# Algebra & Number Theory

Volume 4

2010

No. 7

mathematical sciences publishers

# Algebra & Number Theory

www.jant.org





See inside back cover or www.jant.org for submission instructions.

The subscription price for 2010 is US $140/year for the electronic version, and $200/year (+$30 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Algebra & Number Theory (ISSN 1937-0652) at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

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

PUBLISHED BY

mathematical sciences publishers

http://www.mathscipub.org

A NON-PROFIT CORPORATION

Typeset in LaTeX

ALGEBRA AND NUMBER THEORY 4:7(2010)

# Hochschild cohomology and homology of quantum complete intersections

Steffen Oppermann

We compute the Hochschild cohomology and homology for arbitrary finite-dimensional quantum complete intersections. It turns out that their behavior varies widely, depending on the choice of commutation parameters, and we give precise criteria for when to expect what behavior.

## 1. Introduction

Quantum complete intersections were first discussed by Avramov, Gasharov, and Peeva [Avramov et al. 1997]. Based on the introduction of quantized versions of polynomial rings in [Manin 1987], they introduced the notion of quantum regular sequences.

In this paper we restrict to finite-dimensional quantum complete intersections, that is, algebras of the form $k\langle x_1, \ldots, x_c\rangle/I$, where $I$ is an ideal generated by $x_i^{n_i}$ for some $n_i \in \mathbb{N}_{\geq 2}$, and $x_j x_i - q_{ij} x_i x_j$ for some commutation parameters $q_{ij}$ from the multiplicative group of the field.

In particular, in the case of two variables it is known that the homological behavior of finite-dimensional quantum complete intersections varies greatly, depending on the commutation parameters.

Buchweitz, Green, Madsen, and Solberg [Buchweitz et al. 2005] gave a finite-dimensional quantum complete intersection as the first example of an algebra of infinite global dimension which has finite Hochschild cohomology. This result was generalized in [Bergh and Erdmann 2008], which showed that a finite-dimensional quantum complete intersection of codimension 2 ($c = 2$ in the description above) has an infinite Hochschild cohomology if and only if the commutation parameter is a root of unity.

On the other hand, in [Bergh and Oppermann 2008a] we showed that in the situation that all commutation parameters are roots of unity, the Hochschild cohomology of a quantum complete intersection is as well behaved as in the commutative case:

The author was supported by NFR Storforsk grant no. 167130.
*MSC2000:* primary 16E40; secondary 81R50, 16U80, 16S80.
*Keywords:* Hochschild cohomology, Hochschild homology, quantum complete intersection.

It is a finitely generated $k$ algebra, and any $\operatorname{Ext}^*(M, N)$ for any finite-dimensional modules $M$ and $N$ over the quantum complete intersection is finitely generated as a module over the Hochschild cohomology ring.

This paper gives a general description of the Hochschild cohomology and homology of finite-dimensional quantum complete intersections. Here is an outline:

In Theorems 3.4 and 7.4 we explicitly determine a $k$-basis for the Hochschild cohomology and homology, respectively.

Using these results we study the size of the Hochschild cohomology and homology in the following sense: Let $\mathbb{N}$ be the set of nonnegative integers (i.e., $0 \in \mathbb{N}$). We denote by

$$\gamma(\mathrm{HH}^*(\Lambda)) = \inf\left\{ t \in \mathbb{N} \;\middle|\; \limsup \frac{\dim_k \mathrm{HH}^n(\Lambda)}{n^{t-1}} < \infty \right\}$$

the rate of growth of the Hochschild cohomology (and similarly for the Hochschild homology). In Theorems 4.5 and 8.2, we obtain explicit combinatorial formulas for $\gamma(\mathrm{HH}^*(\Lambda))$ and $\gamma(\mathrm{HH}_*(\Lambda))$. In particular it will be shown (Corollary 4.6) that whenever not all commutation parameters are roots of unity we have $\gamma(\mathrm{HH}^*(\Lambda)) \leq c-2$. For $c=2$ that means that the Hochschild cohomology is finite. This explains why there are essentially only two cases for $c = 2$, while we obtain additional behaviors for larger $c$.

We will also generalize the result of [Bergh and Erdmann 2008] in another way: It will be shown that whenever the commutation parameters are sufficiently generic the Hochschild cohomology of the quantum complete intersection is finite (see Example 6.2).

Finally we will study the multiplicative structure of the Hochschild cohomology ring. It will turn out (Theorem 5.5) that it always contains a subring $\mathcal{S}$ which is finitely generated over $k$, and isomorphic to the quotient of the Hochschild cohomology modulo its nilpotent elements. We will give a criterion for when the entire Hochschild cohomology ring is finitely generated over this subring (Theorem 5.9). We will give examples (Examples 6.4 and 6.5) that all the following behaviors occur (for $c \geq 3$):

- $\mathcal{S} = k$, but $\gamma(\mathrm{HH}^*(\Lambda)) = c-2$.
- $\gamma(\mathcal{S}) = \gamma(\mathrm{HH}^*(\Lambda)) = c-2$, and $\mathrm{HH}^*(\Lambda)$ is finitely generated over $\mathcal{S}$.
- $\gamma(\mathcal{S}) = \gamma(\mathrm{HH}^*(\Lambda)) = c-2$, but $\mathrm{HH}^*(\Lambda)$ is not finitely generated over $\mathcal{S}$.

## 2. Notation and background

Throughout this paper we assume $k$ to be field.

***Quantum complete intersections.*** (See also [Bergh and Erdmann 2008; Bergh and Oppermann 2008a; 2008b].) A finite-dimensional *quantum complete intersection* of codimension $c$ is a $k$-algebra of the form

$$\Lambda_{\boldsymbol{q}}^{\boldsymbol{n}} = \frac{k\langle x_1, \dots, x_c\rangle}{\begin{pmatrix} x_i^{n_i} & \text{for } 1 \le i \le c \\ x_j x_i - q_{ij} x_i x_j & \text{for } 1 \le i < j \le c \end{pmatrix}}$$

with $\boldsymbol{n} = (n_1, \dots, n_c) \in \mathbb{N}_{\ge 2}^c$ and $\boldsymbol{q} = (q_{ij} \mid i < j) \in (k^\times)^{n(n-1)/2}$, where $k^\times$ denotes the multiplicative group $k \setminus \{0\}$. For convenience of notation we also define $q_{ij}$ for $i \ge j$: We set $q_{ii} = 1$ for any $i \in \{1, \dots, c\}$ and $q_{ij} = q_{ji}^{-1}$ for $1 \le j < i \le c$. The relations $x_j x_i - q_{ij} x_i x_j$ for $1 \le j \le i \le c$ are automatically satisfied in $\Lambda_{\boldsymbol{q}}^{\boldsymbol{n}}$.

Note that $\Lambda_{\boldsymbol{q}}^{\boldsymbol{n}}$ is a $\mathbb{Z}^c$-graded algebra by $|x_i| = \operatorname{degree} x_i = e_i$, the $i$-th unit vector. We will denote by $\le$ the partial order on $\mathbb{Z}^c$ defined by comparing vectors componentwise, and by $\mathbf{1} = \sum e_i$ the vector with 1 in every component. With this notation, the dimensions of the graded component of degree $\boldsymbol{d}$ (with $\boldsymbol{d} \in \mathbb{Z}^c$) are

$$\dim(\Lambda_{\boldsymbol{q}}^{\boldsymbol{n}})_{\boldsymbol{d}} = \begin{cases} 1 & \text{if } \mathbf{0} \le \boldsymbol{d} \le \boldsymbol{n} - \mathbf{1}, \\ 0 & \text{otherwise.} \end{cases}$$

For $\boldsymbol{a} \in \mathbb{N}^c$ we will write $\boldsymbol{x}^{\boldsymbol{a}} = x_1^{a_1} \dots x_c^{a_c}$. Note that the multiplication yields something different if we multiply in another order. In particular we do not have $\boldsymbol{x}^{\boldsymbol{a}} \boldsymbol{x}^{\boldsymbol{b}} = \boldsymbol{x}^{\boldsymbol{a}+\boldsymbol{b}}$. By setting

$$\boldsymbol{q}^{\langle \boldsymbol{a} \mid \boldsymbol{b} \rangle} = \prod_{\substack{i,j \in \{1 \dots c\} \\ i < j}} q_{ij}^{a_j b_i}$$

we obtain the multiplication formula $\boldsymbol{x}^{\boldsymbol{a}} \boldsymbol{x}^{\boldsymbol{b}} = \boldsymbol{q}^{\langle \boldsymbol{a} \mid \boldsymbol{b} \rangle} \boldsymbol{x}^{\boldsymbol{a}+\boldsymbol{b}}$.

***Hochschild (co)homology.*** Let $\Lambda$ be a finite-dimensional algebra. We denote by $\Lambda^{\mathrm{en}} = \Lambda \otimes_k \Lambda^{\mathrm{op}}$ the enveloping algebra. Then $\Lambda^{\mathrm{en}}$-modules are $\Lambda$-$\Lambda$ bimodules on which $k$ acts centrally. In particular $\Lambda$ has a natural structure of a $\Lambda^{\mathrm{en}}$-module. Then

$$\mathrm{HH}^*(\Lambda) = \mathrm{Ext}^*_{\Lambda^{\mathrm{en}}}(\Lambda, \Lambda) \quad \text{and} \quad \mathrm{HH}_*(\Lambda) = \mathrm{Tor}_*^{\Lambda^{\mathrm{en}}}(\Lambda, \Lambda)$$

are the Hochschild cohomology and Hochschild homology of $\Lambda$, respectively. With the Yoneda multiplication of extensions $\mathrm{HH}^*$ becomes a $\mathbb{Z}$-graded $k$-algebra, which is graded commutative [Yoneda 1958].

If $\Lambda$ is graded then so is $\Lambda^{\mathrm{en}}$, and $\Lambda$ is a graded $\Lambda^{\mathrm{en}}$-module. It follows that for any $i \in \mathbb{N}$ the Hochschild homology and cohomology groups $\mathrm{HH}_i(\Lambda)$ and $\mathrm{HH}^i(\Lambda)$ are also graded.

***Projective resolutions.*** In order to determine the Hochschild homology and cohomology of a quantum complete intersection $\Lambda = \Lambda_{\boldsymbol{q}}^{\boldsymbol{n}}$ we need to find a projective resolution of $\Lambda$ as a $\Lambda^{\mathrm{en}}$-module. Moreover we want to keep track of the $\mathbb{Z}^c$-grading, so we will need a graded projective resolution.

We have shown in [Bergh and Oppermann 2008a, Lemma 4.5] that such a graded projective resolution can be found by tensoring together the projective resolutions of $k[x_i]/(x_i^{n_i})$ as $(k[x_i]/(x_i^{n_i}))^{\mathrm{en}}$ modules. To simplify the notation we set $\Lambda_i = k[x_i]/(x_i^{n_i})$. Then the graded projective resolution of $\Lambda_i$ as a bimodule is

$$\mathbb{P}_i : \Lambda_i^{\mathrm{en}} \xleftarrow{x_i\otimes 1 - 1\otimes x_i} \Lambda_i^{\mathrm{en}}\langle 1\rangle \xleftarrow{\sum_{k=0}^{n_i-1} x_i^k\otimes x_i^{n_i-1-k}} \Lambda_i^{\mathrm{en}}\langle n_i\rangle \xleftarrow{x_i\otimes 1 - 1\otimes x_i} \Lambda_i^{\mathrm{en}}\langle n_i+1\rangle \longleftarrow \cdots,$$

where $\Lambda_i^{\mathrm{en}}\langle s\rangle$ is the graded module obtained from $\Lambda_i^{\mathrm{en}}$ by increasing the degree of all homogeneous elements by $s$. Note that here all the bimodules are shifted into place in such a way that all the morphisms have degree 0.

With this notation, the total complex

$$\mathrm{Tot}(\mathbb{P}_1 \otimes_k \mathbb{P}_2 \otimes_k \cdots \otimes_k \mathbb{P}_c)$$

is a graded projective resolution of $\Lambda$.

The term in position $\boldsymbol{p} \in \mathbb{N}^c$ of the $c$-tuple complex $\mathbb{P}_1 \otimes_k \mathbb{P}_2 \otimes_k \cdots \otimes_k \mathbb{P}_c$ is

$$\Lambda_1^{\mathrm{en}}\langle \mathfrak{s}(\boldsymbol{p})_1\rangle \otimes \cdots \otimes \Lambda_c^{\mathrm{en}}\langle \mathfrak{s}(\boldsymbol{p})_c\rangle,$$

where for notational compactness we have defined a function $\mathfrak{s} : \mathbb{Z}^c \to \mathbb{Z}^c$ by

$$\mathfrak{s}(\boldsymbol{p})_i = \begin{cases} \frac{1}{2} p_i n_i & \text{if } p_i \text{ is even,} \\ \frac{1}{2}(p_i - 1)n_i + 1 & \text{if } p_i \text{ is odd.} \end{cases}$$

We will also need a left inverse $\mathfrak{p} : \mathbb{Z}^c \to \mathbb{Z}^c$ of $\mathfrak{s}$ given by

$$\mathfrak{p}(\boldsymbol{s}) = \min\{\boldsymbol{p} \in \mathbb{Z}^c \mid \mathfrak{s}(\boldsymbol{p}) \geq \boldsymbol{s}\}.$$

In the $c$-tuple complex $\mathbb{P}_1 \otimes_k \mathbb{P}_2 \otimes_k \cdots \otimes_k \mathbb{P}_c$ all terms are of the form

$$\Lambda_1^{\mathrm{en}}\langle s_1\rangle \otimes_k \cdots \otimes_k \Lambda_c^{\mathrm{en}}\langle s_c\rangle$$

for some $\boldsymbol{s} \in \mathbb{N}^c$. We recall how these are identified with $\Lambda^{\mathrm{en}}\langle \boldsymbol{s}\rangle$.

**Lemma 2.1** [Bergh and Oppermann 2008a, Lemma 4.3]. *For $\boldsymbol{s} \in \mathbb{Z}^c$ we choose an identification*

$$\Lambda_1^{\mathrm{en}}\langle s_1\rangle \otimes_k \cdots \otimes_k \Lambda_c^{\mathrm{en}}\langle s_c\rangle = \Lambda^{\mathrm{en}}\langle \boldsymbol{s}\rangle$$

*such that* $(1\otimes 1) \otimes \cdots \otimes (1\otimes 1)$ *maps to* $1\otimes 1$. *Under such an identification,*

$$(x_1^{a_1} \otimes x_1^{b_1}) \otimes \cdots \otimes (x_c^{a_c} \otimes x_c^{b_c}) \quad \textit{maps to} \quad \frac{\boldsymbol{q}^{\langle \boldsymbol{s} \mid \boldsymbol{s}\rangle}}{\boldsymbol{q}^{\langle \boldsymbol{a}+\boldsymbol{s} \mid \boldsymbol{b}+\boldsymbol{s}\rangle}} \boldsymbol{x}^{\boldsymbol{a}} \otimes \boldsymbol{x}^{\boldsymbol{b}}.$$

**Remark 2.2.** The differentials occurring in the various directions of the $c$-tuple complex being of particular interest, we note that the identification of Lemma 2.1 maps $(1\otimes 1) \otimes \cdots \otimes (1\otimes 1) \otimes (x_i \otimes 1 - 1 \otimes x_i) \otimes (1\otimes 1) \otimes \cdots \otimes (1\otimes 1)$ to

$$\frac{1}{\boldsymbol{q}^{\langle e_i \mid \boldsymbol{s}\rangle}} x_i \otimes 1 - \frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s} \mid e_i\rangle}} 1 \otimes x_i$$

and $(1\otimes 1)\otimes\cdots\otimes(1\otimes 1)\otimes\left(\sum_{j=0}^{n_i-1} x_i^j\otimes x_i^{n_i-1-j}\right)\otimes(1\otimes 1)\otimes\cdots\otimes(1\otimes 1)$ to

$$\sum_{j=0}^{n_i-1}\frac{1}{\boldsymbol{q}^{\langle je_i|\boldsymbol{s}\rangle}\boldsymbol{q}^{\langle \boldsymbol{s}|(n_i-1-j)e_i\rangle}}x_i^j\otimes x_i^{n_i-1-j}=\sum_{j=0}^{n_i-1}\left(\frac{1}{\boldsymbol{q}^{\langle e_i|\boldsymbol{s}\rangle}}\right)^j\left(\frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s}|e_i\rangle}}\right)^{n_i-1-j}x_i^j\otimes x_i^{n_i-1-j}.$$

***Technical notation.*** We need some definitions to keep the notation short in the rest of the paper.

I. We set $Q=(q_{ij})_{ij}$, and think of $Q$ as a (skew symmetric) matrix with entries in the abelian group $k^\times$. That is, $Q$ represents the morphism of abelian groups

$$Q:\mathbb{Z}^c\to(k^\times)^c,\qquad (d_i)_i\mapsto\left(\prod_{j=1}^{c}q_{ij}^{d_j}\right)_i.$$

As usual for matrices we will denote the image of $\boldsymbol{d}\in\mathbb{Z}^c$ under this map by $Q\boldsymbol{d}$, and its $i$-th component by $(Q\boldsymbol{d})_i$.

For $A,B\subseteq\{1,\dots,c\}$, denote by $Q_{A\times B}$ the submatrix containing only the rows indexed by $A$ and the columns indexed by $B$: that is, the matrix representing the composition

$$\mathbb{Z}^B\hookrightarrow\mathbb{Z}^c\xrightarrow{Q}(k^\times)^c\twoheadrightarrow(k^\times)^A.$$

II. We set $\mathcal{R}_i=\begin{cases}\{\zeta\mid\zeta^{n_i}=1\} & \text{if } \operatorname{char}k \text{ divides } n_i,\\ \{\zeta\mid\zeta^{n_i}=1\text{ and }\zeta\neq 1\} & \text{otherwise.}\end{cases}$

III. For a $\mathbb{Z}$-submodule $K$ of $\mathbb{Z}^a$, denote by $\operatorname{pos.rk}K$ the rank of the $\mathbb{Z}$-submodule $K'$ of $K$ generated by $K\cap\mathbb{N}^a$. For example,

$$\operatorname{pos.rk}\left\langle\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\-1\end{pmatrix}\right\rangle=1.$$

## 3. Hochschild cohomology

For $\boldsymbol{d}\in\mathbb{Z}^c$, we will calculate the degree-$\boldsymbol{d}$ part of the Hochschild cohomology. Then we will obtain the entire Hochschild cohomology by adding up these parts.

To find the degree-$\boldsymbol{d}$ part of the cohomology we first have to understand the set

$$\operatorname{Hom}^{\boldsymbol{d}}_{\Lambda^{\mathrm{en}}}(\Lambda^{\mathrm{en}}\langle\boldsymbol{s}\rangle,\Lambda)$$

of degree-$\boldsymbol{d}$ morphisms from the terms of the projective resolution to $\Lambda$.

**Lemma 3.1.** *The set* $\operatorname{Hom}^{\boldsymbol{d}}_{\Lambda^{\mathrm{en}}}(\Lambda^{\mathrm{en}}\langle\boldsymbol{s}\rangle,\Lambda)$ *is nonzero if and only if* $\boldsymbol{0}\le\boldsymbol{s}+\boldsymbol{d}\le\boldsymbol{n}-\boldsymbol{1}$, *and then it is the one-dimensional $k$-vector space generated by*

$$\varphi^{\boldsymbol{s},\boldsymbol{d}}:\Lambda^{\mathrm{en}}\langle\boldsymbol{s}\rangle\to\Lambda,\qquad \boldsymbol{x}^{\boldsymbol{a}}\otimes\boldsymbol{x}^{\boldsymbol{b}}\mapsto\boldsymbol{q}^{\langle\boldsymbol{a}+\boldsymbol{s}+\boldsymbol{d}\,|\,\boldsymbol{b}+\boldsymbol{s}+\boldsymbol{d}\rangle}\boldsymbol{x}^{\boldsymbol{a}+\boldsymbol{s}+\boldsymbol{d}+\boldsymbol{b}}.$$

*Proof.* Clearly any $\Lambda^{\mathrm{en}}$-homomorphism from $\Lambda^{\mathrm{en}}\langle \boldsymbol{s}\rangle$ to any other module is uniquely determined by the image of $1\otimes 1$. If the morphism is to be of degree $\boldsymbol{d}$, this image can only be a scalar multiple of $\boldsymbol{x}^{\boldsymbol{s}+\boldsymbol{d}}$. We choose the image of $1\otimes 1$ to be $\boldsymbol{q}^{\langle \boldsymbol{s}+\boldsymbol{d}\,|\,\boldsymbol{s}+\boldsymbol{d}\rangle}\boldsymbol{x}^{\boldsymbol{s}+\boldsymbol{d}}$ and obtain the formula of the lemma by extending $\Lambda^{\mathrm{en}}$-linearly. $\square$

**Corollary 3.2.** $\dim \operatorname{Hom}^{\boldsymbol{d}}_{\Lambda^{\mathrm{en}}}\big(\Lambda^{\mathrm{en}}\langle \mathfrak{s}(\boldsymbol{p})\rangle,\,\Lambda\big) = \begin{cases} 1 & \text{\textit{if} } \mathfrak{p}(-\boldsymbol{d}) \le \boldsymbol{p} \le \mathfrak{p}(-\boldsymbol{d})+\boldsymbol{1}, \\ 0 & \textit{otherwise.}\end{cases}$

This means that for $\boldsymbol{d}\le \boldsymbol{n}-\boldsymbol{1}$ the $c$-tuple complex $\operatorname{Hom}^{\boldsymbol{d}}_{\Lambda^{\mathrm{en}}}(\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_c,\,\Lambda)$ is concentrated in a cube of sides 1 or 0 (the latter case occurring in directions $i$ with $\mathfrak{p}(-\boldsymbol{d})_i=-1$, i.e., $d_i=n_i-1$), where there is a one-dimensional space in each corner of the cube.

Since by Remark 2.2 these are the terms occurring in the projective resolution, we are in particular interested in what the maps $\varphi^{\boldsymbol{s},\boldsymbol{d}}$ of Lemma 3.1 do to terms of the form

$$\frac{1}{\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{s}\rangle}}x_i\otimes 1-\frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s}\,|\,e_i\rangle}}1\otimes x_i \quad\text{and}\quad \sum_{j=0}^{n_i-1}\Big(\frac{1}{\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{s}\rangle}}\Big)^{j}\Big(\frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s}\,|\,e_i\rangle}}\Big)^{n_i-1-j}x_i^{j}\otimes x_i^{n_i-1-j}.$$

**Lemma 3.3.** *Let* $\boldsymbol{s}$ *and* $\boldsymbol{d}$ *be such that* $\boldsymbol{0}\le\boldsymbol{s}+\boldsymbol{d}\le\boldsymbol{n}-\boldsymbol{1}$, *and let* $i\in\{1\dots c\}$.

(1) *Assume that* $s_i+d_i+1<n_i$. *Then*

$$\varphi^{\boldsymbol{s},\boldsymbol{d}}\bigg(\frac{1}{\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{s}\rangle}}x_i\otimes 1-\frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s}\,|\,e_i\rangle}}1\otimes x_i\bigg)=0$$

*if and only if* $(Q\boldsymbol{d})_i=1$. (For the definition of $Q$ see Technical notation I.)

(2) *Assume that* $s_i+d_i=0$. *Then*

$$\varphi^{\boldsymbol{s},\boldsymbol{d}}\bigg(\sum_{j=0}^{n_i-1}\Big(\frac{1}{\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{s}\rangle}}\Big)^{j}\Big(\frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s}\,|\,e_i\rangle}}\Big)^{n_i-1-j}x_i^{j}\otimes x_i^{n_i-1-j}\bigg)=0$$

*if and only if* $(Q\boldsymbol{d})_i\in\mathcal{R}_i$. (For the definition of $Q$ see Technical notation II.)

*Proof.* We only prove (2). The proof of (1) is a similar and simpler calculation using Lemma 3.1. By that lemma we have

$$\begin{aligned}\varphi^{\boldsymbol{s},\boldsymbol{d}}\bigg(&\sum_{j=0}^{n_i-1}\Big(\frac{1}{\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{s}\rangle}}\Big)^{j}\Big(\frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s}\,|\,e_i\rangle}}\Big)^{n_i-1-j}x_i^{j}\otimes x_i^{n_i-1-j}\bigg)\\ &=\sum_{j=0}^{n_i-1}\Big(\frac{1}{\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{s}\rangle}}\Big)^{j}\Big(\frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s}\,|\,e_i\rangle}}\Big)^{n_i-1-j}\boldsymbol{q}^{\langle je_i+\boldsymbol{s}+\boldsymbol{d}\,|\,(n_i-1-j)e_i+\boldsymbol{s}+\boldsymbol{d}\rangle}\boldsymbol{x}^{\boldsymbol{s}+\boldsymbol{d}+(n_i-1)e_i}.\end{aligned}$$

Upon rearrangement of the right-hand side this becomes

$$\boldsymbol{q}^{\langle \boldsymbol{s}+\boldsymbol{d}\,|\,\boldsymbol{s}+\boldsymbol{d}\rangle} \sum_{j=0}^{n_i-1} (\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{d}\rangle})^j (\boldsymbol{q}^{\langle \boldsymbol{d}\,|\,e_i\rangle})^{n_i-1-j} \boldsymbol{x}^{\boldsymbol{s}+\boldsymbol{d}+(n_i-1)e_i}$$

$$= \underbrace{\boldsymbol{q}^{\langle \boldsymbol{s}+\boldsymbol{d}\,|\,\boldsymbol{s}+\boldsymbol{d}\rangle}}_{\neq 0} \boldsymbol{x}^{\boldsymbol{s}+\boldsymbol{d}+(n_i-1)e_i} \cdot \begin{cases} n_i\, \overbrace{(\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{d}\rangle})^{n_i-1}}^{\neq 0} & \text{if } \boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{d}\rangle} = \boldsymbol{q}^{\langle \boldsymbol{d}\,|\,e_i\rangle}, \\ \dfrac{(\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{d}\rangle})^{n_i} - (\boldsymbol{q}^{\langle \boldsymbol{d}\,|\,e_i\rangle})^{n_i}}{\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{d}\rangle} - \boldsymbol{q}^{\langle \boldsymbol{d}\,|\,e_i\rangle}} & \text{otherwise.} \end{cases}$$

Now the claim follows from the fact that $\dfrac{\boldsymbol{q}^{\langle \boldsymbol{d}\,|\,e_i\rangle}}{\boldsymbol{q}^{\langle e_i\,|\,\boldsymbol{d}\rangle}} = \prod_{j=1}^{c} q_{ij}^{d_j} = (Q\boldsymbol{d})_i$. □

We have shown that the vanishing of the maps on the edges in direction $i$ of the cube $\operatorname{Hom}^{\boldsymbol{d}}_{\Lambda^{\mathrm{en}}}(\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_c, \Lambda)$ does not depend on $\boldsymbol{s}$, that is, if one edge in direction $i$ vanishes, then all vanish. Also, if all the edges in one direction are isomorphisms, the total complex is acyclic. Hence, if we partition the set $\{1, \ldots, c\}$ into subsets

$$\begin{aligned} I_{\max} &= \bigl\{ i \in \{1, \ldots, c\} : n_i = d_i + 1 \bigr\}, \\ I_1 &= \bigl\{ i \in \{1, \ldots, c\} : n_i \mid d_i + 1 \bigr\} \setminus I_{\max}, \\ I_2 &= \bigl\{ i \in \{1, \ldots, c\} : n_i \nmid d_i + 1 \bigr\}, \end{aligned}$$

we have shown the following:

**Theorem 3.4.** *Let $\Lambda = \Lambda^{\boldsymbol{n}}_{\boldsymbol{q}}$ be a quantum complete intersection, and let $\boldsymbol{d} \le \boldsymbol{n} - \boldsymbol{1}$. Then $\mathrm{HH}^{*,\boldsymbol{d}}(\Lambda) \neq 0$ if and only if*

- $(Q\boldsymbol{d})_i \in \mathfrak{R}_i$ *for all* $i \in I_1$, *and*
- $(Q\boldsymbol{d})_i = 1$ *for all* $i \in I_2$.

*In this situation $\mathrm{HH}^{*,\boldsymbol{d}}(\Lambda)$ has the $k$-vector space basis*

$$\bigl\{ E^{\boldsymbol{d}}_{\boldsymbol{p}} \mid \boldsymbol{0} \le \boldsymbol{p} \text{ and } \mathfrak{p}(-\boldsymbol{d}) \le \boldsymbol{p} \le \mathfrak{p}(-\boldsymbol{d}) + \boldsymbol{1} \bigr\}, \tag{$*$}$$

*where $E^{\boldsymbol{d}}_{\boldsymbol{p}}$ is represented by the (degree-$\boldsymbol{d}$) map from the $c$-tuple complex*

$$\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_c$$

*to $\Lambda$ (shifted to position $\boldsymbol{p}$) sending $1 \otimes 1$ to $\boldsymbol{x}^{\boldsymbol{d}+\mathfrak{s}(\boldsymbol{p})}$ in position $\boldsymbol{p}$. In particular,*

$$E^{\boldsymbol{p}}_{\boldsymbol{d}} \text{ has extension degree } \sum_{i=1}^{c} p_i.$$

The assumptions on $\boldsymbol{p}$ in ($*$) just ensure that $\boldsymbol{0} \le \boldsymbol{d} + \mathfrak{s}(\boldsymbol{p}) \le \boldsymbol{n} - \boldsymbol{1}$: in other words, we are in the cube where $\operatorname{Hom}^{\boldsymbol{d}}_{\Lambda^{\mathrm{en}}}(\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_c, \Lambda)$ does not vanish.

Now let us compare this result to the description of $\operatorname{Ext}^*_\Lambda(k,k)$ obtained in [Bergh and Oppermann 2008a]. More precisely: tensoring over $\Lambda$ with $k$ yields a map from the Hochschild cohomology to the Ext-algebra of the $\Lambda$-module $k$. Our aim now is to determine its image. By Theorem 5.3 of that reference, the latter ring has the form

$$\operatorname{Ext}^*_\Lambda(k,k) = \frac{k\langle y_1,\dots,y_c,z_1,\dots,z_c\rangle}{\begin{pmatrix} y_jy_i+q_{ij}y_iy_j & \text{for } i\neq j \\ y_jz_i-q_{ij}^{n_i}z_iy_j & \text{for all } i,j \\ z_jz_i-q_{ij}^{n_in_j}z_iz_j & \text{for all } i,j \\ y_i^2-z_i & \text{if } n_i=2 \\ y_i^2 & \text{if } n_i\neq 2\end{pmatrix}},$$

where $|y_i|=(1,-e_i)$ and $|z_i|=(2,-n_ie_i)$, and where the quotient is by the ideal generated by the polynomials indicated.

**Corollary 3.5.** *The image of the map* $(-\otimes_\Lambda k)_*:\mathrm{HH}^*(\Lambda)\to\operatorname{Ext}^*_\Lambda(k,k)$ *is*

$$\bigoplus_{\boldsymbol{d}\in\mathfrak{D}}\operatorname{Ext}^{*,\boldsymbol{d}}(k,k), \tag{†}$$

*where the sum runs over graded pieces where the corresponding graded piece of the Hochschild cohomology does not vanish*:

$$\mathfrak{D}=\big\{\boldsymbol{d}\in\mathbb{Z}^c:(Q\boldsymbol{d})_i\in\mathfrak{R}_i \textit{ for } i\in I_{\max}\cup I_1,\ (Q\boldsymbol{d})_i=1 \textit{ for } i\in I_2\big\}.$$

*Proof.* By construction the image cannot be bigger than the sum in (†). To see that any $\operatorname{Ext}^{*,\boldsymbol{d}}(k,k)$ with $\boldsymbol{d}\in\mathfrak{D}$ is contained in the image, first note that

$$\dim_k\operatorname{Ext}^{*,\boldsymbol{d}}(k,k)=\begin{cases}1 & \text{if } \forall i: d_i\leq 0 \text{ and } n_i\mid d_i\vee n_i\mid d_i+1,\\ 0 & \text{otherwise.}\end{cases}$$

The condition for $\operatorname{Ext}^{*,\boldsymbol{d}}(k,k)$ not vanishing is equivalent to asking that $\boldsymbol{d}=-\mathfrak{s}(\boldsymbol{p})$ for some $\boldsymbol{p}\in\mathbb{N}^c$. By definition, $E_{\boldsymbol{p}}^{\boldsymbol{d}}$ is represented by a map sending $1\otimes 1$ to 1 in position $\boldsymbol{p}$, and hence it does not vanish when being tensored over $\Lambda$ by $k$. Therefore the image is at least one-dimensional in degree $\boldsymbol{d}$. □

## 4. The rate of growth of the Hochschild cohomology

In this section we study how big the Hochschild cohomology of a finite-dimensional quantum complete intersection is. Our way to measure for measuring the size is the rate of growth, as explained in the following definition.

**Definition 4.1.** Let $X = \coprod_{i=0}^{\infty} X_i$ be an $\mathbb{N}$-graded $k$-module such that the $X_i$ have finite $k$-dimension. The rate of growth of $X$, denoted by $\gamma(X)$, is defined as

$$\gamma(X) = \inf\{t \in \mathbb{N} \mid \exists a \in \mathbb{N} \text{ such that } \forall i : \dim_k X_i \leq a i^t\}.$$

If $X$ is a graded commutative ring that is finitely generated over $k$, then $\gamma(X) = \operatorname{Krull.dim} X$. However this assumption is not always satisfied for the Hochschild cohomology ring of quantum complete intersections (see Sections 5 and 6).

We first decompose the Hochschild cohomology as follows:

**Construction 4.2.** For $G \subseteq \{1, \dots, c\}$ we denote by $\mathrm{HH}_G^*$ the $k$-span of the $E_{\boldsymbol{p}}^{\boldsymbol{d}}$ for which $\boldsymbol{d}$ and $\boldsymbol{n}$ satisfy the condition

$$G = \Big\{ i \in \{1, \dots, c\} \;\Big|\; d_i < n_i - 1 \text{ or } (Q\boldsymbol{d})_i \in \mathcal{R}_i \Big\}.$$

That is, we take all those $E_{\boldsymbol{p}}^{\boldsymbol{d}}$ from Theorem 3.4 such that $G$ contains exactly the indices not in $I_{\max}$ plus those in $I_{\max}$ that fulfill the requirements for elements of $I_1$.

Clearly this yields a decomposition $\mathrm{HH}^*(\Lambda) = \bigoplus_{G \subseteq \{1,\dots,c\}} \mathrm{HH}_G^*$, and hence

$$\gamma\big(\mathrm{HH}^*(\Lambda)\big) = \max_{G \subseteq \{1,\dots,c\}} \gamma(\mathrm{HH}_G^*).$$

**Proposition 4.3.** *For $G \subseteq \{1, \dots, c\}$ the rate of growth of $\mathrm{HH}_G^*$ is*

$$\gamma(\mathrm{HH}_G^*) = \begin{cases} 0 & \textit{if } \mathrm{HH}_G^* = 0, \\ \operatorname{pos.rk} \operatorname{Ker} Q_{G \times G} & \textit{otherwise.} \end{cases}$$

(For the definition of pos.rk see Technical notation III.) *In particular we always have $\gamma(\mathrm{HH}_G^*) \leq |G|$.*

For the proof we will need the following observation.

**Observation 4.4.** *Let $K \leq \mathbb{Z}^a$ be a submodule. The $k$-module with basis $K \cap \mathbb{N}^a$ is $\mathbb{Z}$-graded by $|\boldsymbol{x}| = \sum_{i=1}^{a} x_i$ for $\boldsymbol{x} \in K$. With this grading, its rate of growth is $\gamma(k(K \cap \mathbb{N}^a)) = \operatorname{pos.rk} K$.*

*Proof of Proposition 4.3.* By construction, $\mathrm{HH}_G^*$ has the $k$-basis

$$\Big\{ E_{\boldsymbol{p}}^{\boldsymbol{d}} \mid \boldsymbol{p} \geq \boldsymbol{0},\ \boldsymbol{d} \leq \boldsymbol{n} - \boldsymbol{1},\ \mathfrak{p}(-\boldsymbol{d}) \leq \boldsymbol{p} \leq \mathfrak{p}(-\boldsymbol{d}) + \boldsymbol{1},\ d_i = n_i - 1 \text{ for } i \notin G,$$
$$(Q\boldsymbol{d})_i \notin \mathcal{R}_i \text{ for } i \notin G,\ (Q\boldsymbol{d})_i \in \mathcal{R}_i \text{ for } i \in G \setminus I_2,\ (Q\boldsymbol{d})_i = 1 \text{ for } i \in G \cap I_2 \Big\},$$

and the extension degree of $E_{\boldsymbol{p}}^{\boldsymbol{d}}$ is $\sum_{i=1}^{c} p_i$.

Note that the map $\mathfrak{p}$ is linear up to some rounding. Hence we may calculate the rate of growth with respect to the grading given by $-\sum_{i=1}^{c} d_i$.

Since for any $\boldsymbol{d}$ there are at least one and at most $2^c$ values of $\boldsymbol{p}$ satisfying the conditions of the set above, we may disregard the number of different choices for $\boldsymbol{p}$ for a given $\boldsymbol{d}$.

Next, since $\boldsymbol{d}$ is fixed outside $G$, we may restrict our attention to the $G$ part of the indices. Write $\overline{G} = \{1, \ldots, c\} \setminus G$ and define, for $G \subseteq G' \subseteq \{1, \ldots, c\}$ the set

$$\mathcal{B}_{G'} = \left\{ \boldsymbol{d}_G \in \mathbb{Z}^G \mid \boldsymbol{d}_G \le \boldsymbol{n}_G - \boldsymbol{1},\ Q_{\{i\}\times G}\boldsymbol{d}_G \cdot Q_{\{i\}\times\overline{G}}(\boldsymbol{n}_{\overline{G}} - \boldsymbol{1}) \begin{array}{ll} \notin \mathcal{R}_i & \text{for } i \in G' \setminus G, \\ \in \mathcal{R}_i & \text{for } i \in G \setminus I_2, \\ = 1 & \text{for } i \in I_2 \end{array} \right\}.$$

We need to understand the rate of growth of the $k$-module with basis $\mathcal{B}_{\{1,\ldots,c\}}$. Note that $\mathcal{B}_{\{1,\ldots,c\}} \subseteq \mathcal{B}_G$ (more generally, for $G' \subseteq G''$ we have $\mathcal{B}_{G'} \supseteq \mathcal{B}_{G''}$).

Now $\mathcal{B}_G$ is invariant under adding elements of the set

$$-\Big(\prod_{i\in G} n_i\mathbb{N}\Big) \cap \operatorname{Ker} Q_{G\times G},$$

and contains only finitely many elements which are not obtained from another element by such an addition. Hence, if $\mathcal{B}_G$ is nonempty, the rate of growth of the $k$-module with basis $\mathcal{B}_G$ is identical to the rate of growth of the $k$-module with basis $\mathbb{N}^G \cap \operatorname{Ker} Q_{G\times G}$, which, by Observation 4.4, is $\operatorname{pos.rk} \operatorname{Ker} Q_{G\times G}$.

It follows that $\gamma(\mathrm{HH}^*_G) \le \operatorname{pos.rk} \operatorname{Ker} Q_{G\times G}$.

Now we let $\widehat{G}$ be maximal with $G \subseteq \widehat{G} \subseteq \{1, \ldots, c\}$ such that $\operatorname{pos.rk} \operatorname{Ker} Q_{\widehat{G}\times G} = \operatorname{pos.rk} \operatorname{Ker} Q_{G\times G}$. It follows, as in the discussion above, that if $\mathcal{B}_{\widehat{G}} \neq \varnothing$ then the rate of growth of the $k$-module with basis $\mathcal{B}_{\widehat{G}}$ is $\operatorname{pos.rk} \operatorname{Ker} Q_{G\times G}$.

Finally let $i \notin \widehat{G}$. Using arguments similar to the foregoing, one sees that the rate of growth of the free module with basis $\mathcal{B}_G \setminus \mathcal{B}_{G\cup\{i\}}$ is strictly smaller than $\operatorname{pos.rk} \operatorname{Ker} Q_{G\times G}$.

Since

$$\mathcal{B}_{\{1,\ldots,c\}} = \mathcal{B}_{\widehat{G}} \setminus \Big( \bigcup_{i\notin\widehat{G}} \big(\mathcal{B}_G \setminus \mathcal{B}_{G\cup\{i\}}\big)\Big),$$

it follows that, provided $\mathcal{B}_{\{1,\ldots,c\}} \neq \varnothing$, the rate of growth of the $k$-module with basis $\mathcal{B}_{\{1,\ldots,c\}}$ is $\operatorname{pos.rk} \operatorname{Ker} Q_{G\times G}$. □

Summing up the results for $\mathrm{HH}^*_G$, we have shown:

**Theorem 4.5.** *The rate of growth of the Hochschild cohomology of a finite-dimensional quantum complete intersection is the maximum of* $\operatorname{pos.rk} \operatorname{Ker} Q_{G\times G}$ *over*

$$G = \big\{ i \in \{1, \ldots, c\} \mid d_i < n_i - 1 \text{ or } (Q\boldsymbol{d})_i \in \mathcal{R}_i \big\},$$

*where* $\boldsymbol{d}$ *ranges over elements of* $\mathbb{Z}^c$ *such that* $\boldsymbol{d} \le \boldsymbol{n} - \boldsymbol{1}$, $(Q\boldsymbol{d})_i \in \mathcal{R}_i$ *for all* $i$ *with* $n_i \mid d_i + 1$ *and* $d_i < 0$, *and* $(Q\boldsymbol{d})_i = 1$ *for all* $i$ *with* $n_i \nmid d_i + 1$.

**Corollary 4.6.** *For a finite quantum complete intersection either all* $q_{ij}$ *are roots of unity, or the rate of growth of the Hochschild cohomology is at most* $c - 2$.

*Proof.* Assume not all $q_{ij}$ are roots of unity. Then $\operatorname{rk} \operatorname{Ker} Q \le c-2$, since $Q$ is skew symmetric. Hence $\operatorname{pos.rk} \operatorname{Ker} Q \le c - 2$. Now we consider $G$ with $|G| = c - 1$, that is, $G = \{1, \dots, c\} \setminus \{h\}$ for some $h$. If $\operatorname{rk}\operatorname{Ker} Q_{G\times G} \le c-2$ there is nothing to show, so assume $Q_{G\times G}$ only contains roots of unity. Since $Q$ does not only contain roots of unity there is $i \in G$ such that $q_{ih}$ is not a root of unity. But then $(Q\boldsymbol{d})_i$ cannot be a root of unity for any $\boldsymbol{d} \in \mathbb{Z}^c$ with $d_h = n_h - 1 \neq 0$. Hence this $G$ is not to be considered in the maximum of Theorem 4.5. □

## 5. On the multiplicative structure of the Hochschild cohomology

In this section we will identify a subring $\mathcal{S}$ of the Hochschild cohomology ring, which is a finitely generated commutative $k$-algebra without zero divisors, and is isomorphic to the Hochschild cohomology modulo nilpotent objects. We will completely describe $\mathcal{S}$, determine its Krull dimension, and determine when the entire Hochschild cohomology ring is finitely generated as a module over $\mathcal{S}$.

By Theorem 3.4 we know that the Hochschild cohomology has a $k$-vector space basis

$$\begin{aligned}\big\{E_{\boldsymbol{p}}^{\boldsymbol{d}} \mid \boldsymbol{p} \ge \boldsymbol{0},\ &\mathfrak{p}(-\boldsymbol{d}) \le \boldsymbol{p} \le \mathfrak{p}(-\boldsymbol{d}) + \boldsymbol{1},\\ &(Q\boldsymbol{d})_i \in \mathcal{R}_i \text{ for all } i \text{ with } n_i \mid d_i+1 > 0,\ (Q\boldsymbol{d})_i = 1 \text{ for all } i \text{ with } n_i \nmid d_i+1.\big\}\end{aligned}$$

For simplicity of notation we set $E_{\boldsymbol{p}}^{\boldsymbol{d}} = 0$ whenever $\boldsymbol{d}$ and $\boldsymbol{p}$ do not satisfy the conditions above. Then we always have

$$E_{\boldsymbol{p}}^{\boldsymbol{d}} \cdot E_{\boldsymbol{p}'}^{\boldsymbol{d}'} \in k E_{\boldsymbol{p}+\boldsymbol{p}'}^{\boldsymbol{d}+\boldsymbol{d}'}.$$

**Lemma 5.1.** *Assume $\mathfrak{s}(\boldsymbol{p}) \neq -\boldsymbol{d}$. Then $E_{\boldsymbol{p}}^{\boldsymbol{d}}$ is nilpotent.*

*Proof.* Let $i$ be such that $\mathfrak{s}(\boldsymbol{p})_i > -d_i$. Then

$$\mathfrak{s}(n_i\boldsymbol{p})_i \ge n_i \mathfrak{s}(\boldsymbol{p})_i \ge n_i(1-d_i) \ge n_i - n_i d_i,$$

and hence $(E_{\boldsymbol{p}}^{\boldsymbol{d}})^{n_i} \in k E_{n_i\boldsymbol{p}}^{n_i\boldsymbol{d}} = 0$. □

We are particularly interested in the nonnilpotent elements of the Hochschild cohomology ring. For simplicity of notation, we give the remaining candidates a new name:

$$s_{\boldsymbol{p}} := E_{\boldsymbol{p}}^{-\mathfrak{s}(\boldsymbol{p})}.$$

**Lemma 5.2.** *Let $\boldsymbol{p} \in \mathbb{N}^c$ such that there is $i \in \{1, \dots, c\}$ with $n_i > 2$ and $p_i$ is odd. Then $s_{\boldsymbol{p}}$ is nilpotent.*

*Proof.* A straightforward calculation shows that $(s_{\boldsymbol{p}})^2$ satisfies the assumption of Lemma 5.1. □

Now we set

$$\begin{aligned}\mathcal{S} = {}_k\big\langle s_{\boldsymbol{p}} \mid (Q\mathfrak{s}(\boldsymbol{p}))_i = 1 &\text{ for all } i \text{ with } p_i \text{ even},\\ &(Q\mathfrak{s}(\boldsymbol{p}))_i = -1 \text{ and } n_i = 2 \text{ for all } i \text{ with } p_i \text{ odd}\big\rangle.\end{aligned}$$

By the preceding two lemmas, the composition $\mathcal{S} \hookrightarrow \mathrm{HH}^*(\Lambda) \twoheadrightarrow \dfrac{\mathrm{HH}^*(\Lambda)}{(\text{nilpotence})}$ is onto.

Our next aim is to understand how the elements of $\mathcal{S}$ are multiplied with each other and with the other $E_{\boldsymbol{p}}^{\boldsymbol{d}}$. To do so we lift the map representing $s_{\boldsymbol{p}}$, with $\boldsymbol{p}$ as in the definition of $\mathcal{S}$, to a map of $c$-tuple complexes.

**Lemma 5.3.** *The element $s_{\boldsymbol{p}}$ with $\boldsymbol{p}$ as in the definition of $\mathcal{S}$ is represented by the map of $c$-tuple complexes $\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_c \to (\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_c)[\boldsymbol{p}]$ that sends $1 \otimes 1$ to $(1/\boldsymbol{q}^{\langle \mathfrak{s}(\boldsymbol{r}) \mid \mathfrak{s}(\boldsymbol{p}) \rangle}) \cdot 1 \otimes 1$ in position $\boldsymbol{p}+\boldsymbol{r}$.*

*Proof.* It suffices to verify that the map given in the lemma is a map of $c$-tuple complexes, since then it clearly does the right thing in position $\boldsymbol{p}$. This amounts to checking that the various squares commute — a straightforward, if somewhat tiresome, calculation involving four different cases, according to the parities of the $p_i$ and $r_i$. □

Note that when passing from $c$-tuple complexes to their total complexes some maps need to be multiplied by $-1$. One choice for doing so is to multiply the map in direction $i$ from position $\boldsymbol{p}+e_i$ to $\boldsymbol{p}$ by $\prod_{j<i}(-1)^{p_i}$. With this convention we have the following immediate consequence of Lemma 5.3.

**Corollary 5.4.** *We have $s_{\boldsymbol{p}} E_{\boldsymbol{p}'}^{\boldsymbol{d}} = \dfrac{\prod_{j<i}(-1)^{p_j p'_i}}{\boldsymbol{q}^{\langle \mathfrak{s}(\boldsymbol{p}') \mid \mathfrak{s}(\boldsymbol{p}) \rangle}} E_{\boldsymbol{p}+\boldsymbol{p}'}^{\boldsymbol{d}-\mathfrak{s}(\boldsymbol{p})}$. In particular,*

$$s_{\boldsymbol{p}} s_{\boldsymbol{p}'} = \frac{\prod_{j<i}(-1)^{p_j p'_i}}{\boldsymbol{q}^{\langle \mathfrak{s}(\boldsymbol{p}') \mid \mathfrak{s}(\boldsymbol{p}) \rangle}} s_{\boldsymbol{p}+\boldsymbol{p}'}. \tag{‡}$$

From these results we obtain:

**Theorem 5.5.** *The Hochschild cohomology ring of a quantum complete intersection has a subring $\mathcal{S}$ isomorphic to*

$$_k\Big\langle y_1^{p_1 n_1/2} \dots y_c^{p_c n_c/2} \in k[y_1, \dots y_c] \Big| (Q\mathfrak{s}(\boldsymbol{p}))_i = 1 \textit{ for all } i \textit{ with } p_i \textit{ even}, \\ (Q\mathfrak{s}(\boldsymbol{p}))_i = -1 \textit{ and } n_i = 2 \textit{ for all } i \textit{ with } p_i \textit{ odd} \Big\rangle.$$

*In particular, $\mathcal{S}$ is a finitely generated $k$-algebra without zero-divisors.*

*Moreover the composition*

$$\mathcal{S} \hookrightarrow \mathrm{HH}^*(\Lambda) \twoheadrightarrow \frac{\mathrm{HH}^*(\Lambda)}{(\textit{nilpotence})}$$

*is an isomorphism. Hence $\dfrac{\mathrm{HH}^*(\Lambda)}{(\textit{nilpotence})}$ is a split quotient of $\mathrm{HH}^*(\Lambda)$ and is isomorphic to $\mathcal{S}$.*

*Proof.* That the $s_{\boldsymbol{p}}$ commute can be checked directly, using (‡) in Corollary 5.4. Alternatively, note that, since (‡) implies $s_{\boldsymbol{p}}^2 \neq 0$, either $s_{\boldsymbol{p}}$ lies in the even part of the Hochschild cohomology or char $k = 2$. In both cases it follows from general theory that $s_{\boldsymbol{p}}$ lies in the center of the Hochschild cohomology ring.

Thus $\mathcal{S}$ has the form described in the theorem.

To see that $\mathcal{S}$ is finitely generated as a $k$-algebra we partially order the set $\{y_1^{p_1 n_1/2} \dots y_c^{p_c n_c/2} \in \mathcal{S}\}$ by comparing the exponents componentwise. Since the ideal in $k[y_1, \dots, y_c]$ generated by this set is finitely generated, it follows there are only finitely many minimal elements with respect to this partial order. We claim that they generate $\mathcal{S}$ as a $k$-algebra. Assume that $y_1^{p_1 n_1/2} \dots y_c^{p_c n_c/2} \in \mathcal{S}$ is not minimal. Then one easily sees that $y_1^{p_1 n_1/2} \dots y_c^{p_c n_c/2}$ is the product of two smaller elements of this form (for instance, one of them could be chosen minimal). Iterating we see that any $y_1^{p_1 n_1/2} \dots y_c^{p_c n_c/2} \in \mathcal{S}$ is a product of minimal ones.

The final part of the theorem follows from the comment at the top of page 832 and from Corollary 5.4. $\square$

We conclude this section by giving a precise criterion for when the entire Hochschild cohomology ring is finitely generated over $\mathcal{S}$.

**Lemma 5.6.** *The decomposition* $\mathrm{HH}^*(\Lambda) = \bigoplus_{G \subseteq \{1,\dots,c\}} \mathrm{HH}^*_G$ *of Construction 4.2 respects the $\mathcal{S}$-module structure.*

*Proof.* This follows immediately from the definition of $\mathrm{HH}^*_G$ and the multiplication formula in Corollary 5.4. $\square$

**Proposition 5.7.** *The module $\mathrm{HH}^*_G$ is finitely generated over $\mathcal{S}$ if and only if it is zero or* $\text{pos.rk Ker } Q_{G \times G} = \text{pos.rk Ker } Q_{\{1,\dots,c\} \times G}$.

*Proof.* Clearly we may assume $\mathrm{HH}^*_G \neq 0$. Note that the $s_{\boldsymbol{p}}$ with $p_i \neq 0$ for some $i \in \{1, \dots, c\} \setminus G$ annihilate $\mathrm{HH}^*_G$, and hence $\mathrm{HH}^*_G \neq 0$ is actually a module over the split quotient $\mathcal{S}_G := {}_k\langle s_{\boldsymbol{p}} \in \mathcal{S} \mid \forall i : i \in G \vee p_i = 0\rangle$. By Observation 4.4, $\gamma(\mathcal{S}_G) = \text{pos.rk Ker } Q_{\{1,\dots,c\} \times G}$.

Moreover $\mathcal{S}_G$ acts on $\mathrm{HH}^*_G$ without zero-divisors: Since both $\mathcal{S}_G$ and $\mathrm{HH}^*_G$ are $\mathbb{Z}^c$-graded it suffices to look at graded parts. For these, this is immediate from the multiplication formula in Corollary 5.4. The claim follows. $\square$

**Corollary 5.8.** *For any finite-dimensional quantum complete intersection $\mathrm{HH}^*_{\{1,\dots,c\}}$ is a finitely generated $\mathcal{S}$-module.*

**Theorem 5.9.** *The Hochschild cohomology ring is finitely generated as a module over $\mathcal{S}$ if and only if* $\text{pos.rk Ker } Q_{G \times G} = \text{pos.rk Ker } Q_{\{1,\dots,c\} \times G}$ *for any subset $G \subseteq \{1, \dots, c\}$ for which there exists $\boldsymbol{d} \in \mathbb{Z}^c$ satisfying*

- $\boldsymbol{d} \leq \boldsymbol{n} - \mathbf{1}$,
- $d_i = n_i - 1$ *for all $i \in \{1, \dots, c\} \setminus G$,*
- $(Q\boldsymbol{d})_i \in \mathcal{R}_i$ *for all $i \in G$ with $n_i \mid d_i + 1$,*
- $(Q\boldsymbol{d})_i = 1$ *for all $i \in G$ with $n_i \nmid d_i + 1$.*

## 6. Examples

**Example 6.1** [Bergh and Erdmann 2008]. Let $\Lambda = \Lambda_{q_{12}}^{n_1,n_2}$ be a codimension-2 quantum complete intersection such that $q_{12}$ is not a root of unity. Take $\boldsymbol{d} = (d_1, d_2) \leq (n_1-1, n_2-1)$. Then $\mathrm{HH}^{*,\boldsymbol{d}}(\Lambda)$ does not vanish if and only if for any $i \in \{1, 2\}$ we have $d_i = n_i - 1$ or $n_i \nmid d_i + 1$ and $q_{12}^{d_{3-i}} = 1$. Since $q_{12}$ is not a root of unity this means that for any $i$ with $d_i < n_i - 1$ we have $d_{3-i} = 0$. Therefore the only $\boldsymbol{d}$ that contribute to the Hochschild cohomology are $(n_1-1, n_2-1)$ and $(0, 0)$. For $\boldsymbol{d} = (n_1-1, n_2-1)$ we obtain $I_{\max} = \{1, 2\}$, $I_1 = \varnothing$, $I_2 = \varnothing$, and $\mathfrak{p}(-\boldsymbol{d}) = (-1, -1)$. Hence

$$\mathrm{HH}^{*,\boldsymbol{d}}(\Lambda) = {}_k\big\langle E_{(0,0)}^{(n_1-1,n_2-1)}\big\rangle.$$

For $\boldsymbol{d} = (0, 0)$ we obtain $I_{\max} = \varnothing$, $I_1 = \varnothing$, $I_2 = \{1, 2\}$, and $\mathfrak{p}(-\boldsymbol{d}) = (0, 0)$. Hence

$$\mathrm{HH}^{*,\boldsymbol{d}}(\Lambda) = {}_k\big\langle E_{(0,0)}^{(0,0)}, E_{(0,1)}^{(0,0)}, E_{(1,0)}^{(0,0)}, E_{(1,1)}^{(0,0)}\big\rangle.$$

Summing up we obtain

$$\mathrm{HH}^{*,\boldsymbol{d}}(\Lambda) = {}_k\big\langle E_{(0,0)}^{(n_1-1,n_2-1)}, E_{(0,0)}^{(0,0)}, E_{(0,1)}^{(0,0)}, E_{(1,0)}^{(0,0)}, E_{(1,1)}^{(0,0)}\big\rangle,$$

and hence

$$\dim \mathrm{HH}^*(\Lambda) = (2, 2, 1, 0, \dots).$$

We generalize this example to arbitrary codimensions:

**Example 6.2.** Let $c \geq 2$ and assume the $q_{ij}$ are generic, meaning that $(Q\boldsymbol{d})_i$ is a root of unity only if $d_j = 0$ for all $j \neq i$. Then $\mathrm{HH}^{*,\boldsymbol{d}}(\Lambda) \neq 0$ only for $\boldsymbol{d} = \boldsymbol{n} - \mathbf{1}$ or $\boldsymbol{d} = \mathbf{0}$. Similarly to Example 6.1 we obtain

$$\mathrm{HH}^{*,\boldsymbol{n}-\mathbf{1}}(\Lambda) = {}_k\langle E_{\mathbf{0}}^{\boldsymbol{n}-\mathbf{1}}\rangle, \qquad \mathrm{HH}^{*,\mathbf{0}}(\Lambda) = {}_k\langle E_{\boldsymbol{p}}^{\mathbf{0}} \mid \mathbf{0} \leq \boldsymbol{p} \leq \mathbf{1}\rangle.$$

In particular,

$$\dim \mathrm{HH}^*(\Lambda) = \Big(1 + \binom{c}{0}, \binom{c}{1}, \binom{c}{2}, \binom{c}{c}, \dots\Big).$$

Since the total dimension is finite, the rate of growth $\gamma(\mathrm{HH}^*(\Lambda))$ is 0 and $\mathcal{S} = k$.

Now let us look at the other extreme case.

**Example 6.3** [Bergh and Oppermann 2008a]. Let $c \geq 2$ and let all $q_{ij}$ be roots of unity. Then

$$\text{pos.rk} \operatorname{Ker} Q_{G' \times G} = \operatorname{rk} \operatorname{Ker} Q_{G' \times G} = |G|$$

for any $G, G' \subseteq \{1, \dots, c\}$. Hence $\mathrm{HH}^*(\Lambda)$ is finitely generated over $\mathcal{S}$, and

$$\text{Krull.dim}\, \mathcal{S} = \gamma\big(\mathrm{HH}^*(\Lambda)\big) = c.$$

The final two examples illustrate that when $\gamma(\mathrm{HH}^*(\Lambda)) = c - 2$ very different kinds of behavior can occur.

**Example 6.4.** Let $q \in k^\times$ not be a root of unity and let $c \in \mathbb{N}_{\geq 3}$. Let $\Lambda$ be a quantum complete intersection of codimension $c$, with

$$\begin{aligned} q_{ij} &= 1 && \text{for } i, j < c-1, && q_{i,c-1} = q && \text{for } i < c-1, \\ q_{ic} &= q^{-1} && \text{for } i < c-1, && q_{c-1,c} = q^{-1}. \end{aligned}$$

One easily sees that $\mathcal{S} = k$. A case-by-case study (according to the values of $i$ for which $d_i = n_i - 1$) shows that the subspace $\mathrm{HH}^*_{\{1,\ldots,c-2\}}$ has a finite-dimensional complement in $\mathrm{HH}^*(\Lambda)$. It is nonempty if and only if $n_{c-1} = n_c$, and in that case

$$\gamma\big(\mathrm{HH}^*(\Lambda)\big) = \gamma\big(\mathrm{HH}^*_{\{1,\ldots,c-2\}}\big) = \text{pos.rk Ker}\, \underbrace{Q_{\{1,\ldots,c-2\}\times\{1,\ldots,c-2\}}}_{=0} = c-2.$$

**Example 6.5.** Let $q \in k^\times$ not be a root of unity, let $c \in \mathbb{N}_{\geq 3}$, and (for simplicity) let $\operatorname{char} k \neq 2$. Let $\Lambda$ be a quantum complete intersection of codimension $c$ with

$$\begin{aligned} q_{ij} &= 1 && \text{for } i, j < c-1, && q_{i,c-1} = q && \text{for } i < c-1, \\ q_{ic} &= q^{-1} && \text{for } i < c-1, && q_{c-1,c} = q. \end{aligned}$$

Then
$$\begin{aligned} \mathcal{S} &= {}_k\Big\langle s_{\boldsymbol{p}} \,\Big|\, p_i \text{ even and } \Big(Q\Big(\frac{p_j n_j}{2}\Big)_j\Big)_i = 1 \text{ for all } i \Big\rangle \\ &= {}_k\Big\langle s_{\boldsymbol{p}} \,\Big|\, p_i \text{ even and } \sum_{j=0}^{c-2} p_j n_j = p_{c-1} n_{c-1} = p_c n_c \text{ for all } i \Big\rangle. \end{aligned}$$

In particular,

$$\operatorname{Krull.dim} \mathcal{S} = c-2,$$

and hence also $\gamma(\mathrm{HH}^*(\Lambda)) = c-2$, by Corollary 4.6 .

Similarly to Example 6.4 one sees that $\mathrm{HH}^*_{\{1,\ldots,c-2\}} + \mathrm{HH}^*_{\{1,\ldots,c\}}$ form a subspace of $\mathrm{HH}^*(\Lambda)$ which has a finite-dimensional complement. Since by Corollary 5.8 $\mathrm{HH}^*_{\{1,\ldots,c\}}$ is always finitely generated over $\mathcal{S}$, we only have to look at $\mathrm{HH}^*_{\{1,\ldots,c-2\}}$. As in Example 6.4, one sees that $\mathrm{HH}^*_{\{1,\ldots,c-2\}} \neq 0$ if and only if $n_{c-1} = n_c$. Since

$$\text{pos.rk Ker}\, Q_{\{1,\ldots,c-2\}\times\{1,\ldots,c-2\}} = c-2 \neq 0 = \text{pos.rk Ker}\, Q_{\{1,\ldots,c\}\times\{1,\ldots,c-2\}}$$

it follows that $\mathrm{HH}^*(\Lambda)$ is finitely generated over $\mathcal{S}$ if and only if $n_{c-1} \neq n_c$.

## 7. Hochschild homology

To calculate the Hochschild homology, we proceed as for the Hochschild cohomology. That is, we calculate for any $\boldsymbol{d} \in \mathbb{Z}^c$ the degree-$\boldsymbol{d}$ part of the Hochschild homology. The actual calculations are very similar to the corresponding ones in Section 3, and will therefore be omitted here.

**Observation 7.1.** *The degree-$\boldsymbol{d}$ part $(\Lambda^{\mathrm{en}}\langle \boldsymbol{s}\rangle \otimes_{\Lambda^{\mathrm{en}}} \Lambda)_{\boldsymbol{d}}$ is nonzero if and only if $\boldsymbol{s} \leq \boldsymbol{d} \leq \boldsymbol{s} + \boldsymbol{n} - \boldsymbol{1}$. In that case it is one-dimensional.*

As in the case of the cohomology, it follows that the $c$-tuple complex

$$(\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_c \otimes_{\Lambda^{\mathrm{en}}} \Lambda)_{\boldsymbol{d}}$$

is concentrated in a cube (with sides of length 0 or 1), where there is a one-dimensional space in each corner of the cube.

Next we need to understand what happens to a map $f : \Lambda^{\mathrm{en}}\langle \boldsymbol{s} \rangle \to \Lambda^{\mathrm{en}}\langle \boldsymbol{s}' \rangle$ when it is tensored over $\Lambda^{\mathrm{en}}$ with $\Lambda$.

**Lemma 7.2.** *Let $f : \Lambda^{\mathrm{en}}\langle \boldsymbol{s} \rangle \to \Lambda^{\mathrm{en}}\langle \boldsymbol{s}' \rangle$. Then $f \otimes_{\Lambda^{\mathrm{en}}} \Lambda : \Lambda\langle \boldsymbol{s} \rangle \to \Lambda\langle \boldsymbol{s}' \rangle$ is given by*

$$(f \otimes_{\Lambda^{\mathrm{en}}} \Lambda)(\boldsymbol{x}^{\boldsymbol{a}}) = \sum_i f_2^i \boldsymbol{x}^{\boldsymbol{a}} f_1^i,$$

*where $f(1 \otimes 1) = \sum_i f_1^i \otimes f_2^i$.*

Now we are ready to calculate what tensoring over $\Lambda^{\mathrm{en}}$ with $\Lambda$ does to the maps occurring in the $c$-tuple complex $\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_c$.

**Lemma 7.3.** *Let $\boldsymbol{s}$ and $\boldsymbol{d}$ be such that $\boldsymbol{s} \le \boldsymbol{d} \le \boldsymbol{s} + \boldsymbol{n} - \boldsymbol{1}$ and let $i \in \{1, \dots, c\}$.*

(1) *Assume that $d_i > s_i$. The map $(\Lambda\langle \boldsymbol{s} + e_i \rangle)_{\boldsymbol{d}} \to (\Lambda\langle \boldsymbol{s} \rangle)_{\boldsymbol{d}}$ obtained by tensoring with $\Lambda$ the map over $\Lambda^{\mathrm{en}}$ given by*

$$\Lambda^{\mathrm{en}}\langle \boldsymbol{s} + e_i \rangle \to \Lambda^{\mathrm{en}}\langle \boldsymbol{s} \rangle \quad \textit{mapping} \quad 1 \otimes 1 \mapsto \frac{1}{\boldsymbol{q}^{\langle e_i \mid \boldsymbol{s} \rangle}} x_i \otimes 1 - \frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s} \mid e_i \rangle}} 1 \otimes x_i$$

*and then taking the part of degree $\boldsymbol{d}$ vanishes if and only if $(Q\boldsymbol{d})_i = 1$.*

(2) *Assume that $d_i = s_i + n_i - 1$. The map $(\Lambda\langle \boldsymbol{s} + (n_i - 1)e_i \rangle)_{\boldsymbol{d}} \to (\Lambda\langle \boldsymbol{s} \rangle)_{\boldsymbol{d}}$ obtained by tensoring with $\Lambda$ the map over $\Lambda^{\mathrm{en}}$ given by*

$$\Lambda^{\mathrm{en}}\langle \boldsymbol{s} + e_i \rangle \to \Lambda^{\mathrm{en}}\langle \boldsymbol{s} \rangle$$

*mapping*

$$1 \otimes 1 \mapsto \sum_{j=0}^{n_i - 1} \left( \frac{1}{\boldsymbol{q}^{\langle e_i \mid \boldsymbol{s} \rangle}} \right)^j \left( \frac{1}{\boldsymbol{q}^{\langle \boldsymbol{s} \mid e_i \rangle}} \right)^{n_i - 1 - j} x_i^j \otimes x_i^{n_i - 1 - j}$$

*and then taking the part of degree $\boldsymbol{d}$ vanishes if and only if $(Q\boldsymbol{d})_i \in \mathfrak{R}_i$.*

As for the cohomology, it follows that if the map on one edge of the cube $(\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_c \otimes_{\Lambda^{\mathrm{en}}} \Lambda)_{\boldsymbol{d}}$ vanishes then all parallel maps also vanish.

