Pacific Journal of Mathematics

Volume 281 No. 2 April 2016

PACIFIC JOURNAL OF MATHEMATICS
msp.org/pjm

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

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

The subscription price for 2016 is US $440/year for the electronic version, and $600/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

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

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

PACIFIC JOURNAL OF MATHEMATICS
Vol. 281, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.281.257

THE EISENSTEIN ELEMENTS OF MODULAR SYMBOLS FOR LEVEL PRODUCT OF TWO DISTINCT ODD PRIMES

DEBARGHA BANERJEE AND SRILAKSHMI KRISHNAMOORTHY

We explicitly write down the Eisenstein elements inside the space of modular symbols for Eisenstein series with integer coefficients for the congruence subgroups $\Gamma_0(pq)$ with p and q distinct odd primes, giving an answer to a question of Merel in these cases. We also compute the winding elements explicitly for these congruence subgroups. Our results are explicit versions of the Manin–Drinfeld theorem.

1. Introduction

In his landmark paper on Eisenstein ideals, Mazur studied torsion points of elliptic curves over $\mathbb{Q}$ and gave a list of possible torsion subgroups of elliptic curves (see [Mazur 1977, Theorem 8]). Merel [1996b] wrote down modular symbols for the congruence subgroups $\Gamma_0(p)$ for any odd prime p that correspond to differential forms of the third kind on the modular curves. He then used these modular symbols to give a uniform upper bound on the torsion points of elliptic curves over any number field in terms of its extension degree [Merel 1996a]. The explicit expressions of winding elements for prime level of [Merel 1996b] were used by Calegari and Emerton [2005] to study the ramifications of Hecke algebras at the Eisenstein primes. Several authors afterwards studied the torsion points of elliptic curves over number fields using modular symbols.

In the present paper, we study elements of relative homology groups of the modular curve $X_0(pq)$ that correspond to differential forms of the third kind with p and q distinct odd primes. As a consequence, we give an "effective" proof of the Manin–Drinfeld theorem (Theorem 9) for the special case of the image in $\mathrm{H}_1(X_0(pq), \mathbb{R})$ of the path in $\mathrm{H}_1(X_0(pq), \partial(X_0(pq)), \mathbb{Z})$ joining 0 and $i\infty$. Since the algebraic parts of the special values of the L-function are obtained by integrating differential forms on these modular symbols, our explicit expression of the winding elements should be useful for understanding the algebraic parts of the special values at 1 of the L-functions of the quotient Jacobian of modular curves for the congruence subgroup $\Gamma_0(pq)$ [Agashe 2000].

MSC2010: primary 11F67; secondary 11F11, 11F20, 11F30.
Keywords: Eisenstein series, modular symbols, special values of L-functions.

For $N \in \{p, q, pq\}$, consider the basis E_N of $E_2(\Gamma_0(pq))$ (Section 4) for which all the Fourier coefficients at $i\infty$ belong to $\mathbb{Z}$. The meromorphic differential forms $E_N(z)\,dz$ are of the third kind on the Riemann surface $X_0(pq)$ but of the first kind on the noncompact Riemann surface $Y_0(pq)$.

Let $\xi : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{H}_1(X_0(pq), \text{cusps}, \mathbb{Z})$ be the Manin map (Section 3). For any two coprime integers u and v with $v \geq 1$, let $S(u, v) \in \mathbb{Z}$ be the Dedekind sum (see Section 4.1). If $g \in \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})$ is not of the form $(\pm 1, 1)$, $(\pm 1 \pm kx, 1)$ or $(1, \pm 1 \pm kx)$ with x one of the primes p or q, then we can write it as $(r-1, r+1)$.

Let δ_r be 1 or 0 depending on whether r is odd or even. For any integer k, let $s_k = k + (\delta_k - 1)pq$ be an odd integer. Choose integers s, s' and l, l' such that $l(s_k x + 2) - 2spq = 1$ and $l' s_k x - 2s' pq/x = 1$. Let

$$\gamma_1^{x,k} = \begin{pmatrix} 1+4spq & -2l \\ -4s(s_k x + 2)pq & 1+4spq \end{pmatrix} \quad \text{and} \quad \gamma_2^{x,k} = \begin{pmatrix} 1+4s'pq/x & -2l' \\ -4s'(s_k)pq & 1+4s'pq/x \end{pmatrix}$$

be two matrices (see Lemma 28). For $l = 1, 2$, consider the integers

$$\begin{aligned} P_N(\gamma_l^{x,k}) = \operatorname{sgn}(t(\gamma_l^{x,k}))\Bigg(2\Big(S\big(s(\gamma_l^{x,k}), |t(\gamma_l^{x,k})|N\big) - S\big(s(\gamma_l^{x,k}), |t(\gamma_l^{x,k})|\big)\Big) \\ - S\big(s(\gamma_l^{x,k}), \tfrac{1}{2}|t(\gamma_l^{x,k})|N\big) + S\big(s(\gamma_l^{x,k}), \tfrac{1}{2}|t(\gamma_l^{x,k})|\big)\Bigg) \end{aligned}$$

with

$$s(\gamma_1^{x,k}) = 1 - 4spq(1 + s_k x), \quad t(\gamma_1^{x,k}) = -2(l - 2s(s_k x + 2)pq)$$

and

$$s(\gamma_2^{x,k}) = 1 - 4s'pq\left(s_k - \frac{1}{x}\right), \quad t(\gamma_2^{x,k}) = -2(l' - 2s' s_k pq).$$

Define the function $F_N : \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z}) \to \mathbb{Z}$ by

$$F_N(g) = \begin{cases} 2(S(r, N) - 2S(r, 2N)) & \text{if } g = (r-1, r+1), \\ P_N(\gamma_1^{x,k}) - P_N(\gamma_2^{x,k}) & \text{if } g = (1+kx, 1) \text{ or } g = (-1-kx, 1), \\ -P_N(\gamma_1^{x,k}) + P_N(\gamma_2^{x,k}) & \text{if } g = (1, 1+kx) \text{ or } g = (1, -1-kx), \\ 0 & \text{if } g = (\pm 1, 1). \end{cases}$$

Theorem 1. *The modular symbol*

$$\mathcal{E}_{E_N} = \sum_{g \in \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})} F_{E_N}(g)\xi(g)$$

in $\mathrm{H}_1(X_0(pq), \partial(X_0(pq)), \mathbb{Z})$ *is the Eisenstein element* (Section 5) *corresponding to the Eisenstein series* $E_N \in E_2(\Gamma_0(pq))$.

In [Banerjee 2014], a description is given of Eisenstein elements in terms of certain integrals for $M = p^2$. In this article, we give an explicit description in terms of two matrices $\gamma_1^{x,k}$ and $\gamma_2^{x,k}$. Let $\bar{B}_1 : \mathbb{R} \to \mathbb{R}$ be the periodic first Bernoulli polynomial. For the Eisenstein series E_{pq} (Section 4), we write down the Eisenstein elements more explicitly if $g = (r-1, r+1)$. Replacing p with pq [Merel 1996b, Lemma 4], we write

$$F_{pq}(r-1, r+1) = \sum_{h=0}^{pq-1} \bar{B}_1\left(\frac{hr}{2pq}\right).$$

Recall the concept of the *winding elements* (Definition 37). We write down the explicit expression of the winding elements for the congruence subgroup $\Gamma_0(pq)$.

Corollary 2.

$$(1-pq)e_{pq} = \sum_{x \in (\mathbb{Z}/pq\mathbb{Z})^*} F_{pq}(1, x)\left\{0, \frac{1}{x}\right\}.$$

Note that if $\nu = \gcd(pq-1, 12)$ and $n = (pq-1)/\nu$, then a multiple of winding element ne_{pq} belongs to $\mathrm{H}_1(X_0(pq), \mathbb{Z})$. Manin and Drinfeld proved that the modular symbol $\{0, \infty\}$ belongs to $\mathrm{H}_1(X_0(N), \mathbb{Q})$ using the theory of suitable Hecke operators acting on the modular curve $X_0(N)/\mathbb{Q}$. In this paper, we follow the approach of Merel [1996b, Proposition 11]. Our explicit expression of winding elements should be useful for understanding the algebraic part of the *special values of L-functions* (see [Agashe 2000, p. 26]).

Since Hecke operators are defined over $\mathbb{Q}$, there is a possibility that we can find the Eisenstein elements for the congruence subgroups of odd level in a completely different method without using boundary computations. It is tempting to remark that our method should generalize to the congruence subgroup $\Gamma_0(N)$ at least if N is squarefree and odd. Unfortunately, generalizing our method is equivalent to having an explicit understanding of boundary homologies of modular curves defined over rationals. For instance, if $N = pqr$ with p, q, r three distinct primes then there are eight cusps. Since there are more cusps in these cases, the computation of boundaries becomes much more tedious. One of the authors wishes to tackle the difficulty using the "level" of the cusps in a future article.

2. Acknowledgements

This paper owes its existence to several e-mail communications, encouragements and valuable suggestions from Professor Loïc Merel. We would like to thank IMSc, Chennai, for providing excellent working conditions. Krishnamoorthy would like to thank MPIM, Germany, for the great hospitality during her visit. She was supported by a DST-INSPIRE grant. We wish to thank the anonymous referee for several useful comments that helped us improve the mathematical content of this article.

3. Modular symbols

Let $\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q}) = \overline{\mathbb{H}}$ and let $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ be a congruence subgroup. The topological space $X_\Gamma(\mathbb{C}) = \Gamma \backslash \overline{\mathbb{H}}$ has a natural structure of a smooth compact Riemann surface. Consider the usual projection map $\pi : \overline{\mathbb{H}} \to X_\Gamma(\mathbb{C})$ and recall that it is unramified outside the elliptic points and the set of cusps $\partial(X_\Gamma)$. Both these sets are finite.

3.1. ***Rational structure of the curve $X_0(N)$ defined over $\mathbb{Q}$.*** There is a smooth projective curve $X_0(N)$ defined over $\mathbb{Q}$ for which the space $\Gamma_0(N)\backslash\overline{\mathbb{H}}$ is canonically identified with the set of $\mathbb{C}$-points of the projective curve $X_0(N)$. We are interested in understanding the $\mathbb{Q}$-structure of the compactified modular curve $X_0(N)$.

3.2. ***Classical modular symbols.*** Recall the following fundamental theorem.

Theorem 3 [Manin 1972]. *For $\alpha \in \overline{\mathbb{H}}$, consider the map $c : \Gamma \to \mathrm{H}_1(X_0(N), \mathbb{Z})$ defined by*

$$c(g) = \{\alpha, g\alpha\}.$$

The map c is a surjective group homomorphism which does not depend on the choice of point α. The kernel of this homomorphism is generated by

(1) *the commutator*,

(2) *the elliptic elements*,

(3) *the parabolic elements*

of the congruence subgroup Γ.

In particular, this theorem implies that $\{\alpha, g\alpha\} = 0$ for all $\alpha \in \mathbb{P}^1(\mathbb{Q})$ and $g \in \Gamma$.

3.3. ***Manin map.*** Let $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $R = ST = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)$. The modular group $\mathrm{SL}_2(\mathbb{Z})$ is generated by S and T.

Theorem 4 [Manin 1972]. *Let*

$$\xi : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{H}_1(X_0(pq), \partial(X_0(pq)), \mathbb{Z})$$

be the map that takes a matrix $g \in \mathrm{SL}_2(\mathbb{Z})$ to the class in $\mathrm{H}_1(X_0(pq), \partial(X_0(pq)), \mathbb{Z})$ of the image in $X_0(pq)$ of the geodesic in $\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ joining $g0$ and $g\infty$. Then

- *the map ξ is surjective*;
- *for all $g \in \Gamma_0(pq)\backslash\mathrm{SL}_2(\mathbb{Z})$, we have $\xi(g) + \xi(gS) = 0$ and $\xi(g) + \xi(gR) + \xi(gR^2) = 0$.*

We have a short exact sequence

$$0 \to \mathrm{H}_1(X_0(pq), \mathbb{Z}) \to \mathrm{H}_1(X_0(pq), \partial(X_0(pq)), \mathbb{Z}) \to \mathbb{Z}^{\partial(X_0(pq))} \xrightarrow{\delta'} \mathbb{Z} \to 0.$$

The first map is a canonical injection. The boundary map δ' takes a geodesic, joining the cusps r and s to the formal symbol $[r]-[s]$, and the third map is the sum of the coefficients.

3.4. ***Relative homology group*** $\mathbf{H_1(X_0(pq)-R\cup I, \partial(X_0(pq)), \mathbb{Z})}$**.** Consider the points $i=\sqrt{-1}$ and $\rho=\frac{1}{2}(1+\sqrt{-3})$ on the complex upper half-plane with ν the geodesic joining i and ρ. These are the elliptic points on the Riemann surface $X_0(pq)$. The projection map π is unramified outside cusps and elliptic points.

Say $R=\pi(\mathrm{SL}_2(\mathbb{Z})\rho)$ and let $I=\pi(\mathrm{SL}_2(\mathbb{Z})i)$ be the image of these two sets in $X_0(pq)$. These two sets are disjoint. Consider now the relative homology group $\mathrm{H}_1(Y_0(pq), R\cup I, \mathbb{Z})$. For $g\in\mathrm{SL}_2(\mathbb{Z})$, let $[g]_*$ be the class of $\pi(g\nu)$ in the relative homology group $\mathrm{H}_1(Y_0(pq), R\cup I, \mathbb{Z})$. Let $\rho^*=-\bar{\rho}$ be another point on the boundary of the fundamental domain. The homology groups $\mathrm{H}_1(Y_0(pq), \mathbb{Z})$ are subgroups of $\mathrm{H}_1(Y_0(pq), R\cup I, \mathbb{Z})$. Suppose $z_0\in\mathbb{H}$ is such that $|z_0|=1$ and $\frac{-1}{2}<\mathrm{Re}(z_0)<1$. Let γ be the union of the geodesics in $\mathbb{H}\cup\mathbb{P}^1(\mathbb{Q})$ joining 0 to z_0 and z_0 to $i\infty$. For $g\in\Gamma_0(pq)\backslash\mathrm{SL}_2(\mathbb{Z})$, let $[g]^*$ be the class of $\pi(g\gamma)$ in $\mathrm{H}_1(X_0(pq)-R\cup I, \partial(X_0(pq)), \mathbb{Z})$.

We have an intersection pairing

$$\circ: \mathrm{H}_1(X_0(pq)-R\cup I, \partial(X_0(pq)), \mathbb{Z})\times \mathrm{H}_1(Y_0(pq), R\cup I, \mathbb{Z})\to\mathbb{Z}.$$

Recall the following results.

Proposition 5 [Merel 1996b; 1995, Proposition 1]. *For* $g,h\in\Gamma_0(pq)\backslash\mathrm{SL}_2(\mathbb{Z})$,

$$[g]^*\circ[h]_*=\begin{cases}1 & \textit{if } \Gamma_0(pq)g=\Gamma_0(pq)h,\\ 0 & \textit{otherwise.}\end{cases}$$

Corollary 6 [Merel 1995, Corollary 1]. *The homomorphism of groups* $\mathbb{Z}^{\Gamma_0(pq)\backslash\mathrm{SL}_2(\mathbb{Z})}$ $\to\mathrm{H}_1(Y_0(pq), R\cup I, \mathbb{Z})$ *induced by the map*

$$\xi_0\Big(\sum_g \mu_g g\Big)=\sum_g \mu_g[g]_*$$

is an isomorphism.

The following important property of the intersection pairing will be used later.

Corollary 7 [Merel 1995, Corollary 3]. *For* $g\in\Gamma_0(pq)\backslash\mathrm{SL}_2(\mathbb{Z})$, *let* $\sum_h\mu_h h\in\mathbb{Z}^{\Gamma_0(pq)\backslash\mathrm{SL}_2(\mathbb{Z})}$ *be such that* $\sum_h\mu_h[h]_*$ *is the image of an element of* $\mathrm{H}_1(Y_0(pq), \mathbb{Z})$ *under the canonical injection. We have*

$$[g]^*\circ\Big(\sum_h\mu_h[h]_*\Big)=\mu_g.$$

We have a short exact sequence

$$0 \to \mathrm{H}_1(X_0(pq) - R \cup I, \mathbb{Z}) \\ \to \mathrm{H}_1(X_0(pq) - R \cup I, \partial(X_0(pq)), \mathbb{Z}) \to \mathbb{Z}^{\{\partial(X_0(pq))\}} \xrightarrow{\delta} \mathbb{Z} \to 0.$$

The boundary map δ takes a geodesic, joining the cusps r and s to the formal symbol $[r] - [s]$. Note that $\delta'(\xi(g)) = \delta([g]^*)$ for all $g \in \mathrm{SL}_2(\mathbb{Z})$.

Recall that we have a canonical bijection $\Gamma_0(pq)\backslash\mathrm{SL}_2(\mathbb{Z}) \cong \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})$ given by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \to (c, d)$. Say

$$\alpha_k = \begin{pmatrix} 0 & -1 \\ 1 & k \end{pmatrix}, \quad \beta_r = \begin{pmatrix} -1 & -r \\ p & rp-1 \end{pmatrix} \quad \text{and} \quad \gamma_s = \begin{pmatrix} -1 & -s \\ q & sq-1 \end{pmatrix}.$$

We explicitly write down the elements of $\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})$ as the set

$$\{(1, k), (1, tp), (1, t'q), (p, q), (q, p), (tp, 1), (t'q, 1), (1, 0), (0, 1)\}$$

with $k \in (\mathbb{Z}/pq\mathbb{Z})^*$, $t \in (\mathbb{Z}/q\mathbb{Z})^*$, $t' \in (\mathbb{Z}/p\mathbb{Z})^*$. Observe that $(p, q) = (tp, q) = (p, t'q)$ for all t and t' coprime to pq.

Lemma 8. *The set*

$$\Omega = \{I, \alpha_k, \beta_r, \gamma_s \mid 0 \le k \le pq-1,\ 0 \le r \le p-1,\ 0 \le s \le q-1\}$$

forms a complete set of coset representatives of $\Gamma_0(pq)\backslash\mathrm{SL}_2(\mathbb{Z})$.

Proof. The orbits $\Gamma_0(pq)\alpha_k$, $\Gamma_0(pq)\beta_l$ and $\Gamma_0(pq)\gamma_m$ are disjoint since ab^{-1} does not belong to $\Gamma_0(pq)$ for two distinct matrices a and b from the set Ω. There are $1 + pq + p + q = |\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})|$ coset representatives. □

We list different rational numbers of the form $\alpha(0)$ and $\alpha(\infty)$ with $\alpha \in \Omega$ as equivalence classes of cusps as follows:

0	$1/p$	$1/q$
$\frac{-l}{lp-1}$, $(lp-1, q) = 1$	$\frac{-1}{k}$, $(k, p) > 1$	$\frac{-1}{k}$, $(k, q) > 1$
$\frac{-m}{mq-1}$, $(mq-1, p) = 1$	$\frac{-m}{mq-1}$, $(mq-1, p) > 1$	$\frac{-l}{lp-1}$, $(lp-1, q) > 1$

3.5. *Manin–Drinfeld theorem.* Following [Lang 1995], we briefly recall the statement of the Manin–Drinfeld theorem.

Theorem 9 (Manin–Drinfeld [Drinfeld 1973]). *For a congruence subgroup* Γ *and any two cusps* α *and* β *in* $\mathbb{P}^1(\mathbb{Q})$, *the path*

$$\{\alpha, \beta\} \in \mathrm{H}_1(X_\Gamma, \mathbb{Q}).$$

This theorem can be reformulated in terms of divisor classes on the Riemann surface.

Theorem 10. *Let $a = \sum_i m_i P_i$ be a divisor of degree zero on X. Then a is a divisor of a rational function if and only if there exists a cycle $\sigma \in \mathrm{H}_1(X_\Gamma, \mathbb{Z})$ such that*

$$\int_a \omega = \sum_i m_i \int_{P_0}^{P_i} \omega = \int_\sigma \omega$$

for every $\omega \in \mathrm{H}^0(X_\Gamma, \Omega_{X_\Gamma})$.

As a corollary, we notice that $\{x, y\} \in \mathrm{H}_1(X_\Gamma, \mathbb{Q})$ if and only if there is a positive integer m such that $m(\pi_\Gamma(x) - \pi_\Gamma(y))$ is a divisor of a function. In other words, the degree-zero divisors supported on the cusps are of finite order in the divisor class group. Manin and Drinfeld proved it using the extended action of the usual Hecke operators. In particular, it says that $\{0, \infty\} \in \mathrm{H}_1(X_\Gamma, \mathbb{Q})$ although 0 and ∞ are two inequivalent cusps of X_Γ. Ogg [1974] constructed a certain modular function $X_0(pq)$ whose divisors coincide with degree-zero divisors on the modular curves.

4. Eisenstein series for $\Gamma_0(pq)$ with integer coefficients

Let $\sigma_1(n)$ denote the sum of the positive divisors of n. We consider the series

$$E_2'(z) = 1 - 24\biggl(\sum_n \sigma_1(n) e^{2\pi i n z}\biggr).$$

Let Δ be the Ramanujan cusp form of weight 12. For all $N \in \mathbb{N}$, the function $z \to \Delta(Nz)/\Delta(z)$ is a function on $\mathbb{H}$ invariant under $\Gamma_0(N)$. The logarithmic differential of this function is $2\pi i E_N(z)\,dz$ and E_N is a classical holomorphic modular form of weight two for $\Gamma_0(N)$ with constant term $N - 1$. The differential form $E_N(z)\,dz$ is a differential form of the third kind on $X_0(N)$. The periods (Section 4.1) of these differential forms are in $\mathbb{Z}$.

By [Diamond and Shurman 2005, Theorem 4.6.2], the set $\mathbb{E}_{pq} = \{E_p, E_q, E_{pq}\}$ is a basis of $E_2(\Gamma_0(pq))$.

Lemma 11. *The cusps $\partial(X_0(pq))$ can be identified with the set $\{0, \infty, 1/p, 1/q\}$.*

Proof. If a/c and a'/c' are in $\mathbb{P}^1(\mathbb{Q})$, then

$$\Gamma_0(pq)\frac{a}{c} = \Gamma_0(pq)\frac{a'}{c'} \quad \Longleftrightarrow \quad \begin{pmatrix} ay \\ c \end{pmatrix} \equiv \begin{pmatrix} a' + jc' \\ c'y \end{pmatrix} \pmod{pq}$$

for some j and y such that $\gcd(y, pq) = 1$ (see [Diamond and Shurman 2005, p. 99]). A small check shows that the orbits $\Gamma_0(pq)0$, $\Gamma_0(pq)\infty$, $\Gamma_0(pq)1/p$ and $\Gamma_0(pq)1/q$ are disjoint. $\square$

Let $\mathrm{Div}^0(X_0(pq), \partial(X_0(pq)), \mathbb{Z})$ be the group of degree-zero divisors supported on cusps. For all cusps x, let $e_{\Gamma_0(pq)}(x)$ denote the ramification index of x over $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ and let

$$r_{\Gamma_0(pq)}(x) = e_{\Gamma_0(pq)}(x) a_0(E[x]).$$

By [Stevens 1982, p. 23], there is a canonical isomorphism $\delta : E_2(\Gamma_0(pq)) \to \mathrm{Div}^0(X_0(pq), \partial(X_0(pq)), \mathbb{Z})$ that takes the Eisenstein series E to the divisor

$$\delta(E) = \sum_{x \in \Gamma_0(pq)\backslash\mathbb{P}^1(\mathbb{Q})} r_{\Gamma_0(pq)}(x)[x]. \tag{4-1}$$

Hence, the Eisenstein element is related to the Eisenstein series by the boundary map. In Proposition 34, we prove that the boundary of the Eisenstein element is indeed the boundary of the Eisenstein series. By [Stevens 1985, p. 538], we see that

$$e_{\Gamma_0(pq)}(x) = \begin{cases} q & \text{if } x = 1/p, \\ p & \text{if } x = 1/q, \\ 1 & \text{if } x = \infty, \\ pq & \text{if } x = 0. \end{cases}$$

Since $\sum_{x \in \partial(X_0(pq))} e_{\Gamma_0(pq)}(x) a_0(E[x]) = 0$, we write the corresponding degree-zero divisor as

$$\delta(E) = a_0(E)(\{\infty\} - \{0\}) + q a_0\left(E\left[\frac{1}{p}\right]\right)\left(\left\{\frac{1}{p}\right\} - \{0\}\right) \\ + p a_0\left(E\left[\frac{1}{q}\right]\right)\left(\left\{\frac{1}{q}\right\} - \{0\}\right).$$

4.1. ***Period homomorphisms.*** We now define period homomorphisms for differential forms of the third kind.

Definition 12 (period homomorphism). For $E_N \in \mathbb{E}_{pq}$, the differential forms $E_N(z)\,dz$ are of the third kind on the Riemann surface $X_0(pq)$ but of the first kind on the noncompact Riemann surface $Y_0(N)$. For any $z_0 \in \mathbb{H}$ and $\gamma \in \Gamma_0(pq)$, let $c(\gamma)$ be the class in $\mathrm{H}_1(Y_0(pq), \mathbb{Z})$ of the image in $Y_0(pq)$ of the geodesic in $\mathbb{H}$ joining z_0 and $\gamma(z_0)$. That the class is nonzero follows from Theorem 3. This class is independent of the choice of $z_0 \in \mathbb{H}$. Let $\pi_{E_N}(\gamma) = \int_{c(\gamma)} E_N(z)\,dz$. The map $\pi_{E_N} : \Gamma_0(pq) \to \mathbb{Z}$ is the "period" homomorphism of E_N.

Let $\bar{B}_1(x)$ be the first Bernoulli polynomial of period one defined by

$$\bar{B}_1(0) = 0, \quad \bar{B}_1(x) = x - \tfrac{1}{2}$$

if $x \in (0, 1)$. For any two integers u and v with $v \geq 1$, we define the Dedekind sum

$$S(u, v) = \sum_{t=1}^{v-1} \bar{B}_1\Big(\frac{tu}{v}\Big)\bar{B}_1\Big(\frac{u}{v}\Big).$$

Recall some well-known properties of the period mapping π_{E_N} (see [Mazur 1979, p. 10; Merel 1996b, p. 14]) for the Eisenstein series $E_N \in \mathbb{E}_{pq}$.

Proposition 13. *Let* $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ *be an element of* $\Gamma_0(pq)$. *Then*

(1) π_{E_N} *is a homomorphism* $\Gamma_0(pq) \to \mathbb{Z}$;

(2) *the image of* π_{E_N} *lies in* $\mu\mathbb{Z}$, *where* $\mu = \gcd(N-1, 12)$;

(3) $$\pi_{E_N}(\gamma) = \begin{cases} \dfrac{a+d}{c}(N-1) + 12\,\mathrm{sgn}(c)\Big(S(d, |c|) - S\Big(d, \dfrac{|c|}{N}\Big)\Big) & \textit{if } c \neq 0, \\ \dfrac{b}{d}(N-1) & \textit{if } c = 0; \end{cases}$$

(4) $\pi_{E_N}(\gamma) = \pi_{E_N}\left(\begin{pmatrix} d & c/N \\ Nb & a \end{pmatrix}\right)$.

5. Eisenstein elements

Following [Merel 1996b] and [Merel 1993], we recall the concept of Eisenstein elements of the space of modular symbols. For any natural number $M > 4$, the congruence subgroup $\Gamma_0(M)$ is the subgroup of $\mathrm{SL}_2(\mathbb{Z})$ consisting of all matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ such that $M \mid c$. The congruence subgroup $\Gamma_0(M)$ acts on the upper half-plane $\mathbb{H}$ in the usual way. The quotient space $\Gamma_0(M)\backslash\mathbb{H}$ is denoted by $Y_0(M)$. A priori, these are all Riemann surfaces and hence algebraic curves defined over $\mathbb{C}$. There are models of these algebraic curves defined over $\mathbb{Q}$ and they parametrize elliptic curves with cyclic subgroups of order M. Let $X_0(M)$ be the compactification of the Riemann surface $Y_0(M)$ obtained by adjoining the set of cusps $\partial(X_0(M)) = \Gamma_0(M)\backslash\mathbb{P}^1(\mathbb{Q})$.

Definition 14 (Eisenstein elements). Let $\pi_{E_N} : \mathrm{H}_1(Y_0(pq), \mathbb{Z}) \to \mathbb{Z}$ be the period homomorphism of E_N (Section 4.1). The intersection pairing $\circ$ [Merel 1993] induces a perfect, bilinear pairing

$$\mathrm{H}_1(X_0(pq), \partial(X_0(pq)), \mathbb{Z}) \times \mathrm{H}_1(Y_0(pq), \mathbb{Z}) \to \mathbb{Z}.$$

Since $\circ$ is a nondegenerate bilinear pairing, there is a unique element $\mathcal{E}_{E_N} \in \mathrm{H}_1(X_0(pq), \partial(X_0(pq)), \mathbb{Z})$ such that $\mathcal{E}_{E_N} \circ c = \pi_{E_N}(c)$. The modular symbol $\mathcal{E}_{E_N}$ is the *Eisenstein element* corresponding to the Eisenstein series E_N.

We intersect with the congruence subgroup $\Gamma(2)$ to ensure that the Manin maps become bijective (rather than only surjective), compute the Eisenstein elements for these modular curves, calculate the boundaries and show that these boundaries coincide with the original Eisenstein elements. In the case of $\Gamma_0(p^2)$, although it is difficult to find the Fourier expansion of modular forms at different cusps, fortunately for all $g \in \Gamma_0(p)$ the matrices $g\left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right)g^{-1}$ belong to $\Gamma_0(p^2)$, and hence it is easier to tackle the explicit coset representatives. Unfortunately, for $N = pq$ or $N = p^3$ this is no longer true.

To get around this problem for the congruence subgroup $\Gamma_0(pq)$ with p and q distinct primes, we use the relative homology groups $\mathrm{H}_1(X_0(pq), R \cup I, \mathbb{Z})$. For these relative homology groups, the associated Manin maps are bijective and the push forward of the Eisenstein elements inside the original modular curves turns out to have the same boundary as the original Eisenstein elements. We consider three different homology groups in this paper. In particular, the study of the relative homology group $\mathrm{H}_1(X_0(N), R \cup I, \mathbb{Z})$ to determine the Eisenstein element is a new idea. That these relative homology groups should be useful in the study of modular symbols was discovered by Merel.

Definition 15 (almost Eisenstein elements). For $N \in \{p, q, pq\}$, the differential form $E_N(z)\,dz$ is of the first kind on the Riemann surface $Y_0(pq)$. Since $\circ$ is a nondegenerate bilinear pairing, there is a unique element

$$\mathcal{E}'_{E_N} \in \mathrm{H}_1(X_0(pq) - R \cup I, \partial(X_0(pq)), \mathbb{Z})$$

such that $\mathcal{E}'_{E_N} \circ c = \pi_{E_N}(c)$ for all $c \in \mathrm{H}_1(Y_0(pq), R \cup I, \mathbb{Z})$. We call $\mathcal{E}'_{E_N}$ the *almost Eisenstein element* corresponding to the Eisenstein series E_N.

6. Even Eisenstein elements

6.1. ***Simply connected Riemann surface of genus zero with three marked points.*** Recall that there is only one simply connected (genus zero) compact Riemann surface up to conformal bijections: namely, the Riemann sphere or the projective complex plane $\mathbb{P}^1(\mathbb{C})$. A theorem of Belyi states that every compact, connected, nonsingular algebraic curve X has a model defined over $\overline{\mathbb{Q}}$ if and only if it admits a map to $\mathbb{P}^1(\mathbb{C})$ branched over three points.

Consider the subgroup $\Gamma(2)$ of $\mathrm{SL}_2(\mathbb{Z})$ consisting of all matrices which are the identity modulo the reduction map modulo 2. The Riemann surface $\Gamma(2)$ mod $\overline{\mathbb{H}}$ is a Riemann surface of genus zero, denoted by $X(2)$. Hence, it can be identified with $\mathbb{P}^1(\mathbb{C})$.

The subgroup $\Gamma(2)$ has three cusps $\Gamma(2)0$, $\Gamma(2)1$ and $\Gamma(2)\infty$. Hence, $\Gamma(2)\backslash\overline{\mathbb{H}}$ becomes the simply connected Riemann surface $\mathbb{P}^1(\mathbb{C})$ with the three marked points $\Gamma(2)0$, $\Gamma(2)1$ and $\Gamma(2)\infty$ given by the respective cusps. The modular curve $X_0(pq)$ has no obvious morphism to $X(2)$. So we consider the modular curve X_Γ (Section 6.2). There are two obvious maps π, π' from X_Γ to the compact Riemann surface $X_0(pq)$.

6.2. ***Modular curves with bijective Manin maps.*** For the congruence subgroup $\Gamma = \Gamma_0(pq) \cap \Gamma(2)$, consider the compactified modular curve $X_\Gamma = \Gamma\backslash\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ and let $\pi_\Gamma : \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q}) \to X_\Gamma$ be the canonical surjection.

Let $\pi_0 : \Gamma\backslash\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q}) \to \Gamma(2)\backslash\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ be the map $\pi_0(\Gamma z) = \Gamma(2)z$. The compact Riemann surface $X(2)$ contains three cusps $\Gamma(2)1$, $\Gamma(2)0$ and $\Gamma(2)\infty$. Let $P_- = \pi_0^{-1}(\Gamma(2)1)$ and let P_+ be the union of two sets $\pi_0^{-1}(\Gamma(2)0)$ and $\pi_0^{-1}(\Gamma(2)\infty)$. Consider now the Riemann surface X_Γ with boundary P_+ and P_-.

Let δ_r be 1 or 0 depending on whether r is odd or even. For any integer k, let $s_k = k + (\delta_k - 1)pq$ be an odd integer. Let l and m be two unique integers such that $lq + mp \equiv 1 \pmod{pq}$ with $1 \le l \le p-1$ and $1 \le m \le q-1$. The matrices

$$\alpha'_{pq} = \begin{pmatrix} pq & pq-1 \\ pq+1 & pq \end{pmatrix},$$

$$\alpha'_k = \begin{pmatrix} s_k(pq)^2 & s_k pq - 1 \\ s_k pq + 1 & s_k \end{pmatrix},$$

$$\beta'_r = \begin{pmatrix} -1 & -(r+\delta_r q) \\ p+pq & -1+(r+\delta_r q)(p+pq) \end{pmatrix},$$

$$\gamma'_s = \begin{pmatrix} -1 & -(s+\delta_s pq) \\ q+pq & -1+(s+\delta_s pq)(q+pq) \end{pmatrix}$$

are useful for calculating the boundaries of the Eisenstein elements.

Lemma 16. *The set*

$$\Delta = \{I, \alpha'_k, \beta'_r, \gamma'_s \mid 0 \le k \le pq-1,\ 0 \le r \le q-1,\ 0 \le s \le p-1\} \subset \Gamma(2)$$

forms an explicit set of coset representatives of $\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})$.

Proof. An easy check shows that the orbits $\Gamma_0(pq)\alpha'_k$, $\Gamma_0(pq)\beta'_r$ and $\Gamma_0(pq)\gamma'_s$ are disjoint. Since $|\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})| = pq + p + q + 1$, the result follows. □

The coset representatives in the above lemma are chosen such that $\Gamma_0(pq)\beta_r = \Gamma_0(pq)\beta'_r$ and $\Gamma_0(pq)\gamma_s = \Gamma_0(pq)\gamma'_s$.

Lemma 17. $\Gamma\backslash\Gamma(2)$ *is isomorphic to* $\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})$.

Proof. The explicit coset representatives of Lemma 16 produce the canonical bijection. □

We study the relative homology groups $\mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z})$ and $\mathrm{H}_1(X_\Gamma - P_+, P_-, \mathbb{Z})$. The intersection pairing is a nondegenerate bilinear pairing $\circ : \mathrm{H}_1(X_\Gamma - P_+, P_-, \mathbb{Z}) \times \mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z}) \to \mathbb{Z}$. For $g \in \Gamma\backslash\Gamma(2)$, let $[g]^0$ (respectively $[g]_0$) be the image in X_Γ of the geodesic in $\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ joining $g0$ and $g\infty$ (respectively $g1$ and $g(-1)$). Recall the following fundamental theorems.

Theorem 18 [Merel 1996b]. *Let*

$$\xi_0 : \mathbb{Z}^{\Gamma\backslash\Gamma(2)} \to \mathrm{H}_1(X_\Gamma - P_+, P_-, \mathbb{Z})$$

be the map which takes $g \in \Gamma\backslash\Gamma(2)$ to the element $[g]_0$ and

$$\xi^0 : \mathbb{Z}^{\Gamma\backslash\Gamma(2)} \to \mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z})$$

be the map which takes $g \in \Gamma\backslash\Gamma(2)$ to the element $[g]^0$. The homomorphisms ξ_0 and ξ^0 are isomorphisms.

Theorem 19 [Merel 1996b]. *For $g, g' \in \Gamma(2)$, we have*

$$[g]_0 \circ [g']^0 = \begin{cases} 1 & \textit{if } \Gamma g = \Gamma g', \\ 0 & \textit{otherwise.} \end{cases}$$

The following two lemmas about the set P_- are true for the congruence subgroup $\Gamma_0(N)$ with N odd.

Lemma 20. *We can explicitly write the elements of the set P_- in the form $\Gamma x/y$ with x and y both odd.*

Proof. Suppose that some element of P_- is of the form $\Gamma x/y$ with x and y coprime and y even. Consider the corresponding element in the marked simply connected Riemann surface $X(2)$. The cusp $\Gamma(2)x/y$ is an element such that y is even and p is odd ($\gcd(x, y) = 1$). First, choose p', q' such that $xq' - yp' = 1$ and hence

$$D = \begin{pmatrix} x & p' \\ y & q' \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

Clearly, q' is odd since y is even. If p' is odd then replace the matrix D with DT^{-1} to produce a matrix in $\Gamma(2)$ that takes $i\infty$ to x/y. This contradicts $\Gamma x/y \in P_-$.

If x is even then the projection of $\Gamma x/y$ produces an element of $\Gamma(2)0$. So x is necessarily odd. □

The following lemma is deeply influenced by important results of Manin [1972, Proposition 2.2] and Cremona [1997, Proposition 2.2.3].

Corollary 21. *We can explicitly write the set P_- as $\{\Gamma 1, \Gamma 1/(pq), \Gamma 1/p, \Gamma 1/q\}$.*

Proof. Since $P_- = \pi_0^{-1}(\Gamma(2)1)$, we can write every element of the set P_- as $\Gamma\theta 1$ for some $\theta \in \Delta$ (Lemma 16). Let $\delta \in \{1, p, q, pq\}$. Then every element of P_- can be written as $\Gamma u/(v\delta)$ with $\gcd(u, v\delta) = 1$ and $\gcd(v\delta, pq/\delta) = 1$. Choose an odd integer m and an even integer l such that $lu - mv\delta = 1$. Matrix multiplication shows that

$$\begin{pmatrix} 1 & 0 \\ \delta - 1 & 1 \end{pmatrix}\begin{pmatrix} 1+c & -c \\ c & 1-c \end{pmatrix} = \frac{1}{\delta}$$

and

$$\begin{pmatrix} -m & u+m \\ -l & l+v\delta \end{pmatrix} = \frac{u}{v\delta},$$

and hence

$$A = \begin{pmatrix} 1 & 0 \\ \delta-1 & 1 \end{pmatrix} \begin{pmatrix} 1+c & -c \\ c & 1-c \end{pmatrix} \begin{pmatrix} l+v\delta & -m-u \\ l & -m \end{pmatrix}$$

is a matrix such that $A(u/(v\delta)) = 1/\delta$. The matrix A belongs to Γ if and only if $cv\delta \equiv l' \pmod{pq/\delta}$. Since $v\delta$ is coprime to pq/δ, there is always such a c. Hence, the set P_- consists of the four elements given in the statement of the corollary. $\square$

Let $\pi, \pi' : \Gamma\backslash\overline{\mathbb{H}} \to \Gamma_0(pq)\backslash\overline{\mathbb{H}}$ be the maps $\pi(\Gamma z) = \Gamma_0(pq)z$ and $\pi'(\Gamma z) = \Gamma_0(pq)\frac{1}{2}(z+1)$ respectively. Consider the matrix $h = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right)$. The morphism π' is well defined since the matrix $h\gamma h^{-1}$ belongs to $\Gamma_0(pq)$ for all $\gamma \in \Gamma$. The morphisms π, π' together induce a map

$$\kappa : \mathbb{C}(X_\Gamma) \to \mathbb{C}(X_0(pq))$$

between the function fields of the Riemann surfaces X_Γ given by $\kappa(f(z)) = f(\pi(\Gamma z))^2/f(\pi'(\Gamma z))$. Recall the description of the coordinate chart around a cusp Γx [Miyake 1976] of the Riemann surface X_Γ.

Definition 22. For a cusp y of the congruence subgroup Γ, let Γ_y be the subgroup of Γ fixing y. Let $t \in \mathrm{SL}_2(\mathbb{R})$ be such that $t(y) = i\infty$ and let m be the smallest natural number such that $t\Gamma_y t^{-1}$ is generated by $\left(\begin{smallmatrix} 1 & m \\ 0 & 1 \end{smallmatrix}\right)$. For the modular curve X_Γ, the local coordinate around the point Γy is given by $z \to e^{2\pi i t(z)/m}$.

Example 23. Let $y = 1/\delta$ with δ one of the primes p or q. Then $h(y) = u/\delta$ with $(u, pq) = 1$. Choose integers u', δ' with δ' even such that $u\delta' - u'\delta = 1$; hence $\rho_{h(y)} = \left(\begin{smallmatrix} \delta' & u' \\ -\delta & u \end{smallmatrix}\right)$ is such that $\rho_{h(y)}(h(y)) = i\infty$. We can choose such a $\delta' \in \mathbb{Z}$ since δ is odd.

Matrix multiplication shows that

$$\rho_{h(y)} T^e {\rho_{h(y)}}^{-1} = \begin{pmatrix} 1+e\delta\delta' & e(\delta')^2 \\ -e\delta^2 & 1-e\delta\delta' \end{pmatrix}.$$

Hence, the smallest possible e to ensure $tT^e t^{-1} \subset \Gamma_0(pq)$ is pq/δ.

Example 24. Since $\det(\rho_{h(y)} \circ h) = 2$,

$$t = \begin{pmatrix} \frac{1}{2}l & 0 \\ 0 & 1 \end{pmatrix} \rho_{h(y)} \circ h \in \mathrm{SL}_2(\mathbb{R})$$

and $t(y) = i\infty$. A calculation shows that

$$tT^e t^{-1} = \begin{pmatrix} 1+\frac{1}{2}e\delta\delta' & \frac{1}{4}e\delta'^2 \\ -e\delta^2 & 1-\frac{1}{2}e\delta\delta' \end{pmatrix}.$$

Hence, the smallest possible e to ensure $tT^e t^{-1} \subset \Gamma$ is $e = 2pq/\delta$.

We use the following lemma to construct differential forms of the first kind on the ambient Riemann surface $X_\Gamma - P_+$.

Lemma 25. *Let $f : X_0(pq) \to \mathbb{C}$ be a rational function. The divisors of $\kappa(f)$ are supported on P_+.*

Proof. Suppose f is a meromorphic function on the Riemann surface $X_0(pq)$. Then f is given by g/h with g and h holomorphic functions on $X_0(pq)$. Every element of P_- is of the form $\Gamma 1/\delta$ with $\delta \mid N$. By [Miranda 1995, Proposition 4.1], every holomorphic map on a Riemann surface locally looks like $z \to z^n$.

Consider the morphism π' and the point on the modular curve $\Gamma 1/\delta$. The local coordinates around the points $\Gamma_0(pq)0$, $\Gamma_0(pq)\infty$ and $\Gamma_0(pq)1/p$ are given by $q_0(z) = e^{2\pi i/(-pqz)}$, $q_\infty(z) = e^{2\pi i z}$ and $q_{1/q}(z) = e^{2\pi i z/(p(-qz+1))}$ respectively. In the modular curve X_Γ, the local coordinates around the points of P_- are given by

$$\begin{aligned} q_1(z) &= e^{2\pi i/(2pq(-z+1))}, \\ q_{1/(pq)}(z) &= e^{2\pi i z/(2(-pqz+1))}, \\ q_{1/p}(z) &= e^{2\pi i z/(2q(-pz+1))}, \\ q_{1/q}(z) &= e^{2\pi i z/(2p(-qz+1))}. \end{aligned}$$

Now around the points $\Gamma 1$ and $\Gamma 1/(pq)$ we have the equalities $q_0 \circ \pi = q_1^2$, $q_0 \circ \pi' = q_1^4$ and $q_{1/(pq)} \circ \pi = q_{1/(pq)}^2$, $q_{1/(pq)} \circ \pi' = q_{1/(pq)}^4$.

Let $y = 1/\delta$ with δ one of the primes p or q. The local coordinate chart around the point $\Gamma 1/\delta$ is $z \to e^{2\pi i \rho_{h(x)} \circ h(z)/(4e)}$. The map π' takes it to $e^{2\pi i 2\rho_{h(x)}(h(z))/e}$. For this coordinate chart the map π' is given by $z \to z^4$.

We now consider the map π and a matrix $t = \left(\begin{smallmatrix} 1 & 0 \\ -\delta & 1 \end{smallmatrix}\right)$ such that $t(y) = i\infty$ and $e = pq/\delta$. The local coordinate around the point $\Gamma 1/\delta$ is $z \to e^{2\pi i t(z)/(2e)}$ and the map π takes it to $e^{2\pi i t(z)/e}$. In this coordinate chart, the map π is given by $z \to z^2$. Hence, the function $(f \circ \pi)^2/(f \circ \pi')$ has no zero or pole on P_-. $\square$

Definition 26 (even Eisenstein elements). For $E_N \in \mathbb{E}_{pq}$, let $\lambda_{E_N} : X_0(pq) \to \mathbb{C}$ be the rational function whose logarithmic differential is $2\pi i E_N(z)\, dz = 2\pi i \omega_{E_N}$. Consider the rational function $\lambda_{E_N,2} = (\lambda_{E_N} \circ \pi)^2/(\lambda_{E_N} \circ \pi')$ on X_Γ. By Lemma 25, this function has no zeros and poles in P_-. Let $\kappa^*(\omega_{E_N})$ be the logarithmic differential of the function. Let $\varphi_{E_N}(c) = \int_c \kappa^*(\omega_{E_N})$ be the corresponding period homomorphism $\mathrm{H}_1(X_\Gamma - P_+, P_-, \mathbb{Z}) \to \mathbb{Z}$.

By the nondegeneracy of the intersection pairing, there is a unique element $\mathcal{E}^0_{E_N} \in \mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z})$ such that $\mathcal{E}^0_{E_N} \circ c = \varphi_{E_N}(c)$ for all $c \in \mathrm{H}_1(X_\Gamma - P_+, P_-, \mathbb{Z})$. The modular symbol $\mathcal{E}^0_{E_N}$ is the *even Eisenstein element* corresponding to the Eisenstein series E_N.

For $E_N \in \mathbb{E}_{pq}$, define a function $F_{E_N} : \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z}) \to \mathbb{Z}$ by

$$F_{E_N}(g) = \varphi_{E_N}(\xi_0(g)) = \int_{g(1)}^{g(-1)} \big(2E_N(z) - E_N\big(\tfrac{1}{2}(z+1)\big)\big)\,dz.$$

Remark 27. It is easy to see that for any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma(2)$,

$$h\gamma h^{-1} = \begin{pmatrix} a+c & \frac{1}{2}(b+d-a-c) \\ 2c & d-c \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

For any matrix $\gamma \in \Gamma$, consider the rational numbers

$$\begin{aligned} P_N(\gamma) &= \tfrac{1}{12}(2\pi_{E_N}(\gamma) - \pi_{E_N}(h\gamma h^{-1})), \\ t(\gamma) &= b+d-a-c, \\ s(\gamma) &= a+c. \end{aligned}$$

Lemma 28. *For $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ with $c \neq 0$,*

$$\begin{aligned} P_N(\gamma) = \operatorname{sgn}(t(\gamma))\Big(2\Big(S\big(s(\gamma), |t(\gamma)|pq\big) - S\big(s(\gamma), |t(\gamma)|\big)\Big) \\ - S\big(s(\gamma), \big|\tfrac{1}{2}t(\gamma)\big|pq\big) + S\big(s(\gamma), \tfrac{1}{2}|t(\gamma)|\big)\Big). \end{aligned}$$

In particular, $P_N(\gamma) \in \mathbb{Z}$ for all $\gamma \in \Gamma$.

Proof. Recall the properties of period homomorphism (see Proposition 13). We calculate the corresponding periods:

$$\begin{aligned} \pi_{E_N}(\gamma) &= \pi_E(T\gamma T^{-1}) \\ &= \pi_{E_N}\left(\begin{pmatrix} a+c & -(a+c)+b+d \\ c & -c+d \end{pmatrix}\right) \\ &= \pi_{E_N}\left(\begin{pmatrix} a+c & -(a+c)+b+d \\ c & -c+d \end{pmatrix}\right) \\ &= \pi_{E_N}\left(\begin{pmatrix} d-c & c/N \\ t(\gamma)N & a+c \end{pmatrix}\right). \end{aligned}$$

By Proposition 13, we have

$$\pi_{E_N}(\gamma) = \frac{a+d}{t(\gamma)N}(N-1) + 12\operatorname{sgn}(t(\gamma))\Big(S\big(s(\gamma), |t(\gamma)|N\big) - S\big(s(\gamma), |t(\gamma)|\big)\Big).$$

Similarly,

$$\begin{aligned}
&\pi_{E_N}(h\gamma h^{-1})\\
&\quad=\pi_{E_N}\left(\begin{pmatrix} a+c & \frac{1}{2}(b+d-a-c)\\ 2c & d-c\end{pmatrix}\right)\\
&\quad=\pi_{E_N}\left(\begin{pmatrix} d-c & 2c/N\\ \frac{1}{2}t(\gamma)N^2 & a+c\end{pmatrix}\right)\\
&\quad=\frac{2(a+d)}{t(\gamma)N}(N-1)+12\,\mathrm{sgn}(t(\gamma))\Big(S\big(s(\gamma),\tfrac{1}{2}|t(\gamma)|N\big)-S\big(s(\gamma),\tfrac{1}{2}|t(\gamma)|\big)\Big).
\end{aligned}$$

Hence, we deduce the formula given in the lemma statement. From the formula, we see that $P_N(\gamma)\in\mathbb{Z}$ for all $\gamma\in\Gamma$. □

Let x be one of the primes p or q. Choose integers s, s' and l, l' such that $l(s_k x+2)-2spq=1$ and $l's_k x-2s'pq/x=1$. Let

$$\gamma_1^{x,k}=\begin{pmatrix} 1+4spq & -2l\\ -4s(s_k x+2)pq & 1+4spq\end{pmatrix}\quad\text{and}\quad \gamma_2^{x,k}=\begin{pmatrix} 1+4s'pq/x & -2l'\\ -4s'(s_k)pq & 1+4s'pq/x\end{pmatrix}$$

be two matrices in Γ. Since the integers l and l' are necessarily odd, we have $\gamma_1^{x,k}(1/(s_k x+2))=-1/(s_k x+2)$ and $\gamma_2^{x,k}(1/(s_k x))=-1/(s_k x)$.

Using the formula of Lemma 28, we deduce that

$$s(\gamma_1^{x,k})=1-4spq(1+s_k x),\qquad t(\gamma_1^{x,k})=-2(l-2s(s_k x+2)pq)$$

and

$$s(\gamma_2^{x,k})=1-4s'pq\left(s_k-\frac{1}{x}\right),\qquad t(\gamma_2^{x,k})=-2(l'-2s's_k pq).$$

We can now calculate $P_N(\gamma_1^{x,k})$ and $P_N(\gamma_2^{x,k})$ using Lemma 28.

Proposition 29.

$$F_{E_N}(g)=\begin{cases} 12(S(r,N)-2S(r,2N)) & \text{if } g=(r-1,r+1),\\ 6(P_N(\gamma_1^{x,k})-P_N(\gamma_2^{x,k})) & \text{if } g=(1+kx,1) \text{ or } g=(-1-kx,1),\\ -6(P_N(\gamma_1^{x,k})-P_N(\gamma_2^{x,k})) & \text{if } g=(1,-1-kx) \text{ or } g=(1,1+kx),\\ 0 & \text{if } g=(\pm1,1).\end{cases}$$

Proof. If $g=(r-1,r+1)$ and $E_N\in\mathbb{E}_{pq}$, we get [Merel 1996b, p. 18]

$$F_{E_N}(g)=\varphi_{E_N}(\xi_0(g))=12(S(r,N)-2S(r,2N)).$$

We now find the value of the integrals in the remaining cases. The differential form $k^*(\omega_{E_N})$ is of the first kind on the Riemann surface $X_\Gamma-P_+$. We also note that if $g=(\pm1,1)$, $(\pm1\pm kx,1)$ or $(1,\pm1\pm kx)$ with x one of the primes p or q, then we can't write it as $(r-1,r+1)$.

Since all the Fourier coefficients of the Eisenstein series are real-valued, an argument similar to one in [Merel 1996b, p. 19] shows that $F_{E_N}(s_k x+1, 1) = F_{E_N}(-s_k x-1, 1)$. Consider the path

$$\left\{\frac{1}{s_k x+2}, \frac{-1}{s_k x+2}\right\} = \left\{\frac{1}{s_k x+2}, \frac{1}{s_k x}\right\} + \left\{\frac{1}{s_k x}, \frac{-1}{s_k x}\right\} + \left\{\frac{-1}{s_k x}, \frac{-1}{s_k x+2}\right\}.$$

The rational number $1/(s_k x)$ corresponds to a point of P_- in the Riemann surface X_Γ. The differential form $k^*\omega_{E_N}$ has no zeros and poles on P_-. We deduce that

$$\begin{aligned}\int_{1/(s_k x+2)}^{-1/(s_k x+2)} & k^*(\omega_{E_N}) \\ &= \int_{1/(s_k x+2)}^{1/(s_k x)} k^*(\omega_{E_N}) + \int_{1/(s_k x)}^{-1/(s_k x)} k^*(\omega_{E_N}) + \int_{-1/(s_k x)}^{-1/(s_k x+2)} k^*(\omega_{E_N}) \\ &= 2F_N(s_k x+1, 1) + \int_{1/(s_k x)}^{-1/(s_k x)} k^*(\omega_{E_N}).\end{aligned}$$

Let $\gamma_1^{x,k}$ and $\gamma_2^{x,k}$ be two matrices in Γ such that $\gamma_1^{x,k}(1/(s_k x+2)) = -1/(s_k x+2)$ and $\gamma_2^{x,k}(1/(s_k x)) = -1/(s_k x)$. Then

$$2F_N(s_k x+1, 1) = \int_{1/(s_k x+2)}^{\gamma_1^{x,k}(1/(s_k x+2))} k^*(\omega_{E_N}) - \int_{1/(s_k x)}^{\gamma_2^{x,k}(1/(s_k x))} k^*(\omega_{E_N}).$$

We now prove that $\int_{1/(s_k x)}^{\gamma_2^{x,k}(1/(s_k x))} k^*(\omega_{E_N})$ is independent of the choice of the matrices $\gamma_2^{x,k} \in \Gamma$ that take $1/(s_k x)$ to $-1/(s_k x)$. Suppose $\gamma_2^{x,k}$ and $\gamma'^{x,k}_2$ are two matrices such that $\gamma_2^{x,k}(1/(s_k x)) = \gamma'^{x,k}_2(1/(s_k x)) = -1/(s_k x)$. Since $\gamma_2^{x,k} \in \Gamma$,

$$\varphi_{E_N}(\gamma_2^{x,k}) = \int_{1/(s_k x)}^{\gamma_2^{x,k}(1/(s_k x))} k^*(\omega_{E_N})$$

is independent of the choice of any point in $\mathbb{H} \cup \{-1\}$. By replacing $1/(s_k x)$ with $(\gamma_2^{x,k})^{-1}(\gamma'^{x,k}_2)(1/(s_k x))$, we get that the above integral is the same as

$$\int_{1/(s_k x)}^{\gamma'^{x,k}_2(1/(s_k x))} k^*(\omega_{E_N})$$

and the integral is independent of the choice of exceptional matrices. Similarly, we can prove that

$$\int_{1/(s_k x+2)}^{\gamma^{x,k}(1/(s_k x+2))} k^*(\omega_{E_N})$$

is also independent of the choice of the matrices that take $1/(s_k x+2)$ to $-1/(s_k x+2)$. Since we have already written down two matrices $\gamma_1^{x,k}$ and $\gamma_2^{x,k}$ in Γ such that

$\gamma_1^{x,k}(1/(s_kx+2)) = -1/(s_kx+2)$ and $\gamma_2^{x,k}(1/(s_kx)) = -1/(s_kx)$, we use these matrices to find those integrals.

The above calculation shows that

$$2\pi_{E_N}(\gamma_1^{x,k}) - \pi_{E_N}(h\gamma_1^{x,k}h^{-1}) = 2F_N(s_kx+1, 1) + 2\pi_{E_N}(\gamma_2^{x,k}) - \pi_{E_N}(h\gamma_2^{x,k}h^{-1}).$$

We get

$$\begin{aligned} &F_{E_N}(s_kx+1, 1) \\ &\qquad = \tfrac{1}{2}(2\pi_{E_N}(\gamma_1^{x,k}) - \pi_{E_N}(h\gamma_1^{x,k}h^{-1}) - 2\pi_E(\gamma_2^{x,k}) + \pi_E(h\gamma_2^{x,k}h^{-1})) \\ &\qquad = 6(P_N(\gamma^{x,k}) - P_N(\gamma_2^{x,k})). \end{aligned}$$

Since $F_{E_N}(1+s_kx, 1) = -F_{E_N}(1, -1-s_kx)$, the above equation determines the Eisenstein elements for the Eisenstein series E_N completely. □

From the above lemma, we conclude that $6F_N(g) = F_{E_N}(g)$.

