (85) ci-dessus que l'on a l'égalité

$$e(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-; v^+, \xi^+, v^-, \xi^-) = 2^{-|\text{Jord}_{\text{bp}}(\lambda^+)| - |\text{Jord}_{\text{bp}}(\lambda^-)|} \prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} e^\zeta(i, \text{mult}_{v^-}(i)).$$

La relation (86) se transforme en

$$m_\delta(\rho_1, \rho_2^\zeta) = 2^{-|\text{Jord}_{\text{bp}}(\lambda^+)| - |\text{Jord}_{\text{bp}}(\lambda^-)|} \prod_{i \in \text{Jord}_{\text{bp}}(\lambda)} S^\zeta(i), \quad (88)$$

où on a posé

$$S^\zeta(i) = \sum_{m \in M^\zeta(i)} e^\zeta(i, m).$$

Fixons $i \in \text{Jord}_{\text{bp}}(\lambda)$ et calculons $S^\zeta(i)$. Remarquons qu'en vertu de la condition (81)(b) ci-dessus, de l'égalité $\tau = \tau_{\lambda_1, \lambda_2}$ et de la remarque (1) de 1.13, on a l'égalité

$$(-1)^{\tau^+(i)+\tau^-(i)} = \epsilon^+(i)\epsilon^-(i) \quad \text{si } \text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) \neq 0. \quad (89)$$

La définition des termes $e^\zeta(i, m)$ se simplifie alors dans les deux cas suivants :

$$\text{si } \text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) \neq 0, \quad e^\zeta(i, 1) = 2\epsilon^+(i)(-1)^{\tau^\zeta(i) \text{mult}_{\lambda^-}(i)}; \quad e^\zeta(i, 2) = (-1)^{\tau^\zeta(i) \text{mult}_{\lambda^-}(i)}.$$

On calcule alors :

- (90) (a) si $c^\zeta(i) = 1$ et $\text{mult}_{\lambda^-}(i) = 0$, $S^\zeta(i) = \epsilon^+(i)$;
 (b) si $c^\zeta(i) = 1$ et $\text{mult}_{\lambda^+}(i) = 0$, $S^\zeta(i) = \epsilon^-(i)(-1)^{\tau^-(i)+\tau^\zeta(i)(\text{mult}_{\lambda^-}(i)-1)}$;
 (c) si $c^\zeta(i) = 1$ et $\text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) \neq 0$, $S^\zeta(i) = 2\epsilon^+(i)(-1)^{\text{mult}_{\lambda^-}(i)\tau^\zeta(i)}$;
 (d) si $c^\zeta(i) = 0$ et $\text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) = 0$, $S^\zeta(i) = (-1)^{\text{mult}_{\lambda^-}(i)\tau^\zeta(i)}$;
 (e) si $c^\zeta(i) = 0$ et $\text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) \neq 0$, $S^\zeta(i) = 2(-1)^{\text{mult}_{\lambda^-}(i)\tau^\zeta(i)}$.

On voit que $S^\zeta(i)$ est non nul pour tout i . On déduit de (88) que $m_\delta(\rho_1, \rho_2^\zeta) \neq 0$. Mais cela ne suffit pas à prouver que l'expression (83) est non nulle. Pour cela, montrons que, pour tout $i \in \text{Jord}_{\text{bp}}(\lambda)$, on a l'égalité

$$S^-(i) = \epsilon^+(i)^{\text{mult}_{\lambda^+}(i)} \epsilon^-(i)^{\text{mult}_{\lambda^-}(i)} S^+(i), \quad (91)$$

où, pour simplifier les notations, on a posé $\epsilon^+(i) = 1$ si $i \notin \text{Jord}_{\text{bp}}(\lambda^+)$ et $\epsilon^-(i) = 1$ si $i \notin \text{Jord}_{\text{bp}}(\lambda^-)$. La vérification de cette assertion se fait cas par cas. Traitons seulement le cas où $\text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) \neq 0$. D'après la définition de $c^\zeta(i)$ et la remarque (2) de 1.13, on a la relation $c^+(i) + c^-(i) \equiv \text{mult}_\lambda(i) \pmod{2\mathbb{Z}}$. Supposons d'abord $\text{mult}_\lambda(i)$ pair. Alors $c^+(i) = c^-(i)$. Si ces deux nombres valent 1, on a d'après (90)(c)

$$S^+(i) = 2\epsilon^+(i)(-1)^{\text{mult}_{\lambda^-}(i)\tau^+(i)}, \quad S^-(i) = 2\epsilon^+(i)(-1)^{\text{mult}_{\lambda^-}(i)\tau^-(i)}.$$

D'où $S^-(i) = (-1)^{(\tau^+(i)+\tau^-(i)) \text{mult}_{\lambda^-}(i)} S^+(i)$. En vertu de (89), cela équivaut

$$S^-(i) = \epsilon^+(i)^{\text{mult}_{\lambda^-}(i)} \epsilon^-(i)^{\text{mult}_{\lambda^-}(i)} S^+(i).$$

Mais, puisque $\text{mult}_\lambda(i)$ est pair, $\text{mult}_{\lambda^+}(i)$ et $\text{mult}_{\lambda^-}(i)$ sont de même parité et l'égalité précédente coïncide avec (91). Si $c^+(i) = c^-(i) = 0$, on a d'après (90)(e)

$$S^+(i) = 2(-1)^{\text{mult}_{\lambda^-}(i)\tau^+(i)}, \quad S^-(i) = 2(-1)^{\text{mult}_{\lambda^-}(i)\tau^-(i)},$$

d'où encore $S^-(i) = \epsilon^+(i)^{\text{mult}_{\lambda^-}(i)} \epsilon^-(i)^{\text{mult}_{\lambda^-}(i)} S^+(i)$ et la même conclusion. Supposons maintenant $\text{mult}_\lambda(i)$ impair. Alors $c^+(i) \neq c^-(i)$. Soit $\zeta = \pm$ tel que $c^\zeta(i) = 1$ et $c^{-\zeta}(i) = 0$. D'après (90)(c) et (90)(e), on a

$$S^\zeta(i) = 2\epsilon^+(i)(-1)^{\text{mult}_{\lambda^-}(i)\tau^\zeta(i)}, \quad S^{-\zeta}(i) = 2(-1)^{\text{mult}_{\lambda^-}(i)\tau^{-\zeta}(i)}.$$

D'où

$$S^\zeta(i) = \epsilon^+(i)(-1)^{\text{mult}_{\lambda^-}(i)(\tau^+(i)+\tau^-(i))} S^{-\zeta}(i),$$

ou encore, d'après (89) :

$$S^\zeta(i) = \epsilon^+(i)(\epsilon^+(i)\epsilon^-(i))^{\text{mult}_{\lambda^+}(i)} S^{-\zeta}(i) = \epsilon^+(i)^{1+\text{mult}_{\lambda^+}(i)} \epsilon^-(i)^{\text{mult}_{\lambda^+}(i)} S^{-\zeta}(i).$$

Mais $\text{mult}_{\lambda^+}(i)$ est impair donc $1 + \text{mult}_{\lambda^+}(i)$ est de la même parité que $\text{mult}_{\lambda^+}(i)$. L'égalité précédente coïncide avec (91). Cela démontre (91) dans le cas où $\text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) \neq 0$. On laisse les autres cas au lecteur.

En vertu de (88) et (91), on a l'égalité

$$m_\delta(\rho_1, \rho_2^+) = c m_\delta(\rho_1, \rho_2^-), \quad (92)$$

où

$$c = \left(\prod_{i \in \text{Jord}_{\text{bp}}(\lambda^+)} \epsilon^+(i)^{\text{mult}_{\lambda^+}(i)} \right) \left(\prod_{i \in \text{Jord}_{\text{bp}}(\lambda^-)} \epsilon^-(i)^{\text{mult}_{\lambda^-}(i)} \right).$$

Mais on a vu en [Waldspurger 2018, §1.3(1)] que ce produit déterminait l'indice \sharp : celui-ci est iso si $c = 1$, an si $c = -1$. Autrement dit, $c = \text{sgn}_\sharp$. L'égalité (92) et la non nullité de ses deux membres entraînent la non-nullité de l'expression (83). Cela achève de prouver que $m_{\delta, \sharp}(\mu_1, \eta_1; \mu_2, \eta_2) \neq 0$.

Il reste une dernière condition à prouver, à savoir que $n_2 > 0$ si $\sharp = \text{an}$. Mais, si $n_2 = 0$, les représentations ρ_2^+ et ρ_2^- sont les mêmes : ce sont l'unique représentation du groupe $W_0 = \{1\}$. Donc $m_\delta(\rho_1, \rho_2^+) = m_\delta(\rho_1, \rho_2^-)$ et le calcul ci-dessus entraîne que $\sharp = \text{iso}$. Cela achève la vérification de la condition (72) et en même temps la preuve du théorème 3.3.

Remarque. On peut vérifier directement que, si $\sharp = \text{an}$, le couple (n_1, n_2) fourni par la proposition 1.11, pour notre fonction τ , vérifie $n_2 \neq 0$. En effet, supposons par l'absurde que $n_2 = 0$. A fortiori, l'ensemble d'intervalles de λ_2 est vide donc j et $j+1$ ne sont 2-liés pour aucun $j \geq 1$. Pour j impair, la condition (21) entraîne $\lambda_j = \lambda_{j+1}$. Donc les multiplicités $\text{mult}_{\lambda^+}(i)$ sont toutes paires. Puisque $\sharp = \text{an}$, on a

$$\left(\prod_{i \in \text{Jord}_{\text{bp}}(\lambda^+)} \epsilon^+(i)^{\text{mult}_{\lambda^+}(i)} \right) \left(\prod_{i \in \text{Jord}_{\text{bp}}(\lambda^-)} \epsilon^-(i)^{\text{mult}_{\lambda^-}(i)} \right) = -1.$$

Un $i \in \text{Jord}_{\text{bp}}(\lambda)$ tel que $\text{mult}_{\lambda^+}(i) \text{mult}_{\lambda^-}(i) = 0$ n'intervient pas dans ce produit. Par exemple, si $\text{mult}_{\lambda^-}(i) = 0$, il n'intervient évidemment pas dans le second produit. Il intervient dans le premier par $\epsilon^+(i)^{\text{mult}_{\lambda^+}(i)}$. Mais $\text{mult}_{\lambda^+}(i) = \text{mult}_{\lambda^+}(i)$ est pair et cette contribution vaut 1. En supprimant ces termes et en utilisant que $\text{mult}_{\lambda^+}(i)$ et $\text{mult}_{\lambda^-}(i)$ sont de même parité, on obtient

$$\prod_{i \in \text{Jord}_{\text{bp}}(\lambda^+) \cap \text{Jord}_{\text{bp}}(\lambda^-)} (\epsilon^+(i)\epsilon^-(i))^{\text{mult}_{\lambda^-}(i)} = -1.$$

On peut donc fixer $i \in \text{Jord}_{\text{bp}}(\lambda^+) \cap \text{Jord}_{\text{bp}}(\lambda^-)$ tel que $\epsilon^+(i)\epsilon^-(i) = -1$. D'après (89), on a $\tau^+(i) \neq \tau^-(i)$. Par construction de ces fonctions, cela entraîne qu'il existe un intervalle $\Delta_2 \in \text{Int}(\lambda_2)$ tel que $J(\{i\}) \subset J(\Delta_2)$. A fortiori, $\text{Int}(\lambda_2)$ est non vide, ce qui contredit notre hypothèse $n_2 = 0$.

Index des notations

$\mathbb{C}[X]$	1.1	$k(w)$	2.3	$\rho(\alpha, \beta)$	1.1
cup	1.8	$\kappa_{\pi,0}$	2.3	$\rho^D(\alpha, \beta)$	1.1
$c_{\mathcal{O}}(\pi)$	3.2	$l(\lambda)$	1.1	$\rho_{\lambda,\epsilon}$	1.3, 1.4, 1.5
d	1.2, 1.6, 1.7	$\text{mult}_{\lambda}(i), \text{mult}_{\lambda}(\geq i)$	1.1	ρ_2^+, ρ_2^-	1.12
Δ_{\min}	1.3, 1.4, 1.5	$m(\rho_1, \rho_2^{\xi}; \rho_{\lambda'}, \epsilon', \rho_{\lambda''}, \epsilon'')$	1.12	$S(\lambda)$	1.1
Δ_{\max}	1.4, 1.5	mult!_m	2.1	$S_k(\lambda)$	1.1
δ^+, δ^-	1.13	$\mu(\mathcal{O})$	3.1	\mathfrak{S}_N	1.1
$\delta(\lambda, s, \epsilon)$	3.3	$\mu(\pi)$	3.2	sgn	1.1
E	2.2	Nil_{\sharp}	3.1	sgn_{CD}	1.1
fam	1.3, 1.4, 1.5	\mathbf{Nil}_{\sharp}	3.1	$\mathcal{S}_{N,D}$	1.2
$f_{\text{Lie}}, f_{\text{red}}$	2.2	$\mathfrak{n}(\lambda^+, \lambda^-)$	3.8	symb	1.2
$\phi_{\alpha, \beta', \beta''}$	2.3	$\mathfrak{N}(\lambda^+, \lambda^-)$	3.8	$\text{sp}(\lambda)$	1.3, 1.4, 1.5
$f_{\mathcal{O}_1, \mathcal{O}_2}$	3.1	$\mathcal{O}_{\mathcal{O}_1, \mathcal{O}_2}$	3.1	$\text{sp}(\lambda, \epsilon)$	1.3, 1.4, 1.5
\mathcal{H}	2.3	$\mathcal{P}(N), \mathcal{P}_k(N)$	1.1	$\tau_{\lambda_1, \lambda_2}$	1.10
$h_{\mathcal{O}_1, \mathcal{O}_2}$	3.1	$\mathcal{P}^{\text{symp}}(2n)$	1.3	τ^+, τ^-	1.13
$\text{Int}(\lambda)$	1.3, 1.4, 1.5	$\mathcal{P}^{\text{symp}}(2n)$	1.3	Θ_{π}	2.1
\mathfrak{Int}_d	1.10	$\mathcal{P}^{\text{symp, sp}}(2n)$	1.3	$\Theta_{\pi, \text{cusp}}$	2.1
ind	1.8	$\mathcal{P}^{\text{orth}}(2n+1)$	1.4	$\Theta_{\pi, m, \text{cusp}}$	2.1
$\text{ind}(\lambda_1, \lambda_2)$	1.9	$\mathcal{P}^{\text{orth}}(2n+1)$	1.4	$\Theta_{\pi, \text{cusp}}^M$	2.1
$\text{Int}_{\lambda_1, \lambda_2}$	1.10	$\mathcal{P}^{\text{orth, sp}}(2n+1)$	1.4	$V(\pi)$	3.2
Jord(λ)	1.1	$\mathcal{P}^{\text{orth}}(2n)$	1.5	W_N	1.1
Jord _{bp} (λ)	1.3, 1.4, 1.5	$\mathcal{P}^{\text{orth}}(2n)$	1.5	W_N^D	1.1
Jord _{bp} ^k (λ)	1.3, 1.4, 1.5	$\mathcal{P}^{\text{orth, sp}}(2n)$	1.5	$W(m, N', N'')$	2.3
$J(\Delta)$	1.6, 1.7	$\underline{\mathcal{P}}^{\text{orth}}(2n)$	1.5	$W(m, N', N'')_{\text{ell}}$	2.3
$j_{\min}(\Delta), j_{\max}(\Delta)$	1.6, 1.7	$\underline{\mathcal{P}}^{\text{orth}}(2n)$	1.5	$W_{n_2, \text{iso}}, W_{n_2, \text{an}}$	3.6
$k_{\lambda, \epsilon}$	1.3, 1.4, 1.5	$\mathcal{P}^{\text{orth}}(n)$	1.8	ξ	1.9
$k(r', r''; w)$	2.3	$\mathcal{PP}^{\text{orth}}(n_1, n_2)$	3.6	$\zeta(\lambda)$	1.6, 1.7

Index des notations de [Waldspurger 2018]

$\mathbb{C}[X]$	1.4	D, D^{par}	1.7	\mathcal{F}^L	1.9	$\text{Irr}_{\text{tunip}}$	1.3
$C'_{n'}$	1.5	$\eta(Q)$	1.1	\mathcal{F}^{par}	1.9	$\mathfrak{Irr}_{\text{tunip}}$	1.3
$C''_{n'', \sharp}$	1.5	$\eta^-(Q), \eta^+(Q)$	1.1	\mathcal{F}	2.3	$\text{Irr}_{\text{unip-quad}}$	1.3
$C''_{n''}$	1.5	Ell_{unip}	1.4	$\mathfrak{F}^{\text{par}}$	2.3	$\mathfrak{Irr}_{\text{unip-quad}}$	1.3
$C^{\text{GL}(m)}$	1.5	$\mathfrak{E}\mathfrak{U}_{\text{unip}}$	1.4	G_{iso}	1.1	Jord(λ)	1.3
$\mathbb{C}[\widehat{W}_N]_{\text{cusp}}$	1.8	$\mathfrak{E}\text{ndo}_{\text{tunip}}$	2.1	G_{an}	1.1	Jord _{bp} (λ)	1.3
$D(n)$	1.2	$\mathfrak{E}\text{ndo}_{\text{unip-quad}}$	2.2	Γ	1.8	Jord _{bp} ^k (λ)	1.4
$D_{\text{iso}}(n)$	1.2	$\mathfrak{E}\text{ndo}_{\text{unip-quad}}^{\text{red}}$	2.2	$\mathbf{\Gamma}$	1.8	$K_{n', n''}^{\pm}$	1.2
$D_{\text{an}}(n)$	1.2	$\mathfrak{E}\text{ndo}_{\text{unip, disc}}$	2.4	$\widetilde{\text{GL}}(2n)$	2.1	k	1.9

Index des notations de [Waldspurger 2018] (reprise)

L^*	1.1	$Q_{\text{iso}}, Q_{\text{an}}$	1.1	sgn	1.8
$L_{n',n''}$	1.2	ρ_λ	1.3	sgn_{CD}	1.8
$l(\lambda)$	1.3	\mathcal{R}^{par}	1.5	\mathcal{S}_n	1.11
mult_λ	1.3	$\mathcal{R}^{\text{par, glob}}$	1.5	$\mathfrak{St}_{\text{unip}}$	2.1
\mathfrak{o}	1.1	$\mathcal{R}_{\text{cusp}}^{\text{par}}$	1.5	$\mathfrak{St}_{\text{unip-quad}}$	2.4
$O^+(Q), O^-(Q)$	1.1	$\mathcal{R}_m^{\text{par, glob}}$	1.5	$\mathfrak{St}_{\text{unip, disc}}$	2.4
ϖ	1.1	$\mathcal{R}_m^{\text{par, cusp}}$	1.5	sgn_{iso}	2.6
$\pi_{n',n''}$	1.3	res'_m	1.5	sgn_{an}	2.6
$\mathcal{P}(N)$	1.3	res''_m	1.5	val_F	1.1
$\mathcal{P}^{\text{symp}}(2N)$	1.3	res_m	1.5, 1.8	V_{iso}	1.1
$\mathcal{P}^{\text{symp}}(2N)$	1.3	res_m	1.5	V_{an}	1.1
$\pi(\lambda, s, \epsilon)$	1.3	\mathcal{R}	1.8	W_N, \widehat{W}_N	1.8
$\pi(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-)$	1.3	$\mathcal{R}(\gamma)$	1.8	w_α	1.8
$\pi_{\text{ell}}(\lambda^+, \epsilon^+, \lambda^-, \epsilon^-)$	1.4	$\mathcal{R}(\mathbf{y})$	1.8	$w_{\alpha, \beta}$	1.8
$\text{proj}_{\text{cusp}}$	1.5	$\mathcal{R}^{\text{glob}}$	1.8	$w_{\alpha, \beta', \beta''}$	1.8
$\mathcal{P}(\leq n)$	1.5	$\mathcal{R}_{\text{cusp}}$	1.8	$Z(\lambda)$	1.3
$\mathcal{P}_k(N)$	1.8	Rep	1.9	$Z(\lambda, s)$	1.3
$\Pi(\lambda, s, h)$	2.1	$\rho\iota$	1.10	$Z(\lambda, s)$	1.3
$\Pi^{\text{st}}(\lambda^+, \lambda^-)$	2.4	$S(\lambda)$	1.3	$Z(\lambda, s)^\vee$	1.3
$\mathcal{P}^{\text{symp, disc}}(2n)$	2.4	$\mathfrak{S}_N, \widehat{\mathfrak{S}}_N$	1.8	$ \cdot _F$	1.1

Remerciement

Je remercie vivement le rapporteur qui a lu l'article très soigneusement et a corrigé un nombre humiliant d'erreurs.

Bibliographie

- [Arthur 2013] J. Arthur, *The endoscopic classification of representations: orthogonal and symplectic groups*, Amer. Math. Soc. Colloquium Publ. **61**, Amer. Math. Soc., Providence, RI, 2013. MR Zbl
- [Barbasch et Vogan 1985] D. Barbasch et D. A. Vogan, Jr., "Unipotent representations of complex semisimple groups", *Ann. of Math. (2)* **121**:1 (1985), 41–110. MR Zbl
- [DeBacker 2002] S. DeBacker, "Homogeneity results for invariant distributions of a reductive p -adic group", *Ann. Sci. École Norm. Sup. (4)* **35**:3 (2002), 391–422. MR Zbl
- [Geck et Pfeiffer 2000] M. Geck et G. Pfeiffer, *Characters of finite Coxeter groups and Iwahori–Hecke algebras*, London Math. Soc. Monographs, New Ser. **21**, Oxford Univ. Press, 2000. MR Zbl
- [Lusztig 1990] G. Lusztig, "Green functions and character sheaves", *Ann. of Math. (2)* **131**:2 (1990), 355–408. MR Zbl
- [Lusztig 1992] G. Lusztig, "A unipotent support for irreducible representations", *Adv. Math.* **94**:2 (1992), 139–179. MR Zbl
- [Lusztig 1995] G. Lusztig, "Classification of unipotent representations of simple p -adic groups", *Int. Math. Res. Not.* **1995**:11 (1995), 517–589. MR Zbl

- [Mœglin 1996a] C. Mœglin, “Front d’onde des représentations des groupes classiques p -adiques”, *Amer. J. Math.* **118**:6 (1996), 1313–1346. MR Zbl
- [Mœglin 1996b] C. Mœglin, “Représentations quadratiques unipotentes des groupes classiques p -adiques”, *Duke Math. J.* **84**:2 (1996), 267–332. MR Zbl
- [Mœglin et Renard 2017] C. Mœglin et D. Renard, “Paquets d’Arthur des groupes classiques complexes”, pp. 203–256 dans *Around Langlands correspondences* (Orsay, 2015), édité par F. Brumley et al., *Contemp. Math.* **691**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl arXiv
- [Mœglin et Waldspurger 1987] C. Mœglin et J.-L. Waldspurger, “Modèles de Whittaker dégénérés pour des groupes p -adiques”, *Math. Z.* **196**:3 (1987), 427–452. MR Zbl
- [Mœglin et Waldspurger 2003] C. Mœglin et J.-L. Waldspurger, “Paquets stables de représentations tempérées et de réduction unipotente pour $SO(2n+1)$ ”, *Invent. Math.* **152**:3 (2003), 461–623. MR Zbl
- [Moy et Prasad 1996] A. Moy et G. Prasad, “Jacquet functors and unrefined minimal K -types”, *Comment. Math. Helv.* **71**:1 (1996), 98–121. MR Zbl
- [Waldspurger 2001] J.-L. Waldspurger, *Intégrales orbitales nilpotentes et endoscopie pour les groupes classiques non ramifiés*, Astérisque **269**, Soc. Math. France, Paris, 2001. MR Zbl
- [Waldspurger 2004] J.-L. Waldspurger, “Représentations de réduction unipotente pour $SO(2n+1)$: quelques conséquences d’un article de Lusztig”, pp. 803–910 dans *Contributions to automorphic forms, geometry, and number theory* (Baltimore, 2002), édité par H. Hida et al., Johns Hopkins Univ. Press, Baltimore, 2004. MR Zbl
- [Waldspurger 2016a] J.-L. Waldspurger, “Caractères de représentations de niveau 0”, preprint, 2016. arXiv
- [Waldspurger 2016b] J.-L. Waldspurger, “Représentations de réduction unipotente pour $SO(2n+1)$, II : Endoscopie”, preprint, 2016. arXiv
- [Waldspurger 2018] J.-L. Waldspurger, “Représentations de réduction unipotente pour $SO(2n+1)$, I: Une involution”, *J. Lie Theory* **28**:2 (2018), 381–426. MR

Communicated by Shou-Wu Zhang

Received 2017-02-17 Revised 2018-01-22 Accepted 2018-02-23

jean-loup.waldspurger@imj-prg.fr

CNRS IMJ-PRG, Paris, France

Correspondences without a core

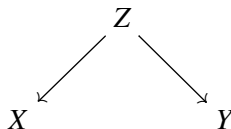
Raju Krishnamoorthy

We study the formal properties of correspondences of curves without a core, focusing on the case of étale correspondences. The motivating examples come from Hecke correspondences of Shimura curves. Given a correspondence without a core, we construct an infinite graph \mathcal{G}_{gen} together with a large group of “algebraic” automorphisms A . The graph \mathcal{G}_{gen} measures the “generic dynamics” of the correspondence. We construct specialization maps $\mathcal{G}_{\text{gen}} \rightarrow \mathcal{G}_{\text{phys}}$ to the “physical dynamics” of the correspondence. Motivated by the abstract structure of the supersingular locus, we also prove results on the number of bounded étale orbits, in particular generalizing a recent theorem of Hallouin and Perret. We use a variety of techniques: Galois theory, the theory of groups acting on infinite graphs, and finite group schemes.

1. Introduction	1173
2. Conventions, notation, and terminology	1177
3. Correspondences and cores	1178
4. A recursive tower	1184
5. The generic graph of a correspondence	1190
6. Symmetric correspondences	1196
7. Specialization of graphs and special orbits	1198
8. Invariant line bundles and invariant sections	1201
9. Clumps	1209
Acknowledgments	1213
References	1213

1. Introduction

To an étale correspondence of complex hyperbolic curves



one can associate a many-valued function $X \dashrightarrow X$. Mochizuki [1998] proved that if this many-valued function has unbounded dynamics, then X , Y , and Z are all complex Shimura curves. Mochizuki uses a highly nontrivial result of Margulis [1991], which characterizes complex Shimura curves via properties of discrete subgroups of $\text{PSL}_2(\mathbb{R})$.

MSC2010: primary 14G35; secondary 05C25, 14H05, 37P55.

Keywords: Shimura curves, special points, correspondences, dynamics.

The most basic examples of Shimura varieties are the modular curves, parametrizing elliptic curves with level structure. A slightly less familiar example comes from moduli spaces of *fake elliptic curves*; these Shimura varieties are projective algebraic curves. It turns out that the modular curves are the only noncompact Shimura curves. See Deligne [1979] for a general introduction to Shimura varieties.

In general, Shimura varieties are quasiprojective algebraic varieties defined over $\overline{\mathbb{Q}}$ [Borovoi 1982; Deligne 1979; Milne 1983; Milne and Shih 1982]. Recent work of Kisin [2010] shows that many Shimura varieties of abelian type have natural integral models, which opens up the possibility of studying their reduction modulo p . PEL-type Shimura varieties are moduli spaces of abelian varieties with manifestly algebraic conditions (i.e., fixing the data of a polarization, endomorphisms, and level). Using the moduli interpretation it is straightforward to define PEL-type Shimura varieties directly over finite fields \mathbb{F}_q , at least for most q . As far as we know, there is not as of yet a direct definition of general non-PEL-type Shimura varieties over \mathbb{F}_q .

Jie Xia has recently taken the simplest example of non-PEL-type Shimura curves, what he calls Mumford curves, and given “direct definitions” over $\overline{\mathbb{F}}_p$ [Xia 2013a; 2013c; 2014]. The most basic example of Mumford curves parametrize abelian 4-folds with certain extra Hodge classes, as in Mumford’s original paper [1969]. Xia proved theorems of the following form: given an abelian scheme $\mathcal{A} \rightarrow X$ over a curve $X/\overline{\mathbb{F}}_p$, there are certain conditions that ensure that the pair (\mathcal{A}, X) is the reduction of a Mumford curve together with its universal abelian scheme over $W(\overline{\mathbb{F}}_p)$.

In Chapter 2 of [Krishnamoorthy 2016], we posed the question of characterizing Shimura curves over \mathbb{F}_q . Unlike Xia, we *did not* assume the existence of an abelian scheme \mathcal{A} over the curves considered. Instead, we took as our starting point Mochizuki’s theorem, which is of group theoretic nature.

Definition. Let $X \leftarrow Z \rightarrow Y$ be a correspondence of curves over k . Then we have an inclusion diagram $k(X) \subset k(Z) \supset k(Y)$ of function fields. We say that the correspondence *has no core* if $k(X) \cap k(Y)$ has transcendence degree 0 over k .

This definition formalizes the phrase “generically unbounded dynamics” (Remark 3.7 and Proposition 5.10). Shimura curves have many étale correspondences without a core. Inspired by Mochizuki’s theorem, we wondered if all étale correspondences of curves without a core are “related to” Shimura curves. Given a smooth projective curve X over \mathbb{F}_q , are there other group theoretic conditions on $\pi_1^{\text{ét}}(X)$ that ensure that X is the reduction modulo p of a classical Shimura curve?

In this article, we explore the formal structure of (étale) correspondences without a core with an aim to understanding the similarities with Hecke correspondences of Shimura curves. We now state the main constructions and results.

Given a correspondence without a core, in Section 5 we construct a pair $(\mathcal{G}_{\text{gen}}, A)$ of an infinite graph together with a large topological group of “algebraic” automorphisms. The graph \mathcal{G}_{gen} roughly measures the “generic dynamics.” In the case of a symmetric l -adic Hecke correspondence of modular curves, \mathcal{G}_{gen} is a tree and the pair $(\mathcal{G}_{\text{gen}}, A)$ is related to the action of $\text{PSL}_2(\mathbb{Q}_l)$ on its building. When \mathcal{G}_{gen} is a tree, we prove that the vertices of \mathcal{G}_{gen} are in bijective correspondence with the maximal open compact subgroups

of a certain subgroup A^{PQ} of A (Corollary 5.21). This perhaps suggests that in this case the action of the topological group A^{PQ} on \mathcal{G}_{gen} is similar to the action of the l -adic linear group $\text{PSL}_2(\mathbb{Q}_l)$ on its building.

Definition. Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of curves over k . A *clump* is a finite set $S \subset Z(\bar{k})$ such that $f^{-1}(f(S)) = g^{-1}(g(S)) = S$. A clump is *étale* if f and g are étale at all points of S .

A clump may be thought of as a “bounded orbit of geometric points.” Hecke correspondences of modular curves over \mathbb{F}_p have a natural étale clump: the supersingular locus.

Theorem 9.6. *Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of curves over a field k without a core. There is at most one étale clump.*

An example: let $l \neq p$. Applying Theorem 9.6 to the Hecke correspondence $Y(1) \leftarrow Y_0(l) \rightarrow Y(1)$ reproves the fact that any two supersingular elliptic curves over $\bar{\mathbb{F}}_p$ are related by an l -primary isogeny. Theorem 9.6 implies a generalization of a theorem of Hallouin and Perret [2014], who came upon it in the analysis of optimal towers in the sense of Drinfeld–Vladut. They use spectral graph theory and an analysis of the singularities of a certain recursive tower. In our language, the hypotheses of their “one clump theorem” are:

- $k \cong \mathbb{F}_q$.
- $X \leftarrow \Gamma \rightarrow X$ is a minimal self-correspondence of type (d, d) .
- \mathcal{H}_{gen} , a certain directed graph where the in- and out-degree of every vertex is d , has no directed cycles.

Our techniques allow one to relax the third condition to “ \mathcal{H}_{gen} is infinite”; in particular, \mathcal{H}_{gen} may have directed cycles. Moreover, our proof works over any field k and is purely algebro-geometric. See the lengthy Remark 9.8 for a full translation/derivation.

We remark that Theorem 9.6 is characteristic-independent and hence applies to correspondences without a core over \mathbb{C} ; in particular, it may be thought of as a result on the dynamics of complex algebraic curves.

Specializing to characteristic 0, we prove the following.

Corollary 9.2. *Let $X \leftarrow Z \rightarrow Y$ be an étale correspondence of projective curves over a field k without a core. Suppose $\text{char}(k) = 0$. Then there are no clumps.*

We unfurl this statement. Think of a symmetric Hecke correspondence $X \leftarrow Z \rightarrow X$ of Shimura curves over \mathbb{C} as a many-valued function from X to X . Then the iterated orbit of any point $x \in X$ under this many-valued function is unbounded. This was likely already known, but we could not find it in the literature. We nonetheless believe our approach is new. We now briefly describe the sections.

In Section 3 we state Mochizuki’s theorem (Theorem 3.10). We then reprise the theme: “are all étale correspondences without a core related to Hecke correspondences of Shimura varieties?” in Question 3.16. The phrase “related to” is absolutely vital, and étale correspondences without a core do not always directly deform to characteristic 0. We will see one example in Remark 3.18 via Igusa level structure. More exotic is Example 3.19 of a central leaf in a Hilbert modular variety; according to a general philosophy of Chai–Oort, these leaves should also be considered Shimura varieties in characteristic p . Unlike in

characteristic 0, however, these may deform in families purely in characteristic p . There are also examples of étale correspondences of curves over $\overline{\mathbb{F}}_p$ without a core using Drinfeld modular curves. We pose a concrete instantiation of Question 3.16 that doesn't mention Shimura varieties at all (Question 3.21).

In Section 4, starting from a correspondence without a core, we use elementary Galois theory to construct an infinite tower of curves W_∞ with “function field” E_∞ . We use this tower to prove that the property of “not having a core” for an étale correspondences specializes in families (Lemma 4.10). As a consequence, there are no global deformations of an étale correspondence of projective hyperbolic curves over \mathbb{C} without a core (Corollary 4.13).

In Section 5, given a correspondence without a core, we construct the pair $(\mathcal{G}_{\text{gen}}, A)$ of an infinite graph together with a large group of “algebraic” automorphisms. The graph \mathcal{G}_{gen} packages the Galois theory of E_∞ and reflects the generic dynamics of the correspondence. We are especially interested in Question 5.17: given an étale correspondence without a core, is \mathcal{G}_{gen} a tree? Using Serre's theory [1977] of groups acting on trees, we prove that in this case the action of A^{PQ} on \mathcal{G}_{gen} shares several properties with the action of the l -adic linear group $\text{PSL}_2(\mathbb{Q}_l)$ on its building (see Proposition 5.19 and Corollary 5.21).

In Section 6 we develop some basic results for symmetric correspondences. We are interested in the following refinement of Question 5.17: given a symmetric étale correspondence without a core, is the pair $(\mathcal{G}_{\text{gen}}, A)$ ∞ -transitive (Question 6.11)? In the case of a symmetric, type (3, 3) correspondence without a core, we are able to verify this using graph theory due to Tutte (Lemma 6.10); in particular, in this case \mathcal{G}_{gen} is a tree.

In Section 7 we construct specialization maps $\mathcal{G}_{\text{gen}} \rightarrow \mathcal{G}_{\text{phys}}$. These roughly specialize the dynamics from the generic point to closed points. When the original correspondence is étale, the maps $\mathcal{G}_{\text{gen}} \rightarrow \mathcal{G}_{\text{phys}}$ are covering spaces of graphs (Lemma 7.2). Motivated by work of Kohel [1996] and Sutherland [2013] on *isogeny volcanoes* of elliptic curves, we speculate on the behavior and asymptotics of these specialization maps in Question 7.5.

The rest of the paper may be read independently. In Section 8 we introduce the notion of an *invariant line bundle* on a correspondence and prove several results about their spaces of sections on (étale) correspondences without a core. In characteristic 0 there are no *invariant pluricanonical differential forms* (Corollary 8.13) on an étale correspondence of projective hyperbolic curves without a core. In characteristic p , however, such forms may exist. The existence of the *Hasse invariant*, a $\text{mod } p$ modular form, is a representative example of the difference. The key to these results is the introduction of the group scheme $\text{Pic}^0(X \leftarrow Z \rightarrow Y)$; when $X \leftarrow Z \rightarrow Y$ does not have a core, we prove that this group scheme is finite (Lemma 8.9). We speculate on the relationship between *invariant differential forms* and $\text{Pic}^0(X \leftarrow Z \rightarrow Y)$ in Question 8.18.

In Section 9 we show that an *étale clump* gives rise to an invariant line bundle together with a line of invariant sections. Using the analysis in Section 8, we prove the two sample theorems above and explicate the relationship between our result and that of Hallouin and Perret. In analogy to the supersingular locus, we wonder if every étale correspondence of projective curves without a core in characteristic p has a clump, equivalently an invariant pluricanonical differential form (Question 9.7).

We briefly comment on Chapter 3 of [Krishnamoorthy 2016] (see also [Krishnamoorthy 2017]). Let $(X \xleftarrow{f} Z \xrightarrow{g} X)$ be a symmetric type $(3, 3)$ étale correspondence without a core over a finite field \mathbb{F}_q . Inspired by the formal similarity between the pair $(\mathcal{G}_{\text{gen}}, A^{PQ})$ and $(\mathcal{T}, \text{PSL}_2(\mathbb{Q}_2))$, where \mathcal{T} is the building of $\text{PGL}_2(\mathbb{Q}_2)$ (i.e., the infinite trivalent tree), we assume that the action of G_P on \mathcal{G}_{gen} is isomorphic to the action of $\text{PSL}_2(\mathbb{Z}_2)$ on \mathcal{T} , a *purely group-theoretic condition*. Call the associated \mathbb{Q}_2 local system \mathcal{L} . Then, using 2-adic group theory we prove that $f^*\mathcal{L} \cong g^*\mathcal{L}$. Suppose further that all Frobenius traces of \mathcal{L} are in \mathbb{Q} . Using a recent breakthrough in the p -adic Langlands correspondence for curves over a finite field due to Abe [2013], we build the following correspondence.

Theorem. *Let C be a smooth, geometrically irreducible, complete curve over \mathbb{F}_q . Suppose q is a square. There is a natural bijection between the following two sets:*

$$\left\{ \begin{array}{l} \overline{\mathbb{Q}}_l\text{-local systems } \mathcal{L} \text{ on } C \text{ such that :} \\ \bullet \mathcal{L} \text{ is irreducible of rank 2.} \\ \bullet \mathcal{L} \text{ has trivial determinant.} \\ \bullet \text{The Frobenius traces are in } \mathbb{Q}. \\ \bullet \mathcal{L} \text{ has infinite image, up to isomorphism.} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{l} p\text{-divisible groups } \mathcal{G} \text{ on } C \text{ such that :} \\ \bullet \mathcal{G} \text{ has height 2 and dimension 1.} \\ \bullet \mathcal{G} \text{ is generically versally deformed.} \\ \bullet \mathcal{G} \text{ has all Frobenius traces in } \mathbb{Q}. \\ \bullet \mathcal{G} \text{ has ordinary and supersingular points,} \\ \text{up to isomorphism.} \end{array} \right\}$$

such that if \mathcal{L} corresponds to \mathcal{G} , then $\mathcal{L} \otimes \mathbb{Q}_l(-\frac{1}{2})$ is compatible with the F -isocrystal $\mathbb{D}(\mathcal{G}) \otimes \mathbb{Q}$.

If \mathcal{G} is everywhere versally deformed on X , Xia [2013b] shows that the pair (X, \mathcal{G}) may be canonically lifted to characteristic 0. In this case the whole correspondence is the reduction modulo p of an étale correspondence of Shimura curves. However, examples coming from Shimura curves with Igusa level structure show that \mathcal{G} may be generically versally deformed without being everywhere versally deformed. For more details, see [Krishnamoorthy 2016; 2017].

2. Conventions, notation, and terminology

We explicitly state conventions and notations. These are in full force unless otherwise stated:

- (1) p is a prime number and q is a power of p .
- (2) \mathbb{F} is a fixed algebraic closure of \mathbb{F}_p .
- (3) A *curve* C over a field k is a quasiprojective geometrically integral scheme of dimension 1 over k . Unless otherwise explicitly stated, we assume $C \rightarrow \text{Spec}(k)$ is smooth.
- (4) A *morphism of curves* $X \rightarrow Y$ over k is a morphism of k -schemes that is nonconstant, finite, and generically separable.
- (5) A smooth curve C over a field k is said to be *hyperbolic* if $\text{Aut}_{\bar{k}}(C_{\bar{k}})$ is finite.
- (6) A *complex Shimura Curve* is a finite étale cover of a one-dimensional complex Shimura variety. (Other authors call such curves *arithmetic curves* or *Shimura-arithmetic curves*.)

- (7) Given a field k , Ω will always be an algebraically closed field of transcendence degree 1 over k .
- (8) In general, X, Y , and Z will be a curves over k , with $M = k(Z)$, $L = k(X)$, and $K = k(Y)$ the function fields. We fix a k -algebra embedding $PQ : k(Z) \hookrightarrow \Omega$ that identifies Ω as an algebraic closure of $k(Z)$.

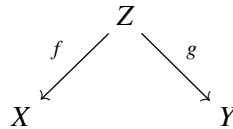
3. Correspondences and cores

Definition 3.1. A smooth curve X over a field k is said to be hyperbolic if $\text{Aut}_{\bar{k}}(X_{\bar{k}})$ is finite.

This is equivalent to the usual criterion of $2g - 2 + r \geq 1$ where g is the geometric genus of the compactification \bar{X} and r is the number of geometric punctures. Over the complex numbers, this is equivalent to X being uniformized by the upper half plane \mathbb{H} .

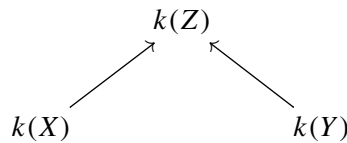
Lemma 3.2. If $X \rightarrow Y$ is a nonconstant morphism of curves over k where Y is hyperbolic, then X is hyperbolic.

Definition 3.3. A correspondence of curves over k is a diagram



of smooth curves over a field k where f and g are finite, generically separable morphisms. We call such a correspondence of type (d, e) if $\deg f = d$ and $\deg g = e$. We call such a correspondence étale if both maps are étale. We call a correspondence minimal if the associated map $Z \rightarrow X \times Y$ is birational onto its image.

To a correspondence we can associate a containment diagram of function fields:



A correspondence is minimal if and only if there is no proper subfield of $k(Z)$ that contains both $k(X)$ and $k(Y)$.

Remark 3.4. Note that we require both f and g to be finite; for instance, strict open immersions are not permitted.

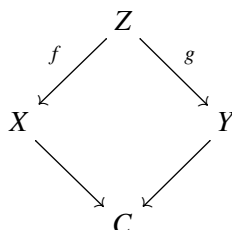
Definition 3.5. We say a correspondence $X \leftarrow Z \rightarrow Y$ of curves over k has a core if the intersection of the two function fields $k(X) \cap k(Y)$ has transcendence degree 1 over k .

Remark 3.6. If a correspondence has a core, then $k(X) \cap k(Y) \subset k(Z)$ is a finite separable field extension. Indeed, suppose it weren't. The morphisms f and g are generically separable. Then at least one of the extensions $k(X) \cap k(Y) \subset k(X)$ or $k(X) \cap k(Y) \subset k(Y)$ is inseparable. Suppose $k(X) \cap k(Y) \subset k(Y)$ is

not separable. Then there exists an element $\lambda \in k(X)$ such that $\lambda \notin k(Y)$ but $\lambda^p \in k(Y)$. But $\lambda \in k(Z)$, so g is not separable, contrary to our original assumption

Suppose $X \leftarrow Z \rightarrow Y$ is a correspondence of curves over k with a core. If X , Y , and Z are projective, we call the smooth projective curve C associated to the field $k(X) \cap k(Y)$ (considered as a field of transcendence degree 1 over k) the *coarse core* of the correspondence if it exists. One may also define the coarse core if X , Y , and Z are affine, see Remark 4.4.

In particular, a correspondence of curves over k has a core if and only if there exists a curve C over k with finite, generically separable maps from X and Y such that the following diagram commutes:



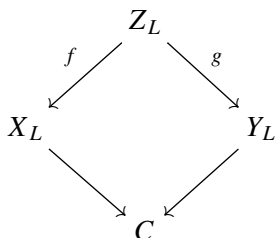
Remark 3.7. Given a correspondence as above, consider the following “many-valued function” $X \dashrightarrow X$ that sends $x \in X$ to the multiset $f(g^{-1}(g(f^{-1}(x))))$, i.e., start with x , take all preimages under f , take the image under g , take all preimages under g and take the image under f . Having a core guarantees that the dynamics of this many-valued function are *uniformly bounded* in the following sense: there exists a positive integer D such that, starting with any point x and iteratively applying the above many-valued function, the size of the image set is no greater than D . In other words, the “orbit” of x under the many-valued function is finite and has size $\leq D$.

We had initially written the following proposition in the case when L/k was an algebraic field extension. The referee explained that this restriction was unnecessary.

Proposition 3.8. *Let $X \leftarrow Z \rightarrow Y$ be a correspondence of curves over k . Let L be a field extension of k and $X_L \leftarrow Z_L \rightarrow Y_L$ the base-changed correspondence of curves over L . Then $X \leftarrow Z \rightarrow Y$ has a core if and only if $X_L \leftarrow Z_L \rightarrow Y_L$ has a core.*

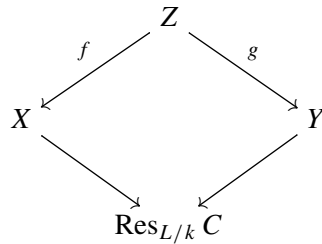
Proof. We may assume X , Y , and Z are projective. If $X \leftarrow Z \rightarrow Y$ has a core, then so does $X_L \leftarrow Z_L \rightarrow Y_L$, so it remains to prove the reverse implication.

First, if $X_L \leftarrow Z_L \rightarrow Y_L$ has a core, then a core exists after a finitely generated field extension, i.e., we may assume that L/k is finitely generated. Call the coarse core C . Now we may spread out the diagram



to a commutative square of projective curves over a finite-type k -scheme U . There is a finite extension k'/k such that $U(k')$ is nonempty. Specializing to such a point $u \in U(k')$, we see that the correspondence $X_{k'} \leftarrow Z_{k'} \rightarrow Y_{k'}$ of curves over k' has a core. In particular, we reduce to the case of L/k being a finite extension.

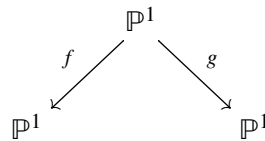
The Weil restriction of scalars $\text{Res}_{L/k} C$ is a smooth, geometrically connected scheme over k by e.g., [Conrad et al. 2010, A.5.9]. The universal property of $\text{Res}_{L/k} C$ yields the following commutative diagram with nonconstant morphisms:



Taking the image of X and Y inside of $\text{Res}_{L/k} C$ allows us to conclude that $X \leftarrow Z \rightarrow Y$ had a core. \square

Remark 3.9. In our conventions, a curve C over a field k is geometrically integral. Therefore k is algebraically closed inside of $k(C)$. If $X \leftarrow Z \rightarrow Y$ is a correspondence of curves over k without a core, then the natural map $k \hookrightarrow k(X) \cap k(Y)$ is therefore an isomorphism, as $k(X) \cap k(Y)$ is a finite extension of k that is contained in $k(X)$.

For most of this article, we focus on correspondences without a core. General correspondences of curves will not have cores. Consider, for instance a correspondence of the form:



For general f and g , the dynamics of the induced many-valued-function $\mathbb{P}^1 \dashrightarrow \mathbb{P}^1$ will be unbounded and hence it will not have a core by Remark 3.7. When we restrict to étale correspondences without cores, there is the following remarkable theorem of Mochizuki (due in large part to Margulis [1991]), which is the starting point of this article.

Theorem 3.10 [Mochizuki 1998]. *If $X \leftarrow Z \rightarrow Y$ is an étale correspondence of hyperbolic curves without a core over a field k of characteristic 0, then $X, Y,$ and Z are all Shimura (a.k.a. Shimura-arithmetic) curves (see [loc. cit., Definitions 2.2 and 2.3]).*

Remark 3.11. Theorem 3.10 in particular implies that if $X \leftarrow Z \rightarrow Y$ is an étale correspondence of complex hyperbolic curves without a core, then all of the curves and maps can be defined over $\overline{\mathbb{Q}}$: all Shimura curves are defined over $\overline{\mathbb{Q}}$ as in Remark 3.14, and the set of nonconstant algebraic morphisms between two complex hyperbolic curves is finite (see also Theorem 4.1 of [Mochizuki 1998]). This will imply there are *no nontrivial (global) deformations* of étale correspondences of projective hyperbolic

curves without a core over \mathbb{C} , see Corollary 4.13. This fails in characteristic p ; we will see examples, explained by Ching-Li Chai, later in Example 3.19.

The proof of Theorem 3.10 comes down to a reduction to $k \cong \mathbb{C}$ by the Lefschetz principle and an explicit description, due to Margulis [1991], of the arithmetic subgroups Γ of $\mathrm{SL}(2, \mathbb{R})$. Given a complex hyperbolic curve C , fix a uniformization $\mathbb{H} \rightarrow C$ to obtain an embedding

$$\Gamma := \pi_1(C) \rightarrow \mathrm{SL}(2, \mathbb{R}).$$

We say $\gamma \in \mathrm{SL}(2, \mathbb{R})$ commensurates Γ if the discrete group $\gamma\Gamma\gamma^{-1}$ is commensurable with Γ , i.e., their intersection is of finite index in both groups. Define $\mathrm{Comm}(\Gamma)$ to be the subgroup commensurating Γ in $\mathrm{SL}(2, \mathbb{R})$ and note that $\Gamma \subset \mathrm{Comm}(\Gamma)$. Margulis has proved that Γ is arithmetic if and only if $[\mathrm{Comm}(\Gamma) : \Gamma] = \infty$, see e.g., Theorem 2.5 of [Mochizuki 1998].

Example 3.12. The commensurator of $\mathrm{SL}(2, \mathbb{Z})$ in $\mathrm{SL}(2, \mathbb{R})$ is $\mathrm{SL}(2, \mathbb{Q})$. The modular curve $Y(1) = [\mathbb{H}/\mathrm{SL}(2, \mathbb{Z})]$ is arithmetic.

Exercise 3.13. The correspondence of nonprojective stacky modular curves $Y(1) \leftarrow Y_0(2) \rightarrow Y(1)$ does not have a core. Here, $Y_0(2)$ is the moduli space of pairs of elliptic curves $(E_1 \xrightarrow{2:1} E_2)$ with a given degree-2 isogeny between them, and the maps send the isogeny to the source and target elliptic curve respectively. Hint: to prove this over the complex numbers, look at the “orbits” of $\tau \in \mathbb{H}$ as in Remark 3.7.

In our conventions, we declared a (*complex*) Shimura curve to be a finite étale cover of a one-dimensional complex Shimura variety. We briefly comment on this.

Remark 3.14. Mochizuki [1998, Definitions 2.2 and 2.3] defines two notions of arithmetic hyperbolic Riemann surface: Margulis arithmeticity and Shimura arithmeticity. Margulis arithmeticity is closer in spirit to the classical definition of a Shimura variety, while Shimura arithmeticity is essentially given by the data of a totally real field F and a quaternion algebra D over F that is split at exactly one of the infinite places. Proposition 2.4 of [loc. cit.] then proves these two definitions are equivalent. If X is an arithmetic curve and $Y \rightarrow X$ is a finite étale cover, then Y is manifestly arithmetic by either definition. In particular, the hyperbolic Riemann surfaces associated to *noncongruence subgroups* of $\mathrm{SL}_2(\mathbb{Z})$ are arithmetic by definition. What we call a complex Shimura curve is precisely what Mochizuki calls an arithmetic curve.

The algebraic curve $\mathbb{P}_{\mathbb{C}}^1 \setminus \{0, 1, \infty\}$ is a degree 2 étale cover of the stack $Y(2) = [\mathbb{H}/\Gamma(2)]$ and is hence a complex Shimura curve under our conventions. Therefore by Belyi’s theorem, for any curve $X/\overline{\mathbb{Q}}$ and any embedding $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$, the complex curve $X_{\mathbb{C}}$ is birational to a complex Shimura curve. Conversely, any complex Shimura curve may be defined over $\overline{\mathbb{Q}}$: every Shimura variety is defined over $\overline{\mathbb{Q}}$ [Borovoi 1982; Deligne 1979; Milne 1983; Milne and Shih 1982] and a finite étale cover of a $\overline{\mathbb{Q}}$ -variety is again defined over $\overline{\mathbb{Q}}$.

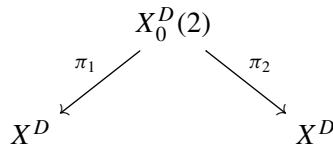
Definition 3.15. Let D be an indefinite nonsplit quaternion algebra over \mathbb{Q} of discriminant d and let \mathcal{O}_D be a fixed maximal order. Let k be a field whose characteristic is prime to d . A *fake elliptic curve with*

multiplication by \mathcal{O}_D is a pair (A, i) of an abelian surface A over k and an injective ring homomorphism $i : \mathcal{O}_D \rightarrow \text{End}_k(A)$.

Pick $t \in \mathcal{O}_D$ such that $t^2 = -d$ and denote by ι the canonical involution on D . The fake elliptic curve (A, i) is endowed with the unique principal polarization such that the Rosati involution induces the following involution on D :

$$x \rightarrow t^{-1}x't.$$

Just as one can construct a modular curve parametrizing elliptic curves, there is a Shimura curve X^D parametrizing fake elliptic curves with multiplication by \mathcal{O}_D . Over the complex numbers, these are compact hyperbolic curves. Explicitly, if one chooses an isomorphism $D \otimes \mathbb{R} \cong M_{2 \times 2}(\mathbb{R})$, look at the image of $\Gamma = \mathcal{O}_D^1$ of elements of \mathcal{O}_D^* of norm 1 (for the standard norm on \mathcal{O}_D) inside of $\text{SL}(2, \mathbb{R})$. This is a discrete subgroup and in fact acts properly discontinuously and cocompactly on \mathbb{H} . The quotient $[\mathbb{H}/\Gamma]$ is the Shimura curve associated to \mathcal{O}_D . There is a notion of isogeny of fake elliptic curves which is required to be compatible with the \mathcal{O}_D structure and the associated “fake degree” of an isogeny. See Buzzard [1997] or Boutot and Carayol [1991] for more details. These definitions allow us to define Hecke correspondences as in the elliptic modular case. For instance, as long as 2 splits in D , one can define the correspondence:



where $X_0^D(2)$ parametrizes pairs of fake elliptic curves $(A_1 \rightarrow A_2)$ with a given isogeny of fake degree 2 between them and π_1 and π_2 are the projections to the source and target respectively. This is an example of an étale correspondence of (stacky) hyperbolic curves without a core. To get an example without orbifold points, one can add auxiliary level structure by picking an open compact subgroup $K \subset \mathbb{A}^f$ of the finite adeles. This correspondence is in fact defined over $\mathbb{Z}[\frac{1}{2S}]$ for an integer S and so may be reduced modulo p for almost all primes.

Motivated by these examples, the orienting question of this article is to explore characteristic p analogs of Mochizuki’s theorem. More specifically, we wish to understand the abstract structure of étale correspondences of hyperbolic curves without a core.

Question 3.16. Let k be a field of characteristic p . If $X \leftarrow Z \rightarrow Y$ is an étale correspondence of hyperbolic curves over k without a core, then is it related to a Hecke correspondence of Shimura varieties or Drinfeld modular varieties?

Remark 3.17. In Corollary 4.12 we will in some sense reduce Question 3.16 to the analogous question with $k = \mathbb{F}$.

The clause “is related to” in Question 3.16 is absolutely vital as we will see in the following examples. Nonetheless we take Question 3.16 as a guiding principle.

Remark 3.18. There are examples of étale correspondence of hyperbolic curves without a core that should not deform to characteristic 0. Consider, for instance, the Hecke correspondence

$$\begin{array}{ccc} & Y_0(2) & \\ \pi_1 \swarrow & & \searrow \pi_2 \\ Y(1) & & Y(1) \end{array}$$

of modular curves over \mathbb{F}_p , with $p \neq 2$. By definition, there is a universal elliptic curve $\mathcal{E} \rightarrow Y(1)$. Let $\mathcal{G} = \mathcal{E}[p^\infty]$ be the associated p -divisible group over $Y(1)$. Note that $\pi_1^* \mathcal{G} \cong \pi_2^* \mathcal{G}$. Let X be the cover of $Y(1)$ that trivializes the finite flat group scheme $\mathcal{G}[p]_{\text{ét}}$ away from the supersingular locus of $Y(1)$. X is branched exactly at the supersingular points. Let Z be the analogous cover of $Y_0(2)$. Then we have an étale correspondence

$$\begin{array}{ccc} & Z & \\ \swarrow & & \searrow \\ X & & X \end{array}$$

which does not have a core (the dynamics of an ordinary point are unbounded) and morally one does not expect this correspondence to lift to characteristic 0. This construction is referred to as adding *Igusa level structure* in the literature: [Ulmer 1990] is a particularly lucid account of this story for modular curves. See Definition 4.8 of [Buzzard 1997] for the analogous construction for Shimura curves parametrizing fake elliptic curves. We take up the example of Igusa curves once again in Example 8.5, from the perspective of the Hasse invariant and the cyclic cover trick.

Modular curves with Igusa level structure still parametrize elliptic curves with some (purely characteristic p) level structure. Ching-Li Chai has provided the following more exotic example which shows that étale correspondence of hyperbolic curves over a field of characteristic p may deform purely in characteristic p .

Example 3.19. Let F be a totally real cubic field and let p be an inert prime. Consider the Hilbert modular threefold \mathcal{X}^F associated to \mathcal{O}_F ; \mathcal{X}^F parametrizes abelian threefolds with multiplication by \mathcal{O}_F . Let X be the reduction of \mathcal{X}^F modulo p and \mathcal{A} be the universal abelian scheme over X . Oort [2004] has constructed a foliation on such Shimura varieties; a leaf of this foliation has the property that the p -divisible group $\mathcal{A}[p^\infty]$ is geometrically constant on the leaf; i.e., if x and y are two geometric points of the leaf, then $\mathcal{A}[p^\infty]_x \cong \mathcal{A}[p^\infty]_y$. We list the possible slopes of a height 6, dimension 3, symmetric p -divisible group:

- (1) $(0, 0, 0, 1, 1, 1)$.
- (2) $(0, 0, \frac{1}{2}, \frac{1}{2}, 1, 1)$.
- (3) $(0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 1)$.
- (4) $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3})$.
- (5) $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

The only slope types that could possibly admit multiplication (up to isogeny) by \mathbb{Q}_{p^3} are 1, 4, and 5 by considerations on the endomorphism algebra. By de Jong–Oort purity, the locus where the slope type $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3})$ occurs inside of X is codimension 1. One can prove that a central leaf with this Newton polygon is a curve. Central leaves of Hilbert modular varieties have the property that they are preserved under l -adic Hecke correspondences and that in fact the l -adic monodromy is as large as possible [Chai 2005]. In particular, a central leaf has many Hecke correspondences. Moreover, as this Newton polygon stratum has dimension 2, the isogeny foliation is one-dimensional and so this Hecke correspondence deforms in a one-parameter family, purely in characteristic p .

We further remark that, in general, there are central leaves of Shimura varieties that are not defined over $\overline{\mathbb{F}}_p$; in particular there are étale correspondences without a core over fields of characteristic p that *do not descend* to the algebraic closure of a prime field, unlike the case of characteristic 0.

Remark 3.20. Chai and Oort have discussed the possibility that central leaves should be considered Shimura varieties in characteristic p . In particular, one could consider the example of a Hecke correspondence of a central leaf to be a Hecke correspondence of Shimura curves. In any case, both the examples of a Hecke correspondence of modular curves with Igusa level structure and a Hecke correspondence of a central leaf of dimension 1 map finitely onto a Hecke correspondence of Shimura varieties which deform to characteristic 0.

Question 3.16 is phrased in [Krishnamoorthy 2016] using only Shimura varieties. Ambrus Pál has informed us that there are examples of étale correspondences of Drinfeld modular curves (i.e., moduli spaces of \mathcal{D} -elliptic modules) over $\overline{\mathbb{F}}(t)$ without a core. All three of these examples have moduli interpretations. Moreover, they all have *many* Hecke correspondences. This motivates the following variant of Question 3.16, which does not mention Shimura/Drinfeld modular varieties at all.

Question 3.21. Let $X \leftarrow Z \rightarrow Y$ be an étale correspondence of hyperbolic curves over k without a core. Do there exist infinitely many minimal étale correspondences between X and Y without a core?

4. A recursive tower

In this section we associate an infinite tower of curves to a correspondence without a core. Properties of this tower will allow us to deduce that “being an étale correspondence without a core” specializes in families (Lemma 4.10); Corollary 4.13 then shows that there are no (global) deformations of an étale correspondence of complex projective hyperbolic curves without a core. We also find a simple criterion for an étale correspondence of projective hyperbolic curves over \mathbb{F} without a core to consist of the reductions modulo p of Shimura curves: if the correspondence lifts to $W(\mathbb{F})$ (Corollary 4.15).

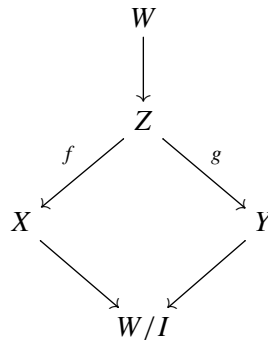
Definition 4.1. Let $f : X \rightarrow Y$ be a finite, nonconstant, generically separable map of curves over a field k . We say f is *finite Galois* if $|\mathrm{Aut}_Y(X)| = \deg(f)$. We say it is *geometrically finite Galois* if $f_{\bar{k}} : X_{\bar{k}} \rightarrow Y_{\bar{k}}$ is Galois.

It is well-known that given a finite, generically separable map of curves over a field k , we may take a Galois closure. In the projective case, this is “equivalent” to taking a Galois closure of the associated extension of function fields, and the affine case follows by the operation of “taking integral closure of the coordinate ring in the extension of function fields.” However, the output of the “Galois closure” operation will not necessarily be a *curve over k* as in our conventions, i.e., it won’t necessarily be a geometrically integral scheme over k , unless k is separably closed. For instance, consider the geometrically Galois morphism $\mathbb{P}_{\mathbb{Q}}^1 \rightarrow \mathbb{P}_{\mathbb{Q}}^1$ given by $t \mapsto t^3$. This is not a Galois extension of fields, and a Galois closure is $\mathbb{P}_{\mathbb{Q}(\zeta_3)}^1$, which is not a geometrically irreducible variety over \mathbb{Q} . In the language of field theory, the field extension $\mathbb{Q} \subset \mathbb{Q}(\zeta_3)(t)$ is not *regular*. Therefore, when we take a Galois closure, we implicitly extended the field k if necessary to ensure that the output is a *curve over k* .

We begin with a simple Galois-theoretic observation related to the existence of a core.