**Theorem 7.4.** *Let $\Lambda = \Lambda_{\boldsymbol{q}}^{\boldsymbol{n}}$ be a quantum complete intersection, and let $\boldsymbol{d} \in \mathbb{N}^c$. Divide the set $\{1 \dots c\}$ into the three parts*

$$I_0 = \{d_i = 0\}, \quad I_1 = \{n_i \mid d_i\} \setminus I_0, \quad I_2 = \{n_i \nmid d_i\}.$$

*Then $\mathrm{HH}_{*,\boldsymbol{d}}(\Lambda) \neq 0$ if and only if*

- *$(Q\boldsymbol{d})_i \in \mathfrak{R}_i$ for any $i \in I_1$, and*
- *$(Q\boldsymbol{d})_i = 1$ for any $i \in I_2$.*

*In this situation* $\mathrm{HH}_{*,\boldsymbol{d}}(\Lambda)$ *has a k-vector space basis*

$$\{T_{\boldsymbol{d}}^{\boldsymbol{p}} \mid \boldsymbol{p} \geq \boldsymbol{0} \text{ and } \mathfrak{p}(\boldsymbol{d}+\boldsymbol{1}) - \boldsymbol{2} \leq \boldsymbol{p} \leq \mathfrak{p}(\boldsymbol{d}+\boldsymbol{1}) - \boldsymbol{1}\}.$$

*Here* $T_{\boldsymbol{d}}^{\boldsymbol{p}}$ *is represented by* $\boldsymbol{x}^{\boldsymbol{d}-\mathfrak{s}(\boldsymbol{p})}$ *in position* $\boldsymbol{p}$. *In particular,* $T_{\boldsymbol{d}}^{\boldsymbol{p}}$ *has torsion degree* $\sum_{i=1}^{c} p_i$.

## 8. The rate of growth of the Hochschild homology

To study the rate of growth of the Hochschild homology, we decompose it similarly to our decomposition of the Hochschild cohomology in Construction 4.2.

**Construction 8.1.** For $G \subseteq \{1, \dots, c\}$ we denote by $\mathrm{HH}_*^G$ the $k$-span of $T_{\boldsymbol{d}}^{\boldsymbol{p}}$ with

$$G = \bigl\{ i \in \{1, \dots, c\} \bigm| d_i > 0 \text{ or } (Q\boldsymbol{d})_i \in \mathcal{R}_i \bigr\}.$$

This yields a decomposition $\mathrm{HH}_*(\Lambda) = \bigoplus_{G \subseteq \{1,\dots,c\}} \mathrm{HH}_*^G$, and hence

$$\gamma\bigl(\mathrm{HH}_*(\Lambda)\bigr) = \max_{G \subseteq \{1,\dots,c\}} \gamma(\mathrm{HH}_*^G).$$

As in the proof of Theorem 4.5, one obtains:

**Theorem 8.2.** *the rate of growth of the Hochschild homology of a finite-dimensional quantum complete intersection, is the maximum of* pos.rk Ker $Q_{G\times G}$ *over*

$$G = \bigl\{ i \in \{1, \dots, c\} \bigm| d_i > 0 \text{ or } (Q\boldsymbol{d})_i \in \mathcal{R}_i \bigr\},$$

*where* $\boldsymbol{d}$ *ranges over element of* $\mathbb{N}^c$ *such that* $(Q\boldsymbol{d})_i \in \mathcal{R}_i$ *for all* $i$ *with* $n_i \mid d_i$ *and* $d_i > 0$, *and* $(Q\boldsymbol{d})_i = 1$ *for all* $i$ *with* $n_i \nmid d_i$.

We conclude this paper by showing that the Hochschild homology of $\Lambda$ is closely related to the Hochschild homologies of certain subalgebras.

For $I \subset \{1, \dots, c\}$, denote by $\Lambda_I$ the subalgebra of $\Lambda$ generated by $x_i$ with $i \in I$. Then $\Lambda_I$ is a split quotient of $\Lambda$ (that is, we have algebra homomorphisms $\Lambda_I \to \Lambda \to \Lambda_I$ whose composition is the identity on $\Lambda_I$). It follows from the functoriality of Hochschild homology that $\mathrm{HH}_*(\Lambda_I)$ can be embedded into $\mathrm{HH}_*(\Lambda)$.

The following theorem shows that the Hochschild homologies of these subalgebras determine the Hochschild homology of $\Lambda$ to a large extent.

**Theorem 8.3.** *Let* $M$ *be the maximum of the rates of growth of* $\mathrm{HH}_*(\Lambda_{\overline{\{i\}}})$, *where* $i \in \{1, \dots, c\}$ *and* $\overline{\{i\}} = \{1, \dots, c\} \setminus \{i\}$. *Then the rate of growth of* $\mathrm{HH}_*(\Lambda)$ *is* $M$ *if* $\mathrm{HH}_*^{\{1,\dots,c\}} = 0$, *and* $\max\{M, \text{pos.rk Ker } Q\}$ *if* $\mathrm{HH}_*^{\{1,\dots,c\}} \neq 0$.

*Proof.* We will need to look at the sets $\mathrm{HH}_*^G$ as well as their analogs for $\mathrm{HH}_*(\Lambda_{\overline{\{i\}}})$. To avoid confusion we write $\mathrm{HH}_*^G(\Lambda)$ and $\mathrm{HH}_*^G(\Lambda_{\overline{\{i\}}})$, respectively, for these vector spaces.

Let $i_0 \in \{1, \dots, c\}$ and $G \subseteq \overline{\{i_0\}}$. It follows from the explicit description of bases in Theorem 7.4 and Construction 8.1 that $\mathrm{HH}_*^G(\Lambda)$ can be identified with a subspace of $\mathrm{HH}_*^G(\Lambda_{\overline{\{i_0\}}})$, and that the set of $T_{\boldsymbol{d}}^{\boldsymbol{p}}$ such that

$$G = \left\{ i \in \overline{\{i_0\}} \;\middle|\; d_i > 0 \text{ or } (Q\boldsymbol{d})_i \in \mathcal{R}_i \right\}$$

and

$$G \neq \left\{ i \in \{1, \dots, c\} \;\middle|\; d_i > 0 \text{ or } (Q\boldsymbol{d})_i \in \mathcal{R}_i \right\}$$

is a basis of the quotient space. This clearly means that

$$\left\{ i \in \{1, \dots, c\} \;\middle|\; d_i > 0 \text{ or } (Q\boldsymbol{d})_i \in \mathcal{R}_i \right\} = G \cup \{i_0\},$$

so the quotient embeds naturally into $\mathrm{HH}_*^{G \cup \{i_0\}}$.

It follows that

$$\gamma\big(\mathrm{HH}_*^G(\Lambda)\big) \le \gamma\big(\mathrm{HH}_*^G(\Lambda_{\overline{\{i_0\}}})\big) \le \max\big\{\gamma\big(\mathrm{HH}_*^G(\Lambda)\big), \gamma\big(\mathrm{HH}_*^{G \cup \{i_0\}}(\Lambda)\big)\big\}.$$

Taking the maximum over all $G$ and $i_0$ proves the theorem. □

## References


[Avramov et al. 1997] L. L. Avramov, V. N. Gasharov, and I. V. Peeva, "Complete intersection dimension", *Inst. Hautes Études Sci. Publ. Math.* 86 (1997), 67–114. MR 99c:13033 Zbl 0918.13008

[Bergh and Erdmann 2008] P. A. Bergh and K. Erdmann, "Homology and cohomology of quantum complete intersections", *Algebra Number Theory* **2**:5 (2008), 501–522. MR 2009h:16010 Zbl 05529335

[Bergh and Oppermann 2008a] P. A. Bergh and S. Oppermann, "Cohomology of twisted tensor products", *J. Algebra* **320**:8 (2008), 3327–3338. MR 2009f:16012 Zbl 1185.16012

[Bergh and Oppermann 2008b] P. A. Bergh and S. Oppermann, "The representation dimension of quantum complete intersections", *J. Algebra* **320**:1 (2008), 354–368. MR 2009c:16020 Zbl 1163.16008

[Buchweitz et al. 2005] R.-O. Buchweitz, E. L. Green, D. Madsen, and Ø. Solberg, "Finite Hochschild cohomology without finite global dimension", *Math. Res. Lett.* **12**:5-6 (2005), 805–816. MR 2006k:16019 Zbl 1138.16003

[Manin 1987] Y. I. Manin, "Some remarks on Koszul algebras and quantum groups", *Ann. Inst. Fourier (Grenoble)* **37**:4 (1987), 191–205. MR 89e:16022 Zbl 0625.58040

[Yoneda 1958] N. Yoneda, "Note on products in Ext", *Proc. Amer. Math. Soc.* **9** (1958), 873–875. MR 31 #233 Zbl 0101.27204



Communicated by Dave Benson

Received 2008-11-26 Revised 2010-08-27 Accepted 2010-10-01



steffen.oppermann@math.ntnu.no *Institutt for matematiske fag, Norwegian University of Science and Technology, 7491 Trondheim, Norway*
http://www.math.ntnu.no/~opperman/

ALGEBRA AND NUMBER THEORY 4:7(2010)

# Meromorphic continuation for the zeta function of a Dwork hypersurface

Thomas Barnet-Lamb

We consider the one-parameter family of hypersurfaces in $\mathbb{P}^5$ over $\mathbb{Q}$ with projective equation $(X_1^5+X_2^5+X_3^5+X_4^5+X_5^5)=5tX_1X_2\dots X_5$, proving that the Galois representations attached to their cohomologies are potentially automorphic, and hence that the zeta function of the family has meromorphic continuation to the whole complex plane.

## 1. Introduction

Harris, Shepherd-Barron, and Taylor have proved in [Harris et al. 2010] a potential modularity theorem, showing that certain Galois representations become automorphic after a sufficiently large totally real base change. In their argument, a key role is played by certain families of hypersurfaces, called Dwork families — in particular, by the part of the cohomology of the family which is invariant under a certain group action. (We will write $\mathfrak{F}$ for the motive given by this part of the cohomology.) The importance of $\mathfrak{F}$ to their argument is reflected in the statement of the theorem they prove: in order to prove an $l$-adic Galois representation $r$ is potentially modular using their theorem, one requires, among other conditions, that the restriction of the residual representation of $r$ to inertia at primes above $l$ be isomorphic to the restriction of the residual representation of some element of the family $\mathfrak{F}$.

They give two applications in their paper. On the one hand, through considerable ingenuity (and the fact that the Dwork family includes the Fermat hypersurface, whose cohomology restricted to inertia is easy to analyze) they are able to deduce that the odd symmetric powers of the cohomology of an elliptic curve over $\mathbb{Q}$ are modular, and (through further ingenuity) to deduce the Sato–Tate conjecture. On the other hand, the form of the condition on the inertial representation makes it very inviting to apply their modularity theorem to $\mathfrak{F}$ itself. It turns out to be fairly simple

The author was partially supported by NSF grant DMS-0600716 and by a Jean E. de Valpine Fellowship.

*MSC2000:* primary 11G40; secondary 11R39, 11F23.

*Keywords:* Dwork hypersurface, potential automorphy, zeta function.

to show that the other conditions of the potential modularity theorem are satisfied, and one can deduce the modularity of $\mathfrak{F}$ and hence the meromorphic continuation and functional equation of the zeta function of this part of the cohomology of the Dwork family.

A very natural question presents itself: is it possible to gain enough control of the other parts of the cohomology of the Dwork family that one can prove meromorphic continuation for the *whole* zeta function? In this paper, I answer this question in the affirmative for $N=5$, and also make some remarks on why further generalization of these methods is likely to be hard absent very significant advances in the technology of lifting theorems. In the analysis here, a key role is played by [Katz 2009], which describes the relative cohomology sheaf of the Dwork family over the base, and its decomposition under the group action alluded to above, in great detail.

***Note.*** Since this paper was written, the technology of potential automorphy has advanced somewhat; in particular, [Barnet-Lamb et al. 2010] proves certain rather general potential automorphy theorems for regular, crystalline, self-dual representations of the Galois group of totally real and CM fields. Using those results, the argument for Proposition 5 — which is perhaps rather involved at present — could be replaced with an appeal to the theorems for compatible systems proved in the new manuscript. The remainder of the present paper, in particular the analysis of the pieces of the cohomology in Section 3, does not seem to be able to be simplified, even with these new results.

## 2. Dwork families

***Note.*** My notation for Dwork families broadly follows [Katz 2009], with $N$ in place of $n$, except that Katz works throughout with sheaves with coefficients in $\overline{\mathbb{Q}}_l$, whereas we will need the flexibility gained by working initially with $\mathbb{Q}_l$ coefficients, and extending to $\overline{\mathbb{Q}}_l$ only as necessary to apply Katz's results. Our notation is not directly comparable with the notation of [Harris et al. 2010].

Let $N$ be a positive integer. Fix a base ring

$$R_0 = \mathbb{Z}[1/N, \mu_N],$$

where $\mu_N$ denotes the $N$-th roots of unity. *For most of this paper, we work over* $\mathbb{Q}(\mu_5)$ *and all Galois representations are representations of subgroups of* $G_{\mathbb{Q}(\mu_5)}$. We will eventually return to working over $\mathbb{Q}$, but when we do so, this will be made explicit. We consider the scheme

$$Y \subset \mathbb{P}^{N-1} \times \mathbb{P}^1$$

over $R_0$ defined by the equations

$$\mu(X_1^N + X_2^N + \cdots + X_N^N) = N\lambda X_1 X_2 \ldots X_N,$$

using $(X_1 : \cdots : X_N)$ and $(\mu : \lambda)$ as coordinates on $\mathbb{P}^{N-1}$ and $\mathbb{P}^1$ respectively. We consider $Y$ as a family of schemes over $\mathbb{P}^1$ by projection to the second factor. We will label points on this $\mathbb{P}^1$ using the affine coordinate $t = \lambda/\mu$, and will write $Y_t$ for the fiber of $Y$ above $t$. From now on (apart from some remarks in the conclusion) we will be concerned exclusively with the case $N = 5$.

The family $Y$ is smooth over the open set $U = \operatorname{Spec} R_0[1/(t^5 - 1)]$ away from the roots of unity. We are interested in the sheaf of relative cohomology of the family $Y$ above the set $U$. Let $l$ be a prime number which splits in $\mathbb{Q}(\mu_5)$.[1] Let $T_0 = U[1/l]$, and form lisse sheaves

$$\mathcal{F}_l^i := R^i \pi_* \mathbb{Q}_l, \tag{1}$$

$$\mathcal{F}^i[l] := R^i \pi_* \mathbb{Z}/l\mathbb{Z} \tag{2}$$

on $T_0$. (We remark that Katz's $\mathcal{F}_l^i$ would be $\mathcal{F}_l^i \otimes \overline{\mathbb{Q}}_l$ in our notation, since he works with algebraically closed coefficients throughout.) As a family of hypersurfaces, much of the cohomology of $Y_\lambda$ is controlled by the hard Lefschetz theorem: for $i \neq N - 2 = 3$, we have

$$\mathcal{F}_l^i = \begin{cases} 0 & (i < 0), \\ 0 & (i > 6), \\ 0 & (i \text{ odd}, \ 0 \leq i \leq 6), \\ \mathbb{Q}_l(-j) & (i = 2j \text{ even}, \ 0 \leq i \leq 6, \ i \neq 3). \end{cases}$$

The contribution to the zeta function from the characters $\mathbb{Q}_l(-j)$ is of course well understood. Thus in order to prove the functional equation for the zeta function of the whole variety, it suffices to control the zeta function of $\mathcal{F}_l^3$. We will refer to the sheaf $\mathcal{F}_l^3$ as $\mathrm{Prim}_l$ from now on. As discussed in the introduction, there is a natural group action on $\mathrm{Prim}_l$, allowing us to break down the cohomology into simpler pieces. Let us now introduce this group action. We will write $\Gamma$ for $(\mu_5)^5$, the 5-fold product of the group of roots of unity, and $\Gamma_W$ for the subgroup of elements $(\zeta_1, \ldots \zeta_5)$ with $\prod_{i=1}^5 \zeta_i = 1$. $\Gamma_W$ acts on $Y$ with $(\zeta_1, \ldots \zeta_5)$ acting via

$$\big((X_1 : \cdots : X_5), t\big) \mapsto \big((\zeta_1 X_1 : \cdots : \zeta_5 X_5), t\big).$$

The image of $\mu_5$ embedded diagonally in $\Gamma$ lies in $\Gamma_W$ and acts trivially under this action. We will write $\Delta$ for this image.

Harris et al. [2010] focus their attention on the invariants under this group action, a sheaf they refer to as $V$. They prove the following theorem:

[1] I believe that this assumption could be dispensed with. However, we will only ever need the theory we are about to develop for one particular choice of $l$, and we will always be able to make this choice such that $l$ splits in $\mathbb{Q}(\mu_5)$. Therefore, I have chosen to make this assumption, since it simplifies the argument. In particular, it means that the sheaves $\mathrm{Prim}_l$ which we define later will have coefficient ring $\mathbb{Q}_l$, rather than an extension field.

**Theorem 1** [Harris et al. 2010, Theorem 4.4]. *Suppose that $t \in \mathbb{Q} - \mathbb{Z}[1/5]$. Then the function $L(V_t, s)$ is defined and has meromorphic continuation to the whole complex plane, satisfying the functional equation*

$$L(V, s) = \epsilon(V, s) L(V, 4 - s).$$

As I have said, our aim in this paper is to analyze the remaining parts of the cohomology and so establish the functional equation for the zeta function of the variety as a whole. As a first step to doing so, let us consider what other parts there actually are.

## 3. The pieces of the cohomology

The character group of $\Gamma$ is $(\mathbb{Z}/5\mathbb{Z})^5$; that of $\Gamma_W$ is $(\mathbb{Z}/5\mathbb{Z})^5/\langle W\rangle$ where we write $W$ for the element $(1, 1, \ldots, 1)$; and that of $\Gamma_W/\Delta$ is $(\mathbb{Z}/5\mathbb{Z})^5_0/\langle W\rangle$ where we write $(\mathbb{Z}/5\mathbb{Z})^5_0$ for $\{(v_1, \ldots, v_5) \in (\mathbb{Z}/5\mathbb{Z})^5 \mid \sum_i v_i = 0\}$. Thus the eigensheaves of $\mathrm{Prim}_l$ under the action described in Section 2 are labeled by elements $v$ of $(\mathbb{Z}/5\mathbb{Z})^5_0/\langle W\rangle$: we may write such an element as $(v_1, \ldots, v_5) \bmod W$ with the $v_i$ elements of $\mathbb{Z}/5\mathbb{Z}$; it will often be convenient to abbreviate this to $[(v_1, \ldots, v_5)]$. Note that our assumption that $\mu_5 \in \mathbb{Q}_l$ is critical here in ensuring that the decomposition into eigensheaves is indeed defined with $\mathbb{Q}_l$ coefficients (and not with coefficients in some extension field). Note also that the labeling is not canonical, but depends on a choice of an identification of the copies of $\mu_5$ in $\mathbb{Q}_l$ and in $R_0$: equivalently, it depends on a choice of embedding $R_0 \hookrightarrow \mathbb{Q}_l$. Having made such a choice, we shall write $\mathrm{Prim}_{l,[(v_1,\ldots,v_5)]}$ for the piece of $\mathrm{Prim}_l$ where $\Gamma_W/\Delta$ acts via $[(v_1, \ldots, v_5)]$. (Thus, for instance, $V_l = \mathrm{Prim}_{l,[(0,0,0,0,0)]}$.) Again, we remark that Katz's $\mathrm{Prim}_{l,v}$ would correspond to our $\mathrm{Prim}_{l,v} \otimes \overline{\mathbb{Q}}_l$.

The obvious action of $S_5$ on $(\mathbb{Z}/5\mathbb{Z})^5$ preserves $(\mathbb{Z}/5\mathbb{Z})^5_0$ and $W$, and hence induces an action of $S_5$ on $(\mathbb{Z}/5\mathbb{Z})^5_0/\langle W\rangle$. Note that, if we permute the $(v_i)$ in this manner, the resulting sheaf $\mathrm{Prim}_{l,v}$ is isomorphic to the original (the isomorphism being induced from the map on $Y$ which permutes the $X_i$ according to the same permutation). Thus to show that all the sheaves $\mathrm{Prim}_{l,v}$ are automorphic it will suffice to consider a set of $v$ representing all the orbits of $(\mathbb{Z}/5\mathbb{Z})^5_0/\langle W\rangle$ under $S_5$.

**Proposition 2.** *All orbits of $(\mathbb{Z}/5\mathbb{Z})^5_0/\langle W\rangle$ under $S_5$ are represented in the list*

$$[(0,1,2,3,4)],\ [(0,0,1,1,3)],\ [(0,0,1,2,2)],\ [(0,0,2,4,4)],$$
$$[(0,0,3,3,4)],\ [(0,0,0,1,4)],\ [(0,0,0,2,3)],\ [(0,0,0,0,0)].$$

*Thus, if $\mathrm{Prim}_{l,v}$ is automorphic for each of these $v \in (\mathbb{Z}/5\mathbb{Z})^5_0$, it is automorphic for all $v$.*

*Proof.* We start with an arbitrary element $v$ of $(\mathbb{Z}/5\mathbb{Z})^5_0/\langle W\rangle$, and pick a representative $(v_1, \ldots, v_n) \in (\mathbb{Z}/5\mathbb{Z})^5_0$. By changing the representative of the congruence class mod $W$, we may ensure that in the list $(v_1, \dots, v_5)$, the 0 occurs at least as often as any other element of $\mathbb{Z}/5\mathbb{Z}$. Then, applying an appropriate permutation to $v$ (and hence to the $v_i$), we may ensure that the $v_i$ increase. (We order congruence classes mod 5 according to the order of their unique representatives in $\{0, \dots, 4\}$.)

Since 0 occurs at least as often as anything else, there must be at least one zero at the beginning of the list $(v_1, \dots, v_5)$. We split into several cases according to the number of zeroes there. Clearly, if there is 1 zero then $v = [(0, 1, 2, 3, 4)]$; if there are 4 or more zeros then $v = [(0, 0, 0, 0, 0)]$; and if there are 3 then the two remaining $v_i$ are 1 and 4 or 2 and 3.

If there are 2 zeroes, then we split into cases according to the value of $v_3$. If, for instance, $v_3 = 1$, then $v_4 + v_5 = 4$, so $v_4, v_5$ must be $\{1, 3\}$, $\{2, 2\}$, or $\{0, 4\}$, the last being impossible since the $v_i$ must increase. The other cases are similar. □

**Proposition 3.** *Assume we have chosen an arbitrary embedding of $R_0$ into $\overline{\mathbb{Q}}_l$. For each $v$ in the following table, the dimension and Hodge–Tate numbers of* $\mathrm{Prim}_{l,v}$ *are as given*:

| $v$ | $\dim \mathrm{Prim}_{l,v}$ | $\mathrm{HT}(\mathrm{Prim}_{l,v})$ |
|---|---|---|
| $[(0, 1, 2, 3, 4)]$ | 0 | $\{\}$ |
| $[(0, 0, 1, 1, 3)]$ | 2 | $\{1, 2\}$ |
| $[(0, 0, 1, 2, 2)]$ | 2 | $\{1, 2\}$ |
| $[(0, 0, 2, 4, 4)]$ | 2 | $\{1, 2\}$ |
| $[(0, 0, 3, 3, 4)]$ | 2 | $\{1, 2\}$ |
| $[(0, 0, 0, 1, 4)]$ | 2 | $\{1, 2\}$ |
| $[(0, 0, 0, 2, 3)]$ | 2 | $\{1, 2\}$ |

*In particular*, $\mathrm{Prim}_{l,[(0,1,2,3,4)]}$ *is zero-dimensional*, *and although the Hodge–Tate numbers depend in principle on the choice of embedding of $R_0$ into $\mathbb{Q}_l$*, *in practice they are independent of this choice.*

*Proof.* Recall that at the beginning of this section we chose a particular embedding $R_0 \to \mathbb{Q}_l$ in order to label the pieces of the cohomology. (We remark that since $R_0 = \mathbb{Z}[1/N, \mu_n]$ and we have a running assumption that $\mu_N \subset \mathbb{Q}_l$, this is the same thing as choosing an embedding $R_0 \hookrightarrow \overline{\mathbb{Q}}_l$.) Katz makes a corresponding choice in [Katz 2009, §1], and the Hodge–Tate numbers at this particular embedding (as well as the dimension, which does not depend on the choice of an embedding) may then be calculated by applying the procedure described in [ibid., Lemma 3.1]. (We will investigate what happens for Hodge–Tate numbers at the other embeddings later.) More precisely, Katz's procedure computes the Hodge–Tate numbers for *his* sheaf $\mathrm{Prim}_{l,v}$, which is our $\mathrm{Prim}_{l,v} \otimes_{\mathbb{Q}_l} \overline{\mathbb{Q}}_l$, but of course the Hodge–Tate numbers of $\mathrm{Prim}_{l,v}$ and $\mathrm{Prim}_{l,v} \otimes_{\mathbb{Q}_l} \overline{\mathbb{Q}}_l$ are the same.

As an example, we will compute the dimension and Hodge–Tate numbers for $v = [(0, 0, 1, 1, 3)]$. We are asked to consider the coset of elements of $(\mathbb{Z}/5\mathbb{Z})^5_0$ representing $v = [(0, 0, 1, 1, 3)]$, that is to say the particular embedding $R_0 \to \mathbb{Q}_l$ which was chosen arbitrarily at the beginning of this section (or in [ibid., §1]) and used to label the pieces of the cohomology

$$\{(0, 0, 1, 1, 3),\ (1, 1, 2, 2, 4),\ (2, 2, 3, 3, 0),\ (3, 3, 4, 4, 1),\ (4, 4, 0, 0, 2)\}.$$

The lemma then tells us that the dimension of $\mathrm{Prim}_l$ can be computed as the number of elements of this set which are *totally nonzero*; that is, contain no 0's. There are two of these. Then, the Hodge–Tate numbers are computed by taking the degrees of the totally nonzero representatives above, where the *degree* of an element $(v_1, \dots, v_n) \in (\mathbb{Z}/5\mathbb{Z})^5_0$ is $\sum_i \tilde{v}_i$, and where (in turn) for each $i$, $\tilde{v}_i$ is the integer representing $v_i$ in the range 0 to 4. Then the multiset of these degrees is the multiset of Hodge–Tate numbers, with each element increased by 1. In our case, the HT numbers are therefore $\{(1+1+2+2+4)/5-1, (3+3+4+4+1)/5-1\} = \{1, 2\}$.

For the other $v$'s, the totally nonzero representatives are as follows:

| $v$ | Totally nonzero representatives |
|---|---|
| $[(0, 0, 1, 2, 2)]$ | $\{(1, 1, 2, 3, 3),\ (2, 2, 3, 4, 4)\}$ |
| $[(0, 0, 2, 4, 4)]$ | $\{(2, 2, 4, 1, 1),\ (4, 4, 1, 3, 3)\}$ |
| $[(0, 0, 3, 3, 4)]$ | $\{(3, 3, 1, 1, 2),\ (4, 4, 2, 2, 3)\}$ |
| $[(0, 0, 0, 1, 4)]$ | $\{(2, 2, 2, 3, 1),\ (3, 3, 3, 4, 2)\}$ |
| $[(0, 0, 0, 2, 3)]$ | $\{(1, 1, 1, 3, 4),\ (4, 4, 4, 1, 2)\}$ |

The result, for the Hodge–Tate numbers at our chosen embedding, follows.

Now, when we change our choice of embedding, the effect is to relabel the various pieces of the cohomology, by multiplying their labels $v$ by an element of $(\mathbb{Z}/5\mathbb{Z})^\times$. It is easy to see from the table that such relabeling sends an eigenspace to another eigenspace where the calculated Hodge–Tate numbers from the algorithm are the same. Hence we are done. □

## 4. Controlling the $L$ functions

We will now try to control the $L$ functions of the two-dimensional pieces we have singled out.

**Lemma 4.** *Let $v$ be taken from the table in Proposition 3. There is a constant $D$ such that if $M$ is an integer divisible only by primes $p > D$ and if $t \in U$ then the map*

$$\pi_1(U, t) \to \mathrm{SL}(\mathrm{Prim}[M]_{v,t})$$

*is surjective.* (*Here* $\mathrm{SL}(\mathrm{Prim}[M]_t)$ *denotes the group of automorphisms of the two-dimensional module* $\mathrm{Prim}[M]_t$ *with determinant* 1.)

*Proof.* The monodromy of $\mathrm{Prim}_l \otimes \overline{\mathbb{Q}}_l$ is Zariski dense in $\mathrm{SL}_2(\mathrm{Prim}_l)$, using [Katz 2009, Lemma 10.3], and remembering that $\mathrm{Sp}_2 = \mathrm{SL}_2$. The same is then seen to hold for $\mathrm{Prim}_l$. We can then deduce the result using [Matthews et al. 1984, Theorem 7.5 and Lemma 8.4] or [Nori 1987, Theorem 5.1]. □

We now proceed to analyze the two-dimensional pieces. We shall write $\mathrm{Prim}_{*,v}$ to mean the motive whose $l$-adic realizations are the $\mathrm{Prim}_{l,v}$ as $l$ varies.

**Proposition 5.** *For each $v$ in the table of Proposition 3 with* $\mathrm{Prim}_{*,v}$ *two-dimensional, and for each $t \in \mathbb{Q}-\mathbb{Z}[1/10]$, the function $L(\mathrm{Prim}_{*,v,t}, s)$ is defined and has meromorphic continuation to the whole complex plane, satisfying the functional equation*

$$L(\mathrm{Prim}_{*,v,t}, s) = \epsilon(\mathrm{Prim}_{*,v,t}, s) L(\mathrm{Prim}_{*,v^*,t}, 4-s),$$

*where we write $v^*$ for $\{5-k \mid k \in v\}$.*

Before proving this, let us briefly remind ourselves of the significance of the words "is defined" in the statement of the theorem. The point is that, for each prime $p$, we wish to construct a local $L$ factor $L_p$, and we do so by looking at our motive's $l$-adic cohomology $\mathrm{Prim}_{l,v,t}$ for some $l \neq p$. Given an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_l$, we can associate a Weil–Deligne representation $\mathrm{WD}(\mathrm{Prim}_{l,v,t}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)})^{\text{F-ss}}$ to this $l$-adic cohomology at $p$, and to this, in turn, we can associate an $L$ factor. To get an unambiguous $L$ factor, we must insist that the Weil–Deligne representation (and hence the $L$ factor) do not depend on the choices we made: that is, the choice of $l$ and of an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_l$. Thus the statement "$L(\mathrm{Prim}_{*,v,t}, s)$ is defined" is saying that for every $p$, the local Weil–Deligne representation at $p$ constructed in this way is independent of these choices.

*Proof of Proposition 5.* Our argument draws heavily on Theorem 3.3 of [Harris et al. 2010]. We first choose $q$ to be a rational prime dividing the denominator of $t$, so that $v_q(t) < 0$ and $q \nmid 10$.

*Step 1.* The goal of this step is to choose certain primes $l$ and $l'$ which will be instrumental to the argument. In order to be in a position to do this we must first analyze the Zariski closure of the image of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\mu_5))$ in the group $\mathrm{GL}(\mathrm{Prim}_{l,v,t})$ of automorphisms of the $\mathbb{Q}_l$ vector space $\mathrm{Prim}_{l,v,t}$. We will write $G_l$ for this image and $G_l^0$ for the connected component of the identity in it.

By [Katz 2009, Lemma 10.1] the local monodromy of $\mathrm{Prim}_{l,v,t} \otimes \overline{\mathbb{Q}}_l$ at $\infty$ is unipotent with a single Jordan block. (Condition 4 of the equivalent conditions in that lemma may be verified by direct inspection of each case in Section 3.) We immediately deduce the same for $\mathrm{Prim}_{l,v,t}$ itself. By the argument used to establish [Harris et al. 2010, Lemma 1.15], and recalling that $v_q(t) < 0$, we conclude that inertia at $q$ acts via a maximal unipotent. Thus $G_0^l$ contains such a maximal unipotent, and hence, by [Scholl 2006, Proposition 3], acts irreducibly.

Moreover, the determinant map to $\mathbb{G}_m$ is dominating. To see this, note that Poincaré duality furnishes us with a perfect pairing between $\mathrm{Prim}_{l,v,t}$ and $\mathrm{Prim}_{l,v^*,t}$ towards $\mathbb{Q}(-3)$, and that $\mathrm{Prim}_{l,v^*,t}$ is the complex conjugate of $\mathrm{Prim}_{l,v,t}$. Thus we have

$$\mathrm{Prim}^c_{l,v,t} = \mathrm{Prim}^\vee_{l,v,t}\epsilon_l^{-3}, \tag{3}$$

which tells us in turn that $(\det \mathrm{Prim}_{l,v,t})(\det \mathrm{Prim}_{l,v,t})^c = \epsilon_l^{-6}$, which would be impossible if the determinant character did not dominate $\mathbb{G}_m$.

Thus by [Katz 1988, Theorem 9.10], we may conclude that $G_l^0$ is $\mathrm{GL}_2$. The main theorem of [Larsen 1995] then tells us that the set of primes $l$ for which we fail to have

$$\mathrm{PSL}_2(\mathrm{Prim}[l]_{v,t}) \subset \Gamma_l \subset \mathrm{PGL}_2(\mathrm{Prim}[l]_{v,t}) \tag{4}$$

has Dirichlet density 0.

By mimicking the argument for Proposition 3.4.2 of [Barnet-Lamb 2008], we can construct a field $F^*(v)$ such that if a prime $l$ splits in $F^*(v)$, the determinant of the natural polarization on $\mathrm{Prim}[l]_v$ coming from Poincaré duality will be a square. (We also take from this paper two notations which we will use a few lines below: the field $F^*(2, 10)$ and the integer $C(2, 10)$, defined respectively in Proposition 3.4.2 and Corollary 2.1.2 there.) Now, since the set of primes for which (4) holds had Dirichlet density 1, the set of primes for which (4) holds and for which $l$ splits in $\mathbb{Q}(\mu_{10})$, in $F^*(v)$ and in $F^*(2, 10)$ has positive density. We may therefore choose $l$ to be such a prime, and in addition insist that

$$v_l(t^5 - 1) = 0$$

and that $l$ be greater than $n$, $C(2, 10)$ and $D$ (see Lemma 4 for the latter).

We choose $l'$ to be a distinct rational prime enjoying the same list of properties. Note that (4) will ensure that the image of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_l))$ in $\mathrm{GL}(\mathrm{Prim}[l]_t)$ is big, via (say) [Clozel et al. 2008, Lemma 2.5.5], and the simplicity of $\mathrm{PSL}_2(\mathbb{F}_l)$ will ensure that $\zeta_l \notin \overline{\mathbb{Q}}^{\ker \mathrm{Prim}[l]_t}$.

*Step 2.* Our next step in the proof is to establish that there exists a CM field $F_1/\mathbb{Q}(\mu_{10})$ and a $t' \in T(F_1)$ such that we have

$$\mathrm{Prim}[l]_{v,t'} \equiv \mathrm{Prim}[l]_{v,t}, \tag{5}$$

$$\mathrm{Prim}[l']_{v,t'}|_{I_{F_{1,\mathfrak{w}}}} \equiv \epsilon_l^{-1} \oplus \epsilon_l^{-2}, \tag{6}$$

$$v_{\mathfrak{q}}(t') < 0 \quad \text{if } \mathfrak{q} \mid q, \tag{7}$$

$$v_{\mathfrak{w}}((t')^5 - 1) = 0 \quad \text{if } \mathfrak{w} \mid ll'. \tag{8}$$

First, pick a point $t'' \in \mathbb{Q}(\mu_{10})^+$ satisfying the following conditions:

- If $\mathfrak{w}|ll'$ then $w((t'')^5 - 1) = 0$.
- If $\mathfrak{w}|l'$ then $\mathrm{Prim}[l']_{t''}|_{I_\mathfrak{w}} \equiv \epsilon_l^{-1} \oplus \epsilon_l^{-2}$.
- $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\mu_{10})^+) \to \mathrm{SL}(\mathrm{Prim}[l']_{t''})$ is surjective.

As in [Harris et al. 2010], the existence of such a $t''$ relies on the form of Hilbert irreducibility with weak approximation; see [Ekedahl 1990]. We may achieve the second condition by taking $t''$ to be $l'$-adically close to zero, since $\mathrm{Prim}[l']_0|_{I_w}$ is $\epsilon_l^{-1} \oplus \epsilon_l^{-2}$. (This last is because we know $\mathrm{Prim}[l']_0|_{I_w}$ to be a direct sum of characters, as in [Deligne et al. 1982], and we know from Proposition 3 that they are crystalline with Hodge–Tate numbers 1, 2.)

We introduce the character $\phi_l$ as is done in Section 3.2 of [Barnet-Lamb 2008]. *All references through the end of this step are to that paper.*[2] We will also consider the $l'$-adic version of $\phi$, as well as the mod $M := ll'$ version; we will abuse notation by writing $\phi$ for all of these.

Now, we follow the argument of the proof of Proposition 3.4.1, with this setup: $l_1 = l$, $l_2 = l'$, $\bar{\rho}_1 = \mathrm{Prim}[l]_{v,t} \otimes \phi^{-1}$, $\bar{\rho}_2 = \mathrm{Prim}[l']_{v,t''} \otimes \phi^{-1}$, the $\mathfrak{q}_j$ are the primes above $q$, and $N = 5$. Details follow:

The first part of the argument works in exactly the same way: we set $M := ll'$, introduce a mod $M$ character $\phi_l$ and a mod $M$ representation $\bar{\rho}_{\mathbb{Z}/M\mathbb{Z}}$, and note that $\mathrm{Prim}[M]$ and $\bar{\rho}_{\mathbb{Z}/M\mathbb{Z}}$ become isomorphic once we disregard the Galois action and keep only the modules with a pairing, using the assumption that $l$ splits in $F^*(v)$.

We next study the determinant of Prim. From the fact that $\psi_1$ maps into the image of geometric monodromy, we can deduce that it is trivial, since we saw above that geometric monodromy was trivial. Thus we deduce that $\det(\mathrm{Prim}[l]_s \otimes \phi^{-1})$ is independent of $s$ — in fact, from an argument analogous to that establishing Lemma 3.2.1, we know that $\det(\mathrm{Prim}[l]_s \otimes \phi^{-1}) = \epsilon_l^{-1}$. Similarly, $\det(\mathrm{Prim}[l]_s \otimes \phi^{-1})$ is independent of $s$.

This tells us that $\det \bar{\rho}_1 = \det \mathrm{Prim}[l]_{v,t} \otimes \phi^{-1}$ and $\det \bar{\rho}_2 = \det \mathrm{Prim}[l']_{v,t''} \otimes \phi^{-1}$, and we can choose an isomorphism $\eta : \det(\mathrm{Prim}[M] \otimes \phi^{-1}) \to \det \bar{\rho}_{\mathbb{Z}/M\mathbb{Z}}$.

In item (1) in the proof of Proposition 3.4.1, the set of automorphisms preserving this fixed isomorphism between determinants is seen to be $\mathrm{SL}(\mathbb{Z}/M\mathbb{Z})$. We know that the monodromy would be dense in this set, since $l, l' > D$.

Finally, in item (3) in the proof, we see that the $\Omega_\mathfrak{w}$ are nonempty sets, since those above $l$ contain points above $t$ and those above $l'$ contain points above $t''$.

It follows that there is a CM field $F_1/\mathbb{Q}(\mu_{10})$ and a $t' \in T_\mathcal{W}(F_1)$ satisfying the conditions (5)–(8) above. (For (6), namely $\mathrm{Prim}[l']_{v,t'}|_{I_{F_\mathfrak{w}}} \equiv \epsilon_l^{-1} \oplus \epsilon_l^{-2}$, use the fact that $\mathrm{Prim}[l']_{v,t'}$ agrees with $\mathrm{Prim}[l']_{t''}$, which was chosen to have this property.)

[2] We depart from the notation there in *not* writing $(\vec{h})$ for the twist by this character, since in fact $h(\sigma) = 1$ for all $\sigma$, so one might think that $(\vec{h})$ means (1), but this is not true: $\phi^{-1}$ is not the cyclotomic character.

*Step 3.* We will now apply Theorem 1.1.3 of [Barnet-Lamb 2008] to show that there exists a CM field $F/F_1/\mathbb{Q}(\mu_{10})$ such that $\mathrm{Prim}_{l',v,t'} \otimes \phi^{-1}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F)}$ is automorphic, by taking $\mathcal{L} = \varnothing$, $N = 10$, and $r = \mathrm{Prim}_{l,v,t'} \otimes \phi^{-1}$. Let us verify the conditions of that theorem in turn. First of all, $l$ splits in $\mathbb{Q}(\mu_{10})$ by our choice of $l$, and $q \nmid 10$ by our by choice of $q$. Then we address the numbered conditions:

(1) *$r$ ramifies only at finitely many primes.* This is trivial, being true for all Galois representations that come from geometry.

(2) $r^c \cong r^\vee \epsilon_l^{-1}$. For the same reason as for (3) (page 846), we have

$$\mathrm{Prim}^c_{l,v,t'} = \mathrm{Prim}^\vee_{l,v,t'} \epsilon_l^{-3},$$

whence we have what we want, since $r = \mathrm{Prim}^\vee_{l,v,t'} \otimes \phi^{-1}$ and $\phi\phi^c = \epsilon_l^{-2}$.

(3) *The Bellaïche–Chenevier sign is* $+1$. This is because the Poincaré duality pairing is symplectic and the multiplier of complex conjugation is odd.

(4) *$r$ is crystalline with the right Hodge–Tate numbers.* This follows immediately from the calculations of the previous proposition, once we note that the twist by $\phi$ changes the Hodge–Tate numbers by 1.

(5) *$r$ is unramified at all the primes of $\mathcal{L}$.* This is vacuous.

(6) $r|_{\mathrm{Gal}(\bar{F}_{v_q}/F_{v_q})}{}^{\mathrm{ss}}$ *is unramified and* $r|_{\mathrm{Gal}(\bar{F}_{v_q}/F_{v_q})}{}^{\mathrm{ss}}$ *has Frobenius eigenvalues* $1, \#k(v_q), \ldots, (\#k(v_q))^{n-1}$. By [Katz 2009, Lemma 10.1] the local monodromy of $\mathrm{Prim}_{l,v,t'} \otimes \overline{\mathbb{Q}}_l$ at $\infty$ is unipotent with a single Jordan block. We immediately deduce the same for $\mathrm{Prim}_{l,v,t'}$. By the argument used to establish [Harris et al. 2010, Lemma 1.15], and recalling that $v_q(t') < 0$, we conclude that inertia at $q$ acts via a maximal unipotent and that the Frobenius eigenvalues are of the form required.

(7) $\det r \equiv \epsilon_l^{-1}$. We saw above that $\det(\mathrm{Prim}[l]_s \otimes \phi^{-1}) = \epsilon_l^{-1}$, as required.

(8) *Let $\bar{r}$ denote the semisimplification of the reduction of $r$, and let $r'$ denote the extension of $r$ to a continuous homomorphism* $\mathrm{Gal}(\bar{F}/F^+) \to \mathcal{G}_n(\overline{\mathbb{Q}}_l)$ (as described in [Clozel et al. 2008, §1]); *then* $\bar{r}'(\mathrm{Gal}(\bar{F}/F(\zeta_l))$ *is big* (in the sense of "big image"). This is true by [Clozel et al. 2008, Lemma 2.5.5], since we chose $t'$ such that $\mathrm{Prim}[l']_{t'} \equiv \mathrm{Prim}[l']_{t''}$, and we chose $t''$ such that $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\mu_{10})^+) \to \mathrm{GL}(\mathrm{Prim}[l']_{t''})$ is surjective.

(9) $\bar{F}^{\ker \mathrm{ad}\, \bar{r}}$ *does not contain* $F(\zeta_l)$. This is true by the simplicity of $\mathrm{PSL}_2(\mathbb{F}_l)$ for $l > 3$, again using the fact that $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\mu_{10})^+) \to \mathrm{GL}(\mathrm{Prim}[l']_{t'})$ is surjective.

(10) *$r$ has the right restriction to inertia.* This was guaranteed by the choice of $t'$, once we note that the twist by $\phi$ changes restriction to inertia by $\epsilon_l$.

(11) *We can choose a polarization with a square determinant.* This follows from the fact that $l'$ splits in $F^*(v)$.

Having got that $\mathrm{Prim}_{l',v,t'} \otimes \phi^{-1}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F)}$ is also automorphic, we deduce that $\mathrm{Prim}_{l,v,t'} \otimes \phi^{-1}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F)}$ is automorphic since $Y_{t'}$ has good reduction at $l$, since we chose $t'$ such that $v_{\mathfrak{w}}((t')^5 - 1) = 0$ for $\mathfrak{w}$ over $l$. Hence also $\mathrm{Prim}_{l,v,t'}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F)}$ is automorphic.

*Step 4.* Next, I claim that $\mathrm{Prim}_{l,v,t}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F)}$ is automorphic, by an appeal to Theorem 4.3.4 of [Clozel et al. 2008]. Conditions (1), (3), and (4) of that theorem are verified just like the corresponding conditions of [Barnet-Lamb 2008, Theorem 1.1.3], while (2) is trivial. We justify the remaining four conditions of Theorem 4.3.4 of [Clozel et al. 2008]:

(5) *$r$ is discrete series somewhere.* We saw in Step 2 that inertia at $q$ acts via a maximal unipotent, which suffices.

(6) *$\bar{F}^{\ker \operatorname{ad} \bar{r}}$ does not contain $F(\zeta_l)$.* True by the remarks just before Step 2.

(7) *$\bar{r}'(\mathrm{Gal}(\bar{F}/F(\zeta_l))$ is big.* Again, true by the remarks just before Step 2.

(8) *The residual representation is automorphic.* Indeed, we have just verified that $\mathrm{Prim}[l]_{v,t'}$ is automorphic, and $\mathrm{Prim}[l]_{v,t'} \equiv \mathrm{Prim}[l]_{v,t}$.

*Step 5.* We now use a rather standard argument to deduce the functional equation of the $L$ function from the potential automorphy which we have just derived. As a virtual representation of $\mathrm{Gal}(F/\mathbb{Q})$, we use Brauer's theorem to write

$$1 = \sum_j a_j \operatorname{Ind}_{\mathrm{Gal}(F/F_j)}^{\mathrm{Gal}\, F/\mathbb{Q}} \chi_j,$$

where the $F_j$ are intermediate fields between $F$ and $\mathbb{Q}$ with $\mathrm{Gal}(F/F_j)$ soluble, $a_j \in \mathbb{Z}$, and for each $j$, $\chi_j : \mathrm{Gal}(F/F_j) \to \mathbb{C}^\times$ is an isomorphism. By solvable base change, since $\mathrm{Prim}_{l,v,t}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F)}$ is automorphic, so is $\mathrm{Prim}_{l,v,t}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F_j)}$ for each $j$; that is, we can find a RAESDC representation $\pi_j$ of weight 0 and type $\{\mathrm{Sp}_n(1)\}_{\{v|q\}}$ such that for *any* rational prime $l^*$ and isomorphism $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$ we have

$$r_{l^*,\iota}(\pi_j) \equiv \mathrm{Prim}_{l^*,v,t}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F_j)},$$

where

$$\mathrm{Prim}_{l^*,v,t} = \sum_j a_j \operatorname{Ind}_{\mathrm{Gal}(F/F_j)}^{\mathrm{Gal}\, F/\mathbb{Q}} r_{l^*,\iota}\big(\pi_j \otimes (\chi_j \circ \mathrm{Art}_{F_j})\big).$$

We deduce, using [Taylor and Yoshida 2007, Theorem 3.2 and Lemma 1.3(2)], that the $L$ function of $\mathrm{Prim}_{*,v,t}$ is defined and that

$$L(\mathrm{Prim}_{*,v,t}) = \prod_j L\big(\pi_j \otimes (\chi_j \circ \mathrm{Art}_{F_j}), s\big)^{a_j},$$

which gives the result, since each of the multiplicands on the right hand side obeys the expected functional equation, whence the left hand side does too. □

We now put together what we know, to control the overall $L$ function of $\mathrm{Prim}_l$.

**Corollary 6.** *The $L$ function of* $\mathrm{Prim}_l$ *has meromorphic continuation to the whole complex plane, for $t \in \mathbb{Q} - \mathbb{Z}[1/10]$.*

*Proof.* Proposition 2 gives us a list of pieces whose $L$ functions we must control. We may control $\mathrm{Prim}_{l,[(0,0,0,0,0)]}$ by Theorem 1 and the rest by Proposition 5. □

As we have set things up, the sheaf $\mathrm{Prim}_l$ has base defined over $\mathbb{Q}(\mu_5)$; but it could also have been defined over $\mathbb{Q}$ (unlike the various pieces $\mathrm{Prim}_{l,v}$, most of which are not defined over $\mathbb{Q}$, as $\mathrm{Gal}(\mathbb{Q}(\mu_5)/\mathbb{Q})$ intermixes the various pieces). From now on, we will consider $\mathrm{Prim}_l$ to have been defined over $\mathbb{Q}$. Recapitulating the last part of the proof of Proposition 5 we get:

**Theorem 7.** *The $L$ function of* $\mathrm{Prim}_{*,t}$ *(now considered to be defined over $\mathbb{Q}$) has meromorphic continuation to the whole complex plane, for $t \in \mathbb{Q} - \mathbb{Z}[1/10]$.*

*Proof.* By Steps 1–4 of the proof of Proposition 5, there are fields $F^{(v)}$ such that $\mathrm{Prim}_{l,v,t}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/F^{(v)})}$ is automorphic for each $v$ in the table in Proposition 3; by the proof of Theorem 1 given in [Harris et al. 2010] the same is true for $v = (0, 0, 0, 0, 0)$, and by Proposition 2, the symmetry of the situation allows us to deduce this for all other $v$. We can modify the proofs of these theorems to ensure that a single field extension $F$ makes all of these representations automorphic simultaneously. (For instance, the proof of [Harris et al. 2010, Theorem 3.1] can handle multiple representations simultaneously; there are no essential difficulties other than those of bookkeeping.) Then the whole sheaf $\mathrm{Prim}_{l,t}$ becomes automorphic when restricted to $G_F$.

We can then use the argument of Step 5 of the proof of Proposition 5 with $\mathrm{Prim}_{*,t}$ taking the place of $\mathrm{Prim}_{*,v,t}$ to deduce the expected functional equation for $L(\mathrm{Prim}_{*,t}, s)$ and thus meromorphic continuation. □

**Corollary 8.** *The zeta function of $Y_t$, for $t \in \mathbb{Q} - \mathbb{Z}[1/10]$, has meromorphic continuation.*

*Proof.* By the remarks preceding Theorem 1, the remaining parts of the cohomology are well understood using the hard Lefschetz theorem. □

## 5. Concluding remarks

We have seen that the zeta function of the hypersurface with projective equation

$$(X_1^5 + X_2^5 + X_3^5 + X_4^5 + X_5^5) = 5t X_1 X_2 \dots X_5$$

has a meromorphic continuation and satisfies the expected functional equation.

It is perhaps natural to wonder whether the techniques used might generalize to more general hypersurfaces of a similar type. For instance, [Harris et al. 2010] shows that the $\Gamma_W/\Delta$ invariants in the cohomology of the variety

$$(X_1^N + X_2^N + X_3^N + X_4^N + \cdots + X_N^N) = NtX_1X_2 \ldots X_N \tag{9}$$

will be automorphic for all odd $N$, so we might wonder whether the result of this paper can be generalized to other $N$'s. The methods of [Katz 2009] work in an even more general context, replacing the monomial $X_1X_2 \ldots X_N$ on the right-hand side of the defining equation with an arbitrary monomial of the required degree, so one might also ask if there are any cases of that form to which we might try to generalize the result of this paper. I feel that a few remarks on these cases may be useful to the reader.

**5.1.** ***Smaller N's in* (9).** It is worth beginning by noting that the cases $N = 1, 2$ are trivial, and the case $N = 3$ is also uninteresting since then (9) describes a family of elliptic curves, and the zeta function is already understood. Thus the only interesting case with smaller $N$ is $N = 4$.

At first sight, it might seem difficult to analyze this case using the methods of this paper, since the result of [Harris et al. 2010] which gives the automorphicity of $\mathrm{Prim}_{l,[(0,\ldots,0)]}$ requires $N$ to be odd. But [Barnet-Lamb 2010] generalizes their methods to cover odd-dimensional cases, and it is then possible to extend the methods of this paper to cover that case, too. In particular, an analysis like that in Section 3 will reveal that all the pieces of the cohomology apart from $\mathrm{Prim}_{l,[(0,0,0,0)]}$ are one- or zero-dimensional, and so trivially automorphic.

I have chosen not to give this argument in full detail, since a very beautiful geometric argument of N. Elkies and M. Schütt (personal communication) tells us that, for the case $N = 4$, each Dwork hypersurface is isogenous to the Kummer surface of a product $E_1 \times E_2$, where $E_1$ and $E_2$ are elliptic curves defined over a quadratic extension of $\mathbb{Q}$, conjugate to each other over $\mathbb{Q}$, and related by a 2-isogeny. This allows one to quite directly see the automorphicity required in this case, and little would be served by giving the full details of the argument above.

**5.2.** ***Larger values of N in* (9).** If we try to extend the methods of this paper to larger $N$, we face the following problem.

**Proposition 9.** (1) *Let $N \geq 8$ be an integer. Then the Hodge–Tate numbers of $\mathrm{Prim}_{l,[(4,N-2,N-2,0,\ldots,0)]}$ include 2 with multiplicity at least 2.*

(2) *Let $N = 6$. Then the Hodge–Tate numbers of $\mathrm{Prim}_{l,[(0,0,0,2,2,2)]}$ include 3 with multiplicity at least 2.*

*Proof.* Again, we use [Katz 2009, Lemma 3.1]. For the first statement, the totally nonzero representatives include $(5, N-1, N-1, 1, \ldots, 1)$ and $(7, 1, 1, 3, \ldots, 3)$.

We have
$$\frac{5+(N-1)+(N-1)+1+\cdots+1}{N}-1=\frac{7+1+1+3+\cdots+3}{N}-1=2,$$
so 2 occurs as a Hodge–Tate weight with multiplicity at least 2.

The proof of the second statement is similar: consider the nonzero representatives (2, 2, 2, 4, 4, 4) and (5, 5, 5, 1, 1, 1). □

Thus in all cases with $N$ even (recall that we need $N$ even for [Harris et al. 2010] to apply[3]) and $N \geq 6$, at least one of the pieces of the cohomology of (9) will have a repeated Hodge–Tate number. At present, apart from some work in the case of two-dimensional Galois representations, there are no modularity lifting theorems for representations with repeated Hodge–Tate numbers, and hence (since such theorems are a key ingredient in proving the potential modularity theorems such as [Taylor 2006; Harris et al. 2010] on which this paper relies) it seems unlikely that the approach of this paper can be extended to cover such cases.

(One might briefly wonder whether some larger algebra of correspondences could be used to cut the cohomology into smaller pieces, small enough that they no longer have repeated Hodge–Tate weights; but this is impossible, since the results of Katz on the monodromy of the cohomology tell us that all the pieces in the decomposition of the cohomology into eigenspaces for $\Gamma_W/\Delta$ cannot be broken up further, as the monodromy acts transitively on each piece.)

**5.3.** ***Other monomials in* (9).** Katz studies the more general equation
$$(X_1^N+X_2^N+X_3^N+\cdots+X_N^N)=N\lambda\prod_i X_i^{w_i}, \tag{10}$$
where $W=(w_1,\ldots,w_N)$ is a sequence of nonnegative integers summing to $N$. It is natural to ask whether the methods of this paper can be extended to any varieties of this form, beyond the cases already considered. Unfortunately, the answer is no.

Let's imagine an analysis based on the same techniques used above. As before, the main challenge would be to analyze the middle-dimensional cohomology, since the rest is determined by hard Lefschetz. We can define $\mathrm{Prim}_l^{N-2}$, as in [Katz 2009], to correspond to the part of the middle-dimensional cohomology not coming from Lefschetz. Following the method above, our next step is to decompose this cohomology into eigensheaves.

The natural group acting on (10) is easily seen to be
$$\Big\{(\zeta_1,\ldots,\zeta_N)\in(\mu_N)^N\;\Big|\;\prod\zeta_i^{w_i}=1\Big\}\Big/\Delta,$$
where $\Delta$, as before, is $\mu_N$ embedded diagonally. The character group in this case is

[3]Even if this were not an obstacle, the $N=7$ case also has a piece of the cohomology with a repeated Hodge–Tate number.

$(\mathbb{Z}/N\mathbb{Z})_0/\langle W\rangle$, where we abuse notation by considering $W$ as a class in $(\mathbb{Z}/N\mathbb{Z})_0$. We will write an element of $(\mathbb{Z}/N\mathbb{Z})_0/\langle W\rangle$ as either $v \bmod W$ or simply $[v]$ and define $\mathrm{Prim}^{N-2}_{l,[v]}$ in a similar manner as before.

Suppose now that we have fixed some $W$. The main challenge in applying the methods to this paper to show that the zeta function of the family (10) is meromorphic will be showing that $\mathrm{Prim}^{N-2}_{l,[v]}$ is automorphic for each $v$. Since this will rely, in the final analysis, on the application of a lifting theorem, we will certainly require that $\mathrm{Prim}^{N-2}_{l,[v]}$ has distinct Hodge numbers for all $v$. This, unfortunately, will never happen except in the cases already considered.

**Proposition 10.** *Suppose $N \geq 3$ and $W \neq (1, 1, \dots, 1)$. Then there exists some $v$ such that the Hodge–Tate numbers of $\mathrm{Prim}^{N-2}_{l,[v]}$ are not all distinct.*

*Proof.* Since $W \neq (1, 1, \dots, 1)$, some $w_i$ — say $w_1$ — vanishes. For $i > 1$, write $h_i$ for $\mathrm{hcf}(w_i, N)$. For each $i > 1$ where $h_i = 1$, set $v_i = 0$; then, for $k \in \mathbb{Z}$,

$$v_i + kw_i \equiv 0 \pmod N \quad\iff\quad k \equiv 0 \pmod N.$$

For $i > 1$ where $h_i > 1$, we can choose $v_i = 1$; then $v_i + kw_i$ will always be $\equiv 1 \bmod h_i$, and we will *never* have $v_i + kw_i \equiv 0 \bmod N$. Finally, $v_1$ is fixed by the condition that the $v_i$ sum to 0. (Note that $\tilde{v}_1 > 1$, since if not $h_i = 1$ for all $i > 1$, hence $v_i > 1$ for all such $i$, which is impossible.)

The elements of $(\mathbb{Z}/N\mathbb{Z})_0$ representing $[v]$ are $v + kW$ for $k \in \mathbb{Z}/N\mathbb{Z}$; it is immediate that the element $v + kW$ is totally nonzero for all $k \in \mathbb{Z}/N\mathbb{Z}$ except $k = 0$, so the totally nonzero representatives are $\{v, v + W, \dots, v + (N-1)W\}$, and the multiset of Hodge–Tate numbers is the multiset

$$\{\deg(v) - 1, \deg(v + W) - 1, \deg(v + 2W) - 1, \dots, \deg(v + (N-1)W - 1\}.$$

If we suppose for contradiction that these numbers are distinct, the $N-1$ integers $\deg(v+W), \deg(v+2W), \dots, \deg(v+(N-1)W)$ are distinct. We now note that the degree of an element of $(\mathbb{Z}/N\mathbb{Z})_0$ is trivially $\leq N-1$, and that, writing $u$ for some $v + kW$, we have $N \deg u \geq \tilde{u}_1 + (N-1)$ since $u$ is totally nonzero), and this in turns equals $\tilde{v}_1 + (N-1) > N$, since $w_1 = 0$. Since $\deg(v + kW) > 1$, we have $N-1$ distinct integers $\deg(v + kW)$ with $1 < \deg(v + kW) \leq N-1$, a contradiction. $\square$

***Acknowledgements.*** I thank my advisor, Richard Taylor, for suggesting this problem to me and for his help in pursuing it. I am grateful to Barry Mazur for helpful suggestions made concerning this material when he examined it as part of my PhD thesis. I also thank the anonymous referee for further helpful suggestions.

## References


[Barnet-Lamb 2008] T. Barnet-Lamb, "Potential automorphy for certain Galois representations to $GL(2n)$", preprint, 2008. To appear in *J. Reine Angew. Math.* arXiv 0811.1586v1

[Barnet-Lamb 2010] T. Barnet-Lamb, "On the potential automorphy of certain odd-dimensional Galois representations", *Compos. Math.* **146**:3 (2010), 607–620. MR 2644929 Zbl 05722159

[Barnet-Lamb et al. 2010] T. Barnet-Lamb, T. Gee, D. Geraghty, and R. Taylor, "Potential automorphy and change of weight", preprint, 2010. arXiv 1010.2561v1

[Clozel et al. 2008] L. Clozel, M. Harris, and R. Taylor, "Automorphy for some $l$-adic lifts of automorphic mod $l$ Galois representations", *Publ. Math. Inst. Hautes Études Sci.* 108 (2008), 1–181. MR 2470687 (2010j:11082) Zbl 1169.11020

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

[Ekedahl 1990] T. Ekedahl, "An effective version of Hilbert's irreducibility theorem", pp. 241–249 in *Séminaire de Théorie des Nombres* (Paris, 1988–1989), edited by C. Goldstein, Progr. Math. **91**, Birkhäuser, Boston, MA, 1990. MR 92f:14018 Zbl 0729.12005

[Harris et al. 2010] M. Harris, N. Shepherd-Barron, and R. Taylor, "A family of Calabi–Yau varieties and potential automorphy", *Ann. of Math.* (2) **171**:2 (2010), 779–813. MR 2630056 Zbl 05712744

[Katz 1988] N. M. Katz, *Gauss sums, Kloosterman sums, and monodromy groups*, Annals of Mathematics Studies **116**, Princeton University Press, 1988. MR 91a:11028 Zbl 0675.14004

[Katz 2009] N. M. Katz, "Another look at the Dwork family", pp. 89–126 in *Algebra, arithmetic, and geometry: in honor of Y. I. Manin*, vol. II, edited by Y. Tschinkel, Progr. Math. **270**, Birkhäuser, Boston, 2009. MR 2641188 Zbl 1195.14015

[Larsen 1995] M. Larsen, "Maximality of Galois actions for compatible systems", *Duke Math. J.* **80**:3 (1995), 601–630. MR 97a:11090 Zbl 0912.11026

[Matthews et al. 1984] C. R. Matthews, L. N. Vaserstein, and B. Weisfeiler, "Congruence properties of Zariski-dense subgroups, I", *Proc. London Math. Soc.* (3) **48**:3 (1984), 514–532. MR 85d:20040 Zbl 0551.20029

[Nori 1987] M. V. Nori, "On subgroups of $\mathrm{GL}_n(\mathbf{F}_p)$", *Invent. Math.* **88**:2 (1987), 257–275. MR 88d:20068 Zbl 0632.20030

[Scholl 2006] A. J. Scholl, "On some $l$-adic representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ attached to noncongruence subgroups", *Bull. London Math. Soc.* **38**:4 (2006), 561–567. MR 2007c:11065 Zbl 1126.11029

[Taylor 2006] R. Taylor, "On the meromorphic continuation of degree two $L$-functions", *Doc. Math.* Extra Vol. (2006), 729–779. MR 2008c:11154 Zbl 1138.11051

[Taylor and Yoshida 2007] R. Taylor and T. Yoshida, "Compatibility of local and global Langlands correspondences", *J. Amer. Math. Soc.* **20**:2 (2007), 467–493. MR 2007k:11193 Zbl 05120650


Communicated by Hélène Esnault
Received 2009-01-16 Revised 2010-08-09 Accepted 2010-08-10

tbl@math.harvard.edu
*Department of Mathematics, Brandeis University, 415 South Street MS 050, Waltham, MA 02453, United States*
http://people.brandeis.edu/~tbl/

ALGEBRA AND NUMBER THEORY 4:7(2010)

msp

# Equations for Chow and Hilbert quotients

Angela Gibney and Diane Maclagan

We give explicit equations for the Chow and Hilbert quotients of a projective scheme $X$ by the action of an algebraic torus $T$ in an auxiliary toric variety. As a consequence we provide geometric invariant theory descriptions of these canonical quotients, and obtain other GIT quotients of $X$ by variation of GIT quotient. We apply these results to find equations for the moduli space $\overline{M}_{0,n}$ of stable genus-zero $n$-pointed curves as a subvariety of a smooth toric variety defined via tropical methods.

## 1. Introduction

When a reductive group $G$ acts linearly on a projective scheme $X$, a fundamental problem is to describe a good notion of a quotient $X/G$. This question frequently arises in the construction and compactification of moduli spaces. In many situations, there is an open subset $U \subset X$ on which $G$ acts freely, such that a scheme $U/G$ exists as a geometric quotient. Constructing the quotient $X/G$ is thus choosing a good compactification of $U/G$. One way to compactify is by forming the Chow quotient $X /\!\!/^{\mathrm{Ch}} G$ or Hilbert quotient $X /\!\!/^{\mathrm{H}} G$ of $X$ by $G$ (see [Kapranov 1993]). These quotients are taken to be the closure of $U/G$ in an appropriate Chow variety or Hilbert scheme. They are natural canonical quotients with proper birational maps to any GIT (geometric invariant theory) quotient. See also [Hu 2005; Keel and Tevelev 2006].

In this paper we treat the case where $G = T^d$ is a $d$-dimensional algebraic torus acting equivariantly on a subscheme $X$ of $\mathbb{P}^m$. Given the ideal $I$ of $X$ as a subscheme of $\mathbb{P}^m$, we describe equations for $X /\!\!/^{\mathrm{Ch}} T^d$ and $X /\!\!/^{\mathrm{H}} T^d$ in the Cox rings of toric subvarieties of the Chow and Hilbert quotients of $\mathbb{P}^m$.

As a first application of our results, we give GIT constructions of $X /\!\!/^{\mathrm{Ch}} T^d$ and $X /\!\!/^{\mathrm{H}} T^d$ and we prove that all GIT quotients of $X$ by $T^d$ can be obtained from the Chow and Hilbert quotients by variation of the GIT.

As a second application we study the action of an $(n-1)$-dimensional torus $T^{n-1}$ on the Grassmannian $G(2, n)$. Here we can take $U$ to be the points with

*MSC2000:* primary 14L30; secondary 14M25, 14L24, 14H10.

*Keywords:* Chow quotient, Hilbert quotient, moduli of curves, space of phylogenetic trees.

nonvanishing Plücker coordinates, and the quotient $U/T^{n-1}$ is the moduli space $M_{0,n}$ of smooth $n$-pointed genus-zero curves. In this case the desired compactification is the celebrated moduli space $\overline{M}_{0,n}$ of stable $n$-pointed genus-zero curves. [Kapranov 1993] showed that $\overline{M}_{0,n}$ is isomorphic to both the Chow and Hilbert quotients of $G(2,n)$ by the $T^{n-1}$-action. We give explicit equations for $\overline{M}_{0,n}$ as a subvariety of a smooth toric variety $X_\Delta$ whose fan is the well-studied space of phylogenetic trees. We show that the equations for $\overline{M}_{0,n}$ in the Cox ring $S$ of $X_\Delta$ are generated by the Plücker relations homogenized with respect to the grading of $S$.

We now describe our results in more detail. The notation $X/\!\!/^\star T^d$ is used to refer to either the Chow or the Hilbert quotient. We assume that no irreducible component of $X$ lies in any coordinate subspace. This means that $X/\!\!/^\star T^d$ is a subscheme of $\mathbb{P}^m/\!\!/^\star T^d$. The quotient $\mathbb{P}^m/\!\!/^\star T^d$ is a not necessarily normal toric variety [Kapranov et al. 1992] whose normalization we denote by $X_{\Sigma^\star}$. By $X/\!\!/_n^\star T^d$ we mean the pullback of $X/\!\!/^\star T^d$ to $X_{\Sigma^\star}$. Our main theorem, in slightly simplified form, is the following. This is proved in Theorems 3.2 and 4.6 and Proposition 4.3.