Lemma 30. *For $E_N \in E_2(\Gamma_0(pq))$, consider the element $\mathcal{E}^0_{E_N} \in \mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z})$ defined by $\mathcal{E}^0_{E_N} = \sum_{g\in\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})} F_{E_N}(g)\xi^0(g)$. For all $c \in \mathrm{H}_1(X_\Gamma - P_+, P_-, \mathbb{Z})$, we have $\mathcal{E}^0_{E_N} \circ c = \varphi_{E_N}(c)$.*

Proof. By Theorem 19, we can write the even Eisenstein element uniquely as

$$\sum_{g\in\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})} H_{E_N}(g)\xi^0(g).$$

By the same theorem, $[g]_0 \circ [h]^0 = 1$ if and only if $\Gamma g = \Gamma h$. The functions H_{E_N} and F_{E_N} coincide since

$$H_{E_N}(g) = \sum_{g\in\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})} H_{E_N}(g)\xi^0(g) \circ \xi_0(g) = \mathcal{E}^0_{E_N} \circ \xi_0(g) = F_{E_N}(g). \qquad \square$$

For the modular curve X_Γ, we have a similar short exact sequence

$$0 \to \mathrm{H}_1(X_\Gamma - P_-, \mathbb{Z}) \to \mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z}) \xrightarrow{\delta^0} \mathbb{Z}^{P_+} \to \mathbb{Z} \to 0.$$

The boundary map δ^0 takes a geodesic, joining the points r and s of P_+ to the formal symbol $[r] - [s]$.

7. Eisenstein elements and winding elements for $\Gamma_0(pq)$

7.1. ***Eisenstein elements for $\Gamma_0(pq)$.*** We first prove an elementary number theoretic lemma. Recall, l and m are two unique integers such that $lq+mp \equiv 1 \pmod{pq}$ with $1 \le l \le p-1$ and $1 \le m \le q-1$.

Lemma 31. *For all k with $1 \le k \le q-1$, we can choose an integer $s(k) \in \mathbb{Z}/q\mathbb{Z}$ such that*

$$(kp, -1) = (p, s(k)p - 1)$$

in $\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})$. *The map* $k \to s(k)$ *is a bijection* $(\mathbb{Z}/q\mathbb{Z})^* \to \mathbb{Z}/q\mathbb{Z} - \{\overline{m}\}$.

Proof. For all k with $1 \le k \le q-1$, let k' be the inverse of k in $(\mathbb{Z}/q\mathbb{Z})^*$. By the Chinese remainder theorem, we choose a unique x with $1 \le x \le pq-1$ such that $x \equiv -1 \pmod p$ and $x \equiv -k' \pmod q$. Observe that x is coprime to both p and q. We write $x = s(k)p - 1$ for a unique $s(k)$ with $0 \le s(k) \le q-1$. Since $\Gamma_0(pq)\backslash \mathrm{SL}_2(\mathbb{Z}) \cong \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})$, we deduce that $(kp, -1) = (xkp, -x) = (-p, -x) = (p, x) = (p, s(k)p - 1)$ in $\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})$.

Consider the map $(\mathbb{Z}/q\mathbb{Z})^* \to \mathbb{Z}/q\mathbb{Z}$ given by $k \to s(k)$. If $lq+mp \equiv 1 \pmod{pq}$ then m is not in the image. This map is one-to-one since $s(k) = s(h)$ implies $k \equiv h \pmod q$. Thus the map $(\mathbb{Z}/q\mathbb{Z})^* \to \mathbb{Z}/q\mathbb{Z} - \{\overline{m}\} k \to s(k)$ is a bijection. □

For all t coprime to pq, consider the set V of all matrices of the form α_t.

Proposition 32. *The boundary of any element*

$$X = \sum_{g \in \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})} F(g)[g]^*$$

in $\mathrm{H}_1(X_0(pq) - R \cup I, \partial(X_0(pq)), \mathbb{Z})$ *is of the form*

$$\delta(X) = A(X)\left[\frac{1}{p}\right] + B(X)\left[\frac{1}{q}\right] + C(X)[\infty] - (A(X) + B(X) + C(X))[0]$$

with

$$\begin{aligned} A(X) &= \sum_{k=0}^{q-1} (F(\beta_k) - F(\beta_k S)), \\ B(X) &= \sum_{i=0}^{p-1} (F(\gamma_i) - F(\gamma_i S)), \\ C(X) &= F(0,1) - F(1,0). \end{aligned}$$

Proof. Choose an explicit coset representative of $\Gamma_0(pq)\backslash \mathrm{SL}_2(\mathbb{Z})$ (see Lemma 8) and write

$$\begin{aligned} X = C(X)[I]^* &+ \sum_{\alpha_t \in V} F(1,t)[\alpha_t]^* + \sum_{k=1}^{q-1} F(1,kp)[\alpha_{kp}]^* \\ &+ \sum_{k=1}^{p-1} F(1,kq)[\alpha_{kq}]^* + \sum_{i=0}^{q-1} F(p, ip-1)[\beta_i]^* + \sum_{j=0}^{p-1} F(q, jq-1)[\beta_j]^*. \end{aligned}$$

According to Lemma 31 for $1 \le k \le q-1$, we have $\alpha_{kp} S = Z\beta_{s(k)}$ for some $Z \in \Gamma_0(pq)$. We deduce that

$$\sum_{k=1}^{q-1} F(1,kp)[\alpha_{kp}]^* + \sum_{i=0}^{q-1} F(p,ip-1)[\beta_i]^*$$
$$= \sum_{k=1}^{q-1} \bigl(F(1,kp)[\alpha_{kp}]^* + F(kp,-1)[\alpha_{kp}S]^*\bigr) + F(\beta_m)[\beta_m]^*$$

and

$$\sum_{k=1}^{p-1} F(1,kq)[\alpha_{kq}]^* + \sum_{j=0}^{p-1} F(q,jq-1)[\gamma_j]^*$$
$$= \sum_{k=1}^{p-1} \bigl(F(1,kq)[\alpha_{kq}]^* + F(kq,-1)[\alpha_{kq}S]^*\bigr) + F(\gamma_l)[\gamma_l]^*.$$

A small check shows that $\delta([\alpha_{kp}]^*) = \delta([\alpha_p]^*)$ and $\delta([\alpha_{kp}]^*) = -\delta([\alpha_{kp}S]^*)$.

We now calculate $\delta([\beta_m]^*)$ and $\delta([\gamma_l]^*)$. Since $lq + mp \equiv 1 \pmod{pq}$ and $-I \in \Gamma_0(pq)$, we get

$$\begin{pmatrix} 1-q(l-1) & m(l-1) \\ (l-1)pq & 1+lq(l-1) \end{pmatrix} \begin{pmatrix} m & -l \\ q & p \end{pmatrix} = \gamma \beta_m S \tag{7-1}$$

and

$$\begin{pmatrix} 1-p(m+1) & -l(m+1) \\ (1+m)pq & 1-mp(l+m) \end{pmatrix} \begin{pmatrix} m & -l \\ q & p \end{pmatrix} = \begin{pmatrix} -1 & -l \\ q & -mp \end{pmatrix} = \gamma_l$$

for some $\gamma \in \Gamma_0(pq)$, and hence we have $\Gamma_0(pq)\beta_m S = \gamma_l$. From $\delta([\beta_m]^*) = \delta([\alpha_q]^* - [\alpha_p]^*)$ and $\delta([\gamma_l]^*) = \delta([\alpha_p]^* - [\alpha_q]^*)$, it is easy to see that

$$\delta\Bigl(\sum_{k=1}^{q-1} F(1,kp)[\alpha_{kp}]^* + \sum_{i=0}^{q-1} F(p,jp-1)[\beta_j]^*\Bigr)$$
$$= \sum_{k=1}^{q-1} \bigl(F(1,kp) - F(kp,-1)\bigr)\delta([\alpha_p]^*) + F(\beta_m)\delta([\beta_m]^*)$$

and

$$\delta\Bigl(\sum_{k=1}^{p-1} F(1,kq)[\alpha_{kq}]^* + \sum_{j=0}^{p-1} F(q,jq-1)[\gamma_j]^*\Bigr)$$
$$= \sum_{k=1}^{p-1} \bigl(F(1,kq) - F(kq,-1)\bigr)\delta([\alpha_q]^*) + F(q,lq-1)\delta([\gamma_l]^*).$$

We have

$$
\begin{aligned}
&F(p, mp-1)\delta([\beta_m]^*) + F(q, lq-1)\delta([\gamma_l]^*)\\
&\qquad\qquad\qquad\qquad = \big(F(\beta_m) - F(\beta_m S)\big)\big(\delta([\alpha_q]^*) - \delta([\alpha_p]^*)\big).
\end{aligned}
$$

Recall, $\delta([\alpha_p]^*) = [0] - [1/p]$ and $\delta([\alpha_q]^*) = [0] - [1/q]$. The above calculation shows that

$$\delta(X) = C(X)\delta([I]^*) + A(X)\delta([\alpha_p]^*) + B(X)\delta([\alpha_q]^*)$$

with

$$
\begin{aligned}
A(X) &= \sum_{k=0}^{q-1}(F(p, kp-1) - F(kp-1, -p)),\\
B(X) &= \sum_{m=0}^{p-1}(F(\gamma_l') - F(\gamma_l' S)),\\
C(X) &= F(I) - F(S).
\end{aligned}
$$

□

We also prove a similar proposition for $\Gamma \subset \Gamma(2)$.

Proposition 33. *The boundary of any element*

$$X = \sum_{g \in \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})} F(g)\xi^0(g)$$

in $\mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z})$ *is of the form*

$$\delta^0(X) = A'(X)\left[\frac{1}{p}\right] + B'(X)\left[\frac{1}{q}\right] + C'(X)[\infty] - (A'(X) + B'(X) + C'(X))[0]$$

with

$$
\begin{aligned}
A'(X) &= \sum_{k=0}^{q-1} F(\beta_k') - \left(\sum_{k=1}^{q-1} F(\alpha_{kp}')\right) - F(\gamma_l'),\\
B'(X) &= \sum_{i=0}^{p-1} F(\gamma_i') - \left(\sum_{k=1}^{p-1} F(\alpha_{kq}')\right) - F(\beta_m'),\\
C'(X) &= F(0, 1) - F(\alpha_{pq}').
\end{aligned}
$$

Proof. This is a straightforward calculation using the coset representatives of $\Gamma\backslash\Gamma(2)$ (see Lemma 16). □

Proposition 34. *For* $E \in \mathbb{E}_{pq}$, *the boundaries of almost Eisenstein elements* $\mathcal{E}_E'$ *in* $\mathrm{H}_1(X_0(pq) - R \cup I, \partial(X_0(pq)), \mathbb{Z})$ *corresponding to the Eisenstein series* E *are* $-\delta(E)$ (Section 4).

Proof. For $E \in \mathbb{E}_{pq}$, let $\mathcal{E}'_E = \sum_{g \in \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})} G_E(g)[g]^*$ be the almost Eisenstein element. According to Proposition 32, we need to calculate $A(\mathcal{E}'_E)$, $B(\mathcal{E}'_E)$ and $C(\mathcal{E}'_E)$.

For all $0 \le k < q-1$, $\beta_k T = \beta_{k+1}$ and $\beta_{q-1} T = \gamma \beta_0$ with

$$\gamma = \begin{pmatrix} 1+pq & q \\ -qp^2 & 1-qp \end{pmatrix}.$$

We have an inclusion $\mathrm{H}_1(Y_0(pq), \mathbb{Z}) \to \mathrm{H}_1(Y_0(pq), R \cup I, \mathbb{Z})$. Since $\{\rho^*, \gamma\rho^*\} = \{\beta_0\rho^*, \gamma\beta_0\rho^*\} = -\sum_{k=0}^{q-1}\{\beta_k\rho, \beta_k\rho^*\}$, we deduce that

$$\begin{aligned}\pi_E(\gamma) &= \int_{z_0}^{\gamma z_0} E(z)\,dz \\ &= \mathcal{E}'_E \circ \{z_0, \gamma z_0\} \\ &= -\mathcal{E}'_E \circ \left(\sum_{k=0}^{q-1}\{\beta_k\rho, \beta_k\rho^*\}\right) \\ &= -\sum_{k=0}^{q-1} \mathcal{E}'_E \circ \{\beta_k\rho, \beta_k\rho^*\}.\end{aligned}$$

Applying Corollary 6, we have

$$\sum_{k=0}^{q-1} \mathcal{E}'_E \circ \{\beta_k\rho, \beta_k\rho^*\} = \sum_{k=0}^{q-1}(G_E(\beta_k) - G_E(\beta_k S)) = -A(\mathcal{E}'_E).$$

Hence, we prove that $A(\mathcal{E}'_E) = -\pi_E(\gamma)$. By interchanging p and q, we have $B(\mathcal{E}'_E) = -\pi_E(\gamma_0)$ for

$$\gamma_0 = \begin{pmatrix} 1+pq & p \\ -pq^2 & 1-qp \end{pmatrix}.$$

We now calculate $\pi_E(\gamma)$ and $\pi_E(\gamma_0)$ using [Stevens 1985]. Recall, $1/p$ is a cusp with $e_{\Gamma_0(pq)}(1/p) = q$. Consider the matrices

$$x = \begin{pmatrix} 1 & -q \\ -p & 1+qp \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} 1 & -p \\ -q & 1+qp \end{pmatrix}.$$

One can easily check that $x\left(\begin{smallmatrix} 1 & q \\ 0 & 1 \end{smallmatrix}\right)x^{-1} = \gamma$ and $y\left(\begin{smallmatrix} 1 & p \\ 0 & 1 \end{smallmatrix}\right)y^{-1} = \gamma_0$. Notice that $x(i\infty) = \Gamma_0(pq)1/p$ and $y(i\infty) = \Gamma_0(pq)1/q$. By [Stevens 1985, p. 524], we deduce that $\pi_E(\gamma) = e_{\Gamma_0(pq)}(1/q)a_0(E[1/p])$ and $\pi_{E_{pq}}(\gamma_0) = e_{\Gamma_0(pq)}(1/p)a_0(E[1/p])$.

According to Proposition 32, the boundary of the almost Eisenstein element corresponding to an Eisenstein series E is

$$\delta(\mathcal{E}'_E) = A(\mathcal{E}'_E)\left[\frac{1}{p}\right] + B(\mathcal{E}'_E)\left[\frac{1}{q}\right] + C(\mathcal{E}'_E)[\infty] - (A(\mathcal{E}'_E) + B(\mathcal{E}'_E) + C(\mathcal{E}'_E))[0]$$

with $A(\mathcal{E}'_E) = qa_0(E[1/p])$, $B(\mathcal{E}'_E) = pa_0(E[1/q])$ and $C(\mathcal{E}'_E) = -(F(I) - F(S))$. Applying Corollary 6 again, we deduce that $F(I) - F(S) = \int_{\rho}^{\rho^*} E(z)\,dz = -a_0(E)$. For $E \in E_2(\Gamma_0(pq))$, the boundary of E is

$$\begin{aligned}\delta(E) &= a_0(E)([\infty] - [0]) \\ &\qquad + qa_0\left(E\left[\frac{1}{p}\right]\right)\left(\left[\frac{1}{p}\right] - [0]\right) + pa_0\left(E\left[\frac{1}{q}\right]\right)\left(\left[\frac{1}{q}\right] - [0]\right) \\ &= \delta(\mathcal{E}'_E).\end{aligned}$$
□

Let β and h be the matrices $\left(\begin{smallmatrix}1 & 2\\ 0 & 1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1 & 1\\ 0 & 2\end{smallmatrix}\right)$ respectively. Let

$$\pi_* : \mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z}) \to \mathrm{H}_1(X_0(pq) - R \cup I, \partial(X_0(pq)), \mathbb{Z})$$

be the isomorphism defined by $\pi_*(\xi_0(g)) = [g]^*$ [Merel 1995, Corollary 1]. It is easy to see that $\delta(\pi_*(X)) = \delta^0(X)$ for all $X \in \mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z})$.

Proposition 35. *For all $E \in \mathbb{E}_{pq}$, let $\mathcal{E}^0_E$ denote the even Eisenstein element in* $\mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z})$ (Section 6). *The boundary of the modular symbol $\pi_*(\mathcal{E}^0_E)$ is $-6\delta(E)$.*

Proof. By Theorem 18, we can explicitly write down the even Eisenstein element $\mathcal{E}^0_E$ in the relative homology group $\mathrm{H}_1(X_\Gamma - P_-, P_+, \mathbb{Z})$ as

$$\mathcal{E}^0_E = \sum_{g \in \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})} F_E(g)\xi_0(g).$$

According to Proposition 33, we need to calculate $A'(\mathcal{E}^0_E)$, $B'(\mathcal{E}^0_E)$ and $C'(\mathcal{E}^0_E)$. For $0 \le k < q - 2$, we have $\beta'_k\beta = \beta'_{k+2}$. A small check shows that $\beta'_{q-1}\beta = \beta'_1$ and $\beta'_{q-2}\beta = \gamma'\beta'_0$ with

$$\gamma' = \begin{pmatrix} 1 + 2pq(1+q) & 2q \\ -2q(p+pq)^2 & 1 - 2pq(1+q) \end{pmatrix} \in \Gamma.$$

As a homology class in $\mathrm{H}_1(X_\Gamma - P_+, P_-, \mathbb{Z})$, we have

$$\begin{aligned}\{-1, \gamma'(-1)\} &= \{\beta'_0(-1), \gamma'\beta'_0(-1)\} \\ &= -\sum_{k=0}^{q-1}\{\beta'_k(1), \beta'_k(-1)\} \\ &= \sum_{k=0}^{q-1}\{\beta'_k(-1), \beta'_k(1)\}.\end{aligned}$$

By the definition of the even Eisenstein elements, we conclude that

$$\begin{aligned}\int_{z_0}^{\gamma' z_0} k^*(\omega_E) &= \mathcal{E}_E^0 \circ \{z_0, \gamma' z_0\}\\ &= -\mathcal{E}_E^0 \circ \sum_{k=0}^{q-1} (\beta'_k(1), \beta'_k(-1))\\ &= -\sum_{k=0}^{q-1} \mathcal{E}_E^0 \circ \{\beta'_k(1), \beta'_k(-1)\}.\end{aligned}$$

It is easy to see that $hASBh^{-1} \in \mathrm{SL}_2(\mathbb{Z})$ for all $A, B \in \Gamma(2)$. Since $[\alpha'_{kq} S] = [\gamma'_{s(k)}]$ in $\mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})$, we have $\kappa' = \alpha'_{kq} S(\gamma'_{s(k)})^{-1} \in \Gamma_0(pq)$ and $h\kappa' h^{-1} \in \Gamma_0(pq)$. We deduce that the differential form

$$k^*(\omega_E) = f(z)\, dz = \bigl(2E(z) - \tfrac{1}{2} E\bigl(\tfrac{1}{2}(z+1)\bigr)\bigr)\, dz$$

is invariant under κ'. According to the above argument,

$$\begin{aligned}F_E(\alpha'_{kq}) &= \int_{\alpha'_{kq}(1)}^{\alpha'_{kq}(-1)} f(z)\, dz\\ &= \int_{\alpha'_{kq} S(-1)}^{\alpha'_{kq} S(1)} f(z)\, dz\\ &= -\int_{\alpha'_{kq} S(1)}^{\alpha'_{kq} S(-1)} f(z)\, dz\\ &= -\int_{\kappa'^{-1}\alpha'_{kq} S(1)}^{\kappa'^{-1}\alpha'_{kq} S(-1)} f(\kappa' z)\, d\kappa' z\\ &= -\int_{\gamma'_{s(k)}(1)}^{\gamma'_{s(k)}(-1)} f(z)\, dz\\ &= -F_E(\gamma'_{s(k)}).\end{aligned} \tag{7-2}$$

A similar calculation shows that $F_E(\gamma'_l) = -F_E(\beta'_m)$ and $F_E(\alpha_{kp}) = -F_E(\beta_{s(k)})$ for some $s(k) \in (\mathbb{Z}/q\mathbb{Z})^*$. Applying Theorem 18, we have

$$\sum_{k=0}^{q-1} F_E(\beta'_k) = \sum_{k=0}^{q-1} \mathcal{E}_E^0 \circ \{\beta'_k(1), \beta'_k(-1)\} = -\int_{z_0}^{\gamma' z_0} k^*(\omega_E).$$

According to the definition of the period π_E of the Eisenstein series $E(z)$ (see Section 4), we get

$$\int_{z_0}^{\gamma' z_0} k^*(\omega_E) = \int_{z_0}^{\gamma' z_0} \bigl(2E(z) - \tfrac{1}{2} E\bigl(\tfrac{1}{2}(z+1)\bigr)\bigr)\, dz = 2\pi_E(\gamma') - \pi_E(h\gamma' h^{-1}).$$

We calculate $\pi_E(\gamma')$ and $\pi_E(h\gamma' h^{-1})$. From Remark 27, we see that

$$h\gamma' h^{-1} = \begin{pmatrix} 1+z & qv^2 \\ -4p^2q(1+q)^2 & 1-z \end{pmatrix}$$

with $v = 1 - p(1+q)$ and $z = 2pqv(1+q)$. Furthermore, the matrix $h\gamma' h^{-1}$ decomposes as

$$h\gamma' h^{-1} = \begin{pmatrix} 1-p(1+q) & \frac{1}{2}p(1+q) \\ -2p(1+q) & 1+p(1+q) \end{pmatrix} \begin{pmatrix} 1 & q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1-p(1+q) & \frac{1}{2}p(1+q) \\ -2p(1+q) & 1+p(1+q) \end{pmatrix}^{-1}.$$

Since the matrix

$$\begin{pmatrix} 1-p(1+q) & \frac{1}{2}p(1+q) \\ -2p(1+q) & 1+p(1+q) \end{pmatrix}^{-1}$$

takes the cusp $i\infty$ to $1/p$, we have $\pi_E(h\gamma' h^{-1}) = qa_0(E[1/p])$. We further decompose γ' as

$$\begin{pmatrix} 1 & -2q \\ -p(1+q) & 1+2pq(1+q) \end{pmatrix} \begin{pmatrix} 1 & 2q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2q \\ -p(1+q) & 1+2pq(1+q) \end{pmatrix}^{-1}.$$

The matrix

$$\begin{pmatrix} 1 & -2q \\ p(1+q) & 1+2pq(1+q) \end{pmatrix}$$

takes the cusp $i\infty$ to $1/p$. We see that $\pi_E(\gamma') = 2qa_0(E[1/p])$ and $\int_{z_0}^{\gamma' z_0} k^*(\omega_E) = 3a_0(E[1/p])$. A simple calculation shows that

$$A'(\mathcal{E}_E^0) = \sum_{k=0}^{q-1} F_E(\beta'_k) - \sum_{k=0}^{q-1} F_E(\alpha'_{kp}) - F_E(\gamma'_m) = 2\sum_{k=0}^{q-1} F_E(\beta'_k) = -6a_0\left(E\left[\frac{1}{p}\right]\right).$$

By interchanging p and q, we get $B'(\mathcal{E}_E^0) = -6a_0(E[1/q])$. Since $\alpha'_{pq} S \in \Gamma_0(pq)$, a calculation similar to (7-2) shows that

$$\begin{aligned} F_E(I) &= -F_E(\alpha_{pq}) \\ &= \int_1^{-1} \left(2E(z) - \tfrac{1}{2}E\left(\tfrac{1}{2}(z+1)\right)\right) dz \\ &= -\int_{-1}^{\beta(-1)} \left(2E(z) - \tfrac{1}{2}E\left(\tfrac{1}{2}(z+1)\right)\right) dz \\ &= -3a_0(E). \end{aligned}$$

We conclude that $C'(\mathcal{E}_E^0) = F_E(I) - F_E(\alpha_{pq}) = -6a_0(E)$ and hence $\delta^0(\mathcal{E}_E^0) = \delta(\mathcal{E}_E^0) = -6\delta(E)$. □

The inclusion map

$$i : (X_0(pq) - R \cup I, \partial(X_0(pq))) \to (X_0(pq), \partial(X_0(pq)))$$

induces an onto map

$$i_* : \mathrm{H}_1(X_0(pq) - R \cup I, \partial(X_0(pq)), \mathbb{Z}) \to \mathrm{H}_1(X_0(pq), \partial(X_0(pq)), \mathbb{Z})$$

with $i_*([g]^*) = \xi(g)$. Note that $\delta([g]^*) = [g0] - [g\infty] = \delta'(\xi(g)) = \delta'(i_*([g]^*))$. From Section 3.4, we have that $\delta(c) = \delta'(i_*(c))$ for all homology classes $c \in \mathrm{H}_1(X_0(pq) - R \cup I, \partial(X_0(pq)), \mathbb{Z})$.

Lemma 36. *The integrals of every holomorphic differential on $X_0(pq)$ over $i_*(\mathcal{E}'_E)$ and $i_*\pi_*(\mathcal{E}^0_E)$ are zero.*

Proof. The proof is a straightforward generalization of [Merel 1996b, Lemma 5]. □

We now prove the main theorem.

Proof of Theorem 1. By [Merel 1995, Corollary 3], we obtain $i_*(\mathcal{E}'_E) \circ c = \mathcal{E}'_E \circ i^* c = \int_c i_*(E(z)\,dz)$. Hence, $i_*(\mathcal{E}'_E)$ is the Eisenstein element inside the space of modular symbols corresponding to E. By Propositions 34 and 35, the boundary of $\pi_*(\mathcal{E}^0_E)$ is the same as the boundary of $6i_*(\mathcal{E}'_E)$.

There is a nondegenerate bilinear pairing $S_2(\Gamma_0(pq)) \times \mathrm{H}_1(X_0(pq), \mathbb{R}) \to \mathbb{C}$ given by $(f, c) = \int_c f(z)\,dz$. Hence, the integrals of holomorphic differentials over $\mathrm{H}_1(X_0(pq), \mathbb{Z})$ are not always zero. By Lemma 36, the integrals of holomorphic differentials over $i_*(\mathcal{E}'_E)$ and $i_*(\pi_*(\mathcal{E}^0_E))$ are always zero. We deduce that

$$\mathcal{E}_E = i_*(\mathcal{E}'_E) = \frac{1}{6} i_*\pi_*(\mathcal{E}^0_E) = \frac{1}{6} \sum_{g \in \mathbb{P}^1(\mathbb{Z}/pq\mathbb{Z})} F_E(g)\xi(g)$$

for $E \in \mathbb{E}_{pq}$. Since $F_N(g) = \frac{1}{6} F_{E_N}(g)$, we obtain the theorem. □

7.2. ***Winding elements of level pq.*** Recall the concept of the *winding element*.

Definition 37 (winding element). Let $\{0, \infty\}$ denote the projection of the path from 0 to ∞ in $\mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ to $X_0(pq)(\mathbb{C})$. We have an isomorphism $\mathrm{H}_1(X_0(pq), \mathbb{Z}) \otimes \mathbb{R} = \mathrm{Hom}_{\mathbb{C}}(\mathrm{H}^0(X_0(pq), \Omega^1), \mathbb{C})$. Let $e_{pq} \in \mathrm{H}_1(X_0(pq), \mathbb{R})$ correspond to the homomorphism $\omega \to -\int_0^\infty \omega$. The modular symbol e_{pq} is called the *winding element*.

The winding elements are the elements of the space of modular symbols whose annihilators define ideals of the Hecke algebras with the L-functions of the corresponding quotients of the Jacobian nonzero. In this paper, we find an explicit expression of the winding element. Let $e_{pq} \in \mathrm{H}_1(X_0(pq), \mathbb{Z}) \otimes \mathbb{R}$ be the winding element. The following lemma will help us write down the winding element explicitly. Since $\sum_{x\in\partial(X_0(pq))} e_{\Gamma_0(pq)}(x)a_0(E[x])=0$, we write

$$\delta(E)=a_0(E)(\{\infty\}-\{0\})+qa_0\Big(E\Big[\frac{1}{p}\Big]\Big)\Big(\Big\{\frac{1}{p}\Big\}-\{0\}\Big)+pa_0\Big(E\Big[\frac{1}{q}\Big]\Big)\Big(\Big\{\frac{1}{q}\Big\}-\{0\}\Big).$$

Lemma 38. *The constant Fourier coefficients of E_{pq} at cusps 0, $1/p$, $1/q$ and ∞ are $\frac{1}{24}(1-pq)/(pq)$, 0, 0 and $\frac{1}{24}(pq-1)$ respectively.*

Proof. We first prove that the constant coefficient for the Fourier expansion of E_{pq} at the cusp $1/p$ is 0. As usual, the constant term of the Fourier expansion of E_{pq} at the cusp $1/p$ is the constant term at ∞ of $E_{pq}[\beta_0]$. Similarly, the constant term of the Fourier expansion of E_{pq} at the cusp $1/q$ is the constant term at ∞ of $E_{pq}[\gamma_0]$. Let Δ be the Ramanujan cusp form of weight 12. We write

$$\frac{d}{dz}\log\Delta(\beta(z))=12\frac{d}{dz}\log(pz+1)+\frac{d}{dz}\log\Delta(z)\quad\text{for } \beta=\begin{pmatrix}1&0\\p&1\end{pmatrix}.$$

A simple calculation shows that

$$\begin{aligned}\Delta\Big(\frac{pqz}{pz+1}\Big)&=\Delta\Big(\begin{pmatrix}q&0\\1&1\end{pmatrix}pz\Big)\\&=\Delta\Big(\begin{pmatrix}q&-1\\1&0\end{pmatrix}\begin{pmatrix}1&1\\0&q\end{pmatrix}pz\Big)\\&=\Delta\Big(\begin{pmatrix}q&-1\\1&0\end{pmatrix}\Big(\frac{pz+1}{q}\Big)\Big)\\&=\Big(\frac{pz+1}{q}\Big)^{12}\Delta\Big(\frac{pz+1}{q}\Big).\end{aligned}$$

By taking logarithmic derivative, we deduce that

$$\frac{d}{dz}\log\Delta\begin{pmatrix}q&-1\\1&0\end{pmatrix}\Big(\frac{pz+1}{q}\Big)=12\frac{d}{dz}\log(pz+1)+\frac{d}{dz}\log\Delta\Big(\frac{pz+1}{q}\Big).$$

Since

$$E_{pq}(z)=\frac{1}{2\pi i}\frac{d}{dz}\log\frac{\Delta(pqz)}{\Delta(z)},$$

the above calculation shows that the constant term of E_{pq} at the cusp $1/p$ is 0. Similarly, the constant term of E_{pq} at the cusp $1/q$ is 0. The constant term of E_{pq} at the cusp ∞ is $\frac{1}{24}(pq-1)$ and at 0 is $\frac{1}{24}(1-pq)/(pq)$. $\square$

Using Lemmas 36 and 38, we have:

Corollary 39.

$$(1-pq)e_{pq}=\sum_{x\in(\mathbb{Z}/pq\mathbb{Z})^*}F_{pq}(1,x)\Big\{0,\frac{1}{x}\Big\}.$$

Remark 40. For the Eisenstein series $E_p \in E_2(\Gamma_0(p))$, $1/p$ represents the cusp ∞ and $1/q$ represents the cusp 0. We deduce that $a_0(E_p[\beta_0]) = \frac{1}{24}(p-1)$ and $a_0(E_p[\gamma_0]) = \frac{1}{24}(1-p)/p$. For the other Eisenstein series $E_q \in E_2(\Gamma_0(q))$, $1/q$ represents the cusp ∞ and $1/p$ represents the cusp 0. We deduce that $a_0(E_q[\gamma_0]) = \frac{1}{24}(q-1)$ and $a_0(E_q[\beta_0]) = \frac{1}{24}(1-q)/q$.

References

[Agashe 2000] A. S. Agashe, *The Birch and Swinnerton-Dyer formula for modular abelian varieties of analytic rank zero*, thesis, University of California, Berkeley, 2000, available at http://search.proquest.com/docview/304623798. MR 2701025

[Banerjee 2014] D. Banerjee, "A note on the Eisenstein elements of prime square level", *Proc. Amer. Math. Soc.* **142**:11 (2014), 3675–3686. MR 3251709 Zbl 06345376

[Calegari and Emerton 2005] F. Calegari and M. Emerton, "On the ramification of Hecke algebras at Eisenstein primes", *Invent. Math.* **160**:1 (2005), 97–144. MR 2007h:11066 Zbl 1145.11314

[Cremona 1997] J. E. Cremona, *Algorithms for modular elliptic curves*, 2nd ed., Cambridge University Press, 1997. MR 99e:11068 Zbl 0872.14041

[Diamond and Shurman 2005] F. Diamond and J. Shurman, *A first course in modular forms*, Graduate Texts in Mathematics **228**, Springer, New York, 2005. MR 2006f:11045 Zbl 1062.11022

[Drinfeld 1973] V. G. Drinfeld, "Две теоремы о модулярных кривых", *Funkcional. Anal. i Priložen.* **7**:2 (1973), 83–84. Translated as "Two theorems on modular curves" in *Funct. Anal. Appl.* **7**:2 (1973), 155–156. MR 47 #6705 Zbl 0285.14006

[Lang 1995] S. Lang, *Introduction to modular forms*, Grundlehren der Mathematischen Wissenschaften **222**, Springer, Berlin, 1995. MR 96g:11037 Zbl 0344.10011

[Manin 1972] J. I. Manin, "Параболические точки и дзета-функции модулярных кривых", *Izv. Akad. Nauk SSSR Ser. Mat.* **36**:1 (1972), 19–66. Translated as "Parabolic points and zeta-functions of modular curves" in *Math. USSR Izv.* **6**:1 (1972), 19–64. MR 47 #3396 Zbl 0243.14008

[Mazur 1977] B. Mazur, "Modular curves and the Eisenstein ideal", *Inst. Hautes Études Sci. Publ. Math.* **47** (1977), 33–186. MR 80c:14015 Zbl 0394.14008

[Mazur 1979] B. Mazur, "On the arithmetic of special values of L functions", *Invent. Math.* **55**:3 (1979), 207–240. MR 82e:14033 Zbl 0426.14009

[Merel 1993] L. Merel, "Intersections sur des courbes modulaires", *Manuscripta Math.* **80**:3 (1993), 283–289. MR 94k:14014 Zbl 0812.14013

[Merel 1995] L. Merel, "Homologie des courbes modulaires affines et paramétrisations modulaires", pp. 110–130 in *Elliptic curves, modular forms, and Fermat's last theorem* (Hong Kong, 1993), edited by J. Coates and S. T. Yau, Series in Number Theory **1**, International Press, Cambridge, MA, 1995. MR 97a:11094 Zbl 0845.11023

[Merel 1996a] L. Merel, "Bornes pour la torsion des courbes elliptiques sur les corps de nombres", *Invent. Math.* **124**:1-3 (1996), 437–449. MR 96i:11057 Zbl 0936.11037

[Merel 1996b] L. Merel, "L'accouplement de Weil entre le sous-groupe de Shimura et le sous-groupe cuspidal de $J_0(p)$", *J. Reine Angew. Math.* **477** (1996), 71–115. MR 97f:11045 Zbl 0859.11036

[Miranda 1995] R. Miranda, *Algebraic curves and Riemann surfaces*, Graduate Studies in Mathematics **5**, American Mathematical Society, Providence, RI, 1995. MR 96f:14029 Zbl 0820.14022

[Miyake 1976] T. Miyake, *Hokei keishiki to seisiiron*, Kinokuniya, Tokyo, 1976. Translated by Yoshitaka Maeda as *Modular forms*, Springer, Berlin, 1989. Corrected 2nd printing in 2006. MR 2006g:11084 Zbl 1159.11014

[Ogg 1974] A. P. Ogg, "Hyperelliptic modular curves", *Bull. Soc. Math. France* **102** (1974), 449–462. MR 51 #514 Zbl 0314.10018

[Stevens 1982] G. Stevens, *Arithmetic on modular curves*, Progress in Mathematics **20**, Birkhäuser, Boston, 1982. MR 87b:11050 Zbl 0529.10028

[Stevens 1985] G. Stevens, "The cuspidal group and special values of L-functions", *Trans. Amer. Math. Soc.* **291**:2 (1985), 519–550. MR 87a:11056 Zbl 0579.10011

Received April 14, 2015. Revised June 25, 2015.

DEBARGHA BANERJEE
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE
PASHAN ROAD
PASHAN
PUNE 411008
INDIA
debargha@iiserpune.ac.in

SRILAKSHMI KRISHNAMOORTHY
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
SARDAR PATEL ROAD
CHENNAI
ADYAR 600036
INDIA
srilakshmi@iitm.ac.in

PACIFIC JOURNAL OF MATHEMATICS
Vol. 281, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.281.287

PRIMITIVELY GENERATED HALL ALGEBRAS

ARKADY BERENSTEIN AND JACOB GREENSTEIN

Dedicated to Professor Anthony Joseph on the occasion of his seventieth birthday

In the present paper we show that Hall algebras of finitary exact categories behave like quantum groups in the sense that they are generated by indecomposable objects. Moreover, for a large class of such categories, Hall algebras are generated by their primitive elements, with respect to the natural comultiplication, even for nonhereditary categories. Finally, we introduce certain primitively generated subalgebras of Hall algebras and conjecture an analogue of "Lie correspondence" for those finitary categories.

1. Introduction

It is well-known that quantum groups are not groups, but rather Hopf algebras, which are similar to enveloping algebras of Lie algebras. Hall–Ringel algebras $H_{\mathscr{A}}$ of finitary exact categories can be regarded, from many points of view, as generalizations of quantum groups. One aspect of this analogy is the following striking result, which we failed to find in the literature.

Theorem 1.1. *The Hall algebra $H_{\mathscr{A}}$ of any finitary exact category $\mathscr{A}$ is generated by isomorphism classes of indecomposable objects in $\mathscr{A}$.*

The authors were partially supported by NSF grants DMS-0800247, DMS-1101507 (Berenstein) and DMS-0654421 (Greenstein) and Simons foundation collaboration grant 245735 (Greenstein).
MSC2010: 16G20, 17B37, 16T15, 16T20.
Keywords: Hall algebra, exact category, PBW property, Nichols algebra.

We prove a refinement of this theorem (Theorem 2.4), which is an analogue of the Poincaré–Birkhoff–Witt property for $H_{\mathscr{A}}$, in §4.3.

However, isomorphism classes of indecomposable objects are not the most efficient as a generating set. For example, if $\mathscr{A}$ is the representation category of a (valued) Dynkin quiver Q, then indecomposables correspond to all positive roots of the simple Lie algebra associated with Q, while $H_{\mathscr{A}}$ can be generated by simple objects (in other words, indecomposables corresponding to simple roots of the Lie algebra). Having this in mind, we introduce minimal generating sets for $H_{\mathscr{A}}$, namely, primitive elements, which generalize these simple root generators.

More precisely, for any finitary exact category $\mathscr{A}$, the Hall algebra $H_{\mathscr{A}}$ has a natural coproduct $\Delta : H_{\mathscr{A}} \to H_{\mathscr{A}} \widehat{\otimes} H_{\mathscr{A}}$ whose image may lie in a suitable completion of the tensor square of $H_{\mathscr{A}}$. Note, however, that the multiplication and Δ are not always compatible, that is, Δ need not be a homomorphism of algebras. The compatibility is guaranteed by *Green's theorem* (see [Green 1995]) for all *hereditary cofinitary* (so that Δ is an "honest" comultiplication rather than a topological one) *abelian* categories $\mathscr{A}$ (see Definition 2.11). This includes all categories $\operatorname{rep}_{\Bbbk} Q$ of finite dimensional representations over a finite field $\Bbbk$ of an acyclic (valued) quiver Q. In a remarkable paper, Sevenhant and Van den Bergh [2001] proved that for $\mathscr{A} = \operatorname{rep}_{\Bbbk} Q$ the Hall algebra $H_{\mathscr{A}}$ is a Nichols algebra in an appropriate braided tensor category (see §2.6 for details) and, in particular, is generated by its space of primitive elements

$$V_{\mathscr{A}} = \{v \in H_{\mathscr{A}} : \Delta(v) = v \otimes 1 + 1 \otimes v\}.$$

We extend this result to a much larger class of categories that we refer to as *profinitary* categories. We introduce profinitary categories in terms of their *Grothendieck monoids* (denoted $\Gamma_{\mathscr{A}}$ for an exact category $\mathscr{A}$, see §2.3 for precise definitions) by requiring that groups of morphisms between any two objects and all Grothendieck equivalence classes are finite. By definition, $H_{\mathscr{A}}$ is naturally graded by $\Gamma_{\mathscr{A}}$ and if $\mathscr{A}$ is profinitary, all homogeneous components $(H_{\mathscr{A}})_\gamma$, $\gamma \in \Gamma_{\mathscr{A}}$ are finite dimensional.

The class of profinitary categories is large enough. For instance, it includes the abelian category $R-\operatorname{fin}$ of all *finite* R-modules M (i.e., finite abelian groups with R-action) for a *finitary* unital ring R, as defined in [Ringel 1990a, §1]. This includes all finitely generated (over $\mathbb{Z}$) unital rings. Moreover, if $\mathscr{A}$ is profinitary, then so is any full subcategory $\mathscr{B} \subset \mathscr{A}$ closed under extensions. The following is the main result of the present work.

Main Theorem 1.2. *For any profinitary and cofinitary exact category $\mathscr{A}$, the Hall algebra $H_{\mathscr{A}}$ is generated by the space $V_{\mathscr{A}}$ of its primitive elements. Moreover, $V_{\mathscr{A}}$ is minimal in the sense that a nonzero element of $V_{\mathscr{A}}$ cannot be expressed as a sum of products of elements of $V_{\mathscr{A}}$.*

We prove Main Theorem 1.2 in §6.4.

Based on the second assertion of Main Theorem 1.2, we can introduce *quasi-Nichols algebras* as both algebras and coalgebras minimally generated by their primitive elements (see Definition 2.17 for details). In particular, it is easy to see (cf. Lemma 2.25) that any Nichols algebra is quasi-Nichols. It is noteworthy that the minimality of $V_{\mathscr{A}}$ has the following nice consequence for constructing primitive elements in $H_{\mathscr{A}}$: once we find a subspace U of $V_{\mathscr{A}}$ such that U generates $H_{\mathscr{A}}$ as an algebra, we must stop because U is the space of *all* primitive elements in $H_{\mathscr{A}}$.

Remark 1.3. Similarly to Grothendieck groups, exact functors induce canonical homomorphisms of Grothendieck monoids. However, even for full embeddings, such homomorphisms need not be injective. On the other hand, unlike the Grothendieck group, the Grothendieck monoid always separates simple objects of the category. For instance, if $\mathscr{A}$ is the category of $\Bbbk$-representations of the quiver $Q = 1 \to 2$ with dimension vectors $(n, 2n)$, $n \in \mathbb{Z}_{\geq 0}$, then $K_0(\mathscr{A}) \cong \mathbb{Z}$, but $\Gamma_{\mathscr{A}}$ is an additive monoid generated by β_1, β_2 subject to the relations $\beta_1 + \beta_2 = 2\beta_1 = 2\beta_2$. The canonical homomorphism $\Gamma_{\mathscr{A}} \to K_0(\mathscr{A})$ is given by $\beta_1 \mapsto 1$, $\beta_2 \mapsto 1$ and thus is not injective (see §3.4 for details.) It should also be noted that in this example $\Gamma_{\mathscr{A}}$ is not a submonoid of the Grothendieck monoid of the category $\operatorname{rep}_{\Bbbk} Q$ since in $\Gamma_{\operatorname{rep}_{\Bbbk} Q}$ both simple objects of $\mathscr{A}$ belong to the same class.

A nice property of profinitary categories is that their Hall algebras always contain primitive elements. If $\mathscr{A}$ is profinitary, then its Grothendieck monoid admits a natural partial order and is generated by its minimal elements with respect to that order (Proposition 2.12). Moreover, for γ minimal the corresponding homogeneous component $(H_{\mathscr{A}})_\gamma$ of $H_{\mathscr{A}}$ is one-dimensional and primitive.

Quite surprisingly, for a profinitary category, cofinitarity is a simple property of its Grothendieck monoid. We say that a monoid Γ is locally finite if for all $\gamma \in \Gamma$, the set $\{(\alpha, \beta) \in \Gamma \times \Gamma : \alpha + \beta = \gamma\}$ is finite.

Theorem 1.4. *A profinitary exact category $\mathscr{A}$ is cofinitary if and only if $\Gamma_{\mathscr{A}}$ is locally finite.*

We prove this theorem in §5.3. As a corollary, we obtain two classes of categories for which profinitarity implies cofinitarity.

Corollary 1.5. (a) *Any full exact subcategory of a profinitary abelian category is cofinitary.*

(b) *Any profinitary exact category whose Grothendieck monoid is finitely generated is cofinitary.*

This corollary is proven in §5.3. Based on the above, we propose the following conjecture.

Conjecture 1.6. *For any profinitary exact category $\mathscr{A}$, its Grothendieck monoid $\Gamma_{\mathscr{A}}$ is locally finite.*

By Theorem 1.4, any category as in the above conjecture is also cofinitary.

This conjecture is nontrivial since there exist profinitary exact categories $\mathscr{A}$ for which any ambient abelian category $\overline{\mathscr{A}}$ (which always exists, see, e.g., [Bühler 2010; Keller 1990]) is not profinitary, and the monoid $\Gamma_{\mathscr{A}}$ need not be finitely generated.

Main Theorem 1.2 and Corollary 1.5(a) imply the following theorem.

Theorem 1.7. *If $\mathscr{A}$ is a profinitary hereditary abelian category, then $H_{\mathscr{A}}$ is a Nichols algebra (see Definition 2.23) of the (braided) space $V_{\mathscr{A}}$ of its primitive elements.*

We prove a refined version of this statement (Theorem 2.26) in §7.2.

The case when $\mathscr{A} = \operatorname{rep}_{\Bbbk} Q$ where Q is a finite acyclic (valued) quiver was established in [Sevenhant and Van Den Bergh 2001, Theorem 1.1], which inspired the present work. If $\mathscr{A}$ is the category of nilpotent representations of $\Bbbk[x]$ for a finite field $\Bbbk$, then Theorem 1.7 recovers the classical result of Zelevinsky [1981] that the Hall–Steinitz algebra is a Hopf algebra (see, e.g., §3.1 for details). More generally, it is well-known that the category $\operatorname{rep}_{\Bbbk} Q$ for *any* finite valued quiver Q is hereditary (see [Gabriel 1973; Hubery 2007]). Therefore, Theorem 1.7 is applicable to such a category as well, that is, $H_{\operatorname{rep}_{\Bbbk} Q}$ is a Nichols algebra. In particular, so is the Hall algebra of the category of finite dimensional modules of the free algebra in n generators over $\Bbbk$.

Furthermore, by definition, $V_{\mathscr{A}}$ is graded by $\Gamma_{\mathscr{A}}$, that is, $V_{\mathscr{A}} = \bigoplus_{\gamma \in \Gamma} (V_{\mathscr{A}})_\gamma$, so $\gamma \in \Gamma_{\mathscr{A}}$ with $(V_{\mathscr{A}})_\gamma \neq 0$ can be thought of as "simple roots" of $\mathscr{A}$. Given $\gamma \in \Gamma_{\mathscr{A}}^+$, define its *multiplicity* m_γ by

$$m_\gamma := \# \operatorname{Ind} \mathscr{A}_\gamma - \dim_{\mathbb{Q}} (V_{\mathscr{A}})_\gamma, \tag{1-1}$$

where $\operatorname{Ind} \mathscr{A}_\gamma = \operatorname{Ind} \mathscr{A} \cap \operatorname{Iso} \mathscr{A}_\gamma$. This definition is justified by the following proposition.

Proposition 1.8. *Let $\mathscr{A}$ be a profinitary cofinitary exact category. Then $m_\gamma \geq 0$ for all $\gamma \in \Gamma_{\mathscr{A}}^+$.*

We prove a more precise version of this result (Proposition 2.20) in §6.5. In particular, Proposition 1.8 implies that if $\operatorname{Ind} \mathscr{A}_\gamma = \varnothing$ then $(V_{\mathscr{A}})_\gamma = 0$, that is, we should look for primitive elements only in those graded components where indecomposables live. Moreover, if $\operatorname{Ind} \mathscr{A}$ is finite, then obviously $V_{\mathscr{A}}$ is finite dimensional and we have an efficient procedure for computing it (see §3).

The term "multiplicity" is justified by the following result, which is an immediate consequence of reformulations [Hua 2000, Theorem 4.1; Deng and Xiao 2003, §4.1] of the famous Kac conjecture [Kac 1980], proved in [Hausel 2010].

Theorem 1.9. *Let Q be an acyclic quiver, $\mathfrak{g}_Q$ be the corresponding Kac–Moody algebra and $\mathscr{A} = \mathrm{rep}_{\Bbbk}(Q)$ where $\Bbbk$ is a finite field with q elements. Then for any $\gamma \in \Gamma_{\mathscr{A}}$ one has*:

(a) $m_\gamma > 0$ *if and only if γ is a nonsimple positive root of $\mathfrak{g}_Q$; in that case, $m_\gamma = \dim(\mathfrak{g}_Q)_\gamma$, that is m_γ is the multiplicity of the root γ in $\mathfrak{g}_Q$.*

(b) $(V_{\mathscr{A}})_\gamma = 0$ *unless γ is simple or imaginary.*

(c) *For any imaginary root γ of $\mathfrak{g}_Q$, $\dim_{\mathbb{Q}}(V_{\mathscr{A}})_\gamma = p_\gamma(q)$ where $p_\gamma \in x\mathbb{Q}[x]$.*

In view of Theorem 1.9(c) and results of [Sevenhant and Van Den Bergh 2001] we define *real simple roots* of $\mathscr{A}$ to be elements $\gamma \in \Gamma_{\mathscr{A}}$ for which $\dim_{\mathbb{Q}}(V_{\mathscr{A}})_\gamma = 1$ and *imaginary simple roots* of $\mathscr{A}$ to be those $\gamma \in \Gamma_{\mathscr{A}}$ with $\dim_{\mathbb{Q}}(V_{\mathscr{A}})_\gamma \geq 2$. For a profinitary category $\mathscr{A}$ we show (Lemma 5.3) that all minimal elements of $\Gamma_{\mathscr{A}} \setminus \{0\}$ are real simple roots.

In fact, the consideration of examples suggests that a stronger version of this statement holds.

Conjecture 1.10. *Let $\mathscr{A}$ be a profinitary and cofinitary exact category. Then each simple imaginary root of $\mathscr{A}$ has nonzero multiplicity.*

Clearly, Theorem 1.9 verifies this conjecture when $\mathscr{A} = \mathrm{rep}_{\Bbbk}(Q)$ for any finite acyclic quiver Q. We provide more supporting evidence in §3. In those cases, $m_\gamma = 1$ quite frequently (see §3.2, §3.3 and §3.4).

Simple real roots are of special interest. Denote by $U_{\mathscr{A}}$ the subalgebra of $H_{\mathscr{A}}$ generated by all $(V_{\mathscr{A}})_\alpha$, where α runs over all real simple roots of $\mathscr{A}$, and refer to it as the *quantum enveloping algebra* of $\mathscr{A}$. The following well-known fact justifies this definition.

Theorem 1.11 [Ringel 1990b]. *If Q is an acyclic valued quiver, then $U_{\mathrm{rep}_{\Bbbk} Q}$ is isomorphic to a quantized enveloping algebra of the nilpotent part of $\mathfrak{g}_Q$.*

Since $[X] \in \mathrm{Iso}\,\mathscr{A}$ is primitive if and only if it is almost simple (see Definition 5.2), the algebra $U_{\mathscr{A}}$ contains the subalgebra $C_{\mathscr{A}}$ of $H_{\mathscr{A}}$ generated by isomorphism classes of all almost simple objects. We call $C_{\mathscr{A}}$ the *composition algebra* of $\mathscr{A}$ since it generalizes the composition algebra of $\mathrm{rep}_{\Bbbk} Q$, which is the subalgebra of $H_{\mathrm{rep}_{\Bbbk} Q}$ generated by isomorphism classes of simple objects. In fact, in the assumptions of the above theorem, $U_{\mathrm{rep}_{\Bbbk} Q} = C_{\mathrm{rep}_{\Bbbk} Q}$. However, it frequently happens that $C_{\mathscr{A}} \subsetneq U_{\mathscr{A}}$ (see §3 for examples). Note the following corollary of Theorem 1.7 and [Andruskiewitsch and Schneider 2002, Corollary 2.3] (see Lemma 2.24).

Corollary 1.12. *If $\mathscr{A}$ is a profinitary hereditary abelian category then both $C_{\mathscr{A}}$ and $U_{\mathscr{A}}$ are Nichols algebras.*

It turns out that there is another algebra $E_{\mathscr{A}}$, which (yet conjecturally) "squeezes" between these two. That is, $E_{\mathscr{A}}$ is generated by elements $e_\gamma \in H_{\mathscr{A}}$, where e_γ is the sum of all isomorphism classes of objects of $\mathscr{A}$ whose image in Γ is γ. Since

$$\operatorname{Exp}_{\mathscr{A}} := \sum_{\gamma \in \Gamma_A} e_\gamma$$

is a group-like element in the completion of $H_{\mathscr{A}}$ with respect to a slightly different coproduct (see [Berenstein and Greenstein 2013, Lemma A.1]), we referred to $\operatorname{Exp}_{\mathscr{A}}$ in [Berenstein and Greenstein 2013] as *the exponential* of $\mathscr{A}$. Hence we sometimes refer to $E_{\mathscr{A}}$ as the exponential algebra of $\mathscr{A}$. By definition, $C_{\mathscr{A}} \subset E_{\mathscr{A}}$.

Conjecture 1.13. *For any profinitary category $\mathscr{A}$ one has*

$$E_{\mathscr{A}} = U_{\mathscr{A}}.$$

In particular, $\operatorname{Exp}_{\mathscr{A}}$ *belongs to the completion of* $U_{\mathscr{A}}$.

In §3 we provide several supporting examples of profinitary categories $\mathscr{A}$ together with the explicit presentations of $H_{\mathscr{A}}$, $U_{\mathscr{A}}$ and $E_{\mathscr{A}}$.

The significance of the conjecture is that it paves the ground for the "Lie correspondence" between the enveloping algebra $U_{\mathscr{A}}$ and the quantum Chevalley group $G_{\mathscr{A}}$ that we introduced in [Berenstein and Greenstein 2013] as an analogue of the corresponding Lie group. That is, Conjecture 1.13 implies that the "tame" part of $G_{\mathscr{A}}$ belongs to the completion of $U_{\mathscr{A}}$.

2. Definitions and main results

2.1. ***Exact categories and Hall algebras.*** All categories are assumed to be essentially small. For such a category $\mathscr{A}$ we denote by $\operatorname{Iso}\mathscr{A}$ the set of isomorphism classes of objects in $\mathscr{A}$. We say that a category $\mathscr{A}$ is Hom-finite if $\operatorname{Hom}_{\mathscr{A}}(X, Y)$ is a finite set for all $X, Y \in \mathscr{A}$.

Let $\mathscr{A}$ be an exact category, in the sense of [Quillen 1973] (see also [Keller 1990; Bühler 2010]). We denote by $\underline{\operatorname{Ext}}^1_{\mathscr{A}}(A, B)$ the set of all isomorphism classes $[X] \in \operatorname{Iso}\mathscr{A}$ such that there exists a short exact sequence

$$B \overset{f}{\rightarrowtail} X \overset{g}{\twoheadrightarrow} A \tag{2-1}$$

(here f is a *monomorphism*, g is an *epimorphism*, f is a kernel of g and g is a cokernel of f). We say that $\mathscr{A}$ is *finitary* if it is Hom-finite and $\underline{\operatorname{Ext}}^1_{\mathscr{A}}(A, B)$ is finite for every $A, B \in \mathscr{A}$.

Following [Hubery 2006] we define Hall numbers for finitary exact categories as follows. For $A, B, X \in \mathscr{A}$ fixed, denote by $\mathcal{E}(A, B)_X$ the set of all short exact sequences (2-1). The group $\operatorname{Aut}_{\mathscr{A}} A \times \operatorname{Aut}_{\mathscr{A}} B$ acts freely on $\mathcal{E}(A, B)_X$ by

$$(\varphi, \psi).(f, g) = (f\varphi^{-1}, \psi g), \quad \varphi \in \operatorname{Aut}_{\mathscr{A}} B, \psi \in \operatorname{Aut}_{\mathscr{A}} A.$$

The Hall number F_{AB}^X is the number of $\mathrm{Aut}_{\mathscr{A}} A \times \mathrm{Aut}_{\mathscr{A}} B$-orbits in $\mathcal{E}(A, B)_X$ and equals

$$F_{AB}^X = \frac{\#\mathcal{E}(A, B)_X}{\#(\mathrm{Aut}_{\mathscr{A}} A \times \mathrm{Aut}_{\mathscr{A}} B)}.$$

Denote

$$H_{\mathscr{A}} = \mathbb{Q}\,\mathrm{Iso}\,\mathscr{A} = \bigoplus_{[X]\in \mathrm{Iso}\,\mathscr{A}} \mathbb{Q}\cdot[X].$$

Proposition 2.1 [Ringel 1990a; Hubery 2006]. *For any finitary exact category $\mathscr{A}$, the space $H_{\mathscr{A}}$ is an associative unital $\mathbb{Q}$-algebra with the product given by*

$$[A]\cdot[B] = \sum_{[C]\in \mathrm{Iso}\,\mathscr{A}} F_{A,B}^C[C]. \tag{2-2}$$

The unity $1 \in H_{\mathscr{A}}$ is the class $[0]$ of the zero object of $\mathscr{A}$.

It is well-known (see, e.g., [Bühler 2010; Keller 1990]) that each exact category $\mathscr{A}$ can be realized as a full subcategory closed under extensions of an abelian category $\overline{\mathscr{A}}$. However, even if $\mathscr{A}$ is finitary, it might be impossible to find an ambient abelian category which is also finitary. On the other hand, any full subcategory of a finitary abelian category closed under extensions is also finitary.

2.2. ***Ordered monoids and the PBW property of Hall algebras.*** Let Λ be an abelian monoid. We say that Λ is *ordered* if there exists a partial order $\lhd$ on Λ^+ such that for $\mu, \mu', \nu, \nu' \in \Lambda^+$, we have

$$\mu \lhd \nu,\ \mu' \unlhd \nu' \Longrightarrow \mu + \mu' \lhd \nu + \nu'.$$

Let $\mathscr{A}$ be a finitary exact category. The set $\mathrm{Iso}\,\mathscr{A}$ is naturally an abelian monoid with the addition operation defined by $[X]+[Y] = [X \oplus Y]$. Every object in $\mathscr{A}$ is a finite direct sum of indecomposable objects (see Lemma 4.9). Thus, in particular, $\mathrm{Ind}\,\mathscr{A}$ generates $\mathrm{Iso}\,\mathscr{A}$ as a monoid. The category $\mathscr{A}$ is said to be Krull–Schmidt if $\mathrm{Iso}\,\mathscr{A}$ is freely generated by $\mathrm{Ind}\,\mathscr{A}$.

Define a relation $\lhd$ on $(\mathrm{Iso}\,\mathscr{A})^+$ by $[M] \lhd [N]$ if

(i) $[N] = [M^+ \oplus M^-]$, and

(ii) there exists a nonsplit short exact sequence $M^- \rightarrowtail M \twoheadrightarrow M^+$.