Lemma 4.2. *Let $X \leftarrow Z \rightarrow Y$ be a correspondence over a field k where Z is hyperbolic. A core exists if and only if there exists a curve W , possibly after replacing k by a finite extension, together with a map $W \rightarrow Z$ such that the composite maps $W \rightarrow X$ and $W \rightarrow Y$ are both finite Galois.*

Proof. Suppose such a curve W existed. Then W is hyperbolic because it maps nontrivially to a hyperbolic curve (Lemma 3.2). The groups $\text{Aut}(W/X)$ and $\text{Aut}(W/Y)$ are both subgroups of $\text{Aut}_k(W)$, which is a finite group because W is hyperbolic. The group I generated by these two Galois groups is therefore finite. Therefore the curve W/I fits into a diagram:



Therefore a core exists. Conversely, if the correspondence has a core, call the coarse core C . Let W be a Galois closure of the map $Z \rightarrow C$, finitely extending the ground field if necessary. Then the composite maps $W \rightarrow X$ and $W \rightarrow Y$ are both finite Galois as desired. \square

Lemma 4.3. *Let $X \leftarrow Z \rightarrow Y$ be a correspondence of (possibly hyperbolic) curves over \mathbb{F} . Then a core exists if and only if there exists a curve W together with a map $W \rightarrow Z$ such that the composite maps $W \rightarrow X$ and $W \rightarrow Y$ are both finite Galois.*

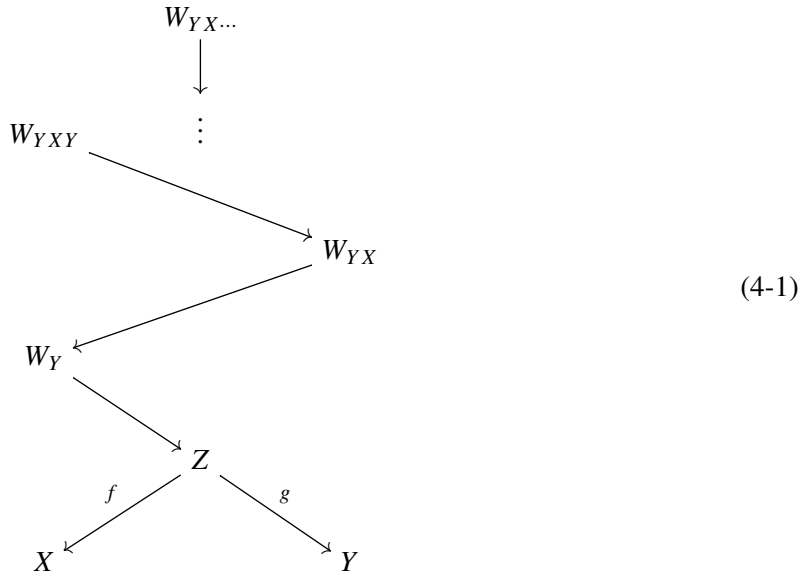
Proof. The proof is almost exactly the same as that of Lemma 4.2: the key observation is that everything in sight (including every element of $\text{Aut}(W/X)$ and $\text{Aut}(W/Y)$) may be defined over some finite field \mathbb{F}_q ; therefore the group they generate inside of $\text{Aut}(W)$ consists of automorphisms defined over \mathbb{F}_q and is hence finite. \square

Remark 4.4. Let $X \leftarrow Z \rightarrow Y$ be a correspondence of affine curves with a core. We prove there exists a curve C together with *finite, generically separable* maps from X and Y making the square commute.

Let $\bar{X} \leftarrow \bar{Z} \rightarrow \bar{Y}$ be the compactified correspondence, with coarse core T . Take a Galois closure \bar{W} of $\bar{Z} \rightarrow T$. Let W be the affine curve associated to the integral closure of $k[Z]$ in $k(\bar{W})$. Then $W \rightarrow X$ and $W \rightarrow Y$ are both finite Galois morphisms of affine curves. In fact, $\text{Aut}(W/X) = \text{Aut}(\bar{W}/\bar{X})$ and likewise for Y . The group I generated by $\text{Aut}(W/X)$ and $\text{Aut}(W/Y)$ inside of $\text{Aut}(W)$ is precisely $\text{Aut}(W/T) = \text{Aut}(\bar{W}/T)$, as T was the coarse core of the projective correspondence. Set $C = W/I$, the affine curve with coordinate ring $k[W]^I$. This C is the *coarse core* of the correspondence of affine curves.

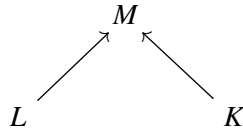
Example 4.5. Let us see the relevance both of Z being hyperbolic in Lemma 4.2 and of the base field being \mathbb{F} in Lemma 4.3. Let $Z = \mathbb{P}_{\mathbb{F}_p(t)}^1$ and consider the following finite subgroups of $\text{PGL}(2, \mathbb{F}_p(t))$: H_1 is generated by the unipotent element $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ and H_2 is generated by the unipotent element $\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$. Quotienting Z gives a correspondence $Z/H_1 \leftarrow Z \rightarrow Z/H_2$. Both arrows are Galois, but there is evidently no core because the subgroup of $\text{PGL}(2, \mathbb{F}_p(t))$ generated by H_1 and H_2 is infinite. Note that for every specialization of $t \in \mathbb{F}$, the correspondence does in fact have a core, for instance by Lemma 4.3.

Let $X \leftarrow Z \rightarrow Y$ be a correspondence of curves without a core where Z is hyperbolic or where $k \cong \mathbb{F}$. We perform the following iterative procedure: take a Galois closure of $Z \rightarrow Y$ and call it W_Y . Because we assumed a core does not exist, the associated map $W_Y \rightarrow X$ cannot be Galois by Lemmas 4.2 and 4.3, respectively. Take a Galois closure of this map and call it W_{YX} . Again, the associated map $W_{YX} \rightarrow Y$ cannot be Galois, so we can take a Galois closure to obtain W_{YXY} . Continuing in the fashion, we get an inverse system of curves $W_{YX\dots}$ over the correspondence.



Note that $W_{YX\dots}$ is Galois over Z . In fact, $W_{YX\dots}$ is Galois over both X and Y because there is a final system of Galois subcovers for each. Note that this procedure may involve algebraic extensions of the field k .

We explicate the based function-field perspective on this construction: let



be the associated diagram of function fields, where $M = k(Z)$, $L = k(X)$, and $K = k(Y)$. Recall that $k \xrightarrow{\sim} L \cap K$; this is exactly the condition that correspondence does not have a core.

Pick an algebraic closure Ω of M , i.e., let Ω be an algebraically closed field of transcendence degree 1 over k and pick *once and for all* an embedding of k -algebras $PQ : M \hookrightarrow \Omega$. (The notation will be justified later, when PQ will correspond to an edge of a graph.) Let E_K be the Galois closure of M/K in Ω . Then E_K/L is no longer Galois by Lemmas 4.2 and Lemma 4.3, respectively. Let E_{KL} be the Galois closure of E_K/L in Ω . Continuing in this fashion, we get an infinite algebraic field extension $E_{KL\dots}$ of M , Galois over both L and K .

Lemma 4.6. *$W_{YX\dots}$ are $W_{XY\dots}$ isomorphic as Z -schemes. That is, by reversing the roles of X and Y we get mutually final systems of Galois covers.*

Proof. Equivalently, we must show that $E_{KL\dots} = E_{LK\dots}$ as subfields of Ω . First, note that $E_K \subset E_{LK}$ because E_K is the minimal extension of E in Ω that is Galois over K . Similarly, $E_{KL} \subset E_{LK}$ because E_{KL} is the minimal extension of E_K in Ω that is Galois over L . Continuing, we see that $E_{KL\dots} \subset E_{LK\dots}$. By symmetry, the reverse inclusion holds as desired. \square

Corollary 4.7. *The field extension $E_{KL\dots} = E_{LK\dots}$ of M , thought of as a subfield of Ω , is characterized by the property that it is the minimal field extension of M inside of Ω that is Galois over both L and K .*

For brevity, we denote the inverse system $W_{XYX\dots}$ by W_∞ . Let E_∞ be the associated function field, considered as a subfield of Ω . In what follows, unless otherwise specified we consider $E_\infty \subset \Omega$ as inclusions of abstract k -algebras.

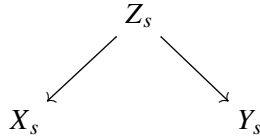
Question 4.8. Does the “field of constants” of E_∞ have finite degree over k ? That is, does $E_\infty \otimes_k \bar{k}$ decompose as an algebra to be the product of finitely many fields? What if the original correspondence is étale?

If $\text{Gal}(k^{\text{sep}}/k)$ is abelian and $\text{Gal}(E_\infty/K)$ has finite abelianization, then Question 4.8 has an affirmative answer. In particular, this applies if $k \cong \mathbb{F}_q$ and $\text{Gal}(E_\infty/K)$ is a semisimple l -adic group. We will see in Proposition 9.10 that if the correspondence has an *étale clump*, then Question 4.8 has an affirmative answer using the following remark.

Remark 4.9. In (4-1), the morphism $W_Y \rightarrow Z$ is Galois. By precomposing with $\text{Aut}(W_Y/Z)$, we equip W_Y with $\text{Aut}(W_Y/Z)$ -many maps to Z . More generally, all curves $W_{YX\dots X}$ will be naturally equipped with $\text{Aut}(W_{YX\dots X}/Z)$ -many maps to Z via precompositions by Galois automorphisms. This will be useful in Remark 7.3, when we try to explicitly understand the curves $W_{YX\dots Y}$.

The following lemma allows us to *specialize étale* correspondences without a core.

Lemma 4.10. *Let S be an irreducible Noetherian scheme with generic point η . Let X, Y , and Z be proper, smooth, geometrically integral curves over S . Suppose Z is “hyperbolic” over S ; that is, the genus of a fiber is at least 2. Let $X \leftarrow Z \rightarrow Y$ be a finite étale correspondence of schemes commuting with the structure maps to S . If s is a geometric point of S such that*



has a core, then $X_\eta \leftarrow Z_\eta \rightarrow Y_\eta$ has a core.

Proof. The property of “having a core” does not change under algebraic field extension by Proposition 3.8. By dévissage, we reduce to the case of $S = \text{Spec}(R)$, where R is a discrete valuation ring with algebraically closed residue field κ . Call the fraction field K . We may further replace R by its integral closure in \bar{K} to get a valuation ring having both the residue field and the fraction field algebraically closed. We do this to not worry about the “extension of ground field” question that is always present when taking a Galois closure.

Call the generic point η and the closed point s . First of all $X_\eta \leftarrow Z_\eta \rightarrow Y_\eta$ is a correspondence of curves over η . Let us suppose it does not have a core. Then the process of iterated Galois closure, as detailed in (4-1), continues endlessly to produce a tower of curves over η . On the other hand, any finite étale morphism has a Galois closure. This implies that we can apply the construction of taking iterated Galois closures to the *finite étale correspondence of schemes* $X \leftarrow Z \rightarrow Y$ to build a tower $W_{YX\dots}$ over S .

As R has an algebraically closed residue field and fraction field, $W_{YX\dots X}$ is a smooth proper curve over S ; in particular the geometric fibers of the morphism $W_{YX\dots X} \rightarrow S$ are irreducible. Moreover, all of the maps $W_{YX\dots Y} \rightarrow Z$ are finite étale. In fact, the fiber of $W_{XY\dots X}$ over the generic point η of $W_{XY\dots X}$ is isomorphic, as a scheme over Z_η , to the corresponding curve in the tower associated to the correspondence $X_\eta \leftarrow Z_\eta \rightarrow Y_\eta$ of curves over η . For instance, $(W_Y)_\eta \rightarrow Y_\eta$ is a Galois closure of the finite étale morphism $Z_\eta \rightarrow Y_\eta$. Therefore, if we could prove $(W_{YX\dots X})_\eta$ were disconnected, we would get a contradiction with the original assumption that $X_\eta \leftarrow Z_\eta \rightarrow Y_\eta$ had no core.

The fact that the maps $Z \rightarrow X, Z \rightarrow Y$, and $W_{YX\dots Y} \rightarrow Z$ are finite étale implies that taking a Galois closure and then restricting to s yields a finite Galois étale cover of Z_s . For example, $(W_Y)_s$ is a (possibly disconnected) finite Galois cover of Y_s that maps surjectively to $(W_s)_{Y_s}$, a Galois closure of the map $Z_s \rightarrow Y_s$.

As the correspondence specialized to s has a core and $Z \rightarrow S$ is assumed to be hyperbolic, Lemma 4.2 implies that there exists a curve $W_{YX\dots Y}$ of our tower over S such that the fiber $(W_{YX\dots Y})_s$ is disconnected. We therefore have a smooth proper curve $W_{XY\dots Y} \rightarrow S$ such that the fiber over s is disconnected. Zariski’s connectedness principle (e.g., Corollary 8.3.6 of [Liu 2002] for the case of curves) implies $(W_{YX\dots Y})_\eta$ is disconnected (this is where we use properness), contradicting our original assumption that $X_\eta \leftarrow Z_\eta \rightarrow Y_\eta$ had no core. □

Remark 4.11. Example 4.5 shows that the argument of Lemma 4.10 *does not work* if the correspondence is not assumed to be étale.

The following corollary shows we may “reduce” the study of Question 3.16 to where $k = \mathbb{F}$.

Corollary 4.12. *Given an étale correspondence of hyperbolic curves $X \leftarrow Z \rightarrow Y$ without a core over a field k of characteristic p , we can specialize to an étale correspondence without a core over \mathbb{F} .*

Proof. By spreading out, we may ensure that we are in the situation of Lemma 4.10. Then the nonexistence of a core implies the same for all of the geometric fibers by Lemma 4.10. \square

Corollary 4.13. *Let S be an integral scheme of finite type over \mathbb{C} with generic point η . Let X, Y , and Z be proper, smooth, geometrically integral, hyperbolic curves over S . Let $X \leftarrow Z \rightarrow Y$ be a finite étale correspondence of schemes commuting with the structure maps to S . Suppose that $X_\eta \leftarrow Z_\eta \rightarrow Y_\eta$ has no core. Then this family of correspondences is étale locally constant.*

Proof. Let \mathcal{M} denote the moduli space of finite étale correspondences of projective curves with the same genera as X, Y , and Z . This is a finite-type moduli “space” (i.e., a DM algebraic stack) that may be defined over $\overline{\mathbb{Q}}$ and has quasiprojective coarse space by the analogous assertions for \mathcal{M}_g for $g \geq 2$. In particular, we obtain an algebraic map

$$u : S \rightarrow \mathcal{M}_{\mathbb{C}}$$

classifying the family $X \leftarrow Z \rightarrow Y$ over S . By Lemma 4.10, for every point $s \in S(\mathbb{C})$, the correspondence $X_s \leftarrow Z_s \rightarrow Y_s$ has no core; therefore Remark 3.11 implies that the image $u(s)$ is a $\overline{\mathbb{Q}}$ -valued point of $\mathcal{M}(\mathbb{C})$. On the other hand, $u(S)$, being the image of a morphism between an irreducible finite-type \mathbb{C} -scheme and a finite-type \mathbb{C} DM-stack, either has image uncountable or a single point by Chevalley’s theorem: a constructible subset of a finite-type DM stack over \mathbb{C} cannot be countably infinite.

We therefore see that the image of u is a single point; as S is reduced, there exists an étale cover $S' \rightarrow S$ such that when we pull $X \leftarrow Z \rightarrow Y$ to S' , the family is locally constant. \square

In other words, Corollary 4.13 says that if $X \leftarrow Z \rightarrow Y$ is an étale correspondence of compact hyperbolic Riemann surfaces without a core, then inside of the moduli space \mathcal{M} of finite étale correspondences of projective curves with the appropriate genera, the moduli point $[X \leftarrow Z \rightarrow Y]$ is isolated.

Lemma 4.10 says that for an étale correspondence of projective hyperbolic curves, the property of “not having a core” specializes. We now show that the property of “not having a core” generalizes, even without the assumption of étale-ness. This is rather useful; it implies that one way to answer Question 3.16 is to directly lift the correspondence to characteristic 0.

Lemma 4.14. *Let $S = \text{Spec}(R)$ be the spectrum of a discrete valuation ring with closed point s and generic point η . Let X, Y , and Z be smooth, projective, geometrically irreducible curves over S and let $X \leftarrow Z \rightarrow Y$ be a correspondence of schemes, commuting with the structure maps to S , that is a correspondence of curves when restricted to s and to η . If over η the correspondence has a core, then over s the correspondence has a core.*

Proof. Let π be a uniformizer of R . Denote by κ residue field of R and by K the fraction field of R . Pick a nonconstant rational function f in the intersection $K(X) \cap K(Y)$ (the intersection takes place in $K(Z)$).

By multiplying by an appropriate power of π , we can guarantee that f extends to rational functions on the special fiber and in fact that f has nonzero reduction in $0 \neq \bar{f} \in \kappa(X_s) \cap \kappa(Y_s)$. Suppose f is constant modulo π , or equivalently that $f \equiv c \pmod{\pi}$ for some $c \in R$. Then $(f - c)/\pi$ may again be reduced modulo π . If $(f - c)/\pi$ is nonconstant on the special fiber, we are done, so suppose not and repeat the procedure. This procedure terminates because our original choice of $f \in K(X)$ was nonconstant and the result will be a nonconstant function in $\kappa(X_s) \cap \kappa(Y_s)$. \square

Corollary 4.15. *Let $X \leftarrow Z \rightarrow Y$ be an étale correspondence of smooth projective hyperbolic curves over \mathbb{F} without a core. If correspondence lifts to a correspondence of curves $\tilde{X} \leftarrow \tilde{Z} \rightarrow \tilde{Y}$ over $W(\mathbb{F})$, then X, Y , and Z are the reductions modulo p of Shimura curves.*

Proof. The lifted correspondence is automatically étale as the étale locus is open. Lemma 4.14 implies that the generic fiber does not have a core. Mochizuki's Theorem 3.10 then implies that \tilde{X}, \tilde{Y} , and \tilde{Z} are all Shimura curves as desired. \square

5. The generic graph of a correspondence

Let $X \leftarrow Z \rightarrow Y$ be a correspondence of curves over k . and let Ω be an algebraically closed field of transcendence degree 1 over k , thought of as a k -algebra. We construct an infinite 2-colored graph $\mathcal{G}_{\text{gen}}^{\text{full}}$, which we call the *full generic graph* of the correspondence. The blue vertices of $\mathcal{G}_{\text{gen}}^{\text{full}}$ are the Ω -valued points of X ; more precisely, a blue vertex is given by a k -algebra homomorphism $k(X) \rightarrow \Omega$. Similarly, the red vertices are the Ω -valued points of Y and the edges are the Ω -valued points of Z . A blue vertex $p : k(X) \hookrightarrow \Omega$ and red vertex $q : k(Y) \hookrightarrow \Omega$ are joined by an edge if there exists an embedding $k(Z) \hookrightarrow \Omega$ that restricts to p and to q on the subfields $k(X)$ and $k(Y)$ respectively. Note that $\text{Aut}_k(\Omega)$ naturally acts on the graph $\mathcal{G}_{\text{gen}}^{\text{full}}$ by postcomposition.

Remark 5.1. The original correspondence is minimal if and only if there are no multiple edges in $\mathcal{G}_{\text{gen}}^{\text{full}}$. (Recall that the morphisms of curves were generically separable by definition.)

Condition 5.2. For the rest of the sections involving graph theory, we suppose that the correspondence $X \leftarrow Z \rightarrow Y$ is minimal in order to get a graph and not a multigraph.

Definition 5.3. Given any subgraph $H \subset \mathcal{G}_{\text{gen}}^{\text{full}}$, we define the subfield $E_H \subset \Omega$ by taking the compositum of the subfields $e(k(Z)) \subset \Omega$, $p(k(X)) \subset \Omega$, and $q(k(Y)) \subset \Omega$ corresponding to all of the edges and vertices e , p , and q in H .

There is no reason to believe that $\mathcal{G}_{\text{gen}}^{\text{full}}$ is connected. We give $\mathcal{G}_{\text{gen}}^{\text{full}}$ a distinguished blue vertex P , red vertex Q , and edge PQ between them by picking the k -embedding

$$PQ : k(Z) \hookrightarrow \Omega$$

and we set the graph \mathcal{G}_{gen} (the *generic graph*) to be the connected component of $\mathcal{G}_{\text{gen}}^{\text{full}}$ containing this distinguished edge. All connected components of $\mathcal{G}_{\text{gen}}^{\text{full}}$ arise in this way and all connected components of

$\mathcal{G}_{\text{gen}}^{\text{full}}$ are isomorphic. We denote by $P(k(X))$ the image of the distinguished blue point P as a subfield of Ω and similarly for $Q(k(Y))$.

Lemma 5.4. *Let $H \subset \mathcal{G}_{\text{gen}}$ be the full subgraph consisting of all vertices of distance at most n from a fixed vertex v ; that is, H is the closed ball $H = B(v, n)$. Then E_H is Galois over E_v .*

Proof. First of all, E_v is the field corresponding to v as in Definition 5.3. We may suppose without loss of generality that v is a blue vertex, so $E_v = v(k(X))$ as v is by definition a k -embedding $k(X)$ to Ω . In other words, v gives Ω the structure of a $k(X)$ -algebra. Now, E_H is the compositum of all of the fields associated to all of the edges and vertices in H in Ω . In particular, if $\mathcal{P} = \{P\}$ is the collection of all paths of length n starting at v , then E_H is the compositum of $(E_P)_{P \in \mathcal{P}}$ inside of Ω . Here each E_P and E_H has a $k(X)$ -algebra structure via v and our goal is to prove that E_H is Galois over $k(X)$ with respect to this algebra structure $v : k(X) \hookrightarrow E_H$.

Consider E_H together with the subfields E_P , for $P \in \mathcal{P}$, as abstract $k(X)$ -algebras. Let ϕ_0 be the original embedding $E_H \hookrightarrow \Omega$. To prove E_H is Galois over $k(X)$, we must show that for every

$$\phi \in \text{Hom}_{k(X)}(E_H, \Omega)$$

the image of ϕ is contained in $\phi_0(E_H)$. Note that ϕ is determined by where all of the E_P are sent. Any ϕ can be obtained from ϕ_0 via an element of $\text{Aut}(\Omega/k(X))$, as Ω is algebraically closed, and so a path P of length n originating at v is sent to another such path P' . In other words, $\phi(E_P) = \phi_0(E_{P'})$ for another path P' of length n originating at v . As E_H was the compositum of all such E_P , it follows that the extension $E_H/k(X)$ is Galois as desired. \square

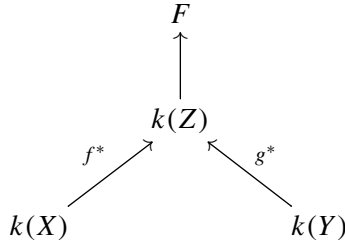
The graph \mathcal{G}_{gen} is a full subgraph of $\mathcal{G}_{\text{gen}}^{\text{full}}$ so, as in Definition 5.3, we can take the associated field $E_{\mathcal{G}_{\text{gen}}} \subset \Omega$ given by the compositum of the subfields of Ω associated to the edges. Let $E \subset \Omega$ be the minimal field extension of $k(Z)$ (with respect to the embedding $PQ : k(Z) \hookrightarrow \Omega$) that is Galois over both $k(X)$ and $k(Y)$. We prove that $E = E_{\mathcal{G}_{\text{gen}}}$ with the next series of results.

Corollary 5.5. *The subfield $E_{\mathcal{G}_{\text{gen}}} \subset \Omega$ is Galois over both $P(k(X))$ and $Q(k(Y))$. Therefore $E \subset E_{\mathcal{G}_{\text{gen}}}$.*

Proof. The connected graph \mathcal{G}_{gen} is the union of the subgraphs $\bigcup_n B(P, n)$ of closed balls of radius n around P , so by Lemma 5.4 the field $E_{\mathcal{G}_{\text{gen}}}$ is Galois over $P(k(X))$. Similarly, $E_{\mathcal{G}_{\text{gen}}}$ is Galois over $Q(k(Y))$. Therefore $E \subset E_{\mathcal{G}_{\text{gen}}}$ as desired. \square

Lemma 5.6. *Let $X \leftarrow Z \rightarrow Y$ be a correspondence of curves over k and embed the function fields into Ω via $PQ : k(Z) \hookrightarrow \Omega$. If there is a subfield $F \subset \Omega$ that is Galois over both $k(X)$ and $k(Y)$, then $E_{\mathcal{G}_{\text{gen}}} \subset F$.*

Proof. We have the following diagram of fields:



where F is Galois over both $k(X)$ and $k(Y)$. The field F is naturally equipped with the structure of a $k(Z)$ algebra. Extend $PQ : k(Z) \hookrightarrow \Omega$ any which way to a $k(Z)$ -algebra embedding $\phi : F \rightarrow \Omega$. Then the image of any edge adjacent to P in \mathcal{G}_{gen} lands inside of the image $\phi(F)$ because F is Galois over $k(X)$. Similarly, the image of any edge adjacent to Q in \mathcal{G}_{gen} lives inside the image of $\phi(F)$.

Let $q \neq Q$ be a vertex adjacent to P . There exists an automorphism $\alpha \in \text{Gal}(\phi(F)/P(k(X)))$ that sends $Q(k(Y))$ to E_q because F is Galois over $k(X)$. Conjugating by α , we deduce that $\phi(F)$ is Galois over E_q and hence the image of all edges emanating from q lie in $\phi(F)$. By propagating, we get that $E_{\mathcal{G}_{\text{gen}}} \subset F$ as desired. □

Corollary 5.7. *We have an equality of fields $E = E_{\mathcal{G}_{\text{gen}}}$, considered as subfields of Ω . Equivalently, $E_{\mathcal{G}_{\text{gen}}}$ is the minimal field extension of $PQ(k(Z))$ inside Ω that is Galois over the fields $P(k(X))$ and $Q(k(Y))$.*

Proof. Combine Lemma 5.6 and Corollary 5.5. □

Corollary 5.8. *Let $X \leftarrow Z \rightarrow Y$ be a correspondence of curves over k without a core with Z hyperbolic or with $k \cong \mathbb{F}$. Then $E_\infty \cong E_{\mathcal{G}_{\text{gen}}}$.*

Proof. The field E_∞ is also the minimal field extension of $PQ(k(Z))$ inside of Ω that is Galois over $P(k(X))$ and $Q(k(Y))$ by Corollary 4.7. □

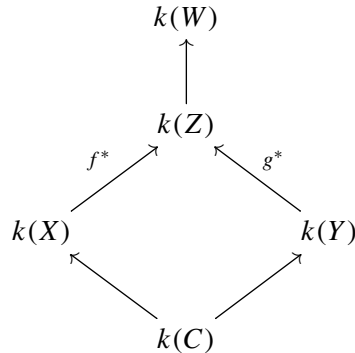
Remark 5.9. One is tempted to make a converse definition to Definition 5.3: given any subfield $E \subset \Omega$ or $E \subset E_\infty$, define $\mathcal{G}_E^{\text{full}}$ and \mathcal{G}_E to be the subgraphs of $\mathcal{G}_{\text{gen}}^{\text{full}}$ and \mathcal{G}_{gen} , respectively, whose points and edges have image inside of E . This definition is rather poorly behaved; for instance if one starts out with a finite connected subgraph $H \subset \mathcal{G}_{\text{gen}}$, takes $E_H \subset E_\infty$, and then looks at the associated graph G_{E_H} , there is no reason to believe that this graph is connected.

The graph \mathcal{G}_{gen} informally reflects the “generic dynamics” of the correspondence. We will see one way of making this precise in Section 7: via a specialization map. Nevertheless, we have the following proposition, which says that a core exists if and only if \mathcal{G}_{gen} is finite (i.e., the “generic dynamics” are bounded), in line with Remark 3.7.

Proposition 5.10. *Let $X \leftarrow Z \rightarrow Y$ be a correspondence of curves over k where Z is hyperbolic or where $k \cong \mathbb{F}$. This correspondence has no core if and only if \mathcal{G}_{gen} is an infinite graph.*

Proof. If \mathcal{G}_{gen} is finite, then $E_{\mathcal{G}_{\text{gen}}}$ is a finite Galois extension of both $k(X)$ and $k(Y)$, so the correspondence has a core by Lemma 4.2 or Lemma 4.3, respectively.

Conversely, if the correspondence had a core, then let C be the coarse core. Let W be a Galois closure of $Z \rightarrow C$. We have the following diagram of fields, where we again fix $PQ : k(Z) \hookrightarrow \Omega$ and any extension $\phi : k(W) \hookrightarrow \Omega$:



Call P , Q , and R the restriction of the algebra embedding PQ to $k(X)$, $k(Y)$, and $k(C)$ respectively. Let v be blue vertex in \mathcal{G}_{gen} adjacent to Q , given by a k -algebra embedding $v : k(X) \hookrightarrow \phi(k(W)) \subset \Omega$ by Lemma 5.6. As $k(W)/k(Y)$ is Galois, there exists an automorphism

$$\alpha \in \text{Gal}(\phi(k(W))/Q(k(Y))) \cong \text{Gal}(k(W)/k(Y))$$

that sends P to v . As $k(C) \subset k(Y)$, this implies that $v|_{k(C)} = R$. By propagating, we see that for every vertex v of \mathcal{G}_{gen} , $v|_{k(C)} = R$. Therefore, for every edge $e \in \mathcal{G}_{\text{gen}}$, thought of as a k -algebra embedding $e : k(Z) \hookrightarrow \Omega$, we have that $e|_{k(C)} = R$. On the other hand, $k(Z)$ is a finite extension of $k(C)$, so there are only finitely many ways to extend R to a k -algebra homomorphism $k(Z) \hookrightarrow \Omega$. Therefore the number of edges is finite, as desired. \square

We record the following easy proposition for later use in proving the surjectivity of the specialization morphism in the case of an étale correspondence without a core.

Proposition 5.11. *For any finite subgraph $H \in \mathcal{G}_{\text{gen}}$, the field E_H is contained in a finite extension F of $PQ(k(Z))$.*

Proof. We have the following two facts:

- E_H lands inside of E_∞ , which is exhausted by fields of the form $E_{KL\dots K}$, by Lemma 5.6.
- E_H is finitely generated over k .

Therefore E_H lands inside of some $E_{KL\dots K}$, a finite extension of $PQ(k(Z))$, as desired. \square

We now analyze the action of various subgroups of $\text{Aut}(E_\infty)$ on \mathcal{G}_{gen} .

Remark 5.12. We take a brief digression into the structure of automorphism groups of fields. Let Ω be any field. We endow the group $\text{Aut}(\Omega)$ with the compact-open topology, considering Ω to be a discrete set. Given any finite subset $S \subset \Omega$, the subgroup $\text{Stab}(S) \subset \text{Aut}(\Omega)$ is an open subgroup and as S ranges these form a neighborhood base of the identity in $\text{Aut}(\Omega)$. If $K \subset \Omega$ is a separable Galois extension with K finitely generated over its prime field, the natural map $\text{Gal}(\Omega/K) \subset \text{Aut}(\Omega)$ is an open embedding

of topological groups; in other words, the topology just defined is compatible with the usual profinite topology on Galois groups.

Note that this procedure generalizes: if $k \subset \Omega$ is a field extension, we may give the group $\text{Aut}_k(\Omega)$ the structure of a topological group, where a neighborhood base of the identity is given by $\text{Stab}(S)$ for finite subsets $S \subset \Omega \setminus k$. However, $\text{Aut}_k(\Omega)$ is not an *open subgroup* of $\text{Aut}(\Omega)$ unless k is finitely generated over its prime field.

Any element $g \in \text{Aut}_k(E_\infty)$ gives a map of graphs $\mathcal{G}_{\text{gen}} \rightarrow \mathcal{G}_{\text{gen}}^{\text{full}}$ by postcomposition: for instance, an edge $e : k(Z) \rightarrow E_\infty \subset \Omega$ is sent to the edge $g \circ e : k(Z) \rightarrow E_\infty \subset \Omega$. In fact, the Galois groups $G_P := \text{Gal}(E_\infty/P(k(X)))$ and $G_Q := \text{Gal}(E_\infty/Q(k(Y)))$ actually act on the connected graph: $g \in G_P$ sends an edge $e : k(Z) \rightarrow E_\infty \subset \Omega$ to $g \circ e : k(Z) \rightarrow E_\infty \subset \Omega$, and $g \circ e$ is an edge of the connected graph \mathcal{G}_{gen} because g fixes P .

Definition 5.13. Let $A \subset \text{Aut}_k(E_\infty)$ the subgroup of $\text{Aut}_k(E_\infty)$ sends \mathcal{G}_{gen} to itself with the induced topology, as in Remark 5.12. Let $A^{PQ} \subset A$ be the subgroup of A generated by G_P and G_Q with the induced topology from A .

Question 5.14. Is $A \hookrightarrow \text{Aut}_k(E_\infty)$ an isomorphism?

Remark 5.15. The topology on A^{PQ} is uniquely determined by declaring the compact subgroups G_P and G_Q to be open.

By definition, A acts faithfully on \mathcal{G}_{gen} : if $g \in A$ acts trivially on \mathcal{G}_{gen} , then it acts trivially on the field generated by all of the vertices and the edges of \mathcal{G}_{gen} , i.e., it is the trivial automorphism of E_∞ . If we give \mathcal{G}_{gen} the discrete topology, A^{PQ} acts continuously on \mathcal{G}_{gen} ; that is, the stabilizer of a vertex is an open subgroup. Let $d = \deg(Z \rightarrow X)$ and $e = \deg(Z \rightarrow Y)$. Then the degree of a blue vertex is d and the degree of a red vertex is e . Moreover, G_P acts transitively on the edges coming out of P by Galois theory and similarly G_Q acts transitively on the edges coming out of Q . By conjugating we see that $A^{PQ} \subset \text{Aut}(\mathcal{G}_{\text{gen}})$ acts transitively on the edges coming out of any vertex. Therefore the group A^{PQ} acts transitively on the edges of \mathcal{G}_{gen} , subject to the constraint that colors of the vertices are preserved. This is recorded in the following corollary.

Corollary 5.16. *In the notation above, A^{PQ} and hence also A act transitively on the edges of \mathcal{G}_{gen} , subject to the constraint that the colors of the vertices are preserved. We say the pair $(\mathcal{G}_{\text{gen}}, A^{PQ})$ is colored-edge-symmetric.*

Question 5.17. If $X \leftarrow Z \rightarrow Y$ is a minimal correspondence with no core, does \mathcal{G}_{gen} have any cycles? What if it is étale?

The graph \mathcal{G}_{gen} being a tree has consequences for the structure of the group A^{PQ} . To state these, we need a theorem of Serre.

Theorem 5.18 (Serre). *Let G be a group acting on a graph X , and let e be an edge of X connecting vertices p and q . Suppose that e is a fundamental domain for the action. Let G_p , G_q , and G_e be the stabilizers in G of p , q , and e respectively. Then the following are equivalent:*

- (1) X is a tree.
- (2) The homomorphism $G_P *_{G_e} G_Q \rightarrow G$ induced by the inclusions $G_P \rightarrow G$ and $G_Q \rightarrow G$ is an isomorphism.

Proof. This is a direct translation of Théorème 6 on Page 48 of [Serre 1977]. \square

Proposition 5.19. *Suppose \mathcal{G}_{gen} is a tree. Then the natural map $G_P *_{G_{PQ}} G_Q \rightarrow A^{PQ}$ is an isomorphism of topological groups.*

Proof. There is no element $a \in A^{PQ}$ that flips any edge e of \mathcal{G}_{gen} because A^{PQ} preserves the coloring. By Corollary 5.16, the segment PQ is a fundamental domain for the action of A^{PQ} on \mathcal{G}_{gen} . Therefore, by Serre's theorem, the fact that \mathcal{G}_{gen} is a tree implies the induced map $G_P *_{G_{PQ}} G_Q \rightarrow A^{PQ}$ is an isomorphism of abstract groups. The group $G_P *_{G_{PQ}} G_Q$ has a natural topology generated by the topologies of G_P and G_Q (because G_{PQ} is an open subgroup of both G_P and G_Q), and endowed with this topology the above map is an isomorphism of topological groups. \square

When \mathcal{G}_{gen} is a tree, we may describe the pair $(\mathcal{G}_{\text{gen}}, A^{PQ})$ in a different way. Given any compact open subgroup $G \subset A^{PQ}$ and any vertex $v \in \mathcal{G}_{\text{gen}}$, the orbit $G.v$ is compact and discrete (as we gave \mathcal{G}_{gen} the discrete topology) and is hence finite. Therefore G acts on a finite subtree T of \mathcal{G}_{gen} and hence the action factors through a finite quotient H of G .

Lemma 5.20. *A finite group H acting on a finite tree T always has a fixed point (though not necessarily a fixed vertex).*

Proof. This is well known and Aaron Bernstein explained the following simple proof to us.

Let the height $h(v)$ of a vertex v be the maximal distance of v to any leaf. Any automorphism of T preserves heights. If there is a unique vertex v of minimal height, we are done, so suppose there is another vertex w of minimal height. Then v and w must be connected by an edge: if the unique path between them contained an intermediate vertex u , then some thought shows that $h(u) < h(v)$. As T is a tree, there can be at most two vertices of minimal height. If there are two, then their midpoint is a fixed point for any automorphism of T . \square

Therefore there must be a point $p \in T$ that is fixed by H ; here T is thought of as a topological space. If p were not a vertex in T , then H would fix the two neighboring vertices of the edge p is on because H respects the coloring of the graph. Therefore H fixes at least one vertex v . On the other hand, given any vertex v , the subgroup G_v fixing v is a compact open subgroup. Therefore, the vertices of \mathcal{G}_{gen} are in natural bijective correspondence with the maximal open compact subgroups G of A^{PQ} .

Corollary 5.21. *If \mathcal{G}_{gen} is a tree, any maximal compact open subgroup G of A^{PQ} is conjugate to either G_P or G_Q .*

Proof. The discussion above shows that every maximal compact open subgroup G of A^{PQ} is G_v for some vertex v of \mathcal{G}_{gen} . The group G_v is conjugate in A^{PQ} to G_P or G_Q by Corollary 5.16. Finally, G_P is not conjugate to G_Q in A^{PQ} because the action of A^{PQ} on \mathcal{G}_{gen} preserves the coloring. \square

Remark 5.22. If \mathcal{G}_{gen} is a tree, then the action of A^{PQ} on \mathcal{G}_{gen} is the conjugation action on the maximal compact subgroups.

We may similarly describe the adjacency relation in \mathcal{G}_{gen} from the group A^{PQ} when \mathcal{G}_{gen} is a tree. Recall our standing assumption that the original correspondence $X \leftarrow Z \rightarrow Y$ is minimal (in order for \mathcal{G}_{gen} to not have multiple edges). As above, we suppose the correspondence is of type (d, e) . Then a blue vertex G_v and a red vertex G_w are joined by an edge if the intersection $G_v \cap G_w$ (inside of A^{PQ}) has index d inside of G_v and index e inside of G_w .

6. Symmetric correspondences

Definition 6.1. A *symmetric correspondence* of curves over k is a self-correspondence $X \xleftarrow{f} Z \xrightarrow{g} X$ over curves over k such that there is an involution $w \in \text{Aut}(Z)$ with $f \circ w = g$, i.e., w swaps f and g . We denote by w^* the induced involution on $k(Z)$.

Note that if the correspondence is minimal, w is unique if it exists. Therefore being symmetric is a property and not a structure of a minimal correspondence.

Lemma 6.2. Let $X \xleftarrow{f} Z \xrightarrow{g} X$ be a symmetric correspondence of curves over k without a core. Suppose Z is hyperbolic or $k \cong \mathbb{F}$. Any $w \in \text{Aut}(Z)$ that swaps f and g lifts to an automorphism \tilde{w} of W_∞ . We denote by \tilde{w}^* the associated automorphism of $E_\infty = k(W_\infty)$.

Proof. We proceed exactly as in the discussion at the beginning of Section 4: let W_f and W_g denote a Galois closures of f and g , respectively. The automorphism w of Z swaps f and g and hence we can choose an isomorphism $w_1 : W_g \rightarrow W_f$ living over w on Z :

$$\begin{array}{ccc} W_g & \xrightarrow{w_1} & W_f \\ \downarrow & & \downarrow \\ W & \xrightarrow{w} & W \end{array}$$

Similarly, we can chose an isomorphism $w_2 : W_{gf} \rightarrow W_{fg}$ living over w on Z , again because w swaps the roles of f and g . Continuing in this fashion, we get an isomorphism of towers

$$\tilde{w} : W_{gf\dots} \rightarrow W_{fg\dots}$$

By Lemma 4.6, $W_{fg\dots}$ is isomorphic to W_{gf} as a procurve over W and we may think of \tilde{w} as an automorphism of W_∞ living over $w \in \text{Aut}(Z)$. □

Remark 6.3. Another way of phrasing Lemma 6.2 is as follows. If $X \xleftarrow{f} Z \xrightarrow{g} X$ is a symmetric correspondence without a core with Z hyperbolic, then for any choice of symmetry w , the following map is (infinite) Galois:

$$W_\infty \rightarrow Z / \langle w \rangle .$$

From this perspective, it is clear that the lift \tilde{w} is not unique.

Definition 6.4. Let $X \xleftarrow{f} Z \xrightarrow{g} X$ be a symmetric correspondence of curves over k without a core where Z is hyperbolic or $k \cong \mathbb{F}$. Pick a symmetry w and a lift \tilde{w} to W_∞ , which exists by Lemma 6.2. Let \tilde{w}^* be the associated automorphism of E_∞ . Define $A^w \subset \text{Aut}_k(E_\infty)$ to be the subgroup generated by A^{PQ} and \tilde{w}^* . We give the subgroup $A^w \subset A$ the induced topology from A .

Remark 6.5. The notation A^w is *a priori* ambiguous as it seems to depend on a choice of lift \tilde{w} . Pick a second lift \tilde{w}' of w . Then $\tilde{w}\tilde{w}'$ fixes Z as w was an involution. In particular, $\tilde{w}^*\tilde{w}'^* \in \text{Gal}(E_\infty/k(Z)) \subset A^{PQ}$, so A^w is independent of the choice of lift of w .

Corollary 6.6. Let $X \xleftarrow{f} Z \xrightarrow{g} X$ be a symmetric correspondence of curves over k without a core with symmetry w . Suppose Z is hyperbolic or $k \cong \mathbb{F}$ and let \tilde{w} be a lift of the symmetry to W_∞ . Then A^w and hence A acts transitively on the oriented edges of \mathcal{G}_{gen} .

Proof. Corollary 5.16 says that A^{PQ} acts transitively on \mathcal{G}_{gen} subject to the constraint that the colors of the vertices are preserved. The automorphism $\tilde{w}^* \in \text{Aut}(E_\infty)$ swaps the points P and Q . By conjugating we get that A^w acts transitively on the edges of \mathcal{G}_{gen} , in the usual sense of remembering the endpoints. \square

Corollary 6.7. Let $X \xleftarrow{f} Z \xrightarrow{g} X$ be a symmetric correspondence of curves over k without a core with symmetry w . Suppose Z is hyperbolic or $k \cong \mathbb{F}$. Then A^{PQ} is a normal subgroup of index 2 inside of A^w .

Proof. Conjugating by \tilde{w}^* swaps G_P and G_Q and hence stabilizes A^{PQ} . Therefore A^{PQ} is normal inside of A^w . Moreover, $(\tilde{w}^*)^2 \in A^{PQ}$, so A^w/A^{PQ} is of order 2. \square

Definition 6.8. Let (G, A) be a pair where G is a connected graph and A is a group of automorphisms of G . (G, A) is said to be (sharply) s -transitive if A acts (sharply) transitively on all s -arcs. (G, A) is said to be ∞ -transitive if it is s -transitive for all $s \geq 1$.

In this language, under the hypotheses of Corollary 6.6 the pair $(\mathcal{G}_{\text{gen}}, A)$ is 1-transitive.

Theorem 6.9 (Tutte). Let G be a connected trivalent graph, A a group of automorphisms of G , and s a positive integer. If (G, A) is s -transitive and not $s+1$ -transitive, then (G, A) is sharply s -transitive.

Proof. The proof is exactly the same as in 7.72 in Tutte's book *Connectivity in Graphs* [1966]. Alternatively, see Djoković and Miller [1980], Theorem 1, for exactly this statement. \square

Lemma 6.10. Let $X \leftarrow Z \rightarrow X$ be a symmetric type $(3, 3)$ correspondence of curves over k without a core with symmetry w . Suppose Z is hyperbolic or $k \cong \mathbb{F}$. Then the pair $(\mathcal{G}_{\text{gen}}, A^w)$ is ∞ -transitive and \mathcal{G}_{gen} is a tree.

Proof. Suppose \mathcal{G}_{gen} had a cycle. The graph \mathcal{G}_{gen} is infinite by Proposition 5.10. Then the pair $(\mathcal{G}_{\text{gen}}, A^w)$ is 1-transitive, so there exists a positive n such that $(\mathcal{G}_{\text{gen}}, A^w)$ is n -transitive but not $n+1$ -transitive. Therefore, to prove \mathcal{G}_{gen} is a tree it suffices to prove that the pair $(\mathcal{G}_{\text{gen}}, A^w)$ is ∞ -transitive.

Suppose $(\mathcal{G}_{\text{gen}}, A^w)$ was not ∞ -transitive. Then there exists a positive integer n such that $(\mathcal{G}_{\text{gen}}, A^w)$ is n -transitive but not $n+1$ -transitive because the graph is infinite, connected and 1-transitive. Theorem 6.9 implies that the pair $(\mathcal{G}_{\text{gen}}, A^w)$ is then sharply n -transitive, i.e., there exists a unique automorphism in A^w sending any n -arc to any other n -arc. Therefore any automorphism in A^w that fixes any given n -arc

must be the identity automorphism. To any n -arc R associate the field E_R which is the field generated by the images of the points and edges inside of E_∞ as in Definition 5.3. Pick the n -arc R through P so that E_R is a finite extension of $P(k(X))$. Note that E_∞ is Galois over E_R . The group $\text{Gal}(E_\infty/E_R)$ acts faithfully on \mathcal{G}_{gen} and fixes R . As $(\mathcal{G}_{\text{gen}}, A^w)$ is sharply n -transitive, the group $\text{Gal}(E_\infty/E_R)$ acts trivially on \mathcal{G}_{gen} . Therefore $E_R = E_\infty$ is a finite extension $k(Z)$, Galois over both $k(X)$ and $k(Y)$, which is a contradiction. \square

Lemma 6.10 poses the following refinement to Question 5.17 on whether or not \mathcal{G}_{gen} is a tree.

Question 6.11. Let $X \leftarrow Z \rightarrow Y$ be a minimal, symmetric, étale correspondence of curves over k without a core. Is the pair $(\mathcal{G}_{\text{gen}}, A^w)$ ∞ -transitive?

We may use Question 6.11 to pose a refinement of Question 3.16

Question 6.12. Let $X \leftarrow Z \rightarrow Y$ be a minimal, symmetric, étale correspondence of projective curves over k without a core. Does the pair $(\mathcal{G}_{\text{gen}}, A^w)$ “look like” the action of SL_2 over a local field on its building?

7. Specialization of graphs and special orbits

Given a correspondence $X \xleftarrow{f} Z \xrightarrow{g} Y$ over a field k , we have defined an undirected 2-colored graph $\mathcal{G}_{\text{gen}}^{\text{full}}$, the full generic graph, using an algebraically closed overfield Ω . In this section we define $\mathcal{G}_{\text{phys}}^{\text{full}}$, the full physical graph, which will be an undirected 2-colored graph, using \bar{k} . The goal of this section is to speculate on the behavior of “specialization maps” $s_z : \mathcal{G}_{\text{gen}} \rightarrow \mathcal{G}_{\text{phys},z}$; informally, if we think of \mathcal{G}_{gen} as the “graph of generic dynamics”, this map specializes to the graph associated to the dynamics of a physical point $z \in Z(\bar{k})$.

Definition 7.1. Given a correspondence $X \xleftarrow{f} Z \xrightarrow{g} Y$ of curves over k , the *full physical graph* $\mathcal{G}_{\text{phys}}^{\text{full}}$ is the following 2-colored graph. The edges are the points $z \in Z(\bar{k})$, the blue vertices are the points $X(\bar{k})$ and the red vertices are the points $Y(\bar{k})$. Adjacent to $z : \text{Spec}(\bar{k}) \rightarrow Z$ is the blue vertex $f \circ z \in X(\bar{k})$ and the red vertex $g \circ z \in Y(\bar{k})$. Given a choice of $z \in Z(\bar{k})$, we denote by $\mathcal{G}_{\text{phys},z}$ the connected component of $\mathcal{G}_{\text{phys}}^{\text{full}}$ that contains z .

Recall the construction of \mathcal{G}_{gen} : pick an edge $PQ \in Z(\Omega)$ of $\mathcal{G}_{\text{gen}}^{\text{full}}$ and define \mathcal{G}_{gen} to be the connected component of $\mathcal{G}_{\text{gen}}^{\text{full}}$ that contains PQ , suppressing the implicit PQ in the notation. The field $E_\infty \subset \Omega$ is the compositum of all of the points and edges of \mathcal{G}_{gen} , thought of as subfields of Ω , by Corollary 5.8. Therefore, an edge e of \mathcal{G}_{gen} yields an element of the set $Z(E_\infty)$. Similarly, a blue vertex v of \mathcal{G}_{gen} yields an element of $X(E_\infty)$ and a red vertex w yields an element of $Y(E_\infty)$.

We spell out exactly what is fixed in the construction of a specialization map. First of all, assume the curves X, Y , and Z are proper over k : this is harmless as any correspondence of curves has a canonical compactification. Pick $z \in Z(\bar{k})$. Then pick a point $\tilde{z} \in W_\infty(\bar{k})$, a geometric point of the scheme W_∞ , i.e., a compatible system of geometric points on the tower defining W_∞ , lying over z . Taking the image of \tilde{z} gives a closed point of the scheme W_∞ , and the ring $\mathcal{O}_{W_\infty, \tilde{z}}$ is a valuation ring because it is the filtered

colimit of valuation rings. Moreover, the fraction field of $\mathcal{O}_{W_\infty, \tilde{z}}$ is E_∞ . The choice of $\tilde{z} : \text{Spec}(\bar{k}) \rightarrow W_\infty$ yields a morphism $\pi : \mathcal{O}_{W_\infty, \tilde{z}} \rightarrow \bar{k}$. We now construct the specialization map

$$s_{\tilde{z}} : \mathcal{G}_{\text{gen}} \rightarrow \mathcal{G}_{\text{phys}}^{\text{full}}$$

Let e be an edge of \mathcal{G}_{gen} . As discussed above, e yields an element of $Z(E_\infty)$. We want to describe $s_{\tilde{z}}(e_\infty)$, the image of e , in $\mathcal{G}_{\text{phys}, z}$. We have the following diagram; the dotted arrow exists uniquely because the structure map $Z \rightarrow \text{Spec}(k)$ is proper:

$$\begin{array}{ccc} \text{Spec}(E_\infty) & \xrightarrow{e} & Z \\ \downarrow & \dashrightarrow & \downarrow \\ \text{Spec}(\bar{k}) & \xrightarrow{\tilde{z}} & \text{Spec}(\mathcal{O}_{W_\infty, \tilde{z}}) \longrightarrow \text{Spec}(k) \end{array}$$

Composing \tilde{z} with the dotted arrow, we get an element $\bar{e} \in Z(\bar{k})$. We set $s_{\tilde{z}}(e) = \bar{e}$. The exact same construction works with (red and blue) vertices, and the result is manifestly a map of graphs. Moreover, as \mathcal{G}_{gen} is connected, so is the image.

Finally, we show that the edge $PQ \in Z(E_\infty)$ is sent to z . The inverse image of $\mathcal{O}_{W_\infty, \tilde{z}}$ under the map $PQ : k(Z) \hookrightarrow E_\infty$ is the valuation ring R of $k(Z)$ corresponding to z . Therefore, when $e = PQ$, the above dotted arrow corresponds to the inclusion $R \hookrightarrow \mathcal{O}_{W_\infty, \tilde{z}}$. As \tilde{z} lives over z , composing this inclusion with π yields $s_{\tilde{z}}(PQ) = z$, as desired. Therefore, we have constructed a map of graphs:

$$s_{\tilde{z}} : \mathcal{G}_{\text{gen}} \rightarrow \mathcal{G}_{\text{phys}, z}$$

Lemma 7.2. *Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be an étale correspondence of projective hyperbolic curves without a core over a field k . Then all of the specialization maps are surjective:*

$$s_{\tilde{z}} : \mathcal{G}_{\text{gen}} \rightarrow \mathcal{G}_{\text{phys}, z}$$

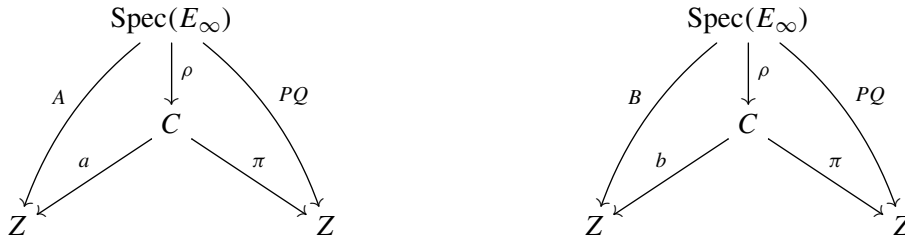
Proof. Because the correspondence is étale, each blue vertex of $\mathcal{G}_{\text{phys}}$ is adjacent to $d = \deg(f)$ edges and each red vertex is adjacent to $e = \deg(g)$ edges. It is therefore equivalent to show that no two adjacent edges of \mathcal{G}_{gen} are sent to the same edge in $\mathcal{G}_{\text{phys}, z}$. Let A and B be two edges sharing the blue vertex p . We want to show that A and B are not sent to the same edge in $\mathcal{G}_{\text{phys}, z}$.

Recall that A and B yield elements of $Z(E_\infty)$ such that $f \circ A = f \circ B = p \in X(E_\infty)$. Proposition 5.11 implies that, after possibly enlarging k , there exists an irreducible curve C together with the maps $\rho : \text{Spec}(E_\infty) \rightarrow C$, $\pi : C \rightarrow Z$, and $a, b : C \rightarrow Z$ such that:

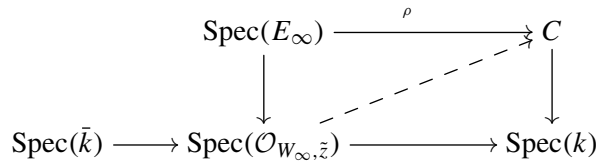
- $\pi \circ \rho = PQ$ considered as elements of $Z(E_\infty)$.
- A and B factor through C via a and b .

In the language of Proposition 5.11, C is the curve associated to a field F of transcendence degree 1 over k , finite over $PQ(k(Z))$, that contains $A(k(Z))$ and $B(k(Z))$. More explicitly, we have the following

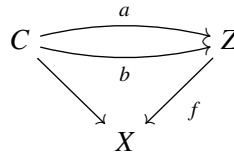
factorizations:



Moreover, the maps π , a , and b are all finite étale. Let us follow the specialization construction. Again, the dotted arrow exists because $C \rightarrow \text{Spec}(k)$ is proper:



This diagram gives us a point $x \in C(\bar{k})$ by composition with the dotted arrow. If A and B are identified under the specialization map, $a(x) = b(x) \in Z(\bar{k})$. Now, $f \circ a = f \circ b$ because A and B shared the vertex ρ , so we have the following diagram:

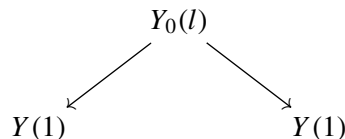


But C is irreducible and the maps a , b , and f are finite étale, so the assumption that $a(x) = b(x)$ implies that $a = b$ and hence $A = B$, as desired. \square

Remark 7.3. The graph $\mathcal{G}_{\text{phys}}$ helps describe the tower W_∞ . In this remark, we suppose all morphisms are unramified at all points specified. For instance, let $\xi_Y \in W_Y(\bar{k})$ map to $z \in Z(\bar{k})$ which maps to $y \in Y(\bar{k})$. Then, as in Remark 4.9, there are naturally $\text{Aut}(W_Y/Z)$ many maps from W_Y to Z and we can look at the images of ξ_Y under these maps. In this way, ξ_Y yields the graph of all edges coming out of y in $\mathcal{G}_{\text{phys},z}$. More generally, a point $\xi_{YX\dots Y} \in W_{YX\dots Y}(\bar{k})$ which maps to $y \in Y(\bar{k})$ under the natural map yields the subgraph of $\mathcal{G}_{\text{phys},z}$ with center y and radius n , where n is the number of letters in the string “ $YX\dots Y$ ”.

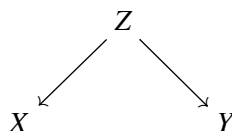
We will use this observation in Proposition 9.10 to show that if an étale clump exists, then Question 4.8 has an affirmative answer.

Consider the Hecke correspondence of open modular curves over \mathbb{F}_p :



The graph \mathcal{G}_{gen} is a tree. For $z \in Y_0(l)(\mathbb{F})$ an ordinary point, $\mathcal{G}_{\text{phys},z}$ has at most one cycle. This follows from the work in David Kohel's thesis [1996], summarized by Andrew Sutherland [2013]. They call this structure an *isogeny volcano*. The cycle comes from the following fact: given an imaginary quadratic field K/\mathbb{Q} , there exists an elliptic curve E/\mathbb{F} with multiplication by the maximal order \mathcal{O}_K . On the other hand, there are only finitely many supersingular points, and in fact Theorem 9.6 implies that if $\mathcal{G}_{\text{phys},z}$ contains one supersingular point it contains all of them.

Definition 7.4. Given an étale correspondence of projective hyperbolic curves



over k without a core, we say a point $z \in Z(\bar{k})$ is *special* if there exists (equivalently for all) $\tilde{z} \in W_\infty(\bar{k})$ over z such that the map $s_{\tilde{z}} : \mathcal{G}_{\text{gen}} \rightarrow \mathcal{G}_{\text{phys},z}$ is not an isomorphism. We say $z \in Z(\bar{k})$ is *generic* if it is not special.

Question 7.5. Let $X \leftarrow Z \rightarrow Y$ be an étale correspondence of projective curves over \mathbb{F}_q without a core.

- (1) Is there always $z \in Z(\mathbb{F})$ that is generic?
- (2) Is there always a special point with unbounded orbit?
- (3) Suppose \mathcal{G}_{gen} is free. For every point $z \in Z(\mathbb{F})$, is $\pi_1(\mathcal{G}_{\text{phys},z})$ finitely generated? If $\mathcal{G}_{\text{phys},z}$ is infinite, does $\mathcal{G}_{\text{phys},z}$ have one cycle?
- (4) What is $\lim_{n \rightarrow \infty} |\{z \in Z(\mathbb{F}_{q^n}) \text{ with } z \text{ generic}\}| / |Z(\mathbb{F}_{q^n})|$?

8. Invariant line bundles and invariant sections

In this section we will need somewhat refined information about abelian varieties and finite group schemes over a field k . Our main reference is [van der Geer and Moonen 2008].

Definition 8.1. Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of curves over k . An *invariant line bundle* \mathcal{L} on the correspondence is a triple $(\mathcal{L}_X, \mathcal{L}_Y, \phi)$ where \mathcal{L}_X is a line bundle on X , \mathcal{L}_Y is a line bundle on Y , and $\phi : f^*\mathcal{L}_X \rightarrow g^*\mathcal{L}_Y$ is an isomorphism of line bundles on Z . The *degree* of an invariant line bundle \mathcal{L} on a correspondence of projective curves is $\deg(f^*\mathcal{L}_X) = \deg(g^*\mathcal{L}_Y)$ on Z . An isomorphism of invariant line bundles $i : \mathcal{L} \rightarrow \mathcal{L}'$ is a pair of isomorphisms $i_X : \mathcal{L}_X \rightarrow \mathcal{L}'_X$ and $i_Y : \mathcal{L}_Y \rightarrow \mathcal{L}'_Y$ that intertwine ϕ and ϕ' when pulled back to Z .

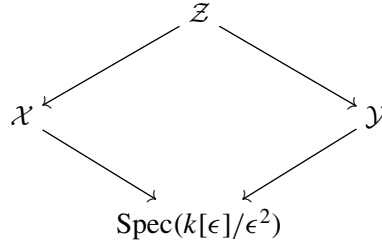
The cohomology of an invariant line bundle \mathcal{L} is defined as follows:

$$H^i(\mathcal{L}) := \{(\xi_X, \xi_Y) \in H^i(X, \mathcal{L}_X) \oplus H^i(Y, \mathcal{L}_Y) \mid f^*(\xi_X) = \phi^* g^*(\xi_Y) \in H^i(Z, f^*\mathcal{L}_X)\}.$$

On a correspondence of projective curves, the group $H^i(\mathcal{L})$ is naturally a finite-dimensional k vector space, and we let $h^i(\mathcal{L}) = \dim_k H^i(\mathcal{L})$. We call elements of $H^0(\mathcal{L})$ *invariant sections*. When it is especially clear from context, we omit the prefix “invariant”.

In general, $\mathcal{O} = (\mathcal{O}_X, \mathcal{O}_Y, 1)$ is an invariant line bundle. Note that if the correspondence is étale, there is a natural invariant line bundle: $\Omega = (\Omega_X^1, \Omega_Y^1, \phi)$; here ϕ is the composition of the canonical isomorphism $f^*\Omega_X^1 \rightarrow \Omega_Z^1$ and the inverse of the canonical isomorphism $g^*\Omega_Y^1 \rightarrow \Omega_Z^1$. We call elements of $H^0(\Omega)$ *invariant differential 1-forms*.

Let \mathcal{T} denote the dual of Ω . Then the first-order (equal-characteristic) deformation space of an étale correspondence of projective curves over k is $H^1(\mathcal{T})$, as we now explain. A first-order deformation is a diagram



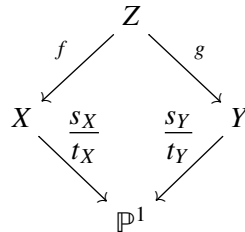
together with an identification of the special fiber with $X \leftarrow Z \rightarrow Y$. The first-order (equal-characteristic) deformation space of a smooth projective curve C/k is naturally isomorphic to $H^1(C, T_C)$. In particular, associated to such a diagram we obtain elements $\xi_X \in H^1(T_X)$, $\xi_Y \in H^1(T_Y)$, and $\xi_Z \in H^1(Z, T_Z)$ associated to \mathcal{X} , \mathcal{Y} , and Z respectively. Now as $Z \rightarrow X$ is finite étale, any deformation \tilde{X} of X naturally induces a deformation \tilde{Z} of Z that maps to \tilde{X} . In the case of first-order deformations, this corresponds to the inclusion

$$H^1(X, T_X) \hookrightarrow H^1(Z, f^*T_X) \cong H^1(Z, T_Z).$$

Putting these facts together we see that $H^1(\mathcal{T})$ is naturally isomorphic to the first-order deformation space of an étale correspondence of projective curves.

Proposition 8.2. *Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of curves over k without a core. Let \mathcal{L} be an invariant line bundle on the correspondence. Then $h^0(\mathcal{L}) \leq 1$.*

Proof. If there were two linearly independent sections $s = (s_X, s_Y)$ and $t = (t_X, t_Y)$, then by taking their ratio we get a map to \mathbb{P}^1 :



Hence there is a core. □

Question 8.3. Let $X \leftarrow Z \rightarrow Y$ be an étale correspondence of projective curves over k without a core:

- If $\text{char}(k) = 0$, is $h^1(\mathcal{T}) = 0$?
- If $\text{char}(k) = p$, what is the maximal possible value of $h^1(\mathcal{T})$ in terms of the genera?

Remark 8.4. As noted in Remark 3.11, in characteristic 0 étale correspondences of projective curves without a core $X \leftarrow Z \rightarrow Y$ do not globally deform. However, it is still perhaps possible that they deform up to finite order: inside of the moduli space \mathcal{M} of finite étale correspondences of projective curves with the given genera the moduli point $[X \leftarrow Z \rightarrow Y]$ is isolated by Corollary 4.13 but \mathcal{M} might not be smooth at this point. Example 3.19 shows that in characteristic p , étale correspondences of projective curves without a core may in fact globally deform.

We will see that, in characteristic 0, there are no invariant sections of any nontrivial invariant line bundle on an étale correspondence without a core. However, they can exist in characteristic p . To better understand this, we briefly review the *cyclic cover trick* for smooth curves. Let T be a smooth curve over k , let \mathcal{L}_T be a line bundle on T . Suppose $s \in H^0(T, \mathcal{L}_T^d)$ for some $d \in \mathbb{N}$ with $(d, \text{char } k) = 1$, such that s is not the power of a section of a smaller power of \mathcal{L}_T . Let \mathcal{A}_s denote the following sheaf of algebras

$$\mathcal{A}_s = \mathcal{O}_T \oplus \mathcal{L}_T^{-1} \oplus \dots \oplus \mathcal{L}_T^{-(d-1)}$$

with multiplication given by the naive multiplication when possible and contraction with s when necessary. The condition that s is not the power of a section implies that \mathcal{A}_s is an irreducible sheaf of algebras. We let $T(s^{1/d}) \rightarrow T$, the d -th *cyclic cover of T by s* , be the normalization of $\text{Spec}_T \mathcal{A}_s$ equipped with the natural map to T . Then $T(s^{1/d})$ is a smooth curve over k . The pullback of \mathcal{L}_T to $T(s^{1/d})$ has a (noncanonical) section, $s^{1/d}$, whose d -th power is s .

We remark that, by construction, the d -th cyclic cover of (T, s) is functorial. In particular, let \mathcal{L} be an invariant line bundle on $X \leftarrow Z \rightarrow Y$ with $s \in H^0(\mathcal{L}^d)$ an invariant section, and suppose that s is not the power of any invariant section of a smaller power of \mathcal{L} . Then we may perform the cyclic cover trick on $(X \leftarrow Z \rightarrow Y, s)$ to obtain

$$\begin{array}{ccc} & Z(s^{1/d}) & \\ & \swarrow \quad \searrow & \\ X(s^{1/d}) & & Y(s^{1/d}) \end{array}$$

The pullback of \mathcal{L} to $X(s^{1/d}) \rightarrow Z(s^{1/d}) \rightarrow Y(s^{1/d})$ has a (noncanonical) invariant section, which we denote by $s^{1/d}$.