**Theorem 1.1.** *Let $T^d \cong (\Bbbk^\times)^d$ act on $\mathbb{P}^m$ and let $X \subset \mathbb{P}^m$ be a $T^d$-equivariant subscheme with corresponding ideal $I(X) = \langle f_1, \ldots, f_g\rangle \subset \Bbbk[x_0, \ldots, x_m]$. Let $X_\Sigma \subset X_{\Sigma^\star}$ be any toric subvariety with $X \subseteq X_\Sigma \subseteq X_{\Sigma^\star}$.*

(1) (Equations) *The ideal $I$ of the Hilbert or Chow quotient $X/\!\!/_n^\star T^d$ in the Cox ring $S = \Bbbk[y_1, \ldots y_r]$ of $X_\Sigma$ can be computed effectively. Explicitly, $I$ is obtained by considering the $f_i$ as polynomials in $y_1, \ldots, y_{m+1}$, homogenizing them with respect to the* $\mathrm{Cl}(X_\Sigma)$*-grading of $S$, and then saturating the result by the product of all the variables in $S$.*

(2) (GIT) *There is a GIT construction of the Chow and Hilbert quotients of $X$, and these are related to the GIT quotients of $X$ by variation of the GIT quotient. This gives equations for all quotients in suitable projective embeddings. Let* $H = \mathrm{Hom}(\mathrm{Cl}(X_\Sigma), \Bbbk^\times)$. *There is a nonzero cone $\mathcal{G} \subset \mathrm{Cl}(X_\Sigma) \otimes \mathbb{R}$ for which $X/\!\!/_n^\star T^d$ is the GIT quotient*

$$X/\!\!/_n^\star T^d = Z(I)/\!\!/_\alpha H$$

*for any rational $\alpha \in$* $\mathrm{relint}(\mathcal{G})$, *where $Z(I)$ is the subscheme of $\mathbb{A}^r$ defined by $I$. For any GIT quotient $X/\!\!/_\beta T^d$ of $X$, there are choices of $\alpha$ outside $\mathcal{G}$ for which $Z(I)/\!\!/_\alpha H = X/\!\!/_\beta T^d$.*

A more precise formulation of the homogenization is given in Theorem 3.2 and Remark 3.3. We explain in Corollary 4.4 how each choice of $\alpha \in \mathrm{relint}(\mathcal{G})$ gives an embedding of $X/\!\!/_n^\star T^d$ into some projective space.

We use tropical algebraic geometry in the spirit of [Tevelev 2007] to embed $\overline{M}_{0,n}$ in a smooth toric variety $X_\Delta$. The combinatorial data describing $\Delta$ and the simple equations for $\overline{M}_{0,n}$ in the Cox ring of $X_\Delta$ are described in the following theorem. Let $[n] = \{1, \ldots, n\}$ and set $\mathcal{I} = \{I \subset [n] : 1 \in I,\ |I| \geq 2,\ |[n] \setminus I| \geq 2\}$. The set $\mathcal{I}$ indexes the boundary divisors of $\overline{M}_{0,n}$.

**Theorem 1.2.** *Let $\Delta$ be the fan in $\mathbb{R}^{\binom{n}{2}-n}$ described in Section 5 (the space of phylogenetic trees). The rays of $\Delta$ are indexed by the set $\mathcal{I}$.*

(1) (Equations) *Equations for $\overline{M}_{0,n}$ in the Cox ring $S = \Bbbk[x_I : I \in \mathcal{I}]$ of $X_\Delta$ are obtained by homogenizing the Plücker relations with respect to the grading of $S$ and then saturating by the product of the variables of $S$. Specifically, the ideal is*

$$I_{\overline{M}_{0,n}} = \Bigg(\Big\langle \prod_{\substack{i,j\in I\\ k,l\notin I}} - \prod_{\substack{i,k\in I\\ j,l\notin I}} x_I + \prod_{\substack{i,l\in I\\ j,k\notin I}} x_I \Big\rangle : \Big(\prod_I x_I\Big)^\infty\Bigg),$$

*where the generating set runs over all $\{i, j, k, l\}$ with $1 \leq i < j < k < l \leq n$, and $x_I = x_{[n]\setminus I}$ if $1 \notin I$.*

(2) (GIT) *There is a nonzero cone $\mathcal{G} \subset \mathrm{Cl}(X_\Delta) \otimes \mathbb{R} \cong \mathrm{Pic}(\overline{M}_{0,n}) \otimes \mathbb{R}$ for which for rational $\alpha \in \mathrm{int}(\mathcal{G})$ we have the GIT construction of $\overline{M}_{0,n}$ as*

$$\overline{M}_{0,n} = Z(I_{\overline{M}_{0,n}}) /\!\!/_\alpha H,$$

*where $Z(I_{\overline{M}_{0,n}}) \subset \mathbb{A}^{|\mathcal{I}|}$ is the affine subscheme defined by $I_{\overline{M}_{0,n}}$, and $H$ is the torus* $\mathrm{Hom}(\mathrm{Cl}(X_\Delta), \Bbbk^\times) \cong (\Bbbk^\times)^{|\mathcal{I}|-\binom{n}{2}+n}$.

(3) (VGIT) *Given $\beta \in \mathbb{Z}^n$ there is $\alpha \in \mathbb{Z}^{|\mathcal{I}|-\binom{n}{2}+n}$ for which*

$$Z(I_{\overline{M}_{0,n}}) /\!\!/_\alpha H = G(2, n) /\!\!/_\beta T^{n-1},$$

*so all GIT quotients of $G(2, n)$ by $T^{n-1}$ can be obtained from $\overline{M}_{0,n}$ by variation of the GIT.*

Statement (3) relates to [Howard et al. 2009], where GIT quotients of $G(2, n)$ by $T^{n-1}$, or equivalently of $(\mathbb{P}^1)^n$ by $\mathrm{Aut}(\mathbb{P}^1)$, were studied.

Keel and Tevelev, in an article titled "Equations for $\overline{M}_{0,n}$" [2009], studied the image of the particular embedding of $\overline{M}_{0,n}$ into a product of projective spaces given by the complete linear series of the very ample divisor $\kappa = K_{\overline{M}_{0,n}} + \sum_{I\in\mathcal{I}} \delta_I$. Theorem 1.2 concerns projective embeddings of $\overline{M}_{0,n}$ corresponding to a full-dimensional subcone of the nef cone of $\overline{M}_{0,n}$, including that given by $\kappa$.

A key idea of this paper is to work in the Cox ring of sufficiently large toric subvarieties of $X_{\Sigma^\star}$. This often allows one to give equations in fewer variables. Also, a truly concrete description of $X_{\Sigma^\star}$ may be cumbersome or impossible, as in the case of $\overline{M}_{0,n}$, but a sufficiently large toric subvariety such as $X_\Delta$ can often be obtained.

We now summarize the structure of the paper. Section 2 contains some tools from toric geometry that will be useful in the rest of the paper. The first part of Theorem 1.1 is proved in Section 3, while the GIT results are proved in Section 4. In Section 5 we explicitly describe the toric variety $X_\Delta$ that contains $\overline{M}_{0,n}$. In Sections 6 and 7 we prove Theorem 1.2. We end the paper with some natural questions about $\overline{M}_{0,n}$ arising from this work.

## 2. Toric tools

In this section we develop some tools to work with toric varieties that will be used in our applications to Chow and Hilbert quotients. We generally follow the notational conventions for toric varieties of [Fulton 1993], with the exception that we do not always require normality. Throughout $\Bbbk$ is an algebraically closed field, and $\Bbbk^\times = \Bbbk \setminus \{0\}$. We denote by $T^d$ an algebraic torus isomorphic to $(\Bbbk^\times)^d$. If $I$ is an ideal in $\Bbbk[x_0, \dots, x_m]$ then $Z(I)$ is the corresponding subscheme of either $\mathbb{A}^{m+1}$ or $\mathbb{P}^m$ depending on the context.

**2A.** ***Producing equations for quotients of subvarieties of tori.*** We first describe how to obtain equations for the quotient of a subvariety of a torus by a subtorus. Let $Y$ be a subscheme of a torus $T^m$ that is equivariant under a faithful action of $T^d$ on $T^m$ given by $(t \cdot x)_j = (\prod_{i=1}^d t_i^{a_{ij}})x_j$, and let $I(Y) \subset \Bbbk[x_1^{\pm 1}, \dots, x_m^{\pm 1}]$ be the ideal of $Y$. Write $A$ for the $d \times m$ matrix with $ij$-th entry $a_{ij}$. Let $D$ be a $(m-d) \times m$ matrix of rank $m-d$ whose rows generate the integer kernel of $A$, so $AD^T = 0$. The matrix $D$ is a *Gale dual* for the $d \times m$ matrix $A = (a_{ij})$ (see [Ziegler 1995, Chapter 6]).

**Proposition 2.1.** *Let $\phi : \Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}] \to \Bbbk[x_1^{\pm 1}, \dots, x_m^{\pm 1}]$ be given by $\phi(z_i) = \prod_{j=1}^m x_j^{D_{ij}}$. Then the ideal of $Y/T^d$ in the coordinate ring $\Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}]$ of $T^m/T^d$ is given by $\phi^{-1}(I(Y))$. This is generated by polynomials $g_1, \dots, g_s$ for which $I(Y) = \langle \phi(g_1), \dots, \phi(g_s) \rangle$.*

*Proof.* The coordinate ring of the quotient $Y/T^d$ is by definition the ring of invariants of $\Bbbk[x_1^{\pm 1}, \dots, x_m^{\pm 1}]/I(Y)$ under the induced action of $T^d$. The $T^d$ action on $T^m$ gives a $\mathbb{Z}^d$-grading of $\Bbbk[x_i^{\pm 1}]$ by setting $\deg x_i = \boldsymbol{a}_i$, where $\boldsymbol{a}_i$ is the $i$-th column of the matrix $A$. Since $T^d$ acts equivariantly on $Y$, the ideal $I(Y)$ is homogeneous with respect to this grading, so $\Bbbk[x_i^{\pm 1}]/I(Y)$ is also $\mathbb{Z}^d$-graded. The ring of invariants is precisely the degree-zero part of this ring.

To prove that this is isomorphic to $k[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}]/\phi^{-1}(I(Y))$, we first define an automorphism of the torus $T^m$ so that $T^d$ is mapped to the subtorus having first $m-d$ coordinates equal to one. Choose any $m \times m$ integer matrix $U$ with determinant one whose first $d$ rows consist of the matrix $D$. This is possible because by the definition of $D$ the cokernel $\mathbb{Z}^m / \operatorname{im}(D^T)$ is torsion-free, so $\mathbb{Z}^m \cong \operatorname{im}(D^T) \oplus \mathbb{Z}^d$.

Then the map $\tilde{\phi} : \Bbbk[z_1^{\pm 1}, \dots, z_m^{\pm 1}] \to \Bbbk[x_1^{\pm 1}, \dots, x_m^{\pm 1}]$ defined by $\tilde{\phi}(z_i) = \prod_{j=1}^m x_j^{U_{ij}}$ determines an automorphism of the torus $T^m$. Note that the map $\phi$ is $\tilde{\phi}$ restricted to the ring $\Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}]$.

The ring $\Bbbk[z_1^{\pm 1}, \dots, z_m^{\pm 1}]$ gets an induced $\mathbb{Z}^d$-grading from the grading on $\Bbbk[x_j^{\pm 1}]$ by setting $\deg z_i = \sum_{j=0}^m U_{ij} \boldsymbol{a}_j$, which is the $i$-th column of $AU^T$. Since the first $m-d$ rows of $U$ are the rows of $D$, and $AD^T = 0$, we thus have $\deg z_i = 0 \in \mathbb{Z}^d$ for $1 \leq i \leq m-d$. The degrees of the $d$ variables $z_{m-d+1}, \dots, z_m$ are linearly independent, since $\operatorname{rank}(AU^T)$ is $d$. This means that the degree zero part of $\Bbbk[z_1^{\pm 1}, \dots, z_m^{\pm 1}]$ is $\Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}]$, which proves that the coordinate ring of $Y/T^d$ is given by $\Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}]/J$, where $J = \tilde{\phi}^{-1}(I(Y)) \cap \Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}]$. The result then follows since $J = \phi^{-1}(I(Y))$. The statement about generators follows from the fact that $\phi$ is injective, since $\tilde{\phi}$ is an isomorphism. $\square$

**Example 2.2.** Let $Y$ be the subscheme of $T^{10}$ defined by the ideal

$$\begin{aligned} I = \langle & x_{12}x_{34} - x_{13}x_{24} + x_{14}x_{23},\ x_{12}x_{35} - x_{13}x_{25} + x_{15}x_{23}, \\ & x_{12}x_{45} - x_{14}x_{25} + x_{15}x_{24},\ x_{13}x_{45} - x_{14}x_{35} + x_{15}x_{34},\ x_{23}x_{45} - x_{24}x_{35} + x_{25}x_{34}\rangle \\ \subseteq {} & \Bbbk[x_{ij}^{\pm 1} : 1 \leq i < j \leq 5]. \end{aligned}$$

This is the intersection with the torus of $\mathbb{A}^{10}$ of the affine cone over the Grassmannian $G(2,5)$ in its Plücker embedding into $\mathbb{P}^9$. The torus $T^5$ acts equivariantly on $Y$ by $t \cdot x_{ij} = t_i t_j x_{ij}$, giving rise to matrices

$$A = \begin{pmatrix} 1&1&1&1&0&0&0&0&0&0 \\ 1&0&0&0&1&1&1&0&0&0 \\ 0&1&0&0&1&0&0&1&1&0 \\ 0&0&1&0&0&1&0&1&0&1 \\ 0&0&0&1&0&0&1&0&1&1 \end{pmatrix}, \qquad D = \begin{pmatrix} 0&1&-1&0&-1&1&0&0&0&0 \\ 0&1&0&-1&-1&0&1&0&0&0 \\ 1&0&-1&0&-1&0&0&1&0&0 \\ 1&0&0&-1&-1&0&0&0&1&0 \\ 1&1&-1&-1&-1&0&0&0&0&1 \end{pmatrix},$$

where the columns are ordered $\{12, 13, \dots, 35, 45\}$. The map $\phi : \Bbbk[z_1^{\pm 1}, \dots, z_5^{\pm 1}] \to \Bbbk[x_{ij}^{\pm 1} : 1 \leq i < j \leq 5]$ is given by $\phi(z_1) = x_{13}x_{24}/x_{14}x_{23}$, $\phi(z_2) = x_{13}x_{25}/x_{15}x_{23}$, $\phi(z_3) = x_{12}x_{34}/x_{14}x_{23}$, $\phi(z_4) = x_{12}x_{35}/x_{15}x_{23}$, and $\phi(z_5) = x_{12}x_{13}x_{45}/x_{14}x_{15}x_{23}$.

The ideal $\phi^{-1}(I)$ is then

$$\begin{aligned} &\langle z_3 - z_1 + 1, z_4 - z_2 + 1, z_5 - z_2 + z_1, z_5 - z_4 + z_3, z_5 - z_1 z_4 + z_2 z_3\rangle \\ &\qquad = \langle z_3 - z_1 + 1, z_4 - z_2 + 1, z_5 - z_2 + z_1\rangle \subset \Bbbk[z_1^{\pm 1}, \dots, z_5^{\pm 1}]. \end{aligned}$$

For this it is essential that we work in the Laurent polynomial ring; for example, $\phi(z_5 - z_2 + z_1) = x_{13}/(x_{14}x_{15}x_{23})(x_{12}x_{45} - x_{14}x_{25} + x_{15}x_{24})$. The variety of $\phi^{-1}(I)$ is the moduli space $M_{0,5}$. Note that this shows that $M_{0,5}$ is a complete intersection in $T^5$, cut out by three linear equations. This example is continued in Example 3.1 and Sections 5, 6, and 7.

**2B.** ***Producing equations for closures in toric varieties.*** The next proposition describes how to find the ideal of the closure of a subvariety of a torus in a toric variety. We use the notation $T^{m-d}$ for ease of connection with the rest of this section, but there is no requirement that this torus be obtained as a quotient.

Recall that the Cox ring of a normal toric variety $X_\Sigma$ (see [Cox 1995; Mustaţă 2002]) is the polynomial ring $S = \Bbbk[y_1, \dots, y_r]$, where $r = |\Sigma(1)|$ is the number of rays of $\Sigma$. It is graded by the divisor class group of $X_\Sigma$, so that $\deg y_i = [D_i]$, where $[D_i] \in \mathrm{Cl}(X_\Sigma)$ is the class of the torus-invariant divisor $D_i$ associated to the $i$-th ray $\rho_i$ of $\Sigma$. An ideal $I \subset S$ determines an ideal sheaf $\tilde{I}$ on $X_\Sigma$, and thus a closed subscheme of $X_\Sigma$, and conversely, every ideal sheaf on $X_\Sigma$ is of the form $\tilde{I}$ for some ideal $I$ of $S$ (Theorem 1.1 of [Mustaţă 2002] removes the need for the simplicial hypothesis in [Cox 1995, Theorem 3.7]). The sheaf $\tilde{I}$ is given on an affine chart $U_\sigma$ of $X_\Sigma$ by $I_\sigma = (IS_{\prod_{i \notin \sigma} y_i})_{\mathbf{0}}$. The correspondence between ideals in $S$ and closed subschemes of $X_\Sigma$ is not bijective, but for any closed subscheme $Z \subset X_\Sigma$ there is a largest ideal $I(Z) \subset S$ with $\widetilde{I(Z)} = \mathcal{I}_Z$.

Recall also that if $I$ is an ideal in a ring $R$ and $y \in R$, then $(I : y^\infty) = \{r \in R : ry^k \in I \text{ for some } k > 0\}$. Geometrically this removes irreducible components supported on the variety of $y$.

**Proposition 2.3.** *Let $X_\Sigma$ be an $(m-d)$-dimensional toric variety with Cox ring $S = \Bbbk[y_1, \dots, y_r]$. Set $y = \prod_{i=1}^r y_i$ so that*

$$\rho : \Bbbk[T^{m-d}] = \Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}] \longrightarrow (S_y)_0$$

*is the isomorphism given by the inclusion of the torus $T^{m-d}$ into $X_\Sigma$. If $Y \subset T^{m-d}$ is given by ideal $I(Y) = \langle f_1, \dots, f_s \rangle \subset \Bbbk[T^{m-d}]$, then the ideal $I$ for the closure $\overline{Y}$ of $Y$ in $X_\Sigma$ is $(\rho(I(Y))S_y) \cap S$, which is*

$$I = \left(\left\langle \widetilde{\rho(f_i)} : 1 \le i \le s \right\rangle : y^\infty\right),$$

*where $\widetilde{\rho(f_i)}$ is obtained by clearing the denominator of $\rho(f_i)$.*

*Proof.* Let $J = \rho(I(Y)) \subset (S_y)_0$. The closure $\overline{Y}$ of $Y$ in $X_\Sigma$ is the smallest closed subscheme of $X_\Sigma$ containing $Y$. Since the torus $T^{m-d}$ is the affine toric variety corresponding to the cone $\sigma$ consisting of just the origin in $\Sigma$, for this $\sigma$ we have $I_\sigma = (IS_y)_{\mathbf{0}}$ for any ideal $I \subseteq S$. As the correspondence between subschemes $Z$ of $X_\Sigma$ and ideals $I(Z)$ of $S$ is inclusion reversing, we have that $I(\overline{Y})$ is the largest ideal $I$ in $S$ with $(IS_y)_0 \subseteq J$ for which $I = I(Z)$ for some subscheme $Z$ of $X_\Sigma$.

There is a monomial, and thus a unit, in any degree $a$ for which $(S_y)_a$ is nonzero, so if $K$ is an ideal in $S_y$, then $K_0 S_y = K$, and $(K \cap S)S_y = K$. Thus $((JS_y \cap S)S_y)_0 = (JS_y)_0 = J$, and if $I$ is any homogeneous ideal in $S$ with $(IS_y)_0 \subset J$ then $I \subseteq IS_y \cap S \subset JS_y \cap S$. Thus to show that $I = JS_y \cap S$ is the ideal of $I(\overline{Y})$, we need only show that $I$ is of the form $I(Z)$ for some subscheme $Z$ of

$X_\Sigma$. Indeed, let $I'$ be the largest ideal in $S$ with $\tilde{I}' = \tilde{I}$. Then by construction we have $(I'S_y)_0 = (IS_y)_0 = J$, so by above $I' \subseteq I$, and thus $I = I'$. This shows that $I(\overline{Y}) = JS_y \cap S$.

Suppose now that $I(Y)$ is generated by $\{f_1, \dots, f_s\} \subset \Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}]$. Then $JS_y$ is generated by $\{\rho(f_1), \dots, \rho(f_s)\}$. The denominator of each $\rho(f_i)$ is a monomial, which is a unit in $S_y$, so $JS_y$ is also generated by the polynomials $\widetilde{\rho(f_i)}$ obtained by clearing the denominators in the $\rho(f_i)$. The result follows from the observation that if $K$ is an ideal in $S_y$, generated by $\{g_1, \dots, g_s\} \subset S$, then $K \cap S = (\langle g_1, \dots, g_s \rangle : y^\infty)$. See, for example, [Eisenbud 1995, Exercise 2.3]. $\square$

**2C.** ***Sufficiently large subvarieties via tropical geometry.*** A key idea of this paper is to work in toric varieties whose fan has few cones. This is made precise in the following definition.

**Definition 2.4.** A toric subvariety $X_\Delta$ of a toric variety $X_\Sigma$ with torus $T^m$ is *sufficiently large* with respect to a subvariety $Y \subseteq X_\Sigma$ if the fan $\Delta$ contains all cones of $\Sigma$ corresponding to $T^m$-orbits of $X_\Sigma$ that intersect $Y$.

Tropical geometry provides the tools to compute whether a given toric subvariety $X_\Delta \subseteq X_\Sigma$ is sufficiently large for $Y \subseteq X_\Sigma$. We now review the version we use in this paper. Let $Y \subset T^m \cong (\Bbbk^\times)^m$ be a subscheme defined by the ideal $I = I(Y) \subset S := \Bbbk[x_1^{\pm 1}, \dots, x_m^{\pm 1}]$. Given a vector $w \in \mathbb{R}^m$ we can compute the leading form $\mathrm{in}_w(f)$ of a polynomial $f \in S$, which is the sum of those terms $c_u x^u$ in $f$ with $w \cdot u$ minimal. The initial ideal $\mathrm{in}_w(J)$ is $\langle \mathrm{in}_w(f) : f \in J \rangle$. Let $K$ be any algebraically closed field extension of $\Bbbk$ with a nontrivial valuation $\mathrm{val} : K^\times \to \mathbb{R}$ such that $\mathrm{val}(\Bbbk) = 0$. We denote by $V_K(I)$ the set $\{u \in (K^\times)^m : f(u) = 0 \text{ for all } f \in I\}$.

**Definition/Theorem 2.5.** *Let $Y$ be a subvariety of $T^m$. The tropical variety of $Y$, denoted* $\mathrm{Trop}(Y)$, *is the closure in $\mathbb{R}^m$ of the set*

$$\big\{(\mathrm{val}(u_1), \dots, \mathrm{val}(u_m)) \in \mathbb{R}^m : (u_1, \dots, u_m) \in V_K(J)\big\}.$$

*This equals the set*

$$\{w \in \mathbb{R}^m : \mathrm{in}_w(I) \neq \langle 1 \rangle\}.$$

*There is a polyhedral fan $\Sigma$ whose support is* $\mathrm{Trop}(Y)$.

Versions of this result appear in [Speyer and Sturmfels 2004, Theorem 2.1] and [Einsiedler et al. 2006; Draisma 2008; Jensen et al. 2008; Payne 2009]. We consider here only the "constant coefficient" case, where the coefficients of polynomials generating defining ideal $I$ live in $\Bbbk$, so have valuation zero. This guarantees there is the structure of a fan on $\mathrm{Trop}(Y)$, rather than a polyhedral complex. We note that we follow the conventions for tropical varieties as tropicalizations of usual varieties as in, for example, [Speyer and Sturmfels 2004; Gathmann 2006] rather than the more intrinsic definition used by [Mikhalkin 2006]. For readers familiar with these works we emphasize that we follow the min convention for the tropical semiring rather than the max convention of [Gathmann 2006].

The key result is the following fundamental lemma and its immediate corollary.

**Lemma 2.6** [Tevelev 2007, Lemma 2.2]. *Let $Y$ be a subvariety of the torus $T^m$, and let $\mathcal{F}$ be an $l$-dimensional cone in $\mathbb{Q}^m$ whose rays are spanned by part of a basis for $\mathbb{Z}^m \subseteq \mathbb{Q}^m$. Let $U_{\mathcal{F}} = \mathbb{A}^l \times (\Bbbk^\times)^{m-l}$ be the corresponding affine toric variety. Then the closure $\overline{Y}$ of $Y$ in $U_{\mathcal{F}}$ intersects the closed orbit of $U_{\mathcal{F}}$ if and only if the interior of the cone $\mathcal{F}$ intersects the tropical variety of $Y$.*

**Corollary 2.7.** *Let $Y$ be a subvariety of the torus $T^m$. Let $X_\Sigma$ be an $m$-dimensional toric variety with dense torus $T^m$ and fan $\Sigma \subseteq \mathbb{R}^m$. Let $\operatorname{Trop}(Y) \subseteq \mathbb{R}^m$ be the tropical variety of $Y \subseteq T^m$. Then the closure $\overline{Y}$ of $Y$ in $X_\Sigma$ intersects the $T^m$-orbit of $X_\Sigma$ corresponding to a cone $\sigma \subseteq \Sigma$ if and only if $\operatorname{Trop}(Y)$ intersects the interior of $\sigma$.*

*Thus a toric subvariety $X_\Delta$ of $X_\Sigma$ with $\Delta$ a subfan of $\Sigma$ is sufficiently large with respect to $\overline{Y}$ exactly when $\Delta$ contains every cone of $\Sigma$ whose relative interior intersects the* $\operatorname{Trop}(Y)$.

*Proof.* Since $\Sigma$ is not assumed to be smooth, or even simplicial, we first resolve singularities. Let $\pi : X_{\Sigma'} \to X_\Sigma$ be a toric resolution of singularities, so $\Sigma'$ is a refinement of the fan $\Sigma$, and let $\overline{Y}'$ be the strict transform of $\overline{Y}$ in $X_{\Sigma'}$, which is the closure of $Y$ in $X_{\Sigma'}$. It suffices to prove the corollary for $\overline{Y}' \subseteq X_{\Sigma'}$, as $\overline{Y}$ intersects the orbit corresponding to a cone $\sigma \in \Sigma$ if and only if $\overline{Y}'$ intersects the orbit corresponding to a cone $\sigma' \in \Sigma'$ with $\sigma' \subseteq \sigma$. Since the orbit corresponding to a cone $\sigma' \in \Sigma'$ is the closed orbit of the corresponding $U_{\sigma'}$, this result now follows from Lemma 2.6. $\square$

**Example 2.8.** Let $\Sigma_1$ and $\Sigma_2$ be the two complete fans with rays as pictured in Figure 1. If $Y \subset T^2$ is a subvariety with tropicalization the dotted line shown in both figures, then the shaded fans define sufficiently large toric subvarieties with respect to the respective closures of $Y$ in $X_{\Sigma_1}$ and $X_{\Sigma_2}$.

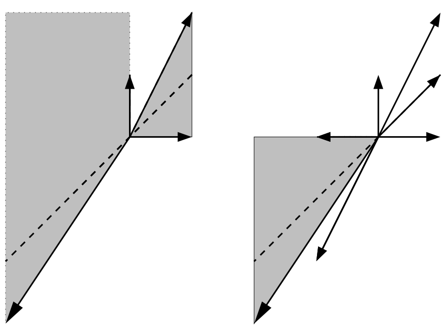


**Figure 1.** Sufficient toric subvarieties of $X_\Sigma$ with respect to $\overline{Y}$.

## 3. Equations for Chow and Hilbert quotients

Given equations for a $T^d$-equivariant subscheme $X \subset \mathbb{P}^m$, we show in this section how to effectively compute generators for the ideal of the Chow or Hilbert quotient in the Cox ring of a suitably chosen toric variety.

**3A.** ***Definition of Chow and Hilbert quotients.*** We first recall the definition of the Chow and Hilbert quotients $X/\!\!/^{\mathrm{Ch}} T^d$ and $X/\!\!/^{\mathrm{H}} T^d$ of a projective variety by the action of a $d$-dimensional algebraic torus $T^d$. We assume that $X$ is equivariantly embedded into a projective space $\mathbb{P}^m$ with no irreducible component contained in a coordinate subspace.

Identifying $T^d$ with $(\Bbbk^\times)^{d+1}/(\Bbbk^\times)$ so that points in $T^d$ are equivalence classes $t = [t_0 : \cdots : t_d]$, we can write

$$T^d \times \mathbb{P}^m \to \mathbb{P}^m, \quad (t, [x_0, \ldots, x_m]) \mapsto \left[\left(\prod_{j=0}^{d} t_j^{a_{j0}}\right) x_0, \ldots, \left(\prod_{j=0}^{d} t_j^{a_{jm}}\right) x_m\right], \tag{1}$$

and let $A$ be the $(d+1) \times (m+1)$ matrix with $ij$-th entry $a_{ij}$ for $0 \le i \le d$ and $0 \le j \le m$. We assume that the $T^d$-action on $\mathbb{P}^m$ is faithful, so $A$ has rank $d+1$. The compatibility of the action with the diagonal $\Bbbk^\times$ action on $(\Bbbk^\times)^{d+1}$ and $(\Bbbk^\times)^{m+1}$ means that all column sums of $A$ agree, so the row space of $A$ contains the all-ones vector. Let $X_A \subset \mathbb{P}^m$ be the closure of the $T^d$-orbit of $e = [1 : \cdots : 1] \in T^m$. Then $X_A$ is a toric variety with associated torus $T^d$ and corresponding toric ideal $I_A = \langle x^u - x^v : Au = Av \rangle \subset \Bbbk[x_0, \ldots, x_m]$ (see [Gel'fand et al. 1994, Chapter 5] and [Sturmfels 1996, Chapter 4]).

We identify $T^m$ with the quotient $(\Bbbk^\times)^{m+1}/(\Bbbk^\times)$, so a point on $T^m$ is an equivalence class $s = [s_0 : \cdots : s_m]$. Then $T^m$ acts on $\mathbb{P}^m$ by $s \cdot x = [s_0 x_0 : \cdots : s_m x_m]$ and every point in the geometric quotient $T^m/T^d$ corresponds to an orbit of $T^d$ whose closure in $\mathbb{P}^m$ is a $d$-dimensional subscheme of $\mathbb{P}^m$ having ideal $I_{sA} = \langle s^v x^u - s^u x^v : Au = Av \rangle \subset \Bbbk[x_0, \ldots, x_m]$, with $s \in T^m$. These orbit closures all have the same Hilbert polynomial, and so define closed points on the same connected component of $\mathrm{Hilb}(\mathbb{P}^m)$, and there is an induced morphism $\phi_H : T^m/T^d \to \mathrm{Hilb}(\mathbb{P}^m)$. The Hilbert quotient $X/\!\!/^{\mathrm{H}} T^d$ is defined to be the closure in $\mathrm{Hilb}(\mathbb{P}^m)$ of $\phi_H((X \cap T^m)/T^d)$. Since $\mathbb{P}^m/\!\!/^{\mathrm{H}} T^d$ is the closure of $\phi_H(T^m/T^d)$ in $\mathrm{Hilb}(\mathbb{P}^m)$, $X/\!\!/^{\mathrm{H}} T^d \subset \mathbb{P}^m/\!\!/^{\mathrm{H}} T^d$.

Analogously, there is a morphism $\phi_{\mathrm{Ch}} : T^m/T^d \to \mathrm{Chow}(\mathbb{P}^m)$ (see [Kollár 1996, Sections 1.3 and 1.4]). The Chow quotient $X/\!\!/^{\mathrm{Ch}} T^d$ is defined to be the closure in $\mathrm{Chow}(\mathbb{P}^m)$ of $\phi_{\mathrm{Ch}}((X \cap T^m)/T^d)$ and $X/\!\!/^{\mathrm{Ch}} T^d \subset \mathbb{P}^m/\!\!/^{\mathrm{Ch}} T^d$.

We remark that while the definitions given here appear to depend on the choice of projective embedding, the Chow and Hilbert quotients of $X$ are in fact independent of this choice. See [Kapranov 1993] for a more intrinsic formulation. See also [Białynicki-Birula and Sommese 1987] for original work on the Hilbert quotient.

For the most part the Chow and Hilbert quotients can be treated uniformly, and we use the notation $X /\!\!/^{\star} T^d$ to denote either of $X /\!\!/^{\mathrm{Ch}} T^d$ or $X /\!\!/^{\mathrm{H}} T^d$.

The following example will be developed more in Section 5.

**Example 3.1.** Consider the action of the $(n-1)$-dimensional torus

$$T^{n-1} \cong (\Bbbk^{\times})^n / \Bbbk^{\times}$$

on $\mathbb{P}_{\mathbb{C}}^{\binom{n}{2}-1}$ given by

$$[t_1 : \cdots : t_n] \cdot [\{x_{ij}\}_{1 \leq i < j \leq n}] = [\{t_i t_j x_{ij}\}_{1 \leq i < j \leq n}]. \tag{2}$$

Recall that the Plücker embedding of the Grassmannian $G(2,n)$ into $\mathbb{P}^{\binom{n}{2}-1}$ is given by taking a subspace $V$ to $V \wedge V$, or by taking a $2 \times n$ matrix representing a choice of basis for $V$ to its vector of $2 \times 2$ minors. The ideal $I_{2,n}$ of $G(2,n)$ in the homogeneous coordinate ring of $\mathbb{P}^{\binom{n}{2}-1}$ is generated by the set of Plücker equations $\langle p_{ijkl} = x_{ij}x_{kl} - x_{ik}x_{jl} + x_{il}x_{jk} : 1 \leq i < j < k < l \leq n \rangle$ and hence is $T^{n-1}$-equivariant. Let $G^0(2,n)$ be the open set inside $G(2,n)$ consisting of those points with nonvanishing Plücker coordinates, corresponding to those two-planes that do not have a nonzero intersection with any two coordinate hyperplanes. The torus $T^{n-1}$ acts freely on $G^0(2,n)$ and all orbits have maximal dimension. The moduli space $M_{0,n}$ parametrizing smooth $n$-pointed rational curves is equal to the geometric quotient $G^0(2,n)/T^{n-1}$. By [Kapranov 1993, Theorem 4.1.8]

$$G(2,n) /\!\!/^{\mathrm{Ch}} T^{n-1} = G(2,n) /\!\!/^{\mathrm{H}} T^{n-1} = \overline{M}_{0,n}.$$

**3B.** ***Chow and Hilbert quotients of projective spaces.*** The Chow and the Hilbert quotients of $\mathbb{P}^m$ by $T^d$ are both not necessarily normal toric varieties [Kapranov et al. 1992]. We next describe the fans $\Sigma^{\mathrm{Ch}}$ and $\Sigma^{\mathrm{H}}$ associated to the normalizations of $\mathbb{P}^m /\!\!/^{\mathrm{Ch}} T^d$ and $\mathbb{P}^m /\!\!/^{\mathrm{H}} T^d$. The fan $\Sigma^{\mathrm{Ch}}$ is the *secondary fan* of the matrix $A$ given in (1). Top-dimensional cones of the secondary fan correspond to regular triangulations of the vector configuration determined by the columns of $A$. These are also indexed by radicals of initial ideals of $I_A$ by [Sturmfels 1996, Theorem 8.3]. See [Gel'fand et al. 1994] for a description of the secondary fan. The fan $\Sigma^{\mathrm{H}}$ is the *saturated Gröbner fan*, whose cones are indexed by the saturation of initial ideals of $I_A$ with respect to the irrelevant ideal $\langle x_0, \ldots, x_m \rangle$. See [Bayer and Morrison 1988; Mora and Robbiano 1988] for the original work on the Gröbner fan, and [Sturmfels 1996; Maclagan and Thomas 2007] for expositions. When we want to refer to both fans simultaneously we use the notation $\Sigma^{\star}$.

We denote by $\overline{N}$ the common lattice of the fans $\Sigma^{\mathrm{Ch}}$ and $\Sigma^{\mathrm{H}}$. Note that $\overline{N} = N/N'$, where $N \cong \mathbb{Z}^{m+1}$ and $N'$ is the integer row space of the matrix $A$. The torus $T^{d+1}$ with $T^d = T^{d+1}/\Bbbk^{\times}$ is equal to $N' \otimes \Bbbk^{\times}$, and $T^{m-d} = T^{m+1}/T^{d+1} \cong \overline{N} \otimes \Bbbk^{\times}$.

The images of the basis elements $\boldsymbol{e}_i \in \mathbb{Z}^{m+1} \cong N$ correspond to rays of $\Sigma^{\star}$. We can thus identify these rays with the columns of the Gale dual $D$ of the matrix

$A$ (see Section 2A). By the construction of $D$ the integer column space of $D$ is $\mathbb{Z}^{m-d} \cong \overline{N}$.

**3C.** ***Equations for Chow and Hilbert quotients.*** In this section we show how to give equations for the Chow or Hilbert quotient of $X$ in the Cox ring of a toric variety $X_\Sigma$. Throughout this section $X$ is a $T^d$-equivariant subscheme of $\mathbb{P}^m$ with no irreducible component lying in any coordinate subspace, and we choose $X_\Sigma$ to be a sufficiently large toric subvariety of the normalization $X_{\Sigma^\star}$ of $\mathbb{P}^m /\!\!/^\star T^d$ with respect to the pullback of $X /\!\!/^\star T^d$ to $X_{\Sigma^\star}$. We recall that the notation $X /\!\!/^\star T^d$ stands for either the Chow or Hilbert quotient, and any statement using this notation is short-hand for two separate results; one for the Chow quotient, and one for the Hilbert quotient. We denote by $I(X) \subseteq \Bbbk[x_0, \ldots, x_m]$ the saturated ideal defining $X$. Let the $(d+1) \times (m+1)$ matrix $A$ record the weights of the $T^d$ action, and let $D$ be its Gale dual (see (1)). Let $R$ be the matrix whose columns are the first integer lattice points on the rays of $\Sigma$. Since the columns of $D$ span the lattice $\overline{N}$ of $\Sigma^\star$, one has that

$$R = DV,$$

for some $(m+1) \times r$ matrix $V$. We denote by $X /\!\!/_n^\star T^d$ the pullback of $X /\!\!/^\star T^d$ to the normalization $X_{\Sigma^\star}$ of $\mathbb{P}^m /\!\!/^\star T^d$.

**Theorem 3.2.** *Let $T^d$ act on $\mathbb{P}^m$ and let $X \subset \mathbb{P}^m$ be a $T^d$-equivariant subscheme with $I(X) = \langle f_1, \ldots, f_g \rangle \subseteq \Bbbk[x_0, \ldots, x_m]$. Let $X_\Sigma$ be a toric subvariety of $X_{\Sigma^\star}$ containing $X /\!\!/_n^\star T^d$, and let $S = \Bbbk[y_1, \ldots y_r]$ be the Cox ring of $X_\Sigma$. Let*

$$\nu : \Bbbk[x_0, \ldots, x_m] \to \Bbbk[y_1^{\pm 1}, \ldots, y_r^{\pm 1}] \quad \textit{be given by} \quad \nu(x_i) = \prod_{j=1}^{r} y_j^{V_{ij}}.$$

*The ideal of $X /\!\!/_n^\star T^d$ in $S$ is $I = \nu(I(X)) \cap S$, and is obtained by clearing denominators in $\{\nu(f_i) : 1 \le i \le g\}$, and then saturating the result by the product of all the variables in $S$. If $X /\!\!/^\star T^d$ is irreducible and normal, it is isomorphic to $X /\!\!/_n^\star T^d$ and $I$ is the ideal of $X /\!\!/^\star T^d$ in $S$.*

**Remark 3.3.** When $X_\Sigma$ contains all rays of $\Sigma^\star$ corresponding to columns of $D$, and $D$ has no repeated columns, we can write $V = (I \mid C^T)$, where $I$ is the $(m+1) \times (m+1)$ identity matrix and $C$ is an integer $(r-m-1) \times (m+1)$ matrix. Thus the map $\nu$ may be thought of as homogenizing the ideal $I(X)$ with respect to the grading of $S$. This is the formulation assumed in Theorem 1.1(1).

*Proof of Theorem 3.2.* The proof of the first claim proceeds in three parts. First, using Proposition 2.1, we find equations in $\Bbbk[z_1^{\pm 1}, \ldots, z_{m-d}^{\pm 1}]$, the coordinate ring of the torus $T^m/T^d$ of $X_\Sigma$, for $(X \cap T^m)/T^d$. Second, we use the isomorphism $\rho : \Bbbk[z_1^{\pm 1}, \ldots, z_{m-d}^{\pm 1}] \to S_{(\prod_{i=1}^r y_i)}$, and apply Proposition 2.3 to obtain generators for the ideal in $S$ of the closure of $(X \cap T^m)/T^d$ in $X_{\Sigma^\star}$. Since this is $X /\!\!/_n^\star T^d$, we last check that this is the same as the ideal obtained by clearing denominators and saturating.

The key diagram is the following, where $U$ is as in the proof of Proposition 2.1 and the $i$ are inclusions:

$$\begin{tikzcd}
\Bbbk[z_1^{\pm 1}, \dots, z_{m+1}^{\pm 1}] \arrow[r, "\sim", "U^T"'] & \Bbbk[x_0^{\pm 1}, \dots, x_m^{\pm 1}] \cong \Bbbk[M] \\
\Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}] \arrow[u, "i"] \arrow[r, "\sim", "D^T"'] \arrow[d, "\sim"', "R^T"] & \Bbbk[x_0^{\pm 1}, \dots, x_m^{\pm 1}]_0 \arrow[u, "i"'] \arrow[dl, "\sim", "V^T"'] \\
(S_y)_0 &
\end{tikzcd}$$

The content of Proposition 2.1 is that the ideal in $\Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}]$ of $(X \cap T^m)/T^d$ is given by $i^{-1}(U^{T^{-1}}(I(X)))$, which is well-defined since $U^{T^{-1}}(I(X))$ is in the image of $i$. The ideal $i^{-1}(U^{T^{-1}}(I(X)))$ is taken under the map $D^T$ to the degree zero part of the ideal $I(X)$ in $\Bbbk[x_0^{\pm 1}, \dots, x_m^{\pm 1}]$, since $U$ is chosen in Proposition 2.1 to have its first $d+1$ rows equal to $D$. We thus see that the choice of identification of the lattice $\bar{N}$ of the fan $\Sigma$ with $\mathbb{Z}^{m-d}$ identifies $\Bbbk[\bar{M}]$ with $\Bbbk[z_1^{\pm 1}, \dots, z_{m-d}^{\pm 1}]$, where $\bar{M}$ is the lattice dual to $\bar{N}$. This means that the isomorphism $\Bbbk[\bar{M}] \cong (S_y)_0$ given in [Cox 1995, Lemma 2.2] is given by the matrix $R^T$, so the function $\rho$ of Proposition 2.3 is given by $\rho(z_i) = \prod_{j=1}^{r} y_j^{R_{ij}}$. It thus follows from Proposition 2.3 that the ideal in $S$ of the closure of $(X \cap T^m)/T^d$ in $X_\Sigma$ is given by applying $R^T \circ i^{-1} \circ U^{T^{-1}}$ to the generators of $I(X)$, clearing denominators, and then saturating by the product of the variables of $S$.

To complete the proof, it thus suffices to observe that $R^T \circ i^{-1} \circ U^{T^{-1}}$ restricted to the degree zero part of $\Bbbk[x_0^{\pm 1}, \dots, x_m^{\pm 1}]$ is given by the matrix $V^T$. This follows from the fact that $R^T = V^T D^T$ and the bottom triangle of the above commutative diagram is made of three isomorphisms.

We now consider the case where $X /\!\!/^{\star} T^d$ is irreducible and normal, and show that it isomorphic to $X /\!\!/_n^{\star} T^d$. Set $Y = X /\!\!/^{\star} T^d$, and $Z = \mathbb{P}^m /\!\!/^{\star} T^d$. Let $\tilde{Y}$ and $\tilde{Z} = X_{\Sigma^\star}$ denote the respective normalizations. Begin by reducing to the case where $Z = \operatorname{Spec}(A)$ and $Y = \operatorname{Spec}(A/I)$ are irreducible affine schemes with $I$ an ideal in a ring $A$. Note that $Y$ intersects the torus $T^{m-d}$ of the not necessarily normal toric variety $Z$, and thus intersects the smooth locus of $Z$. Since $Y$ is irreducible, this means that the map $Y \times_Z \tilde{Z} \to Y$ is birational, and so $A$ and $A/I$ have the same total quotient ring. Normalization is a finite morphism, and the pullback of a finite morphism is finite, so $\tilde{A}/I\tilde{A}$ is integral over $A/I$. The isomorphism $\tilde{A}/I\tilde{A} \cong A/I$ then follows from the fact that $Y$ is normal, so $A/I$ is integrally closed. Finally since all maps are inclusions, everything glues to prove $Y \times_Z \tilde{Z} \cong Y$. $\square$

**Remark 3.4.** The choice of the matrix $V$ in Theorem 3.2 is not unique. This does not affect the computation, however, as the induced map $\nu$ is unique when restricted to the degree zero part of $\Bbbk[x_0^{\pm 1}, \dots, x_m^{\pm 1}]$.

**Example 3.5.** Theorem 3.2 lets us compute equations for $\overline{M}_{0,5}$. Kapranov's description of $\overline{M}_{0,5}$ as the Chow or Hilbert quotient of the Grassmannian $G(2,5)$ by the $T^4$ action, described in Example 3.1, gives embeddings of $\overline{M}_{0,5}$ into the normalizations of $\mathbb{P}^9 /\!\!/^{\mathrm{Ch}} T^4$ and $\mathbb{P}^9 /\!\!/^{\mathrm{H}} T^4$. These are both five-dimensional toric varieties whose rays include the columns of the matrix $D$ of Example 2.2, plus ten additional rays, being the columns of the matrix $-D$. Let $\Delta$ be the two-dimensional fan with rays the columns $\boldsymbol{d}_{ij}$ of $D$, and cones $\operatorname{pos}(\boldsymbol{d}_{ij}, \boldsymbol{d}_{kl} : \{i, j\} \cap \{k, l\} = \varnothing)$. By direct computation, or by Theorem 5.7, the fan $\Delta$ is a subfan of the fan of both $\Sigma^{\mathrm{Ch}}$ and $\Sigma^{\mathrm{H}}$ that defines a sufficiently large toric subvariety. For this fan we have $R = D$, so $V$ is the $\binom{10}{2} \times \binom{10}{2}$ identity matrix.

The ideal in $\Bbbk[x_{ij} : 1 \leq i < j \leq 5]$ defining $G(2,5)$ as a subscheme of $\mathbb{P}^9$ is generated by the Plücker equations given in Example 2.2. The Cox ring of $\mathbb{P}_\Delta$ is $\Bbbk[y_{ij} : 1 \leq i < j \leq 5]$, and the map $\nu$ of Theorem 3.2 is the identity map $\nu(x_{ij}) = y_{ij}$. The ideal of the normal irreducible variety $\overline{M}_{0,5}$ in the Cox ring of $\mathbb{P}_\Delta$ is thus generated by the Plücker relations. The saturation step is unnecessary in this case as this ideal is prime.

## 4. GIT constructions of Chow/Hilbert quotients

In this section we give a GIT construction of the Chow and Hilbert quotients of $X$. This follows from their description in the Cox ring of a toric variety in Section 3. We also recover all GIT quotients of $X$ by $T^d$ by variation of the GIT quotient. As before $X$ is a $T^d$-equivariant subscheme of $\mathbb{P}^m$ with no irreducible component lying in any coordinate hyperplane.

Let $X_\Sigma$ be a sufficiently large toric subvariety of $X_{\Sigma^\star}$. We assume in addition that $X_\Sigma$ has a torsion-free divisor class group, which can be guaranteed, for example, by taking $X_\Sigma$ to contain the rays of $\Sigma^\star$ corresponding to columns of the matrix $D$ (see Section 2A). Let $I$ be the ideal of $X /\!\!/_n^\star T^d$ in the Cox ring of $X_\Sigma$. We next define a cone which will index our choices of GIT quotient.

**Definition 4.1.** Let $\{[D_i] : 1 \leq i \leq |\Sigma(1)|\}$ be the set of classes of the torus invariant divisors on $X_\Sigma$. Set

$$\mathscr{G}(X_\Sigma) = \bigcap_{\sigma \in \Sigma} \operatorname{pos}([D_i] : i \notin \sigma).$$

We note that $\mathscr{G}(X_\Sigma)$ is the cone in $\mathrm{Cl}(X_\Sigma) \otimes \mathbb{R}$ spanned by divisor classes $[D] \in \mathrm{Cl}(X_\Sigma)$ for which $[D]$ is globally generated.

**Lemma 4.2.** *The cone $\mathscr{G}(X_\Sigma)$ has positive dimension.*

*Proof.* Since $\Sigma$ is a subfan of the fan $\Sigma^\star$ we can number the rays of $\Sigma^\star$ so that the first $p$ live in $\Sigma$, while the last $s$ do not. If $i : X_\Sigma \to X_{\Sigma^\star}$ is the inclusion map, then the pullback from $\mathrm{Cl}(X_{\Sigma^\star})$ to $\mathrm{Cl}(X_\Sigma)$ is given by $i^\star\big(\big[\sum_{i=1}^{p+s} a_i D_i\big]\big) = \big[\sum_{i=1}^{p} a_i D_i\big]$. Since $X_{\Sigma^\star}$ is a projective toric variety, its nef cone $\mathcal{G}(X_{\Sigma^\star}) = \bigcap_{\sigma \in \Sigma^\star} \mathrm{pos}([D_i] : i \notin \sigma)$ is a positive-dimensional cone that can be viewed as living in $\mathrm{Cl}(X_{\Sigma^\star}) \otimes \mathbb{R}$. The image of $\mathcal{G}(X_{\Sigma^\star})$ under the map $i^*$ is contained in $\mathcal{G}(X_\Sigma)$. We now show that some nonzero vector in $\mathcal{G}(X_{\Sigma^\star})$ is taken to a nonzero vector under $i^*$. Indeed, otherwise every element of $\mathcal{G}(X_{\Sigma^\star})$ could be written in the form $\sum_{i=p+1}^{p+s} a_i[D_i]$, so $\mathcal{G}(X_{\Sigma^\star}) \subseteq \mathrm{pos}([D_i] : p+1 \leq i \leq s)$. But by the relation between the cones of $\mathcal{G}(X_{\Sigma^\star})$ and chambers of the chamber complex [Billera et al. 1990], this means that the cone spanned by all the rays of $\Sigma$ lies in $\Sigma^\star$. However $X_\Sigma$ was assumed to be sufficiently large, which means that $\Sigma$ contains all cones of $\Sigma^\star$ intersecting the tropical variety of $X$, so the balancing condition on tropical varieties [Speyer 2005, Theorem 2.5.1] implies that the rays of $\Sigma$ positively span the entire space $\mathbb{R}^{m-d}$. This would mean that the cone spanned by the rays of $\Sigma$ was all of $\mathbb{R}^{m-d}$, contradicting it lying in the fan $\Sigma^\star$. We thus conclude that there are elements $v \in \mathcal{G}(X_{\Sigma^\star})$ with $i^*(v) \neq 0$, so $\mathcal{G}(X_\Sigma)$ is a positive-dimensional cone. $\square$

Let $l = |\Sigma(1)| - \dim(X_\Sigma)$ be the rank of $\mathrm{Cl}(X_\Sigma)$. Let $H$ be the algebraic torus $\mathrm{Hom}(\mathrm{Cl}(X_\Sigma), \Bbbk^\times) \cong (\Bbbk^\times)^l$. We can regard $\mathrm{Cl}(X_\Sigma) \otimes \mathbb{R}$ as the space of (real) characters of the torus $H$, and $\mathcal{G}(X_\Sigma)$ as a subcone of the character space. The torus $H$ acts on $\mathbb{A}^r$ by $h \cdot x_i = h([D_i])x_i$. Recall that the torus of $X_\Sigma$ is $T^m/T^d$. We denote by $\mathrm{relint}(\mathcal{G}(X_\Sigma))$ the relative interior of the cone $\mathcal{G}(X_\Sigma)$.

Let $r = |\Sigma(1)|$. Recall that for $\alpha \in \mathbb{Z}^l$ the GIT quotient $Y /\!\!/_\alpha H$ of an affine variety $Y \subset \mathbb{A}^r$ is

$$Y /\!\!/_\alpha H = \mathrm{Proj}\big(\bigoplus_{j \geq 0} (\Bbbk[x_1, \dots, x_r]/I(Y))_{j\alpha}\big),$$

where the $\mathbb{Z}^l$ grading on the polynomial ring comes from the $H$-action on $\mathbb{A}^r$. For $\alpha \in \mathbb{Q}^l$ we define $Y /\!\!/_\alpha H$ to be $Y /\!\!/_{s\alpha} H$ for any integral multiple $s\alpha$.

**Proposition 4.3.** *Let $Y \subseteq \mathbb{A}^r$ be the subscheme defined by the ideal $I \subseteq \mathrm{Cox}(X_\Sigma)$ of $X /\!\!/_n^\star T^d$. For rational $\alpha \in \mathrm{relint}(\mathcal{G}(X_\Sigma))$ we have*

$$X /\!\!/_n^\star T^d = Y /\!\!/_\alpha H.$$

*Proof.* It follows from the results of [Cox 1995] and the chamber complex description of the secondary fan that $X_{\Sigma'} = \mathbb{A}^r /\!\!/_\alpha H$ is a projective toric variety whose fan $\Sigma'$ has same rays as $\Sigma$ and contains $\Sigma$ as a subfan. The dense torus of $X_{\Sigma'}$ is also $T^m/T^d$. The quotient $Y /\!\!/_\alpha H$ is a subvariety of $X_{\Sigma'}$. Let $(Y /\!\!/_\alpha H)^0 = (Y /\!\!/_\alpha H) \cap T^m/T^d = (Y \cap T^r)/H = (X \cap T^m)/T^d$, where the last equality follows from the fact that $T^r/H \cong T^m/T^d$. Then $Y /\!\!/_\alpha H$ is the closure of $(Y /\!\!/_\alpha H)^0$ in $X_{\Sigma'}$. By Corollary 2.7, to show that $Y /\!\!/_\alpha H$ is the closure of $(Y /\!\!/_\alpha H)^0$ inside $X_\Sigma$ it suffices to show that the tropical variety of $(X \cap T^m)/T^d \subseteq T^m/T^d$ is contained in the support of $\Sigma$. This follows from the hypothesis that $X_\Sigma$ is sufficiently large, again by Corollary 2.7. Since $X /\!\!/_n^\star T^d$ is the closure of $(X \cap T^m)/T^d$ in $X_\Sigma$, the proposition follows. $\square$

This GIT description gives projective embeddings of $X /\!\!/_n^\star T^d$, as we now describe. Let $S = \Bbbk[x_1, \dots, x_r]$ be the Cox ring of $X_\Sigma$. For $\alpha \in \mathcal{G}(X_\Sigma)$, write $S^\alpha$ for the subring $\bigoplus_{j \geq 0} S_{j\alpha}$ of $S$. Note that $S^\alpha$ has a standard $\mathbb{Z}$-grading, by setting $\deg f = j$ for all $f \in S_{j\alpha}$. If $J$ is an ideal in $S$, write $J^\alpha$ for the ideal $J \cap S^\alpha$ of $S^\alpha$. The GIT description above gives that

$$X /\!\!/_n^\star T^d = \operatorname{Proj}(S^\alpha / I^\alpha).$$

**Corollary 4.4.** *The GIT description of $X /\!\!/_n^\star T^d$ gives a projective embedding of $X /\!\!/_n^\star T^d$ with the pullback of $\mathcal{O}(1)$ on $\mathbb{P}^N$ equal to $\pi^*(\alpha)$ for a large enough multidegree $\alpha \in \mathcal{G}(X_\Sigma)$, where $\pi$ is the embedding of $X /\!\!/_n^\star T^d$ into $X_\Sigma$.*

*Proof.* Assume $\alpha$ is large enough so that $S^\alpha / I^\alpha$ is generated in degree one. If not, we can replace $\alpha$ by $l\alpha$ for $l \gg 0$. Let $x^{u_0}, \dots, x^{u_N} \in S_\alpha$ generate $S^\alpha / I^\alpha$, and define a map $\phi : \Bbbk[z_0, \dots, z_N] \to S^\alpha / I^\alpha$ by $\phi(z_i) = x^{u_i}$. Let $J = \ker \phi$. Then $X /\!\!/_n^\star T^d = \operatorname{Proj}(\Bbbk[z_0, \dots, z_N]/J)$, and by construction the pullback of $\mathcal{O}(1)$ is $\pi^*(\alpha)$. $\square$

**Example 4.5.** We continue the discussion of $\overline{M}_{0,5}$ begun in Examples 3.1 and 3.5. Example 3.5 shows that $\overline{M}_{0,5}$ is the GIT quotient of the affine cone over the Grassmannian $G(2,5)$ by a five-dimensional torus, so $\overline{M}_{0,5} = G(2,5) /\!\!/_\alpha T^4$ for all $\alpha$ in the relative interior of $\mathcal{G}(X_\Delta)$. The cone $\mathcal{G}(X_\Delta)$ is the intersection of the cones $\operatorname{pos}(\boldsymbol{d}_{pq} : \{p,q\} \neq \{i,j\}, \{k,l\})$ for all choices of $i, j, k, l$ distinct, where $\boldsymbol{d}_{pq}$ is the column of the matrix $D$ of Example 2.2 indexed by $\{p,q\}$. This cone thus contains the vector $\alpha = (2,2,2,2,2)$ in its relative interior. The Cox ring of $X_\Delta$, $S = \Bbbk[x_{ij} : 1 \leq i < j \leq 5]$, has an action of the symmetric group $S_5$ permuting the variables, and $S^\alpha$ is generated in degree one by monomials $x_{12}^2 x_{34} x_{35} x_{45}$ and $x_{12} x_{23} x_{34} x_{45} x_{15}$ and their $S_5$-orbits, which are of size 10 and 12 respectively. Thus this gives an embedding of $\overline{M}_{0,5}$ into $\mathbb{P}^{21}$. The ideal $I^\alpha$ is then cut out by linear equations induced from multiples of the Plücker relations such as $x_{15}x_{12}x_{34}x_{23}x_{45} - x_{15}x_{12}x_{34}x_{24}x_{35} + x_{15}x_{12}x_{34}^2 x_{25}$ and its $S_5$-orbit. Thus $\overline{M}_{0,5}$ is cut out as a subscheme of $\mathbb{P}^{21}$ by these linear relations and the binomial relations coming from the kernel of the surjective map $\Bbbk[z_0, \dots, z_{21}] \to S^\alpha$.

We now show that other GIT quotients of $X$ can be obtained from $X /\!\!/_n^\star T^d$ by variation of the GIT quotient (VGIT). We assume that $\Sigma$ contains all the rays of $\Sigma^\star$ corresponding to rays of the Gale dual matrix $D$ (see Section 2A), and also that $D$ has no repeated columns. This means that we can write $R = (D \mid DC^T) = D(I \mid C^T)$ for some $s \times (m+1)$ matrix $C$. Let $S = \operatorname{Cox}(X_\Sigma) = \Bbbk[x_0, \dots, x_m, y_1, \dots, y_s]$, where $s = r - (m+1) = l - (d+1)$ and $\tilde{S} = \Bbbk[x_0, \dots, x_m]$. Grade $\tilde{S}$ by $\deg x_i = \boldsymbol{a}_i$, where $\boldsymbol{a}_i$ is the $i$-th column of the matrix $A$. Our assumption on $\Sigma$ means that the grading matrix for $S$ can be written in block form as

$$\left(\begin{array}{c|c} A & 0 \\ \hline -C & I_s \end{array}\right). \tag{3}$$

We write $\mathbb{N}A$ for the subsemigroup of $\mathbb{Z}^{d+1}$ generated by the columns of the matrix $A$. Let $\pi : \mathbb{Z}^l \to \mathbb{Z}^{d+1}$ be the projection onto the first $d+1$ coordinates, and let $\pi_2 : \mathbb{Z}^l \to \mathbb{Z}^s$ be the projection onto the last $s = l - (d+1)$ coordinates. Recall the homomorphism $\nu : \Bbbk[x_0, \dots, x_m] \to \Bbbk[x_0, \dots, x_m, y_1^{\pm 1}, \dots, y_s^{\pm 1}]$ given in the statement of Theorem 3.2. Since $D$ does not have repeated columns this is given by $\nu(x^u) = x^u y^{Cu}$. Again $Y = Z(I)$ is the subscheme of $\mathbb{A}^{m+1+s}$ defined by the ideal $I$ of $X/\!\!/_n^{\star} T^d$.

**Theorem 4.6.** *With the notation given above, fix $\beta \in \mathbb{N}A$. Then if $\alpha \in \mathbb{Z}^l$ satisfies $\pi(\alpha) = \beta$, and $\alpha_i \geq -\min\{(Cu)_i : Au = \beta, u \in \mathbb{Q}^{m+1}_{\geq 0}\}$ for $1 \leq i \leq s$, then*

$$Y/\!\!/_{\alpha} H \cong X/\!\!/_{\pi(\alpha)} T^d.$$

*Proof.* We will show that the map $\nu$ described above induces an isomorphism between $\tilde{S}^{\beta}/I(X)^{\beta}$ and $S^{\alpha}/I^{\alpha}$.

The map $\nu$ sends a monomial $x^u$ to $x^u y^{Cu}$. Define $\nu' : \tilde{S}^{\beta} \to S^{\alpha}$ by setting $\nu'(x^u) = x^u y^{l\pi_2(\alpha)+Cu}$, when $\deg x^u = l\beta$. By construction if $\deg x^u = l\beta$ then $\deg \nu'(x^u) = l\alpha$, and the assumption on $\alpha$ implies that $l\pi_2(\alpha) + Cu \in \mathbb{N}^{l-d-1}$, so the map is well-defined. Note also that $\nu'$ is injective and surjective, as if $x^u y^v \in S_{l\alpha}$, then $\deg x^u = l\beta$, and $v$ must equal $Cu$. We denote by $\bar{\nu}'$ the induced map from $\tilde{S}^{\beta}$ to $S^{\alpha}/I^{\alpha}$, and let $J = \ker \bar{\nu}'$. Since $\bar{\nu}'$ is surjective as $\nu'$ is, it remains to show that $J = I(X)^{\beta}$. Since $\nu'$ is a graded homomorphism, $J$ is a homogeneous ideal, so it suffices to check that each homogeneous polynomial in $J$ lies in $I(X)$ and vice versa. Recall from Theorem 3.2 that $I = (\nu(I(X))S_y) \cap S$, where $y = \prod_{i=0}^m x_i \prod_{j=1}^s y_j$. If $f = \sum_u c_u x^u \in (I(X))_{l\beta}$, then $\nu(f) = \sum_u c_u x^u y^{Cu}$, so $\nu'(f) = \sum c_u x^u y^{l\pi_2(\alpha)+Cu} = y^{l\pi_2(\alpha)}\nu(f) \in I_{l\alpha}$, and thus $I(X)^{\beta} \subseteq J$.

Suppose now that $f \in J_{l\beta}$, so $\nu'(f) \in I_{l\alpha}$, and thus $\nu'(f) = \sum_i h_i \nu(g_i)$ for $g_i \in I(X)$ and $h_i \in S_y$, where we may assume that the $h_i$ are constant multiples of Laurent monomials, and the $g_i$ are homogeneous. Write $f = \sum_u c_u x^u$, $g_i = \sum_v d_{v,i} x^v$, and $h_i = c_i x^{u_i^+}/x^{u_i^-} y^{v_i}$, where $u_i^+, u_i^- \in \mathbb{N}^{m+1}$, and $v_i \in \mathbb{Z}^s$. Note that $\nu(c_i x^{u_i^+} g_i) = c_i x^{u_i^+} y^{Cu_i^+} \nu(g_i)$, so by changing $v_i$ and $g_i$ we may assume that $c_i = 1$ and $u_i^+ = 0$ for all $i$. Also, note that the map $\nu$ lifts to a map from $\Bbbk[x_0^{\pm 1}, \dots, x_m^{\pm 1}]$ to $\Bbbk[x_0^{\pm 1}, \dots, x_m^{\pm 1}, y_1^{\pm 1}, \dots, y_s^{\pm 1}]$, and $I'(X) := I(X)\Bbbk[x_0^{\pm 1}, \dots, x_m^{\pm 1}] \cap \tilde{S} = I(X)$ (as no irreducible component of $X$ is contained in a coordinate subspace). Also $\nu(I'(X))S_y = \nu(I(X))S_y$, so we may also assume that $u_i^- = 0$ for all $i$. Thus we have $\nu'(f) = \sum_i y^{v_i}\nu(g_i)$ where $v_i \in \mathbb{Z}^s$, and $g_i \in I'(X)_{l\beta}$ with $\tilde{g} = \sum_i g_i \in \tilde{S}$

because $\tilde{g} = \nu'(f)\big|_{y_i=1}$ and thus $\tilde{g} \in (I(X))_\beta$. Since $\deg \nu'(f) = l\alpha$, and each $g_i$ is homogeneous, we have $\deg(y^{v_i}\nu(g_i)) = l\alpha$, so $\nu'(g_i) = y^{v_i}\nu(g_i)$, and thus $\nu'(\tilde{g}) = \nu'(f)$. From this it follows that $\tilde{g} = f$, since $\nu'$ is injective by construction, and thus $f \in I(X)_{l\beta}$, so $J = I(X)^\beta$ as required. □

**Corollary 4.7.** *The Chow and Hilbert quotients $X /\!\!/^{\mathrm{Ch}} T^d$ and $X /\!\!/^{\mathrm{H}} T^d$ and the GIT quotients $X /\!\!/_\alpha T^d$ are related by variation of the GIT quotient.*

*Proof.* It remains to check that an $\alpha$ satisfying the hypotheses of Theorem 4.6 actually exists. This means checking that the set $\{(Cu)_i : Au = \beta, u \in \mathbb{Q}^{m+1}_{\geq 0}\}$ is bounded below so the minimum exists. This follows from the fact that $\{u : Au = \beta, u \in \mathbb{Q}^{m+1}_{\geq 0}\}$ is a polytope, since the all-ones vector is in the row space of $A$. Thus the lower bound is the minimum of a linear functional on a polytope, which is finite. □

## 5. A toric variety containing $\overline{M}_{0,n}$

In the remainder of the paper we apply the previous theorems to obtain equations for $\overline{M}_{0,n}$. Using the construction given in Example 3.1, $\overline{M}_{0,n}$ is a subvariety of the normalization of the Chow or Hilbert quotient of $\mathbb{P}^{\binom{n}{2}-1}$ by $T^{n-1}$. In this section we describe a smooth normal toric variety $X_\Delta$ that is a sufficiently large toric subvariety of both $X_{\Sigma^{\mathrm{Ch}}}$ and $X_{\Sigma^{\mathrm{H}}}$. We prove in Theorem 5.7 that $\overline{M}_{0,n}$ is contained in $X_\Delta$.

We first define a simplicial complex and set of vectors that form the underlying combinatorics of $\Delta$. Recall that $[n] = \{1, \dots, n\}$, and for $I \subset [n]$ we have $I^c = [n] \setminus I$.

**Definition 5.1.** Let $\mathcal{I} = \{I \subset [n] : 1 \in I,\ |I| \geq 2,\ |I^c| \geq 2\}$. The simplicial complex $\tilde{\Delta}$ has vertices $\mathcal{I}$ and $\sigma \subseteq \mathcal{I}$ is a simplex of $\tilde{\Delta}$ if for all $I$ and $J \in \sigma$ we have $I \subseteq J$, $J \subseteq I$, or $I \cup J = [n]$.