By abuse of notation, we also denote by $\lhd$ the transitive closure of this relation.

We say that a partial (pre)order $\prec$ on a set Λ is *inductive* if there exists a function $f : \Lambda \to \mathbb{Z}_{\geq 0}$ such that $\lambda \prec \mu \Longrightarrow f(\lambda) < f(\mu)$. It is obvious that an inductive preorder is a partial order.

Theorem 2.2. *Let $\mathscr{A}$ be a* Hom*-finite exact category. Then* $(\mathrm{Iso}\,\mathscr{A}, \lhd)$ *is an ordered monoid and $\lhd$ is inductive with the function $f : \mathrm{Iso}\,\mathscr{A} \to \mathbb{Z}_{\geq 0}$ given by*

$$f([M]) = \#\,\mathrm{End}_{\mathscr{A}} M.$$

Remark 2.3. If the category $\mathscr{A}$ is finitary, one can show that the assertion holds with f replaced by the function $[M] \mapsto \#\operatorname{Ext}^1_{\mathscr{A}}(M, M)$, $[M] \in \operatorname{Iso}\mathscr{A}$.

We prove this theorem in §4.2. It is used as the key ingredient in a proof of the following theorem, which generalizes [Guo and Peng 1997, Theorem 3.1] and establishes the (weak) PBW property of Hall algebras.

Theorem 2.4. *Let $\mathscr{A}$ be a finitary exact category. Then for any total order on the set* Ind $\mathscr{A}$ *of isomorphism classes of indecomposable objects in $\mathscr{A}$, $H_{\mathscr{A}}$ is spanned, as a $\mathbb{Q}$-vector space, by ordered monomials on* Ind $\mathscr{A}$*. Moreover, if $\mathscr{A}$ is Krull–Schmidt, then such monomials form a basis of $H_{\mathscr{A}}$.*

We prove this theorem in §4.3. After [Joyce 2007; Riedtmann 1994], this further extends an analogy between Hall algebras of finitary categories and universal enveloping algebras.

2.3. ***The Grothendieck monoid and grading.*** Define the relation $\equiv$ on the monoid $\operatorname{Iso}\mathscr{A}$ by

$$[X] \equiv [Y] \iff [X], [Y] \in \underline{\operatorname{Ext}}^1_{\mathscr{A}}(M, N) \text{ for some } M, N \in \mathscr{A}.$$

This relation is clearly symmetric and reflexive, hence its transitive closure is an equivalence relation on $\operatorname{Iso}\mathscr{A}$ which we also denote by $\equiv$. The additivity of $\operatorname{Ext}^1_{\mathscr{A}}(A, B) := \bigcup_X \mathcal{E}(A, B)_X / \operatorname{Aut}_{\mathscr{A}} X$ in both A and B yields the following lemma.

Lemma 2.5. *The relation $\equiv$ is a congruence relation on* $\operatorname{Iso}\mathscr{A}$*, that is, $[X] \equiv [Y]$, $[X'] \equiv [Y']$ implies that $[X \oplus X'] \equiv [Y \oplus Y']$.*

Definition 2.6. The Grothendieck monoid $\Gamma_{\mathscr{A}}$ of $\mathscr{A}$ is the quotient of $\operatorname{Iso}\mathscr{A}$ by the congruence $\equiv$.

Given an object M in $\mathscr{A}$, we denote its image in $\Gamma_{\mathscr{A}}$ by $|M|$. For all $\gamma \in \Gamma_{\mathscr{A}}$, set

$$\operatorname{Iso}\mathscr{A}_\gamma = \{[X] \in \operatorname{Iso}\mathscr{A} : |X| = \gamma\}.$$

We refer to $\operatorname{Iso}\mathscr{A}_\gamma$ as a Grothendieck class in $\mathscr{A}$, and write $\operatorname{Ind}\mathscr{A}_\gamma = \operatorname{Ind}\mathscr{A} \cap \operatorname{Iso}\mathscr{A}_\gamma$.

The following fact is obvious.

Lemma 2.7. *For any finitary exact category $\mathscr{A}$, the assignment $[M] \mapsto |M|$ defines a grading of the Hall algebra $H_{\mathscr{A}}$ of $\mathscr{A}$ by the Grothendieck monoid $\Gamma_{\mathscr{A}}$.*

Remark 2.8. After Grothendieck, one defines the Grothendieck group $K_0(\mathscr{A})$ of $\mathscr{A}$ as the universal abelian group generated by $\Gamma_{\mathscr{A}}$. Note that the canonical homomorphism of monoids $\Gamma_{\mathscr{A}} \to K_0(\mathscr{A})$ can be very far from injective. One example was already provided in the introduction. Perhaps the most extreme example is the following. Let $\mathscr{A} = \operatorname{Vect}_{\Bbbk}$ be the category of *all* $\Bbbk$-vector spaces over some field $\Bbbk$. Then $\Gamma_{\mathscr{A}}$ identifies with the monoid of cardinal numbers. In particular, if V is infinite dimensional and W is finite dimensional then $|V| = |V| + |V| = |W| + |V|$. This implies that in $K_0(\mathrm{Vect}_{\Bbbk})$, $|U| = 0$ for every object U of $\mathrm{Vect}_{\Bbbk}$, that is, $K_0(\mathrm{Vect}_{\Bbbk}) = 0$.

Also, while $K_0(\mathscr{A})$ can contain elements of finite order, this never occurs in $\Gamma_{\mathscr{A}}$. Indeed, since $[0] \in \underline{\mathrm{Ext}}^1_{\mathscr{A}}(A, B)$ implies that $A = B = 0$ and the direct sum of two nonzero objects is clearly nonzero, we immediately obtain the following lemma.

Lemma 2.9. *For any exact category $\mathscr{A}$, zero is the only invertible element of the Grothendieck monoid $\Gamma_{\mathscr{A}}$.*

2.4. ***Profinitary and cofinitary categories.*** Let Γ be an abelian monoid. Define a relation $\preceq$ on Γ by $\alpha \preceq \beta$ if $\beta = \alpha + \gamma$ for some $\gamma \in \Gamma$. This relation is clearly an additive preorder and $0 \preceq \gamma$ for any $\gamma \in \Gamma$. The following lemma is obvious.

Lemma 2.10. *The preorder $\preceq$ is a partial order on Γ if and only if the equality $\alpha + \beta + \gamma = \alpha$ for $\alpha, \beta, \gamma \in \Gamma$ implies that $\alpha = \alpha + \beta = \alpha + \gamma$. In that case, 0 is the only invertible element of Γ.*

We say that Γ is *naturally ordered* if $\preceq$ is a partial order.

Definition 2.11. We say that a Hom-finite exact category $\mathscr{A}$ is

(i) *profinitary* if $\mathrm{Iso}\,\mathscr{A}_\gamma$ is a finite set for all $\gamma \in \Gamma_{\mathscr{A}}$, and

(ii) *cofinitary* (cf. [Kapranov et al. 2012]) if for every $[X] \in \mathrm{Iso}\,\mathscr{A}$, the set

$$\{([A], [B]) \in \mathrm{Iso}\,\mathscr{A} \times \mathrm{Iso}\,\mathscr{A} : [X] \in \underline{\mathrm{Ext}}^1_{\mathscr{A}}([A], [B])\}$$

is finite.

Since $\mathcal{E}(M, N)_X$ identifies with a subset of $\mathrm{Hom}_{\mathscr{A}}(N, X) \times \mathrm{Hom}_{\mathscr{A}}(X, M)$, any profinitary category is necessarily finitary.

Proposition 2.12. *Let $\mathscr{A}$ be a profinitary category. Then $\Gamma_{\mathscr{A}}$ is naturally ordered and is generated by its minimal elements.*

A proof of this proposition is given in §5.2.

Remark 2.13. One can characterize profinitary categories as follows. If $\mathscr{A}$ is Hom-finite and its Grothendieck monoid is locally finite, as defined before Theorem 1.4, and $\mathrm{Ind}\,\mathscr{A}_\gamma$ is finite for all $\gamma \in \Gamma_{\mathscr{A}}$, then $\mathscr{A}$ is profinitary.

Theorem 2.14. *Any profinitary abelian category has the finite length property, hence is Krull–Schmidt.*

We prove this theorem in §5.3. This result, together with Theorem 1.4, yields Corollary 1.5(a).

Remark 2.15. The finite length property in an abelian category $\mathscr{A}$ is much stronger than the Krull–Schmidt property. For instance, the Grothendieck monoid of an abelian category with the finite length property is freely generated by classes of simple objects and the canonical homomorphism $\Gamma_{\mathscr{A}} \to K_0(\mathscr{A})$ is injective. On the other hand, the category of coherent sheaves on $\mathbb{P}^1$ is Krull–Schmidt, but lacks the finite length property and each Grothendieck class $\operatorname{Iso}\mathscr{A}_\gamma$, $\gamma \neq 0$ is infinite.

2.5. ***Comultiplication and primitive generation.*** Let $\mathscr{A}$ be any Hom-finite exact category. Define a linear map $\Delta : H_{\mathscr{A}} \to H_{\mathscr{A}} \widehat{\otimes} H_{\mathscr{A}}$ by

$$\Delta([C]) = \sum_{[A],[B]\in \operatorname{Iso}\mathscr{A}} F_C^{A,B} \cdot [A] \otimes [B] \,, \tag{2-3}$$

where $H_{\mathscr{A}} \widehat{\otimes} H_{\mathscr{A}}$ is the completion of the usual tensor product with to the $\Gamma_{\mathscr{A}}$-grading and $F_C^{A,B}$ is the dual Hall number given by

$$F_C^{A,B} = \frac{\#(\operatorname{Aut}_{\mathscr{A}} A \times \operatorname{Aut}_{\mathscr{A}} B)}{\# \operatorname{Aut}_{\mathscr{A}} C} \, F_{B,A}^C .$$

It follows from Riedtmann's formula [1994] that

$$F_C^{A,B} = \frac{\# \operatorname{Ext}^1_{\mathscr{A}}(B, A)_C}{\# \operatorname{Hom}_{\mathscr{A}}(B, A)},$$

where $\operatorname{Ext}^1_{\mathscr{A}}(B, A)_C = \mathcal{E}(B, A)_C / \operatorname{Aut}_{\mathscr{A}} C$. Also define a linear map $\varepsilon : H_{\mathscr{A}} \to \mathbb{Q}$ by

$$\varepsilon([C]) = \delta_{[0],[C]}. \tag{2-4}$$

The following fact is obvious.

Lemma 2.16. (a) *$H_{\mathscr{A}}$ is a topological coalgebra with respect to the above comultiplication and counit.*

(b) *If $\mathscr{A}$ is cofinitary then $H_{\mathscr{A}}$ is an ordinary coalgebra, that is, the image of the comultiplication Δ lies in $H_{\mathscr{A}} \otimes H_{\mathscr{A}}$.*

For any coalgebra C with unity denote by $\operatorname{Prim}(C)$ the set of all primitive elements, i.e.,

$$\operatorname{Prim}(C) = \{c \in C : \Delta(c) = c \otimes 1 + 1 \otimes c\}.$$

Definition 2.17. Let A be both a unital algebra and a coalgebra over a field $\mathbb{F}$. We say that A is a *quasi-Nichols algebra* if A decomposes as $\mathbb{F} \oplus V \oplus \big(\sum_{r>1} V^r\big)$ where $V = \operatorname{Prim}(A)$.

The following is the main result of the paper (Main Theorem 1.2) and is proven in §6.4.

Theorem 2.18. *Let $\mathscr{A}$ be a profinitary and cofinitary exact category. Then the Hall algebra $H_{\mathscr{A}}$ is quasi-Nichols.*

This theorem has the following useful corollary, which we prove in §6.5.

Corollary 2.19. *Let*

$$P = \ker\varepsilon \cdot \ker\varepsilon = \mathbb{Q}\{[M][N] : [M], [N] \in (\operatorname{Iso}\mathscr{A})^{+}\}, \qquad P_\gamma := P \cap (H_{\mathscr{A}})_\gamma. \tag{2-5}$$

Then $P = \sum_{k\geq 2} \operatorname{Prim}(H_{\mathscr{A}})^k = \sum_{k\geq 2} (\mathbb{Q}\operatorname{Ind}\mathscr{A})^k$ and $(H_{\mathscr{A}})_\gamma = \operatorname{Prim}(H_{\mathscr{A}})_\gamma \oplus P_\gamma$ for all $\gamma \in \Gamma^{+}_{\mathscr{A}}$.

A natural question is to compute dimensions of $\operatorname{Prim}(H_{\mathscr{A}})_\gamma$, $\gamma \in \Gamma^{+}_{\mathscr{A}}$. The following is a refinement of Proposition 1.8.

Proposition 2.20. *In the notation* (1-1) *we have*

$$m_\gamma = \dim_{\mathbb{Q}}(P_\gamma \cap \mathbb{Q}\operatorname{Ind}\mathscr{A}_\gamma)$$

for all $\gamma \in \Gamma^{+}_{\mathscr{A}}$. In particular, if $\operatorname{Ind}\mathscr{A}_\gamma \subset P_\gamma$ then $\operatorname{Prim}(H_{\mathscr{A}})_\gamma = 0$.

We prove Proposition 2.20 in §6.5, as well as the following observation, which is useful for computing primitive elements.

Lemma 2.21. *Each primitive element contains at least one isomorphism class $[X] \in \operatorname{Ind}\mathscr{A}$ in its decomposition with respect to the basis $\operatorname{Iso}\mathscr{A}$ of $H_{\mathscr{A}}$. In other words, $\operatorname{Prim}(H_{\mathscr{A}}) \cap \mathbb{Q}(\operatorname{Iso}\mathscr{A} \setminus \operatorname{Ind}\mathscr{A}) = \{0\}$.*

2.6. ***Hereditary categories and Nichols algebras.*** Let Γ be an abelian monoid and let $\mathscr{C}_\Gamma$ be the tensor category of Γ-graded vector spaces $V = \bigoplus_{\gamma\in\Gamma} V_\gamma$ over a field $\mathbb{F}$. The following fact can be easily checked.

Lemma 2.22. *For each bicharacter $\chi : \Gamma \times \Gamma \to \mathbb{F}^\times$ the category $\mathscr{C}_\Gamma$ is a braided tensor category $(\mathscr{C}_\Gamma, \Psi)$ with the braiding $\Psi_{U,V} : U \otimes V \to V \otimes U$ for objects U, V in $\mathscr{C}_\Gamma$ given by*

$$\Psi_{U,V}(u \otimes v) = \chi(\gamma, \delta)\, v \otimes u,$$

for any $u \in U_\gamma$, $v \in V_\delta$, $\gamma, \delta \in \Gamma$.

By a slight abuse of notation, given a bicharacter $\chi : \Gamma \times \Gamma \to \mathbb{F}^\times$ we denote this braided tensor category $\mathscr{C}_\Gamma$ by $\mathscr{C}_\chi$.

Now let $\mathscr{A}$ be a finitary *hereditary* category, i.e., $\operatorname{Ext}^i_{\mathscr{A}}(M, N) = 0$ for $i > 1$ and all $M, N \in \mathscr{A}$. Let $\chi_{\mathscr{A}} : \Gamma \times \Gamma \to \mathbb{Q}^\times$ be the bicharacter given by

$$\chi_{\mathscr{A}}(|M|, |N|) = \frac{\#\operatorname{Ext}^1_{\mathscr{A}}(M, N)}{\#\operatorname{Hom}_{\mathscr{A}}(M, N)}.$$

The bicharacter $\chi_{\mathscr{A}}$ is easily seen to be well-defined because it is just the (multiplicative) Euler form.

Nichols algebras were formally defined in [Andruskiewitsch and Schneider 2002].

Definition 2.23 [Andruskiewitsch and Schneider 2002, Definition 2.1]. Let $(\mathscr{C}, \Psi)$ be a braided $\mathbb{F}$-linear tensor category with a braiding Ψ. Let V be an object in $(\mathscr{C}, \Psi)$. A graded bialgebra with unity $B = \bigoplus_{n\geq 0} B_n$ in $(\mathscr{C}, \Psi)$ is called a *Nichols algebra* of V if $B_0 = \mathbb{F}$, $B_1 = V$ and B is generated, as an algebra, by $B_1 = \operatorname{Prim}(B)$.

For each object V of a braided tensor category $(\mathscr{C}, \Psi)$, the tensor algebra $T(V)$ is a graded bialgebra (even a Hopf algebra) in $(\mathscr{C}, \Psi)$ with the coproduct determined by requiring each $v \in V$ to be primitive and the grading defined by assigning degree 1 to elements of V. It is well-known [Andruskiewitsch and Schneider 2002, Proposition 2.2] that the Nichols algebra of V is unique up to an isomorphism and is the quotient of $T(V)$ by the maximal graded bi-ideal $\mathfrak{I}$ of $T(V)$ which is an object in $(\mathscr{C}, \Psi)$ and satisfies $\mathfrak{I} \cap V = \{0\}$. Henceforth we denote the Nichols algebra of V by $\mathcal{B}(V)$.

The following is proved in [Andruskiewitsch and Schneider 2002, Corollary 2.3].

Lemma 2.24. *The assignment $V \mapsto \mathcal{B}(V)$ defines a functor from $(\mathscr{C}, \Psi)$ to the category of bialgebras in $(\mathscr{C}, \Psi)$. Moreover, for any morphism $f : U \to V$ in $(\mathscr{C}, \Psi)$, the kernel of the corresponding homomorphism $\mathcal{B}(f)$ is the (bi-)ideal in $\mathcal{B}(U)$ generated by* $\ker f \subset U$.

The following fact is immediate from the definitions.

Lemma 2.25. *Let B be a bialgebra in $(\mathscr{C}, \Psi)$ which is a quasi-Nichols algebra. Then B is Nichols if and only if $\sum_{r\geq 2} (\operatorname{Prim}(B))^r$ is direct.*

The following extends the main result of [Sevenhant and Van Den Bergh 2001].

Theorem 2.26. *For any profinitary hereditary abelian category $\mathscr{A}$, the Hall algebra $H_{\mathscr{A}}$ is isomorphic to the Nichols algebra $\mathcal{B}(V_{\mathscr{A}})$ in the category $\mathscr{C}_{\chi_{\mathscr{A}}}$, where $V_{\mathscr{A}} =$* $\operatorname{Prim}(H_{\mathscr{A}})$.

We prove this theorem in §7.2.

Remark 2.27. In fact, the original result of [Sevenhant and Van Den Bergh 2001, Theorem 1.1] follows from Theorem 2.26. The classification of diagonally braided Nichols algebras was obtained in [Andruskiewitsch and Schneider 2002, §5] and, in particular, generalizes some results of [Sevenhant and Van Den Bergh 2001].

3. Examples

In this section we construct primitive elements in several Hall algebras and provide supporting evidence for Conjectures 1.13 and 1.10. Throughout this section we write $\overline{\Delta}(x) = \Delta(x) - x \otimes 1 - 1 \otimes x$ (thus, x is primitive if and only if $x \in \ker \overline{\Delta}$). Needless to say, every (almost) simple object S satisfies $\overline{\Delta}([S]) = 0$ so we focus only on nonsimple primitive elements. In this section, $\Bbbk$ always denotes a finite field with q elements and all categories are assumed to be $\Bbbk$-linear.

3.1. ***Classical Hall–Steinitz algebra.*** Let R be a principal ideal domain such that $R/\mathfrak{m}$ is a finite field for any maximal ideal $\mathfrak{m}$ of R. Let $\mathscr{A} = \mathscr{A}(\mathfrak{m})$ be the full subcategory of finite length R-modules M satisfying $\mathfrak{m}^r M = 0$ for some $r \geq 0$. Then for each $r > 0$, there exists a unique, up to an isomorphism, indecomposable object $\mathcal{I}_r = R/\mathfrak{m}^r \in \mathscr{A}$. More generally, given a partition $\lambda = (\lambda_1 \geq \cdots \geq \lambda_k > 0)$, set $\mathcal{I}_\lambda = \mathcal{I}_{\lambda_1} \oplus \cdots \oplus \mathcal{I}_{\lambda_k}$ and write $\ell(\lambda) = k$.

Since the Euler form of $\mathscr{A}$ is identically zero and $\mathscr{A}$ is hereditary, $H_{\mathscr{A}}$ is an ordinary Hopf algebra (the braiding is trivial). The Grothendieck monoid of $\mathscr{A}$ being $\mathbb{Z}_{\geq 0}$, the algebra $H_{\mathscr{A}}$ is $\mathbb{Z}_{\geq 0}$-graded. We now provide a new (very short) proof of the following classical result.

Theorem 3.1 [Macdonald 1979; Zelevinsky 1981]. *The Hall algebra $H_{\mathscr{A}}$ is commutative and cocommutative and is freely generated by the $[\mathcal{I}_n]$, $n > 0$. Moreover, $H_{\mathscr{A}}$ is freely generated by its primitive elements $\mathcal{P}_n$, $n > 0$.*

Proof. It is easy to see, using duality, that $H_{\mathscr{A}}$ is commutative, hence cocommutative. Let $\mathscr{P}$ be the set of all partitions. Given a partition $\lambda = (\lambda_1 \geq \cdots \geq \lambda_r > 0) \in \mathscr{P}$, let $M_\lambda = [\mathcal{I}_{\lambda_1}] \cdots [\mathcal{I}_{\lambda_r}]$. By Theorem 2.4, the set $\{M_\lambda\}_{\lambda \in \mathscr{P}}$ is a basis of $H_{\mathscr{A}}$, hence $H_{\mathscr{A}}$ is freely generated by the isomorphism classes of indecomposables $[\mathcal{I}_n]$, $n > 0$. Since $H_{\mathscr{A}}$ is commutative, $P = \ker\varepsilon \cdot \ker\varepsilon$ is spanned by the M_λ with $\ell(\lambda) \geq 2$, hence $\mathbb{Q}\operatorname{Ind}\mathscr{A} \cap P = \{0\}$ and by Proposition 2.20, $\dim_{\mathbb{Q}} \operatorname{Prim}(H_{\mathscr{A}})_n = \#\operatorname{Ind}\mathscr{A}_n = 1$ for all $n > 0$. Thus, for each $n > 0$ we have a unique, up to a scalar, nonzero primitive element $\mathcal{P}_n$ in $(H_{\mathscr{A}})_n$. The dimension considerations and Theorem 2.18 immediately imply that $H_{\mathscr{A}}$ is freely generated by the $\mathcal{P}_n$, $n > 0$. □

This theorem has the following nice corollary.

Corollary 3.2. *For all $n > 0$, let $x_n \in (H_{\mathscr{A}})_n \setminus \mathbb{Q}(\operatorname{Iso}\mathscr{A}_n \setminus \operatorname{Ind}\mathscr{A}_n)$. Then $\{x_n\}_{n>0}$ freely generates $H_{\mathscr{A}}$. In particular, $E_{\mathscr{A}} = H_{\mathscr{A}}$.*

The elements $\mathcal{P}_n$ can be computed explicitly (see, e.g., [Hubery 2005, §5]), namely

$$\mathcal{P}_n = \sum_{\lambda \vdash n} \Bigg(\prod_{j=1}^{\ell(\lambda)-1} (1 - q^j) \Bigg) [\mathcal{I}_\lambda],$$

where $q = |R/\mathfrak{m}|$.

Under the isomorphism $\psi : H_{\mathscr{A}} \to \operatorname{Sym}$, $[\mathcal{I}_\lambda] \mapsto q^{-n(\lambda)} P_\lambda(x; q^{-1})$ [Macdonald 1979; Zelevinsky 1981], where Sym is the algebra of symmetric polynomials in infinitely many variables and $P_\lambda(x; t)$ is the Hall–Littlewood polynomial, the image of $\mathcal{P}_n$ is the n-th power sum p_n. As shown in [Zelevinsky 1981], the p_n are primitive elements in Sym with the comultiplication defined by

$$\Delta(e_n) = \sum_{i=0}^{n} e_i \otimes e_{n-i},$$

where e_r is the r-th elementary symmetric polynomial, which equals $q^{-\binom{r}{2}} \psi([\mathcal{I}_{(1^r)}])$. Note also that $\psi\big(\sum_{\lambda\vdash n}[\mathcal{I}_\lambda]\big)$ is the n-th complete symmetric function h_n.

Since $C_{\mathscr{A}} = \mathbb{Q}[\mathcal{P}_1]$, we have $C_{\mathscr{A}} \subsetneq H_{\mathscr{A}}$. Since $\dim_{\mathbb{Q}} \operatorname{Prim}(H_{\mathscr{A}})_n = 1$ for all $n > 0$, it follows that $U_{\mathscr{A}} = H_{\mathscr{A}}$. Thus, $C_{\mathscr{A}} \subsetneq E_{\mathscr{A}} = U_{\mathscr{A}} = H_{\mathscr{A}}$.

3.2. *Homogeneous tubes.* Let $\mathscr{A}$ be the category of finite dimensional $\Bbbk$-representations of a tame acyclic quiver Q. Then $\mathscr{A}$ decomposes into a triple of subcategories of preprojective, preinjective and regular representations (see [Auslander et al. 1995, Chapter VIII]) which we denote, as in [Berenstein and Greenstein 2013, §5], by $\mathscr{A}_-$, $\mathscr{A}_+$ and $\mathscr{A}_0$, respectively. The category $\mathscr{A}_0$ can be further decomposed into the so-called stable tubes, that is, components of the Auslander–Reiten quiver of $\mathscr{A}$ on which the Auslander translation acts as an autoequivalence of finite order, called the *rank* of the tube. It is well-known that rank 1, or homogeneous, tubes are parametrized by the set $\Bbbk\mathbb{P}^1$ of homogeneous prime ideals in $\Bbbk[x, y]$. Given a homogeneous prime ideal ρ, let $\deg \rho$ be the degree of a generator of that ideal and denote by $\mathscr{T}_\rho$ the corresponding rank 1 tube. Then $\mathscr{T}_\rho$ is equivalent to the category of nilpotent representations of $\mathbb{K}[x]$ where $[\mathbb{K} : \Bbbk] = \deg \rho$ and its Hall algebra is isomorphic to the classical Hall–Steinitz algebra. Thus, for each $r > 0$, $\mathscr{T}_\rho$ contains a unique indecomposable $\mathcal{I}_r(\rho)$ of length r. Given a partition $\lambda = (\lambda_1 \geq \cdots \geq \lambda_k > 0)$, let $\mathcal{I}_\lambda(\rho) = \mathcal{I}_{\lambda_1}(\rho) \oplus \cdots \oplus \mathcal{I}_{\lambda_k}(\rho)$. By §3.1 the elements

$$\mathcal{P}_n(\rho) = \sum_{\lambda\vdash n} \bigg(\prod_{j=1}^{\ell(\lambda)-1} (1 - q^{j \deg \rho}) \bigg) [\mathcal{I}_\lambda(\rho)]$$

are primitive in $H_{\mathscr{T}_\rho}$. Let $\mathscr{A}_{0,h}$ be the full subcategory of homogeneous objects in $\mathscr{A}_0$ (cf. [Dlab and Ringel 1976, Theorem 3.5]). Since $H_{\mathscr{A}_{0,h}}$ is isomorphic to the tensor product of the $H_{\mathscr{T}_\rho}$ as a bialgebra, this gives all primitive elements in $H_{\mathscr{A}_{0,h}}$. The Grothendieck monoid of $\mathscr{A}_{0,h}$ equals the direct sum of infinitely many copies (indexed by $\rho \in \Bbbk\mathbb{P}^1$) of $\mathbb{Z}_{\geq 0}$.

However, the elements $\mathcal{P}_n(\rho)$ are not primitive in $H_{\mathscr{A}}$ since an object in $\mathscr{A}_{0,h}$ can have preprojective subobjects and preinjective quotients. They can be used to construct primitive elements in $H_{\mathscr{A}}$.

Conjecture 3.3. *The elements*

$$\mathcal{P}_n(\rho) - \frac{1}{N(\deg \rho)} \sum_{\rho' \in \Bbbk\mathbb{P}^1 : \deg \rho' = \deg \rho} \mathcal{P}_n(\rho'),$$

are primitive in $H_{\mathscr{A}}$, *where* $N(d)$ *is the number of elements of* $\Bbbk\mathbb{P}^1$ *of degree* d (*that is*, $N(1) = |\Bbbk| + 1$ *while* $N(d)$, $d > 1$, *is the number of irreducible monic polynomials of degree* d *in one variable*).

This formula can be easily checked in small cases (see, for example, §3.8) or for the Kronecker quiver, using the results of [Szántó 2006]. Since $F_M^{I,P} = 0$ for all $P \in \mathscr{A}_-$ and $I \in \mathscr{A}_+$, the above conjecture is an immediate consequence of the next conjecture.

Conjecture 3.4.[1] *Let* $I \in \mathscr{A}_+$ *and* $P \in \mathscr{A}_-$. *Then for any partition* λ *we have* $F^{P,I}_{\mathcal{I}_\lambda(\rho)} = F^{P,I}_{\mathcal{I}_\lambda(\rho')}$ *where* $\rho, \rho' \in \Bbbk\mathbb{P}^1$ *with* $\deg\rho = \deg\rho'$.

This is known to hold in some special cases (see for example [Szántó 2006; Hubery 2004]).

In the category $\mathscr{A}$, we have $C_{\mathscr{A}} = E_{\mathscr{A}} = U_{\mathscr{A}} \subsetneq H_{\mathscr{A}}$. On the other hand, for $\mathscr{A}_0$ we have $C_{\mathscr{A}_0} \subsetneq E_{\mathscr{A}_0} = U_{\mathscr{A}_0} = H_{\mathscr{A}_0}$ and similarly for each homogeneous tube.

3.3. ***A tame valued quiver.*** Consider now the valued quiver $1 \xrightarrow{(4,1)} 2$. Let $\Bbbk_2$ be a field extension of $\Bbbk_1 = \Bbbk$ of degree 4. Note that $\Bbbk_2$ contains precisely $q^4 - q^2$ elements of degree 4 over $\Bbbk$ and $q^2 - q$ elements of degree 2. A representation of this quiver is a triple (V_1, V_2, f) where V_i is a $\Bbbk_i$-vector space and $f \in \operatorname{Hom}_{\Bbbk}(V_1, V_2)$. Finally, a morphism $(V_1, V_2, f) \to (W_1, W_2, g)$ is a pair (φ_1, φ_2) where $\varphi_i \in \operatorname{Hom}_{\Bbbk_i}(V_i, W_i)$ and $g \circ \varphi_1 = \varphi_2 \circ g$.

The smallest indecomposable regular representation is $(\Bbbk_1^2, \Bbbk_2, f)$, where f is injective. Thus, f is given by a pair $(\lambda, \mu) \in \Bbbk_2 \times \Bbbk_2$ which is linearly independent over $\Bbbk$ (this pair is the image under f of the standard basis of $\Bbbk_1^2$). It is easy to see that, up to an isomorphism, such a pair can be assumed to be of the form $(\lambda, 1)$ where $\lambda \in \Bbbk_2 \setminus \Bbbk_1$. Denote the resulting representation by $E_1(\lambda)$. A morphism $f : E_1(\lambda) \to E_1(\lambda')$ is uniquely determined by a matrix $\varphi_1 = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M_2(\Bbbk)$ and $\varphi_2 \in \Bbbk_2$ and we have

$$(b\lambda' + d)\lambda = a\lambda' + c.$$

If λ has degree 4 over $\Bbbk$ then $\operatorname{End}_{\mathscr{A}} E_1(\lambda) \cong \Bbbk$ and $\operatorname{Aut}_{\mathscr{A}} E_1(\lambda) \cong \Bbbk^\times$. Otherwise, $\operatorname{End}_{\mathscr{A}} E_1(\lambda) \cong L$ and $\operatorname{Aut}_{\mathscr{A}} E_1(\lambda) \cong L^\times$ where $[L : \Bbbk] = 2$. It follows that all $E_1(\lambda)$ with $\deg_{\Bbbk} \lambda = 2$ are isomorphic, since the stabilizer of such a λ in $\mathrm{GL}(2, \Bbbk)$ has index $q^2 - q$, and that there are q nonisomorphic representations $E_1(\lambda)$ with $\deg_{\Bbbk} \lambda = 4$. It is easy to see that for any $\lambda \in \Bbbk_2 \setminus \Bbbk_1$ we have $(q^2-1)(q-1)$ short exact sequences

$$0 \to P_1 \to E_1(\lambda) \to S_1 \to 0$$

[1] After the present paper was accepted for publication, we were informed that a proof of Conjecture 3.4 was announced in [Deng and Ruan 2015].

and $q(q^4-1)(q^2-1)(q-1)$ short exact sequences

$$0 \to S_2 \to E_1(\lambda) \to S_1^{\oplus 2} \to 0.$$

As a result, we conclude that

$$\bar{\Delta}(E_1(\lambda)) = \frac{(q^2-1)(q-1)}{|\operatorname{Aut}_{\mathscr{A}} E_1(\lambda)|}\big([P_1]\otimes[S_1] + q(q^4-1)[S_2]\otimes[S_1^{\oplus 2}]\big),$$

hence

$$\mathcal{P}_1(\lambda) := E_1(\lambda) - \frac{1}{(q+1)|\operatorname{Aut}_{\mathscr{A}} E_1(\lambda)|}\sum_{\mu\in(\Bbbk_2\setminus\Bbbk_1)/\operatorname{GL}(2,\Bbbk)} |\operatorname{Aut}_{\mathscr{A}} E_1(\mu)| E_1(\mu)$$

is primitive, and these are all primitive elements of degree $2\alpha_1+\alpha_2$ in $H_{\mathscr{A}}$. There is precisely one linear relation among them, namely

$$\sum_{\lambda\in(\Bbbk_2\setminus\Bbbk_1)/\operatorname{GL}(2,\Bbbk)} |\operatorname{Aut}_{\mathscr{A}} E_1(\lambda)|\mathcal{P}_1(\lambda) = 0.$$

In this case, like in §3.2, $C_{\mathscr{A}} = E_{\mathscr{A}} = U_{\mathscr{A}} \subsetneq H_{\mathscr{A}}$ which supports Conjecture 1.13. Also, $\dim_{\mathbb{Q}} \operatorname{Prim}(H_{\mathscr{A}})_{2\alpha_1+\alpha_2} = q$ and $m_{2\alpha_1+\alpha_2} = 1$.

3.4. ***Hereditary categories defined by submonoids.*** The next two examples are special cases of the following construction. Consider a submonoid Γ_0 of the Grothendieck monoid Γ of an abelian category $\mathscr{A}$, and define a full subcategory $\mathscr{A}(\Gamma_0)$ of $\mathscr{A}$ whose objects X satisfy $|X| \in \Gamma_0$. By construction, $\mathscr{A}(\Gamma_0)$ is closed under extensions and hence is exact.

First, let $\mathscr{A}$ be the category of $\Bbbk$-representations of the quiver $1 \to 2$. Then $\Gamma_{\mathscr{A}}$ is freely generated by $\alpha_i = |S_i|$ where the S_i, $i = 1, 2$ are simple objects. Fix $r > 0$. Let $\Gamma_r = \mathbb{Z}_{\geq 0}(\alpha_1 + r\alpha_2)$ and set $\mathscr{B}_r = \mathscr{A}(\Gamma_r)$. Let $P_1 = I_2$ be the projective cover of S_1 and the injective envelope of S_2 in $\mathscr{A}$. Then in $H_{\mathscr{A}}$ we have

(3-1) $[S_1][S_2]=[S_2][S_1]+[P_1], \quad [S_1][P_1]=q[P_1][S_1], \quad [P_1][S_2]=q[S_2][P_1].$

Every object in $\mathscr{B}_r$ is isomorphic to $S_1^{\oplus a}\oplus P_1^{\oplus b}\oplus S_2^{\oplus(ra+(r-1)b)}$, $a, b \geq 0$. The only simple objects in $\mathscr{B}_r$, up to an isomorphism, are $X_1 = S_1\oplus S_2^{\oplus r}$ and $X_2 = S_2^{\oplus r-1}\oplus P_1$. Then $[X_1]$ is a nonzero multiple of $E_1 = [S_2]^r[S_1]$, and $[X_2]$ of $E_2 = [S_2]^{r-1}[P_1]$. In particular, the E_i are primitive elements of $H_{\mathscr{B}_r}$. Using (3-1) we can show that E_1 and E_2 satisfy the relation

$$E_2E_1 = q^{r-1}E_1E_2 - [r-1]_q E_2^2,$$

where $[s]_q = 1 + \cdots + q^{s-1}$. The Grothendieck monoid of $\mathscr{B}_r$ is generated by $\beta_i = |X_i|$, $i = 1, 2$, subject to the relation $\beta_1 + \beta_2 = 2\beta_1 = 2\beta_2$ (thus $\Gamma_{\mathscr{B}_r}$ does not coincide with Γ_r and is not even a submonoid of $\Gamma_{\mathscr{A}}$). It is not hard to check that E_1 and E_2 generate $H_{\mathscr{B}_r}$ and hence form a basis of $\operatorname{Prim}(H_{\mathscr{B}_r})$.

In this case we have $C_{\mathscr{B}_r} = U_{\mathscr{B}_r} = E_{\mathscr{B}_r} = H_{\mathscr{B}_r}$ and so Conjecture 1.13 holds.

A more complicated example is obtained as follows. Let $\mathscr{A}$ be the category of $\Bbbk$-representations of the quiver $1 \to 0 \leftarrow 2$. As in the previous example, $\Gamma_{\mathscr{A}}$ is freely generated by $\alpha_i = |S_i|$, $0 \le i \le 2$. Let $\Gamma_\circ = \{s\alpha_0 + r\alpha_1 + r\alpha_2 : r, s \in \mathbb{Z}_{\ge 0}\}$ and let $\mathscr{B} = \mathscr{A}(\Gamma_\circ)$. Let P_i be the projective cover of S_i in $\mathscr{A}$ and I_i be its injective envelope. Thus, $I_1 = S_1$, $I_2 = S_2$, $|I_0| = \alpha_0 + \alpha_1 + \alpha_2$, $|P_1| = \alpha_0 + \alpha_1$, $P_0 = S_0$ and $|P_2| = \alpha_2 + \alpha_0$. The simple objects in $\mathscr{B}$ are $S_1 \oplus S_2$ and S_0, while the nonsimple indecomposable objects are

$$P_1 \oplus S_2, \quad P_2 \oplus S_1, \quad P_1 \oplus P_2, \quad I_0.$$

The Grothendieck monoid of $\mathscr{B}$ is freely generated by $\beta_1 = |S_1 \oplus S_2|$ and $\beta_0 = |S_0|$. Clearly, $Y_1 = [S_1 \oplus S_2]$ and $Y_0 = [S_0]$ are primitive in $H_{\mathscr{B}}$. We also have two linearly independent primitive elements of degree $\beta_1 + \beta_0$, say

$$\begin{aligned} Z_1 &= [I_0] - (q-1)[P_1 \oplus S_2], \\ Z_2 &= [I_0] - (q-1)[P_2 \oplus S_1]. \end{aligned}$$

Then

$$[Z_1, Z_2] = 0, \quad [Y_1, Z_1]_q = [Y_1, Z_2]_q = 0, \quad [Z_1, Y_0]_q = [Z_2, Y_0]_q = 0,$$

and

$$[Y_1, [Y_1, Y_0]]_{q^2} = Y_1(Z_1 + Z_2), \quad [[Y_1, Y_0], Y_0]_q, Y_0]_{q^2} = 0,$$

where $[a, b]_t = ab - tba$. Here $C_{\mathscr{B}} = E_{\mathscr{B}} = U_{\mathscr{B}} \subsetneq H_{\mathscr{B}}$ which again supports Conjecture 1.13. Also, we have a unique imaginary simple root $\beta_1 + \beta_0$, and $\dim_{\mathbb{Q}} \operatorname{Prim}(H_{\mathscr{B}})_{\beta_1+\beta_0} = 2$ while $m_{\beta_1+\beta_0} = 1$.

3.5. ***Sheaves on projective curves.*** Consider the category $\mathscr{A}$ of coherent sheaves on $\mathbb{P}^1(\Bbbk)$ (cf. [Burban and Schiffmann 2012; Kapranov 1997; Baumann and Kassel 2001]). Following [Baumann and Kassel 2001], $\mathscr{A}$ is equivalent to the category with objects (M', M'', ϕ) where M' is a $\Bbbk[z]$-module, M'' is a $\Bbbk[z^{-1}]$-module and ϕ is an isomorphism of $\Bbbk[z, z^{-1}]$-modules $M'_z \to M''_{z^{-1}}$. In particular, for any $n \in \mathbb{Z}$, we have an indecomposable object $\mathcal{O}(n) = (\Bbbk[z], \Bbbk[z^{-1}], \phi_n)$ where $\phi_n \in \operatorname{Aut} \Bbbk[z, z^{-1}]$ is multiplication by z^{-n}. We have (cf. [Baumann and Kassel 2001])

$$\dim_{\Bbbk} \operatorname{Hom}_{\mathscr{A}}(\mathcal{O}(m), \mathcal{O}(n)) = \max(0, n - m + 1)$$

and any nonzero morphism $\mathcal{O}(m) \to \mathcal{O}(n)$ is injective.

Consider now the full subcategory $\mathscr{A}_{\mathrm{lc}}$ of locally free coherent sheaves on $\mathbb{P}^1$. Any object in $\mathscr{A}_{\mathrm{lc}}$ is isomorphic to a direct sum of objects of the form $\mathcal{O}(m)$ and these are precisely the indecomposables in $\mathscr{A}_{\mathrm{lc}}$. The Grothendieck monoid of $\mathscr{A}_{\mathrm{lc}}$ identifies with $\{(0, 0)\} \cup \mathbb{Z}_{\ge 0} \times \mathbb{Z}$ with $|\mathcal{O}(n)| = (1, n)$. Note that $\mathscr{A}_{\mathrm{lc}}$ has no simple objects. The category $\mathscr{A}_{\mathrm{lc}}$ is closed under extensions and hence is exact. Since

$\mathcal{A}_{\mathrm{lc}}$ is Krull–Schmidt, its Hall algebra has a basis consisting of ordered monomials on $X_m := [\mathcal{O}(m)]$ for *any* total order on $\mathbb{Z}$. Since $m < n$ implies that $\mathcal{O}(n)/\mathcal{O}(m)$ is not an object in $\mathscr{A}_{\mathrm{lc}}$, it follows that $\mathcal{O}(m)$ is almost simple, hence X_m is primitive for all $m \in \mathbb{Z}$. Thus, $H_{\mathscr{A}_{\mathrm{lc}}}$ is primitively generated. By [Baumann and Kassel 2001, Theorem 10(iii)] the defining relations in $H_{\mathscr{A}_{\mathrm{lc}}}$ are

$$X_n X_m = q^{n-m+1} X_m X_n + (q^2 - 1) q^{n-m-1} \sum_{a=1}^{\lfloor (n-m)/2 \rfloor} X_{m+a} X_{n-a}, \quad m < n.$$

However, Theorem 2.18 does not apply to the Hall algebra of $\mathscr{A}$ or $\mathscr{A}_{\mathrm{lc}}$ since the categories $\mathscr{A}$ or even $\mathscr{A}_{\mathrm{lc}}$ are neither profinitary nor cofinitary. For example, every object $\mathcal{O}(m) \oplus \mathcal{O}(n)$, $m > n$ appears as the middle term of a short exact sequence

$$0 \to \mathcal{O}(n-a) \to \mathcal{O}(m) \oplus \mathcal{O}(n) \to \mathcal{O}(m+a) \to 0$$

for all $a \geq 0$.

On the other hand, the Hall algebra of the subcategory of torsion sheaves is isomorphic to the Hall algebra of the regular subcategory for the valued quiver $1 \xrightarrow{(2,2)} 2$, or, equivalently, the Kronecker quiver.

It should be noted that the Hall algebra of the subcategory of preprojective modules $\mathscr{B}_+$ in the category $\mathscr{B}$ of $\Bbbk$-representations of the Kronecker quiver is isomorphic to the subalgebra of $H_{\mathscr{A}_{\mathrm{lc}}}$ generated by the X_m for $m > 0$. Indeed, $\Gamma_{\mathscr{B}_+} \cong \mathbb{Z}_{\geq 0}$, and for each $k > 0$ there is a unique preprojective indecomposable Q_k with $|Q_k| = k$. It is easy to see, by grading considerations, that Q_k is primitive. Then the $[Q_k]$, $k \geq 0$ can be shown to satisfy exactly the same relations as the X_n (see [Szántó 2006, Theorem 4.2]). In this case we have

$$C_{\mathscr{B}_+} \subsetneq U_{\mathscr{B}_+} = E_{\mathscr{B}_+} = H_{\mathscr{B}_+}.$$

This situation can be generalized as follows. Let X be a smooth projective curve and let $\mathscr{A}$ be the category of coherent sheaves on X. Let $\mathscr{A}_{\mathrm{lc}}^{\geq d}$ be the full subcategory of $\mathscr{A}$ whose objects are locally free sheaves of positive rank and of degree $\geq d$. Since the rank and the degree are additive on short exact sequences, this subcategory is closed under extensions. Since for a coherent sheaf $\mathcal{F}$ the possible degrees of its subsheaves of rank r are bounded above (cf. [Kapranov et al. 2012, Proposition 2.5]), for any fixed pair (r, d) there are finitely many subsheaves of $\mathcal{F}$ of rank r and degree d. We conclude that the category $\mathscr{A}_{\mathrm{lc}}^{\geq d}$ is cofinitary and profinitary, hence Theorem 2.18 applies and the Hall algebra of $\mathscr{A}_{\mathrm{lc}}^{\geq d}$ is generated by its primitive elements. Results on primitive elements in this algebra can be found in [Kapranov et al. 2012, §3.2]. Note that $\mathscr{A}_{\mathrm{lc}}$ is Krull–Schmidt, hence its Hall algebra is PBW on indecomposables.

3.6. ***Nonhereditary categories of finite type.*** Let $\mathscr{A}$ be the category of $\Bbbk$-representations of the quiver

$$
\begin{tikzcd}
 & 1 \arrow[dl, "a_{12}"'] \arrow[dr, "a_{13}"] & \\
2 \arrow[dr, "a_{24}"'] & & 3 \arrow[dl, "a_{34}"] \\
 & 4 &
\end{tikzcd}
$$

satisfying the relation $a_{24}a_{12} = 0$. This category has 14 isomorphism classes of indecomposable objects, 12 of them having different images in $\Gamma_{\mathscr{A}}$ and the two remaining ones, namely the projective cover P_1 of S_1 and the injective envelope I_4 of S_4, having the same image $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$ (as before, $\alpha_i = |S_i|$).

Let S_{ij} and S_{ijk} be the unique, up to an isomorphism, indecomposables with $|X| = \alpha_i + \alpha_j$ and $|X| = \alpha_i + \alpha_j + \alpha_k$, respectively. Then $[S_{ij}], [S_{ijk}] \in P$ follows easily, hence $\operatorname{Prim}(H_{\mathscr{A}})_{\alpha_i+\alpha_j} = 0 = \operatorname{Prim}(H_{\mathscr{A}})_{\alpha_i+\alpha_j+\alpha_k}$ by Proposition 2.20. Let us show that $\operatorname{Prim}(H_{\mathscr{A}})_{\alpha_1+\alpha_2+\alpha_3+\alpha_4} = 0$; then the only primitive elements are those in $\operatorname{Prim}(H_{\mathscr{A}})_{\alpha_i}$, $1 \le i \le 4$.

For every object M with $|M| = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$, except P_1, I_4 and $S_2 \oplus S_{134}$, there exists a pair of objects A, B such that $F_N^{A,B} = 0$ unless $[N] = [M]$. This implies that $\operatorname{Prim}(H_{\mathscr{A}})_{\alpha_1+\alpha_2+\alpha_3+\alpha_4}$ is contained in the linear span of $[P_1]$, $[I_4]$ and $[S_2 \oplus S_{134}]$. We have (with $h = |\Bbbk^\times| = q - 1$)

$$
\begin{aligned}
\bar{\Delta}([S_2 \oplus S_{134}]) &= [S_{134}] \otimes [S_2] + [S_2] \otimes [S_{134}] \\
&\qquad + h\big([S_2 \oplus S_{34}] \otimes [S_1] + [S_2 \oplus S_4] \otimes [S_{13}] \\
&\qquad\qquad + [S_{34}] \otimes [S_1 \oplus S_2] + [S_4] \otimes [S_2 \oplus S_{13}]\big), \\
\bar{\Delta}([I_4]) &= h\big([S_{134}] \otimes [S_2] + [S_{234}] \otimes [S_1] + [S_{24}] \otimes [S_{13}]\big) \\
&\qquad + h^2\big([S_{34}] \otimes [S_1 \oplus S_2] + [S_4] \otimes [S_2 \oplus S_{13}]\big), \\
\bar{\Delta}([P_1]) &= h\big([S_{34}] \otimes [S_{12}] + [S_2] \otimes [S_{134}] + [S_4] \otimes [S_{123}]\big) \\
&\qquad + h^2\big([S_2 \oplus S_{34}] \otimes [S_1] + [S_2 \oplus S_4] \otimes [S_{13}]\big).
\end{aligned}
$$

It is now clear that $\operatorname{Prim}(H_{\mathscr{A}})_{\alpha_1+\alpha_2+\alpha_3+\alpha_4} = 0$.

Let $E_i = [S_i]$, $1 \le i \le 4$. To write a presentation of $H_{\mathscr{A}}$, it is useful to introduce $Z = [P_1] + [I_4] - (q-1)[S_2 \oplus S_{134}]$. We obtain

$$
\begin{aligned}
&[E_i, [E_i, E_j]]_q = 0 = [[E_i, E_j], E_j]_q \\
&\qquad\qquad \text{for } (i, j) \in \{(1,2), (1,3), (2,4), (3,4)\}, \\
&[E_2, E_3] = 0 = [E_1, E_4],
\end{aligned}
\tag{3-2}
$$

and also

$$[E_4, [E_1, E_2]] = 0, \quad [E_1, Z]_q = 0 = [Z, E_4]_q, \quad [E_2, Z] = 0 = [E_3, Z],$$

where

$$Z = [E_1, [E_2, [E_3, E_4]]_q] - [E_4, [E_3, [E_2, E_1]]_q].$$

If we consider the category of representations of the same quiver satisfying the relation $a_{24}a_{12} = a_{34}a_{13}$, its Hall algebra's subspace of primitive elements is spanned by the E_i, $1 \le i \le 4$ which satisfy (3-2), as well as

$$\begin{aligned} [E_4, [E_1, E_2]] &= 0 = [E_4, [E_1, E_3]] \\ [E_1, [E_2, [E_3, E_4]]] &= [E_4, [E_3, [E_1, E_2]]] \\ &= [E_4, [E_2, [E_1, E_3]]] = [E_1, [E_3, [E_2, E_4]]]. \end{aligned}$$

In both cases $C_{\mathscr{A}} = E_{\mathscr{A}} = U_{\mathscr{A}} = H_{\mathscr{A}}$.

3.7. ***Special pairs of objects and primitive elements.*** Before we consider the next group of examples, we make the following observation. Suppose that we have a pair of indecomposable objects $X \not\cong Y$ in $\mathscr{A}$ satisfying $\operatorname{Hom}_{\mathscr{A}}(X, Y) = 0 = \operatorname{Hom}_{\mathscr{A}}(Y, X)$, $\operatorname{End}_{\mathscr{A}} X \cong \operatorname{End}_{\mathscr{A}} Y \cong \Bbbk$ is a field and

$$\dim_{\Bbbk} \operatorname{Ext}^1_{\mathscr{A}}(X, Y) = \dim_{\Bbbk} \operatorname{Ext}^1_{\mathscr{A}}(Y, X) = 1.$$

Then there exist unique $[Z_{YX}], [Z_{XY}] \in \operatorname{Iso} \mathscr{A}$ such that

$$\underline{\operatorname{Ext}}^1_{\mathscr{A}}(X, Y) = \{[X \oplus Y], [Z_{XY}]\}, \qquad \underline{\operatorname{Ext}}^1_{\mathscr{A}}(Y, X) = \{[X \oplus Y], [Z_{YX}]\}.$$

Let $\mathscr{B} = \mathscr{A}(X, Y)$ be the minimal additive full subcategory of $\mathscr{A}$ containing X and Y and closed under extensions. Then in $H_{\mathscr{B}}$ we have

$$\begin{aligned} \bar{\Delta}([Z_{YX}]) &= (q-1)[X] \otimes [Y], \\ \bar{\Delta}([Z_{XY}]) &= (q-1)[Y] \otimes [X], \\ \bar{\Delta}([X \oplus Y]) &= [X] \otimes [Y] + [Y] \otimes [X], \end{aligned}$$

and so

$$[Z_{XY}] + [Z_{YX}] - (q-1)[X \oplus Y]$$

is primitive in $H_{\mathscr{B}}$. Indeed, $|\operatorname{Ext}^1_{\mathscr{A}}(Y, X)_{Z_{YX}}| = q - 1 = |\operatorname{Ext}^1_{\mathscr{A}}(X, Y)_{Z_{XY}}|$ and so by Riedtmann's formula,

$$F^{X,Y}_{Z_{YX}} = q - 1 = F^{Y,X}_{Z_{XY}}, \quad F^{X,Y}_{X\oplus Y} = F^{Y,X}_{X\oplus Y} = 1.$$

This element need not be primitive in $H_{\mathscr{A}}$ but is often useful for computations.

3.8. ***A rank 2 tube.*** Let $\mathscr{A} = \mathrm{rep}_{\Bbbk}(Q)$ where Q is a valued acyclic quiver of tame type. Let τ be the Auslander–Reiten translation and consider a regular component of the Auslander–Reiten quiver which is a tube of rank 2 (that is, for every indecomposable object M in that component we have $\tau^2(M) \cong M$). The smallest example is provided by the quiver

$$\begin{tikzcd} & 2 \arrow[dr] & \\ 1 \arrow[ur] \arrow[rr] & & 3 \end{tikzcd}$$

and the Auslander–Reiten component containing S_2.

Let X be a simple object in our tube. Then $\tau(X)$ is also simple and both satisfy $\mathrm{End}_{\mathscr{A}} X \cong \mathrm{End}_{\mathscr{A}} \tau(X) \cong \Bbbk$. Furthermore,

$$\mathrm{Ext}^1_{\mathscr{A}}(X, \tau(X)) \cong \mathrm{Hom}_{\mathscr{A}}(\tau(X), \tau(X)), \quad \mathrm{Ext}^1_{\mathscr{A}}(\tau(X), X) \cong \mathrm{Hom}_{\mathscr{A}}(X, X),$$

and so $X, \tau(X)$ satisfy the assumptions of §3.7. Thus, we obtain a primitive element of degree $|X| + |\tau(X)|$ in the Hall algebra of our tube given by

$$Z_{X,\tau(X)} + Z_{\tau(X),X} - (q-1)[X \oplus Y].$$

For the quiver shown above, with $X = S_2$ we have

$$Y = \tau(X) = \begin{tikzcd} & 0 \arrow[dr, "0"] & \\ \Bbbk \arrow[ur, "0"] \arrow[rr, "1"] & & \Bbbk \end{tikzcd}$$

while

$$Z_{YX} = \begin{tikzcd} & \Bbbk \arrow[dr, "0"] & \\ \Bbbk \arrow[ur, "1"] \arrow[rr, "1"] & & \Bbbk \end{tikzcd}, \qquad Z_{XY} = \begin{tikzcd} & \Bbbk \arrow[dr, "1"] & \\ \Bbbk \arrow[ur, "0"] \arrow[rr, "1"] & & \Bbbk \end{tikzcd}.$$

However, in $H_{\mathscr{A}}$ we have

$$\bar{\Delta}_{\mathscr{A}}(Z_{YX} + Z_{XY} - (q-1)[X \oplus Y]) = (q-1)([S_3] \otimes [I_2] + [P_2] \otimes [S_1])$$

where I_2 is the injective envelope of S_2 and P_2 is its projective cover. Other indecomposable objects with the same image in $\Gamma_{\mathscr{A}}$ are, up to an isomorphism,

$$E_1(\lambda) = \begin{tikzcd} & \Bbbk \arrow[dr, "1"] & \\ \Bbbk \arrow[ur, "1"] \arrow[rr, "\lambda"] & & \Bbbk \end{tikzcd}, \qquad \lambda \in \Bbbk,$$

and we have

$$\bar{\Delta}(E_1(\lambda)) = (q-1)([S_3] \otimes [I_2] + [P_2] \otimes [S_1]).$$

This gives $q-1$ linearly independent primitive elements

$$\mathcal{P}_1(\lambda) = E_1(\lambda) - \frac{1}{q}\sum_{\mu\in\Bbbk} E_1(\mu)$$

and one more primitive element

$$[Z_{YX}] + [Z_{XY}] - (q-1)[X\oplus Y] - \frac{1}{q}\sum_{\lambda\in\Bbbk} E_1(\lambda).$$

Thus, in this case $m_{\alpha_1+\alpha_2+\alpha_3} = 2$ and $\dim \operatorname{Prim}(H_{\mathscr{A}})_{\alpha_1+\alpha_2+\alpha_3} = q$.

In general, primitive elements in Hall algebras corresponding to nonhomogeneous tubes were computed in [Hubery 2005]. It should be noted that they are not primitive in $H_{\mathscr{A}}$ but, conjecturally, can be used to construct primitive elements in a way similar to that shown above.

3.9. ***Cyclic quivers with relations.*** Let $\mathscr{A}$ be the category of representations of the quiver

$$1 \underset{a_{21}}{\overset{a_{12}}{\rightleftarrows}} 2$$

satisfying the relation $a_{21}a_{12} = 0$. The three nonsimple indecomposable objects are, up to an isomorphism,

$$S_{12}:\ \Bbbk \underset{0}{\overset{1}{\rightleftarrows}} \Bbbk\ , \qquad S_{21}:\ \Bbbk \underset{1}{\overset{0}{\rightleftarrows}} \Bbbk\ , \qquad S_{212}:\ \Bbbk \underset{(0\,1)}{\overset{\binom{1}{0}}{\rightleftarrows}} \Bbbk^2\ .$$

The object S_{12} is the projective cover of S_1 while S_{21} is its injective envelope. Thus,

$$\bar{\Delta}([S_{12}]) = (q-1)[S_2]\otimes[S_1], \quad \bar{\Delta}([S_{21}]) = (q-1)[S_1]\otimes[S_2]$$

and so

$$Z = [S_{12}] + [S_{21}] - (q-1)[S_1\oplus S_2] \tag{3-3}$$

is the unique, up to a scalar, primitive element in $|S_1| + |S_2|$. Let $E_1 = [S_1]$ and $E_2 = [S_2]$. Then $\operatorname{Prim}(H_{\mathscr{A}})$ is spanned by E_1, E_2 and Z and

$$[E_1, Z] = [E_2, Z] = 0$$

and

$$[E_1, [E_1, E_2]_q]_{q^{-1}} = (1-q^{-1})E_1 Z, \qquad [E_2, [E_2, [E_2, E_1]]_q]_{q^{-1}} = 0$$

is a presentation of $H_{\mathscr{A}}$.