Example 8.5. Consider a Hecke correspondence of (open) modular curves over \mathbb{F} . Then the Hasse invariant H yields an invariant section of an invariant line bundle; if $p > 2$, this invariant line bundle is $\Omega^{(p-1)/2}$. Recall that the divisor of H is the supersingular locus. The Hasse invariant similarly exists on a Hecke correspondence of moduli spaces of fake elliptic curves. Therefore, there are examples of invariant sections of nontrivial invariant line bundles on étale correspondences of projective curves over \mathbb{F} without a core.

In these cases, the “Igusa level structure” construction of Remark 3.18 is precisely the $(\frac{p-1}{2})$ -st cyclic cover construction associated to the invariant section H of $\Omega^{(p-1)/2}$. In particular, the induced correspondences of Igusa curves have an invariant differential form coming from “ $H^{2/(p-1)}$ ”. See [Ulmer

1990] for a brief introduction to the Hasse invariant and the Igusa construction and Chapter 1 of [Katz 1973] for a more thorough explication of modular forms.

Given a correspondence of projective curves, $X \leftarrow Z \rightarrow Y$, there are induced maps $f^*: \text{Pic}(X) \rightarrow \text{Pic}(Z)$ and $g^*: \text{Pic}(Y) \rightarrow \text{Pic}(Z)$ between the Picard schemes; both of these maps have finite (though not necessarily reduced) kernel. Restricting, there are induced maps $f^*: \text{Pic}^0(X) \rightarrow \text{Pic}^0(Z)$ and $g^*: \text{Pic}^0(Y) \rightarrow \text{Pic}^0(Z)$. We denote by $f^* \text{Pic}^0(X) \cap g^* \text{Pic}^0(Y)$ the scheme-theoretic intersection of the image of these two maps in $\text{Pic}^0(Z)$; note that this group scheme need not be reduced in positive characteristic.

Definition 8.6. Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of projective curves over k . The *Picard scheme of $X \leftarrow Z \rightarrow Y$* is the closed subgroup scheme of $\text{Pic}(X) \times \text{Pic}(Y)$ given by

$$\text{Pic}(X \leftarrow Z \rightarrow Y) := \ker(\text{Pic}(X) \times \text{Pic}(Y) \xrightarrow{f^* - g^*} \text{Pic}(Z)).$$

Similarly, $\text{Pic}^0(X \leftarrow Z \rightarrow Y) := \ker(\text{Pic}^0(X) \times \text{Pic}^0(Y) \xrightarrow{f^* - g^*} \text{Pic}^0(Z)).$

Remark 8.7. The scheme $\text{Pic}^0(X \leftarrow Z \rightarrow Y)$ need not be reduced in positive characteristic. As usual, if Z has a k -rational point, then $\text{Pic}(X \leftarrow Z \rightarrow Y)(k)$ is isomorphic the group of isomorphism classes of invariant line bundles on $X \leftarrow Z \rightarrow Y$. Finally, $\text{Pic}(X \leftarrow Z \rightarrow Y) / \text{Pic}^0(X \leftarrow Z \rightarrow Y) \hookrightarrow \mathbb{Z}$ via the degree map on Z .

We note that $\text{Pic}(X \leftarrow Z \rightarrow Y) \rightarrow \text{Pic}(X)$ and $\text{Pic}(X \leftarrow Z \rightarrow Y) \rightarrow \text{Pic}(Y)$ both have finite kernels. Moreover,

$$\text{Pic}^0(X \leftarrow Z \rightarrow Y) \rightarrow f^* \text{Pic}^0(X) \cap g^* \text{Pic}^0(Y) \subset \text{Pic}^0(Z)$$

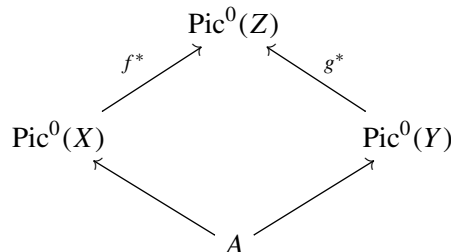
has finite kernel. The following theorem will be very useful for us.

Theorem 8.8. *Let A be an abelian variety over a field k and let $G \hookrightarrow A$ be a closed subgroup scheme. Then the connected reduced group subscheme $G_{\text{red}}^0 \hookrightarrow A$ is an abelian subvariety.*

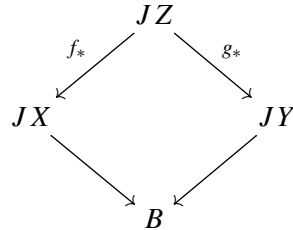
Proof. This is Proposition 5.31 in [van der Geer and Moonen 2008]. □

Lemma 8.9. *Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of projective curves over k without a core. Then $\text{Pic}(X \leftarrow Z \rightarrow Y)$ has no positive-dimensional abelian subvarieties. In particular, $\text{Pic}^0(X \leftarrow Z \rightarrow Y)$ is a finite group scheme over k .*

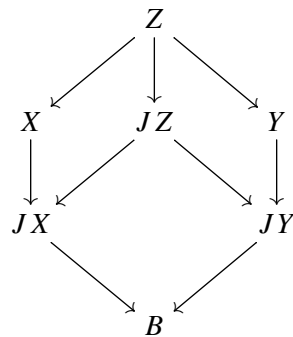
Proof. We may suppose all of the curves have genus at least 1. To prove the first statement, we show that there is no abelian variety A with finite maps fitting into the following diagram:



By dualizing, this is equivalent to showing that there is no abelian variety B with nonconstant surjective maps fitting into the following diagram:



(While $\text{Pic}^0(Z)$ is canonically principally polarized, we write the dual as JZ to remember the Albanese functoriality.) Suppose such a B fitting into the diagram existed. We will prove that the correspondence has a core. Choose a point $z \in Z(k)$ (extend k if necessary) and let $x = f(z)$ and $y = g(z)$. Then we have Abel–Jacobi maps which yield a morphism of *correspondences*:



i.e., the above diagram commutes. Moreover, under the Abel–Jacobi map, Z spans JZ as a group and likewise with X and Y . Therefore the induced maps $X \rightarrow B$ and $Y \rightarrow B$ are nonconstant. In particular, their image in B is a curve; therefore $X \leftarrow Z \rightarrow Y$ has a core.

We now prove that $\text{Pic}^0(X \leftarrow Z \rightarrow Y)$ is finite. If $\text{Pic}^0(X \leftarrow Z \rightarrow Y)$ were not finite, then it would be a positive-dimensional group subscheme of $\text{Pic}^0(X) \times \text{Pic}^0(Y)$. Then $A = \text{Pic}^0(X \leftarrow Z \rightarrow Y)_{\text{red}}^0$ is a closed, reduced, connected subgroup scheme of an abelian variety over k and is hence an abelian variety by Theorem 8.8. □

Corollary 8.10. *Let $X \leftarrow Z \rightarrow Y$ be an étale correspondence of projective hyperbolic curves over k without a core. Let \mathcal{L} be an invariant line bundle of positive degree. Then there exists $j, k \in \mathbb{N}$ with $\mathcal{L}^j \cong \Omega^k$.*

Proof. The set of degrees of invariant line bundles is a subgroup of \mathbb{Z} , so $\Omega^{-n} \otimes \mathcal{L}^m$ has degree 0 for some $m, n \in \mathbb{N}$. As our correspondence doesn't have a core, $\Omega^{-n} \otimes \mathcal{L}^m$ is torsion by Lemma 8.9. Therefore there exists $j, k \in \mathbb{N}$ such that $\mathcal{L}^j \cong \Omega^k$. □

Corollary 8.10 shows that, for étale correspondences of projective curves without a core, Ω plays

a special role. We will now see several striking consequences of Lemma 8.9 and Corollary 8.10 in characteristic 0.

Corollary 8.11. *Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of projective curves over k without a core. Suppose $\text{char}(k) = 0$. Then $f^*H^1(X, \mathcal{O}_X) \cap g^*H^1(Y, \mathcal{O}_Y) = 0$ inside of $H^1(Z, \mathcal{O}_Z)$ and $f^*H^0(X, \Omega_X^1) \cap g^*H^0(Y, \Omega_Y^1) = 0$ inside of $H^0(Z, \Omega_Z^1)$.*

Proof. The vector space $H^1(X, \mathcal{O}_X)$ is the tangent space at the identity of $\text{Pic}^0(X)$. Moreover, the vector space $f^*H^1(X, \mathcal{O}_X) \cap g^*H^1(Y, \mathcal{O}_Y)$ is the tangent space at the identity of $f^*\text{Pic}^0(X) \cap g^*\text{Pic}^0(Y)$, a closed subgroup of $\text{Pic}^0(Z)$. As the characteristic is 0, $f^*\text{Pic}^0(X) \cap g^*\text{Pic}^0(Y)$ is reduced and hence the connected component of the identity of $f^*\text{Pic}^0(X) \cap g^*\text{Pic}^0(Y)$ is an abelian variety. Lemma 8.9 implies that this abelian variety has dimension 0 and hence $f^*H^1(X, \mathcal{O}_X) \cap g^*H^1(Y, \mathcal{O}_Y) = 0$.

By the Lefschetz principle, we may suppose $k \cong \mathbb{C}$. If C is a smooth projective complex curve, $H_{\text{sing}}^1(C(\mathbb{C}), \mathbb{C}) \cong H^0(C, \Omega_C^1) \oplus H^1(C, \mathcal{O}_C)$ and $\overline{H^1(C, \mathcal{O}_C)} = H^0(C, \Omega_C^1)$ by Hodge symmetry. Therefore

$$f^*H^0(X, \Omega_X^1) \cap g^*H^0(Y, \Omega_Y^1) = \overline{f^*H^1(X, \mathcal{O}_X) \cap g^*H^1(Y, \mathcal{O}_Y)}$$

inside of $H_{\text{sing}}^1(Z(\mathbb{C}), \mathbb{C})$. The fact that $\dim f^*H^1(X, \mathcal{O}_X) \cap g^*H^1(Y, \mathcal{O}_Y) = 0$ implies the result. □

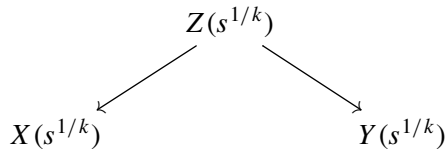
Corollary 8.12. *Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of projective curves over \mathbb{C} without a core. Then $f^*H_{\text{sing}}^1(X, \mathbb{Z}) \cap g^*H_{\text{sing}}^1(Y, \mathbb{Z}) = 0$ inside of $H_{\text{sing}}^1(Z, \mathbb{Z})$.*

Proof. This is immediate from Corollary 8.11 and the fact that pulling back H_{sing}^1 under f and g induces a morphism of integral Hodge structures. □

Corollary 8.13. *Let $X \leftarrow Z \rightarrow Y$ be an étale correspondence of projective curves over k without a core. Suppose $\text{char}(k) = 0$. Let \mathcal{L} be a nontrivial invariant line bundle. Then $h^0(\mathcal{L}) = 0$.*

Proof. We may suppose $\deg \mathcal{L} > 0$. Then there exists $j, k \in \mathbb{N}$ such that $\mathcal{L}^j \cong \Omega^k$ by Corollary 8.10. It therefore suffices to prove that no positive power of Ω has a section.

Suppose $s \in H^0(\Omega^k)$ is not the power of any smaller-degree invariant pluricanonical form on $X \leftarrow Z \rightarrow Y$. Then we may apply the cyclic-cover trick to obtain an étale correspondence



with an invariant differential form. This contradicts Corollary 8.11. □

We know, via the example of the Hasse invariant (see Example 8.5), that Corollary 8.13 is false in characteristic p . By examining the argument of Corollary 8.11, we see that the characteristic 0 hypothesis is used twice. First, we used that fact that all group schemes are reduced to argue that $h^1(\mathcal{O}) = 0$. Second, we used the Lefschetz principle and Hodge theory, namely $H^0(X, \Omega_X^1) = \overline{H^1(X, \mathcal{O}_X)}$, to relate $h^1(\mathcal{O})$ to $h^0(\Omega)$.

We further investigate the failure of Corollary 8.13 in characteristic p . To do this, we briefly recall a few facts about (commutative) finite group schemes. Let k be a field of characteristic p and let G be a finite group scheme over k . We denote by G^0 the connected component of the identity. There is the connected-étale sequence

$$1 \rightarrow G^0 \rightarrow G \rightarrow G^{\text{ét}} \rightarrow 1$$

which splits if k is perfect. The space of invariant differentials on G , a k -vector space denoted by $\omega_{G/k}$, may be identified with the cotangent space at the origin of G (see 3.14, 3.15 of [van der Geer and Moonen 2008]). We denote by $G[p]$ the p -torsion of G . Then the embedding $G[p] \hookrightarrow G$ induces an isomorphism on the level of (co)tangent spaces at the identity (e.g., see the proof of 4.47 of [loc. cit.])

Remark 8.14. The nomenclature “invariant differential” is slightly overloaded; we use this phrase to refer to a (left-) invariant differential form on a group scheme. When we use “invariant differential form”, we mean a section of $H^0(\Omega)$ on an étale correspondence. We trust that this is not too confusing.

We record the following fact, surely well known, for lack of a reference.

Lemma 8.15. *Let $f : A \rightarrow B$ be a surjective morphism of abelian varieties over a field k . Then f is separable if and only if the pullback map $f^* : H^0(B, \Omega_B^1) \rightarrow H^0(A, \Omega_A^1)$ is injective.*

Proof. Let $K = \ker(f)$ be the kernel of f . Then we have the following inclusion of group schemes

$$(K^0)_{\text{red}} \subset K^0 \subset K$$

Now, $(K^0)_{\text{red}}$ is a closed, reduced, connected subgroup scheme of an abelian variety over k ; hence it is an abelian variety by Theorem 8.8. Therefore $A/(K^0)_{\text{red}}$ exists as an abelian variety (see Section 9.5 of [Polishchuk 2003] or Example 4.40 in [van der Geer and Moonen 2008]). Similarly, A/K^0 , a quotient of $A/(K^0)_{\text{red}}$ by the finite group scheme $K^0/(K^0)_{\text{red}}$ exists as an abelian variety. We have the following commutative diagram:

$$\begin{array}{ccccc}
 K & \longrightarrow & A & \longrightarrow & B \\
 & & \searrow & & \uparrow \\
 & & & & A/K^0 \\
 & & & & \uparrow \\
 & & & & A/(K^0)_{\text{red}}
 \end{array}$$

where the right vertical arrows are isogenies. In particular $A/K^0 \rightarrow B$ is a separable isogeny and $A/(K^0)_{\text{red}} \rightarrow A/K^0$ is a purely inseparable isogeny. By looking at tangent spaces, we see K^0 is nonreduced if and only if the pullback map $H^0(B, \Omega_B^1) \rightarrow H^0(A/(K^0)_{\text{red}}, \Omega_{A/(K^0)_{\text{red}}}^1)$ is not injective. On the other hand the short exact sequence of abelian varieties over k

$$0 \rightarrow (K^0)_{\text{red}} \rightarrow A \rightarrow A/(K^0)_{\text{red}} \rightarrow 0$$

shows that the pullback map $H^0(A/(K^0)_{\text{red}}, \Omega_{A/(K^0)_{\text{red}}}^1) \rightarrow H^0(A, \Omega_A^1)$ is injective. Therefore $f^* : H^0(B, \Omega_B^1) \rightarrow H^0(A, \Omega_A^1)$ is injective if and only if K^0 is reduced, i.e., if and only if f is a separable morphism. \square

Corollary 8.16. *Let $f : C \rightarrow D$ be a generically separable, finite morphism of projective curves over k . Then $f_* : JC \rightarrow JD$ is separable.*

Proof. Choose an element c of $C(k)$ (after possibly extending k) and let $d = f(c)$. Then we have the following commutative diagram:

$$\begin{array}{ccc} C & \longrightarrow & JC \\ \downarrow & & \downarrow \\ D & \longrightarrow & JD \end{array}$$

where the horizontal arrows are the Abel–Jacobi maps associated to c and d respectively. Pulling back along these Abel–Jacobi maps yields isomorphisms $H^0(JC, \Omega_{JC}^1) \rightarrow H^0(C, \Omega_C^1)$ and $H^0(JD, \Omega_{JD}^1) \rightarrow H^0(D, \Omega_D^1)$, compatible with pulling back along f and f_* . As f was assumed to be generically separable, we obtain that

$$(f_*)^* : H^0(JD, \Omega_{JD}^1) \rightarrow H^0(JC, \Omega_{JC}^1)$$

is injective. Now apply Lemma 8.15. \square

Let k be a field of characteristic p and let $X \leftarrow Z \rightarrow Y$ be a correspondence of projective curves over k without a core. Suppose $\text{Pic}^0(X \leftarrow Z \rightarrow Y)$ is nontrivial. We have the following diagram:

$$\begin{array}{ccc} & \text{Pic}^0(Z) & \\ \nearrow & & \nwarrow \\ \text{Pic}^0(X) & & \text{Pic}^0(Y) \\ \nwarrow & & \nearrow \\ & \text{Pic}^0(X \leftarrow Z \rightarrow Y) & \end{array}$$

Let $G = \text{Pic}^0(X \leftarrow Z \rightarrow Y)[p]$. Take p -torsion and apply Cartier duality to obtain the diagram:

$$\begin{array}{ccc} & JZ[p] & \\ \swarrow & & \searrow \\ JX[p] & & JY[p] \\ \swarrow & & \searrow \\ & \check{G} & \end{array} \tag{8-1}$$

Now, if A is any abelian variety over k , the natural inclusion $A[p] \hookrightarrow A$ induces an isomorphism on the level of invariant differentials: $H^0(A, \Omega_A^1) \cong \omega_{A[p]/k}$. On the other hand, pulling back differential

1-forms under $f_* : JZ \rightarrow JX$ and $g_* : JZ \rightarrow JY$ is an injective operation by Corollary 8.16. Therefore pulling back invariant differentials is injective:

$$\omega_{JX[p]/k} \hookrightarrow \omega_{JZ[p]/k} \quad \text{and} \quad \omega_{JY[p]/k} \hookrightarrow \omega_{JZ[p]/k}.$$

Pick $z \in Z(k)$ and set $x = f(z)$ and $y = g(z)$. Then using the compatible Abel–Jacobi maps, we obtain that the following two vector spaces are isomorphic:

$$\{(\eta_X, \eta_Y) \in H^0(X, \Omega_X^1) \oplus H^0(Y, \Omega_Y^1) \mid f^*\eta_X = g^*\eta_Y\} \cong \{(s, t) \in \omega_{JX[p]/k} \oplus \omega_{JY[p]/k} \mid f^*s = g^*t\}.$$

Corollary 8.17. *Let $X \leftarrow Z \rightarrow Y$ be an étale correspondence of projective curves over k without a core. Then:*

- $h^0(\Omega) = \dim_k \{(s, t) \in \omega_{JX[p]/k} \oplus \omega_{JY[p]/k} \mid f^*s = g^*t\}$.
- If the map $T_e JX[p] \rightarrow T_e \check{G}$ is nonzero, then $h^0(\Omega) = 1$.
- The dimension of the image of $T_e JX[p] \rightarrow T_e \check{G}$ is no greater than 1.

Proof. The first part follows from the above discussion. If the map $T_e JX[p] \rightarrow T_e \check{G}$ is nonzero, then the pullback map $\omega_{\check{G}/k} \rightarrow \omega_{JX[p]/k} \hookrightarrow \omega_{JZ[p]/k}$ has nonzero image. By the commutativity of (8-1) there exists a pair $(s, t) \in \omega_{JX[p]/k} \oplus \omega_{JY[p]/k}$ such that the pullbacks to $\omega_{JZ[p]/k}$ agree. Hence there exists an invariant differential form on $X \leftarrow Z \rightarrow Y$. The dimension of the image of the map $\omega_{\check{G}/k} \rightarrow \omega_{JX[p]/k}$ is at most 1 because $h^0(\Omega) \leq 1$. In particular, the dimension of the image of $T_e JX[p] \rightarrow T_e \check{G}$ is at most 1. \square

Question 8.18. Let $X \leftarrow Z \rightarrow Y$ be an étale correspondence of projective curves over k without a core. Suppose $\text{char}(k) = p$. If $h^0(\Omega) = 1$, is the Cartier dual of $\text{Pic}^0(X \leftarrow Z \rightarrow Y)$ nonreduced?

9. Clumps

Definition 9.1. Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of curves over a field k . A *clump* S is a finite set of \bar{k} points $S \subset Z(\bar{k})$ such that $f^{-1}(f(S)) = g^{-1}(g(S)) = S$. An *étale clump* is a clump S such that f and g are étale at all points of S .

If $X \xleftarrow{f} Z \xrightarrow{g} Y$ has a core, then as in Remark 3.7 every $z \in Z(\bar{k})$ is contained in a clump. In the language of Remark 3.7, a clump is a *finite union of bounded orbits of geometric points*.

Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of curves over k of type (d, e) . Given an étale clump S , we now construct a natural invariant line bundle $\mathcal{L}(S)$ together with a one-dimensional subspace $V_S \subset H^0(\mathcal{L}(S))$ of invariant sections. (This line bundle may only be defined after a finite extension of k .) Think of S as an effective divisor on Z where all of the coefficients of the points are 1. Then f_*S is an effective divisor on X , all of whose coefficients are exactly d because f is étale at all points of S and has degree d . Therefore $\frac{1}{d}f_*S$ makes sense as an effective divisor on X ; it is the divisor associated to the finite set $f(S) \subset X$. The associated line bundle $\mathcal{L}_X(S)$ on X comes equipped with a natural one-dimensional space of sections $W_X \subset H^0(X, \mathcal{L}_X(S))$ with the following defining property: $\text{div}(w) = \frac{1}{d}f_*S$ for any $w \in W_X$.

Moreover, $f^*\mathcal{L}_X(S)$ is isomorphic to the line bundle associated with the divisor S . Similarly we obtain a line bundle $\mathcal{L}_Y(S)$ on Y with a natural one-dimensional space of sections W_Y . We set

$$\mathcal{L}(S) := (\mathcal{L}_X(S), \mathcal{L}_Y(S), \phi)$$

for any choice of isomorphism ϕ between the pullbacks. The vector space $H^0(\mathcal{L}(S))$ has a natural line V_S of invariant sections, given by f^*W_X and g^*W_Y ; in particular $h^0(\mathcal{L}(S)) \geq 1$.

Corollary 9.2. *Let $X \leftarrow Z \rightarrow Y$ be a étale correspondence of projective curves over k without a core. Suppose $\text{char } k = 0$. Then there are no clumps.*

Proof. A clump S is automatically étale and hence yields a nontrivial invariant line bundle $\mathcal{L}(S)$ such that $h^0(\mathcal{L}(S)) \geq 1$. This contradicts Corollary 8.13. \square

Remark 9.3. Corollary 9.2 shows that there is no direct analog of the supersingular locus in characteristic 0 for the following reason: Hecke orbits are big. This provides another conceptual reason why there is no canonical lift for supersingular elliptic curves.

Corollary 9.4. *A Hecke correspondence of compactified modular curves over \mathbb{C} is ramified at at least one of the cusps.*

Proof. The cusps are a clump. Hecke correspondences are unramified on open modular curves; if the compactified correspondence were unramified at all of the cusps, then the cusps would form a clump on an étale correspondence of projective curves without a core, contradicting Corollary 9.2. \square

Remark 9.5. The hypothesis of Corollary 9.2 implies that X , Y , and Z are Shimura curves by Theorem 3.10. This corollary was probably known, but we could not find a reference. Similarly, Corollary 9.4 admits a direct approach, but we find our method conceptually appealing.

Theorem 9.6. *Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be a correspondence of curves over a field k without a core. There is at most one étale clump.*

Proof. It is harmless to compactify the correspondence, so we assume X , Y , and Z are all projective. Suppose there were two étale clumps, S and T . As in the discussion above, they give rise to positive invariant line bundles $\mathcal{L}(S)$ and $\mathcal{L}(T)$ together with lines $V_S \subset H^0(\mathcal{L}(S))$ and $V_T \subset H^0(\mathcal{L}(T))$. There exists $m, n \in \mathbb{N}$ such that $\mathcal{L}(S)^m \otimes \mathcal{L}(T)^{-n}$ has degree 0. Lemma 8.9 implies that $\text{Pic}^0(X \leftarrow Z \rightarrow Y)$ is a finite group scheme over k ; in particular, $\mathcal{L}(S)^m \otimes \mathcal{L}(T)^{-n}$ is a torsion line bundle. Therefore there exists $j, k \in \mathbb{N}$ such that $\mathcal{L}(S)^j \cong \mathcal{L}(T)^k$.

The divisor of any element of $V_S^{\otimes j}$ is a positive multiple of S , and similarly the divisor of any element of $V_T^{\otimes k}$ is a positive multiple of T . In particular, if $S \neq T$, then the spaces $V_S^{\otimes j}$ and $V_T^{\otimes k}$ would be different lines inside of $H^0(\mathcal{L}(S)^j) \cong H^0(\mathcal{L}(T)^k)$. This would imply that $h^0(\mathcal{L}(S)^j) \geq 2$, contradicting Proposition 8.2. \square

Question 9.7. Let k be a field of characteristic p . Let $X \xleftarrow{f} Z \xrightarrow{g} Y$ be an étale correspondence of projective curves over k without a core. Is there always a clump? Equivalently, is there always an invariant pluricanonical differential form?

Remark 9.8. Theorem 9.6 generalizes the main theorem of [Hallouin and Perret 2014] (see the Introduction and Theorem 19 of that article), and the proof technique is completely different. In particular, those authors use the Perron–Frobenius theorem from spectral graph theory. We provide a detailed description of how to derive their result from ours.

Let $k \cong \mathbb{F}_q$ and let X be a smooth projective (geometrically irreducible) curve over k . Hallouin and Perret consider correspondences $\Gamma \subset X \times X$, with the assumption that Γ is absolutely irreducible and of type (d, d) . Let $\mathcal{T}(X, \Gamma)$ be the sequence of curves $(C_n)_{n \geq 1}$ defined as follows:

$$C_n = \{(P_1, P_2, \dots, P_n) \in X^n \mid (P_i, P_{i+1}) \in \Gamma \text{ for each } i = 1, \dots, n - 1\}.$$

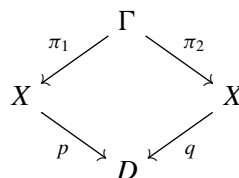
Let $\mathcal{G}_\infty(X, \Gamma)$, the *geometric graph*, be the graph whose vertices are the geometric points $X(\mathbb{F})$ and for which there is an oriented edge from $P \in X(\mathbb{F})$ to $Q \in X(\mathbb{F})$ if $(P, Q) \in \Gamma$. Theorem 19 of [loc. cit.] states that if the C_n are irreducible for all $n \geq 1$, then $\mathcal{G}_\infty(X, \Gamma)$ has *at most one* finite d -regular subgraph. As the correspondence is of type (d, d) , every finite d -regular subgraph of $\mathcal{G}_\infty(X, \Gamma)$ induces an étale clump $S_\Gamma \subset \Gamma(\mathbb{F})$ with the following “symmetry” property: $\pi_1(S_\Gamma) = \pi_2(S_\Gamma)$. We call S_Γ a *symmetric étale clump* and set $S_X = \pi_1(S_\Gamma) = \pi_2(S_\Gamma)$.

To understand their hypotheses, we first make the following definition. Let Ω be an algebraically closed field of transcendence degree 1 over k . Let $\mathcal{H}_{\text{gen}}^{\text{full}}$ be the following *directed graph*: the vertices are elements of $X(\Omega)$ and the edges are $\Gamma(\Omega)$. The source of an edge e is $\pi_1(e) \in X(\Omega)$ and the target of e is $\pi_2(e) \in X(\Omega)$. As usual, this graph is generally not connected and all connected components are isomorphic: we let \mathcal{H}_{gen} be any connected component. Every vertex of the graph \mathcal{H}_{gen} has in-degree and out-degree d . The hypothesis that C_n is irreducible for all n is equivalent to \mathcal{H}_{gen} having *no directed cycles*. Note that this implies, but is not equivalent to, \mathcal{H}_{gen} being infinite.

There is of course a surjective “collapsing” map $\mathcal{G}_{\text{gen}}^{\text{full}} \rightarrow \mathcal{H}_{\text{gen}}^{\text{full}}$ for a self correspondence $X \leftarrow \Gamma \rightarrow X$. One may make this a map of directed graphs by giving the following orientation to edges in the 2-colored graph $\mathcal{G}_{\text{gen}}^{\text{full}}$: an edge e between a blue vertex v and a red vertex w has the orientation $v \rightarrow w$. This map does *not necessarily* yield a surjective map $\mathcal{G}_{\text{gen}} \rightarrow \mathcal{H}_{\text{gen}}$; in particular, \mathcal{G}_{gen} can be finite with \mathcal{H}_{gen} infinite (e.g., see Elkies’ Example 9.9).

We now derive their result from ours. Let us assume, as they implicitly do, that \mathcal{H}_{gen} has no directed cycles. There are two possibilities: either $X \leftarrow \Gamma \rightarrow X$ has no core (i.e., \mathcal{G}_{gen} is infinite by Proposition 5.10), or $X \leftarrow \Gamma \rightarrow X$ has a core (i.e., \mathcal{G}_{gen} is finite by Proposition 5.10).

In the first case, Theorem 9.6 directly applies. In the second case, we will derive their theorem from ours. We first note that it is sufficient to prove the theorem after replacing Γ by its normalization, i.e., we may assume Γ is smooth. Call the coarse core D . We have the following diagram:



As D is the coarse core, Γ is the normalization of a component of $X \times_{p,D,q} X$. A symmetric étale clump S_Γ of $X \leftarrow \Gamma \rightarrow X$ yields unique étale clump S_X for the correspondence $D \leftarrow X \rightarrow D$. In particular, if we show that $D \leftarrow X \rightarrow D$ has at most one étale clump, we will have proven $X \leftarrow \Gamma \rightarrow X$ has at most one symmetric étale clump and we will have succeeded in deriving their theorem from ours.

We need only prove that $D \leftarrow X \rightarrow D$ has no core. This is where we use the irreducibility of all of the C_n . Note that C_n is birational to $\Gamma \times_{\pi_2, X, \pi_1} \Gamma \times \cdots \times_{\pi_2, X, \pi_1} \Gamma$ and $\lim_{n \rightarrow \infty} \deg(C_n \rightarrow D) = \infty$. On the other hand, Γ is birational to a component of $X \times_{p,D,q} X$. Therefore C_n is birational to an irreducible component

$$X \times_{p,D,q} X \times \cdots \times_{p,D,q} X$$

with increasing degree over D as $n \rightarrow \infty$. We now argue this cannot happen if $D \leftarrow X \rightarrow D$ had a core.

If $D \leftarrow X \rightarrow D$ has a core, we can find a curve $W \rightarrow X$ that is finite Galois over both compositions to D by Lemma 4.3. If E is any irreducible component of $X \times_{p,D,q} X \times \cdots \times_{p,D,q} X$, then

$$\deg(E \rightarrow D) \leq \deg(W \rightarrow D).$$

As the C_n are birational to irreducible of components of $X \times_{p,D,q} X \times \cdots \times_{p,D,q} X$ and $\deg(C_n \rightarrow D)$ goes to ∞ as $n \rightarrow \infty$, we see that $D \leftarrow X \rightarrow D$ has no core. Therefore Theorem 9.6 applies.

We remark that this argument only requires that there are components of C_n whose degree over D goes to ∞ as $n \rightarrow \infty$. In particular, we only need that \mathcal{H}_{gen} is an infinite graph.

Example 9.9. Consider the symmetric modular correspondence $Y(1) \leftarrow Y_0(2) \rightarrow Y(1)$ over \mathbb{F} . Then points of the form $\{(P_1, P_2, P_1) \mid (P_1, P_2) \in Y_0(2)\}$ are an irreducible component of C_3 . Therefore C_3 is not irreducible and their theorem does not directly apply. Note that \mathcal{G}_{gen} is a tree, by direct computation or Lemma 6.10. However, one can massage the correspondence, à la [Elkies 1997], to obtain the one-clump theorem for this correspondence using their method: it is equivalent to prove that there is only one clump for the correspondence

$$Y_0(2) \leftarrow Y_0(4) \rightarrow Y_0(2).$$

Here $Y_0(4)$ parametrizes pairs of elliptic curves equipped with a cyclic degree 4 isogeny between them $[E_1 \rightarrow E_2]$. This cyclic isogeny is uniquely the composition $E_1 \rightarrow E' \rightarrow E_2$, and the two maps to $Y_0(2)$ send this isogeny to $[E_1 \rightarrow E']$ and $[E' \rightarrow E_2]$ respectively. Note that this correspondence has a core: $Y(1)$, where $[E_1 \rightarrow E']$ and $[E' \rightarrow E_2]$ are both sent to $[E']$. Hallouin and Perret’s theorem applies to this correspondence. This correspondence has the property that \mathcal{G}_{gen} is finite (because there is a core) but \mathcal{H}_{gen} is infinite. For more details, see [Hallouin and Perret 2014] or Section 2.5 of [Krishnamoorthy 2016].

We describe a simple consequence of having a clump, providing a partial affirmative answer to Question 4.8.

Proposition 9.10. *Let $X \leftarrow Z \rightarrow Y$ be a correspondence of curves without a core with Z hyperbolic. If an étale clump exists, then the degree of the maximal “field of constants” of E_∞ is finite over k .*

Proof. If an étale clump exists, then all of the points of the clump are defined over a finite extension of fields k'/k . There are therefore k' -valued points of all of the curves $W_{YX\dots Y}$, as in Remark 7.3. This implies that all of the $W_{YX\dots Y}$ and hence W_∞ and E_∞ have field of constants contained in k' . The field of constants of E_∞ is then finite over k as desired. \square

Acknowledgments

This work is an extension of Chapter 2 of my PhD thesis at Columbia University. I am very grateful to Johan de Jong, my former thesis advisor, for guiding this project and for countless inspiring discussions. Ching-Li Chai read my thesis very carefully and provided many illuminating corrections and remarks, especially Example 3.19; I thank him. I also thank Aaron Bernstein, Ashwin Deopurkar, Remy van Dobben de Bruyn, H el ene Esnault, Ambrus Pal, and especially Philip Engel for interesting conversations on the topic of this article. Finally, I thank the referee, who read the article quite thoroughly and provided many helpful corrections, comments, and suggestions. During my time at Freie Universit at Berlin I have been funded by an NSF postdoctoral fellowship, Grant No. DMS-1605825.

References

- [Abe 2013] T. Abe, “Langlands correspondence for isocrystals and the existence of crystalline companions for curves”, preprint, 2013. arXiv
- [Borovoi 1982] M. V. Borovoi, “The Langlands conjecture on the conjugation of Shimura varieties”, *Funktsional. Anal. i Prilozhen.* **16**:4 (1982), 61–62. In Russian; translation in *Funct. Anal. Appl.* **16**:4 (1982), 292–294. MR
- [Boutot and Carayol 1991] J.-F. Boutot and H. Carayol, “Uniformisation p -adique des courbes de Shimura: les th eor emes de  Cerednik et de Drinfeld”, pp. 45–158 in *Courbes modulaires et courbes de Shimura* (Orsay, 1987/1988), Ast erisque **196-197**, Math. Soc. France, Paris, 1991. MR Zbl
- [Buzzard 1997] K. Buzzard, “Integral models of certain Shimura curves”, *Duke Math. J.* **87**:3 (1997), 591–612. MR Zbl
- [Chai 2005] C.-L. Chai, “Monodromy of Hecke-invariant subvarieties”, *Pure Appl. Math. Q.* **1**:2 (2005), 291–303. MR Zbl
- [Conrad et al. 2010] B. Conrad, O. Gabber, and G. Prasad, *Pseudo-reductive groups*, New Mathematical Monographs **17**, Cambridge Univ. Press, 2010. MR Zbl
- [Deligne 1979] P. Deligne, “Vari et es de Shimura: interpr etation modulaire, et techniques de construction de mod eles canoniques”, pp. 247–289 in *Automorphic forms, representations and L-functions, II* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, RI, 1979. MR Zbl
- [Djokovi c and Miller 1980] D.  . Djokovi c and G. L. Miller, “Regular groups of automorphisms of cubic graphs”, *J. Combin. Theory Ser. B* **29**:2 (1980), 195–230. MR Zbl
- [Elkies 1997] N. D. Elkies, “Explicit modular towers”, pp. 23–32 in *Proc. 35th Annual Allerton Conference on Communication, Control and Computing* (Urbana, IL, 1997), edited by T. Ba sar and A. Vardy, Univ. Illinois Urbana-Champaign, 1997. Zbl
- [van der Geer and Moonen 2008] G. van der Geer and B. Moonen, “Abelian varieties”, preliminary book draft, 2008, Available at <https://www.math.ru.nl/~bmoonen/research.html#bookabvar>.
- [Hallouin and Perret 2014] E. Hallouin and M. Perret, “Recursive towers of curves over finite fields using graph theory”, *Mosc. Math. J.* **14**:4 (2014), 773–806. MR Zbl
- [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
- [Kisin 2010] M. Kisin, “Integral models for Shimura varieties of abelian type”, *J. Amer. Math. Soc.* **23**:4 (2010), 967–1012. MR Zbl

- [Kohel 1996] D. R. Kohel, *Endomorphism rings of elliptic curves over finite fields*, Ph.D. thesis, University of California, Berkeley, 1996. Available at <https://search.proquest.com/docview/304241260>.
- [Krishnamoorthy 2016] S. Krishnamoorthy, *Dynamics, graph theory, and Barsotti–Tate groups: variations on a theme of Mochizuki*, Ph.D. thesis, Columbia University, 2016. Available at <https://search.proquest.com/docview/1783085609>.
- [Krishnamoorthy 2017] R. Krishnamoorthy, “Rank 2 local systems, Barsotti–Tate groups, and Shimura curves”, preprint, 2017. arXiv
- [Liu 2002] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Grad. Texts in Math. **6**, Oxford Univ. Press, 2002. MR Zbl
- [Margulis 1991] G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, *Ergebnisse der Mathematik* (3) **17**, Springer, 1991. MR Zbl
- [Milne 1983] J. S. Milne, “The action of an automorphism of \mathbb{C} on a Shimura variety and its special points”, pp. 239–265 in *Arithmetic and geometry, I*, edited by M. Artin and J. Tate, *Progr. Math.* **35**, Birkhäuser, Boston, 1983. MR Zbl
- [Milne and Shih 1982] J. S. Milne and K.-y. Shih, “Conjugates of Shimura varieties”, pp. 280–356 in *Hodge cycles, motives, and Shimura varieties*, *Lecture Notes in Math.* **900**, Springer, 1982. Zbl
- [Mochizuki 1998] S. Mochizuki, “Correspondences on hyperbolic curves”, *J. Pure Appl. Algebra* **131**:3 (1998), 227–244. MR Zbl
- [Mumford 1969] D. Mumford, “A note of Shimura’s paper ‘Discontinuous groups and abelian varieties’”, *Math. Ann.* **181**:4 (1969), 345–351. MR Zbl
- [Oort 2004] F. Oort, “Foliations in moduli spaces of abelian varieties”, *J. Amer. Math. Soc.* **17**:2 (2004), 267–296. MR Zbl
- [Polishchuk 2003] A. Polishchuk, *Abelian varieties, theta functions and the Fourier transform*, *Cambridge Tracts in Math.* **153**, Cambridge Univ. Press, 2003. MR Zbl
- [Serre 1977] J.-P. Serre, *Arbres, amalgames, SL_2* , *Astérisque* **46**, Soc. Math. France, Paris, 1977. Translated as *Trees*, Springer, 1980. MR Zbl
- [Sutherland 2013] A. V. Sutherland, “Isogeny volcanoes”, pp. 507–530 in *ANTS X: Proc. Tenth Algorithmic Number Theory Symposium* (San Diego, 2012), edited by E. W. Howe and K. S. Kedlaya, *Open Book Ser.* **1**, Math. Sci. Publ., Berkeley, CA, 2013. MR Zbl
- [Tutte 1966] W. T. Tutte, *Connectivity in graphs*, *Mathematical Expositions* **15**, Univ. Toronto Press, 1966. MR Zbl
- [Ulmer 1990] D. L. Ulmer, “On universal elliptic curves over Igusa curves”, *Invent. Math.* **99**:1 (1990), 377–391. MR Zbl
- [Xia 2013a] J. Xia, “Crystalline Hodge cycles and Shimura curves in positive characteristics”, preprint, 2013. arXiv
- [Xia 2013b] J. Xia, “On the deformation of a Barsotti–Tate group over a curve”, preprint, 2013. arXiv
- [Xia 2013c] J. Xia, “Tensor decomposition of isocrystals characterizes Mumford curves”, preprint, 2013. arXiv
- [Xia 2014] J. Xia, “ l -adic monodromy and Shimura curves in positive characteristics”, preprint, 2014. arXiv

Communicated by Bjorn Poonen

Received 2017-05-10 Revised 2018-01-16 Accepted 2018-03-29

raju@math.columbia.edu

Freie Universität Berlin, Germany

Local topological algebraicity with algebraic coefficients of analytic sets or functions

Guillaume Rond

We prove that any complex or real analytic set or function germ is topologically equivalent to a germ defined by polynomial equations whose coefficients are algebraic numbers.

The problem of the algebraicity of analytic sets or mappings is an old subject of study. It is known that the germ of a coherent analytic set with an isolated singularity is analytically equivalent to the germ of an algebraic set [Kucharz 1986; Tougeron 1976]. But in the general case the germ of an analytic set is not even locally diffeomorphic to the germ of an algebraic set [Whitney 1965]. On the other hand, considering a weaker equivalence relation, T. Mostowski [1984] proved that the germ of an analytic set is always homeomorphic to the germ of an algebraic set and this has been generalized to analytic function germs [Bilski et al. 2017]. For practical, effective and sometimes even theoretical purposes (for instance see [Budur and Wang 2017]) it is often not possible to handle coefficients that are transcendental numbers and so it is important to work with polynomial equations whose coefficients are rational or algebraic numbers. But it is well known that a small perturbation of the coefficients of polynomial equations defining an algebraic set germ or an algebraic function germ can drastically change the topology of the germ.

The goal of this paper is to extend the results of [Bilski et al. 2017] by proving that any complex or real analytic set or function germ is homeomorphic to an algebraic germ defined over the algebraic numbers. Our main result is the following one:

Theorem 1. *Let $\mathbb{K} = \mathbb{R}$ or \mathbb{C} . Let $(V, 0) \subset (\mathbb{K}^n, 0)$ be an analytic set germ and $g : (V, 0) \rightarrow (\mathbb{K}, 0)$ be an analytic function germ. Then there is a homeomorphism*

$$h : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^n, 0)$$

such that

- (i) $(h(V), 0)$ is the germ of an algebraic subset of \mathbb{K}^n defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$,
- (ii) $g \circ h^{-1}$ is the germ of a polynomial function defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$.

Moreover when we consider the particular case where there is no function germ g but only the set germ $(V, 0)$ we can be more precise about the nature of the homeomorphism:

MSC2010: primary 32S05; secondary 11G35, 13F25, 14B07, 32A05, 32B10, 32S15.

Keywords: local topological type, local algebraicity.

Theorem 2. *Let $\mathbb{K} = \mathbb{R}$ or \mathbb{C} . Let $(V, 0) \subset (\mathbb{K}^n, 0)$ be an analytic set germ. Then there is a homeomorphism $h : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^n, 0)$ such that*

- (i) *$h(V)$ is the germ of an algebraic subset of \mathbb{K}^n defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$,*
- (ii) *$(V, 0)$ is Whitney equisingular with $(h(V), 0)$,*
- (iii) *h is subanalytic and arcanalytic.*

The proof of our main result is based on the approach introduced in [Mostowski 1984] and extended in [Bilski et al. 2017]. For instance the idea to prove Theorem 2 in the case where $(V, 0)$ is a hypersurface germ is to use a version of the nested Artin–Płoski–Popescu approximation theorem, which we prove in this paper (see Theorem 12), in order to construct a regular Zariski equisingular deformation of $(V, 0)$ such that one of the fibers is the germ of a Nash hypersurface defined over $\overline{\mathbb{Q}}$. By a refinement of a theorem of Varčenko [1972] due to A. Parusiński and L. Păunescu [2017] such a deformation is a Whitney equisingular deformation so it is topologically trivial and the trivialization is subanalytic and arcanalytic. Then we use the Artin–Mazur theorem to transform our germ of a Nash set into the germ of an algebraic set (still defined over $\overline{\mathbb{Q}}$) by a local diffeomorphism. For Theorem 1 the idea is to apply essentially the same procedure to the graph of g , and the main difference concerns the part where we transform a Nash function germ into an algebraic function germ since the Artin–Mazur theorem is not sufficient to do this transformation. This part requires the construction of a particular deformation of the Nash set germ which is topologically trivial thanks to the Thom–Mather isotopy lemma. The paper is organized as follows: The first and main part is devoted to giving an algebraic statement concerning complex-coefficient algebraic power series that are solutions of algebraic equations with coefficients in $\overline{\mathbb{Q}}$. It shows that such solutions are \mathbb{C} -points of a family of algebraic solutions defined over $\overline{\mathbb{Q}}$ (see Theorem 7). In the next parts we apply this statement to prove Theorem 2 and then Theorem 1, essentially by proving that the approach used in [Bilski et al. 2017] remains valid in our situation.

Remark 3. Let us mention that B. Teissier [1990] provided an example of the germ of a complex algebraic surface in $(\mathbb{C}^3, 0)$ defined by a polynomial equation with coefficients in $\mathbb{Q}[\sqrt{5}]$ which is not Whitney equisingular to the germ of an algebraic set defined over \mathbb{Q} . So we cannot replace $\overline{\mathbb{Q}}$ by \mathbb{Q} in the statement of Theorem 2.

Remark 4. It is known that the germ of an analytic set is not always diffeomorphic to the germ of an algebraic set (see [Whitney 1965]). Let us mention that in general the germ of an algebraic set is neither diffeomorphic to the germ of an algebraic set defined over $\overline{\mathbb{Q}}$. For instance let us consider the germ of the curve $(V, 0) \subset (\mathbb{R}^2, 0)$ defined by the equation

$$xy(x - y)(x - \xi y),$$

where $\xi \in \mathbb{R}$ is a transcendental number. Indeed $(V, 0)$ is the union of four lines whose cross-ratio is ξ . If $(V, 0)$ were diffeomorphic to the germ of an algebraic set $(W, 0)$ defined over $\overline{\mathbb{Q}}$, the differential of the diffeomorphism germ would induce a bijective linear map between the tangent spaces at 0 of V and W .

Such a linear map preserves the cross-ratio, so the tangent space of $(W, 0)$ at 0 would be the union of four lines whose cross-ratio is equal to ξ and this would not be possible since $(W, 0)$ would be defined by algebraic equations with coefficients in $\overline{\mathbb{Q}}$. This example extends to the case where $\mathbb{K} = \mathbb{C}$ much as was done in [Bilski et al. 2017, Example 6.2].

Remark 5. In general for a given analytic map germ $g : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^m, 0)$ there is no germ of a homeomorphism $h : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^n, 0)$ such that $g \circ h$ is the germ of a polynomial map: just take $g : (\mathbb{K}, 0) \rightarrow (\mathbb{K}^2, 0)$ given by $g(x) = (x, e^x)$ (see [Bilski et al. 2017, Example 6.3]). In particular Theorem 1 cannot be extended to analytic map germs $(\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^m, 0)$. But in general, even if $g : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^m, 0)$ is the germ of a polynomial map there is no germ of a homeomorphism $h : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^n, 0)$ such that $g \circ h$ is the germ of a polynomial map defined over $\overline{\mathbb{Q}}$. Indeed let $g : (\mathbb{C}, 0) \rightarrow (\mathbb{C}^2, 0)$ be defined by $g(x) = (x, \xi x)$ where $\xi \in \mathbb{C}$ is a transcendental number. If there were a homeomorphism germ $h : (\mathbb{C}, 0) \rightarrow (\mathbb{C}, 0)$ such that $g \circ h$ is the germ of a polynomial map defined over $\overline{\mathbb{Q}}$ then both $h(x)$ and $\xi h(x)$ would be algebraic over $\overline{\mathbb{Q}}[x]$. But this would imply that ξ is algebraic over $\overline{\mathbb{Q}}$ which is not possible.

Remark 6. In Theorem 1 we do not know if the germ of a homeomorphism can be chosen to be arcanalytic or subanalytic. Indeed the proof of this result goes as follows: First we construct a Zariski equisingular deformation of the graphs of g and of the function germs defining $(V, 0)$ with the graphs of a Nash function germ \tilde{g} and of function germs defining the germ of a Nash set $(\tilde{V}, 0)$. Using the Artin–Mazur theorem we can reduce the situation to the case where $(\tilde{V}, 0)$ is the germ of an algebraic set and \tilde{g} is a unit u times a polynomial function germ P . Then, in [Bilski et al. 2017] a Thom stratification of the deformation

$$(t, x) \rightarrow (1 - t)u(0)P(x) + tu(x)P(x)$$

is constructed which shows (by the Thom–Mather isotopy lemma) that the function germ $\tilde{g} = uP$ is homeomorphic to the function germ P . But the Thom–Mather isotopy lemma does not provide an arcanalytic or subanalytic homeomorphism in general.

Notation and terminology. We will denote by x and y the vectors of indeterminates (x_1, \dots, x_n) and (y_1, \dots, y_m) . The notation x^i denotes the vector of indeterminates (x_1, \dots, x_i) for any $i \leq n$. When $\mathbb{K} = \mathbb{R}$ or \mathbb{C} , we denote by $\mathbb{K}\{x\}$ the ring of convergent power series with coefficients in \mathbb{K} , and by $\mathbb{K}\langle x \rangle$ the ring of algebraic power series with coefficients in \mathbb{K} . This means that $\mathbb{K}\langle x \rangle$ is the subring of $\mathbb{K}[[x]]$ whose elements are algebraic over $\mathbb{K}[x]$. We have $\mathbb{K}\langle x \rangle \subset \mathbb{K}\{x\}$; i.e., every algebraic power series is convergent.

Let $\mathbb{K} = \mathbb{C}$ or \mathbb{R} . Let Ω be an open subset of \mathbb{K}^n and let f be an analytic function on Ω . We say that f is a *Nash function* at $p \in \Omega$ if its Taylor expansion at p is an algebraic power series. An analytic function on Ω is a Nash function if it is a Nash function at every point of Ω . An analytic mapping $\varphi : \Omega \rightarrow \mathbb{K}^N$ is a *Nash mapping* if all its components are Nash functions on Ω . A subset X of Ω is called a *Nash subset* of Ω if for every $p \in \Omega$ there exist an open neighborhood U of p in Ω and Nash functions f_1, \dots, f_s on U such that $X \cap U = \{z \in U \mid f_1(z) = \dots = f_s(z) = 0\}$. A germ X_p of a set X at $p \in \Omega$ is a *Nash germ* if

there exists an open neighborhood U of p in Ω such that $X \cap U$ is a Nash subset of U . A Nash function germ is said to be defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$ if it satisfies a nontrivial polynomial equation with coefficients in $\overline{\mathbb{Q}} \cap \mathbb{K}$. This is equivalent to saying that its Taylor expansion at a $\overline{\mathbb{Q}} \cap \mathbb{K}$ -point is an algebraic power series whose coefficients are in $\overline{\mathbb{Q}} \cap \mathbb{K}$, i.e., an element of $(\overline{\mathbb{Q}} \cap \mathbb{K})[[x]]$. A Nash set is said to be defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$ if it is locally defined by Nash function germs defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$.

1. An approximation result

We begin by stating the main result of this part:

Theorem 7. *Let $f(x, y) \in \overline{\mathbb{Q}}\langle x \rangle[y]^p$ and let us consider a solution $y(x) \in \mathbb{C}\langle x \rangle^m$ of*

$$f(x, y(x)) = 0.$$

Then there exist a new set of indeterminates $t = (t_1, \dots, t_r)$, a vector of algebraic power series

$$y(t, x) = \sum_{\alpha \in \mathbb{N}^n} y_\alpha(t) x^\alpha \in \overline{\mathbb{Q}}\langle t, x \rangle^m$$

and $\mathbf{t} = (t_1, \dots, t_r) \in \mathbb{C}^r$ belonging to the domain of convergence of all the $y_\alpha(t)$ such that

$$y(x) = y(\mathbf{t}, x) \quad \text{and} \quad f(x, y(\mathbf{t}, x)) = 0.$$

Remark 8. This theorem is not true if we replace $\overline{\mathbb{Q}}$ by \mathbb{Q} . For instance let x and y be single indeterminates and set $f = y^2 - 2x^2$. Then there is no algebraic power series $y(x, t) \in \mathbb{Q}\langle x, t \rangle$ such that

$$y(x, t)^2 - 2x^2 = 0$$

but we have

$$f(x, \sqrt{2}x) = 0.$$

Proof of Theorem 7. If $y(x) \in \overline{\mathbb{Q}}\langle x \rangle^m$ then we take $r = 0$ and there is nothing to prove. Let us assume that $y(x) \in \mathbb{C}\langle x \rangle^m \setminus \overline{\mathbb{Q}}\langle x \rangle^m$. By Lemma 10 given below we may assume that there exist $y'(t, u, v, x) \in \overline{\mathbb{Q}}\langle t, u, v, x \rangle^m \cap \overline{\mathbb{Q}}[t, u, v][[x]]^m$, where $t = (t_1, \dots, t_r)$ and u and v are single indeterminates, and $\mathbf{t} \in \mathbb{C}^r$, $\mathbf{u} \in \mathbb{C}$, $\mathbf{v} \in \mathbb{C}$ such that

$$y(x) = y'(\mathbf{t}, \mathbf{u}, \mathbf{v}, x). \tag{1}$$

Moreover we may assume that t_1, \dots, t_r are algebraically independent over $\overline{\mathbb{Q}}$, $\mathbf{u} = 1/R(t_1, \dots, t_r)$ for some polynomial $R \in \overline{\mathbb{Q}}[t_1, \dots, t_r]$ such that $R(t_1, \dots, t_r) \neq 0$, and \mathbf{v} is finite over $\mathbb{L} := \overline{\mathbb{Q}}(t_1, \dots, t_r)$. Let $P(t_1, \dots, t_r, v) \in \overline{\mathbb{Q}}(t)[v]$ be the monic polynomial of minimal degree in v such that

$$P(t_1, \dots, t_r, \mathbf{v}) = 0.$$

Let $D \subset \mathbb{C}^r$ be the discriminant locus of $P(t, v)$ seen as a polynomial in v (i.e., D is the locus of points $q \in \mathbb{C}^r$ such that q is a pole of one of the coefficients of P or such that $P(q, v)$ has at least one multiple

root). Since $P(t_1, \dots, t_r, v)$ has no multiple roots in an algebraic closure of \mathbb{L} , the point t is not in D . Then there exist $\mathcal{U} \subset \mathbb{C}^r \setminus D$ a simply connected open neighborhood of p and analytic functions

$$w_i : \mathcal{U} \rightarrow \mathbb{C}, \quad i = 1, \dots, d,$$

such that

$$P(t, v) = \prod_{i=1}^d (v - w_i(t))$$

and $w_1(t_1, \dots, t_r) = v$. Moreover the $t \mapsto w_i(t)$ are algebraic functions over $\overline{\mathbb{Q}}[t]$. In particular the Taylor series of w_1 at a point of $\mathcal{U} \cap \overline{\mathbb{Q}}^r$ is an algebraic power series with algebraic coefficients. Since the polynomial R is not vanishing at p the function

$$t \in \mathbb{C}^r \setminus \{R = 0\} \mapsto \frac{1}{R(t)}$$

is also an analytic function which is algebraic over $\overline{\mathbb{Q}}[t]$ and so its Taylor series at a point of $\overline{\mathbb{Q}}^r \setminus \{R = 0\}$ is an algebraic power series with algebraic coefficients. Let $q := (q_1, \dots, q_r) \in \overline{\mathbb{Q}}^r \cap \mathcal{U} \setminus \{R = 0\}$ such that t belongs to an open polydisc Δ centered at q and such that $\Delta \subset \mathcal{U} \setminus \{R = 0\}$. We denote by $\varphi_1(t)$ and $\varphi_2(t) \in \overline{\mathbb{Q}}\langle t \rangle$ the Taylor series of $t \mapsto 1/R(t)$ and w_1 at q . For simplicity we can make a translation and assume that q is the origin of \mathbb{C}^r . In particular the series $\varphi_1(t)$ and $\varphi_2(t)$ are convergent at t . We have

$$f(x, y'(t_1, \dots, t_r, u, v, x)) = 0$$

or equivalently

$$f(x, y'(t_1, \dots, t_r, \varphi_1(t_1, \dots, t_r), \varphi_2(t_1, \dots, t_r), x)) = 0.$$

The function

$$(t, x) \mapsto F(t, x) := f(x, y'(t_1, \dots, t_r, \varphi_1(t_1, \dots, t_r), \varphi_2(t_1, \dots, t_r), x))$$

is an algebraic function over $\overline{\mathbb{Q}}[t, x]$. So if $F(t, x) \not\equiv 0$ there exists an algebraic function $(t, x) \mapsto g(t, x)$ such that

$$(t, x) \mapsto g(t, x)F(t, x)$$

is a nonzero polynomial function. Indeed if

$$a_0(t, x)T^e + a_1(t, x)T^{e-1} + \dots + a_e(t, x)$$

is a polynomial of minimal degree having $F(t, x)$ as a root then $a_e(t, x) \not\equiv 0$ and we can choose

$$g(t, x) := -a_0(t, x)F(t, x)^{e-1} - a_1(t, x)F(t, x)^{e-2} + \dots - a_{e-1}(t, x)$$

so we have

$$g(t, x)F(t, x) = a_e(t, x).$$

Since $F(t, x) = 0$ we have $a_e(t, x) = 0$ but t_1, \dots, t_r, x being algebraically independent over $\overline{\mathbb{Q}}$ we obtain that $a_e(t, x) \equiv 0$ which is a contradiction. Thus we have

$$F(t, x) = f(x, y'(t_1, \dots, t_r, \varphi_1(t_1, \dots, t_r), \varphi_2(t_1, \dots, t_r), x)) = 0.$$

This proves the theorem if we define

$$y(t, x) = y'(t, \varphi_1(t), \varphi_2(t), x).$$

All the series $y_\alpha(t)$ are then convergent at t , since $\varphi_1(t)$ and $\varphi_2(t)$ are convergent power series at t and since $y'(t, u, v, x) \in \overline{\mathbb{Q}}\langle t, u, v, x \rangle^m \cap \overline{\mathbb{Q}}[t, u, v][[x]]^m$. \square

Remark 9. Let us assume that $f(x, y) \in \mathbb{Q}\langle x \rangle[y]^p$. In the proof of Theorem 7 let us assume that $r = 0$, i.e., the coefficients of $y(x)$ belong to a finite field extension of \mathbb{Q} . In this case the analytic function w_1 is a constant function whose value is in $\overline{\mathbb{Q}} \setminus \mathbb{Q}$. This is why we need to work with the algebraically closed field $\overline{\mathbb{Q}}$ and not only with \mathbb{Q} .

Lemma 10. *Let $f \in \mathbb{C}\langle x \rangle^m \setminus \overline{\mathbb{Q}}\langle x \rangle^m$. Then there exist complex numbers t_1, \dots, t_r, u and v with $r \geq 1$ and $F \in \overline{\mathbb{Q}}\langle t, u, v, x \rangle^m$, where $t = (t_1, \dots, t_r)$ and u and v are single indeterminates, such that*

- $F \in \overline{\mathbb{Q}}[t, u, v][[x]]^m$,
- $f(x) = F(t_1, \dots, t_r, u, v, x)$,
- the extension $\overline{\mathbb{Q}} \rightarrow \overline{\mathbb{Q}}(t_1, \dots, t_r)$ is purely transcendental,
- $u = 1/R(t_1, \dots, t_r)$ for some polynomial $R \in \overline{\mathbb{Q}}[t_1, \dots, t_r]$ with $R(t_1, \dots, t_r) \neq 0$,
- v is finite over $\overline{\mathbb{Q}}(t_1, \dots, t_r)$.

Proof. Let \mathbb{K} be the field extension of $\overline{\mathbb{Q}}$ generated by the coefficients of the minimal polynomials of the components of f . Then the coefficients of the components of f belong to a finite field extension of \mathbb{K} (see for instance [Cutkosky and Kashcheyeva 2008]). Let us replace \mathbb{K} by this finite field extension. There exists a purely transcendental finitely generated field extension $\overline{\mathbb{Q}} \rightarrow \mathbb{L}$ such that $\mathbb{L} \rightarrow \mathbb{K}$ is finite. By enlarging \mathbb{K} we may assume that $\mathbb{L} \rightarrow \mathbb{K}$ is normal. By the primitive element theorem $\mathbb{K} = \mathbb{L}(a)$ for some $a \in \mathbb{C}$ algebraic over \mathbb{L} . Let us write $f(x) = (f_1(x), \dots, f_m(x))$. We can write

$$f_i(x) = \sum_{k=0}^{d-1} a^k f_{i,k}(x) \quad \text{for } i = 1, \dots, m,$$

where d is the degree of a over \mathbb{L} and the $f_{i,k}(x)$ are power series with coefficients in \mathbb{L} . Let us denote by

$$a_1 = a, a_2, \dots, a_d$$

the conjugates of a over \mathbb{L} . Since $f_i(x)$ is algebraic over $\mathbb{K}[x]$ and $\mathbb{L} \rightarrow \mathbb{K}$ is an algebraic extension, $f_i(x)$ is algebraic over $\mathbb{L}[x]$. Let $P_i(x, y) = \sum_{\alpha, l} p_{\alpha, l} x^\alpha y^l \in \mathbb{L}[x, y]$ be a nonzero vanishing polynomial of $f_i(x)$ and let σ be a \mathbb{L} -automorphism of \mathbb{K} such that $\sigma(a) = a_j$ for some j . It induces a $\mathbb{L}[[x]]$ -automorphism of

$\mathbb{K}[[x]]$ defined by $\sigma(\sum_{\alpha} c_{\alpha}x^{\alpha}) = \sum_{\alpha} \sigma(c_{\alpha})x^{\alpha}$. Then we have

$$0 = \sigma(P_i(x, f_i(x))) = \sum_{\alpha,l} \sigma(p_{\alpha,l})x^{\alpha} \sigma(f_i(x))^l = \sum_{\alpha,l} \sigma(p_{\alpha,l})x^{\alpha} \left(\sum_{k=0}^{d-1} a_j^k f_{i,k}(x) \right)^l.$$

Thus for every $i = 1, \dots, m$ and $j = 1, \dots, d$ the power series

$$\sum_{k=0}^{d-1} a_j^k f_{i,k}(x) \in \mathbb{K}[[x]]$$

is algebraic over $\mathbb{K}[x]$. Let M be the (nonsingular) $d \times d$ Vandermonde matrix associated to the a_j . Then we have

$$\tilde{f}_i(x) = M \bar{f}_i(x),$$

where $\tilde{f}_i(x)$ is the vector whose entries are the $\sum_{k=0}^{d-1} a_j^k f_{i,k}(x)$ for $j = 1, \dots, d$, and $\bar{f}_i(x)$ is the vector whose entries are the $f_{i,k}(x)$. Then $\bar{f}_i(x) = M^{-1} \tilde{f}_i(x)$; thus the $f_{i,k}(x)$ are algebraic over $\mathbb{K}[x]$ and so over $\mathbb{L}[x]$. This shows that $f_{i,k}(x) \in \mathbb{L}\langle x \rangle$ for every i and k .