Let $\mathscr{E} = \{ij : 2 \leq i < j \leq n, ij \neq 23\}$ be an indexing set for a basis of $\mathbb{R}^{\binom{n}{2}-n}$. For each $I \in \mathcal{I}$, define the vector

$$\boldsymbol{r}_I = (R_{ij,I})_{ij \in \mathscr{E}} \in \mathbb{Z}^{\binom{n}{2}-n}, \quad R_{ij,I} = \begin{cases} 1 & \text{if } |I \cap \{ij\}| = 0,\ |I \cap \{23\}| > 0, \\ -1 & \text{if } |I \cap \{ij\}| > 0,\ |I \cap \{23\}| = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Form the $\left(\binom{n}{2} - n\right) \times |\mathcal{I}|$ matrix $R$ with $I$-th column $\boldsymbol{r}_I$.

Note that the set $\mathcal{I}$ also labels the boundary divisors $\delta_I$ of $\overline{M}_{0,n}$, and a simplex $\sigma$ lies in $\tilde{\Delta}$ precisely if the corresponding boundary divisors intersect nontrivially.

**Proposition 5.2.** *The collection of cones $\{\mathrm{pos}(\boldsymbol{r}_I : I \in \sigma) : \sigma \in \tilde{\Delta}\}$ is a polyhedral fan $\Delta$ in $\mathbb{R}^{\binom{n}{2}-n}$ of dimension $n-3$. The associated toric variety $X_\Delta$ is smooth, and the support of $\Delta$ is the tropical variety of $M_{0,n} \subset T^{\binom{n}{2}-n}$.*

This fan is well known in the literature as the *space of phylogenetic trees* [Speyer and Sturmfels 2004, Section 4, Buneman 1974; Vogtmann 1990; Billera et al. 2001]. To prove Proposition 5.2 and for results in the remainder of this section, we use the following notation.

**Definition 5.3.** Fix $n \geq 6$. Let $A_n$ be the $n \times \binom{n}{2}$-dimensional matrix with $ij$-th column equal to $e_i + e_j$ and let $D$ be the $\left(\binom{n}{2} - n\right) \times \binom{n}{2}$ matrix

$$D = (D_{ij,kl})_{\substack{ij \in \mathscr{E} \\ 1 \leq k < l \leq n}}, \quad D_{ij,kl} = \begin{cases} 1 & \text{if } kl = ij, \text{ or } kl \in \{12, 13\}, |\{ij\} \cap \{kl\}| = \varnothing, \\ -1 & \text{if } kl = 23, \text{ or } kl \in \{1i, 1j\}, |\{23\} \cap \{kl\}| = \varnothing, \\ 0 & \text{otherwise.} \end{cases}$$

Let $C$ be the $\left(|\mathscr{I}| - \binom{n}{2}\right) \times \binom{n}{2}$ matrix with rows indexed by $I \in \mathscr{I}$ with $3 \leq |I| \leq n-3$, columns indexed by $\{ij : 1 \leq i < j \leq n\}$, and entries

$$C_{I,ij} = \begin{cases} 1 & \text{if } i, j \in I, \\ 0 & \text{otherwise.} \end{cases}$$

A straightforward calculation from the definition of $D$ shows that

$$R = D(I \mid C^T). \tag{4}$$

The matrix $A_n$ is the vertex-edge incidence matrix for the complete graph on $n$ vertices and $D$ is its Gale dual, so $DA_n^T = 0$. Note that the square $\left(\binom{n}{2} - n\right) \times \left(\binom{n}{2} - n\right)$ submatrix of $D$ with columns indexed by $\mathscr{E}$ is the identity matrix. In particular, $D$ has rank $\binom{n}{2} - n$.

*Proof of Proposition 5.2.* Let the lattice $L$ be the integer row space of $A_n$. The affine cone over the Grassmannian $G(2, n)$ in its Plücker embedding is a subvariety of $\mathbb{A}^{\binom{n}{2}}$ and we denote by $AG^0(2, n)$ its intersection with $T^{\binom{n}{2}}$. In [Speyer and Sturmfels 2004] it is shown that the tropical variety of $AG^0(2, n)$ is a $(2n-3)$-dimensional fan in $\mathbb{R}^{\binom{n}{2}}$ with lineality space $L \otimes \mathbb{R}$, and the simplicial complex corresponding to the fan structure on $\operatorname{Trop}(AG^0(2, n))$ is $\tilde{\Delta}$. Specifically, the cone corresponding to the cone $\sigma \in \tilde{\Delta}$ is $\operatorname{pos}(\boldsymbol{e}_I : I \in \sigma) + L$, where

$$\boldsymbol{e}_I = \sum_{i \in I, j \notin I} \boldsymbol{e}_{ij} \in \mathbb{Z}^{\binom{n}{2}}.$$

Recall from Example 3.1 that $M_{0,n} = G^0(2, n)/T^{n-1}$. Using Proposition 2.1, one can show that for any $X \subseteq T^m$ that is invariant under the action of a torus $T^d \subset T^m$, the tropical variety of $X/T^d \subset T^m/T^d$ is equal to the quotient of the tropical variety of $X \subset T^m$ by the tropical variety of $T^d$. Thus $\operatorname{Trop}(M_{0,n}) = \operatorname{Trop}(G^0(2, n))/\operatorname{Trop}(T^{n-1}) = \operatorname{Trop}(AG^0(2, n))/L$. To prove the first part of the proposition it then suffices to show that the image of $\boldsymbol{e}_I$ in $\mathbb{R}^{\binom{n}{2}}/L$ is $\boldsymbol{r}_I$. Since the matrix $D$ of Definition 5.3 is a Gale dual for $A_n$, the map $\mathbb{R}^{\binom{n}{2}} \to \mathbb{R}^{\binom{n}{2}}/L \cong \mathbb{R}^{\binom{n}{2}-n}$

given by sending the basis vector $\boldsymbol{e}_{ij}$ of $\mathbb{R}^{\binom{n}{2}}$ to $-\frac{1}{2}\boldsymbol{d}_{ij}$ is an isomorphism, where $\boldsymbol{d}_{ij}$ is the $ij$-th column of $D$. The image of $\boldsymbol{e}_I$ under this map is then $-\frac{1}{2}\sum_{i\in I, j\notin I}\boldsymbol{d}_{ij}$. Since $\sum_{1\leq i<j\leq n}\boldsymbol{d}_{ij}=\boldsymbol{0}$, this is $\frac{1}{2}\sum_{i,j\in I}\boldsymbol{d}_{ij}+\frac{1}{2}\sum_{i,j\in I^c}\boldsymbol{d}_{ij}$. Let $l_i$ denote the $i$-th row of the matrix $A_n$, and note that $\sum_{i\in I} l_i-\sum_{i\notin I} l_i=\sum_{i,j\in I}\boldsymbol{e}_{ij}-\sum_{i,j\in I^c}\boldsymbol{e}_{ij}$, so $\sum_{i,j\in I}\boldsymbol{d}_{ij}=\sum_{i,j\in I^c}\boldsymbol{d}_{ij}$. Thus the image of $\boldsymbol{e}_I=\sum_{i,j\in I}\boldsymbol{d}_{ij}=\sum_{i,j\in I^c}\boldsymbol{d}_{ij}$. Since this is equal to $(D(I\mid C^T))_I=\boldsymbol{r}_I$, we conclude that $\boldsymbol{r}_I$ is the image of $\boldsymbol{e}_I$. Note that this map is induced from a map of lattices, as the relevant lattice in $\mathbb{R}^{\binom{n}{2}}$ is the index two sublattice of $\mathbb{Z}^{\binom{n}{2}}$ with even coordinate sum.

The fact that $\Delta$ is simplicial of dimension $n-3$ is due to [Robinson and Whitehouse 1996]. Recall that a fan is smooth if for each cone the intersection of the lattice with the linear span of that cone is generated by the first lattice points on each ray of the cone. The fact that the fan is smooth follows from the work of Feichtner and Yuzvinsky. In [Feichtner 2006] it is shown that the fan $\Delta$ is the one associated to the nested set complex for a related hyperplane arrangement, while in [Feichtner and Yuzvinsky 2004, Proposition 2] it is shown that the fans associated to nested set complexes are smooth. □

**Notation 5.4.** Let $A_n$ and $D$ be the matrices described in Definition 5.3. Recall that the *chamber complex* $\Sigma(D)$ is the polyhedral fan in $\mathbb{R}^{\binom{n}{2}-n}$ subdividing the cone spanned by the columns of $D$ obtained by intersecting all simplicial cones spanned by columns of $D$. This is equal to the secondary fan of $A_n$, and thus to the fan of the toric variety $X_{\Sigma^{\mathrm{Ch}}}$ (see [Billera et al. 1990]). Recall also that the regular subdivision $\Delta_w$ of the configuration of the columns $\boldsymbol{a}_{ij}$ of $A_n$ corresponding to a vector $w\in\mathbb{R}^{\binom{n}{2}}$ has $\operatorname{pos}(\boldsymbol{a}_{ij}:ij\in\sigma)$ as a cell for $\sigma\subseteq\{ij:1\leq i<j\leq n\}$ if and only if there is some $\boldsymbol{c}\in\mathbb{R}^n$ such that $\boldsymbol{c}\cdot\boldsymbol{a}_{ij}=w_{ij}$ for $ij\in\sigma$, and $\boldsymbol{c}\cdot\boldsymbol{a}_{ij}<w_{ij}$ for $ij\notin\sigma$ (see [Gel'fand et al. 1994; Sturmfels 1996; Maclagan and Thomas 2007] for background on regular subdivisions). We denote by $\mathbb{N}A_n$ the subsemigroup of $\mathbb{N}^n$ generated by the columns of $A_n$, and by $\mathbb{R}_{\geq 0}A_n$ the cone in $\mathbb{R}^n$ whose rays are the positive spans of the columns of $A_n$.

**Proposition 5.5.** *The fan $\Delta$ is a subfan of the fan $X_{\Sigma^{\mathrm{Ch}}}$, so $X_\Delta$ is a toric subvariety of $X_{\Sigma^{\mathrm{Ch}}}$.*

*Proof.* Let $\sigma$ be a top-dimensional cone of $\Delta$. Then $\sigma=\operatorname{pos}(\boldsymbol{r}_{I_1},\dots,\boldsymbol{r}_{I_{n-3}})$, where $I_i\subseteq I_j$, $I_j\subseteq I_i$, or $I_i\cup I_j=[n]$. There is a trivalent (phylogenetic) tree $\tau$ with $n$ labeled leaves such that the $I_j$ correspond to the splits obtained by deleting internal edges of $\tau$.

Since $\sigma$ is a cone in $V=\mathbb{R}^{\binom{n}{2}}/\operatorname{row}(A)$, we can chose a lift $w\in\tilde{V}=\mathbb{R}^{\binom{n}{2}}$ for a vector in $\sigma$, and thus consider the regular subdivision $\Delta_w$ of the configuration $A_n$, which does not depend on the choice of lift.

To show that $\sigma$ is a cone in the chamber complex of $D$ we first characterize the subdivision $\Delta_w$ coming from a lift $w$ of a vector in $\sigma$. To an internal vertex $v$ of the phylogenetic tree $\tau$ we associate the set $\mathscr{C}_v$ of pairs $ij$ such that the path in $\tau$ between the vertices labeled $i$ and $j$ passes through $v$, and the cone $C_v = \operatorname{pos}(\boldsymbol{e}_{ij} : ij \in \mathscr{C}_v) \subseteq \mathbb{R}^n$. The cone $C_v$ is obtained by taking the cone over the polytopes $M(\cong_v)$ of [Kapranov 1993, Remark 1.3.7].

We claim that for any lift $w \in \mathbb{R}^{\binom{n}{2}}$ of a vector in $\sigma$ the subdivision $\Delta_w$ has cones $C_v$ for $v$ an internal vertex of $\tau$. To see this, for each $\boldsymbol{r}_I \in \sigma$ set $I' = [n] \setminus I$ if the path from leaf $i$ to leaf $j$ in $\tau$ passes through $v$ for some $i, j \in I$, and $I' = I$ otherwise. This ensures that $ij \notin \mathscr{C}_v$ for $i, j \in I'$. If $\sum_{i=1}^{n-3} a_i \boldsymbol{r}_{I_i}$ is a point in $\sigma$ with $a_i > 0$ for all $i$, we can choose the lift $w = \sum_{i=1}^{n-3} a_i \sum_{i,j \in I_i'} \boldsymbol{e}_{ij}$. Since $w_{ij} > 0$ for $ij \notin \mathscr{C}_v$, and $w_{ij} = 0$ for $ij \in \mathscr{C}_v$, taking $\boldsymbol{c} = \boldsymbol{0}$ we see that $C_v$ is a face of $\Delta_w$.

This shows that the given collection are cones in the subdivision $\Delta_w$. That they cover all of the cone generated by the $\boldsymbol{a}_{ij}$ is [Kapranov 1993, Claim 1.3.9]. For the reader's convenience we give a self-contained proof. To show that we are not missing anything, it suffices to show that any $\boldsymbol{v} \in \mathbb{N}A_n$ lies in $C_v$ for some $C_v$, which will show that the $C_v$ cover $\mathbb{R}_{\geq 0}A_n$. If $\boldsymbol{v}$ lies in $\mathbb{N}A_n$ then there is a graph $\Gamma$ on $n$ vertices with degree sequence $\boldsymbol{v}$. We may assume that $\Gamma$ has the largest edge sum out of all graphs with degree sequence $\boldsymbol{v}$, where an edge $ij$ has weight the number of internal edges in the path between $i$ and $j$ in the tree $\tau$. This means that if $ij$ and $kl$ are two edges of $\Gamma$ with $|\{i, j, k, l\}| = 4$, the paths in $\tau$ corresponding to these two edges must cross, as otherwise we could get a larger weight by replacing these two edges by the pair with the same endpoints that do cross. We claim that there is then some vertex $v \in \tau$ for which the path in $\tau$ corresponding to each edge in $\Gamma$ passes through $v$, which will show that $\boldsymbol{v} \in C_v$. The statement is trivial if $\Gamma$ has at most two edges, since any two paths must intersect. The set of collections of edges of $\Gamma$ for which the corresponding paths share a common vertex forms a simplicial complex, so if the claim is false, we can find a subgraph $\Gamma'$ of $\Gamma$ for which there is no vertex of $\tau$ through which all of the corresponding paths pass, but every proper subgraph of $\Gamma'$ has the desired property ($\Gamma'$ is a minimal nonface of the simplicial complex). The subgraph $\Gamma'$ must have at least three edges. Pick three edges $e_1, e_2, e_3$ of $\Gamma'$, and let $v_i$ for $i \in \{1, 2, 3\}$ be a vertex of $\tau$ for which the path corresponding to each vertex of $\Gamma'$ except $e_i$ passes. Since $\tau$ is a tree one of these vertices lies in the path between the other two; without loss of generality we assume that $v_2$ lies between $v_1$ and $v_3$. But the path corresponding to $e_2$ passes through $v_1$ and $v_3$ while avoiding $v_2$, a contradiction.

This shows that the subdivision corresponding to a vector in the interior of $\sigma$ is the same subdivision $\Delta_\sigma$ for any vector in $\sigma$, and that $\sigma$ lies inside the cone $\bigcap_v \operatorname{pos}(\boldsymbol{r}_{ij} : ij \notin \mathscr{C}_v)$ of the chamber complex of $D$. To finish the proof, we show that this cone lies in $\sigma$. The cone $\sigma$ has the following facet description: for each quadruple $i, j, k, l$ two of the pairs of paths $\{ij, kl\}$, $\{ik, jl\}$, and $\{il, jk\}$ in the tree $\tau$ have the same combined length, and one pair is shorter. Without loss of generality we assume that $ij, kl$ is the shorter pair. This gives the inequality $w_{ij} + w_{kl} \geq w_{ik} + w_{jl} = w_{il} + w_{jk}$. The set of these inequalities as $\{i, j, k, l\}$ ranges over all 4-tuples of $[n]$ gives a description of the lift in $\mathbb{R}^{\binom{n}{2}}$ of the cone $\sigma$ (see [Speyer and Sturmfels 2004, Theorem 4.2]). Note that $\operatorname{row}(A)$ lies in this cone, so to show that a vector $\boldsymbol{v}$ in the relative interior of $\bigcap_v \operatorname{pos}(\boldsymbol{r}_{ij} : ij \notin \mathscr{C}_v)$ lies in $\sigma$, it suffices to show that for each inequality there is some lift of $\boldsymbol{v}$ to $\mathbb{R}^{\binom{n}{2}}$ that satisfies that inequality. Given such a $\boldsymbol{v}$, and a 4-tuple $\{i, j, k, l\}$ giving the inequality $w_{ij} + w_{kl} \geq w_{ik} + w_{jl} = w_{il} + w_{jk}$, pick a vertex $v$ in $\tau$ that lies in all of the paths $ik$, $jl$, $il$, and $jk$. Then since $\boldsymbol{v} \in \operatorname{relint}(\operatorname{pos}(\boldsymbol{r}_{ij} : ij \notin \mathscr{C}_v))$, there is a lift $w$ of $\boldsymbol{v}$ with $w_{ik} = w_{jl} = w_{il} = w_{jk} = 0$, and $w_{ij}, w_{kl} \geq 0$, which satisfies the given inequality. We conclude that $\bigcap_v \operatorname{pos}(\boldsymbol{r}_{ij} : ij \notin \mathscr{C}_v) \subseteq \sigma$, and thus we have equality, so $\sigma$ is a cone in the chamber complex of $D$. □

Recall that an open cone $\sigma \in \Sigma^{\mathrm{H}}$ consists of $w$ for which the saturation of $\operatorname{in}_w(I_A)$ is constant, where we here follow the standard convention that the leading form $\operatorname{in}_w(f)$ of a polynomial $f$ consists of terms of largest $w$-degree.

**Proposition 5.6.** *The fan $\Delta$ is a subfan of the fan $\Sigma^{\mathrm{H}}$, so $X_\Delta$ is a toric subvariety of $X_{\Sigma^{\mathrm{H}}}$.*

*Proof.* Continuing with the notation from the proof of Proposition 5.5, we first show that $\operatorname{in}_w(I_{A_n})$ is constant for all lifts $w$ of vectors in a maximal cone $\sigma$ of $\Delta$ corresponding to a phylogenetic tree $\tau$. A planar representation of $\tau$ with the vertices on a circle determines a circular order on $[n]$. Without loss of generality we may assume that this is the standard increasing order. Draw the complete graph $K_n$ on the circle with the same order. By [Sturmfels 1996, Theorem 9.1] a reduced Gröbner basis $\mathscr{B}$ for $I_{A_n}$ is given by binomials of the form $x_{ij}x_{kl} - x_{ik}x_{jl}$, where the edges $ik$ and $jl$ of $K_n$ cross but the edges $ij$ and $kl$ do not. The open Gröbner cone corresponding to this is $\mathscr{C} = \{w \in \mathbb{R}^{\binom{n}{2}} : w_{ij} + w_{kl} > w_{ik} + w_{jl}\}$. Note that, as in the proof of Proposition 5.5, the lift of any vector in $\sigma$ lies on the boundary of $\mathscr{C}$, and so there is a term order $\prec$ for which the initial ideal $\operatorname{in}_\prec(\operatorname{in}_w(I_{A_n}))$ equals the ideal $\langle x_{ij}x_{kl} : ij, kl \text{ do not cross in } K_n\rangle$, where $w$ is a lift of any vector in $\sigma$. Thus by [Sturmfels 1996, Corollary 1.9], a Gröbner basis for $I_{A_n}$ with respect to the weight order given by such a $w$ is obtained by taking the initial terms with respect to $w$ of the Gröbner basis $\mathscr{B}$. For a binomial $x_{ij}x_{kl} - x_{ik}x_{jl}$, where the edges $ik$ and $jl$ of $K_n$ cross but the edges $ij$ and $kl$ do not, either the paths $ik$ and $jl$ also cross in $\tau$, or they do not. If the paths do cross, the lift $w$ of a vector in $\sigma$ has $w_{ik} + w_{jl} = w_{ij} + w_{jl}$, and if they do not there is an internal edge of $\tau$ in the paths $ik$ and $jl$ but not the paths $ij$ and $kl$, so if $w$ lies in the interior of $\sigma$ we have $w_{ik} + w_{jl} < w_{ij} + w_{jl}$. This means that the initial ideal is determined by the tree $\tau$, so $\operatorname{in}_w(I_{A_n})$ is constant for all lifts $w$ of vectors in $\sigma$. This shows that $\sigma$ is contained in a cone $\sigma'$ of $\Sigma^{\mathrm{H}}$.

It remains to show that $\sigma = \sigma'$. This follows from Proposition 5.5, which shows that $\sigma$ is a cone in the secondary fan of $I_{A_n}$, since the Gröbner fan refines the secondary fan [Sturmfels 1996, Proposition 8.15]. The fan $\Sigma^{\mathrm{H}}$ is obtained from the Gröbner fan of $I_{A_n}$ by amalgamating cones corresponding to initial ideals with the same saturation with respect to $\langle x_{ij} : 1 \leq i < j \leq n \rangle$, so the result follows, since ideals with the same saturation have the same radical and thus the cones live in the same secondary cone. □

We are now able to show that $\overline{M}_{0,n}$ is a subvariety of $X_\Delta$. As described in Example 3.1 the moduli space $\overline{M}_{0,n}$ is both the Chow and Hilbert quotient of the Grassmannian $G(2,n)$ by the torus $T^{n-1}$.

**Theorem 5.7.** *The toric variety $X_\Delta$ is the union of those $T^{\binom{n}{2}-n}$-orbits of $X_{\Sigma^\star}$ intersecting the closure of $M_{0,n}$ in this $X_{\Sigma^\star}$. The closure of $M_{0,n} \subseteq T^{\binom{n}{2}-n}$ inside $X_\Delta$ is equal to $\overline{M}_{0,n}$.*

We remark that the second part of this result was originally observed by related methods in [Tevelev 2007, Theorem 5.5].

*Proof of Theorem 5.7.* By Propositions 5.5 and 5.6 we know that $\Delta$ is a subfan of both $\Sigma^{\mathrm{Ch}}$ and $\Sigma^{\mathrm{H}}$, and by Proposition 5.2 we know that the support of $\Delta$ is the tropical variety of $M_{0,n} \subset T^{\binom{n}{2}-n}$. Corollary 2.7 says that the closure of $M_{0,n} \subset T^{\binom{n}{2}-n}$ inside $X_{\Sigma^\star}$ intersects the orbit corresponding to a cone $\sigma \in \Sigma^\star$ if and only the tropical variety of $M_{0,n}$ intersects the interior of $\sigma$. Thus the orbits of $X_{\Sigma^\star}$ intersecting this closure are precisely the orbits in the toric subvariety $X_\Delta$. This proves the first assertion of the theorem. For the second, note that the closure of $M_{0,n}$ in $X_\Delta$ is $G(2,n)/\!\!/_n^\star T^{n-1}$, and since $\overline{M}_{0,n} = G(2,n)/\!\!/^\star T^{n-1}$ is smooth and irreducible, this is equal to $\overline{M}_{0,n}$ by Theorem 3.2. □

**Remark 5.8.** We emphasize that although we know that the fans $\Sigma^{\mathrm{Ch}}$ and $\Sigma^{\mathrm{H}}$ are the secondary and saturated Gröbner fans respectively, we do not even know how many rays each has. So while in theory one could describe equations for $\overline{M}_{0,n}$ in the Cox rings of these toric varieties, this is not possible in practice. On the other hand, $\Delta$ has a completely explicit description for all $n$. In particular, this makes the Cox ring of $X_\Delta$ accessible and enables us to derive equations for $\overline{M}_{0,n}$ inside it.

## 6. Equations for $\overline{M}_{0,n}$

In this section we apply Theorem 3.2 to give equations for $\overline{M}_{0,n}$ in the Cox ring of the toric variety $X_\Delta$ described in Section 5.

Recall that $\mathcal{I} = \{I \subset [n] : 1 \in I, |I|, |I^c| \geq 2\}$, and there is a ray $\operatorname{pos}(\boldsymbol{r}_I)$ of $\Delta$ for each $I \in \mathcal{I}$. The Cox ring of $X_\Delta$ is $S = \Bbbk[x_I : I \in \mathcal{I}]$, with $\deg x_I = [D_I] \in \mathrm{Cl}(X_\Delta)$, where $D_I$ is the torus-invariant divisor corresponding to the ray through $\boldsymbol{r}_I$. We construct the grading matrix as follows. Set $b = |\mathcal{I}| - \binom{n}{2}$ for $n \geq 5$. Let $G$ be the $(n+b) \times |\mathcal{I}|$ matrix that is given in block form by

$$G = \left(\begin{array}{c|c} A_n & 0 \\ \hline -C & I_b \end{array}\right), \tag{5}$$

where 0 denotes the $n \times b$ zero matrix, $A_n$ and $C$ are as in Definition 5.3, and $I_b$ is the $b \times b$ identity matrix. For $n = 4$ we set $G = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$.

As the following lemma shows, this form of $G$ is consistent with the choice of grading matrix for the Cox ring of $X_\Delta$ suggested in (3). The lemma also shows that $\mathrm{Cl}(X_\Delta) \cong \mathrm{Pic}(\overline{M}_{0,n})$, so we first recall the description of $\mathrm{Pic}(\overline{M}_{0,n})$. For $I \subset [n]$ with $1 \notin I$ the notation $\boldsymbol{e}_I$ means the basis element $\boldsymbol{e}_{I^c}$ of $\mathbb{Z}^{|\mathcal{I}|}$.

**Proposition 6.1** [Keel 1992, Theorem 1]. *Let $W$ be the sublattice of $\mathbb{Z}^{|\mathcal{I}|}$ spanned by the vectors*

$$w_{ijkl} = \prod_{\substack{i,j \in I \\ k,l \notin I}} \boldsymbol{e}_I - \prod_{\substack{i,l \in I \\ j,k \notin I}} \boldsymbol{e}_I,$$

*where $\{i, j, k, l\} \subseteq [n]$ has size four. The Picard group of $\overline{M}_{0,n}$ is isomorphic to $\mathbb{Z}^{|\mathcal{I}|}/W$.*

**Lemma 6.2.** *For $n \geq 5$ the divisor class group of $X_\Delta$ is isomorphic to $\mathbb{Z}^{b+n}$, with the image of $[D_I]$ under this isomorphism equal to the column $\boldsymbol{g}_I$ of the matrix $G$ indexed by $I \in \mathcal{I}$. We have $\mathrm{Cl}(X_\Delta) \cong \mathrm{Pic}(\overline{M}_{0,n})$, with the isomorphism taking $[D_I]$ to the boundary divisor $\delta_I$.*

*Proof.* That $\mathrm{Cl}(X_\Delta) \cong \mathbb{Z}^{b+n}$ follows from the short exact sequence (†) computing the class group of a toric variety [Fulton 1993, p. 63], together with Proposition 5.2, since smooth toric varieties have torsion-free divisor class groups. To see that the image of $[D_I]$ is $\boldsymbol{g}_I$, the $I$-th column of $G$, it suffices to show that the matrix $G$ is a Gale dual for the matrix $R$, so the exact sequence (†) is

$$0 \to M \xrightarrow{R^T} \mathbb{Z}^{|\mathcal{I}|} \xrightarrow{G} \mathrm{Cl}(X_\Delta) \to 0.$$

Now $GR^T = G(I \mid C^T)^T D^T$, so

$$GR^T = \left(\begin{array}{c|c} A_n & 0 \\ \hline -C & I \end{array}\right) \left(\begin{array}{c} I \\ \hline C \end{array}\right) D^T = \left(\begin{array}{c} AD^T \\ \hline 0 \end{array}\right) = 0.$$

Finally, to show that $\mathrm{Cl}(X_\Delta) \cong \mathrm{Pic}(\overline{M}_{0,n})$, since $\mathrm{Cl}(X_\Delta) \cong \mathbb{Z}^{|\mathcal{I}|}/\operatorname{im}_{\mathbb{Z}} R^T$ and $\mathrm{Pic}(\overline{M}_{0,n}) \cong \mathbb{Z}^{|\mathcal{I}|}/W$, it suffices to show that $W = \operatorname{im}_{\mathbb{Z}} R^T$. Since $\operatorname{im}_{\mathbb{Z}} R^T = \ker_{\mathbb{Z}} G$, both lattices $W$ and $\operatorname{im} R^T$ are saturated, and $\operatorname{rank} W = \operatorname{rank} R^T = \binom{n}{2} - n$, it suffices to show that each $w_{ijkl}$ lies in the kernel of $G$. Now $w_{ijkl}$ restricted to the sets $I$ with $|I| = 2$ or $|I| = n-2$ is $\boldsymbol{e}_{ij} + \boldsymbol{e}_{kl} - \boldsymbol{e}_{il} - \boldsymbol{e}_{jk}$, which lies in $\ker A_n$. Restricting $w_{ijkl}$ to $\{\boldsymbol{e}_{st} : s, t \in I\} \cup \{\boldsymbol{e}_I\}$ we see that $\left(-C \,\middle|\, I_b\right) w_{ijkl} = 0$. For example, if

$1, i, j, k, l \in I$ then the sum is $-1-1+1+1+0=0$. Thus $w_{ijkl} \in \ker G$ as required. □

For $n=4$ we have $\mathrm{Cl}(X_\Delta) \cong \mathrm{Pic}(\overline{M}_{0,4}) \cong \mathbb{Z}$, and the image of the $\boldsymbol{g}_{ij}$ is $1 \in \mathbb{Z}$, which is also equal to each $[D_{ij}]$.

We begin by proving the first part of Theorem 1.2. Recall that

$$\mathcal{I} = \{I \subset [n] : 1 \in I,\ |I| \geq 2,\ |I^c| \geq 2\}.$$

If $1 \notin I \subseteq [n]$ then by $x_I$ we mean the variable $x_{[n]\setminus I}$ in the Cox ring $S = \Bbbk[x_I : I \in \mathcal{I}]$ of $X_\Delta$.

**Theorem 6.3.** *For $n \geq 5$ the equations for $\overline{M}_{0,n}$ in the Cox ring of $X_\Delta$ are obtained by homogenizing the Plücker relations with respect to the grading of $S$ and then saturating by the product of the variables of $S$. Specifically, the ideal is*

$$I(\overline{M}_{0,n}) = \Big(\Big\langle \prod_{\substack{ij\in I\\ kl\notin I}} - \prod_{\substack{ik\in I\\ jl\notin I}} x_I + \prod_{\substack{il\in I\\ jk\notin I}} x_I \Big\rangle : \Big(\prod_I x_I\Big)^\infty\Big),$$

*where the generating set runs over all $\{i, j, k, l\}$ with $1 \leq i < j < k < l \leq n$.*

Before proving the theorem, we first find equations for the intersection $M_{0,n}$ of $\overline{M}_{0,n}$ with the torus $\mathbb{T} = T^{\binom{n}{2}-1}/T^{n-1} \cong (\Bbbk^\times)^{\binom{n}{2}-n}$ of the toric variety $X_\Delta$. The coordinates for $\mathbb{T}$ are labeled by $\mathscr{E} = \{ij : 2 \leq i < j \leq n, ij \neq 23\}$. Recall that the ideal of $G(2,n) \subset \mathbb{P}^{\binom{n}{2}-1}$ is generated by the Plücker relations $p_{ijkl} = x_{ij}x_{kl} - x_{ik}x_{jl} + x_{il}x_{jk}$:

$$I_{2,n} = \langle x_{ij}x_{kl} - x_{ik}x_{jl} + x_{il}x_{jk} : 1 \leq i < j < k < l \leq n\rangle.$$

**Proposition 6.4.** *The intersection $M_{0,n} = \big(Z(I_{2,n}) \cap T^{\binom{n}{2}}\big)/T^n \subseteq \mathbb{T}$ is cut out by the equations*

$$J = \langle z_{kl} - z_{2l} + z_{2k} : 3 \leq k \leq l \leq n\rangle \subseteq \Bbbk[z_{ij}^{\pm 1} : ij \in \mathscr{E}],$$

*where we set $z_{23} = 1$.*

*Proof.* We first show that $\big(Z(I_{2,n}) \cap T^{\binom{n}{2}}\big)/T^n$ is defined by the ideal

$$J' = \langle z_{ij}z_{kl} - z_{ik}z_{jl} + z_{il}z_{jk} : 1 \leq i < j < k < l \leq n\rangle \subseteq \Bbbk[z_{ij}^{\pm 1} : ij \in \mathscr{E}],$$

where we set $z_{ij} = 1$ when $ij \notin \mathscr{E}$. The content of Proposition 2.1 is that the relevant ideal is $\phi^{-1}(I_{2,n})$, where $\phi : \Bbbk[z_{ij}^{\pm 1} : ij \in \mathscr{E}] \to \Bbbk[x_{ij}^{\pm 1} : 1 \leq i < j \leq n]$ is given by $\phi(z_{ij}) = \prod_{kl} x_{kl}^{D_{ij,kl}}$.

Since the map $\phi$ is an injection, to show that $J'$ is the desired ideal we just need to show that the generators of $J'$ are taken to a generating set for $I_{2,n} \subseteq \Bbbk[x_{ij}^{\pm 1}]$

by $\phi$. When $ij \in \mathscr{E}$ we have

$$\phi(z_{ij}) = \begin{cases} (x_{ij}x_{12}x_{13})/(x_{1i}x_{1j}x_{23}) & \text{if } i, j \geq 4, \\ (x_{2j}x_{13})/(x_{1j}x_{23}) & \text{if } i = 2, \\ (x_{3j}x_{12})/(x_{1j}x_{23}) & \text{if } i = 3. \end{cases}$$

The proof breaks down into several cases, depending on how many of $1, 2, 3$ lie in $\{i, j, k, l\}$. For example, if $1, 2, 3 \notin \{i, j, k, l\}$, then $z_{ij}z_{kl} - z_{ik}z_{jl} + z_{il}z_{jk} = (x_{ij}x_{kl} - x_{ik}x_{jl} + x_{il}x_{jk})(x_{12}x_{13})^2/(x_{1i}x_{1j}x_{1k}x_{1l}x_{23}^2)$. The other cases to check are:

- $i = 1, 2, 3 \notin \{j, k, l\}$,
- $i = 1$, $j = 2$, $3 \notin \{k, l\}$, or $j = 3$, $2 \notin \{k, l\}$,
- $i = 1$, $j = 2$, $k = 3$,
- $i = 2$, $1, 3 \notin \{j, k, l\}$ or $j = 3$, $1, 2 \notin \{i, k, l\}$, and
- $i = 2$, $j = 3$, $1 \notin \{k, l\}$.

In every case we see that the polynomial $\phi(z_{ij}z_{kl} - z_{ik}z_{jl} + z_{il}z_{jk})$ is equal to $x_{ij}x_{kl} - x_{ik}x_{jl} + x_{il}x_{jk}$ times a monomial in the $x_{mn}$. This shows that $\phi$ takes a generating set for $J'$ to a generating set for $I_{2,n} \subset \Bbbk[x_{ij}^{\pm 1} : 1 \leq i < j \leq n]$, and thus $J'$ is the ideal of $\big(Z(I_{2,n}) \cap T^{\binom{n}{2}}\big)/T^n$.

To see that $J = J'$, it suffices to show that all other generators of $J'$ lie in the ideal generated by these linear ones. Indeed,

$$\begin{aligned} &z_{ij}z_{kl} - z_{ik}z_{jl} + z_{il}z_{jk} \\ &\quad = z_{ij}(z_{kl} - z_{2l} + z_{2k}) - z_{ik}(z_{jl} - z_{2l} + z_{2j}) + z_{il}(z_{jk} - z_{2k} + z_{2j}) \\ &\qquad + (z_{2l} - z_{2k})(z_{ij} - z_{2j} + z_{2i}) + (z_{2k} - z_{2j})(z_{il} - z_{2l} + z_{2i}) + (z_{2j} - z_{2l})(z_{ik} - z_{2k} + z_{2i}). \end{aligned}$$

□

**Remark 6.5.** Consider the ideal $\tilde{J} \subset \Bbbk[z_{ij} : ij \in \mathscr{E} \cup \{23\}]$ obtained by homogenizing the ideal $J$ by adding the variable $z_{23}$. The variety $Z(\tilde{J}) \subseteq \mathbb{P}^{\binom{n-1}{2}-1}$ is the linear subspace equal to the row space of the $(n-1) \times \binom{n-1}{2}$ matrix $\tilde{A}$ whose columns are the positive roots of the root system $A_{n-2}$. Specifically, the rows of $\tilde{A}$ are indexed by $2, \ldots, n$, and the columns are indexed by $\{\{i, j\} : 2 \leq i < j \leq n\}$, with $\tilde{A}_{i,\{j,k\}}$ equal to 1 if $i = j$, $-1$ if $i = k$, and zero otherwise. The variety $M_{0,n}$ is the intersection of $Z(\tilde{J}) \subseteq \mathbb{P}^{\binom{n-1}{2}-1}$ with the torus $T^{\binom{n-1}{2}-1}$ of $\mathbb{P}^{\binom{n-1}{2}-1}$. This exhibits $M_{0,n}$ as a hyperplane complement and as a very affine variety in its intrinsic torus [Tevelev 2007].

*Proof of Theorem 6.3.* By Theorem 5.7 $X_\Delta$ is a sufficiently large toric subvariety of $X_{\Sigma^\star}$, so Theorem 3.2 describes how to get equations for $\overline{M}_{0,n}$ inside $X_\Delta$. Equation (4) defines $V = \big(I_{\binom{n}{2}} \mid C^T\big)$, so the map $\nu : \Bbbk[x_{ij} : 1 \leq i < j \leq n] \to \Bbbk[x_I^{\pm 1} : I \in \mathscr{I}]$

of Theorem 3.2 is given by $\nu(x_{ij}) = x_{ij}\prod_{i,j\in I} x_I$, where the product is over $I \in \mathcal{I}$ with $1 \in I$, and $3 \le |I| \le n-3$. Thus, for $i$, $j$, $k$, and $l$ distinct, we have

$$\nu(x_{ij}x_{kl}) = x_{ij}x_{kl}\prod_{ij\in I} x_I \prod_{kl\in I} x_I,$$

with the same restrictions on the products. Using the convention $x_I = x_{[n]\setminus I}$, we can write this as

$$\nu(x_{ij}x_{kl}) = x_{ij}x_{kl}\prod_{\substack{ij\in I\\ kl\notin I}} x_I \prod_{\substack{1\in I\\ |\{i,j,k,l\}\cap I|=3}} x_I \prod_{1,i,j,k,l\in I} x_I^2,$$

where here the products are over all $I \in \mathcal{I}$ with $3 \le |I| \le n-3$, and there is no restriction that $1 \in I$ in the first product. Write

$$M_{ijkl} = \prod_{\substack{1\in I\\ |\{i,j,k,l\}\cap I|=3}} x_I \prod_{1,i,j,k,l\in I} x_I^2.$$

Then

$$\nu(p_{ijkl}) = M_{ijkl}\Big(\prod_{\substack{ij\in I\\ kl\notin I}} x_I - \prod_{\substack{ik\in I\\ jl\notin I}} x_I + \prod_{\substack{il\in I\\ jk\notin I}} x_I\Big).$$

Note that $\nu(p_{ijkl})$ is already a polynomial, so there is no need to clear denominators. Thus by Theorem 3.2 the ideal $I_{\overline{M}_{0,n}}$ in the Cox ring of $X_\Delta$ is given by

$$\begin{aligned} I_{\overline{M}_{0,n}} &= \Big(\Big\langle\Big(\prod_{\substack{ij\in I\\ lk\notin I}} x_I - \prod_{\substack{ik\in I\\ jl\notin I}} x_I + \prod_{\substack{il\in I\\ jk\notin I}} x_I\Big)M_{ijkl}\Big\rangle : \Big(\prod_{I\in\mathcal{I}} x_I\Big)^\infty\Big)\\ &= \Big(\Big\langle\prod_{\substack{ij\in I\\ kl\notin I}} x_I - \prod_{\substack{ik\in I\\ jl\notin I}} x_I + \prod_{\substack{il\in I\\ jk\notin I}} x_I\Big\rangle : \Big(\prod_{I\in\mathcal{I}} x_I\Big)^\infty\Big), \end{aligned}$$

where the generating sets run over all $\{i, j, k, l\}$ with $1 \le i < j < k < l \le n$. □

**Example 6.6.** (1) The case $n = 5$ is covered in Examples 2.2, 3.5, and 4.5.

(2) When $n=6$, the ideal $I_{2,6}$ is generated by the equations $p_{ijkl} = x_{ij}x_{kl} - x_{ik}x_{jl} + x_{il}x_{jk}$, where $1 \le i < j < k < l \le 6$ in the ring $\Bbbk[x_{ij} : 1 \le i < j \le 6]$. The Cox ring of $X_\Delta$ is $\Bbbk[x_I : I \in \mathcal{I}]$. Applying the change of coordinates given by the matrix $R$ the $p_{ijkl}$ become $\tilde{p}_{ijkl} = x_{ij}x_{kl}x_{ijm}x_{ijn} - x_{ik}x_{jl}x_{ikm}x_{ikn} + x_{il}x_{jk}x_{ilm}x_{iln}$, where $\{i, j, k, l, m, n\} = \{1, 2, 3, 4, 5, 6\}$. The ideal $I_{\overline{M}_{0,6}}$ has additional generators

$$\begin{aligned} q_{ij} = x_{ik}x_{jk}x_{ijk}^2x_{lm}x_{ln}x_{mn} - x_{il}x_{jl}x_{ijl}^2x_{km}x_{kn}x_{mn}\\ + x_{im}x_{jm}x_{ijm}^2x_{kl}x_{kn}x_{ln} - x_{in}x_{jn}x_{ijn}^2x_{kl}x_{km}x_{lm}, \end{aligned}$$

for $1 \le i < j \le 6$, where again $\{i, j, k, l, m, n\} = \{1, \dots, 6\}$.

**Remark 6.7.** When $n=4$ we can still follow the recipe of Section 3 to obtain the equations for $\overline{M}_{0,4}\cong\mathbb{P}^1$ inside $X_{\Sigma^*}\cong\mathbb{P}^2$. The Grassmannian $G(2,5)\subseteq\mathbb{P}^5$ is the hypersurface $Z(x_{12}x_{34}-x_{13}x_{24}+x_{14}x_{23})$. In this case $M_{0,4}=Z(z_{34}-z_{24}+1)$ in $T^2=\operatorname{Spec}(\Bbbk[z_{24}^{\pm 1},z_{34}^{\pm 1}])$. Also

$$R=\begin{pmatrix}1&0&-1&-1&0&1\\0&1&-1&-1&1&0\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\\0&0&0\end{pmatrix},\quad\text{so } V=\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\end{pmatrix}.$$

Thus $\nu(x_{12}x_{34}-x_{13}x_{24}+x_{14}x_{23})=y_1-y_2+y_3\subseteq\Bbbk[y_1,y_2,y_3]=\operatorname{Cox}(\mathbb{P}^2)$.

## 7. VGIT and the effective cone of $\overline{M}_{0,n}$

An important invariant of a projective variety $Y$ is its pseudoeffective cone $\overline{\operatorname{Eff}}(Y)$, and one of the primary goals of Mori theory is to understand the decomposition of this cone into Mori chambers. In this section we prove Theorem 7.1, which is the second part of Theorem 1.2. We identify a subcone $\mathcal{G}$ of the effective cone of $\overline{M}_{0,n}$ for which one has that $\overline{M}_{0,n}$ can be constructed as a GIT quotient of an affine variety with linearization determined by a given $D\in\mathcal{G}$. This prompts Question 7.4(3), which asks whether the cone $E$ of divisors spanned by boundary classes is a Mori dream region of the effective cone.

Let $H=\operatorname{Hom}(\operatorname{Cl}(X_\Delta),\Bbbk^\times)\cong(\Bbbk^\times)^{b+n}$. The torus $H$ acts on $\mathbb{A}^{|\mathcal{I}|}$ with weights given by the columns of the matrix $G$ of (5). Recall the cone

$$\mathcal{G}(X_\Delta)=\bigcap_{\sigma\in\Delta}\operatorname{pos}([D_I]:I\notin\sigma)$$

from Definition 4.1. Let $i:\overline{M}_{0,n}\to X_\Delta$ be the inclusion of Theorem 5.7. The pullback $i^*(\mathcal{G}(X_\Delta))\subset N^1(\overline{M}_{0,n})\otimes\mathbb{R}$ is a subcone of the nef cone of $\overline{M}_{0,n}$.

**Theorem 7.1.** (1) *For rational $\alpha\in\operatorname{int}(\mathcal{G}(X_\Delta))$ we have the GIT construction of $\overline{M}_{0,n}$ as*

$$\overline{M}_{0,n}=Z(I_{\overline{M}_{0,n}})/\!\!/_\alpha H,$$

*where $Z(I_{\overline{M}_{0,n}})\subset\mathbb{A}^{|\mathcal{I}|}$ is the affine subscheme defined by $I_{\overline{M}_{0,n}}$.*

(2) *Let $n\geq 5$. Given $\beta\in\mathbb{N}A_n$ there is $\alpha\in\mathbb{N}G$ for which*

$$Z(I_{\overline{M}_{0,n}})/\!\!/_\alpha H=G(2,n)/\!\!/_\beta T^{n-1},$$

*so all GIT quotients of $G(2,n)$ by $T^{n-1}$ can be obtained from $\overline{M}_{0,n}$ by variation of the GIT.*

*Proof.* The first part of the theorem is a direct application of Proposition 4.3. For the second, note that for $n \geq 5$ there are no repeated columns in the matrix $D$, so the result follows from Theorem 4.6 and Corollary 4.7. □

**Remark 7.2.** (1) The second part of the theorem is still true for $n = 4$, as $\overline{M}_{0,4} \cong \mathbb{P}^1$, which is also equal to one of the GIT quotients of $G(2,4)$.

(2) When $n = 5$, $I_{\overline{M}_{0,5}} = I_{2,5}$, so we see that $\overline{M}_{0,5} = G(2,5)/\!\!/_\alpha T^4$ for $\alpha \in \operatorname{int}(\mathscr{G}(X_\Delta))$. In this case the second part of the theorem is a tautology.

(3) For larger $n$, since $A$ and $C$ are nonnegative matrices, expression

$$-\min\{(Cu)_i : Au = \beta, u \in \mathbb{Q}^{m+1}_{\geq 0}\}$$

of Theorem 4.6 is a nonpositive integer, and thus we can choose $\alpha = (\beta, 0)$.

**Remark 7.3.** A natural problem is to give a combinatorial description of these equations for $\overline{M}_{0,n}$, similar to that given for the GIT quotient $G(2,n)/\!\!/_\beta T^{n-1}$ in [Howard et al. 2009]. Generators for the corresponding ring can still be described by graphs on $n$ vertices, but an added complication is that their Kempe lemma is not true; the generators corresponding to noncrossing graphs no longer give a basis for the degree-one part of the ring.

Hu and Keel Hu and Keel 2000 Mori dream spaces are introduced, which are varieties whose effective cones are polyhedral and for which the Mori chamber decomposition breaks this cone into a finite number of polyhedral pieces. They prove that if a variety $Y$ is a Mori dream space, then there is an embedding of $Y$ into a projective toric variety $X_\Sigma$, and thus a GIT construction of $Y$, so that $Y$ and $X_\Sigma$ have isomorphic Picard groups and effective cones, and all small $\mathbb{Q}$-factorial modifications of $Y$ can be obtained by variation of the GIT quotient from $Y$. They also define the weaker notion of subcone $C \subseteq \overline{\operatorname{Eff}}(Y)$ being a Mori dream region. This holds if

$$R = \bigoplus_{D \in C} H^0(V, D)$$

is finitely generated.

Hu and Keel raised the question of whether $\overline{M}_{0,n}$ is a Mori dream space. Theorem 5.7 shows that $\overline{M}_{0,n}$ embeds into $X_\Delta$ and Lemma 6.2 shows that $\operatorname{Pic}(\overline{M}_{0,n}) \cong \operatorname{Cl}(X_\Delta)$. However, for $n \geq 6$, Keel and Vermeire showed that the cone $E$ generated by the boundary divisors of $\overline{M}_{0,n}$ is a proper subcone of $\operatorname{Eff}(\overline{M}_{0,n})$ [Vermeire 2002], so the effective cones of $\overline{M}_{0,n}$ and $X_\Delta$ differ. On the other hand, Castravet [2009] showed $\overline{M}_{0,6}$ is indeed a Mori dream space. We believe that $E$ may be a Mori dream region. Indeed, the GIT chambers of $Z(I_{\overline{M}_{0,n}})/\!\!/_\alpha H$ divide $E$ into polyhedral chambers, each of which corresponds to a different compactification of $M_{0,n}$. The chamber containing $\mathscr{G}(X_\Delta)$ corresponds to the compactification $\overline{M}_{0,n}$.

**Question 7.4.** (1) Let $\overline{M}_{0,6} \to X_\Sigma$ be the embedding of $\overline{M}_{0,6}$ into a toric variety of dimension 24 with isomorphic Picard group and effective cone guaranteed by the Mori dream space construction. Let $X_{\Sigma'}$ be the toric subvariety obtained from $X_\Sigma$ by removing $T^{24}$-orbits of $X_\Sigma$ not intersecting $\overline{M}_{0,6}$. Is $X_{\Sigma'}$ obtained from $X_\Delta$ in a natural way (such as by tropical modifications) that generalizes to $n > 6$?

(2) Is there a toric embedding $\overline{M}_{0,n} \to X_\Sigma$ with $\overline{M}_{0,n}$ and $X_\Sigma$ having isomorphic Picard groups and effective cones that can be obtained from $X_\Delta$ by tropical modifications? This would support the conjecture that $\overline{M}_{0,n}$ is a Mori dream space.

(3) Let $E$ be the closed subcone of $\mathrm{Eff}(\overline{M}_{0,n})$ spanned by the boundary divisors, let $S$ be the Cox ring of $X_\Delta$ and $I_{\overline{M}_{0,n}}$ the ideal of $\overline{M}_{0,n}$ in $S$. Is

$$(S/I_{\overline{M}_{0,n}})_D \cong H^0(\overline{M}_{0,n}, D)$$

for all $D \in E$? This would imply that $E$ is a Mori dream region for $\overline{M}_{0,n}$.

***Acknowledgements.*** We thank Klaus Altmann for helpful discussions about Chow quotients. We also thank Sean Keel and Jenia Tevelev for useful conversations. The authors were partially supported by NSF grants DMS-0509319 (Gibney) and DMS-0500386 (Maclagan).

## References

[Bayer and Morrison 1988] D. Bayer and I. Morrison, "Standard bases and geometric invariant theory, I: Initial ideals and state polytopes", *J. Symbolic Comput.* **6** (1988), 209–217. MR 90e:13001 Zbl 0675.13014

[Białynicki-Birula and Sommese 1987] A. Białynicki-Birula and A. J. Sommese, "A conjecture about compact quotients by tori", pp. 59–68 in *Complex analytic singularities*, edited by T. Suwa and P. Wagreich, Adv. Stud. Pure Math. **8**, North-Holland, Amsterdam, 1987. MR 88j:32042 Zbl 0616.14037

[Billera et al. 1990] L. J. Billera, P. Filliman, and B. Sturmfels, "Constructions and complexity of secondary polytopes", *Adv. Math.* **83**:2 (1990), 155–179. MR 92d:52028 Zbl 0714.52004

[Billera et al. 2001] L. J. Billera, S. P. Holmes, and K. Vogtmann, "Geometry of the space of phylogenetic trees", *Adv. in Appl. Math.* **27**:4 (2001), 733–767. MR 2002k:05229 Zbl 0995.92035

[Buneman 1974] P. Buneman, "A note on the metric properties of trees", *J. Combinatorial Theory Ser. B* **17** (1974), 48–50. MR 51 #218 Zbl 0286.05102

[Castravet 2009] A.-M. Castravet, "The Cox ring of $\overline{M}_{0,6}$", *Trans. Amer. Math. Soc.* **361**:7 (2009), 3851–3878. MR 2009m:14037 Zbl 1172.14010

[Cox 1995] D. A. Cox, "The homogeneous coordinate ring of a toric variety", *J. Algebraic Geom.* **4**:1 (1995), 17–50. MR 95i:14046 Zbl 0846.14032

[Draisma 2008] J. Draisma, "A tropical approach to secant dimensions", *J. Pure Appl. Algebra* **212**:2 (2008), 349–363. MR 2008j:14102 Zbl 1126.14059

[Einsiedler et al. 2006] M. Einsiedler, M. Kapranov, and D. Lind, "Non-Archimedean amoebas and tropical varieties", *J. Reine Angew. Math.* **601** (2006), 139–157. MR 2007k:14038 Zbl 1115.14051

[Eisenbud 1995] D. Eisenbud, *Commutative algebra*, Graduate Texts in Mathematics **150**, Springer, New York, 1995. MR 97a:13001 Zbl 0819.13001

[Feichtner 2006] E. M. Feichtner, "Complexes of trees and nested set complexes", *Pacific J. Math.* **227**:2 (2006), 271–286. MR 2008a:05275 Zbl 1177.05031

[Feichtner and Yuzvinsky 2004] E. M. Feichtner and S. Yuzvinsky, "Chow rings of toric varieties defined by atomic lattices", *Invent. Math.* **155**:3 (2004), 515–536. MR 2004k:14009 Zbl 1083.14059

[Fulton 1993] W. Fulton, *Introduction to toric varieties*, Ann. Math. Studies **131**, Princeton University Press, 1993. MR 94g:14028 Zbl 0813.14039

[Gathmann 2006] A. Gathmann, "Tropical algebraic geometry", *Jahresber. Deutsch. Math.-Verein.* **108**:1 (2006), 3–32. MR 2007e:14088 Zbl 1109.14038

[Gel'fand et al. 1994] I. M. Gel'fand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, resultants, and multidimensional determinants*, Birkhäuser, Boston, 1994. MR 1264417 (95e:14045) Zbl 0827.14036

[Howard et al. 2009] B. Howard, J. Millson, A. Snowden, and R. Vakil, "The equations for the moduli space of $n$ points on the line", *Duke Math. J.* **146**:2 (2009), 175–226. MR 2009m:14070 Zbl 1161.14033

[Hu 2005] Y. Hu, "Topological aspects of Chow quotients", *J. Differential Geom.* **69**:3 (2005), 399–440. MR 2006f:14054 Zbl 1087.14032

[Hu and Keel 2000] Y. Hu and S. Keel, "Mori dream spaces and GIT", *Michigan Math. J.* **48** (2000), 331–348. MR 2001i:14059 Zbl 1077.14554

[Jensen et al. 2008] A. N. Jensen, H. Markwig, and T. Markwig, "An algorithm for lifting points in a tropical variety", *Collect. Math.* **59**:2 (2008), 129–165. MR 2009a:14077 Zbl 1151.13021

[Kapranov 1993] M. M. Kapranov, "Chow quotients of Grassmannians, I", pp. 29–110 in *I. M. Gel'fand Seminar*, Adv. Soviet Math. **16**, Amer. Math. Soc., Providence, RI, 1993. MR 95g:14053 Zbl 0811.14043

[Kapranov et al. 1992] M. M. Kapranov, B. Sturmfels, and A. V. Zelevinsky, "Chow polytopes and general resultants", *Duke Math. J.* **67**:1 (1992), 189–218. MR 93e:14062 Zbl 0780.14027

[Keel 1992] S. Keel, "Intersection theory of moduli space of stable $n$-pointed curves of genus zero", *Trans. Amer. Math. Soc.* **330**:2 (1992), 545–574. MR 92f:14003 Zbl 0768.14002

[Keel and Tevelev 2006] S. Keel and J. Tevelev, "Geometry of Chow quotients of Grassmannians", *Duke Math. J.* **134**:2 (2006), 259–311. MR 2007m:14017 Zbl 1107.14026

[Keel and Tevelev 2009] S. Keel and J. Tevelev, "Equations for $\overline{M}_{0,n}$", *Internat. J. Math.* **20**:9 (2009), 1159–1184. MR 2011a:14059 Zbl 1187.14031

[Kollár 1996] J. Kollár, *Rational curves on algebraic varieties*, Ergebnisse der Mathematik (3) **32**, Springer, Berlin, 1996. MR 98c:14001 Zbl 0877.14012

[Maclagan and Thomas 2007] D. Maclagan and R. R. Thomas, "Computational algebra and combinatorics of toric ideals", pp. 1–106 in *Commutative algebra and combinatorics* (Allahabad, 2003), vol. 1, edited by R. V. Gurjar et al., Ramanujan Math. Soc. Lect. Notes Ser. **4**, Ramanujan Math. Soc., Mysore, 2007. MR 2009c:13069

[Mikhalkin 2006] G. Mikhalkin, "Tropical geometry and its applications", pp. 827–852 in *International Congress of Mathematicians*, vol. II, edited by J. L. V. Marta Sanz-Solé, Javier Soria and J. Verdera, Eur. Math. Soc., Zürich, 2006. MR 2008c:14077 Zbl 1103.14034

[Mora and Robbiano 1988] T. Mora and L. Robbiano, "The Gröbner fan of an ideal", *J. Symbolic Comput.* **6**:2-3 (1988), 183–208. MR 90d:13004 Zbl 0668.13017

[Mustaţă 2002] M. Mustaţă, "Vanishing theorems on toric varieties", *Tohoku Math. J.* (2) **54**:3 (2002), 451–470. MR 2003e:14013 Zbl 1092.14064

[Payne 2009] S. Payne, "Fibers of tropicalization", *Math. Z.* **262** (2009), 301–311. MR 2010c:14081 Zbl 05565288

[Robinson and Whitehouse 1996] A. Robinson and S. Whitehouse, "The tree representation of $\Sigma_{n+1}$", *J. Pure Appl. Algebra* **111**:1-3 (1996), 245–253. MR 97g:55010 Zbl 0865.55010

[Speyer 2005] D. E. Speyer, *Tropical geometry*, Ph.D. thesis, University of California, Berkeley, 2005, Available at http://tinyurl.com/29947uz. MR 2707751

[Speyer and Sturmfels 2004] D. Speyer and B. Sturmfels, "The tropical Grassmannian", *Adv. Geom.* **4**:3 (2004), 389–411. MR 2005d:14089 Zbl 1065.14071

[Sturmfels 1996] B. Sturmfels, *Gröbner bases and convex polytopes*, University Lecture Series **8**, American Mathematical Society, Providence, RI, 1996. MR 97b:13034

[Tevelev 2007] J. Tevelev, "Compactifications of subvarieties of tori", *Amer. J. Math.* **129**:4 (2007), 1087–1104. MR 2008f:14068 Zbl 1154.14039

[Vermeire 2002] P. Vermeire, "A counterexample to Fulton's conjecture on $\overline{M}_{0,n}$", *J. Algebra* **248**:2 (2002), 780–784. MR 2002k:14043 Zbl 1039.14014

[Vogtmann 1990] K. Vogtmann, "Local structure of some Out$(F_n)$-complexes", *Proc. Edinburgh Math. Soc.* (2) **33**:3 (1990), 367–379. MR 92d:57002 Zbl 0694.20021

[Ziegler 1995] G. M. Ziegler, *Lectures on polytopes*, Graduate Texts in Mathematics **152**, Springer, New York, 1995. MR 96a:52011 Zbl 0823.52002

Communicated by Ravi Vakil
Received 2009-05-29 Revised 2010-02-17 Accepted 2010-05-05

agibney@math.uga.edu *Department of Mathematics, University of Georgia, Athens, GA 30602, United States*

D.Maclagan@warwick.ac.uk *Mathematics Institute, Zeeman Building, University of Warwick, Coventry, CV4 7AL, United Kingdom*

ALGEBRA AND NUMBER THEORY 4:7(2010)

# Haglund–Haiman–Loehr type formulas for Hall–Littlewood polynomials of type $B$ and $C$

Cristian Lenart

In previous work we showed that two apparently unrelated formulas for the Hall–Littlewood polynomials of type $A$ are, in fact, closely related. The first is the tableau formula obtained by specializing $q = 0$ in the Haglund–Haiman–Loehr formula for Macdonald polynomials. The second is the type $A$ instance of Schwer's formula (rephrased and rederived by Ram) for Hall–Littlewood polynomials of arbitrary finite type; Schwer's formula is in terms of so-called alcove walks, which originate in the work of Gaussent and Littelmann and of the author with Postnikov on discrete counterparts to the Littelmann path model. We showed that the tableau formula follows by "compressing" Ram's version of Schwer's formula. In this paper, we derive new tableau formulas for the Hall–Littlewood polynomials of type $B$ and $C$ by compressing the corresponding instances of Schwer's formula.

## 1. Introduction

Hall–Littlewood polynomials are at the center of many recent developments in representation theory and algebraic combinatorics. They were originally defined in type $A$, as a basis for the algebra of symmetric functions depending on a parameter $t$. This basis interpolates between two fundamental bases: the one of Schur functions, at $t = 0$, and the one of monomial functions, at $t = 1$. Besides the original motivation for defining Hall–Littlewood polynomials, which comes from the Hall algebra [Littlewood 1961], there are many other applications; see for example [Lenart 2011] and the references therein.

Macdonald [1971] showed that there is a formula for the spherical functions corresponding to a Chevalley group over a $p$-adic field that generalizes the formula for

Cristian Lenart was partially supported by the National Science Foundation grant DMS-0701044.
*MSC2000:* primary 05E05; secondary 33D52.
*Keywords:* Hall–Littlewood polynomials, Macdonald polynomials, alcove walks, Schwer's formula, the Haglund–Haiman–Loehr formula.

the Hall–Littlewood polynomials. Thus, the Macdonald spherical functions generalize the Hall–Littlewood polynomials to all root systems, and the two names are used interchangeably in the literature. There are two families of Hall–Littlewood functions of arbitrary type, called $P$ and $Q$, which form dual bases for the Weyl group invariants. The $P$-polynomials specialize to the Weyl characters at $t=0$. The transition matrix between Weyl characters and $P$-polynomials is given by Lusztig's $t$-analog of weight multiplicities (Kostka–Foulkes polynomials of arbitrary type), which are certain affine Kazhdan–Lusztig polynomials [Kato 1982; Lusztig 1983]. On the combinatorial side, we have the Lascoux–Schützenberger formula [1979] for the Kostka–Foulkes polynomials in type $A$, but no generalization of this formula to other types is known. Other applications of the type $A$ Hall–Littlewood polynomials that extend to arbitrary type are those related to fermionic multiplicity formulas [Ardonne and Kedem 2007] and affine crystals [Lecouvey and Shimozono 2007]. We refer to [Nelsen and Ram 2003; Stembridge 2005] for surveys on Hall–Littlewood polynomials of arbitrary type.

Macdonald [1992; 2000] defined a remarkable family of orthogonal polynomials depending on parameters $q, t$, which bear his name. These polynomials generalize the spherical functions for a $p$-adic group, the Jack polynomials, and the zonal polynomials. At $q = 0$, the Macdonald polynomials specialize to the Hall–Littlewood polynomials, and thus they further specialize to the Weyl characters (upon setting $t = 0$ as well). There has been considerable interest recently in the combinatorics of Macdonald polynomials. This stems in part from a combinatorial formula for the ones corresponding to type $A$, which is due to Haglund, Haiman, and Loehr [Haglund et al. 2005]. This formula is in terms of fillings of Young diagrams, and uses two statistics, called inv and maj, on such fillings. The Haglund–Haiman–Loehr formula has already found important applications, such as new proofs of the positivity theorem for Macdonald polynomials, which states that the two-parameter Kostka–Foulkes polynomials have nonnegative integer coefficients. One of these proofs is based on Hecke algebras [Grojnowski and Haiman 2007], while the other is purely combinatorial and leads to a positive formula for the two-parameter Kostka–Foulkes polynomials [Assaf 2010]. Moreover, in the one-parameter case (that is, when $q = 0$), the Haglund–Haiman–Loehr formula was used to give a concise derivation of the Lascoux–Schützenberger formula for the Kostka–Foulkes polynomials of type $A$ [Haglund et al. 2005, Section 7].

An apparently unrelated development, at the level of arbitrary finite root systems, led to Schwer's formula [2006], rephrased and rederived by Ram [2006], for the Hall–Littlewood polynomials of arbitrary type. The latter formulas are in terms of so-called alcove walks, which originate in the work of Gaussent and Littelmann [2005] and of the author with Postnikov [Lenart and Postnikov 2007; 2008] on discrete counterparts to the Littelmann path model [Littelmann 1994; 1995]. Schwer's formula was recently generalized by Ram and Yip [2011] to a similar formula for the Macdonald polynomials. The generalization consists in the fact that the latter formula is in terms of alcove walks with both "positive" and "negative" foldings, whereas in the former only "positive" foldings appear.

In [Lenart 2011], we related Schwer's formula to the Haglund–Haiman–Loehr formula. More precisely, we showed that we can group the terms in the type $A$ instance of Schwer's formula (in fact, we used Ram's version of it) for $P_\lambda(x;t)$ into equivalence classes, such that the sum in each equivalence class is a term in the Haglund–Haiman–Loehr formula for $q=0$. An equivalence class consists of all the terms corresponding to alcove walks that produce the same filling of a Young diagram $\lambda$ (indexing the Hall–Littlewood polynomial) via a simple construction. In fact, we first considered the case when the partition $\lambda$ has no two parts identical (that is, it is a regular weight); the general case, which displays additional complexity, was considered in the Appendix to the same paper, written with Lubovsky. The work referring to a regular weight $\lambda$ was then extended in [Lenart 2009], by showing that the type $A$ instance of the Ram–Yip formula for Macdonald polynomials compresses, in a similar way, to a formula analogous to the Haglund–Haiman–Loehr one, but with fewer terms.

In this paper we extend the results in [Lenart 2011] to types $B$ and $C$. More precisely, we derive new formulas for the Hall–Littlewood polynomials of type $B$ and $C$ indexed by regular weights in terms of fillings of Young diagrams; we do this by compressing the corresponding instances of Schwer's formula (in fact, we again use Ram's version of it). Note that no tableau formula for the Hall–Littlewood or Macdonald polynomials exists beyond type $A$ so far. Our approach provides a natural way to obtain such formulas, and suggests that this method could be further extended to type $D$ (this case is slightly more complex than types $B$ and $C$, as seen below), as well as to Macdonald polynomials; these problems are currently explored, as is the compression in the case of a Hall–Littlewood polynomial indexed by a nonregular weight (by extending the type $A$ result in the Appendix of [Lenart 2011]). Our formula is more complex than the corresponding one in type $A$ (that is, the Haglund–Haiman–Loehr formula at $q=0$). However, the statistic we use is, in the case of some special fillings, completely similar to the Haglund–Haiman–Loehr inversion statistic (which is the more intricate of their two statistics). The naturality of our formula is also supported by the fact that the Kashiwara–Nakashima tableaux [1994] of type $B$ and $C$ are, essentially, the surviving fillings in this formula when we set $t=0$. We also note that the passage from (Ram's version of) Schwer's formula to ours results in a considerably larger reduction in the number of terms in type $B$ and $C$ compared to type $A$. In terms of applications, it would be very interesting to see whether our formula could be used to derive, in the spirit of [Haglund et al. 2005, Section 7], a positive combinatorial formula for Lusztig's $t$-analog of weight multiplicities in type $B$ and $C$, which has been long sought.