Now let $\mathscr{A}$ be the category of representations of the same quiver satisfying the relations $a_{21}a_{12} = 0 = a_{12}a_{21}$. In this case, we have four indecomposable objects S_1, S_2, S_{12} and S_{21} which coincide with the ones listed above. Thus, S_{ij} is the injective envelope of S_i and the projective cover of S_j, $\{i, j\} = \{1, 2\}$. As before, we have a unique nonsimple primitive element given by the same formula (3-3). The following provides a presentation for $H_{\mathscr{A}}$:

$$[E_1, [E_1, E_2]_q]_{q^{-1}} = (1-q^{-1})E_1 Z,$$
$$[E_2, [E_2, E_1]_q]_{q^{-1}} = (1-q^{-1})E_2 Z,$$
$$[E_1, Z] = [E_2, Z] = 0.$$

In both examples, we have $C_{\mathscr{A}} \subsetneq U_{\mathscr{A}} = E_{\mathscr{A}} = H_{\mathscr{A}}$ which contributes supporting evidence for Conjecture 1.13. Note also that in this case $m_\gamma = 1$ for $\gamma = |S_1| + |S_2|$.

4. The PBW property of Hall algebras and proof of Theorem 2.4

4.1. ***Rings filtered and graded by ordered monoids.*** Let $(\Lambda, \lhd)$ be an ordered abelian monoid, as defined in §2.2. We write $\mu \unlhd \nu$ if either $\mu = \nu$ or $\mu \neq \nu$ and $\mu \lhd \nu$.

Definition 4.1. We say that a unital ring $\mathcal{H}$ is Λ-filtered if $\mathcal{H}$ contains a family of abelian subgroups $\mathcal{H}^{\unlhd\lambda}$, $\lambda \in \Lambda^+$, such that for all $\lambda, \mu \in \Lambda^+$,

(i) $1_{\mathcal{H}} \in \mathcal{H}^{\unlhd\lambda}$ and $\lambda \unlhd \mu \implies \mathcal{H}^{\unlhd\lambda} \subset \mathcal{H}^{\unlhd\mu}$;

(ii) $\mathcal{H} = \sum_{\lambda\in\Lambda^+} \mathcal{H}^{\unlhd\lambda}$;

(iii) $\mathcal{H}^{\unlhd\lambda} \cdot \mathcal{H}^{\unlhd\mu} \subset \mathcal{H}^{\unlhd(\lambda+\mu)}$.

This definition is similar to that in [Polishchuk and Positselski 2005, §4.7]; however, we do not require the ring $\mathcal{H}$ to admit a $\mathbb{Z}_{\geq 0}$-grading compatible with Λ.

Given $\lambda \in \Lambda^+$, let

$$\mathcal{H}^{\lhd\lambda} = \begin{cases} R & \text{if } \lambda \text{ is minimal,} \\ \sum_{\mu\lhd\lambda} \mathcal{H}^{\unlhd\mu} & \text{if } \lambda \text{ is not minimal,} \end{cases}$$

where $R = \bigcap_{\lambda\in\Lambda^+} \mathcal{H}^{\unlhd\lambda}$. Note that R is a subring of $\mathcal{H}$ and that each $\mathcal{H}^{\unlhd\lambda}$, hence $\mathcal{H}^{\lhd\lambda}$, is an R-bimodule. We have

$$\mathcal{H}^{\lhd\lambda} \cdot \mathcal{H}^{\unlhd\mu} \subset \mathcal{H}^{\lhd(\lambda+\mu)}, \quad \mathcal{H}^{\unlhd\lambda} \cdot \mathcal{H}^{\lhd\mu} \subset \mathcal{H}^{\lhd(\lambda+\mu)}. \tag{4-1}$$

Define the abelian group $\operatorname{gr}_\Lambda \mathcal{H}$ by

$$\operatorname{gr}_\Lambda \mathcal{H} = R \oplus \bigoplus_{\lambda\in\Lambda} \overline{\mathcal{H}}^\lambda, \quad \overline{\mathcal{H}}^\lambda := \mathcal{H}^{\unlhd\lambda}/\mathcal{H}^{\lhd\lambda}.$$

Lemma 4.2. *The abelian group* $\operatorname{gr}_\Lambda \mathcal{H}$ *is a* Λ*-graded unital ring with the multiplication given by*

$$(x + \mathcal{H}^{\lhd\lambda}) \bullet (y + \mathcal{H}^{\lhd\mu}) = x \cdot y + \mathcal{H}^{\lhd(\mu+\nu)}, \quad \textit{for } x \in \mathcal{H}^{\unlhd\lambda},\ y \in \mathcal{H}^{\unlhd\mu}$$

and $r \bullet (x + \mathcal{H}^{\lhd\lambda}) = rx + \mathcal{H}^{\lhd\lambda}$, $(x + \mathcal{H}^{\lhd\lambda}) \bullet r = xr + \mathcal{H}^{\lhd\lambda}$ *for all* $x \in \mathcal{H}^{\unlhd\lambda}$, $r \in R$.

Proof. By construction, the multiplication by elements of R is well-defined. Using (4-1), we obtain, for all $x \in \mathcal{H}^{\trianglelefteq\lambda}$, $y \in \mathcal{H}^{\trianglelefteq\lambda}$,

$$(x+\mathcal{H}^{\lhd\lambda})\bullet(y+\mathcal{H}^{\lhd\mu}) \subset x\cdot y+\mathcal{H}^{\trianglelefteq\lambda}\cdot\mathcal{H}^{\lhd\mu}+\mathcal{H}^{\lhd\lambda}\cdot\mathcal{H}^{\trianglelefteq\mu}+\mathcal{H}^{\lhd\lambda}\cdot\mathcal{H}^{\lhd\mu} \subset x\cdot y+\mathcal{H}^{\lhd(\lambda+\mu)}.$$

Thus, $\bullet$ is well-defined. The distributivity and the associativity follow from those in $\mathcal{H}$. Then the ring $\mathcal{H}$ is graded by Λ by construction. It remains to observe that 1_R is the unity of $\operatorname{gr}_\Lambda \mathcal{H}$. □

Corollary 4.3. *For any Λ-filtered ring $\mathcal{H}$ and any collection $\lambda_1, \dots, \lambda_k \in \Lambda$, we have*

$$\mathcal{H}^{\trianglelefteq\lambda_1}\cdots\mathcal{H}^{\trianglelefteq\lambda_k}/(\mathcal{H}^{\trianglelefteq\lambda_1}\cdots\mathcal{H}^{\trianglelefteq\lambda_k}\cap\mathcal{H}^{\lhd(\lambda_1+\cdots+\lambda_k)}) = \overline{\mathcal{H}}^{\lambda_1}\bullet\cdots\bullet\overline{\mathcal{H}}^{\lambda_k}.$$

Let $\Lambda_{\min}$ be the set of minimal, with respect to the partial order $\trianglelefteq$, elements of Λ^+. We say that Λ is *optimal* if it is generated by $\Lambda_{\min}$.

Recall that an $\mathbb{F}$-algebra A is generated over its subalgebra A_0 by a subspace $A_1 \subset A$ if A_1 is an A_0-bimodule and there exists a surjective homomorphism $T_{A_0}(A_1) \to A$ which restricts to the identity on $A_0 + A_1$. Let $(\Lambda, \trianglelefteq)$ be an optimal monoid and for any subset $\Lambda_\circ$ of $\Lambda_{\min}$ define

$$\mathcal{H}_\circ := \sum_{\lambda\in\Lambda_\circ}\mathcal{H}^{\trianglelefteq\lambda}, \quad \overline{\mathcal{H}}_\circ := \bigoplus_{\lambda\in\Lambda_\circ}\overline{\mathcal{H}}^{\lambda}.$$

Lemma 4.4. *Let $(\Lambda, \trianglelefteq)$ be an optimal monoid and let $\mathcal{H}$ be a Λ-filtered ring. Let $\Lambda_\circ \subset \Lambda_{\min}$ be a generating set for Λ as a monoid. If $\mathcal{H}_\circ$ generates $\mathcal{H}$ then $\overline{\mathcal{H}}_\circ$ generates* $\operatorname{gr}_\Lambda \mathcal{H}$ *over R.*

Proof. Given $x \in \mathcal{H}$, define $\nu(x) = \min\{k \ge 0 : x \in \mathcal{H}_\circ^k\}$ where $\mathcal{H}_\circ^0 = R = \overline{\mathcal{H}}_\circ^0$. Since $\operatorname{gr}_\Lambda \mathcal{H}$ is Λ-graded, it is sufficient to prove that for every $\bar{x} \in \overline{\mathcal{H}}^\lambda$, $\lambda \in \Lambda^+$ we have $\bar{x} \in \overline{\mathcal{H}}_\circ^{\bullet k}$ for some k. Take $x \in \mathcal{H}^{\trianglelefteq\lambda} \setminus \mathcal{H}^{\lhd\lambda}$ such that $x + \mathcal{H}^{\lhd\lambda} = \bar{x}$. Let $k = \nu(x)$. Then

$$x \in \sum \mathcal{H}^{\trianglelefteq\lambda_1}\cdots\mathcal{H}^{\trianglelefteq\lambda_k},$$

where the sum is taken over all $(\lambda_1, \dots, \lambda_k) \in \Lambda_\circ^k$ such that $\lambda_1 + \cdots + \lambda_k = \lambda$. Using Corollary 4.3 we conclude that $\bar{x} \in \sum \overline{\mathcal{H}}^{\lambda_1}\bullet\cdots\bullet\overline{\mathcal{H}}^{\lambda_k} \subset \overline{\mathcal{H}}_\circ{}^{\bullet k}$. □

Proposition 4.5. *Suppose that $(\Lambda, \trianglelefteq)$ is optimal and $\lhd$ is inductive. Let $\Lambda_\circ \subset \Lambda_{\min}$ be a generating set for Λ. If $\overline{\mathcal{H}}_\circ$ generates* $\operatorname{gr}_\Lambda \mathcal{H}$ *over R then $\mathcal{H}_\circ$ generates $\mathcal{H}$.*

Proof. Define

$$\bar{\nu}(\bar{x}) = \min\{k \ge 0 : \bar{x} \in \overline{\mathcal{H}}_\circ^{\bullet k}\}$$

for all $\bar{x} \in \operatorname{gr}_\Lambda \mathcal{H}$. We prove by induction on $f(\lambda)$, $\lambda \in \Lambda^+$ that for every $x \in \mathcal{H}^{\trianglelefteq\lambda}$, we have $x \in \mathcal{H}_\circ^k$ for some $k \ge 0$. This is sufficient since every $x \in \mathcal{H}$ belongs to the sum of finitely many $\mathcal{H}^{\trianglelefteq\lambda}$.

The induction base is obvious since for $\lambda \in \Lambda_\circ$ we can take $k = 1$. Suppose that $x \in \mathcal{H}^{\trianglelefteq\lambda}$ for some $\lambda \in \Lambda^+ \setminus \Lambda_\circ$. If $x \in \mathcal{H}^{\trianglelefteq\mu}$ for some $\mu \lhd \lambda$ then we are done by the induction hypothesis. Therefore, we may assume that $x \in \mathcal{H}^{\trianglelefteq\lambda} \setminus \mathcal{H}^{\lhd\lambda}$ hence $\bar{x} := x + \mathcal{H}^{\lhd\lambda} \neq 0$ in $\operatorname{gr}_\Lambda \mathcal{H}$. Let $k = \bar{\nu}(\bar{x})$. Then

$$\bar{x} \in \sum \overline{\mathcal{H}}^{\lambda_1} \bullet \cdots \bullet \overline{\mathcal{H}}^{\lambda_k},$$

where the sum is taken over $(\lambda_1, \ldots, \lambda_k) \in \Lambda_\circ^k$ such that $\lambda_1 + \cdots + \lambda_k = \lambda$. Then

$$x \in \sum_{\substack{(\lambda_1,\ldots,\lambda_k)\in\Lambda_\circ^k \\ \lambda_1+\cdots+\lambda_k=\lambda}} \mathcal{H}^{\trianglelefteq\lambda_1} \cdots \mathcal{H}^{\trianglelefteq\lambda_k} + \mathcal{H}^{\lhd\lambda} \subset \mathcal{H}_\circ^k + \mathcal{H}^{\lhd\lambda}.$$

hence $x = x' + x''$ where $x' \in \mathcal{H}_\circ^k$, $x'' \in \mathcal{H}^{\lhd\lambda}$. Then using the definition of $\mathcal{H}^{\lhd\lambda}$ we can write $x'' = x_1'' + \cdots + x_\ell''$, where $x_j'' \in \mathcal{H}^{\trianglelefteq\mu_j}$ with $\mu_j \lhd \lambda$, $1 \le j \le \ell$. Since $f(\mu_j) < f(\lambda)$, by the induction hypothesis $x_j'' \in \mathcal{H}_\circ^{k_j'}$ for some $k_j' \ge 1$ with $1 \le j \le \ell$. Then $x \in \mathcal{H}_\circ^{\max(k,k_1',\ldots,k_\ell')}$. □

Proposition 4.6 (weak PBW property). *Let $(\Lambda, \trianglelefteq)$ be an optimal monoid, let $\lhd$ be inductive and let $\Lambda_\circ \subset \Lambda_{\min}$ be a subset which generates Λ as a monoid. Let $\mathcal{H}$ be a Λ-filtered ring. Suppose that there exists a total order $\le$ on $\Lambda_\circ$ such that*

$$\operatorname{gr}_\Lambda \mathcal{H} = \sum_{k\ge 0} \sum_{\lambda_1\le\cdots\le\lambda_k\in\Lambda_\circ^k} \overline{\mathcal{H}}^{\lambda_1} \bullet \cdots \bullet \overline{\mathcal{H}}^{\lambda_k}.$$

Then

$$\mathcal{H} = \sum_{k\ge 0} \sum_{\lambda_1\le\cdots\le\lambda_k\in\Lambda_\circ^k} (\mathcal{H}^{\trianglelefteq\lambda_1}) \cdots (\mathcal{H}^{\trianglelefteq\lambda_k}).$$

Proof. The argument is similar to the proof of Proposition 4.5. Let

$$\mathcal{H}^{\langle k\rangle} = \sum_{(\lambda_1\le\cdots\le\lambda_k)\in\Lambda_\circ^k} (\mathcal{H}^{\trianglelefteq\lambda_1}) \cdots (\mathcal{H}^{\trianglelefteq\lambda_k}),$$

We prove, by induction on $f(\lambda)$, $\lambda \in \Lambda^+$ that for all $x \in \mathcal{H}^{\trianglelefteq\lambda}$ there exists $k \ge 0$ such that $x \in \mathcal{H}^{\langle k\rangle}$. If $\lambda \in \Lambda_\circ$ then $x \in \mathcal{H}^{\langle 1\rangle}$ and we are done. Otherwise,

$$x + \mathcal{H}^{\lhd\lambda} \in \sum_{\substack{(\lambda_1\le\cdots\le\lambda_k)\in\Lambda_\circ^k \\ \lambda_1+\cdots+\lambda_k=\lambda}} \overline{\mathcal{H}}^{\lambda_1} \bullet \cdots \bullet \overline{\mathcal{H}}^{\lambda_k},$$

which implies that $x \in \mathcal{H}^{\langle k\rangle} + \mathcal{H}^{\lhd\lambda}$. Since $\mathcal{H}^{\lhd\lambda} = \sum_{\mu\lhd\lambda} \mathcal{H}^{\trianglelefteq\mu}$, we then have $x = x' + x_1'' + \cdots + x_\ell''$ where $x' \in \mathcal{H}^{\langle k\rangle}$ and $x_j'' \in \mathcal{H}^{\trianglelefteq\mu_j}$ for $\mu_j \lhd \lambda$ and $1 \le j \le \ell$. Applying the induction hypothesis to the x_j'' we conclude that $x_j'' \in \mathcal{H}^{\langle k_j\rangle}$ for some k_j, $1 \le j \le \ell$, hence $x \in \mathcal{H}^{\langle\max(k,k_1,\ldots,k_\ell)\rangle}$. □

We now consider a special case which we will later apply to Hall algebras.

Corollary 4.7. *Let* $(\Lambda, \trianglelefteq)$ *be an optimal monoid and let* $\triangleleft$ *be an inductive order. Let* $\mathcal{H}$ *be a unital* $\mathbb{F}$*-algebra with a basis* $\{[\lambda] : \lambda \in \Lambda\}$ *such that* $[0] = 1_{\mathcal{H}}$ *and*

$$[\lambda] \cdot [\mu] \in \mathbb{F}^{\times}[\lambda + \mu] + \sum_{\nu \triangleleft \lambda+\mu} \mathbb{F}[\nu].$$

for all $\lambda, \mu \in \Lambda$. *Then for any subset* $\Lambda_\circ$ *of* $\Lambda_{\min}$ *which generates* Λ *as a monoid, the set* $[\Lambda_\circ] := \{[\lambda] : \lambda \in \Lambda_\circ\}$ *generates* $\mathcal{H}$ *as an algebra. Moreover, for any total order on* $\Lambda_\circ$, *the set* $\mathbf{M}([\Lambda_\circ])$ *of ordered monomials in* $[\Lambda_\circ]$ *spans* $\mathcal{H}$ *as an* $\mathbb{F}$*-vector space. Finally, if* Λ *is freely generated by* $\Lambda_\circ$ *then* $\mathbf{M}([\Lambda_\circ])$ *is a basis of* $\mathcal{H}$.

Proof. Clearly, $\mathcal{H}$ is Λ-filtered with $\mathcal{H}^{\trianglelefteq\lambda} = \mathbb{F}\{[\mu] : \mu \trianglelefteq \lambda\}$. In particular, $R = \mathbb{F}\cdot[0] = \mathbb{F}$. Then $\operatorname{gr}_\Lambda \mathcal{H}$ has a basis $\{\overline{[\lambda]} : \lambda \in \Lambda\}$ and

$$\overline{[\lambda]} \cdot \overline{[\mu]} \in \mathbb{F}^{\times}\overline{[\lambda + \mu]}, \tag{4-2}$$

hence $\overline{[\lambda + \mu]} \in \mathbb{F}^{\times}\overline{[\lambda]} \cdot \overline{[\mu]}$.

Let $\leq$ be any total order on $\Lambda_\circ$. Given $\lambda \in \Lambda$, we can write $\lambda = \lambda_1 + \cdots + \lambda_r$ with $\lambda_i \in \Lambda_\circ$, $1 \leq i \leq r$ and $\lambda_1 \leq \cdots \leq \lambda_r$. By (4-2) we have $\overline{[\lambda]} \in \mathbb{F}^{\times}\overline{[\lambda_1]} \cdots \overline{[\lambda_r]}$. Taking into account that $\mathcal{H}^{\trianglelefteq\lambda} = \mathbb{F} + \mathbb{F}[\lambda]$ for $\lambda \in \Lambda_\circ$, we see that all assumptions of Proposition 4.6 are satisfied. □

4.2. ***Proof of Theorem 2.2.*** The key ingredient of our argument is the following result.

Proposition 4.8. *Let* $\mathscr{A}$ *be a* Hom*-finite exact category. Then for any short exact sequence*

$$M^{-} \overset{f_-}{\rightarrowtail} M \overset{f_+}{\twoheadrightarrow} M^{+}, \tag{4-3}$$

we have

$$e([M]) \leq e([M^{+} \oplus M^{-}]), \tag{4-4}$$

where $e([X]) := \#\operatorname{End}_{\mathscr{A}} X$ *for* $[X] \in \operatorname{Iso}\mathscr{A}$. *Moreover, if* (4-4) *is an equality then* (4-3) *splits.*

Proof. We need to prove that the following inequalities hold for every N in $\mathscr{A}$:

$$\begin{aligned} &\#\operatorname{Hom}_{\mathscr{A}}(N, M) \leq \#\operatorname{Hom}(N, M^{+} \oplus M^{-}), \\ &\#\operatorname{Hom}_{\mathscr{A}}(M, N) \leq \#\operatorname{Hom}(M^{+} \oplus M^{-}, N). \end{aligned} \tag{4-5}$$

To prove the first inequality, recall (see, e.g., [Buchsbaum 1959; Yoneda 1954]) that for every N in $\mathscr{A}$, (4-3) induces a long exact sequence of finite abelian groups

$$\begin{aligned} 0 \to \operatorname{Hom}_{\mathscr{A}}(N, M^{-}) \to \operatorname{Hom}_{\mathscr{A}}(N, M) \to \operatorname{Hom}_{\mathscr{A}}(N, M^{+}) \overset{\delta_*}{\to} \\ \operatorname{Ext}^1_{\mathscr{A}}(N, M^{-}) \to \operatorname{Ext}^1_{\mathscr{A}}(N, M) \to \operatorname{Ext}^1_{\mathscr{A}}(N, M^{+}) \to \cdots . \end{aligned}$$

Truncating this sequence yields an exact sequence

$$(4\text{-}6)\quad 0 \to \operatorname{Hom}_{\mathscr{A}}(N, M^-) \to \operatorname{Hom}_{\mathscr{A}}(N, M) \to \operatorname{Hom}_{\mathscr{A}}(N, M^+) \xrightarrow{\delta_*} \operatorname{Im}\delta_* \to 0.$$

Then, computing the multiplicative Euler characteristic of (4-6), we obtain

$$(4\text{-}7)\qquad \begin{aligned}\#\operatorname{Hom}_{\mathscr{A}}(N, M^-)\cdot\#\operatorname{Hom}_{\mathscr{A}}(N, M^+) &= \#\operatorname{Hom}_{\mathscr{A}}(N, M)\cdot\#\operatorname{Im}\delta_*\\ &\geq \#\operatorname{Hom}_{\mathscr{A}}(N, M),\end{aligned}$$

which immediately yields the first inequality in (4-5).

To prove the second inequality, recall that for all N in $\mathscr{A}$, (4-3) induces a long exact sequence of abelian finite groups

$$\begin{aligned}0 \to \operatorname{Hom}_{\mathscr{A}}(M^+, N) \to \operatorname{Hom}_{\mathscr{A}}(M, N) \to \operatorname{Hom}_{\mathscr{A}}(M^-, N) \xrightarrow{\delta^*}&\\ \operatorname{Ext}^1_{\mathscr{A}}(M^+, N) \to \operatorname{Ext}^1_{\mathscr{A}}(M, N) \to \operatorname{Ext}^1_{\mathscr{A}}(M^-, N) \to \cdots&\,.\end{aligned}$$

Similarly, truncating this sequence yields

$$0 \to \operatorname{Hom}_{\mathscr{A}}(M^+, N) \to \operatorname{Hom}_{\mathscr{A}}(M, N) \to \operatorname{Hom}_{\mathscr{A}}(M^-, N) \xrightarrow{\delta^*} \operatorname{Im}\delta^* \to 0,$$

and the argument identical to the above gives

$$\#\operatorname{Hom}_{\mathscr{A}}(M, N) \leq \#\operatorname{Hom}_{\mathscr{A}}(M^+, N)\cdot\#\operatorname{Hom}_{\mathscr{A}}(M^-, N)$$

which is equivalent to the second inequality in (4-5).

Combining the first inequality in (4-5) with $N = M^+ \oplus M^-$ and the second inequality in (4-5) with $N = M$ we obtain

$$\begin{aligned}e([M]) = \#\operatorname{End}_{\mathscr{A}} M &\leq \#\operatorname{Hom}_{\mathscr{A}}(M^+ \oplus M^-, M)\\ &\leq \#\operatorname{End}_{\mathscr{A}} M^+ \oplus M^- = e([M^+ \oplus M^-]).\end{aligned}$$

To prove the last assertion, it suffices to show, in view of the above chain of inequalities, that $\#\operatorname{Hom}_{\mathscr{A}}(M^+ \oplus M^-, M) = \#\operatorname{End}_{\mathscr{A}} M^+ \oplus M^-$ implies that (4-3) splits. Indeed, using the additivity of $\operatorname{Hom}_{\mathscr{A}}$ in the first argument we rewrite the latter equality as

$$\begin{aligned}&\#\operatorname{Hom}_{\mathscr{A}}(M^+, M)\cdot\#\operatorname{Hom}_{\mathscr{A}}(M^-, M)\\ &\qquad = \#\operatorname{Hom}_{\mathscr{A}}(M^+, M^+ \oplus M^-)\cdot\#\operatorname{Hom}_{\mathscr{A}}(M^-, M^+ \oplus M^-).\end{aligned}$$

This and (4-5) taken with $N = M^-$ imply

$$\#\operatorname{Hom}_{\mathscr{A}}(M^+, M) \geq \#\operatorname{Hom}_{\mathscr{A}}(M^+, M^+ \oplus M^-),$$

which, together with (4-5) with $N = M^+$, yield

$$\#\operatorname{Hom}_{\mathscr{A}}(M^+, M) = \#\operatorname{Hom}_{\mathscr{A}}(M^+, M^+ \oplus M^-).$$

The last equality and (4-7) taken with $N = M^+$ imply that $E_0 = \operatorname{Im}\delta_* = 0$, hence the natural map $\operatorname{Hom}_{\mathscr{A}}(M^+, M) \to \operatorname{End}_{\mathscr{A}} M^+$ is surjective. Therefore, there exists $g \in \operatorname{Hom}_{\mathscr{A}}(M^+, M)$ such that $f_+ \circ g = 1_{M^+}$, hence (4-3) splits. □

Recall that $\lhd$ is the preorder defined as the transitive closure of the relation

$$[M] \lhd [M^- \oplus M^+] \iff \exists \text{ a nonsplit short exact sequence } M^- \rightarrowtail M \twoheadrightarrow M^+$$

(cf. §2.2). By Proposition 4.8, $\lhd$ is an inductive preorder with the function mapping $[X]$ to $e([X])$, hence is an inductive partial order.

It remains to prove that the order $\lhd$ is compatible with the addition in $\operatorname{Iso}\mathscr{A}$. Indeed, note that for any X in $\mathscr{A}$, the short exact sequence (4-3) yields a short exact sequence

$$M^- \oplus X \xrightarrow{\begin{pmatrix} f_- & 0 \\ 0 & 1_X \end{pmatrix}} M \oplus X \xrightarrow{(f_+,0)} M^+, \tag{4-8}$$

hence $[M\oplus X] \trianglelefteq [M^- \oplus M^+ \oplus X]$. If $[M] \lhd [M^- \oplus M^+]$, that is, (4-3) is nonsplit, then clearly (4-8) is also nonsplit, so $[M\oplus X] \lhd [M^- \oplus M^+ \oplus X]$. Taking transitive closure implies that $[M\oplus X]\lhd[N\oplus X]$ for all $[M], [N] \in \operatorname{Iso}\mathscr{A}$ such that $[M]\lhd[N]$ and for all $[X] \in \operatorname{Iso}\mathscr{A}$. This completes the proof of Theorem 2.2. □

4.3. *Proof of Theorem 2.4.* We are now going to apply the machinery developed in §4.1. We begin by proving that $(\operatorname{Iso}\mathscr{A}, \lhd)$ is optimal.

Lemma 4.9. *Let $\mathscr{A}$ be an exact* Hom*-finite category. Then every object X in $\mathscr{A}$ is a finite direct sum of indecomposable objects and the number of indecomposable summands of X is bounded above by $\#\operatorname{End}_{\mathscr{A}} X$.*

Proof. Let X be a nonzero object in $\mathscr{A}$. Write $X = X_1 \oplus \cdots \oplus X_s$ for some $s > 0$, where all the X_i are nonzero. Then $\#\operatorname{End}_{\mathscr{A}} X \geq \sum_{j=1}^{s} \#\operatorname{End}_{\mathscr{A}} X_i \geq s$. Let k be the maximal positive integer s such that X can be written as a direct sum of s nonzero objects. The maximality of k immediately implies that each summand is indecomposable. □

Remark 4.10. It should be noted that the Krull–Schmidt theorem does not have to hold in this generality. For example, the full subcategory of the category of $\Bbbk$-representations of the quiver $1 \to 0 \leftarrow 2$, with the dimension vector satisfying $\dim_{\Bbbk} V_1 = \dim_{\Bbbk} V_2$, is not Krull–Schmidt.

Corollary 4.11. *The monoid* $\operatorname{Iso}\mathscr{A}$ *is generated by* $\operatorname{Ind}\mathscr{A}$ *and is optimal with respect to $\trianglelefteq$.*

Proof. The first assertion is immediate from the lemma. To prove the second, observe that if $[N]$ is not minimal, then $[M] \lhd [N]$ for some $[M] \in \operatorname{Iso}\mathscr{A}$ and so N is decomposable. Thus, every $[X] \in \operatorname{Ind}\mathscr{A}$ is minimal with respect to the partial order $\trianglelefteq$, hence $\operatorname{Iso}\mathscr{A}$ is generated by its minimal elements. □

Proof of Theorem 2.4. Since $(\operatorname{Iso}\mathscr{A}, \lhd)$ is optimal and $\lhd$ is inductive, the algebra $H_{\mathscr{A}}$ satisfies the assumptions of Corollary 4.7 with $\Lambda = \operatorname{Iso}\mathscr{A}$ and $\Lambda_\circ = \operatorname{Ind}\mathscr{A}$. Therefore, for any total order on $\operatorname{Ind}\mathscr{A}$, ordered monomials on $\operatorname{Ind}\mathscr{A}$ span $H_{\mathscr{A}}$. Finally, if $\mathscr{A}$ is Krull–Schmidt, $\operatorname{Iso}\mathscr{A}$ is freely generated by $\operatorname{Ind}\mathscr{A}$, hence ordered monomials on $\operatorname{Ind}\mathscr{A}$ form a basis of $H_{\mathscr{A}}$. □

5. The Grothendieck monoid of a profinitary category

5.1. *Almost simple objects.* We will repeatedly need the following obvious description of the defining relation of the Grothendieck monoid.

Lemma 5.1. *Suppose that* $[X] \neq [Y] \in (\operatorname{Iso}\mathscr{A})^+$ *and* $|X| = |Y|$. *Then there exist* $[X_i] \in (\operatorname{Iso}\mathscr{A})^+$, $0 \le i \le r$ *and* $[A_i], [B_i] \in (\operatorname{Iso}\mathscr{A})^+$, $1 \le i \le r$ *such that* $[X_0] = [X]$, $[X_r] = [Y]$ *and* $[X_{i-1}], [X_i] \in \underline{\operatorname{Ext}}^1_{\mathscr{A}}(A_i, B_i)$, $1 \le i \le r$.

Definition 5.2. We say that an object $X \neq 0$ in an exact category $\mathscr{A}$ is *almost simple* if there is no nontrivial short exact sequence $Y \rightarrowtail X \twoheadrightarrow Z$ (or, equivalently, $[X] \in \underline{\operatorname{Ext}}^1_{\mathscr{A}}(A, B) \Longrightarrow \{[A], [B]\} = \{[X], [0]\}$) and *simple* if it has no proper nonzero subobjects.

Clearly, in an abelian category these notions coincide. Note that an almost simple object is always indecomposable. Let $S_{\mathscr{A}} \subset \operatorname{Iso}\mathscr{A}$ be the set of isomorphism classes of almost simple objects. The definition (2-3) of comultiplication Δ implies that

$$F_X^{AB} = \begin{cases} 1 & \text{if } \{[A], [B]\} = \{[X], [0]\}, \\ 0 & \text{otherwise,} \end{cases}$$

hence

$$S_{\mathscr{A}} \subset \operatorname{Prim}(H_{\mathscr{A}}). \tag{5-1}$$

Let Γ be an abelian monoid. Observe that the elements of $\Gamma^+ \setminus (\Gamma^+ + \Gamma^+)$ are precisely the minimal elements of Γ^+ in the preorder $\preceq$ (cf. §2.4).

Lemma 5.3. *Let* $\mathscr{A}$ *be an exact category. Then the restriction of the canonical homomorphism of monoids* $\phi_{\mathscr{A}} : \operatorname{Iso}\mathscr{A} \to \Gamma_{\mathscr{A}}$ *to* $S_{\mathscr{A}}$ *is a bijection*

$$S_{\mathscr{A}} \to \Gamma^+_{\mathscr{A}} \setminus (\Gamma^+_{\mathscr{A}} + \Gamma^+_{\mathscr{A}}). \tag{5-2}$$

In particular, if $\mathscr{A}$ *is* Hom*-finite, then* $(H_{\mathscr{A}})_\gamma$ *equals* $\operatorname{Prim}(H_{\mathscr{A}})_\gamma$ *and is one-dimensional for all* $\gamma \in \Gamma^+_{\mathscr{A}} \setminus (\Gamma^+_{\mathscr{A}} + \Gamma^+_{\mathscr{A}})$.

Proof. Lemma 5.1 implies that for $[X] \in S_{\mathscr{A}}$, we have $|X| = |Y|$ if and only if $[X] = [Y]$. This shows that the restriction of $\phi_{\mathscr{A}}$ to $S_{\mathscr{A}}$ is injective. Furthermore, if $|X| = |Y| + |Z| = |Y \oplus Z|$ for some nonzero $[Y], [Z]$ then $[X] = [Y \oplus Z]$, which is a contradiction since X is indecomposable. Thus, $\operatorname{Im}\phi_{\mathscr{A}} \subset \Gamma^+_{\mathscr{A}} \setminus (\Gamma^+_{\mathscr{A}} + \Gamma^+_{\mathscr{A}})$ and so the map in (5-2) is well-defined. Finally, if $[X] \notin S_{\mathscr{A}}$, then $[X] \in \underline{\operatorname{Ext}}^1_{\mathscr{A}}(A, B)$

with $|A|, |B| \in \Gamma^+_{\mathscr{A}}$, hence $|X| = |A| + |B| \in \Gamma^+_{\mathscr{A}} + \Gamma^+_{\mathscr{A}}$. Thus, the preimage of $\Gamma^+_{\mathscr{A}} \setminus (\Gamma^+_{\mathscr{A}} + \Gamma^+_{\mathscr{A}})$ is contained in $S_{\mathscr{A}}$ hence $\phi_{\mathscr{A}}|_{S_{\mathscr{A}}}$ is surjective.

In particular, $\dim_{\mathbb{Q}}(H_{\mathscr{A}})_\gamma = \#\operatorname{Iso}\mathscr{A}_\gamma = 1$ for all $\gamma \in \Gamma^+_{\mathscr{A}} \setminus (\Gamma^+_{\mathscr{A}} + \Gamma^+_{\mathscr{A}})$. The equality $(H_{\mathscr{A}})_\gamma = \operatorname{Prim}(H_{\mathscr{A}})_\gamma$ now follows from (5-1). □

Remark 5.4. Note that we can have $\dim_{\mathbb{Q}}(H_{\mathscr{A}})_\gamma = 1$ even for $\gamma \in \Gamma^+ + \Gamma^+$. For example, if S, S' are simple objects with $\operatorname{Ext}^1_{\mathscr{A}}(S, S') = \operatorname{Ext}^1_{\mathscr{A}}(S', S) = 0$, then $(H_{\mathscr{A}})_{|S|+|S'|} = \mathbb{Q}[S \oplus S'] = \mathbb{Q}[S][S']$. However, in that case $\operatorname{Prim}(H_{\mathscr{A}})_\gamma = 0$.

5.2. ***Proof of Proposition 2.12.*** Let

$$\Gamma^f_{\mathscr{A}} = \{\gamma \in \Gamma_{\mathscr{A}} : \#\operatorname{Iso}\mathscr{A}_\gamma < \infty\}.$$

Thus, $\mathscr{A}$ is profinitary if $\Gamma_{\mathscr{A}} = \Gamma^f_{\mathscr{A}}$. Note, however, that $\Gamma^f_{\mathscr{A}}$ need not be a submonoid of $\Gamma_{\mathscr{A}}$. Since $\#\operatorname{Iso}\mathscr{A}_\gamma = 1$ for $\gamma \in \Gamma^+_{\mathscr{A}}$ minimal, all minimal elements of $\Gamma_{\mathscr{A}}$ are contained in $\Gamma^f_{\mathscr{A}}$. Given $\gamma \in \Gamma^f_{\mathscr{A}}$, let $s_\gamma = \max_{[X] \in \operatorname{Iso}\mathscr{A}_\gamma} \#\operatorname{End}_{\mathscr{A}} X$.

Proposition 2.12 is a special case of the following proposition.

Proposition 5.5. *Let $\mathscr{A}$ be a* Hom*-finite exact category. Then the restriction of the preorder $\preceq$ to $\Gamma^f_{\mathscr{A}}$ is a partial order. Moreover, $\Gamma^f_{\mathscr{A}}$ is contained in the submonoid of $\Gamma_{\mathscr{A}}$ generated by its minimal elements.*

Proof. We need the following lemma.

Lemma 5.6. *Let $\gamma \in \Gamma^f_{\mathscr{A}} \setminus \{0\}$. Then γ can be written as a sum of finitely many minimal elements of $\Gamma^+_{\mathscr{A}}$ and the number of summands in any such presentation is bounded by s_γ.*

Proof. The proof is almost identical to that of Lemma 4.9. Write $\gamma = \gamma_1 + \cdots + \gamma_s$ for some $\gamma_i \in \Gamma^+_{\mathscr{A}}$. Take $X_i \in \mathscr{A}$ with $|X_i| = \gamma_i$ and let $X = X_1 \oplus \cdots \oplus X_s$. Then s cannot exceed the maximal possible number of indecomposable summands of X which, by Lemma 5.3, is bounded above by $\#\operatorname{End}_{\mathscr{A}} X \le s_\gamma$. Let k be the maximal integer s such that γ can be written as a sum of s elements of $\Gamma^+_{\mathscr{A}}$. Then the maximality of k implies that each summand is minimal. □

It follows from Lemma 5.6 that for $\alpha \in \Gamma^f_{\mathscr{A}}$, $\alpha = \alpha + \beta$ implies that $\beta = 0$. Then $\alpha + \beta + \gamma = \alpha$ implies that $\beta + \gamma = 0$, hence $\beta = \gamma = 0$ since 0 is the only invertible element of $\Gamma_{\mathscr{A}}$. The first assertion of the proposition now follows from Lemma 2.10, while the second is immediate from Lemma 5.6. □

5.3. ***Proofs of Theorems 1.4, 2.14 and Corollary 1.5.*** We begin with Theorem 1.4.

Proof of Theorem 1.4. Since $[A \oplus B] \in \underline{\mathrm{Ext}}^1_{\mathscr{A}}(A, B)$, Definition 2.11 implies that a profinitary category $\mathscr{A}$ is cofinitary if and only if for any $\gamma \in \Gamma_{\mathscr{A}}$ the set

$$\begin{aligned}\mathcal{E}_\gamma &:= \{([A], [B]) \in \operatorname{Iso}\mathscr{A} \times \operatorname{Iso}\mathscr{A} : |A| + |B| = \gamma\} \\ &= \bigcup_{[X]\in\operatorname{Iso}\mathscr{A}\,:\,|X|=\gamma} \{([A], [B]) \in \operatorname{Iso}\mathscr{A} \times \operatorname{Iso}\mathscr{A} : [X] \in \underline{\mathrm{Ext}}^1_{\mathscr{A}}(A, B)\}\end{aligned}$$

is finite. On the other hand,

$$\mathcal{E}_\gamma = \bigcup_{\alpha,\beta\in\Gamma_{\mathscr{A}}\,:\,\alpha+\beta=\gamma} \operatorname{Iso}\mathscr{A}_\alpha \times \operatorname{Iso}\mathscr{A}_\beta.$$

Therefore, $\mathcal{E}_\gamma$ is finite if and only if $\{(\alpha, \beta) \in \Gamma_{\mathscr{A}} \times \Gamma_{\mathscr{A}} : \alpha + \beta = \gamma\}$ is finite. □

Now we proceed to prove Theorem 2.14. Given an object $X \in \mathscr{A}$, an admissible flag on X is a sequence of objects $X_0 = X, X_1, \dots, X_s = 0$ together with short exact sequences $X_i \rightarrowtail X_{i-1} \twoheadrightarrow Y_i$, $1 \le i \le s$. An admissible flag is said to be a *composition series* if the Y_i are almost simple for all $1 \le i \le s$.

Proposition 5.7. *Let $\mathscr{A}$ be a profinitary exact category. Suppose that $\gamma \in \Gamma_{\mathscr{A}} \setminus \{0\}$. Then $X \in \mathscr{A}$ with $|X| = \gamma$ admits a composition series. Moreover, the length of any composition series of X is bounded above by s_γ.*

Proof. We use induction on the partially ordered set $(\Gamma_{\mathscr{A}}, \preceq)$ (see Proposition 5.5). If $\gamma \in \Gamma_{\mathscr{A}}$ is minimal then X with $|X| = \gamma$ is almost simple by (5-2), and hence admits a composition series. Suppose the assertion is established for all $\gamma' \prec \gamma$ and γ is not minimal. Then X with $|X| = \gamma$ is not almost simple, hence there exists a short exact sequence $X'' \rightarrowtail X \overset{h}{\twoheadrightarrow} X'$ with $|X'|, |X''| \prec |X|$. By the induction hypothesis there exists a short exact sequence $Y'' \rightarrowtail X' \overset{g}{\twoheadrightarrow} Y$ with Y almost simple. Let $Y_1 = Y$. Then we have a short exact sequence

$$X_1 \rightarrowtail X \overset{gh}{\twoheadrightarrow} Y_1$$

where $|X_1| \prec |X|$. Therefore, X_1 admits a composition series by the induction hypothesis, which establishes the first assertion of the lemma. The second assertion is immediate from Lemma 5.6 since $|X| = |Y_1| + \cdots + |Y_s|$. □

Proof of Theorem 2.14. If $\mathscr{A}$ is profinitary and abelian, then the composition series from Proposition 5.7 is a composition series in the usual sense since all almost simple objects are simple. Theorem 2.14 is now immediate. □

Proof of Corollary 1.5. Since a full exact subcategory of a cofinitary exact category is also cofinitary, to prove (a), it suffices to consider the case when $\mathscr{A}$ is a profinitary abelian category. Note that the uniqueness of composition factors in an abelian category with the finite length property (see, e.g., [Joyce 2006, Theorem 2.7]) implies that $\Gamma_{\mathscr{A}}$ is freely generated by its minimal elements. It remains to apply

Theorem 1.4. To prove (b), note that by Lemma 5.6, $\Gamma_{\mathscr{A}}$ is finitely generated if and only if it contains finitely many minimal elements $\gamma_1, \dots, \gamma_n$. Again by Lemma 5.6, the number of decompositions of $\gamma \in \Gamma_{\mathscr{A}}$ as $\gamma = \sum_{i=1}^n c_i \gamma_i$, $c_i \in \mathbb{Z}_{\ge 0}$ is bounded above by $\binom{s_\gamma + n}{n}$, which is the number of n-tuples $(c_1, \dots, c_n) \in \mathbb{Z}_{\ge 0}^n$ with $\sum_{i=1}^n c_i \le s_\gamma$. The assertion is now immediate from Theorem 1.4. $\square$

6. Coalgebras in tensor categories and proof of the main theorem

6.1. ***Quasiprimitive elements and coideals.*** Let $\mathbb{F}$ be a field of characteristic zero. Let H_0 be a bialgebra over $\mathbb{F}$ and let $\mathscr{C}$ be the category of left H_0-comodules. Given $V \in \mathscr{C}$, we denote the left coaction of H_0 by $\delta_V : V \to H_0 \otimes V$ and, using Sweedler-like notation, write

$$\delta_V(v) = v^{(-1)} \otimes v^{(0)}, \quad v \in V.$$

The category $\mathscr{C}$ is an $\mathbb{F}$-linear tensor category with the unit object $\mathbb{F}$, the tensor product $A \otimes B = A \otimes_{\mathbb{F}} B$ of objects $A, B \in \mathscr{C}$ acquiring a left H_0-comodule structure via

$$\delta_{A\otimes B}(a \otimes b) = a^{(-1)} b^{(-1)} \otimes a^{(0)} \otimes b^{(0)},$$

for all $a \in A$, $b \in B$.

By definition, a coalgebra in $\mathscr{C}$ is an object $C \in \mathscr{C}$ together with morphisms $\Delta \in \operatorname{Hom}_{\mathscr{C}}(C, C \otimes C)$ and $\varepsilon \in \operatorname{Hom}_{\mathscr{C}}(C, \mathbb{F})$ satisfying the usual axioms. For any coalgebra C in $\mathscr{C}$, denote by $C_0 = \operatorname{Corad}_{\mathscr{C}}(C)$ the sum of all simple finite dimensional subcoalgebras of C in $\mathscr{C}$ and refer to it as the *coradical* of C in $\mathscr{C}$. Clearly, C_0 is a subcoalgebra of C in $\mathscr{C}$. Denote also

$$C_1 = \operatorname{QPrim}_{\mathscr{C}}(C) = \Delta^{-1}(C \otimes C_0 + C_0 \otimes C)$$

and refer to it as the quasiprimitive set of C. Then C_1 is a $\mathscr{C}$-subobject of C. It is well-known (see [Sweedler 1969, Corollary 9.1.7]) that

$$\Delta(C_1) \subset C_1 \otimes C_0 + C_0 \otimes C_1.$$

In particular, if $C_0 = \mathbb{F}$ then $\operatorname{QPrim}_{\mathscr{C}}(C) = \mathbb{F} \oplus \operatorname{Prim}(C)$. More generally, we have the following lemma which extends a well-known result (cf. [Montgomery 1993, Theorem 5.2.2; Sweedler 1969, §9.1]).

Lemma 6.1. *Any coalgebra C in $\mathscr{C}$ admits an increasing coradical filtration by subcoalgebras $C_k \subset C$ in $\mathscr{C}$, $k \ge 0$, defined by $C_0 = \operatorname{Corad}_{\mathscr{C}}(C)$, $C_1 = \operatorname{QPrim}_{\mathscr{C}}(C)$ and*

$$C_k = \Delta^{-1}(C \otimes C_{k-1} + C_0 \otimes C)$$

for $k > 1$. Moreover, $\Delta(C_k) = \sum_{i=0}^k C_i \otimes C_{k-i}$. $\square$

A coideal in C is a $\mathscr{C}$-subobject I of C satisfying

$$\Delta(I) \subset C \otimes I + I \otimes C.$$

Proposition 6.2. *Let C be a coalgebra in $\mathscr{C}$. Then for any nonzero coideal I in $\mathscr{C}$ one has*

$$I \cap \mathrm{QPrim}_{\mathscr{C}}(C) \neq \{0\}.$$

Proof. For each $k \geq 0$ denote $I_k := I \cap C_k$. If $I_0 \neq \{0\}$, then we are done since $I_0 \subset C_0 \subset C_1$. Assume that $I_0 = 0$. Since $C_0 \subset C_1 \subset \cdots$ is a filtration, there exists a unique $k \geq 1$ such that $I_{k-1} = 0$ and $I_k \neq 0$. Then

$$\Delta(I_k) \subset C_0 \otimes I_k + I_k \otimes C_0.$$

Since C_1 is the maximal subobject V of C with the property $\Delta(V) \subset C_0 \otimes V + V \otimes C_0$, it follows that $I_k \subset C_1$ and so $k = 1$. Thus, $I_1 = I \cap C_1 = I \cap \mathrm{QPrim}_{\mathscr{C}}(C) \neq \{0\}$. □

6.2. *Invariant pairing.* Given two objects A, B in $\mathscr{C}$, a pairing $\langle \cdot, \cdot \rangle : A \otimes B \to \mathbb{F}$ is called H_0-*invariant* if

$$a^{(-1)} \langle a^{(0)}, b \rangle = b^{(-1)} \langle a, b^{(0)} \rangle$$

for all $a \in A$, $b \in B$.

The following example plays a fundamental role in the sequel.

Example 6.3. Let Γ be an abelian monoid. Its monoidal algebra $H_0 = \mathbb{F}\Gamma$ has a natural coalgebra structure, with the elements of Γ being group-like. Then a left H_0-comodule V is in fact a Γ-graded vector space, since $V = \bigoplus_{\gamma \in \Gamma} V_\gamma$ where $V_\gamma = \{v \in V : \delta_V(v) = \gamma \otimes v\}$. It is easy to see that a pairing $\langle \cdot, \cdot \rangle : A \otimes B \to \mathbb{F}$ is H_0-invariant if and only if $\langle A_\gamma, B_{\gamma'} \rangle = 0$, $\gamma \neq \gamma' \in \Gamma$.

Lemma 6.4. *Let $\langle \cdot, \cdot \rangle : A \otimes B \to \mathbb{F}$ be an H_0-invariant pairing between objects A and B of $\mathscr{C}$. Then for any subobject A_0 of A in $\mathscr{C}$, its right orthogonal complement*

$$A_0{}^{\perp} = \{b \in B : \langle A_0, b \rangle = 0\}$$

is a subobject of B in $\mathscr{C}$. Likewise, for any subobject B_0 of B in $\mathscr{C}$, its left orthogonal complement

$${}^{\perp}B_0 = \{a \in A : \langle a, B_0 \rangle = 0\}$$

is a subobject of A in $\mathscr{C}$.

Proof. We prove the first assertion only, the argument for the second one being similar. Given $b \in A_0{}^{\perp}$, write $\delta_B(b) = \sum_i h_i \otimes b_i$ where the $h_i \in H_0$ are linearly independent and $b_i \in B$. Since the pairing is invariant, we have for all $a \in A_0$

$$\sum_i h_i \langle a, b_i \rangle = a^{(-1)} \langle a^{(0)}, b \rangle = 0$$

since $\delta_A(a) = a^{(-1)} \otimes a^{(0)} \in H_0 \otimes A_0$. Therefore, $\langle a, b_i \rangle = 0$ for all i, hence $b_i \in {A_0}^{\perp}$ and so $\delta_B(b) \in H_0 \otimes {A_0}^{\perp}$. □

We now prove that an H_0-invariant pairing between nonisomorphic simple objects in $\mathscr{C}$ must be identically zero. For that purpose, it will be convenient to introduce the dual picture. Let $H_0^* = \mathrm{Hom}_{\mathbb{F}}(H_0, \mathbb{F})$. Then H_0^* is an associative $\mathbb{F}$-algebra via $f \cdot g = (f \otimes g) \circ \Delta_{H_0}$ for all $f, g \in H_0^*$, where $\Delta_{H_0} : H_0 \to H_0 \otimes H_0$ is the comultiplication on H_0 (hereafter we identify $\mathbb{F} \otimes_{\mathbb{F}} V$ with V via the canonical isomorphism). Then a left H_0-comodule V is naturally a left H_0^*-module via $f \triangleright v = (f \otimes 1)\delta_V(v)$, for all $f \in H_0^*$ and $v \in V$. This yields a fully faithful functor from the category $\mathscr{C}$ to the category of left H_0^*-modules. In particular, $V \cong V'$ in $\mathscr{C}$ if and only if they are isomorphic as H_0^*-modules. If $\langle \cdot , \cdot \rangle : A \otimes B \to \mathbb{F}$ is an H_0-invariant pairing, then for all $a \in A$, $b \in B$ and $f \in H_0^*$ we have

$$
\begin{aligned}
\text{(6-1)} \qquad \langle f \triangleright a, b \rangle &= f(a^{(-1)}) \langle a^{(0)}, b \rangle \\
&= f(b^{(-1)}) \langle a, b^{(0)} \rangle = \langle a, f \triangleright b \rangle.
\end{aligned}
$$

Finally, note that V is a simple H_0-comodule if and only if it is simple as a left H_0^*-module.

Proposition 6.5. *Let A and B be simple objects in $\mathscr{C}$. Let $\langle \cdot , \cdot \rangle : A \otimes B \to \mathbb{F}$ be a nonzero H_0-invariant pairing. Then $A \cong B$ in $\mathscr{C}$.*

Proof. Given $a \in A$, let $J_a = \mathrm{Ann}_{H_0^*} a = \{f \in H_0^* : f \triangleright a = 0\}$. We need the following technical result.

Lemma 6.6. *Let A, B be objects in $\mathscr{C}$ and let $\langle \cdot , \cdot \rangle : A \otimes B \to \mathbb{F}$ be an H_0-invariant pairing such that ${}^{\perp}B = 0$. If B is simple, then $J_a \subset \mathrm{Ann}_{H_0^*} B$ for all $a \in A$, $a \neq 0$. Moreover, if A is also simple, then $J_a = \mathrm{Ann}_{H_0^*} A$.*

Proof. Let $a \in A$, $a \neq 0$ and take $f \in J_a$. It follows from (6-1) that for all $b \in B$, $0 = \langle f \triangleright a, b \rangle = \langle a, f \triangleright b \rangle$. Thus, $\langle a, J_a \triangleright B \rangle = 0$, hence $a \in {}^{\perp}(J_a \triangleright B)$. Since ${}^{\perp}B = 0$, this implies that $J_a \triangleright B$ is a proper H_0^*-submodule of B, hence $J_a \triangleright B = 0$ by the simplicity of B.

Suppose now that A is also simple. Then J_a is a maximal left ideal for all $a \neq 0$. If $J_a \neq J_{a'}$ for some $a, a' \in A$ then $J_a + J_{a'} = H_0^* \ni 1$, hence $B = 0$, which contradicts the simplicity of B. Thus, $J_a = J_{a'}$ for all $a, a' \in A$ and so

$$
\mathrm{Ann}_{H_0^*} A = \bigcap_{a' \in A} J_{a'} = J_a. \qquad \square
$$

Since A, B are simple and the form $\langle \cdot , \cdot \rangle$ is H_0-invariant and nonzero, ${}^{\perp}B = 0$ by Lemma 6.4. Then $\mathrm{Ann}_{H_0^*} A \subset \mathrm{Ann}_{H_0^*} B$ by Lemma 6.6. Let $R = H_0^* / \mathrm{Ann}_{H_0^*} A$. Then both A and B are R-modules in a natural way and are simple as such. Moreover, $A \cong B$ as H_0-comodules if and only if $A \cong B$ as R-modules. Furthermore, by definition of R and Lemma 6.6 every nonzero element of R acts on A by an injective $\mathbb{F}$-linear endomorphism. Since A is a simple H_0-comodule, it is finite dimensional (see, e.g., [Montgomery 1993, Corollary 5.1.2]). Thus, each nonzero element of R acts on A by an $\mathbb{F}$-automorphism. This implies that R is a division algebra, hence admits a unique, up to an isomorphism, simple finite dimensional module, and so $A \cong B$ as R-modules. Therefore, $A \cong B$ as objects in $\mathscr{C}$. □

Remark 6.7. It can be shown that R is a field, since for all $f, g \in H_0^*$ we have

$$\langle fg \triangleright a, b\rangle = \langle g \triangleright a, f \triangleright b\rangle = \langle a, (gf) \triangleright b\rangle = \langle gf \triangleright a, b\rangle.$$

Hence, since both A and B are simple, $fg - gf \in \operatorname{Ann}_{H_0{}^*} A$.

Denote by $\mathscr{C}^f$ the full subcategory of $\mathscr{C}$ whose objects are direct sums of simple comodules with finite multiplicities. Thus, an object V of $\mathscr{C}^f$ can be written as $V = \bigoplus_{\boldsymbol{i}\in \boldsymbol{I}} V_{\boldsymbol{i}}$ where each $V_{\boldsymbol{i}}$ is a finite direct sum of isomorphic simple subcomodules of V, and hence by [Montgomery 1995] is finite dimensional.

Lemma 6.8. *Suppose that $V = \bigoplus_{\boldsymbol{i}\in \boldsymbol{I}} V_{\boldsymbol{i}} \in \mathscr{C}^f$ admits an H_0-invariant bilinear form $\langle \cdot, \cdot \rangle : V \otimes V \to \mathbb{F}$. Then for any subobject U of V in $\mathscr{C}$,*

$$U^\perp \supset \bigoplus_{\boldsymbol{i}\in \boldsymbol{I}} U_{\boldsymbol{i}}^\perp,$$

where $U_{\boldsymbol{i}}^\perp = \{v \in V_{\boldsymbol{i}} : \langle U \cap V_{\boldsymbol{i}}, v\rangle = 0\}$.

Proof. By Proposition 6.5, $\langle V_{\boldsymbol{i}}, V_{\boldsymbol{j}}\rangle = 0$ if $\boldsymbol{i} \neq \boldsymbol{j}$. The assertion is now immediate. □

6.3. *Quasiprimitive generators.*

Definition 6.9. Let $(A, \cdot, 1)$ be a unital algebra and (B, Δ, ε) be a coalgebra in $\mathscr{C}$. We say that an H_0-invariant pairing $\langle \cdot, \cdot \rangle : A \otimes B \to \mathbb{F}$ is compatible with $(A, \cdot, 1)$ and (B, Δ, ε) if

$$\langle a \cdot a', b\rangle = \langle a \otimes a', \Delta(b)\rangle, \quad \varepsilon(b) = \langle 1, b\rangle$$

for all $a, a' \in A$, $b \in B$, where $\langle \cdot, \cdot \rangle : (A \otimes A) \otimes (B \otimes B) \to \mathbb{F}$ is defined by

$$\langle a \otimes a', b \otimes b'\rangle = \langle a, b'\rangle\langle a', b\rangle.$$

The main ingredient in our proof of Theorem 2.18 is the following result.

Theorem 6.10. *Let A be an algebra (denoted by $(A, \cdot, 1)$) and a coalgebra (denoted by (A, Δ, ε)) in $\mathscr{C}^f$. Let $\langle \cdot, \cdot \rangle : A \otimes A \to \mathbb{F}$ be a compatible pairing between $(A, \cdot, 1)$ and (A, Δ, ε) satisfying $\langle a, a\rangle \neq 0$ for all $a \in A \setminus \{0\}$. Then $(A, \cdot, 1)$ is generated by $A_1 = \operatorname{QPrim}(A, \Delta, \varepsilon)$.*

Proof. Let B be a $\mathscr{C}$-subalgebra of A. Since $A \in \mathscr{C}^f$ and B is its subobject, $A = \bigoplus_{\boldsymbol{i}} A_{\boldsymbol{i}}$ and $B = \bigoplus_{\boldsymbol{i}} B_{\boldsymbol{i}}$ where $B_{\boldsymbol{i}} = A_{\boldsymbol{i}} \cap B$. By Proposition 6.5, $\langle A_{\boldsymbol{i}}, A_{\boldsymbol{j}} \rangle = 0$ for all $\boldsymbol{i} \neq \boldsymbol{j}$. We claim that $B^\perp$ is a coideal of A in $\mathscr{C}$.