Let t_1, \dots, t_r be a transcendence basis of $\mathbb{L}/\overline{\mathbb{Q}}$. Then by Lemma 11 given below we have

$$f_{i,k} = \sum_{\alpha \in \mathbb{N}^n} \frac{S_{i,k,\alpha}(t_1, \dots, t_r)}{R_{i,k}(t_1, \dots, t_r)^{|\alpha|}} x^{\alpha}$$

for some polynomials $S_{i,k,\alpha}$ and $R_{i,k} \in \overline{\mathbb{Q}}[t_1, \dots, t_r]$. By replacing each $R_{i,k}$ by $\prod_{j,l} R_{j,l}$ and multiplying every $S_{i,k,\alpha}$ by $\prod_{(j,l) \neq (i,k)} R_{j,l}^{|\alpha|}$ we may assume that $R_{i,k} = R_{i',k'} = R$ for every (i, k) and (i', k') . The power series

$$f_{i,k}^*(x) = f_{i,k}(R(t_1, \dots, t_r)x_1, \dots, R(t_1, \dots, t_r)x_n) = \sum_{\alpha \in \mathbb{N}^n} S_{i,k,\alpha}(t_1, \dots, t_r) x^{\alpha}$$

belongs to $\overline{\mathbb{Q}}(t_1, \dots, t_r)\langle x \rangle$ since $f_{i,k}(x) \in \overline{\mathbb{Q}}(t_1, \dots, t_r)\langle x \rangle$. Thus we have

$$f_{i,k}^* = F_{i,k}(t_1, \dots, t_r, x)$$

with, for every i and k ,

$$F_{i,k} := \sum_{\alpha \in \mathbb{N}^n} S_{k,\alpha}(t_1, \dots, t_r) x^{\alpha} \in \overline{\mathbb{Q}}[t_1, \dots, t_r][[x]]$$

where the t_i are new indeterminates. Moreover let $P_{i,k}(t_1, \dots, t_r, x, y) \in \overline{\mathbb{Q}}[t_1, \dots, t_r, x, y]$, where y is a new indeterminate, be a nonzero polynomial with $P_{i,k}(t, x, f_{i,k}^*(x)) = 0$. Since $F_{i,k} \in \overline{\mathbb{Q}}[t_1, \dots, t_r][[x]]$ for every k , we can write

$$P_{i,k}(t_1, \dots, t_r, x, F_{i,k}(t_1, \dots, t_r, x)) = \sum_{\beta \in \mathbb{N}^n} P_{i,k,l}(t_1, \dots, t_r) x^{\beta}$$

for some polynomials $P_{i,k,\beta} \in \overline{\mathbb{Q}}[t_1, \dots, t_r]$. Thus $P_{i,k,\beta}(t_1, \dots, t_r) = 0$ for every i, k and β , but since t_1, \dots, t_r are algebraically independent over $\overline{\mathbb{Q}}$ we have

$$P_{i,k,\beta}(t_1, \dots, t_r) = 0$$

for every i, k and β so

$$P_{i,k}(t_1, \dots, t_r, x, F_{i,k}(t_1, \dots, t_r, x)) = 0$$

and this implies that $F_{i,k} \in \overline{\mathbb{Q}}(t_1, \dots, t_r, x)$. In particular if u denotes a new indeterminate we have

$$F_{i,k}(t_1, \dots, t_r, ux_1, \dots, ux_n) \in \overline{\mathbb{Q}}(t_1, \dots, t_r, u, x) \cap \overline{\mathbb{Q}}[t_1, \dots, t_r, u][[x]] \quad \text{for all } k.$$

Finally we set $t = (t_1, \dots, t_r)$ and

$$F_i(t, x) = \sum_{k=0}^{d-1} v^k F_{i,k}(t_1, \dots, t_r, ux_1, \dots, ux_n)$$

where v denotes a new indeterminate. Thus the result is proven with F the vector whose components are the F_i and $u = 1/R(t_1, \dots, t_r)$ and $v = a$. □

The following version of Eisenstein lemma is essentially [Tougeron 1990, Lemma 2.2] and the proof is the same — but we give it here for the convenience of the reader:

Lemma 11 (Eisenstein lemma). *Let $f \in \overline{\mathbb{Q}}(t_1, \dots, t_r)(x)$ be an algebraic power series where the $t_i \in \mathbb{C}$ are algebraically independent over $\overline{\mathbb{Q}}$. Then there exist a polynomial $R(t) \in \overline{\mathbb{Q}}[t]$ and polynomials $S_\alpha(t) \in \overline{\mathbb{Q}}[t]$ for every $\alpha \in \mathbb{N}^n$, where $t = (t_1, \dots, t_r)$ is a vector of new indeterminates, such that*

$$f(x) = \sum_{\alpha \in \mathbb{N}^n} \frac{S_\alpha(t_1, \dots, t_r)}{R(t_1, \dots, t_r)^{|\alpha|}} x^\alpha.$$

Proof of Lemma 11. Let $P(x, y) \in \overline{\mathbb{Q}}(t_1, \dots, t_r)[x, y]$ be a minimal polynomial of f , i.e., a generator of the kernel of the ring morphism:

$$\begin{aligned} \overline{\mathbb{Q}}(t_1, \dots, t_r)[x, y] &\rightarrow \overline{\mathbb{Q}}(t_1, \dots, t_r)[[x]], \\ p(x, y) &\mapsto p(x, f(x)). \end{aligned}$$

Let us set

$$e := \text{ord}_x \left(\frac{\partial P}{\partial y}(x, f(x)) \right).$$

We have $e < \infty$ since $P(x, y)$ is a minimal polynomial of $f(x)$. Let us write $f = \sum_{\alpha \in \mathbb{N}^n} f_\alpha(t)x^\alpha$, where $t = (t_1, \dots, t_r)$ and $f_\alpha(t) \in \overline{\mathbb{Q}}(t)$. Let $b(t) \in \overline{\mathbb{Q}}[t]$ be a common denominator of the $f_\alpha(t)$ for $|\alpha| \leq 2e + 1$. Then Lemma 11 is satisfied by f if and only if it is satisfied by $b(t)f$. Thus we may replace f by $b(t)f$. In this case a minimal polynomial of $b(t)f$ is

$$P'(x, y) := b(t)^{\text{deg}_y(P)} P \left(x, \frac{y}{b(t)} \right).$$

Moreover by multiplying $P'(x, y)$ by an element of $\overline{\mathbb{Q}}(t_1, \dots, t_r)$ we may assume that $P'(x, y) \in \overline{\mathbb{Q}}[t_1, \dots, t_r][x, y]$. Then we have

$$e = \text{ord}_x \left(\frac{\partial P'}{\partial y}(x, b(t)f(x)) \right).$$

Thus we may replace f by $b(t)f$ and assume that $f_\alpha(t) \in \overline{\mathbb{Q}}[t]$ for $|\alpha| \leq 2e + 1$.

We define

$$P^*(u, x, y) := P(ux_1, \dots, ux_n, y) \in \overline{\mathbb{Q}}[t_1, \dots, t_r, x_1, \dots, x_n][u, y]$$

and

$$f^*(u, x) := f(ux_1, \dots, ux_n)$$

where u is a new indeterminate. Then $P^*(u, x, f^*(u, x)) = 0$ so $f^* \in \overline{\mathbb{Q}}(t, x)\langle u \rangle$. Let us denote by $f^{*(2e+1)}(u)$ the $(2e + 1)$ -truncation of $f^*(u)$ (i.e., we remove from $f^*(u)$ all the monomials which are divisible by u^{2e+2}). Then

$$P^*(u, x, f^{*(2e+1)}) \in (u)^{2e+2} \quad \text{and} \quad \frac{\partial P^*}{\partial y}(u, x, f^{*(2e+1)}) \in (u)^e \setminus (u)^{e+1}.$$

Let us set

$$y = u^{e+1}y' + f^{*(2e+1)}$$

where y' is a new indeterminate. Then

$$P^*(u, x, y) = P^*(u, x, f^{*(2e+1)}) + \frac{\partial P^*}{\partial y}(u, x, f^{*(2e+1)})u^{e+1}y' + u^{2e+2}y'^2Q(u, y')$$

for some polynomial Q . Thus the equation $P^*(u, x, y) = 0$ is equivalent to

$$\frac{P^*(u, x, f^{*(2e+1)})}{u^{2e+1}} + \frac{\frac{\partial P^*}{\partial y}(u, x, f^{*(2e+1)})}{u^e}y' + uy'^2Q(u, y') = 0. \tag{2}$$

Since $f_\alpha(t) \in \overline{\mathbb{Q}}[t]$ for $|\alpha| \leq 2e + 1$ we have $f^{*(2e+1)} \in \overline{\mathbb{Q}}[t, x, u]$. Since the coefficients in (2) are polynomials in u, x and the $f_\alpha(t)$ for $|\alpha| \leq 2e + 1$, they belong to $\overline{\mathbb{Q}}[t, x, u]$. Let $R(t, x) \in \overline{\mathbb{Q}}[t, x]$ be defined by

$$R(t, x) = \left(\frac{\frac{\partial P^*}{\partial y}(u, x, f^{*(2e+1)})}{u^e} \right)_{|u=0}.$$

Since $\text{ord}_u(P^*(u, x, f^{*(2e+1)})/u^{2e+1}) \geq 1$ we see that $R(t, x)^2$ divides

$$\frac{P^*(R(t, x)^2u', x, f^{*(2e+1)}(R(t, x)^2u'))}{(R(t, x)^2u')^{2e+1}}$$

where u' is a new indeterminate. Thus by replacing y' by $R(t, x)y''$ and u by $R(t, x)^2u'$ in (2), and dividing by $R(t, x)^2$ we conclude that (2) is equivalent to

$$A_1(t, x, u') + (1 + u'A_2(t, x, u'))y'' + u'y''^2A_3(t, x, u') = 0, \tag{3}$$

where the A_i belongs to $\overline{\mathbb{Q}}[\mathbf{t}, x, u']$. By the implicit function theorem (or Hensel’s lemma) this equation has a unique solution in $\overline{\mathbb{Q}}\langle \mathbf{t}, x, u' \rangle$ which is necessarily $g^* := f^*(R(\mathbf{t}, x)^2 u')/R(\mathbf{t}, x)$. Moreover we can also apply Hensel’s lemma to this equation to see that it has a unique solution in the completion of $\overline{\mathbb{Q}}[\mathbf{t}, x, u']$ with respect to the ideal generated by u' , i.e., in the ring $\overline{\mathbb{Q}}[\mathbf{t}, x][[u']]$. Thus

$$g^* \in \overline{\mathbb{Q}}[\mathbf{t}, x][[u']] \cap \overline{\mathbb{Q}}\langle \mathbf{t}, x, u' \rangle.$$

In particular the coefficients g_k^* defined by $g^*(u') = \sum_{k \geq 0} g_k^* u'^k$ are polynomials over $\overline{\mathbb{Q}}$ depending on the t_i and the x_j . Moreover

$$f^*(u) = \sum_{k \geq 0} \frac{g_k^*(\mathbf{t}, x)}{R(\mathbf{t}, x)^{2k-1}} u^k.$$

On the other hand we have

$$f^*(u) = \sum_{k \geq 0} \left(\sum_{|\alpha|=k} f_\alpha(\mathbf{t}) x^\alpha \right) u^k;$$

hence

$$\sum_{|\alpha|=k} f_\alpha(\mathbf{t}) x^\alpha = \frac{g_k^*(\mathbf{t}, x)}{R(\mathbf{t}, x)^{2k-1}} \quad \text{for all } k \in \mathbb{N}. \tag{4}$$

For every $\alpha \in \mathbb{N}^n$, let us write $f_\alpha(\mathbf{t}) = h_\alpha(\mathbf{t})/l_{|\alpha|}(\mathbf{t})$ where $h_\alpha(\mathbf{t}), l_{|\alpha|}(\mathbf{t}) \in \overline{\mathbb{Q}}[\mathbf{t}]$ and $l_{|\alpha|}(\mathbf{t})$ is coprime with $\sum_{|\alpha|=k} h_\alpha(\mathbf{t}) x^\alpha$. Then $l_{|\alpha|}(\mathbf{t})$ divides $R(\mathbf{t}, x)^{2|\alpha|-1}$. Let $r(\mathbf{t})$ be the greatest divisor of $R(\mathbf{t}, x)$ belonging to $\overline{\mathbb{Q}}[\mathbf{t}]$. Then there exists $d_{|\alpha|}(\mathbf{t}) \in \overline{\mathbb{Q}}[\mathbf{t}]$ such that $l_{|\alpha|}(\mathbf{t}) d_{|\alpha|}(\mathbf{t}) = r(\mathbf{t})^{2|\alpha|-1}$. Thus

$$f_\alpha(\mathbf{t}) = \frac{h_\alpha(\mathbf{t}) d_{|\alpha|}(\mathbf{t})}{r(\mathbf{t})^{2|\alpha|-1}}.$$

This proves the lemma. □

Theorem 7 allows us to prove the following version of the nested Artin–Płoski–Popescu approximation theorem:

Theorem 12. *Let $f(x, y) \in \overline{\mathbb{Q}}\langle x \rangle[y]^p$ and let us consider a solution $y(x) \in \mathbb{C}\{x\}^m$ of*

$$f(x, y(x)) = 0.$$

Let us assume that $y_i(x)$ depends only on $(x_1, \dots, x_{\sigma(i)})$ where $i \mapsto \sigma(i)$ is an increasing function. Then there exist two sets of indeterminates $z = (z_1, \dots, z_s)$ and $t = (t_1, \dots, t_r)$, an increasing function τ , convergent power series $z_i(x) \in \mathbb{C}\{x\}$ vanishing at 0 such that $z_1(x), \dots, z_{\tau(i)}(x)$ depend only on $(x_1, \dots, x_{\sigma(i)})$, complex numbers $t_1, \dots, t_r \in \mathbb{C}$ and an algebraic power series vector solution $y(t, x, z) \in \overline{\mathbb{Q}}\langle t, x, z \rangle^m$ of

$$f(x, y(t, x, z)) = 0$$

such that

$$y_i(t, x, z) \in \overline{\mathbb{Q}}\langle t, x_1, \dots, x_{\sigma(i)}, z_1, \dots, z_{\tau(i)} \rangle \quad \text{for every } i,$$

$y(\mathbf{t}, x, z)$ is well defined and $y(x) = y(\mathbf{t}, x, z(x))$.

Proof. By [Bilski et al. 2017, Theorem 1.2] there exist a new set of indeterminates $z = (z_1, \dots, z_s)$, an increasing function τ , convergent power series $z_i(x) \in \mathbb{C}\{x\}$ vanishing at 0 such that $z_1(x), \dots, z_{\tau(i)}(x)$ depend only on $(x_1, \dots, x_{\sigma(i)})$, and an algebraic power series vector solution $y(x, z) \in \mathbb{C}\langle x, z \rangle^m$ of

$$f(x, y(x, z)) = 0$$

such that

$$y_i(x, z) \in \mathbb{C}\langle x_1, \dots, x_{\sigma(i)}, z_1, \dots, z_{\tau(i)} \rangle \quad \text{for every } i$$

and $y(x) = y(x, z(x))$. Then we apply Theorem 7 to the vector $y(x, z)$. □

2. Proof of Theorem 2

The proof is similar to the proof of [Bilski et al. 2017, Theorem 1.2] and so we will refer several times to this paper for details. For convenience x^{n-1} will denote the vector of indeterminates (x_1, \dots, x_{n-1}) and, more generally, x^i will denote the vector of indeterminates (x_1, \dots, x_i) . Firstly we consider the case $\mathbb{K} = \mathbb{C}$. Let $g_1, \dots, g_k \in \mathbb{C}\{x\}$ be the defining equations of $(V, 0)$. By a linear change of coordinates we may assume that the g_i are Weierstrass polynomials in x_n :

$$g_s(x) = x_n^{r_s} + \sum_{j=1}^{r_s} a_{n-1,s,j}(x^{n-1})x_n^{r_s-j} \quad \text{for all } s = 1, \dots, k$$

and

$$\text{mult}_0(g_s) = r_s \quad \text{for all } s = 1, \dots, k. \tag{5}$$

Then the $a_{n-1,s,j}$ are arranged in a row vector $a_{n-1} \in \mathbb{C}\{x^{n-1}\}^{p_n}$ with $p_n = \sum_s r_s$. Let f_n denote the product of the g_s . Let $\Delta_{n,i}$ denote the i -th generalized discriminant of f_n seen as a polynomial in x_n (see [Bilski et al. 2017, 4.2]). This is a polynomial depending on a_{n-1} . Then let $\Delta_{n,j_n}(a_{n-1})$ be the first nonvanishing generalized discriminant. After a linear change of coordinates in x_1, \dots, x_{n-1} we may assume, by the Weierstrass preparation theorem, that

$$\Delta_{n,j_n}(a_{n-1}) = u_{n-1}(x^{n-1}) \left(x_{n-1}^{p_{n-1}} + \sum_{j=1}^{p_{n-1}} a_{n-2,j}(x^{n-2})x_{n-1}^{p_{n-1}-j} \right),$$

where $u_{n-1}(0) \neq 0$ and for all j , $a_{n-2,j}(0) = 0$. We carry on with this construction (exactly as in [Bilski et al. 2017, 4.2]) and define a sequence of Weierstrass polynomials $f_i(x^i)$ for $i = 1, \dots, n - 1$ such that $f_i = x_i^{p_i} + \sum_{j=1}^{p_i} a_{i-1,j}(x^{i-1})x_i^{p_i-j}$ is the Weierstrass polynomial associated to the first nonidentically zero generalized discriminant $\Delta_{i+1,j_{i+1}}(a_i)$ of f_{i+1} , where a_i denotes the vector $(a_{i,1}, \dots, a_{i,p_{i+1}})$:

$$\Delta_{i+1,j_{i+1}}(a_i) = u_i(x^i) \left(x_i^{p_i} + \sum_{j=1}^{p_i} a_{i-1,j}(x^{i-1})x_i^{p_i-j} \right), \quad i = 0, \dots, n - 1. \tag{6}$$

Thus the vector of power series a_i satisfies

$$\Delta_{i+1,k}(a_i) \equiv 0 \quad \text{for } k < j_{i+1} \text{ and } i = 0, \dots, n - 1. \tag{7}$$

In particular $\Delta_{1,j_1}(a_0)$ is a constant. Then we use Theorem 12 to see that there exist two sets of indeterminates $z = (z_1, \dots, z_s)$ and $t = (t_1, \dots, t_r)$, an increasing function τ , convergent power series $z_i(x) \in \mathbb{C}\{x\}$ vanishing at 0, complex numbers $t_1, \dots, t_r \in \mathbb{C}$, algebraic power series $u_i(t, x^i, z) \in \overline{\mathbb{Q}}\langle t, x^i, z_1, \dots, z_{\tau(i)} \rangle$ and vectors of algebraic power series

$$a_i(t, x^i, z) \in \overline{\mathbb{Q}}\langle t, x^i, z_1, \dots, z_{\tau(i)} \rangle^{P_i}$$

such that the following hold:

- (a) $z_1(x), \dots, z_{\tau(i)}(x)$ depend only on (x_1, \dots, x_i) .
- (b) $a_i(t, x^i, z)$ and $u_i(t, x^i, z)$ are solutions of (6) and (7).
- (c) $a_i(x^i) = a_i(t, x^i, z(x^i))$ and $u_i(x^i) = u_i(t, x^i, z(x^i))$.

Let \bar{u}_i be the constant coefficient of $u_i(x^i)$. Because $z(0) = 0$ we have $\bar{u}_i = u_i(t, 0, 0)$ and $u_i(t, 0, 0) \in \overline{\mathbb{Q}}\langle t \rangle$. In particular $u_i(t, 0, 0) \neq 0$. Let $\gamma : \mathcal{U} \rightarrow \mathbb{C}^r$ be the analytic map defined by

$$\gamma(\lambda) = (1 - \lambda)\mathbf{q} + \lambda\mathbf{t}$$

where \mathcal{U} is an open connected neighborhood of the closed unit disc in \mathbb{C} and $\mathbf{q} \in \overline{\mathbb{Q}}^r$. Because $u_i(t, 0, 0) \neq 0$ and $\overline{\mathbb{Q}}$ is dense in \mathbb{C} we may choose \mathbf{q} close enough to \mathbf{t} such that

$$u_i(\gamma(\lambda), 0, 0) \neq 0 \quad \text{for all } i \text{ and } \lambda \in \mathcal{U}.$$

Again because $\overline{\mathbb{Q}}$ is dense in \mathbb{C} we can find $\mathbf{q} \in \overline{\mathbb{Q}}$ close enough to \mathbf{t} such that the following are, for all $\lambda \in \mathcal{U}$, well defined convergent power series in x :

$$F_n(\lambda, x) := \prod_s G_s(\lambda, x), \quad \text{for } G_s(\lambda, x) := x_n^{r_s} + \sum_{j=1}^{r_s} a_{n-1,s,j}(\gamma(\lambda), x^{n-1}, \lambda z(x^{n-1}))x_n^{r_s-j}$$

and, for $i = 1, \dots, n - 1$,

$$F_i(\lambda, x) := x_i^{P_i} + \sum_{j=1}^{P_i} a_{i-1,j}(\gamma(\lambda), x^{i-1}, \lambda z(x^{i-1}))x_i^{P_i-j} \quad \text{and} \quad u_i(\gamma(\lambda), x^i, z(x^i)).$$

Finally we set $F_0 \equiv 1$. Because $u_i(\gamma(\lambda), 0, z(0)) \neq 0$, the family $F_i(\lambda, x)$ satisfies the assumptions of [Parusiński and Păunescu 2017, Theorem 3.3] with $|\lambda| \leq 1$, i.e., the family is Zariski equisingular. Moreover by (5) we have

$$\text{mult}_0(G_s) = r_s \quad \text{for all } s = 1, \dots, k,$$

so the family is Zariski equisingular with transverse projections (see [Parusiński and Păunescu 2017, Definition 4.1]). So by Theorem 4.3 of that work this family is a regular Zariski equisingular family and by Theorem 7.1 of that work it is Whitney equisingular. Thus $\{F_n(0, x) = 0\}$ and $\{F_n(1, x) = 0\} = \{f_n(x) = 0\}$ are homeomorphic and the homeomorphism between them can be chosen to be subanalytic and arcanalytic. We have $F_n(0, x) \in \overline{\mathbb{Q}}\langle x \rangle$ thus, by [Bilski et al. 2017, Theorem 3.2], we may assume that $(V, 0)$ is the germ of a Nash set defined over $\overline{\mathbb{Q}}$. When $\mathbb{K} = \mathbb{R}$ we may also assume that $(V, 0)$ is the germ of a Nash

set defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$. This follows from the complex case by the same argument used in the proof of [Bilski et al. 2017, Corollary 4.1]. Then we conclude with the following theorem:

Theorem 13. *Let $(V, 0) \subset (\mathbb{K}^n, 0)$ be a Nash set germ defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$. Then there exists a local Nash diffeomorphism $h : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^n, 0)$ such that $h(V)$ is the germ of an algebraic subset of \mathbb{K}^n defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$.*

Proof. This follows from Proposition 14 given below which is a slight modification of [Bochnak and Kucharz 1984, Proposition 2]. Indeed let $f : U \rightarrow \mathbb{K}^m$ be a Nash function such that $f^{-1}(0) = V$. Then by Proposition 14 we have $V = s^{-1}(\varphi^{-1}(0))$. But $s : U \rightarrow s(U)$ is a Nash diffeomorphism by Proposition 14ii. So we set $h = s$ and $h(V)$ is an algebraic set equal to $\varphi^{-1}(0)$, again by using the notations of Proposition 14. \square

Proposition 14. *Let $f : U \rightarrow \mathbb{K}^m$ be a Nash map defined on an open connected set $U \subset \mathbb{K}^n$ by algebraic power series with coefficients in $\overline{\mathbb{Q}} \cap \mathbb{K}$. Then there exist an algebraic set $X \subset \mathbb{K}^n \times \mathbb{K}^N$, a polynomial map $\varphi : X \rightarrow \mathbb{K}^m$ and a Nash map $s : U \rightarrow \mathbb{K}^n \times \mathbb{K}^N$ satisfying the following properties:*

- (i) $s(U) \subset \text{Reg}(X)$ is a connected component of $p^{-1}(U) \cap X$, where $p : \mathbb{K}^n \times \mathbb{K}^N \rightarrow \mathbb{K}^n$ is the first projection.
- (ii) $p \circ s = \text{Id}_U$.
- (iii) $f = \varphi \circ s$.
- (iv) *the coefficients of the polynomials defining X and φ are in $\overline{\mathbb{Q}} \cap \mathbb{K}$.*

Proof. The existence of X , φ and s satisfying (i), (ii) and (iii) are given by [Bochnak and Kucharz 1984, Proposition 2] in the general case where f is defined by algebraic power series with coefficients in \mathbb{K} . In fact X is the normalization of the Zariski closure of the graph of f and φ is the restriction to X of a generic linear map $\mathbb{K}^{n+N} \rightarrow \mathbb{K}^m$. In particular, since f is assumed to be defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$, we have that X is defined by polynomial equations with coefficients in $\overline{\mathbb{Q}} \cap \mathbb{K}$. Because φ is generic we can choose such a φ with coefficients in $\overline{\mathbb{Q}} \cap \mathbb{K}$ since this field is dense in \mathbb{K} . \square

3. Proof of Theorem 1

The proof is similar to the proof of [Bilski et al. 2017, Theorem 1.3] and so once again we will refer several times to this paper for some details. We begin by considering the case $\mathbb{K} = \mathbb{C}$. Let g_1, \dots, g_p be power series defining $(V, 0)$. Let us replace n by $n - 1$ to assume that $(V, 0) \subset (\mathbb{C}^{n-1}, 0)$ and let (x_2, \dots, x_n) denote the coordinates in \mathbb{C}^{n-1} . Let us set $g_0 := g$. After a linear change of coordinates in x_2, \dots, x_n (i.e., preserving x_1) we have

$$\prod_{m=0}^p (x_1 - g_m(x_2, \dots, x_n))$$

is x_n -regular. Thus we may write

$$\prod_{m=0}^p (x_1 - g_m(x_2, \dots, x_n)) = u_n(x) \left(x_n^{p_n} + \sum_{j=1}^{p_n} a_{n-1,j} (x^{n-1}) x_n^{p_n-j} \right),$$

where $u_n(0) \neq 0$ and $a_{n-1,j}(0) = 0$. We set

$$f_n(x) = x_n^{p_n} + \sum_{j=1}^{p_n} a_{n-1,j} (x^{n-1}) x_n^{p_n-j}$$

so that

$$u_n(x) f_n(x) = \prod_{m=0}^p \left(x_1 - \sum_{k=2}^n x_k b_{m,k} (x_2, \dots, x_n) \right), \quad (8)$$

with $g_m = \sum_{k=2}^n x_k b_{m,k}$ for some power series $b_{m,k}$ since $g_m(0) = 0$ for every m . We denote by $b \in \mathbb{C}\{x\}^{p(n-1)}$ and $a_{n-1} \in \mathbb{C}\{x^{n-1}\}^{p_n}$ the vector of the coefficients $b_{m,k}$ and $a_{n-1,j}$, respectively. Again we denote by $\Delta_{n,i}$ the generalized discriminants of f_n which are polynomials in a_{n-1} . Let j_n be the positive integer such that

$$\Delta_{n,i}(a_{n-1}) \equiv 0 \quad \text{for } i < j_n,$$

and $\Delta_{n,j_n}(a_{n-1}) \neq 0$. After a linear change of coordinates (x_2, \dots, x_{n-1}) we may write

$$\Delta_{n,j_n}(a_{n-1}) = u_{n-1}(x^{n-1}) x_1^{q_{n-1}} \left(x_{n-1}^{p_{n-1}} + \sum_{j=1}^{p_{n-1}} a_{n-2,j} (x^{n-2}) x_{n-1}^{p_{n-1}-j} \right),$$

where $u_{n-1}(0) \neq 0$ and $a_{n-2,j}(0) = 0$. We set

$$f_{n-1} = x_{n-1}^{p_{n-1}} + \sum_{j=1}^{p_{n-1}} a_{n-2,j} (x^{n-2}) x_{n-1}^{p_{n-1}-j}$$

and the vector of its coefficients $a_{n-2,j}$ is denoted by $a_{n-2} \in \mathbb{C}\{x^{n-2}\}^{p_{n-1}}$. Let j_{n-1} be the positive integer such that

$$\Delta_{n-1,k}(a_{n-2}) \equiv 0 \quad \text{for all } k < j_{n-1} \quad \text{and} \quad \Delta_{n-1,j_{n-1}}(a_{n-2}) \neq 0.$$

Then again we divide $\Delta_{n-1,j_{n-1}}$ by the maximal power of x_1 and, after a linear change of coordinates (x_2, \dots, x_{n-2}) , we denote by $f_{n-2}(x^{n-2})$ the associated Weierstrass polynomial.

We carry on with this construction and define a sequence of Weierstrass polynomials $f_i(x^i)$, for $i = 1, \dots, n-1$, such that $f_i = x_i^{p_i} + \sum_{j=1}^{p_i} a_{i-1,j} (x^{i-1}) x_i^{p_i-j}$ is the Weierstrass polynomial associated to the first nonidentically zero generalized discriminant $\Delta_{i,j_i}(a_{i+1})$ of f_{i+1} , divided by the maximal power of x_1 , where $a_i = (a_{i,1}, \dots, a_{i,p_i})$:

$$\Delta_{i+1,j_{i+1}}(a_i) = u_i(x^i) x_1^{q_i} \left(x_i^{p_i} + \sum_{j=1}^{p_i} a_{i-1,j} (x^{i-1}) x_i^{p_i-j} \right) \quad \text{for } i = 0, \dots, n-1. \quad (9)$$

Thus the vector of power series a_i satisfies

$$\Delta_{i+1,k}(a_{i-1}) \equiv 0 \quad \text{for } k < j_{i+1} \text{ and } i = 0, \dots, n-1. \tag{10}$$

Then we use Theorem 12 to see that there exist two sets of indeterminates $z = (z_1, \dots, z_s)$ and $t = (t_1, \dots, t_r)$, an increasing function τ , convergent power series $z_i(x) \in \mathbb{C}\{x\}$ vanishing at 0, complex numbers $t_1, \dots, t_r \in \mathbb{C}$, algebraic power series $u_i(t, x^i, z) \in \overline{\mathbb{Q}}\langle t, x^i, z_1, \dots, z_{\tau(i)} \rangle$ and vectors of algebraic power series

$$b(t, x, z) \in \overline{\mathbb{Q}}\langle t, x, z \rangle^{p(n-1)} \quad \text{and} \quad a_i(t, x^i, z) \in \overline{\mathbb{Q}}\langle t, x^{(i)}, z_1, \dots, z_{\tau(i)} \rangle^{p_i},$$

such that the following hold:

- (a) $z_1(x), \dots, z_{\tau(i)}(x)$ depend only on (x_1, \dots, x_i) .
- (b) $a_i(t, x^i, z)$, $u_i(t, x^i, z)$ and $b(t, x, z)$ are solutions of (8), (9) and (10).
- (c) $a_i(x^i) = a_i(t, x^i, z(x^i))$, $u_i(x^i) = u_i(t, x^i, z(x^i))$ and $b(x) = b(t, x, z(x))$.

Then we repeat what we did in the proof of Theorem 2. Let \bar{u}_i be the constant coefficient of $u_i(x^i)$. Because $z(0) = 0$ we have $\bar{u}_i = u_i(t, 0, 0)$ and $u_i(t, 0, 0) \in \overline{\mathbb{Q}}\langle t \rangle$. In particular $u_i(t, 0, 0) \neq 0$. Let $\gamma : \mathcal{U} \rightarrow \mathbb{C}^r$ be the analytic map defined by

$$\gamma(\lambda) = (1 - \lambda)\mathbf{q} + \lambda\mathbf{t}$$

where \mathcal{U} is an open connected neighborhood of the closed unit disc in \mathbb{C} and $\mathbf{q} \in \overline{\mathbb{Q}}^r$. Because $u_i(t, 0, 0) \neq 0$ and $\overline{\mathbb{Q}}$ is dense in \mathbb{C} we may choose \mathbf{q} close enough to \mathbf{t} such that

$$u_i(\gamma(\lambda), 0, 0) \neq 0 \quad \text{for all } i \text{ and } \lambda \in \mathcal{U}.$$

Again because $\overline{\mathbb{Q}}$ is dense in \mathbb{C} we can find $\mathbf{q} \in \overline{\mathbb{Q}}$ close enough to \mathbf{t} such that the following are, for all $\lambda \in \mathcal{U}$, well defined convergent power series in x :

$$F_i(\lambda, x) := x_i^{p_i} + \sum_{j=1}^{p_i} a_{i-1,j}(\gamma(\lambda), x^{i-1}, \lambda z(x^{i-1}))x_i^{p_i-j} \quad \text{for } i = 0, \dots, n,$$

$$u_i(\gamma(\lambda), x^i, z(x^i)) \quad \text{for } i = 1, \dots, n-1.$$

We have

$$u_n(\gamma(\lambda), x, \lambda z(x))F_n(\lambda, x) = \prod_{m=0}^p \left(x_1 - \sum_{k=2}^n x_k b_{m,k}(\gamma(\lambda), x, \lambda z(x)) \right).$$

By the implicit function theorem or the Weierstrass preparation theorem we have

$$x_1 - \sum_{k=2}^n x_k b_{m,k}(\gamma(\lambda), x, \lambda z(x)) = v_m(\lambda, x)(x_1 - G_m(\lambda, x_2, \dots, x_n)),$$

where $v_m(\lambda, x) \in \mathbb{C}\{\lambda, x\}$, $G_m(\lambda, x_2, \dots, x_n) \in \mathbb{C}\{\lambda, x_2, \dots, x_n\}$ and $v_m(0, 0) \neq 0$. Because

$$x_1 - \sum_{k=2}^n x_k b_{m,k}(\gamma(0), x, 0) \in \overline{\mathbb{Q}}\langle x \rangle$$

we have

$$v_m(0, x), G_m(0, x_2, \dots, x_n) \in \overline{\mathbb{Q}}\langle x \rangle$$

by unicity in the Weierstrass preparation theorem. We set

$$\hat{g}_m(y) := G_m(0, y) \quad \text{for } m = 0, \dots, p$$

where $y = (y_1, \dots, y_{n-1})$ is a new vector of indeterminates. Then, for both cases $\mathbb{K} = \mathbb{C}$ or \mathbb{R} , we conclude exactly as in [Bilski et al. 2017] (see the end of 5.4 and Proposition 5.3 in that work) to show that there is a homeomorphism $h : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^n, 0)$ such that $(h(V), 0)$ is a germ of a Nash subset of \mathbb{K}^n defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$ and $g \circ h$ is the germ of a Nash function defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$.

Then we deduce from Proposition 14 the following analogue of [Bilski et al. 2017, Theorem 5.4]:

Corollary 15. *Let g_i be algebraic powers series with coefficients in $\overline{\mathbb{Q}} \cap \mathbb{K}$ defining Nash function germs $g_i : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}, 0)$. Then there exist a Nash diffeomorphism $h : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^n, 0)$ and Nash units $u_i : (\mathbb{K}^n, 0) \rightarrow \mathbb{K}$, $u_i(0) \neq 0$, such that, for every i , $u_i(x)g_i(h(x))$ are polynomial function germs defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$.*

Proof. We have the following fact: let $(Y, 0) \subset (\mathbb{K}^n, 0)$ be a Nash set germ defined by algebraic power series with coefficients in $\overline{\mathbb{Q}} \cap \mathbb{K}$. Then there exists a Nash diffeomorphism $h : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^n, 0)$ such that for every irreducible analytic component W of $(Y, 0)$, the ideal of functions vanishing on $h(W)$ is generated by polynomials with coefficients in $\overline{\mathbb{Q}} \cap \mathbb{K}$. This follows from Proposition 14 by applying word for word the proof of “(i) \implies (iv)” in [Bochnak and Kucharz 1984, Theorem 5]. Thus when $\mathbb{K} = \mathbb{C}$ this fact applied to the germ $(Y, 0)$ defined by the products of the g_i proves the theorem. When $\mathbb{K} = \mathbb{R}$ we conclude as done for this case in the proof of [Bilski et al. 2017, Theorem 5.4]. \square

Let us recall that we have shown that there is a homeomorphism $h : (\mathbb{K}^n, 0) \rightarrow (\mathbb{K}^n, 0)$ such that $(h(V), 0)$ is the germ of a Nash subset of \mathbb{K}^n defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$ and $g \circ h$ is the germ of a Nash function defined over $\overline{\mathbb{Q}} \cap \mathbb{K}$. So we conclude the proof of Theorem 1 by using Theorem 13 and [Bilski et al. 2017, Theorem 5.5].

Acknowledgements

This work originates from discussions with Adam Parusiński. I wish to thank him warmly for the many fruitful discussions we had on this problem and his helpful remarks and valuable suggestions concerning the earlier versions of this paper. I also thank the referee for their useful and suitable comments and remarks.

References

- [Bilski et al. 2017] M. Bilski, A. Parusiński, and G. Rond, “Local topological algebraicity of analytic function germs”, *J. Algebraic Geom.* **26**:1 (2017), 177–197. MR Zbl
- [Bochnak and Kucharz 1984] J. Bochnak and W. Kucharz, “Local algebraicity of analytic sets”, *J. Reine Angew. Math.* **352** (1984), 1–14. MR Zbl

- [Budur and Wang 2017] N. Budur and B. Wang, “Local systems on analytic germ complements”, *Adv. Math.* **306** (2017), 905–928. MR Zbl
- [Cutkosky and Kashcheyeva 2008] S. D. Cutkosky and O. Kashcheyeva, “Algebraic series and valuation rings over nonclosed fields”, *J. Pure Appl. Algebra* **212**:8 (2008), 1996–2010. MR Zbl
- [Kucharz 1986] W. Kucharz, “Power series and smooth functions equivalent to a polynomial”, *Proc. Amer. Math. Soc.* **98**:3 (1986), 527–533. MR Zbl
- [Mostowski 1984] T. Mostowski, “Topological equivalence between analytic and algebraic sets”, *Bull. Polish Acad. Sci. Math.* **32**:7-8 (1984), 393–400. MR Zbl
- [Parusiński and Păunescu 2017] A. Parusiński and L. Păunescu, “Arc-wise analytic stratification, Whitney fibering conjecture and Zariski equisingularity”, *Adv. Math.* **309** (2017), 254–305. MR Zbl
- [Teissier 1990] B. Teissier, “Un exemple de classe d’équisingularité irrationnelle”, *C. R. Acad. Sci. Paris Sér. I Math.* **311**:2 (1990), 111–113. MR Zbl
- [Tougeron 1976] J.-C. Tougeron, “Solutions d’un système d’équations analytiques réelles et applications”, *Ann. Inst. Fourier (Grenoble)* **26**:3 (1976), 109–135. MR Zbl
- [Tougeron 1990] J.-C. Tougeron, “Sur les racines d’un polynôme à coefficients séries formelles”, pp. 325–363 in *Real analytic and algebraic geometry* (Trento, Italy, 1988), edited by M. Galbiati and A. Tognoli, Lecture Notes in Math. **1420**, Springer, 1990. MR Zbl
- [Varčenko 1972] A. N. Varčenko, “Theorems on the topological equisingularity of families of algebraic varieties and families of polynomial mappings”, *Izv. Akad. Nauk SSSR Ser. Mat.* **36** (1972), 957–1019. In Russian; translated in *Math. USSR-Izv.* **6** (1972), 949–1008. MR Zbl
- [Whitney 1965] H. Whitney, “Local properties of analytic varieties”, pp. 205–244 in *Differential and combinatorial topology*, edited by S. S. Cairns, Princeton Univ. Press, 1965. Reprinted as pp. 497–536 in his *Collected papers*, Birkhäuser, Boston, 1992. MR Zbl

Communicated by János Kollár

Received 2017-06-09 Revised 2017-10-03 Accepted 2018-01-20

guillaume.rond@univ-amu.fr

Aix-Marseille Université CNRS, Marseille, France

Polynomial bound for the nilpotency index of finitely generated nil algebras

Mátyás Domokos

Working over an infinite field of positive characteristic, an upper bound is given for the nilpotency index of a finitely generated nil algebra of bounded nil index n in terms of the maximal degree in a minimal homogenous generating system of the ring of simultaneous conjugation invariants of tuples of n -by- n matrices. This is deduced from a result of Zubkov. As a consequence, a recent degree bound due to Derksen and Makam for the generators of the ring of matrix invariants yields an upper bound for the nilpotency index of a finitely generated nil algebra that is polynomial in the number of generators and the nil index. Furthermore, a characteristic free treatment is given to Kuzmin's lower bound for the nilpotency index.

1. Introduction

Throughout this note \mathbb{F} stands for an infinite field of positive characteristic. All vector spaces, tensor products and algebras are taken over \mathbb{F} . The results of this paper are valid in arbitrary characteristic, but they are known in characteristic zero (in fact stronger statements hold in characteristic zero, see Formanek [1991], giving in particular an account of relevant works of Razmyslov [1974] and Procesi [1976]).

Write $\mathcal{F}_m := \mathbb{F}\langle x_1, \dots, x_m \rangle$ for the free associative \mathbb{F} -algebra with identity 1 on m generators x_1, \dots, x_m , and let \mathcal{F}_m^+ be its ideal generated by x_1, \dots, x_m (so \mathcal{F}_m^+ is the free nonunitary associative algebra of rank m). For a positive integer n denote by $I_{n,m}$ the ideal in \mathcal{F}_m generated by $\{a^n \mid a \in \mathcal{F}_m^+\}$. A theorem of Kaplansky [1946] asserts that if a finitely generated associative algebra satisfies the polynomial identity $x^n = 0$, then it is nilpotent. Equivalently, there exists a positive integer d such that for all $i_1, \dots, i_d \in \{1, \dots, m\}$ the monomial $x_{i_1} \cdots x_{i_d}$ belongs to $I_{n,m}$. Denote by $d_{\mathbb{F}}(n, m)$ the minimal such d . In other words, $d_{\mathbb{F}}(n, m)$ is the minimal positive integer d such that all \mathbb{F} -algebras that are generated by m elements and satisfy the polynomial identity $x^n = 0$ satisfy also the polynomial identity $y_1 \cdots y_d = 0$. This is a notable quantity of noncommutative ring theory; Jacobson [1945] reduced the Kurosh problem for finitely generated algebraic algebras of bounded degree to the case of nil algebras of bounded degree. We mention also that proving nilpotency of nil rings under various conditions is a natural target for ring theorists, see for example the paper of Guralnick, Small and Zelmanov [2010].

This research was partially supported by National Research, Development and Innovation Office, NKFIH K 119934.

MSC2010: primary 16R10; secondary 13A50, 15A72, 16R30.

Keywords: nil algebra, nilpotent algebra, matrix invariant, degree bound.

The number $d_{\mathbb{F}}(n, m)$ is tightly connected with a quantity appearing in commutative invariant theory defined as follows. Consider the generic matrices

$$X_r = (x_{ij}(r))_{1 \leq i, j \leq n}, \quad r = 1, \dots, m.$$

These are elements in the algebra $A^{n \times n}$ of $n \times n$ matrices over the mn^2 -variable commutative polynomial algebra $A = \mathbb{F}[x_{ij}(r) \mid 1 \leq i, j \leq n, 1 \leq r \leq m]$. The general linear group $\mathrm{GL}_n(\mathbb{F})$ acts on A via \mathbb{F} -algebra automorphisms; for $g \in \mathrm{GL}_n(\mathbb{F})$ we have that $g \cdot x_{ij}(r)$ is the (i, j) -entry of the matrix $g^{-1}X_r g$. Set $R_{n,m} = A^{\mathrm{GL}_n(\mathbb{F})}$, the subalgebra of $\mathrm{GL}_n(\mathbb{F})$ -invariants. This is the algebra of polynomial invariants under simultaneous conjugation of m -tuples of $n \times n$ matrices. The polynomial ring A is graded in the standard way, and since the $\mathrm{GL}_n(\mathbb{F})$ -action preserves the grading, the subalgebra $R_{n,m}$ is generated by homogeneous elements. Being the algebra of invariants of a reductive group, $R_{n,m}$ is finitely generated by the Hilbert–Nagata theorem (see for example [Newstead 1978]). We write $\beta_{\mathbb{F}}(n, m)$ for the minimal positive integer d such that the \mathbb{F} -algebra $R_{n,m}$ is generated by elements of degree at most d . The main result of the present note is the following inequality:

Theorem 1.1.
$$d_{\mathbb{F}}(n, m) \leq \beta_{\mathbb{F}}(n, m + 1).$$

Remark 1.2. In the reverse direction it was shown in [Domokos 2002, Theorem 3] that for $n \geq 2$ we have

$$\beta_{\mathbb{F}}(n, m) \leq \left\lfloor \frac{n}{2} \right\rfloor d_{\mathbb{F}}(n, m).$$

Theorem 1.1 is derived from a theorem of Zubkov [1996] (for which Lopatin [2013] gave versions and improvements), see Theorem 2.1. Using a result of Ivanyos, Qiao and Subrahmanyam [2017], Derksen and Makam [2017b] found strong bounds on the degrees of invariants defining the null-cone of m -tuples of $n \times n$ matrices under simultaneous conjugation, and derived from this the following upper bound on $\beta_{\mathbb{F}}(n, m)$:

Theorem 1.3 [Derksen and Makam 2017a, Theorem 1.4]. *We have the inequality*

$$\beta_{\mathbb{F}}(n, m) \leq (m + 1)n^4.$$

Given this result Derksen and Makam conjectured [2017a, Conjecture 1.5] that there exists an upper bound on $d_{\mathbb{F}}(n, m)$ that is polynomial in n and m . Combining Theorem 1.1 and Theorem 1.3 we obtain the following affirmative answer to this conjecture:

Corollary 1.4.
$$d_{\mathbb{F}}(n, m) \leq (m + 2)n^4.$$

Remark 1.5. Corollary 1.4 is a drastic improvement of the earlier known general upper bounds on $d_{\mathbb{F}}(n, m)$:

- (1) $d_{\mathbb{F}}(n, m) \leq n^6 m^{n+1}$ by Belov [1992].
- (2) $d_{\mathbb{F}}(n, m) \leq \frac{1}{6} n^6 m^n$ by Klein [2000].
- (3) $d_{\mathbb{F}}(n, m) \leq 2^{18} mn^{12 \log_3(n) + 28}$ by Belov and Kharitonov [2012].

It is easy to see that $d_{\mathbb{F}}(2, m) \leq m + 1$. We note that for the case $n = 3$ exact results on $d_{\mathbb{F}}(3, m)$ were obtained by Lopatin [2005]. Moreover, Lopatin [2012] proved that if $\text{char}(\mathbb{F}) > \frac{n}{2}$ then $d_{\mathbb{F}}(n, m) \leq n^{1+\log_2(3m+2)}$ and $d_{\mathbb{F}}(n, m) \leq 2^{2+n/2}m$.

Remark 1.6. When $\text{char}(\mathbb{F}) > n^2 + 1$, we have $\beta_{\mathbb{F}}(n, m) \leq n^2$. Indeed, the proof presented by Formanek [1986] (following the original arguments of Razmyslov [1974] and Procesi [1976]) for the zero characteristic case of the corresponding inequality goes through without essential changes when $\text{char}(\mathbb{F}) > n^2 + 1$. Thus by Theorem 1.1 we get that $d_{\mathbb{F}}(n, m) \leq n^2$ when $\text{char}(\mathbb{F}) > n^2 + 1$.

In Section 3 we show that the following lower bound for $d_{\mathbb{F}}(n, m)$ due to E. N. Kuzmin [1975] when $\text{char}(\mathbb{F}) = 0$ or $\text{char}(\mathbb{F}) > n$ holds in arbitrary characteristic:

Theorem 1.7. *The monomial $x_2x_1x_2x_1^2x_2x_1^3 \cdots x_2x_1^{n-1}$ is not contained in the ideal $I_{n,2}$. In particular, for $m \geq 2$ we have $d_{\mathbb{F}}(n, m) \geq n(n + 1)/2$.*

Remark 1.8. It is well known that when $0 < \text{char}(\mathbb{F}) \leq n$, the element $x_1x_2 \cdots x_m$ is not contained in $I_{n,m}$, see for example [Nagata 1952, 5. Remark (I)]. So in this case for $m \geq 2$ we have

$$\max \left\{ m + 1, \frac{n(n + 1)}{2} \right\} \leq d_{\mathbb{F}}(n, m) \leq (m + 2)n^4.$$

2. Identities of matrices with forms

The map $x_i \mapsto X_i$ ($i = 1, \dots, m$) extends to a unique \mathbb{F} -algebra homomorphism $\varphi_1 : \mathcal{F}_m \rightarrow A^{n \times n}$. We have $\varphi_1(1) = I$, the $n \times n$ identity matrix. Consider the commutative polynomial algebra

$$\mathcal{P}_{n,m} = \mathbb{F}[s_l(a) \mid a \in \mathcal{F}_m^+, l = 1, \dots, n]$$

generated by the infinitely many commuting indeterminates $s_l(a)$. Define the \mathbb{F} -algebra homomorphism

$$\varphi_2 : \mathcal{P}_{n,m} \rightarrow R_{n,m}, \quad \varphi_2(s_l(a)) = \sigma_l(\varphi_1(a)),$$

where for $B \in A^{n \times n}$ we have

$$\det(tI + B) = \sum_{l=0}^n t^l \sigma_{n-l}(B),$$

so $\sigma_l(B)$ is the sum of the principal $l \times l$ minors of B . A theorem of Donkin [1992] asserts that φ_2 is surjective onto $R_{n,m}$. Combining φ_1 and φ_2 we get an \mathbb{F} -algebra homomorphism

$$\varphi : \mathcal{P}_{n,m} \otimes \mathcal{F}_m \rightarrow A^{n \times n}, \quad b \otimes a \mapsto \varphi_2(b)\varphi_1(a).$$

The subalgebra $C_{n,m} = \varphi(\mathcal{P}_{n,m} \otimes \mathcal{F}_m)$ is called the *algebra of matrix concomitants*. It can be interpreted as the algebra of $\text{GL}_n(\mathbb{F})$ -equivariant polynomial maps $(\mathbb{F}^{n \times n})^m \rightarrow \mathbb{F}^{n \times n}$, where $\text{GL}_n(\mathbb{F})$ acts on $\mathbb{F}^{n \times n}$ by conjugation and on the space $(\mathbb{F}^{n \times n})^m$ of m -tuples of matrices by simultaneous conjugation. For $a \in \mathcal{F}_m^+$

define an element $\chi_n(a)$ in $\mathcal{P}_{n,m} \otimes \mathcal{F}_m$ as follows:

$$\chi_n(a) = \sum_{l=0}^n (-1)^l s_l(a) \otimes a^{n-l}$$

(where $s_0(a) = 1$). We need the following result of Zubkov [1996] (see also Lopatin [2013, Theorem 2.4]):

Theorem 2.1 [Zubkov 1996]. *The ideal $\ker(\varphi)$ is generated by*

$$\{b \otimes 1, \chi_n(a) \mid b \in \ker(\varphi_2), a \in \mathcal{F}_m^+\}.$$

Remark 2.2. The papers [Zubkov 1996; Lopatin 2013] use different commutative polynomial algebras than our $\mathcal{P}_{n,m}$, however, it is straightforward that Theorem 2.1 is an immediate consequence of the versions stated there. We note that those works give descriptions of the ideal $\ker(\varphi_2)$ as well. A self-contained approach to the theorem of Zubkov can be found in the recent book by De Concini and Procesi [2017].

Denote by $\eta : C_{n,m} \rightarrow C_{n,m}/R_{n,m}^+ C_{n,m}$ the natural surjection (ring homomorphism), where $R_{n,m}^+$ is the sum of the positive degree homogeneous components of $R_{n,m}$.

Corollary 2.3. *The kernel of $\eta \circ \varphi_1$ is the ideal $I_{n,m} = (a^n \mid a \in \mathcal{F}_m^+)$ in \mathcal{F}_m .*

Proof. We have $\ker(\eta \circ \varphi_1) = \ker(\eta \circ \varphi) \cap \mathcal{F}_m$ (where we identify \mathcal{F}_m with the subalgebra $1 \otimes \mathcal{F}_m$ in $\mathcal{P}_{n,m} \otimes \mathcal{F}_m$). The ideal $(s_l(a) \otimes 1 \mid a \in \mathcal{F}_m^+, 1 \leq l \leq n)$ is mapped surjectively onto $R_{n,m}^+ C_{n,m}$ by [Donkin 1992]. Therefore we have

$$\ker(\eta \circ \varphi) = \varphi^{-1}(R_{n,m}^+ C_{n,m}) = \ker(\varphi) + (s_l(a) \otimes 1 \mid a \in \mathcal{F}_m^+, 1 \leq l \leq n) = (s_l(a) \otimes 1, 1 \otimes a^n \mid a \in \mathcal{F}_m^+, 1 \leq l \leq n),$$

(the last equality follows from Theorem 2.1 and the fact that $1 \otimes a^n - \chi_n(a)$ belongs to $(s_l(a) \otimes 1 \mid a \in \mathcal{F}_m^+, 1 \leq l \leq n)$). Obviously the ideal $(s_l(a) \otimes 1, 1 \otimes a^n \mid a \in \mathcal{F}_m^+, 1 \leq l \leq n)$ intersects \mathcal{F}_m in $I_{n,m}$. \square

Remark 2.4. Corollary 2.3 implies that the relatively free algebra $\mathcal{F}_m/I_{n,m}$ is isomorphic to $C_{n,m}/R_{n,m}^+ C_{n,m}$. When $\text{char}(\mathbb{F}) = 0$, this statement is due to Procesi [1976, Corollary 4.7].

The algebras $R_{n,m}$ and $C_{n,m}$ are \mathbb{Z}^m -graded:

$$\deg_m(X_{i_1} \cdots X_{i_d}) = (\alpha_1, \dots, \alpha_m) \text{ where } \alpha_k = |\{j \mid i_j = k\}|$$

and

$$\deg_m(\sigma_l(X_{i_1} \cdots X_{i_d})) = l \cdot \deg_m(X_{i_1} \cdots X_{i_d}).$$

Proof of Theorem 1.1. Set $d = \beta_{\mathbb{F}}(n, m + 1)$. We have to show that $x_{i_1} \cdots x_{i_d} \in I_{n,m}$ for all $i_1, \dots, i_d \in \{1, \dots, m\}$. Recall that by [Donkin 1992] the algebra $R_{n,m+1}$ is generated by the elements $\sigma_l(W)$, where W is a word in X_1, \dots, X_{m+1} , and $l \in \{1, \dots, n\}$. The total degree of the element $\text{Tr}(X_{i_1} \cdots X_{i_d} X_{m+1}) \in R_{n,m+1}$ is strictly greater than $\beta_{\mathbb{F}}(n, m + 1)$, whence we have a relation

$$\text{Tr}(X_{i_1} \cdots X_{i_d} X_{m+1}) = \sum_{\lambda \in \Lambda} a_{\lambda} f_{\lambda}, \tag{1}$$

where Λ is a finite index set, $a_\lambda \in \mathbb{F}$, and each $f_\lambda \in R_{n,m+1}$ is a product $f_\lambda = \sigma_{l_1}(W_1) \cdots \sigma_{l_r}(W_r)$ with $r \geq 2$ and W_1, \dots, W_r nonempty words in X_1, \dots, X_{m+1} . The \mathbb{Z}^{m+1} -multidegree of $\text{Tr}(X_{i_1} \cdots X_{i_d} X_{m+1})$ is

$$\text{deg}_{m+1}(\text{Tr}(X_{i_1} \cdots X_{i_d} X_{m+1})) = (\text{deg}_m(\text{Tr}(X_{i_1} \cdots X_{i_d})), 1).$$

The terms f_λ are all \mathbb{Z}^{m+1} -homogeneous, whence we may assume that each has the above \mathbb{Z}^{m+1} -degree (since the other possible terms on the right-hand side of (1) must cancel each other). It follows that for each f_λ exactly one of its factors $\sigma_{l_1}(W_1), \dots, \sigma_{l_r}(W_r)$ has \mathbb{Z}^{m+1} -degree of the form $(\alpha_1, \dots, \alpha_m, 1)$, say this is $\sigma_{l_1}(W_1)$, and the remaining factors have \mathbb{Z}^{m+1} -degree of the form $(\gamma_1, \dots, \gamma_m, 0)$. Necessarily we have $l_1 = 1$ and so $\sigma_{l_1}(W_1) = \text{Tr}(X_{m+1}Z)$ for some (possibly empty) word Z in X_1, \dots, X_m , and W_2, \dots, W_r are nonempty words in X_1, \dots, X_m . Set

$$g_\lambda = \sigma_{l_2}(W_2) \cdots \sigma_{l_r}(W_r)Z \in C_{n,m},$$

and note that $f_\lambda = \text{Tr}(g_\lambda X_{m+1})$. Using linearity of $\text{Tr}(-)$ relation (1) can be written as

$$\text{Tr}\left(X_{m+1}\left(X_{i_1} \cdots X_{i_d} - \sum_{\lambda \in \Lambda} a_\lambda g_\lambda\right)\right) = 0 \in R_{n,m+1}. \tag{2}$$

Substituting $X_{m+1} \mapsto E_{ij}$ (the matrix whose (i, j) -entry is 1 and all other entries are 0) we get from (2) that the (j, i) -entry of $X_{i_1} \cdots X_{i_d} - \sum_{\lambda \in \Lambda} a_\lambda g_\lambda$ is 0. This holds for all (i, j) , thus we have the equality

$$X_{i_1} \cdots X_{i_d} = \sum_{\lambda} a_\lambda g_\lambda. \tag{3}$$

The right-hand side of (3) is obviously contained in $R_{n,m}^+ C_{n,m}$, therefore it follows from (3) that the element $x_{i_1} \cdots x_{i_d} \in \mathcal{F}_m$ belongs to the kernel of $\eta \circ \varphi_1$. Thus by Corollary 2.3 we conclude that $x_{i_1} \cdots x_{i_d} \in I_{n,m}$. \square

3. Lower bound

Kuzmin’s proof of the case $\text{char}(\mathbb{F}) = 0$ or $\text{char}(\mathbb{F}) > n$ of Theorem 1.7 (it is presented also in the survey of Drensky in [Drensky and Formanek 2004]) uses crucially Lemma 3.1 below, relating the *complete linearization of x^n* , namely

$$P_n(x_1, \dots, x_n) = \sum_{\pi \in \text{Sym}\{1, \dots, n\}} x_{\pi(1)} x_{\pi(2)} \cdots x_{\pi(n)} \in \mathcal{F}_n.$$

Lemma 3.1. *If $\text{char}(\mathbb{F}) = 0$ or $\text{char}(\mathbb{F}) > n$, then $I_{n,m}$ is spanned as an \mathbb{F} -vector space by the elements $P_n(w_1, \dots, w_n)$, where w_1, \dots, w_n range over all nonempty monomials in x_1, \dots, x_m .*

Remark 3.2. The assumption on $\text{char}(\mathbb{F})$ in Lemma 3.1 is necessary, its statement obviously fails if $0 < \text{char}(\mathbb{F}) \leq n$ (as it can be easily seen already in the special case $m = 1$). Now we modify the arguments of Kuzmin to obtain Theorem 1.7 in a characteristic free manner. It turns out that although Lemma 3.1 can not be applied, the main combinatorial ideas of Kuzmin’s proof do work.

Consider the free \mathbb{Z} -algebra $\mathcal{Z} = \mathbb{Z}\langle x, y \rangle^+$ without unity. Write \mathcal{M} for the set of nonempty monomials (words) in x and y . For a positive integer k write $\mathcal{Z}(k)$ for the \mathbb{Z} -submodule of \mathcal{Z} generated by the $w \in \mathcal{M}$ whose total degree in y is $k-1$. It will be convenient to use the following notation, for $(a_1, \dots, a_k) \in \mathbb{N}_0^k$ set

$$[a_1, \dots, a_k] = x^{a_1} y x^{a_2} y \cdots y x^{a_k} \in \mathcal{M}.$$

The symmetric group $S_k = \text{Sym}\{1, \dots, k\}$ acts on the right linearly on $\mathcal{Z}(k)$, extending linearly the permutation action on $\mathcal{Z}(k) \cap \mathcal{M}$ given by

$$[a_1, \dots, a_k]^\pi = [a_{\pi(1)}, \dots, a_{\pi(k)}] \quad \text{for } \pi \in S_k.$$

Let \mathcal{B} denote the \mathbb{Z} -submodule of \mathcal{Z} generated by all the elements $[a_1, \dots, a_k]$ ($k \in \mathbb{N}$) such that $a_i \geq n$ for some $i \in \{1, \dots, k\}$ or $a_i = a_j$ for some $1 \leq i < j \leq k$, and by all the elements of the form $[a_1, \dots, a_k] + [a_1, \dots, a_k]^{(ij)}$ where (ij) denotes the transposition interchanging i and j for $1 \leq i < j \leq k$. We shall use the following obvious properties of \mathcal{B} :

Lemma 3.3. (i) *The \mathbb{Z} -submodule $\mathcal{B} \cap \mathcal{Z}(k)$ of $\mathcal{Z}(k)$ is S_k -stable.*

(ii) *We have the inclusions $y\mathcal{B} \subset \mathcal{B}$, $\mathcal{Z}y\mathcal{B} \subset \mathcal{B}$, $\mathcal{B}y \subset \mathcal{B}$, and $\mathcal{B}y\mathcal{Z} \subset \mathcal{B}$.*

(iii) *Let k be a positive integer, $u_1, \dots, u_{k-1} \in \mathcal{M}$ monomials such that $u_i \in y\mathcal{Z} \cap \mathcal{Z}y$ or $u_i = y$ for $i = 1, \dots, k-1$. Then \mathcal{B} contains the image of the \mathbb{Z} -module map on $\mathcal{B} \cap \mathcal{Z}(k)$ given by*

$$[a_1, \dots, a_k] \mapsto x^{a_1} u_1 x^{a_2} u_2 x^{a_3} \cdots u_{k-1} x^{a_k}.$$

(iv) *For any positive integer a , the \mathbb{Z} -submodule \mathcal{B} of \mathcal{Z} is preserved by the derivation δ_a on \mathcal{Z} defined by $\delta_a(x) = x^a$ and $\delta_a(y) = 0$.*

(v) *The factor \mathcal{Z}/\mathcal{B} is a free \mathbb{Z} -module freely generated by the images under the natural surjection $\mathcal{Z} \rightarrow \mathcal{Z}/\mathcal{B}$ of the monomials*

$$\widehat{\mathcal{M}} = \{[a_1, \dots, a_k] \mid k \in \mathbb{N} \text{ and } 0 \leq a_1 < a_2 < \cdots < a_k \leq n-1\}.$$

Proof. Statements (i), (ii), (iii), (iv) are immediate consequences of the construction of \mathcal{B} . To prove (v) note that $\mathcal{Z} = \bigoplus \mathcal{Z}(c_1, \dots, c_k)$ where the direct sum is taken over $k \in \mathbb{N}$ and $0 \leq c_1 \leq \cdots \leq c_k$, and $\mathcal{Z}(c_1, \dots, c_k)$ stands for the \mathbb{Z} -submodule generated by $[c_1, \dots, c_k]^\pi$ as π ranges over S_k . Moreover, $\mathcal{B} = \bigoplus \mathcal{B}(c_1, \dots, c_k)$ where $\mathcal{B}(c_1, \dots, c_k) = \mathcal{B} \cap \mathcal{Z}(c_1, \dots, c_k)$. Now $\mathcal{Z}(c_1, \dots, c_k) \subset \mathcal{B}$ if $c_i = c_j$ for some $i \neq j$ or if $c_i \geq n$ for some i . It is also clear that for $0 \leq a_1 < \cdots < a_k$ we have $\mathcal{Z}(a_1, \dots, a_k) = \mathbb{Z} \cdot [a_1, \dots, a_k] + \mathcal{B}(a_1, \dots, a_k)$, so the monomials in $\widehat{\mathcal{M}}$ generate the \mathbb{Z} -module \mathcal{Z} modulo \mathcal{B} . Suppose that some nontrivial \mathbb{Z} -linear combination of the elements in $\widehat{\mathcal{M}}$ belongs to \mathcal{B} . The above direct sum decompositions of \mathcal{Z} and \mathcal{B} imply then that there exist $q, k \in \mathbb{N}$, and $0 \leq a_1 < \cdots < a_k \leq n-1$ such that $q[a_1, \dots, a_k] \in \mathcal{B}(a_1, \dots, a_k)$. This means that

$$q[a_1, \dots, a_k] = \sum_{i=1}^s \varepsilon_i (w_i + w_i^{\pi_i}), \tag{4}$$

where $\varepsilon_i = \pm 1$, $w_i \in \mathcal{Z}(a_1, \dots, a_k) \cap \mathcal{M}$ and $\pi_i \in S_k$ is a transposition for $i = 1, \dots, s$. Suppose that s in (4) is minimal. Without loss of generality we may assume that $w_1 = [a_1, \dots, a_k]$ and $\varepsilon_1 = 1$. The word $w_1^{\pi_1}$ must be canceled by some summand $\varepsilon_i(w_i + w_i^{\pi_i})$ with $i \geq 2$ on the right-hand side of (4), so after a possible renumbering we have $\varepsilon_2(w_2 + w_2^{\pi_2}) = -(w_1^{\pi_1} + w_1^{\pi_1\pi_2})$. Now the term $-w_1^{\pi_1\pi_2}$ must be canceled by w_1 or by some summand $\varepsilon_i(w_i + w_i^{\pi_i})$ with $i \geq 3$. It means that the right-hand side of (4) has a subsum of the form

$$(w_1 + w_1^{\pi_1}) - (w_1^{\pi_1} + w_1^{\pi_1\pi_2}) + (w_1^{\pi_1\pi_2} + w_1^{\pi_1\pi_2\pi_3}) - \dots + (-1)^{r-1}(w_1^{\pi_1 \dots \pi_{r-1}} + w_1^{\pi_1 \dots \pi_r}), \tag{5}$$

where $w_1^{\pi_1 \dots \pi_r} = w_1$. This latter equality forces that $\pi_1 \dots \pi_r$ is the identity permutation, so r is even, and then the sum (5) is zero. So all these terms can be omitted from (4). This contradicts the minimality of s . This shows that $q[a_1, \dots, a_k]$ is not contained in \mathcal{B} . □

Lemma 3.4. *Let k be a positive integer, $a_1 \leq a_2 \leq \dots \leq a_k \in \mathbb{N}_0$, and $r \in \mathbb{N}_0$ with $a_1 + k + r > n$. Then*

$$\sum_{c_1 + \dots + c_k = r} \sum_{\pi \in S_k} [a_1 + c_{\pi(1)}, \dots, a_k + c_{\pi(k)}] \in \mathcal{B}. \tag{6}$$

Proof. Apply induction on k . In the case $k = 1$ the element in question in (6) is x^{a_1+r} , which belongs to \mathcal{B} by the assumption $a_1 + 1 + r > n$. Suppose next that $k > 1$, and the statement of the lemma holds for smaller k . The terms $[a_1 + d_1, \dots, a_k + d_k]$ in the sum (6) can be grouped into three classes:

- (A) $a_1 + d_1 < a_2$.
- (B) $a_1 + d_1 = a_2 + d_2$.
- (C) $a_1 + d_1 \geq a_2$ and $a_1 + d_1 \neq a_2 + d_2$.