## 2. The tableau formula in type $C$

Let us start by recalling the Weyl group of type $B/C$, viewed as the group of signed permutations $B_n$. Such permutations are bijections $w$ from

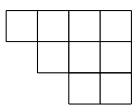
$$[\bar{n}] := \{1 < 2 < \cdots < n < \bar{n} < \overline{n-1} < \cdots < \bar{1}\}$$

to $[\bar{n}]$ satisfying $w(\bar{\imath}) = \overline{w(i)}$. Here $\bar{\imath}$ is viewed as $-i$, so $\bar{\bar{\imath}} = i$. We use the window notation $w = w(1)\dots w(n)$. Given $1 \le i < j \le n$, we denote by $(i, j)$ the reflection that transposes the entries in positions $i$ and $j$ (upon right multiplication). Similarly, we denote by $(i, \bar{\jmath})$, again for $i < j$, the transposition of entries in positions $i$ and $j$ followed by the sign change of those entries. Finally, we denote by $(i, \bar{\imath})$ the sign change in position $i$. Given $w$ in $B_n$, we define

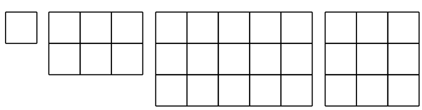
$$\begin{aligned} \ell_+(w) &:= \left|\{(k,l) : 1 \le k < l \le n,\ w(k) > w(l)\}\right|, \\ \ell_-(w) &:= \left|\{(k,l) : 1 \le k \le l \le n,\ w(k) > \overline{w(l)}\}\right|. \end{aligned} \tag{2-1}$$

Then the length of $w$ is given by $\ell(w) := \ell_+(w) + \ell_-(w)$.

Let $\lambda$ be a partition corresponding to a regular weight in type $C_n$ for $n \ge 2$, that is, $\lambda = (\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0)$ with $\lambda_i \in \mathbb{Z}$. We identify $\lambda$ with its Young (or Ferrers) diagram, as usual, but we draw this diagram in "Japanese style" (as opposed to the more common English or French styles), that is, we embed it in the third quadrant, where $n = 3$:

$\lambda = (4, 3, 2) =$  .

Consider the shape $\hat{\lambda}$ obtained from $\lambda$ by replacing each column of height $k$ with $k$ or $2k-1$ (adjacent) copies of it, depending on the given column being the rightmost one or not. In this example, we have

$\hat{\lambda} = (12, 11, 8) =$  .

Here $\hat{\lambda}$ is shown divided into rectangular blocks, each of which corresponds to a column of $\lambda$; the heights of the blocks (from right to left) are given by the conjugate partition $\lambda' = (3, 3, 2, 1)$.

We are representing a filling $\sigma$ of $\hat{\lambda}$ as a concatenation of columns $C_{ij}$ and $C'_{ik}$, where $i = 1, \dots, \lambda_1$, while for a given $i$ we have $j = 1, \dots, \lambda'_i$ if $i > 1$, $j = 1$ if

$i=1$, and $k=2,\dots,\lambda_i'$; the columns $C_{ij}$ and $C_{ik}'$ have height $\lambda_i'$. More precisely, we let

$$\sigma=\mathscr{C}^{\lambda_1}\dots\mathscr{C}^1, \tag{2-2}$$

where

$$\mathscr{C}^i := \begin{cases} C_{i2}'\dots C_{i,\lambda_i'}'C_{i1}\dots C_{i\lambda_i'} & \text{if } i>1,\\ C_{i2}'\dots C_{i,\lambda_i'}'C_{i1} & \text{if } i=1.\end{cases}$$

Note that the leftmost column is $C_{\lambda_1,1}$ and the rightmost column is $C_{11}$.

**Example 2.1.** The following is a filling for the partition considered above, where we use the same division into blocks as above:

$$\begin{array}{c|c|ccc|ccccc|ccc}
 & \mathscr{C}^4 & & \mathscr{C}^3 & & & & \mathscr{C}^2 & & & & \mathscr{C}^1 & \\
 & C_{41} & C_{32}' & C_{31} & C_{32} & C_{22}' & C_{23}' & C_{21} & C_{22} & C_{23} & C_{12}' & C_{13}' & C_{11} \\
\hline
\sigma = & \bar{1} & \bar{1} & \bar{1} & 3 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\
 & & \bar{2} & \bar{2} & \bar{2} & 3 & 3 & 2 & 2 & 2 & 2 & 2 & 2 \\
 & & & & & \bar{1} & \bar{1} & \bar{3} & \bar{3} & \bar{3} & 3 & 3 & 3
\end{array}.$$

Essentially, the description (2-2) of a filling of $\hat{\lambda}$ says that the column to the right of $C_{ij}$ is $C_{i,j+1}$, whereas the column to the right of $C_{ik}'$ is $C_{i,k+1}'$. Here we are assuming that the mentioned columns exist, up to the conventions

$$C_{i,\lambda_i'+1} = \begin{cases} C_{i-1,2}' & \text{if } i>1 \text{ and } \lambda_{i-1}'>1,\\ C_{i-1,1} & \text{if } i>1 \text{ and } \lambda_{i-1}'=1,\end{cases} \qquad C_{i,\lambda_i'+1}' = C_{i1}. \tag{2-3}$$

The entry in position $i$, counted from the top, in some column $C$ is denoted by $C(i)$. We also write $C[i,j]$ for the portion of column $C$ consisting of the entries in positions $i,i+1,\dots,j$; this is empty if $i>j$.

We consider the set $\mathscr{F}(\lambda)$ of fillings of $\hat{\lambda}$ with entries in $[\bar{n}]$ that satisfy the following conditions:

(1) The rows are weakly decreasing from left to right.

(2) No column contains two entries $a,b$ with $a=\pm b$.

(3) Each column after the first is related to its left neighbor as indicated in the next paragraph. (Essentially, consecutive columns differ by a *signed cycle*, that is, a composition $(r_1,\bar{j})\dots(r_p,\bar{j})$, where $1\le r_1<\dots<r_p<j$; furthermore, $j$ varies from 1 to the length of the column in question, as we consider the columns from left to right.)

Here we let the reflections in $B_n$ act on columns $C$ like they do on signed permutations; for instance, $C(a,\bar{b})$ is the column obtained from $C$ by transposing the entries in positions $a,b$ and by changing their signs. Let us first explain the passage from some column $C_{ij}$ to $C_{i,j+1}$. There exist positions $1\le r_1<\dots<r_p<j$ (possibly $p=0$) such that $C_{i,j+1}$ differs from $D=C_{ij}(r_1,\bar{j})\dots(r_p,\bar{j})$ only in

position $j$, while $C_{i,j+1}(j) \le D(j)$. To include the case $j = \lambda_i'$ in this description, just replace $C_{i,j+1}$ everywhere by $C_{i,j+1}[1, \lambda_i']$ and use the conventions (2-3). Let us now explain the passage from some column $C_{ik}'$ to $C_{i,k+1}'$. There exist positions $1 \le r_1 < \cdots < r_p < k$ (possibly $p = 0$) such that $C_{i,k+1}' = C_{ik}'(r_1, \overline{k}) \ldots (r_p, \overline{k})$. This description includes the case $k = \lambda_i'$, based on the conventions (2-3).

Note that the filling $\sigma$ in Example 2.1 satisfies the above conditions. Indeed, conditions (1) and (2) are clearly verified. Then compare, for instance,

$$C_{33}[1,2] = C_{22}'[1,2] = \begin{array}{|c|}\hline 2 \\ \hline 3 \\ \hline\end{array} \quad \text{and} \quad D = C_{32}(1,\overline{2}) = \begin{array}{|c|}\hline 3 \\ \hline 2 \\ \hline\end{array}(1,\overline{2}) = \begin{array}{|c|}\hline 2 \\ \hline 3 \\ \hline\end{array};$$

they only differ in position 2, while $C_{22}'(2) = 3 < D(2) = \overline{3}$. Similarly, we have

$$C_{24}' = C_{21} = C_{23}'(1,\overline{3})(2,\overline{3}).$$

Also note that, while the rows are weakly decreasing (from left to right), the columns need not be always increasing or always decreasing (compare $C_{32}' = C_{31}$ with the other columns).

Let us now define the content of a filling. For this purpose, we first associate with a filling $\sigma$ a *compressed* version of it, namely the filling $\overline{\sigma}$ of the partition $2\lambda$. This is defined as follows:

$$\overline{\sigma} = \overline{\mathscr{C}}^{\lambda_1} \ldots \overline{\mathscr{C}}^{1}, \quad \text{where} \quad \overline{\mathscr{C}}^{i} := C_{i2}' C_{i1}, \tag{2-4}$$

where the conventions (2-3) are used again. Now define $\operatorname{ct}\sigma = (c_1, \ldots, c_n)$, where $c_i$ is half the difference between the number of occurrences of the entries $i$ and $\bar{\imath}$ in $\overline{\sigma}$. Sometimes, this vector is written in terms of the coordinate vectors $\varepsilon_i$:

$$\operatorname{ct}\sigma = c_1\varepsilon_1 + \cdots + c_n\varepsilon_n = \tfrac{1}{2}\sum_{b\in\overline{\sigma}} \varepsilon_{\overline{\sigma}(b)}; \tag{2-5}$$

here the last sum is over all boxes $b$ of $\overline{\sigma}$, and we set $\varepsilon_{\bar{\imath}} := -\varepsilon_i$. In our running example, we have

$$\overline{\sigma} = \begin{array}{|c|c|c|c|c|c|c|c|}\hline \overline{1} & \overline{1} & \overline{1} & \overline{1} & 2 & 1 & 1 & 1 \\ \hline \multicolumn{1}{c}{} & \multicolumn{1}{c|}{} & \overline{2} & \overline{2} & 3 & 2 & 2 & 2 \\ \cline{3-8} \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c}{} & \multicolumn{1}{c|}{} & \overline{1} & \overline{3} & 3 & 3 \\ \cline{5-8}\end{array},$$

so $\operatorname{ct}\sigma = (-1, 1, 1)$.

We now define two statistics on fillings in $\mathscr{F}(\lambda)$ that will be used in our compressed formula for Hall–Littlewood polynomials. Intervals refer to the totally ordered set $[\overline{n}]$. Let

$$\sigma_{ab} := \begin{cases} 1 & \text{if } a, b \ge \overline{n}, \\ 0 & \text{otherwise.} \end{cases} \tag{2-6}$$

Given a word $w$, we use the notation $N_{ab}(w)$ for the number of entries in $w$ contained in the interval $(a, b)$.

Given two columns $D$, $C$ of the same height $d$ such that $D \geq C$ componentwise, we will define two statistics $N(D, C)$ and $\operatorname{des}(D, C)$ in some special cases, as specified below.

**Case 0.** If $D = C$, then $N(D, C) := 0$ and $\operatorname{des}(D, C) := 0$.

**Case 1.** Assume that $C = D(r, \bar{\jmath})$ with $r < j$. Let $a := D(r)$ and $b := D(j)$. In this case, we set

$$N(D, C) := N_{\bar{b}a}(D[r+1, j-1]) + \left|(\bar{b}, a) \setminus \{\pm D(i) \,:\, i = 1, \ldots, j\}\right| + \sigma_{ab},$$

and $\operatorname{des}(D, C) := 1$.

**Case 2.** Assume that $C = D(r_1, \bar{\jmath}) \ldots (r_p, \bar{\jmath})$, where $1 \leq r_1 < \cdots < r_p < j$. Let $D_i := D(r_1, \bar{\jmath}) \ldots (r_i, \bar{\jmath})$ for $i = 0, \ldots, p$, so that $D_0 = D$ and $D_p = C$. We define

$$N(D, C) := \sum_{i=1}^{p} N(D_{i-1}, D_i), \quad \operatorname{des}(D, C) := p.$$

For instance, in the example above, we have

$$N(C'_{23}C_{21}) = N\left(\begin{array}{|c|c|}\hline 2 & 1 \\ \hline 3 & 2 \\ \hline \bar{1} & \bar{3} \\ \hline\end{array}\right) = N\left(\begin{array}{|c|c|}\hline 2 & 1 \\ \hline 3 & 3 \\ \hline \bar{1} & \bar{2} \\ \hline\end{array}\right) + N\left(\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 3 & 2 \\ \hline \bar{2} & \bar{3} \\ \hline\end{array}\right) = N_{12}\left(\boxed{3}\right) + N_{23}(\varnothing) = 0,$$

and $\operatorname{des}(C'_{23}C_{21}) = 2$.

**Case 3.** Assume that $C$ differs from $D' := D(r_1, \bar{\jmath}) \ldots (r_p, \bar{\jmath})$ with $1 \leq r_1 < \cdots < r_p < j$ (possibly $p = 0$) only in position $j$, while $C(j) < D'(j)$. We define

$$N(D, C) := N(D, D') + N_{C(j), D'(j)}(D[j+1, d]), \quad \operatorname{des}(D, C) := p + 1.$$

For instance, in our running example, we have

$$N(C_{31}C_{32}) = N\left(\begin{array}{|c|c|}\hline \bar{1} & 3 \\ \hline \bar{2} & \bar{2} \\ \hline\end{array}\right) = N_{3\bar{1}}\left(\boxed{\bar{2}}\right) = 1,$$

and $\operatorname{des}(C_{31}C_{32}) = 1$.

If the height of $C$ is larger than the height $d$ of $D$ (necessarily by 1), and $N(D, C[1, d])$ can be computed as above, we let $N(D, C) := N(D, C[1, d])$ and $\operatorname{des}(D, C) := \operatorname{des}(D, C[1, d])$. For instance, we have

$$N(C_{32}C'_{22}) = N\left(\begin{array}{|c|c|}\hline 3 & 2 \\ \hline \bar{2} & 3 \\ \hline\end{array}\right) = N\left(\begin{array}{|c|c|}\hline 3 & 2 \\ \hline \bar{2} & \bar{3} \\ \hline\end{array}\right) + N_{3\bar{3}}(\varnothing) = N_{23}(\varnothing) = 0,$$

and $\operatorname{des}(C_{32}C'_{22}) = 2$.

Given a filling $\sigma$ in $\mathcal{F}(\lambda)$ with columns $C_m, \dots, C_1$, we set

$$N(\sigma) := \sum_{i=1}^{m-1} N(C_{i+1}, C_i) + \ell_+(C_1);$$

here $\ell_+(C_1)$ is defined as in (2-1). We also set

$$\operatorname{des}\sigma := \sum_{i=1}^{m-1} \operatorname{des}(C_{i+1}, C_i).$$

Note that $\operatorname{des}\sigma$ essentially counts the descents in the rows of $\sigma$. In our running example, we have $N(\sigma) = 1$ and $\operatorname{des}\sigma = 6$.

We can now state our new formula for the Hall–Littlewood polynomials of type $C$, which follows as a corollary of our main result, Theorem 4.6. A completely similar formula in type $B$ is discussed in Section 5. We refer to Proposition 2.4 and Remarks 4.7 for more insight into our formula. In particular, the Kashiwara–Nakashima tableaux of type $C$ are, essentially, the surviving fillings in this formula when we set $t = 0$. Furthermore, in some special cases, the statistic $N(\sigma)$ is completely similar to the Haglund–Haiman–Loehr inversion statistic (the more intricate of their two statistics); more precisely, this happens when the related chains in Bruhat order contain no reflections of type $B$, that is $(i, \bar{j})$, where $i$ and $j$ are less than the height of the corresponding column of the filling (see Proposition 2.4).

**Theorem 2.2.** *Given a regular weight $\lambda$, we have*

$$P_\lambda(X; t) = \sum_{\sigma \in \mathcal{F}(\lambda)} t^{N(\sigma)} (1-t)^{\operatorname{des}\sigma} x^{\operatorname{ct}\sigma}, \tag{2-7}$$

*where $x^{(c_1,\dots,c_n)} := x_1^{c_1} \dots x_n^{c_n}$.*

**Example 2.3.** Consider the simplest case, namely $n = 2$ and $\lambda = (2, 1)$. This leads to considering fillings of the shape $(3, 2)$ with elements in $[\bar{2}]$, namely

$$\begin{array}{|c|c|c|} \hline e & c & a \\ \hline \end{array}$$
$$\begin{array}{c|c|c|} & d & b \\ \hline \end{array}.$$

The fillings need to satisfy the following conditions:

- $a \le c \le e$, $b \le d$.
- $a \ne \pm b$.
- either $c = a$ and $d = b$, or $c = \bar{b}$ and $d = \bar{a}$.

For $i \in \{1, 2\}$, let $n_i$ be half the difference between the number of $i$'s and $\bar{i}$'s in the multiset $\{a, b, c, d, e, e\}$. Given a proposition $A$, we let $\chi(A)$ be 1 or 0, depending on the logical value of $A$ being true or false. Then

$$P_{(2,1)}(x_1, x_2; t) = \sum_{(a,b,c,d,e)} t^{\chi(a>b)+\chi(a,b\leq 2,\, a\neq c)}(1-t)^{\chi(a\neq c)+\chi(c\neq e)} x_1^{n_1} x_2^{n_2}.$$

It turns out that there are 27 terms in this sum, versus 70 terms in (Ram's version of) Schwer's formula. For instance, the terms contributing to the coefficient of $x_2$ correspond to the fillings

$$\begin{array}{|c|c|c|}\hline \bar{1} & 1 & 1 \\ \hline \multicolumn{1}{c|}{} & 2 & 2 \\ \cline{2-3}\end{array}\,, \qquad \begin{array}{|c|c|c|}\hline \bar{1} & 2 & 2 \\ \hline \multicolumn{1}{c|}{} & 1 & 1 \\ \cline{2-3}\end{array}\,, \qquad \begin{array}{|c|c|c|}\hline 2 & 2 & 1 \\ \hline \multicolumn{1}{c|}{} & \bar{1} & \bar{2} \\ \cline{2-3}\end{array}\,;$$

the associated polynomials in $t$ are

$$1-t, \qquad t(1-t), \qquad 1-t,$$

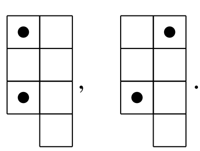
respectively. Note that these polynomials are obtained by compressing 3, 2, and 2 terms in Schwer's formula, respectively. By symmetry, the coefficients of $x_1$, $x_2$, $x_1^{-1}$, and $x_2^{-1}$ in $P_{(2,1)}(x_1, x_2; t)$ are all $(t+2)(1-t)$. Other fillings have an even larger number of terms in Schwer's formula corresponding to them, such as

$$\begin{array}{|c|c|c|}\hline \bar{1} & \bar{2} & \bar{2} \\ \hline \multicolumn{1}{c|}{} & \bar{1} & \bar{1} \\ \cline{2-3}\end{array}\,,$$

which has 7; in other words, the associated polynomial in $t$, namely $1-t$, which contributes to the coefficient of $x_1^{-2}x_2^{-1}$, is the sum of 7 polynomials of the form $t^r(1-t)^s$ in Schwer's formula. In conclusion, we have

$$\begin{aligned} P_{(2,1)}(x_1, x_2; t) = x_1^2x_2 + x_1x_2^2 + x_1^2x_2^{-1} + x_1x_2^{-2} + x_1^{-1}x_2^2 + x_1^{-2}x_2 \\ + x_1^{-1}x_2^{-2} + x_1^{-2}x_2^{-1} + (t+2)(1-t)(x_1 + x_2 + x_1^{-1} + x_2^{-1}). \end{aligned}$$

In order to relate our statistic $N(\sigma)$ to the Haglund–Haiman–Loehr *inversion statistic* and to compare our formula to its type $A$ counterpart (see [Haglund et al. 2005, Proposition 8.1] or [Lenart 2011, Theorem 2.10]), let us recall some definitions from [Haglund et al. 2005; Lenart 2011]. We start by considering fillings $\tau$ of the shape $\lambda$ with entries in $[\bar{n}]$, which are again displayed in Japanese style, as a sequence of columns $\tau = C_{\lambda_1} \dots C_1$; here $C_i$ is a sequence $(C_i(1), \dots, C_i(\lambda_i'))$, so the entry in cell $u = (i, j)$ is $\tau(u) = C_j(i)$. Two cells $u, v \in \lambda$ are said to *attack* each other if they are in one of the following two relative positions:



An *inversion* of $\tau$ is a pair of attacking cells $(u, v)$ that have one of the following two relative positions, where $a := \tau(u) < b := \tau(v)$:

$$\begin{array}{|c|c|}\hline a & \\ \hline & \\ \hline b & \\ \hline \multicolumn{1}{c|}{} & \\ \cline{2-2}\end{array}\,, \qquad \begin{array}{|c|c|}\hline & b \\ \hline & \\ \hline a & \\ \hline \multicolumn{1}{c|}{} & \\ \cline{2-2}\end{array}\,.$$

The Haglund–Haiman–Loehr statistic $\operatorname{inv}\tau$ is defined as the number of inversions of $\tau$. The *descent statistic*, denoted $\operatorname{des}\tau$ (which is similar to des for fillings of $\hat{\lambda}$ defined above, as seen below), is the number of cells $u = (i, j)$ with $j \neq 1$ and $\tau(u) > \tau(v)$, where $v = (i, j-1)$. As usual, let

$$n(\lambda) := \sum_i (i-1)\lambda_i ,$$

and assume that $\tau$ has the following two properties: (i) $\tau(u) \neq \tau(v)$ whenever $u$ and $v$ attack each other; and (ii) $\tau$ is weakly decreasing in rows. Then it was shown in [Lenart 2011, Proposition 2.12] that the so-called *complementary inversion statistic* $\operatorname{cinv}\tau := n(\lambda) - \operatorname{inv}\tau$ counts the triples of cells filled with $a < b < c$ that have the following relative position (here the cell supposed to contain $c$ might be outside the shape $\lambda$, in which case we only require $a < b$):

$$\begin{array}{|c|c|}\hline & b \\ \hline & \\ \hline c & a \\ \hline \multicolumn{1}{c|}{} & \\ \cline{2-2}\end{array}\,.$$

**Proposition 2.4.** *Let $\sigma$ in $\mathcal{F}(\lambda)$ be a filling satisfying the properties that $C'_{i,j+1} = C'_{i,j}$ for all $i$ and $j = 2, \ldots, \lambda'_i$; and that $C_{i,j+1}$ differs from $C_{ij}$ at most in position $j$, for all $i$ and $j = 1, \ldots, \lambda'_i$. Let $\tilde{\sigma}$ be the filling of $\lambda$ given by*

$$\tilde{\sigma} := C_{\lambda_1,1} C_{\lambda_1 - 1,1} \ldots C_{11}.$$

*Then $N(\sigma) = \operatorname{cinv}\tilde{\sigma}$ and $\operatorname{des}\sigma = \operatorname{des}\tilde{\sigma}$.*

Before presenting the proof, let us exhibit an example.

**Example 2.5.** For the partition $\lambda = (4, 3, 2)$ considered above, a filling satisfying the conditions in Proposition 2.4 is

$$\begin{array}{ccccc}
 & \mathscr{C}^4 & \mathscr{C}^3 & \mathscr{C}^2 & \mathscr{C}^1 \\
 & \begin{array}{c} C_{41} \end{array} & \begin{array}{ccc} C'_{32} & C_{31} & C_{32} \end{array} & \begin{array}{ccccc} C'_{22} & C'_{23} & C_{21} & C_{22} & C_{23} \end{array} & \begin{array}{ccc} C'_{12} & C'_{13} & C_{11} \end{array} \\
\sigma = & \begin{array}{|c|}\hline \overline{1} \\ \hline \multicolumn{1}{c}{} \\ \multicolumn{1}{c}{} \end{array} & \begin{array}{|c|c|c|}\hline \overline{2} & \overline{2} & 1 \\ \hline 3 & 3 & 3 \\ \hline \multicolumn{3}{c}{} \end{array} & \begin{array}{|c|c|c|c|c|}\hline 1 & 1 & 1 & 1 & 1 \\ \hline 2 & 2 & 2 & 2 & 2 \\ \hline \overline{3} & \overline{3} & \overline{3} & \overline{3} & \overline{3} \\ \hline \end{array} & \begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline 2 & 2 & 2 \\ \hline 3 & 3 & 3 \\ \hline \end{array}\,.
\end{array}$$

We have

$$\tilde{\sigma}=\begin{array}{|c|c|c|c|}\hline \bar{1}&\bar{2}&1&1\\ \hline \multicolumn{1}{c|}{}&3&2&2\\ \cline{2-4} \multicolumn{2}{c|}{}&\bar{3}&3\\ \cline{3-4}\end{array}\,.$$

It is easy to check that $N(\sigma)=\operatorname{cinv}\tilde{\sigma}=1$ and $\operatorname{des}\sigma=\operatorname{des}\tilde{\sigma}=4$. Indeed, the only triple of cells in $\tilde{\sigma}$ contributing to statistic cinv, which in this case is missing one cell, is formed by the cells in the unique column of height 2.

*Proof of Proposition 2.4.* The equality $\operatorname{des}\sigma=\operatorname{des}\tilde{\sigma}$ is clear, so we concentrate on the other equality. Let $m:=\lambda_1$ be the number of columns of $\lambda$, and let $C_m=C_{m1},\dots,C_1=C_{11}$ be the columns of $\tilde{\sigma}$, of lengths $c_m:=\lambda'_m,\dots,c_1:=\lambda'_1$; let $C'_k:=C_k[1,c_{k+1}]$, for $k=1,\dots,m-1$. We refer to a pair $(i,j)$ with $1\le i<j\le c_k$ and $C_k(i)>C_k(j)$ as a (type $A$) inversion in $C_k$. It is easy to see that $\tilde{\sigma}$ satisfies the properties considered above: (i) $\tilde{\sigma}(u)\neq\tilde{\sigma}(v)$ whenever $u$ and $v$ attack each other; (ii) $\tilde{\sigma}$ is weakly decreasing in rows. We start by evaluating $N(\mathscr{C}^k C_{k-1,1})$, with $\mathscr{C}^k$ as in (2-2). By definition, $N(\mathscr{C}^k C_{k-1,1})=\sum_{i=1}^{c_k-1}N_{C_{k-1}(i),C_k(i)}(C_k[i+1,c_k])$. This is the number of inversions $(i,j)$ in $C_k$ for which $C_{k-1}(i)<C_k(j)$. If $(i,j)$ is an inversion in $C_k$ not satisfying the previous condition, then $C_{k-1}(i)>C_k(j)$ (by property (i) of $\tilde{\sigma}$), and thus $(i,j)$ is an inversion in $C'_{k-1}$ (by property (ii) of $\tilde{\sigma}$). Moreover, the only inversions of $C'_{k-1}$ that do not arise in this way are those counted by the statistic $\operatorname{cinv}(C_kC'_{k-1})$, so

$$N(\mathscr{C}^k C_{k-1,1})=\ell_+(C_k)-(\ell_+(C'_{k-1})-\operatorname{cinv}(C_kC'_{k-1})).$$

We conclude that

$$N(\sigma)-\ell_+(C_1)=\sum_{k=2}^{m}\ell_+(C_k)-\ell_+(C'_{k-1})+\operatorname{cinv}(C_kC'_{k-1}).$$

Now recall that $\lambda$ has no two parts identical. We clearly have $c_m=1$, so $\ell_+(C_m)=0$. Therefore,

$$\begin{aligned}N(\sigma)=\sum_{k=2}^{m}\ell_+(C_{k-1})-\ell_+(C'_{k-1})+\operatorname{cinv}(C_kC'_{k-1})&=\sum_{k=2}^{m}\operatorname{cinv}(C_kC_{k-1})\\&=\operatorname{cinv}\tilde{\sigma}.\end{aligned}$$ □

## 3. Background on Ram's version of Schwer's formula

We recall some background information on finite root systems and affine Weyl groups.

**3.1. *Root systems.*** Let $\mathfrak{g}$ be a complex semisimple Lie algebra, and $\mathfrak{h}$ a Cartan subalgebra, whose rank is $r$. Let $\Phi\subset\mathfrak{h}^*$ be the corresponding irreducible *root system*, $\mathfrak{h}^*_{\mathbb{R}}\subset\mathfrak{h}^*$ the real span of the roots, and $\Phi^+\subset\Phi$ the set of positive roots. Let $\alpha_1,\dots,\alpha_r\in\Phi^+$ be the corresponding *simple roots*. We denote by $\langle\cdot,\cdot\rangle$ the nondegenerate scalar product on $\mathfrak{h}_{\mathbb{R}}^*$ induced by the Killing form. Given a root $\alpha$, we consider the corresponding *coroot* $\alpha^\vee := 2\alpha/\langle\alpha,\alpha\rangle$ and reflection $s_\alpha$.

Let $W$ be the corresponding *Weyl group*, whose Coxeter generators are denoted, as usual, by $s_i := s_{\alpha_i}$. The length function on $W$ is denoted by $\ell(\cdot)$. The *Bruhat graph* on $W$ is the directed graph with edges $u \to w$, where $w = us_\beta$ for some $\beta \in \Phi^+$, and $\ell(w) > \ell(u)$; we usually label such an edge by $\beta$ and write $u \xrightarrow{\beta} w$. The *reverse Bruhat graph* is obtained by reversing the directed edges above. The *Bruhat order* on $W$ is the transitive closure of the relation corresponding to the Bruhat graph.

The *weight lattice* $\Lambda$ is given by

$$\Lambda := \{\lambda \in \mathfrak{h}_{\mathbb{R}}^* : \langle\lambda,\alpha^\vee\rangle \in \mathbb{Z} \text{ for any } \alpha \in \Phi\}. \tag{3-1}$$

The weight lattice $\Lambda$ is generated by the *fundamental weights* $\omega_1,\dots,\omega_r$, which form the dual basis to the basis of simple coroots, that is, $\langle\omega_i,\alpha_j^\vee\rangle = \delta_{ij}$. The set $\Lambda^+$ of *dominant weights* is given by

$$\Lambda^+ := \{\lambda \in \Lambda : \langle\lambda,\alpha^\vee\rangle \geq 0 \text{ for any } \alpha \in \Phi^+\}.$$

The subgroup of $W$ stabilizing a weight $\lambda$ is denoted by $W_\lambda$, and the set of minimum coset representatives in $W/W_\lambda$ by $W^\lambda$. Let $\mathbb{Z}[\Lambda]$ be the group algebra of the weight lattice $\Lambda$, which has a $\mathbb{Z}$-basis of formal exponents $\{x^\lambda : \lambda \in \Lambda\}$ with multiplication $x^\lambda \cdot x^\mu := x^{\lambda+\mu}$.

Given $\alpha \in \Phi$ and $k \in \mathbb{Z}$, we denote by $s_{\alpha,k}$ the reflection in the affine hyperplane

$$H_{\alpha,k} := \{\lambda \in \mathfrak{h}_{\mathbb{R}}^* : \langle\lambda,\alpha^\vee\rangle = k\}. \tag{3-2}$$

These reflections generate the *affine Weyl group* $W_{\mathrm{aff}}$ for the *dual root system* $\Phi^\vee := \{\alpha^\vee : \alpha \in \Phi\}$. The hyperplanes $H_{\alpha,k}$ divide the real vector space $\mathfrak{h}_{\mathbb{R}}^*$ into open regions, called *alcoves*. The *fundamental alcove* $A_\circ$ is given by

$$A_\circ := \{\lambda \in \mathfrak{h}_{\mathbb{R}}^* : 0 < \langle\lambda,\alpha^\vee\rangle < 1 \text{ for all } \alpha \in \Phi^+\}.$$

**3.2.** ***Alcove walks.*** We say that two alcoves $A$ and $B$ are *adjacent* if they are distinct and have a common wall. Given two such alcoves, we write $A \xrightarrow{\beta} B$ if the common wall is of the form $H_{\beta,k}$ and the root $\beta \in \Phi$ points in the direction from $A$ to $B$.

**Definition 3.1.** An *alcove path* is a sequence of alcoves such that any two consecutive ones are adjacent. We say that an alcove path $(A_0, A_1, \dots, A_m)$ is *reduced* if $m$ is the minimal length of all alcove paths from $A_0$ to $A_m$.

We need the following generalization of alcove paths.

**Definition 3.2.** An *alcove walk* is a sequence

$$\Omega = (A_0, F_1, A_1, F_2, \ldots, F_m, A_m, F_\infty)$$

such that $A_0, \ldots, A_m$ are alcoves; $F_i$ is a codimension-one common face of the alcoves $A_{i-1}$ and $A_i$, for $i = 1, \ldots, m$; and $F_\infty$ is a vertex of the last alcove $A_m$. The weight $F_\infty$ is called the *weight* of the alcove walk, and is denoted by $\mu(\Omega)$.

The *folding operator* $\phi_i$ is the operator that acts on an alcove walk by leaving its initial segment from $A_0$ to $A_{i-1}$ intact and by reflecting the remaining tail in the affine hyperplane containing the face $F_i$. In other words, we define

$$\phi_i(\Omega) := (A_0, F_1, A_1, \ldots, A_{i-1}, F_i' = F_i, A_i', F_{i+1}', A_{i+1}', \ldots, A_m', F_\infty');$$

here $A_j' := \rho_i(A_j)$ for $j \in \{i, \ldots, m\}$, $F_j' := \rho_i(F_j)$ for $j \in \{i, \ldots, m\} \cup \{\infty\}$, and $\rho_i$ is the affine reflection in the hyperplane containing $F_i$. Note that any two folding operators commute. An index $j$ such that $A_{j-1} = A_j$ is called a *folding position* of $\Omega$. Let $\mathrm{fp}(\Omega) := \{j_1 < \cdots < j_s\}$ be the set of folding positions of $\Omega$. If this set is empty, $\Omega$ is called *unfolded*. We define the operator "unfold", producing an unfolded alcove walk, by

$$\mathrm{unfold}(\Omega) = \phi_{j_1} \ldots \phi_{j_s}(\Omega).$$

**Definition 3.3.** An alcove walk $\Omega = (A_0, F_1, A_1, F_2, \ldots, F_m, A_m, F_\infty)$ is called *positively folded* if, for any folding position $j$, the alcove $A_{j-1} = A_j$ lies on the positive side of the affine hyperplane containing the face $F_j$.

We now fix a dominant weight $\lambda$ and a reduced alcove path

$$\Pi := (A_0, A_1, \ldots, A_m)$$

from $A_\circ = A_0$ to its translate $A_\circ + \lambda = A_m$. Assume that

$$A_0 \xrightarrow{\beta_1} A_1 \xrightarrow{\beta_2} \cdots \xrightarrow{\beta_m} A_m,$$

where $\Gamma := (\beta_1, \ldots, \beta_m)$ is a sequence of positive roots. This sequence, which determines the alcove path, is called a $\lambda$*-chain* (of roots). Two equivalent definitions of $\lambda$-chains (in terms of reduced words in affine Weyl groups, and an interlacing condition) can be found in [Lenart and Postnikov 2007, Definition 5.4] and [Lenart and Postnikov 2008, Definition 4.1 and Proposition 4.4]; note that the $\lambda$-chains considered in the these papers are obtained by reversing the ones in this paper. We also let $r_i := s_{\beta_i}$, and let $\hat{r}_i$ be the affine reflection in the common wall of $A_{i-1}$ and $A_i$, for $i = 1, \ldots, m$; in other words, $\hat{r}_i := s_{\beta_i, l_i}$, where $l_i := |\{j \le i \,:\, \beta_j = \beta_i\}|$ is the cardinality of the corresponding set. Given

$$J = \{j_1 < \cdots < j_s\} \subseteq [m] := \{1, \ldots, m\},$$

we define the Weyl group element $\phi(J)$ and the weight $\mu(J)$ by

$$\phi(J) := r_{j_1} \dots r_{j_s}, \quad \mu(J) := \hat{r}_{j_1} \dots \hat{r}_{j_s}(\lambda). \tag{3-3}$$

Given $w \in W$, we define the alcove path $w(\Pi) := (w(A_0), w(A_1), \dots, w(A_m))$. Consider the set of alcove paths

$$\mathcal{P}(\Gamma) := \{w(\Pi) : w \in W^\lambda\}.$$

We identify any $w(\Pi)$ with the obvious unfolded alcove walk of weight

$$\mu(w(\Pi)) := w(\lambda).$$

Let us now consider the set of alcove walks

$$\mathcal{F}_+(\Gamma) := \{\text{positively folded alcove walks } \Omega : \text{unfold}(\Omega) \in \mathcal{P}(\Gamma)\}.$$

We can encode an alcove walk $\Omega$ in $\mathcal{F}_+(\Gamma)$ by the pair $(w, J)$ in $W^\lambda \times 2^{[m]}$, where

$$\text{fp}(\Omega) = J \quad \text{and} \quad \text{unfold}(\Omega) = w(\Pi).$$

Clearly, we can recover $\Omega$ from $(w, J)$ with $J = \{j_1 < \cdots < j_s\}$ by

$$\Omega = \phi_{j_1} \dots \phi_{j_s}(w(\Pi)).$$

Let $\mathcal{A}(\Gamma)$ be the image of $\mathcal{F}_+(\Gamma)$ under the map $\Omega \mapsto (w, J)$. We call a pair $(w, J)$ in $\mathcal{A}(\Gamma)$ an *admissible pair*, and the subset $J \subseteq [m]$ in this pair a *$w$-admissible subset*.

**Proposition 3.4** [Lenart 2011]. *If $\Omega \mapsto (w, J)$, then $\mu(\Omega) = w(\mu(J))$. Moreover,*

$$\begin{aligned}\mathcal{A}(\Gamma) = \big\{(w, J) \in W^\lambda \times 2^{[m]} : J = \{j_1 < \cdots < j_s\},\\ w > wr_{j_1} > \cdots > wr_{j_1} \dots r_{j_s} = w\phi(J)\big\};\end{aligned} \tag{3-4}$$

*where the decreasing chain is in the Bruhat order on the Weyl group, its steps not being covers necessarily.*

The formula for the Hall–Littlewood $P$-polynomials in [Schwer 2006] was rederived in [Ram 2006] in a slightly different version, based on positively folded alcove walks. Based on Proposition 3.4, we now restate the latter formula in terms of admissible pairs.

**Theorem 3.5** [Ram 2006; Schwer 2006]. *Given a dominant weight $\lambda$, we have*

$$P_\lambda(X; t) = \sum_{(w,J) \in \mathcal{A}(\Gamma)} t^{(1/2)(\ell(w)+\ell(w\phi(J))-|J|)} (1-t)^{|J|} x^{w(\mu(J))}. \tag{3-5}$$

## 4. Specializing Ram's version of Schwer's formula to type $C$

We now restrict ourselves to the root system of type $C_n$. We can identify the space $\mathfrak{h}_{\mathbb{R}}^*$ with $V := \mathbb{R}^n$, the coordinate vectors being $\varepsilon_1, \dots, \varepsilon_n$. The root system $\Phi$ can be represented as $\Phi = \{\pm\varepsilon_i \pm \varepsilon_j : 1 \le i < j \le n\} \cup \{\pm 2\varepsilon_i : 1 \le i \le n\}$. The simple roots are $\alpha_i = \varepsilon_i - \varepsilon_{i+1}$, for $i = 1, \dots, n-1$ and $\alpha_n = 2\varepsilon_n$. The fundamental weights are $\omega_i = \varepsilon_1 + \cdots + \varepsilon_i$, for $i = 1, \dots, n$. The weight lattice is $\Lambda = \mathbb{Z}^n$. A dominant weight $\lambda = \lambda_1\varepsilon_1 + \cdots + \lambda_{n-1}\varepsilon_{n-1} + \lambda_n\varepsilon_n$ is identified with the partition $(\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{n-1} \ge \lambda_n \ge 0)$ of length at most $n$. A dominant weight is regular if all these inequalities are strict: that is, the corresponding partition has all parts distinct and nonzero. We fix such a partition $\lambda$ for the remainder of this paper.

The corresponding Weyl group $W$ is the group of signed permutations $B_n$. For simplicity, we use the same notation for roots and the corresponding reflections (see Section 2). For instance, given $1 \le i < j \le n$, we denote by $(i, j)$ the positive root $\varepsilon_i - \varepsilon_j$, by $(i, \bar{j})$ the positive root $\varepsilon_i + \varepsilon_j$, and by $(i, \bar{\imath})$ the positive root $2\varepsilon_i$.

Let
$$\Gamma(k) := \Gamma'_2 \dots \Gamma'_k \Gamma_1(k) \dots \Gamma_k(k),$$
where
$$\begin{aligned} \Gamma'_j &:= \big((1, \bar{j}), \quad (2, \bar{j}), \quad \dots, (j-1, \bar{j})\big); \\ \Gamma_j(k) &:= \big((1, \bar{j}), \quad (2, \bar{j}), \quad \dots, (j-1, \bar{j}), \\ &\qquad (j, \overline{k+1}), (j, \overline{k+2}), \dots, (j, \bar{n}), \quad (j, \bar{j}), \\ &\qquad (j, n), \quad (j, n-1), \quad \dots, (j, k+1)\big). \end{aligned}$$

**Lemma 4.1.** *$\Gamma(k)$ is an $\omega_k$-chain.*

*Proof.* We use the criterion for $\lambda$-chains given in [Lenart and Postnikov 2008, Definition 4.1 and Proposition 4.4] (see also Proposition 10.2 of the same reference). This criterion says that a chain of roots $\Gamma$ is a $\lambda$-chain if and only if it satisfies the following conditions:

(R1) The number of occurrences of any positive root $\alpha$ in $\Gamma$ is $\langle \lambda, \alpha^\vee \rangle$.

(R2) For each triple of positive roots $(\alpha, \beta, \gamma)$ with $\gamma^\vee = \alpha^\vee + \beta^\vee$, the subsequence of $\Gamma$ consisting of $\alpha, \beta, \gamma$ is a concatenation of pairs $(\gamma, \alpha)$ and $(\gamma, \beta)$ (in any order).

Letting $\lambda = \omega_k = \varepsilon_1 + \cdots + \varepsilon_k$, the first condition is easily checked; for instance, a root $(a, \bar{b})$ appears twice in $\Gamma(k)$ if $a < b \le k$, once if $a \le k < b$, and zero times otherwise. For the second condition, we use a case by case analysis, as follows, where $a < b < c$:

(1) $\alpha = (a, b),\ \beta = (b, c),\ \gamma = (a, c)$; (2) $\alpha = (a, b),\ \beta = (b, \bar{c}),\ \gamma = (a, \bar{c})$;

(3) $\alpha = (a, c),\ \beta = (b, \bar{c}),\ \gamma = (a, \bar{b})$; (4) $\alpha = (b, c),\ \beta = (a, \bar{c}),\ \gamma = (a, \bar{b})$;

(5) $\alpha = (a, b),\ \beta = (b, \bar{b}),\ \gamma = (a, \bar{a})$; (6) $\alpha = (a, \bar{a}),\ \beta = (b, \bar{b}),\ \gamma = (a, \bar{b})$.

Case (1) is the same as in type $A$. Cases (2)–(4) each have the three subcases $k \geq c$, $b \leq k < c$, and $a \leq k < b$; while cases (5) and (6) each have the two subcases $k \geq b$ and $a \leq k < b$. For instance, if $b \leq k < c$ in case (3), the subsequence of $\Gamma(k)$ consisting of $\alpha, \beta, \gamma$ is $((a, \overline{b}), (a, c), (a, \overline{b}), (b, \overline{c}))$. $\square$

Hence, we can construct a $\lambda$-chain as a concatenation $\Gamma := \Gamma^{\lambda_1} \dots \Gamma^1$, where

$$\Gamma^i = \Gamma(\lambda'_i) = \Gamma'_{i2} \dots \Gamma'_{i,\lambda'_i} \Gamma_{i1} \dots \Gamma_{i,\lambda'_i} \quad \text{and} \quad \Gamma_{ij} = \Gamma_j(\lambda'_i), \;\; \Gamma'_{ij} = \Gamma'_j. \tag{4-1}$$

This $\lambda$-chain is fixed for the remainder of this paper. Thus, we can replace the notation $\mathscr{A}(\Gamma)$ with $\mathscr{A}(\lambda)$.

**Example 4.2.** Consider $n = 3$ and $\lambda = (3, 2, 1)$, for which we have the $\lambda$-chain below. The factorization of $\Gamma$ into subchains is indicated with vertical bars, while the double vertical bars separate the subchains corresponding to different columns. The underlined pairs are only relevant in Example 4.3 below.

$$\begin{aligned}\Gamma &= \Gamma_{31} \,\|\, \Gamma'_{22}\Gamma_{21}\Gamma_{22} \,\|\, \Gamma'_{12}\Gamma'_{13}\Gamma_{11}\Gamma_{12}\Gamma_{13} \\ &= \big((1,\overline{2}), \underline{(1,\overline{3})}, (1,\overline{1}), (1,3), (1,2) \,\| \\ &\qquad \underline{(1,\overline{2})} \,|\, (1,\overline{3}), (1,\overline{1}), (1,3) \,|\, (1,\overline{2}), (2,\overline{3}), \underline{(2,\overline{2})}, \underline{(2,3)} \,\| \\ &\qquad (1,\overline{2}) \,|\, (1,\overline{3}), (2,\overline{3}) \,|\, (1,\overline{1}) \,|\, (1,\overline{2}), (2,\overline{2}) \,|\, (1,\overline{3}), (2,\overline{3}), (3,\overline{3})\big).\end{aligned} \tag{4-2}$$

We represent the Young diagram of $\lambda$ inside a broken $3 \times 2$ rectangle, as below. In this way, a reflection in $\Gamma$ can be viewed as swapping entries and/or changing signs in the two parts of each column, or only the top part.

$$\begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline \multicolumn{1}{c|}{} & 2 & 2 \\ \cline{2-3} \multicolumn{1}{c}{} & \multicolumn{1}{c|}{} & 3 \\ \cline{3-3} \end{array}$$

$$\begin{array}{|c|c|} \cline{1-1} 2 & \multicolumn{1}{c}{} \\ \hline 3 & 3 \\ \hline \end{array}$$

Given the $\lambda$-chain $\Gamma$ above, in Section 3.2 we considered subsets

$$J = \{j_1 < \cdots < j_s\}$$

of $[m]$, where $m$ is the length of $\Gamma$. Instead of $J$, it is now convenient to use the subsequence $T$ of the roots in $\Gamma$ whose positions are in $J$. This is viewed as a concatenation with distinguished factors $T_{ij}$ and $T'_{ik}$ induced by the factorization (4-1) of $\Gamma$.

All the notions defined in terms of $J$ are now redefined in terms of $T$. As such, from now on we will write $\phi(T)$, $\mu(T)$, and $|T|$, the latter being the size of $T$; see (3-3). If $J$ is a $w$-admissible subset for some $w$ in $B_n$, we will also call the corresponding $T$ a *$w$-admissible sequence*, and $(w, T)$ an admissible pair. We will use the notation $\mathscr{A}(\Gamma)$ and $\mathscr{A}(\lambda)$ accordingly. We denote by $wT_{\lambda_1,1}\dots T_{ij}$ and $wT_{\lambda_1,1}\dots T'_{ik}$ the permutations obtained from $w$ via right multiplication by the transpositions in $T_{\lambda_1,1},\dots,T_{ij}$ and $T_{\lambda_1,1},\dots,T'_{ik}$, considered from left to right. This agrees with the convention of using pairs to denote both roots and the corresponding reflections. As such, $\phi(T)$ can now be written simply as $T$.

**Example 4.3.** We continue Example 4.2, by picking the admissible pair $(w,J)$ with $w=\bar{1}\,\bar{2}\,\bar{3}\in B_3$ and $J=\{2,6,12,13\}$ (see the underlined positions in (4-2)). Thus, we have

$$T=T_{31}\parallel T'_{22}T_{21}T_{22}\parallel T'_{12}T'_{13}T_{11}T_{12}T_{13}=((1,\bar{3})\parallel(1,\bar{2})\mid\,\mid(2,\bar{2}),(2,3)\parallel\,\mid\,\mid\,\mid\,\mid\,).$$

The corresponding decreasing chain in Bruhat order is the following, where the swapped entries are shown in bold (we represent permutations as broken columns starting with $w$, as discussed in Example 4.2, and we display the splitting of the chain into subchains induced by the splitting of $T$ just given):

$$\begin{array}{c}\boxed{\mathbf{\bar 1}}\\ \\ \boxed{\bar 2}\\ \boxed{\mathbf{\bar 3}}\end{array} > \begin{array}{c}\boxed{3}\\ \\ \boxed{\bar 2}\\ \boxed{1}\end{array} \parallel \begin{array}{c}\boxed{\mathbf{3}}\\ \boxed{\mathbf{\bar 2}}\\ \\ \boxed{1}\end{array} > \begin{array}{c}\boxed{2}\\ \boxed{\bar 3}\\ \\ \boxed{1}\end{array} \mid \begin{array}{c}\boxed{2}\\ \boxed{\bar 3}\\ \\ \boxed{1}\end{array} \mid \begin{array}{c}\boxed{2}\\ \boxed{\mathbf{\bar 3}}\\ \\ \boxed{1}\end{array} > \begin{array}{c}\boxed{2}\\ \boxed{\mathbf{3}}\\ \\ \boxed{\mathbf{1}}\end{array} > \begin{array}{c}\boxed{2}\\ \boxed{1}\\ \\ \boxed{3}\end{array} \parallel \begin{array}{c}\boxed{2}\\ \boxed{1}\\ \boxed{3}\end{array} \mid \begin{array}{c}\boxed{2}\\ \boxed{1}\\ \boxed{3}\end{array} \mid \begin{array}{c}\boxed{2}\\ \boxed{1}\\ \boxed{3}\end{array} \mid \begin{array}{c}\boxed{2}\\ \boxed{1}\\ \boxed{3}\end{array} \mid \begin{array}{c}\boxed{2}\\ \boxed{1}\\ \boxed{3}\end{array}.$$

Given a (not necessarily admissible) pair $(w,T)$, with $T$ split into factors $T_{ij}$ and $T'_{ik}$ as above, we consider the permutations

$$\pi_{ij}=\pi_{ij}(w,T):=wT_{\lambda_1,1}\dots T_{i,j-1},\qquad \pi'_{ik}=\pi'_{ik}(w,T):=wT_{\lambda_1,1}\dots T'_{i,k-1};$$

when undefined, $T_{i,j-1}$ and $T'_{i,k-1}$ are given by conventions similar to (2-3), based on the corresponding factorization (4-1) of the $\lambda$-chain $\Gamma$. In particular, $\pi_{\lambda_1,1}=w$.

**Definition 4.4.** The *filling map* is the map $f$ from pairs $(w,T)$, not necessarily admissible, to fillings $\sigma=f(w,T)$ of the shape $\hat{\lambda}$, defined (based on the notation (2-2)) by

$$C_{ij}=\pi_{ij}[1,\lambda'_i],\qquad C'_{ik}=\pi'_{ik}[1,\lambda'_i]. \tag{4-3}$$

**Example 4.5.** Given $(w,T)$ as in Example 4.3, we have

$$f(w,T)=\begin{array}{|c|c|c|c|c|c|c|}\hline \bar1 & 3 & 2 & 2 & 2 & 2 & 2\\ \hline & \bar2 & \bar3 & \bar3 & 1 & 1 & 1\\ \hline & & & & 3 & 3 & 3\\ \hline\end{array}.$$

The following theorem describes the way in which our tableau formula (2-7) is obtained by compressing Ram's version of Schwer's formula (3-5). Recall that $\lambda$ is a regular weight, so $B_n^\lambda=B_n$, and thus the pairs $(w,J)$ in $\mathscr{A}(\lambda)$ are only subject to the decreasing chain condition in (3-4); this fact is implicitly used in the proof of the theorem.

**Theorem 4.6.** (i) *We have* $f(\mathcal{A}(\lambda)) = \mathcal{F}(\lambda)$.

(ii) *Given any* $\sigma \in \mathcal{F}(\lambda)$ *and* $(w, T) \in f^{-1}(\sigma)$, *we have* $\operatorname{ct} f(w, T) = w(\mu(T))$.

(iii) *The following compression formula holds*:

$$\sum_{(w,T)\in f^{-1}(\sigma)} t^{\frac{1}{2}(\ell(w)+\ell(wT)-|T|)}(1-t)^{|T|} = t^{N(\sigma)}(1-t)^{\operatorname{des}\sigma}. \tag{4-4}$$

*Proof.* We start with part (i). That we have $f(\mathcal{A}(\lambda)) \subseteq \mathcal{F}(\lambda)$ is clear from the definition of the set of fillings $\mathcal{F}(\lambda)$ in Section 2 and the construction (4-1) of the fixed $\lambda$-chain $\Gamma$. Vice versa, given a filling $\sigma$ in $\mathcal{F}(\lambda)$, it is not hard to construct an admissible pair $(w, T)$ in $f^{-1}(\sigma)$. We will assign to the columns $C_{ij}$ and $C'_{ij}$ signed permutations $\rho_{ij}$ and $\rho'_{ij}$ in $B_n$ recursively, starting with $\rho_{11} := C_{11}$; in parallel, we construct the reverse $\operatorname{rev} T$ of the mentioned chain of roots $T$, and conclude by letting $w := \rho_{\lambda_1,1}$. Each time we pass to the left neighbor $C'_{ik}$ of a column $C'_{i,k+1} = C'_{ik}(r_1, \overline{k}) \dots (r_p, \overline{k})$, we append to the part of $\operatorname{rev} T$ already constructed the roots $(r_p, \overline{k}), \dots, (r_1, \overline{k})$ and let $\rho'_{ik} := \rho'_{i,k+1}(r_p, \overline{k}) \dots (r_1, \overline{k})$. We proceed similarly when passing to the left neighbor $C_{ij}$ of a column $C_{i,j+1}$, where $C_{i,j+1}$ differs from $D = C_{ij}(r_1, \overline{j}) \dots (r_p, \overline{j})$ only in position $j$; the only difference is that, in this case, we start by applying to $\rho_{i,j+1}$ and appending to $\operatorname{rev} T$ the reflection that exchanges the entry $C_{i,j+1}(j)$ with $D(j)$, and then we proceed as above.

Parts (ii) and (iii) of the theorem are proved in Sections 6 and 7. □

**Remarks 4.7.** (i) The Kashiwara–Nakashima tableaux [1994] of shape $\lambda$ index the basis elements of the irreducible representation of $\mathfrak{sp}_{2n}$ of highest weight $\lambda$. It is shown in Proposition 4.8 below that these tableaux correspond precisely to the surviving fillings in our formula (2-7) when we set $t = 0$.

(ii) In (4-4), in general, we cannot replace the filling map $f$ with the map $\overline{f}$, sending $(w, T)$ to the compressed version $\overline{f(w, T)}$ of $f(w, T)$. Indeed, consider $n = 2$, $\lambda = (3, 2)$, and the following filling of $2\lambda = (6, 4)$, which happens to be the "doubled" version of a Kashiwara–Nakashima tableau:

$$\overline{\sigma} = \begin{array}{|c|c|c|c|c|c|} \hline \overline{2} & \overline{2} & \overline{2} & \overline{2} & 1 & 1 \\ \hline \multicolumn{2}{c|}{} & \overline{1} & \overline{1} & 2 & 2 \\ \cline{3-6} \end{array}.$$

If $(w, T) \in \overline{f}^{-1}(\overline{\sigma})$, we must have $w = \overline{21}$ and

$$T \subseteq \Gamma_{21}\Gamma_{22} = \big((1, \overline{1}) | (1, \overline{2}), (2, \overline{2})\big),$$

where the full $\lambda$-chain factors as follows:

$$\Gamma = \Gamma_{31} \parallel \Gamma'_{22}\Gamma_{21}\Gamma_{22} \parallel \Gamma'_{12}\Gamma_{11}\Gamma_{12}.$$

There are two elements $(w, T^1)$ and $(w, T^2)$ in $\overline{f}^{-1}(\overline{\sigma})$, namely

$$T^1 = \big((1, \overline{2})\big) \quad \text{and} \quad T^2 = \big((1, \overline{1}), (1, \overline{2}), (2, \overline{2})\big).$$

But we have

$$\sum_{(w,T)\in \overline{f}^{-1}(\overline{\sigma})} t^{(1/2)(\ell(w)+\ell(wT)-|T|)}(1-t)^{|T|} = t(1-t)+(1-t)^3 = (1-t)(1-t+t^2).$$

In general, this sum has several factors not of the form $t$ or $(1-t)$, which are hard to control.

(iii) To measure the compression phenomenon, we define the *compression factor* $c(\lambda)$ as in [Lenart 2011], as the ratio of the number of terms in Ram's version of Schwer's formula for $\lambda$ and the number of terms in the tableau formula. The compression factor is considerably larger in type $C$. For instance, for $\lambda = (3, 2, 1, 0)$ in $C_4$ we have 23,495 terms in the compressed formula, while $c(\lambda) = 44.9$.

**Proposition 4.8.** *The map $\sigma \mapsto \overline{\sigma}$ defined by (2-4) is a bijection between the fillings $\sigma$ in $\mathcal{F}(\lambda)$ with $N(\sigma) = 0$ and the "doubled" versions of the type $C$ Kashiwara–Nakashima tableaux of shape $\lambda$.*

*Proof.* Adamczak and the author [2009] proved that for each type $C$ Kashiwara–Nakashima tableau of shape $\lambda$ there is a unique admissible pair $(w, T)$ whose associated chain in Bruhat order is saturated and ends at the identity, such that the compressed version $\overline{\sigma}$ of $\sigma = f(w, T)$ is the "doubled" version of the given tableau. It follows that the term associated to $(w, T)$ in (3-5) is $t^0(1-t)^{|T|}x^{w(\mu(T))}$, and therefore $N(\sigma) = 0$, by (4-4). On the other hand, since $P_\lambda(x; 0)$ is the irreducible character indexed by $\lambda$, which is expressed in terms of Kashiwara–Nakashima tableaux, we conclude that all $\sigma$ in $\mathcal{F}(\lambda)$ with $N(\sigma) = 0$ arise in this way. □

## 5. The tableau formula in type $B$

We now restrict ourselves to the root system of type $B_n$. This can be represented as $\Phi = \{\pm\varepsilon_i \pm \varepsilon_j : 1 \le i < j \le n\} \cup \{\pm\varepsilon_i : 1 \le i \le n\}$. The simple roots are $\alpha_i = \varepsilon_i - \varepsilon_{i+1}$, for $i = 1, \ldots, n-1$ and $\alpha_n = \varepsilon_n$. The fundamental weights are $\omega_i = \varepsilon_1 + \cdots + \varepsilon_i$, for $i = 1, \ldots, n-1$ and $\omega_n = \frac{1}{2}(\varepsilon_1 + \cdots + \varepsilon_n)$. A dominant weight $\lambda = \alpha_1\omega_1 + \cdots + \alpha_n\omega_n$, where $\alpha_i \in \mathbb{Z}_{\ge 0}$, is identified with the partition $\mu = (n^{\alpha_n}, \ldots, 1^{\alpha_1})$; we let $\ell(\mu) := \alpha_1 + \cdots + \alpha_n$, and write $\mu = (\mu_1, \ldots, \mu_{\ell(\mu)})$. A dominant weight is regular if $\alpha_i > 0$ for all $i$. Let us now fix such a weight $\lambda$.

The corresponding Weyl group $W$ is the same group of signed permutations $B_n$ considered above. For simplicity, we again use the same notation for roots and the corresponding reflections; see Section 2. The pairs $(i, j)$ and $(i, \overline{j})$ have the same meaning as in type $C$, whereas $(i)$ denotes the positive root $\varepsilon_i$. Note that, as a reflection in $B_n$, $(i)$ is the same as $(i, \overline{\imath})$ in type $C$.

The canonical $\omega_k$-chains and $\lambda$-chains are very similar to those in type $C$. If $k<n$, let

$$\Gamma(k) := \Gamma_1' \dots \Gamma_k' \Gamma_1(k) \dots \Gamma_k(k),$$

where

$$\begin{aligned}
\Gamma_j' &:= \big((1,\bar{\jmath}), \quad (2,\bar{\jmath}), \quad \dots, (j-1,\bar{\jmath}), (j)\big);\\
\Gamma_j(k) &:= \big((1,\bar{\jmath}), \quad (2,\bar{\jmath}), \quad \dots, (j-1,\bar{\jmath}),\\
&\qquad (j,\overline{k+1}), (j,\overline{k+2}), \dots, (j,\bar{n}),\\
&\qquad (j),\\
&\qquad (j,n), \quad (j,n-1), \dots, (j,k+1)\big).
\end{aligned}$$

On the other hand, we let

$$\Gamma(n) := \Gamma_1' \dots \Gamma_n' = \Gamma_1(n) \dots \Gamma_n(n).$$

As in the type $C$ case, we can prove that $\Gamma(k)$ is an $\omega_k$-chain for any $k$. Hence, we can construct a $\lambda$-chain as a concatenation $\Gamma := \Gamma^{\ell(\mu)} \dots \Gamma^1$, where $\Gamma^i = \Gamma(\mu_i)$.

The filling map is defined as in Definition 4.4. This gives rise to fillings

$$\sigma = \mathscr{C}^{\ell(\mu)} \dots \mathscr{C}^1,$$

where each $\mathscr{C}^i$ is a concatenation of columns of height $\mu_i$:

$$\mathscr{C}^i := \begin{cases} C_{i1}' \dots C_{i,\mu_i}' C_{i1} \dots C_{i\mu_i} & \text{if } \mu_i < n,\\ C_{i1} \dots C_{i,\mu_i} & \text{if } i \neq 1 \text{ and } \mu_i = n,\\ C_{11} & \text{if } i = 1. \end{cases}$$

The fillings are subject to the same conditions (1)–(3) as in type $C$ in Section 2, where condition (3) is made more precise below. In fact, the $\lambda$-chain $\Gamma$ above specifies the way in which each column is related to its left neighbor. Essentially, everything is similar to type $C$, except for a small difference in the passage from some column $C_{ik}'$ to $C_{i,k+1}'$. Namely, there exist positions $1 \le r_1 < \dots < r_p < k$ (possibly $p=0$) such that $C_{i,k+1}' = C_{ik}'(r_1,\bar{k}) \dots (r_p,\bar{k})$, like in type $C$, or $C_{i,k+1}' = C_{ik}'(r_1,\bar{k}) \dots (r_p,\bar{k})(k)$, in which case we also require $C_{i,k+1}'(k) \le n$.

The weight of a filling and the statistics $N(\sigma)$ and $\operatorname{des}\sigma$ are defined completely similarly to type $C$. The only minor addition is the definition of $N(D,C)$ and $\operatorname{des}(D,C)$ when $C = D(r_1,\bar{k}) \dots (r_p,\bar{k})(k)$. With the notation in Case 2 of the definition of $N(D,C)$, we set

$$N(D,C) := N(D,D_p) + N(D_p,C), \quad \operatorname{des}(D,C) := p+1.$$

Here $N(D,D_p)$ is defined in Case 2, whereas $N(D_p,C)$ is given by the second formula in (7-1); more precisely,

$$N(D_p,C) := \tfrac{1}{2}\big|(\bar{a},a) \setminus \{\pm D_p(i) : i = 1,\dots,k\}\big|,$$

where $a := D_p(k)$.

Given these constructions, the proof of the following theorem is completely similar to its counterparts in type $C$, since no new situations arise.

**Theorem 5.1.** *Theorems 2.2 and 4.6 hold in type $B$, with the appropriate constructions explained above.*

**Remark 5.2.** The situation in type $D$ is slightly more complex. In this case, applying the preceding ideas leads to an analog of the compression formula (4-4) that contains factors of the form $1 - t^k$ with $k > 1$ in the right side. However, these factors are not hard to control, while no extra factors appear.

## 6. Proof of Theorem 4.6(ii)

Recall the $\lambda$-chain $\Gamma$ in Section 4. Let us write $\Gamma = (\beta_1, \dots, \beta_m)$, as in Section 3.2. As such, we recall the hyperplanes $H_{\beta_k, l_k}$ and the corresponding affine reflections $\hat{r}_k = s_{\beta_k, l_k} = s_{\beta_k} + l_k \beta_k$.

Now fix a signed permutation $w$ in $B_n$ and a subset $J = \{j_1 < \cdots < j_s\}$ of $[m]$ (not necessarily $w$-admissible). Let $\Pi$ be the alcove path corresponding to $\Gamma$, and define the alcove walk $\Omega$ as in Section 3.2, by

$$\Omega := \phi_{j_1} \dots \phi_{j_s}(w(\Pi)).$$

Given $k$ in $[m]$, let $i = i(k)$ be the largest index in $[s]$ for which $j_i < k$, and let $\gamma_k := w r_{j_1} \dots r_{j_i}(\beta_k)$. Then the hyperplane containing the face $F_k$ of $\Omega$ (see Definition 3.2) is of the form $H_{\gamma_k, m_k}$; in other words,

$$H_{\gamma_k, m_k} = w \hat{r}_{j_1} \dots \hat{r}_{j_i}(H_{\beta_k, l_k}).$$

Our first goal is to describe $m_k$ purely in terms of the filling associated to $(w, J)$.

Let $\hat{t}_k$ be the affine reflection in the hyperplane $H_{\gamma_k, m_k}$. Note that

$$\hat{t}_k = w \hat{r}_{j_1} \dots \hat{r}_{j_i} \hat{r}_k \hat{r}_{j_i} \dots \hat{r}_{j_1} w^{-1}.$$

Thus, we can see that

$$w \hat{r}_{j_1} \dots \hat{r}_{j_i} = \hat{t}_{j_i} \dots \hat{t}_{j_1} w.$$

Let $T = ((a_1, b_1), \dots, (a_s, b_s))$ be the subsequence of $\Gamma$ indexed by the positions in $J$; see Section 4. Let $T^i$ be the initial segment of $T$ with length $i$, let $w_i := w T^i$, and let $\sigma_i := \overline{f(w, T^i)}$; see (2-4). In particular, $\sigma_0$ is the filling with all entries in row $i$ equal to $w(i)$, and $\sigma := \sigma_s = \overline{f(w, T)}$. The columns of a filling of $2\lambda$ are numbered left to right by $2\lambda_1$ to 1. We split each segment $\Gamma^k$ of $\Gamma$ into two parts: the head, which is a concatenation of $\Gamma'_{k\cdot}$, and the tail, which is a concatenation of $\Gamma_{k\cdot}$; see (4-1). We say that the head corresponds to column $2k$ of the Young diagram $2\lambda$, whereas the tail corresponds to column $2k-1$ (see the construction of $f(w, T)$ in Section 4 and (2-4)). If $\beta_{j_{i+1}} = (a_{i+1}, b_{i+1}) = (a, b)$ falls in the segment of $\Gamma$ corresponding to column $p$ of $2\lambda$, then $\sigma_{i+1}$ is obtained from $\sigma_i$ by replacing the entry $w_i(a)$ with $w_i(b)$ in the columns $p-1, \ldots, 1$ of $\sigma_i$, as well as, possibly, the entry $w_i(\overline{b})$ with $w_i(\overline{a})$ in the same columns.

Now fix a position $k$, and consider $i = i(k)$ and the roots $\beta_k$, $\gamma := \gamma_k$, as above, where $\gamma_k$ might be negative. Assume that $\beta_k$ falls in the segment of $\Gamma$ corresponding to column $q$ of $2\lambda$. Given a filling $\phi$, we denote by $\phi^{[p]}$ the part of $\phi$ consisting of columns $2\lambda_1, 2\lambda_1-1, \ldots, p$, and by $\phi^{(p,q]}$ the part consisting of columns $p-1,\ p-2, \ldots, q$. We also recall the definition (2-5) and conventions related to the content of a filling; this definition now applies to any filling of some Young diagram.

**Proposition 6.1.** *With the same notation*, *we have*

$$m_k = \langle \mathrm{ct}\,\sigma^{[q]}, \gamma^\vee \rangle.$$

*Proof.* We apply induction on $i$, which starts at $i = 0$, when the verification is straightforward. We now proceed from $j_1 < \cdots < j_i < k$, where $i = s$ or $k \le j_{i+1}$, to $j_1 < \cdots < j_{i+1} < k$, and we freely use the notation above.

Assume that $\beta_{j_{i+1}}$ falls in the segment of $\Gamma$ corresponding to column $p$ of $2\lambda$, where $p \ge q$.