Indeed, for any $\boldsymbol{i}$, $\boldsymbol{j}$ we have

$$\{0\} = \langle B_{\boldsymbol{i}} \cdot B_{\boldsymbol{j}}, B^\perp \rangle = \langle B_{\boldsymbol{i}} \otimes B_{\boldsymbol{j}}, \Delta(B^\perp) \rangle.$$

Thus, $\Delta(B^\perp) \subset \bigoplus_{\boldsymbol{i},\boldsymbol{j}} (B_{\boldsymbol{i}} \otimes B_{\boldsymbol{j}})^\perp$ where $(B_{\boldsymbol{i}} \otimes B_{\boldsymbol{j}})^\perp = \{z \in A_{\boldsymbol{j}} \otimes A_{\boldsymbol{i}} : \langle B_{\boldsymbol{i}} \otimes B_{\boldsymbol{j}}, z \rangle = 0\}$. We need the following simple fact from linear algebra.

Lemma 6.11. *Let U, V be finite dimensional vector spaces over $\mathbb{F}$ and $U' \subset U$, $V' \subset V$ their subspaces. Then*:

(a) $U' \otimes V' = (U' \otimes V) \cap (U \otimes V')$;

(b) *For any subspaces V_1, V_2 of V,*

$$(V_1 \cap V_2)^\perp = V_1^\perp + V_2^\perp, \quad V_1^\perp \cap V_2^\perp = (V_1 + V_2)^\perp,$$

where $W^\perp = \{f \in V^* : f(W) = 0\}$ for any subspace $W \subset V$;

(c) $(U' \otimes V')^\perp = V'^\perp \otimes U^* + V^* \otimes U'^\perp$, *where we canonically identify* $(U \otimes V)^*$ *with* $V^* \otimes U^*$.

Proof. Parts (a) and (b) are easily checked. For (c), note that $(U' \otimes V)^\perp = V^* \otimes U'^\perp$ and $(U \otimes V')^\perp = V'^\perp \otimes U^*$. Hence by parts (a), and (b)

$$\begin{aligned}(U' \otimes V')^\perp &= ((U' \otimes V) \cap (U \otimes V'))^\perp \\ &= (U' \otimes V)^\perp + (U \otimes V')^\perp = V^* \otimes U'^\perp + V'^\perp \otimes U^*. \qquad \square\end{aligned}$$

Since $A_{\boldsymbol{k}}$ is finite dimensional and the restriction of $\langle \cdot, \cdot \rangle$ to $A_{\boldsymbol{k}}$ is nondegenerate, we naturally identify $A_{\boldsymbol{k}}^*$ with $A_{\boldsymbol{k}}$ via $a \mapsto f_a$, where $f_a(a') = \langle a', a \rangle$. Then, applying Lemma 6.11(c) with $U = A_{\boldsymbol{i}}$, $V = A_{\boldsymbol{j}}$, $U' = B_{\boldsymbol{i}}$ and $V' = B_{\boldsymbol{j}}$, we obtain

$$(B_{\boldsymbol{i}} \otimes B_{\boldsymbol{j}})^\perp = A_{\boldsymbol{j}} \otimes B_{\boldsymbol{i}}^\perp + B_{\boldsymbol{j}}^\perp \otimes A_{\boldsymbol{i}}, \tag{6-2}$$

where $B_{\boldsymbol{k}}^\perp = \{a \in A_{\boldsymbol{k}} : \langle B_{\boldsymbol{k}}, a \rangle = 0\}$. We conclude that

$$\Delta(B^\perp) \subset \bigoplus_{\boldsymbol{i},\boldsymbol{j}} (B_{\boldsymbol{i}} \otimes B_{\boldsymbol{j}})^\perp \subset A \otimes B^\perp + B^\perp \otimes A.$$

To complete the proof of the claim, observe that $\varepsilon(B^\perp) = \langle 1, B^\perp \rangle = 0$.

Now we complete the proof of Theorem 6.10. Let B be the subalgebra of A in $\mathscr{C}$ generated by the subobject $A_1 = \mathrm{QPrim}_{\mathscr{C}}(A)$ of A, and suppose that $B \neq A$. Then, by the above claim, the orthogonal complement $I = B^\perp$ is a coideal of A in $\mathscr{C}$. By Lemma 6.8, $I \supset \bigoplus_{\boldsymbol{i}} B_{\boldsymbol{i}}^\perp \neq \{0\}$ because $B_{\boldsymbol{i}} \neq A_{\boldsymbol{i}}$ for some $\boldsymbol{i}$. Therefore, $I \neq \{0\}$ and so $I \cap A_1 \neq \{0\}$ by Proposition 6.2. Yet $I \cap B = \{0\}$ since $\langle x, x \rangle \neq 0$ for all $x \in A$, hence $I \cap A_1 = \{0\}$ and we obtain a contradiction. Thus, $B = A$. $\square$

6.4. ***Proof of Theorems 1.2 and 2.18.*** Let $\mathbb{F} = \mathbb{Q}$ and define Green's pairing $\langle\,\cdot\,,\,\cdot\,\rangle : H_{\mathscr{A}} \otimes H_{\mathscr{A}} \to \mathbb{Q}$ (cf. [Green 1995]) by

$$\langle [A], [B] \rangle = \frac{\delta_{[A],[B]}}{|\operatorname{Aut}_{\mathscr{A}}(A)|} \tag{6-3}$$

for any $[A], [B] \in \operatorname{Iso}\mathscr{A}$.

Clearly, this pairing is positive definite and symmetric. We extend $\langle\,\cdot\,,\,\cdot\,\rangle$ to a symmetric bilinear form on $H_{\mathscr{A}} \otimes H_{\mathscr{A}}$ by

$$\langle [A] \otimes [B], [C] \otimes [D] \rangle = \langle [A], [D] \rangle \langle [B], [C] \rangle$$

for any $[A], [B], [C], [D] \in \operatorname{Iso}\mathscr{A}$.

Lemma 6.12. *Let $\mathscr{A}$ be a cofinitary category. Then* (6-3) *is a compatible pairing* (*in the sense of Definition 6.9*) *between the Hall algebra $H_{\mathscr{A}}$ and the coalgebra* $(H_{\mathscr{A}}, \Delta, \varepsilon)$.

Proof. We abbreviate $\Gamma = \Gamma_{\mathscr{A}}$ and let $\mathscr{C} = \mathscr{C}_\Gamma$ be the category of Γ-graded vector spaces or, equivalently, $\mathbb{Q}\Gamma$-comodules (cf. Example 6.3). It follows immediately from Example 6.3 that the pairing (6-3) is $\mathbb{Q}\Gamma$-invariant.

It remains to prove the compatibility in the sense of Definition 6.9, that is,

$$\langle [A] \cdot [B], [C] \rangle = \langle [A] \otimes [B], \Delta([C]) \rangle$$

for all $[A], [B], [C] \in \operatorname{Iso}\mathscr{A}$. Indeed,

$$\begin{aligned}
\langle [A] \cdot [B], [C] \rangle = \frac{F^C_{A,B}}{|\operatorname{Aut}_{\mathscr{A}}(C)|} &= \frac{F^{B,A}_C}{|\operatorname{Aut}_{\mathscr{A}}(B)|\,|\operatorname{Aut}_{\mathscr{A}}(A)|} \\
&= \sum_{[A'],[B']} F^{B',A'}_C \langle [A], [A'] \rangle \langle [B], [B'] \rangle \\
&= \sum_{[B'],[A']} F^{B',A'}_C \langle [A] \otimes [B], [B'] \otimes [A'] \rangle \\
&= \langle [A] \otimes [B], \Delta([C]) \rangle.
\end{aligned}$$

□

Proof of Theorems 1.2 and 2.18. Suppose that $\mathscr{A}$ is profinitary and cofinitary. Since for each $\gamma \in \Gamma = \Gamma_{\mathscr{A}}$, $(H_{\mathscr{A}})_\gamma$ is finite dimensional and hence is a finite direct sum of isomorphic simple left $\mathbb{Q}\Gamma$-comodules, $H_{\mathscr{A}} \in \mathscr{C}^f_\Gamma$. Then, clearly, $A = H_{\mathscr{A}}$ and the pairing (6-3) satisfy all the assumptions of Theorem 6.10. Therefore, $H_{\mathscr{A}}$ is generated by $A_1 = \operatorname{QPrim}(H_{\mathscr{A}}, \Delta, \varepsilon)$ in $\mathscr{C}_\Gamma$.

Our next step is to show that $A_1 = \operatorname{Prim}(H_{\mathscr{A}}, \Delta, \varepsilon)$, which gives the first assertion of Main Theorem 1.2. For that, we need the following result.

Lemma 6.13. *Let $C = \bigoplus_{\gamma\in\Gamma} C_\gamma$ be a coalgebra in the category $\mathscr{C}_\Gamma$. Assume that for every $\gamma \in \Gamma^+$, there exists $h_\gamma \in \mathbb{Z}_{>0}$ such that γ cannot be written as a sum of* *more than h_γ elements of* Γ^+. *Then* $\mathrm{Corad}_{\mathscr{C}}(C) \subset C_{\mathbf{0}}$ *where* $\mathbf{0}$ *is the zero element of* Γ.

Proof. First, observe that $\mathbf{0}$ is the only invertible element of Γ, since otherwise $\mathbf{0} = \alpha + \beta$ for some $\alpha, \beta \in \Gamma^+$ and so $\alpha = (n+1)\alpha + n\beta$ for any $n \in \mathbb{Z}_{>0}$, which is a contradiction. Since for any subcoalgebra $D = \bigoplus_{\gamma \in \Gamma} D_\gamma$ of C in $\mathscr{C}_\Gamma$

$$\Delta(D_\gamma) \subset \bigoplus_{\gamma', \gamma'' \in \Gamma \,:\, \gamma = \gamma' + \gamma''} D_{\gamma'} \otimes D_{\gamma''},$$

it follows that $\Delta(D_{\mathbf{0}}) \subset D_{\mathbf{0}} \otimes D_{\mathbf{0}}$. Therefore, $D_{\mathbf{0}}$ is a subcoalgebra of D.

We claim that $D = 0$ if and only if $D_{\mathbf{0}} = 0$. Indeed, if $D_{\mathbf{0}} = 0$, since for the k-th iterated comultiplication Δ^k we have

$$\Delta^k(D_\gamma) \subset \sum_{\gamma_0, \dots, \gamma_k \in \Gamma \,:\, \gamma_0 + \cdots + \gamma_k = \gamma} D_{\gamma_0} \otimes \cdots \otimes D_{\gamma_k},$$

it follows that $\Delta^{h_\gamma}(D_\gamma) = 0$, since then in each summand we must have $\gamma_i = 0$ for some $0 \le i \le h_\gamma$ by the assumptions of the lemma. This implies that $D_\gamma = 0$ for all $\gamma \in \Gamma$, hence $D = 0$. The converse is obvious.

Thus, if D is a simple subcoalgebra of C, then $D_{\mathbf{0}} \ne 0$ and so $D = D_{\mathbf{0}}$. □

By Lemma 5.6, $\Gamma_{\mathscr{A}}$ satisfies the assumptions of Lemma 6.13, with $h_\gamma \le s_\gamma$, hence $\mathrm{Corad}_{\mathscr{C}}(H_{\mathscr{A}}) = \mathbb{Q}$ and $\mathrm{QPrim}_{\mathscr{C}}(H_{\mathscr{A}}) = \mathbb{Q} \oplus \mathrm{Prim}(H_{\mathscr{A}})$. This proves the first assertion of Main Theorem 1.2. It remains to prove the second assertion (and thus complete the proof of Theorem 2.18), namely, that $\mathrm{Prim}(H_{\mathscr{A}})$ is a minimal generating space of $H_{\mathscr{A}}$. We need the following result.

Lemma 6.14. *Suppose A is both a unital algebra and coalgebra with* $\Delta(1) = 1 \otimes 1$. *Assume that A admits a compatible pairing* $\langle \cdot\,, \cdot \rangle : A \otimes A \to \mathbb{F}$, *in the sense of Definition 6.9, such that* $\langle a, 1 \rangle = \varepsilon(a)$ *for all* $a \in A$. *Let* $V = \mathrm{Prim}(A)$. *Then* $1 \notin V$ *and* $\left\langle \sum_{k \ge 2} V^k, \mathbb{F} \oplus V \right\rangle = 0$.

Proof. Since $v \in V$ is primitive, $\varepsilon(v) = 0$. Furthermore, we show that $\varepsilon : A \to \mathbb{F}$ is a homomorphism of algebras. Indeed, given $a, a' \in A$, we have

$$\varepsilon(aa') = \langle aa', 1 \rangle = \langle a \otimes a', \Delta(1) \rangle = \langle a \otimes a', 1 \otimes 1 \rangle = \langle a', 1 \rangle \langle a, 1 \rangle = \varepsilon(a)\varepsilon(a').$$

This immediately implies that $\varepsilon(V^\ell) = 0$ and $\langle V^\ell, V^0 \rangle = 0$, $\ell > 0$. Finally, let $v \in V$ and $x, y \in \ker \varepsilon$. Then

$$\langle xy, v \rangle = \langle x \otimes y, \Delta(v) \rangle = \langle x \otimes y, v \otimes 1 + 1 \otimes v \rangle = \langle y, v \rangle \varepsilon(x) + \varepsilon(y) \langle x, v \rangle = 0. \tag{6-4}$$

Let $\ell > 1$. Since $V^\ell \subset V \cdot V^{\ell-1}$ and $V^k \subset \ker \varepsilon$ for all $k > 0$, it follows that $\langle V^\ell, V \rangle = 0$. □

Let $V_{\mathscr{A}} = \operatorname{Prim}(H_{\mathscr{A}})$ and $(H_{\mathscr{A}})_{>1} = \sum_{r\ge 2} V_{\mathscr{A}}^r$. From Lemma 6.14, we have $\langle (H_{\mathscr{A}})_{>1}, \mathbb{Q}\oplus V_{\mathscr{A}}\rangle = 0$. Since the pairing $\langle\,\cdot\,,\cdot\,\rangle$ on $H_{\mathscr{A}}$ is symmetric positive definite, $(H_{\mathscr{A}})_{>1}\cap(\mathbb{Q}\oplus V_{\mathscr{A}}) = \{0\}$, hence the sum $(\mathbb{Q}\oplus V_{\mathscr{A}}) + (H_{\mathscr{A}})_{>1}$ is direct. This proves the second assertion of Main Theorem 1.2 and completes the proof of Theorem 2.18. □

6.5. *Proof of Corollary 2.19 and estimates for primitive elements.*

Proof of Corollary 2.19. Let $H_{\mathscr{A}}^+ = \ker\varepsilon$ and let $R\subset H_{\mathscr{A}}^+$ be a generating space for $H_{\mathscr{A}}$. Then $(H_{\mathscr{A}}^+)^\ell = \sum_{k\ge\ell} R^k$, $\ell\ge 1$. Taking $R = \mathbb{Q}\operatorname{Ind}\mathscr{A}$ (Theorem 1.1) and $R = \operatorname{Prim}(H_{\mathscr{A}})$ (Theorem 2.18) we conclude that

$$P = (H_{\mathscr{A}}^+)^2 = \sum_{k\ge 2}\operatorname{Prim}(H_{\mathscr{A}})^k = \sum_{k\ge 2}(\mathbb{Q}\operatorname{Ind}\mathscr{A})^k.$$

On the other hand, $H_{\mathscr{A}}^+ = \operatorname{Prim}(H_{\mathscr{A}}) + P$ and $P\cap\operatorname{Prim}(H_{\mathscr{A}}) = \{0\}$ by Lemma 6.14. Therefore, $H_{\mathscr{A}}^+ = \operatorname{Prim}(H_{\mathscr{A}})\oplus P$. The graded version is immediate. □

Proof of Proposition 2.20 and Lemma 2.21. We need the following obvious fact from linear algebra.

Lemma 6.15. *Let U be a finite dimensional $\mathbb{F}$-vector space and U_1, U_1', U_2 be subspaces of U such that $U = U_1 + U_2 = U_1' + U_2$. If $U_1\cap U_2 = \{0\}$, then* $\dim_{\mathbb{F}} U_1 = \dim_{\mathbb{F}} U_1' - \dim_{\mathbb{F}}(U_1'\cap U_2)$.

Taking into account Corollary 2.19, we apply this lemma with $U = (H_{\mathscr{A}})_\gamma$, $U_1' = \mathbb{Q}\operatorname{Ind}\mathscr{A}_\gamma$, $U_2 = P_\gamma$ and $U_1 = \operatorname{Prim}(H_{\mathscr{A}})_\gamma$ to obtain

$$\dim_{\mathbb{Q}}\operatorname{Prim}(H_{\mathscr{A}})_\gamma = \#\operatorname{Ind}\mathscr{A}_\gamma - \dim_{\mathbb{Q}}(P_\gamma\cap\mathbb{Q}\operatorname{Ind}\mathscr{A}_\gamma),$$

which yields Proposition 2.20.

To prove Lemma 2.21, note that $\mathbb{Q}(\operatorname{Iso}\mathscr{A}\setminus\operatorname{Ind}\mathscr{A}) = (\mathbb{Q}\operatorname{Ind}\mathscr{A})^\perp$. Thus,

$$\begin{aligned}\operatorname{Prim}(H_{\mathscr{A}})_\gamma\cap\mathbb{Q}(\operatorname{Iso}\mathscr{A}\setminus\operatorname{Ind}\mathscr{A}) &\subset (\mathbb{Q}\operatorname{Ind}\mathscr{A}_\gamma)^\perp\cap P_\gamma^\perp\\ &= (\mathbb{Q}\operatorname{Ind}\mathscr{A}_\gamma + P_\gamma)^\perp = (H_{\mathscr{A}})_\gamma^\perp = 0\end{aligned}$$

by Lemma 6.11(b) and Corollary 2.19. □

7. Proof of Theorem 2.26

7.1. ***Diagonally braided categories.*** We call a bialgebra H_0 coquasitriangular if it has a skew Hopf self-pairing $\mathcal{R}: H_0\otimes H_0\to\mathbb{Q}$. Let $\mathscr{C}$ be the category of left H_0-comodules. This category is braided via the commutativity constraint $\Psi_{U,V}: U\otimes V\to V\otimes U$ for all objects U, V of $\mathscr{C}$ defined by

$$\Psi_{U,V}(u\otimes v) = \mathcal{R}(u^{(-1)}, v^{(-1)})\cdot v^{(0)}\otimes u^{(0)}$$

for all $u \in U$, $v \in V$, where we use the Sweedler-like notation for the coactions $\delta_U(u) = u^{(-1)} \otimes u^{(0)}$ and $\delta_V(v) = v^{(-1)} \otimes v^{(0)}$. We will write $\mathscr{C}_{\mathcal{R}}$ to emphasize that $\mathscr{C}$ is a braided category.

Remark 7.1. The category $\mathscr{C}_\chi$ introduced in Lemma 2.22 is equivalent to the category of H_0-comodules, where $H_0 = \mathbb{Q}\Gamma$ is the monoidal algebra of Γ and $\mathcal{R}|_{\Gamma\times\Gamma} = \chi$.

Our present aim is to prove the following result.

Theorem 7.2. *Let B be a bialgebra in $\mathscr{C}_{\mathcal{R}}$.*

(a) *The space $V = \mathrm{Prim}(B)$ is a subobject of B in $\mathscr{C}_{\mathcal{R}}$.*

(b) *Suppose that B admits a compatible pairing, in the sense of Definition 6.9, such that $\langle b, 1\rangle = \varepsilon(b)$ and $\langle b, b\rangle \neq 0$ for all $b \in B \setminus \{0\}$. Then the canonical inclusion $V \hookrightarrow B$ extends to an injective homomorphism*

$$\mathcal{B}(V) \to B \tag{7-1}$$

of bialgebras in $\mathscr{C}_{\mathcal{R}}$. In particular, if B is generated by V, then (7-1) *is an isomorphism.*

Proof. Part (a) is a special case of the following simple fact.

Lemma 7.3. *If C is a coalgebra in $\mathscr{C}_{\mathcal{R}}$ with unity, then $V := \mathrm{Prim}(C)$ is a subobject of C in $\mathscr{C}_{\mathcal{R}}$.*

Proof. Denote by $\delta_C : C \to H_0 \otimes C$ the left coaction of H_0 on C. All we have to show is that $\delta_C(V) \subset H_0 \otimes V$. Fix a basis $\{b_i\}$ of H_0 and let $v \in \mathrm{Prim}(C)$. Write

$$\delta_C(v) = \sum_i b_i \otimes v_i, \quad v_i \in C.$$

Since $\Delta : C \to C \otimes C$ is a morphism of left H_0-comodules,

$$(1 \otimes \Delta) \circ \delta_C(v) = \delta_C(v \otimes 1) + \delta_C(1 \otimes v).$$

Taking into account that $\delta_C(1) = 1 \otimes 1$, we obtain

$$\sum_i b_i \otimes \Delta(v_i) = \sum_i b_i \otimes v_i \otimes 1 + \sum_i b_i \otimes 1 \otimes v_i \ ,$$

which implies that

$$\Delta(v_i) = v_i \otimes 1 + 1 \otimes v_i,$$

that is, $v_i \in V$ for all i. □

Now we prove (b). Denote by B' the subalgebra of B generated by $V = \mathrm{Prim}(B)$. It is sufficient to show that $B' = \mathcal{B}(V)$. We need the following result.

Proposition 7.4. $B' = \bigoplus_{k\geq 0} V^k$, *hence B' is a graded algebra.*

Proof. We prove that $\langle V^\ell, V^k\rangle = 0$ for all $0 \le k < \ell$ by induction on the pairs (k, ℓ) with $k < \ell$, ordered lexicographically. The induction base for $k = 0, 1$ is established in Lemma 6.14. Now, fix $\ell > 2$ and suppose that $\langle V^s, V^r\rangle = 0$ for all $r < s < \ell$. Let $1 < k < \ell$. Since Δ is a homomorphism of algebras,

$$\Delta(V^k) \subset (V \otimes 1 + 1 \otimes V)^k \subset \sum_{i=0}^{k} V^{k-i} \otimes V^i,$$

hence

$$\begin{aligned}\langle V^\ell, V^k\rangle \subset \langle V \otimes V^{\ell-1}, \Delta(V^k)\rangle &\subset \sum_{i=0}^{k}\langle V \otimes V^{\ell-1}, V^{k-i} \otimes V^i\rangle \\ &= \sum_{i=0}^{k}\langle V, V^i\rangle\langle V^{\ell-1}, V^{k-i}\rangle = \langle V, V\rangle\langle V^{\ell-1}, V^{k-1}\rangle = \{0\}\end{aligned}$$

by the inductive hypothesis. It remains to show that the sum $\sum_{k\ge 0} V^k$ is direct, which is an immediate consequence of the following obvious fact.

Lemma 7.5. *Let U_i, $i \in \mathbb{Z}_{\ge 0}$, be subspaces of an $\mathbb{F}$-vector space U with a bilinear form $\langle\,\cdot\,,\cdot\,\rangle : U \otimes U \to \mathbb{F}$ such that $\langle U_j, U_i\rangle = 0$ if $j > i$ and $\langle u, u\rangle \ne 0$ for all $u \in U \setminus \{0\}$. Then the sum $\sum_i U_i$ is direct.* □

This completes the proof of Proposition 7.4. □

Since $B'_0 = \mathbb{Q}$ and $B'_1 = V = \mathrm{Prim}(B') = \mathrm{Prim}(B)$, B' is the Nichols algebra of V by Definition 2.23. Theorem 7.2 is therefore proved. □

7.2. *Proof of Theorem 2.26.* We need the following reformulation of Green's celebrated theorem for Hall algebras ([Green 1995]; see also [Walker 2011]).

Proposition 7.6. *Let $\mathscr{A}$ be a finitary and cofinitary hereditary abelian category. Then the Hall algebra $H_{\mathscr{A}}$ is a bialgebra in $\mathscr{C}_{\chi_{\mathscr{A}}}$ with the coproduct Δ given by* (2-3) *and the counit ε given by* (2-4).

Proof. For every $[C], [C'] \in \mathrm{Iso}\,\mathscr{A}$ we have

$$\begin{aligned}\Delta([C])\Delta([C']) &= \Bigl(\sum_{[A],[B]} F_C^{A,B}\cdot[A]\otimes[B]\Bigr)\Bigl(\sum_{[A'],[B']} F_{C'}^{A',B'}\cdot[A']\otimes[B']\Bigr)\\ &= \sum_{[A],[B],[A'],[B']} F_C^{A,B}F_{C'}^{A',B'}\cdot\frac{|\mathrm{Ext}^1_{\mathscr{A}}(B, A')|}{|\mathrm{Hom}_{\mathscr{A}}(B, A')|}[A][A']\otimes[B][B']\\ &= \sum_{\substack{[A],[B],[A'],\\ [B'],[A''],[B'']}} F_C^{A,B}F_{C'}^{A',B'}F_{A,A'}^{A''}F_{B,B'}^{B''}\frac{|\mathrm{Ext}^1_{\mathscr{A}}(B, A')|}{|\mathrm{Hom}_{\mathscr{A}}(B, A')|}\cdot[A'']\otimes[B'']\end{aligned}$$

On the other hand,

$$\Delta([C][C']) = \sum_{[C'']} F_{C,C'}^{C''} \Delta([C'']) = \sum_{[C''],[A''],[B'']} F_{C,C'}^{C''} F_{C''}^{A'',B''} \cdot [A''] \otimes [B''].$$

We need the following lemma.

Lemma 7.7 ([Green 1995, Theorem 2], see also [Schiffmann 2012]). *If $\mathscr{A}$ is a finitary and cofinitary hereditary abelian category, then for any objects A'', B'', C, C' of $\mathscr{A}$ one has*

$$\sum_{[A],[A'],[B],[B'] \in \operatorname{Iso} \mathscr{A}} \frac{|\operatorname{Ext}^1_{\mathscr{A}}(B, A')|}{|\operatorname{Hom}_{\mathscr{A}}(B, A')|} \cdot F_{B,B'}^{B''} F_{A,A'}^{A''} F_C^{A,B} F_{C'}^{A',B'} = \sum_{[C''] \in \operatorname{Iso} \mathscr{A}} F_{C,C'}^{C''} F_{C''}^{A'',B''} \tag{7-2}$$

This immediately implies that $\Delta([C])\Delta([C']) = \Delta([C][C'])$. □

Theorem 2.26 now follows from Proposition 7.6 and Theorem 7.2. □

Acknowledgments

We are grateful to Bernhard Keller, Dylan Ruppel and Vadim Vologodsky for stimulating discussions. An important part of this work was done during Berenstein's visit to the MSRI in the framework of the "Cluster algebras" program and he thanks the Institute and the organizers for their hospitality and support. Greenstein is indebted to the Department of Mathematics of Uppsala University and especially Volodymyr Mazorchuk for their hospitality and support. In the final stage of work on this paper we benefited from the hospitality of Institut des Hautes Études Scientiques (IHES), which we gratefully acknowledge.

References

[Andruskiewitsch and Schneider 2002] N. Andruskiewitsch and H.-J. Schneider, "Pointed Hopf algebras", pp. 1–68 in *New directions in Hopf algebras*, edited by S. Montgomery and H.-J. Schneider, Mathematical Sciences Research Institute Publications **43**, Cambridge University Press, 2002. MR 2003e:16043 Zbl 1011.16025

[Auslander et al. 1995] M. Auslander, I. Reiten, and S. O. Smalø, *Representation theory of Artin algebras*, Cambridge Studies in Advanced Mathematics **36**, Cambridge University Press, 1995. MR 96c:16015 Zbl 0834.16001

[Baumann and Kassel 2001] P. Baumann and C. Kassel, "The Hall algebra of the category of coherent sheaves on the projective line", *J. Reine Angew. Math.* **533** (2001), 207–233. MR 2002g:14019 Zbl 0967.18005

[Berenstein and Greenstein 2013] A. Berenstein and J. Greenstein, "Quantum Chevalley groups", pp. 71–102 in *Noncommutative birational geometry, representations and combinatorics*, edited by

A. Berenstein and V. Retakh, Contemporary Mathematics **592**, American Mathematical Society, Providence, RI, 2013. MR 3087940 Zbl 1321.18006

[Buchsbaum 1959] D. A. Buchsbaum, "A note on homology in categories", *Ann. of Math.* (2) **69** (1959), 66–74. MR 25 #3974 Zbl 0084.26801

[Bühler 2010] T. Bühler, "Exact categories", *Expo. Math.* **28**:1 (2010), 1–69. MR 2011e:18020 Zbl 1192.18007

[Burban and Schiffmann 2012] I. Burban and O. Schiffmann, "On the Hall algebra of an elliptic curve, I", *Duke Math. J.* **161**:7 (2012), 1171–1231. MR 2922373 Zbl 1286.16029

[Deng and Ruan 2015] B. Deng and S. Ruan, "Hall polynomials for tame type", preprint, 2015. arXiv 1512.03504

[Deng and Xiao 2003] B. Deng and J. Xiao, "A new approach to Kac's theorem on representations of valued quivers", *Math. Z.* **245**:1 (2003), 183–199. MR 2004k:16032 Zbl 1053.16010

[Dlab and Ringel 1976] V. Dlab and C. M. Ringel, *Indecomposable representations of graphs and algebras*, Memoirs of the American Mathematical Society **6**:173, American Mathematical Society, Providence, RI, 1976. MR 56 #5657 Zbl 0332.16015

[Gabriel 1973] P. Gabriel, "Indecomposable representations, II", pp. 81–104 in *Symposia Mathematica, XI: Convegno di Algebra Commutativa, INDAM* (Rome, 1971), Academic Press, London, 1973. MR 49 #5132 Zbl 0276.16001

[Green 1995] J. A. Green, "Hall algebras, hereditary algebras and quantum groups", *Invent. Math.* **120**:2 (1995), 361–377. MR 96c:16016 Zbl 0836.16021

[Guo and Peng 1997] J. Y. Guo and L. Peng, "Universal PBW-basis of Hall–Ringel algebras and Hall polynomials", *J. Algebra* **198**:2 (1997), 339–351. MR 99b:16033 Zbl 0893.16005

[Hausel 2010] T. Hausel, "Kac's conjecture from Nakajima quiver varieties", *Invent. Math.* **181**:1 (2010), 21–37. MR 2011d:14033 Zbl 1198.16016

[Hua 2000] J. Hua, "Counting representations of quivers over finite fields", *J. Algebra* **226**:2 (2000), 1011–1033. MR 2001j:16017 Zbl 0972.16006

[Hubery 2004] A. Hubery, "The composition algebra of an affine quiver", preprint, 2004. arXiv math/0403206

[Hubery 2005] A. Hubery, "Symmetric functions and the centre of the Ringel–Hall algebra of a cyclic quiver", *Math. Z.* **251**:3 (2005), 705–719. MR 2006i:16019 Zbl 1084.16010

[Hubery 2006] A. Hubery, "From triangulated categories to Lie algebras: a theorem of Peng and Xiao", pp. 51–66 in *Trends in representation theory of algebras and related topics*, edited by J. A. de la Peña and R. Bautista, Contemporary Mathematics **406**, American Mathematical Society, Providence, RI, 2006. MR 2008f:16033 Zbl 1107.16021

[Hubery 2007] A. Hubery, "Ringel–Hall algebras", lecture notes, 2007, available at http://wwwmath.uni-paderborn.de/~hubery/RHAlg.pdf.

[Joyce 2006] D. Joyce, "Configurations in abelian categories, I: Basic properties and moduli stacks", *Adv. Math.* **203**:1 (2006), 194–255. MR 2007b:14023 Zbl 1102.14009

[Joyce 2007] D. Joyce, "Configurations in abelian categories, II: Ringel–Hall algebras", *Adv. Math.* **210**:2 (2007), 635–706. MR 2008f:14022 Zbl 1119.14005

[Kac 1980] V. G. Kac, "Infinite root systems, representations of graphs and invariant theory", *Invent. Math.* **56**:1 (1980), 57–92. MR 82j:16050 Zbl 0427.17001

[Kapranov 1997] M. M. Kapranov, "Eisenstein series and quantum affine algebras", in *Algebraic geometry, 7*, Itogi Nauki i Tekhniki, Seriya Sovremennaya Matematika i Ee Prilozheniya. Tematicheskie Obzory **41**, VINITI, Moscow, 1997. Translated in *J. Math. Sci* (*New York*) **84**:5 (1997), 1311–1360. MR 98i:11033 Zbl 0929.11015

[Kapranov et al. 2012] M. M. Kapranov, O. Schiffmann, and E. Vasserot, "The Hall algebra of a curve", preprint, 2012. arXiv 1201.6185v2

[Keller 1990] B. Keller, "Chain complexes and stable categories", *Manuscripta Math.* **67**:4 (1990), 379–417. MR 91h:18006 Zbl 0753.18005

[Macdonald 1979] I. G. Macdonald, *Symmetric functions and Hall polynomials*, Clarendon/Oxford University Press, New York, 1979. MR 84g:05003 Zbl 0487.20007

[Montgomery 1993] S. Montgomery, *Hopf algebras and their actions on rings*, CBMS Regional Conference Series in Mathematics **82**, American Mathematical Society, Providence, RI, 1993. MR 94i:16019 Zbl 0793.16029

[Montgomery 1995] S. Montgomery, "Indecomposable coalgebras, simple comodules, and pointed Hopf algebras", *Proc. Amer. Math. Soc.* **123**:8 (1995), 2343–2351. MR 95j:16046 Zbl 0836.16024

[Polishchuk and Positselski 2005] A. Polishchuk and L. Positselski, *Quadratic algebras*, University Lecture Series **37**, American Mathematical Society, Providence, RI, 2005. MR 2006f:16043 Zbl 1145.16009

[Quillen 1973] D. Quillen, "Higher algebraic K-theory, I", pp. 85–147 in *Algebraic K-theory, I: Higher K-theories* (Seattle, WA, 1972), edited by H. Bass, Lecture Notes in Mathematics **341**, Springer, Berlin, 1973. MR 49 #2895 Zbl 0292.18004

[Riedtmann 1994] C. Riedtmann, "Lie algebras generated by indecomposables", *J. Algebra* **170**:2 (1994), 526–546. MR 96e:16013 Zbl 0841.16018

[Ringel 1990a] C. M. Ringel, "Hall algebras", pp. 433–447 in *Topics in algebra, part 1* (Warsaw, 1988), edited by S. Balcerzyk et al., Banach Center Publications **26**:1, PWN, Warsaw, 1990. MR 93f:16027 Zbl 0778.16004

[Ringel 1990b] C. M. Ringel, "Hall algebras and quantum groups", *Invent. Math.* **101**:3 (1990), 583–591. MR 91i:16024 Zbl 0735.16009

[Schiffmann 2012] O. Schiffmann, "Lectures on Hall algebras", pp. 1–141 in *Geometric methods in representation theory, II* (Grenoble, 2008), edited by M. Brion, Séminaires et Congrès **24**, part 2, Société Mathématique de France, Paris, 2012. MR 3202707 Zbl 1309.18012 arXiv math.RT/0611617

[Sevenhant and Van Den Bergh 2001] B. Sevenhant and M. Van Den Bergh, "A relation between a conjecture of Kac and the structure of the Hall algebra", *J. Pure Appl. Algebra* **160**:2-3 (2001), 319–332. MR 2002f:17031 Zbl 1025.17006

[Sweedler 1969] M. E. Sweedler, *Hopf algebras*, W. A. Benjamin, New York, 1969. MR 40 #5705 Zbl 0194.32901

[Szántó 2006] C. Szántó, "Hall numbers and the composition algebra of the Kronecker algebra", *Algebr. Represent. Theory* **9**:5 (2006), 465–495. MR 2007g:16024 Zbl 1161.16011

[Walker 2011] C. D. Walker, *A categorification of Hall algebras*, thesis, University of California, Riverside, 2011, available at http://search.proquest.com/docview/883977400. MR 2912198

[Yoneda 1954] N. Yoneda, "On the homology theory of modules", *J. Fac. Sci. Univ. Tokyo. Sect. I.* **7** (1954), 193–227. MR 16,947d Zbl 0058.01902

[Zelevinsky 1981] A. V. Zelevinsky, *Representations of finite classical groups: a Hopf algebra approach*, Lecture Notes in Mathematics **869**, Springer, Berlin, 1981. MR 83k:20017 Zbl 0465.20009

Received March 23, 2015.

ARKADY BERENSTEIN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OREGON
EUGENE, OR 97403-1222
UNITED STATES
arkadiy@uoregon.edu

JACOB GREENSTEIN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
RIVERSIDE, CA 92521-0135
UNITED STATES
jacob.greenstein@ucr.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 281, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.281.333

GENERALIZED SPLINES ON ARBITRARY GRAPHS

SIMCHA GILBERT, JULIANNA TYMOCZKO AND SHIRA VIEL

Let G be a graph whose edges are labeled by ideals of a commutative ring. We introduce a *generalized spline*, which is a vertex labeling of G by elements of the ring so that the difference between the labels of any two adjacent vertices lies in the corresponding edge ideal. Generalized splines arise naturally in combinatorics (*algebraic splines* of Billera and others) and in algebraic topology (certain equivariant cohomology rings, described by Goresky, Kottwitz, and MacPherson, among others). The central question of this paper asks when an arbitrary edge-labeled graph has nontrivial generalized splines. The answer is "always", and we prove the stronger result that the module of generalized splines contains a free submodule whose rank is the number of vertices in G. We describe the module of generalized splines when G is a tree, and give several ways to describe the ring of generalized splines as an intersection of generalized splines for simpler subgraphs of G. We also present a new tool which we call the *GKM matrix*, an analogue of the incidence matrix of a graph, and end with open questions.

1. Introduction

The goal of this paper is to generalize and extend combinatorial constructions that have become increasingly important in many areas of algebraic geometry and topology, as well as to establish a firm combinatorial footing for these constructions. Given a commutative ring R with identity, an arbitrary graph $G = (V, E)$, and a function $\alpha : E \to \{\text{ideals } I \subseteq R\}$, we will define a ring of *generalized splines*. This paper

(1) proves foundational results about generalized splines;

(2) completely analyzes the ring of generalized splines for trees and shows families of generalized splines for arbitrary cycles;

We are very grateful for the contributions of Arielle McCoy and Tia Pilaroscia, and for helpful conversations with Matthias Franz, Ruth Haas, and Hal Schenck. We are indebted to an anonymous referee for very useful comments.
This work was partially supported by NSF grant DMS-0611020 through the Center for Women in Mathematics at Smith College. Tymoczko was partially supported by a Sloan Fellowship and by NSF grants DMS-1101170 and DMS-0801554.
MSC2010: 05C78, 05E15, 55N25.
Keywords: splines, GKM theory, equivariant cohomology, algebraic graph theory.

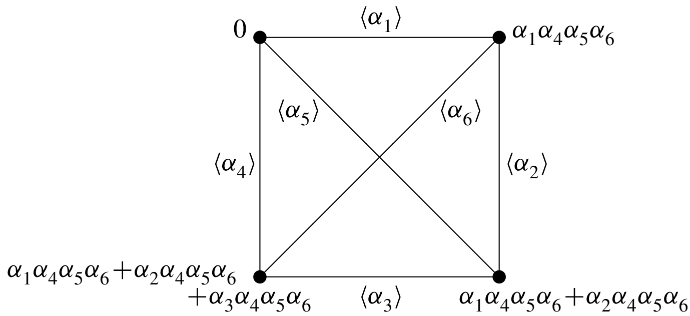
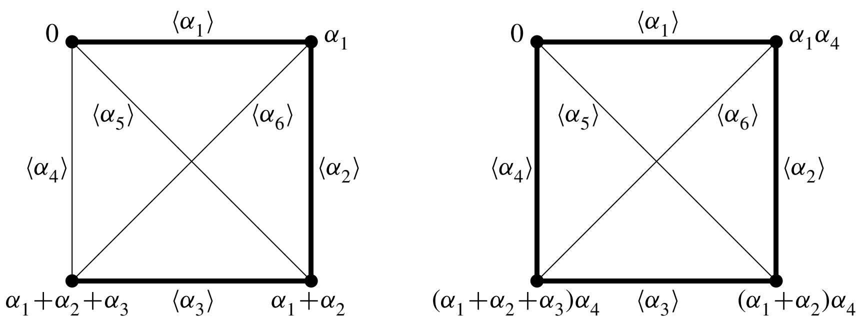


Figure 1. Example of a generalized spline on K_4.



(a) Spline on subgraph $(P_4, \alpha|_{P_4})$. (b) Spline on subgraph $(C_4, \alpha|_{C_4})$.

Figure 2. Nonexamples of generalized splines on K_4.

(3) produces an R-submodule within the ring of generalized splines that has rank $|V|$, as long as R is an integral domain; and

(4) shows that the study of generalized splines for arbitrary graphs can be reduced to the case of different subgraphs, especially cycles or trees.

Generalized splines as we define them are a subring of a product of copies of R:

Definition 1.1. The ring of generalized splines R_G of the pair (G, α) is defined by

$$R_G = \{\boldsymbol{p} \in R^{|V|} : \text{for each edge } e = uv, \text{ the difference } \boldsymbol{p}_u - \boldsymbol{p}_v \in \alpha(e)\}.$$

Figures 1 and 2 display examples and nonexamples of elements of R_{K_4} in the case when each ideal $\alpha(e)$ is generated by a single ring element (given inside $\langle \cdot \rangle$). The vertices are labeled with elements of R_{K_4} and the collection of vertex labels in Figure 1 is a generalized spline. Note that Figures 2(a) and 2(b) are not generalized splines on K_4 but *are* generalized splines on the subgraphs in bold. These examples hold for any ring R and any choice of elements $\alpha_1, \alpha_2, \ldots, \alpha_6 \in R$ to generate the ideals $\alpha(e)$.

The name *generalized spline* comes from one of the important constructions that we extend. Historically, engineers modeled complicated objects like ships or cars by identifying important points of the vehicle and then attaching thin strips of wood (called *splines*) at those points to approximate the entire hull.

Mathematically, a spline is a collection of polynomials on the faces of a polyhedral complex that agree (modulo a power of a linear form) on the intersections of two faces. We refer to this classical tradition as the *analytic* approach to splines; it studies the vector space $C_k^r(D)$, where D is a simplicial complex, r is the order of smoothness to which the polynomials agree over faces, and k is the maximal degree of a polynomial supported on a maximal face. Splines are used in approximation theory and numerical analysis, with applications in data interpolation, to create smooth curves in computer graphics, to find numerical solutions to partial differential equations, and for other applications [Bartels 1984; Chui and Lai 1985; 1990].

In the analytic tradition, mathematicians seek individual splines satisfying particular properties as well as characterizations of the space of splines associated to a given object — for instance, the dimension [Alfeld 1986; 1987; Alfeld and Schumaker 1987; 1990; Chui and He 1989; Gmelig Meyling and Pfluger 1985; Schumaker 1979; 1984a; 1984b] or basis [Alfeld et al. 1987a; 1987b; Morgan and Scott 1975; Schumaker 1988] for a space of splines. Alfeld and Schumaker's work is both representative and epitomic: a seminal result of theirs proved a bound on the dimension of $C_k^r(D)$ when D is a planar simplicial complex and $k \geq 3r+1$ [1987].

Billera [1988] pioneered the study of what some call *algebraic* splines, introducing methods from homological and commutative algebra to prove a conjecture of Strang on the dimension of $C_k^1(D)$ when D is a generic planar simplicial complex. In the abstract algebraic setting, mathematicians generalize the class of geometric objects associated to splines (e.g., Schumaker [1984b], Billera and Rose [1991], and McDonald and Schenck [2009] study piecewise polynomials on a polyhedral complex rather than just a simplicial complex) and study algebraic invariants of modules other than dimension or bases (e.g., the more fundamental question of freeness [Haas 1991; Billera and Rose 1992; Yuzvinsky 1992; Dalbec and Schenck 2001; DiPasquale 2012], or more algebraically involved questions like computing coefficients of the Hilbert polynomial [Billera and Rose 1991; Schenck and Stillman 1997; Schenck 1997; McDonald 2007; McDonald and Schenck 2009], identifying the syzygy module of the span of the edge ideals [Schumaker 1979; Rose 1995; 2004], or analyzing algebraic varieties associated to the piecewise polynomials [Wang 2000; Zhu and Wang 2005; 2011]). Billera and Rose [1991] introduced a description of splines in terms of the dual graph of the polyhedral complex that is equivalent to the piecewise-polynomial definition for so-called *hereditary* complexes. Many others used Billera and Rose's approach in later research [McDonald and Schenck 2009; Rose 1995; 2004], and it is our starting point.

In what we might call the *topological* approach to splines, geometers and topologists recently and independently rediscovered splines as equivariant cohomology rings of toric and other algebraic varieties (though they rarely use the name "splines") [Brion 1996; Payne 2006; Bahri et al. 2009; Schenck 2012]. Goresky, Kottwitz, and MacPherson developed a combinatorial construction of equivariant cohomology called *GKM theory* [Goresky et al. 1998], which can be used for any algebraic variety X with an appropriate torus action. Unknowingly, they described precisely the dual-graph construction of splines: GKM theory builds a graph G_X whose vertices are the T-fixed points of X and whose edges are the one-dimensional orbits of X. Each edge of this graph is associated to a principle ideal $\langle \alpha_e \rangle$ in a polynomial ring, coming from the weight α_e of the torus action on the one-dimensional torus orbits in X. The GKM ring associated to the pair (G_X, α) agrees with what we call the ring of generalized splines for (G_X, α). The main theorem of GKM theory asserts that under appropriate conditions, this GKM ring is in fact isomorphic to the equivariant cohomology ring $H_T^*(X; \mathbb{C})$. (Their work relies on earlier work of many others, including a much more general result of Chang and Skjelbred [1974] that points to one way to extend this work topologically to cases in which the ideals $\langle \alpha_e \rangle$ are no longer principal.) We omit details of the topological background here because there are several excellent surveys [Knutson and Tao 2003; Tymoczko 2005; Holm 2008]. However, GKM theory is a powerful tool in Schubert calculus [Goldin and Tolman 2009; Knutson and Tao 2003], symplectic geometry [Goldin and Tolman 2009; Guillemin et al. 2006; Harada et al. 2005], representation theory [Fiebig 2011], and other fields. (In some of these applications, the *ring* structure of splines is more important than the module structure.)

We note that the most powerful results in each of these approaches are not replicable using other approaches. For instance, Mourrain and Villamizar [2013] recently used the algebraic approach to try to re-prove Alfeld and Schumaker's results, but could not attain their bound.

Our definition of generalized splines allows us to do several things that weren't possible from the algebraic or geometric perspectives:

- *We give a lower bound for one of the central questions of classical splines.* Corollary 5.2 proves that every collection of generalized splines over an integral domain has a free submodule of rank $|V|$, producing a lower bound for the dimension of the ring of splines R_G whenever R_G is a free module over R. This significantly generalizes work of Guillemin and Zara in the GKM context [2003, Theorem 2.1].

- *We streamline earlier combinatorial constructions of splines.* Our construction isolates and highlights the algebraic structures used in previous work on splines. In our language, algebraic splines assume that the ideals $\alpha(e)$ are principal and that the generators for the ideals $\alpha(e)$ satisfy some coprimality conditions. A classical condition like "piecewise polynomials meet with order r smoothness at an edge e" corresponds to using the edge ideal $\alpha(e)^{r+1}$ instead of $\alpha(e)$.

From the geometric point of view, we owe much to a series of papers by Guillemin and Zara [2001; 2003] whose goal is to construct geometric properties of GKM manifolds from a strictly combinatorial viewpoint. Yet their combinatorial model imposes more restrictions than the classical definition of splines — conditions that are natural (and necessary!) for any geometric application.

- *We expand the family of objects on which splines are defined to arbitrary graphs.* Our work shows that graphs that have no reasonable geometric interpretation nonetheless are central to the analysis of splines. Theorem 6.1 decomposes the ring of splines for a graph G in terms of the splines for subgraphs of G; Corollary 6.2 specializes Theorem 6.1 to spanning trees, whose splines are completely described in Theorem 4.1; and Theorem 6.3 decomposes the ring of splines for G in terms of a particular collection of subcycles and subtrees of G. Cycles play a similarly key role in Rose's description of the syzygies of spline ideals [1995; 2004] (see also [Schumaker 1979]). Yet neither trees nor cycles are geometrically meaningful from a GKM perspective. (See [Handschy et al. 2014] and [Bowden et al. 2015] for a deeper investigation of generalized splines on cycles.)

- *We expand the family of rings on which splines are defined.* This gives a convenient language to describe simultaneously the GKM constructions for equivariant cohomology and equivariant K-theory. Moreover, generalized splines over integers have interesting connections to elementary number theory [Handschy et al. 2014].

- *We provide the natural language for further generalizations of splines.* Our construction of generalized splines extends even more: label each vertex of the graph G by a (possibly distinct) R-module M_v and label each edge by a module M_e together with homomorphisms $M_v \to M_e$ for each vertex v incident to the edge e. Geometrically, this corresponds to Braden and MacPherson's construction of equivariant intersection homology [2001], also used by Fiebig in representation-theoretic contexts [2011]. We discuss this and other open questions in Section 7.

The rest of this paper is structured as follows. Section 2 establishes essential results for generalized splines that were first shown in special cases like equivariant cohomology and algebraic splines. We highlight Theorem 2.12 and Corollary 2.13, which generalize and strengthen Rose's result [1995] that for certain polyhedral complexes, the syzygies B of the spline ideal are a direct summand of the splines $R_G \cong R \oplus B$. Corollary 2.13 uses this in Rose's special case to show that the syzygies of the ideal generated by the image of α are isomorphic as a module to the collection of generalized splines whose restriction to a particular fixed point is zero. This relates the algebraically natural question of finding syzygies of splines to the question of finding a particular, geometrically natural kind of basis for the module of splines. Section 3 describes a tool analogous to the incidence matrix of a graph that we call a *GKM matrix*. Section 4 completely characterizes the generalized splines for trees in terms of a minimal set of free generators for the ring of generalized splines.

One of our central questions is, when does an edge-labeled graph have nontrivial generalized splines? The answer (essentially always, as in Theorem 5.1) is actually more refined. Corollary 5.2 explicitly constructs a free R-submodule of the generalized splines on G of rank $|V|$. When R is an integral domain and the generalized splines form a free R-module (as is the case for GKM theory), we conclude that the rank of the R-module of generalized splines is at least $|V|$.

Section 5 uses analogues of a shelling order (in combinatorics) or a "flow-up basis" (in geometry) to identify R-submodules of the generalized splines. Section 6 characterizes generalized splines differently: in terms of the intersections of the generalized splines formed by various subgraphs. This allows us to reframe the definition of generalized splines as an intersection of very simple graphs (Theorem 6.1) and to reduce the number of intersections needed by using certain spanning trees (Corollary 6.2). Finally, Theorem 6.3 analyzes the GKM matrix directly to decompose the ring of generalized splines on G as an intersection of the generalized splines for particular subcycles of G.

2. Definitions and foundational results

In this section, we formalize a collection of definitions which were stated implicitly in the introduction. We then give foundational results describing the structure of the ring of generalized splines, including key methods to construct the ring and to build new generalized splines from existing ones.

We begin with a quick overview of our notational conventions.

- G: a graph, defined as a set of vertices V and edges E. Assumed throughout to be finite with no multiple edges between vertices.
- R: a commutative ring with identity 1.
- $\mathcal{I}$: the set of ideals in R.
- α: an edge-labeling function on G that assigns a nonzero element of $\mathcal{I}$ to each edge in E. See Definition 2.1.
- (G, α): an edge-labeled graph.
- $\alpha(e_{i,j}) = \alpha(v_i v_j) = I_{e_{i,j}}$: the image of the edge $e_{i,j} = v_i v_j$ under the map α.
- $\alpha_{i,j}$: an arbitrary element of the ideal $\alpha(e_{i,j})$. When $\alpha(e_{i,j})$ is principal, $\alpha_{i,j}$ often denotes the generator.
- R_G: the ring of generalized splines on (G, α). See Definition 2.3.

- $\boldsymbol{p}$: a generalized spline; $\boldsymbol{p} = (\boldsymbol{p}_{v_1}, \boldsymbol{p}_{v_2}, \ldots, \boldsymbol{p}_{v_{|V|}})$ denotes an element of $\bigoplus_{v \in V} R$. See Definition 2.3.
- $\boldsymbol{p}_v$: the coordinate of $\boldsymbol{p}$ corresponding to vertex $v \in V$. An element of R.
- M_G: the (possibly extended) GKM matrix for the graph G. See Definition 3.1.

The first definition describes the combinatorial setup of our work: a graph whose edges are labeled by ideals in a ring R. The ring R is always assumed to be a commutative ring with identity, though in later sections we occasionally add more conditions.

Definition 2.1. Let $G = (V, E)$ be a graph and let R be a commutative ring with identity. An *edge-labeling function* is a map $\alpha : E \to \{\text{ideals } I \subseteq R\}$ from the set of edges of G to the set of nonzero ideals in R. An *edge-labeled graph* is a pair (G, α) of a graph G together with an edge-labeling function of E. We often refer to the set of ideals in R as $\mathcal{I}$.

We now precisely define the compatibility condition that we use on the edges.

Definition 2.2. Let $G = (V, \alpha)$ be an edge-labeled graph. An element $\boldsymbol{p} \in \bigoplus_{v \in V} R$ satisfies the *GKM condition* at an edge $e = uv$ if $\boldsymbol{p}_u - \boldsymbol{p}_v \in \alpha(e)$.

In GKM theory and in the theory of algebraic splines, the ring R is a polynomial ring in n variables. The ideal $\alpha(e)$ is the principal ideal generated by a linear form in GKM theory, and by a power of a linear form in the theory of algebraic splines.

We build the ring of generalized splines by imposing the GKM condition at every edge in the graph.

Definition 2.3. Let (G, α) be an edge-labeled graph. The *ring of generalized splines* is

$$R_{G,\alpha} = \Big\{ \boldsymbol{p} \in \bigoplus_{v \in V} R \text{ such that } \boldsymbol{p} \text{ satisfies the GKM condition at each edge } e \in E \Big\}.$$

Each element of $R_{G,\alpha}$ is called a *generalized spline*. When there is no risk of confusion, we write R_G.

We first confirm that in fact R_G is a ring.

Proposition 2.4. *R_G is a ring with unit $\mathbf{1}$ defined by $\mathbf{1}_v = 1$ for each vertex $v \in V$.*

Proof. By definition R_G is a subset of the product ring $\bigoplus_{v \in V} R$ so we need only confirm that the identity is in R_G and that R_G is closed under addition and multiplication. The operations are componentwise addition and multiplication since R_G is in $\bigoplus_{v \in V} R$. The identity in $\bigoplus_{v \in V} R$ is the generalized spline $\mathbf{1}$ defined by $\mathbf{1}_v = 1$ for each vertex $v \in V$. This satisfies the GKM condition at each edge because for

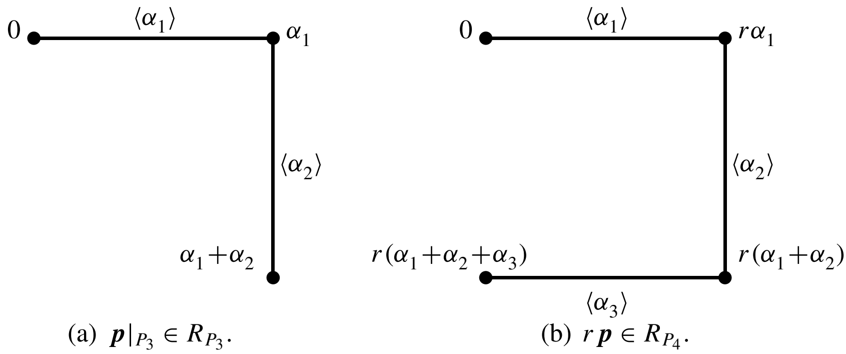


Figure 3. New generalized splines from old.

every edge $e = uv$ we have $\mathbf{1}_u - \mathbf{1}_v = 0$ and 0 is in each ideal $\alpha(e)$. The set R_G is closed under addition because if $\boldsymbol{p}, \boldsymbol{q} \in R_G$ then for each edge $e = uv$ we have

$$(\boldsymbol{p}+\boldsymbol{q})_u - (\boldsymbol{p}+\boldsymbol{q})_v = (\boldsymbol{p}_u + \boldsymbol{q}_u) - (\boldsymbol{p}_v + \boldsymbol{q}_v) = (\boldsymbol{p}_u - \boldsymbol{p}_v) + (\boldsymbol{q}_u - \boldsymbol{q}_v),$$

which is in $\alpha(e)$ by the GKM condition. Similarly, the set R_G is closed under multiplication because if $\boldsymbol{p}, \boldsymbol{q} \in R_G$ then for each edge $e = uv$ we have

$$\begin{aligned}(\boldsymbol{pq})_u - (\boldsymbol{pq})_v &= (\boldsymbol{p}_u\boldsymbol{q}_u) - (\boldsymbol{p}_v\boldsymbol{q}_v)\\ &= (\boldsymbol{p}_u\boldsymbol{q}_u - \boldsymbol{p}_v\boldsymbol{q}_u) + (\boldsymbol{p}_v\boldsymbol{q}_u - \boldsymbol{p}_v\boldsymbol{q}_v)\\ &= \boldsymbol{q}_u(\boldsymbol{p}_u - \boldsymbol{p}_v) + \boldsymbol{p}_v(\boldsymbol{q}_u - \boldsymbol{q}_v),\end{aligned}$$

which is in $\alpha(e)$ by the GKM condition. □

The generalized splines R_G also form an R-module: multiplication by r corresponds to scaling each polynomial in the spline $\boldsymbol{p}$ or equivalently to multiplication by $r\mathbf{1}$. Figure 3(b) demonstrates the R-module structure of R_{P_4}: multiplying $\boldsymbol{p}$ by an arbitrary element $r \in R$ produces the spline $r\boldsymbol{p} = (r\boldsymbol{p}_{v_1}, r\boldsymbol{p}_{v_2}, r\boldsymbol{p}_{v_3}) \in R_{P_4}$.

One major question we study is whether there are *nontrivial* generalized splines.

Definition 2.5. A *nontrivial generalized spline* is an element $\boldsymbol{p} \in R_G$ that is not in the principal ideal $R\mathbf{1}$.

In other words, we ask whether the R-module R_G contains at least two linearly independent elements. We answer this question completely (and more strongly) in Theorem 5.1 and Corollary 5.6: yes, except in the trivial cases when G consists of a single point or R is zero.