The sum of the terms of type (A) is a sum of expressions of the form

$$x^{a_1+d_1} y \sum_{c_2 + \dots + c_k = r - d_1} \sum_{\pi \in \text{Sym}\{2, \dots, k\}} [a_2 + c_{\pi(2)}, \dots, a_k + c_{\pi(k)}]. \tag{7}$$

Here $a_2 + (k - 1) + (r - d_1) \geq a_1 + k + r > n$, hence by the induction hypothesis

$$\sum_{c_2 + \dots + c_k = r - d_1} \sum_{\pi \in \text{Sym}\{2, \dots, k\}} [a_2 + c_{\pi(2)}, \dots, a_k + c_{\pi(k)}]$$

belongs to \mathcal{B} . Now by Lemma 3.3(ii) we conclude that the element in (7) belongs to \mathcal{B} . The terms of type (B) belong to \mathcal{B} by construction of \mathcal{B} . Finally, a term $[a_1 + d_1, \dots, a_k + d_k]$ of type (C) can be paired off with the term $[a_1 + e_1, a_2 + e_2, a_3 + d_3, \dots, a_k + d_k]$ where $e_1 = a_2 - a_1 + d_2$ and $e_2 = a_1 - a_2 + d_1$ (so this is also of type (C)), and the sum of these two terms belongs to \mathcal{B} by construction of \mathcal{B} . □

Corollary 3.5. *Let k be a positive integer, $(a_1, \dots, a_k) \in \mathbb{N}_0^k$, and $r \in \mathbb{N}_0$ with $r + k > n$. Then*

$$\sum_{c_1 + \dots + c_k = r} \sum_{\pi \in S_k} [a_1 + c_{\pi(1)}, \dots, a_k + c_{\pi(k)}] \in \mathcal{B}.$$

Proof. Take a permutation $\rho \in S_k$ such that $a_{\rho(1)} \leq \dots \leq a_{\rho(k)}$. Applying ρ to the element in the statement we get

$$\sum_{c_1+\dots+c_k=r} \sum_{\pi \in S_k} [a_{\rho(1)} + c_{\pi(1)}, \dots, a_{\rho(k)} + c_{\pi(k)}],$$

which belongs to $\mathcal{B} \cap \mathcal{Z}(k)$ by Lemma 3.4. Our statement follows by Lemma 3.3(i). □

Lemma 3.6. *Suppose $1 \leq k \leq n + 1$, $w_1, \dots, w_{k-1} \in \mathcal{M}$ are monomials having positive degree in y , and $a, b \in \mathbb{N}_0$. Then*

$$x^a P_n(w_1, \dots, w_{k-1}, x, \dots, x) x^b \in \mathcal{B}. \tag{8}$$

Proof. We have $w_i = x^{a_i} u_i x^{b_i}$ where $a_i, b_i \in \mathbb{N}_0$ and $u_i \in y\mathcal{Z} \cap \mathcal{Z}y$ or $u_i = y$ ($i = 1, \dots, k - 1$). Then the element in (8) is

$$\sum_{\rho \in S_{k-1}} \left((n - k + 1)! \sum_{c_1+\dots+c_k=n-k+1} \sum_{\pi \in S_k} x^{d_1+c_{\pi(1)}} u_{\rho(1)} x^{d_2+c_{\pi(2)}} u_{\rho(2)} \dots x^{d_{k-1}+c_{\pi(k-1)}} u_{\rho(k-1)} x^{d_k+c_{\pi(k)}} \right),$$

where $d_1 = a + a_{\rho(1)}$, $d_2 = a_{\rho(2)} + b_{\rho(1)}$, $d_3 = a_{\rho(3)} + b_{\rho(2)}$, $d_{k-1} = a_{\rho(k-1)} + b_{\rho(k-2)}$ and $d_k = b_{\rho(k-1)} + b$. The summand corresponding to $\rho \in S_{k-1}$ in the outer sum is contained in \mathcal{B} by Corollary 3.5 and Lemma 3.3(iii). □

Lemma 3.7. *For any $w_1, \dots, w_n \in \mathcal{M}$, $w_0, w_{n+1} \in \mathcal{M} \cup \{1\}$ we have*

$$w_0 P_n(w_1, \dots, w_n) w_{n+1} \in \mathcal{B}.$$

Proof. By Lemma 3.3(ii) it is sufficient to deal with the case $w_0 = x^a$, $w_{n+1} = x^b$. We may assume that w_1, \dots, w_{k-1} have positive degree in y and $w_{k-1+j} = x^{c_j}$ for $j = 1, \dots, n - k + 1$. If $n - k + 1 = 0$ or all the $c_j = 1$ then we are done by Lemma 3.6. Suppose next that $n - k + 1 > 0$, $c_1, \dots, c_l > 1$ with $l \geq 1$, and $c_{l+1} = \dots = c_{n-k+1} = 1$. By induction on l we show that $x^a P_n(w_1, \dots, w_{k-1}, x^{c_1}, \dots, x^{c_l}, x, \dots, x) x^b \in \mathcal{B}$. By the induction hypothesis (or by Lemma 3.6 when $l = 1$)

$$f = x^a P_n(w_1, \dots, w_{k-1}, x^{c_1}, \dots, x^{c_{l-1}}, x, \dots, x) x^b \in \mathcal{B},$$

hence by Lemma 3.3(iv) $\delta_{c_l}(f) \in \mathcal{B}$. We have

$$\begin{aligned} \delta_{c_l}(f) &= ax^{a+c_l-1} P_n(w_1, \dots, w_{k-1}, x^{c_1}, \dots, x^{c_{l-1}}, x, \dots, x) x^b \\ &+ \sum_{i=1}^{k-1} x^a P_n(w_1, \dots, \delta_{c_l}(w_i), \dots, w_{k-1}, x^{c_1}, \dots, x^{c_{l-1}}, x, \dots, x) x^b \\ &+ \sum_{j=1}^{l-1} c_j x^a P_n(w_1, \dots, w_{k-1}, x^{c_1}, \dots, x^{c_j+c_l-1}, \dots, x^{c_{l-1}}, x, \dots, x) x^b \\ &+ (n - k - l + 2) x^a P_n(w_1, \dots, w_{k-1}, x^{c_1}, \dots, x^{c_l}, x, \dots, x) x^b \\ &+ bx^a P_n(w_1, \dots, w_{k-1}, x^{c_1}, \dots, x^{c_{l-1}}, x, \dots, x) x^{b+c_l-1}. \end{aligned}$$

All terms other than $(n - k - l + 2)x^a P_n(w_1, \dots, w_{k-1}, x^{c_1}, \dots, x^{c_l}, x, \dots, x)x^b$ on the right-hand side above belong to \mathcal{B} by the induction hypothesis. Taking into account that \mathcal{Z}/\mathcal{B} is torsion free by Lemma 3.3(v) we conclude the desired inclusion

$$x^a P_n(w_1, \dots, w_{k-1}, x^{c_1}, \dots, x^{c_l}, x, \dots, x)x^b \in \mathcal{B}. \quad \square$$

For $\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{N}_0^m$ denote by $P_\lambda(x_1, \dots, x_m) \in \mathbb{Z}\langle x_1, \dots, x_m \rangle$ the multihomogeneous component of $(x_1 + \dots + x_m)^n$ having \mathbb{Z}^m -degree λ .

Corollary 3.8. *For any $m \in \mathbb{N}$, $w_1, \dots, w_m \in \mathcal{M}$, $w_0, w_{m+1} \in \mathcal{M} \cup \{1\}$ and for any $\lambda \in \mathbb{N}_0^m$ we have that*

$$w_0 P_\lambda(w_1, \dots, w_m) w_{m+1} \in \mathcal{B}.$$

Proof. We have the equality

$$P_\lambda(x_1, \dots, x_m) = \frac{1}{\prod_{i=1}^m (\lambda_i!)} P_n(\underbrace{x_1, \dots, x_1}_{\lambda_1}, \dots, \underbrace{x_m, \dots, x_m}_{\lambda_m}).$$

Therefore the statement follows from Lemma 3.7 by Lemma 3.3(v). □

Proposition 3.9. *The ideal $I_{n,2}$ is contained in the subspace $\mathbb{F} \otimes_{\mathbb{Z}} \mathcal{B}$ of $\mathbb{F}\langle x, y \rangle$.*

Proof. The ideal $I_{n,2}$ is spanned as an \mathbb{F} -vector space by elements of the form

$$w_0(c_1 w_1 + \dots + c_m w_m)^n w_{m+1},$$

where the w_i are monomials in x and y and they have positive total degree for $i = 1, \dots, m$, and $c_1, \dots, c_m \in \mathbb{F}$. Since we have the equality

$$(c_1 w_1 + \dots + c_m w_m)^n = \sum_{\lambda \in \mathbb{N}_0^m, \lambda_1 + \dots + \lambda_m = n} c_1^{\lambda_1} \dots c_m^{\lambda_m} P_\lambda(w_1, \dots, w_m),$$

our statement follows from Corollary 3.8. □

Proof of Theorem 1.7. By Lemma 3.3(v) the monomials

$$\{x^{a_1} y x^{a_2} y x^{a_3} \dots y x^{a_k} \mid 0 \leq a_1 < a_2 < \dots < a_k \leq n - 1\}$$

are linearly independent in $\mathcal{F}_2 = \mathbb{F}\langle x, y \rangle$ modulo the subspace $\mathbb{F} \otimes_{\mathbb{Z}} \mathcal{B}$. Since $\mathbb{F} \otimes_{\mathbb{Z}} \mathcal{B}$ contains the ideal $I_{n,2}$ by Proposition 3.9, our statement follows. □

References

- [Belov 1992] A. J. Belov, “Some estimations for nilpotence of nil-algebras over a field of an arbitrary characteristic and height theorem”, *Comm. Algebra* **20**:10 (1992), 2919–2922. MR Zbl
- [Belov and Kharitonov 2012] A. Y. Belov and M. I. Kharitonov, “Subexponential estimates in Shirshov’s height theorem”, *Mat. Sb.* **203**:4 (2012), 81–102. In Russian; translation in *Sb. Math.* **203**:3-4 (2012), 534–553. MR Zbl
- [De Concini and Procesi 2017] C. De Concini and C. Procesi, *The invariant theory of matrices*, University Lecture Series **69**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

- [Derksen and Makam 2017a] H. Derksen and V. Makam, “Generating invariant rings of quivers in arbitrary characteristic”, *J. Algebra* **489** (2017), 435–445. MR Zbl
- [Derksen and Makam 2017b] H. Derksen and V. Makam, “Polynomial degree bounds for matrix semi-invariants”, *Adv. Math.* **310** (2017), 44–63. MR Zbl
- [Domokos 2002] M. Domokos, “Finite generating system of matrix invariants”, *Math. Pannon.* **13**:2 (2002), 175–181. MR Zbl
- [Donkin 1992] S. Donkin, “Invariants of several matrices”, *Invent. Math.* **110**:2 (1992), 389–401. MR Zbl
- [Drensky and Formanek 2004] V. Drensky and E. Formanek, *Polynomial identity rings*, Birkhäuser, Basel, 2004. MR Zbl
- [Formanek 1986] E. Formanek, “Generating the ring of matrix invariants”, pp. 73–82 in *Ring theory* (Antwerp, 1985), edited by F. M. J. Van Oystaeyen, Lecture Notes in Math. **1197**, Springer, 1986. MR Zbl
- [Formanek 1991] E. Formanek, *The polynomial identities and invariants of $n \times n$ matrices*, CBMS Regional Conf. Ser. in Math. **78**, Amer. Math. Soc., Providence, RI, 1991. MR Zbl
- [Guralnick et al. 2010] R. M. Guralnick, L. W. Small, and E. Zelmanov, “Nil subrings of endomorphism rings of finitely generated modules over affine PI-rings”, *J. Algebra* **324**:11 (2010), 3044–3047. MR Zbl
- [Ivanyos et al. 2017] G. Ivanyos, Y. Qiao, and K. V. Subrahmanyam, “Non-commutative Edmonds’ problem and matrix semi-invariants”, *Comput. Complexity* **26**:3 (2017), 717–763. MR Zbl
- [Jacobson 1945] N. Jacobson, “Structure theory for algebraic algebras of bounded degree”, *Ann. of Math. (2)* **46** (1945), 695–707. MR Zbl
- [Kaplansky 1946] I. Kaplansky, “On a problem of Kurosch and Jacobson”, *Bull. Amer. Math. Soc.* **52** (1946), 496–500. MR Zbl
- [Klein 2000] A. A. Klein, “Bounds for indices of nilpotency and nility”, *Arch. Math. (Basel)* **74**:1 (2000), 6–10. MR Zbl
- [Kuzmin 1975] E. N. Kuzmin, “О теореме Нагаты–Хигмана”, pp. 101–107 in *Математические структуры. Вычислительная математика. Математическое моделирование*, edited by B. Sendov, Bulgarian Acad. Sci., Sofia, 1975.
- [Lopatin 2005] A. A. Lopatin, “Relatively free algebras with the identity $x^3 = 0$ ”, *Comm. Algebra* **33**:10 (2005), 3583–3605. MR Zbl
- [Lopatin 2012] A. A. Lopatin, “On the nilpotency degree of the algebra with identity $x^n = 0$ ”, *J. Algebra* **371** (2012), 350–366. MR Zbl
- [Lopatin 2013] A. A. Lopatin, “Matrix identities with forms”, *J. Pure Appl. Algebra* **217**:11 (2013), 2056–2075. MR Zbl
- [Nagata 1952] M. Nagata, “On the nilpotency of nil-algebras”, *J. Math. Soc. Japan* **4** (1952), 296–301. MR Zbl
- [Newstead 1978] P. E. Newstead, *Introduction to moduli problems and orbit spaces*, Tata Inst. Fundamental Res. Lectures on Math. and Phys. **51**, Tata Inst. Fundamental Res., Bombay, 1978. MR Zbl
- [Procesi 1976] C. Procesi, “The invariant theory of $n \times n$ matrices”, *Advances in Math.* **19**:3 (1976), 306–381. MR Zbl
- [Razmyslov 1974] Y. P. Razmyslov, “Trace identities of full matrix algebras over a field of characteristic zero”, *Izv. Akad. Nauk SSSR Ser. Mat.* **38**:4 (1974), 723–756. In Russian; translation in *Math. USSR-Izv.* **8**:4 (1974), 727–760. MR Zbl
- [Zubkov 1996] A. N. Zubkov, “On a generalization of the Razmyslov–Procesi theorem”, *Algebra i Logika* **35**:4 (1996), 433–457. In Russian; translation in *Algebra Logic* **35**:4 (1996), 241–254. MR Zbl

Communicated by Michel Van den Bergh

Received 2017-06-27 Accepted 2018-03-29

domokos.matyas@renyi.mta.hu

MTA Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences,
Budapest, Hungary

Arithmetic functions in short intervals and the symmetric group

Brad Rodgers

We consider the variance of sums of arithmetic functions over random short intervals in the function field setting. Based on the analogy between factorizations of random elements of $\mathbb{F}_q[T]$ into primes and the factorizations of random permutations into cycles, we give a simple but general formula for these variances in the large q limit for arithmetic functions that depend only upon factorization structure. From this we derive new estimates, quickly recover some that are already known, and make new conjectures in the setting of the integers.

In particular we make the combinatorial observation that any function of this sort can be explicitly decomposed into a sum of functions u and v , depending on the size of the short interval, with u making a negligible contribution to the variance, and v asymptotically contributing diagonal terms only.

This variance evaluation is closely related to the appearance of random matrix statistics in the zeros of families of L -functions and sheds light on the arithmetic meaning of this phenomenon.

1. Historical background and motivation

The purpose of this paper is to explore a connection between two well-known phenomena in number theory: that the zeros of a family of L -functions distribute like the eigenvalues of a random matrix and that the prime factors of a random integer distribute like the cycles of a random permutation. We use this connection to give a general yet simple description for the statistical behavior of sums of arithmetic functions over short intervals. The results that we ultimately prove will make use of a function field analogy: they concern arithmetic functions defined on $\mathbb{F}_q[T]$ rather than the integers and we will require that $q \rightarrow \infty$. We begin this section however with a discussion of some historical conjectures and heuristics from the integers that motivate what follows. A statement of the most important results we prove may be found at the beginning of Section 3 — our main results are Theorems 3.1 and 3.2 along with Corollary 3.5. Key use is made of a combinatorial variant of the explicit formula of Weil, Theorem 7.1, which may be of independent interest.

We recall the following conjectures:

Conjecture 1.1 [Good and Churchhouse 1968]. *As $X \rightarrow \infty$, for $H = X^\delta$ with $\delta \in (0, 1)$,*

$$\frac{1}{X} \int_X^{2X} \left(\sum_{x \leq n \leq x+H} \mu(n) \right)^2 dx \sim \frac{6}{\pi^2} H.$$

MSC2010: primary 11M50; secondary 11N37, 11T55.

Keywords: arithmetic in function fields, random matrices, the symmetric group.

Conjecture 1.2 [Goldston and Montgomery 1987]. *As $X \rightarrow \infty$ for $H = X^\delta$ with $\delta \in (0, 1)$,*

$$\frac{1}{X} \int_X^{2X} \left(\sum_{x \leq n \leq x+H} \Lambda(n) - H \right)^2 dx \sim H(\log X - \log H).$$

In both conjectures, we consider random $x \in [X, 2X]$ and seek to compute the variance of the sum of an arithmetic function, $\mu(n)$ or $\Lambda(n)$, over the random short interval $[x, x + H]$. Here $\mu(n)$ is the Möbius function, which oscillates around the value 0, and $\Lambda(n)$ is the von Mangoldt function which has an average value of 1, by the prime number theorem. Similar conjectures can be made for, for instance, the higher order von Mangoldt functions $\Lambda_j(n)$ [Rodgers 2015] or the k -fold divisor function $d_k(n)$ [Keating et al. 2018], the latter of which is conjectured to display a very curious series of “phase changes” as the parameter δ varies. These conjectures are known to be closely related to the conjectural phenomenon that the zeros of families of L -functions tends to distribute like the eigenvalues of certain random matrices (see [Katz and Sarnak 1999] for an exposition on the latter phenomenon).

In the past few years, beginning with the work of Keating and Rudnick [2014], function field variants of these conjectures have been proved. (In some cases the function field theorems have in fact motivated new conjectures.) In order to state these function field results, we make use of a well-known dictionary between the integers \mathbb{Z} and the ring of polynomials over a finite field, that is $\mathbb{F}_q[T]$. To review this dictionary and fix some of our notation:

- The collection of monic polynomials, \mathcal{M} , takes the place of positive integers.
- The degree, $\deg(f)$, of $f \in \mathcal{M}$ takes the place of $\log n$ for $n \in \mathbb{N}$.
- The collection of degree n monic polynomials, \mathcal{M}_n , takes the place of integers lying in a dyadic interval $[X, 2X]$.
- Irreducible polynomials take the role of primes.
- For $f \in \mathcal{M}$ and $h < \deg(f)$, the set $I(f; h) := \{g \in \mathcal{M} : \deg(f - g) \leq h\}$ is a short interval around the polynomial f , playing the role of $[x, x + H]$. (Here h may be thought of as corresponding to $\log H$.)

The reader should verify that $|\mathcal{M}_n| = q^n$, while $|I(f; h)| = q^{h+1}$. (Note that in the notation above, we have suppressed a dependence on the parameter q .)

This set up is explained more extensively in, for instance, the ICM address of Rudnick [2014] or the book of Rosen [2015]. We have the following analogues of Conjectures 1.1 and 1.2:

Theorem 1.3 [Rudnick 2014; Bae et al. 2015]. *For fixed $0 \leq h \leq n - 5$, as $q \rightarrow \infty$,*

$$\frac{1}{q^n} \sum_{f \in \mathcal{M}_n} \left| \sum_{g \in I(f; h)} \mu(g) \right|^2 \sim q^{h+1}. \quad (1)$$

Theorem 1.4 [Keating and Rudnick 2014]. *For fixed $0 \leq h \leq n - 5$, as $q \rightarrow \infty$,*

$$\frac{1}{q^n} \sum_{f \in \mathcal{M}_n} \left| \sum_{g \in I(f; h)} \Lambda(g) - q^{h+1} \right|^2 \sim q^{h+1}(n - h - 2). \quad (2)$$

For $g \in \mathcal{M}$, the Möbius function $\mu(g)$ is defined in analogy with the integers by $\mu(g) = (-1)^\ell$ if g is squarefree (that is g has no repeated factors) and $g = P_1 \cdots P_\ell$ in its prime factorization, and $\mu(g) = 0$ if g is squareful¹ (that is g is not squarefree). Likewise $\Lambda(g) = \deg(P)$ if $g = P^k$ for a prime P and a power $k \geq 1$, and $\Lambda(g) = 0$ otherwise.

We introduce a notation to write these results more succinctly. For a function $\eta : \mathcal{M}_n \rightarrow \mathbb{C}$, we define its mean value by

$$\mathbb{E}_{f \in \mathcal{M}_n} \eta(f) := \frac{1}{q^n} \sum_{f \in \mathcal{M}_n} \eta(f), \tag{3}$$

and its variance by

$$\text{Var}_{f \in \mathcal{M}_n} (\eta(f)) := \frac{1}{q^n} \sum_{f \in \mathcal{M}_n} |\eta(f) - \mathbb{E}_{\mathcal{M}_n} \eta|^2. \tag{4}$$

Note that both the mean value and variance typically depend on the size of the field q . As a test of notation, the reader may easily verify that

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} 1 \right) = 0. \tag{5}$$

Likewise we see that (1) may be rewritten

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} \mu(g) \right) \sim q^{h+1}, \tag{6}$$

and (2)

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} \Lambda(g) \right) \sim q^{h+1}(n - h - 2), \tag{7}$$

as $q \rightarrow \infty$.

We may add another recent result to this list, due to Keating, the author, Roditty-Gershon, and Rudnick [Keating et al. 2018], for the k -fold divisor function, which is defined in analogy with the integers by $d_k(f) := |\{(a_1, \dots, a_k) \in \mathcal{M}^k : f = a_1 \cdots a_k\}|$.

Theorem 1.5. *For fixed positive integer k , and fixed $0 \leq h \leq n - 5$, as $q \rightarrow \infty$,*

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} d_k(g) \right) = q^{h+1} \mathcal{I}_k(n, n - h - 2) + O(q^{h+1/2}), \tag{8}$$

where $\mathcal{I}_k(m, N)$ is the count of lattice points $(x_{ij}) \in (\mathbb{Z})^{k^2}$ satisfying each of the following conditions:

- (i) $0 \leq x_{ij} \leq N$ for all $1 \leq i, j \leq k$.
- (ii) $x_{11} + \cdots + x_{kk} = m$.

¹There is a closely related terminology “square-full”, which means something quite different — namely that for prime P , if $P \mid g$, we have $P^2 \mid g$ also. The distinction is important to keep in mind. Square-full numbers will not play a role in this paper.

(iii) *The array x_{ij} is weakly decreasing across columns and down rows. That is,*

$$\begin{array}{cccc}
 x_{11} & \geq & x_{12} & \geq \cdots \geq & x_{1k} \\
 |V & & |V & & |V \\
 x_{21} & \geq & x_{22} & \geq \cdots \geq & x_{2k} \\
 |V & & |V & & |V \\
 \vdots & & \vdots & \ddots & \vdots \\
 |V & & |V & & |V \\
 x_{k1} & \geq & x_{k2} & \geq \cdots \geq & x_{kk}.
 \end{array}$$

Of the evaluations (5)–(8), only (5) may be proved easily (in fact trivially). Nonetheless, the estimate in (6), while deep, at least has a heuristic meaning that is easy to understand; it is just the claim that in expanding the variance into a sum over two indices, the Möbius function is so oscillatory that off-diagonal terms make no contribution. That is, (6) may be understood heuristically in the following way:

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} \mu(g) \right) = \frac{q^{h+1}}{q^n} \sum_{\substack{g_1, g_2 \\ \deg(g_1 - g_2) \leq h}} \mu(g_1)\mu(g_2) \approx \frac{q^{h+1}}{q^n} \sum_{\substack{g_1, g_2 \in \mathcal{M}_n \\ g_1 = g_2}} \mu(g_1)\mu(g_2).$$

See for instance [Ng 2008] for a broader application of this heuristic in the setting of the integers, and see [Carmon and Rudnick 2014; Carmon 2015] for estimates of off-diagonal sums of the Möbius function in the function field setting.

The evaluation of the k -fold divisor function in (8) is obviously of a more complicated sort, even heuristically. In particular it may be seen that $\mathcal{I}_k(n; n - h - 2)$ is a piecewise polynomial, and for $k \geq 3$ as h ranges from 0 to $n - 5$, it exhibits several phase changes in its behavior in various ranges of h (see [Keating et al. 2018, §4]). The arithmetic reason for these phase changes in particular is rather mysterious.

Nonetheless, we make the following claim: (8) may be understood arithmetically as nothing more complicated than a combination of the phenomena that give rise to (5) and (6). For any degree n and short interval size h , we will observe that we may decompose

$$d_k(f) = u(f) + v(f),$$

where u and v are arithmetic functions, with $u(f)$ regular enough within the specified short intervals that (in analogy with (5)),

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} u(g) \right) = o(q^{h+1}), \tag{9}$$

while $v(f)$ is oscillatory and

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} v(g) \right) \sim q^{h+1} \cdot \frac{1}{q^{n+1}} \sum_{g \in \mathcal{M}_n} |v(g)|^2. \tag{10}$$

That is, as with the Möbius function, only diagonal terms contribute to its variance, in analogy with (6).

From Cauchy–Schwarz, it follows that

$$\mathrm{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} d_k(g) \right) \sim q^{h+1} \cdot \frac{1}{q^n} \sum_{g \in \mathcal{M}_n} |v(g)|^2.$$

This decomposition is explicit, based on symmetric function theory, and is given below — the quantity $\mathcal{I}_k(n; n - h - 2)$ may be recovered from it. That (9) holds for our function u will be a relatively shallow fact (having to do with the number of zeros of a certain family of L -functions), and one may think of u as being the largest piece of d_k with enough regularity that (9) holds for this reason. The intricacies of the variance estimate in (8) may thus be thought of as resulting from the fact that this decomposition changes for various values of n and h .

Such a decomposition is not limited to the k -fold divisor function. Any arithmetic functions whose value depends only upon the factorization type of its argument may be decomposed in this way and the variance of its sum over short intervals may thus be evaluated. What we mean by factorization type is defined formally below; roughly this is the size of all prime factors, listed with multiplicity. The functions $\mu(f)$, $\Lambda(f)$, $\Lambda_j(f)$, and $d_k(f)$ are all examples to which the result may be applied.

The evaluation of variance for such a general class of function is closely related to the known phenomena that the zeros of L -functions distribute like the eigenvalues of a random matrix. Indeed, the result we prove may be seen to be an equivalent restatement of an equidistribution result of Katz, Theorem 4.2 below. (We make use of Katz’s theorem in our proof, and so we *do not* arrive at an independent proof of it however.)

We will use this general variance evaluation to recover several of the results that have been mentioned above with relatively little extra work and to derive new results that seem difficult by other means. New conjectures in the setting of the integers are put forward based on these results. Perhaps of particular interest, we consider sums of the function $\omega(n)$, counting prime factors; based on a function field model, we conjecture that the variance of sums of this function is somewhat smaller than a naive heuristic would lead one to believe.

In addition to yielding a pleasant general formula, the decomposition results of this paper help elucidate why random matrix universality should make an appearance in number theory. A complementary perspective as to the arithmetic reasons for the appearance of random matrix theory in number theory, dealing with the integers themselves, has appeared in the work of Bogomolny and Keating [1995; 1996] and in work of Conrey and Keating [2015a; 2015b; 2015c; 2016]. It would be very interesting to see if the combinatorial decompositions in the present paper can be extended to the setting of the integers in a way consistent with various conjectures that have been made there.

We finally note a recent application of our main results to algebraic geometry proper; by combining Theorem 3.1 with other work of their own, Hast and Matei [2016] have given a geometric interpretation of this result. Indeed, it may be possible and it would be interesting to prove Theorem 3.1 of this paper directly through algebro-geometric means.

2. The symmetric group and factorization type

2A. The decomposition described in Section 1 and the corresponding estimates for variance hinge upon a well-known analogy between the prime factors of a random integer or element of $\mathbb{F}_q[T]$ and the cycles of a random permutation. (Later an application of symmetric function theory to the zeros of L -functions will play an equally important and dual role.)

We begin by recalling how it is that factorizations over $\mathbb{F}_q[T]$ resemble the cycles of permutations.

Recall that \mathcal{M}_n , the collection of monic polynomials of degree n , consists of q^n elements. Recall also that a partition λ of a positive integer n is defined to be a sequence of nonincreasing positive integers $(\lambda_1, \lambda_2, \dots, \lambda_k)$ such that for $|\lambda| := \lambda_1 + \dots + \lambda_k$ we have $|\lambda| = n$. We will also use the notation $\lambda \vdash n$ to indicate that λ is a partition of n .

Definition 2.1. For an element f of \mathcal{M}_n that is squarefree, if f has prime factorization $f = P_1 P_2 \cdots P_k$ with $\deg P_1 \geq \deg P_2 \geq \dots$, we define the *factorization type* to be the partition of n given by

$$\tau_f = (\deg P_1, \dots, \deg P_k).$$

For f that is not squarefree (i.e., squareful) we adopt the convention that $\tau_f = \emptyset$ (the empty partition).

In the above definition we have fixed our attention on the squarefrees because as $q \rightarrow \infty$ nearly all elements of \mathcal{M}_n are squarefree [Carlitz 1932, §6]; see also [Rosen 2002, Proposition 2.3] or [Weiss 2013, Theorem 4.1]:

$$\frac{1}{q^n} \#\{f \in \mathcal{M}_n : f \text{ squarefree}\} = 1 - O\left(\frac{1}{q}\right). \tag{11}$$

Note that likewise any element σ of the symmetric group \mathfrak{S}_n on n elements can be written uniquely as a product of disjoint cycles: $\sigma = \sigma_1 \sigma_2 \cdots \sigma_k$. Denote the lengths of the cycles by $|\sigma_i|$. For instance $|(245)| = 3$, where we have used cycle notation to represent the permutation.

Definition 2.2. For an element $\sigma \in \mathfrak{S}_n$, with $\sigma = \sigma_1 \sigma_2 \cdots \sigma_k$ and $|\sigma_1| \geq |\sigma_2| \geq \dots$ we define the *cycle type* to be the partition of n given by

$$\tau_\sigma = (|\sigma_1|, \dots, |\sigma_k|).$$

It is well known that as $q \rightarrow \infty$ the distribution over \mathcal{M}_n of factorization types tends to the distribution of cycle types in \mathfrak{S}_n [Andrade et al. 2015]:

Proposition 2.3. For a partition $\lambda \vdash n$,

$$\lim_{q \rightarrow \infty} \mathbb{P}_{f \in \mathcal{M}_n}(\tau_f = \lambda) = \mathbb{P}_{\sigma \in \mathfrak{S}_n}(\tau_\sigma = \lambda).$$

Here and in what follows we have used elementary probabilistic notation, for instance

$$\mathbb{P}_{f \in \mathcal{M}_n}(\tau_f = \lambda) := \frac{1}{q^n} \#\{f \in \mathcal{M}_n : \tau_f = \lambda\}.$$

There is a well-known expression for the probability that a random permutation has a cycle structure λ , due to Cauchy. We use the standard partition frequency notation $\lambda = \langle 1^{m_1} 2^{m_2} \dots j^{m_j} \rangle$; this means for $\lambda = (\lambda_1, \lambda_2, \dots)$, that m_1 of the parts of λ are equal to 1, m_2 are equal to 2, etc. So if $\tau_\sigma = \langle 1^{m_1} 2^{m_2} \dots j^{m_j} \rangle$, σ has m_1 1-cycles, m_2 2-cycles, etc. With this notation, Cauchy's result is that

$$\mathbb{P}_{\sigma \in \mathfrak{S}_n}(\tau_\sigma = \lambda) = \mathbf{p}(\lambda), \quad \text{where } \mathbf{p}(\lambda) := \prod_{i=1}^j \frac{1}{i^{m_i} m_i!}. \tag{12}$$

It is worth mentioning a recent result of Andrade, Bary-Soroker, and Rudnick [Andrade et al. 2015] that has generalized this picture. They show that the factorization types of a random polynomial f and a shift $f + \alpha$ become independent as $q \rightarrow \infty$:

Theorem 2.4 (Andrade, Bary-Soroker, and Rudnick). *For partitions λ and $\nu \vdash n$, uniformly for $\deg(\alpha) < n$,*

$$\mathbb{P}_{f \in \mathcal{M}_n}(\tau_f = \lambda, \tau_{f+\alpha} = \nu) = \mathbf{p}(\lambda)\mathbf{p}(\nu) + O(q^{-1/2}).$$

In fact they demonstrate this independence even for multiple shifts: the factorization types of $f + \alpha_1, f + \alpha_2, \dots, f + \alpha_k$ become independent as well.

2B. In this paper we will be concerned with the distribution of arithmetic functions $a : \mathcal{M} \mapsto \mathbb{C}$ such that $a(f)$ depends only upon the size and exponents of the prime factors of f . To make a more formal definition, if f has prime factorization $P_1^{e_1} \dots P_k^{e_k}$, with P_1, \dots, P_k monic primes, we call the data $(\deg P_1, e_1; \dots; \deg P_k, e_k)$, the *extended factorization type* of f . We will be concerned with functions a such that $a(f)$ is defined for all monic $f \in \mathbb{F}_q[T]$ for all q and such that the value $a(f)$ depends only on the extended factorization type of f ; we call such functions *factorization functions*. The class of factorization functions includes, for instance, the Möbius function $\mu(f)$, the von Mangoldt function $\Lambda(f)$, the count-of-divisors function $d(f)$, the indicator function of degree n polynomials $\mathbf{1}[\deg(f) = n]$, the indicator function of squarefree polynomials $\mu(n)^2$, etc. It does not include Dirichlet characters, for instance.

It is evident that for each n , the linear space of factorization functions supported on degree n polynomials is of finite dimension. The space of factorization functions supported on degree n squarefree polynomials is likewise of (smaller) finite dimension. In invoking the symmetric group, Proposition 2.3 and Theorem 2.4 suggest that the space of factorization functions has an important basis that may provide useful information: namely the irreducible characters of \mathfrak{S}_n .

In describing how such characters may be applied to elements of $\mathbb{F}_q[T]$, we suppose the reader is familiar with the most basic outlines of representation theory over \mathfrak{S}_n , along the lines of, for instance, Chapter 4 of [Fulton and Harris 1991]. We recall that the space of class functions of \mathfrak{S}_n are those functions $a(\sigma)$ with a value depending only on the cycle type the permutation σ and that a basis for such functions is given by the irreducible characters, for which we use the notation²

$$X^\lambda(\sigma).$$

²We use the letter X rather than the more traditional χ to distinguish these characters from Dirichlet characters which will make an appearance later on.

If σ has cycle type τ , sometimes instead of $X^\lambda(\sigma)$ we write $X^\lambda(\tau)$, since X^λ depends only on cycle type. Such characters are indexed by partitions $\lambda \vdash n$, and there is a one-to-one correspondence between irreducible characters of \mathfrak{S}_n and partitions of n . These characters satisfy the orthogonality relation:

$$\mathbb{E}_{\sigma \in \mathfrak{S}_n} X^{\lambda_1}(\sigma) X^{\lambda_2}(\sigma) = \delta_{\lambda_1 = \lambda_2}. \tag{13}$$

For an element $f \in \mathbb{F}_q[T]$, for $\lambda \vdash n$, we define

$$X^\lambda(f) := \begin{cases} X^\lambda(\tau_f) & \text{if } \deg(f) = n \text{ and } f \text{ is squarefree,} \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see from this definition that for any factorization function a , there exists a unique decomposition

$$a(f) = \sum_{\lambda} \hat{a}_\lambda X^\lambda(f) + b(f), \tag{14}$$

where $b(f)$ is a function supported on the squarefuls, \hat{a}_λ are constants that depend on the function a and are defined by this relation, and the sum is over all partitions. (Note that for any particular f of degree n , the sum in (14) will be a finite sum over $\lambda \vdash n$, all other terms in the summand being 0.)

Note that from Proposition 2.3 and the orthogonality relation (13) we may equivalently define the coefficients \hat{a}_λ for $\lambda \vdash n$ by

$$\hat{a}_\lambda := \lim_{q \rightarrow \infty} \frac{1}{q^n} \sum_{f \in \mathcal{M}_n} a(f) X^\lambda(f).$$

For instance, since $X^{(n)}$ is the trivial character, we have $\mathbb{E}_{f \in \mathcal{M}_n} a(f) \rightarrow \hat{a}_{(n)}$, as $q \rightarrow \infty$.³

Hast and Matei [2016, Theorem 4.4] have considered a class of functions called arithmetic functions of von Mangoldt type, which is similar to the class of factorization functions defined here (see [Hast and Matei 2016] for details of the definition). For this class of functions, Hast and Matei prove what may be thought of as a first-order short interval analogue of Andrade, Bary-Soroker, and Rudnick’s result in Theorem 2.4. Rewritten in the notation used above:

Theorem 2.5. *For a fixed arithmetic function of von Mangoldt type $a(f)$, and fixed $n \geq 4, 1 \leq h \leq n - 3$,*

$$\mathbb{E}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} a(g) \right)^2 = q^{2(h+1)} (\mathbb{E}_{f \in \mathcal{M}_n} a(f))^2 + O(q^{(h+1)}), \tag{15}$$

as $q \rightarrow \infty$.

This is sufficient to recover the upper bound of $O(q^{h+1})$ for the variance computed by Theorem 1.4 of Keating and Rudnick, though not the constant $n - h - 2$.

This result of Hast and Matei is of interest perhaps especially because their methods are rather different than ours — in particular they do not require any of the facts about L -functions that we will make use of in what follows. Other related recent papers with a perspective similar to Hast and Matei’s, making use of

³Alternatively, if $\lambda \vdash n$, and $A : S_n \rightarrow \mathbb{C}$ is the function induced by a , then \hat{a}_λ is also equal to the Fourier coefficient of A .

the connection between polynomials over a finite field and the symmetric group to investigate arithmetic functions defined on \mathcal{M} , include [Church et al. 2014; Gadish 2017].

3. A statement of main results

3A. We are now in a position to state our main results.

Theorem 3.1. For $a(f)$ a fixed factorization function, and fixed h and n with $0 \leq h \leq n - 5$,

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} a(g) \right) = q^{h+1} \sum_{\substack{\lambda \vdash n \\ \lambda_1 \leq n-h-2}} |\hat{a}_\lambda|^2 + O(q^{h+1/2}). \tag{16}$$

Here the coefficients \hat{a}_λ are defined by the expansion (14), and the sum in (16) is over all partitions $(\lambda_1, \lambda_2, \dots)$ of n such that λ_1 (and therefore every λ_i) is no more than $n - h - 2$.

In (16), the implied constant of the error term depends on h, n and the factorization function itself, so the result is only of interest as $q \rightarrow \infty$.

In Section 9 we compute the coefficients in the expansion (14) for the factorization functions $\mu(f)$, $\Lambda(f)$, $\Lambda_j(f)$ and $d_k(f)$. These expansions, applied in Theorem 3.1 are sufficient to recover estimates for the variance of sums of these arithmetic functions over short intervals which we have cited in Theorems 1.3, 1.4, and 1.5.

Likewise we consider the arithmetic functions $\omega(f)$, counting the number of prime factors of f , and likewise the function $\mu(f)\omega(f)$. The short interval variance of these functions is computed in Section 9 by using Theorem 3.1, and this leads us to make a conjecture in the setting of the integers which seems perhaps somewhat surprising.

Note also that Theorem 3.1 gives us a nontrivial upper bound for the variance of arithmetic functions supported on the squarefrees, though the upper bound is one which may be far from optimal. Work of Keating and Rudnick [2016] and Roditty-Gershon [2017] considers some related questions about the squarefrees (and indeed square-fulls) more carefully to get asymptotics, not only upper bounds.

The variance evaluation in Theorem 3.1 comes in part from a combinatorial analysis of random matrix integrals. In particular the already mentioned function field equidistribution theorem of Katz plays an important role in the proof.

A likewise central role is played by a combinatorial analogue of the explicit formula of Weil, relating the zeros of an L -function to certain arithmetic functions. In particular, in Section 7 and especially Theorem 7.1 we show that Schur functions of zeros of L -functions are closely related to the characters $X^\lambda(f)$ defined above.

We note the conjectural appearance of the symmetric group in other closely related contexts, for example in Dehaye’s work [≥ 2018] on moments of the Riemann zeta function. It would be of interest to pursue this connection further.

3B. The same result may be stated perhaps more strikingly along the lines advertised in Section 1. Let \mathcal{F}_n be the linear space of factorization functions supported on \mathcal{M}_n , and define \mathcal{U}_n^h to be the subspace of

factorization functions for which variance is negligible; that is,

$$\mathcal{U}_n^h := \left\{ u \in \mathcal{F}_n : \lim_{q \rightarrow \infty} \text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} u(g) \right) = o(q^{h+1}) \right\}. \tag{17}$$

We may endow \mathcal{F}_n with an inner product: for $a_1, a_2 \in \mathcal{F}_n$, we define

$$\langle a_1, a_2 \rangle := \lim_{q \rightarrow \infty} \frac{1}{q^n} \sum_{f \in \mathcal{M}_n} a_1(f) \overline{a_2(f)}. \tag{18}$$

This inner product is degenerate, but only on factorization functions supported on the squarefuls. If we decompose $\mathcal{F}_n = \mathcal{G}_n \oplus \mathcal{B}_n$, where \mathcal{G}_n is the space of factorization functions supported on squarefree monic polynomials of degree n , and \mathcal{B}_n is the space supported on squarefuls, then the equidistribution of factorization types imply that this is a proper inner product when restricted to \mathcal{G}_n .

We will show that $\mathcal{B}_n \subseteq \mathcal{U}_n^h$, and so if we define \mathcal{V}_n^h to be the orthogonal complement to \mathcal{U}_n^h inside \mathcal{G}_n , we have

$$\mathcal{F}_n = \mathcal{U}_n^h \oplus \mathcal{V}_n^h.$$

We will observe the following restatement of Theorem 3.1,

Theorem 3.2. *Let $0 \leq h \leq n - 5$ be fixed and v be a fixed factorization function from the subspace \mathcal{V}_n^h . Then*

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} v(g) \right) = q^{h+1} \langle v, v \rangle + O(q^{h+1/2}).$$

That is, for \mathcal{V}_n^h , only diagonal terms contribute to the variance, while by definition for \mathcal{U}_n^h the variance is of lower order.

Thus, this theorem implies an estimate for the variance of an arbitrary factorization function $a \in \mathcal{F}_n$, since there is a unique decomposition $a = u + v$ with $u \in \mathcal{U}_n^h$ and $v \in \mathcal{V}_n^h$. Indeed,

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} u(g) \right) = o(q^{h+1}),$$

so (using Cauchy–Schwarz to bound covariance),

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} a(g) \right) = q^{h+1} \langle v, v \rangle + o(q^{h+1}). \tag{19}$$

The spaces \mathcal{U}_n^h and \mathcal{V}_n^h can be characterized explicitly.

Proposition 3.3. *We have*

$$\mathcal{U}_n^h = \mathcal{A}_n^h \oplus \mathcal{B}_n,$$

where

$$\mathcal{A}_n^h := \text{span}\{X^\lambda(f) : \lambda \vdash n, \lambda_1 \geq n - h - 1\},$$

$$\mathcal{B}_n := \{b(f) : b \in \mathcal{F}_n \text{ is supported on squareful elements}\}.$$

Furthermore

$$\mathcal{V}_n^h = \text{span}\{X^\lambda(f) : \lambda \vdash n, \lambda_1 \leq n - h - 2\}.$$

This explicit decomposition is what connects Theorems 3.1 and 3.2. It is worthwhile to emphasize once again an interpretation of this result; the determination that the variance of short interval sums of functions lying in \mathcal{U}_n^h is negligible will be a relatively simple fact to verify — we will see that functions lying in this space are forced to be regular across short intervals owing in the end to a paucity of zeros of L -functions. The theorem tells us that outside this first obstruction, factorization functions otherwise behave in an oscillatory fashion, akin to the Möbius function, when summed in a short interval.

There is another appealing way to write this decomposition, based on a suggestion by J. Ellenberg:

Proposition 3.4. *Define the space \mathcal{U}_n^h as in the start of this subsection. Then \mathcal{U}_n^h consists of functions $u(f)$ that can be written in the form*

$$u(f) = \sum_{\substack{\delta \mid f \\ \deg(\delta) \leq h+1}} \alpha(\delta) + b(f), \quad \text{for all } f \in \mathcal{M}_n, \tag{20}$$

where $\alpha(\delta)$ is a factorization function and $b(f)$ is a factorization function supported on the squarefuls.

Here the sum is over all monic polynomials δ dividing f with degree no more than $h + 1$.

Indeed, it will again follow quite easily that for all factorization functions that can be represented as truncated divisor sums in this way, the value of their sums over short intervals will remain basically constant no matter the choice of short interval, so that these sums have negligible variance. The space \mathcal{V}_n^h remains defined as the complement of \mathcal{U}_n^h , and so an interpretation of this decomposition remains the same — outside an “easy-to-find” obstruction, functions otherwise behave in an oscillatory fashion when summed in a short interval.

As a corollary of Theorem 3.2 and Proposition 3.4, we have

Corollary 3.5. *For $a(f)$ a fixed factorization function and fixed h and n with $0 \leq h \leq n - 5$,*

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} a(g) \right) = q^{h+1} \inf_{\alpha \in \mathcal{F}} \left\| a(f) - \sum_{\substack{\delta \mid f \\ \deg(\delta) \leq h+1}} \alpha(\delta) \right\|^2 + O(q^{h+1/2}), \tag{21}$$

where \mathcal{F} is the space of all factorization functions and $\|\cdot\|$ is the norm induced by the inner product (18).

Rather curiously, the minimization problem arising in the computing the right-hand side of (21) has some similarity to those which arise in connection to the Selberg sieve.

We turn to a proof of these decompositions and Theorem 3.2 in Section 11.

3C. Because Theorem 3.1 allows us to compute variances for general factorization functions, it is also straightforward to use it to compute covariance. We record a general formula for covariance in Section 10 and draw out some interesting consequences that appear to be new in the literature.

3D. A similar set of results could be developed for factorization functions in arithmetic progressions rather than short intervals, though we don’t do so here.

3E. In the next two sections we recall some background material regarding Dirichlet characters, L -functions, and symmetric function theory. We turn to the substantial portion of the proof of Theorem 3.1 in Section 8.

4. Background on Dirichlet characters and zeros of L -functions

4A. We recall a few of the basic facts about Dirichlet characters defined over $\mathbb{F}_q[T]$ that we will use. Our notation is the same as that from [Rosen 2002; Keating and Rudnick 2014; Rudnick 2014; Rodgers 2015; Keating et al. 2018] and a reader familiar with the facts from any one of those may skip this section and refer back to it as it is referenced.

In $\mathbb{F}_q[T]$, we will make use of the family of primitive even characters modulo the element T^M for powers $M \geq 1$. We call a character χ *even* if for all $c \in \mathbb{F}_q$ and all $f \in \mathbb{F}_q[T]$, we have $\chi(cf) = \chi(f)$. Recall that the number of Dirichlet characters modulo T^M is

$$\Phi(T^M) = q^M \left(1 - \frac{1}{q}\right), \tag{22}$$

the number of primitive Dirichlet characters is

$$\Phi_{\text{prim}}(T^M) = q^M \left(1 + O\left(\frac{1}{q}\right)\right), \tag{23}$$

the number of even Dirichlet characters is

$$\Phi^{\text{ev}}(T^M) = q^{M-1}, \tag{24}$$

and the number of even primitive characters is

$$\Phi_{\text{prim}}^{\text{ev}}(T^M) = q^{M-1} \left(1 + O\left(\frac{1}{q}\right)\right). \tag{25}$$

We recall that the L -function of a Dirichlet character χ is defined for $|u| < 1/q$ by

$$\mathcal{L}(u, \chi) := \sum_{f \text{ monic}} \chi(f) u^{\deg(f)} = \prod_{\substack{P \text{ monic,} \\ \text{irreducible}}} \frac{1}{1 - \chi(P) u^{\deg(P)}},$$

and that for χ nontrivial that $\mathcal{L}(u, \chi)$ is a polynomial in u , defined for $|u| \geq 1/q$ by analytic continuation. The Riemann hypothesis, in this context a theorem of Weil [1967], states that all roots of $\mathcal{L}(u, \chi)$ lie on the circles $|u| = q^{-1/2}$ or $|u| = 1$. If χ is a nontrivial character modulo a polynomial Q of degree M , then $\mathcal{L}(u, \chi)$ has no more than $M - 1$ roots, and as a well-known consequence of this and the Riemann hypothesis,

$$\sum_{f \in \mathcal{M}_n} \Lambda(f) \chi(f) = O_M(q^{n/2}). \tag{26}$$

4B. In the case that χ is a primitive character we can succinctly say more. In this case for χ modulo T^M , the polynomial $\mathcal{L}(u, \chi)$ has exactly $M - 1$ roots. Define the function λ_χ to be 1 if χ is even, and 0 otherwise. When χ is even, $\mathcal{L}(u, \chi)$ has a simple zero at $u = 1$, otherwise all zeros of this polynomial lie on the circle $|u| = q^{-1/2}$. We can record this information in a single equation; we have for primitive

characters χ and $N := \deg Q - 1 - \lambda_\chi$,

$$\mathcal{L}(u, \chi) = (1 - \lambda_\chi u) \prod_{j=1}^N (1 - q^{1/2} e^{i2\pi\vartheta_j} u) = (1 - \lambda_\chi u) \det(1 - q^{1/2} u \Theta_\chi), \tag{27}$$

where $e^{i2\pi\vartheta_1}, \dots, e^{i2\pi\vartheta_N}$ lie on the unit circle and are determined by the character χ and

$$\Theta_\chi := \text{diag}(e^{i2\pi\vartheta_1}, \dots, e^{i2\pi\vartheta_N})$$

is known as the unitarized Frobenius matrix. From logarithmic differentiation we also have the *explicit formula*,

$$\sum_{f \in \mathcal{M}_n} \Lambda(f) \chi(f) = -q^{n/2} \text{Tr } \Theta_\chi^n - \lambda_\chi. \tag{28}$$

To control the distribution of zeros, a theorem of Katz will be important for us, as it has been in all investigations of this sort since Keating and Rudnick’s [2014]. We let $\text{PU}(m)$ be the projective unitary group, the quotient of the unitary group $U(m)$ by unit modulus scalars, endowed with Haar measure, and $\text{PU}(m)^\#$ be the space of conjugacy classes of $\text{PU}(m)$, with inherited measure.

Theorem 4.1 [Katz 2013, Theorem 8.1]. *Fix $M \geq 5$. Over the family of even primitive characters $\chi \pmod{T^M}$, the conjugacy classes of the unitarized Frobenii Θ_χ become equidistributed in $\text{PU}(M - 2)^\#$ as $q \rightarrow \infty$.*

More computationally the meaning of the theorem is as follows: for any continuous class function $\phi : U(M - 2) \rightarrow \mathbb{C}$ such that $\phi(e^{i2\pi\theta} g) = \phi(g)$ for all unit scalars $e^{i2\pi\theta}$ and unitary matrices g , we have

$$\lim_{q \rightarrow \infty} \mathbb{E}_{\chi(T^M)_{\text{prim, ev}}} \phi(\Theta_\chi) = \int_{U(M-2)} \phi(g) dg,$$

as $q \rightarrow \infty$, where for typographical reasons we have written

$$\mathbb{E}_{\chi(T^M)_{\text{prim, ev}}} \phi(\Theta_\chi) := \frac{1}{\Phi_{\text{prim}}^{\text{ev}}(T^M)} \sum_{\substack{\chi(T^M) \\ \text{prim, ev}}} \phi(\Theta_\chi).$$

Indeed, Katz also considers the rate of convergence in this result, at least for a sufficiently simple function ϕ .

Theorem 4.2 [Katz 2013, Theorem 8.2]. *Fix $M \geq 5$. For a fixed class function $\phi : U(M - 2) \rightarrow \mathbb{C}$ as described above such that the map induced by ϕ from $\text{PU}(M - 2)$ to \mathbb{C} is a linear combination of irreducible characters of $\text{PU}(M - 2)$:*

$$\mathbb{E}_{\chi(T^M)_{\text{prim, ev}}} \phi(\Theta_\chi) = \int_{U(M-2)} \phi(g) dg + O(q^{-1/2}).$$

4C. The reason we will be interested in characters modulo T^M is the following involution used by Keating and Rudnick.

We let \mathcal{P}_n be the collection of degree n polynomials in $\mathbb{F}_q[T]$ and $\mathcal{P}_n^\natural := \{f \in \mathcal{P}_n : (f, T) = 1\}$. Equivalently \mathcal{P}_n^\natural is the collection of degree n polynomials with a constant coefficient that is nonzero. Our

involution is the mapping $f \mapsto f^*$ from \mathcal{P}_n^{\natural} to itself defined by

$$(a_0 + a_1T^1 + \cdots + a_nT^n)^* = a_n + a_{n-1}T + \cdots + a_0T^n. \quad (29)$$

It is straightforward to check that for f with nonzero constant coefficient,

$$(f^*)^* = f,$$

and for f and g with nonzero constant coefficient,

$$(fg)^* = f^*g^*.$$

If we extend the definition of factorization type to \mathcal{P}_n , so that for $f \in \mathcal{P}_n$ and for that scalar $c \in \mathbb{F}_q$ such that $cf \in \mathcal{M}_n$, the factorization type of f is defined to be the factorization type of cf , it follows that for $f \in \mathcal{P}_n^{\natural}$,

$$\tau_f = \tau_{f^*}. \quad (30)$$

This involution is useful for us because for $g_1, g_2 \in \mathcal{P}_n^{\natural}$,

$$\deg(g_1 - g_2) \leq h$$

if and only if

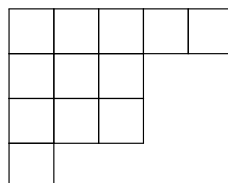
$$g_1^* - g_2^* \equiv 0 \pmod{T^{n-h}}.$$

This equivalence is easily checked. It is because of this that we may use Dirichlet characters and their L -functions to study short interval sums.

5. Background on symmetric function theory

5A. We recall some notation and well-known facts from symmetric function theory that we will use in what follows. A standard reference and introduction to the material we recall here is [Stanley 1999, Chapter 7].

We have already defined partitions and discussed their basic notation in Section 2A. One additional way to represent partitions is as a Young diagram. This is an array of left-justified boxes, with the number of boxes in each row weakly decreasing. For a partition λ , the Young diagram corresponding to λ has λ_1 boxes in its first row, λ_2 boxes in its second row, etc. For instance, the Young diagram with shape $(5, 3, 3, 1)$ is as follows:



The *dual partition* λ' is defined to be $(\lambda'_1, \lambda'_2, \dots)$ where λ'_i is the number of boxes in the i -th column of the Young diagram corresponding to λ . So in our example above $(5, 3, 3, 1)' = (4, 3, 3, 1, 1)$.

The *length of a partition*, $\ell(\lambda)$, is defined to be k , where $\lambda = (\lambda_1, \dots, \lambda_k)$. For instance $\ell(5, 3, 3, 1) = 4$.

Young diagrams may be used to write down a relatively simple expression for characters of the symmetric group in the form of the famous Murnaghan–Nakayama rule. We quickly recall it here, taking from the presentation in [Stanley 1999, §7.17], which is recommended for those who have not seen this result before. As a prerequisite, we define *Young tableaux of shape* λ to be arrays of numbers, weakly increasing across rows and down columns, written in the squares of a Young diagram of λ . A *border strip tableau of shape* λ and *type* τ is a Young tableau such that among the entries the number i occurs exactly τ_i times, and for each i the set of squares in which i has been written form a *border strip*—that is, a connected collection of squares with no square upward and to the left of any others. The *height* of a border strip is one less than the number of rows that contain it, and the height $h(T)$ of a tableau T composed of border strips is the sum of the heights of the border strips.

Theorem 5.1 (Murnaghan–Nakayama rule). *For λ a partition of n and τ the type of a permutation from \mathfrak{S}_n*

$$X^\lambda(\tau) = \sum_T (-1)^{h(T)}, \tag{31}$$

where the sum is over all border strip tableaux T of shape λ and type τ .

Remark. A reader unfamiliar with characters of the symmetric group but nonetheless comfortable with the statement of the Murnaghan–Nakayama rule may take (31) as their definition the symmetric group’s characters.

5B. We will need to work with symmetric polynomials in m variables. Two bases for these polynomials that will be important for us are the power sum symmetric functions and Schur functions. Both bases are indexed by partitions.

For *power sum symmetric functions* in the variables $\omega_1, \dots, \omega_m$ we recall the definition that for an integer n ,

$$p_n = p_n(\omega_1, \dots, \omega_m) := \omega_1^n + \dots + \omega_m^n,$$

and for a partition $\lambda = (\lambda_1, \dots, \lambda_k)$, we define

$$p_\lambda := p_{\lambda_1} \cdots p_{\lambda_k}.$$

It is an elementary fact [Stanley 1999, Corollary 7.7.2] that any symmetric polynomial in the variables $\omega_1, \dots, \omega_m$ can be expressed uniquely as a linear combination of the functions p_λ .

Schur functions in the variables $\omega_1, \dots, \omega_m$ have the following as their classical definition. For a partition λ with $\ell(\lambda) \leq m$, set

$$s_\lambda = s_\lambda(\omega_1, \dots, \omega_m) := \frac{\det(\omega_i^{\lambda_j+m-j})_{i,j=1}^m}{\det(\omega_i^{n-j})_{i,j=1}^m}.$$

If $\ell(\lambda) < m$, we extend λ with 0's in the extra places so that the above definition still makes sense — i.e., $\lambda = (\lambda_1, \dots, \lambda_k, 0, \dots, 0)$. If $\ell(\lambda) > m$, we set $s_\lambda = 0$.

It is well-known (though not completely obvious at first glance) that s_λ defined as above is a symmetric polynomial with integer coefficients. As with power sums, any symmetric polynomial in the variables $\omega_1, \dots, \omega_m$ can be expressed uniquely as a linear combination of the functions s_λ . Proofs of these facts may be found in [Stanley 1999, Chapter 7].

For these symmetric polynomials we have the following important identities:

Theorem 5.2 (Frobenius). *For $\lambda \vdash n$,*

$$s_\lambda = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} X^\lambda(\sigma) p_\sigma = \sum_{\nu \vdash n} p(\nu) X^\lambda(\nu) p_\nu. \quad (32)$$

Likewise,

$$p_\nu = \sum_{\lambda \vdash n} X^\lambda(\nu) s_\lambda. \quad (33)$$

Proof. Equation (32) is Theorem 7.17.3 of [Stanley 1999], while (33) is Corollary 7.17.4. □

We can also express s_λ in terms of the elementary symmetric functions, defined by

$$e_n = e_n(\omega_1, \dots, \omega_m) := \sum_{i_1 < \dots < i_n} \omega_{i_1} \cdots \omega_{i_n},$$

with the conventions $e_0 = 1$ and $e_n(\omega_1, \dots, \omega_m) = 0$ for $n > m$.

Theorem 5.3 (Jacobi–Trudi). *For $\lambda_1 \leq k$,*

$$s_\lambda = \det(e_{\lambda'_i - i + j})_{i,j=1}^k.$$

Proof. This is a special case of Corollary 7.16.2 of [Stanley 1999]. □

Remark. This is often known as the *dual* Jacobi–Trudi identity because there is an equivalent formula in terms of the complete homogeneous symmetric functions; see [Stanley 1999, Theorem 7.16.1].

5C. One of the many results that is derived in the literature from Theorem 5.2 is an identity for characters of the symmetric group indexed by partitions that are dual to each other. We cite it here because we will use it later.

Proposition 5.4. *For $\sigma \in \mathfrak{S}_n$ and $\lambda \vdash n$,*

$$X^{\lambda'}(\sigma) = (-1)^{n-\ell(\sigma)} X^\lambda(\sigma).$$

Here $\ell(\sigma) := \ell(\tau_\sigma)$.

Proof. This is example 2 of section I.7 in [Macdonald 1995]. □

5D. One of the reasons we are interested in Schur functions is their appearance in random matrix theory. It is well known that they satisfy the following orthogonality relation.

Theorem 5.5. *For partitions λ and ν ,*

$$\int_{U(m)} s_\lambda(g) \overline{s_\nu(g)} dg = \delta_{\lambda\nu} \cdot \delta_{\ell(\lambda), \ell(\nu) \leq m}.$$

Moreover, if λ and ν are partitions of the same number (that is $|\lambda| = |\nu|$)

$$\int_{\text{PU}(m)} s_\lambda(g) \overline{s_\nu(g)} dg = \delta_{\lambda\nu} \cdot \delta_{\ell(\lambda), \ell(\nu) \leq m}.$$

Here $s_\lambda(g)$ and $s_\nu(g)$ are Schur functions whose entries are the m eigenvalues of the matrix g . A more or less self-contained proof may be found in [Gamburd 2007] as well as in more standard texts on representation theory.

6. A basis for factorization functions, and a bound for character sums

6A. We turn in this section to a proof of Theorem 3.1. Our strategy will be a familiar one, similar in its broad outlines to the original proof of Keating and Rudnick. By making use of the involution described in Section 4, we transfer a short interval sum to an average over sums of Dirichlet characters against factorization functions. These are in turn evaluated by using an equidistribution result of Katz and the combinatorial analysis of Section 7. This combinatorial analysis is perhaps the most important observation of the paper. In terms of technique, some new issues arise that have not appeared in the past just because we work with factorization functions in general.

6B. We begin by noting some ways to build factorization functions out of simpler functions. For two arithmetic functions ϕ_1 and ϕ_2 we define the convolution in the usual way,

$$\phi_1 \star \phi_2(f) := \sum_{\substack{f_1 f_2 = f \\ f_1, f_2 \in \mathcal{M}}} \phi_1(f_1) \phi_2(f_2).$$

It is clear that if ϕ_1 and ϕ_2 are factorization functions, then $\phi_1 \star \phi_2$ will be a factorization function as well.