We need to compute

$$w\hat{r}_{j_1} \ldots \hat{r}_{j_{i+1}}(H_{\beta_k, l_k}) = \hat{t}_{j_{i+1}} \ldots \hat{t}_{j_1} w(H_{\beta_k, l_k}) = \hat{t}_{j_{i+1}}(H_{\gamma,m}),$$

where $m = \langle \mathrm{ct}\,\sigma_i^{[q]}, \gamma^\vee \rangle$, by induction. Let $\gamma' := \gamma_{j_{i+1}}$, and $\hat{t}_{j_{i+1}} = s_{\gamma',m'}$, where $m' = \langle \mathrm{ct}\,\sigma_i^{[p]}, (\gamma')^\vee \rangle$, by induction. We use the formula

$$s_{\gamma',m'}(H_{\gamma,m}) = H_{s_{\gamma'}(\gamma), m - m'\langle \gamma', \gamma^\vee \rangle}.$$

Thus, the proof is reduced to showing that

$$m - m'\langle \gamma', \gamma^\vee \rangle = \langle \mathrm{ct}\,\sigma_{i+1}^{[q]}, s_{\gamma'}(\gamma^\vee) \rangle.$$

An easy calculation, based on the information above, shows that the latter equality is nontrivial only if $p > q$, in which case it is equivalent to

$$\langle \mathrm{ct}\,\sigma_{i+1}^{(p,q]} - \mathrm{ct}\,\sigma_i^{(p,q]}, \gamma^\vee \rangle = \langle \gamma', \gamma^\vee \rangle \langle \mathrm{ct}\,\sigma_{i+1}^{(p,q]}, (\gamma')^\vee \rangle.$$

This equality is a consequence of

$$\mathrm{ct}\,\sigma_{i+1}^{(p,q]} = s_{\gamma'}(\mathrm{ct}\,\sigma_i^{(p,q]}),$$

which follows from the construction of $\sigma_{i+1}$ from $\sigma_i$ explained above. □

*Proof of Theorem 4.6(ii).* We apply induction on the size of $T$, using freely the notation above. We prove the statement for $T = (\beta_{j_1}, \ldots, \beta_{j_{s+1}})$, assuming it holds for $T^s = (\beta_{j_1}, \ldots, \beta_{j_s})$. We have

$$w(\mu(T)) = w\hat{r}_{j_1} \ldots \hat{r}_{j_{s+1}}(\lambda) = \hat{t}_{j_{s+1}} \ldots \hat{t}_{j_1} w(\lambda) = \hat{t}_{j_{s+1}}(\mathrm{ct}\,\sigma_s),$$

by induction. We need to show that

$$\hat{t}_{j_{s+1}}(\mathrm{ct}\,\sigma_s) = \mathrm{ct}\,\sigma_{s+1}. \tag{6-1}$$

Let $\gamma := \gamma_{j_{s+1}}$ and assume that $\beta_{j_{s+1}}$ falls in the segment of $\Gamma$ corresponding to column $p$ of $2\lambda$. Based on Proposition 6.1, (6-1) is rewritten as

$$s_\gamma(\mathrm{ct}\,\sigma_s) + \langle \mathrm{ct}\,\sigma_s^{[p]}, \gamma^\vee \rangle \gamma = \mathrm{ct}\,\sigma_{s+1}. \tag{6-2}$$

Decomposing $\mathrm{ct}\,\sigma_s$ as $\mathrm{ct}\,\sigma_s^{[p]} + \mathrm{ct}\,\sigma_s^{(p,1]}$ (using the notation above), and $\mathrm{ct}\,\sigma_{s+1}$ in a similar way, (6-2) would follow from

$$\begin{aligned} s_\gamma(\mathrm{ct}\,\sigma_s^{[p]}) + \langle \mathrm{ct}\,\sigma_s^{[p]}, \gamma^\vee \rangle \gamma &= \mathrm{ct}\,\sigma_{s+1}^{[p]}, \\ s_\gamma(\mathrm{ct}\,\sigma_s^{(p,1]}) &= \mathrm{ct}\,\sigma_{s+1}^{(p,1]}. \end{aligned}$$

The first equality is clear since $\sigma_s^{[p]} = \sigma_{s+1}^{[p]}$, while the second one follows from the construction of $\sigma_{s+1}$ from $\sigma_s$ explained above. □

## 7. Proof of Theorem 4.6(iii)

We start by recalling some basic facts about the group $B_n$ and some notation from Section 2. We will use the following notation related to a word $w = w_1 \ldots w_l$ with integer letters, which is sometimes identified with its set of letters

$$w[i,j] := w_i \ldots w_j, \quad N_{ab}(w) := |(a,b) \cap w|, \quad N_{ab}(\pm w) := N_{ab}(w) + N_{ab}(-w),$$

where $-w := \overline{w_1} \ldots \overline{w_l}$. Given words $w^1, \ldots, w^p$, we let

$$N_{ab}(w^1, \ldots, w^p) := N_{ab}(w^1) + \cdots + N_{ab}(w^p).$$

We also let

$$\tau_{ab} := \begin{cases} 1 & \text{if } a, b \le n, \\ 0 & \text{otherwise.} \end{cases}$$

With this notation, given a signed permutation $w$ in $B_n$ and $1 \le i < j \le n$, $a := w(i)$, $b := w(j)$, we have

$$\begin{aligned} \frac{\ell(w(i,j)) - \ell(w) - 1}{2} &= N_{ab}(w[i,j]), \\ \frac{\ell(w(i,\bar{\imath})) - \ell(w) - 1}{2} &= N_{a\bar{a}}(w[i,n]), \\ \frac{\ell(w(i,\bar{\jmath})) - \ell(w) - 1}{2} &= N_{a\bar{b}}\big(w[i,j-1], \pm w[j+1,n]\big) + \tau_{ab}, \end{aligned} \tag{7-1}$$

assuming that the left side is nonnegative (that is, that we go up in Bruhat order upon applying the indicated reflection); these facts are used implicitly throughout.

Given a chain of roots $\Delta$, we define $\mathcal{A}^r(\Delta)$ as in (3-4) except that we impose an increasing chain condition and $w \in W$. The following simple lemma will be useful throughout, for splitting chains into subchains.

**Lemma 7.1.** *Consider $(w, T)$ with $T$ written as a concatenation $S_1 \dots S_p$. Let $w_i := wS_1 \dots S_i$, so $w_0 = w$. Then*

$$\tfrac{1}{2}(\ell(w)+\ell(wT)-|T|)=\tfrac{1}{2}(\ell(w_{p-1})+\ell(w_p)-|S_p|)+\sum_{i=1}^{p-1}\tfrac{1}{2}(\ell(w_{i-1})-\ell(w_i)-|S_i|).$$

Let $\Delta$ be the chain

$$\begin{aligned}\Delta := \big(\,&(1, p{+}1), (1, p{+}2),\ \dots, (1, n),\\ &(1, \overline{1}),\\ &(1, \bar{n}),\quad (1, \overline{n-1}), \dots, (1, \overline{p+1})\,\big).\end{aligned}$$

**Proposition 7.2.** *Consider a signed permutation $w$ in $B_n$ with $a := w(1)$, a position $1 \le p \le n$, and a value $b \in \{\pm a\} \cup (\pm w[p{+}1, n])$ such that $b \ge a$. Then*

$$\sum_{\substack{T:(w,T)\in\mathcal{A}^r(\Delta)\\ wT(1)=b}} t^{\frac{1}{2}(\ell(wT)-\ell(w)-|T|)}(1-t)^{|T|} = t^{N_{ab}(w[2,p])}(1-t)^{1-\delta_{ab}}, \tag{7-2}$$

*where $\delta_{ab}$ is the Kronecker delta.*

*Proof.* Given $s \in \{\overline{1}, \pm(p+1), \dots, \pm n\}$, we let $\Delta_s$ be the subchain of $\Delta$ starting with $(1, s)$. We also let

$$S(w, s) := \sum_{\substack{T:(w,T)\in\mathcal{A}^r(\Delta_s)\\ wT(1)=b}} t^{\frac{1}{2}(\ell(wT)-\ell(w)-|T|)}(1-t)^{|T|}.$$

We consider three cases: $b = w(q)$, $b = \overline{w(q)}$, and $b = \bar{a}$. The proof in the first case is identical to that of the analogous result for type $A$, namely [Lenart 2011, Proposition 5.3].

<u>Case (ii): $b = \overline{w(q)}$.</u> Let $c := w(q) = \overline{b}$ and $p < q \le s$. We start by showing that

$$S(w, \bar{s}) = t^{N_{a\bar{c}}(w[2,q-1],\,w[q+1,s],\,\pm w[s+1,n])+\tau_{ac}}(1-t). \tag{7-3}$$

We use induction on $s$, which starts at $s = q$. For $s > q$, let $w^1 := w[2, q{-}1]$, $w^2 := w[q{+}1, s{-}1]$, $w^3 := w[s{+}1, n]$, and $d := w(s)$. The sum $S(w, \bar{s})$ splits into two sums: over $s$ such that $(1, \bar{s}) \notin T$ and such that $(1, \bar{s}) \in T$. By induction, the first sum is

$$S(w, \overline{s-1}) = t^{N_{a\bar{c}}(w^1, w^2, \pm d w^3)+\tau_{ac}}(1-t).$$

Again by induction, if $a < \bar{d} < \bar{c}$, then the second sum is

$$t^{N_{a\bar{d}}(w^1cw^2,\pm w^3)+\tau_{ad}}(1-t)S(w(1,\bar{s}),\overline{s-1})$$
$$= t^{N_{a\bar{d}}(w^1cw^2,\pm w^3)+N_{\bar{d}\bar{c}}(w^1,w^2,\pm\bar{a}w^3)+\tau_{ad}+\tau_{\bar{d}c}}(1-t)^2;$$

otherwise, it is empty. Adding up the two sums into which $S(w,\bar{s})$ splits, we obtain

$$t^{N_{a\bar{c}}(w^1,w^2d,\pm w^3)+\tau_{ac}}(1-t).$$

This last claim rests on the easily verified facts that if $a < \bar{d} < \bar{c}$, then

$$\tau_{ad}+\tau_{\bar{d}c}=\tau_{ac}, \quad N_{a\bar{d}}(c)+N_{\bar{d}\bar{c}}(\bar{a})=N_{a\bar{c}}(d).$$

Still assuming that $c = w(q) = \bar{b}$ and $p < q$, we now show that

$$S(w,\bar{1}) = t^{N_{a\bar{c}}(w[2,q-1],w[q+1,n])+\tau'_{ac}}(1-t), \tag{7-4}$$

where

$$\tau'_{ac} := \begin{cases} 1 & \text{if } a < c \le n, \\ 0 & \text{otherwise.} \end{cases}$$

Let $w^1 := w[2,q-1]$, as before, and let $w^2 := w[q+1,n]$. The sum $S(w,\bar{1})$ splits into two sums, depending on whether $(1,\bar{1}) \notin T$ or and $(1,\bar{1}) \in T$. By (7-3), the first sum is

$$S(w,\bar{n}) = t^{N_{a\bar{c}}(w^1,w^2)+\tau_{ac}}(1-t).$$

Again by (7-3), if $c < a \le n$, then the second sum is

$$t^{N_{a\bar{a}}(w^1cw^2)}(1-t)S(w(1,\bar{1}),\bar{n}) = t^{N_{a\bar{a}}(w^1cw^2)+N_{\bar{a}\bar{c}}(w^1,w^2)+\tau_{\bar{a}c}}(1-t)^2;$$

otherwise, it is empty. Adding up the two sums into which $S(w,\bar{s})$ splits, we obtain

$$t^{N_{a\bar{c}}(w^1,w^2)+\tau'_{ac}}(1-t).$$

Assuming that $c = w(q) = \bar{b}$ and $p < q < s$, we now show that

$$S(w,s) = t^{N_{a\bar{c}}(w[2,q-1],w[q+1,s-1])+\tau'_{ac}}(1-t). \tag{7-5}$$

We use decreasing induction on $s$. As before, we let $w^1 := w[2,q-1]$, $w^2 := w[q+1,s-1]$, and $d := w(s)$. The sum $S(w,s)$ splits into two sums, depending on whether $(1,s) \notin T$ or $(1,s) \in T$. By induction, the first sum is

$$S(w,s+1) = t^{N_{a\bar{c}}(w^1,w^2d)+\tau'_{ac}}(1-t).$$

Again by induction, if $a < d < \bar{c}$, then the second sum is

$$t^{N_{ad}(w^1cw^2)}(1-t)S(w(1,s),s+1) = t^{N_{ad}(w^1cw^2)+N_{d\bar{c}}(w^1,w^2a)+\tau'_{dc}}(1-t)^2;$$

otherwise, it is empty. (In both calculations, induction works by substituting $\overline{1}$ for $n+1$ when $s=n$, and by using (7-4) in this case.) Adding up the two sums into which $S(w,s)$ splits, we obtain

$$t^{N_{a\bar{c}}(w^1,w^2)+\tau'_{ac}}(1-t).$$

This last claim rests on the easily verified fact that if $a<d<\bar{c}$, then

$$N_{ad}(c)+\tau'_{dc}=\tau'_{ac}.$$

Case (iii): $b=\bar{a}$. We need to show that

$$S(w,\, p{+}1)=t^{N_{a\bar{a}}(w[2,p])}(1-t). \tag{7-6}$$

We use decreasing induction on $p$, which starts at $p=n$; in this case $\Delta$ only contains the pair $(1,\overline{1})$, so the convention of substituting $\overline{1}$ for $n+1$ works well here too. For $p<n$, we let $d:=w(p+1)$. The sum $S(w,\, p{+}1)$ splits into two sums, depending on whether $(1,\, p{+}1)\notin T$ or $(1,\, p{+}1)\in T$. By induction, the first sum is

$$S(w,\, p{+}2)=t^{N_{a\bar{a}}(w[2,p]d)}(1-t).$$

If $a<d<\bar{a}$, then by (7-5) of case (ii), the second sum is

$$t^{N_{ad}(w[2,p])}(1-t)S\big(w(1,\, p{+}1),\, p{+}2\big)=t^{N_{ad}(w[2,p])+N_{d\bar{a}}(w[2,p])+\tau'_{da}}(1-t)^2;$$

otherwise, it is empty. Adding up the two sums into which $S(w,\, p{+}1)$ splits, we obtain the desired result.

Case (ii) (continued). Assuming that $c=w(q)=\overline{b}$ and $p<q$, we now show that

$$S(w,q)=t^{N_{a\bar{c}}(w[2,q-1])}(1-t). \tag{7-7}$$

The sum $S(w,q)$ splits into two sums, depending on whether $(1,q)\notin T$ or $(1,q)\in T$. By (7-5) of case (ii), the first sum is

$$S(w,q{+}1)=t^{N_{a\bar{c}}(w[2,q-1])+\tau'_{ac}}(1-t).$$

If $a<c\le n$, then by (7-6) of case (iii), the second sum is

$$t^{N_{ac}(w[2,q-1])}(1-t)S(w(1,q)),q{+}1)=t^{N_{ac}(w[2,q-1])+N_{c\bar{c}}(w[2,q])}(1-t)^2;$$

otherwise, it is empty. Adding up the two sums into which $S(w,q)$ splits, we obtain the desired result.

The final step in case (ii) is to prove that

$$S(w,\, p{+}1)=t^{N_{a\bar{c}}(w[2,p])}(1-t). \tag{7-8}$$

This can be done by decreasing induction on $p$, starting with $p = q-1$, which is the case proved in (7-7). The procedure is completely similar to the ones above, and, in fact, to the one for type $A$ in [Lenart 2011, Proposition 5.3]. $\square$

Consider the chain

$$\begin{aligned}\Phi := \Gamma_1(n)\dots\Gamma_n(n) = \big(&(1,\bar{1}),\\ &(1,\bar{2}),\,(2,\bar{2}),\\ &\qquad\dots\\ &(1,\bar{n}),\,(2,\bar{n}),\ \dots,\,(n-1,\bar{n})\big).\end{aligned} \tag{7-9}$$

We denote by $\Phi_{ij}$ the subchain of $\Phi$ starting with $(i,\bar{\jmath})$. Given a signed permutation $w$, recall the definition (2-1) of $\ell_+(w)$ and $\ell_-(w)$. Given $(i,j)$ with $1 \le i \le j \le n$, we also define

$$\begin{aligned}\ell_-^{ij}(w) &:= |\{(k,l) : (k,\bar{l}) \in \Phi \setminus \Phi_{ij}\,,\ w(k) > \overline{w(l)}\}|,\\ \overline{\ell}_-^{ij} &:= \ell_-(w) - \ell_-^{ij}(w).\end{aligned} \tag{7-10}$$

**Proposition 7.3.** *Fix $(i,j)$ with $1 \le i \le j \le n$ and a signed permutation $w$ in $B_n$. We have*

$$\sum_{T:(w,T)\in\mathscr{A}(\Phi_{ij})} t^{\frac{1}{2}(\ell(w)+\ell(wT)-|T|)}(1-t)^{|T|} = t^{\ell_+(w)+\ell_-^{ij}(w)}. \tag{7-11}$$

*In particular, if this sum is over $(w,T) \in \mathscr{A}(\Phi)$, then the right side is $t^{\ell_+(w)}$.*

*Proof.* Let us denote the sum in the left side of (7-11) by $S(w,i,j)$, and the corresponding sum over $(w,T) \in \mathscr{A}(\Phi_{ij} \setminus \{(i,\bar{\jmath})\})$ by $S'(w,i,j)$. We view the chain $\Phi$ as a total order on the pairs $(i,\bar{\jmath})$, with $(1,\bar{1})$ being the smallest pair. With this in mind, we use decreasing induction on pairs $(i,\bar{\jmath})$. Given such a pair, if $w(i) < \overline{w(j)}$, then the induction step is clear, so assume the contrary. We can now split $S(w,i,j)$ into two sums, depending on whether $(i,\bar{\jmath}) \notin T$ or $(i,\bar{\jmath}) \in T$. By induction, the first sum is

$$S'(w,i,j) = t^{1+\ell_+(w)+\ell_-^{ij}(w)}.$$

By induction and Lemma 7.1, the second sum is

$$\begin{aligned}&(1-t)t^{\frac{1}{2}(\ell(w)-\ell(w(i,\bar{\jmath}))-1)}S'(w(i,\bar{\jmath}),i,j)\\ &\qquad\qquad = (1-t)t^{\frac{1}{2}(\ell(w)-\ell(w(i,\bar{\jmath}))-1)+\ell_+(w(i,\bar{\jmath}))+\ell_-^{ij}(w(i,\bar{\jmath}))}.\end{aligned}$$

The induction step is completed once we show that

$$\ell_+(w) + \ell_-^{ij}(w) = \tfrac{1}{2}\,(\ell(w) - \ell(w(i,\bar{\jmath})) - 1) + \ell_+(w(i,\bar{\jmath})) + \ell_-^{ij}(w(i,\bar{\jmath})).$$

This equality can be rewritten as

$$\Delta\ell_+(w) + \Delta\ell_-^{ij}(w) - 1 = \Delta\overline{\ell}_-^{ij}(w),$$

where $\Delta\ell_+(w) := \ell_+(w) - \ell_+(w(i,\bar{j}))$, and similarly for the other two variations. To prove this, we first fix $k$ between $i$ and $j$, and analyze the contribution to the three variations of the pairs $(i,k)$ and $(k,j)$; see (2-1) and (7-10). For simplicity, let $a := w(i)$, $b := w(k)$, and $c := w(j)$, where $a > \bar{c}$. The nonzero contributions are as follows:

- the pair $(i,k)$ contributes 1 to $\Delta\ell_+(w)$ if $a > b > \bar{c}$;
- the pair $(k,j)$ contributes $-1$ to $\Delta\ell_+(w)$ if $\bar{a} < b < c$, which is equivalent to $a > \bar{b} > \bar{c}$;
- the pair $(i,k)$ contributes 1 to $\Delta\ell_-^{ij}(w)$ if $a > \bar{b} > \bar{c}$;
- the pair $(k,j)$ contributes 1 to $\Delta\overline{\ell}_-^{ij}(w)$ if $a > b > \bar{c}$.

The second and third contributions cancel out, whereas the first and fourth are equal. The analysis is completely similar if $k < i$ or $k > j$. The pair $(i,j)$ only contributes 1 to $\Delta\ell_-^{ij}(w)$. As far as the pairs $(i,i)$ and $(j,j)$ are concerned, the contribution of the first one to $\Delta\ell_-^{ij}(w)$ and of the second one to $\Delta\overline{\ell}_-^{ij}(w)$ are both equal to $\sigma_{ac}$; see (2-6). □

*Proof of Theorem 4.6(iii).* Fix a filling $\sigma$ in $\mathcal{F}(\lambda)$ with columns $C_{ij}$ and $C'_{ij}$, as explained in Section 2. Recall the chain $\Phi := \Gamma_1(n)\ldots\Gamma_n(n) = \Gamma_{11}\ldots\Gamma_{1n}$ in (7-9). By splitting the $\lambda$-chain $\Gamma$ into the tail $\Phi$ and its complement, and by using Lemma 7.1, the sum in the left side of (4-4) can be rewritten as

$$\begin{aligned}\sum_{(w,T)\in f^{-1}(\sigma)} & t^{\frac{1}{2}(\ell(w)+\ell(wT)-|T|)}(1-t)^{|T|} \\ &= \Bigg(\sum_{\substack{(w,T)\in f^{-1}(\sigma)\\ T_{11}=\cdots=T_{1n}=\varnothing}} t^{(1/2)(\ell(w)-\ell(wT)-|T|)}(1-t)^{|T|}\Bigg) \\ &\qquad\times\Bigg(\sum_{T:(C_{11},T)\in\mathcal{A}(\Phi)} t^{\frac{1}{2}(\ell(C_{11})+\ell(C_{11}T)-|T|)}(1-t)^{|T|}\Bigg). \qquad (7\text{-}12)\end{aligned}$$

Here the column $C_{11}$, which has height $n$, is viewed as a signed permutation in $B_n$. By Proposition 7.3, the second bracket is $t^{\ell_+(C_{11})}$.

To evaluate the first bracket, we reverse all chains. Let us start by recalling the construction (4-1) of the $\lambda$-chain $\Gamma$, and in particular the order in which the subchains $\Gamma_{ij}$ and $\Gamma'_{ij}$ are concatenated (including the conventions in Section 2 related to $\Gamma_{i,j+1}$ and $\Gamma'_{i,j+1}$). We denote by $\Gamma^r_{ij}$ and $(\Gamma'_{ij})^r$ the corresponding reverse chains. Also recall that we defined $\mathcal{A}^r(\cdot)$ as in (3-4) except that we imposed an increasing chain condition and $w \in W$. We consider pairs $(w_{ij}, S_{ij})$ in $\mathscr{A}^r(\Gamma_{ij}^r)$ and $(w'_{ij}, S'_{ij})$ in $\mathscr{A}^r((\Gamma'_{ij})^r)$, where $w_{ij}$ and $w'_{ij}$ are defined by

$$w_{ij} := C_{11} S'_{1,\lambda'_1} \dots S_{i,j+1}, \quad w'_{ij} := C_{11} S'_{1,\lambda'_1} \dots S_{i,j+1},$$

where the concatenation order for $S_{ij}$ and $S'_{ij}$ comes from that for $\Gamma_{ij}$ and $\Gamma'_{ij}$; in particular, $w'_{1,\lambda'_1} = C_{11}$. Given this notation, we define the sum

$$\Sigma_{ij} := \sum_{\substack{S_{ij} : (w_{ij}, S_{ij}) \in \mathscr{A}^r(\Gamma_{ij}^r) \\ w_{ij} S_{ij}[1,\lambda'_i] = C_{ij}}} t^{\frac{1}{2}(\ell(w_{ij} S_{ij}) - \ell(w_{ij}) - |S_{ij}|)} (1-t)^{|S_{ij}|},$$

and the similar sum $\Sigma'_{ij}$. We can now evaluate the first bracket in the right side of (7-12):

$$\sum_{\substack{(w,T) \in f^{-1}(\sigma) \\ T_{11} = \cdots = T_{1n} = \varnothing}} t^{\frac{1}{2}(\ell(w) - \ell(wT) - |T|)} (1-t)^{|T|} = \Sigma_{\lambda_1,1} \dots \Sigma'_{ij} \dots \Sigma_{ij} \dots \Sigma'_{1,\lambda'_1}.$$

In fact, we first write the sum in the left side as an iterated sum, which factors in the way shown above because $\Sigma_{ij}$ only depends on $w_{ij}[1, \lambda'_i] = C_{i,j+1}[1, \lambda'_i]$ (rather than the whole $w_{ij}$), by Proposition 7.2.

We conclude the proof by calculating the sum $\Sigma_{ij}$, the calculation for $\Sigma'_{ij}$ being similar but simpler. For simplicity, let $d := \lambda'_i$, $w = w_{ij}$, $C := C_{i,j+1}[1, \lambda'_i]$, and $D := C_{ij}$. Assume that $C$ differs from $D' := D(r_1, \bar{\jmath}) \dots (r_p, \bar{\jmath})$ with $1 \le r_1 < \cdots < r_p < j$ (possibly $p = 0$) only in position $j$. Let $\Gamma_{ij}^r = \Delta\Delta'$, where

$$\begin{aligned}
\Delta := \big( &(j, d+1), (j, d+2), \ \dots, \ (j, n), \\
&(j, \bar{\jmath}), \\
&(j, \bar{n}), \ (j, \overline{n-1}), \ \dots, \ (j, \overline{d+1}) \big), \\
\Delta' := \big( &(j-1, \bar{\jmath}), \ \dots, \ (2, \bar{\jmath}), (1, \bar{\jmath}) \big).
\end{aligned}$$

Correspondingly, the chains $S_{ij}$ split into a head $S$, which can vary, and a fixed tail

$$S' := ((r_p, \bar{\jmath}), \dots, (r_1, \bar{\jmath})).$$

We have

$$\Sigma_{ij} = t^e (1-t)^p \sum_{\substack{S : (w,S) \in \mathscr{A}^r(\Delta) \\ wS(j) = D'(j)}} t^{\frac{1}{2}(\ell(wS) - \ell(w) - |S|)} (1-t)^{|S|},$$

where $e := \frac{1}{2}(\ell(wSS') - \ell(wS) - p)$. By Proposition 7.2, the sum in the right side is

$$t^{N_{C(j), D'(j)}(D[j+1,d])} (1-t);$$

note that this sum is missing when $D' = C$, which is another possibility. The exponent $e$ is calculated based on (7-1). □

## References


[Adamczak and Lenart 2009] W. Adamczak and C. Lenart, "The alcove path model and Young tableaux", preprint, 2009, available at http://www.albany.edu/∼lenart/articles/tableaux-v3.pdf.

[Ardonne and Kedem 2007] E. Ardonne and R. Kedem, "Fusion products of Kirillov–Reshetikhin modules and fermionic multiplicity formulas", *J. Algebra* **308**:1 (2007), 270–294. MR 2008a:17028 Zbl 1122.17014

[Assaf 2010] S. Assaf, "Dual equivalence graphs, I: A combinatorial proof of LLT and Macdonald positivity", preprint, 2010. arXiv 1005.3759

[Gaussent and Littelmann 2005] S. Gaussent and P. Littelmann, "LS galleries, the path model, and MV cycles", *Duke Math. J.* **127**:1 (2005), 35–88. MR 2006c:20092 Zbl 1078.22007

[Grojnowski and Haiman 2007] I. Grojnowski and M. Haiman, "Affine Hecke algebras and positivity of LLT and Macdonald polynomials", preprint, 2007, available at http://math.berkeley.edu/∼mhaiman/ftp/llt-positivity/new-version.pdf.

[Haglund et al. 2005] J. Haglund, M. Haiman, and N. Loehr, "A combinatorial formula for Macdonald polynomials", *J. Amer. Math. Soc.* **18**:3 (2005), 735–761. MR 2006g:05223a Zbl 1061.05101

[Kashiwara and Nakashima 1994] M. Kashiwara and T. Nakashima, "Crystal graphs for representations of the $q$-analogue of classical Lie algebras", *J. Algebra* **165**:2 (1994), 295–345. MR 95c:17025 Zbl 0808.17005

[Kato 1982] S.-i. Kato, "Spherical functions and a $q$-analogue of Kostant's weight multiplicity formula", *Invent. Math.* **66**:3 (1982), 461–468. MR 84b:22030 Zbl 0498.17005

[Lascoux and Schützenberger 1979] A. Lascoux and M.-P. Schützenberger, "Croissance des polynômes de Foulkes–Green", *C. R. Acad. Sci. Paris Sér. A-B* **288**:2 (1979), A95–A98. MR 80d:20052 Zbl 0398.20052

[Lecouvey and Shimozono 2007] C. Lecouvey and M. Shimozono, "Lusztig's $q$-analogue of weight multiplicity and one-dimensional sums for affine root systems", *Adv. Math.* **208**:1 (2007), 438–466. MR 2008d:17022 Zbl 0498.17005

[Lenart 2009] C. Lenart, "On combinatorial formulas for Macdonald polynomials", *Adv. Math.* **220**:1 (2009), 324–340. MR 2009h:05211 Zbl 05376877

[Lenart 2011] C. Lenart, "Hall–Littlewood polynomials, alcove walks, and fillings of Young diagrams", *Discrete Math.* **311**:4 (2011), 258–275.

[Lenart and Postnikov 2007] C. Lenart and A. Postnikov, "Affine Weyl groups in $K$-theory and representation theory", *Int. Math. Res. Not.* **2007**:12 (2007), Art. ID rnm038, 65. MR 2008j:14105 Zbl 1137.14037

[Lenart and Postnikov 2008] C. Lenart and A. Postnikov, "A combinatorial model for crystals of Kac–Moody algebras", *Trans. Amer. Math. Soc.* **360**:8 (2008), 4349–4381. MR 2010a:17037 Zbl 05308768

[Littelmann 1994] P. Littelmann, "A Littlewood–Richardson rule for symmetrizable Kac–Moody algebras", *Invent. Math.* **116**:1-3 (1994), 329–346. MR 95f:17023 Zbl 0805.17019

[Littelmann 1995] P. Littelmann, "Paths and root operators in representation theory", *Ann. of Math.* (2) **142**:3 (1995), 499–525. MR 96m:17011 Zbl 0858.17023

[Littlewood 1961] D. E. Littlewood, "On certain symmetric functions", *Proc. London Math. Soc.* (3) **11** (1961), 485–498. MR 24 #A173 Zbl 0099.25102

[Lusztig 1983] G. Lusztig, "Singularities, character formulas, and a $q$-analog of weight multiplicities", pp. 208–229 in *Analysis and topology on singular spaces, II, III* (Luminy, 1981), Astérisque **101**, Soc. Math. France, Paris, 1983. MR 85m:17005 Zbl 0561.22013

[Macdonald 1971] I. G. Macdonald, *Spherical functions on a group of $p$-adic type*, Publications of the Ramanujan Institute **2**, Ramanujan Institute, 1971. MR 55 #8261 Zbl 0302.43018

[Macdonald 1992] I. G. Macdonald, "Schur functions: theme and variations", pp. 5–39 in *Séminaire Lotharingien de Combinatoire* (Saint-Nabor, 1992), edited by J. Zeng, Publ. Inst. Rech. Math. Av. **498**, Univ. Louis Pasteur, Strasbourg, 1992. MR 95m:05245 Zbl 0889.05073

[Macdonald 2000] I. G. Macdonald, "Orthogonal polynomials associated with root systems", Séminaire Lotharingien de Combinatoire B45a, 2000, available at http://www.emis.de/journals/SLC/wpapers/s45macdonald.pdf. MR 2002a:33021 Zbl 1032.33010

[Nelsen and Ram 2003] K. Nelsen and A. Ram, "Kostka–Foulkes polynomials and Macdonald spherical functions", pp. 325–370 in *Surveys in combinatorics* (Bangor, 2003), edited by C. D. Wensley, London Math. Soc. Lecture Note Ser. **307**, Cambridge Univ. Press, Cambridge, 2003. MR 2004h:05126 Zbl 1036.05049

[Ram 2006] A. Ram, "Alcove walks, Hecke algebras, spherical functions, crystals and column strict tableaux", *Pure Appl. Math. Q.* **2**:4 (2006), 963–1013. MR 2008a:20013 Zbl 1127.20005

[Ram and Yip 2011] A. Ram and M. Yip, "A combinatorial formula for Macdonald polynomials", *Adv. Math.* **226**:1 (2011), 309–331.

[Schwer 2006] C. Schwer, "Galleries, Hall–Littlewood polynomials, and structure constants of the spherical Hecke algebra", *Int. Math. Res. Not.* **2006** (2006), Art. ID 75395, 31. MR 2007m:22018 Zbl 1121.05121

[Stembridge 2005] J. Stembridge, "Kostka–Foulkes polynomials of general type", Lecture notes for the Generalized Kostka Polynomials Workshop at the American Institute of Mathematics, Palo Alto, CA, 2005, available at http://www.math.lsa.umich.edu/~jrs/papers/kostka.ps.gz.

Communicated by Georgia Benkart
Received 2009-07-16 Revised 2010-07-11 Accepted 2010-10-13

lenart@albany.edu

*Department of Mathematics and Statistics,*
*State University of New York at Albany,*
*1400 Washington Avenue, Albany, NY 12222, United States*
http://www.albany.edu/~lenart

ALGEBRA AND NUMBER THEORY 4:7(2010)

# On exponentials of exponential generating series

Roland Bacher

After identification of the algebra of exponential generating series with the shuffle algebra of ordinary formal power series, the exponential map

$$\exp_! : X\mathbb{K}[[X]] \longrightarrow 1 + X\mathbb{K}[[X]]$$

for the associated Lie group with multiplication given by the shuffle product is well-defined over an arbitrary field $\mathbb{K}$ by a result going back to Hurwitz. The main result of this paper states that $\exp_!$ and its reciprocal map $\log_!$ induce a group isomorphism between the subgroup of rational, respectively algebraic series of the additive group $X\mathbb{K}[[X]]$ and the subgroup of rational, respectively algebraic series in the group $1 + X\mathbb{K}[[X]]$ endowed with the shuffle product, if the field $\mathbb{K}$ is a subfield of the algebraically closed field $\overline{\mathbb{F}}_p$ of characteristic $p$.

## 1. Introduction

The equality

$$\left(\sum_{n=0}^{\infty} \alpha_n \frac{X^n}{n!}\right)\left(\sum_{n=0}^{\infty} \beta_n \frac{X^n}{n!}\right) = \sum_{n=0}^{\infty}\sum_{m=0}^{n} \binom{n+m}{n} \alpha_n \beta_m \frac{X^{n+m}}{(n+m)!} \tag{1}$$

shows that we can define an algebra structure on the vector space

$$\mathscr{E}(\mathbb{K}) = \left\{ \sum_{n=0}^{\infty} \alpha_n \frac{X^n}{n!} \;\middle|\; \alpha_0, \alpha_1, \ldots \in \mathbb{K} \right\}$$

of formal exponential generating series with coefficients $\alpha_0, \alpha_1, \ldots$ in an arbitrary field or ring $\mathbb{K}$. For the sake of simplicity, we will work only over fields. The expression $\alpha_n/n!$ should be considered formally, since the numerical value of $n!$ is zero over a field of positive characteristic $p \le n$.

*MSC2000:* primary 11B85; secondary 11B73, 11E08, 11E76, 22E65.

*Keywords:* Bell numbers, exponential function, shuffle product, formal power series, divided powers, rational series, algebraic series, homogeneous form, automaton sequence.

Motivation for this work is given by the fact that the formula (1) allows us to define the shuffle product

$$\sum_{n=0}^{\infty} \gamma_n X^n = \Big(\sum_{n=0}^{\infty} \alpha_n X^n\Big) \sqcup\!\sqcup \Big(\sum_{n=0}^{\infty} \beta_n X^n\Big)$$

of two formal power series $\sum_{n=0}^{\infty} \alpha_n X^n$ and $\sum_{n=0}^{\infty} \beta_n X^n$ by setting

$$\gamma_n = \sum_{k=0}^{n} \binom{n}{k} \alpha_k \beta_{n-k}. \tag{2}$$

The definition of the shuffle product arises in the theory of divided powers; see, for example, [Berthelot and Ogus 1978, Definition 3.1]. The main properties needed in this paper are already in [Hurwitz 1899]. I have the impression that the main results of the present paper, given by Theorems 1.1 and 1.3, do not fit very well into the theory of divided powers: they are based on an interplay between ordinary power series (used for defining rationality and algebraicity) and exponential power series (used for defining an analogue of the exponential map in positive characteristic). A special instance of this exponential map is a standard ingredient for divided powers [Berthelot and Ogus 1978, Appendix A, Proposition A1], but ordinary formal power series do not seem to play a significant role there.

Definition (2) is also a particular case of a shuffle product defined more generally for formal power series in several noncommuting variables. The associated shuffle algebras arise, for example, in the study of free Lie algebras [Reutenauer 1993], Hopf algebras and polyzetas [Zagier 1994; Cartier 2002], formal languages [Berstel and Reutenauer 1988], etc.

I became interested in this subject through the study of the properties of the algebra of recurrence matrices, a subset of sequences of matrices displaying a kind of self-similarity structure used in [Bacher 2006; 2008] for studying reductions of the Pascal triangle modulo suitable Dirichlet characters. Such recurrence matrices are closely related to automata groups and complex dynamical systems; see, for example, [Nekrashevych 2005] for details. Over a finite field, they can be identified with rational formal power series in several noncommuting variables (the underlying algebras are however very different) and it is thus natural to investigate properties of other possible products preserving these sets. The main results of this paper, Theorems 1.1 and 1.3 (and their effective analogues, Theorems 1.5 and 1.6), deal with properties of the shuffle product for formal power series in one variable that have gone unnoticed in the existing literature, as far as I am aware.

We denote by

$$\mathfrak{m}_{\mathscr{E}} = \Big\{ \sum_{n=1}^{\infty} \alpha_n \frac{X^n}{n!} \;\Big|\; \alpha_1, \alpha_2, \dots \in \mathbb{K} \Big\} \subset \mathscr{E}(\mathbb{K})$$

the maximal ideal of the local algebra $\mathscr{E}(\mathbb{K})$. A straightforward computation, already known to Hurwitz [1899], shows that $a^n/n!$ is always well-defined for $a \in \mathfrak{m}_{\mathscr{E}}$. Endowing $\mathbb{K}$ with the discrete topology and $\mathscr{E}(\mathbb{K})$ with the topology given by coefficientwise convergence, the functions

$$\exp(a) = \sum_{n=0}^{\infty} \frac{a^n}{n!} \quad \text{and} \quad \log(1+a) = -\sum_{n=1}^{\infty} \frac{(-a)^n}{n}$$

are always defined for $a \in \mathfrak{m}_{\mathscr{E}}$.

Switching back to ordinary generating series

$$A = \sum_{n=1}^{\infty} \alpha_n X^n, \quad B = \sum_{n=1}^{\infty} \beta_n X^n \in \mathfrak{m},$$

contained in the maximal ideal $\mathfrak{m} = X\mathbb{K}[[X]]$, of (ordinary) formal power series, we write

$$\exp_! A = 1 + B$$

if

$$\exp\left(\sum_{n=1}^{\infty} \alpha_n \frac{X^n}{n!}\right) = 1 + \sum_{n=1}^{\infty} \beta_n \frac{X^n}{n!}.$$

It is easy to see that $\exp_!$ defines a one-to-one map between $\mathfrak{m}$ and $1+\mathfrak{m}$ with reciprocal map

$$1 + B \mapsto A = \log_!(1+B).$$

It satisfies

$$\exp_!(A+B) = \exp_! A \sqcup\!\sqcup \exp_! B$$

for all $A, B \in \mathfrak{m}$, where the shuffle product

$$\left(\sum_{n=0}^{\infty} \alpha_n X^n\right) \sqcup\!\sqcup \left(\sum_{n=0}^{\infty} \beta_n X^n\right) = \sum_{n,m=0}^{\infty} \binom{n+m}{n} \alpha_n \beta_m X^{n+m}$$

corresponds to the ordinary product of the associated exponential generating series. The map $\exp_!$ defines thus an isomorphism between the additive group $(\mathfrak{m}, +)$ and the *special shuffle group* $(1+\mathfrak{m}, \sqcup\!\sqcup)$ with group-law given by the shuffle product. It coincides with the familiar exponential map from the Lie algebra $\mathfrak{m}$ into the special shuffle group, considered as an infinite-dimensional Lie group.

It follows from [Fliess 1974] that rational, respectively algebraic elements form a subgroup in $(1+\mathfrak{m}, \sqcup\!\sqcup)$ if one works over a subfield of $\overline{\mathbb{F}}_p$. It is thus natural to consider the corresponding subgroups (under the reciprocal map $\log_!$ of the Lie exponential $\exp_! : \mathfrak{m} \mapsto 1+\mathfrak{m}$) in the isomorphic additive group $(\mathfrak{m}, +)$ forming the Lie algebra of $(1+\mathfrak{m}, \sqcup\!\sqcup)$. The answer, which is the main result of this paper, is surprisingly simple: the corresponding subgroup is exactly the subgroup of all rational, respectively algebraic elements in the additive group $\mathfrak{m}$. We have thus:

**Theorem 1.1.** *Let $\mathbb{K}$ be a subfield of the algebraically closed field $\overline{\mathbb{F}}_p$ of positive characteristic $p$. Given a series $A \in \mathfrak{m} = X\mathbb{K}[[X]]$, the following two assertions are equivalent*:

- *$A$ is rational.*
- $\exp_!$ *$A$ is rational.*

Theorem 1.1 fails in characteristic zero: the series

$$\log_!(1-X) = -\sum_{n=1}^{\infty}(n-1)!\,X^n$$

is obviously transcendental. (This series also shows that Theorem 1.3 does not hold in characteristic zero.)

**Example 1.2.** The Bell numbers $B_0, B_1, B_2, \ldots$ (see [Comtet 1970, pp. 45–46] or [Stanley 1999, Example 5.2.4]) are the natural integers defined by

$$\sum_{n=0}^{\infty} B_n \frac{x^n}{n!} = e^{e^x-1},$$

and have combinatorial interpretations.

Since $e^x - 1$ is the exponential generating series of the sequence $0, 1, 1, \ldots$, we have $\sum_{n=0}^{\infty} B_n x^n = \exp_!(x/(1-x))$ for the ordinary generating series

$$\sum_{n=0}^{\infty} B_n x^n = 1 + x + 2x^2 + 5x^3 + 15x^4 + 52x^5 + 203x^6 + 877x^7 + 4140x^8 + \cdots$$

of the Bell numbers.

The reduction of $\sum_{n=0}^{\infty} B_n x^n$ modulo a prime $p$ is thus always a rational element of $\mathbb{F}_p[[x]]$. A few such reductions are

$$\frac{1}{1+x+x^2} \pmod 2, \qquad \frac{1+x+x^2}{1-x^2-x^3} \pmod 3, \qquad \frac{1+x+2x^2-x^4}{1-x^4-x^5} \pmod 5.$$

**Theorem 1.3.** *Let $\mathbb{K}$ be a subfield of the algebraically closed field $\overline{\mathbb{F}}_p$ of positive characteristic $p$. Given a series $A \in \mathfrak{m} = X\mathbb{K}[[X]]$, the following two assertions are equivalent*:

- *$A$ is algebraic.*
- $\exp_!$ *$A$ is algebraic.*

Theorems 1.1 and 1.3 are the main results of this paper and can be restated as follows.

**Corollary 1.4.** *Over a subfield $\mathbb{K} \subset \overline{\mathbb{F}}_p$, the group isomorphism*

$$\exp_! : (\mathfrak{m}, +) \longrightarrow (1+\mathfrak{m}, \sqcup\!\sqcup)$$

*restricts to an isomorphism between the subgroups of rational elements in* $(\mathfrak{m}, +)$ *and in* $(1+\mathfrak{m}, \sqcup\!\sqcup)$.

*It restricts also to an isomorphism between the subgroups of algebraic elements in* $(\mathfrak{m}, +)$ *and in* $(1+\mathfrak{m}, \sqcup\!\sqcup)$.

*In particular, the subgroup of rational, respectively algebraic elements in the shuffle group* $(1+\mathfrak{m}, \sqcup\!\sqcup)$ *is a Lie-group whose Lie algebra* (*over* $\mathbb{K} \subset \overline{\mathbb{F}}_p$) *is given by the additive subgroup of all rational, respectively algebraic elements in* $(\mathfrak{m}, +)$.

Theorems 1.1 and 1.3 can be made more precise as follows.

Given a rational series $A \in \mathbb{K}[[X]]$ represented by a reduced fraction $f/g$, where $f, g$ with $g \neq 0$ are two coprime polynomials of degrees $\deg f$ and $\deg g$, we set $\|A\| = \max(1+\deg f, \deg g)$; see also Proposition 2.1 for a well-known equivalent description of $\|A\|$.

**Theorem 1.5.** *We have*

$$\|\exp_! A\| \leq p^{q^{\|A\|}} \quad \textit{and} \quad \|\log_!(1+A)\| \leq 1 + \|1+A\|^p$$

*for a rational series $A$ in $\mathfrak{m} \subset \overline{\mathbb{F}}_p[[X]]$ having all its coefficients in a finite subfield $\mathbb{F}_q \subset \overline{\mathbb{F}}_p$ containing $q = p^e$ elements.*

The bounds for $\log_!$ (and the analogous bounds in the algebraic case) can be improved; see Proposition 7.1.

Theorem 1.5 could be called an effective version of Theorem 1.1: given a rational series represented by $f/g \in \mathfrak{m}$ with $f, g \in \overline{\mathbb{F}}_p[X]$, Theorem 1.1 ensures the existence of polynomials $u, v$ such that $\exp_!(f/g) = u/v$. Theorem 1.5 shows that $u$ and $v$ are of degree at most $p^{q^{\|f/g\|}}$. They can thus be recovered as suitable Padé approximants from the series development of $\exp_!(f/g)$ up to order $2p^{q^{\|f/g\|}}$. Experimentally, the number $\|\exp_! A\|$ is generally much smaller.

Since the bounds for $\log_!$ are better than for $\exp_!$, the determination of the rational series $B = \exp_! A$ with $A \in \mathfrak{m}$ rational is best done as follows: start by “guessing” the rational series $B$ and check (or improve the guess for $B$ in case of failure) that $A = \log_!(B)$ using the bounds for $\log_!$.

Given a prime $p$ and a formal power series $C = \sum_{n=0}^{\infty} \gamma_n X^n$ in $\mathbb{K}[[X]]$ with coefficients in a subfield $\mathbb{K}$ of $\overline{\mathbb{F}}_p$, we define for $f \in \mathbb{N}$, $k \in \mathbb{N}$, $k < p^f$ the series

$$C_{k,f} = \sum_{n=0}^{\infty} \gamma_{k+np^f} X^n.$$

The vector space $\mathcal{K}(C) = \mathbb{K}C + \sum_{k,f} \mathbb{K}C_{k,f}$ spanned by $C$ and by all series of the form $C_{k,f}$, $k \in \{0, \dots, p^f - 1\}$, $f \in \{1, 2, \dots\}$ is called the *$p$-kernel* of $C$. We denote its dimension by $\kappa(C) = \dim \mathcal{K}(C)$.

Algebraic series in $\mathbb{K}[[X]]$ for $\mathbb{K}$ a subfield of $\overline{\mathbb{F}}_p$ are characterized by a theorem of Christol [Allouche and Shallit 2003, Theorem 12.2.5] stating that a series $C$ in $\overline{\mathbb{F}}_p[[X]]$ is algebraic if and only if its $p$-kernel $\mathcal{K}(C)$ is of finite dimension $\kappa(C) < \infty$. We have $\kappa(A + B) \le \kappa(A) + \kappa(B)$, and an algebraic series $A \in \overline{\mathbb{F}}_p[[X]]$ has a minimal polynomial of degree at most $p^{\kappa(A)}$ with respect to $A$.

**Theorem 1.6.** *We have*

$$\kappa(\exp_! A) \le q^{\kappa(A)-1} p^{q^{\kappa(A)}} \quad \textit{and} \quad \kappa(\log_!(1+A)) \le 1 + 4(\kappa(1+A))^p$$

*for a nonzero algebraic series $A$ in $\mathfrak{m} \subset \overline{\mathbb{F}}_p[[X]]$ having all its coefficients in a finite subfield $\mathbb{F}_q \subset \overline{\mathbb{F}}_p$ containing $q = p^e$ elements.*

Considerations similar to those made after Theorem 1.5 are valid and Theorem 1.6 can be turned into an algorithmically effective version of Theorem 1.3.

A map $\mu : \mathcal{V} \longrightarrow \mathcal{W}$ between two $\mathbb{K}$-vector spaces is a homogeneous form of degree $d$ if $l \circ \mu : \mathcal{V} \longrightarrow \mathbb{K}$ is homogeneous of degree $d$ (given by a homogeneous polynomial of degree $d$ with respect to coordinates) for all linear forms $l : \mathcal{W} \longrightarrow \mathbb{K}$.

A useful ingredient for proving Theorems 1.1, 1.3 and their effective versions is the following characterization of $\log_!$:

**Proposition 1.7.** *Over a field $\mathbb{K} \subset \overline{\mathbb{F}}_p$, the application* $\log_! : 1 + \mathfrak{m} \longrightarrow \mathfrak{m}$ *extends to a homogeneous form of degree $p$ from $\mathbb{K}[[X]]$ into* $\mathfrak{m}$.

**Example 1.8.** In characteristic 2, we have

$$\log_!\left(\sum_{n=0}^{\infty} \alpha_n X^n\right) = \sum_{n=0}^{\infty} \alpha_{2^n}^2 X^{2^{n+1}} + \sum_{0 \le i < j} \binom{i+j}{i} \alpha_i \alpha_j X^{i+j},$$

for $\sum_{n=0}^{\infty} \alpha_n X^n$ in $1 + X\overline{\mathbb{F}}_2[[X]]$.

**Remark 1.9.** Defining $f_!$ as

$$f_!\left(\sum_{n=1}^{\infty} \alpha_n X^n\right) = \sum_{n=1}^{\infty} \beta_n X^n$$

if

$$f\left(\sum_{n=1}^{\infty} \alpha_n \frac{X^n}{n!}\right) = \sum_{n=1}^{\infty} \beta_n \frac{X^n}{n!},$$

Theorems 1.1, 1.3, 1.5 and 1.6 have analogues for the functions $\sin_!$ and $\tan_!$ (and for their reciprocal functions $\arcsin_!$ and $\arctan_!$).

The rest of the paper is organized as follows. In Sections 2–6, we recall a few definitions and well-known facts that are essentially standard knowledge in the theory of divided powers; see [Berthelot and Ogus 1978] or the original work [Roby 1963; 1965]. Section 7 contains the proofs for all results mentioned above.

In a second part, starting at Section 8, we generalize Theorems 1.1 and 1.5 to formal power series in several noncommuting variables.

## 2. Rational and algebraic elements in $\mathbb{K}[[X]]$

This section recalls a few well-known facts concerning rational and algebraic elements in the algebra $\mathbb{K}[[X]]$ of formal power series.

We denote by $\tau : \mathbb{K}[[X]] \longrightarrow \mathbb{K}[[X]]$ the shift operator

$$\tau\left(\sum_{n=0}^{\infty} \alpha_n X^n\right) = \sum_{n=0}^{\infty} \alpha_{n+1} X^n$$

acting on formal power series. The following well-known result characterizes rational series:

**Proposition 2.1.** *A formal power series $A = \sum_{n=0}^{\infty} \alpha_n X^n$ of $\mathbb{K}[[X]]$ is rational if and only if the series $A, \tau(A), \tau^2(A), \dots, \tau^k(A) = \sum_{n=0}^{\infty} \alpha_{n+k} X^n, \dots$ span a finite-dimensional vector space in $\mathbb{K}[[X]]$.*

*More precisely, the vector space spanned by $A, \tau(A), \tau^2(A), \dots, \tau^i(A), \dots$ has dimension $\|A\| = \max(1 + \deg f, \deg g)$ if $f/g$, with $f, g \in \mathbb{K}[X]$, is a reduced expression of a rational series $A$.*

The function $A \mapsto \|A\|$ satisfies the inequality

$$\|A + B\| \leq \|A\| + \|B\|$$

for rational series $A, B$ in $\mathbb{K}[[X]]$. As a particular case, we have

$$\|A\| - 1 \leq \|1{+}A\| \leq \|A\| + 1.$$

Given a prime $p$ and a formal power series $C = \sum_{n=0}^{\infty} \gamma_n X^n$ in $\overline{\mathbb{F}}_p[[X]]$, we denote by $\kappa(C) \in \mathbb{N} \cup \{\infty\}$ the dimension of its $p$-kernel

$$\mathcal{K}(C) = \mathbb{K}C + \sum_{f,k} \overline{\mathbb{F}}_p C_{k,f},$$

spanned by $C$ and all series of the form

$$C_{k,f} = \sum_{n=0}^{\infty} \gamma_{k+np^f} X^n,$$

with $k \in \mathbb{N}$ such that $k < p^f$ for $f \in \{1, 2, \dots\}$.

Algebraic series of $\mathbb{K}[[X]]$, for $\mathbb{K}$ a subfield of the algebraic closure $\overline{\mathbb{F}}_p$ of finite prime characteristic $p$, are characterized by the following theorem of Christol [1979] (see also [Allouche and Shallit 2003, Theorem 12.2.5]):

**Theorem 2.2.** *A formal power series $C = \sum_{n=0}^{\infty} \gamma_n X^n$ of $\overline{\mathbb{F}}_p[[X]]$ is algebraic if and only if the dimension $\kappa(C) = \dim \mathcal{K}(C)$ of its $p$-kernel $\mathcal{K}(C)$ is finite.*

Finiteness of $\kappa(C)$ amounts to recognizability of $C$, which has the following well-known consequence.

**Corollary 2.3.** *An algebraic series of $\overline{\mathbb{F}}_p[[X]]$ has all its coefficients in a finite subfield of $\overline{\mathbb{F}}_p$.*

**Proposition 2.4.** *Let $C = \sum_{n=0}^{\infty} \gamma_n X^n$ be an algebraic series with coefficients in a subfield $\mathbb{K} \subset \overline{\mathbb{F}}_p$.*

(i) $\mathcal{K}(\tau(C)) \subset \mathcal{K}(C) + \tau(\mathcal{K}(C))$, *which implies* $\kappa(\tau(C)) \leq 2\kappa(C)$.

(ii) $\mathcal{K}(C) \subset \mathbb{K} + \mathcal{K}(\tau(C)) + X\mathcal{K}(\tau(C))$, *which implies* $\kappa(C) \leq 1 + 2\kappa(\tau(C))$.

*Proof.* Assertion (i) follows from an iterated application of the easy computations

$$(\tau(C))_{k,1} = C_{k+1,1},$$

if $0 \leq k < p-1$ and

$$(\tau(C))_{p-1,1} = \tau(C_{0,1}).$$

The proof of assertion (ii) is similar. □

## 3. The shuffle algebra

This section recalls mostly well-known results concerning shuffle products of elements in the set $\mathbb{K}[[X]]$ of formal power series over a commutative field $\mathbb{K}$, which is arbitrary unless specified otherwise.

The *shuffle product*

$$A \sqcup\!\sqcup\, B = C = \sum_{n=0}^{\infty} \gamma_n X^n$$

of $A = \sum_{n=0}^{\infty} \alpha_n X^n$ and $B = \sum_{n=0}^{\infty} \beta_n X^n$ is defined by

$$\gamma_n = \sum_{k=0}^{n} \binom{n}{k} \alpha_k \beta_{n-k},$$

and corresponds to the usual product $ab = c$ of the associated exponential generating series

$$a = \sum_{n=0}^{\infty} \alpha_n \frac{X^n}{n!}, \quad b = \sum_{n=0}^{\infty} \beta_n \frac{X^n}{n!}, \quad c = \sum_{n=0}^{\infty} \gamma_n \frac{X^n}{n!}.$$

The *shuffle algebra* is the algebra $(\mathbb{K}[[X]], ⧢)$ obtained by endowing the vector space $\mathbb{K}[[X]]$ of ordinary generating series with the shuffle product. By construction, the shuffle algebra is isomorphic to the algebra $\mathscr{E}(\mathbb{K})$ of exponential generating series. In characteristic zero, the trivial identity

$$\sum_{n=0}^{\infty}\alpha_n X^n=\sum_{n=0}^{\infty}(n!\alpha_n)\frac{X^n}{n!}$$

gives an isomorphism between the usual algebra $\mathbb{K}[[X]]$ of ordinary generating series and the shuffle algebra $(\mathbb{K}[[X]], ⧢)$.

The identity

$$\Big(\sum_{n\geq 0}\lambda^n X^n\Big) ⧢ \Big(\sum_{n\geq 0}\mu^n X^n\Big)=\sum_{n\geq 0}(\lambda+\mu)^n X^n,$$

equivalent to $e^{\lambda X}e^{\mu X}=e^{(\lambda+\mu)X}$, implies that the convergence radius of the shuffle product of two complex series with strictly positive convergence radii $\rho_1, \rho_2$ is at least the harmonic mean $1/(1/\rho_1+1/\rho_2)$ of $\rho_1$ and $\rho_2$.

**Proposition 3.1.** *The shift operator*

$$\tau\Big(\sum_{n=0}^{\infty}\alpha_n X^n\Big)=\sum_{n=0}^{\infty}\alpha_{n+1}X^n$$

*acts as a derivation on the shuffle algebra.*

*Proof.* The map $\tau$ is clearly linear. The computation

$$\begin{aligned}\tau\Big(\sum_{i,j\geq 0}\binom{i+j}{i}\alpha_i\beta_j X^{i+j}\Big)&=\sum_{\substack{i,j\geq 0\\ i+j\geq 1}}\binom{i+j}{i}\alpha_i\beta_j X^{i+j-1}\\&=\sum_{\substack{i,j\geq 0\\ i+j\geq 1}}\Big(\binom{i+j-1}{i-1}+\binom{i+j-1}{j-1}\Big)\alpha_i\beta_j X^{i+j-1}\end{aligned}$$

shows that $\tau$ satisfies the Leibniz rule $\tau(A ⧢ B)=\tau(A) ⧢ B+A ⧢ \tau(B)$. □

Proposition 3.1 is trivial and well-known in characteristic zero: the usual derivation $d/dX$ acts obviously as the shift operator on the algebra $\mathscr{E}(\mathbb{K})$ of exponential generating series over a field of characteristic zero.

The following two results seem to be due to Fliess [1974, Proposition 6].

**Proposition 3.2.** *Shuffle products of rational power series are rational.*

*More precisely, we have*

$$\|A ⧢ B\|\leq\|A\|\,\|B\|,$$

*for two rational series $A, B$ in $\mathbb{K}[[X]]$.*

*Proof.* Proposition 3.1 implies $\tau^n(A \sqcup\!\sqcup B) = \sum_{k=0}^{n} \binom{n}{k} \tau^k(A) \sqcup\!\sqcup \tau^{n-k}(B)$. The series $\tau^n\ (A \sqcup\!\sqcup B)$ belongs thus to the vector space spanned by shuffle products with factors in the vector spaces $\sum_{n\geq 0} \mathbb{K}\tau^n(A)$ and $\sum_{n\geq 0} \mathbb{K}\tau^n(B)$. This implies the inequality. Proposition 2.1 ends the proof. □

**Proposition 3.3.** *Shuffle products of algebraic series in* $\overline{\mathbb{F}}_p[[X]]$ *are algebraic.*

*More precisely, we have*

$$\kappa(A \sqcup\!\sqcup B) \leq \kappa(A)\kappa(B).$$

*Proof.* Denoting by $C_{k,f}$ the series

$$C_{k,f} = \sum_{n=0}^{\infty} \gamma_{k+np^f} X^n$$

associated to a series $C = \sum_{n=0}^{\infty} \gamma_n X^n$, as in Section 2, and by $\kappa(C)$ the dimension of the vector space $\mathcal{K}(C) = \mathbb{K}C + \sum_{k,f} \overline{\mathbb{F}}_p C_{k,f}$, Lucas's identity [1878]

$$\binom{n}{k} \equiv \prod_{i\geq 0} \binom{\nu_i}{\kappa_i} \pmod{p},$$

for $n = \sum_{i\geq 0} \nu_i p^i$ and $k = \sum_{i\geq 0} \kappa_i p^i$ with $\nu_i, \kappa_i \in \{0, \ldots, p-1\}$, implies

$$(A \sqcup\!\sqcup B)_{k,1} = \sum_{i=0}^{k} \binom{k}{i} A_{i,1} \sqcup\!\sqcup B_{k-i,1},$$

for $k = 0, \ldots, p-1$. Iteration of this formula shows that $(A \sqcup\!\sqcup B)_{k,f}$ (for arbitrary $k, f \in \mathbb{N}$ such that $k < p^f$) belongs to the vector space spanned by shuffle products with factors in the vector spaces $\mathcal{K}(A)$ and $\mathcal{K}(B)$ of dimension $\kappa(A)$ and $\kappa(B)$.

Christol's Theorem (Theorem 2.2) ends the proof. □

**Remark 3.4.** Given a subfield $\mathbb{K}$ of $\overline{\mathbb{F}}_p$, let $\mathcal{A} \subset \mathbb{K}[[X]]$ denote a vector space of finite dimension $a = \dim \mathcal{A}$ containing the $p$-kernel $\mathcal{K}(A)$ of every element $A \in \mathcal{A}$.

We consider an element $B = A_1 \sqcup\!\sqcup A_2 \sqcup\!\sqcup \cdots \sqcup\!\sqcup A_k$ given by the shuffle product of $k$ series $A_1, \ldots, A_k \in \mathcal{A}$. Expressing all elements $A_1, A_2, \ldots$ as linear combinations of elements in a fixed basis of $\mathcal{A}$ and using commutativity of the shuffle product, the proof of Proposition 3.3 shows that the inequality $\kappa(B) \leq \kappa(A_1)\kappa(A_2)\cdots \leq a^k = (\dim \mathcal{A})^k$ can be improved to

$$\kappa(B) \leq \binom{k+a-1}{a-1},$$

where the binomial coefficient gives the dimension of the vector space of homogeneous polynomials of degree $k$ in $a$ (commuting) variables $X_1, X_2, \ldots, X_a$.

## 4. The special shuffle group

We call the group of units of the shuffle algebra $(\mathbb{K}[[X]], ⧢)$ the *shuffle group*. Its elements are given by the set $\mathbb{K}^* + X\mathbb{K}[[X]]$ underlying the multiplicative unit group. The shuffle group is the direct product of the unit group $\mathbb{K}^*$ of $\mathbb{K}$ with the *special shuffle group* $(1 + X\mathbb{K}[[X]], ⧢)$.

The inverse in the shuffle group of $1 - A \in (1 + X\mathbb{K}[[X]], ⧢)$ is given by

$$\sum_{n=0}^{\infty} A^{⧢^n} = 1 + A + A ⧢ A + A ⧢ A ⧢ A + \cdots,$$

where $A^{⧢^0} = 1$ and $A^{⧢^{n+1}} = A ⧢ A^{⧢^n}$ for $n \geq 1$.

The trivial identity $X ⧢ X^n = \binom{n+1}{1} X^{n+1} = (n+1)X^{n+1} \in \mathbb{K}[[X]]$ implies $(1 - X) ⧢ \left(\sum_{n=0}^{\infty} n! X^n\right) = 1$. Invertible rational (analytical) power series have thus generally a transcendental (nonanalytical) shuffle inverse over the complex numbers.

**Proposition 4.1.** *The special shuffle group* $(1 + X\mathbb{K}[[X]], ⧢)$ *is isomorphic to an infinite-dimensional* $\mathbb{F}_p$*-vector space if the field* $\mathbb{K}$ *is of positive characteristic* $p$.

Proposition 4.1 shows that $(1 + X\mathbb{K}[[X]], ⧢)$ is not isomorphic to the multiplicative group structure on $1 + X\mathbb{K}[[X]]$ if $\mathbb{K}$ is of positive characteristic.

*Proof of Proposition 4.1.* It follows from the fact that $\exp_!$ is a group isomorphism between the $\mathbb{F}_p$-vector space $\mathfrak{m}$ and the special shuffle group. □

Proposition 4.1 follows also as a special case from Proposition 8.1. This yields a different proof, which is not based on properties of $\exp_!$.

**Remark 4.2.** One can show that a rational fraction $A \in 1 + X\mathbb{C}[[X]]$ has a rational inverse for the shuffle product if and only if $A = 1/(1 - \lambda X)$ with $\lambda \in \mathbb{C}$. (Compute $A ⧢ B = 1$ using the decomposition into simple fractions of the rational series $A, B$.)

## 5. The exponential and the logarithm for exponential generating functions

Hurwitz showed that $1/(k!)a^k$ is well-defined for $a \in \mathfrak{m}_{\mathscr{E}}$ with coefficients in an arbitrary field or commutative ring [Hurwitz 1899, Satz 1]. We give a different proof of this fact, that implies that $\exp_!$ and $\log_!$ are well-defined over fields of positive characteristic.

**Proposition 5.1.** *For all natural numbers* $j, k \geq 1$, *the set* $\{1, \ldots, jk\}$ *can be partitioned in exactly*

$$\frac{(jk)!}{(j!)^k k!} \tag{3}$$

*different ways into $k$ unordered disjoint subsets of $j$ elements. In particular, the rational number in (3) is an integer for all natural numbers $j, k$ such that $j \geq 1$.*

*Proof.* The multinomial coefficient $(jk)!/(j!)^k$ counts the number of ways of partitioning $\{1, \ldots, jk\}$ into an ordered sequence of $k$ disjoint subsets containing all $j$ elements. Dividing by $k!$ removes the order on these $k$ subsets.

This proves that the formula defines an integer for all $j, k \geq 1$, and integrality obviously also holds for $k = 0$ and $j \geq 1$. □

**Remark 5.2.** A slightly different proof of Proposition 5.1 follows from the observation that $(jk)!/((j!)^k k!)$ is the index in the symmetric group over $jk$ elements of the subgroup formed by all permutations stabilizing a partition of the set $\{1, \ldots, jk\}$ into $k$ disjoint subsets of $j$ elements.

A different proof is given by the formula

$$\frac{(jk)!}{(j!)^k k!} = \prod_{n=1}^{k} \binom{nj-1}{j-1},$$

easily shown using induction on $k$; see [Berthelot and Ogus 1978, Section 3] (which contains a small misprint).

**Proposition 5.3.** *For any natural integer $k \in \mathbb{N}$, there exist polynomials $P_{k,n} \in \mathbb{N}[\alpha_1, \ldots, \alpha_n]$ such that*

$$\frac{1}{k!}\left(\sum_{n=1}^{\infty} \alpha_n \frac{X^n}{n!}\right)^k = \sum_{n=0}^{\infty} P_{k,n}(\alpha_1, \alpha_2, \ldots, \alpha_n)\frac{X^n}{n!}.$$

*Proof.* The contribution of a monomial

$$\alpha_1^{j_1}\alpha_2^{j_2}\ldots\alpha_s^{j_s}\frac{X^{\sum_{i=1}^s ij_i}}{\left(\sum_{i=1}^s ij_i\right)!},$$

with $j_1 + j_2 + \cdots + j_s = k$, to $\frac{1}{k!}\left(\sum_{n=1}^{\infty} \alpha_n X^n/n!\right)^k$ is given by

$$\frac{1}{k!}\,\frac{k!}{(j_1)!(j_2)!\ldots(j_s)!}\,\frac{\left(\sum_{i=1}^s ij_i\right)!}{\prod_{i=1}^s (i!)^{j_i}} = \left(\prod_{i=1}^{s}\frac{(ij_i)!}{(i!)^{j_i}(j_i)!}\right)\frac{\left(\sum_{i=1}^s ij_i\right)!}{\prod_{i=1}^s (ij_i)!},$$

where the last expression is a product of a natural integer by Proposition 5.1 and of a multinomial coefficient. It is thus a natural integer. □

**Corollary 5.4.** *For $a = \sum_{n=1}^{\infty} \alpha_n X^n/n!$, the formulae*

$$\exp\left(\sum_{n=1}^{\infty} \alpha_n \frac{X^n}{n!}\right) = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} P_{k,n}(\alpha_1, \ldots, \alpha_n)\frac{X^n}{n!}$$

*and*

$$\log\left(1+\sum_{n=1}^{\infty}\alpha_n\frac{X^n}{n!}\right)=\sum_{k=1}^{\infty}\sum_{n=0}^{\infty}(-1)^{k+1}(k-1)!\,P_{k,n}(\alpha_1,\dots,\alpha_n)\frac{X^n}{n!}$$

*define the exponential function and the logarithm of an exponential generating series in* $a\in\mathfrak{m}_{\mathscr{E}}$ *and* $1+a\in 1+\mathfrak{m}_{\mathscr{E}}$*, respectively, over an arbitrary field* $\mathbb{K}$*. These functions are one-to-one and mutually reciprocal.*

The following result shows that the functions $\exp_!$ and $\log_!$ behave as expected under the derivation $\tau:\sum_{n=0}^{\infty}\alpha_n X^n\mapsto\sum_{n=0}^{\infty}\alpha_{n+1}X^n$ of the shuffle algebra.