If some edge labels were zero then the ring of splines could be trivial for trivial algebraic reasons: for instance, if all edge labels of G are zero then the only elements of R_G are trivial splines. This is why $\alpha(e)$ is always nonzero in Definition 2.1.

Definition 2.6. Let (G, α) and (G', α') be edge-labeled graphs with respect to R. A *homomorphism of edge-labeled graphs* $\phi : (G, \alpha) \to (G', \alpha')$ is a graph homomorphism $\phi_1 : G \to G'$ together with a ring *automorphism* $\phi_2 : R \to R$ so that for each edge $e \in E_G$ we have $\phi_2(\alpha(e)) = \alpha'(\phi_1(e))$:

$$\begin{array}{ccc} E_G & \xrightarrow{\phi_1} & E_{G'} \\ \Big\downarrow{\scriptstyle \alpha} & & \Big\downarrow{\scriptstyle \alpha'} \\ \mathcal{I} & \xrightarrow{\phi_2} & \mathcal{I} \end{array} \tag{1}$$

An *isomorphism of edge-labeled graphs* is a homomorphism of edge-labeled graphs whose underlying graph homomorphism is in fact an isomorphism.

We stress that the map ϕ_2 is a ring *automorphism*. This ensures that ϕ_2 induces a map on the set of ideals $\phi_2 : \mathcal{I} \to \mathcal{I}$ and that the diagram in (1) is well defined. The content of the definition is that the diagram commutes.

Many interesting homomorphisms of edge-labeled graphs arise when $\phi_2 : R \to R$ is the identity homomorphism. Indeed, when R is the integers, this is essentially the only case. However, some rings R have very interesting automorphisms: for instance, when R is a polynomial on n variables, the symmetric group on n letters acts on R by permuting variables. This induces an important action in equivariant cohomology, which is substantively different from a closely related action induced by the identity ring automorphism [Tymoczko 2008]. Our first proposition confirms that the ring of generalized splines is invariant under edge-labeled isomorphisms. More precisely, when two graphs are edge-labeled isomorphic, any generalized spline for one graph will be a generalized spline for the other.

Proposition 2.7. *If $\phi : (G, \alpha) \to (G', \alpha')$ is an isomorphism of edge-labeled graphs then ϕ induces an isomorphism of the corresponding rings of generalized splines $\phi_* : R_G \cong R_{G'}$ according to the rule that $\phi_*(\boldsymbol{p})_{\phi_1(u)} = \phi_2(\boldsymbol{p}_u)$ for each $u \in V_G$.*

Proof. By definition of generalized splines,

$$\boldsymbol{p} \in R_G \iff \boldsymbol{p}_u - \boldsymbol{p}_v \in \alpha(e) \text{ for each edge } e = uv \text{ in } E_G.$$

The map $\phi_2 : R \to R$ is an automorphism of rings, so the GKM conditions imply

$$\boldsymbol{p} \in R_G \iff \phi_2(\boldsymbol{p}_u) - \phi_2(\boldsymbol{p}_v) \in \phi_2(\alpha(e)) \text{ for each edge } e = uv \text{ in } E_G. \tag{2}$$

The map ϕ_1 is an isomorphism between the underlying graphs G and G', so e is an edge in G if and only if $\phi_1(e)$ is an edge in G'. Incorporating the fact that $\alpha'(\phi_1(e)) = \phi_2(\alpha(e))$ for each edge $e \in E_G$, this means (2) is equivalent to

$$\phi_2(\boldsymbol{p}_u) - \phi_2(\boldsymbol{p}_v) \in \alpha'(\phi_1(e)) \text{ for each edge } e' = \phi_1(u)\phi_1(v) \text{ in } E_{G'}. \tag{3}$$

We have that (3) is equivalent to $\phi_*(\boldsymbol{p}) \in R_{G'}$, so we conclude that $\boldsymbol{p}$ is a generalized spline in R_G if and only if $\phi_*(\boldsymbol{p})$ is in $R_{G'}$. □

The next proposition verifies that a generalized spline for the pair (G, α) is a generalized spline for every subgraph of G.

Proposition 2.8. *Let (G, α) be an edge-labeled graph and $G' = (V', E')$ a subgraph of G. Let $(G', \alpha|_{E'})$ be the edge-labeled graph whose function $\alpha|_{E'}$ denotes the restriction of α to the edge set of G'. If $\boldsymbol{p}$ is a generalized spline for (G, α) then $\boldsymbol{p}|_{V'} \in \bigoplus_{v \in V'} R$ is a generalized spline for $(G', \alpha|_{E'})$.*

Proof. Let $G' \subseteq G$ as in the hypothesis, let $\boldsymbol{p}$ be a generalized spline for (G, α), and consider the subcollection $\boldsymbol{p}|_{V'}$ obtained by restricting $\boldsymbol{p}$ to the vertex set $V' \subseteq V$ of G'. For any edge uv in G' the corresponding edge uv is in E since $E' \subseteq E$. This implies that $\boldsymbol{p}_u - \boldsymbol{p}_v \in \alpha(uv)$ by the GKM condition for (G, α). Since the edge-labeling function for G' is the restriction $\alpha|_{E'}$ to the edges in $E' \subseteq E$, we conclude that the GKM condition is satisfied at every edge of G'. It follows that $\boldsymbol{p}|_{V'}$ is a generalized spline for $(G', \alpha|_{E'})$. □

Example 2.9. Consider the generalized spline on the bold P_4 in Figure 2(a) with edges labeled as in Figure 1. Removing a leaf and its incident edge from P_4 gives the subgraph P_3 in Figure 3(a). The generalized spline for P_4 still satisfies the GKM condition at every vertex on the subgraph. Thus $\boldsymbol{p}|_{P_3}$ is a generalized spline for P_3.

The next proposition shows that the special case when one of the edges is associated to the unit ideal $\alpha(e) = R$ is equivalent to a kind of restriction as in Proposition 2.8. In this case, the edge e can be erased without affecting the ring of generalized splines.

Proposition 2.10. *Suppose that the edge-labeled graph (G, α) has an edge e with $\alpha(e) = R$. Let $G' = (V_G, E - \{e\})$ be the graph G with edge e erased, and let $\alpha' : E - \{e\} \to \mathcal{I}$ be the restriction $\alpha' = \alpha|_{E-\{e\}}$. Then*

$$R_G = R_{G'}.$$

Proof. Proposition 2.8 says that every generalized spline of G is a generalized spline of G', since G' is a subgraph of G with the same vertex set whose labeling agrees on shared edges. Hence $R_G \subseteq R_{G'}$. To prove the converse, suppose $\boldsymbol{p}$ is a generalized spline for (G', α'). The GKM condition guarantees that $\boldsymbol{p}_u - \boldsymbol{p}_v \in \alpha(uv)$ for every edge $uv \in E - \{e\}$. In addition, if u_0, v_0 are the endpoints of the edge e, then $\boldsymbol{p}_{u_0} - \boldsymbol{p}_{v_0} \in R$ is vacuously true. Since $\alpha(e) = R$ we conclude that the GKM condition is satisfied for the edge e as well. So $\boldsymbol{p} \in R_G$ and $R_{G'} = R_G$. □

We may build generalized splines from disjoint unions of graphs by taking the direct sum of the respective generalized splines.

Proposition 2.11. *If $G = G_1 \cup G_2$ is the union of two disjoint graphs then the ring of splines is $R_G = R_{G_1} \oplus R_{G_2}$.*

Proof. Rearranging the GKM conditions gives

$$\begin{aligned} R_G &= \Big\{ \boldsymbol{p} \in \bigoplus_{v \in V} R \text{ such that } \boldsymbol{p} \text{ satisfies the GKM condition at each edge } e \in E(G) \Big\} \\ &= \Big\{ \boldsymbol{p} \in \bigoplus_{v \in V(G_1)} R \text{ such that } \boldsymbol{p}_v - \boldsymbol{p}_u \in \alpha(uv) \text{ for all } uv \in E(G_1) \Big\} \\ &\qquad \oplus \Big\{ \boldsymbol{p} \in \bigoplus_{v \in V(G_2)} R \text{ such that } \boldsymbol{p}_v - \boldsymbol{p}_u \in \alpha(uv) \text{ for all } uv \in E(G_2) \Big\} \\ &= R_{G_1} \oplus R_{G_2} \end{aligned}$$

because the vertex sets of G_1 and G_2 are disjoint. □

Another approach to constructing generalized splines is to build them one vertex at a time. The next result decomposes the R-module of generalized splines into a direct sum of the trivial generalized splines and the generalized splines that are zero at a particular vertex.

Theorem 2.12. *Suppose that G is a connected graph with edge-labeling function $\alpha : V \to \mathcal{I}$. Fix a vertex $v \in V$. Then every generalized spline $\boldsymbol{p} \in R_G$ can be written uniquely as $\boldsymbol{p} = r\mathbf{1} + \boldsymbol{p}^{\boldsymbol{v}}$ where $\boldsymbol{p}^{\boldsymbol{v}}$ is a generalized spline satisfying $\boldsymbol{p}^{\boldsymbol{v}}_v = 0$ and $r \in R$ satisfies $r = \boldsymbol{p}_v$. In other words, if $M = \langle \boldsymbol{p} : \boldsymbol{p}_v = 0 \rangle$ then $R_G \cong R\mathbf{1} \oplus M$ as R-modules.*

Proof. The trivial generalized spline $\mathbf{1}$ is in R_G by Proposition 2.4. Let $r \in R$ be the element $r = \boldsymbol{p}_v$. Then define $\boldsymbol{p}^{\boldsymbol{v}}$ to be the generalized spline $\boldsymbol{p}^{\boldsymbol{v}} = \boldsymbol{p} - r\mathbf{1}$. (There is a unique element in the ring R_G that satisfies this equation.) By construction,

$$\boldsymbol{p}^{\boldsymbol{v}}_v = \boldsymbol{p}_v - r\mathbf{1}_v = r - r = 0. \qquad \square$$

The previous result could lead us to consider R-module bases of generalized splines; see the open questions in Section 7. Instead, we combine it with a result of Rose's to relate the generalized splines that vanish at a particular vertex to the syzygies of the module generated by the edge ideals. (Schumaker also implicitly considered syzygies in an earlier work on splines [1979].)

Corollary 2.13. *Suppose G is the dual graph of a hereditary polyhedral complex Δ and that R is the polynomial ring $\mathbb{R}[x_1, x_2, \dots, x_d]$. For each edge e in G, let ℓ_e be an affine form generating the polynomials vanishing on the intersection of faces in Δ corresponding to e. Define α to be the function $\alpha(e) = \langle \ell_e^{r+1} \rangle$ for each edge e and let*

$$B = \Big\{ (b_1, \dots, b_{|E|}) \in R^{|E|} : \textit{for all cycles } C \textit{ in } G, \\ \textit{the linear combination } \sum_{e \in C} b_e \ell_e^{r+1} = 0 \Big\}.$$

Then $M \cong B$ as R-modules.

Proof. Under these conditions, Rose proved that $R_G \cong R \oplus B$ as R-modules [1995, Theorem 2.2]. From the previous claim, we conclude $M \cong B$ as desired. $\square$

We close this section by describing the relationship between the ring of generalized splines associated to an edge-labeling α and the ring of generalized splines associated to the edge-labeling $r\alpha$ obtained by scaling.

Theorem 2.14. *Suppose that (G, α) is a connected edge-labeled graph. Fix an element $r \in R$ and define the edge-labeling function $r\alpha : E \to \mathcal{I}$ by $r\alpha(e) = rI_e$ for each edge $e \in E$. Choose a vertex $v_0 \in V$ and define $M = \langle \boldsymbol{p} : \boldsymbol{p}_{v_0} = 0\rangle$. If R is an integral domain then*

$$R_{G,r\alpha} = R\mathbf{1} \oplus rM.$$

Proof. Theorem 2.12 showed that $R_{G,\alpha} = R\mathbf{1} \oplus M$. The multiple $rR_{G,\alpha}$ belongs to $R_{G,r\alpha}$ by definition, so $rM \subseteq R_{G,r\alpha}$. We also know the intersection $rM \cap R\mathbf{1}$ is zero since the only element of $R\mathbf{1}$ whose restriction to v_0 vanishes is the zero spline. So $R_{G,r\alpha} \supseteq R\mathbf{1} \oplus rM$.

We now prove the opposite containment. Suppose $\boldsymbol{p}' \in R_{G,r\alpha}$ and suppose $\boldsymbol{p} = \boldsymbol{p}' - \boldsymbol{p}'_{v_0}\mathbf{1}$. (Note that $\boldsymbol{p}$ satisfies the GKM condition for $(G, r\alpha)$ at each edge.) We will prove that $\boldsymbol{p} \in rM$. We split the argument into two pieces: showing that $\boldsymbol{p}$ is divisible by r *at each vertex*, and then showing that $\boldsymbol{p}$ satisfies the GKM conditions of rM.

To begin, we prove by induction that if v_k is connected to v_0 by a path of length k then $\boldsymbol{p}_{v_k} \in rR$ is in the principal ideal generated by r. The unique path of length zero is our base case, and the element $\boldsymbol{p}_{v_0} = 0 \in rR$ by construction. Suppose the claim is true for paths of length $k-1$ and let v_k be a vertex connected to v_0 by a path of length k. Then v_k is adjacent to a vertex v_{k-1} which is connected to v_0 by a path of length $k-1$. We know $\boldsymbol{p}_{v_{k-1}} \in rR$ by the inductive hypothesis, and $\boldsymbol{p}_{v_k} - \boldsymbol{p}_{v_{k-1}} \in rI_{e_k}$ for the edge $e_k = v_{k-1}v_k$ by the GKM condition. The sum $rI_{e_k} + rR \subseteq rR$ since ideals are closed under addition, so $\boldsymbol{p}_{v_k} \in rR$ as desired. By induction and because G is connected, we conclude that $\boldsymbol{p}_v \in rR$ for all $v \in V$.

We just showed that each ring element $\boldsymbol{p}$ is divisible by r. For each vertex v, let $\boldsymbol{q}_v$ be the ring element with $\boldsymbol{p}_v = r\boldsymbol{q}_v$ and collect the $\boldsymbol{q}_v$ into the element $\boldsymbol{q} \in R^{|V|}$. We ask whether $\boldsymbol{q} \in M$. To answer this, we need to know whether for each edge $e = uv$ we have $\boldsymbol{q}_u - \boldsymbol{q}_v \in I_e$. We know that $\boldsymbol{p}_u - \boldsymbol{p}_v \in rI_e$ by the GKM condition. Let $x = \boldsymbol{q}_u - \boldsymbol{q}_v \in R$ to isolate the underlying algebraic question: If $rx \in rI_e$ then is $x \in I_e$? The answer is yes when R is an integral domain: if $rx \in rI_e$ then we can find $y \in I_e$ with $rx = ry$. Hence $r(x-y) = 0$, which implies $x = y$ as long as R is an integral domain. $\square$

3. The GKM matrix

The results in the previous section allow us to build new generalized splines from existing ones. To construct generalized splines from scratch we need a systematic method for recording and analyzing GKM conditions. We do this by representing GKM conditions in matrix form. This section shows how to construct GKM matrices and gives several examples.

Our definition of the GKM matrix assumes the graph G is directed. Remark 3.5 shows that changing the directions on the edges of G does not affect the solution space of the matrix, so we generally omit orientations from our figures and our discussion.

Definition 3.1. The *GKM matrix* of the directed, edge-labeled graph (G, α) is an $|E| \times |V|$ matrix constructed so that the row corresponding to each directed edge $e = uv \in E$ has

- 1 in the column corresponding to u,
- -1 in the column corresponding to v, and
- 0 otherwise.

An *extended GKM matrix* of the pair (G, α) is an $|E| \times (|V| + 1)$ matrix whose first $|V|$ columns are the GKM matrix, and whose last entry in the row corresponding to edge e is any element $\alpha_e \in \alpha(e)$. When there is no risk of confusion, we refer to an extended GKM matrix as simply the GKM matrix.

For instance, if $\alpha(e) = \langle \alpha_{e_1}, \dots, \alpha_{e_m} \rangle$ is finitely generated, we could write the last entry in the row corresponding to e as $q_{e_1}\alpha_{e_1} + \cdots + q_{e_m}\alpha_{e_m}$ for arbitrary $q_{e_i} \in R$. In particular, if the ideal $\alpha(e)$ is principal and $\alpha(e) = \langle \alpha_e \rangle$ then we typically write the last column of the extended GKM matrix as the vector $(q_e\alpha_e)_{e \in E}$ for arbitrary coefficients $q_e \in R$.

Remark 3.2. Using this language, we can reframe the syzygy module of spline ideals that Rose defined and that we saw in Corollary 2.13. (See also [Schumaker 1979].) In our context, the syzygy module is essentially the collection of elements $q_e \in \alpha(e)$ from the edge ideals so that $\sum_{e \in C} q_e = 0$ for each cycle C in G. In other words, it describes a collection of elements $q_e \in \alpha(e)$ for which the extended GKM matrix represents a homogeneous system of equations. This condition appears naturally as we analyze the ring R_G further in Theorem 6.3.

Generally we consider q_e to be a parameter that takes values in R, as in the following proposition, which follows immediately from the construction of the GKM matrix.

Proposition 3.3. *Let M_G denote the GKM matrix of (G, α). Then the spline $\boldsymbol{p} \in R^{|V|}$ is a generalized spline for (G, α) if and only if there is an extended GKM matrix $[M_G|\boldsymbol{v}]$ for which $\boldsymbol{p}$ is a solution.*

Proof. The matrix M_G is constructed to record the GKM condition at every edge $e_{i,j} \in E(G)$. Hence a spline $\boldsymbol{p} = (\boldsymbol{p}_{v_1}, \dots, \boldsymbol{p}_{v_{|V|}}) \in R^{|V|}$ is a generalized spline for (G, α) if and only if $M_G \boldsymbol{p} = \boldsymbol{v}$ for some vector $\boldsymbol{v} = (\alpha_e)_{e \in E}$. This is equivalent to saying the spline $\boldsymbol{p}$ is a solution to the system $[M_G|\boldsymbol{v}]$ for some extended GKM matrix, as claimed. □

We can now manipulate M_G to obtain systems of equations that are equivalent to the original GKM conditions on G. We state the following corollary simply to stress this fundamental linear algebra property.

Corollary 3.4. *If $[M'|\boldsymbol{v}']$ is obtained from $[M|\boldsymbol{v}]$ by a series of reversible row or column operations, then the solution set in $R^{|V|}$ to $[M'|\boldsymbol{v}']$ is the same as that of $[M|\boldsymbol{v}]$.*

Reversible operations correspond to invertible matrices in $\mathrm{GL}_{|V|}(R)$. For instance, multiplying a row by x is not reversible for the ring $R = \mathbb{C}[x]$ since $1/x$ is not in R. However, multiplying a row by x is reversible when $R = \mathbb{C}(x)$.

Remark 3.5. Changing the direction of a given edge in G amounts to multiplying the corresponding row in M_G by -1, a reversible operation. Hence while the definition of the GKM matrix for the pair (G, α) requires a directed graph, the actual direction chosen is irrelevant to the solution set given by Proposition 3.3.

Example 3.6. We start with the path P_3 from Figure 3(a). Its extended GKM matrix is

$$M_{P_3} = \left[\begin{array}{ccc|c} 1 & -1 & 0 & q_1\alpha_1 \\ 0 & 1 & -1 & q_2\alpha_2 \end{array}\right],$$

whose rows may be added to obtain the equivalent system

$$\left[\begin{array}{ccc|l} 1 & 0 & -1 & q_2\alpha_2 + q_1\alpha_1 \\ 0 & 1 & -1 & q_2\alpha_2 \end{array}\right].$$

If $\boldsymbol{p} = (\boldsymbol{p}_{v_1}, \boldsymbol{p}_{v_2}, \boldsymbol{p}_{v_3}) \in R_{P_3}$ then the system has dependent variables $\boldsymbol{p}_{v_1}$ and $\boldsymbol{p}_{v_2}$ and independent variable $\boldsymbol{p}_{v_3}$. All solutions may be written in the form

$$\begin{aligned} \boldsymbol{p}_{v_1} &= \boldsymbol{p}_{v_3} + q_2\alpha_2 + q_1\alpha_1, \\ \boldsymbol{p}_{v_2} &= \boldsymbol{p}_{v_3} + q_2\alpha_2, \end{aligned}$$

where $\boldsymbol{p}_{v_3}$, q_1, and q_2 are freely chosen elements of R. Setting $\boldsymbol{p}_{v_3} = 0$, $q_1 = 1$, and $q_2 = 1$ yields the generalized spline in Figure 3(a).

The following generalization will be a central part of our proof of Theorem 3.8.

Example 3.7. Consider the path P_n on n vertices:

$$\begin{tikzcd} P_n = & \underset{v_1}{\bullet} \arrow[r, dash, "\alpha(e_{1,2})"] & \underset{v_2}{\bullet} \arrow[r, dash] & \cdots \arrow[r, dash] & \underset{v_{n-1}}{\bullet} \arrow[r, dash, "\alpha(e_{n-1,n})"] & \underset{v_n}{\bullet} \end{tikzcd}$$

The GKM matrix for this path is

$$\left[\begin{array}{ccccccc|c} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & \alpha_{1,2} \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 & \alpha_{2,3} \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 & \alpha_{3,4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 & \alpha_{n-1,n} \end{array}\right],$$

where $\alpha_{i,i+1} \in \alpha(e_{i,i+1})$ are arbitrarily chosen. As before, we can row-reduce the GKM matrix by setting row i to be the sum $\sum_{k=i}^{n}(\text{row } k)$ for each $1 \le i \le n$. We obtain an equivalent system of rank $n-1$ in which $\boldsymbol{p}_{v_n}$ is the only free variable in the set $\{\boldsymbol{p}_{v_i} : i = 1, \ldots, n\}$. (This system is of maximal rank since an $(n-1) \times (n+1)$ system of equations can have at most one free variable among the $\boldsymbol{p}_{v_i}$.) Figure 4 shows this equivalent system:

$$\left[\begin{array}{ccccccc|l} 1 & 0 & 0 & 0 & \cdots & 0 & -1 & \alpha_{n-1,n} + \cdots + \alpha_{3,4} + \alpha_{2,3} + \alpha_{1,2} \\ 0 & 1 & 0 & 0 & \cdots & 0 & -1 & \alpha_{n-1,n} + \cdots + \alpha_{3,4} + \alpha_{2,3} \\ 0 & 0 & 1 & 0 & \cdots & 0 & -1 & \alpha_{n-1,n} + \cdots + \alpha_{3,4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \quad\vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 & \alpha_{n-1,n} \end{array}\right]$$

Figure 4. A system equivalent to the GKM matrix for P_n.

The linear combinations that occur in the last column of the matrix in Figure 4 can be used to construct generalized splines for more complicated graphs as well. For instance, the next result builds on this description of paths to describe a collection of (usually) nontrivial generalized splines for the cycle C_n.

Theorem 3.8. *Let C_n be a finite edge-labeled cycle given by vertices $v_1, v_2, \ldots, v_n$ in order. Define the vector $\boldsymbol{p} \in R^{|V|}$ by*

$$\begin{bmatrix} \boldsymbol{p}_{v_1} \\ \boldsymbol{p}_{v_2} \\ \boldsymbol{p}_{v_3} \\ \vdots \\ \boldsymbol{p}_{v_{n-1}} \\ \boldsymbol{p}_{v_n} \end{bmatrix} = \boldsymbol{p}_{v_1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} + \alpha_{1,n} \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{1,2} \\ \alpha_{2,3} \\ \alpha_{3,4} \\ \vdots \\ \alpha_{n-2,n-1} \\ \alpha_{n-1,n} \end{bmatrix} \tag{4}$$

with arbitrary choices of $\boldsymbol{p}_{v_1} \in R$, $\alpha_{i,i+1} \in \alpha(e_{i,i+1})$, *and* $\alpha_{1,n} \in \alpha(e_{1,n})$. *Then* $\boldsymbol{p}$ *is a generalized spline for* C_n. *The spline* $\boldsymbol{p}$ *is nontrivial exactly when* $\alpha_{1,n}$ *and at least one of the* $\alpha_{i,i+1}$ *are nonzero.*

Proof. We check that $\boldsymbol{p} \in R^n$ satisfies the GKM condition at every edge of C_n. For all i with $2 \le i \le n-1$ we have

$$\begin{aligned}
&\boldsymbol{p}_{v_{i+1}} - \boldsymbol{p}_{v_i} \\
&\quad = (\boldsymbol{p}_{v_1} + \alpha_{1,n}(\alpha_{1,2} + \cdots + \alpha_{i-1,i} + \alpha_{i,i+1})) - (\boldsymbol{p}_{v_1} + \alpha_{1,n}(\alpha_{1,2} + \cdots + \alpha_{i-1,i})) \\
&\quad = \alpha_{1,n}\alpha_{i,i+1},
\end{aligned}$$

which is in $\alpha(e_{i,i+1})$ by assumption on $\alpha_{i,i+1}$. It remains to check that the GKM condition is satisfied at edges $e_{1,2}$ and $e_{1,n}$. At edge $e_{1,2}$ we have

$$\boldsymbol{p}_{v_2} - \boldsymbol{p}_{v_1} = (\boldsymbol{p}_{v_1} + \alpha_{1,n}\alpha_{1,2}) - \boldsymbol{p}_{v_1} = \alpha_{1,n}\alpha_{1,2},$$

which is in the ideal $\alpha(e_{1,2})$. At edge $e_{1,n}$ we have

$$\boldsymbol{p}_{v_n} - \boldsymbol{p}_{v_1} = (\boldsymbol{p}_{v_1} + \alpha_{1,n}(\alpha_{1,2} + \cdots + \alpha_{n-1,n})) - \boldsymbol{p}_{v_1} = \alpha_{1,n}(\alpha_{1,2} + \cdots + \alpha_{n-1,n}),$$

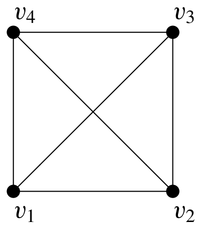
which is in the ideal $\alpha(e_{1,n})$. Hence $\boldsymbol{p}$ is a generalized spline for C_n. The spline $\boldsymbol{p}$ is nontrivial if and only if the second term is nonzero, namely, when $\alpha_{1,n}$ and at least one of the $\alpha_{i,i+1}$ are nonzero. □

Theorem 3.8 actually does more: it identifies a collection of generalized splines for C_n that are linearly independent for many choices of R. Indeed, we can write the generalized splines from Theorem 3.8 in parametric form:

$$\begin{bmatrix} \boldsymbol{p}_{v_1} \\ \boldsymbol{p}_{v_2} \\ \boldsymbol{p}_{v_3} \\ \boldsymbol{p}_{v_4} \\ \vdots \\ \boldsymbol{p}_{v_n} \end{bmatrix} = \boldsymbol{p}_{v_1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \alpha_{1,n}\alpha_{1,2} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \alpha_{1,n}\alpha_{2,3} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \cdots + \alpha_{1,n}\alpha_{n-1,n} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{5}$$

with coefficients $\boldsymbol{p}_{v_1} \in R$ and $\alpha_{i,i+1} \in \alpha(e_{i,i+1}) = I_{i,i+1}$ for all $1 \le i \le n-1$. The vectors $[1, 1, 1, \ldots, 1]^T$, $[0, 1, 1, \ldots, 1]^T$, $\ldots$, $[0, 0, 0, \ldots, 1]^T$ are linearly independent in R^n but are not necessarily elements of R_{C_n}. If R is an integral domain then for any fixed choices of $\alpha_{i,j} \in \alpha(e_{i,j}) = I_{i,j}$ the vectors $[1, 1, 1, \ldots, 1]^T$, $\alpha_{1,n}\alpha_{1,2}[0, 1, 1, \ldots, 1]^T, \ldots, \alpha_{1,n}\alpha_{n-1,n}[0, 0, 0, \ldots, 1]^T$ are both linearly independent and in R_{C_n}.

We will use these kinds of splines — which arise naturally when considering the GKM matrix — repeatedly in subsequent sections of the paper.



$$
\begin{aligned}
\boldsymbol{p}_{v_1} - \boldsymbol{p}_{v_2} &\in \alpha(e_{1,2}) = \langle \alpha_{1,2} \rangle \\
\boldsymbol{p}_{v_1} - \boldsymbol{p}_{v_3} &\in \alpha(e_{1,3}) = \langle \alpha_{1,3} \rangle \\
\boldsymbol{p}_{v_1} - \boldsymbol{p}_{v_4} &\in \alpha(e_{1,4}) = \langle \alpha_{1,4} \rangle \\
\boldsymbol{p}_{v_2} - \boldsymbol{p}_{v_3} &\in \alpha(e_{2,3}) = \langle \alpha_{2,3} \rangle \\
\boldsymbol{p}_{v_2} - \boldsymbol{p}_{v_4} &\in \alpha(e_{2,4}) = \langle \alpha_{2,4} \rangle \\
\boldsymbol{p}_{v_3} - \boldsymbol{p}_{v_4} &\in \alpha(e_{3,4}) = \langle \alpha_{3,4} \rangle
\end{aligned}
$$

Figure 5. GKM conditions for K_4 whose ideals are all principal.

$$
M_{K_4} = \left[\begin{array}{cccc|l}
1 & 0 & 0 & -1 & q_{1,4}\alpha_{1,4} \\
0 & 1 & 0 & -1 & q_{2,3}\alpha_{2,3} \\
0 & 0 & 1 & -1 & q_{3,4}\alpha_{3,4} \\
0 & 0 & 0 & 0 & q_{1,2}\alpha_{1,2} - q_{1,4}\alpha_{1,4} + q_{2,4}\alpha_{2,4} \\
0 & 0 & 0 & 0 & q_{1,3}\alpha_{1,3} - q_{1,4}\alpha_{1,4} + q_{3,4}\alpha_{3,4} \\
0 & 0 & 0 & 0 & q_{2,3}\alpha_{2,3} - q_{2,4}\alpha_{2,4} + q_{3,4}\alpha_{3,4}
\end{array}\right]
$$

Figure 6. A system equivalent to the extended GKM matrix for K_4 when all ideals are principal.

Example 3.9. We return to the case of the complete graph K_4 whose ideals $\alpha(e)$ are all principal. By Definition 2.3, the tuple $\boldsymbol{p} = (\boldsymbol{p}_{v_1}, \boldsymbol{p}_{v_2}, \boldsymbol{p}_{v_3}, \boldsymbol{p}_{v_4})$ is a generalized spline for K_4 if and only if it satisfies the GKM conditions in Figure 5.

The difference $\boldsymbol{p}_{v_i} - \boldsymbol{p}_{v_j}$ is in the ideal $\alpha(e_{i,j}) = \langle \alpha_{i,j} \rangle$ if and only if the difference $\boldsymbol{p}_{v_i} - \boldsymbol{p}_{v_j} = q_{i,j}\alpha_{i,j}$ for some $q_{i,j} \in R$, so we represent these GKM conditions by the following matrix equation (the coefficient matrix is the GKM matrix):

$$
\begin{bmatrix}
1 & -1 & 0 & 0 \\
1 & 0 & -1 & 0 \\
1 & 0 & 0 & -1 \\
0 & 1 & -1 & 0 \\
0 & 1 & 0 & -1 \\
0 & 0 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
\boldsymbol{p}_{v_1} \\ \boldsymbol{p}_{v_2} \\ \boldsymbol{p}_{v_3} \\ \boldsymbol{p}_{v_4}
\end{bmatrix}
= \left[q_{1,2}, q_{1,3}, q_{1,4}, q_{2,3}, q_{2,4}, q_{3,4} \right]
\begin{bmatrix}
\alpha_{1,2} \\ \alpha_{1,3} \\ \alpha_{1,4} \\ \alpha_{2,3} \\ \alpha_{2,4} \\ \alpha_{3,4}
\end{bmatrix}.
$$

After several invertible row operations in which we add various rows to other rows, we obtain an equivalent system of equations such as that given in Figure 6.

4. Generalized splines for trees

We will now use the GKM matrix to describe all generalized splines for trees. We start by describing the generalized splines for paths, using the same argument as that for trees but without the notational technicalities.

Figure 4 shows a matrix that is row-equivalent to the GKM matrix for the path (P_n, α). The solutions can be written in parametric form as

$$\begin{bmatrix} \boldsymbol{p}_{v_1} \\ \boldsymbol{p}_{v_2} \\ \boldsymbol{p}_{v_3} \\ \boldsymbol{p}_{v_4} \\ \vdots \\ \boldsymbol{p}_{v_{n-1}} \\ \boldsymbol{p}_{v_n} \end{bmatrix} = \boldsymbol{p}_{v_n} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} + \alpha_{n-1,n} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{bmatrix} + \cdots + \alpha_{3,4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + \alpha_{2,3} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + \alpha_{1,2} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix},$$

where the coefficients $\boldsymbol{p}_{v_n}$ and $\alpha_{i,i+1}$ for all $1 \le i \le n-1$ are chosen arbitrarily from the sets R and $\alpha(e_{i,i+1}) = I_{i,i+1}$ respectively. By Corollary 3.4, this gives precisely the collection of generalized splines for the path P_n.

When R is an integral domain, this also gives linearly independent vectors in R_{P_n} (for any choices of $\alpha_{i,i+1} \in I_{i,i+1}$):

$$\mathcal{B}_{R_{P_n}} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} \alpha_{n-1,n} \\ \alpha_{n-1,n} \\ \alpha_{n-1,n} \\ \alpha_{n-1,n} \\ \vdots \\ \alpha_{n-1,n} \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} \alpha_{3,4} \\ \alpha_{3,4} \\ \alpha_{3,4} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \alpha_{2,3} \\ \alpha_{2,3} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \alpha_{1,2} \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \right\}. \tag{6}$$

Morally speaking, this decomposition describes something very close to a basis for the generalized splines — as long as we can write a basis for the ideals $I_{i,i+1}$. For instance, when each ideal $I_{i,i+1}$ is principal and $\alpha_{i,i+1}$ denotes the generator of $I_{i,i+1}$ for each $1 \le i \le n-1$, then these vectors form a basis for R_{P_n}. In general, we won't be able to find a basis for R_G because we can't even necessarily find bases for the ideals $I_{i,i+1}$. Even when R is a polynomial ring, we need all of the technical tools developed in the theory of Gröbner bases to compute bases of ideals in R.

However, we can find generators for the splines on trees. We reformulate the essential property of this basis from the point of view of trees. Observe that $\boldsymbol{p} \in R_{P_n}$ must satisfy the following property for any $v_i, v_j \in V(P_n)$ with $i < j$:

$$\boldsymbol{p}_{v_j} = \boldsymbol{p}_{v_i} + \sum_{k=i}^{j-1} \alpha_{k,k+1} \quad \text{for some } \alpha_{k,k+1} \in I_{k,k+1}. \tag{7}$$

Trees are more complicated than paths, so describing the general result precisely is more complicated. The main idea is similar to the one above, though. It relies on the fact that there is exactly one path between any two vertices in a tree, as well as on (7).

Theorem 4.1. *Let $T = (V, E, \alpha)$ be a finite edge-labeled tree. The tuple $\boldsymbol{p} \in R^{|T|}$ is a generalized spline $\boldsymbol{p} \in R_T$ if and only if given any two vertices $v_i, v_j \in V$ we may write*

$$\boldsymbol{p}_{v_j} = \boldsymbol{p}_{v_i} + \alpha_{i,i_1} + \cdots + \alpha_{i_{m-1},i_m} + \alpha_{i_m,j} \quad \textit{for some } \alpha_{l,k} \in \alpha(e_{l,k}) = I_{l,k}, \tag{8}$$

where $v_i, v_{i_1}, \ldots, v_{i_m}, v_j$ are the vertices in the unique path connecting v_i and v_j in the tree T. Furthermore $\boldsymbol{p}$ is nontrivial if and only if at least one of the $\alpha_{l,k}$ is nonzero.

Proof. We proceed via induction on $|V|$. The base case $|V| = 1$ is trivial since $E = \varnothing$. We also prove the case $|V| = 2$, namely, when T is a path on two vertices. Denote the vertices of T by v_1 and v_2 and the edge set by $E = \{e_{1,2}\}$. Now let $\boldsymbol{p} = (p_{v_1}, p_{v_2}) \in R^2$. By Definition 2.3 we know $\boldsymbol{p} \in R_T$ if and only if $\boldsymbol{p}_{v_1} - \boldsymbol{p}_{v_2} \in I_{1,2}$. We rewrite this as $\boldsymbol{p}_{v_1} = \boldsymbol{p}_{v_2} + \alpha_{1,2}$ for some choice of $\alpha_{1,2} \in I_{1,2}$. In other words $\boldsymbol{p}$ is a generalized spline for T if and only if $\boldsymbol{p}$ satisfies (8) for all pairs of vertices in $V = \{v_1, v_2\}$. Furthermore $\boldsymbol{p}$ is nontrivial if and only if $\boldsymbol{p}_{v_1} \neq \boldsymbol{p}_{v_2}$ or equivalently $\alpha_{1,2} \neq 0$.

Assume the theorem holds for every tree with at most n vertices and let $T' = (V', E', \alpha)$ with $|V'| = n+1$. Suppose $\boldsymbol{p} \in R^{|V'|}$ satisfies (8) for all pairs of vertices in V' and let $e_{h,k} \in E'$ be an arbitrary edge. Since v_h and v_k are adjacent in T' we know $\boldsymbol{p}_k = \boldsymbol{p}_h + \alpha_{h,k}$ for some $\alpha_{h,k} \in I_{h,k}$ by (8). Rewriting this condition, we obtain $\boldsymbol{p}_k - \boldsymbol{p}_h \in I_{h,k}$. Since $e_{h,k}$ was arbitrary we conclude $\boldsymbol{p} \in R_{T'}$.

Conversely, suppose that $\boldsymbol{p} \in R_{T'}$. We show that $\boldsymbol{p}$ satisfies (8) for all vertices in V'. Without loss of generality, label the vertices of T' so that v_{n+1} is a leaf adjacent to v_n. Choose arbitrary $v_i, v_j \in V'$ and let $v_i, v_{i_1}, \ldots, v_{i_m}, v_j$ denote the vertices in the unique path connecting v_i and v_j in T'. Let T denote the subgraph $T \subseteq T'$ induced by $v_i, v_{i_1}, \ldots, v_{i_m}, v_j$. The graph T is a tree itself, since it is a connected subgraph of a tree. The restriction of $\boldsymbol{p}$ to the vertices in T is a generalized spline for T by Proposition 2.8. If T has at most n vertices then the inductive hypothesis implies that $\boldsymbol{p}$ satisfies (8) for the pair v_i, v_j. If T has $n+1$ vertices then T is a path of length $n+1$. Figure 4 shows a system equivalent to the GKM matrix in this case. The first row of this matrix describes the equation

$$\boldsymbol{p}_{v_j} = \boldsymbol{p}_{v_i} + \alpha_{i,i_1} + \cdots + \alpha_{i_{m-1},i_m} + \alpha_{i_m,j}$$

for some set $\alpha_{l,k} \in \alpha(e_{l,k}) = I_{l,k}$. In other words, this graph also satisfies (8), proving our claim.

Finally, the spline $\boldsymbol{p}$ is nontrivial if and only if there exists some pair of vertices $v_i, v_j \in V'$ such that $\boldsymbol{p}_{v_i} \neq \boldsymbol{p}_{v_j}$. This is equivalent to saying that the coefficients $\alpha_{i,i_1}, \alpha_{i_1,i_2}, \ldots, \alpha_{i_{m-1},i_m}, \alpha_{i_m,j}$ associated to the path $v_i, v_{i_1}, \ldots, v_{i_m}, v_j$ are not all equal to 0, by (8). Equivalently there exists a pair l, k with $\alpha_{l,k} \neq 0$ as desired. $\square$

5. Existence of generalized splines and lower bounds on the rank of R_G

We now address a fundamental question: Do nontrivial generalized splines exist for an arbitrary edge-labeled graph (G, α)? We solved this question in the case of edge-labeled cycles (C_n, α) in Theorem 3.8. The answer in that case (yes) leads naturally to a stronger result: Equation (5) actually identifies a collection of generalized splines that are linearly independent when R is an integral domain. The condition that R be an integral domain is crucial, as Bowden and Tymoczko show in forthcoming work [2015].

Similarly, we will answer the existence question for generalized splines on arbitrary (G, α) (yes, unless G consists of a single vertex) by constructing a collection of generalized splines that are linearly independent when R is an integral domain. This provides a lower bound on the rank of R_G as an R-module when R_G is a free R-module, and constructs a collection of generators associated to vertices when the ideal $\alpha(e)$ is principal for each edge e. All of these hypotheses are satisfied for the generalized splines used to construct equivariant cohomology and equivariant K-theory, where constructing bases is an important and well-studied question [Guillemin and Zara 2001; Goldin and Tolman 2009]. Geometrically, Theorem 5.1 and Corollary 5.2 partially extend existing results on flow-up classes in equivariant cohomology, since we broaden the class of varieties for which we can construct linearly independent rank-n collections of flow-up classes. The result is new for equivariant K-theory. We note, however, that our flow-up classes are generally not a basis for R_G.

Corollary 5.2 proves that each R_G contains a free submodule of rank n as a special (and simpler) case of Theorem 5.1.

Theorem 5.1. *Let (G, α) be a finite edge-labeled graph. Fix any subgraph G' of G and let $\boldsymbol{p}$ be a generalized spline for $(G', \alpha|_{G'})$. Let $N_{G'} = \prod_S \alpha_{i,j}$, where each $\alpha_{i,j}$ is a nonzero element of the ideal $\alpha(v_i v_j)$ and the product is taken over the set S of edges incident to a vertex in G' but not in G', namely,*

$$S = \{\alpha_{i,j} : v_i v_j \in E(G - G') \text{ and } v_i \in V(G') \text{ or } v_j \in V(G')\}.$$

Then the vector $\boldsymbol{q}$ defined by

$$\boldsymbol{q}_{v_i} = \begin{cases} N_{G'} \boldsymbol{p}_{v_i} & \text{if } v_i \in V(G'), \\ 0 & \text{if } v_i \notin V(G') \end{cases}$$

is a generalized spline for G.

Proof. For each edge $v_i v_j \in E(G)$, there are three possibilities:

(1) Both v_i and v_j are in $V(G')$. Then $\boldsymbol{p}_{v_i} - \boldsymbol{p}_{v_j}$ satisfies the GKM condition in G'. Thus $\boldsymbol{q}_{v_i} - \boldsymbol{q}_{v_j} = N_{G'}(\boldsymbol{p}_{v_i} - \boldsymbol{p}_{v_j})$ satisfies the GKM condition for v_i, v_j in G since $\alpha(v_i v_j)$ is an ideal and $N_{G'} \in R$.

(2) Neither v_i nor v_j is in $V(G')$. Then the difference $\boldsymbol{q}_{v_i} - \boldsymbol{q}_{v_j} = 0 - 0$ vacuously satisfies the GKM condition for v_i, v_j in G.

(3) Exactly one of v_i, v_j is in $V(G')$. Suppose that $v_i \in V(G')$ and $v_j \notin V(G')$. Consider the difference $\boldsymbol{q}_{v_i} - \boldsymbol{q}_{v_j} = N_{G'}(\boldsymbol{p}_{v_i} - \boldsymbol{p}_{v_j})$. The factor $N_{G'}$ is in the ideal $\alpha(v_i v_j)$ by definition of $N_{G'}$ and by definition of ideals. Hence the product $N_{G'}(\boldsymbol{p}_{v_i} - \boldsymbol{p}_{v_j})$ satisfies the GKM condition for v_i, v_j in G. □

The next corollary constructs classes that look like what are called "flow-up" classes in geometric applications. Given a partial order on the vertices of G, a *flow-up class* associated to the vertex v is a generalized spline $\boldsymbol{p}^{\boldsymbol{v}}$ so that for each vertex u with $u \not> v$ the spline satisfies $\boldsymbol{p}^{\boldsymbol{v}}_u = 0$. (In geometric applications, flow-up classes satisfy additional conditions as well.) These classes occur naturally in geometric applications: the partial order comes from a suitably generic one-dimensional torus action on the variety (and hence on the graph), and the spline is the cohomology class associated to the subvariety that flows into the vertex v. The most famous examples of flow-up classes occur in flag varieties and Grassmannians, where they are known as Schubert classes and where they in fact form a basis for the ring of generalized splines (equivariant cohomology rings, in the geometric context).

Our motivation for the next sequence of corollaries comes from these geometric applications. In those cases, the ideals $\alpha(e)$ for each edge e are principal. If some ideals were not principal, the results that follow could be refined to construct a larger free submodule of R_G.

We now construct a rank-n free submodule of the generalized splines for an arbitrary edge-labeled graph (G, α) using a collection of linearly independent flow-up classes. The reader interested only in the special case of this corollary could prove it directly by taking G' to be a single vertex.

Corollary 5.2. *Let R be an integral domain and (G, α) a connected edge-labeled graph on n vertices. Then R_G contains a free R-submodule of rank n.*

Proof. Enumerate the vertices in $V(G)$ as $v_1, v_2, \ldots, v_n$. For each v_i define G'_i to be the subgraph consisting of exactly vertex v_i. Clearly $\boldsymbol{p} = \boldsymbol{1}$ is a generalized spline for $(G'_i, \alpha|_{G'_i})$ for all $1 \le i \le n$. Then Theorem 5.1 yields generalized splines $\{\boldsymbol{q}_i : i = 1, \ldots, n\}$ for G, where $\boldsymbol{q}_{i_{v_j}} = \delta_{ij} N_{G'_i}$ and $N_{G'_i} = \prod_{j \ne i} \alpha_{i,j}$ for arbitrarily chosen $0 \ne \alpha_{i,j} \in \alpha(v_i v_j)$. We show that this set is linearly independent in the R-module R_G. Suppose $\sum_{i=1}^n c_i \boldsymbol{q}_i = \boldsymbol{0}$ for coefficients $c_i \in R$. For each $1 \le j \le n$, evaluation at v_j yields

$$\sum_{i=1}^{n} c_i \boldsymbol{q}_{i_{v_j}} = \sum_{i=1}^{n} c_i \delta_{ij} N_{G'_i} = c_j N_{G'_j} = 0. \tag{9}$$

Since R is an integral domain and each $\alpha_{i,j} \neq 0$ it follows that $N_{G'_j} \neq 0$ for all j. Hence (9) implies $c_j = 0$ for all $1 \leq j \leq n$ so that $\{\boldsymbol{q}_i : i = 1, \dots, n\}$ is linearly independent in R_G and therefore spans a free R-submodule of rank n. □

The next corollary makes note of a particular choice for the scaling factor $N_{G'}$ in Theorem 5.1 that can be useful in the kinds of examples that arise in geometric applications. All of the hypotheses hold in typical geometric applications (equivariant cohomology with field coefficients, equivariant K-theory with field coefficients, and classical algebraic splines).

Corollary 5.3. *Fix an edge-labeled graph (G, α) and let R be a unique factorization domain. Suppose that for each edge e the ideal $\alpha(e)$ is principal and choose a generator $\alpha_{i,j}$ for each edge $e = v_i v_j$. Then for any subgraph G' of G we may apply Theorem 5.1 by choosing*

$$N_{G'} = \operatorname{lcm}\{\alpha_{i,j} : v_i v_j \in E(G - G') \textit{ and } v_i \in V(G') \textit{ or } v_j \in V(G')\}.$$

The next two corollaries of Theorem 5.1 address particular ways to construct (nontrivial) generalized splines for G from subgraphs of G.

Corollary 5.4. *If G contains any subgraph G' for which $R_{G'}$ contains a nontrivial generalized spline then R_G also contains a nontrivial generalized spline.*

Example 5.5. We can construct generalized splines for the edge-labeled graph (K_4, α) given in Figure 1 using these corollaries. The vertex in the upper-left corner is v_1 and the other vertices, clockwise around the square, are v_2, v_3, v_4. Let C_4 denote the Hamiltonian cycle determined by ordering the vertices $v_1 v_2 v_3 v_4$, and let

$$N_{C_4} = \operatorname{lcm}\{\alpha(v_1 v_3), \alpha(v_2 v_4)\} = \operatorname{lcm}\{\alpha_5, \alpha_6\}$$

with the labeling in Figure 1. Theorem 3.8 constructed many nontrivial generalized splines for C_4, including

$$\boldsymbol{p} = \begin{bmatrix} 0 \\ \alpha(v_1 v_4)\alpha(v_1 v_2) \\ \alpha(v_1 v_4)(\alpha(v_1 v_2) + \alpha(v_2 v_3)) \\ \alpha(v_1 v_4)(\alpha(v_1 v_2) + \alpha(v_2 v_3) + \alpha(v_3 v_4)) \end{bmatrix} = \begin{bmatrix} 0 \\ \alpha_4 \alpha_1 \\ \alpha_4(\alpha_1 + \alpha_2) \\ \alpha_4(\alpha_1 + \alpha_2 + \alpha_3) \end{bmatrix}.$$

The corollaries show that the multiple $N_{C_4} \boldsymbol{p}$ is a generalized spline for K_4.

Corollary 5.6. *Let R be an integral domain. If G contains at least two vertices then R_G contains a nontrivial generalized spline.*

Proof. The vertex set V has at least two vertices, so V has a proper subset. Let G' denote a subgraph of G induced by any proper subset of V. Choose the unit $\mathbf{1} \in R_{G'}$ for the spline $\boldsymbol{p}$ in Theorem 5.1. The factor $N_{G'}$ is nonzero because R is an integral domain. □

6. Decomposing R_G as an intersection

This section describes two ways to express R_G as an intersection of rings R_{G_i} for simpler graphs G_i. Both are inspired by the GKM matrix, which allows us to recognize and manipulate the GKM conditions for various subgraphs of G.

In the first decomposition, we essentially reorganize the GKM matrix and identify the GKM matrices associated to subgraphs of G inside the GKM matrix for G. When these subgraphs are the edges themselves, we recover the result that the generalized splines are the intersection of the GKM conditions on all edges independently. We can alternatively take these subgraphs to be trees, whose generalized splines we identified completely in Section 4; this reduces the number of intersections needed to calculate R_G.

In the other decomposition, we row-reduce the GKM matrix in a natural way to demonstrate that R_G is the intersection of the generalized splines for a particular collection of subcycles of G. This demonstrates how the combinatorial perspective can contribute to the study of generalized splines and GKM theory: cycles are subgraphs that do not arise from geometric considerations but are natural in this more general combinatorial setting. It also reinforces Rose's results [1995; 2004] showing the importance of cycles in studying splines. Handschy, Melnick, and Reinders [2014] identify a basis for generalized splines with integer coefficients over cycles in forthcoming work. Bowden, Cao, Hagen, King, and Reinders [2015] give a simpler basis for generalized splines over cycles whose edge labels satisfy a coprimality condition; this allows them to identify the ring structure of the generalized splines completely.

We begin by expressing the ring of generalized splines as an intersection of generalized splines for subgraphs.

Theorem 6.1. *Let (G, α) be an edge-labeled graph. Suppose $G_1, G_2, \ldots, G_k$ are a collection of spanning subgraphs of G whose union is G, in the sense that $V(G_i) = V(G)$ for all i and $\bigcup_{i=1}^k E(G_i) = E(G)$. Let $\alpha_i = \alpha|_{G_i}$ be the edge labelings given by restriction for each i. Then*

$$R_G = \bigcap_{i=1}^{k} R_{G_i}.$$

Proof. Proposition 2.8 showed that R_G is contained in $R_{G'}$ for each spanning subgraph G' of G, and in particular is contained in R_{G_i} for each subgraph G_i. Conversely, suppose $\boldsymbol{p}$ is contained in $\bigcap_{i=1}^k R_{G_i}$. Every edge $v_j v_k \in E(G)$ is contained in the edge set of (at least) one of the subgraphs, say G_i. The spline $\boldsymbol{p}$ is a generalized spline for G_i by hypothesis, so the GKM condition is satisfied at $v_j v_k$ in G_i and hence in G as well. □

Figure 7. Two spanning trees whose generalized splines determine R_{C_3}.

Theorem 6.1 generalizes the definition of R_G. Indeed, for each edge $e \in E(G)$, consider the subgraph $G_e = (V(G), \{e\})$. The ring R_{G_e} is exactly the subring of $R^{|V(G)|}$ defined by applying the GKM condition at just the edge e. Theorem 6.1 says

$$R_G = \bigcap_{e \in E(G)} R_{G_e},$$

namely, that the generalized splines on G are formed by imposing the GKM condition on every edge of G simultaneously.

The next corollary uses another common family of subgraphs: spanning trees. We completely identified the generalized splines for trees in Theorem 4.1. Thus, the corollary expresses the ring of generalized splines using far fewer intersections than in the original GKM formulation. Calculating intersections of subrings is subtle, so this corollary reduces the computational complexity of identifying the ring of generalized splines.

Corollary 6.2. *If G can be written as a union of spanning trees $T_1, T_2, \dots, T_m$ (whose edges are not necessarily disjoint) and if $\alpha_i = \alpha|_{T_i}$ is the edge labeling given by restriction for each i then*

$$R_G = \bigcap_{i=1}^{m} R_{T_i}.$$

Figure 7 shows an example using the 3-cycle and principal-ideal edge labels. In this case R_G can be expressed as the intersection of just two rings of generalized splines, each of which is completely known. In fact, Theorem 4.1 says that the generalized splines for the two marked paths have the form

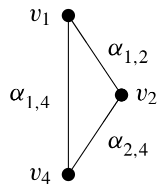
$$(p_1,\ p_1 + \alpha_{1,4} p_4 + \alpha_{2,4} p_2,\ p_1 + \alpha_{1,4} p_4)$$

and

$$(q_1,\ q_1 + \alpha_{1,2} q_2,\ q_1 + \alpha_{1,2} q_2 + \alpha_{2,4} q_4)$$

for free choices of elements $p_1, p_2, p_4, q_1, q_2, q_4 \in R$. The intersection of these two sets is R_{C_3}.

Given a connected graph G, we could also use Theorem 6.1 to describe R_G in terms of the generalized splines for cycles as follows. Fix a spanning tree T



$$\leftrightarrow \left[\begin{array}{ccc|c} 1 & -1 & 0 & q_{1,2}\alpha_{1,2} \\ 0 & 1 & -1 & q_{2,4}\alpha_{2,4} \\ 1 & 0 & -1 & q_{1,4}\alpha_{1,4} \end{array}\right] \leftrightarrow \left[\begin{array}{ccc|l} 1 & -1 & 0 & q_{1,2}\alpha_{1,2} \\ 0 & 1 & -1 & q_{2,4}\alpha_{2,4} \\ 0 & 0 & 0 & q_{1,4}\alpha_{1,4} - q_{1,2}\alpha_{1,2} - q_{2,4}\alpha_{2,4} \end{array}\right]$$

Figure 8. A triangle, its extended GKM matrix, and a row-reduction.

for G. For each edge $e \in E(G) - E(T)$ let C_e denote the unique cycle contained in $T \cup \{e\}$. (This cycle exists and is unique by a classical result in graph theory [West 2001, pp. 68–69].) Let C'_e be the graph containing the cycle C_e as one connected component and the rest of the vertices of G as the other connected components. Then

$$R_G = R_T \cap \bigcap_{e \in E(G)-E(T)} R_{C'_e} \tag{10}$$

by Theorem 6.1.

However, a natural row-reduction of the GKM matrix of G proves this intersection directly. To motivate our approach, we return to the complete graph on four vertices with principal-ideal edge labels from Example 3.9. The system of equations in Figure 6 is consistent precisely when $\boldsymbol{q} = (q_{1,2}, q_{1,3}, q_{1,4}, q_{2,3}, q_{3,4}) \in R^5$ satisfies the homogeneous system of equations

$$\begin{aligned} q_{1,2}\alpha_{1,2} - q_{1,4}\alpha_{1,4} + q_{2,4}\alpha_{2,4} &= 0, \\ q_{1,3}\alpha_{1,3} - q_{1,4}\alpha_{1,4} + q_{3,4}\alpha_{3,4} &= 0, \\ q_{2,3}\alpha_{2,3} - q_{2,4}\alpha_{2,4} + q_{3,4}\alpha_{3,4} &= 0. \end{aligned} \tag{11}$$

Figure 8 shows the edge-labeled 3-cycle v_1, v_2, v_4 of Figure 7, its extended GKM matrix, and a natural row-reduction of its extended GKM matrix. The equation that remains is (up to sign) the same as that which occurs in (11). In fact, the entire system in (11) arises from the equations (up to sign) for the three subcycles induced by the vertices

- v_1, v_2, v_4,
- v_1, v_3, v_4, and
- v_2, v_3, v_4.

The next theorem generalizes this example. We also see it, together with Remark 3.2, as a first step towards generalizing Rose's work on syzygies of edge ideals [1995; 2004].

Theorem 6.3. *Suppose that (G, α) is an edge-labeled graph on n vertices. Fix a spanning tree T for G. For each edge $e \in E(G) - E(T)$ let C_e denote the unique cycle contained in $T \cup \{e\}$. Then the extended GKM matrix for G is equivalent to an extended GKM matrix for T, followed for each edge $e \in E(G) - E(T)$ by a row that is zero except in the last column, which is $\sum_{e' \in C_e} q_{e'}$ where $q_{e'}$ are arbitrary elements of $\alpha(e')$.*

Proof. Choose a spanning tree T for the graph G. We assume without loss of generality that the first $n-1$ rows of the GKM matrix for G correspond to the edges in T. The first $n-1$ rows of the GKM matrix of G thus consist of the GKM matrix for T, by construction.

Consider each of the other rows in turn. Each row corresponds to an edge e in G but not T. We now describe an invertible row operation to eliminate all nonzero entries from the first n columns of the row corresponding to e and describe R_G more precisely. Denote the edges of the cycle C_e by $e_1 = e = v_{i_1}v_{i_2}$, $e_2 = v_{i_2}v_{i_3}$, $\ldots$, $e_k = v_{i_k}v_{i_1}$. Let $c_j \in \{\pm 1\}$ be the entry in the row corresponding to e_j and the column corresponding to vertex v_{i_j} for each $2 \leq j \neq n$. Denote the e_j-th row of the GKM matrix by r_{e_j}. The sum of the scaled rows,

$$\sum_{j=2}^{k} c_j r_{e_j},$$

has 1 in column v_{i_2}, -1 in column v_{i_1}, 0 in the rest of the first n columns, and $\sum_{j=2}^{k} c_j q_j$ in the last column, all by the definition of the GKM matrix. Finally we add $\sum_{j=2}^{k} c_j r_{e_j}$ to the row corresponding to e. This leaves 0 in the first n columns of row e and $q_e + \sum_{j=2}^{k} c_j q_j$ in the last entry of the row.