For integers $m, e \geq 1$ we define the factorization function

$$\iota_{m,e}(f) = \begin{cases} 1 & \text{if } f = P^e \text{ with } \deg(P) = m, \\ 0 & \text{otherwise.} \end{cases}$$

Thus $\iota_{m,e}$ is the indicator function of e -th powers of m -th degree primes and is supported on \mathcal{M}_{me} . We generalize it in the following way: for an array $(\mathbf{m}, \mathbf{e}) = (m_1, e_1; m_2, e_2; \dots; m_\ell, e_\ell)$ we define

$$\iota_{(\mathbf{m}, \mathbf{e})} = \iota_{m_1, e_1} \star \iota_{m_2, e_2} \star \dots \star \iota_{m_\ell, e_\ell}. \tag{34}$$

Proposition 6.1. *Any factorization function supported on \mathcal{M}_n is a linear combination of the functions $\iota_{(\mathbf{m}, \mathbf{e})}$. (Necessarily $m_1 e_1 + \dots + m_\ell e_\ell = n$).*

Proof. Let $\mathcal{M}_{n,L}$ be the collection of elements of \mathcal{M}_n with extended factorization type $(m_1, e_1; \dots; m_\ell, e_\ell)$ for $\ell \leq L$ and $\mathcal{F}_{n,L}$ be the collection of factorization function supported on $\mathcal{M}_{n,L}$. We suppose the proposition is true for all factorization functions supported on $\mathcal{M}_{r,L}$ with $r \leq n$ and show that it is true for $\mathcal{M}_{n,L+1}$. Since it is obviously true (check) for $\mathcal{M}_{n,1}$ for all n , this will verify the claim by induction.

We introduce indicator functions $I_{(m,e)}$ of the extended factorization type (m, e) ; that is for $f \in \mathcal{M}$, we set $I_{(m_1,e_1;\dots;m_\ell,e_\ell)}(f) = 1$ if f has extended factorization type $(m_1, e_1; \dots; m_\ell, e_\ell)$ and we set $I_{(m_1,e_1;\dots;m_\ell,e_\ell)}(f) = 0$ otherwise. Clearly

$$\mathcal{F}_{n,L+1} = \text{span}\{I_{(m_1,e_1;\dots;m_\ell,e_\ell)} : m_1e_1 + \dots + m_\ell e_\ell = n, \ell \leq L + 1\},$$

so to prove our claim we need only show that each

$$I_{(m_1,e_1;\dots;m_{L+1},e_{L+1})} \tag{35}$$

is a linear combination of functions $\iota_{(m,e)}$. Suppose ν of the terms $(m_1, e_1), \dots, (m_L, e_L)$ in (35) are equal to (m_{L+1}, e_{L+1}) . (We allow ν to be 0.) By inspection of elements of $\mathcal{M}_{n,L+1}$ we see that

$$I_{(m_1,e_1;\dots;m_L,e_L)} \star \iota_{m_{L+1},e_{L+1}} - (\nu + 1)I_{(m_1,e_1;\dots;m_{L+1},e_{L+1})} \tag{36}$$

is supported on $\mathcal{M}_{n,L}$. By inductive hypothesis then (36) is a linear combination of terms $\iota_{(m,e)}$. Likewise by inductive hypothesis, $I_{(m_1,e_1;\dots;m_L,e_L)}$ is a linear combination of such terms, so $I_{(m_1,e_1;\dots;m_L,e_L)} \star \iota_{m_{L+1},e_{L+1}}$ will be as well. Returning to (36), since $\nu + 1 \neq 0$, this shows that $I_{(m_1,e_1;\dots;m_{L+1},e_{L+1})}$ is therefore a linear combination of such terms, so that as claimed all factorization functions on $\mathcal{M}_{n,L+1}$ are linear combinations of such terms also. □

6C. We have indicated that we must work with Dirichlet characters modulo T^M for some power M . Note that for any nontrivial Dirichlet character χ modulo T^M , we have, by excluding powers of primes from the sum in the first line below and using the Riemann hypothesis in the form (26) in the second,

$$\sum_{f \in \mathcal{M}_n} \iota_{n,1}(f)\chi(f) = \frac{1}{n} \sum_{f \in \mathcal{M}_n} \Lambda(f)\chi(f) + O(q^{n/2}) = O_M(q^{n/2}).$$

Thus for any $e \geq 2$, as long as $\chi^e \neq \chi_0$,

$$\sum_{f \in \mathcal{M}_{me}} \iota_{m,e}(f)\chi(f) = \sum_{f \in \mathcal{M}_m} \iota_{m,1}(f)\chi^e(f) = O_M(q^{m/2}).$$

For $e \geq 3$, trivially

$$\sum_{f \in \mathcal{M}_{me}} \iota_{m,e}(f)\chi(f) = O(q^m).$$

Note that for $m \geq 1, e \geq 2$, we have $\frac{m}{2} \leq \frac{me}{2} - \frac{1}{2}$, and for $m \geq 1, e \geq 3$, we have likewise $m \leq \frac{me}{2} - \frac{1}{2}$. Thus combining the two estimates above, we see that unless $\chi^2 = \chi_0$, we have for $e \geq 2$,

$$\sum_{f \in \mathcal{M}_{me}} \iota_{m,e}(f)\chi(f) = O(q^{me/2-1/2}).$$

Hence recalling the definition (34), unless $\chi^2 = \chi_0$, if some $e_i \geq 2$,

$$\sum_{f \in \mathcal{M}_n} \iota_{m,e}(f)\chi(f) = O_{M,n}(q^{n/2-1/2}), \tag{37}$$

where $n = m_1e_1 + \dots + m_k e_k$.

We have thus obtained

Lemma 6.2. *If b is a fixed factorization function supported on the squarefuls, for χ a Dirichlet character modulo T^M , as long as $\chi^2 \neq \chi_0$,*

$$\sum_{f \in \mathcal{M}_n} b(f)\chi(f) = O_{M,n}(q^{n/2-1/2}).$$

Proof. For such b , the function $b(f)\mathbf{1}_{\mathcal{M}_n}(f)$ is necessarily a linear combination of functions $\iota_{m,e}$, where in each case some $e_i \geq 2$. □

In the case that $\chi^2 = \chi_0$, we may genuinely have a worse bound, but it is easy to see in the same way that as long as $\chi \neq \chi_0$ for $\chi \pmod{T^M}$, the bound in Lemma 6.2 may be replaced by $O_{M,n}(q^{n/2})$. Indeed, for such an estimate, it is easy to see that we have no need that our factorization function be supported on the squarefuls as it was in Lemma 6.2.

Lemma 6.3. *If a is a fixed factorization function, for χ a nontrivial Dirichlet character modulo T^M ,*

$$\sum_{f \in \mathcal{M}_n} a(f)\chi(f) = O_{n,M}(q^{n/2}).$$

Note that a character satisfies $\chi^2 = \chi_0$ only if it is real. Fortunately there are not many real characters modulo T^M .

Lemma 6.4. *Over $\mathbb{F}_q[T]$, the number of nontrivial real characters modulo T^M is $O(1)$ if $2 \nmid q$, and $O(q^{\lfloor M/2 \rfloor})$ if $2 \mid q$.*

Proof. Let \widehat{G} be the group of characters. Real characters χ are characterized by having $\chi^2 = \chi_0$. As $\widehat{G} \cong (\mathbb{F}_q[T]/(T^M))^*$, the number of real characters is equal to the number of $f \in \mathbb{F}_q[T]$ with $(f, T^M) = 1$ and $\deg(f) < M$ such that

$$f^2 \equiv 1 \pmod{T^M}. \tag{38}$$

Yet if $2 \nmid q$, we have $(f - 1, f + 1) = 1$ and so (38) implies $f \equiv \pm 1 \pmod{T^M}$, which is satisfied by only two such f . Hence in this case there are at most two real characters modulo T^M , and thus at most one nontrivial real character.

If $2 \mid q$, the situation is more complicated. If $f = a_0 + \dots + a_{M-1}T^{M-1}$, we have

$$f^2 = a_0^2 + a_1^2T^2 + \dots + a_{M-1}^2T^{2(M-1)},$$

so that each solution $f^2 \equiv 1 \pmod{T^M}$ entails $\lfloor (M - 1)/2 \rfloor + 1$ linear equations,

$$a_0^2 = 1, a_1^2 = 0, \dots, a_{\lfloor (M-1)/2 \rfloor}^2 = 0$$

of which there is only one solution. The remaining $M - 1 - \lfloor (M - 1)/2 \rfloor = \lfloor M/2 \rfloor$ coefficients $a_{\lfloor (M-1)/2 \rfloor + 1}, \dots, a_{M-1}$ may vary freely, but this leads to only $q^{\lfloor M/2 \rfloor}$ different solutions. \square

Remark. I thank Ofir Gorodetsky for suggesting this proof of Lemma 6.4 to me.

7. Schur functions of zeros

7A. We have noted the explicit formula (28), which establishes a correspondence between the von Mangoldt function $\Lambda(f)$ and the traces of powers of unitarized Frobenius matrices. Written another way, let χ be a primitive character modulo T^m . For $p_n(\Theta_\chi)$ the symmetric power sum of the unitarized zeros $\{e^{i2\pi\vartheta_1}, \dots, e^{i2\pi\vartheta_{m-2}}\}$ of $\mathcal{L}(u, \chi)$, the explicit formula is just the statement that

$$p_n(\Theta_\chi) = \frac{-1}{q^{n/2}} \sum_{f \in \mathcal{M}_n} \Lambda(f)\chi(f) + O(1/q^{n/2}) = \frac{-n}{q^{n/2}} \sum_{\substack{P \in \mathcal{M}_n \\ \text{prime}}} \chi(P) + O(q^{-1/2}) \tag{39}$$

for $\chi^2 \neq \chi_0$. (We require $\chi^2 \neq \chi$ in order to absorb higher prime powers into the error term.) By multiplying these power sums together, from unique factorization and a simple counting argument, it follows that for the partition $\nu = \langle 1^{m_1} 2^{m_2} \dots j^{m_j} \rangle$, with $\nu \vdash n$,

$$p_\nu(\Theta_\chi) = \frac{1}{q^{n/2}} \prod_{i=1}^j i^{m_i} m_i! \sum_{f \in \mathcal{M}_n} \mathbf{1}_\nu(\tau_f) \mu(f) \chi(f) + O(q^{-1/2}).$$

We have used the Riemann hypothesis bound (26) to retain this error term from (39). Note that the coefficient $\prod i^{m_i} m_i!$ here is $1/p(\nu)$, defined in (12) from our introductory remarks about the symmetric group. By applying the Frobenius formula, Theorem 5.2, we see that for the Schur function with arguments $\{e^{i2\pi\vartheta_1}, \dots, e^{i2\pi\vartheta_{m-2}}\}$,

$$s_\lambda(\Theta_\chi) = \frac{1}{q^{n/2}} \sum_{f \in \mathcal{M}_n} \mu(f) X^\lambda(f) \chi(f) + O(q^{-1/2}).$$

Because $\mu(f) X^\lambda(f) = (-1)^{\ell(\tau_f)} X^\lambda(\tau_f) = (-1)^n X^{\lambda'}(\tau_f)$ by Proposition 5.4, we have thus shown:

Theorem 7.1. *For χ a primitive character modulo T^m with $\chi^2 \neq \chi_0$,*

$$s_\lambda(\Theta_\chi) = \frac{(-1)^n}{q^{n/2}} \sum_{f \in \mathcal{M}_n} X^{\lambda'}(f) \chi(f) + O_{n,m}(q^{-1/2}).$$

7B. Note that in the above theorem, there is no explicit reference to the degree m of the polynomial T^m . Nonetheless, if χ is primitive and even, $s_\lambda(\Theta_\chi)$ is a polynomial in $m - 2$ variables, and so we must have $s_\lambda(\Theta_\chi) = 0$ for $\ell(\lambda) > m - 2$. We have thus observed

Corollary 7.2. *If $\ell(\lambda') = \lambda_1 > m - 2$,*

$$\sum_{f \in \mathcal{M}_n} X^{\lambda'}(f) \chi(f) = O(q^{(n-1)/2}),$$

uniformly for χ a primitive even character modulo T^m .

Remark. A similar statement to the above can of course be written down for odd primitive characters.

As another consequence of Theorem 7.1,

Corollary 7.3. *For partitions $\lambda, \nu \vdash n$ and $m \geq 5$,*

$$\mathbb{E}_{\chi(T^m)_{\text{prim, ev}}} \left(\frac{1}{q^{n/2}} \sum_{f \in \mathcal{M}_n} X^\lambda(f) \chi(f) \right) \overline{\left(\frac{1}{q^{n/2}} \sum_{f \in \mathcal{M}_n} X^\nu(f) \chi(f) \right)} = \delta_{\lambda\nu} \cdot \delta_{\ell(\lambda'), \ell(\nu') \leq m-2} + O(q^{-1/2}). \quad (40)$$

Proof. By Theorem 7.1, the left-hand side of (40) can be written

$$\frac{1}{\Phi_{\text{prim}}^{\text{ev}}(T^m)} \sum_{\substack{\chi(T^m) \\ \text{prim, ev}}} (s_{\lambda'}(\Theta_\chi) + O(q^{-1/2})) \overline{(s_{\nu'}(\Theta_\chi) + O(q^{-1/2}))} + O\left(\frac{q^{\lfloor m/2 \rfloor}}{\Phi_{\text{prim}}^{\text{ev}}(T^m)}\right), \quad (41)$$

using Lemmas 6.3 and 6.4 to bound the contribution of characters with $\chi^2 = \chi_0$. For $m \geq 5$, recalling the value of $\Phi_{\text{prim}}^{\text{ev}}(T^m)$ given in (25), we certainly have

$$\frac{q^{\lfloor m/2 \rfloor}}{\Phi_{\text{prim}}^{\text{ev}}(T^m)} = O(q^{-1/2}),$$

and using the equidistribution Theorem 4.2 to treat the main term, we see that (41) reduces to

$$\int_{U(M-2)} s_{\lambda'} \overline{s_{\nu'}} dg + O(q^{-1/2}).$$

(Note that the symmetric polynomial $s_{\lambda'} \overline{s_{\nu'}}$, homogeneous under unimodular multiplication, is a linear combination of characters of $\text{PU}(M-2)$.) This agrees with the right-hand side of (40) by the orthonormality of Schur functions (Theorem 5.5). □

7C. We will later need the following result, which is essentially the “easy” case of Corollary 7.3.

Lemma 7.4. *For a_1 and a_2 , factorization functions supported on \mathcal{M}_n , and m sufficiently large (depending on n),*

$$\lim_{q \rightarrow \infty} \mathbb{E}_{\chi(T^m)_{\text{prim, ev}}} \left(\frac{1}{q^{n/2}} \sum_{f \in \mathcal{M}_n} a_1(f) \chi(f) \right) \overline{\left(\frac{1}{q^{n/2}} \sum_{f \in \mathcal{M}_n} a_2(f) \chi(f) \right)} = \langle a_1, a_2 \rangle, \quad (42)$$

with the inner product defined by (18).

Proof. This is not a deep result, following from nothing more sophisticated than orthogonality relations for characters averaged in this way.

Nonetheless, it is less work for us at this point to make use of Corollary 7.3 and note the following, for $m \geq \min(5, n+2)$: if a_1 or a_2 is supported on the squarefuls, then (42) is true (with the right-hand side obviously equal to 0), owing to Lemmas 6.2 and 6.3 (with contributions of characters $\chi^2 = \chi_0$ in the average dealt with as in the proof of Corollary 7.3). Moreover, if these functions are characters of the symmetric group, $a_1(f) = X^\lambda(f)$ and $a_2(f) = X^\nu(f)$, then (42) is true by Corollary 7.3. Since any factorization function can be written as a linear combination of characters and some function supported on the squarefuls, this verifies (42) in general. □

8. A proof of Theorem 3.1

8A. Because we will be using characters modulo powers of T , we must work with polynomials f that are coprime to T . We recall our definition \mathcal{P}_n^\natural and make a similar definition for monic polynomials

$$\mathcal{P}_n^\natural := \{f \in \mathcal{P}_n : f(0) \neq 0\} \quad \text{and} \quad \mathcal{M}_n^\natural := \{f \in \mathcal{M}_n : f(0) \neq 0\}.$$

In addition we define for $f \in \mathcal{M}_n$,

$$\tilde{a}(f) := a(f) - E(a; n), \quad \text{with } E(a; n) := \frac{1}{|\mathcal{M}_n|} \sum_{g \in \mathcal{M}_n} a(g)$$

and for $f \in \mathcal{M}_n^\natural$,

$$\tilde{a}^\natural(f) := a(f) - E^\natural(a; n), \quad \text{with } E^\natural(a; n) := \frac{1}{|\mathcal{M}_n^\natural|} \sum_{g \in \mathcal{M}_n^\natural} a(g).$$

With these conventions, our proof of Theorem 3.1 may be broken into five pieces.

Step 1: In the first place, we reduce the variance of short interval sums, restricted to \mathcal{M}_n^\natural , to a sum over Dirichlet characters.

Lemma 8.1. *For any factorization function a ,*

$$\sum_{f \in \mathcal{M}_n} \left| \sum_{\substack{g \in I(f; h) \\ g \in \mathcal{M}_n^\natural}} \tilde{a}^\natural(g) \right|^2 = \frac{q^{h+1}(q-1)}{\Phi(T^{n-h})} \sum_{\substack{\chi \neq \chi_0(T^{n-h}) \\ \text{even}}} \left| \sum_{g \in \mathcal{M}_n} a(g)\chi(g) \right|^2.$$

for $0 \leq h \leq n - 1$.

The proof is a straightforward modification of Steps 1 and 2 in [Rodgers 2015], and we refer the reader to that paper for details. In summary: one transfers the short interval sum to a sum over Dirichlet characters by making use of the involution described in Section 4 of this paper.

Step 2: We next bound the sums in Lemma 8.1 for all factorization functions that are supported on the squarefuls.

Lemma 8.2. *For a fixed factorization b function supported on the squarefuls,*

$$\sum_{f \in \mathcal{M}_n} \left| \sum_{\substack{g \in I(f; h) \\ g \in \mathcal{M}_n^\natural}} \tilde{b}^\natural(g) \right|^2 = O_{n,h}(q^h q^n). \tag{43}$$

for $0 \leq h \leq n - 4$.

Proof. Clearly by Lemma 8.1 we need only show that

$$\frac{q-1}{\Phi(T^{n-h})} \sum_{\substack{\chi \neq \chi_0(T^{n-h}) \\ \text{even}}} \left| \sum_{g \in \mathcal{M}_n} b(g)\chi(g) \right|^2 = O_{n,h}(q^{n-1}). \tag{44}$$

From Lemma 6.2, we note that for nonreal characters χ modulo T^{n-h} , uniformly

$$\left| \sum_{g \in \mathcal{M}_n} b(f)\chi(f) \right| = O_{n,h}(q^{n/2-1/2}),$$

while from Lemma 6.4 there are at most $O(q^{(n-h)/2})$ real nontrivial characters, and for such a character by Lemma 6.3 this sum is $O_{n,h}(q^{n/2})$. Hence the left-hand side of (44) is at most

$$\frac{q-1}{\Phi(T^{n-h})} (\Phi_{\text{ev}}(T^{n-h}) \cdot O_{n,h}(q^{n-1}) + O_{n,h}(q^n q^{(n-h)/2})) = O_{n,h}(q^{n-1}). \quad \square$$

In a similar same manner, we obtain a more general bound for factorization functions that needn't be supported on the squarefuls.

Lemma 8.3. *For a fixed factorization function a ,*

$$\sum_{f \in \mathcal{M}_n} \left| \sum_{\substack{g \in I(f;h) \\ f \in \mathcal{M}_n^{\natural}}} \tilde{a}^{\natural}(g) \right|^2 = O_{n,h}(q^{h+1}q^n),$$

for $0 \leq h \leq n$.

Proof. This follows from Lemmas 8.1 and 6.3. □

Step 3: We show that the variances of sums over \mathcal{M}_n^{\natural} we have computed in Lemma 8.1 are not far from those of sums over \mathcal{M}_n , which we are ultimately after.

Lemma 8.4. *For a fixed factorization function a ,*

$$\sum_{f \in \mathcal{M}_n} \left| \sum_{g \in I(f;h)} \tilde{a}(g) \right|^2 = \sum_{f \in \mathcal{M}_n} \left| \sum_{\substack{g \in I(f;h) \\ f \in \mathcal{M}_n^{\natural}}} \tilde{a}^{\natural}(g) \right|^2 + O_{h,n}(q^{h+1/2}q^n),$$

for $0 \leq h \leq n$.

Note, in comparison with the error term, that we expect the left-hand side to usually be of order $q^n q^{h+1}$.

Proof. We make use of a mapping of polynomials $f \mapsto f^{[i]}$ defined by

$$(a_0 + a_1T + \cdots + a_nT^n)^{[i]} = a_i + a_{i+1}T + \cdots + a_nT^{n-i},$$

so that if $T^i \mid f$,

$$f = T^i f^{[i]}.$$

For $f \in \mathcal{M}_n$, we may partition $I(f; h)$ into the disjoint union

$$I(f; h) = \left(\bigcup_{i=0}^h \{T^i g \in I(f; h) : g \in \mathcal{M}_{n-i}^{\natural}\} \right) \cup \{T^{h+1} f^{[h+1]}\}. \quad (45)$$

For $g \in \mathcal{M}_{n-i}^{\natural}$, we define the function

$$a_{[i]}(g) := a(T^i g),$$

with

$$\tilde{a}_{[i]}(g) := a_{[i]}(g) - E^{\natural}(a_{[i]}; n - i), \quad \text{with} \quad E^{\natural}(a_{[i]}; n - i) = \frac{1}{|\mathcal{M}_{n-i}^{\natural}|} \sum_{g \in \mathcal{M}_{n-i}^{\natural}} a_{[i]}(g).$$

From the partitioning (45), it is easy to see that for $f \in \mathcal{M}_n$,

$$\sum_{g \in I(f; h)} a(g) = \sum_{\substack{g \in I(f; h) \\ g \in \mathcal{M}_n^{\natural}}} a(g) + \sum_{\substack{g \in I(f^{[1]}; h-1) \\ g \in \mathcal{M}_{n-1}^{\natural}}} a_{[1]}(g) + \cdots + \sum_{\substack{g \in I(f^{[h]}; 0) \\ g \in \mathcal{M}_{n-h}^{\natural}}} a_{[h]}(g) + \underbrace{a(T^{h+1} f^{[h+1]})}_{=O_{n,h}(1)}. \quad (46)$$

Using that

$$|\mathcal{M}_n^{\natural}| = q^{n-1}(q - 1), \quad \text{and} \quad |\{g \in I(f; h) : g \in \mathcal{M}_n^{\natural}\}| = q^h(q - 1),$$

one may verify (with a little work, but straightforwardly) that

$$\sum_{g \in I(f; h)} E(a; n) = \frac{q^{h+1}}{q^n} \sum_{g \in \mathcal{M}_n} a(g) = \sum_{k=0}^h \sum_{\substack{g \in I(f^{[k]}; h-k) \\ g \in \mathcal{M}_{n-k}^{\natural}}} E^{\natural}(a_{[k]}; n - k) + \underbrace{\frac{q^{h+1}}{q^n} \sum_{g \in \mathcal{M}_{n-h-1}} a(T^{h+1} g)}_{=O_{n,h}(1)}. \quad (47)$$

Thus combining (46) and (47), we have uniformly for $f \in \mathcal{M}_n$,

$$\sum_{g \in I(f; h)} \tilde{a}(g) = \sum_{\substack{g \in I(f; h) \\ g \in \mathcal{M}_n^{\natural}}} \tilde{a}^{\natural}(g) + \sum_{\substack{g \in I(f^{[1]}; h-1) \\ g \in \mathcal{M}_{n-1}^{\natural}}} \tilde{a}_{[1]}^{\natural}(g) + \cdots + \sum_{\substack{g \in I(f^{[h]}; 0) \\ g \in \mathcal{M}_{n-h}^{\natural}}} \tilde{a}_{[h]}^{\natural}(g) + O_{n,h}(1). \quad (48)$$

For each i , the function $a_{[i]}$ defined on \mathcal{M}_{n-i} extends uniquely to a factorization function defined on all of \mathcal{M}_{n-i} . Hence, using Lemma 8.3 to pass to the second line below,

$$\sum_{f \in \mathcal{M}_n} \left| \sum_{\substack{g \in I(f^{[i]}; h-i) \\ g \in \mathcal{M}_{n-i}^{\natural}}} \tilde{a}_{[i]}^{\natural}(g) \right|^2 = q^i \sum_{f \in \mathcal{M}_{n-i}} \left| \sum_{g \in I(f; h-i)} \tilde{a}_{[i]}^{\natural}(g) \right|^2 \ll_{n,h} q^i q^{h-i+1} q^{n-i}.$$

This quantity is no more than $q^h q^n$ for $i \geq 1$, and for $i = 0$ it is of course equal to $q^{h+1} q^n$.

Therefore, squaring the identity (48) and summing over $g \in \mathcal{M}_n$, then using Cauchy–Schwarz and (49) to bound all terms but one on the right,

$$\sum_{f \in \mathcal{M}_n} \left| \sum_{g \in I(f; h)} \tilde{a}(g) \right|^2 = \sum_{f \in \mathcal{M}_n} \left| \sum_{\substack{g \in I(f; h) \\ g \in \mathcal{M}_n^{\natural}}} \tilde{a}^{\natural}(g) \right|^2 + O_{n,h}(q^n q^{h+1/2}),$$

as claimed. □

Step 4: Recall the “factorization Fourier expansion” (14):

$$a(f) = A(f) + b(f), \tag{49}$$

with

$$A(f) := \sum_{\lambda} \hat{a}_{\lambda} X^{\lambda}(f),$$

where $b(f)$ is a function supported on the squarefuls. We use this to reduce variance for the function $a(f)$ to finding the covariance of characters $X^{\lambda}(f)$.

We introduce the shorthand, for partitions $\lambda, \nu \vdash n$,

$$\Delta_{\lambda, \nu}(m) := \mathbb{E}_{\chi_{\text{prim, ev}}(T^m)} \left(\sum_{f \in \mathcal{M}_n} X^{\lambda}(f) \chi(f) \right) \overline{\left(\sum_{g \in \mathcal{M}_n} X^{\nu}(g) \chi(g) \right)}. \tag{50}$$

Note that by Corollary 7.3, for $m \geq 5$,

$$\Delta_{\lambda, \nu}(m) = \delta_{\lambda \nu} \delta_{\ell(\lambda'), \ell(\nu') \leq m-2} + O(q^{-1/2}) = \delta_{\lambda \nu} \delta_{\lambda_1, \nu_1 \leq m-2} + O(q^{-1/2}). \tag{51}$$

Lemma 8.5. *For a fixed factorization function a , with $0 \leq h \leq n - 4$,*

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f; h)} a(g) \right) = q^{h+1} \sum_{\mu, \nu \vdash n} \Delta_{\lambda, \nu}(n-h) \hat{a}_{\lambda} \overline{\hat{a}_{\nu}} + O_{n, h}(q^{h+1/2}). \tag{52}$$

Proof. The variance in (52) is given by

$$\frac{1}{q^n} \sum_{f \in \mathcal{M}_n} \left| \sum_{g \in I(f; h)} \tilde{a}(g) \right|^2 = \frac{1}{q^n} \sum_{f \in \mathcal{M}_n} \left| \sum_{\substack{g \in I(f; h) \\ g \in \mathcal{M}_n^{\text{even}}}} \tilde{A}^{\natural}(g) \right|^2 + O_{n, h}(q^{h+1/2}),$$

where we have reduced to a sum of terms $\tilde{A}^{\natural}(g)$ by using Lemma 8.4 and then Lemmas 8.2 and 8.3 to absorb a sum of terms $b^{\natural}(g)$ into the error term.

In turn from Lemma 8.1,

$$\begin{aligned} \frac{1}{q^n} \sum_{f \in \mathcal{M}_n} \left| \sum_{g \in I(f; h)} \tilde{A}^{\natural}(g) \right|^2 &= \frac{q^{h+1}(q-1)}{q^n \Phi(T^{n-h})} \sum_{\substack{\chi \neq \chi_0(T^{n-h}) \\ \text{even}}} \left| \sum_{g \in \mathcal{M}_n} A(g) \chi(g) \right|^2 \\ &= \frac{q^{h+2}}{q^n q^{n-h}} \left(\sum_{\substack{\chi(T^{n-h}) \\ \text{prim, ev}}} \left| \sum_{g \in \mathcal{M}_n} A(g) \chi(g) \right|^2 + O_{n, h}(q^n \cdot q^{n-h-2}) \right). \end{aligned}$$

The second line has followed by taking nonprimitive even characters from the sum and bounding their contribution by Lemma 6.3. The above quantity simplifies to

$$q^{h+1} \mathbb{E}_{\chi_{\text{prim, ev}}(T^{n-h})} \left| \sum_{g \in \mathcal{M}_n} A(g) \chi(g) \right|^2 + O_{n, h}(q^h),$$

and Lemma 8.5 follows by expanding $A(g)$ into a linear combination of characters X^λ (recall A is defined by (49)) and then expanding the square above. □

With this lemma in place, Theorem 3.1 now follows by applying (51).

9. Factorization Fourier expansions

9A. We list some examples of the expansion (14) for the arithmetic functions we considered in Section 1. In this way we recover Theorems 1.3, 1.4, and 1.5, estimating variance over short intervals of the Möbius function, the von Mangoldt function, and the k -fold divisor function. We also consider the function ω , which as usual counts distinct prime factors, and this leads to a new result for the variance of $\omega(f)$ and $\mu(f)\omega(f)$ summed over short intervals.

Proposition 9.1. For $f \in \mathcal{M}_n$,

$$\mu(f) = (-1)^n X^{(1^n)}(f).$$

Remark. Applied to Theorem 3.1 this recovers Theorem 1.3, for $\mu(f)$.

Proof. Both $\mu(f)$ and $X^{(1^n)}(f)$ will be zero unless f is squarefree. But for $f = p_1 \cdots p_\ell$, with all factors distinct, $\mu(f) = (-1)^\ell$, while it may be checked $X^{(1^n)}(f) = (-1)^{\deg(p_1)-1} \cdots (-1)^{\deg(p_\ell)-1}$. As $(-1)^{\deg(p_1)} \cdots (-1)^{\deg(p_\ell)} = (-1)^n$, this verifies the claim. □

Proposition 9.2. For $f \in \mathcal{M}_n$,

$$\mu(f)^2 = X^{(n)}(f).$$

Proof. As $X^{(n)}$ is the trivial character, this is clear. □

Proposition 9.3. For $f \in \mathcal{M}_n$,

$$\Lambda(f) = \sum_{r=1}^n (-1)^{n-r} X^{(r, 1^{n-r})}(f) + b(f),$$

for a function $b(f)$ that is supported on the squarefuls.

Remark. This recovers Theorem 1.4, for $\Lambda(f)$.

If we define the function,

$$\Lambda_j(f) := \sum_{\substack{g|f \\ g \text{ monic}}} \mu(g) \deg(f/g)^j, \tag{53}$$

the proposition above is special case of:

Proposition 9.4. For $f \in \mathcal{M}_n$,

$$\Lambda_j(f) = \sum_{r=1}^n (-1)^{n-r} (r^j - (r-1)^j) X^{(r, 1^{n-r})}(f) + b(f),$$

for a function $b(f)$ that is supported on the squarefuls.

Remark. This recovers an estimate for the covariance of almost-primes in short intervals, proved in [Rodgers 2015].

Note that, using (53),

$$\Lambda_j(f) = \sum_{r=1}^n (r^j - (r-1)^j) \sum_{\substack{g|f \\ \deg(g) \leq n-r \\ g \text{ monic}}} \mu(g),$$

so we have that Proposition 9.4 is a corollary of:

Proposition 9.5. For $f \in \mathcal{M}_n$, with $n = r + s$ and $0 \leq s < n$,

$$\sum_{\substack{g|f \\ \deg(g) \leq s \\ g \text{ monic}}} \mu(g) = (-1)^s X^{(r,1^s)}(f) + b(f),$$

for a function $b(f)$ that is supported on the squarefuls.

Proof. We will need to make use of the Murnaghan–Nakayama rule, quoted in Theorem 5.1.

We may suppose that f is squarefree (otherwise the proposition is trivial), and let $f = p_1 \cdots p_\ell$ with $\deg p_i = \tau_i$, $\tau_1 \geq \tau_2 \geq \cdots$. We apply the Murnaghan–Nakayama rule to the type $\tau_f = (\tau_1, \dots, \tau_\ell)$ and Young diagram of $(r, 1^s)$. For any border-strip tableau, let $I \subset \{2, \dots, \ell\}$ be the collection of numbers that appear in rows 2 through s of the Young diagram of $(r, 1^s)$. Writing

$$\tau_I := \sum_{i \in I} \tau_i,$$

to form a valid border-strip tableau, it is easy to see that we require only that $\tau_I \leq s$ and $\tau_1 + \tau_I \geq s + 1$. Hence, applying the rule,

$$X^{(r,1^s)}(f) = \sum_{\substack{I \subset \{2, \dots, \ell\} \\ \tau_I \leq s \\ \tau_1 + \tau_I \geq s+1}} (-1)^{\tau_I - |I|} (-1)^{(s+1) - \tau_I - 1} = (-1)^s \sum_{\substack{I \subset \{2, \dots, \ell\} \\ s - \tau_1 < \tau_I \leq k}} (-1)^{|I|}. \tag{54}$$

Yet,

$$\sum_{\substack{g|f \\ \deg(g) \leq s \\ g \text{ monic}}} \mu(g) = \sum_{J \subset \{1, \dots, \ell\}} (-1)^{|J|} \tag{55}$$

By breaking the right-hand sum into parts for which 1 is an element of J or not, we see that (55) is equal to

$$\sum_{\substack{I \subset \{2, \dots, \ell\} \\ \tau_I \leq s}} (-1)^{|I|} + \sum_{\substack{I \subset \{2, \dots, \ell\} \\ \tau_I + \tau_1 \leq s}} (-1)^{|I|+1} = \sum_{\substack{I \subset \{2, \dots, \ell\} \\ s - \tau_1 < \tau_I \leq s}} (-1)^{|I|}.$$

Comparing this with (54) yields the result. □

Proposition 9.6. For $f \in \mathcal{M}_n$,

$$d_k(f) = \sum_{\substack{\lambda \vdash n \\ \ell(\lambda) \leq k}} s_\lambda(\underbrace{1, \dots, 1}_k) X^\lambda(f) + b(f), \tag{56}$$

for a function $b(f)$ that is supported on the squarefuls. Moreover, we have the following equivalent expressions for $s_\lambda(1, \dots, 1)$:

$$(i) \quad s_\lambda(\underbrace{1, \dots, 1}_k) = \prod_{1 \leq i < j \leq k} \frac{\lambda_i - \lambda_j + j - i}{j - i}, \tag{57}$$

with the convention that $\lambda_{\ell(\lambda)+1} = \dots = \lambda_k = 0$ if $\ell(\lambda) < k$.

$$(ii) \quad s_\lambda(\underbrace{1, \dots, 1}_k) = \text{GT}_k(\lambda), \tag{58}$$

where $\text{GT}_k(\lambda)$ is the number of triangular arrays of nonnegative integers

$$\begin{array}{ccccccc} x_1^{(1)} & & x_2^{(1)} & & \dots & & x_k^{(1)} \\ & \ddots & & \ddots & & \ddots & \\ & & x_1^{(k-1)} & & x_2^{(k-1)} & & \\ & & & & x_1^{(k)} & & \end{array}$$

with entries weakly decreasing left-to-right down diagonals and weakly increasing left-to-right up diagonals (that is, $x_j^{(i)} \geq x_j^{(i+1)} \geq x_{j+1}^{(i)}$), and in the top row, $x_i^{(1)} = \lambda_i$, with again the convention if $\ell(\lambda) < k$ that $\lambda_{\ell(\lambda)+1} = \dots = \lambda_k = 0$.

$$(iii) \quad s_\lambda(\underbrace{1, \dots, 1}_k) = \prod_{u \in \lambda} \frac{k + c(u)}{h(u)}, \tag{59}$$

where the product is over all squares u of the Young diagram of λ , and where if we label the squares u by the coordinates (i, j) with $1 \leq j \leq \lambda_i$, the content $c(u)$ is defined by

$$c(u) = i - j,$$

and the hook length $h(u)$ is defined by

$$h(u) = \lambda_i + \lambda'_j - i - j + 1.$$

(See [Stanley 1999, p. 373] for a lengthier account of these definitions.)

Remark. Using the representation (ii), this recovers the variance of the k -fold divisor function given in Theorem 1.5.

Proof. It will again be sufficient to consider f squarefree. We note that for p prime, $d_k(p) = k$, so for $f = p_1 \cdots p_\ell$ with all prime factors distinct,

$$d_k(f) = k^\ell = k^{\ell(\tau)},$$

where τ is the factorization type of f . On the other hand,

$$k^{\ell(\tau)} = p_\tau \underbrace{(1, \dots, 1)}_k = \sum_{\lambda \vdash n} s_\lambda \underbrace{(1, \dots, 1)}_k X^\lambda(\tau), \tag{60}$$

by Theorem 5.2 of Frobenius. This proves (56).

For the formula given in (i), note that for $\ell(\lambda) > k$, we have $s_\lambda(1, \dots, 1) = 0$, while for $\ell(\lambda) \leq k$, the identity (57) is [Fulton 1997, Example 6, Chapter 6].

For the formula given in (ii), note that $s_\lambda(1, \dots, 1)$ is equal to the number of semistandard Young tableaux of shape λ with entries 1 through k (see [Stanley 1999, §7.10]), and by a well-known bijection (again, see [Stanley 1999, §7.10]) this is equal to $\text{GT}_k(\lambda)$. (For readers familiar with the terminology, $\text{GT}_k(\lambda)$ is a count of Gelfand–Tsetlin patterns.)

For the formula given in (iii), this is Corollary 7.21.4 of [Stanley 1999]. □

Proposition 9.7. *Let $\omega(f)$ be the number of distinct primes that divide f . Then for $f \in \mathcal{M}_n$,*

$$\omega(f) = H_n X^{(n)}(f) + \sum_{\lambda} (-1)^v \left(\frac{1}{\lambda_2 + v} - \frac{1}{\lambda_1 + v + 1} \right) X^{(\lambda_1, \lambda_2, 1^v)}(f) + b(f), \tag{61}$$

where the sum is over all partitions $\lambda = (\lambda_1, \lambda_2, 1^v) \vdash n$ with $\lambda_2 \geq 1$ and $v \geq 0$, where $b(f)$ is a function supported on the squarefuls, and where

$$H_n := \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}.$$

Remark. The mean value as $q \rightarrow \infty$ of $\omega(f)$ for $\deg(f) = n$ is H_n . Because $X^{(n)}(f) = 1$ for all squarefree f , the expression (61) may be thought of as characterizing the oscillation of $\omega(f)$ around this value.

Proof. We use the identity (60) from the last proof, along with the representation (59) for $s_\lambda(1, \dots, 1)$. Taken together these imply for $\tau \vdash n$ and positive integer k ,

$$k^{\ell(\tau)} = \sum_{\lambda \vdash n} \prod_{u \in \lambda} \frac{k + c(u)}{h(u)} X^\lambda(\tau). \tag{62}$$

Though we have only demonstrated (62) for an integer k , both the left- and right-hand side of this identity are polynomials in k , and therefore (62) must hold for all $k \in \mathbb{C}$. Differentiating (62) and setting $k = 1$ requires some slightly tedious book-keeping, but is otherwise straightforward and gives us

$$\ell(\tau) = H_n X^{(n)}(\tau) + \sum_{\lambda} (-1)^v \left(\frac{1}{\lambda_2 + v} - \frac{1}{\lambda_1 + v + 1} \right) X^{(\lambda_1, \lambda_2, 1^v)}(\tau). \tag{63}$$

Applying this to the factorization types of $f \in \mathcal{M}_n$ gives the proposition. □

Proposition 9.8. For $f \in \mathcal{M}_n$,

$$\mu(f)\omega(f) = (-1)^n \left[H_n X^{(1^n)}(f) + \sum (-1)^v \left(\frac{1}{j+v+1} - \frac{1}{i+j+v+2} \right) X^{(v+2, 2^j, 1^i)}(f) \right] + b(f), \tag{64}$$

where the sum is over all partitions $(v + 2, 2^j, 1^i) \vdash n$, with $i, j, v \geq 0$, and $b(f)$ is a function supported on the squarefals.

Proof. For f squarefree with factorization type τ , note that $\mu(f)\omega(f) = (-1)^{\ell(\tau)} \ell(\tau)$. But by applying Proposition 5.4 to the identity (63), we may decompose $(-1)^{\ell(\tau)} \ell(\tau)$ into a sum over irreducible characters associated to dual partitions. This decomposition yields (64). \square

9B. By applying Theorem 3.1 to Propositions 9.7 and 9.8 we straightforwardly obtain the following results:

Corollary 9.9. For fixed $0 \leq h \leq n - 5$,

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f; h)} \omega(g) \right) = q^{h+1} \sum_{1 \leq \lambda_2 \leq \lambda_1 \leq n-h-2} \sum_{\lambda_1 + \lambda_2 \leq n} \left(\frac{1}{n - \lambda_1} - \frac{1}{n - \lambda_2 + 1} \right)^2 + O(q^{h+1/2}). \tag{65}$$

Corollary 9.10. For fixed $0 \leq h \leq n - 5$,

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f; h)} \mu(g)\omega(g) \right) = q^{h+1} \left[H_n^2 + \sum_{h+2 \leq i+2j \leq n-2} \sum \left(\frac{1}{n-i-j-1} - \frac{1}{n-j} \right)^2 \right] + O(q^{h+1/2}). \tag{66}$$

9C. Because the double-indexed sum in the asymptotic formula of (66) remains bounded for $n \rightarrow \infty$ and $h \sim \delta n$ with $\delta > 0$, and because $H_n \sim \log n = \log \deg(f)$ for $f \in \mathcal{M}_n$, one may think of Corollary 9.10 as a function field analogue of the following conjecture over the integers (which is intuitive enough on its own):

Conjecture 9.11. For $H = X^\delta$ with fixed $\delta \in (0, 1)$, as $X \rightarrow \infty$, we have

$$\frac{1}{X} \int_X^{2X} \left(\sum_{x \leq n \leq n+H} \mu(n)\omega(n) \right)^2 dx \sim H(\log \log X)^2.$$

Corollary 9.9 has a rather more striking interpretation. In (65) the double indexed sum remains bounded for $h \sim \delta n$ with $\delta \in (0, 1)$ fixed; indeed the reader may check that

$$\sum_{\substack{1 \leq \lambda_2 \leq \lambda_1 \leq n-h-2 \\ \lambda_1 + \lambda_2 \leq n}} \left(\frac{1}{n - \lambda_1} - \frac{1}{n - \lambda_2 + 1} \right)^2 \sim p(\delta) < +\infty \tag{67}$$

as $n \rightarrow \infty$ for⁴

$$p(\delta) := \int_{\substack{x+y \geq 1 \\ \delta \leq x \leq y \leq 1}} \left(\frac{1}{x} - \frac{1}{y} \right)^2 dx dy.$$

⁴One can further reduce the integral to see

$$p(\delta) = \begin{cases} \log((1-\delta)/\delta) + \delta - \text{Li}_2(1-\delta) + \text{Li}_2(\delta) - \log(1-\delta) \log(\delta) & \text{for } \delta \leq 1/2, \\ (1-\delta)/\delta - (1-\delta) - \log(\delta)^2 & \text{for } \delta > 1/2, \end{cases}$$

where Li_2 is the dilogarithm. Note the phase change at $\delta = \frac{1}{2}$.

Because this is bounded it is reasonable to suppose:

Conjecture 9.12. For $H = X^\delta$ with fixed $\delta \in (0, 1)$ as $X \rightarrow \infty$ we have

$$\frac{1}{X} \int_X^{2X} \left(\sum_{x \leq n \leq x+H} \omega(n) \right)^2 dx - \left(\frac{1}{X} \int_X^{2X} \sum_{x \leq n \leq x+H} \omega(n) dx \right)^2 = O_\delta(H). \tag{68}$$

There is a sense in which an estimate of the sort in Conjecture 9.12 would be surprising, since the Erdős–Kac theorem [1940] predicts that diagonal terms make a contribution of size $H \log \log X$. Clearly that $\delta \in (0, 1)$ remain fixed is important for anything like Conjecture 9.12 to be true — the consideration of diagonal terms shows that we cannot have such an estimate if $\delta \rightarrow 0$ as $X \rightarrow \infty$. Nonetheless the function field analogy remains, and it would be interesting to study in greater depth whether Conjecture 9.12 is true.⁵

Rather more ambitiously, one may even guess that the right-hand side of (68) can be replaced by

$$p(\delta)H + o_\delta(H).$$

10. Covariance

10A. In analogy with the definition of variance, (4), we define the covariance of two arithmetic functions η_1 and η_2 by

$$\text{Covar}_{f \in \mathcal{M}_n}(\eta_1(f), \eta_2(f)) := \frac{1}{q^n} \sum_{f \in \mathcal{M}_n} (\eta_1(f) - \mathbb{E}_{\mathcal{M}_n} \eta_1) \overline{(\eta_2(f) - \mathbb{E}_{\mathcal{M}_n} \eta_2)}.$$

Because Theorem 3.1 holds for a general factorization function a , it implies by a standard argument a corresponding result for covariance.

Theorem 10.1. For $a(f)$ and $b(f)$ fixed factorization functions and for fixed $0 \leq h \leq n - 5$,

$$\text{Covar}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} a(g), \sum_{g \in I(f;h)} b(g) \right) = q^{h+1} \sum_{\substack{\lambda \vdash n \\ \lambda_1 \leq n-h-2}} \hat{a}_\lambda \overline{\hat{b}_\lambda} + O(q^{h+1/2}).$$

One consequence of this is worthwhile to draw out. Since $\mu(g) = X^{(1^n)}(g)$, we see directly that:

Corollary 10.2. For $a(f)$ a fixed factorization function and for fixed $0 \leq h \leq n - 5$,

$$\text{Covar}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} a(g), \sum_{g \in I(f;h)} \mu(g) \right) = q^{h+1} \hat{a}_{(1^n)} + O(q^{h+1/2}) = q^{h+1} \cdot \frac{1}{q^n} \sum_{f \in \mathcal{M}_n} \mu(g) a(g) + o(q^{h+1}).$$

That is to say, the Möbius function oscillates to such an extent that in estimating its short-interval-sum covariance against any factorization function, only diagonal terms contribute. It is easy to see that (up to values on the squarefuls) the Möbius function is unique among factorization functions in this regard.

⁵Andrew Granville (personal communication) has shown a variant of this conjecture is true for a restricted range of δ , when $\omega(n)$ is replaced by $\omega_y(n)$, a count of prime factors of n less than $y = X^{1/2-\epsilon}$.

For example, Corollary 10.2 implies

$$\text{Covar}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} \Lambda(g), \sum_{g \in I(f;h)} \mu(g) \right) \sim -q^{h+1},$$

as $q \rightarrow \infty$. Over the integers we have the following analogy:

Conjecture 10.3. For $H = X^\delta$ with $\delta \in (0, 1)$,

$$\frac{1}{X} \int_X^{2X} \left(\sum_{x \leq n \leq x+H} \Lambda(n) - H \right) \left(\sum_{x \leq n \leq x+H} \mu(n) \right) dx \sim -H,$$

as $X \rightarrow \infty$.

11. Decompositions: proofs of Theorem 3.2 and Corollary 3.5

11A. We now turn to the decomposition of the space of factorization functions \mathcal{F} into \mathcal{U}_n^h and \mathcal{V}_n^h and the corresponding evaluation of variance described in Theorem 3.2. Recall that \mathcal{U}_n^h is the linear space of functions defined by (17) and \mathcal{V}_n^h is orthogonal complement supported on squarefuls. We first demonstrate the explicit characterization of the spaces \mathcal{U}_n^h and \mathcal{V}_n^h given by Proposition 3.3.

Proof of Proposition 3.3. Let \mathcal{A}_n^h and \mathcal{B}_n^h be as in the proposition and

$$\mathcal{C}_n^h := \text{span}\{X^\lambda(f) : \lambda \vdash n, \lambda_1 \leq n - h - 2\}.$$

Note that \mathcal{C}_n^h is supported on the squarefrees, and

$$\mathcal{F} = (\mathcal{A}_n^h \oplus \mathcal{B}_n^h) \oplus \mathcal{C}_n^h.$$

Moreover, by the equidistribution of factorization types and cycles types and the orthogonality of characters X^λ , \mathcal{A}_n^h is orthogonal to \mathcal{C}_n^h .

Theorem 3.1 implies that $\mathcal{A}_n^h \oplus \mathcal{B}_n^h \subset \mathcal{U}_n^h$, and likewise that $\mathcal{C}_n^h \cap \mathcal{U}_n^h = \{0\}$, so that no function outside of $\mathcal{A}_n^h \oplus \mathcal{B}_n^h$ lies in \mathcal{U}_n^h ; that is, $\mathcal{A}_n^h \oplus \mathcal{B}_n^h = \mathcal{U}_n^h$. \mathcal{V}_n^h , defined to be the orthogonal complement supported on squarefuls, is thus identical with \mathcal{C}_n^h , which proves the proposition. □

Proof of Theorem 3.2. Note that for $v \in \mathcal{V}_n^h$ with

$$v(f) = \sum_{\lambda_1 \leq n-h-2} \hat{v}_\lambda X^\lambda(f),$$

we have

$$\langle v, v \rangle = \lim_{q \rightarrow \infty} \frac{1}{q^n} \sum_{f \in \mathcal{M}_n} v(f) \overline{v(f)} = \sum_{\lambda_1 \leq n-h-2} |\hat{v}_\lambda|^2,$$

by again making use of the equidistribution of factorization types and cycle types (Proposition 2.3). Combined with Theorem 3.1, this gives the result. □

11B. We now turn to Proposition 3.4 and Corollary 3.5.

Proof of Proposition 3.4. We note first that for any factorization function α , it is simple to see that

$$w(f) := \sum_{\substack{\delta \mid f \\ \deg(\delta) \leq h+1}} \alpha(\delta), \quad (\text{defined for } f \in \mathcal{M}_n)$$

lies in \mathcal{U}_n^h . (Recall that \mathcal{U}_n^h is defined by (17).) For in this case, for any $f \in \mathcal{M}_n$,

$$\sum_{g \in I(f;h)} q(g) = \sum_{\deg(\delta) \leq h+1} \alpha(\delta) \sum_{\substack{g \in I(f;h) \\ \delta \mid g}} 1 = \sum_{\deg(\delta) \leq h+1} \alpha(\delta) q^{h+1-\deg(\delta)}.$$

This does not depend on f , so that

$$\text{Var}_{f \in \mathcal{M}_n} \left(\sum_{g \in I(f;h)} w(g) \right) = 0,$$

implying $w \in \mathcal{U}_n^h$. Since we already know any factorization function $b \in \mathcal{F}_n$ supported on the squarefuls lies in the linear space \mathcal{U}_n^h , and function of the form $w(f) + b(f)$ must therefore lie in \mathcal{U}_n^h .

Hence to complete the proof of the proposition, we need only show that all functions in \mathcal{U}_n^h are of this form. Having already characterized \mathcal{U}_n^h in terms of characters of the symmetric group in Proposition 3.3, we will have done so if we show that for $\lambda \vdash n$ with $\lambda_1 \geq n - h - 1$, there exists a factorization function α and a factorization function b supported on the squarefuls such that

$$X^\lambda(f) = \sum_{\substack{\delta \mid f \\ \deg(\delta) \leq h+1}} \alpha(\delta) + b(f), \quad (\text{for all } f \in \mathcal{M}_n).$$

The remainder of this proof is devoted to a demonstration in four steps of this claim.

Step 1: Let m be arbitrary. For an even primitive character χ modulo T^m , from the identity

$$\left(1 - \frac{u}{\sqrt{q}}\right) \prod_{j=1}^{m-2} (1 - ue^{i2\pi\vartheta_j}) = \mathcal{L}\left(\frac{u}{\sqrt{q}}, \chi\right) = \sum_{n \geq 0} u^n \frac{1}{q^{n/2}} \sum_{f \in \mathcal{M}_n} \chi(f),$$

we have the following expression for elementary symmetric functions in the normalized roots of the \mathcal{L} -function:

$$e_n(\Theta_\chi) = \frac{(-1)^n}{q^{n/2}} \sum_{f \in \mathcal{M}_n} \chi(f) + O_{n,m}(q^{-1/2}). \tag{69}$$

Step 2: We note for $n_1 + \dots + n_k = n$,

$$\begin{aligned} e_{n_1}(\Theta_\chi) \cdots e_{n_k}(\Theta_\chi) &= \frac{(-1)^n}{q^{n/2}} \left(\sum_{f_1 \in \mathcal{M}_{n_1}} \chi(f_1) + O_{n,m}(q^{-1/2}) \right) \cdots \left(\sum_{f_k \in \mathcal{M}_{n_k}} \chi(f_k) + O_{n,m}(q^{-1/2}) \right) \\ &= \frac{(-1)^n}{q^{n/2}} \sum_{\substack{f_1 \in \mathcal{M}_{n_1} \\ g \in \mathcal{M}_{n_2+\dots+n_k}}} \chi(f_1 g) \alpha(g) + O_{n,m}(q^{-1/2}), \end{aligned}$$

where

$$\alpha(g) := \sum_{\substack{f_2 \cdots f_k = g \\ f_2 \in \mathcal{M}_2, \dots, f_k \in \mathcal{M}_k}} 1$$

is a factorization function supported on $\mathcal{M}_{n_2 + \dots + n_k}$. In particular, we have that if $n_1 \geq n - h - 1$ (so $n_2 + \dots + n_k \leq h + 1$) then

$$e_{n_1} \cdots e_{n_k} = \frac{(-1)^n}{q^{n/2}} \sum_{f \in \mathcal{M}_n} \chi(f) \sum_{\delta | f} \alpha(\delta) + O_{n,m}(q^{-1/2}), \tag{70}$$

for a factorization function $\alpha(\delta)$ supported on the set of δ with $\deg(\delta) \leq h + 1$.

Step 3: From an expansion of the determinant in the Jacobi–Trudi identity, we see for $\lambda \vdash n$ that $s_{\lambda'}$ is a linear combination of terms $e_{n_1} \cdots e_{n_k}$ with $n_1 + \dots + n_k = n$ and (from the top row of the determinant) $n_1 \geq \lambda_1$ always. Hence via step 2, if $\lambda_1 \geq n - h - 1$,

$$s_{\lambda'}(\Theta_{\chi}) = \frac{(-1)^n}{q^{n/2}} \sum_{f \in \mathcal{M}_n} \chi(f) \sum_{\delta | f} \alpha(\delta) + O_{n,m}(q^{-1/2}), \tag{71}$$

for a factorization function $\alpha(\delta)$ supported on δ with $\deg(\delta) \leq h + 1$, since linear combinations of terms of the form $\sum_{\delta | f} \alpha(\delta)$ remain of this form.

Yet from Theorem 7.1

$$s_{\lambda'}(\Theta_{\chi}) = \frac{(-1)^n}{q^{n/2}} \sum_{f \in \mathcal{M}_n} X^{\lambda}(f) \chi(f) + O(q^{-1/2}). \tag{72}$$

Hence pairing (71) and (72) we have

$$\frac{1}{q^{n/2}} \sum_{f \in \mathcal{M}_n} \chi(f) \left(X^{\lambda}(f) - \sum_{\substack{\delta | f \\ \deg(\delta) \leq h+1}} \alpha(\delta) \right) = O_{n,m}(q^{-1/2}). \tag{73}$$

Step 4: In (73), m is arbitrary; take m sufficiently large depending on n , with the intention of using Lemma 7.4. We have upon squaring and averaging,

$$\mathbb{E}_{\chi(T^m)_{\text{prim, ev}}} \left| \frac{1}{q^{n/2}} \sum_{f \in \mathcal{M}_n} \chi(f) \left(X^{\lambda}(f) - \sum_{\substack{\delta | f \\ \deg(\delta) \leq h+1}} \alpha(\delta) \right) \right|^2 \rightarrow 0,$$

as $q \rightarrow \infty$. But then from Lemma 7.4,

$$\left\| X^{\lambda}(f) - \sum_{\substack{\delta | f \\ \deg(\delta) \leq h+1}} \alpha(\delta) \right\| = 0,$$

for $\|\cdot\|$ the norm induced by our inner product. Since this inner product is nondegenerate on functions supported on the squarefrees, we must have

$$X^\lambda(f) = \sum_{\substack{\delta|f \\ \deg(\delta) \leq h+1}} \alpha(\delta) + b(f),$$

for some function $b(f)$ supported on the squarefuls, as claimed. □

Proof of Corollary 3.5. This follows immediately from Theorem 3.2 and Proposition 3.4. For in the identity (19), the function $v(f)$ is a projection of the function $a(f)$ to the subspace \mathcal{V}_n^h , but then

$$\langle v, v \rangle = \|\text{Proj}_{\mathcal{V}_n^h}(a)\|^2 = \inf_{u \in \mathcal{U}_n^h} \|a - u\|^2 = \inf_{\alpha \in \mathcal{F}} \left\| a(f) - \sum_{\substack{\delta|f \\ \deg(\delta) \leq h+1}} \alpha(\delta) \right\|^2.$$

□

11C. It is worthwhile to reflect one last time on the dichotomy between \mathcal{U}_n^h and \mathcal{V}_n^h . Theorem 7.1 gives us another way to characterize them. \mathcal{U}_n^h is just the collection of those factorization functions u for which

$$\sum_{f \in \mathcal{M}_n} u(f)\chi(f) = O(q^{n/2-1/2}), \tag{74}$$

uniformly for all even primitive characters modulo T^{n-h} . The reason that Theorem 7.1 implies (74) is very simply that $\mathcal{L}(u, \chi)$ always has $n - h - 2$ nontrivial zeros. Contrariwise, Theorem 3.2 and Proposition 3.3 tell us that for those factorization functions which do not have enough structure to belong to \mathcal{U}_n^h their variance may be computed according to the most naive heuristic of randomness. Indeed, one last reformulation of Theorem 3.2 may be seen to be the following: for $v_1, v_2 \in \mathcal{V}_n^n$,

$$\mathbb{E}_{\chi(T^{n-h})} \sum_{\substack{\text{prim. ev} \\ f, g \in \mathcal{M}_n \\ f \neq g}} v_1(f)\chi(f)\overline{v_2(g)\chi(g)} = o(q^n). \tag{75}$$

It would be interesting to see whether a modification of this picture is consistent with conjectures that have been made in other settings (e.g., in the fixed q large n limit, or over number fields), or indeed with statistics in orthogonal and symplectic families.

Acknowledgments

For helpful discussions, I thank a number of people, including Efrat Bank, Dan Bump, Reda Chhaibi, Paul-Olivier Dehaye, Andrew Granville, Jeff Lagarias, Zeev Rudnick, Will Sawin, and especially Ofir Gorodetsky, who made a number of careful suggestions and corrections to an earlier draft that have greatly improved the paper. I also want to thank Jordan Ellenberg, Daniel Hast, and Vlad Matei; the decomposition of Proposition 3.4 came out of discussions with them during a visit to Madison. A discussion on the website MathOverflow, available at <http://mathoverflow.net/q/233167>, was useful for

finding a reference. Part of this work was done while I was a postdoctoral fellow at the University of Zürich, and I thank that institution for its hospitality. Finally I thank the anonymous referee for a very attentive reading of the manuscript and a number of corrections and comments.

References

- [Andrade et al. 2015] J. C. Andrade, L. Bary-Soroker, and Z. Rudnick, “Shifted convolution and the Titchmarsh divisor problem over $\mathbb{F}_q[t]$ ”, *Philos. Trans. Roy. Soc. A* **373**:2040 (2015), art. id. 20140308. MR Zbl
- [Bae et al. 2015] S. Bae, B. Cha, and H. Jung, “Möbius function in short intervals for function fields”, *Finite Fields Appl.* **34** (2015), 235–249. MR Zbl
- [Bogomolny and Keating 1995] E. B. Bogomolny and J. P. Keating, “Random matrix theory and the Riemann zeros, I: Three- and four-point correlations”, *Nonlinearity* **8**:6 (1995), 1115–1131. MR Zbl
- [Bogomolny and Keating 1996] E. B. Bogomolny and J. P. Keating, “Random matrix theory and the Riemann zeros, II: n -point correlations”, *Nonlinearity* **9**:4 (1996), 911–935. MR Zbl
- [Carlitz 1932] L. Carlitz, “The arithmetic of polynomials in a Galois field”, *Amer. J. Math.* **54**:1 (1932), 39–50. MR Zbl
- [Carmon 2015] D. Carmon, “The autocorrelation of the Möbius function and Chowla’s conjecture for the rational function field in characteristic 2”, *Philos. Trans. Roy. Soc. A* **373**:2040 (2015), art. id. 20140311. MR Zbl
- [Carmon and Rudnick 2014] D. Carmon and Z. Rudnick, “The autocorrelation of the Möbius function and Chowla’s conjecture for the rational function field”, *Q. J. Math.* **65**:1 (2014), 53–61. MR Zbl
- [Church et al. 2014] T. Church, J. S. Ellenberg, and B. Farb, “Representation stability in cohomology and asymptotics for families of varieties over finite fields”, pp. 1–54 in *Algebraic topology: applications and new directions*, edited by U. Tillmann et al., *Contemp. Math.* **620**, Amer. Math. Soc., Providence, RI, 2014. MR Zbl
- [Conrey and Keating 2015a] B. Conrey and J. P. Keating, “Moments of zeta and correlations of divisor-sums, I”, *Philos. Trans. Roy. Soc. A* **373**:2040 (2015), art. id. 20140313. MR Zbl
- [Conrey and Keating 2015b] B. Conrey and J. P. Keating, “Moments of zeta and correlations of divisor-sums, II”, pp. 75–85 in *Advances in the theory of numbers*, edited by A. Alaca et al., *Fields Inst. Commun.* **77**, Fields Inst. Res. Math. Sci., Toronto, 2015. MR Zbl
- [Conrey and Keating 2015c] B. Conrey and J. P. Keating, “Moments of zeta and correlations of divisor-sums, III”, *Indag. Math. (N.S.)* **26**:5 (2015), 736–747. MR Zbl
- [Conrey and Keating 2016] B. Conrey and J. P. Keating, “Pair correlation and twin primes revisited”, *Proc. A.* **472**:2194 (2016), art. id. 20160548. MR Zbl
- [Dehaye \geq 2018] P.-O. Dehaye, “Combinatorics of lower order terms in the moments conjecture for the Riemann zeta function”, in preparation.
- [Erdős and Kac 1940] P. Erdős and M. Kac, “The Gaussian law of errors in the theory of additive number theoretic functions”, *Amer. J. Math.* **62** (1940), 738–742. MR Zbl
- [Fulton 1997] W. Fulton, *Young tableaux: with applications to representation theory and geometry*, *Lond. Math. Soc. Student Texts* **35**, Cambridge Univ. Press, 1997. MR Zbl
- [Fulton and Harris 1991] W. Fulton and J. Harris, *Representation theory*, *Graduate Texts in Math.* **129**, Springer, 1991. MR Zbl
- [Gadish 2017] N. Gadish, “A trace formula for the distribution of rational G -orbits in ramified covers, adapted to representation stability”, *New York J. Math.* **23** (2017), 987–1011. MR Zbl
- [Gamburd 2007] A. Gamburd, “Some applications of symmetric functions theory in random matrix theory”, pp. 143–170 in *Ranks of elliptic curves and random matrix theory*, edited by J. B. Conrey et al., *Lond. Math. Soc. Lecture Note Ser.* **341**, Cambridge Univ. Press, 2007. MR Zbl
- [Goldston and Montgomery 1987] D. A. Goldston and H. L. Montgomery, “Pair correlation of zeros and primes in short intervals”, pp. 183–203 in *Analytic number theory and Diophantine problems* (Stillwater, OK, 1984), edited by A. C. Adolphson et al., *Progr. Math.* **70**, Birkhäuser, Boston, 1987. MR Zbl

- [Good and Churchhouse 1968] I. J. Good and R. F. Churchhouse, “The Riemann hypothesis and pseudorandom features of the Möbius sequence”, *Math. Comp.* **22**:104 (1968), 857–861. MR Zbl
- [Hast and Matei 2016] D. Hast and V. Matei, “Higher moments of arithmetic functions in short intervals: a geometric perspective”, preprint, 2016. arXiv
- [Katz 2013] N. M. Katz, “Witt vectors and a question of Keating and Rudnick”, *Int. Math. Res. Not.* **2013**:16 (2013), 3613–3638. MR Zbl
- [Katz and Sarnak 1999] N. M. Katz and P. Sarnak, “Zeroes of zeta functions and symmetry”, *Bull. Amer. Math. Soc. (N.S.)* **36**:1 (1999), 1–26. MR Zbl
- [Keating and Rudnick 2014] J. P. Keating and Z. Rudnick, “The variance of the number of prime polynomials in short intervals and in residue classes”, *Int. Math. Res. Not.* **2014**:1 (2014), 259–288. MR Zbl
- [Keating and Rudnick 2016] J. Keating and Z. Rudnick, “Squarefree polynomials and Möbius values in short intervals and arithmetic progressions”, *Algebra Number Theory* **10**:2 (2016), 375–420. MR Zbl
- [Keating et al. 2018] J. P. Keating, B. Rodgers, E. Roditty-Gershon, and Z. Rudnick, “Sums of divisor functions in $\mathbb{F}_q[t]$ and matrix integrals”, *Math. Z.* **288**:1-2 (2018), 167–198. MR Zbl
- [Macdonald 1995] I. G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed., Oxford Univ. Press, 1995. MR Zbl
- [Ng 2008] N. Ng, “The Möbius function in short intervals”, pp. 247–258 in *Anatomy of integers*, edited by J.-M. De Koninck et al., CRM Proc. Lecture Notes **46**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl
- [Rodgers 2015] B. Rodgers, “The covariance of almost-primes in $\mathbb{F}_q[T]$ ”, *Int. Math. Res. Not.* **2015**:14 (2015), 5976–6004. MR Zbl
- [Roditty-Gershon 2017] E. Roditty-Gershon, “Square-full polynomials in short intervals and in arithmetic progressions”, *Res. Number Theory* **3** (2017), art. id. 3. MR Zbl
- [Rosen 2002] M. Rosen, *Number theory in function fields*, Graduate Texts in Math. **210**, Springer, 2002. MR Zbl
- [Rudnick 2014] Z. Rudnick, “Some problems in analytic number theory for polynomials over a finite field”, pp. 443–459 in *Proceedings of the International Congress of Mathematicians, II* (Seoul, 2014), edited by S. Y. Jang et al., Kyung Moon Sa, Seoul, 2014. MR Zbl
- [Stanley 1999] R. P. Stanley, *Enumerative combinatorics, II*, Cambridge Studies in Adv. Math. **62**, Cambridge Univ. Press, 1999. MR Zbl
- [Weil 1967] A. Weil, *Basic number theory*, Die Grundlehren der Math. Wissenschaften **144**, Springer, 1967. MR Zbl
- [Weiss 2013] B. L. Weiss, “Probabilistic Galois theory over p -adic fields”, *J. Number Theory* **133**:5 (2013), 1537–1563. MR Zbl

Communicated by Peter Sarnak

Received 2017-09-30 Revised 2018-01-29 Accepted 2018-03-18

rbrad@umich.edu

*Department of Mathematics, University of Michigan, Ann Arbor, MI,
United States*

Cohomology for Drinfeld doubles of some infinitesimal group schemes

Eric M. Friedlander and Cris Negron

Consider a field k of characteristic $p > 0$, the r -th Frobenius kernel $\mathbb{G}_{(r)}$ of a smooth algebraic group \mathbb{G} , the Drinfeld double $D\mathbb{G}_{(r)}$ of $\mathbb{G}_{(r)}$, and a finite dimensional $D\mathbb{G}_{(r)}$ -module M . We prove that the cohomology algebra $H^*(D\mathbb{G}_{(r)}, k)$ is finitely generated and that $H^*(D\mathbb{G}_{(r)}, M)$ is a finitely generated module over this cohomology algebra. We exhibit a finite map of algebras $\theta_r : H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$, which offers an approach to support varieties for $D\mathbb{G}_{(r)}$ -modules. For many examples of interest, θ_r is injective and induces an isomorphism of associated reduced schemes. For M an irreducible $D\mathbb{G}_{(r)}$ -module, θ_r enables us to identify the support variety of M in terms of the support variety of M viewed as a $\mathbb{G}_{(r)}$ -module.