**Proposition 5.5.** *For all* $A\in\mathfrak{m}=X\mathbb{K}[[X]]$ *over an arbitrary field* $\mathbb{K}$*, we have*

$$\tau(\exp_! A)=(\exp_! A)\sqcup\tau(A)$$

*and*

$$\tau(\log_!(1+A))=(1+A)^{\sqcup^{-1}}\sqcup\tau(A),$$

*where* $(1+A)^{\sqcup^{-1}}$ *denotes the shuffle inverse of* $(1+A)$.

*Proof.* Proposition 3.1 implies the formal identities

$$\tau\left(\sum_{n=0}^{\infty}\frac{A^{\sqcup^n}}{n!}\right)=\sum_{n=0}^{\infty}n\frac{A^{\sqcup^{n-1}}}{n!}\sqcup\tau(A)=\left(\sum_{n=0}^{\infty}\frac{A^{\sqcup^n}}{n!}\right)\sqcup\tau(A),$$

for $A\in\mathfrak{m}$. By Proposition 5.1, this identity holds over the ring $\mathbb{Z}$ and thus over an arbitrary commutative field. This establishes the formula for $\exp_!$.

For $\log_!$ we get similarly

$$\tau\left(-\sum_{n=1}^{\infty}\frac{(-A)^{\sqcup^n}}{n}\right)=\sum_{n=1}^{\infty}n\frac{(-A)^{\sqcup^{n-1}}}{n}\sqcup\tau(A)=\left(\sum_{n=0}^{\infty}(-A)^{\sqcup^n}\right)\sqcup\tau(A),$$

which implies the result, by Proposition 5.1 and by the trivial identity

$$(1+A)^{\sqcup^{-1}}=\sum_{n=0}^{\infty}(-A)^{\sqcup^n},$$

for the shuffle inverse $(1+A)^{\sqcup^{-1}}$ of $1+A\in 1+\mathfrak{m}$. □

## 6. The logarithm as a $p$-homogeneous form over $\overline{\mathbb{F}}_p[[x]]$

Given a fixed prime number $p$, Proposition 4.1 implies that there exist polynomials $Q_{p,n}\in\mathbb{N}[\alpha_0,\dots,\alpha_n]$ for $n\geq 1$ such that

$$\left(\sum_{n=0}^{\infty}\alpha_n X^n\right)^{\sqcup^p}=\alpha_0^p+p\sum_{n=1}^{\infty}Q_{p,n}(\alpha_0,\dots,\alpha_n)X^n.$$

The polynomials $Q_{p,n}$ are homogeneous of degree $p$ with respect to the variables $\alpha_0,\dots,\alpha_n$, and we denote by

$$\mu_p\Big(\sum_{n=0}^{\infty}\alpha_n X^n\Big)=\sum_{n=1}^{\infty}Q_{p,n}(\alpha_0,\dots,\alpha_n)X^n$$

the $p$-homogeneous form defined by the ordinary generating series of the polynomials $Q_{p,1}, Q_{p,2},\dots$.

**Proposition 6.1.** *The restriction of $\mu_p$ to $1+\mathfrak{m}\subset\overline{\mathbb{F}}_p[[X]]$ coincides with the function $\log_!$.*

*Proof.* We have

$$\tau(\mu_p(1+A))=(1+A)^{\sqcup\sqcup^{p-1}}\sqcup\sqcup\,\tau(1+A)$$

for $A$ in $\mathfrak{m}$, where $\tau\big(\sum_{n=0}^{\infty}\alpha_n X^n\big)=\sum_{n=0}^{\infty}\alpha_{n+1}X^n$ is the shift operator of Proposition 3.1. This identity defines the restriction of the $p$-homogeneous form $\mu_p$ to $1+\mathfrak{m}$. Proposition 5.5 and the identity

$$(1+A)^{\sqcup\sqcup^{p-1}}\sqcup\sqcup\,(1+A)=1$$

show that the function $\log_!$ satisfies the same equation

$$\tau(\log_!(1+A))=(1+A)^{\sqcup\sqcup^{p-1}}\sqcup\sqcup\,\tau(1+A).$$

Since both series $\mu_p(1+A)$ and $\log_!(1+A)$ are without constant term, the equality $\tau(\mu_p(1+A))=\tau(\log_!(1+A))$ implies $\mu_p(1+A)=\log_!(1+A)$. □

## 7. Proofs

**Proposition 7.1.** *If $A$ in $X\overline{\mathbb{F}}_p[[X]]$ is rational (respectively algebraic), then the formal power series $\log_!(1+A)$ is rational (respectively algebraic).*

*More precisely,*

$$\|\log_!(1+A)\|\le 1+\binom{p+\|1+A\|-1}{p}\le 1+\|1+A\|^p$$

*for $A$ rational in $\mathfrak{m}=X\overline{\mathbb{F}}_p[[X]]$, and*

$$\kappa(\log_!(1+A))\le 1+4\kappa(A)\binom{p+\kappa(1+A)-2}{p-1}\le 1+4(\kappa(1+A))^p$$

*for $A$ algebraic in $\mathfrak{m}$.*

**Proposition 7.2.** *If $A$ in $X\overline{\mathbb{F}}_p[[X]]$ is rational (respectively algebraic), then $\exp_! A$ is rational (respectively algebraic).*

*More precisely, denoting by $q=p^e$ the cardinality of a finite field $\mathbb{F}_q\subset\overline{\mathbb{F}}_p$ containing all coefficients of $A$,*

$$\|\exp_! A\|\le p^{q^{\|A\|}}$$

*for $A$ rational in $\mathfrak{m}$, and*

$$\kappa(\exp_! A) \leq q^{\kappa(A)-1} p^{q^{\kappa(A)}}$$

*for $A$ algebraic and nonzero in $\mathfrak{m}$.*

Theorems 1.1, 1.3, 1.5 and 1.6 are now simple reformulations of Propositions 7.1 and 7.2.

*Proof of Proposition 7.1.* Apply the identity $(1+A)^{\sqcup\!\sqcup p} = 1$, which follows from Proposition 4.1, to

$$\tau(\log_!(1+A)) = (1+A)^{\sqcup\!\sqcup -1} \sqcup\!\sqcup \tau(A)$$

of Proposition 5.5, to establish

$$\tau(\log_!(1+A)) = (1+A)^{\sqcup\!\sqcup p-1} \sqcup\!\sqcup \tau(A),$$

already encountered in the proof of Proposition 6.1. This shows

$$\|\tau(\log_!(1+A)\| \leq \|1+A\|^{p-1}\|\tau(A)\| \leq \|1+A\|^p$$

and implies

$$\|\log_!(1+A)\| \leq 1 + \|1+A\|^p.$$

This proves the cruder inequality in the rational case. The finer inequality follows from the fact that all $p$ factors of

$$(1+A)^{\sqcup\!\sqcup p-1} \sqcup\!\sqcup \tau(A) = \tau(\log_!(1+A))$$

belong to a common vector space of dimension $\|1+A\|$ that is closed for the shift map. The details are the same as for Remark 3.4.

For algebraic $A$ we have similarly

$$\begin{aligned}\kappa(\tau(\log_!(1+A))) &\leq (\kappa(1+A))^{p-1}\,\kappa(\tau(A)) = (\kappa(1+A))^{p-1}\,\kappa(\tau(1+A)) \\ &\leq (\kappa(1+A))^{p-1} 2\kappa(1+A) \leq 2(\kappa(1+A))^p,\end{aligned}$$

using Proposition 2.4(i). This shows

$$\kappa(\log_!(1+A)) \leq 1 + 2\kappa(\tau(\log_!(1+A))) \leq 1 + 4(\kappa(1+A))^p,$$

by Proposition 2.4(ii), and ends the proof for the cruder inequality.

The finer inequality follows from Proposition 2.4 and Remark 3.4. □

Given a vector space $\mathcal{V} \subset \mathbb{K}[[X]]$ containing $\mathbb{K}$, we denote by $\Gamma(\mathcal{V})$ the shuffle subgroup generated by all elements of $\mathcal{V} \cap (1 + X\mathbb{K}[[X]])$.

**Lemma 7.3.** *Every element of a vector space $\mathcal{V} \subset \mathbb{K}[[X]]$ containing the field $\mathbb{K}$ of constants can be written as a linear combination of elements in $\Gamma(\mathcal{V})$.*

*Proof.* We have the identity

$$A = (1 - \epsilon(A) + A) + (\epsilon(A) - 1),$$

where $\epsilon\left(\sum_{n=0}^{\infty} \alpha_n X^n\right) = \alpha_0$ is the augmentation map and where $(1 - \epsilon(A) + A)$ and the constant $(\epsilon(A) - 1)$ are both in $\mathbb{K}\Gamma(\mathcal{V})$ for $A \in \mathcal{V}$. $\square$

*Proof of Proposition 7.2 for A rational.* Corollary 2.3 shows that we can work over a finite subfield $\mathbb{K} = \mathbb{F}_q$ of $\overline{\mathbb{F}}_p$ consisting of $q = p^e$ elements.

Given a rational series $A$ in $\mathfrak{m} = X\mathbb{K}[[X]]$, we denote by $\Gamma_A$ the shuffle subgroup generated by all elements of the set

$$\left\{ \bigcup_{n=0}^{\infty} \left(\tau^n(A) + \mathbb{K}\right) \right\} \cap \{1 + X\mathbb{K}[[X]]\}.$$

This generating set of $\Gamma_A$ contains at most $q^{\|A\|}$ elements. Proposition 4.1 implies thus that $\Gamma_A$ is a finite group having at most $p^{q^{\|A\|}}$ elements. The subalgebra $\mathbb{K}[\Gamma_A] \subset \mathbb{K}[[X]]$ spanned by all elements of $\Gamma_A$ is thus of dimension $\leq p^{q^{\|A\|}}$. The identity

$$\tau(\exp_! A) = \exp_! A \sqcup\!\sqcup\, \tau(A)$$

of Proposition 5.5 and the fact that the derivation $\tau$ of $\mathbb{K}[[X]]$ restricts to a derivation of the subalgebra $\mathbb{K}[\Gamma_A]$ show the inclusion

$$\tau^n(\exp_! A) \in \exp_! A \sqcup\!\sqcup\, \mathbb{K}[\Gamma_A]$$

for all $n \in \mathbb{N}$ by Lemma 7.3. This ends the proof, since the right side is a $\mathbb{K}$-vector space of dimension at most $p^{q^{\|A\|}}$. $\square$

**Proposition 7.4.** *We have, for every prime number $p$ and for all natural integers $j, k$ such that $j \geq 1$, the identity*

$$\frac{(jk)!}{(j!)^k k!} \equiv \frac{(pjk)!}{((pj)!)^k k!} \pmod{p}.$$

*Proof.* The fraction on the right side yields the cardinality of the set $\mathcal{E}$ of all partitions of $\{1, \ldots, pjk\}$ into $k$ subsets of $pj$ elements. Consider the group $G$ generated by the $jk$ cycles of length $p$ of the form $(i, i+jk, i+2jk, \ldots, i+(p-1)jk)$ for $i = 1, \ldots, jk$. The group $G$ has $p^{jk}$ elements and acts on the set of partitions by preserving their type defined as the multiset of cardinalities of all involved parts. In particular, it acts by permutation on the set $\mathcal{E}$. A partition $P \in \mathcal{E}$ is a fixpoint for $G$ if and only if every part of $P$ is a union of $G$-orbits. Choosing a bijection between $\{1, \ldots, jk\}$ and $G$-orbits of $\{1, \ldots, pjk\}$, fixpoints of $\mathcal{E}$ are in bijection with partitions of the set $\{1, \ldots, jk\}$ into $k$ subsets of $j$ elements. The number of fixpoints of the $G$-action on $\mathcal{E}$ equals thus $(jk)!/((j!)^k k!)$. Since $G$ is a $p$-group, the cardinalities of all nontrivial $G$-orbits of $\mathscr{E}$ are strictly positive powers of $p$. This ends the proof. $\square$

**Corollary 7.5.** $\exp_!$ *and* $\log_!$ *commute with the "Frobenius substitution"*

$$\varphi\left(\sum_{n=0}^{\infty}\alpha_n X^n\right)=\sum_{n=0}^{\infty}\alpha_n X^{pn}$$

*for series in* $X\overline{\mathbb{F}}_p[[X]]$ *and* $1+X\overline{\mathbb{F}}_p[[X]]$, *respectively.*

This implies $(\exp_! A)_{0,f}=\exp_! A_{0,f}$, where $C_{k,f}=\sum_{n=0}^{\infty}\gamma_{k+np^f}X^n$ for $C=\sum_{n=0}^{\infty}\gamma_n X^n$.

**Lemma 7.6.** $(B\sqcup\!\sqcup C)_{0,1}=B_{0,1}\sqcup\!\sqcup C_{0,1}$.

*Proof.* Follows from the identity $\binom{pn}{k}\equiv 0 \pmod p$ if $k\not\equiv 0 \pmod p$. $\square$

*Proof of Proposition 7.2 for A algebraic.* We work again over a finite subfield $\mathbb{K}=\mathbb{F}_q\subset\overline{\mathbb{F}}_p$ containing all coefficients of $A$. Let $\Gamma_A$ denote the shuffle subgroup generated by all elements in

$$(\mathscr{K}(A)+\mathbb{K})\cap(1+X\mathbb{K}[[X]]),$$

where

$$\mathscr{K}(A)=\mathbb{K}A+\sum_{k,f}\mathbb{K}A_{k,f}$$

denotes the $p$-kernel of $A$. We denote by $\mathbb{K}[\Gamma_A]\subset(\mathbb{K}[[X]],\sqcup\!\sqcup)$ the shuffle subalgebra of dimension at most $p^{q^{\kappa(A)}}$ spanned by all elements of the group $\Gamma_A\subset(1+X\mathbb{K}[[X]],\sqcup\!\sqcup)$.

Using the convention $A_{0,0}=A$, we have, for $B\in\mathbb{K}[\Gamma(A)]$ and for $k$ such that $0\le k<p$,

$$\begin{aligned}(\exp_!(A_{0,f})\sqcup\!\sqcup B)_{k,1}&=(\tau^k(\exp_!(A_{0,f})\sqcup\!\sqcup B))_{0,1}\\&=\left(\sum_{j=0}^{k}\binom{k}{j}\tau^j(\exp_! A_{0,f})\sqcup\!\sqcup\tau^{k-j}(B)\right)_{0,1}\\&=\sum_{j=0}^{k}\binom{k}{j}(\tau^j(\exp_! A_{0,f}))_{0,1}\sqcup\!\sqcup B_{k-j,1},\end{aligned}$$

where the last equality is due to Lemma 7.6 (and to the equality $(\tau^k(C))_{0,1}=C_{k,1}$ for $0\le k<p$).

Proposition 5.5 gives $\tau(\exp_! A_{0,f})=(\exp_! A_{0,f})\sqcup\!\sqcup\tau(A_{0,f})$; iterating this identity shows that $\tau^j(\exp_! A_{0,f})$ is of the form $(\exp_! A_{0,f})\sqcup\!\sqcup F$, where $F$ is a linear combination of shuffle products involving at most $j$ factors of the set $\{\tau(A_{0,f}), \tau^2(A_{0,f}), \dots, \tau^j(A_{0,f})\}$. Applying Lemma 7.6, we get

$$(\tau^j(\exp_! A_{0,f}))_{0,1} = (\exp_{0,f+1}(A)) \sqcup\!\sqcup F_{0,1}.$$

An iterated application of Lemma 7.6 now shows that $F_{0,1}$ is a linear combination of shuffle products involving at most $j$ factors in $\{A_{1,f+1}, \dots, A_{j,f+1}\}$. We thus have $F_{0,1} \in \mathbb{K}[\Gamma_A]$, by Lemma 7.3, and we get the inclusion

$$(\exp_!(A_{0,f}) \sqcup\!\sqcup \mathbb{K}[\Gamma_A])_{k,1} \subset \exp_!(A_{0,f+1}) \sqcup\!\sqcup \mathbb{K}[\Gamma_A]$$

for all $f \in \mathbb{N}$ and for all $k \in \{0, \dots, p-1\}$.

Setting
$$E_A = \{\exp_! B \mid B \in \mathcal{K}(A) \cap X\mathbb{K}[[X]]\},$$
we have the inclusion

$$\mathcal{K}(\exp_! A) \subset E_A \sqcup\!\sqcup \mathbb{K}[\Gamma_A] \subset \mathbb{K}[E_A] \sqcup\!\sqcup \mathbb{K}[\Gamma_A],$$

where $\mathcal{K}(\exp_! A)$ denotes the $p$-kernel of $\exp_! A$. This implies

$$\kappa(\exp_! A) \leq \dim \mathbb{K}[E_A] \dim \mathbb{K}[\Gamma_A].$$

We suppose now $A$ that is nonzero. The vector space $\mathcal{K}(A) \cap X\mathbb{K}[[X]]$ is thus of codimension 1 in $\mathcal{K}(A)$. The image $E_A$ of $\mathcal{K}(A) \cap X\mathbb{K}[[X]]$ under the group isomorphism $\exp_! : (X\mathbb{K}[[X]], +) \mapsto (1 + X\mathbb{K}[[X]], \sqcup\!\sqcup)$ is hence a subgroup of cardinality $q^{\kappa(A)-1}$ in $(1 + X\mathbb{K}[[X]], \sqcup\!\sqcup)$. We have thus

$$\kappa(\exp_! A) \leq \dim \mathbb{K}[E_A] \dim \mathbb{K}[\Gamma_A] \leq q^{\kappa(A)-1} p^{q^{\kappa(A)}},$$

which ends the proof. □

## 8. Power series in free noncommuting variables

This and the next section recall a few basic and well-known facts concerning (rational) power series in free noncommuting variables; see, for instance, [Stanley 1999] or [Berstel and Reutenauer 1988]. Sometimes, however, we use a different terminology, motivated by [Bacher 2008].

We denote by $\mathcal{X}^*$ the free monoid on a finite set $\mathcal{X} = \{X_1, \dots, X_k\}$. We write 1 for the identity element and we use a boldface capital $\boldsymbol{X}$ for a noncommutative monomial $\boldsymbol{X} = X_{i_1} X_{i_2} \dots X_{i_l} \in \mathcal{X}^*$. We denote by

$$A = \sum_{\boldsymbol{X} \in \mathcal{X}^*} (A, \boldsymbol{X})\boldsymbol{X} \in \mathbb{K}\langle\langle X_1, \dots, X_k \rangle\rangle$$

a noncommutative formal power series, where $\mathcal{X}^* \ni \boldsymbol{X} \mapsto (A, \boldsymbol{X}) \in \mathbb{K}$ stands for the coefficient function.

We denote by $\mathfrak{m} \subset \mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$ the maximal ideal consisting of formal power series without constant coefficient, and by $\mathbb{K}^* + \mathfrak{m} = \mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle \setminus \mathfrak{m}$ the unit group of the algebra $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$ consisting of all (multiplicatively) invertible elements in $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$. The unit group is isomorphic to the direct product $\mathbb{K}^* \times (1 + \mathfrak{m})$, where $\mathbb{K}^*$ is the central subgroup consisting of nonzero constants and where $1 + \mathfrak{m}$ denotes the multiplicative subgroup given by the affine subspace formed by power series with constant coefficient 1. We have $(1-A)^{-1} = 1 + \sum_{n=1}^{\infty} A^n$ for the multiplicative inverse $(1-A)^{-1}$ of an element $1 - A \in 1 + \mathfrak{m}$.

***The shuffle algebra.*** The *shuffle product* $\boldsymbol{X} \sqcup\!\sqcup \boldsymbol{X}'$ of two noncommutative monomials $\boldsymbol{X}, \boldsymbol{X}' \in \mathscr{X}^*$ of degrees $a = \deg \boldsymbol{X}$ and $b = \deg \boldsymbol{X}'$ (for the obvious grading given by $\deg X_1 = \cdots = \deg X_k = 1$) is the sum of all $\binom{a+b}{a}$ monomials of degree $a + b$ obtained by shuffling in all possible ways the linear factors (elements of $\mathscr{X}$) involved in $\boldsymbol{X}$ with the linear factors of $\boldsymbol{X}'$. A monomial involved in $\boldsymbol{X} \sqcup\!\sqcup \boldsymbol{X}'$ can be thought of as a monomial of degree $a + b$ whose linear factors are colored by two colors with $\boldsymbol{X}$ corresponding to the product of all linear factors of the first color and $\boldsymbol{X}'$ corresponding to the product of the remaining linear factors. The shuffle product $\boldsymbol{X} \sqcup\!\sqcup \boldsymbol{X}'$ can also be recursively defined by $\boldsymbol{X} \sqcup\!\sqcup 1 = 1 \sqcup\!\sqcup \boldsymbol{X} = \boldsymbol{X}$ and

$$(\boldsymbol{X} X_s) \sqcup\!\sqcup (\boldsymbol{X}' X_t) = (\boldsymbol{X} \sqcup\!\sqcup (\boldsymbol{X}' X_t)) X_s + ((\boldsymbol{X} X_s) \sqcup\!\sqcup \boldsymbol{X}') X_t,$$

where $X_s, X_t \in \mathscr{X} = \{X_1, \ldots, X_k\}$ are monomials of degree 1.

Extending the shuffle product in the obvious way to formal power series endows the vector space $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$ with an associative and commutative algebra structure called the *shuffle algebra*. In the case of one variable $X = X_1$, we recover the definition of Section 3.

The group $\mathrm{GL}_k(\mathbb{K})$ acts on the vector space $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$ by linear substitutions. This action induces an automorphism of the multiplicative (noncommutative) algebra-structure or of the (commutative) shuffle algebra-structure underlying the vector space $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$.

Substitution of all variables $X_j$ of formal power series in $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$ by $X$ (or more generally by arbitrary not necessarily equal formal power series without constant term) yields a homomorphism of (shuffle) algebras into the commutative (shuffle) algebra $\mathbb{K}[[X]]$.

The commutative unit group (set of invertible elements for the shuffle product) of the shuffle algebra, given by the set $\mathbb{K}^* + \mathfrak{m}$, is isomorphic to the direct product $\mathbb{K}^* \times (1 + \mathfrak{m})$, where $1 + \mathfrak{m}$ is endowed with the shuffle product. The inverse of an element $1 - A \in (1 + \mathfrak{m}, \sqcup\!\sqcup)$ is given by

$$\sum_{n=0}^{\infty} A^{\sqcup\!\sqcup^n} = 1 + A + A \sqcup\!\sqcup A + A \sqcup\!\sqcup A \sqcup\!\sqcup A + \cdots .$$

The following result generalizes Proposition 4.1:

**Proposition 8.1.** *Over a field of positive characteristic* $p$, *the subgroup* $1+\mathfrak{m}$ *of the shuffle group is an infinite-dimensional* $\mathbb{F}_p$*-vector space.*

*Proof.* Contributions to a $p$-fold shuffle product $A_1 \sqcup\!\sqcup A_2 \sqcup\!\sqcup \cdots \sqcup\!\sqcup A_p$ are given by monomials with linear factors colored by $p$ colors $\{1, \ldots, p\}$ keeping track of their "origin" with coefficients given by the product of the corresponding "monochromatic" coefficients in $A_1, \ldots, A_p$. A permutation of the colors $\{1, \ldots, p\}$ (and in particular, a cyclic permutation of all colors) leaves such a contribution invariant if $A_1 = \cdots = A_p$. Coefficients of strictly positive degree in $A^{\sqcup\!\sqcup^p}$ are thus zero in characteristic $p$. □

As in the one-variable case, one can prove that

$$\frac{1}{k!}A^{\sqcup\!\sqcup^k}$$

is defined over an arbitrary field $\mathbb{K}$ for $A \in \mathfrak{m}$. Monomials contributing to $A^{\sqcup\!\sqcup^k}$ can be considered as colored by $k$ colors and the $k!$ possible color-permutations yield identical contributions.

For $A \in \mathfrak{m}$, we denote by

$$\exp_! A = \sum_{n=0}^{\infty} \frac{1}{n!} A^{\sqcup\!\sqcup^n}$$

the resulting exponential map from the Lie algebra $\mathfrak{m}$ into the infinite-dimensional commutative Lie group $(1+\mathfrak{m}, \sqcup\!\sqcup)$. As expected, its reciprocal function is

$$\log_!(1+A) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} A^{\sqcup\!\sqcup^n}.$$

In the case of a field $\mathbb{K}$ of positive characteristic $p$, the function $\log_!$ is again given by the restriction to $1+\mathfrak{m}$ of a $p$-homogeneous form $\mu_p$.

The form $\mu_p$ has all its coefficients in $\mathbb{N}$ and is again defined by the equality

$$A^{\sqcup\!\sqcup^p} = (A, 1)^p + p\mu_p(A)$$

over $\mathbb{Z}$. It can thus be defined over an arbitrary field.

## 9. Rational series

We say that a formal power series $A$ is *rational* if it belongs to the smallest subalgebra in $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$ that contains the free associative algebra $\mathbb{K}\langle X_1, \ldots, X_k\rangle$ of noncommutative polynomials and intersects the group of multiplicative units of $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$ in a subgroup.

Given a monomial $\boldsymbol{T} \in \mathscr{X}^*$, we denote by

$$\rho(\boldsymbol{T}) : \mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle \longrightarrow \mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$$

the linear application defined by

$$\rho(\boldsymbol{T})A = \sum_{\boldsymbol{X}\in\mathscr{X}^*} (A, \boldsymbol{X}\boldsymbol{T})\boldsymbol{X}$$

for $A = \sum_{\boldsymbol{X}\in\mathscr{X}^*}(A, \boldsymbol{X})\boldsymbol{X}$ in $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$. The identity

$$\rho(\boldsymbol{T})(\rho(\boldsymbol{T}')A) = \rho(\boldsymbol{T}\boldsymbol{T}')A$$

shows that we have a representation

$$\rho : \mathscr{X}^* \longrightarrow \mathrm{End}(\mathbb{K}\langle\langle\mathscr{X}\rangle\rangle)$$

of the free monoid $\mathscr{X}^*$ on $\mathscr{X}$. The *recursive closure* $\overline{A}$ of a power series $A$ is the vector space spanned by its orbit $\rho(\mathscr{X}^*)A$ under $\rho(\mathscr{X}^*)$. We call the dimension $\dim \overline{A}$ of $\overline{A}$ the *complexity* of $A$.

We call a subspace $\mathscr{A} \subset \mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$ *recursively closed* if it contains the recursive closure of all its elements.

Rational series coincide with series of finite complexity by a theorem of Schützenberger [Berstel and Reutenauer 1988, Theorem 1].

**Remark 9.1.** In the case of one variable, the complexity $\dim \overline{A}$ of a reduced nonzero rational fraction $A = f/g$ with $f \in \mathbb{K}[X]$ and $g \in 1 + X\mathbb{K}[X]$ equals $\dim \overline{A} = \max(1 + \deg f, \deg g)$.

**Remark 9.2.** The (generalized) *Hankel matrix* $H = H(A)$ of

$$A = \sum_{\boldsymbol{X}\in\mathscr{X}^*} (A, \boldsymbol{X})\boldsymbol{X} \in \mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$$

is the infinite matrix with rows and columns indexed by the free monoid $\mathscr{X}^*$ of monomials and entries $H_{\boldsymbol{X}\boldsymbol{X}'} = (A, \boldsymbol{X}\boldsymbol{X}')$. The rank of $H$ is given by the complexity $\dim \overline{A}$ of $A$, and $\overline{A}$ corresponds to the column-span of $H$.

Given subspaces $\mathscr{A}, \mathscr{B}$ of $\mathbb{K}\langle\langle\mathscr{X}\rangle\rangle$, we denote by $\mathscr{A} \sqcup\!\sqcup \mathscr{B}$ the vector space spanned by all products $A \sqcup\!\sqcup B$ with $A \in \mathscr{A}$ and $B \in \mathscr{B}$.

**Proposition 9.3.** *We have the inclusion* $\overline{A \sqcup\!\sqcup B} \subset \overline{A} \sqcup\!\sqcup \overline{B}$ *for the shuffle product* $A \sqcup\!\sqcup B$ *of* $A, B \in \mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$.

**Corollary 9.4** [Fliess 1974, Proposition 4]. *We have*

$$\dim(\overline{A \sqcup\!\sqcup B}) \leq \dim \overline{A}\, \dim \overline{B}$$

*for the shuffle product* $A \sqcup\!\sqcup B$ *of* $A, B \in \mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$. *In particular, shuffle products of rational elements in* $\mathbb{K}\langle\langle X_1, \ldots, X_k\rangle\rangle$ *are rational.*

*Proof of Proposition 9.3.* For $Y \in \overline{A}$, $Z \in \overline{B}$ and $X$ in $\{X_1, \dots, X_k\}$, the recursive definition of the shuffle product given in Section 8 shows

$$\rho(X)(Y \sqcup\!\sqcup Z) = (\rho(X)Y) \sqcup\!\sqcup Z + Y \sqcup\!\sqcup (\rho(X)Z).$$

We thus have the inclusions $\rho(X)(Y \sqcup\!\sqcup Z) \in \overline{A} \sqcup\!\sqcup Z + Y \sqcup\!\sqcup \overline{B} \subset \overline{A} \sqcup\!\sqcup \overline{B}$, which show that the vector space $\overline{A} \sqcup\!\sqcup \overline{B}$ is recursively closed. Proposition 9.3 follows now from the inclusion $A \sqcup\!\sqcup B \in \overline{A} \sqcup\!\sqcup \overline{B}$. $\square$

**Remark 9.5.** Similar arguments show that the set of rational elements in $\mathbb{K}\langle\langle X_1, \dots, X_k\rangle\rangle$ is also closed under the ordinary product (and multiplicative inversion of invertible series), Hadamard product and composition (where one considers $A \circ (B_1, \dots, B_k)$ with $A \in \mathbb{K}\langle\langle X_1, \dots, X_k\rangle\rangle$ and $B_1, \dots, B_k \in \mathfrak{m} \subset \mathbb{K}\langle\langle X_1, \dots, X_k\rangle\rangle$).

**Remark 9.6.** The shuffle inverse of a rational element in $\mathbb{K}^* + \mathfrak{m}$ is in general not rational in characteristic 0. An exception is given by geometric progressions

$$\Big(1 - \sum_{j=1}^{k} \lambda_j X_j\Big)^{-1} = \sum_{n=0}^{\infty} \Big(\sum_{j=1}^{k} \lambda_j X_j\Big)^{n},$$

since we have

$$\frac{1}{1-\sum_{j=1}^{k} \lambda_j X_j} \sqcup\!\sqcup \frac{1}{1-\sum_{j=1}^{k} \mu_j X_j} = \frac{1}{1-\sum_{j=1}^{k} (\lambda_j+\mu_j) X_j}$$

corresponding to $e^{\lambda X} e^{\mu X} = e^{(\lambda+\mu)X}$ in the one-variable case.

There are no other such elements in $1 + \mathfrak{m} \subset \mathbb{K}[[X]]$; see Remark 4.2. I do not know whether the maximal rational shuffle subgroup of $1 + \mathfrak{m} \subset \mathbb{C}\langle\langle X_1, \dots, X_k\rangle\rangle$ (defined as the set of all rational elements in $1 + \mathfrak{m}$ with rational inverse for the shuffle product) contains other elements if $k \geq 2$.

**Remark 9.7.** Any finite set of rational elements in $\mathbb{K}\langle\langle X_1, \dots, X_k\rangle\rangle$ over a field $\mathbb{K}$ of positive characteristic is included in a unique minimal finite-dimensional recursively closed subspace of $\mathbb{K}\langle\langle X_1, \dots, X_k\rangle\rangle$ that intersects the shuffle group $(\mathbb{K}^* + \mathfrak{m}, \sqcup\!\sqcup)$ in a subgroup.

## 10. Main result for generating series in noncommuting variables

The following statement is our main result in a noncommutative framework.

**Theorem 10.1.** *Let $\mathbb{K}$ be a subfield of $\overline{\mathbb{F}}_p$. Given a noncommutative formal power series $A \in \mathfrak{m} \subset \mathbb{K}\langle\langle \mathscr{X}\rangle\rangle$, the following two assertions are equivalent*:

- *$A$ is rational.*
- $\exp_!$ *$A$ is rational.*

*More precisely, we have for a rational series $A$ in $\mathfrak{m}$ the inequalities*

$$\dim \overline{\log_!(1+A)} \le 1 + \bigl(\dim \overline{1+A}\bigr)^p$$

*and*

$$\dim \overline{\exp_! A} \le p^{q^{\dim \bar{A}}},$$

*where $q = p^e$ is the cardinality of a finite field $\mathbb{F}_q$ containing all coefficients of $A$.*

*Proof.* The identity

$$\log_!(1+A) = \sum_{X \in \mathscr{X}} \Bigl((1+A)^{\sqcup\!\sqcup^{p-1}} \sqcup\!\sqcup\, \rho(X)A\Bigr)X$$

and Corollary 9.4 show

$$\dim \overline{\log_!(1+A)} \le 1 + \bigl(\dim \overline{1+A}\bigr)^p.$$

For the opposite direction we denote by $\mathbb{K} = \mathbb{F}_q$ a finite subfield of $\overline{\mathbb{F}}_p$ containing all coefficients of $A$. We have

$$\overline{\exp_! A} \subset \exp_! A \sqcup\!\sqcup\, \mathbb{K}[\Gamma(A)],$$

where $\mathbb{K}[\Gamma(A)]$ is the shuffle subalgebra of dimension at most $p^{q^{\dim \bar{A}}}$ spanned by all elements of the group $\Gamma$ generated by all elements of the form

$$(\bar{A} + \mathbb{K}) \cap (1 + \mathfrak{m}).$$

This implies the inequality $\dim \overline{\exp_! A} \le p^{q^{\dim \bar{A}}}$, which ends the proof. $\square$

## Acknowledgements

I thank Jean-Paul Allouche, Michel Brion, Alexei Panchishkin, Tanguy Rivoal, Boris Venkov and an anonymous referee for their interest and helpful remarks.

## References

[Allouche and Shallit 2003] J.-P. Allouche and J. Shallit, *Automatic sequences*, Cambridge University Press, 2003. MR 2004k:11028 Zbl 1086.11015

[Bacher 2006] R. Bacher, "La suite de Thue–Morse et la catégorie Rec", *C. R. Math. Acad. Sci. Paris* **342**:3 (2006), 161–164. MR 2006g:11050 Zbl 1148.11012

[Bacher 2008] R. Bacher, "Determinants related to Dirichlet characters modulo 2, 4 and 8 of binomial coefficients and the algebra of recurrence matrices", *Internat. J. Algebra Comput.* **18**:3 (2008), 535–566. MR 2009c:15006 Zbl 1190.15005

[Berstel and Reutenauer 1988] J. Berstel and C. Reutenauer, *Rational series and their languages*, EATCS Monographs on Theoretical Computer Science **12**, Springer, Berlin, 1988. Revised edition available at http://tagh.de/tom/wp-content/uploads/berstelreutenauer2008.pdf. MR 90e:68054 Zbl 0668.68005

[Berthelot and Ogus 1978] P. Berthelot and A. Ogus, *Notes on crystalline cohomology*, Princeton University Press, 1978. MR 58 #10908 Zbl 0383.14010

[Cartier 2002] P. Cartier, "Fonctions polylogarithmes, nombres polyzêtas et groupes pro-unipotents", pp. 137–173 in *Séminaire Bourbaki* 2000/2001 (Exposé 885), Astérisque **282**, Soc. Mat. de France, Paris, 2002. MR 2004i:19005

[Christol 1979] G. Christol, "Ensembles presque periodiques $k$-reconnaissables", *Theoret. Comput. Sci.* **9**:1 (1979), 141–145. MR 80e:68141 Zbl 0402.68044

[Comtet 1970] L. Comtet, *Analyse combinatoire, II*, Le Mathématicien **5**, Presses Universitaires de France, Paris, 1970. MR 41 #6697 Zbl 0221.05002

[Fliess 1974] M. Fliess, "Sur divers produits de séries formelles", *Bull. Soc. Math. France* **102** (1974), 181–191. MR 50 #7125 Zbl 0313.13021

[Hurwitz 1899] A. Hurwitz, "Über die Entwickelungscoefficienten der lemniskatischen Functionen", *Math. Ann.* **51** (1899), 196–226. JFM 29.0385.02

[Lucas 1878] E. Lucas, "Sur les congruences des nombres eulériens et les coefficients différentiels des fonctions trigonométriques suivant un module premier", *Bull. Soc. Math. France* **6** (1878), 49–54. MR 1503769 JFM 10.0139.04

[Nekrashevych 2005] V. Nekrashevych, *Self-similar groups*, Mathematical Surveys and Monographs **117**, American Mathematical Society, Providence, RI, 2005. MR 2006e:20047 Zbl 1087.20032

[Reutenauer 1993] C. Reutenauer, *Free Lie algebras*, London Mathematical Society Monographs. New Series **7**, Oxford University Press, 1993. MR 94j:17002 Zbl 0798.17001

[Roby 1963] N. Roby, "Lois polynomes et lois formelles en théorie des modules", *Ann. Sci. École Norm. Sup.* (3) **80** (1963), 213–348. MR 28 #5091 Zbl 0117.02302

[Roby 1965] N. Roby, "Les algèbres à puissances divisées", *Bull. Sci. Math.* (2) **89** (1965), 75–91. MR 33 #1348 Zbl 0145.04503

[Stanley 1999] R. P. Stanley, *Enumerative combinatorics, II*, Cambridge Studies in Advanced Mathematics **62**, Cambridge University Press, 1999. MR 2000k:05026 Zbl 0928.05001

[Zagier 1994] D. Zagier, "Values of zeta functions and their applications", pp. 497–512 in *First European Congress of Mathematics* (Paris, 1992), vol. 2, edited by A. Joseph et al., Progr. Math. **120**, Birkhäuser, Basel, 1994. MR 96k:11110 Zbl 0822.11001

Communicated by Hendrik W. Lenstra
Received 2009-08-24 Revised 2010-07-13 Accepted 2010-10-17

Roland.Bacher@ujf-grenoble.fr *Université Grenoble I, CNRS UMR 5582, Institut Fourier, 100, rue des Maths, Boîte Postale 74, 38402 St. Martin d'Hères, France*

ALGEBRA AND NUMBER THEORY 4:7(2010)

# On families of $\varphi, \Gamma$-modules

Kiran Kedlaya and Ruochuan Liu

Berger and Colmez (2008) formulated a theory of families of overconvergent étale $(\varphi, \Gamma)$-modules associated to families of $p$-adic Galois representations over $p$-adic Banach algebras. In contrast with the classical theory of $(\varphi, \Gamma)$-modules, the functor they obtain is not an equivalence of categories. In this paper, we prove that when the base is an affinoid space, every family of (overconvergent) étale $(\varphi, \Gamma)$-modules can locally be converted into a family of $p$-adic representations in a unique manner, providing the "local" equivalence. There is a global mod $p$ obstruction related to the moduli of residual representations.

## Introduction

Berger and Colmez [2008] introduced a theory of families of overconvergent étale $(\varphi, \Gamma)$-modules associated to families of $p$-adic Galois representations over $p$-adic Banach algebras. The $p$-adic families of local Galois representations emerging from number theory are usually over rigid analytic spaces. So we are mainly interested in the case where the bases are reduced affinoid spaces. However, even in this case, the functor of Berger and Colmez is far from an equivalence of categories, in contrast with the classical theory of $(\varphi, \Gamma)$-modules. This was first noticed by Chenevier [Berger and Colmez 2008, remarque 4.2.10]: if the base is the $p$-adic unit circle $M(\mathbb{Q}\langle X, Y\rangle/(XY-1))$, then it is easy to see that the free rank-1 overconvergent étale $(\varphi, \Gamma)$-module $D$ with a basis $e$ such that $\varphi(e) = Ye$ and $\gamma(e) = e$ for $\gamma \in \Gamma$ does not come from a family of $p$-adic representations over the same base.

On the other hand, in his proof of the density of crystalline representations, Colmez [2008, proposition 5.2] proved that for certain families of rank-2 triangular étale $(\varphi, \Gamma)$-modules, one can locally convert such a family into a family of $p$-adic representations using his theory of *Espaces Vectoriels de dimension finie* (it is clear that we can also convert Chenevier's example locally). Moreover, Colmez

While writing this paper, Kedlaya was supported by NSF CAREER grant DMS-0545904, the MIT NEC Research Support Fund, and the MIT Cecil and Ida Green Career Development Professorship, while Liu was a postdoctoral fellow of Foundation Sciences Mathématiques de Paris.

*MSC2000:* primary 11F80; secondary 11S20.

*Keywords:* $p$-adic Galois representations, $(\varphi, \Gamma)$-modules.

remarked [2008, remarque 5.3(2)]: *On aurait pu aussi utiliser une version «en famille» des théorèmes à la Dieudonné–Manin de Kedlaya. Il y a d'ailleurs une concordance assez frappante entre ce que permettent de démontrer ces théorèmes de Kedlaya et la théorie des Espaces Vectoriels de dimension finie.*

Unfortunately, as noted in [Liu 2008], there is no family version of Kedlaya's slope filtrations theorem in general, because the slope polygons of families of Frobenius modules are not necessarily locally constant. Nonetheless, one may still ask to what extent one can convert a globally étale family of $(\varphi, \Gamma)$-modules back into a Galois representation. As Chenevier's example shows, this cannot be done in general over an affinoid base. The best one can hope for in general is the following theorem, which extends a result of Dee [2001]. (In the statement, the distinction between a $(\varphi, \Gamma)$-module and a family of $(\varphi, \Gamma)$-modules is that the former is defined as a module over a ring, whereas the latter is defined as a coherent sheaf over a rigid analytic space.)

**Theorem 0.1.** *Let $S$ be a Banach algebra over $\mathbb{Q}_p$ of the form $R \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$, where $R$ is a complete noetherian local domain of characteristic 0 whose residue field is finite over $\mathbb{F}_p$. Then for any finite extension $K$ of $\mathbb{Q}_p$, the categories of $S$-linear representations of $G_K$, of étale $(\varphi, \Gamma)$-modules over $\mathbf{B}^\dagger_K \widehat{\otimes}_{\mathbb{Q}_p} S$, and of families of étale $(\varphi, \Gamma)$-modules over $\mathbf{B}^\dagger_{\mathrm{rig},K} \widehat{\otimes}_{\mathbb{Q}_p} S$ are all equivalent.*

For instance, if $S$ is an affinoid algebra and we are given an étale $(\varphi, \Gamma)$-module over $\mathbf{B}^\dagger_K \widehat{\otimes}_{\mathbb{Q}_p} S$, we recover a linear representation over each residue disc of $S$ (and every affinoid subdomain of such a disc), but these representations may not glue. This is what happens in Chenevier's example, because the mod $p$ representations cannot be uniformly trivialized. In fact, the obstruction to converting a $(\varphi, \Gamma)$-module back into a representation exists purely at the residual level; it suggests a concrete realization of the somewhat murky notion of "moduli of residual (local) representations".

By combining Theorem 0.1 with the results of [Liu 2008], we obtain a result that applies when only one fiber of the $(\varphi, \Gamma)$-module is known to be étale. (Beware that the natural analogue of this statement in which the rigid analytic point $x$ is replaced by a Berkovich point is trivially false.)

**Theorem 0.2.** *Let $S$ be an affinoid algebra over $\mathbb{Q}$, and let $M_S$ be a family of $(\varphi, \Gamma)$-modules over $\mathbf{B}^\dagger_{\mathrm{rig},K} \widehat{\otimes}_{\mathbb{Q}_p} S$. If $M_x$ is étale for some $x \in M(S)$, then there exists an affinoid neighborhood $M(B)$ of $x$ and a $B$-linear representation $V_B$ of $G_K$ whose associated $(\varphi, \Gamma)$-module is isomorphic to $M_S \widehat{\otimes}_S B$. Moreover, $V_B$ is unique for this property.*

To prove the Fontaine–Colmez theorem, Berger [2008] constructed a morphism from the category of filtered $(\varphi, N)$-modules to the category of $(\varphi, \Gamma)$-modules. It should be possible to generalize Berger's construction to families of filtered $(\varphi, N)$-modules; upon doing so, one would get a family version of the Fontaine–Colmez theorem by Theorem 0.2. That is, one would know that a weakly admissible family of filtered $(\varphi, N)$-modules over an affinoid base (with trivial $\varphi$-action on the base) becomes admissible in a neighborhood of each rigid analytic point.

## 1. Rings of $p$-adic Hodge theory

We begin by introducing some of the rings used in $p$-adic Hodge theory. This is solely to fix notation; we do not attempt to expose the constructions in any detail. For that, see for instance [Berger 2004]. Here, whenever a ring is defined whose notation includes a boldface $\mathbf{A}$, the same notation with $\mathbf{A}$ replaced by $\mathbf{B}$ will indicate the result of inverting $p$.

Let $\mathbb{C}_p$ be a completed algebraic closure of $\mathbb{Q}$, with valuation subring $\mathbb{O}_{\mathbb{C}_p}$ and $p$-adic valuation $v_p$ normalized with $v_p(p)=1$. Let $\overline{v}_p : \mathbb{O}_{\mathbb{C}_p}/(p) \to [0,1) \cup \{+\infty\}$ be the semivaluation obtained by truncation. Define $\widetilde{\mathbf{E}}^+$ to be the ring of sequences $(x_n)_{n=0}^{\infty}$ in $\mathbb{O}_{\mathbb{C}_p}/(p)$ such that $x_{n+1}^p = x_n$ for all $n$. Define a function

$$v_{\mathbf{E}} : \widetilde{\mathbf{E}}^+ \to [0, +\infty]$$

by sending the zero sequence to $+\infty$, and sending each nonzero sequence $(x_n)$ to the common value of $p^n \overline{v}_p(x_n)$ for all $n$ with $x_n \neq 0$. This gives a valuation under which $\widetilde{\mathbf{E}}^+$ is complete. Moreover, if we put $\widetilde{\mathbf{E}} = \mathrm{Frac}(\widetilde{\mathbf{E}}^+)$, and let $\epsilon = (\epsilon_n)$ be an element of $\widetilde{\mathbf{E}}^+$ with $\epsilon_0 = 1$ and $\epsilon_1 \neq 1$, then $\widetilde{\mathbf{E}}$ is a completed algebraic closure of $\mathbb{F}_p((\epsilon - 1))$.

Let $\widetilde{\mathbf{A}}$ be the $p$-typical Witt ring $W(\widetilde{\mathbf{E}})$, which is the unique complete discrete valuation ring with maximal ideal $(p)$ and residue field $\widetilde{\mathbf{E}}$. For each positive integer $n$, $W(\widetilde{\mathbf{E}})/p^n W(\widetilde{\mathbf{E}})$ inherits a topology from the valuation topology on $\widetilde{\mathbf{E}}$, under which it is complete. We call the inverse limit of these the *weak topology* on $\widetilde{\mathbf{A}}$. We similarly obtain a weak topology on $\widetilde{\mathbf{B}}$.

For any $n \geq 0$, we let $\mu_{p^n}$ denote the set of $p^n$-th roots of unity in $\overline{\mathbb{Q}}_p$, and let $\mu_{p^\infty} = \bigcup_{n \geq 0} \mu_{p^n}$. For $K$ a finite extension of $\mathbb{Q}$, let

$$K_\infty = K(\mu_{p^\infty}), \quad H_K = \mathrm{Gal}(\overline{K}/K_\infty), \quad \Gamma = \Gamma_K = \mathrm{Gal}(K_\infty/K), \quad K_0' = \mathbb{Q}_p^{\mathrm{ur}} \cap K_\infty.$$

Put $\pi = [\epsilon] - 1$, where brackets denote the Teichmüller lift. Using the completeness of $\widetilde{\mathbf{A}}$ for the weak topology, we may embed $\mathbb{Z}_p((\pi))$ into $\widetilde{\mathbf{A}}$. Let $\mathbf{A}$ be the $p$-adic completion of the integral closure of $\mathbb{Z}_p((\pi))$ in $\widetilde{\mathbf{A}}$, and put $\mathbf{A}_K = \mathbf{A}^{H_K}$. These rings carry actions of $G_K$ that are continuous for the weak topology on the rings and the profinite topology on $G_K$. They also carry endomorphisms $\varphi$ (which are weakly and $p$-adically continuous) induced by the Witt vector Frobenius on $\widetilde{\mathbf{A}}$.

For $s>0$, the subset

$$\widetilde{\mathbf{A}}^{\dagger,s}=\Big\{x\in\widetilde{\mathbf{A}}\,\Big|\,x=\sum_{k\in\mathbb{Z}}p^k[x_k],\,v_{\widetilde{\mathbf{E}}}(x_k)+\frac{psk}{p-1}\ge 0,\,\lim_{k\to+\infty}v_{\widetilde{\mathbf{E}}}(x_k)+\frac{psk}{p-1}=+\infty\Big\}$$

is a subring of $\widetilde{\mathbf{A}}$ that is complete for the valuation

$$w_s(x)=\inf_k\Big\{v_{\widetilde{\mathbf{E}}}(x_k)+\frac{psk}{p-1}\Big\}.$$

Put

$$\widetilde{\mathbf{B}}^{\dagger}=\bigcup_{s>0}\widetilde{\mathbf{B}}^{\dagger,s},\quad \mathbf{B}_K^{\dagger,s}=\mathbf{B}_K\cap\widetilde{\mathbf{B}}^{\dagger,s},\quad \mathbf{B}_K^{\dagger}=\bigcup_{s>0}\mathbf{B}_K^{\dagger,s},$$
$$\mathbf{A}_K^{\dagger,s}=\mathbf{A}_K\cap\widetilde{\mathbf{A}}^{\dagger,s},\quad \mathbf{A}_K^{\dagger}=\mathbf{A}\cap\mathbf{B}_K^{\dagger}.$$

(This last ring is strictly larger than $\bigcup_{s>0}\mathbf{A}_K^{\dagger,s}$.) These rings carry an action of $\varphi$. with the proviso that $\varphi$ gtakes a ring with a superscript of $s$ to the corresponding ring with $s$ replaced by $ps$. For $n$ a positive integer, write

$$\mathbf{A}_{K,n}^{\dagger,s}=\varphi^{-n}(\mathbf{A}_K^{\dagger,p^ns}).$$

Let $\widetilde{\mathbf{B}}_{\mathrm{rig}}^{\dagger,s}$ be the Fréchet completion of $\widetilde{\mathbf{B}}^{\dagger,s'}$ under the valuations $w_{s'}$ for all $s'\ge s$, and put $\widetilde{\mathbf{B}}_{\mathrm{rig}}^{\dagger}=\bigcup_{s>0}\widetilde{\mathbf{B}}_{\mathrm{rig}}^{\dagger,s}$. Similarly, let $\mathbf{B}_{\mathrm{rig},K}^{\dagger,s}$ be the Fréchet completion of $\mathbf{B}_K^{\dagger,s'}$ under the valuations $w_{s'}$ for all $s'\ge s$, and put $\mathbf{B}_{\mathrm{rig},K}^{\dagger}=\bigcup_{s>0}\mathbf{B}_{\mathrm{rig},K}^{\dagger,s}$. It turns out that $(\mathbf{B}_{\mathrm{rig}}^{\dagger,s})^{H_K}=\mathbf{B}_{\mathrm{rig},K}^{\dagger,s}$.

Some of these rings admit more explicit descriptions, as follows. It turns out that $\mathbf{B}_K$ is isomorphic to the $p$-adic local field

$$\mathscr{E}_{K_0'}=\Big\{f=\sum_{i=-\infty}^{+\infty}a_iT^i\,\Big|\,a_i\in K_0',\inf_i\{v_p(a_i)\}>-\infty,\lim_{i\to-\infty}v_p(a_i)=+\infty\Big\}$$

with valuation $w(f)=\min_{i\in\mathbb{Z}}v_p(a_i)$ and imperfect residue field $k'((T))$, where $k'$ is the residue field of $K_0'$. There is no distinguished such isomorphism in general (except for $K=\mathbb{Q}_p$, where one may take $T=\pi$), but suppose we fix a choice. Then $\mathbf{B}$ corresponds to the completion of the maximal unramified extension of $\mathbf{B}_K$. For $s\gg 0$ (depending on $K$ and the choice of the isomorphism $\mathbf{B}_K\cong\mathscr{E}_{K_0'}$), $\mathbf{B}_K^{\dagger,s}$ corresponds to the subring $\mathscr{E}_{K_0'}^s$ of $\mathscr{E}_{K_0'}$ defined as

$$\mathscr{E}_{K_0'}^s=\Big\{f=\sum_{i=-\infty}^{+\infty}a_iT^i\,\Big|\,a_i\in K_0',\inf_i\{v_p(a_i)\}>-\infty,\lim_{i\to-\infty}i+\frac{ps}{p-1}v_p(a_i)=+\infty\Big\},$$

that is, the bounded Laurent series in $T$ convergent on the annulus $0<v_p(T)\le 1/s$. Meanwhile, $\mathbf{B}_{\mathrm{rig},K}^{\dagger,s}$ corresponds to the ring

$$\mathscr{R}_{K_0'}^s=\Big\{f=\sum_{i=-\infty}^{+\infty}a_iT^i\,\Big|\,a_i\in K_0',\lim_{i\to+\infty}i+rv_p(a_i)=+\infty\ \text{for all}\ r>0,$$
$$\lim_{i\to-\infty}i+\frac{ps}{p-1}v_p(a_i)=+\infty\Big\},$$

that is, the unbounded Laurent series in $T$ convergent on $0 < v_p(T) \leq 1/s$. The union $\mathcal{R}_{K_0'} = \bigcup_{s>0} \mathcal{R}^s_{K_0'}$ is commonly called the *Robba ring* over $K_0'$.

## 2. $p$-adic representations and $(\varphi, \Gamma)$-modules

We next introduce $p$-adic representations and the objects of semilinear algebra used to describe them. Fix a finite extension $K$ of $\mathbb{Q}_p$. For $R$ a topological ring, we will mean by an *$R$-linear representation* a finite $R$-module equipped with a continuous linear action of $G_K$. (We will apply additional adjectives like "free", which are to be passed through to the underlying $R$-module.) Fontaine [1990] constructed a functor giving an equivalence of categories between $\mathbb{Q}$-linear representations and certain linear (or rather semilinear) algebraic data, as follows. (We may extend to $L$-linear representations for finite extensions $L$ of $\mathbb{Q}$ by restricting the coefficient field to $\mathbb{Q}$ and then keeping track of the $L$-action separately.)

An *étale $\varphi$-module* over $\mathbf{A}_K$ is a finite module $N$ over $\mathbf{A}_K$ equipped with a semilinear action of $\varphi$ such that the $\mathbf{A}_K$-linear map $\varphi^* N \to N$ induced by the $\varphi$-action is an isomorphism. An *étale $\varphi$-module* over $\mathbf{B}_K$ is a finite module $M$ over $\mathbf{B}_K$, equipped with a semilinear action of $\varphi$, that contains an $\mathbf{A}_K$-lattice $N$ (that is, a finite $\mathbf{A}_K$-submodule such that the induced map $N \otimes_{\mathbf{A}_K} \mathbf{B}_K \to M$ is an isomorphism) that forms an étale $\varphi$-module over $\mathbf{A}_K$. An *étale $(\varphi, \Gamma)$-module* over $\mathbf{A}_K$ or $\mathbf{B}_K$ is an étale $\varphi$-module equipped with a semilinear action of $\Gamma$ which commutes with the $\varphi$-action and is continuous for the profinite topology on $\Gamma$ and the weak topology on $\mathbf{A}_K$. Note that an étale $(\varphi, \Gamma)$-module over $\mathbf{B}_K$ may contain an $\mathbf{A}_K$-lattice that forms an étale $\varphi$-module over $\mathbf{A}_K$ but is not stable under $\Gamma$; on the other hand, the images of such a lattice under $\Gamma$ span another lattice which forms an étale $\varphi$-module over $\mathbf{A}_K$.

For $T$ a $\mathbb{Z}$-linear representation, define $\mathrm{D}(T) = (\mathbf{A} \otimes_{\mathbb{Z}_p} T)^{H_K}$; this gives an $\mathbf{A}_K$-module equipped with commuting semilinear actions of $\varphi$ and $\Gamma$. Similarly, for $V$ a $\mathbb{Q}$-linear representation, define $\mathrm{D}(V) = (\mathbf{B} \otimes_{\mathbb{Q}_p} V)^{H_K}$.

**Theorem 2.1** [Fontaine 1990]. *The functor $T \mapsto \mathrm{D}(T)$ (resp. $V \mapsto \mathrm{D}(V)$) is an equivalence from the category of $\mathbb{Z}$-linear representations (resp. $\mathbb{Q}$-linear representations) of $G_K$ to the category of étale $(\varphi, \Gamma)$-modules over $\mathbf{A}_K$ (resp. $\mathbf{B}_K$); a quasiinverse functor is given by $D \mapsto (\mathbf{A} \otimes_{\mathbf{A}_K} D)^{\varphi=1}$ (resp. $D \mapsto (\mathbf{B} \otimes_{\mathbf{B}_K} D)^{\varphi=1}$).*

Dee [2001] extended Fontaine's results to families of $\mathbb{Z}$-representations, as follows. Let $R$ be a complete noetherian local ring whose residue field $k_R$ is finite over $\mathbb{F}_p$, equipped with the topology defined by its maximal ideal $\mathfrak{m}_R$; we may then view $R$ as a topological $\mathbb{Z}$-algebra. We form the completed tensor product $R \widehat{\otimes}_{\mathbb{Z}} \mathbf{A}$ by completing the ordinary tensor product for the ideal $p\mathbf{A} + \mathfrak{m}_R$, and similarly with $\mathbf{A}$ replaced by $\mathbf{A}_K$.

We define $(\varphi, \Gamma)$-modules and étale $(\varphi, \Gamma)$-modules over $R \mathbin{\widehat{\otimes}}_{\mathbb{Z}} \mathbf{A}_K$ by analogy with the definitions over $\mathbf{A}_K$. For $T_R$ an $R$-representation, define $\mathrm{D}(T_R) = ((R \mathbin{\widehat{\otimes}}_{\mathbb{Z}} \mathbf{A}) \otimes_R T_R)^{H_K}$.

**Theorem 2.2** [Dee 2001]. *The functor*

$$T_R \mapsto \mathrm{D}(T_R)$$

*is an equivalence from the category of $R$-representations to the category of étale $(\varphi, \Gamma)$-modules over $R \mathbin{\widehat{\otimes}}_{\mathbb{Z}} \mathbf{A}_K$; a quasiinverse functor is given by*

$$D \mapsto ((R \mathbin{\widehat{\otimes}}_{\mathbb{Z}} \mathbf{A}) \otimes_{R \mathbin{\widehat{\otimes}}_{\mathbb{Z}} \mathbf{A}_K} D)^{\varphi=1}.$$

We next cite a refinement of Fontaine's result. We define $(\varphi, \Gamma)$-modules and étale $(\varphi, \Gamma)$-modules over $\mathbf{A}_K^{\dagger}$ and $\mathbf{B}_K^{\dagger}$ by analogy with the definitions over $\mathbf{A}_K$ and $\mathbf{B}_K$. For $V$ a $\mathbb{Q}_p$-linear representation, define $\mathrm{D}_K^{\dagger,r}(V) = (\mathbf{B}^{\dagger,r} \otimes_{\mathbb{Q}_p} V)^{H_K}$ (where $\mathbf{B}^{\dagger,r} = \mathbf{B} \cap \widetilde{\mathbf{B}}^{\dagger,r}$) and $\mathrm{D}_K^{\dagger}(V) = \bigcup_{r>0} \mathrm{D}_K^{\dagger,r}(V) = (\mathbf{B}^{\dagger} \otimes_{\mathbb{Q}_p} V)^{H_K}$.

**Theorem 2.3** [Cherbonnier and Colmez 1998]. *For any $\mathbb{Q}_p$-linear representation $V$, there exists $r(V) > 0$ such that*

$$\mathrm{D}_K(V) = \mathbf{B}_K \otimes_{\mathbf{B}_K^{\dagger,r}} \mathrm{D}_K^{\dagger,r}(V) \quad \textit{for all} \quad r \geqslant r(V).$$

*Equivalently, $\mathrm{D}_K^{\dagger}(V)$ is an étale $(\varphi, \Gamma)$-module over $\mathbf{B}_K^{\dagger}$ of dimension $\dim_{\mathbb{Q}} V$. Therefore $V \mapsto \mathrm{D}_K^{\dagger}(V)$ is an equivalence from the category of $p$-adic representations of $G_K$ to the category of étale $(\varphi, \Gamma)$-modules over $\mathbf{B}_K^{\dagger}$. Furthermore, $\mathrm{D}_K^{\dagger}(V)$ is the unique maximal étale $(\varphi, \Gamma)$-submodule of $\mathrm{D}_K(V)$ over $\mathbf{B}_K^{\dagger}$.*

Berger and Colmez [2008] extended these results to families of $p$-adic representations. However, unlike Dee's families, the families considered by Berger and Colmez are over Banach algebras over $\mathbb{Q}$. (Berger and Colmez were forced to make a freeness hypothesis on the representation space; we relax this hypothesis later in the case of an affinoid algebra. See Definition 3.12.)

For $S$ a commutative Banach algebra over $\mathbb{Q}_p$, let $\mathbb{O}_S$ be the ring of elements of $S$ of norm at most 1, and let $I_S$ be the ideal of elements of $\mathbb{O}_S$ of norm strictly less than 1. Note that it makes sense to form a completed tensor product with $S$ or $\mathbb{O}_S$ when the other tensorand carries a norm under which it is complete — for example, for the rings $\widetilde{\mathbf{A}}^{\dagger,s}, \mathbf{A}_{L,n}^{\dagger,s}, \widetilde{\mathbf{B}}^{\dagger,s}, \mathbf{B}_L^{\dagger,s}$ using the norm corresponding to the valuation $w_s$.

**Proposition 2.4** [Berger and Colmez 2008, proposition 4.2.8]. *Let $S$ be a commutative Banach algebra over $\mathbb{Q}_p$. Let $T_S$ be a free $\mathbb{O}_S$-linear representation of rank $d$. Let $L$ be a finite Galois extension of $K$ such that $G_L$ acts trivially on $T_S/12pT_S$. Then there exists $n(L, T_S) \geq 0$ such that for $n \geq n(L, T_S)$,*

$$(\mathbb{O}_S \mathbin{\widehat{\otimes}}_{\mathbb{Z}_p} \tilde{\mathbf{A}}^{\dagger,(p-1)/p}) \otimes_{\mathbb{O}_S} T_S$$

*has a unique sub-*$(\mathbb{O}_S \widehat{\otimes}_{\mathbb{Z}_p} \mathbf{A}_{L,n}^{\dagger,(p-1)/p})$*-module* $\mathrm{D}_{L,n}^{\dagger,(p-1)/p}(T_S)$ *that is free of rank* $d$*, is fixed by* $H_L$*, has a basis almost invariant under* $\Gamma_L$ *(that is, for each* $\gamma \in \Gamma_L$*, the matrix of action of* $\gamma - 1$ *on the basis has positive valuation), and satisfies*

$$(\mathbb{O}_S \widehat{\otimes}_{\mathbb{Z}_p} \tilde{\mathbf{A}}^{\dagger,(p-1)/p}) \otimes_{\mathbb{O}_S \widehat{\otimes}_{\mathbb{Z}_p} \mathbf{A}_{L,n}^{\dagger,(p-1)/p}} \mathrm{D}_{L,n}^{\dagger,(p-1)/p}(T_S) = (\mathbb{O}_S \widehat{\otimes} \tilde{\mathbf{A}}^{\dagger,(p-1)/p}) \otimes_{\mathbb{O}_S} T_S.$$

**Theorem 2.5** [Berger and Colmez 2008, théorème 4.2.9]. *Let* $S$ *be a commutative Banach algebra over* $\mathbb{Q}_p$*. Let* $V_S$ *be an* $S$*-linear representation admitting a free Galois-stable* $\mathbb{O}_S$*-lattice* $T_S$*. There exists an* $s(V_S) \geq 0$ *such that for any* $s \geq s(V_S)$*, we may define*

$$\mathrm{D}_K^{\dagger,s}(V_S) = ((S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_L^{\dagger,s}) \otimes_{\mathbb{O}_S \widehat{\otimes}_{\mathbb{Z}_p} \mathbf{A}_L^{\dagger,s(V_S)}} \varphi^n(\mathrm{D}_{L,n}^{\dagger,p-1/p}(T_S)))^{H_K}$$

*for some* $L$ *and* $n$*, so that the construction does not depend on the choices of* $T_S$*,* $L$*, and* $n$*, and the following statements hold.*

(a) *The* $(S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger,s})$*-module* $\mathrm{D}_K^{\dagger,s}(V_S)$ *is locally free of rank* $d$*.*

(b) *The natural map* $\mathrm{D}_K^{\dagger,s}(V_S) \otimes_{S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger,s}} (S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger,s}) \to V_S \otimes_S (S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger,s})$ *is an isomorphism.*

(c) *For any maximal ideal* $\mathfrak{m}_x$ *of* $S$*, for* $V_x = V_S \otimes_S (S/\mathfrak{m}_x)$*, the natural map* $\mathrm{D}_K^{\dagger,s}(V_S) \otimes_S (S/\mathfrak{m}_x) \to \mathrm{D}_K^{\dagger,s}(V_x)$ *is an isomorphism.*

We write $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger} = \bigcup_{s>0}(S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger,s})$ and $S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger} = \bigcup_{s>0}(S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger,s})$. (Note that $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger}$ does not necessarily embed into $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K$, due to the incompatibility between the topologies used for the completed tensor products.) We then put

$$\mathrm{D}_K^{\dagger}(V_S) = \mathrm{D}_K^{\dagger,s}(V_S) \otimes_{S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger,s}} (S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger}).$$

We may recover $V_S$ from $\mathrm{D}_K^{\dagger}(V_S)$ as follows.