The elements q_e and q_j are arbitrary elements of their respective ideals and c_j is a unit in R for each j so the set of all possible $q_e + \sum_{j=2}^{k} c_j q_j$ is the same as the set of all possible $\sum_{e' \in C_e} q_{e'}$. The result follows. □

The last corollary uses this information to describe the generalized splines for G in terms of the generalized splines for cycles, as promised.

Corollary 6.4. *Suppose that (G, α) is an edge-labeled graph on n vertices. Fix a spanning tree T for G. For each edge $e \in E(G) - E(T)$ let C_e denote the unique cycle contained in $T \cup \{e\}$ together with the other vertices in G. Then*

$$R_G = R_T \cap \bigcap_{e \in E(G) - E(T)} R_{C'_e}.$$

Proof. Consider an edge e outside of the spanning tree T and its corresponding cycle C_e. The previous theorem showed that the submatrix of an extended GKM matrix for G given by the rows indexed by the edges $e' \in E(C_e)$ forms an extended GKM matrix for the cycle C_e. The vector $\boldsymbol{p} \in R^{|V|}$ solves an extended GKM matrix for G if and only if it simultaneously solves the corresponding extended GKM matrices for T and all of the C_e for $e \in E(G) - E(T)$. □

7. Open questions

We end with several open questions, extending some of the major research problems for splines and GKM theory to the context of generalized splines.

Most research into what we call generalized splines focuses on particular examples, whether because of explicit hypotheses (e.g., a particular choice of the ring R, the graph G, or the edge-labeling function α) or implicit hypotheses (e.g., that edge labels be principal). Special cases remain very important, both for applications and for data to build the general theory.

Question 7.1. Identify R_G in important special cases: for instance, when all edge labels $\alpha(e)$ are principal ideals; or when R is a particular ring (integers, polynomial rings, ring of Laurent polynomials); or when G is a particular graph or family of graphs (cycles, complete graphs, bipartite graphs, hypercubes).

Splines on complete graphs are particularly important for approximation theory, where they appear as the Alfeld split of a simplex (for a proof see [Tymoczko 2015, Section 3.1]).

Billera asked the following question, seeking an interpretation of r-smoothness in the context of equivariant cohomology. We extend Billera's question to ask about the analogue of r-smoothness for generalized splines over arbitrary rings.

Question 7.2. Let (G, α) be an edge-labeled graph. Define the function $\alpha^r : E \to \mathcal{I}$ by the condition that for each edge e the image $\alpha^r(e)$ is the r-th power $(\alpha(e))^r$. The r-smooth generalized splines are the elements of the ring R_{G,α^r}. We ask how the r-smooth generalized splines compare for various r. Billera asks for a geometric interpretation of r-smoothness in the context of equivariant cohomology rings.

As a module, the generalized splines R_G can also be viewed as group representations: for instance, the group of automorphisms of the graph G that preserve the edge labeling naturally induces a representation on R_G. Representations obtained in this and similar ways are often intrinsically interesting [Fiebig 2011; Tymoczko 2008] and can also be a powerful tool with which to approach other questions in this section [Tymoczko 2008].

Question 7.3. Given a specific automorphism group, what are the induced representations on R_G (in terms of irreducible representations, say)? For what families of graphs are there nontrivial representations on R_G?

Propositions 2.8 and 2.10 and Sections 5 and 6 all use combinatorial aspects of graphs to analyze the ring of generalized splines. More recently, Handschy, Melnick, and Reinders [Handschy et al. 2014] and Bowden, Cao, Hagen, King, and Reinders [Bowden et al. 2015] have used deletion and contraction to study splines on cycles. We believe that these are special cases of a more general relationship between the underlying combinatorics and geometry.

Question 7.4. How do classical graph-theoretic constructions (such as deletion-contraction) affect the algebraic structure of splines R_G?

Theorems 2.12, 4.1 and 5.1 are part of a larger program to identify useful bases for splines and GKM modules [Haas 1991; Goldin and Tolman 2009; Guillemin and Zara 2003]. The next question extends that program to generalized splines.

Question 7.5. Given a graph G, find a minimal generating set (or basis, if R is an integral domain) for the generalized splines R_G. If G is a particular family of graphs (cycles, complete graphs, etc.), can we find a minimal generating set (or basis) for R_G?

More specifically, geometers think about bases with certain "upper-triangularity" properties that arise in many important examples, like Schubert classes, Białynicki-Birula classes, and the canonical classes of [Knutson and Tao 2003] and [Goldin and Tolman 2009] (see also [Harada and Tymoczko 2010]). Theorem 5.1 is an initial step in constructing *flow-up bases* for generalized splines.

Question 7.6. What is the right definition for a flow-up class in the module of generalized splines? Under what conditions is there a flow-up basis for the generalized splines?

Answering the previous question may require further extending generalized splines so that the vertices are labeled by different modules M_v rather than a fixed ring R, as described in Section 1. Characterizing those splines would have immediate implications in geometric applications like computing equivariant intersection homology.

Question 7.7. Which of the results in this paper extend to generalized splines over modules? Is there an algorithm or an explicit formula to construct flow-up basis classes for generalized spines over modules?

References

[Alfeld 1986] P. Alfeld, "On the dimension of multivariate piecewise polynomials", pp. 1–23 in *Numerical analysis* (Dundee, 1985), edited by D. F. Griffiths and G. A. Watson, Pitman Research Notes in Mathematics Series **140**, Longman Scientific and Technical, Harlow, 1986. MR 88d:41014 Zbl 0655.41010

[Alfeld 1987] P. Alfeld, "A case study of multivariate piecewise polynomials", pp. 149–159 in *Geometric modeling: algorithms and new trends* (Albany, NY, 1985), edited by G. E. Farin, Society for Industrial and Applied Mathematics, Philadelphia, 1987. MR 936451

[Alfeld and Schumaker 1987] P. Alfeld and L. L. Schumaker, "The dimension of bivariate spline spaces of smoothness r for degree $d \geq 4r+1$", *Constr. Approx.* **3**:2 (1987), 189–197. MR 88e:41025 Zbl 0646.41008

[Alfeld and Schumaker 1990] P. Alfeld and L. L. Schumaker, "On the dimension of bivariate spline spaces of smoothness r and degree $d = 3r + 1$", *Numer. Math.* **57**:6-7 (1990), 651–661. MR 91h:65022 Zbl 0725.41012

[Alfeld et al. 1987a] P. Alfeld, B. Piper, and L. L. Schumaker, "An explicit basis for C^1 quartic bivariate splines", *SIAM J. Numer. Anal.* **24**:4 (1987), 891–911. MR 88i:41014 Zbl 0658.65008

[Alfeld et al. 1987b] P. Alfeld, B. Piper, and L. L. Schumaker, "Minimally supported bases for spaces of bivariate piecewise polynomials of smoothness r and degree $d \geq 4r + 1$", *Comput. Aided Geom. Design* **4**:1-2 (1987), 105–123. MR 88d:65019 Zbl 0668.41011

[Bahri et al. 2009] A. Bahri, M. Franz, and N. Ray, "The equivariant cohomology ring of weighted projective space", *Math. Proc. Cambridge Philos. Soc.* **146**:2 (2009), 395–405. MR 2010a:57054 Zbl 1205.14022

[Bartels 1984] R. H. Bartels, "Splines in interactive computer graphics", pp. 1–29 in *Numerical analysis* (Dundee, 1983), edited by D. F. Griffiths, Lecture Notes in Mathematics **1066**, Springer, Berlin, 1984. MR 86b:65011 Zbl 0544.65009

[Billera 1988] L. J. Billera, "Homology of smooth splines: generic triangulations and a conjecture of Strang", *Trans. Amer. Math. Soc.* **310**:1 (1988), 325–340. MR 89k:41010 Zbl 0718.41017

[Billera and Rose 1991] L. J. Billera and L. L. Rose, "A dimension series for multivariate splines", *Discrete Comput. Geom.* **6**:2 (1991), 107–128. MR 92g:41010 Zbl 0725.13011

[Billera and Rose 1992] L. J. Billera and L. L. Rose, "Modules of piecewise polynomials and their freeness", *Math. Z.* **209**:4 (1992), 485–497. MR 93f:52017 Zbl 0891.13004

[Bowden and Tymoczko 2015] N. Bowden and J. S. Tymoczko, "Splines mod m", preprint, 2015. arXiv 1501.02027

[Bowden et al. 2015] N. Bowden, S. Hagen, M. King, and S. Reinders, "Bases and structure constants of generalized splines with integer coefficients on cycles", preprint, 2015. arXiv 1502.00176

[Braden and MacPherson 2001] T. Braden and R. MacPherson, "From moment graphs to intersection cohomology", *Math. Ann.* **321**:3 (2001), 533–551. MR 2003g:14030 Zbl 1077.14522

[Brion 1996] M. Brion, "Piecewise polynomial functions, convex polytopes and enumerative geometry", pp. 25–44 in *Parameter spaces* (Warsaw, 1994), edited by P. Pragacz, Banach Center Publications **36**, Polish Academy of Science, Warsaw, 1996. MR 99h:14052 Zbl 0878.14035

[Chang and Skjelbred 1974] T. Chang and T. Skjelbred, "The topological Schur lemma and related results", *Ann. of Math.* (2) **100** (1974), 307–321. MR 51 #11552 Zbl 0249.57023

[Chui and He 1989] C. K. Chui and T. X. He, "On the dimension of bivariate superspline spaces", *Math. Comp.* **53**:187 (1989), 219–234. MR 90c:41014 Zbl 0668.41013

[Chui and Lai 1985] C. K. Chui and M. J. Lai, "On bivariate vertex splines", pp. 84–115 in *Multivariate approximation theory, III* (Oberwolfach, 1985), edited by W. Schempp and K. Zeller, Internationale Schriftenreihe zur Numerischen Mathematik **75**, Birkhäuser, Basel, 1985. MR 88m:41002 Zbl 0588.65009

[Chui and Lai 1990] C. K. Chui and M. J. Lai, "Multivariate vertex splines and finite elements", *J. Approx. Theory* **60**:3 (1990), 245–343. MR 91e:65022 Zbl 0702.41017

[Dalbec and Schenck 2001] J. P. Dalbec and H. K. Schenck, "On a conjecture of Rose", *J. Pure Appl. Algebra* **165**:2 (2001), 151–154. MR 2002i:41046 Zbl 1089.41029

[DiPasquale 2012] M. R. DiPasquale, "Shellability and freeness of continuous splines", *J. Pure Appl. Algebra* **216**:11 (2012), 2519–2523. MR 2927183 Zbl 1300.13012

[Fiebig 2011] P. Fiebig, "Sheaves on affine Schubert varieties, modular representations, and Lusztig's conjecture", *J. Amer. Math. Soc.* **24**:1 (2011), 133–181. MR 2012a:20072 Zbl 1270.20053

[Gmelig Meyling and Pfluger 1985] R. H. J. Gmelig Meyling and P. R. Pfluger, "On the dimension of the spline space $S_2^1(\Delta)$ in special cases", pp. 180–190 in *Multivariate approximation theory, III* (Oberwolfach, 1985), edited by W. Schempp and K. Zeller, Internationale Schriftenreihe zur Numerischen Mathematik **75**, Birkhäuser, Basel, 1985. MR 88f:65024 Zbl 0567.41015

[Goldin and Tolman 2009] R. F. Goldin and S. Tolman, "Towards generalizing Schubert calculus in the symplectic category", *J. Symplectic Geom.* **7**:4 (2009), 449–473. MR 2010k:53143 Zbl 1185.57024

[Goresky et al. 1998] M. Goresky, R. Kottwitz, and R. MacPherson, "Equivariant cohomology, Koszul duality, and the localization theorem", *Invent. Math.* **131**:1 (1998), 25–83. MR 99c:55009 Zbl 0897.22009

[Guillemin and Zara 2001] V. Guillemin and C. Zara, "1-skeleta, Betti numbers, and equivariant cohomology", *Duke Math. J.* **107**:2 (2001), 283–349. MR 2002j:53110 Zbl 1020.57013

[Guillemin and Zara 2003] V. Guillemin and C. Zara, "The existence of generating families for the cohomology ring of a graph", *Adv. Math.* **174**:1 (2003), 115–153. MR 2004f:53099 Zbl 1020.57014

[Guillemin et al. 2006] V. Guillemin, T. S. Holm, and C. Zara, "A GKM description of the equivariant cohomology ring of a homogeneous space", *J. Algebraic Combin.* **23**:1 (2006), 21–41. MR 2008b:55010 Zbl 1096.53050

[Haas 1991] R. Haas, "Module and vector space bases for spline spaces", *J. Approx. Theory* **65**:1 (1991), 73–89. MR 92b:41021 Zbl 0732.41014

[Handschy et al. 2014] M. Handschy, J. Melnick, and S. Reinders, "Integer generalized splines on cycles", preprint, 2014. arXiv 1409.1481

[Harada and Tymoczko 2010] M. Harada and J. S. Tymoczko, "Poset pinball, GKM-compatible subspaces, and Hessenberg varieties", preprint, 2010. arXiv 1007.2750

[Harada et al. 2005] M. Harada, A. Henriques, and T. S. Holm, "Computation of generalized equivariant cohomologies of Kac–Moody flag varieties", *Adv. Math.* **197**:1 (2005), 198–221. MR 2006h:53086 Zbl 1110.55003

[Holm 2008] T. S. Holm, "Act globally, compute locally: group actions, fixed points, and localization", pp. 179–195 in *Toric topology*, edited by M. Harada et al., Contemporary Mathematics **460**, American Mathematical Society, Providence, RI, 2008. MR 2009h:53191 Zbl 1155.53051

[Knutson and Tao 2003] A. Knutson and T. Tao, "Puzzles and (equivariant) cohomology of Grassmannians", *Duke Math. J.* **119**:2 (2003), 221–260. MR 2006a:14088 Zbl 1064.14063

[McDonald 2007] T. McDonald, "Splines with boundary conditions", *Comput. Math. Appl.* **54**:9-10 (2007), 1234–1239. MR 2009b:65021 Zbl 1140.41001

[McDonald and Schenck 2009] T. McDonald and H. K. Schenck, "Piecewise polynomials on polyhedral complexes", *Adv. in Appl. Math.* **42**:1 (2009), 82–93. MR 2009k:41013 Zbl 1178.41008

[Morgan and Scott 1975] J. Morgan and R. Scott, "A nodal basis for C^1 piecewise polynomials of degree $n \geq 5$", *Math. Comp.* **29**:131 (1975), 736–740. MR 51 #11930 Zbl 0307.65074

[Mourrain and Villamizar 2013] B. Mourrain and N. Villamizar, "Homological techniques for the analysis of the dimension of triangular spline spaces", *J. Symbolic Comput.* **50** (2013), 564–577. MR 2996896 Zbl 06124247

[Payne 2006] S. Payne, "Equivariant Chow cohomology of toric varieties", *Math. Res. Lett.* **13**:1 (2006), 29–41. MR 2007f:14052 Zbl 1094.14036

[Rose 1995] L. L. Rose, "Combinatorial and topological invariants of modules of piecewise polynomials", *Adv. Math.* **116**:1 (1995), 34–45. MR 97b:13036 Zbl 0837.57020

[Rose 2004] L. L. Rose, "Graphs, syzygies, and multivariate splines", *Discrete Comput. Geom.* **32**:4 (2004), 623–637. MR 2005g:41024 Zbl 1069.41012

[Schenck 1997] H. K. Schenck, *Homological methods in the theory of splines*, thesis, Cornell University, Ithaca, NY, 1997, available at http://search.proquest.com/docview/304343601. MR 2695686

[Schenck 2012] H. K. Schenck, "Equivariant Chow cohomology of nonsimplicial toric varieties", *Trans. Amer. Math. Soc.* **364**:8 (2012), 4041–4051. MR 2912444 Zbl 06191303

[Schenck and Stillman 1997] H. K. Schenck and M. Stillman, "Local cohomology of bivariate splines", *J. Pure Appl. Algebra* **117/118** (1997), 535–548. MR 99d:13011 Zbl 0902.41010

[Schumaker 1979] L. L. Schumaker, "Lower bounds for spline approximation", pp. 213–223 in *Approximation theory: VIth semester, Stefan Banach International Mathematics Center* (Warsaw, 1975), edited by Z. Ciesielski, Banach Center Publications **4**, Polish Academy of Science, Warsaw, 1979. MR 81h:41011 Zbl 0418.41005

[Schumaker 1984a] L. L. Schumaker, "Bounds on the dimension of spaces of multivariate piecewise polynomials", *Rocky Mountain J. Math.* **14**:1 (1984), 251–264. MR 85h:41091 Zbl 0601.41034

[Schumaker 1984b] L. L. Schumaker, "On spaces of piecewise polynomials in two variables", pp. 151–197 in *Approximation theory and spline functions* (St. John's, Newfoundland, 1983), edited by S. P. Singh et al., NATO Advanced Science Institute Series C Mathematical and Physical Sciences **136**, Reidel, Dordrecht, 1984. MR 86j:41009 Zbl 0553.41018

[Schumaker 1988] L. L. Schumaker, "Dual bases for spline spaces on cells", *Comput. Aided Geom. Design* **5**:4 (1988), 277–284. MR 90a:41013 Zbl 0652.41012

[Tymoczko 2005] J. S. Tymoczko, "An introduction to equivariant cohomology and homology, following Goresky, Kottwitz, and MacPherson", pp. 169–188 in *Snowbird lectures in algebraic geometry*, edited by R. Vakil, Contemporary Mathematics **388**, American Mathematical Society, Providence, RI, 2005. MR 2006m:55019

[Tymoczko 2008] J. S. Tymoczko, "Permutation actions on equivariant cohomology of flag varieties", pp. 365–384 in *Toric topology*, edited by M. Harada et al., Contemporary Mathematics **460**, American Mathematical Society, Providence, RI, 2008. MR 2010i:14090 Zbl 1147.14024

[Tymoczko 2015] J. S. Tymoczko, "Splines in geometry and topology", *Comput. Aided Geom. Design* (online publication November 2015).

[Wang 2000] R.-H. Wang, "Multivariate spline and algebraic geometry", *J. Comput. Appl. Math.* **121**:1-2 (2000), 153–163. MR 1780047 Zbl 0960.41008

[West 2001] D. B. West, *Introduction to graph theory*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 2001. MR 96i:05001 Zbl 0845.05001

[Yuzvinsky 1992] S. Yuzvinsky, "Modules of splines on polyhedral complexes", *Math. Z.* **210**:2 (1992), 245–254. MR 93h:52015 Zbl 0848.52003

[Zhu and Wang 2005] C.-G. Zhu and R.-H. Wang, "Piecewise semialgebraic sets", *J. Comput. Math.* **23**:5 (2005), 503–512. MR 2006d:14071 Zbl 1084.14058

[Zhu and Wang 2011] C.-G. Zhu and R.-H. Wang, "The correspondence between multivariate spline ideals and piecewise algebraic varieties", *J. Comput. Appl. Math.* **236**:5 (2011), 793–800. MR 2853504 Zbl 1236.65012

Received March 4, 2015. Revised August 20, 2015.

SIMCHA GILBERT
DEPARTMENT OF MATHEMATICS AND STATISTICS
SMITH COLLEGE
44 COLLEGE LANE
NORTHAMPTON, MA 01063
UNITED STATES
sgilbert@smith.edu

JULIANNA TYMOCZKO
DEPARTMENT OF MATHEMATICS AND STATISTICS
SMITH COLLEGE
44 COLLEGE LANE
NORTHAMPTON, MA 01063
UNITED STATES
jtymoczko@smith.edu

SHIRA VIEL
DEPARTMENT OF MATHEMATICS
NORTH CAROLINA STATE UNIVERSITY
2108 SAS HALL
RALEIGH, NC 27695
UNITED STATES
spolste@ncsu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 281, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.281.365

GOOD TRACES FOR NOT NECESSARILY SIMPLE DIMENSION GROUPS

DAVID HANDELMAN

Akin's notion of good measure, introduced to classify measures on Cantor sets, has been translated to dimension groups and traces by Bezuglyi and the author, but emphasizing the simple (minimal dynamical system) case. Here we permit nonsimplicity. Goodness of tensor products of large classes of non-good traces (measures) is established. We also determine the pure faithful good traces on the dimension groups associated to xerox-type actions on AF C*-algebras; the criteria turn out to involve algebraic geometry and number theory.

We also deal with a coproduct of dimension groups, wherein, despite expectations, goodness of direct sums is nontrivial. In addition, we verify a conjecture of Bezuglyi and Handelman (2014) concerning good subsets of Choquet simplices, in the finite-dimensional case.

Introduction and definitions

Akin [1999; 2005] (see also [Akin et al. 2008], among others) introduced and studied the notion of good measures in connection with the classification of (probability) measures on Cantor sets up to homeomorphism. With the development in [Putnam

Supported in part by a Discovery grant from NSERC.
MSC2010: primary 19K14; secondary 06F20, 46L80, 37A55, 37A35, 46A55, 28E99.
Keywords: good measure, trace, dimension group, affine representation, Newton polyhedron, approximately divisible, order ideal, simplicial group.

1989; Herman et al. 1992; Giordano et al. 1995], among others, of classification and construction of minimal actions with respect to strong orbit and orbit equivalence via Vershik maps and ordered Grothendieck groups of AF C*-algebras, this and related properties were translated into the language of (traces on) dimension groups (a class of partially ordered abelian groups) in [Bezuglyi and Handelman 2014], henceforth abbreviated [BeH 2014]. In particular, the characterizations therein of goodness of traces on simple dimension groups provided relatively easy constructions of good and non-good measures on minimal systems. For more details, see the discussion in the introduction to [BeH 2014].

Recent work (e.g., [Medynets 2006; Frick and Ormes 2013; Petersen 2012]) has extended Vershik action(s) to nonminimal systems, and correspondingly to nonsimple dimension groups. Here we give computable criteria for goodness in the general (approximately divisible) case, and then use the criteria to give a surprising result that tensor products of (some) non-good traces are good; this applies to the ugly traces of [BeH 2014]. We also completely determine the pure faithful traces on fixed point algebras under xerox actions of tori: the latter include Pascal's triangle and variations corresponding to spatially and temporally homogeneous random walks with finite support on the lattice $\mathbb{Z}^d$.

From [Handelman 1985, Theorem III.3], the pure faithful traces correspond to points $r = (r_i)$ in the strictly positive orthant of $\mathbb{R}^d$; those that are good are precisely the algebraic points that satisfy two number-theoretic conditions, which in the case that $d = 1$, reduce to (i) no other algebraic conjugate of $r = r_1$ is positive and (ii) if the leading and terminal coefficients of the polynomial implementing the random walk are a_0 and a_k, then there exists s such that a_0^s/r and $a_k^s r$ are both algebraic integers.

We also deal with a strict form of direct sum of dimension groups, determining when the corresponding sum of traces is good; there are some surprises here, as the direct sum can be good without either one being good. We find for each m, a collection of simple dimension groups with traces, (G_i, τ_i), such that for any strict direct sum of m or fewer distinct summands, $\bigoplus_{i \in S} G_i$, the sum of the traces is not good, but for any sum of more than m direct summands, the sum *is* good.

We then consider good sets of traces. The first problem is the definition; it should be consistent with the current definition in the simple case and in the singleton case, and we discuss various possibilities; finally, we settle on one. We show that for the class of dimension groups considered above (arising from random walks on $\mathbb{Z}^d$), with any reasonable definition, the notion is surprisingly restrictive, and even order unit goodness turns out to be sensitive to the Newton polyhedra of the polynomials (unlike the case for single traces).

There are three appendices. The first discusses connections with dynamical systems, mostly for simple dimension groups. The second characterizes order unit good traces on simplicial groups, and the resulting characterization suggests that there are no effective criteria for goodness or order unit goodness when there are discrete traces, in contrast to the approximately divisible situation discussed in the rest of this article. Appendix C verifies, in the case of a finite-dimensional trace space, a conjecture made in [BeH 2014, Section 7] concerning the structure of good subsets relative to a simplex.

Definitions. A partially ordered abelian group G with positive cone G^+ is *unperforated* if whenever n is a positive integer and $g \in G$, then $ng \in G^+$ entails $g \in G^+$. An *order unit* for G is an element $u \in G^+$ such that for all $g \in G$, there exists a positive integer K such that $-Ku \leq g \leq Ku$. A *trace* (formerly, *state*) is a nonzero positive group homomorphism $\tau : G \to \mathbb{R}$; if $\tau(u) = 1$ and u is an order unit, we say τ is normalized (with respect to u). The trace τ is *faithful* if $\ker \tau \cap G^+ = \{0\}$ (this is *much* weaker than being one-to-one, and corresponds to faithfulness of the corresponding measure when there is a dynamical system nearby).

When (G, u) is a partially ordered abelian group with order unit, we may form $S(G, u)$, the compact convex set of normalized traces, equipped with the weak (or point-open) topology. We denote by $\operatorname{Aff} S(G, u)$ the Banach space of continuous convex-linear (affine) real-valued functions on $S(G, u)$. There is a natural representation $G \to \operatorname{Aff} S(G, u)$, given by $g \mapsto \hat{g}$, where $\hat{g}(\tau) = \tau(g)$. We call this the *affine representation of* (G, u). If h in $\operatorname{Aff} S(G, u)$ is strictly positive (as a function on $S(G, u)$), we write $h \gg 0$. When G is unperforated, we may use the notation $g \gg 0$ or $0 \ll g$ to indicate that g is an order unit; this is consistent, as $\hat{g} \gg 0$ if and only if g is an order unit.

If $Y \subset S(G, u)$, we define $Y^{\vdash} = \{h \in \operatorname{Aff} S(G, u) \mid h|Y \equiv 0\}$; when $Y = \{\tau\}$, a singleton, we abbreviate this to $\tau^{\vdash}$. In this case, $\tau^{\vdash}$ is a codimension-one subspace of $\operatorname{Aff} S(G, u)$ and is an order ideal if and only if τ is pure. Following the convention in [BeH 2014], we signal purity with the replacement notation $\tau^{\perp}$.

If (G, u) is an unperforated ordered abelian group, we say G is *approximately divisible* if its range in $\operatorname{Aff} S(G, u)$ is norm-dense; for dimension groups with order unit, this is equivalent to $\tau(G)$ being dense in $\mathbb{R}$ for all pure traces τ, or equivalently, for all order units $g \in G$, there exist order units a, b of G such that $g = 2a + 3b$ (and there are many other equivalent formulations) [Handelman 2014, Corollary 6.2].

When I is a subgroup (typically an order ideal) of a partially ordered abelian group G, we say *I has its own order unit w* or *w is a relative order unit of I* if $w \in I$ is an order unit of I with respect to the relative ordering inherited from G. This is to emphasize the fact that w is *not* an order unit for G, merely for I.

If G is an unperforated ordered abelian group, we say G is *nearly divisible* if for every order ideal (I, w) which has its own order unit, (I, w) is approximately divisible; from the discussion above, an equivalent form not referring to order ideals is that for all $g \in G^+$, there exist $a, b \in G^+$ such that $g = 2a + 3b$ and $g \leq ka, kb$ for some positive integer k. This appears to be a new concept.

For example, if $G = H \otimes U$, where H is a partially ordered unperforated abelian group and U is a noncyclic subgroup of the rationals $\mathbb{Q}$, then G is nearly divisible, and it is approximately divisible if it has an order unit. We will see plenty of nearly divisible examples that are not of this type in later sections.

A trace on G is *discrete* if its image $\tau(G)$ is a cyclic (that is, discrete) subgroup of $\mathbb{R}$. An alternative characterization of approximately divisible, for dimension groups, is that (G, u) admit no discrete traces; for nearly divisible, the characterization is that no nonzero order ideal with order unit admits a discrete trace.

For general relevant results on partially ordered abelian groups, especially dimension groups, see [Goodearl 1986].

An *interval* in a partially ordered group G is a subset of the form $[0, b] := \{g \in G \mid 0 \leq g \leq b\}$ for some $b \in G^+$.

Following [BeH 2014], and based on Akin's notion for measures on Cantor sets, a trace $\tau : G \to \mathbb{R}$ is *good* (as a trace of G) if for all $b \in G^+$, we have $\tau([0, b]) = [0, \tau(b)] \cap \tau(G)$; that is, if $a' \in G$ and $0 \leq \tau(a') \leq \tau(b)$, there exists $a \in [0, b]$ such that $a - a' \in \ker \tau$. If (G, u) is a partially ordered abelian group with order unit, we say τ is *order unit good* if in the definition of good, we restrict b to be an order unit.

1. Characterization of goodness

Order unit goodness is relatively easy to characterize when (G, u) is an approximately divisible dimension group [BeH 2014, Proposition 1.7]: τ is order unit good if and only if the image of $\ker \tau$ in $\operatorname{Aff} S(G, u)$ is dense in

$$\tau^{\vdash} := \{h \in \operatorname{Aff} S(G, u) \mid h(\tau) = 0\}$$

(the latter is a closed codimension-one subspace of $\operatorname{Aff} S(G, u)$). This makes examples and non-examples relatively easy to construct. There is a corresponding characterization for goodness, which we shall simplify a bit, and use to actually do something.

Proposition 1.1. *Suppose (G, u) is a dimension group with order unit. Let τ be a faithful trace of G. Then τ is good if and only if for all nonzero order ideals with order unit (I, w), both $\tau(I) = \tau(G)$ and $\tau|I$ is order unit good. If τ is pure, then sufficient for goodness is that there exist an order ideal I such that $\tau|I$ is good and $\tau(I^+) = \tau(G^+)$.*

Remark. Necessity is shown in [BeH 2014, Proposition 4.2]; although the statement hypothesizes that τ be pure, this is not used in the proof (it *is* used there in the proof of sufficiency); also shown there was that if τ is good, then $\tau|I$ is good (as a trace on the order ideal I), and this implies (in the case that I is approximately divisible) that $\tau|I$ is order unit good, just from the definitions.

Remark. It is always possible to reduce to the case that τ is faithful by factoring out the maximal order ideal J contained in $\ker\tau$ [BeH 2014, Lemma 4.4]. In this case, the criteria apply to G/J (replacing G). This would make the statement somewhat more complicated.

Proof. Proof of necessity is given in [BeH 2014, Proposition 4.2], requiring neither purity of τ nor approximate divisibility.

Conversely, suppose $a \in G$, $b \in G^+$ and $0 < \tau(a) < \tau(b)$. Form the order ideal I generated by b, that is, $I = \{c \in G \mid \exists N \in \mathbb{N}$ such that $-Nb \le g \le Nb\}$. Then I is an order ideal with its own order unit, b. Since $\tau(I^+) = \tau(G^+)$, we have $\tau(I) = \tau(G)$, and thus there exists $a_1 \in I$ such that $\tau(a_1) = \tau(a)$. Now order unit goodness of $\tau|I$ yields $a' \in I$ such that $\tau(a') = \tau(a_1) = \tau(a)$ and $0 \le a' \le b$, verifying goodness of τ.

The final statement is just the sufficiency condition of [BeH 2014, Proposition 4.2]. □

Let G be a dimension group, and let I and J be order ideals thereof. Then $H := I + J$ (the set of sums of elements in I and J) and $I \cap J$ are both order ideals. Most of the following are variations on [BeH 2014, Lemma 1.3]. As in [BeH 2014], an element v of G^+ is *τ-good* or *τ-order unit good* if $\tau([0, v]) = [0, \tau(v)] \cap \tau(G)$.

Lemma 1.2. *Suppose G is a dimension group, and I and J each have relative order units, w, y respectively. Then:*

(a) *$I + J$ is an order ideal of G with a relative order unit.*

(b) *Let τ be a trace on G such that* $\ker\tau \cap G^+ = \{0\}$ *and $\tau(I) \cap \tau(J)$ is dense in $\mathbb{R}$. If $\tau|I$ and $\tau|J$ are good (as traces on I and J respectively), then $\tau|(I + J)$ is good.*

(c) *If $I + J$ is approximately divisible, then every order unit b of $I + J$ can be written in the form $b = u + v$, where u, v are relative order units for I, J respectively.*

(d) *If v is τ-order unit good (with respect to I) and w is τ-order unit good (with respect to J), and $\tau(I) \cap \tau(J)$ is dense in $\mathbb{R}$, then $v + w$ is τ-order unit good with respect to $I + J$.*

(e) *Suppose that each of I, J and $I + J$ is approximately divisible, and τ is a trace on $I + J$ such that each of $\tau|I$ and $\tau|J$ is order unit good, and $\tau(I) \cap \tau(J)$ is dense in $\mathbb{R}$. Then τ is order unit good as a trace of $I + J$.*

Remark. Part (c) can fail if approximate divisibility is dropped; for example, take $G = \mathbb{Z}^3$ with the usual simplicial ordering, let I be the order ideal generated by $(1, 1, 0)$ and let J be the order ideal generated by $(0, 1, 1)$; then $I + J = G$ and the order unit $(1, 1, 1)$ cannot be realized as a sum of relative order units from I and J respectively.

Proof. (a) That $I+J$ is an order ideal is ancient; see, e.g., [Goodearl 1986]. If w and y are respective order units for I and J, then $z := w+y$ is an order unit for $I+J$. To see this, let $f \in (I+J)^+$; for dimension groups, $(I+J)^+ = I^+ + J^+$, hence we can find $e \in I^+$ and $g \in J^+$ such that $f = e+g$. Since there exist positive integers k, k' such that $e \leq kw$ and $g \leq k'v$, we have $f \leq k''z$, where $k'' = \max\{k, k'\}$.

(b) Select $b \in G^+$ and $a \in G$ such that $\tau(a) < \tau(b)$. We may write $b = i + j$, where $i \in I^+$ and $j \in J^+$. Then $\tau(i), \tau(j) > 0$. We may write $\tau(a) = r + s$, where $r \in \tau(I)$ and $s \in \tau(J)$.

Assume $\tau(a) \geq \tau(i)$. By density of $\tau(I) \cap \tau(J)$, given

$$0 < \epsilon < \min\{\tau(i), \tau(b) - \tau(a)\},$$

there exists $\delta \in \tau(I) \cap \tau(J)$ such that $\tau(i) - \epsilon < r + \delta < \tau(i)$. Then $s - \delta = \tau(a) - r - \delta$ satisfies

$$\tau(a) - \tau(i) + \epsilon > s - \delta > \tau(a) - \tau(i) > 0.$$

Hence we can write $\tau(a) = (r+\delta) + (s-\delta)$, where the parenthesized terms are respectively in the intervals $(0, \tau(i))$ and $(0, \tau(a) - \tau(i) + \epsilon)$. However, $\epsilon < \tau(b) - \tau(a)$ entails $\tau(a) - \tau(i) + \epsilon < \tau(b) - \tau(i) = \tau(j)$. Since $\pm\delta \in \tau(I \cap J)$, we may thus find $a_1 \in I$ and $a_2 \in J$ such that $0 < \tau(a_1) < \tau(i)$ and $0 < \tau(a_2) < \tau(j)$. Since each of $\tau|I$ and $\tau|J$ is good, there exist $c_1 \in [0, i]$ (the interval in I) and $c_2 \in [0, j]$ such that $\tau(c_1) < \tau(i)$ and $\tau(c_2) < \tau(j)$. Hence we have $c := c_1 + c_2 \in [0, b]$ and $\tau(c) = \tau(c_1) + \tau(c_2) < \tau(i) + \tau(j) = \tau(b)$, verifying goodness in this case.

Reversing the roles of i and j, the same conclusion results if $\tau(a) \geq \tau(j)$, so we are reduced to the case that $\tau(a) < \min\{\tau(i), \tau(j)\}$. If $\tau(a) = 0$, there is nothing to do (except set $c = 0$). Otherwise, choose $0 < \epsilon < \tau(a)/2$, find real $\delta \in \tau(I \cap J)$ such that $\tau(a)/2 - \epsilon < \delta + r < \tau(a)/2$, and consider $\tau(a) = (r+\delta) + (s-\delta)$; then $r + \delta \in (0, \tau(a)/2) \subset (0, \tau(i))$, so $s - \delta \in (\tau(a)/2, \tau(a)) \subset (0, \tau(j))$. Now we can proceed as in the previous paragraph.

(c) Now let b be an order unit of $I+J$. By approximate divisibility of $I+J$, the range of $I+J$ in $\operatorname{Aff} S(I+J, b)$ is dense; hence given $\epsilon > 0$, we may find $b_0 \in I+J$ such that $(1/2 - \epsilon)\mathbf{1} < \hat{b}_0 < \mathbf{1/2}$ (where $\hat{\ }$ refers only to the representation on $S(I+J, b)$, that is, $\hat{b} = \mathbf{1}$). Let $\epsilon < 1/8$, so that $\hat{b}_0 \gg 0$ and thus b_0 is an order unit of $I+J$, and moreover, $2b_0 \leq b$, and $b - b_0$ is also an order unit for $I+J$.

Now consider the set $\mathcal{S} := \{c \in I^+ \mid c \leq b_0\}$. This is directed, as if $c, c' \in \mathcal{S}$, then we have $c, c' \leq b_0, c+c'$; interpolating, we obtain c'' such that $c, c' \leq c'' \leq b_0, c+c'$; as $c + c' \in I$, it follows that $c'' \in I$, so $c'' \in \mathcal{S}$. As there exists k such that $w \leq kb_0$, we can write $w = \sum_{i=1}^k w_i$, where $w_i \in I^+$ and each $w_i \leq b_0$. Then $w_i \in \mathcal{S}$, so there exists $u_0 \in I^+$ such that $w_i \leq u_0 \leq b_0$ for all i. Since $\sum w_i = w$ is an order unit for I, we know that ku_0 is an order unit for I, and thus u_0 is too. Hence there exists an order unit u_0 of I such that $u_0 \leq b_0$.

Since $b-b_0$ is also an order unit for $I+J$, applying the same process to J instead of I yields an order unit v_0 of J such that $v_0 \le b-b_0$. Thus $u_0+v_0 \le b_0+(b-b_0)=b$. The element $b-(u_0+v_0)$ is in the positive cone of $I+J$, so it can be written as $b-(u_0+v_0)=c+d$, where $c \in I^+$ and $d \in J^+$. This yields $b=(u_0+c)+(v_0+d)$; setting $u=u_0+c$, we see that $u \in I^+$ and is larger than an order unit for I, and so is itself an order unit for I; similarly $v=v_0+d$ is an order unit for J.

(d)–(e) Select an order unit b for $I+J$, and $a \in I+J$ such that $0<\tau(a)<\tau(b)$. By (c), we may write $b=u+v$, where u and v are order units for I and J respectively. We can write $a=r+s$, where $r \in I$ and $s \in J$, and set $t=\tau(u)$ (as $\tau|I$ is order unit good, it does not vanish identically, hence $t>0$), so that $\tau(v)=\tau(b)-t$, which is again positive. Now proceed as in the proof of (b). □

The density requirement on $\tau(I)\cap\tau(J)$ is essential.

Lemma 1.3. *Suppose that u and v are elements of G^+, and let τ be a trace such that each is τ-order unit good on the order ideals they generate, $I(u)$ and $I(v)$ respectively.*

- (a) *If $u+v$ is τ-order unit good on $I(u)+I(v)=I(u+v)$ and $\tau(I(u))+\tau(I(v))$ is dense in $\mathbb{R}$, then $\tau(I(u))\cap\tau(I(v))\neq\{0\}$.*
- (b) *If, additionally, both $\tau(I(u))$ and $\tau(I(v))$ are dense subgroups of $\mathbb{R}$, then so is $\tau(I(u))\cap\tau(I(v))$.*

Proof. Suppose the intersection consists of just 0. We may find positive real numbers $s \in \tau(I(u))$ and $t \in \tau(I(v))$ such that $s>\tau(u)$, $t>\tau(v)$, and $0<r:=s-t<\tau(u+v)$ (since the value group is dense). By order unit goodness, there exists a such that $0 \le a \le u+v$ and $\tau(a)=r$. Riesz decomposition entails $a=a_1+a_2$, where $0\le a_1 \le u$ and $0 \le a_2 \le v$. Set $s'=\tau(a_1)\ge 0$ and $t'=\tau(a_2)\ge 0$. Then $s-t=s'+t'$, so $s-s'=t+t'$. The intersection consisting of 0 forces $s=s'$ and $t=-t'$; the latter forces $t=t'=0$, a contradiction.

Now suppose the intersection is nonzero and not dense. Then it is cyclic, so there exists $x \in \mathbb{R}$, which we may assume to be positive, such that $\tau(I(u))\cap\tau(I(v))=x\mathbb{Z}$. We may find $0<s,t<x$ with $s \in \tau(I(u))$ and $t \in \tau(I(v))$ such that $0<r:=s-t$. Find $a \le u+v$ as above with $r=\tau(a)$, similarly decompose $a=a_1+a_2$, and define s', t' as in the preceding paragraph. We deduce $s-s'=t+t'$, hence there exists an integer m such that $s-s'=mx=t+t'$; as $t,t'\ge 0$, we have $m \ge 0$, but as $s<x$, we have $m<1$, hence $m=0$. This forces $t=t'=0$, again a contradiction. □

Corollary 1.4. *Let G be a nearly divisible dimension group with a faithful trace τ. Suppose that I and J are order ideals with their own order units such that each of $\tau|I$, $\tau|J$, and $\tau|(I+J)$ is order unit good. Then $\tau(I)\cap\tau(J)$ is a dense subgroup of $\mathbb{R}$.*

Proof. Since τ is faithful, $\tau|I$ and $\tau|J$ are nonzero, and since every trace on an order ideal with order unit is nondiscrete (as the order ideals are approximately divisible by definition), it follows that $\tau(I)$ and $\tau(J)$ are dense. Now Lemma 1.3(b) applies. □

Let (G, u) be a dimension group. Let $\mathcal{J}$ be a collection of nonzero order ideals, each with their own order unit, such that every order ideal of G with order unit can be expressed as a sum of order ideals from $\mathcal{J}$ (such a sum can always be made finite, as the order ideal has an order unit); then we say $\mathcal{J}$ is a *generating set of order ideals of* G.

The criteria in Lemma 1.2 for goodness can be reduced to that on a generating set of order ideals. This will make the computations of Section 4 much simpler.

Lemma 1.5. *Let* (G, u) *be a nearly divisible dimension group, let* $\mathcal{J}$ *be a generating set of order ideals of* G, *and let* τ *be a faithful trace of* G. *For* τ *to be a good trace of* G, *it is sufficient that it satisfy*

(i) *for all* $J \in \mathcal{J}$, *we have* $\tau(J) = \tau(G)$ *and*

(ii) *for all* $J \in \mathcal{J}$, *we have* $\tau|J$ *is an order unit good trace of* J.

Proof. We can express a nonzero order ideal I with order unit as $I = \sum J_\alpha$ for some $J_\alpha \in \mathcal{J}$. Thus $\tau(I) = \sum \tau(J_\alpha) = \tau(G)$.

Since I has an order unit, the sum can be made finite; now we apply induction (on the number of summands) to Lemma 1.2(d); this verifies the second property in Proposition 1.1. □

Verifying the criteria for goodness and related properties is much simpler when the partially ordered abelian group is an ordered ring having 1 as an order unit.

Lemma 1.6. *Let* $(R, 1)$ *be a* (*commutative*) *partially ordered commutative ring with* 1 *as order unit. If* R *is approximately divisible*, *then it is nearly divisible.*

Proof. Approximate divisibility implies the existence of order units u and v such that $1 = 2u + 3v$; for any $r \in R^+ \setminus \{0\}$, we thus have $r = 2(ru) + 3(rv)$. From $1 \leq ku, kv$ for some positive integer k, we deduce $r \leq k(ru), k(rv)$, verifying the definition of nearly divisible. □

The following is implicit in the proof of [BeH 2014, Corollary 7.12].

Lemma 1.7. *Let* $(R, 1)$ *be a partially ordered* (*commutative*) *unperforated ring with* 1 *as order unit*, *that is*, *an approximately divisible dimension group. Let* τ *be a faithful pure trace. Then* τ *is order unit good if and only if for all* $\sigma \in \partial_e S(R, 1) \setminus \{\tau\}$, *we have* $\sigma(\ker \tau) \neq \{0\}$.

Proof. Since 1 is an order unit of the partially ordered ring, $X := \partial_e S(R, 1)$ is compact and consists precisely of the normalized multiplicative traces of R; moreover, $\operatorname{Aff} S(R, 1) = C(X, \mathbb{R})$ with the affine representation reinterpreted as $\tilde{g}(\phi) = \phi(g)$ for $\phi \in X$ (note the use of $\tilde{}$ rather than $\hat{}$, to distinguish them). By approximate divisibility, the image of R is dense in $C(X, \mathbb{R})$. If A is any ideal of R, then its closure in $C(X, \mathbb{R})$ is an ideal therein, and hence is of the form $\mathrm{Ann}(Y) := \{f \in C(X, \mathbb{R}) \mid f|Y \equiv 0\}$ for a unique compact subset Y of X.

Since τ is pure, it is multiplicative, and therefore $\ker\tau$ is an ideal of R (*not* an order ideal, unless $\ker\tau = 0$, as $\ker\tau \cap R^+ = \{0\}$ is the definition of faithfulness). The closure of the image of $\ker\tau$ in $C(X, \mathbb{R})$ can thus be written in the form $\mathrm{Ann}(Y)$ for some compact subset Y.

If τ is order unit good, then $\mathrm{Ann}(Y)$ is $\mathrm{Ann}(\{\tau\})$ (corresponding to $\tau^\perp$ in $\mathrm{Aff}\, S(R, 1)$), from which it follows that $Y = \{\tau\}$. Hence if $\sigma \in X \setminus \{\tau\}$, there exists a continuous $f : X \to [0, 1]$ such that $f(\tau) = 0$ but $f(\sigma) = 1$; then $f \in \mathrm{Ann}(\{\tau\})$, hence there exist $g_n \in R$ such that $g_n \in \ker\tau$ and $\tilde{g}_n \to f$ uniformly. Applying σ, there exists n such that $\sigma(g_n) \neq 0$, so that $\sigma(\ker\tau) \neq \{0\}$.

Conversely, suppose for every $\sigma \in X \setminus \{\tau\}$, we have $\sigma(\ker\tau) \neq \{0\}$. Then $\sigma \notin Y$, hence $Y = \{\tau\}$, so that the closure of the image of $\ker\tau$ is codimension one in $C(X, \mathbb{R})$, hence equal to $\tau^\perp$ in $\mathrm{Aff}\, S(G, u)$. Thus τ is order unit good. □

2. Tensor products

If G and H are partially ordered abelian groups, we may form the tensor product (as $\mathbb{Z}$-modules) $G \otimes_{\mathbb{Z}} H$ (usually, we delete the subscripted $\mathbb{Z}$); it is equipped with a cone which makes it into a partially ordered group, $\left\{\sum g_i \otimes h_i \;\middle|\; g_i \in G^+ \text{ and } h_i \in H^+\right\}$ [Goodearl and Handelman 1986, Proposition 2.1]. If both are dimension groups, then so is $G \otimes H$, and if u, v are respectively order units for G, H, then $u \otimes v$ is an order unit for $G \otimes H$. If σ, τ are respective (normalized) traces on (G, u) and (H, v), then $\sigma \otimes \tau$ (defined in the obvious way) is a (normalized) trace of $(G \otimes H, u \otimes v)$.

Appendix A informally discusses connections between tensor products of dimension groups and products of $\mathbb{Z}$-actions on Cantor minimal systems.

A special case occurs when we form the divisible hull of a dimension group, $G \otimes \mathbb{Q}$, the rational vector space that G generates. Then τ extends to a trace $G \otimes \mathbb{Q}$ in the obvious way, denoted $\tau \otimes 1_{\mathbb{Q}}$. In general, τ being order unit good or good implies the corresponding property for $\tau \otimes 1_{\mathbb{Q}}$, but the converse fails practically generically. As a special case, we [BeH 2014] defined a trace τ to be *ugly* if $\tau \otimes 1_{\mathbb{Q}}$ is good and $\ker\tau$ has *discrete* image in (the Banach space) $\mathrm{Aff}\, S(G, u)$. Ugly traces exist in profusion.

In Akin's original context of measures on Cantor sets, he showed that (what amounts to) the tensor product of good traces is good; in the context of simple dimension groups or more generally for approximately divisible dimension groups, the tensor product of order unit good traces was shown to be order unit good [BeH 2014, Proposition 5.2]. Here, we show a somewhat surprising result for order unit goodness: if (G, u) and (H, v) are approximately divisible, and both $\sigma \otimes 1_{\mathbb{Q}}$ and

$\tau \otimes 1_{\mathbb{Q}}$ are order unit good on their respective groups, then $\sigma \otimes \tau$ is order unit good (as a trace on $G \otimes H$). This means that the tensor product has a stronger property (in general) than its constituents. In particular, the tensor product of ugly traces is at least order unit good.

A weaker notion is *refinability*; again based on Akin's definition in the dynamical situation, and translated to partially ordered groups: a trace τ on (G, u) is refinable if whenever $b \in G^+ \setminus \ker \tau$ and $\{a_i\}$ is a finite subset of G^+ such that $\tau(b) = \sum \tau(a_i)$, there exist $\{a_i'\} \subset G^+$ such that $b = \sum a_i'$ and $\tau(a_i) = \tau(a_i')$. Surprisingly, the corresponding tensor product results actually fail for refinability (even though the set of refinable traces is a dense G_δ in the trace space).

Using the criterion of Proposition 1.1, we then obtain a corresponding criterion for goodness of the tensor product (G and H are nearly divisible, $\sigma \otimes 1_{\mathbb{Q}}$ and $\tau \otimes 1_{\mathbb{Q}}$ are good, and a condition that guarantees the value groups on the order ideals is the same as the full value group).

Proposition 2.1. *Let (G, u) and (H, v) be approximately divisible dimension groups with traces σ and τ respectively. If each of $\sigma \otimes 1_{\mathbb{Q}}$ and $\tau \otimes 1_{\mathbb{Q}}$ on $G \otimes \mathbb{Q}$ and $H \otimes \mathbb{Q}$ respectively is order unit good, then the trace on $(G \otimes H, u \otimes v)$ given by $\sigma \otimes \tau$ is order unit good.*

If we only require that $\sigma \otimes \tau \otimes 1_{\mathbb{Q}}$ (a trace on $G \otimes H \otimes \mathbb{Q}$) be order unit good (in place of each of $\sigma \otimes 1_{\mathbb{Q}}$ and $\tau \otimes 1_{\mathbb{Q}}$ being good), the conclusion is false; an example will be given later (Example 2.6).

We require a number of elementary results about tensor products. Here the tensors will be over one of the rings $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$; *torsion-free (module)* means torsion-free abelian group when the underlying ring is $\mathbb{Z}$; otherwise, it just means vector space over the relevant field.

Lemma 2.2. *Let A and B be torsion-free modules, and $A' \subset A$, $B' \subset B$ submodules such that A/A' and B/B' are torsion-free.*

(a) *The kernel of the map $A \otimes B \to A \otimes (B/B')$ is $A \otimes B'$.*

(b) *The kernel of the map $A \otimes B \to (A/A') \otimes (B/B')$ is $A \otimes B' + A' \otimes B$.*

Proof. (a) One inclusion is obvious. Because the quotient is torsion-free, $A \otimes B/B'$ is torsion-free. We have an induced map $(A \otimes B)/(A \otimes B') \to A \otimes (B/B')$. If z is in the kernel, find a nonzero integer n such that nz has a representative in $A \otimes B$ of least length (as n varies over nonzero integers), say $nz = \sum a_i \otimes b_i + (A \otimes B')$. Then $\{a_i\}$ is rationally linearly independent; hence the image, $n\bar{z}$, yields $0 = \sum a_i \otimes (b_i + B')$. Since B/B' is torsion-free, this easily implies all $b_i + B' = 0$ (tensor with $\mathbb{Q}$ if necessary, so we are working over a field, then use a basis for $B'\mathbb{Q}$, extended to $B\mathbb{Q}$). (This proof works for all fields.)

(b) First, $A \otimes B/(A \otimes B')$ is naturally isomorphic to $A \otimes B/B'$ by (a). Then another application of (a) with the order reversed yields a natural isomorphism $(A\otimes(B/B'))/(A'\otimes(B/B')) \cong (A/A')\otimes(B/B')$. Then the kernel of the first map is $A\otimes B'$, and that of the second is $A'\otimes(B/B')$, which pulls back to $A\otimes B'+A'\otimes B$. □

Proof of Proposition 2.1. We will show that the closure of the image of $\ker \sigma \otimes \tau$ in $\operatorname{Aff} S(G \otimes H, u \otimes v)$ is $(\sigma \otimes \tau)^{\vdash}$; by [BeH 2014, Proposition 1.7], $\sigma \otimes \tau$ is order unit good.

First, we identify the product $\operatorname{Aff} S(G, u) \otimes_{\mathbb{R}} \operatorname{Aff} S(H, v)$ with a subspace of $\operatorname{Aff} S(G \otimes H, u \otimes v)$ in the obvious way. Standard results (e.g., pure traces are pure tensors) yield that it is a dense subspace.

We note that $(\ker \sigma) \otimes H + G \otimes (\ker \tau) \subseteq \ker \sigma \otimes \tau$. It easily follows that the closure of the image of $(\ker \sigma) \otimes H$ contains everything in $y \otimes \operatorname{Aff} S(H, v)$ (real tensors), where y varies over the image of $\ker \sigma$ (in $\sigma^{\vdash} \subset \operatorname{Aff} S(G, u)$). For y fixed, $y \otimes \operatorname{Aff} S(H, v)$ is a real vector space, and this means that we can rewrite it as $y\mathbb{R} \otimes \operatorname{Aff} S(H, v)$ (just approximate real multiples of $\hat{v}$ by elements of $\widehat{H}$, and transfer through the tensor product). Taking finite sums, we see that the closure of the image of $\ker \sigma \otimes H$ includes the closure of $\operatorname{Im}(\ker \sigma)\mathbb{Q} \otimes \operatorname{Aff} S(H, v)$.

Now $\sigma \otimes 1_{\mathbb{Q}}$ being order unit good implies $(\ker \sigma) \otimes \mathbb{Q}$ has dense image in $\sigma^{\vdash}$ (in $\operatorname{Aff} S(G, u)$). If e is an element of $G \otimes \mathbb{Q}$, there exists a nonzero integer m such that $me \in G$. If in addition, $\sigma \otimes 1_{\mathbb{Q}}(e) = 0$, then $\sigma(me) = 0$; thus $\ker(\sigma \otimes 1_{\mathbb{Q}}) \subseteq (\ker \sigma)\mathbb{Q}$ (the reverse inclusion is trivial, but never needed).

Thus the closure of the image of $(\ker \sigma) \otimes H$ contains $\operatorname{Im}(\ker \sigma)\mathbb{Q} \otimes \operatorname{Aff} S(H, v)$, which in turn contains the closure of $\operatorname{Im}(\ker \sigma)\mathbb{Q} \otimes \operatorname{Aff} S(H, v)$, and thus includes $\sigma^{\vdash} \otimes \operatorname{Aff} S(H, v)$.

Similarly, the closure of the image of $G \otimes \ker \tau$ contains $\operatorname{Aff} S(G, u) \otimes \tau^{\vdash}$. Set $A = \operatorname{Aff} S(G, u)$, $A' = \sigma^{\vdash}$, $B = \operatorname{Aff} S(H, v)$, and $B' = \tau^{\vdash}$; then each is a Banach space, and (A/A') and (B/B') are both one-dimensional, and the closure of the image of $\ker(\sigma \otimes \tau)$ contains $A' \otimes B + A \otimes B'$.

By (b) above, $(A \otimes B)/(A' \otimes B + A \otimes B')$ is one-dimensional. Let $W = A' \otimes B + A \otimes B'$ and $Z = \operatorname{Aff} S(G, u) \otimes \operatorname{Aff} S(H, v)$, so that W is a codimension-one subspace of Z. It is now an easy exercise to show that when we complete Z to $\operatorname{Aff} S(G\otimes H, u\otimes v)$, the closure, $\overline{W}$, is of at most codimension one. (This is a general Banach space result; if $\overline{W} \neq \bar{z}$, then $W = \overline{W} \cap Z$ as W is codimension one in Z; choose $z \in Z \setminus W$; the functional sending $z \mapsto 1$ and $W \mapsto 0$ is continuous (essentially the closed graph theorem), and hence extends to a bounded linear functional p on $\overline{W}$; we may write arbitrary $y \in \bar{Z}$ as $\lim y_n$; then $y_n = p(y_n)z + (y_n - p(y_n)z)$, and thus by continuity, $y = p(y)z + (y - p(y)z)$, and $y - p(y)z$ is in $\overline{W}$, hence $z + \overline{W} = \bar{z}$.)

In particular, the closure of the image of $\ker \sigma \otimes \tau$ in $\operatorname{Aff} S(G \otimes H, u \otimes v)$ is codimension one. As it is contained in $(\sigma \otimes \tau)^{\vdash}$, which is proper, it follows that the image of $\ker \sigma \otimes \tau$ is dense in $(\sigma \otimes \tau)^{\vdash}$. □

This explains a phenomenon exemplified in [BeH 2014, Example 9]. Let G be a critical dimension group of rank $k+1$ (that is, a free rank $k+1$ abelian group densely embedded in $\mathbb{R}^k$, and equipped with the strict ordering therefrom [Handelman 1982]). Then we say G is *basic* (as a critical group) if it is *order-isomorphic* to the subgroup of $\mathbb{R}^k$ spanned by $\{e_i; \sum \alpha_j e_j\}$, where $\{e_i\}$ is the standard basis and $\{1, \alpha_1, \ldots, \alpha_k\}$ is linearly independent over the rationals (this guarantees density of the subgroup). Every critical group is *topologically isomorphic* to a group of the latter form.

For basic critical groups, every pure trace is ugly, as is immediate from the definitions. Hence if G_i are basic critical groups (and there is more than one), then all of the pure traces of their tensor product (a simple dimension group) $\bigotimes G_i$ are good. In [BeH 2014, Example 9], an example was given of a basic critical group of rank three, for which all pure traces on $G \otimes G$ are good. We also asked whether the pure traces on $G \otimes G \otimes G$ are good, and now we know that the answer is yes.

It is plausible that among critical groups, basic ones are characterized by all pure traces being ugly; this is false, but is close to being true [Handelman 2013a, Proposition 7.4]. There are lots of critical groups for which all or some are bad, hence not ugly [BeH 2014, Section 2]. It can also happen that if both σ, τ are bad traces (a trace τ is *bad* if $\ker \tau$ consists of the infinitesimal elements of the group [BeH 2014]), then $\sigma \otimes \tau$ is good; but it can also arise that $\sigma \otimes \tau$ is not even ugly.