1. Introduction

For a Hopf algebra A over a field k , we denote by $H^*(A, k) = \text{Ext}_A^*(k, k)$ the Hopf cohomology and we denote by $H^*(A, M) = \text{Ext}_A^*(k, M)$ the cohomology of A with values in a finite dimensional A -module M . The goal of this paper is to prove the following conjecture for an interesting class of examples.

Conjecture (the finite generation conjecture). *For any finite dimensional Hopf algebra A , and finite dimensional A -module M , the cohomology $H^*(A, k)$ is a finitely generated k -algebra and $H^*(A, M)$ is a finitely generated module over $H^*(A, k)$.*

The conjecture has existed as a question at least since the 1990's (see e.g., [Friedlander and Suslin 1997]), and was recently stated explicitly in the work of Etingof and Ostrik [2004]. In the finite characteristic setting, the conjecture was verified for cocommutative Hopf algebras in the work of Friedlander and Suslin [1997]. This followed earlier work of Friedlander and Parshall [1986] on the cohomology of restricted enveloping algebras. More recently, Drupieski [2016] generalized these results to finite super groups (i.e., cocommutative Hopf algebras in the symmetric category of $\mathbb{Z}/2\mathbb{Z}$ -graded vector spaces).

For a commutative Hopf algebra A over a field of characteristic p , one can arrive at the desired finite generation result from the existence of an abstract algebra isomorphism $A \cong k[\mathbb{Z}/p^1\mathbb{Z}] \otimes \cdots \otimes k[\mathbb{Z}/p^l\mathbb{Z}]$ whenever A is local, and the fact that the cohomology $H^*(A, M)$ only depends on the algebra structure of A .

In characteristic 0 most of the work to date has focused on pointed Hopf algebras. Ginzburg and Kumar [1993] (see also [Bendel et al. 2014]) showed that small quantum groups have finitely generated

Friedlander's work was partially supported by the Simons Foundation. Negron was supported by NSF Postdoctoral Research Fellowship DMS-1503147.

MSC2010: primary 57T05; secondary 20G10, 20G40.

Keywords: Hopf cohomology, Drinfeld doubles, finite group schemes.

cohomology. Mastnak, Pevtsova, Schauenburg, and Witherspoon [2010] verified the finite generation conjecture for most pointed Hopf algebras with abelian group of grouplikes. Such Hopf algebras were classified by Andruskiewitsch and Schneider [2002], and can be understood broadly as deformations of small quantum groups. For results concerning pointed Hopf algebras with nonabelian grouplikes one can see [Ştefan and Vay 2016].

In this work we consider Drinfeld doubles of finite group schemes in characteristic $p > 0$. We recall that the Drinfeld double DG of a finite group scheme G is the smash product

$$DG = \mathbb{O}(G) \# kG$$

of the group algebra kG of G acting via the adjoint action on the algebra $\mathbb{O}(G) = (kG)^*$ of functions on G . The coalgebra structure on DG is the product structure $\mathbb{O}(G)^{\text{cop}} \otimes kG$, where the cop superscript indicates that we take the opposite comultiplication. The Drinfeld double DG is neither commutative nor cocommutative (unless G is commutative) and rarely pointed. For some examples of the computational and theoretical significance of the double one can see [Etingof and Gelaki 2002; Etingof 2002; Kashina et al. 2006; Ng and Schauenburg 2007; Shimizu 2017].

Our finite generation results for Drinfeld doubles apply to other Hopf algebras thanks to various general properties of the Drinfeld double construction. For example, $\text{rep}(DA) \cong \text{rep}(D(A^*))$ and for any cocycle twist σ (see [Montgomery 2004]), $\text{rep}(DA) \cong \text{rep}(D(A_\sigma))$ [Majid and Oeckl 1999; Benkart et al. 2010].

Let us now fix k a field of finite characteristic p . We assume additionally that p is odd, although most of our results will still hold when $p = 2$ (see Section 4C). Recall that the r -th Frobenius kernel $\mathbb{G}_{(r)}$ is the group scheme theoretic kernel of the r -th Frobenius map $F^r : \mathbb{G} \rightarrow \mathbb{G}^{(r)}$ (see Section 2). We refer the reader to [Sullivan 1978; Cline 1987; Jantzen 2003] for some discussion of the important role Frobenius kernels play in the modular representation theory of algebraic groups.

We prove the following:

Theorem (Theorems 5.3 and 5.6). *Consider the r -th Frobenius kernel $\mathbb{G}_{(r)}$ of a smooth algebraic group \mathbb{G} . The cohomology of the double $H^*(D\mathbb{G}_{(r)}, k)$ is a finitely generated k algebra. Moreover, for any finite dimensional $D\mathbb{G}_{(r)}$ -module M , the cohomology $H^*(D\mathbb{G}_{(r)}, M)$ is a finitely generated $H^*(D\mathbb{G}_{(r)}, k)$ -module.*

Our approach utilizes associations between deformation theory and Hopf cohomology. We show that the deformation $\mathbb{G}_{(r+1)}$ of $\mathbb{G}_{(r)}$ produces a natural map $\sigma_{\mathbb{G}} : \mathfrak{g}^{(r)} \rightarrow H^2(\mathbb{O}(\mathbb{G}_{(r)}), k)$, where $\mathfrak{g} = \text{Lie}(\mathbb{G})$. The map $\sigma_{\mathbb{G}}$ has a natural lift to the cohomology of the double $\sigma_{\mathbb{D}} : \mathfrak{g}^{(r)} \rightarrow H^2(D\mathbb{G}_{(r)}, k)$, which is again constructed in a deformation theoretic manner. The smoothness hypothesis of \mathbb{G} plays an important role in our proof. Namely, we apply an argument which uses in an essential way the structure of $\mathbb{G}_{(r+1)}$ as a flat extension of $\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}$ to obtain cohomology classes via deformation theory.

In proving the above theorem, we construct a finite algebra map

$$\theta_r : H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$$

(using σ_D and the inflation $H^*(\mathbb{G}_{(r)}, k) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$) with associated map of reduced affine schemes

$$\Theta_r : |D\mathbb{G}_{(r)}| \rightarrow |k\mathbb{G}_{(r)}| \times (\mathfrak{g}^*)^{(r)} \tag{1}$$

(see Theorem 5.3). Here, and elsewhere, we employ the usual notation $S(V[n])$ for the symmetric algebra over k of the k -vector space V placed in degree n , and we use the notation $|A| = \text{Spec } H^{\text{ev}}(A, k)_{\text{red}}$ for the reduced spectrum of the cohomology of a Hopf algebra A .

For many classical algebraic groups \mathbb{G} we are able to deduce additional information concerning Θ_r as formulated in the following theorem.

Theorem (Corollary 6.11). *Let \mathbb{G} be a general linear group, simple algebraic group, Borel subgroup in a simple algebraic group, or a unipotent subgroup in a semisimple algebraic group which is normalized by a maximal torus. Suppose that p is very good for \mathbb{G} , or that $p > \text{cl}(\mathbb{G})$ in the unipotent case:*

- *If $p > \dim \mathbb{G} + 1$ then the map Θ_r of (1) is an isomorphism for all r .*
- *For arbitrary p , the map Θ_r is an isomorphism whenever r is such that $p^r > 2 \dim \mathbb{G}$.*

The key observation we use in proving the above theorem is that the hypotheses guarantee the existence of a quasilogarithm $L : \mathbb{G} \rightarrow \mathfrak{g}$ [Kazhdan and Varshavsky 2006]. This leads to a grading on the Drinfeld double $D\mathbb{G}_{(r)}$ which greatly simplifies the analysis of the Lyndon–Hochschild–Serre spectral sequence we use to investigate the cohomology of $D\mathbb{G}_{(r)}$. The “very good” condition on p is a mild condition which we review in Section 6. In the unipotent case, the integer $\text{cl}(\mathbb{G})$ is the nilpotence class of \mathbb{G} , which is always less than $\dim(\mathbb{G})$. The theorem implies an equality of dimensions $\dim |D\mathbb{G}_{(r)}| = \dim |k\mathbb{G}_{(r)}| + \dim \mathbb{G}$ for such classical groups.

We also consider the support variety $|D\mathbb{G}_{(r)}|_M$ associated to a $D\mathbb{G}_{(r)}$ -module M . The support variety for M is defined as the closed, reduced, subscheme in $|D\mathbb{G}_{(r)}|$ defined by the kernel of the algebra map

$$- \otimes M : H^{\text{ev}}(D\mathbb{G}_{(r)}, k) \rightarrow \text{Ext}_{D\mathbb{G}_{(r)}}^{\text{ev}}(M, M).$$

Theorem (Corollary 7.6). *Suppose \mathbb{G} is as in the statement of the previous theorem. If $p > \dim \mathbb{G} + 1$ or $p^r > 2 \dim \mathbb{G}$, then for any irreducible $D\mathbb{G}_{(r)}$ -module M the map Θ_r of (1) restricts to an isomorphism of schemes*

$$\Theta_{r,M} : |D\mathbb{G}_{(r)}|_M \xrightarrow{\sim} |k\mathbb{G}_{(r)}|_M \times (\mathfrak{g}^*)^{(r)}.$$

We supplement the preceding results by extending many of them to relative Drinfeld doubles (see Section 5C).

Organization. In Section 3, we discuss associations between deformations and Hopf cohomology, and produce the aforementioned maps $\sigma_{\mathbb{G}}$ and σ_D . In Section 4 we prove that the algebra map $S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(\mathbb{G}_{(r)}, k)_{\text{red}}$ induced by $\sigma_{\mathbb{G}}$ is an isomorphism. We use the lifting σ_D , in conjunction with the inflation map $H^*(\mathbb{G}_{(r)}, k) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$, to establish the finite generation of cohomology for the double $D\mathbb{G}_{(r)}$ in Section 5. Section 6 is dedicated to an analysis of classical groups at large primes. Section 7 is dedicated to support varieties.

2. Finite group schemes and the Frobenius

We fix from this point on a field k of finite characteristic p . We assume $p \neq 2$ (see Section 4C). A “scheme” is a scheme of finite type over k and $\otimes = \otimes_k$. All schemes considered in this work will be affine. Throughout, by an algebraic group we mean an affine group scheme which is reduced, absolutely irreducible, and of finite type over k .

For an affine group scheme G , a rational (left) G -representation is a (right) comodule over the coordinate algebra $\mathbb{O}(G)$. A G -algebra is an $\mathbb{O}(G)$ -comodule algebra (i.e., an algebra R which is a rational G -representation in such a way that the multiplication $R \otimes R \rightarrow R$ is a map of G -representations). We let $H^*(G, M)$ denote the rational group cohomology of G with coefficients in M . If G is a finite group scheme with Hopf algebra kG (the “group algebra” of G), then $H^*(G, M) = H^*(kG, M)$.

In this section we review some standard information on Frobenius maps and Frobenius kernels. One can see Jantzen’s book [2003] for a more detailed presentation.

2A. Frobenius maps and Frobenius kernels. Let $\phi^r : k \rightarrow k$ be the p^r -th power map on k , $\lambda \mapsto \lambda^{p^r}$. Given an affine group scheme G we let $G^{(r)}$ denote the fiber product of G with $\text{Spec}(k)$ along ϕ^r ,

$$\begin{array}{ccc} G^{(r)} & \longrightarrow & G \\ \downarrow & & \downarrow \\ \text{Spec}(k) & \xrightarrow{(\phi^r)^*} & \text{Spec}(k). \end{array}$$

By functoriality of the pullback we see that $(-)^{(r)}$ provides a functor on the category of group schemes over k . There is a natural map of group schemes $F^r : G \rightarrow G^{(r)}$ over k given explicitly as follows: $\mathbb{O}(G^{(r)}) = \mathbb{O}(G) \otimes_{\phi^r} k \rightarrow \mathbb{O}(G)$ sends $f \otimes_{\phi^r} \lambda$ to $\lambda \cdot f^{p^r}$.

- Definition 2.1.** (i) The above map $F^r : G \rightarrow G^{(r)}$ is called the r -th Frobenius map.
 (ii) The r -th Frobenius kernel $G_{(r)}$ of G is the group scheme theoretic kernel of the r -th Frobenius map, $1 \rightarrow G_{(r)} \rightarrow G \xrightarrow{F^r} G^{(r)}$.
 (iii) We say G is of height $\leq r$ if $G = G_{(r)}$.

The closed subgroup scheme $G_{(r)}$ in G is the spectrum of the quotient Hopf algebra

$$\mathbb{O}(G_{(r)}) = \mathbb{O}(G) / (f^{p^r} : f \in m_G),$$

where m_G is the maximal ideal corresponding to the identity in G . Whence we see that an affine group scheme G is of height $\leq r$ if and only if $f^{p^r} = 0$ for each $f \in m_G$.

Example 2.2. For G a height 1 group scheme, we have $G = \text{Spec}(u(\mathfrak{g})^*)$ where \mathfrak{g} is the restricted Lie algebra for G and $u(\mathfrak{g})$ is the restricted enveloping algebra. This association gives a natural bijection between height 1 group schemes and finite dimensional restricted Lie algebras.

Example 2.3. Consider GL_n . This is the spectrum of the Hopf algebra

$$\mathbb{O}(GL_n) = k[x_{ij}, \det^{-1} : 1 \leq i, j \leq n]$$

with comultiplication $\Delta(x_{ij}) = \sum_{k=1}^n x_{ik} \otimes x_{kj}$, counit $\epsilon(x_{ij}) = \delta_{ij}$, and antipode given by the adjoint formula for the inverse of a matrix. The Frobenius kernels in this case are given by

$$\mathbb{O}(GL_{n(r)}) = k[x_{ij} : 1 \leq i, j \leq n] / (x_{ij}^{p^r} - \delta_{ij}).$$

Note that in the above presentation of the Frobenius kernel the determinant is already invertible.

2B. Frobenius twists of representations. For a rational G -representation V we let $V^{(r)}$ denote the new G -representation which is the vector space $k \otimes_{\phi^r} V$ along with the G -action given by the composite

$$G \xrightarrow{\phi^r} G^{(r)} \rightarrow GL(V)^{(r)} = GL(V^{(r)}).$$

The tensor product $k \otimes_{\phi^r}$ here denotes base change along ϕ^r . As a comodule, $V^{(r)}$ has right $\mathbb{O}(G)$ -coaction given by

$$\rho^{(r)}(c \otimes v) = \sum_i (c \otimes v_{i_0}) \otimes v_{i_1}^{p^r},$$

where the initial coaction of $\mathbb{O}(G)$ on V is given by $\rho(v) = \sum_i v_{i_0} \otimes v_{i_1}$. (In the above equation $c \in k$ and $v \in V$.) We call $V^{(r)}$ the r -th Frobenius twist of V . The proof of the following lemma is immediate from the observation that the composition $\mathbb{O}(G^{(r)}) \rightarrow \mathbb{O}(G) \rightarrow \mathbb{O}(G_{(r)})$ factors through the counit for $\mathbb{O}(G^{(r)})$.

Lemma 2.4. *For G of height $\leq r$ and V any rational G -representation, G acts trivially on the r -th Frobenius twist $V^{(r)}$.*

We also employ a natural isomorphism of G -representations $(V^*)^{(r)} \xrightarrow{\sim} (V^{(r)})^*$ given by the formula $c \otimes f \mapsto (c' \otimes v \mapsto cc' f(v)^{p^r})$.

3. Deformations of Frobenius kernels and cohomology

We fix a positive integer r and consider the r -th Frobenius kernel $\mathbb{G}_{(r)}$ of a smooth linear algebraic group over k , a field of odd characteristic $p > 0$. We denote by $D\mathbb{G}_{(r)}$ the Drinfeld double of the Hopf algebra $k\mathbb{G}_{(r)}$. This is the smash product $D\mathbb{G}_{(r)} = \mathbb{O}(\mathbb{G}_{(r)}) \# k\mathbb{G}_{(r)}$ of the coordinate algebra $\mathbb{O}(\mathbb{G}_{(r)})$ with the group algebra $k\mathbb{G}_{(r)}$ with respect to the right adjoint action of $\mathbb{G}_{(r)}$ on itself [Montgomery 1993, Corollary 10.3.10].

The adjoint action of $\mathbb{G}_{(r)}$ on itself corresponds to the $\mathbb{O}(\mathbb{G}_{(r)})$ -coaction $\rho(f) = \sum_i f_{i_2} \otimes S(f_{i_1}) f_{i_3}$ specifically, and subsequent $k\mathbb{G}_{(r)}$ -action $\xi \cdot f = \sum_i f_{i_2} \xi(S(f_{i_1}) f_{i_3})$. The Hopf structure on $D\mathbb{G}_{(r)}$ is the unique one so that the two inclusions $\mathbb{O}(\mathbb{G}_{(r)})^{\text{cop}} \rightarrow D\mathbb{G}_{(r)}$ and $k\mathbb{G}_{(r)} \rightarrow D\mathbb{G}_{(r)}$ are maps of Hopf algebras.

We proceed to construct cohomology classes in $H^2(D\mathbb{G}_{(r)}, k)$ which will enable our proof of finite generation in Section 5. Our construction involves deformations of $\mathbb{O}(\mathbb{G}_{(r)})$ and $D\mathbb{G}_{(r)}$, in particular the embedding $\mathbb{G}_{(r)} \rightarrow \mathbb{G}_{(r+1)}$ which we view as a deformation of $\mathbb{G}_{(r)}$ parametrized by $\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}$. This deformation leads to classes in the Hochschild cohomology group $HH^2(D\mathbb{G}_{(r)}, k)$ and thereby classes in $H^2(D\mathbb{G}_{(r)}, k)$.

3A. Hochschild cohomology and deformations. We recall that the Hochschild cohomology of an algebra R with coefficients in an R -bimodule M is defined as

$$HH^*(R, M) \equiv \text{Ext}_{R \otimes R^{\text{op}}}^*(R, M),$$

and $HH^*(R) = HH^*(R, R)$. Thus, $HH^*(R, M)$ is functorial with respect to maps $M \rightarrow N$ of R -bimodules. Moreover, we have the well-known surjection (see [Ginzburg and Kumar 1993, §5.6] or [Pevtsova and Witherspoon 2009, §7])

$$HH^*(R) \cong H^*(R, R^{\text{ad}}) \twoheadrightarrow H^*(R, k) \equiv \text{Ext}_R^*(k, k), \quad \text{if } R \text{ is a Hopf algebra}$$

(using the fact that $k \rightarrow R^{\text{ad}}$ splits). We further recall that a (infinitesimal) deformation \mathcal{R} of an algebra R parametrized by a scheme $\text{Spec}(A)$ is a flat A -algebra, where A is an Artin local (commutative) algebra with residue field k , equipped with a fixed isomorphism $\mathcal{R} \otimes_A k \xrightarrow{\sim} R$. Given any map $A \rightarrow A'$ of such Artinian local algebras and a deformation \mathcal{R} parametrized by $\text{Spec}(A)$, we can extend \mathcal{R} along $A \rightarrow A'$ to get a deformation $\mathcal{R} \otimes_A A'$ parametrized by $\text{Spec}(A')$. Two deformations \mathcal{R} and \mathcal{R}' parametrized by $\text{Spec}(A)$ are said to be isomorphic if there is an A -algebra isomorphism $l : \mathcal{R} \rightarrow \mathcal{R}'$ fitting into a diagram

$$\begin{array}{ccc} \mathcal{R} & \xrightarrow{l} & \mathcal{R}' \\ & \searrow & \swarrow \\ & R & \end{array} .$$

A special role is played by deformations parametrized by $\text{Spec}(k[\varepsilon])$, where $k[\varepsilon] \equiv k[t]/t^2$ is the Artin local algebra of “dual numbers”.

Theorem 3.1 [Gerstenhaber 1964]. *There is a naturally constructed bijection*

$$\{\text{deformations of } R \text{ parametrized by } \text{Spec}(k[\varepsilon])\} / \cong \xrightarrow{\sim} HH^2(R). \tag{2}$$

The domain of the above bijection has a natural linear structure under which (2) is a linear isomorphism. Let us explain some of the details of Gerstenhaber’s result.

Consider a deformation \mathcal{R} of R over $\text{Spec}(k[\varepsilon])$. By choosing a $k[\varepsilon]$ -linear isomorphism $R[\varepsilon] \equiv R \otimes k[\varepsilon] \cong \mathcal{R}$ the deformation \mathcal{R} may be identified with the $k[\varepsilon]$ -module $R[\varepsilon]$ equipped with a multiplication

$$a \cdot_{\mathcal{R}} b = ab + F_{\mathcal{R}}(a, b)\varepsilon, \quad a, b \in R \subset R \otimes k[\varepsilon].$$

The function $F_{\mathcal{R}} : R \otimes R \rightarrow R$ defines a 2-cocycle in the standard Hochschild cochain complex

$$C^*(R) = 0 \rightarrow R \rightarrow \text{Hom}_k(R, R) \rightarrow \text{Hom}_k(R \otimes R, R) \rightarrow \text{Hom}_k(R^{\otimes 3}, R) \rightarrow \dots$$

This determines a map from deformations to $HH^2(R)$. To define the inverse map, one simply uses a 2-cocycle in the standard Hochschild cochain complex to define a multiplication on $R[\varepsilon]$. The addition of isoclasses of deformations $[\mathcal{R}] + [\mathcal{R}']$ corresponds to addition of the functions $F_{\mathcal{R}} + F_{\mathcal{R}'}$ and scaling $c[\mathcal{R}]$ corresponds to scaling the function $cF_{\mathcal{R}}$.

The following lemma should be standard.

Lemma 3.2. *Let \mathcal{R} be an (infinitesimal) deformation of R parametrized by $S = \text{Spec}(A)$. Then there is a k -linear mapping*

$$\Sigma_{\mathcal{R}} : T_p S \rightarrow HH^2(R)$$

which sends an element $\xi \in T_p S = \text{Hom}_{\text{Alg}}(A, k[\varepsilon])$ to the class corresponding to the deformation $\mathcal{R} \otimes_A k[\varepsilon]$, where we change base via ξ .

In the statement of the above lemma p is the unique point in S .

Proof. Given $\xi \in T_p S$ we let $\text{Def}_{\xi} = \mathcal{R} \otimes_A k[\varepsilon]$ denote the corresponding deformation. For the proof we identify the tangent space $T_p S$ with k -linear maps $m_A \rightarrow k$ which vanish on m_A^2 , where m_A is the unique maximal ideal of A . We adopt an A -linear identification $\mathcal{R} = R \otimes A$, and write the multiplication on \mathcal{R} as $r \cdot_{\mathcal{R}} r' = rr' + E(r, r')$, where $r, r' \in R$ and E is a linear function $E : R \otimes R \rightarrow R \otimes m_A$.

If we take $F_{\xi} = (1 \otimes \xi)E$, for $\xi \in T_p S$, then the multiplication on the base change Def_{ξ} is given by $r \cdot_{\xi} r' = rr' + F_{\xi}(r, r')\varepsilon$. Whence we have an equality in Hochschild cohomology

$$\Sigma_{\mathcal{R}}(\xi) = [\text{Def}_{\xi}] = [F_{\xi}] \in HH^2(R).$$

By the definition of F_{ξ} we see that $F_{c\xi+c'\xi'} = cF_{\xi} + c'F_{\xi'}$. It follows that the map $\Sigma_{\mathcal{R}} : T_p S \rightarrow HH^2(R)$ is k -linear. □

Definition 3.3. Given a deformation \mathcal{R} of a Hopf algebra R parametrized by S , we let

$$\sigma_{\mathcal{R}} : T_p S \rightarrow H^2(R, k)$$

denote the composite $T_p S \xrightarrow{\Sigma_{\mathcal{R}}} HH^2(R) \rightarrow H^2(R, k)$, where $\Sigma_{\mathcal{R}}$ is as in Lemma 3.2.

3B. Cohomology classes for the coordinate algebra via deformations. For the remainder of this section, we fix \mathbb{G} a smooth (affine) algebraic group of dimension n and a positive integer r . We take $\mathfrak{g} = \text{Lie}(\mathbb{G}) = \text{Lie}(\mathbb{G}_{(s)})$ for any $s \geq 1$; in particular, $\mathfrak{g} = \text{Lie}(\mathbb{G}_{(r)})$. We shall view $\mathbb{O}(\mathbb{G}_{(r+1)})$ as a deformation of $\mathbb{O}(\mathbb{G}_{(r)})$ parametrized by $\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}$. One sees this geometrically using the pullback square

$$\begin{array}{ccc} \mathbb{G}_{(r)} & \longrightarrow & \mathbb{G}_{(r+1)} \\ \downarrow & & \downarrow \\ \text{Spec}(k) & \longrightarrow & \mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}. \end{array} \tag{3}$$

Proposition 3.4. *The extension $\mathbb{O}(\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}) \rightarrow \mathbb{O}(\mathbb{G}_{(r+1)})$ is a deformation of $\mathbb{O}(\mathbb{G}_{(r)})$ parametrized by $\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)} \cong \mathbb{G}_{(1)}^{(r)}$. We refer to this deformation of $\mathbb{O}(\mathbb{G}_{(r)})$ as \mathbb{O}_{nat} .*

Proof. The isomorphism $\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)} \cong \mathbb{G}_{(1)}^{(r)}$ is induced by the Frobenius $\mathbb{G}_{(r+1)} \rightarrow \mathbb{G}_{(r)}$, and can be found in [Jantzen 2003, Proposition I.9.5]. The fact that $\mathbb{O}(\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}) \rightarrow \mathbb{O}(\mathbb{G}_{(r+1)})$ is a deformation of $\mathbb{O}(\mathbb{G}_{(r)})$ follows easily from the diagram (3). □

Take $\mathbb{O} = \mathbb{O}(\mathbb{G}_{(r)})$. Note that $\mathfrak{g}^{(r)} = T_1\mathbb{G}_{(1)}^{(r)}$. We get from Lemma 3.2 and \mathbb{O}_{nat} a canonical linear map $\sigma_{\mathbb{O}} = \sigma_{\mathbb{O}_{\text{nat}}} : \mathfrak{g}^{(r)} \rightarrow H^2(\mathbb{O}, k)$ and induced algebra map

$$\sigma'_{\mathbb{O}} : S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(\mathbb{O}, k),$$

where $S(-)$ denotes the symmetric algebra. From the identification $H^1(\mathbb{O}, k) = T_1\mathbb{G}_{(r)} = \mathfrak{g}$, in conjunction with $\sigma'_{\mathbb{O}}$, we get yet another algebra map

$$\wedge^*(\mathfrak{g}[1]) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(\mathbb{O}, k). \tag{4}$$

In Section 4 below we will prove the following proposition.

Proposition 3.5. *The algebra map (4) is an isomorphism of $\mathbb{G}_{(r)}$ -algebras. In particular, $\sigma'_{\mathbb{O}} : S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(\mathbb{O}(\mathbb{G}_{(r)}), k)$ is an isomorphism modulo nilpotents.*

The $\mathbb{G}_{(r)}$ -action on the product $\wedge^*(\mathfrak{g}[1]) \otimes S(\mathfrak{g}^{(r)}[2])$ is induced by the adjoint action on \mathfrak{g} and the trivial action on its twist $\mathfrak{g}^{(r)}$.

Remark 3.6. We can easily establish an *abstract* algebra isomorphism between $\wedge^*(\mathfrak{g}) \otimes S(\mathfrak{g}^{(r)}[2])$ and the cohomology $H^*(\mathbb{O}, k)$ as follows. As verified in [Waterhouse 1979, Theorem 14.4], the fact that $\mathbb{G}_{(r)}$ is connected implies that there is an isomorphism $\mathbb{O} \cong k[x_1, \dots, x_n]/(x_1^{p^{e_1}}, \dots, x_n^{p^{e_n}})$ for some $n, e_1, \dots, e_n > 0$. The well-known computation of $H^*(k[x]/(x^{p^e}), k) \simeq H^*(\mathbb{Z}/p^e, k)$ and the Künneth theorem thus implies the asserted isomorphism. The significance of Proposition 3.5 is that we may use the deformation map $\sigma_{\mathbb{O}}$ to arrive at such an isomorphism. We will see below that $\sigma_{\mathbb{O}}$ admits a lift to the cohomology of the double $D\mathbb{G}_{(r)}$. The existence of such a lift is an essential point in the proof that the cohomology of the double is finitely generated.

3C. Cohomology classes for the double via deformations. Since $\mathbb{G}_{(r)}$ acts trivially on the quotient $\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}$ we see that the image of the inclusion

$$\mathbb{O}(\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}) \rightarrow \mathbb{O}(\mathbb{G}_{(r+1)}) = \mathbb{O}_{\text{nat}}$$

is in the $\mathbb{G}_{(r)}$ -invariants. Hence the induced inclusion into the smash product

$$\mathbb{O}(\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}) \rightarrow \mathbb{O}_{\text{nat}} \# k\mathbb{G}_{(r)}$$

has central image, where $\mathbb{G}_{(r)}$ acts via the adjoint action on \mathbb{O}_{nat} . Furthermore, the reduction

$$(\mathbb{O}_{\text{nat}} \# k\mathbb{G}_{(r)}) \otimes_{\mathbb{O}(\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)})} k$$

recovers the double $D\mathbb{G}_{(r)}$. Whence we have that the above smash product is a deformation of the double parametrized by $\mathbb{G}_{(1)}^{(r)} \cong \mathbb{G}_{(r+1)}/\mathbb{G}_{(r)}$. We denote this deformation $D_{\text{nat}} = \mathbb{O}_{\text{nat}} \# k\mathbb{G}_{(r)}$.

The deformation D_{nat} induces a map to cohomology

$$\sigma_{\text{D}} \equiv \sigma_{D_{\text{nat}}} : \mathfrak{g}^{(r)} \rightarrow H^2(D\mathbb{G}_{(r)}, k)$$

and subsequent graded algebra morphism

$$\sigma'_D : S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(D\mathbb{G}_{(r)}, k).$$

Proposition 3.7. *The triangle*

$$\begin{array}{ccc} & H^2(D\mathbb{G}_{(r)}, k) & \\ \sigma_D \nearrow & & \searrow \text{res} \\ \mathfrak{g}^{(r)} & \xrightarrow{\sigma_{\mathbb{O}}} & H^2(\mathbb{O}(\mathbb{G}_{(r)}), k) \end{array} \tag{5}$$

commutes.

Proof. Take $\mathbb{O} = \mathbb{O}(\mathbb{G}_{(r)})$ and $D = D\mathbb{G}_{(r)}$. The diagram (5) follows from the diagram

$$\begin{array}{ccc} \mathbb{O}_{\text{nat}} & \xrightarrow{\text{incl}} & D_{\text{nat}} \\ \downarrow & & \downarrow \\ \mathbb{O} & \xrightarrow{\text{incl}} & D, \end{array}$$

where the top map is one of $\mathbb{O}(\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)})$ -algebras and the vertical maps are given by applying $(-)\otimes_{\mathbb{O}(\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)})} k$. In particular, the commutative square implies that the maps E^D and $E^{\mathbb{O}}$ from the proof of Lemma 3.2 can be chosen in a compatible manner so that $E^D|_{\mathbb{O}\otimes\mathbb{O}} = E^{\mathbb{O}}$. Hence the resulting Hopf 2-cocycles \bar{F}^D_{ξ} and $\bar{F}^{\mathbb{O}}_{\xi}$, corresponding to an element $\xi \in \mathfrak{g}^{(r)}$, are such that

$$\text{res}(\sigma_D(\xi)) = \text{res}([\bar{F}^D_{\xi}]) = [\bar{F}^D_{\xi}|_{\mathbb{O}\otimes\mathbb{O}}] = [\bar{F}^{\mathbb{O}}_{\xi}] = \sigma_{\mathbb{O}}(\xi). \quad \square$$

Corollary 3.8. *The map $\sigma_{\mathbb{O}} : \mathfrak{g}^{(r)} \rightarrow H^2(\mathbb{O}(\mathbb{G}_{(r)}), k)$ from Section 3B has image in the invariants $H^2(\mathbb{O}(\mathbb{G}_{(r)}), k)^{\mathbb{G}_{(r)}}$.*

Proof. The restriction $H^*(D, k) \rightarrow H^*(\mathbb{O}, k)$ is induced by the cochain inclusion

$$\text{Hom}^*_D(P, k) = \text{Hom}^*_{\mathbb{O}}(P, k)^{\mathbb{G}_{(r)}} \rightarrow \text{Hom}^*_{\mathbb{O}}(P, k),$$

where P is any resolution of k over D . Hence the lifting of Proposition 3.7 implies that $\sigma_{\mathbb{O}}$ has image in the $\mathbb{G}_{(r)}$ -invariants. □

We can consider also the inflation $H^*(\mathbb{G}_{(r)}, k) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$ induced by the Hopf projection $D\mathbb{G}_{(r)} \rightarrow k\mathbb{G}_{(r)}$. This inflation, in conjunction with the algebra map σ'_D , represents contributions to the cohomology of the double coming from the two constituent factors $k\mathbb{G}_{(r)}$ and \mathbb{O} .

Definition 3.9. We let

$$\theta_r : H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$$

denote the product of the inflation from $H^*(\mathbb{G}_{(r)}, k)$ and σ'_D .

We will find in Section 5 that the map θ_r is finite. It will follow that the cohomology of the double is finitely generated.

4. Proof of Proposition 3.5

For a deformation \mathcal{R} of an algebra R parametrized by $S = \text{Spec}(A)$, we view the tangent space $T_p S$ as the first cohomology $H^1(A, k)$. (Both of which are identified with algebra maps to the dual numbers $\text{Hom}_{\text{Alg}}(A, k[\varepsilon])$.) So $\sigma_{\mathcal{R}}$ will appear as

$$\sigma_{\mathcal{R}} : H^1(A, k) \rightarrow H^2(R, k).$$

In the case $\mathbb{G} = \mathbb{G}_a$, we will see that the map $\sigma_{\mathbb{G}}$ induced by the deformation \mathbb{G}_{nat} (which we denote by \mathcal{L} in this case) behaves like the Bockstein map for the integral cohomology of a cyclic group with coefficients in \mathbb{F}_p . In particular, it picks out an algebra generator in second cohomology. From this observation we will deduce Proposition 3.5 for general smooth \mathbb{G} .

4A. Generalized (higher) Bocksteins for \mathbb{G}_a . In this subsection, we consider the special case $\mathbb{G} = \mathbb{G}_a$, the additive group (whose coordinate algebra is a polynomial algebra on one variable). Consider the deformation $\mathbb{O}(\mathbb{G}_{a(r+1)}) = k[t]/(t^{p^{r+1}})$ of $\mathbb{O}(\mathbb{G}_{a(r)}) = k[t]/(t^{p^r})$ parametrized by $\mathbb{O}(\mathbb{G}_{a(1)}^{(r)}) = (k[t]/(t^p)) \otimes_{\phi^r} k$. To ease notation take $\mathcal{Z} = \mathbb{O}(\mathbb{G}_{a(r+1)})$, $Z = \mathbb{O}(\mathbb{G}_{a(r)})$ and $Z' = \mathbb{O}(\mathbb{G}_{a(1)}^{(r)})$. The deformation \mathcal{Z} produces a map

$$\sigma_{\mathcal{Z}} : H^1(Z', k) \rightarrow H^2(Z, k).$$

We let $\alpha \in H^1(Z', k) = \text{Hom}_{\text{Alg}}(Z', k[\varepsilon])$ denote the class given by the projection $Z' \rightarrow k[\varepsilon]$, $t \mapsto \varepsilon$.

Definition 4.1. Take $\beta \equiv \sigma_{\mathcal{Z}}(\alpha) \in H^2(Z, k)$. We say that β is the (higher order) Bockstein of the class $\alpha \in H^1(Z', k)$.

Recall that for $i \geq 0$ and $q > 1$ the cohomology $H^i(k[t]/(t^q), k)$ is 1 dimensional. (One can see this directly from the minimal, periodic, resolution of k .) Hence $H^1(Z', k)$ and $H^2(Z, k)$ are 1 dimensional.

Lemma 4.2. *The map $\sigma_{\mathcal{Z}} : H^1(Z', k) \rightarrow H^2(Z, k)$ is a linear isomorphism. In particular, β is nonzero.*

Proof. It suffices to show that the image β of $\alpha \in H^1(Z', k)$ is nonzero. Consider the base change $\mathcal{Z} \otimes_{Z'} k[\varepsilon]$ via α , and the $k[\varepsilon]$ -linear identification $\mathcal{Z} \otimes_{Z'} k[\varepsilon] \cong Z[\varepsilon]$ given by

$$Z[\varepsilon] \xrightarrow{\sim} \mathcal{Z} \otimes_{Z'} k[\varepsilon], \quad t^i \mapsto t^i \otimes 1, t^i \varepsilon \mapsto t^i \otimes \varepsilon.$$

This induces a multiplication $z \cdot_{\alpha} z' = zz' + F_{\alpha}(z, z')\varepsilon$ on $Z[\varepsilon]$, where F_{α} is a Hochschild 2-cocycle. We have then $[\mathcal{Z} \otimes_{Z'} k[\varepsilon]] = [F_{\alpha}] \in HH^2(Z)$ and the corresponding Hopf cohomology class is $[\bar{F}_{\alpha}] \in H^2(Z, k)$, where \bar{F}_{α} is the composite of F_{α} with the counit $\bar{F}_{\alpha} = \epsilon F_{\alpha}$.

We want to show that $\beta = \sigma_{\mathcal{Z}}(\alpha) = [\bar{F}_{\alpha}]$ is nonzero (i.e., that \bar{F}_{α} is not a coboundary). One sees directly that $\bar{F}_{\alpha}(t^l, t^m) = \delta_{l+m, p^r}$, and in particular $\bar{F}_{\alpha}(t^i, t^{p^r-i}) = 1$. One also sees that the differential of any degree 1 function $f \in \text{Hom}_k(Z, k)$ in the Hopf cochain complex for Z is such that

$$d(f)(t^i, t^{p^r-i}) = \pm f(t^{p^r}) = \pm f(0) = 0.$$

Therefore \bar{F}_{α} cannot be a coboundary, and the cohomology class $\beta = [\bar{F}_{\alpha}]$ is nonzero. □

We can consider now the n -th tensor product $\mathcal{Z}^{\otimes n}$ as a deformation of $Z^{\otimes n}$, parametrized by $\text{Spec}((Z')^{\otimes n})$. We let \mathfrak{g}_a denote the Lie algebra of \mathbb{G}_a so that

$$(\mathfrak{g}_a^{(r)})^n = H^1((Z')^{\otimes n}, k) = \text{Hom}_{\text{Alg}}((Z')^{\otimes n}, k[\varepsilon]),$$

with each element $\sum_{i=1}^n c_i \alpha_i \in (\mathfrak{g}_a^{(r)})^n$ corresponding to the algebra map

$$\sum_i c_i \alpha_i : (Z')^{\otimes n} \rightarrow k[\varepsilon], \quad t_i \mapsto c_i \varepsilon.$$

Here α_i is the basis vector for the i -th copy of $\mathfrak{g}_a^{(r)}$, defined as above, and t_i is the generator of the i -th factor in $(Z')^{\otimes n}$.

Proposition 4.3. *The map $\sigma_{\mathcal{Z}^{\otimes n}} : (\mathfrak{g}_a^{(r)})^n \rightarrow H^2(Z^{\otimes n}, k)$ induces an injective graded k -algebra map*

$$\sigma'_{\mathcal{Z}^{\otimes n}} : S((\mathfrak{g}_a^{(r)})^n[2]) \rightarrow H^*(Z^{\otimes n}, k)$$

which is an isomorphism modulo nilpotents.

Proof. We claim that the reduction $\sigma_{\text{red}} : (\mathfrak{g}_a^{(r)})^n \rightarrow H^2(Z^{\otimes n}, k)_{\text{red}}$ is injective. (Here by $H^2(Z^{\otimes n}, k)_{\text{red}}$ we mean the degree 2 portion of the reduced algebra and by σ_{red} we mean the composite of $\sigma_{\mathcal{Z}^{\otimes n}}$ with the reduction.) It suffices to show that for any nonzero $\underline{c} = \sum_i c_i \alpha_i$ there is an index j such that restriction along the factor $Z_j \rightarrow Z^{\otimes n}$ produces a nonzero element in the cohomology $H^*(Z_j, k)_{\text{red}}$, via the composite

$$(\mathfrak{g}_a^{(r)})^n \xrightarrow{\sigma} H^*(Z^{\otimes n}, k)_{\text{red}} \rightarrow H^*(Z_j, k)_{\text{red}} \cong k[\beta_j].$$

For any such $\underline{c} \in (\mathfrak{g}_a^{(r)})^n$ let $\text{Def}_{\underline{c}}$ denote the corresponding deformation $\mathcal{Z}^{\otimes n} \otimes_{(Z')^{\otimes n}} k[\varepsilon]$, where we change base along the corresponding map $(Z')^{\otimes n} \rightarrow k[\varepsilon]$, $t_i \mapsto c_i \varepsilon$.

Consider such a nonzero \underline{c} and take j such that the j -th entry c_j is nonzero. We claim that the image of the corresponding class $\sigma_{\mathcal{Z}^{\otimes n}}(\underline{c}) \in H^2(Z^{\otimes n}, k)$ in $H^2(Z_j, k)$ is exactly the class $c_j \beta_j \in H^2(Z_j, k)$. One way to see this is to note that the Hochschild 2-cocycle corresponding to $\text{Def}_{\underline{c}}$ is a function $F_{\underline{c}} : Z^{\otimes n} \otimes Z^{\otimes n} \rightarrow Z^{\otimes n}$ with restriction

$$F_{\underline{c}} : Z_j \otimes Z_j \rightarrow Z^{\otimes n}, \quad t_j^l \otimes t_j^m \mapsto c_j t_j^{(l+m)-p^r},$$

where a negative power is considered to be 0. (This is just as in Lemma 4.2.) Composing with the counit produces the function $\bar{F}_{c_j} : t^l \otimes t^m \mapsto c_j \delta_{l+m, p^r}$. The function \bar{F}_{c_j} is equal to $c_j \bar{F}_{\alpha_j}$, where \bar{F}_{α_j} is as in the proof of Lemma 4.2, and we can consult the proof of Lemma 4.2 again to see that $[\bar{F}_{c_j}] = c_j [\bar{F}_{\alpha_j}] = c_j \beta_j$.

Upon choosing coordinates of $Z^{\otimes n}$ to obtain the identification

$$H^*(Z^{\otimes n}, k)_{\text{red}} \cong (\otimes_{i=1}^n H^*(Z_i, k))_{\text{red}} \cong k[\beta_1, \dots, \beta_n],$$

we easily see that the reduced algebra has dimension n in degree 2. So injectivity of σ_{red} implies that σ_{red} is an isomorphism. Consequently, the algebra map σ'_{red} (the multiplicative extension of σ_{red}) is an isomorphism. As a consequence, $\sigma'_{\mathcal{Z}^{\otimes n}}$ must be injective as well. □

4B. The proof of Proposition 3.5. We retain our notations \mathcal{Z} and Z from above, and take also $\mathbb{O} = \mathbb{O}(\mathbb{G}_{(r)})$.

Proof of Proposition 3.5. The identification of $H^1(\mathbb{O}, k)$ with $\text{Hom}_k(m_G/m_G^2, k)$ implies that $\mathfrak{g} = H^1(\mathbb{O}, k)$. Invariance of the image of $\mathfrak{g}^{(r)}$ follows from Corollary 3.8. Whence the algebra map $\wedge^*(\mathfrak{g}[1]) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(\mathbb{O}, k)$ of (4) is one of $\mathbb{G}_{(r)}$ -algebras. It remains to show that the map is a (linear) isomorphism.

Since \mathbb{G} is smooth, we can choose complete local coordinates $\{x_i\}_i$ at the identity to get algebra presentations

$$\mathbb{O}_{\text{nat}} = \mathbb{O}(\mathbb{G}_{(r+1)}) = k[x_1, \dots, x_n]/(x_i^{p^{r+1}}) \quad \text{and} \quad \mathbb{O} = k[x_1, \dots, x_n]/(x_i^{p^r}).$$

Whence we have an algebra isomorphism $Z^{\otimes n} \xrightarrow{\sim} \mathbb{O}$, $t_i \mapsto x_i$, under which the deformations $\mathcal{Z}^{\otimes n}$ and \mathbb{O}_{nat} can be identified. Thus the maps $\sigma_{\mathcal{Z}^{\otimes n}}$ and $\sigma_{\mathbb{O}}$ are also identified, and we see that $\sigma'_0 : S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(\mathbb{O}, k)$ is an isomorphism modulo nilpotents by Proposition 4.3.

Since we know abstractly that

$$H^*(\mathbb{O}, k) = \wedge^*(H^1(\mathbb{O}, k)) \otimes S(V) = \wedge^*(\mathfrak{g}) \otimes S(V),$$

for any vector space complement V to $\wedge^2 \mathfrak{g}$ in $H^2(\mathbb{O}, k)$, it suffices to show that $\sigma_{\mathbb{O}}(\mathfrak{g}^{(r)})$ is a complement to the second wedge power of \mathfrak{g} . However, this follows from the facts that σ'_0 is an isomorphism modulo nilpotents and that the kernel of the reduction $H^2(\mathbb{O}, k) \rightarrow H^2(\mathbb{O}, k)_{\text{red}}$ is exactly $\wedge^2 H^1(\mathbb{O}, k) = \wedge^2 \mathfrak{g}$. \square

4C. In characteristic 2. Suppose $\text{char}(k) = 2$ and let \mathbb{G} be a smooth algebraic group over k . Consider the r -th Frobenius kernel $\mathbb{G}_{(r)}$ with $r > 1$. In this case we have an algebra identification

$$\mathbb{O}(\mathbb{G}_{(r)}) = k[x_1, \dots, x_n]/(x_1^{2^r}, \dots, x_n^{2^r}) = \otimes_{i=1}^n k[x_i]/(x_i^{2^r}).$$

Furthermore, since $H^*(k[x]/(x^{2^r}), k) = k[a, b]/(a^2)$, where $\deg(a) = 1$ and $\deg(b) = 2$, we see that all elements in $H^1(\mathbb{O}(\mathbb{G}_{(r)}), k)$ are square zero. Hence we can construct an algebra map

$$\wedge^*(\mathfrak{g}[1]) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(\mathbb{O}(\mathbb{G}_{(r)}), k) \tag{6}$$

via the identification $\mathfrak{g} = H^1(\mathbb{O}(\mathbb{G}_{(r)}), k)$ and the deformation map $\sigma_{\mathbb{O}}$, just as before. The above proof of Proposition 3.5 can now be repeated verbatim to arrive at

Proposition 4.4. *When $\text{char}(k) = 2$ and $r > 1$, the algebra map (6) is an isomorphism of $\mathbb{G}_{(r)}$ -algebras.*

Under these same hypotheses all proofs in Sections 5–7 also apply verbatim. Hence we are able to deal with these cases without any deviation in our presentation.

Remark 4.5. When $\text{char}(k) = 2$ and $r = 1$, the algebra map $S(\mathfrak{g}[1]) \rightarrow H^*(\mathbb{O}, k)$ induced by the identification $\mathfrak{g} = H^1(\mathbb{O}, k)$ is an isomorphism. The methods employed in the proof of Proposition 3.5 show that, in this case,

$$\sigma'_0 : S(\mathfrak{g}^{(1)}[2]) \rightarrow H^*(\mathbb{O}, k)$$

is the Frobenius.

Now, in degree 2 we have an exact sequence of $\mathbb{G}_{(1)}$ -representations $0 \rightarrow \mathfrak{g}^{(1)} \rightarrow S^2(\mathfrak{g}) = H^2(\mathbb{C}, k) \rightarrow M \rightarrow 0$, where $M = \text{coker}(\sigma_{\mathbb{C}})$. The possible failure of this sequence to split over $\mathbb{G}_{(1)}$ obstructs our proof of Theorem 5.3 below. In particular, it is not apparent how one can construct the complement Γ to $S(\mathfrak{g}^{(1)}[2])$ employed in the proof of the aforementioned theorem.

5. Finite generation of cohomology

We consider a smooth algebraic group \mathbb{G} and an integer $r > 0$. As always, \mathbb{G} is assumed to be affine of finite type over k . In Theorems 5.3 and 5.6 below, we prove finite generation of cohomology for the Drinfeld double $D\mathbb{G}_{(r)} \equiv \mathbb{C}(\mathbb{G}_{(r)}) \# k\mathbb{G}_{(r)}$ of the r -th Frobenius kernel $\mathbb{G}_{(r)}$. Our technique is to use the Grothendieck spectral sequence [Grothendieck 1957] as in [Friedlander and Suslin 1997].

5A. A spectral sequence for the cohomology of the double. We begin with a general result.

Proposition 5.1. *Let $F : \mathcal{A} \rightarrow \mathcal{B}$ and $G : \mathcal{B} \rightarrow \mathcal{C}$ be additive, left exact functors between abelian categories with enough injectives and suppose that F sends injective objects of \mathcal{A} to injective objects of \mathcal{B} . Assume further that \mathcal{A}, \mathcal{B} and \mathcal{C} have tensor products and that F and G are equipped with natural maps $F(V) \otimes F(V') \rightarrow F(V \otimes V')$ and $G(W) \otimes G(W') \rightarrow G(W \otimes W')$. Then for any pairing $V \otimes V' \rightarrow V''$ there exists a pairing of Grothendieck spectral sequence*

$$\{R^s G(R^t(F(V)))\} \Rightarrow R^{s+t}(G \circ F)(V) \otimes \{R^{s'} G(R^{t'}(F(V')))\} \Rightarrow R^{s'+t'}(G \circ F)(V') \\ \rightarrow \{R^{s''} G(R^{t''}(F(V'')))\} \Rightarrow R^{s''+t''}(G \circ F)(V'').$$

Proof. The Grothendieck spectral sequence for the composition of left exact functors between abelian categories with enough injectives, $G \circ F : \mathcal{A} \rightarrow \mathcal{C}$, arises from a Massey exact couple. Namely, one takes an injective resolution $V \rightarrow I^*$ of an object of V of \mathcal{A} , and then takes a Cartan–Eilenberg resolution $F(I^*) \rightarrow J^{*,*}$ of the cochain complex $F(I^*)$; $J^{*,*}$ is a double complex of injective objects of \mathcal{B} which not only gives an injective resolution of each $F(I^n)$ but also of each $H^n(F(I^*))$. Then the Massey exact couple is given by “triples” $(i : D \rightarrow D, j : D \rightarrow E, k : E \rightarrow D)$,

$$\dots \xrightarrow{k} D = \bigoplus_p H^{p+q}(F^{p+1}(\text{Tot}(G(J^{*,*})))) \xrightarrow{i} \bigoplus_p D = H^{p+q}(F^p(\text{Tot}(G(J^{*,*})))) \\ \xrightarrow{j} E = \bigoplus_{p,q} H^{p+q}(F^p(\text{Tot}(G(J^{*,*}))) / F^{p+1}(\text{Tot}(G(J^{*,*})))) \xrightarrow{k} \dots,$$

where $F^p(\text{Tot}(G(J^{*,*}))) = \text{Tot}(G(\bigoplus_{i \geq p} (J^{i,*})))$.

Assuming that \mathcal{A}, \mathcal{B} and \mathcal{C} have tensor products, a pairing of objects in \mathcal{A} gives rise to a pairing of Massey exact couples. Namely, given injective resolutions $V \rightarrow I^*, V' \rightarrow I'^*, V'' \rightarrow I''^*$ and a pairing $V \otimes V' \rightarrow V''$, then the usual extension argument for the injective complex I''^* tells us that there is a map of cochain complexes $\text{Tot}(I^* \otimes I'^*) \rightarrow I''^*$, unique up to chain homotopy, extending this pairing. This, in turn, determines a pairing of bicomplexes $G(J^{*,*}) \otimes G(J'^{*,*}) \rightarrow G(J''^{*,*})$ and thus of

filtered total complexes. The pairing on exact couples takes the expected form using the natural map $\bigoplus_{s+t=n}(H^s(C^*) \otimes H^t(C'^*)) \rightarrow H^{s+t}(C^* \otimes C'^*)$.

Massey [1954] gives sufficient conditions for a pairing of exact couples to determine a pairing of spectral sequences (see also [Friedlander and Suslin 1997]). The essential condition is Massey’s condition μ_n for each $n \geq 0$: for $z \otimes z'$ bihomogeneous in $E \otimes E'$ and any $x \otimes x'$ bihomogeneous in $D \otimes D'$ such that $k(z) = i^n(x)$, $k(z') = (i')^n(x')$, there exists $x'' \in D''$ with $k''(z \cdot z') = (i'')^n(x'')$ and $j''(x'') = j(x) \cdot z' + (-1)^{\deg(z)} z \cdot j'(x')$. In our context, $z \otimes z' \in H^{p+q}(F^p/F^{p+1}) \otimes H^{p'+q'}(F^{p'}/F^{p'+1})$ and $x \otimes x' \in H^{p+q+1}(F^{p+n+1}) \otimes H^{p'+q'+1}(F^{p'+n+1})$. To satisfy condition μ_n , we take $x'' \in H^{p+q+p'+q'}(F^{p+p'+n+1})$ to be the image of $x \otimes x'$ given by the pairing map. □

Recall that $D\mathbb{G}_{(r)}/\mathcal{O}(\mathbb{G}_{(r)})$ is $k\mathbb{G}_{(r)}$ -Galois as in [Montgomery 1993]. One can view this property as the condition that $D\mathbb{G}_{(r)}$ is a $k\mathbb{G}_{(r)}$ torsor (in the context of Hopf algebras) over $\mathcal{O}(\mathbb{G}_{(r)})$: there is a natural bijection $D\mathbb{G}_{(r)} \otimes_{\mathcal{O}(\mathbb{G}_{(r)})} D\mathbb{G}_{(r)} \rightarrow D\mathbb{G}_{(r)} \otimes k\mathbb{G}_{(r)}$. By normality of $\mathcal{O}(\mathbb{G}_{(r)})$ in $D\mathbb{G}_{(r)}$, for any $D\mathbb{G}_{(r)}$ -module V on which $\mathcal{O}(\mathbb{G}_{(r)})$ acts trivially we have $V^{D\mathbb{G}_{(r)}} = V^{\mathbb{G}_{(r)}}$. Hence the invariants functor for $D\mathbb{G}_{(r)}$ factors

$$\text{Hom}_{D\mathbb{G}_{(r)}}(k, -) = \text{Hom}_{k\mathbb{G}_{(r)}}(k, -) \circ \text{Hom}_{\mathcal{O}(\mathbb{G}_{(r)})}(k, -) : \text{rep}(D\mathbb{G}_{(r)}) \rightarrow \text{Vect}.$$

Proposition 5.2. *The above composition of functors leads to a Grothendieck spectral sequence of k -algebras*

$$E_2^{s,t}(k) = H^s(\mathbb{G}_{(r)}, H^t(\mathcal{O}(\mathbb{G}_{(r)}), k)) \Rightarrow H^{s+t}(D\mathbb{G}_{(r)}, k). \tag{7}$$

For any $D\mathbb{G}_{(r)}$ -module M , the above composition of functors leads to Grothendieck spectral sequence

$$E_2^{s,t}(M) = H^s(\mathbb{G}_{(r)}, H^t(\mathcal{O}(\mathbb{G}_{(r)}), M)) \Rightarrow H^{s+t}(D\mathbb{G}_{(r)}, M), \tag{8}$$

which is a spectral sequence of modules over (7).

Proof. The equalities

$$\text{Hom}_{\mathcal{O}(\mathbb{G}_{(r)})}(k, (D\mathbb{G}_{(r)})^*) = \text{Hom}_{\mathcal{O}(\mathbb{G}_{(r)})}(D\mathbb{G}_{(r)}, k) = \text{Hom}_k(k\mathbb{G}_{(r)}, k) = (k\mathbb{G}_{(r)})^*$$

imply $\text{Hom}_{\mathcal{O}(\mathbb{G}_{(r)})}(k, (D\mathbb{G}_{(r)})^*)$ is projective as well as injective as a $k\mathbb{G}_{(r)}$ -module (because a $k\mathbb{G}_{(r)}$ module is projective if and only if it is injective [Jantzen 2003; Montgomery 1993]). Since

$$\text{Hom}_{\mathcal{O}(\mathbb{G}_{(r)})}(k, (D\mathbb{G}_{(r)})^*) = (k\mathbb{G}_{(r)})^*,$$

we conclude that $\text{Hom}_{\mathcal{O}(\mathbb{G}_{(r)})}(k, -)$ sends injective $D\mathbb{G}_{(r)}$ -modules to injective $k\mathbb{G}_{(r)}$ -modules. Consequently, Grothendieck’s construction of the spectral sequence for a composition of left exact functors applies to the composition $\text{Hom}_{k\mathbb{G}_{(r)}}(k, -) \circ \text{Hom}_{\mathcal{O}(\mathbb{G}_{(r)})}(k, -)$, and this spectral sequence takes the form (7) when applied to k and the form (8) when applied to M .

The algebra structure on (7) and the module structure on (8) follow from the multiplicative structure established in Proposition 5.1 in view of the pairing $k \otimes k \rightarrow k$ (multiplication of k) and $k \otimes M \rightarrow M$ (pairing with the trivial module) in the category $\text{rep}(D\mathbb{G}_{(r)})$. □

5B. Finite generation. We can now prove that $H^*(D\mathbb{G}_{(r)}, k)$ is a finitely generated algebra. This will be followed by Theorem 5.6, establishing our general finite generation theorem. Recall that an algebra map $A \rightarrow B$ is called finite if B is a finite module over A . Recall also the map θ_r of Definition 3.9.

Theorem 5.3. *Let \mathbb{G} be a smooth algebraic group over a field k of positive characteristic, let $r > 0$ be a positive integer, and let $D\mathbb{G}_{(r)} \equiv \mathbb{O}(\mathbb{G}_{(r)})\#k\mathbb{G}_{(r)}$ denote the Drinfeld double of the r -th Frobenius kernel of \mathbb{G} .*

Then the graded k -algebra map

$$\theta_r : H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$$

is finite.

Consequently:

- $H^*(D\mathbb{G}_{(r)}, k)$ is a finitely generated k -algebra.
- $H^*(D\mathbb{G}_{(r)}, M)$ is a finite $H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2])$ -module and hence a finite $H^*(D\mathbb{G}_{(r)}, k)$ -module for any finite dimensional $D\mathbb{G}_{(r)}$ -module M whose restriction to $\mathbb{O}(\mathbb{G}_{(r)})$ has trivial action.

Proof. Take $\mathbb{O} = \mathbb{O}(\mathbb{G}_{(r)})$ and $C^* = H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2])$. This proof is an adaption of the proof of Theorem 1.1 of [Friedlander and Suslin 1997]. Let $\{E_r^{s,t}, r \geq 2\}$ denote the spectral sequence $\{E_r^{s,t}(k), r \geq 2\}$ of Proposition 5.2.

Observe that $H^*(\mathbb{O}, k) = S(\mathfrak{g}^{(r)}[2]) \otimes \Gamma$; here, $S(\mathfrak{g}^{(r)}[2])$ has trivial $\mathbb{G}_{(r)}$ -action and $\Gamma \equiv \wedge^*(H^1(\mathbb{O}, k))$ is finite dimensional. Thus, $E_2^{*,*} = H^*(\mathbb{G}_{(r)}, H^*(\mathbb{O}, k))$ equals $H^*(\mathbb{G}_{(r)}, \Gamma) \otimes S(\mathfrak{g}^{(r)}[2])$, since $M \mapsto H^0(\mathbb{G}_{(r)}, M \otimes V)$ is the composite of $H^0(\mathbb{G}_{(r)}, -)$ and the exact functor $- \otimes V$ for any trivial $\mathbb{G}_{(r)}$ -module V . We equip $H^*(\mathbb{G}_{(r)}, H^*(\mathbb{O}, k)) = H^*(\mathbb{G}_{(r)}, \Gamma) \otimes S(\mathfrak{g}^{(r)}[2])$ with the “external tensor product module structure” for the algebra $C^* = H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2])$.

By Theorem 1.1 of [Friedlander and Suslin 1997], $H^*(\mathbb{G}_{(r)}, \Gamma)$ is a finite $H^*(\mathbb{G}_{(r)}, k)$ -module. It follows that $H^*(\mathbb{G}_{(r)}, H^*(\mathbb{O}, k))$ is a finite C^* -module. We identify this C^* -module structure on $H^*(\mathbb{G}_{(r)}, H^*(\mathbb{O}, k))$ as that given by the coproduct $\phi \otimes \psi$ of two maps ϕ and ψ associated to the spectral sequence: The first is the map

$$\phi : S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(D\mathbb{G}_{(r)}, k) = E_\infty^* \rightarrow E_\infty^{0,*} \subset E_2^{0,*} \subset E_2^{*,*}$$

given by Proposition 3.7. The second is the natural map

$$H^{\text{ev}}(\mathbb{G}_{(r)}, k) \subset E_2^{*,0} \subset E_2^{*,*}.$$

We have thus verified the hypotheses of Lemma 1.6 of [Friedlander and Suslin 1997], enabling us to conclude that $H^*(D\mathbb{G}_{(r)}, k)$ is a finite module over the finitely generated algebra C^* and thus is itself finitely generated.