**Lemma 2.6.** $(S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger})^{\varphi=1} = S.$

*Proof.* We reduce at once to the case where $S$ is countably topologically generated over $\mathbb{Q}_p$. In this case, by [Bosch et al. 1984, Proposition 2.7.2/3], we can find a *Schauder basis* of $S$ over $\mathbb{Q}$; in other words, there exists an index set $I$ such that $S$ is isomorphic as a topological $\mathbb{Q}$-vector space to the Banach space

$$l_0^{\infty}(I, \mathbb{Q}) = \{(a_i)_{i \in I} \mid a_i \in \mathbb{Q}, a_i \to 0\}.$$

(The supremum norm need only be equivalent to the Banach norm on $S$; the two need not be equal.) We can then write $S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger}$, as a topological $\mathbb{Q}$-vector space, as

$$l_0^{\infty}(I, \widetilde{\mathbf{B}}^{\dagger}) = \{(a_i)_{i \in I} \mid a_i \in \widetilde{\mathbf{B}}^{\dagger}, a_i \to 0\}.$$

In this presentation, the $\varphi$-action carries $(a_i)_{i\in I}$ to $(\varphi(a_i))_{i\in I}$. It is then clear that $(S\,\widehat{\otimes}_{\mathbb{Q}_p}\widetilde{\mathbf{B}}^\dagger)^{\varphi=1} = (l_0^\infty(I,\widetilde{\mathbf{B}}^\dagger))^{\varphi=1} = l_0^\infty(I,\mathbb{Q}) = S$. □

**Proposition 2.7.** $(\mathrm{D}_K^\dagger(V_S)\otimes_{S\,\widehat{\otimes}_{\mathbb{Q}_p}\mathbf{B}_K^\dagger}(S\,\widehat{\otimes}_{\mathbb{Q}_p}\widetilde{\mathbf{B}}^\dagger))^{\varphi=1} = V_S$.

*Proof.* From Theorem 2.5(b) we get

$$\mathrm{D}_K^\dagger(V_S)\otimes_{S\,\widehat{\otimes}_{\mathbb{Q}_p}\mathbf{B}_K^\dagger}(S\,\widehat{\otimes}_{\mathbb{Q}_p}\widetilde{\mathbf{B}}^\dagger) = V_S\otimes_S(S\,\widehat{\otimes}_{\mathbb{Q}_p}\widetilde{\mathbf{B}}^\dagger).$$

By Lemma 2.6, it follows that

$$(\mathrm{D}_K^\dagger(V_S)\otimes_{S\,\widehat{\otimes}_{\mathbb{Q}_p}\mathbf{B}_K^\dagger}(S\,\widehat{\otimes}_{\mathbb{Q}_p}\widetilde{\mathbf{B}}^\dagger))^{\varphi=1} = V_S\otimes_S(S\,\widehat{\otimes}_{\mathbb{Q}_p}\widetilde{\mathbf{B}}^\dagger)^{\varphi=1} = V_S. \quad \square$$

This suggests that the object $\mathrm{D}_K^\dagger(V_S)$ merits the following definition.

**Definition 2.8.** Define a $(\varphi,\Gamma)$-*module* over $S\,\widehat{\otimes}_{\mathbb{Q}_p}\mathbf{B}_K^\dagger$ to be a finite locally free module over $S\,\widehat{\otimes}_{\mathbb{Q}_p}\mathbf{B}_K^\dagger$, equipped with commuting continuous $(\varphi,\Gamma)$-actions such that $\varphi^*D_S\to D_S$ is an isomorphism. We say a $(\varphi,\Gamma)$-module $M_S$ over $S\,\widehat{\otimes}_{\mathbb{Q}_p}\mathbf{B}_K^\dagger$ is *étale* if it admits a finite $(\varphi,\Gamma)$-stable $(\mathbb{O}_S\,\widehat{\otimes}_{\mathbb{Z}_p}\mathbf{A}_K^\dagger)$-submodule $N_S$ such that $\varphi^*N_S\to N_S$ is an isomorphism and the induced map

$$N_S\otimes_{\mathbb{O}_S\,\widehat{\otimes}_{\mathbb{Z}_p}\mathbf{A}_K^\dagger} S\,\widehat{\otimes}_{\mathbb{Q}_p}\mathbf{B}_K^\dagger \to M_S$$

is an isomorphism. In this language, Theorem 2.5 implies that $\mathrm{D}_K^\dagger(V_S)$ is an étale $(\varphi,\Gamma)$-module over $S\,\widehat{\otimes}_{\mathbb{Q}_p}\mathbf{B}_K^\dagger$.

## 3. Gluing on affinoid spaces

Throughout this section, let $S$ denote an affinoid algebra over $\mathbb{Q}_p$. We explain how to perform gluing for finite modules over $S\,\widehat{\otimes}_{\mathbb{Q}_p}\mathbf{B}_K^\dagger$. We start with some basic notions from [Bosch et al. 1984].

**Definition 3.1.** Let $M(S)$ be the set of maximal ideals of $S$, that is, the affinoid space associated to $S$. For $X$ a subset of $M(S)$, an *affinoid subdomain* of $X$ is a subset $U$ of $X$ for which there exists a morphism $S\to S'$ of affinoid algebras such that the induced map $M(S')\to M(S)$ is universal for maps from an affinoid space to $M(S)$ landing in $U$. The algebra $S'$ is then unique up to unique isomorphism, and the resulting map $M(S)\to U$ is a bijection.

The set $M(S)$ carries two canonical $G$-topologies, defined as follows. In the *weak $G$-topology*, the admissible open sets are the affinoid subdomains, and the admissible coverings are the finite coverings. In the *strong $G$-topology*, the admissible open sets are the subsets $U$ of $M(S)$ admitting a covering by affinoid subdomains such that the induced covering of any affinoid subdomain of $U$ can be refined to a finite cover by affinoid subdomains, and the admissible coverings are the ones whose restriction to any affinoid subdomain can be refined to a finite cover by affinoid subdomains. The categories of sheaves on these two topologies are equivalent, because the strong $G$-topology is *slightly finer* than the weak one [Bosch et al. 1984, §9.1].

We need a generalization of the Tate and Kiehl theorems on coherent sheaves on affinoid spaces.

**Definition 3.2.** For $A$ a commutative Banach algebra over $\mathbb{Q}_p$, define the presheaf $\mathscr{A}$ on the weak $G$-topology of $M(S)$ by declaring that

$$\mathscr{A}(M(S')) = S' \widehat{\otimes}_{\mathbb{Q}_p} A.$$

**Lemma 3.3.** *For $A$ a commutative Banach algebra over $\mathbb{Q}_p$, the presheaf $\mathscr{A}$ is a sheaf for the weak $G$-topology of $M(S)$, and hence extends uniquely to the strong $G$-topology.*

*Proof.* Since every finite covering of an affinoid space by affinoid subdomains can be refined to a Laurent covering, it is enough to check the sheaf condition for Laurent coverings [Bosch et al. 1984, Proposition 8.2.2/5]. This reduces to checking for coverings of the form

$$M(S) = M(S\langle f\rangle) \cup M(S\langle f^{-1}\rangle)$$

for $f \in S$. The claim then is that the sequence

$$0 \to S \widehat{\otimes}_{\mathbb{Q}_p} A \to (S\langle f\rangle \widehat{\otimes}_{\mathbb{Q}_p} A) \times (S\langle f^{-1}\rangle \widehat{\otimes}_{\mathbb{Q}_p} A) \xrightarrow{d^0} S\langle f, f^{-1}\rangle \widehat{\otimes}_{\mathbb{Q}_p} A \to 0$$

is exact; this follows from the corresponding assertion for $A = \mathbb{Q}_p$, for which see [Bosch et al. 1984, §8.2.3]. □

From now on, we consider only the strong $G$-topology on $M(S)$.

**Definition 3.4.** For $A$ a commutative Banach algebra over $\mathbb{Q}_p$, an $\mathscr{A}$-module $N$ on $M(S)$ is *coherent* if there exists an admissible covering $\{M(S_i)\}_{i\in I}$ of $M(S)$ by affinoid subdomains such that for each $i \in I$, we have

$$N|_{M(S_i)} = \operatorname{coker}(\varphi : \mathscr{A}^m|_{M(S_i)} \to \mathscr{A}^n|_{M(S_i)})$$

for some morphism $\varphi$ of $\mathscr{A}$-modules. By Lemma 3.3, this is equivalent to requiring $N|_{M(S_i)}$ to be the sheaf associated to some finitely presented $(S_i \widehat{\otimes}_{\mathbb{Q}_p} A)$-module.

**Lemma 3.5.** *Let $A$ be a commutative Banach algebra over $\mathbb{Q}_p$ such that for each Tate algebra $T_n$ over $\mathbb{Q}_p$, $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ is noetherian. Then for any coherent $\mathscr{A}$-module $N$ on $M(S)$, the first Čech cohomology $\check{H}^1(N)$ vanishes.*

*Proof.* As in Lemma 3.3, it suffices to check vanishing of the first Čech cohomology computed on a cover of $M(S)$ of the form

$$M(S) = M(S\langle f\rangle) \cup M(S\langle f^{-1}\rangle)$$

for some $f \in S$, such that $N$ is represented on each of the two covering subsets by a finite module. For this, we may follow the proof of [Bosch et al. 1984, Lemma 9.4.3/5] verbatim. (The noetherian condition is needed so that the invocation of [Bosch et al. 1984, Proposition 3.7.3/3] within the proof of [Bosch et al. 1984, Lemma 9.4.3/5] remains valid.) $\square$

To recover an analogue of Kiehl's theorem, however, we need an extra condition.

**Proposition 3.6.** *Let $A$ be a commutative Banach algebra over $\mathbb{Q}_p$ such that for each Tate algebra $T_n$ over $\mathbb{Q}_p$, $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ is noetherian and the map*

$$\mathrm{Spec}(T_n \widehat{\otimes}_{\mathbb{Q}_p} A) \to \mathrm{Spec}(T_n)$$

*carries $M(T_n \widehat{\otimes}_{\mathbb{Q}_p} A)$ to $M(T_n)$. Then any coherent $\mathscr{A}$-module $N$ on $M(S)$ is associated to a finite $(S \widehat{\otimes}_{\mathbb{Q}_p} A)$-module.*

*Proof.* There must exist a finite covering of $M(S)$ by affinoid subdomains $M(S_1)$, $\dots, M(S_n)$ such that $N|_{M(S_i)}$ is associated to a finite $(S_i \widehat{\otimes}_{\mathbb{Q}_p} A)$-module $N_i$. As in [Bosch et al. 1984, Lemma 9.4.3/6], we may deduce from Lemma 3.5 that for each $\mathfrak{m} \in M(S_i)$, the map $N(M(S)) \to (N/\mathfrak{m}N)(M(S_i))$ is surjective. By the hypothesis on $A$, each maximal ideal of $S_i \widehat{\otimes}_{\mathbb{Q}_p} A$ lies over a maximal ideal of $S_i$; we may thus deduce that $N(M(S)) \otimes_S S_i$ surjects onto $N(M(S_i))$. Since the latter is a finite $S_i \widehat{\otimes}_{\mathbb{Q}} A$-module, we can choose finitely many elements of $N(M(S))$ that generate all of the $N(M(S_i))$. That is, we have a surjection $\mathscr{A}^n \to N$ for some $n$; repeating the argument for the kernel of this map yields the claim. $\square$

To use the above argument, we need to prove a variant of the Nullstellensatz; for simplicity, we restrict to the case where $K$ is discretely valued (the case of interest in this paper). We first prove a finite generation result using ideas from the theory of Gröbner bases.

**Lemma 3.7.** *Let $K$ be a complete discretely valued field extension of $\mathbb{Q}_p$. Let $A$ be a commutative Banach algebra over $K$ such that $A$ has the same set of nonzero norms as $K$, and the ring $\mathcal{O}_A/I_A$ is noetherian. Then $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ is also noetherian.*

*Proof.* Equip the monoid $\mathbb{Z}^n_{\geq 0}$ with the componentwise partial ordering $\leq$ and the lexicographic total ordering $\preceq$. That is, $(x_1, \dots, x_n) \leq (y_1, \dots, y_n)$ if $x_i \leq y_i$ for all $i$, whereas $(x_1, \dots, x_n) \preceq (y_1, \dots, y_n)$ if there exists an index $i \in \{1, \dots, n+1\}$ such that $x_j = y_j$ for $j < i$, and either $i = n+1$ or $x_i \leq y_i$. Recall that $\leq$ is a well partial ordering and that $\preceq$ is a well total ordering; in particular, any sequence in $\mathbb{Z}^n_{\geq 0}$ has a subsequence that is weakly increasing under both orderings.

For $I = (i_1, \dots, i_n) \in \mathbb{Z}^n_{\geq 0}$, write $t^I$ for $t_1^{i_1} \cdots t_n^{i_n}$. We represent each element $x \in T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ as a formal sum $\sum_I x_I t^I$ with $x_I \in A$, such that for each $\epsilon > 0$, there exist only finitely many indices $I$ with $|x_I| \leq \epsilon$. For $x$ nonzero, define the *degree* of $x$, denoted $\deg(x)$, to be the maximal index $I$ under $\preceq$ among those indices maximizing $|x_I|$. Define the *leading coefficient* of $x$ to be the coefficient $x_{\deg(x)}$.

Let $J$ be any ideal of $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$. We apply a Buchberger-type algorithm to construct a generating set $m_1, \dots, m_k$ for $J$, as follows. Start with the empty list (that is, $k = 0$). As long as possible, given $m_1, \dots, m_k$, choose an element $m_{k+1}$ of $J \cap \mathbb{O}_A$ with leading coefficient $a_{k+1}$, for which we *cannot* choose $I_1, \dots, I_k \in \mathbb{Z}^n_{\geq 0}$ and $b_1, \dots, b_k \in \mathbb{O}_A$ satisfying both

(a) $\deg(m_{k+1}) = \deg(m_i t^{I_i})$ whenever $b_i \neq 0$ and

(b) $a_{k+1} - a_1 b_1 - \cdots - a_k b_k \in I_A$.

In particular, we must have $|a_{k+1}| = 1$.

We claim this process must terminate. Suppose the contrary; then there must exist a sequence of indices $i_1 < i_2 < \cdots$ such that $\deg(m_{i_1}) \leq \deg(m_{i_2}) \leq \cdots$. Then the sequence of ideals $(a_{i_1}), (a_{i_1}, a_{i_2}), \dots$ in $\mathbb{O}_A / I_A$ must be strictly increasing, but this violates the hypothesis that $\mathbb{O}_A / I_A$ is noetherian. Hence the process terminates.

Let $|\cdot|_1$ denote the 1-Gauss norm on $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$. We now write each element of $J$ as a linear combination of $m_1, \dots, m_k$ using a form of the Buchberger division algorithm. Start with some nonzero $x \in J$ and put $x_0 = x$. Given $x_l \in J$, if $x_l = 0$, put $y_{l,1} = \cdots = y_{l,k} = 0$ and $x_{l+1} = 0$. Otherwise, choose $\lambda \in A^\times$ with $|\lambda x_l|_1 = 1$. By the construction of $m_1, \dots, m_k$, there must exist $I_1, \dots, I_k \in \mathbb{Z}^n_{\geq 0}$ and $b_1, \dots, b_k \in \mathbb{O}_A$ satisfying conditions (a) and (b) with $m_{k+1}$ replaced by $\lambda x_l$. Put $y_{l,i} = \lambda^{-1} b_i t^{I_i}$ and $x_{l+1} = x_l - y_{l,1} m_1 - \cdots - y_{l,k} m_k$.

If $|x_{l+1}|_1 = |x_l|_1$, we must have $\deg(x_{l+1}) \prec \deg(x_l)$. Since $\preceq$ is a well ordering, we must have $|x_{l'}|_1 < |x_l|_1$ for some $l' > l$. Since $K$ is discretely valued and $A$ has the same group of nonzero norms as $K$, we conclude that $|x_l|_1 \to 0$ as $l \to \infty$.

Since $|y_{l,i}|_1 \leq |x_l|_1$, we may set $y_i = \sum_{l=0}^\infty y_{l,i}$ to get elements of $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ such that $x = y_1 m_1 + \cdots + y_k m_k$. This proves that $J$ is always finitely generated, so $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ is noetherian. $\square$

Next we establish an analogue of the Nullstellensatz by combining the previous argument with an idea of Munshi [May 2003].

**Lemma 3.8.** *Take $K$ and $A$ as in Lemma 3.7, but suppose also that the intersection of the nonzero prime ideals of $A$ is zero. Then for any maximal ideal $\mathfrak{m}$ of $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$, the intersection $\mathfrak{m} \cap A$ is nonzero.*

*Proof.* Suppose on the contrary that $\mathfrak{m}$ is a maximal ideal of $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ such that $\mathfrak{m} \cap A = 0$. Since $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ is noetherian by Lemma 3.7, $\mathfrak{m}$ is closed by [Bosch et al. 1984, Proposition 3.7.2/2]. Hence $\mathfrak{m} + A$ is also a closed subspace of $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$. Let $\psi : T_n \widehat{\otimes}_{\mathbb{Q}_p} A \to (T_n \widehat{\otimes}_{\mathbb{Q}_p} A)/A$ be the canonical projection; it is a bounded surjective morphism of Banach spaces with kernel $A$. Put $V = \psi(\mathfrak{m} + A)$; since

$\mathfrak{m} + A = \psi^{-1}(V)$, the open mapping theorem [Bosch et al. 1984, §2.8.1] implies that $V$ is closed. Hence $\psi$ induces a bounded bijective map $\mathfrak{m} \to V$ between two Banach spaces; by the open mapping theorem again, the inverse of $\psi$ is also bounded.

Using the power series representation of elements of $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$, let us represent $(T_n \widehat{\otimes}_{\mathbb{Q}_p} A)/A$ as the set of series in $T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ with zero constant term. We may then represent $\psi$ as the map that subtracts off the constant term. Define the *nonconstant degree* of $x \in T_n \widehat{\otimes}_{\mathbb{Q}_p} A$ as $\deg'(x) = \deg(\psi(x))$, and define the *leading nonconstant coefficient* of $x$ to be the coefficient $x_{\deg'(x)}$.

We construct $m_1, \dots, m_k \in \mathfrak{m}$ using the following modified Buchberger algorithm. As long as possible, choose an element $m_{k+1}$ of $\mathfrak{m} \cap \mathcal{O}_A$ with nonconstant leading coefficient $a_{k+1}$, for which we cannot choose $I_1, \dots, I_k \in \mathbb{Z}^n_{\geq 0}$ and $b_1, \dots, b_k \in \mathcal{O}_A$ satisfying both

(a) $\deg'(m_{k+1}) = \deg'(m_i t^{I_i})$ whenever $b_i \neq 0$ and

(b) $a_{k+1} - a_1 b_1 - \cdots - a_k b_k \in I_A$.

Again, this algorithm must terminate.

By the hypothesis on $A$, we can choose a nonzero prime ideal $\mathfrak{p}$ of $A$ not containing the product $a_1 \cdots a_k$. By our earlier hypothesis that $\mathfrak{m} \cap A = 0$, we have $\mathfrak{m} \cap \mathfrak{p} = 0$. Hence $\mathfrak{m} + \mathfrak{p}(T_n \widehat{\otimes}_{\mathbb{Q}_p} A)$ is the unit ideal, so we can find $x_0 \in \mathfrak{p}(T_n \widehat{\otimes}_{\mathbb{Q}_p} A)$ such that $1 + x_0 \in \mathfrak{m}$.

We now perform a modified division algorithm. Given $x_l \in \mathfrak{p}(T_n \widehat{\otimes}_{\mathbb{Q}_p} A)$ such that $1 + x_l \in \mathfrak{m}$, we cannot have $x_l \in A$. We may thus choose $\lambda \in A^\times$ with $|\psi(\lambda x_l)|_1 = 1$. By the construction of $m_1, \dots, m_k$, there must exist $I_1, \dots, I_k \in \mathbb{Z}^n_{\geq 0}$ and $b_1, \dots, b_k \in A$ satisfying conditions (a) and (b) with $m_{k+1}$ replaced by $\lambda x_l$. Put $y_{l,i} = \lambda^{-1} b_i t^{I_i}$ and $x_{l+1} = x_l - y_{l,1} m_1 - \cdots - y_{l,k} m_k$.

As in the proof of Lemma 3.7, we see that $|\psi(x_l)|_1 \to 0$ as $l \to \infty$. Since $\psi$ has bounded inverse, we also conclude that $|x_l|_1 \to 0$ as $l \to \infty$. However, since $\mathfrak{m}$ is closed, this yields the contradiction $1 \in \mathfrak{m}$. We conclude that $\mathfrak{m} \cap A \neq 0$, as desired. □

**Lemma 3.9.** *For any Tate algebra $T_n$ over $\mathbb{Q}_p$, any rational $s > 0$, and any complete discretely valued field extension $K$ of $\mathbb{Q}_p$, $T_n \widehat{\otimes}_{\mathbb{Q}_p} \mathscr{E}^s_K$ is noetherian and each of its maximal ideals has residue field finite over $K$. In particular, every maximal ideal of $T_n \widehat{\otimes}_{\mathbb{Q}_p} \mathscr{E}^s_K$ lies over a maximal ideal of $T_n$.*

*Proof.* The Banach norm on $\mathscr{E}^s_K$ is the maximum of the $p$-adic norm and the norm induced by $w_s$. By enlarging $K$, we may assume that the nonzero values of this norm are all achieved by elements of $K$. In this case, we check that $A = \mathscr{E}^s_K$ satisfies the hypotheses of Lemma 3.8. First, the nonzero norms of elements of $A$ are all realized by units of the form $\lambda t^i$ with $\lambda \in K^\times$ and $i \in \mathbb{Z}$. Second, the residue ring $\mathbb{O}_A/I_A$ is isomorphic to a Laurent polynomial ring over a field, which is noetherian. Third, for each nonzero element $x$ of $A$, we can construct $y \in A$ whose Newton polygon has no slopes in common with that of $x$; this implies that $x$ and $y$ generate the unit ideal (see, for example, [Kedlaya 2005b, §2.6]), so any maximal ideal containing $y$ fails to contain $x$. Hence the intersection of the nonzero prime ideals of $A$ is zero; moreover, the quotient of $A$ by any nonzero ideal is finite over $K$. We may thus apply Lemma 3.8 to deduce the claim. □

By combining Proposition 3.6 with Lemma 3.9, we deduce the following. (The second assertion follows from the first because for a coherent module, local freeness can be checked at each maximal ideal.)

**Proposition 3.10.** *For any $s > 0$ and any finite extension $K$ of $\mathbb{Q}_p$, for $A = \mathscr{E}^s_K$, any coherent $\mathscr{A}$-module $\mathscr{V}$ on $M(S)$ is associated to a finite $(S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} A)$-module $V$. Moreover, $\mathscr{V}$ is locally free if and only if $V$ is.*

Using this, we may extend Theorem 2.5 for affinoid algebras, to eliminate the hypothesis requiring a free Galois-stable lattice. We first handle the case where $V_S$ is itself free.

**Theorem 3.11.** *Let $S$ be an affinoid algebra over $\mathbb{Q}_p$. Let $V_S$ be a free $S$-linear representation. There exists $s(V_S) \geq 0$ such that for $s \geq s(V_S)$, we may construct a $(S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^{\dagger,s}_K)$-module $\mathrm{D}^{\dagger,s}_K(V_S)$ satisfying the following conditions.*

- *The $(S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^{\dagger,s}_K)$-module $\mathrm{D}^{\dagger,s}_K(V_S)$ is locally free of rank $d$.*
- *The natural map $\mathrm{D}^{\dagger,s}_K(V_S) \otimes_{S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^{\dagger,s}_K} (S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger,s}) \to V_S \otimes_S (S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger,s})$ is an isomorphism.*
- *For any maximal ideal $\mathfrak{m}_x$ of $S$, for $V_x = V_S \otimes_S (S/\mathfrak{m}_x)$, the natural map $\mathrm{D}^{\dagger,s}_K(V_S) \otimes_S (S/\mathfrak{m}_x) \to \mathrm{D}^{\dagger,s}_K(V_x)$ is an isomorphism.*
- *The construction is functorial in $V_S$, compatible with passage from $K$ to a finite extension, and compatible with Theorem 2.5 in case $V_S$ admits a Galois-stable free lattice.*

*Proof.* Let $T_S$ be any free $\mathbb{O}_S$-lattice in $V_S$. Since the Galois action is continuous, there exists a finite Galois extension $L$ of $K$ such that $G_L$ carries $T_S$ into itself. For such $L$, for $s$ sufficiently large, $\mathrm{D}^{\dagger,s}_L(V_S)$ is locally free of rank $d$ by Theorem 2.5; moreover, it carries an action of $\mathrm{Gal}(L/K)$. If we restrict scalars on this module back to $S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^{\dagger,s}_K$, then $\mathrm{D}^{\dagger,s}_K(V_S)$ appears as a direct summand; this summand is then finite projective, and hence locally free (since $T_n \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathscr{E}^s_K$ is noetherian by Lemma 3.9). Moreover, the $\mathrm{Gal}(L/K)$-action on $\mathrm{D}^{\dagger,s}_L(V_S)$ allows us to extend the $\Gamma_L$-action on $\mathrm{D}^{\dagger,s}_K(V_S)$ to a $\Gamma_K$-action. This yields the desired assertions. □

**Definition 3.12.** Let $S$ be an affinoid algebra over $\mathbb{Q}_p$. Let $V_S$ be a locally free $S$-linear representation; we can then choose a finite covering of $M(S)$ by affinoid subdomains $M(S_1), \dots, M(S_n)$ such that $V_i = V_S \otimes_S S_i$ is free over $S_i$ for each $i$. We may then apply Theorem 3.11 to $V_i$ to produce $\mathrm{D}_K^{\dagger,s}(V_i)$ for $s$ sufficiently large.

By Proposition 3.10, these glue to form a finite $(S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger,s})$-module $\mathrm{D}_K^{\dagger,s}(V_S)$, which satisfies the analogues of the assertions of Theorem 3.11. We may then define

$$\mathrm{D}_K^{\dagger}(V_S) = \mathrm{D}^{\dagger,s}(V_S) \otimes_{S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger,s}} (S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger}),$$

and this will be an étale $(\varphi, \Gamma)$-module over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger}$. The analogue of Proposition 2.7 will also carry over.

**Remark 3.13.** Chenevier has pointed out that Theorem 3.11 is also an easy consequence of [Chenevier 2009, Lemme 3.18]. That lemma implies that for $S$ an affinoid algebra over $\mathbb{Q}_p$ and $V_S$ a locally free $S$-linear representation, there exist an affine formal scheme $\mathrm{Spf}(R)$ of finite type over $\mathbb{Z}$, equipped with an isomorphism $R \otimes_{\mathbb{Z}} \mathbb{Q}_p \cong S$, and a locally free $R$-linear representation $T_R$ admitting an isomorphism $T_R \otimes_{\mathbb{Z}_p} \mathbb{Q}_p \cong V_S$. This makes it possible to glue the Berger–Colmez theorem by doing so on a suitable formal model of $S$.

## 4. Local coefficient algebras

Here we show that in a restricted setting, it is possible to invert the $(\varphi, \Gamma)$-module functor $D_K^{\dagger}$.

**Definition 4.1.** By a *coefficient algebra*, we mean a commutative Banach algebra $S$ over $\mathbb{Q}$ satisfying the following conditions.

- The norm on $S$ restricts to the norm on $\mathbb{Q}$.
- For each maximal ideal $\mathfrak{m}$ of $S$, the residue field of $\mathfrak{m}$ is finite over $\mathbb{Q}$.
- The Jacobson radical of $S$ is zero; in particular, $S$ is reduced.

For instance, any reduced affinoid algebra over $\mathbb{Q}$ is a coefficient algebra.

By a *local coefficient algebra*, we mean a coefficient algebra $S$ of the form $R \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$, where $R$ is a complete noetherian local domain of characteristic 0 with residue field finite over $\mathbb{F}_p$. For instance, if $S$ is a reduced affinoid algebra over $\mathbb{Q}_p$ equipped with the spectral norm, and $R$ is the completion of $\mathbb{O}_S$ at a maximal ideal, then $R \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ is a local coefficient algebra.

One special property of local coefficient algebras is the following. (Compare the discussion preceding Lemma 2.6.)

**Proposition 4.2.** *Let $R$ be a complete noetherian local domain of characteristic 0 with residue field finite over $\mathbb{F}_p$, and let $S$ be the local coefficient algebra $R \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$.*

(1) *We may naturally identify $(R \widehat{\otimes}_{\mathbb{Z}_p} \mathbf{A}_K) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ with the $p$-adic completion of $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^{\dagger}$.*

(2) *We may naturally identify* $(R \widehat{\otimes}_{\mathbb{Z}_p} \widetilde{\mathbf{A}}_K) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ *with a subring of the* $p$*-adic completion of* $S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger}_K$.

*Proof.* Let $P_{1,n,s}, P_{2,n,s}, P_{3,n,s}$ denote the completed tensor products

$$(R/p^n R) \widehat{\otimes}_{\mathbb{Z}_p} (\mathbf{A}^{\dagger,s}/p^n \mathbf{A}^{\dagger,s})$$

formed using the following choices for the topologies on the two sides.

- For $P_{1,n,s}$, use on the left side the discrete topology, and on the right side the topology induced by $w_s$.
- For $P_{2,n,s}$, use on the left side the topology induced by $\mathfrak{m}_R$, and on the right side the topology induced by $w_s$.
- For $P_{3,n,s}$, use on the left side the topology induced by $\mathfrak{m}_R$, and on the right side the discrete topology.

These constructions relate to our original question as follows. If we take the inverse limit of the $P_{1,n,s}$ as $n \to \infty$, then invert $p$, then take the union over all choices of $s$, we recover $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}$. If we take the union of the $P_{3,n,s}$ over all choices of $s$, then take the inverse limit as $n \to \infty$, and finally invert $p$, we recover $(R \widehat{\otimes}_{\mathbb{Z}_p} \mathbf{A}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$.

To establish (1), it thus suffices to check that the natural maps $P_{1,n,s} \to P_{2,n,s}$ and $P_{3,n,s} \to P_{2,n,s}$ are both bijections. Put $A = \mathfrak{m}_R(R/p^n R)$ and $I = \mathfrak{m}_R A$. Put $B = \mathbf{A}^{\dagger,s}/p^n \mathbf{A}^{\dagger,s}$ and choose an ideal of definition $J \subseteq B$ for the topology induced by $w_s$. In this notation, $A$ is $I$-adically complete and separated, $B$ is $J$-adically complete and separated, and both $A$ and $B$ are flat over $\mathbb{Z}/p^n\mathbb{Z}$. Put $C = A \otimes_{\mathbb{Z}/p^n\mathbb{Z}} B$. The $IC$-adic completion of $C$ is then the inverse limit over $m$ of the quotients $C/I^m C = (A/I^m A) \otimes_{\mathbb{Z}/p^n\mathbb{Z}} B$. Since $B$ is flat over $\mathbb{Z}/p^n\mathbb{Z}$ and $A/I^m A$ has finite cardinality, the completeness of $B$ with respect to $J$ implies the completeness of $C/I^m C$ with respect to $J(C/I^m C)$. It follows that $C$ is complete with respect to $IC + JC$, which means that $P_{1,n,s} \to P_{2,n,s}$ is a bijection. Similarly, we may argue that $P_{3,n,s} \to P_{2,n,s}$ is bijective using the fact that $B/J^m B$ is of finite cardinality.

This yields (1). The whole argument carries over in the case of (2) except for the finiteness of $B/J^m B$; hence in this case, we only have that $P_{1,n,s} \to P_{2,n,s}$ is a bijection and $P_{3,n,s} \to P_{2,n,s}$ is injective. $\square$

**Theorem 4.3.** *Let* $S$ *be a local coefficient algebra. Let* $M_S$ *be an étale* $(\varphi, \Gamma)$*-module over* $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}_K$*, and put*

$$V_S = (M_S \otimes_{S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}_K} (S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger}))^{\varphi=1}.$$

*Then* $V_S$ *is an* $S$*-linear representation for which the natural map* $\mathrm{D}^{\dagger}_K(V_S) \to M_S$ *is an isomorphism.*

*Proof.* By Proposition 4.2(1), we may identify the $p$-adic completion of $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^\dagger$ with $(R \widehat{\otimes}_{\mathbb{Z}_p} \mathbf{A}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$. This allows us to define

$$V'_S = \big(M_S \otimes_{S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^\dagger} ((R \widehat{\otimes}_{\mathbb{Z}_p} \widetilde{\mathbf{A}}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p)\big)^{\varphi=1}.$$

By Theorem 2.2, the natural map

$$V'_S \otimes_{(R \widehat{\otimes}_{\mathbb{Z}_p} \mathbf{A}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p} ((R \widehat{\otimes}_{\mathbb{Z}_p} \widetilde{\mathbf{A}}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p) \to M_S \otimes_{S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}_K^\dagger} ((R \widehat{\otimes}_{\mathbb{Z}_p} \widetilde{\mathbf{A}}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p)$$

is an isomorphism.

By Proposition 4.2(2), we may identify $(R \widehat{\otimes}_{\mathbb{Z}_p} \widetilde{\mathbf{A}}) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$ with a subring of the $p$-adic completion of $S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^\dagger$. Using this identity, we may argue as in [Kedlaya 2008, Proposition 1.2.7] to show that $V'_S \subseteq V_S$, which is enough to establish the desired result. $\square$

## 5. A lifting argument

While one cannot invert the functor $\mathrm{D}_K^\dagger$ for an arbitrary $S$, one can give a partial result.

**Lemma 5.1.** *For any commutative Banach algebra $S$ over $\mathbb{Q}_p$, any $s > 0$, and any $x \in S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger,s}$, the equation*

$$y - \varphi^{-1}(y) = x$$

*has a solution $y \in S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^{\dagger,s}$. More precisely, we may choose $y$ such that $v_p(y) \geq v_p(x)$ and $w_s(y) \geq w_s(x)$.*

*Proof.* For $S = \mathbb{Q}_p$, the existence of a solution $y \in \widetilde{\mathbf{B}}$ follows from the fact that $\widetilde{\mathbf{B}}$ is a complete discretely valued field with algebraically closed residue field. Write $x = \sum_k p^k [x_k]$ and $y = \sum_k p^k [y_k]$. We claim that $y$ can be chosen such that for each $k$,

$$\inf\{v_{\widetilde{\mathbf{E}}}(y_\ell) : \ell \leq k\} \geq \inf\{v_{\widetilde{\mathbf{E}}}(x_\ell) : \ell \leq k\},$$

which yields the desired results. This choice can be made because for any $\bar{x} \in \widetilde{\mathbf{E}}$, the equation $\bar{y} - \bar{y}^{1/p} = \bar{x}$ always has a solution $\bar{y} \in \widetilde{\mathbf{E}}$ with

$$v_{\widetilde{\mathbf{E}}}(\bar{y}) \geq \begin{cases} v_{\widetilde{\mathbf{E}}}(\bar{x}) & \text{if } v_{\widetilde{\mathbf{E}}}(\bar{x}) \leq 0, \\ p v_{\widetilde{\mathbf{E}}}(\bar{x}) & \text{if } v_{\widetilde{\mathbf{E}}}(\bar{x}) \geq 0. \end{cases}$$

For general $S$, write $x$ as a convergent sum $\sum_i u_i \otimes x_i$ with $u_i \in S$ and $x_i \in \widetilde{\mathbf{B}}^{\dagger,s}$. For each $i$, let $y_i \in \widetilde{\mathbf{B}}^{\dagger,s}$ be a solution of $y_i - \varphi^{-1}(y_i) = x_i$ with $w_s(y_i) \geq w_s(x_i)$. Then the sum $y = \sum_i u_i \otimes y_i$ converges with the desired effect. $\square$

**Theorem 5.2.** *Let $S$ be a commutative Banach algebra over $\mathbb{Q}_p$. Let $M_S$ be a free étale $(\varphi, \Gamma)$-module over $S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^\dagger_K$. Suppose that there exists a basis of $M_S$ on which $\varphi - 1$ acts via a matrix whose entries have positive $p$-adic valuation. Then*

$$V_S = (M_S \otimes_{S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^\dagger_K} (S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^\dagger))^{\varphi=1}$$

*is a free $S$-linear representation for which the natural map $\mathrm{D}^\dagger_K(V_S) \to M_S$ is an isomorphism.*

*Proof.* Choose a basis of

$$M'_S = M_S \otimes_{S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^\dagger_K} (S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^\dagger)$$

on which $\varphi - 1$ acts via a matrix $A$ whose entries belong to $S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^{\dagger,s}_K$ for some $s > 0$ and have $p$-adic valuation bounded below by $c > 0$. We may apply Lemma 5.1 to choose a matrix $X$ such that $X$ has entries in $S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^{\dagger,s}_K$ with $p$-adic valuation bounded below by $c$, $\min_{i,j}\{w_s(X_{i,j})\} \geq \min_{i,j}\{w_s(A_{i,j})\}$, and $X - \varphi^{-1}(X) = A$. We can thus change basis to get a new basis of $M'_S$ on which $\varphi - 1$ acts via the matrix

$$(I_n - \varphi^{-1}(X))^{-1}(I_n + A)(I_n - X) - I_n,$$

whose entries have valuation bounded below by $2c$. If we repeat this process, we get a sequence of matrices $X_1, X_2, \dots$ such that $w_s(X_i)$ is bounded below, and the $p$-adic valuation of $X_i$ is at least $ci$. It follows that $w_{s'}(X_i)$ tends to infinity for any $s' > s$, so the product $(I_n + X_1)(I_n + X_2) \cdots$ converges in $S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^{\dagger,s'}_K$ and defines a basis of $M'_S$ fixed by $\varphi$. This proves the claim. □

**Remark 5.3.** The hypothesis about the basis of $M_S$ is needed in Theorem 5.2 for the following reason. For $R$ an arbitrary $\mathbb{F}_p$-algebra, if $\varphi$ acts as the identity on $R$ and as the $p$-power Frobenius on $\widetilde{\mathbf{E}}$, given an invertible square matrix $A$ over $R \otimes_{\mathbb{F}_p} \widetilde{\mathbf{E}}$, we cannot necessarily solve the matrix equation $U^{-1} A \varphi(U) = A$ for an invertible matrix $U$ over $R \otimes_{\mathbb{F}_p} \widetilde{\mathbf{E}}$. For instance, in Chenevier's example, there is no solution of the equation $\varphi(z) = Yz$.

One may wish to view the collection of isomorphism classes of $(\varphi, \Gamma)$-modules over $R \otimes_{\mathbb{F}_p} \mathbb{F}_p((\epsilon - 1))$, for $R$ an $\mathbb{F}_p$-algebra, as the "$R$-valued points of the moduli space of mod $p$ representations of $G_{\mathbb{Q}_p}$." To replace $\mathbb{Q}_p$ with $K$, one should replace $\mathbb{F}_p((\epsilon - 1))$ with the $H_K$-invariants of its separable closure.

## 6. Families of $(\varphi, \Gamma)$-modules and étale models

We turn from $(\varphi, \Gamma)$-modules over $S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^\dagger_K$ to those over $S \mathbin{\widehat{\otimes}}_{\mathbb{Q}_p} \mathbf{B}^\dagger_{\mathrm{rig},K}$. In the absolute case, these have important applications to the study of de Rham representations, as shown by Berger; see for instance [2004]. In the relative case, however, they do not form a robust enough category to be useful; it is better to pass to a more geometric notion. For this, we must restrict to the case where $S$ is an affinoid algebra.

**Definition 6.1.** Let $K$ be a finite extension of $\mathbb{Q}_p$, and let $S$ be an affinoid algebra over $K$. Recall that $\mathcal{R}^s_K$ denotes the ring of Laurent series with coefficients in $K$ in a variable $T$ convergent on the annulus $0 < v_p(T) \leq 1/s$, and that $\mathcal{R}_K = \bigcup_{s>0} \mathcal{R}^s_K$. By a *vector bundle* over $S \widehat{\otimes}_K \mathcal{R}^s_K$, we will mean a coherent locally free sheaf over the product of this annulus with $M(S \otimes_K K)$ in the category of rigid analytic spaces over $K$. (In case $S$ is disconnected, we insist that the rank be constant, not just locally constant.) By a vector bundle over $S \widehat{\otimes}_K \mathcal{R}_K$, we will mean an object in the direct limit as $s \to \infty$ of the categories of vector bundles over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}^s_K$.

Recall that for $s$ sufficiently large, we can produce an isomorphism

$$\mathbf{B}^{\dagger,s}_{\mathrm{rig},K} \cong \mathcal{R}^s_{K'_0}.$$

We thus obtain the notion of a vector bundle over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger,s}_{\mathrm{rig},K}$, dependent on the choice of the isomorphism. However, the notion of a vector bundle over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}_{\mathrm{rig},K}$ does not depend on any choices.

**Remark 6.2.** For $S = K$ discretely valued, every vector bundle over $\mathcal{R}^s_K$ is freely generated by global sections [Kedlaya 2005a, Theorem 3.4.1]. On the other hand, for $S$ an affinoid algebra over $\mathbb{Q}_p$, we do not know whether any vector bundle over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}^s_K$ is $S$-locally free; this does not follow from [Lütkebohmert 1977], which only applies to closed annuli.

**Definition 6.3.** Let $K$ be a finite extension of $\mathbb{Q}_p$, and let $S$ be an affinoid algebra over $\mathbb{Q}_p$. By a *family of* $(\varphi, \Gamma)$*-modules* over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}_{\mathrm{rig},K}$, we mean a vector bundle $V$ over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}_{\mathrm{rig},K}$ equipped with an isomorphism $\varphi^* V \to V$, viewed as a semilinear $\varphi$-action, and a semilinear $\Gamma$-action commuting with the $\varphi$-action. Call a family of $(\varphi, \Gamma)$-modules over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}_{\mathrm{rig},K}$ *étale* if it arises by base extension from an étale $(\varphi, \Gamma)$-module over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}_{K}$, called an *étale model* of the family.

It turns out that étale models are unique when they exist. To check this without any reducedness hypothesis on $S$, we need a generalization of the fact that a reduced affinoid algebra embeds into a product of complete fields [Berkovich 1990, Proposition 2.4.4].

**Lemma 6.4.** *Let $K$ be a finite extension of $\mathbb{Q}_p$, and let $S$ be an affinoid algebra over $\mathbb{Q}_p$. Then there exists a strict inclusion $S \to \prod_{i=1}^n A_i$ of topological rings, in which each $A_i$ is a finite connected algebra over a complete discretely valued field.*

*Proof.* Let $T$ be the multiplicative subset of $\mathcal{O}_S$ consisting of elements whose images in $\mathcal{O}_S/I_S$ are not zero divisors. For any $s \in S$ and $t \in T$, we have $|st| = |s||t|$, so the norm on $S$ extends uniquely to the localization $S[T^{-1}]$. The completion of this localization has the desired form. $\square$

**Proposition 6.5.** *Let $K$ be a finite extension of $\mathbb{Q}_p$, and let $S$ be an affinoid algebra over $\mathbb{Q}_p$. Then the natural base change functor from étale $(\varphi, \Gamma)$-modules over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^\dagger_K$ to families of $(\varphi, \Gamma)$-modules over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^\dagger_{\mathrm{rig},K}$ is fully faithful. In fact, this holds even without the $\Gamma$-action.*

*Proof.* Note that if we replace $S$ by a complete discretely valued field $L$, we may deduce the analogous claim by [Kedlaya 2005b, Theorem 6.3.3] after translating notations. (Families of $(\varphi, \Gamma)$-modules over $\mathbf{B}^\dagger_{\mathrm{rig},K}$ are finite free over $\mathbf{B}^\dagger_{\mathrm{rig},K}$, by Remark 6.2.) In fact, if we replace $S$ by a finite algebra over $L$, we may make the same deduction by restricting scalars to $L$. We may thus deduce the original claim by embedding $S$ into a product of finite algebras over complete discretely valued fields using Lemma 6.4. $\square$

**Corollary 6.6.** *Let $K$ be a finite extension of $\mathbb{Q}_p$, and let $S$ be an affinoid algebra over $\mathbb{Q}_p$. Then an étale model of a family of $(\varphi, \Gamma)$-modules over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^\dagger_{\mathrm{rig},K}$ is unique if it exists.*

**Definition 6.7.** Let $S$ be an affinoid algebra over $\mathbb{Q}_p$. Let $V_S$ be a locally free $S$-linear representation. We define $\mathrm{D}^\dagger_K(V_S)$ as in Definition 3.12, then put

$$\mathrm{D}^\dagger_{\mathrm{rig},K}(V_S) = \mathrm{D}^\dagger_K(V_S) \otimes_{S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^\dagger_K} (S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^\dagger_{\mathrm{rig},K}).$$

This is an étale $(\varphi, \Gamma)$-module over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^\dagger_{\mathrm{rig},K}$, from which we may recover $V_S$ by taking

$$V_S = (\mathrm{D}^\dagger_{\mathrm{rig},K}(V_S) \otimes_{S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^\dagger_{\mathrm{rig},K}} (S \widehat{\otimes}_{\mathbb{Q}_p} \widetilde{\mathbf{B}}^\dagger_{\mathrm{rig}}))^{\varphi=1}.$$

We may now obtain Theorem 0.1 by combining Theorem 4.3 (via Definition 3.12) with Proposition 6.5.

## 7. Local étaleness

We now turn to Theorem 0.2 of the introduction. Given what we already have proven, this can be obtained by invoking some results from [Liu 2008]. For the convenience of the reader, we recall these results in detail.

**Lemma 7.1.** *Let $K$ be a finite extension of $\mathbb{Q}_p$, and let $S$ be an affinoid algebra over $K$. For any $x \in M(S)$ and $\lambda > 0$, there exists an affinoid subdomain $M(B)$ of $M(S)$ containing $x$ such that if $f \in S$ vanishes at $x$, then $|f(y)| \leq \lambda |f|_S$ for any $y \in M(B)$.*

*Proof.* We first prove the lemma for $S = T_n = K\langle x_1, \dots, x_n\rangle$, the $n$-dimensional Tate algebra over $K$. It is harmless to enlarge $K$, so we may suppose without loss of generality that $x$ is the origin $x_1 = \dots = x_n = 0$. Choosing a rational number

$\lambda' < \lambda$, the affinoid subdomain $\{(x_1, \dots, x_n) \in M(S) \,\big|\, |x_1| \le \lambda', \dots, |x_n| \le \lambda'\}$ satisfies the required property.

For general $S$, the reduction $\overline{S} = \mathcal{O}_S/\mathfrak{m}_K\mathcal{O}_S$ is a finite type scheme over the residue field $k$ of $K$. For $n$ sufficiently large, we take a surjective $k$-algebra homomorphism $\overline{\alpha} : k[\overline{x}_1, \dots, \overline{x}_n] \twoheadrightarrow \overline{S}$. We lift $\overline{\alpha}$ to a $K$-affinoid algebra homomorphism $\alpha : K\langle x_1, \dots, x_n\rangle \to S$ by mapping $x_i$ to a lift of $\overline{\alpha}(\overline{x}_i)$ in $\mathcal{O}_S$. Then it follows from Nakayama's lemma that $\alpha$ maps $\mathcal{O}_K\langle x_1, \dots, x_n\rangle$ onto $\mathcal{O}_S$. Let $\alpha$ also denote the induced map from $M(S)$ to $M(K\langle x_1, \dots, x_n\rangle)$. By the case of $K\langle x_1, \dots, x_n\rangle$, we can find an affinoid neighborhood $M(B)$ of $\alpha(x)$ satisfying the required property for $\lambda/p$. Now for any nonzero $f \in S$ vanishing at $x$, we choose $c \in \mathbb{Q}$ such that $|c| \le |f|_S \le p|c|$, yielding $pf/c \in \mathcal{O}_S$. Pick $f' \in \mathcal{O}_K\langle x_1, \dots, x_n\rangle$ such that $\alpha(f') = pf/c$. Then $f'(\alpha(x)) = (pf/c)(x) = 0$ implies that $|f'(y)| \le (\lambda/p)|f'|_{T_n} \le \lambda/p$ for any $y \in M(B)$. Then for any $y \in \alpha^{-1}(M(B))$, we have $|pf(y)|/|c| = |f'(\alpha(y))| \le \lambda/p$, yielding $|f(y)| \le \lambda|c| \le \lambda|f|_S$. Hence $\alpha^{-1}(M(B))$ is an affinoid neighborhood of $x$ satisfying the property we need. $\square$

**Definition 7.2.** For $S$ a commutative Banach algebra over $\mathbb{Q}_p$ and $I$ a subinterval of $\mathbb{R}$, let $\mathcal{R}_S^I$ be the ring of Laurent series over $S$ in the variable $T$ convergent for $v(T)^{-1} \in I$. Let $v_S$ be the valuation on $S$, and for $s \in I$ and $x = \sum_i x_i T^i \in \mathcal{R}_S^I$ put

$$w_s(x) = \inf_i \{i + s v_S(x_i)\}.$$

Put $\mathcal{R}_S^s = \mathcal{R}_S^{[s,+\infty)}$, which we may identify with the completed tensor product $S \,\widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}_{\mathbb{Q}_p}^s$ for the Fréchet topology on the right, and put $\mathcal{R}_S = \bigcup_{s>0} \mathcal{R}_S^s$. Let $\mathcal{R}_S^{\mathrm{int},s}$ be the subring of $\mathcal{R}_S^s$ consisting of series with coefficients in $\mathcal{O}_S$.

**Lemma 7.3** (based on [Kedlaya 2005b, Lemma 6.1.1]). *Let $K$ be a finite extension of $\mathbb{Q}_p$, and let $S$ be an affinoid algebra over $K$. Pick $s_0 > 0$. Let $\varphi : \mathcal{R}_S^{s_0/p} \to \mathcal{R}_S^{s_0}$ be a map of the form $\sum_i c_i T^i \mapsto \sum_i \varphi_S(c_i) W^i$, where $\varphi_S : S \to S$ is an isometry and $W \in \mathcal{R}_S^{s_0}$ satisfies $w_{s_0}(W - T^p) > w_{s_0}(T^p)$. For some $s \ge s_0$, suppose $D$ is an invertible $n \times n$ matrix over $\mathcal{R}_S^{[s,s]}$, and put $h = -w_s(D) - w_s(D^{-1})$; it is clear that $h \ge 0$. Let $F$ be an $n \times n$ matrix over $\mathcal{R}_S^{[s,s]}$ such that $w_s(FD^{-1} - I_n) \ge c + h/(p-1)$ for a positive number $c$. Then for any positive integer $k$ satisfying $2(p-1)sk \le c$, there exists an invertible $n \times n$ matrix $U$ over $\mathcal{R}_S^{[s/p,s]}$ such that $U^{-1}F\varphi(U)D^{-1} - I_n$ has entries in $p^k\mathcal{R}_S^{\mathrm{int},s}$ and $w_s(U^{-1}F\varphi(U)D^{-1} - I_n) \ge c + h/(p-1)$.*

*Proof.* For $i \in \mathbb{R}$, $s > 0$, $f = \sum_{j=-\infty}^{+\infty} a_j T^j \in \mathcal{R}_S$, we set

$$v_i(f) = \min\{j : v_S(a_j) \le i\} \quad \text{and} \quad v_{i,s}(f) = v_i(f) + si.$$

It is clear that $v_{i,s}(f) \ge w_s(f)$. (If $S$ is a field, these quantities are similar to $v_i^{\mathrm{naive}}$, $v_{i,r}^{\mathrm{naive}}$ in [Kedlaya 2005b, p. 458], albeit with a slightly different normalization.)

We define a sequence of invertible matrices $U_0, U_1, \ldots$ over $\mathcal{R}_S^{[s/p,s]}$ and a sequence of matrices $F_0, F_1, \ldots$ over $\mathcal{R}_S^{[s,s]}$ as follows. Set $U_0 = I_n$. Given $U_l$, put $F_l = U_l^{-1} F\varphi(U_l)$. Suppose

$$F_l D^{-1} - I_n = \sum_{m=-\infty}^{\infty} V_m T^m,$$

where the $V_m$'s are $n\times n$ matrices over $S$. Let $X_l = \sum_{v_S(V_m)<k} V_m T^m$, and put $U_{l+1} = U_l(I_n + X_l)$. Set

$$c_l = \inf_{i\leq k-1} \{v_{i,s}(F_l D^{-1} - I_n) - h/(p-1)\}.$$

We now prove by induction that $c_l \geq ((l+1)/2)c$, $w_s(F_l D^{-1} - I_n) \geq c + h/(p-1)$ and $U_l$ is invertible over $\mathcal{R}_S^{[s/p,s]}$ for any $l \geq 0$. This is obvious for $l = 0$. Suppose that the claim is true for some $l \geq 0$. Then for any $t \in [s/p, s]$, since

$$c_l \geq \frac{l+1}{2} c \geq (p-1)sk,$$

we have

$$w_t(X_l) \geq w_s(X_l) - (s-t)k \geq (c_l + h/(p-1)) - (s-t)k > 0.$$

Hence $U_{l+1}$ is also invertible over $\mathcal{R}_S^{[s/p,s]}$. Furthermore, we have

$$\begin{aligned} w_s(D\varphi(X_l)D^{-1}) &\geq w_s(D) + w_s(\varphi(X_l)) + w_s(D^{-1}) = p w_{s/p}(X_l) - h \\ &> p\Big(c_l + \frac{h}{p-1}\Big) - h - (p-1)sk = pc_l + \frac{h}{p-1} - (p-1)sk \\ &\geq c_l + \tfrac{1}{2}c + \frac{h}{p-1} + (\tfrac{1}{2}c - (p-1)sk) \geq \frac{l+2}{2}c + \frac{h}{p-1}, \end{aligned}$$

since $c_l \geq \frac{l+1}{2}c$ by the inductive assumption. Note that

$$\begin{aligned} F_{l+1}D^{-1} - I_n &= (I_n + X_l)^{-1} F_l D^{-1}(I_n + D\varphi(X_l)D^{-1}) - I_n \\ &= ((I_n + X_l)^{-1} F_l D^{-1} - I_n) + (I_n + X_l)^{-1}(F_l D^{-1}) D\varphi(X_l) D^{-1}. \end{aligned}$$

Since $w_s(F_l D^{-1}) \geq 0$ and $w_s((I_n + X_l)^{-1}) \geq 0$, we have

$$w_s((I_n + X_l)^{-1}(F_l D^{-1}) D\varphi(X_l) D^{-1}) \geq \frac{l+2}{2}c + \frac{h}{p-1}.$$

Write

$$\begin{aligned} (I_n + X_l)^{-1} F_l D^{-1} - I_n &= (I_n + X_l)^{-1}(F_l D^{-1} - I_n - X_l) \\ &= \sum_{j=0}^{\infty} (-X_l)^j (F_l D^{-1} - I_n - X_l). \end{aligned}$$

For $j \geq 1$, we have

$$w_s((-X_l)^j(F_l D^{-1} - I_n - X_l)) \geq c + c_l + \frac{2h}{p-1} > \frac{l+2}{2}c + \frac{h}{p-1}.$$

By the definition of $X_l$, we also have $v_i(F_l D^{-1} - I_n - X_l) = \infty$ for $i < k$ and $w_s(F_l D^{-1} - I_n - X_l) \geq c + h/(p-1)$. Putting these together, we get that

$$v_{i,s}(F_{l+1}D^{-1} - I_n) \geq \frac{l+2}{2}c + \frac{h}{p-1}$$

for any $i < k$, that is, $c_{l+1} \geq \dfrac{l+2}{2}c$, and that $w_s(F_{l+1}D^{-1} - I_n) \geq c + \dfrac{h}{p-1}$. The induction step is finished.

Now since $w_t(X_l) \geq c_l + h/(p-1) - (p-1)ps/k$ for $t \in [s/p, s]$, and $c_l \to \infty$ as $l \to \infty$, the sequence $U_l$ converges to a limit $U$, which is an invertible $n \times n$ matrix over $\mathcal{R}_S^{[s/p,s]}$ satisfying $w_s(U^{-1}F\varphi(U)D^{-1} - I_n) \geq c + h/(p-1)$. Furthermore,

$$\begin{aligned} v_{m,s}(U^{-1}F\varphi(U)D^{-1} - I_n) &= \lim_{l\to\infty} v_{m,s}(U_l^{-1}F\varphi(U_l)D^{-1} - I_n) \\ &= \lim_{l\to\infty} v_{m,s}(F_{l+1}D^{-1} - I_n) = \infty, \end{aligned}$$

for any $m < k$. Therefore $U^{-1}F\varphi(U)D^{-1} - I_n$ has entries in $p^k\mathcal{R}_S^{\mathrm{int},s}$. $\square$

**Theorem 7.4.** *Let $S$ be an affinoid algebra over $\mathbb{Q}_p$, and let $M_S$ be a family of $(\varphi, \Gamma)$-modules over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}_{\mathrm{rig},K}$, such that for some $x \in M(S)$ whose residue field is contained in $S$, the fiber $M_x$ of $M_S$ over $x$ is étale. Then there exist an affinoid neighborhood $M(B)$ of $x$ and a finite extension $L$ of $K$ such that the base extension $M_B$ of $M_S$ to $B \widehat{\otimes}_{\mathbb{Q}_p} \mathbf{B}^{\dagger}_{\mathrm{rig},L}$ has an étale model in which the entries of the matrix of $\varphi - 1$ have positive $p$-adic valuation.*

*Proof.* Because Proposition 6.5 does not require the $\Gamma$-action, it suffices to construct an étale model just for the $\varphi$-action. Choose an isomorphism

$$\mathbf{B}^{\dagger,s_0}_{\mathrm{rig},K} \cong \mathcal{R}^{s_0}_{K_0'}$$

for some $s_0 > 0$, via which $\varphi$ induces a map from $\mathcal{R}^{s_0/p}_{K_0'}$ to $\mathcal{R}^{s_0}_{K_0'}$ satisfying

$$w_{s_0}(\varphi(T) - T^p) > w_{s_0}(T^p).$$

Choose $s \geq s_0$ such that $M_S$ is represented by a vector bundle $V_S$ over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}^{s/p}_{K_0'}$ equipped with an isomorphism $\varphi^* V_S \to V_S$ of vector bundles over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}^{s}_{K_0'}$.

By hypothesis, $M_x$ is étale. After increasing $s$, we may therefore assume that $M_x$ admits a basis $e_x$ on which $\varphi$ acts via an invertible matrix over $\mathcal{R}^{\mathrm{int},s}_{S/\mathfrak{m}_x}$. Lift this matrix to a matrix $D$ over $\mathcal{R}^{\mathrm{int},s}_S$, using the inclusion $S/\mathfrak{m}_x \hookrightarrow S$ which was assumed to exist. By enlarging $K$, we can ensure that $D - 1$ has positive $p$-adic valuation (by first doing so modulo $\mathfrak{m}_x$).

By results of Lütkebohmert [1977, Sätze 1 and 2], the restriction of $V_S$ to $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}_{K_0'}^{[s/p,s]}$ is $S$-locally free. By replacing $M(S)$ with an affinoid subdomain containing $x$, we may reduce to the case where this restriction admits a basis $e_S$. Let $A$ be the matrix via which $\varphi$ acts on this basis; it has entries in $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}_{K_0'}^{[s,s]}$. Let $V$ be a matrix over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}_{K_0'}^{[s/p,s]}$ lifting (again using the inclusion $S/\mathfrak{m}_x \hookrightarrow S$) the change-of-basis matrix from the mod-$\mathfrak{m}_x$ reduction of $e_S$ to $e_x$.

By Lemma 7.1, we can shrink $S$ so as to make $D$ invertible over $\mathcal{R}_S^{\mathrm{int},s}$. We can also force $V$ to become invertible, and we may make $V^{-1}A\varphi(V) - D$ as small as desired. We may thus put ourselves in a position to apply Lemma 7.3 with $F = V^{-1}A\varphi(V)$, to produce an invertible $n \times n$ matrix $U$ over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}_{K_0'}^{[s/p,s]}$ such that

$$W = U^{-1}F\varphi(U)D^{-1} - I_n$$

has entries in $p\mathcal{O}_S \widehat{\otimes}_{\mathbb{Z}_p} \mathcal{R}_{K_0'}^{\mathrm{int},s}$ and $w_s(W) > 0$.

Changing basis from $e_S$ via the matrix $VU$ gives another basis $e_S'$ of $V_S$ over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}_{K_0'}^{[s/p,s]}$, on which $\varphi$ acts via the matrix $W + I_n$.

We may change the basis $e_S'$ using $(W + I_n)D$ to get a new basis of $V_S$ over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}_{K_0'}^{[s,ps]}$; since the matrix $(W + I_n)D$ is invertible over $\mathcal{O}_S \widehat{\otimes}_{\mathbb{Z}_p} \mathcal{R}_{K_0'}^{\mathrm{int},s}$, it is also the case that the basis $e_S'$ also generates $V_S$ over $S \widehat{\otimes}_{\mathbb{Q}_p} \mathcal{R}_{K_0'}^{[s,ps]}$. Repeating the argument, we can deduce that $e_S'$ is actually a basis of $V_S$ generating an étale model. This proves the claim. □

Combining Theorem 5.2 with Theorem 7.4 yields Theorem 0.2. Note that before applying Theorem 7.4, we must first extend scalars from $S$ to $S \otimes_{\mathbb{Q}_p} L$ for $L = S/\mathfrak{m}_x$; we then use Galois descent for the action of $\mathrm{Gal}(L/\mathbb{Q}_p)$ to recover a statement about $S$ itself.

**Remark 7.5.** Unfortunately, there is no natural extension of Theorem 7.4 to the Berkovich analytic space $\mathcal{M}(S)$ associated to $S$. For instance, take $K = \mathbb{Q}_p$, $S = \mathbb{Q}_p\langle y\rangle$, and let $M_S$ be free of rank 2 with the action of $\varphi$ given by the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & y/p \end{pmatrix}$$

(in which $T$ does not appear). The locus of $x \in M(S)$ where $M_x$ is étale is precisely the disc $|y| \le |p|$, which does not correspond to an open subset of $\mathcal{M}$.

On the other hand, it may still be the case that $M_S$ is étale if and only if $M_x$ is étale (in an appropriate sense) for each $x \in \mathcal{M}(S)$.

**Remark 7.6.** It should be possible to generalize Berger's construction [2008] to families of filtered $(\varphi, N)$-modules. With such a generalization, one would deduce immediately from Theorem 7.4 that any family of weakly admissible $(\varphi, N)$-modules over an affinoid base (with trivial $\varphi$-action on the base) arises from a Galois representation in a neighborhood of any given rigid analytic point. However, in view of Remark 7.5, we cannot make the corresponding assertion for Berkovich points.

**Remark 7.7.** The families of $(\varphi, \Gamma)$-modules considered here are "arithmetic" in the sense that $\varphi$ acts trivially on the base $S$. They correspond to "arithmetic" families of Galois representations, such as the $p$-adic families arising in the theory of $p$-adic modular forms. There is also a theory of "geometric" families of $(\varphi, \Gamma)$-modules, in which $\varphi$ acts as a Frobenius lift on the base $S$. These correspond to representations of arithmetic fundamental groups via the work of Faltings, Andreatta, Brinon, Iovita, et al. In this theory, one does expect the étale locus to be open, as in [Hartl 2006, Theorem 5.2]. One also expects a family of $(\varphi, \Gamma)$-modules to be globally étale if and only if it is étale over each Berkovich point (but not if it is only étale over each rigid point, as shown by the Rapoport–Zink spaces). We hope to consider this question in subsequent work.

## Acknowledgments

Kedlaya thanks Siegfried Bosch for helpful comments. Liu thanks Pierre Colmez for the helpful conversation in spring of 2008 that inspired this project, Laurent Berger for answering many questions patiently, and Marie-France Vignéras for arranging his visit to the Institut de Mathématiques de Jussieu in spring of 2008.

## References

[Berger 2004] L. Berger, "An introduction to the theory of $p$-adic representations", pp. 255–292 in *Geometric aspects of Dwork theory*, vol. 1, edited by A. Adolphson et al., de Gruyter, Berlin, 2004. MR 2005h:11265 Zbl 1118.11028

[Berger 2008] L. Berger, "Équations différentielles $p$-adiques et $(\phi, N)$-modules filtrés", pp. 13–38 in *Représentations p-adiques de groupes p-adiques, I: Représentations galoisiennes et* $(\phi, \Gamma)$-*modules*, edited by L. Berger et al., Astérisque **319**, 2008. MR 2010d:11056 Zbl 1168.11019

[Berger and Colmez 2008] L. Berger and P. Colmez, "Familles de représentations de de Rham et monodromie $p$-adique", pp. 303–337 in *Représentations p-adiques de groupes p-adiques, I: Représentations galoisiennes et* $(\phi, \Gamma)$-*modules*, edited by L. Berger et al., Astérisque **319**, 2008. MR 2010g:11091 Zbl 1168.11020

[Berkovich 1990] V. G. Berkovich, *Spectral theory and analytic geometry over non-Archimedean fields*, Mathematical Surveys and Monographs **33**, American Mathematical Society, Providence, RI, 1990. MR 91k:32038 Zbl 0715.14013

[Bosch et al. 1984] S. Bosch, U. Güntzer, and R. Remmert, *Non-Archimedean analysis*, Grundlehren der Mathematischen Wissenschaften **261**, Springer, Berlin, 1984. MR 86b:32031

[Chenevier 2009] G. Chenevier, "Une application des variétés de Hecke des groups unitaires", preprint, 2009, Available at http://www.math.polytechnique.fr/~chenevier/articles/famgal.pdf.

[Cherbonnier and Colmez 1998] F. Cherbonnier and P. Colmez, "Représentations $p$-adiques surconvergentes", *Invent. Math.* **133**:3 (1998), 581–611. MR 2000d:11146 Zbl 0928.11051

[Colmez 2008] P. Colmez, "Représentations triangulines de dimension 2", pp. 213–258 in *Représentations p-adiques de groupes p-adiques, I: Représentations galoisiennes et* $(\phi, \Gamma)$*-modules*, edited by L. Berger et al., Astérisque **319**, 2008. MR 2010f:11173 Zbl 1168.11022

[Dee 2001] J. Dee, "$\Phi$-$\Gamma$ modules for families of Galois representations", *J. Algebra* **235**:2 (2001), 636–664. MR 2001m:12012 Zbl 0984.11062

[Fontaine 1990] J.-M. Fontaine, "Représentations $p$-adiques des corps locaux, I", pp. 249–309 in *The Grothendieck Festschrift*, vol. 2, edited by P. Cartier et al., Progr. Math. **87**, Birkhäuser, Boston, 1990. MR 92i:11125 Zbl 0743.11066

[Hartl 2006] U. Hartl, "On a conjecture of Rapoport and Zink", preprint, 2006. arXiv math/0605254

[Kedlaya 2005a] K. S. Kedlaya, "Local monodromy of $p$-adic differential equations: an overview", *Int. J. Number Theory* **1**:1 (2005), 109–154. MR 2006g:12013 Zbl 1107.12005

[Kedlaya 2005b] K. S. Kedlaya, "Slope filtrations revisited", *Doc. Math.* **10** (2005), 447–525. MR 2007c:14016 Zbl 1081.14028

[Kedlaya 2008] K. S. Kedlaya, "Slope filtrations for relative Frobenius", pp. 259–301 in *Représentations p-adiques de groupes p-adiques, I: Représentations galoisiennes et* $(\phi, \Gamma)$*-modules*, edited by L. Berger et al., Astérisque **319**, 2008. MR 2010c:14024 Zbl 1168.11053

[Liu 2008] R. Liu, "Slope filtrations in families", preprint, 2008. To appear in *Journal de l'Institut de Math. de Jussieu*.

[Lütkebohmert 1977] W. Lütkebohmert, "Vektorraumbündel über nichtarchimedischen holomorphen Räumen", *Math. Z.* **152**:2 (1977), 127–143. MR 55 #3336

[May 2003] J. P. May, "Munshi's proof of the Nullstellensatz", *Amer. Math. Monthly* **110**:2 (2003), 133–140. MR 1952440 Zbl 1053.13010

Communicated by Brian Conrad
Received 2009-12-10 Accepted 2010-01-10

kedlaya@math.mit.edu — *Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States*
http://www-math.mit.edu/~kedlaya/

ruochuan@umich.edu — *Department of Mathematics, University of Michigan, 1844 East Hall, 530 Church Street, Ann Arbor, MI 48109-1043, United States*
http://www-personal.umich.edu/~ruochuan

## Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

**Originality**. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

**Language**. Articles in ANT are usually in English, but articles written in other languages are welcome.

**Required items**. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

**Format**. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

**References**. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

**Figures**. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@mathscipub.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 4 No. 7 2010