Now, we consider a finite dimensional $D\mathbb{G}_{(r)}$ -module M whose restriction to \mathbb{O} has trivial action. Then $E_2^{*,*}(M) = H^*(\mathbb{G}_{(r)}, H^*(\mathbb{O}, M))$ equals $H^*(\mathbb{G}_{(r)}, \Gamma \otimes M) \otimes S(\mathfrak{g}^{(r)}[2])$ which is a finite C^* -module by another application of Theorem 1.1 of [Friedlander and Suslin 1997] (this time, for the finite dimensional $k\mathbb{G}_{(r)}$ -module $\Gamma \otimes M$). Since $\{E_r^{*,*}(M)\}$ is a module over $\{E_r^{*,*}\}$ by Proposition 5.1, Lemma 1.6 of

[Friedlander and Suslin 1997] applies once again to imply that $H^*(D\mathbb{G}_{(r)}, M)$ is finite as a C^* -module and thus as a $H^*(D\mathbb{G}_{(r)}, k)$ -module. □

Recall our notation $|A| \equiv \text{Spec } H^{\text{ev}}(A, k)_{\text{red}}$ from the introduction.

Corollary 5.4. *We have the inequality*

$$\dim |D\mathbb{G}_{(r)}| \leq \dim |k\mathbb{G}_{(r)}| + \dim \mathbb{G}.$$

Proof. Since θ_r is finite, by Theorem 5.3, the induced map on affine spectra

$$|D\mathbb{G}_{(r)}| \rightarrow \text{Spec}(H^{\text{ev}}(\mathbb{G}_{(r)}, k)_{\text{red}} \otimes S(\mathfrak{g}^{(r)}[2])) \cong |k\mathbb{G}_r| \times \mathbb{A}^d$$

has finite fibers, where $d = \dim(\mathbb{G})$. □

Proposition 5.5. *Let G be an infinitesimal group scheme. If V is a simple module for DG , then V restricts to a trivial $\mathbb{O}(G)$ -module.*

Proof. The maximal ideal m in $\mathbb{O}(G)$ is nilpotent and is normalized by the action of G on $\mathbb{O}(G)$. Hence, the ideal $kG \cdot m \subset DG$ is also nilpotent, and therefore contained in the Jacobson radical of DG . We conclude that restricting along the projection $DG \rightarrow DG/(kG \cdot m) = kG$ determines a bijection $\text{Irrep}(kG) \rightarrow \text{Irrep}(DG)$. □

In the following theorem, we implicitly use the following fact for any Noetherian k -algebra C : a C -module M is Noetherian if and only if it is finitely generated (as a C -module).

Theorem 5.6. *As in Theorem 5.3, let \mathbb{G} be a smooth algebraic group over a field k of positive characteristic, let $r > 0$ be a positive integer, and let $D\mathbb{G}_{(r)} \equiv \mathbb{O}(\mathbb{G}_{(r)})\#k\mathbb{G}_{(r)}$ denote the Drinfeld double of the r -th Frobenius kernel of \mathbb{G} .*

If M is a finite dimensional $D\mathbb{G}_{(r)}$ -module, then $H^(D\mathbb{G}_{(r)}, M)$ is finitely generated as a $H^*(D\mathbb{G}_{(r)}, k)$ -module.*

Proof. By Theorem 5.3 and Proposition 5.5, $H^*(D\mathbb{G}_{(r)}, M)$ is finitely generated over $H^*(D\mathbb{G}_{(r)}, k)$ whenever M is an irreducible $D\mathbb{G}_{(r)}$ -module. More generally, we proceed by induction on the length of a composition series for M as a $D\mathbb{G}_{(r)}$ -module. Consider a short exact sequence $0 \rightarrow N \rightarrow M \rightarrow Q \rightarrow 1$ of finite dimensional $D\mathbb{G}_{(r)}$ -modules with N irreducible and assume our induction hypothesis applies to Q . Let $V \subset H^*(D\mathbb{G}_{(r)}, M)$ denote the image of $H^*(D\mathbb{G}_{(r)}, N)$ and let $W \subset H^*(D\mathbb{G}_{(r)}, Q)$ denote the image of $H^*(D\mathbb{G}_{(r)}, M)$. Since $H^*(D\mathbb{G}_{(r)}, k)$ is Noetherian, V is a Noetherian $H^*(D\mathbb{G}_{(r)}, k)$ -module since it is a quotient of the Noetherian $H^*(D\mathbb{G}_{(r)}, k)$ -module $H^*(D\mathbb{G}_{(r)}, N)$; moreover, W is a Noetherian $H^*(D\mathbb{G}_{(r)}, k)$ -module since it is a submodule of the Noetherian $H^*(D\mathbb{G}_{(r)}, k)$ -module $H^*(D\mathbb{G}_{(r)}, Q)$. Granted the short exact sequence $0 \rightarrow V \rightarrow H^*(D\mathbb{G}_{(r)}, M) \rightarrow W \rightarrow 0$ of $H^*(D\mathbb{G}_{(r)}, k)$ -modules, we conclude that $H^*(D\mathbb{G}_{(r)}, M)$ is also Noetherian as a $H^*(D\mathbb{G}_{(r)}, k)$ -module. □

5C. Cohomology of relative doubles. Given an inclusion of finite dimensional Hopf algebras $A \rightarrow B$, we can form the relative double $D(B, A)$, which is the vector space $B^* \otimes A$ along with multiplication given by the same formula as for the standard double. Rather, we give $D(B, A)$ the unique Hopf structure so that the vector space inclusion $D(B, A) \rightarrow D(B)$ is a map of Hopf algebras. The relative double can be of technical importance, especially in tensor categorical settings (see for example [Gelaki et al. 2009; Etingof et al. 2011]).

For a closed subgroup $G \rightarrow \mathbb{G}_{(r)}$ we write $D(\mathbb{G}_{(r)}, G)$ for the relative double

$$D(\mathbb{G}_{(r)}, G) = D(k\mathbb{G}_{(r)}, kG) = \mathbb{O}(\mathbb{G}_{(r)}) \# kG,$$

where the smash product is taken relative to the adjoint action of G on $\mathbb{O}(\mathbb{G}_{(r)})$.

Dually, for a quotient $B \rightarrow C$ of finite dimensional Hopf algebras we define the relative double $D(C, B)$ as the vector space $C^* \otimes B$ along with the unique Hopf structure so that the inclusion $D(C, B) \rightarrow D(B^*)$ is a map of Hopf algebras. For a group scheme quotient $\mathbb{G}_{(r)} \rightarrow G'$ we write

$$D(G', \mathbb{G}_{(r)}) = \mathbb{O}(G') \# k\mathbb{G}_{(r)}. \tag{9}$$

From [Radford 1993, (11)–(12)], we see that $D(G', \mathbb{G}_{(r)})$ is identified with the relative double $D(kG', k\mathbb{G}_{(r)})$.

Theorem 5.7. *Let \mathbb{G} be a smooth algebraic group. Consider an arbitrary closed subgroup scheme G in $\mathbb{G}_{(r)}$, and the relative double $D(\mathbb{G}_{(r)}, G)$. Then:*

- *The cohomology $H^*(D(\mathbb{G}_{(r)}, G), k)$ is a finitely generated algebra.*
- *If M is a finite dimensional $D(\mathbb{G}_{(r)}, G)$ -module, then $H^*(D(\mathbb{G}_{(r)}, G), M)$ is a finitely generated module over $H^*(D(\mathbb{G}_{(r)}, G), k)$.*

The same finite generation results hold for the relative doubles $D(\mathbb{G}_{(r)}/\mathbb{G}_{(s)}, \mathbb{G}_{(r)})$, for $s \leq r$.

Sketch proof. Consider a closed subgroup $G \rightarrow \mathbb{G}_{(r)}$. We have the sequence $\mathbb{O}(\mathbb{G}_{(r)}) \rightarrow D(\mathbb{G}_{(r)}, G) \rightarrow kG$, from which we derive Grothendieck spectral sequences as in (7) and (8). We need to exhibit a finitely generated algebra of permanent cocycles in the E_2 -page of the spectral sequence

$$E_2^{s,t}(k) = H^s(G, H^t(\mathbb{O}(\mathbb{G}_{(r)}), k)) \Rightarrow H^{s+t}(D(\mathbb{G}_{(r)}, G), k)$$

over which $E_2^{*,*}$ is a finite module. Just as in the proof of Theorem 5.3, it suffices to show that the image of the embedding $\sigma_{\mathbb{O}} : \mathfrak{g}^{(r)} \rightarrow H^2(\mathbb{O}(\mathbb{G}_{(r)}), k)$ from Section 3B consists entirely of permanent cocycles in $E_2^{*,*}$. The deformation $D_{\text{nat}} = D(\mathbb{G}_{(r+1)}, G)$ provides a lifting $\sigma_D : \mathfrak{g}^{(r)} \rightarrow H^2(D(\mathbb{G}_{(r)}, G), k)$ of $\sigma_{\mathbb{O}}$, which verifies permanence of the cocycles $\mathfrak{g}^{(r)} \subset H^2(\mathbb{O}(\mathbb{G}_{(r)}), k)$. We can now argue as in the proof of Theorem 5.3 to establish finite generation.

In the case of a quotient $\mathbb{G}_{(r)}/\mathbb{G}_{(s)} \cong \mathbb{G}_{(r-s)}^{(s)}$, we have the deformation $\mathbb{O}_{\text{nat}} = \mathbb{O}(\mathbb{G}_{(r+1)}/\mathbb{G}_{(s)})$ of $\mathbb{O}(\mathbb{G}_{(r)}/\mathbb{G}_{(s)})$ and the deformation $D_{\text{nat}} = \mathbb{O}_{\text{nat}} \# k\mathbb{G}_{(r)}$ of the relative double $D(\mathbb{G}_{(r)}/\mathbb{G}_{(s)}, \mathbb{G}_{(r)})$. These deformations provide an inclusion

$$\sigma_{\mathbb{O}} : \mathfrak{g}^{(r)} \rightarrow H^2(\mathbb{O}(\mathbb{G}_{(r)}/\mathbb{G}_{(s)}), k)$$

and a lifting $\sigma_D : \mathfrak{g}^{(r)} \rightarrow H^2(\mathbf{D}(\mathbb{G}_{(r)}/\mathbb{G}_{(s)}, \mathbb{G}_{(r)}), k)$ of $\sigma_{\mathbb{G}}$. We employ $\sigma_{\mathbb{G}}$ and σ_D , and again argue as in Theorem 5.3, to establish finite generation. \square

Remark 5.8. For a general quotient $p : \mathbb{G}_{(r)} \rightarrow G'$, we expect that finite generation of cohomology for the relative double $\mathbf{D}(G', \mathbb{G}_{(r)})$ can be proved via the same deformation theoretic approach as above. If we take $K = \ker(p)$, the necessary deformation in this case should be provided by the quotient scheme $\mathbb{G}_{(r+1)}/K$. Some care needs to be taken, however, in dealing with the arbitrary nature of the subgroup K .

Remark 5.9. In the notation of [Gelaki et al. 2009, §2B], the relative double $\mathbf{D}(\mathbb{G}_{(r)}, G)$ has representation category isomorphic to the relative center $Z_{\mathcal{C}}(\mathcal{M})$ where $\mathcal{C} = \text{rep}(\mathbb{G}_{(r)})$, $\mathcal{M} = \text{rep}(G)$, and the \mathcal{C} -action on \mathcal{M} is given by the restriction functor $\text{rep}(\mathbb{G}_{(r)}) \rightarrow \text{rep}(G)$. Similarly, for a quotient $\mathbb{G}_{(r)} \rightarrow G'$, we have $\mathbf{D}(G', \mathbb{G}_{(r)}) \cong Z_{\mathcal{D}}(\mathcal{N})$ where $\mathcal{D} = \text{corep}(k\mathbb{G}_{(r)})$ and $\mathcal{N} = \text{corep}(kG')$.

In the final two sections of this paper we provide analyses of the spectrum of cohomology and support for the (usual) double $\mathbf{D}\mathbb{G}_{(r)}$. These analyses are valid for the relative doubles $\mathbf{D}(\mathbb{G}_{(r)}, G)$ as well. In particular, one replaces $k\mathbb{G}_{(r)}$ with kG and repeats the arguments verbatim. As we would like to emphasize the double $\mathbf{D}\mathbb{G}_{(r)}$, we choose not to make explicit reference to the relative settings therein.

6. Spectrum of cohomology for classical groups

By Theorem 5.3, the cohomology of the double $\mathbf{D}\mathbb{G}_{(r)}$ is finite over the image of $H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2])$, under the map θ_r of Definition 3.9. The map θ_r then induces a finite scheme map

$$\Theta_r : |\mathbf{D}\mathbb{G}_{(r)}| \rightarrow |k\mathbb{G}_{(r)}| \times (\mathfrak{g}^*)^{(r)}, \tag{10}$$

where $|A| = \text{Spec } H^{\text{ev}}(A)_{\text{red}}$.

In this section we show that Θ_r is an isomorphism when \mathbb{G} is one of many classical algebraic groups with either p sufficiently large for p or r sufficiently large relative to the dimension of \mathbb{G} . Our results follow from an analysis of algebraic groups which admit a quasilogarithm.

Remark 6.1. The schemes $|k\mathbb{G}_{(r)}|$ have been extensively studied and, in conjunction with support varieties of $\mathbb{G}_{(r)}$ -representations, provide one means of approaching modular representation theory. One can see the survey [Pevtsova 2013] for example.

6A. Quasilogarithms. Let \mathbb{G} be an algebraic group with Lie algebra $\mathfrak{g} = \text{Lie}(\mathbb{G})$. We let \mathbb{G} act on itself and its Lie algebra \mathfrak{g} via the adjoint action. The following definition is adapted from [Kazhdan and Varshavsky 2006].

Definition 6.2. A quasilogarithm for \mathbb{G} is a \mathbb{G} -equivariant map $L : \mathbb{G} \rightarrow \mathfrak{g}$ of k -schemes such that $L(1) = 0$ and the differential $d_1 L : T_1\mathbb{G} \rightarrow T_0\mathfrak{g}$ is the identity on \mathfrak{g} .

The information of a quasilogarithm for \mathbb{G} is exactly the information of a \mathbb{G} -linear splitting $\mathfrak{g}^* \rightarrow m_{\mathbb{G}}$ of the projection $m_{\mathbb{G}} \rightarrow m_{\mathbb{G}}/m_{\mathbb{G}}^2 = \mathfrak{g}^*$, where $m_{\mathbb{G}}$ is the maximal ideal corresponding to the identity of \mathbb{G} . Let us give some examples.

Proposition 6.3. *The general linear group GL_n admits a quasilogarithm.*

Proof. The augmentation ideal m_{GL} is generated by the functions $x_{ij} - \delta_{ij}$. Take V to be the span of these functions $k\{x_{ij} - \delta_{ij} : 1 \leq i, j \leq n\}$. The sequence $V \rightarrow m_{GL} \rightarrow \mathfrak{gl}_n^*$ provides a linear isomorphism between V and \mathfrak{gl}_n^* .

For the comultiplication on $\mathcal{O}(GL_n)$ we have $\Delta(x_{ij}) = \sum_k x_{ik} \otimes x_{kj}$. Thus for the adjoint coaction ρ restricted to V we will have $\rho(V) \subset (k1_{\mathcal{O}} \oplus V) \otimes \mathcal{O}(GL_n)$. Since m_{GL} is preserved by the adjoint coaction, and $V \subset m_{GL}$, we will also have $\rho(V) \subset m_{GL} \otimes \mathcal{O}(GL_n)$. Taking the intersection of these two subspaces gives $\rho(V) \subset V \otimes \mathcal{O}(GL_n)$. Thus we see that V is a subcomodule of $\mathcal{O}(GL_n)$ under the adjoint coaction. The aforementioned sequence then provides a GL_n -linear isomorphism $V \rightarrow \mathfrak{gl}_n^*$. Taking the inverse $\mathfrak{gl}_n^* \rightarrow V \subset m_{GL}$ provides the desired quasilogarithm. \square

We can also address many simple algebraic groups. An odd prime p is *very good* for a simple algebraic group \mathbb{G} if p does not divide n for \mathbb{G} of type A_{n-1} , if $p \neq 3$ for \mathbb{G} of type E_6, E_7, F_4, G_2 , and $p \neq 3, 5$ for \mathbb{G} of type E_8 . For convenience we extend the notion of a very good prime to GL_n , in which case all primes will be considered very good.

Corollary 6.4 (cf. [Bezrukavnikov et al. 2016, Lemma C3]). *If \mathbb{G} is a simple algebraic group for which p is very good, then \mathbb{G} admits a quasilogarithm. Furthermore, any Borel subgroup \mathbb{B} in such a \mathbb{G} also admits a quasilogarithm.*

Proof. In this case there exists an integer n and an embedding $i : \mathbb{G} \rightarrow GL_n$ such that the differential $d_1 i : \mathfrak{g} \rightarrow \mathfrak{gl}_n$ admits a \mathbb{G} -equivariant splitting $\tau : \mathfrak{gl}_n \rightarrow \mathfrak{g}$, by a result of Garibaldi [2009, Proposition 8.1]. Composing with a quasilogarithm L for GL_n produces a quasilogarithm L' for \mathbb{G} ,

$$\mathbb{G} \longrightarrow GL_n \xrightarrow{L} \mathfrak{gl}_n \xrightarrow{\tau} \mathfrak{g}.$$

By [Kazhdan and Varshavsky 2006, Lemma 1.8.3], the restriction of L' to any Borel subgroup \mathbb{B} will provide a quasilogarithm for \mathbb{B} . \square

Consider a semisimple algebraic group \mathbb{G} and a unipotent subgroup \mathbb{U} in \mathbb{G} which is normalized by a maximal torus. We let $cl(\mathbb{U})$ denote the nilpotence class of a \mathbb{Q} -form of \mathbb{U} in a \mathbb{Q} -form of \mathbb{G} (see [Seitz 2000]). For example, if we consider $\mathbb{G} = SL_n$ and \mathbb{U} the unipotent subgroup of upper triangular matrices, then $cl(\mathbb{U}) = n - 1$. The following result is covered in work of Seitz [2000, Proposition 5.2].

Proposition 6.5. *Let \mathbb{G} be semisimple and \mathbb{U} be a unipotent subgroup in \mathbb{G} which is normalized by a maximal torus. If $p > cl(\mathbb{U})$ then \mathbb{U} admits a quasilogarithm.*

The main principle here is quite simple. Under this restriction on p , the usual exponent on the \mathbb{Q} -form $\exp_{\mathbb{Q}} : \mathfrak{u}_{\mathbb{Q}} \rightarrow \mathbb{U}_{\mathbb{Q}}$ is an isomorphism defined over $\mathbb{Z}_{(p)}$, and hence induces an isomorphism $\exp_k : \mathfrak{u} \rightarrow \mathbb{U}$ over k . We define L as the inverse $L = \exp_k^{-1}$. Equivariance of L under the adjoint \mathbb{U} -action follows from $\mathbb{U}_{\mathbb{Q}}$ -invariance of $\exp_{\mathbb{Q}}$.

6B. Induced gradings on the double. Consider an algebraic group \mathbb{G} with a fixed quasilogarithm L . From L we get a map of $\mathbb{G}_{(r)}$ -algebras $S(\mathfrak{g}^*) \rightarrow \mathbb{O}(\mathbb{G}_{(r)})$ via the composition $S(\mathfrak{g}^*) \xrightarrow{L^*} \mathbb{O}(\mathbb{G}) \rightarrow \mathbb{O}(\mathbb{G}_{(r)})$, for each r . Since each $x \in \mathfrak{g}^*$ maps into the augmentation ideal in $\mathbb{O}(\mathbb{G})$, there is furthermore an induced $\mathbb{G}_{(r)}$ -algebra map $l_r : S(\mathfrak{g}^*)/I_r \rightarrow \mathbb{O}(\mathbb{G}_{(r)})$, where I_r is the ideal generated by the p^r -th powers of elements in \mathfrak{g}^* . In other words, I_r is the ideal generated by the image of the augmentation ideal under the r -th Frobenius. We can now take a smash product to arrive at an algebra map

$$\mathcal{L}_r : (S(\mathfrak{g}^*)/I_r) \# k\mathbb{G}_{(r)} \rightarrow D\mathbb{G}_{(r)}. \tag{11}$$

We note that the algebra $S(\mathfrak{g}^*)/I_r$ is graded, since the ideal I_r is generated by the homogenous elements x^{p^r} , $x \in \mathfrak{g}^*$. Furthermore, under this grading $k\mathbb{G}_{(r)}$ acts by graded endomorphisms. Hence the smash product $(S(\mathfrak{g}^*)/I_r) \# k\mathbb{G}_{(r)}$ is graded with \mathfrak{g}^* in degree 1 and $k\mathbb{G}_{(r)}$ in degree 0. This point will be of some significance below.

Lemma 6.6. *Suppose \mathbb{G} is smooth and admits a quasilogarithm L . Then for any $r > 0$ the above map $\mathcal{L}_r : (S(\mathfrak{g}^*)/I_r) \# k\mathbb{G}_{(r)} \rightarrow D\mathbb{G}_{(r)}$ is an isomorphism of algebras.*

Proof. Recall that $\dim(\mathbb{G}) = \dim(\mathfrak{g})$ whenever \mathbb{G} is smooth (see [Jantzen 2003, I.7.17(1)]). The localization at the distinguished maximal ideals of $S(\mathfrak{g}^*)$ and $\mathbb{O}(\mathbb{G})$, $S(\mathfrak{g}^*)_0 \rightarrow \mathbb{O}(\mathbb{G})_1$, is a local map of regular, local k -algebras of dimension $\dim \mathfrak{g}$ which induces an isomorphism on corresponding maximal ideals modulo their squares. Thus, L induces an isomorphism of complete local rings $\hat{L}_1 : \widehat{S(\mathfrak{g}^*)} \xrightarrow{\sim} \widehat{\mathbb{O}_{\mathbb{G},1}}$ (see e.g., [Matsumura 1989, proof of Lemma 10.28.1]). We mod out by the images of the maximal ideals under the r -th Frobenius to arrive at an isomorphism

$$S(\mathfrak{g}^*)/I_r = \widehat{S(\mathfrak{g}^*)}/\hat{I}_r \xrightarrow{\sim} \widehat{\mathbb{O}_{\mathbb{G},1}}/(f^{p^r} : f \in \hat{m}_{\mathbb{G}}) = \mathbb{O}(\mathbb{G})/(f^{p^r} : f \in m_{\mathbb{G}}) = \mathbb{O}(\mathbb{G}_{(r)}).$$

One can check on elements to see that the above isomorphism is exactly l_r . Thus, $l_r : S(\mathfrak{g}^*)/I_r \rightarrow \mathbb{O}(\mathbb{G}_{(r)})$ and hence $\mathcal{L}_r : (S(\mathfrak{g}^*)/I_r) \# k\mathbb{G}_{(r)} \rightarrow D\mathbb{G}_{(r)}$ are isomorphisms. \square

As a consequence of Lemma 6.6, we see that when \mathbb{G} is smooth and admits a quasilogarithm the double $D\mathbb{G}_{(r)}$ inherits a grading induced by \mathcal{L}_r . This grading is such that $k\mathbb{G}_{(r)}$ lies in degree 0 and $\mathcal{L}_r(\mathfrak{g}^*)$ lies in degree 1. The coordinate algebra $\mathbb{O}(\mathbb{G}_{(r)})$ will be a graded subalgebra in the double, with $\mathbb{O}(\mathbb{G}_{(r)})_0 = k$ and $\mathbb{O}(\mathbb{G}_{(r)})_1 = l_r(\mathfrak{g}^*)$.

We now consider the algebras $\mathbb{O}(\mathbb{G}_{(r)})$ and $D\mathbb{G}_{(r)}$ as graded (Noetherian, locally finite) algebras. As with any Noetherian graded algebra, the cohomologies $\text{Ext}_{\mathbb{O}(\mathbb{G}_{(r)})}^*(M, N)$ and $\text{Ext}_{D\mathbb{G}_{(r)}}^*(M, N)$ of finitely generated graded modules inherit natural gradings, in addition to the cohomological gradings. In particular, the cohomologies $H^*(\mathbb{O}(\mathbb{G}_{(r)}), k)$ and $H^*(D\mathbb{G}_{(r)}, k)$ will be graded. (See e.g., [Artin et al. 1990].) We call this extra grading on cohomology the *internal grading*.

Lemma 6.7. *Let \mathbb{G} be smooth with a fixed quasilogarithm. Consider $H^*(\mathbb{O}(\mathbb{G}_{(r)}), k)$ with its induced internal grading. Under the isomorphism $\wedge^*(\mathfrak{g}) \otimes S(\mathfrak{g}^{(r)}[2]) \cong H^*(\mathbb{O}(\mathbb{G}_{(r)}), k)$ of Proposition 3.5, \mathfrak{g} is identified with a subspace of internal degree 1 and $\mathfrak{g}^{(r)}$ is identified with a subspace of internal degree p^r .*

Proof. The algebra $\mathbb{O} = \mathbb{O}(\mathbb{G}_{(r)})$ is connected graded and generated in degree 1. Hence $\mathfrak{g} \cong H^1(\mathbb{O}, k)$ is concentrated in degree 1 (see [Artin et al. 1990]).

Under the gradings induced by the quasilogarithm, the reduction

$$\mathbb{O}_{\text{nat}} = \mathbb{O}(\mathbb{G}_{(r+1)}) \rightarrow \mathbb{O}$$

is a homogeneous map, and each deformation $\text{Def}_\xi = \mathbb{O}_{\text{nat}} \otimes_{\mathbb{O}(\mathbb{G}_{(r+1)}/\mathbb{G}_{(r)})} k[\varepsilon]$ associated to an element $\xi \in \mathfrak{g}^{(r)}$ is graded, where we take $\deg(\varepsilon) = p^r$. By choosing any graded $k[\varepsilon]$ -linear identification $\mathbb{O}[\varepsilon] \cong \text{Def}_\xi$ we see that the associated function $F_\xi : \mathbb{O} \otimes \mathbb{O} \rightarrow \mathbb{O}$, which is defined by the equation $a \cdot_\xi b = ab + F(a, b)\varepsilon$, is such that $\deg(F(a, b)) = \deg(a \otimes b) - p^r$. So the Hochschild 2-cocycle $F_\xi \in \text{Hom}_k(\mathbb{O} \otimes \mathbb{O}, \mathbb{O})$ is degree p^r , as is its image $\bar{F}_\xi \in \text{Hom}_k(\mathbb{O} \otimes \mathbb{O}, k)$. It follows that $\sigma_{\mathbb{O}}(\xi) = [\bar{F}_\xi] \in H^2(\mathbb{O}, k)$ is a homogeneous element of degree p^r . \square

6C. Spectra of cohomology. Recall the map Θ_r from (10) and the definition $|A| = \text{Spec } H^{\text{ev}}(A, k)_{\text{red}}$.

Theorem 6.8. *Suppose \mathbb{G} is a smooth algebraic group which admits a quasilogarithm. If r is such that $p^r > \dim(\mathbb{G})$, then*

$$\theta_r : H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(\text{D}\mathbb{G}_{(r)}, k)$$

is finite and injective. Consequently, the scheme map

$$\Theta_r : |\text{D}\mathbb{G}_{(r)}| \rightarrow |k\mathbb{G}_{(r)}| \times (\mathfrak{g}^*)^{(r)}$$

is finite and surjective, and furthermore $\dim|\text{D}\mathbb{G}_{(r)}| = \dim|k\mathbb{G}_{(r)}| + \dim \mathbb{G}$.

Proof. We freely use the notation of the proof of Theorem 5.3, and omit the shift [2] in the symmetric algebra to ease notation. According to Lemma 6.6, $\text{D}\mathbb{G}_{(r)}$ inherits a natural algebra grading via the isomorphism \mathcal{L}_r of (11), \mathbb{O} is a graded subalgebra, and the exact sequence $1 \rightarrow \mathbb{O} \rightarrow \text{D}\mathbb{G}_{(r)} \rightarrow k\mathbb{G}_{(r)} \rightarrow 1$ is a sequence of graded algebra maps, where $k\mathbb{G}_{(r)}$ is taken to be entirely in degree 0. In this case the spectral sequence of Proposition 5.2 inherits an internal grading so that all differentials are homogeneous of degree 0.

The internal grading at the E_2 -page is such that the degree on each $E_2^{ij} = H^i(\mathbb{G}_{(r)}, H^j(\mathbb{O}))$ is induced by the degree on $H^j(\mathbb{O})$. In particular, each summand $\wedge^{j_1} \mathfrak{g} \otimes S^{j_2}(\mathfrak{g}) \subset H^{j_1+j_2}(\mathbb{O})$ is of internal degree $j_1 + p^r j_2$, by Lemma 6.7, and the corresponding summands in the decomposition

$$H^i(\mathbb{G}_{(r)}, H^j(\mathbb{O}, k)) = H^i\left(\mathbb{G}_{(r)}, \bigoplus_{j_1+2j_2=j} \wedge^{j_1}(\mathfrak{g}) \otimes S^{j_2}(\mathfrak{g}^{(r)})\right) = \bigoplus_{j_1+2j_2=j} H^i(\mathbb{G}_{(r)}, \wedge^{j_1}(\mathfrak{g})) \otimes S^{j_2}(\mathfrak{g}^{(r)})$$

are of respective degrees $j_1 + j_2 p^r$.

Since $\dim \mathfrak{g} < p^r$, the index j_1 is such that $0 \leq j_1 < p^r$. Hence the degree $p^r \mathbb{Z}$ portion of the E_2 -page is exactly the prescribed subalgebra of permanent cocycles

$$(E_2^{i,j})_{p^r \mathbb{Z}} = H^i(\mathbb{G}_{(r)}, k) \otimes S^{j/2}(\mathfrak{g}^{(r)}) \Rightarrow (H^*(\text{D}\mathbb{G}_{(r)}, k))_{p^r \mathbb{Z}}, \tag{12}$$

where $S^{j/2}(\mathfrak{g}^{(r)})$ is taken to be 0 when j is odd.

By homogeneity of the differentials, and the fact that all of the elements of degrees $p^r\mathbb{Z}$ in $E_2^{*,*}$ are cocycles by (12), we see that no elements of degrees $p^r\mathbb{Z}$ are coboundaries. One can make the same argument at each subsequent page of the spectral sequence to find that the map $H^i(\mathbb{G}_{(r)}, k) \otimes S^t(\mathfrak{g}^{(r)}) \rightarrow E_s^{i,2t}$ is injective for all i, t , and s . It follows that $\text{gr } \theta_r : H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}) \rightarrow E_\infty^{*,*}$ is injective.

Injectivity of the associated graded map $\text{gr } \theta_r$ implies that $\theta_r : H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$ is injective. By Theorem 5.3, θ_r is also finite. After taking even degrees and reducing,

$$\theta_{\text{red}}^{\text{ev}} : H^{\text{ev}}(\mathbb{G}_{(r)}, k)_{\text{red}} \otimes S(\mathfrak{g}^{(r)}) \rightarrow H^*(D\mathbb{G}_{(r)}, k)_{\text{red}}$$

remains injective and finite. In particular, $\theta_{\text{red}}^{\text{ev}}$ is an integral extension. Thus, the map on spectra induced by θ^{ev} is finite and surjective [Matsumura 1989, Theorem 9.3]. The asserted computation of dimension follows. □

Note that the dimension of $|\mathbb{O}(\mathbb{G}_{(r)})|$ is equal to $\dim \mathbb{G}$, by Proposition 3.5 (and [Jantzen 2003, I.7.17(1)]). Hence the equality of dimensions of Theorem 6.8 can also be written as

$$\dim |D\mathbb{G}_{(r)}| = \dim |k\mathbb{G}_{(r)}| + \dim |\mathbb{O}(\mathbb{G}_{(r)})|.$$

Under stronger assumptions on p or r we can significantly strengthen the conclusion of Theorem 6.8. Indeed one can leverage the internal grading on the given spectral sequence, as in the proof of Theorem 6.8, to show that Θ_r is an isomorphism in such circumstances.

Theorem 6.9. *Suppose \mathbb{G} is a smooth algebraic group which admits a quasilogarithm. Suppose additionally that r is such that $p^r > 2 \dim \mathbb{G}$. Then the image of the injective algebra map*

$$\theta_r : H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$$

admits an $H^(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2])$ -module complement J which consists entirely of nilpotent elements in $H^*(D\mathbb{G}_{(r)}, k)$. Furthermore, the induced map on reduced spectra*

$$\Theta_r : |D\mathbb{G}_{(r)}| \rightarrow |k\mathbb{G}_{(r)}| \times (\mathfrak{g}^*)^{(r)}$$

is an isomorphism.

Proof. Fix a quasilogarithm on \mathbb{G} , and consider the induced gradings on cohomology. Note that we can consider all of our \mathbb{Z} -graded spaces as $\mathbb{Z}/p^r\mathbb{Z}$ -graded spaces, via the projection $\mathbb{Z} \rightarrow \mathbb{Z}/p^r\mathbb{Z}$. For convenience, we employ $\mathbb{Z}/p^r\mathbb{Z}$ -gradings in this proof. For an element $a \in \mathbb{Z}/p^r\mathbb{Z}$ we let \tilde{a} denote the unique representative of a in $\{0, \dots, p^r - 1\}$.

Just as in Lemma 6.7, one can check that the natural map $\sigma_D : \mathfrak{g}^{(r)} \rightarrow H^2(D\mathbb{G}_{(r)}, k)$ has image in degree $p^r = 0$ with respect to the $\mathbb{Z}/p^r\mathbb{Z}$ -grading on cohomology. We also have that the inflation $H^*(\mathbb{G}_{(r)}, k) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$ has image entirely in degree 0, since the projection $D\mathbb{G}_{(r)} \rightarrow k\mathbb{G}_{(r)}$ is graded with $k\mathbb{G}_{(r)}$ entirely in degree 0. By the same spectral sequence calculation as was given in the proof of Theorem 6.8, we find that

$$\theta_r : H^*(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}) \rightarrow H^*(D\mathbb{G}_{(r)}, k)$$

is an isomorphism onto the degree 0 portion of the cohomology of $D\mathbb{G}_{(r)}$.

Under the induced $\mathbb{Z}/p^r\mathbb{Z}$ -grading on the spectral sequence $\{E_s^{*,*}\}$ of the proof of Theorem 6.8 we have

$$(E_2^{i,j})_0 = H^i(\mathbb{G}_{(r)}, k) \otimes S^{j/2}(\mathfrak{g}^{(r)}) \quad \text{and} \quad (E_2^{i,j})_a = 0 \text{ for each } a = \dim \mathbb{G} + 1, \dots, p^r - 1.$$

This implies that $(H^*(D\mathbb{G}_{(r)}, k))_a = 0$ for each such a . Hence, any homogenous element $\xi \in H^*(D\mathbb{G}_{(r)}, k)$ of nonzero internal degree $\deg(\xi)$ satisfies $\xi^m = 0$, where

$$m = \begin{cases} \lfloor p^r / \widetilde{\deg(\xi)} \rfloor & \text{if } \widetilde{\deg(\xi)} \nmid p^r, \\ p^r / \widetilde{\deg(\xi)} - 1 & \text{if } \widetilde{\deg(\xi)} \mid p^r, \end{cases}$$

since $\widetilde{\deg(\xi^m)}$ will be among $\dim \mathbb{G} + 1, \dots, p^r - 1$. Said another way, the subspace J spanned by elements of nonzero degree is contained in the nilradical, and the inclusion

$$(H^{\text{ev}}(D\mathbb{G}_{(r)}, k)_0)_{\text{red}} \rightarrow H^{\text{ev}}(D\mathbb{G}_{(r)})_{\text{red}}$$

is therefore an isomorphism. Since θ_r is an isomorphism onto the degree 0 portion of cohomology, it follows that

$$\theta_{\text{red}}^{\text{ev}} : H^{\text{ev}}(\mathbb{G}_{(r)}, k)_{\text{red}} \otimes S(\mathfrak{g}^{(r)}) \rightarrow H^{\text{ev}}(D\mathbb{G}_{(r)}, k)_{\text{red}}$$

is an isomorphism. We take spectra to find that Θ_r is an isomorphism. □

Theorem 6.10. *Suppose \mathbb{G} is a smooth algebraic group which admits a quasilogarithm, and that $p > \dim \mathbb{G} + 1$. Then the image of θ_r in $H^*(D\mathbb{G}_{(r)}, k)$ has a complement J which consists entirely of nilpotents, just as in Theorem 6.9. Furthermore, the map*

$$\Theta_r : |D\mathbb{G}_{(r)}| \rightarrow |k\mathbb{G}_{(r)}| \times (\mathfrak{g}^*)^{(r)}$$

is an isomorphism for all r .

Proof. Our argument will be similar to that of Theorem 6.9. Via the projection $\mathbb{Z} \rightarrow \mathbb{Z}/p\mathbb{Z}$ we get $\mathbb{Z}/p\mathbb{Z}$ -gradings on the spectral sequence $\{E_r^{*,*}\}$ and the cohomology $H^*(D\mathbb{G}_{(r)}, k)$. We have, under these $\mathbb{Z}/p\mathbb{Z}$ -gradings, that

$$(E_2^{i,j})_0 = H^i(\mathbb{G}_{(r)}, k) \otimes S^{j/2}(\mathfrak{g}^{(r)}), \quad (E_2^{i,j})_{p-1} = (E_2^{i,j})_{-1} = 0,$$

and that θ_r is an isomorphism onto the degree 0 portion of cohomology $H^*(D\mathbb{G}_{(r)}, k)_0$. Consider now any homogeneous element $\xi \in H^*(D\mathbb{G}_{(r)}, k)$ of degree $d \neq 0$. Since $\mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ is a field there is a positive integer $d' \in \mathbb{Z}$ which reduces to $-d^{-1} \pmod p$. We then find that $\xi^{d'} = 0$, since

$$\deg(\xi^{d'}) = -d^{-1}d = -1 \quad \text{and} \quad H^*(D\mathbb{G}_{(r)}, k)_{-1} = 0.$$

Hence the subspace J of element of nonzero degree is contained in the nilradical. Just as before, this implies that θ_r induces an isomorphism

$$\theta_{\text{red}}^{\text{ev}} : H^{\text{ev}}(\mathbb{G}_{(r)}, k)_{\text{red}} \otimes S(\mathfrak{g}^{(r)}) \rightarrow H^{\text{ev}}(D\mathbb{G}_{(r)}, k)_{\text{red}},$$

and that Θ_r is an isomorphism as well. □

One considers the examples of Section 6A to arrive at

Corollary 6.11. *Let \mathbb{G} be a general linear group, simple algebraic group, Borel subgroup in a simple algebraic group, or a unipotent subgroup in a semisimple algebraic group which is normalized by a maximal torus. Suppose that p is very good for \mathbb{G} , or that $p > \text{cl}(\mathbb{G})$ in the unipotent case:*

- *If $p > \dim \mathbb{G} + 1$ then Θ_r is an isomorphism for all r .*
- *For arbitrary p satisfying the hypothesis, the map Θ_r is an isomorphism whenever r is such that $p^r > 2 \dim \mathbb{G}$.*

7. Results for support varieties

For a Hopf algebra A and finite dimensional A -module M we let $|A|_M$ denote the support variety for M . This is the closed, reduced, subscheme in $|A|$ defined by the kernel of the algebra map

$$- \otimes M : H^{\text{ev}}(A, k) \rightarrow \text{Ext}_A^{\text{ev}}(M, M). \tag{13}$$

In this section we consider the support $|\text{D}\mathbb{G}_{(r)}|_M$ associated to a finite dimensional $\text{D}\mathbb{G}_{(r)}$ -module M . We show that there is a finite scheme map

$$\Theta_r^M : |\text{D}\mathbb{G}_{(r)}|_M \rightarrow |k\mathbb{G}_{(r)}|_M \times (\mathfrak{g}^*)^{(r)}$$

for any M with trivial restriction to $\mathbb{O}(\mathbb{G}_{(r)})$, and that Θ_r^M is an isomorphism whenever M is irreducible and \mathbb{G} is a classical group at a large prime or large r .

7A. Generalities for support varieties. Let A be a Hopf algebra and M be a finite dimensional A -module. Under the natural identification

$$\text{Ext}_A^*(M, M) = \text{Ext}_A^*(k, M \otimes M^*) = H^*(A, M \otimes M^*),$$

(13) corresponds to the mapping

$$\text{coev}_*^M : H^{\text{ev}}(A, k) \rightarrow H^{\text{ev}}(A, M \otimes M^*)$$

induced by the coevaluation $\text{coev}^M : k \rightarrow M \otimes M^*$ [Etingof et al. 2015, Proposition 2.10.8]. The algebra structure on $H^{\text{ev}}(A, M \otimes M^*)$ is induced by the algebra structure on $M \otimes M^* \cong \text{End}_k(M, M)$. By [Mac Lane 1963, Theorem VII.4.1] (see also [Suarez-Alvarez 2004]) the image of $H^{\text{ev}}(A, k)$ lies in the center of $H^{\text{ev}}(A, M \otimes M^*)$.

For \mathbb{G} smooth and M any finite dimensional $\text{D}\mathbb{G}_{(r)}$ -module, θ_r produces an algebra map

$$f_{r,M} : H^{\text{ev}}(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^{\text{ev}}(\text{D}\mathbb{G}_{(r)}, M \otimes M^*). \tag{14}$$

Explicitly, $f_{r,M}$ is the composite

$$H^{\text{ev}}(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) \xrightarrow{\theta_r} H^{\text{ev}}(\text{D}\mathbb{G}_{(r)}, k) \xrightarrow{\text{coev}_*^M} H^{\text{ev}}(\text{D}\mathbb{G}_{(r)}, M \otimes M^*).$$

By the definition of $f_{r,M}$, one sees that the reduced subscheme in $|k\mathbb{G}_{(r)}| \times (\mathfrak{g}^*)^{(r)}$ defined by the kernel of $f_{r,M}$ is exactly the image of $|\mathrm{D}\mathbb{G}_{(r)}|_M$ under $\Theta_r : |\mathrm{D}\mathbb{G}_{(r)}| \rightarrow |k\mathbb{G}_{(r)}| \times (\mathfrak{g}^*)^{(r)}$.

By the material of Section 6 we understand that Θ_r is often an isomorphism. However, by finiteness of Θ_r in general, we can adapt an argument of [Friedlander and Parshall 1987, Proposition 1.5] in all circumstances to arrive at:

Proposition 7.1. *A finite dimensional $\mathrm{D}\mathbb{G}_{(r)}$ -module M is projective (or, equivalently, injective) as a $\mathrm{D}\mathbb{G}_{(r)}$ -module if and only if $\Theta_r(|\mathrm{D}\mathbb{G}_{(r)}|_M) = \{0\}$.*

Proof. One simply repeats the proof of [Friedlander and Parshall 1987, Proposition 1.5], using the fact that $\mathrm{rep}(\mathrm{D}\mathbb{G}_{(r)})$ is a Frobenius category [Larson and Sweedler 1969]. □

For the remainder of the section we seek to give a more precise description of the support $|\mathrm{D}\mathbb{G}_{(r)}|_M$ for a finite dimensional $\mathrm{D}\mathbb{G}_{(r)}$ -module M whose restriction to $\mathcal{O}(\mathbb{G}_{(r)})$ is trivial (and thus arises as the restriction along the quotient $\mathrm{D}\mathbb{G}_{(r)} \rightarrow kG_{(r)}$ of a $kG_{(r)}$ module which we also denote by M). By Proposition 5.5, this condition is satisfied by any irreducible $\mathrm{D}\mathbb{G}_{(r)}$ -module. Whenever M satisfies this condition, there is a natural inflation map $H^*(\mathbb{G}_{(r)}, M) \rightarrow H^*(\mathrm{D}\mathbb{G}_{(r)}, M)$.

In the statement of the following lemma, we consider the algebra map

$$\theta_{r,M} : H^{\mathrm{ev}}(\mathbb{G}_{(r)}, M \otimes M^*) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^{\mathrm{ev}}(\mathrm{D}\mathbb{G}_{(r)}, M \otimes M^*)$$

induced by the inflation $H^{\mathrm{ev}}(\mathbb{G}_{(r)}, M \otimes M^*) \rightarrow H^{\mathrm{ev}}(\mathrm{D}\mathbb{G}_{(r)}, M \otimes M^*)$ and the map from $S(\mathfrak{g}^{(r)}[2])$ defined via σ'_D as above.

Lemma 7.2. *For any finite dimensional $\mathrm{D}\mathbb{G}_{(r)}$ -module M whose restriction to $\mathcal{O}(\mathbb{G}_{(r)})$ is trivial, the following diagram commutes*

$$\begin{array}{ccc} H^{\mathrm{ev}}(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) & \xrightarrow{\mathrm{coev}_*^M \otimes \mathrm{id}_S} & H^{\mathrm{ev}}(\mathbb{G}_{(r)}, M \otimes M^*) \otimes S(\mathfrak{g}^{(r)}[2]) \\ \theta_r \downarrow & \dashrightarrow^{f_{r,M}} & \downarrow \theta_{r,M} \\ H^{\mathrm{ev}}(\mathrm{D}\mathbb{G}_{(r)}, k) & \xrightarrow{\mathrm{coev}_*^M} & H^{\mathrm{ev}}(\mathrm{D}\mathbb{G}_{(r)}, M \otimes M^*). \end{array}$$

Proof. It suffices to prove that the two maps

$$H^{\mathrm{ev}}(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) \rightrightarrows H^{\mathrm{ev}}(\mathrm{D}\mathbb{G}_{(r)}, M \otimes M^*)$$

agree on the factors $H^{\mathrm{ev}}(\mathbb{G}_{(r)}, k)$ and $S(\mathfrak{g}^{(r)}[2])$ independently. The two restrictions to $S(\mathfrak{g}^{(r)}[2])$ are equal, since they are both defined as the composition

$$S(\mathfrak{g}^{(r)}[2]) \xrightarrow{\sigma'_D} H^*(\mathrm{D}\mathbb{G}_{(r)}, k) \xrightarrow{\mathrm{coev}_*^M} H^*(\mathrm{D}\mathbb{G}_{(r)}, M \otimes M^*).$$

So we need only establish commutativity of the diagram

$$\begin{array}{ccc}
 H^{\text{ev}}(\mathbb{G}_{(r)}, k) & \xrightarrow{\text{coev}_*^M} & H^{\text{ev}}(\mathbb{G}_{(r)}, M \otimes M^*) \\
 \text{inf} \downarrow & & \downarrow \text{inf} \\
 H^{\text{ev}}(\text{D}\mathbb{G}_{(r)}, k) & \xrightarrow{\text{coev}_*^M} & H^{\text{ev}}(\text{D}\mathbb{G}_{(r)}, M \otimes M^*),
 \end{array}$$

which follows by functoriality of the inflation map. □

Proposition 7.3. *For any finite dimensional $\text{D}\mathbb{G}_{(r)}$ -module M whose restriction to $\mathbb{O}(\mathbb{G}_{(r)})$ is trivial (for example, if M is irreducible), the restriction of $\Theta_r : |\text{D}\mathbb{G}_{(r)}|_M \rightarrow |k\mathbb{G}_{(r)}| \times (\mathfrak{g}^*)^{(r)}$ to $|\text{D}\mathbb{G}_{(r)}|_M$ factors through the closed subscheme $|k\mathbb{G}_{(r)}|_M \times (\mathfrak{g}^*)^{(r)}$, determining a finite map of schemes*

$$\Theta_{r,M} : |\text{D}\mathbb{G}_{(r)}|_M \rightarrow |k\mathbb{G}_{(r)}|_M \times (\mathfrak{g}^*)^{(r)}.$$

Proof. The image of $\Theta_r|_{|\text{D}\mathbb{G}_{(r)}|_M}$ is the closed subscheme defined by the kernel of $f_{r,M}$. By Lemma 7.2, $f_{r,M}$ factors through the product map

$$\text{coev}_*^M \otimes \text{id}_S : H^{\text{ev}}(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^{\text{ev}}(\mathbb{G}_{(r)}, M \otimes M^*) \otimes S(\mathfrak{g}^{(r)}[2]),$$

and hence

$$\ker(\text{coev}_*^M) \otimes S(\mathfrak{g}^{(r)}[2]) \subset \ker(f_{r,M}).$$

It follows that $\Theta_r|_{|\text{D}\mathbb{G}_{(r)}|_M}$ factors through $|k\mathbb{G}_{(r)}|_M \times (\mathfrak{g}^*)^{(r)}$. □

7B. Support varieties for classical groups. We now consider irreducible modules and classical groups. We fix \mathbb{G} a smooth algebraic group.

Lemma 7.4. *Suppose \mathbb{G} admits a quasilogarithm and that V is an irreducible $\text{D}\mathbb{G}_{(r)}$ -module. Suppose additionally that $p^r > \dim \mathbb{G}$. Then the map*

$$\theta_{r,V} : H^{\text{ev}}(\mathbb{G}_{(r)}, V \otimes V^*) \otimes S(\mathfrak{g}^{(r)}[2]) \rightarrow H^{\text{ev}}(\text{D}\mathbb{G}_{(r)}, V \otimes V^*)$$

is injective.

Proof. Take $\mathbb{O} = \mathbb{O}(\mathbb{G}_{(r)})$. It suffices to show that the associated graded map $\text{gr} \theta_{r,V}$ is injective, under some filtration.

We consider the Grothendieck spectral sequence

$$E_2^{i,j} = H^i(\mathbb{G}_{(r)}, H^j(\mathbb{O}, V \otimes V^*)) \Rightarrow H^{i+j}(\text{D}\mathbb{G}_{(r)}, V \otimes V^*)$$

induced by the sequence $1 \rightarrow \mathbb{O} \rightarrow \text{D}\mathbb{G}_{(r)} \rightarrow k\mathbb{G}_{(r)} \rightarrow 1$. Recall, from Proposition 5.5, that \mathbb{O} acts trivially on V and V^* . Whence we may rewrite the above spectral sequence as

$$E_2^{i,j} = H^i(\mathbb{G}_{(r)}, (\wedge^{j_1} \mathfrak{g}) \otimes V \otimes V^*) \otimes S^{j_2}(\mathfrak{g}^{(r)}[2]) \Rightarrow H^{i+j}(\text{D}\mathbb{G}_{(r)}, V \otimes V^*).$$

Since \mathbb{O} acts trivially on V and V^* , the $D\mathbb{G}_{(r)}$ -module $V \otimes V^*$ is graded and concentrated in degree 0, under the \mathbb{Z} -grading on $D\mathbb{G}_{(r)}$ induced by any quasilogarithm on \mathbb{G} . Now one can argue just as in the proof of Theorem 6.8, using the grading on the above spectral sequence induced by the quasilogarithm, to conclude that $\theta_{r,V}$ is injective. \square

Theorem 7.5. *Suppose \mathbb{G} admits a quasilogarithm and that V is an irreducible $D\mathbb{G}_{(r)}$ -module. Then the scheme map*

$$\Theta_{r,V} : |D\mathbb{G}_{(r)}|_V \rightarrow |k\mathbb{G}_{(r)}|_V \times (\mathfrak{g}^*)^{(r)}$$

is finite and surjective. Furthermore, when $p > \dim \mathbb{G} + 1$ or $p^r > 2 \dim \mathbb{G}$ the map $\Theta_{r,V}$ is an isomorphism.

Proof. Finiteness follows by finiteness of Θ_r . So we need only check surjectivity to establish the first claim. We omit the shift [2] in the symmetric algebra to ease notation. As discussed above, the image of $\Theta_{r,V}$ is the subscheme associated to the kernel of the algebra map

$$f_{r,V} : H^{\text{ev}}(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}) \rightarrow H^{\text{ev}}(D\mathbb{G}_{(r)}, V \otimes V^*),$$

which was defined at (14). Now, by Lemma 7.2, we have that $f_{r,V}$ factors as the composite of

$$\text{coev}_*^V \otimes \text{id}_S : H^{\text{ev}}(\mathbb{G}_{(r)}, k) \otimes S(\mathfrak{g}^{(r)}) \rightarrow H^{\text{ev}}(\mathbb{G}_{(r)}, V \otimes V^*) \otimes S(\mathfrak{g}^{(r)})$$

with

$$\theta_{r,V} : H^{\text{ev}}(\mathbb{G}_{(r)}, V \otimes V^*) \otimes S(\mathfrak{g}^{(r)}) \rightarrow H^{\text{ev}}(D\mathbb{G}_{(r)}, V \otimes V^*).$$

By Lemma 7.4, $\theta_{r,V}$ is injective. Hence it follows that $\ker(f_{r,V}) = \ker(\text{coev}_*^V) \otimes S(\mathfrak{g}^{(r)})$ and subsequently

$$\Theta_{r,V}(|D\mathbb{G}_{(r)}|_V) = |k\mathbb{G}_{(r)}|_V \times (\mathfrak{g}^*)^{(r)}.$$

The fact that $\Theta_{r,V}$ is an isomorphism when $p > \dim \mathbb{G} + 1$ or $p^r > 2 \dim \mathbb{G}$ follows from the fact that Θ_r is an isomorphism in these cases, by Theorems 6.9 and 6.10. \square

We apply the theorem in the classical settings to find

Corollary 7.6. *Let \mathbb{G} be a general linear group, simple algebraic group, Borel subgroup in a simple algebraic group, or a unipotent subgroup in a semisimple algebraic group which is normalized by a maximal torus. Suppose that p is very good for \mathbb{G} , or that $p > \text{cl}(\mathbb{G})$ in the unipotent case:*

- *If $p > \dim \mathbb{G} + 1$ then $\Theta_{r,V}$ is an isomorphism for every r and irreducible $D\mathbb{G}_{(r)}$ -module V .*
- *For arbitrary p satisfying the hypothesis, and r such that $p^r > 2 \dim \mathbb{G}$, the map $\Theta_{r,V}$ is an isomorphism for every irreducible $D\mathbb{G}_{(r)}$ -module V .*

Proof. By Propositions 6.3 and 6.5, and Corollary 6.4, the group \mathbb{G} admits a quasilogarithm. Hence we may apply Theorem 7.5. \square

Acknowledgments

Thanks to Roman Bezrukavnikov, Robert Guralnick, Julia Pevtsova, Julia Plavnik, and Sarah Witherspoon for helpful conversations. We are particularly grateful to the referee for detailed, constructive comments.

References

- [Andruskiewitsch and Schneider 2002] N. Andruskiewitsch and H.-J. Schneider, “Pointed Hopf algebras”, pp. 1–68 in *New directions in Hopf algebras*, edited by S. Montgomery and H.-J. Schneider, Math. Sci. Res. Inst. Publ. **43**, Cambridge Univ. Press, 2002. MR Zbl
- [Artin et al. 1990] M. Artin, J. Tate, and M. Van den Bergh, “Some algebras associated to automorphisms of elliptic curves”, pp. 33–85 in *The Grothendieck Festschrift, I*, edited by P. Cartier et al., Progr. Math. **86**, Birkhäuser, Boston, 1990. MR Zbl
- [Bendel et al. 2014] C. P. Bendel, D. K. Nakano, B. J. Parshall, and C. Pillen, *Cohomology for quantum groups via the geometry of the nullcone*, Mem. Amer. Math. Soc. **1077**, Amer. Math. Soc., Providence, RI, 2014. MR Zbl
- [Benkart et al. 2010] G. Benkart, M. Pereira, and S. Witherspoon, “Yetter–Drinfeld modules under cocycle twists”, *J. Algebra* **324**:11 (2010), 2990–3006. MR Zbl
- [Bezrukavnikov et al. 2016] R. Bezrukavnikov, D. Kazhdan, and Y. Varshavsky, “On the depth r Bernstein projector”, *Selecta Math. (N.S.)* **22**:4 (2016), 2271–2311. MR Zbl
- [Cline 1987] E. Cline, “Simulating algebraic geometry with algebra, III: The Lusztig conjecture as a TG_1 -problem”, pp. 149–161 in *The Arcata Conference on Representations of Finite Groups* (Arcata, CA, 1986), edited by P. Fong, Proc. Sympos. Pure Math. **47**, part 2, Amer. Math. Soc., Providence, RI, 1987. MR Zbl
- [Drupieski 2016] C. M. Drupieski, “Cohomological finite-generation for finite supergroup schemes”, *Adv. Math.* **288** (2016), 1360–1432. MR Zbl
- [Etingof 2002] P. Etingof, “On Vafa’s theorem for tensor categories”, *Math. Res. Lett.* **9**:5-6 (2002), 651–657. MR Zbl
- [Etingof and Gelaki 2002] P. Etingof and S. Gelaki, “On the quasi-exponent of finite-dimensional Hopf algebras”, *Math. Res. Lett.* **9**:2-3 (2002), 277–287. 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. 2011] P. Etingof, D. Nikshych, and V. Ostrik, “Weakly group-theoretical and solvable fusion categories”, *Adv. Math.* **226**:1 (2011), 176–205. 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
- [Friedlander and Parshall 1986] E. M. Friedlander and B. J. Parshall, “Cohomology of Lie algebras and algebraic groups”, *Amer. J. Math.* **108**:1 (1986), 235–253. MR Zbl
- [Friedlander and Parshall 1987] E. M. Friedlander and B. J. Parshall, “Geometry of p -unipotent Lie algebras”, *J. Algebra* **109**:1 (1987), 25–45. MR Zbl
- [Friedlander and Suslin 1997] E. M. Friedlander and A. Suslin, “Cohomology of finite group schemes over a field”, *Invent. Math.* **127**:2 (1997), 209–270. MR Zbl
- [Garibaldi 2009] S. Garibaldi, “Vanishing of trace forms in low characteristics”, *Algebra Number Theory* **3**:5 (2009), 543–566. MR Zbl
- [Gelaki et al. 2009] S. Gelaki, D. Naidu, and D. Nikshych, “Centers of graded fusion categories”, *Algebra Number Theory* **3**:8 (2009), 959–990. MR Zbl
- [Gerstenhaber 1964] M. Gerstenhaber, “On the deformation of rings and algebras”, *Ann. of Math. (2)* **79** (1964), 59–103. MR Zbl
- [Ginzburg and Kumar 1993] V. Ginzburg and S. Kumar, “Cohomology of quantum groups at roots of unity”, *Duke Math. J.* **69**:1 (1993), 179–198. MR Zbl
- [Grothendieck 1957] A. Grothendieck, “Sur quelques points d’algèbre homologique”, *Tôhoku Math. J. (2)* **9**:2 (1957), 119–183. MR Zbl
- [Jantzen 2003] J. C. Jantzen, *Representations of algebraic groups*, 2nd ed., Mathematical Surveys and Monographs **107**, Amer. Math. Soc., Providence, RI, 2003. MR Zbl
- [Kashina et al. 2006] Y. Kashina, Y. Sommerhäuser, and Y. Zhu, *On higher Frobenius–Schur indicators*, Mem. Amer. Math. Soc. **855**, Amer. Math. Soc., Providence, RI, 2006. MR Zbl

- [Kazhdan and Varshavsky 2006] D. Kazhdan and Y. Varshavsky, “Endoscopic decomposition of certain depth zero representations”, pp. 223–301 in *Studies in Lie theory*, edited by J. Bernstein et al., Progr. Math. **243**, Birkhäuser, Boston, 2006. MR Zbl
- [Larson and Sweedler 1969] R. G. Larson and M. E. Sweedler, “An associative orthogonal bilinear form for Hopf algebras”, *Amer. J. Math.* **91**:1 (1969), 75–94. MR Zbl
- [Mac Lane 1963] S. Mac Lane, *Homology*, Grundlehren der Math. Wissenschaften **114**, Springer, New York, 1963. MR Zbl
- [Majid and Oeckl 1999] S. Majid and R. Oeckl, “Twisting of quantum differentials and the Planck scale Hopf algebra”, *Comm. Math. Phys.* **205**:3 (1999), 617–655. MR Zbl
- [Massey 1954] W. S. Massey, “Products in exact couples”, *Ann. of Math. (2)* **59**:3 (1954), 558–569. MR Zbl
- [Mastnak et al. 2010] M. Mastnak, J. Pevtsova, P. Schauenburg, and S. Witherspoon, “Cohomology of finite-dimensional pointed Hopf algebras”, *Proc. Lond. Math. Soc. (3)* **100**:2 (2010), 377–404. MR Zbl
- [Matsumura 1989] H. Matsumura, *Commutative ring theory*, 2nd ed., Cambridge Studies in Adv. Math. **8**, Cambridge Univ. Press, 1989. MR Zbl
- [Montgomery 1993] S. Montgomery, *Hopf algebras and their actions on rings*, CBMS Regional Conf. Series in Math. **82**, Amer. Math. Soc., Providence, RI, 1993. MR Zbl
- [Montgomery 2004] S. Montgomery, “Algebra properties invariant under twisting”, pp. 229–243 in *Hopf algebras in noncommutative geometry and physics* (Brussels, 2002), edited by S. Caenepeel and F. Van Oystaeyen, Lect. Notes in Pure and Appl. Math. **239**, Dekker, New York, 2004. MR Zbl
- [Ng and Schauenburg 2007] S.-H. Ng and P. Schauenburg, “Frobenius–Schur indicators and exponents of spherical categories”, *Adv. Math.* **211**:1 (2007), 34–71. MR Zbl
- [Pevtsova 2013] J. Pevtsova, “Representations and cohomology of finite group schemes”, pp. 231–261 in *Advances in representation theory of algebras* (Bielefeld, 2012), edited by D. J. Benson et al., Eur. Math. Soc., Zürich, 2013. MR Zbl
- [Pevtsova and Witherspoon 2009] J. Pevtsova and S. Witherspoon, “Varieties for modules of quantum elementary abelian groups”, *Algebr. Represent. Theory* **12**:6 (2009), 567–595. MR Zbl
- [Radford 1993] D. E. Radford, “Minimal quasitriangular Hopf algebras”, *J. Algebra* **157**:2 (1993), 285–315. MR Zbl
- [Seitz 2000] G. M. Seitz, “Unipotent elements, tilting modules, and saturation”, *Invent. Math.* **141**:3 (2000), 467–502. MR Zbl
- [Shimizu 2017] K. Shimizu, “Integrals for finite tensor categories”, preprint, 2017. arXiv
- [Ştefan and Vay 2016] D. Ştefan and C. Vay, “The cohomology ring of the 12-dimensional Fomin–Kirillov algebra”, *Adv. Math.* **291** (2016), 584–620. MR Zbl
- [Suarez-Alvarez 2004] M. Suarez-Alvarez, “The Hilton–Heckmann argument for the anti-commutativity of cup products”, *Proc. Amer. Math. Soc.* **132**:8 (2004), 2241–2246. MR Zbl
- [Sullivan 1978] J. B. Sullivan, “Representations of the hyperalgebra of an algebraic group”, *Amer. J. Math.* **100**:3 (1978), 643–652. MR Zbl
- [Waterhouse 1979] W. C. Waterhouse, *Introduction to affine group schemes*, Grad. Texts in Math. **66**, Springer, 1979. MR Zbl

Communicated by Susan Montgomery

Received 2017-10-09 Revised 2018-02-12 Accepted 2018-03-29

ericmf@usc.edu

*Department of Mathematics, University of Southern California,
Los Angeles, CA, United States*

negronc@mit.edu

*Department of Mathematics, Massachusetts Institute of Technology,
Cambridge, MA, 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 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 L^AT_EX but submissions in other varieties of T_EX, 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 BibT_EX 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. 5 2018

Semistable Chow–Hall algebras of quivers and quantized Donaldson–Thomas invariants HANS FRANZEN and MARKUS REINEKE	1001
Certain abelian varieties bad at only one prime ARMAND BRUMER and KENNETH KRAMER	1027
Characterization of Kollár surfaces GIANCARLO URZÚA and JOSÉ IGNACIO YÁÑEZ	1073
Représentations de réduction unipotente pour $SO(2n + 1)$, III: Exemples de fronts d’onde JEAN-LOUP WALDSPURGER	1107
Correspondences without a core RAJU KRISHNAMOORTHY	1173
Local topological algebraicity with algebraic coefficients of analytic sets or functions GUILLAUME ROND	1215
Polynomial bound for the nilpotency index of finitely generated nil algebras MÁTYÁS DOMOKOS	1233
Arithmetic functions in short intervals and the symmetric group BRAD RODGERS	1243
Cohomology for Drinfeld doubles of some infinitesimal group schemes ERIC M. FRIEDLANDER and CRIS NEGRON	1281



1937-0652(2018)12:5;1-C