Now suppose that (G, u) and (H, v) are nearly divisible, and σ, τ are normalized traces on G, H respectively such that $\sigma \otimes 1_{\mathbb{Q}}$ and $\tau \otimes 1_{\mathbb{Q}}$ are both good. We expect to obtain that $\sigma \otimes \tau$ is a good trace on $G \otimes H$.

Lemma 2.3. *Let (G, u) and (H, v) be dimension groups with order unit. Then*:

(a) *$G \otimes H$ is approximately divisible if and only if at least one of G or H is.*

(b) *$G \otimes H$ is nearly divisible if and only if at least one of G or H is.*

Proof. (a) Suppose G is approximately divisible. Every pure trace of $(G \otimes H, u \otimes v)$ is of the form $\sigma \otimes \tau$ [Goodearl and Handelman 1986, Lemma 4.1], where σ, τ are pure traces of G, H respectively. Then $(\sigma \otimes \tau)(G \otimes H)$ is $\sigma(G) \cdot \tau(H)$ (the set of sums of terms of the form $\sigma(g) \cdot \tau(h)$); as $\sigma(G)$ is dense, obviously so is $\sigma(G) \cdot \tau(H)$, so that $G \otimes H$ has no discrete pure traces, and is thus approximately divisible. The same argument applies if instead H is approximately divisible.

If neither G nor H is approximately divisible, then there exists a discrete trace σ of G and a discrete trace τ of H; as these are normalized (at u, v respectively), $\sigma(G) = (1/n)\mathbb{Z}$ and $\tau(H) = (1/m)\mathbb{Z}$ for some positive integers m and n; then $(\sigma \otimes \tau)(G \otimes H) = (1/mn)\mathbb{Z}$, which is discrete. Hence $G \otimes H$ admits a discrete trace, and thus is not approximately divisible.

(b) Select $a = \sum g_i \otimes h_i \in (G \otimes H)^+$; from the definition of the ordering on the tensor product, we can assume each of g_i and h_i are positive in their respective groups. By definition, we can write $g_i = 2a_i + 3b_i$, where $0 \leq g_i \leq ka_i, kb_i$ for some positive integer k; since the sum is finite, we can take the same integer k for all i. Set $c_1 = \sum a_i \otimes h_i$ and $c_2 = \sum b_i \otimes h_i$. Then $a = 2c_1 + 3c_2$; moreover, $\sum g_i \otimes h_i \leq k \sum a_i \otimes h_i$, that is, $a \leq kc_1$, and similarly $a \leq kc_2$.

If neither G nor H is nearly divisible, there exist an order ideal of G with its own order unit, (I, w) together with a discrete trace (of I) ϕ, and an order ideal of H with its own order unit, (J, y) and a discrete trace on it, ψ. Then $\phi \otimes \psi$ is a discrete trace (as above) of $I \otimes J$; this being an order ideal of $G \otimes H$, the latter is not nearly divisible. □

Lemma 2.4. *Let G and H be nearly divisible, having faithful traces σ and τ respectively such that $\sigma \otimes 1_{\mathbb{Q}}$ and $\tau \otimes 1_{\mathbb{Q}}$ are good as traces on $G \otimes \mathbb{Q}$, $H \otimes \mathbb{Q}$ respectively.*

(a) *Let (I, w) be an order ideal of G with its own order unit, and let (J, y) be an order ideal of H with its own order unit. Then $(\sigma \otimes \tau)|(I \otimes J)$ is order unit good.*

(b) *Suppose for each order ideal I of G, $\sigma(I) = \sigma(G)$, and similarly, for each order ideal J of H, we have $\tau(J) = \tau(H)$. Then for every nonzero order ideal L of $G \otimes H$, we have $(\sigma \otimes \tau)(L) = (\sigma \otimes \tau)(G \otimes H)$.*

(c) *Suppose the hypotheses of (b) apply. Let (L, e) be an arbitrary order ideal of $G \otimes H$ with its own order unit. Then $(\sigma \otimes \tau)|L$ is order unit good.*

Proof. (a) Each of the restrictions of $\sigma \otimes 1_{\mathbb{Q}}$ and $\tau \otimes 1_{\mathbb{Q}}$ to $I \otimes \mathbb{Q}$ and $J \otimes \mathbb{Q}$ respectively is good, hence order unit good, and thus $(\sigma \otimes \tau)|(I \otimes J)$ is an order unit good trace of $I \otimes J$.

(b) First, if $L = I \otimes J$ (where I and J are nonzero order ideals in G and H respectively), then $(\sigma \otimes \tau)(I \otimes J)$ is the subgroup of $\mathbb{R}$ generated by all terms of the form $\sigma(a) \cdot \tau(b)$, where $a \in I$ and $b \in J$, and $(\sigma \otimes \tau)(G \otimes H)$ has the same form, except a and b are allowed to vary over G and H respectively. Since for all $a \in G$, there exists $a' \in I$ such that $\sigma(a') = \sigma(a)$, and similarly for τ, the two groups are equal.

If $e \in L^+$, then by the definition of the tensor product ordering, we can write $e = \sum g_i \otimes h_i$. For an element x in the positive cone of a dimension group, let $I(x)$ be the order ideal it generates; then it is easy to check (since sums of order ideals are again order ideals in a dimension group) that $L = I(e) = \sum I(g_i) \otimes I(h_i)$; in particular, L contains a tensor product of order ideals, so the previous paragraph applies.

(c) Every $e \in (G \otimes H)^+$ can be written in the form $e = \sum g_i \otimes h_i$ with $g_i \in G^+$ and $h_i \in H^+$. By (a), the restriction of $\sigma \otimes \tau$ to each of $I(g_i) \otimes I(h_i)$ is order unit good. Since $\sigma \otimes \tau(L) = (\sigma \otimes \tau)(G \otimes H)$, for any nonzero order ideal L of $G \otimes H$, we may apply Lemma 1.2(e) (the intersection of the value groups is dense), so the restriction of $\sigma \otimes \tau$ to L is order unit good. □

Proposition 2.5. *Suppose that* (G, u, σ) *and* (H, v, τ) *are nearly divisible dimension groups with faithful trace having the following properties*:

(i) *For all nonzero order ideals* I *and* J *of* G *and* H *respectively*, $\sigma(I) = \sigma(G)$ *and* $\tau(J) = \tau(H)$.

(ii) *Each of* $\sigma \otimes 1_{\mathbb{Q}}$ *and* $\tau \otimes 1_{\mathbb{Q}}$ *is good on* $G \otimes \mathbb{Q}$, $H \otimes \mathbb{Q}$ *respectively.*

Then $\sigma \otimes \tau$ *is a good trace of* $G \otimes H$.

Proof. This follows from Lemma 2.3, Lemma 2.4, and Proposition 1.1. □

Even in the simple case, it is not true that goodness of $\tau \otimes \tau \otimes \mathbf{1}_{\mathbb{Q}}$ (a trace on $G \otimes G \otimes \mathbb{Q}$) implies $\tau \otimes \tau$ is good. In fact, the next example illustrates something more drastic.

A weaker property than goodness is refinability: a trace $\tau : G \to \mathbb{R}$ is *refinable* if whenever $b \in G^+$ and there exist $\{a_i\} \subset G^+$ such that $\tau(b) = \sum \tau(a_i)$, then there exist $a_i' \in G^+$ such that $\tau(a_i) = \tau(a_i')$ and $b = \sum a_i'$. Good traces are refinable [BeH 2014, Lemma 7.3]. Following [BeH 2014], a trace τ is *bad* if $\operatorname{Inf} G = \ker \tau$ and τ is not the only normalized trace. It is trivial that bad traces are refinable when $\operatorname{Inf} G = \{0\}$ [BeH 2014].

More interestingly, when there is more than one trace, bad traces are generic; in fact, they constitute a dense G_δ of $S(G, u)$, merely under the assumption that G is countable [Giordano et al. $\geq$ 2016]. Because of this, one would expect refinability to be even better behaved under tensor products than goodness. This is not the case.

Example 2.6. There exists a simple dimension group G with a pure trace τ with the following properties:

(a) τ is bad, and thus is refinable.

(b) $\tau \otimes \tau \otimes 1_{\mathbb{Q}}$, a trace on $G \otimes G \otimes \mathbb{Q}$, is good.

(c) The trace $\tau \otimes \tau : G \otimes G \to \mathbb{R}$ is not even refinable.

(d) The trace $\tau \otimes \tau \otimes \tau \otimes \tau$ on $G^{\otimes 4}$ is good.

Proof. Let α be real, quartic, and integral (that is, it satisfies a monic degree-four irreducible polynomial with integer coefficients), and let β be a real number not satisfying any degree-four polynomial over the rationals (in particular, $\beta \notin \mathbb{Q}(\alpha)$, where the latter is the field generated over the rationals by α). Let G be the subgroup of $\mathbb{R}^2$ generated by $\{(1, 1), (\alpha, \beta), (\alpha^2, \beta^2)\}$. The three 2×2 determinants are $\{\beta - \alpha, \beta^2 - \alpha^2, \alpha\beta^2 - \alpha^2\beta\}$; since $\beta \neq \alpha$, this set is rationally linearly independent (rational linear independence of $\{1, \alpha + \beta, \alpha\beta\}$ follows from $\beta \notin \mathbb{Q}(\alpha)$). Thus G is dense in $\mathbb{R}^2$, so with the strict ordering it inherits from the latter, it will be a simple dimension group.

Let $\tau : G \to \mathbb{R}$ be the projection onto the first coordinate. This is a pure trace, and moreover, $\ker \tau = \{0\}$, so that τ is bad, and thus is refinable.

Now make the identifications

$$(1,0)\otimes(1,0)\mapsto(1,0,0,0),\quad (1,0)\otimes(0,1)\mapsto(0,1,0,0),$$
$$(0,1)\otimes(1,0)\mapsto(0,0,1,0),\quad (0,1)\otimes(0,1)\mapsto(0,0,0,1).$$

This yields order isomorphisms $\mathbb{Z}^2\otimes\mathbb{Z}^2\to\mathbb{Z}^4$ and $\mathbb{R}^2\otimes\mathbb{R}^2\to\mathbb{R}^4$. Then $G\otimes G$, being a simple dimension group, inherits the strict ordering on $\mathbb{R}^4$, and is spanned by the nine elements

$$\begin{aligned} &a=(1,1,1,1), && b=(\alpha,\beta,\alpha,\beta), && c=(\alpha^2,\beta^2,\alpha^2,\beta^2),\\ &d=(\alpha,\alpha,\beta,\beta), && e=(\alpha^2,\alpha\beta,\alpha\beta,\beta^2), && f=(\alpha^3,\alpha\beta^2,\alpha^2\beta,\beta^2),\\ &g=(\alpha^2,\alpha^2,\beta^2,\beta^2), && h=(\alpha^3,\alpha^2\beta,\alpha\beta^2,\beta^3), && j=(\alpha^4,\alpha^2\beta^2,\alpha^2\beta^2,\beta^4). \end{aligned}$$

The trace $\tau\otimes\tau$ identifies with the projection on the first coordinate, which we will call σ. Since α satisfies an irreducible polynomial of degree four (and therefore of no less degree), say $\alpha^4=A\alpha^3+B\alpha^2+C\alpha+D$, with $A, B, C, D\in\mathbb{Z}$, it follows that

$$\ker\sigma=\langle d-b, e-c, g-e, h-f, j-Af-Bc-Cb-Da\rangle.$$

(Since $\sigma(G\otimes G)\subset\mathbb{Z}[\alpha]$ and the latter is rank four, the kernel has rank five; this reduces the problem to showing the cokernel of the group on the right is torsion-free, which is routine.) The first four of the generators are of the form $(0,*,*,0)$, and it is easy to verify that the group that they span is dense in $0\oplus\mathbb{R}^2\oplus 0$. The last generator has nonzero fourth coordinate (since β satisfies no fourth degree polynomial); call it γ. Thus the closure of $\ker\sigma$ is $\{0\}\oplus\mathbb{R}^2\oplus\gamma\mathbb{Z}$.

In particular, the closure of $\ker\sigma$ is not a real vector space, so $\sigma=\tau\otimes\tau$ is not good; moreover, it is not even refinable, since $\phi(\ker\sigma)$ is cyclic and nonzero, where ϕ is the projection onto the fourth coordinate [BeH 2014, Proposition B.5].

On the other hand, $(\ker\sigma)\otimes\mathbb{Q}$ is dense in $\{0\}\oplus\mathbb{R}^3$, and since $\ker(\sigma\otimes 1_{\mathbb{Q}})$ contains $(\ker\sigma)\otimes\mathbb{Q}$, it follows that $\sigma\otimes 1_{\mathbb{Q}}=\tau\otimes\tau\otimes 1_{\mathbb{Q}}$ is a good trace on $G\otimes\mathbb{Q}$. By Proposition 2.1, $\tau\otimes\tau\otimes\tau\otimes\tau$ is a good trace on $G\otimes G\otimes G\otimes G$. □

Left open are the properties of $\tau\otimes\tau\otimes\tau$. Short of computing with a $\mathbb{Z}$-basis consisting of 27 elements, there did not seem to be any method of attack (the kernel has rank 23; however, the infinitesimal subgroup is substantial, too).

3. Examples from xerox actions of tori on UHF algebras

We characterize the good faithful pure traces on the dimension groups arising from xerox-product-type actions of tori on UHF C*-algebras. It turns out that there is a surprising number-theoretic component.

Appendix A points out strong analogies between Bernoulli measures and the traces discussed here.

Form the Laurent polynomial ring in d variables over the integers, $\mathbb{Z}[x_i^{\pm 1}]$, and let $\mathbb{Z}[x_i^{\pm 1}]^+$ denote the set of those with only nonnegative coefficients. As in [Handelman 1985; 1987], we adopt monomial notation; that is, for $w \in \mathbb{Z}^d$, define $x^w = x_1^{w(1)} \cdot x_2^{w(2)} \cdot \dots \cdot x_d^{w(d)}$. For any $f \in \mathbb{Z}[x_i^{\pm 1}]$, we denote the coefficient of x^w in f by (f, x^w) (inner product notation, which is consistent with the origins of the work), and we set $\operatorname{Log} f := \{w \in \mathbb{Z}^d \mid (f, x^w) \neq 0\}$. Let $P = \sum a_w x^w \in \mathbb{Z}[x_i^{\pm 1}]^+$ (where $a_w \in \mathbb{Z}^+$), and form the ring $R_P = \mathbb{Z}[\{x^w/P\}_{w \in \operatorname{Log} P}]$. Equipped with the partial ordering generated additively and multiplicatively by $\{x^w/P \mid w \in \operatorname{Log} P\}$, this is a dimension group and an ordered ring with 1 as order unit, and many more properties (marked with bullets below). We may also form $\mathbb{Z}[x_i^{\pm 1}, 1/P]$ (a subring of the field of fractions of the Laurent polynomial ring). It also has a partial ordering given by $\{f/P^k \mid \exists N$ such that $P^N f$ has no negative coefficients$\}$. The restriction of this to R_P yields the original ordering.

This arose from the following construction. Let $n = P(1, 1, 1, \dots, 1)$, and form $\mathcal{A} = \otimes M_n\mathbb{C}$ (the UHF C*-algebra). The Laurent polynomial P is the character of an n-dimensional representation of the torus $\mathbb{T}^d$, say given by $z \mapsto \operatorname{diag}(z^w)$ (one for each w that appears in P, with repetitions as indicated by the multiplicities, that is, the coefficients). This yields a map $\pi : \mathbb{T}^d \to M_n\mathbb{C}$ with nonzero entries along the diagonal. Form $\phi := \otimes \operatorname{Ad} \pi : \mathbb{T}^d \to \operatorname{Aut} \mathcal{A}$, and the corresponding fixed point subrings, $\mathcal{A}^{\phi(\mathbb{T}^d)}$, and $\mathcal{A} \times_\phi \mathbb{T}^d$, the latter the C*-crossed product. Then $(\mathrm{K}_0(\mathcal{A}^{\phi(\mathbb{T}^d)}), [1])$ is a naturally ordered ring isomorphic to R_P and $\mathrm{K}_0(\mathcal{A} \times_\phi \mathbb{T}^d)$ is similarly isomorphic to the ordered ring $\mathbb{Z}[x_i^{\pm 1}, 1/P]$. This will play a role in what follows.

Renault [1980] determined the positive cone and analyzed (inter alia) the structure of R_P when $P = 1 + x$. The pure (ergodic) traces thereon were determined by Orey (in terms of the simple random walk) in the mid-1960s.

We normally assume that P is projectively faithful; that is, $\operatorname{Log} P - \operatorname{Log} P$ generates (as an abelian group) the standard copy of $\mathbb{Z}^d$ in $\mathbb{R}^d$ (we can reduce to this case anyway). This has the effect that whenever $v \in \operatorname{Log} P^k \cap \operatorname{int} \operatorname{cvx} \operatorname{Log} P^k$ for some positive integer k, it follows that x^v/P^k belongs to R_P and $R_P[(x^v/P^k)^{-1}] = \mathbb{Z}[x_i^{\pm 1}, 1/P]$; i.e., the larger ring is obtained by inverting x^v/P^k.

We call an element of the form x^w/P with $w \in \operatorname{Log} P$ a *formal monomial* in R_P. (It can happen that $x^w/P \in R_P$ even if $w \notin \operatorname{Log} P$ — e.g., if $w + \operatorname{Log} P^k \subseteq \operatorname{Log} P^{k+1}$ for some k. This is not significant for what follows.)

In addition to the obvious facts about R_P (it is a commutative, finitely generated — hence noetherian — domain), the following results are known:

• $R_P = \{g/P^k \mid g \in \mathbb{Z}[x],\ \operatorname{Log} g \subset \operatorname{Log} P^k\}$, R_P is a partially ordered ring with 1 as an order unit, and it is a dimension group [Handelman 1985, Section I].

• All sums and finite intersections of order ideals are order ideals (this is true for all dimension groups) [Goodearl 1986].

• Products of order ideals are order ideals (this is not generally true for dimension groups that are commutative partially ordered domains having 1 as an order unit) [Handelman 1985].

• Every order ideal is an ideal (this is true in every partially ordered commutative ring in which 1 is an order unit) [Handelman 1985, Proposition I.2].

• If f is a formal monomial, then $f R_P$ (the ideal generated by f) is an order ideal [Handelman 1987, Proposition II.2A].

• Every order ideal is the finite sum of ideals, $\sum f_i R_P$, where f_i are formal monomials, and all such sums are order ideals [Handelman 1987, p. 19].

• If f is a formal monomial and $a \in R_P$, then $fa \in R_P^+$ implies $a \in R_P^+$ (this follows from the definitions); the conclusion is also true if we replace *formal monomial* by *order unit,* a result that is very special for R_P [Handelman 1987, Proposition II.5].

• The pure traces are exactly the multiplicative ones (this is true for any partially ordered ring with 1 as an order unit); the pure faithful traces are exactly those of the form $\tau_r(g/P^k) = g(r)/P^k(r)$, where $r = (r_i)$ is a strictly positive d-tuple in $\mathbb{R}^d$, and these extend in the obvious way to positive homomorphisms $\tau_r : \mathbb{Z}[x_i^{\pm 1}; 1/P] \to \mathbb{R}$ (warning: although the ring $\mathbb{Z}[x_i^{\pm 1}; 1/P]$ is partially ordered, 1 is not an order unit for it) [Handelman 1985, Theorem III.3].

• The weighted moment map/Legendre transform corresponding to P implements a homeomorphism $\partial_e S(R_P, 1) \to \operatorname{cvx} \operatorname{Log} P$ (the latter is the *Newton polytope* of P) sending the faithful pure traces onto the interior; unexpectedly, the set of *pure* traces admits a type of convex structure; in particular, the faces correspond to traces that factor through quotients in a particularly nice way [Handelman 1987, Theorem IV.1].

• In general, R_P is not a pure polynomial ring; only rarely does it have unique factorization [Handelman 1987, Theorem A.8A].

Now let us consider the following property of a faithful pure trace $\tau \equiv \tau_r$:

(1) for every nonzero order ideal I, we have $\tau_r(I) = \tau_r(R_P)$.

By Proposition 1.1, this is one of the two necessary conditions for τ_r to be a good trace.

Here $r = (r_i) \in (\mathbb{R}^d)^{++}$ as described above. First we note that $\{f R_P\}$ (as f varies over all products of formal monomials) is a generating set of order ideals with order unit (they are given as ring ideals, but in fact are order ideals by the properties above, and every order ideal is a finite sum of these). Necessary and sufficient for (1) to hold is simply that it hold for all ideals of the form $I_w = (x^w/P)R_P$ (where $w \in \operatorname{Log} P$, a finite set). To see this, note that $\tau_r(I_w) = (r^w/P(r))\tau(R_P)$, hence $\tau_r(I_w) = \tau_r(R_P)$ if and only if $P(r)/r^w \in \tau_r(R_P)$; thus if this holds for all $w \in \operatorname{Log} P$, then each of $P(r)/r^w$ belong to $\tau(R_P)$, and hence all their products do;

this means that for every formal monomial f, $1/\tau_r(f)$ belongs to $\tau_r(R_P)$, hence $\tau_r(fR_P)=\tau(R_P)$.

The upshot of this is that τ_r satisfies (1) if and only if for all $w \in \operatorname{Log} P$, we have $P(r)/r^w \in \tau_r(R_P)$. The latter is simply $\mathbb{Z}[r^w/P(r)]_{w\in \operatorname{Log} P}$. So we deduce this:

Lemma 3.1. *For $r \in (\mathbb{R}^d)^{++}$, τ_r satisfies* (1) *if and only if for all $v \in \operatorname{Log} P$, $P(r)/r^v \in \mathbb{Z}[r^w/P(r)]_{w\in \operatorname{Log} P}$.*

This is a fairly drastic condition, even when $d=1$ and $P=1+x$ or $2+3x$.

For $r \in (\mathbb{R}^d)^{++}$ and $P \in \mathbb{Z}[x_i^{\pm 1}]^+$, let $R_r=\mathbb{Z}[\{r^w/P(r)\}_{w\in \operatorname{Log} P}]$; this is exactly $\tau_r(R_P)$, and is a finitely generated unital subring of $\mathbb{R}$. The next lemma says that r satisfies (1) if and only if when we extend τ_r all the way up to $\mathbb{Z}[x_1^{\pm 1},\dots,x_d^{\pm 1}, P^{-1}]$, the image of τ_r does not increase — something we should have expected, in terms of the original definition.

Lemma 3.2. *Let $r=(r_i)\in(\mathbb{R}^d)^{++}$ and $P \in \mathbb{Z}[x_i^{\pm 1}]^+$ be projectively faithful. Then r satisfies* (1) *if and only if $R_r=\mathbb{Z}[r_i^{\pm 1};\, P(r)^{-1}]$.*

Proof. We may construct R_P by beginning with $\mathbb{Z}[x_i^{\pm 1}]$ (the Laurent polynomial ring) instead of $\mathbb{Z}[x_i]$; this is how it was originally constructed in [Handelman 1985; 1987]. By replacing P by x^vP^t for some $v \in \mathbb{Z}^d$ and positive integer t (this has no effect on R_P, up to order isomorphism), we can arrange for $\mathbf{0}$ to be in the interior of $\operatorname{cvx}\operatorname{Log} P$ and in $\operatorname{Log} P$. Then $1/P \in R_P$ and we may invert $1/P$, creating $R_P[P]=\mathbb{Z}[x_i^{\pm 1};\, P^{-1}]$ [Handelman 1987]. Let $I=(1/P)R_P$; this is an order ideal [Handelman 1987, p. 19], and $\mathbb{Z}[x_i^{\pm 1};\, P^{-1}]=\bigcup_{j\in\mathbb{Z}^+} P^jR_P$.

If r satisfies (1) with respect to P, then applying it to I, we obtain

$$\tau_r(I)=\tau_r(1/P)\tau_r(R)=(1/P(r))\tau_r(R)=(1/P(r))R_r.$$

By hypothesis, this is R_r, so $P(r)$ is a unit in R_r. Thus $\tau_r(P^jR_P)=P^j(r)R_r\subset R_r$. Taking the union, we obtain $\tau_r(\mathbb{Z}[x_i^{\pm 1};\, P^{-1}])\subseteq R_r$, and the reverse inclusion is trivial.

Conversely, suppose $R_r=\tau_r(\mathbb{Z}[x_i^{\pm 1};\, P^{-1}])$. Then $\tau_r(x_i^{\pm 1})=r_i^{\pm 1}$ and $\tau_r(P^{\pm 1})=P^{\pm 1}(r)$ belong to R_r and are invertible therein. Thus if f is any formal monomial, $\tau_r(f)$ is a product of terms of the form $r^w/P(r)$, and hence is invertible in R_r. Thus if I is an order ideal, it contains a formal monomial, and $\tau_r(I)$ contains an invertible element in R_r, and so $\tau_r(I)=R_r=\tau_r(R_P)$. Thus r satisfies (1). □

In other words, (1) holds if and only if the range of evaluation at r on R_P is the same as the range of the evaluation on the much larger ring $\mathbb{Z}[x_i^{\pm 1}, 1/P]$.

Now we consider what (1) means in the special case that $d=1$.

Let A be a unital subring of $\mathbb{C}$, the complexes. A complex number r is *integral over A* (or *r is an A-algebraic integer*) if it satisfies a monic polynomial with coefficients from A; equivalently, $r \in A[r^{-1}]$. The number r is an A-algebraic unit if it satisfies a monic polynomial with coefficients from A whose constant term is invertible in A; equivalently, $A[r] = A[r^{-1}]$. If $A = \mathbb{Z}$, we just write *integral* (adjective) or *algebraic integer* (noun). If $A = \mathbb{Q}$, these notions coincide, and we just say r is *algebraic.* The *degree* of an integral or algebraic element is the degree of its minimal polynomial (over A).

Lemma 3.3. *Let P be a projectively faithful element of $\mathbb{Z}[x]^+$ with smallest and largest degree coefficients a_0 and a_k respectively. If $r \in \mathbb{R}^{++}$ satisfies* (1) *with respect to P, then there exist nonnegative integers s and t such that a_0^s/r and $a_k^t r$ are integral.*

Proof. Write $P = a_0 + \sum_{0<i<k} a_i x^i + a_k x^k$, where a_i are nonnegative integers (some can be zero, but we still need $\gcd(\{i \mid a_i \neq 0\} \cup \{k\}) = 1$). From $P(r) \in \mathbb{Z}[\{r^j/P(r)\}_{j \in \mathrm{Log}\, P}]$, we deduce an equation of the form $P(r)^{m+1} = p(r)$, where $p \in \mathbb{Z}[x]$ and $\deg p \leq \deg P^m = km$. The leading term of this expression is $a_k^{m+1} r^{(m+1)k}$, and so r satisfies a monic polynomial with coefficients from $A = \mathbb{Z}[a_k^{-1}]$. It follows that $a_k^t r$ is integral for all sufficiently large s.

Replacing P by its *reversal* (also called *reciprocal*) $\widetilde{P}$ (defined by $\widetilde{P}(x) = P(x^{-1})x^k$), and redoing the process yields the other form, that a_0^s/r is integral. □

The following is true if we weaken the hypotheses on P to be projectively faithful (instead of requiring all the intermediate coefficients to be strictly positive). The modifications to the proof will muddy an already-complicated but elementary argument; so we just outline it afterwards. We can replace P by any power of itself, without changing anything, so the no-gaps condition is just that the second largest and second smallest terms have nonzero coefficients.

Proposition 3.4. *Let $r \in \mathbb{R}^{++}$ and $P \in \mathbb{Z}[x]^+$ be $\sum_{i=0}^k a_i x^i$, where all $a_i \neq 0$. Let a_0 and a_k be the coefficients of the least and greatest degree terms in P. Let $R_r = \mathbb{Z}[\{r^i/P(r)\}_{i \in \mathrm{Log}\, P}]$. Then the following are equivalent*:

(i) *r satisfies* (1) *with respect to P.*

(ii) *There exist nonnegative integers s and t such that both $a_k^s r$ and a_0^t/r are algebraic integers.*

(iii) $R_r = \mathbb{Z}[r^{\pm 1}, P(r)^{\pm 1}]$.

(iv) *For all $j \in \mathrm{Log}\, P$, we have $P(r)/r^j \in R_r$.*

Proof. We begin with (ii) implies (iv). Without loss of generality, we may assume $P = a_0 + \sum_{0<i<k} a_i x^i + a_k x^k$.

If c is an algebraic integer, then $\mathbb{Z}[c]$ is free on the $\mathbb{Z}$-basis $\{1, c, c^2, \ldots, c^{e-1}\}$, where e is the degree of c (this is an alternative definition of integrality); in particular, for every positive integer u, we can write $c^u = \sum_{i=0}^{e-1} b_i c^i$; in other words, there exists a polynomial $p \in \mathbb{Z}[x]$ of degree at most $e - 1$ such that $c^u = p(c)$.

Apply this to $c = a_k^s r$; for each positive integer u, we can write $(a_k^s r)^u = p_u(a_k^s r) = q_u(r)$, where $\deg q_u \leq e-1$. Multiplying this by $r^{u(s-1)}$, we obtain $(a_k r)^{us} = r^{u(s-1)} q_u$; setting $Q_u = x^{u(s-1)} q_u$, we have $(a_k r)^{us} = Q_u(r)$, where $Q_u \in \mathbb{Z}[x]$ and $\deg Q_u = u(s-1) + \deg q_u \leq u(s-1) + e - 1$. Hence (multiplying by an additional r^j), for every $j = 0, 1, 2, \ldots$, we have $Q_{u,j} \in \mathbb{Z}[x]$ such that $\deg Q_{u,j} = u(s-1) + j$ and $(a_k r)^{us+j} = Q_{u,j}(r)$. We will subsequently choose u to be fairly large.

Now let N be a (large) positive integer, and consider the k leading coefficients of P^N, that is, the coefficients of the terms $x^{kN}, x^{kN-1}, x^{kN-2}, \ldots, x^{kN-k+1}$. They are respectively divisible by $a_k^N, a_k^{N-1}, \ldots, a_k^{N-k+1}$ (as is trivially easy to see). Hence we may find integers b_i (with $b_0 = 1$) such that

$$P^N - \sum_{i=0}^{k-1} (a_k x)^{N-i} x^{N(k-1)} b_i := G$$

is a polynomial of degree at most $Nk - k$. Assume (as we may) that $N - k = us$ for some integer u. Replace each $(a_k x^{N-i})$ by $Q_{u,k-i}$; this has no effect on the value at r. Setting $H = \sum_{i=0}^{k-1} b_i Q_{u,k-i} x^{N(k-1)}$, we have $P^N(r) = (G+H)(r)$. Then

$$\begin{aligned} \deg(G+H) &\leq \max\{\deg G, \deg H\} \\ &\leq \max\{Nk-k, \max_i\{\deg Q_{u,k-i} + Nk - N\}\} \\ &\leq \max\{Nk-k, u(s-1)+e-1+Nk-N\} \\ &= \max\{Nk-k, Nk-N+e-1+N-k-u\} \\ &\leq \max\{Nk-k, Nk-k-u+e-1\}. \end{aligned}$$

We can choose $u \geq e-1$ at the outset, and so guarantee that $\deg(G+H) \leq Nk-k$. Thus $P(r) = (G+H)(r)/P^{N-1}(r)$. For every $0 \leq i \leq k$, we have $r^i/P(r) \in R_r$, and since $\deg(G+H) \leq Nk-k = \deg P^{N-1}$, we obtain $P(r) \in R_r$.

Now form the reversal of P, given by $\widetilde{P}(x) = P(x^{-1})x^k$; this reverses the roles of a_k and a_0, and the same process (using a_0^t/r being integral) yields, after translating back, $P(r)/r^k \in R_r$. From $P(r) \in R_r$, we obtain $r^i = (r^i/P(r)) \cdot P(r) \in R_r$ for $i \in \operatorname{Log} P$, and thus for all $i \geq 0$. Since $P(r)/r^k \in R_r$, we deduce $r^{-k} \in R_r$, hence $r^{-j} \in R_r$ for all $j \geq 0$; thus $P(r)/r^j \in R_r$.

Now (i) implies (ii) was done in the previous lemma, and the equivalence of (i), (iii), and (iv) follows from the general results preceding this. □

To prove the result when P is only projectively faithful, we can still write $P = a_0 + \sum_{1 \leq i \leq k-1} a_i x^i + a_k x^k$, where $\gcd\{i \mid a_i \neq 0\} = 1$ (equivalent after translation to projective faithfulness). Then it is elementary, and presumably well-known, that there exists M such that for all N, we have $(P^N, x^i) \neq 0$ if $M < i < kN - M$. Now in the construction above, make sure that when the multiplications by powers of r take place, the exponent lands in the interval where all the coefficients are guaranteed to be nonzero (we are of course free to take arbitrary large powers of P).

A strange consequence is that when the hypotheses on P are satisfied, the set of r such that τ_r satisfies (1) is closed under multiplication; this follows immediately from (ii), but not obviously from any of the other equivalent properties.

Multiplicativity does not appear to extend to more than one variable. For example, if $P = 2 + 3x + 5y$, and we restrict to $r = (m, n)$ with positive integer coordinates, it is tedious but routine to see that τ_r satisfies (1) with respect to P if and only if for all primes p and q,

$$p \mid m \Longrightarrow p \mid (2 + 5n) \quad \text{and} \quad q \mid n \Longrightarrow q \mid (2 + 3m).$$

For example, (7, 1), (3, 11), $(2^i, 2^j)$ (where both $i, j > 0$) satisfy these conditions, but (14, 2) does not. There may be another, more appropriate, notion of multiplication with respect to which the set is closed.

Another general property concerns approximate divisibility. Let $K = \operatorname{cvx} \operatorname{Log} P$; this is a compact convex polytope. Let $e \in K$ be an extreme point (we do not use the usual term, *vertex,* because this might be confused with lattice point); then $v \in \operatorname{Log} P$, and there is a pure trace σ^v associated with v, given by $\sigma^v(g/P^k) = (g, x^{kv})/(P, x^v)^k$ (this can also be obtained as the limit along a path of τ_r), via l'Hôpital's rule, as in [Handelman 1985, Section III] (especially just before III.3).

Since every order ideal of R_P is of the form $\sum f_i R_P$ (finite sum), if we assume that R_P is approximately divisible, then R_P is nearly divisible. Thus every order ideal has its own order unit and is approximately divisible. If τ is faithful, then $\tau(I \cap J) \neq 0$ (no finite intersections of order ideals can be zero since they are also ideals in a domain), and $I \cap J$ is itself approximately divisible, hence $\tau(I \cap J)$ is dense in $\mathbb{R}$. Thus for any faithful trace that is order unit good for R_P, its restriction to any nonzero order ideal is also order unit good.

Thus we have the following.

Lemma 3.5. *The ordered ring R_P is approximately divisible if and only if for all extreme points v of $K = \operatorname{cvx} \operatorname{Log} P$, we have $(P, x^v) > 1$.*

Lemma 3.6. *Let $P = \sum \lambda_w x^w \in \mathbb{Z}[x_i^{\pm 1}]^+$ with $(P, x^v) > 1$ for all extreme points of $K = \operatorname{cvx} \operatorname{Log} P$.*

(a) *Then R_P is nearly divisible*

(b) *If τ is a faithful trace that is order unit good for R_P, then its restriction to any nonzero ideal is order unit good for that ideal.*

If we replace R_P by $S_P := R_P \otimes \mathbb{Q} = \mathbb{Q}[x^w/P]$, then it is divisible, which is of course stronger than nearly divisible, so that (a) holds automatically (without the hypothesis on the coefficients at extreme points), and (b) also holds by the same arguments.

Proposition 3.7. *Let $r = (r_i) \in (\mathbb{R}^d)^{++}$, and let $P \in \mathbb{Z}[x_i^{\pm 1}]^+$ be projectively faithful.*

(a) *The pure trace τ_r on R_P is good if and only if*

(i) *τ_r is order unit good for R_P and*

(ii) *for all $v \in \operatorname{Log} P$, we have $P(r)/r^v \in \mathbb{Z}[r^w/P(r)]_{w \in \operatorname{Log} P}$.*

(b) *The pure trace τ_r on S_P is good if and only if*

(i) *τ_r is order unit good for R_P.*

Remark. Note the absence of (ii) from (b), and the appearance of R_P in (b)(i). It is known (along the same lines as in [BeH 2014, Proposition 5.10]), that if τ_r is order unit good (for either coefficient ring), then each r_i is algebraic. Since $\mathbb{Q}[r_1, \ldots, r_d]$ is thus a field, (ii) is redundant in (b).

Proof. We show that if τ_r is order unit good (which means that the closure of the image of $\ker \tau_r$ in $\operatorname{Aff} S(R, 1)$ is exactly $\tau_r^{\perp} = \{h \in \operatorname{Aff} S(R, 1) \mid h(\tau_r) = 0\}$), then its restriction to any order ideal is also order unit good. It suffices to do this for $I = f R_P$, where f is a formal monomial.

The map $R_P \to f R_P$, given by $r \mapsto f r$, is an order-isomorphism of R_P modules (this of course uses the fact that $f r \geq 0$ in R_P entails $r \geq 0$). Using f as an order unit for I, the map on traces $\tau \mapsto \tau/\tau(f)$ (restricted to those τ such that $\tau(f) \neq 0$) sends $\tau_r \to \tau_r/\tau_r(f) = \tau'$, and $\ker \tau' = \ker \tau_r \cap f R_P = f \cdot \ker \tau_r$ (since $f(r) \neq 0$). The map between R_P modules induces an affine homeomorphism between $S(R_P, 1)$ and $S(I, f)$, sending τ_r to τ', and it easily follows that τ' is order unit good. But τ' is just the normalization of $\tau | I$, hence the latter is order unit good.

The rest follows from the preceding results. □

In one variable, we can show that τ_r is order unit good if and only if none of the algebraic conjugates of r (except itself) are positive real. In more than one variable, the situation is far more complicated, and there is no decisive theorem (yet).

Example. Let $d = 1$ and $P = 1 + x$; then we can rewrite $R_P = \mathbb{Z}[1/P, x/P] = \mathbb{Z}[1 - X, X]$, where $X = x/(1 + x)$, and the positive cone translates to $\langle X, 1 - X \rangle$. This goes back to Renault [1980]. The translation, however, obscures some of the features, as we will see. First, R_P has two discrete pure traces, $\tau_0 = \sigma^0$ and $\tau_\infty = \sigma^1$ (0 and 1 are the extreme points of the convex set $\operatorname{cvx} \operatorname{Log} P = [0, 1]$), so it is not approximately divisible. However, it is interesting to calculate the condition that $\tau_r(I) = \tau_r(R_P)$ for all nonzero order ideals.

By Proposition 3.7 above, this amounts to $1 + r,\ 1 + 1/r \in \mathbb{Z}[1/(1+r), r/(1+r)]$; as $r/(1+r) = 1 - 1/(1+r)$, the condition (1) is equivalent to $1 + r^{\pm 1} \in \mathbb{Z}[1/(1+r)]$. Now for a real number s, the condition $s \in \mathbb{Z}[1/s]$ is equivalent to s being an algebraic integer (that is, satisfying a monic integer polynomial). Hence we infer that if (1) holds for τ_r, then r has to be an algebraic unit (that is, not only is its minimal polynomial over the integers monic, but the constant term must be ± 1 as well). Conversely, if r is an algebraic unit, then the desired membership property holds.

We conclude that τ_r satisfies (1) if and only if r is an algebraic unit.

In particular, if r is an integer, then τ_r satisfies (1) if and only if $r = 1$ (we are restricting ourselves to actual traces, hence excluding negative values for r).

The translation, $X = x/(1+x)$ converts r to $r/(1+r)$; then of course $\tau(X)$ is a fractional linear transformation of an algebraic unit, but this characterization is not as pleasant as the pretranslation version. □

Let $V \subset \mathbb{C}^d$. For A a subring of $\mathbb{C}$, define $I_A(V)$ to be the ideal in the polynomial ring $A[x_1, \dots, x_d]$ consisting of polynomials that vanish at all points of V. Given an ideal I of $A[x_1, \dots, x_d]$, define $Z_A(I)$ to be the common zero set (in $\mathbb{C}^d$) of all elements of I. The *variety generated by V over A* is simply $Z_A I_A(V)$. If $A = \mathbb{Z}$, we drop the subscript.

We say $r = (r_i) \in (\mathbb{R}^d)^{++}$ is *really isolated* if $ZI(\{r\}) \cap (\mathbb{R}^d)^{++} = \{r\}$. For example, if $d = 1$, then r is really isolated if r is algebraic and all algebraic conjugates of r other than r itself are *not* positive real. In general, r is really isolated means that the slice of the variety generated by r (or more simply, the Zariski closure of $\{r\}$) by the positive orthant contains only r.

The argument in [BeH 2014, Proposition 5.10] shows that if r is really isolated (or more generally, $\{r\}$ is an isolated point in $(\mathbb{R}^d)^{++} \cap ZI(\{r\})$), then all of its coordinates are algebraic (there is an assumption in [BeH 2014] concerning interior points which is automatic here). We remind the reader that we have assumed that P is projectively faithful, which implies in particular, that its Newton polytope contains a d-ball.

The condition that r be really isolated appears in [BeH 2014, Examples 5 and 10], for which the relevant dimension groups are remotely related to the ones appearing here.

Proposition 3.8. *Suppose R_P is approximately divisible, and τ is a pure faithful trace. Then:*

(a) *τ is an order unit good trace of R_P if and only if $\tau = \tau_r$, where $r \in (\mathbb{R}^d)^{++}$ is really isolated.*

(b) *τ_r is a good trace of R_P if and only if r is really isolated and for all $v \in \operatorname{Log} P$, we have $P(r)/r^v \in \mathbb{Z}[\{r^w/P(r)\}_{w \in \operatorname{Log} P}]$.*

(c) *τ_r is a good trace of $R_P \otimes \mathbb{Q}$ if and only if r is really isolated.*

Proof. Every pure faithful trace of R_P is of the form τ_r for (a unique) r in the positive orthant.

If r is not really isolated, then there exists $r' \in (\mathbb{R}^d)^{++}$ such that every polynomial that vanishes at r also vanishes at r'. Suppose $a := g/P^k \in R_P$; we may assume $\operatorname{Log} g \subseteq \operatorname{Log} P^k$. If $\tau_r(a) = 0$, then $g(r) = 0$, hence $g(r') = 0$, whence $\tau_{r'}(a) = 0$;

thus with $\sigma = \tau_{r'}$, we have $\sigma \in \partial_e S(R, 1) \setminus \{\tau_r\}$ such that $\sigma \,|\, \ker \tau_r \equiv 0$. Hence τ_r is not order unit good. The same of course applies with $R_P \otimes \mathbb{Q}$ in place of R_P.

Conversely, suppose that r is really isolated, but there exists $\sigma \in \partial_e S(R, 1) \setminus \{\tau_r\}$ such that $\sigma \,|\, \ker \tau_r = 0$. Then σ cannot be faithful (as otherwise, $\sigma = \tau_{r'}$ for some $r' \in (\mathbb{R}^d)^{++}$, and $r' \in ZI(\{r\})$). Consider $S = R_P \otimes \mathbb{Q}$, and let T_r, Σ be the extension to S of τ_r and Σ (both extend, since the ranges are torsion-free abelian groups). Then $T_r(S) = \mathbb{Q}[r^w/P(r)]$, which is a field (since the coordinates are algebraic, so are all the $r^w/P(r)$). Then $\ker T_r$ is a field, so $\ker T_r$ is a maximal ideal. Also, $\ker T_r \cap R_P = \ker \tau_r$ and $\ker \Sigma \cap R_P = \sigma$. If $\ker \tau_r \subseteq \ker \sigma$, then $\ker T_r \subset \ker \Sigma$, but maximality of $\ker T_r$ implies $\ker T_r = \ker \Sigma$, and thus $\ker \tau_r = \ker \sigma$. However, since σ is not faithful, $\ker \sigma$ contains a positive nonzero element of R_P, whereas $\ker \tau_r$ does not, a contradiction.

Hence if r is really isolated, then $\sigma \in \partial_e S(R_P, 1) \setminus \{\tau_r\}$ implies $\sigma(\ker \tau_r) \neq 0$, and by Lemma 1.7 above, this implies τ_r is order unit good. The same of course applies to T_r as a trace on S_P. This yields (a), and contributes to (c).

Part (b) now follows from preceding results in this section.

Part (c) comes from $\mathbb{Q}[r^w/P(r)]$ being a field (which in turn arises because the coordinates of r are algebraic), so that condition (1) is automatic. □

A particular consequence is that the set of good pure faithful traces of $S_P = R_P \otimes \mathbb{Q}$ is the same for all choices (with d fixed) of faithfully projective $P \in \mathbb{Z}[x_i]^+$ (or $P \in \mathbb{Q}[x_i]^+$), whereas for R_P, there is dependence on P.

When $d = 1$, the conditions for τ_r to be good are precisely that no distinct algebraic conjugate of r be positive and the integrality condition, (ii), of Proposition 3.4.

Example. Let $d = 1$ and $P = 2+3x$. By Proposition 3.4, the positive real number r satisfies (1) if and only if there exists s such that both $2^s/r$ and $3^s r$ are integral. Let $K = \mathbb{Q}(r)$, and $\mathbb{Z}_K$ denote the ring of integers in K. The fractional ideal $r\mathbb{Z}_K$ factors as $\prod \mathcal{P}_i / \prod \mathcal{Q}_j$ (where $\mathcal{P}_i$ and $\mathcal{Q}_j$ are prime ideals in $\mathbb{Z}_K$, and we allow repetitions; the products might also be over the empty set). The intersections $\mathcal{P}_i \cap \mathbb{Z}$ and $\mathcal{Q}_j \cap \mathbb{Z}$ determine primes in $\mathbb{Z}$, denoted respectively p_i and q_j. Then (1) is equivalent to $p_i = 2$ and $q_j = 3$ for all i and j. Hence τ_r is good for R_P if and only if no nonidentity algebraic conjugate is positive and the prime factorization of the fractional ideal $r\mathbb{Z}_K$ consists of primes sitting over 2 in the numerator and over 3 in the denominator.

In this section, we have restricted ourselves to pure *faithful* traces; this is a technical convenience. By the comment after Proposition 1.1, we can factor out the largest order ideal contained in the kernel of a trace, and in the case that the dimension group is R_P, these correspond to quotients corresponding to faces of the Newton polytope [Handelman 1985, Section VII]. This amounts to a reduction to a lower dimensional lattice and vector space, that is, a polynomial in fewer variables.

There are related naturally occurring classes of dimension groups whose pure traces can be similarly analyzed. For example, for the matrix-valued random walks appearing in [Handelman 2009], in nondegenerate cases, the pure faithful traces are similarly parameterized by the positive orthant (the nonfaithful traces are generically terrible, but can be analyzed in reasonable cases). An example appears in [Petersen 2012], where very specific local limit asymptotics were used to derive the one-parameter family (indexed by the unit interval) of pure traces. In fact, that random walk can be represented as $M = \left(\begin{smallmatrix} 1+x & x \\ 1 & 0 \end{smallmatrix}\right)$, and in this very simple case, via [Handelman 2009], we can write down the pure traces parameterized by $[0, \infty]$ (the endpoints corresponding to the two nonfaithful pure traces) via the large eigenvalue function. Alternatively (in the notation of [Handelman 2009]), it is elementary that $(1+x)\widehat{M}^{-1}$ is an order unit in $E_b(G_M)$, so on setting $P = 1+x$, we can view M/P as a matrix with entries in R_P without changing the pure trace space. This yields a parameterization of the pure traces by those of R_P (again via the large eigenvalue function, an algebraic function), which are indexed by the unit interval.

4. Direct sums and goodness

For (noncyclic) simple dimension groups, there is a notion of direct sum (corresponding to coproduct; see [BeH 2014, Appendix B] for a discussion). This actually extends to nearly divisible dimension groups. Let G and H be dimension groups. Form the group direct sum $K = G \oplus H$, and impose on it the *strict ordering* given by the positive cone

$$K^+ := \bigl\{(g, h) \bigm| g \in G^+ \setminus \{0\} \text{ and } h \in H^+ \setminus \{0\}\bigr\} \cup \{(0, 0)\}.$$

We see immediately that K is an unperforated partially ordered group; we denote it by $G \oplus_s H$, although we frequently suppress the subscript s. In general, K need not be a dimension group (as a simple example, if G is simple and $H = \mathbb{Z}$, then K is a simple partially ordered abelian group with a discrete trace, and hence cannot be a dimension group [Goodearl 1986, Proposition 4.22]).

A partially ordered abelian group G is *prime* if the intersection of any two nonzero order ideals contains a nonzero positive element; for dimension groups, this definition simplifies to “the intersection of any nonzero order ideals is nonzero”. Here is a natural generalization of [Effros et al. 1980, Corollary 1.2].

Lemma 4.1. *Let G and H be dimension groups. Then the strict direct sum $K = G \oplus_s H$ is a dimension group if and only if both of the following conditions hold*:

(a) *Both G and H are prime.*

(b) *Both G and H are nearly divisible.*

Proof. Assume (a) and (b). The unperforation and directedness of K are trivial, so it suffices to prove Riesz decomposition. Suppose $0 \le a \le b+c$, where $a = (g, h)$, $b = (e, f)$, and $c = (k, l)$ and all of a, b, c are in $K^+ \setminus \{0\}$ (if any of a, b, or c is zero, the decomposition is immediate). This entails all of g, h, e, f, k, l are nonzero positive elements of their respective groups, and moreover, that either $a = b+c$ (in which case, the decomposition condition is satisfied) or both $g < e+k$ and $h < f+l$ in their respective groups. Assume the latter.

By Riesz decomposition in G, there exist $0 \le e_1 \le e$ and $0 \le k_1 \le k$ such that $g = e_1 + k_1$. If $e = e_1$, then $k_1 < k$, and the intersection of the order ideals $\langle e\rangle \cap \langle k - k_1\rangle$ is thus nonzero, and contains a nonzero positive element, z. The order ideal $\langle z\rangle$ has z as an order unit, hence (by (b)) is approximately divisible, and it follows immediately that there exists a positive, nonzero x such that $x < e, k - k_1$. Then we can write $g = (e_1 - x) + (k_1 + x)$. Now $e - (e_1 - x) > 0$, and $k - (k_1 + x) > 0$.

Now suppose that $e_1 = 0$, so that $g = k_1$. By the same procedure as in the previous paragraph, there exists a nonzero positive x such that $x < e, k_1$, so we can write $g = x + (k_1 - x)$, with $x < e$ and $k_1 - x < k$.

This leaves the case that $0 < e_1 < e$. If k_1 is zero or k, we reverse the roles of e and k and apply the preceding, so that in all cases, we can find nonzero positive $x_i \in G$ such that $g = x_1 + x_2$, with $0 < x_1 < e$ and $0 < x_2 < k$.

By applying the preceding to H in place of G, we obtain $y_i \in H$ such that $h = y_1 + y_2$ with $0 < y_1 < f$ and $0 < y_2 < l$. Then $g = (x_1, y_1) + (x_2, y_2)$ is the desired decomposition.

Conversely, suppose that K satisfies Riesz decomposition. If H were not prime, we could find nonzero y, z in H^+ such that $0 \le h \le y, z$ implies $h = 0$. Consider, for $g \in G^+ \setminus \{0\}$, $(5g, y) \le (3g, y) + (3g, z)$: this holds in K; hence if Riesz decomposition applies, we can write $(5g, y) = (3g, y_1) + (3g, z_1)$ with the latter two terms in K^+, and at the very least $y_1 \le y$ and $z_i \le z$. In particular $y = y_1 + z_1$, but since $z_1 \le y, z$, we have $z_1 = 0$, so that $y_1 = y$. But this entails $(3g, 0)$ is in the positive cone of K, which is a contradiction, since $g > 0$. Now the same argument applies to G, so both have to be prime.

Suppose that K is a dimension group; we can also assume that G admits an order unit. Now assume that H is not nearly divisible. Then H admits an order ideal with its own order unit, (I, w), that has a discrete pure trace; call it τ. Then $\ker \tau$ is a maximal order ideal of H; call it T. Since G has an order unit, it has a maximal order ideal, J. Consider $J \oplus_s T$. This is an order ideal of K, and it is easy to verify that $K/(J \oplus T)$ is order isomorphic to $K' := (G/J) \oplus_s \mathbb{Z}$. Since both pieces are simple, K' is simple, but admits a discrete trace (projection on the second coordinate). Thus K' is not a dimension group. On the other hand, since K is a dimension group, and the quotient of it by an order ideal is also a dimension group, we have a contradiction. Hence H must be nearly divisible. The same applies with H replaced by G. $\square$

If $K = G \oplus_s H$, and σ and τ are traces on G and H respectively, we consider the possibility that $\phi := \sigma \oplus \tau$ (defined by $(g, h) \mapsto \sigma(g) + \tau(h)$) is good or order unit good. This turns out to be surprisingly interesting. Iteration of this process yields some weird examples.

Lemma 4.2 (a consequence of the method of proof of [BeH 2014, Proposition 1.7]). *Suppose (K, w) is an approximately divisible dimension group with order unit, and ϕ is an order unit good trace. Then whenever $a \in G$, $b \in G^{++}$ and $0 < \phi(a) < \phi(b)$, for all $\epsilon > 0$, there exists $a' \in [0, b]$ such that $\phi(a') = a$ and $\|\hat{a}' - \hat{b}\sigma(a)/\sigma(b)\| < \epsilon$.*

Proof. Approximate divisibility implies density of G in $\operatorname{Aff} S(G, u)$. Set $j = \sigma(b)\hat{b}/\sigma(a)$, so that $j(\sigma) = \sigma(a)$ and $\inf j = \sigma(a)\sigma(b)^{-1} \inf \hat{b}$. There exists $g_n \in G$ such that $\hat{g}_n \to j$ uniformly. If for infinitely many n, we have $g_n(\sigma) = \sigma(a)$, we are done (taking large enough n). Otherwise, select $\sigma(a)(\sigma(b)2)^{-1} \inf b > \epsilon > 0$ and $\|\hat{g}_n - j\| < \epsilon$. Then $|\sigma(g_n) - \sigma(a)| < \epsilon$ provided n is sufficiently large; if $\sigma(g_n) > \sigma(a)$, set $c_n = g_n - a$. There exists an order unit z_n such that $0 < \sigma(c_n)\mathbf{1} < \hat{z}_n < 2\epsilon$. By order unit goodness, there exists $v_n \ll z_n$ such that $\sigma(c_n) = \sigma(v_n)$, and of course, $\|v_n\| \le \|\hat{z}_n\| < 2\epsilon$. Then $g_n - v_n$ has image within 3ϵ of j, and it is easy to check that $g_n - v_n$ is strictly positive, and hence is an order unit.

If instead, $\sigma(a) > \sigma(c_n)$ for infinitely many n, we obtain a corresponding $c_n = g_n - a$ and $v_n \ll z_n$, and this time, $g_n + v_n$ has all the right properties. In both cases, by taking n sufficiently large, we make the error terms go to zero, and hence obtain the a' as one of $g_n \pm v_n$. □

In the following, the function ψ need not be a group homomorphism.

Lemma 4.3. *Suppose G and H are nearly divisible dimension groups, each with order unit, and respective trace σ and τ. Let $K = G \oplus H$ with the strict ordering, and suppose that the trace on K, $\phi := \sigma \oplus \tau$, is order unit good. Then provided the following condition holds, σ is order unit good as a trace on G:*

- *There exists a function $\psi : \tau^{-1}(\sigma(G) \cap \tau(H)) \to \sigma^{-1}(\sigma(G) \cap \tau(H))$ that is pseudonorm continuous with the additional property that $\sigma\psi = \tau$.*

Remark. As we will see below, without the weird extra condition, the result fails.

Proof. Select an order unit b in G, and a in G such that $0 < \sigma(a) < \sigma(b)$. As H is approximately divisible, there is a sequence of order units (h_n) in H such that $h_n \to 0$ (with respect to the pseudonorm topology on H; equivalently, as functions on $S(H, v)$, $\hat{h}_n$ converges uniformly to zero). There also exists δ in G such that $\sigma(b - a)/4 < \hat{\delta} < \min\{\sigma(b - a)/2, \inf_{\theta \in S(G,u)} \theta(b)/2\}$ uniformly on $S(G, u)$. Then $B_n := (b - \delta, h_n)$ are order units of $G \oplus H$, and $\phi(a, 0) < \phi(B_n) = \sigma(b) - \sigma(\delta) + \tau(h_n)$.

Since ϕ is order unit good and each B_n is an order unit, there exist (a_n, z_n) such that $0 \ll a_n \ll b - \delta$ and $0 \ll z_n \ll h_n$, with $\phi((a_n, z_n)) = \sigma(a)$, and by the previous lemma, $\inf_{S(G,u)} \hat{a}_n$ is bounded below (as $n \to \infty$); in particular, $\|z_n\|_H \to 0$ and $\sigma(a_n) + \tau(z_n) = \sigma(a)$. Thus $z_n \in \tau^{-1}(\sigma(G) \cap \tau(H))$, so we may consider the sequence $\psi(z_n) \in \sigma^{-1}(\sigma(G) \cap \tau(H))$. Since ψ is pseudonorm continuous, $\widehat{\psi(z_n)} \to 0$ uniformly on $S(G, u)$.

Consider $a_n + \psi(z_n)$; its value at σ is $\sigma(a_n) + \sigma(\psi(z_n)) = \sigma(a_n) + \tau(z_n) = \sigma(a)$. If we choose n sufficiently large so that $\|\widehat{\psi(z_n)}\| < \inf \delta$, then $a_n + \psi(z_n) \ll b - \delta + \delta = b$. In addition, we can also choose n sufficiently large so that $\inf \widehat{\psi(z_n)} > -\inf_{S(G,u)} \hat{a}_n$, by the uniform boundedness below of the a_n (there is no guarantee that $\psi(z_n)$ is positive). Then $a_n + \psi(z_n)$ is an order unit in the interval $[0, b]$ and we are done. □

One advantage of not requiring normalization of σ and τ is that we can replace them by any positive scalar multiples in testing for order unit goodness of $\lambda\sigma \oplus \mu\tau$; the first hypotheses are unchanged, but the second translates to density of $(\lambda\sigma(G)) \cap (\mu\tau(G))$ in $\mathbb{R}$. In the following, we cannot apply earlier results directly, since $G \oplus 0$ is not an order ideal of $G \oplus H$ (strict ordering).

Lemma 4.4. *Suppose that σ is a trace on G, τ is a trace on H, and $\sigma \oplus \tau = \phi$ is order unit good for $K = G \oplus H$ with the strict ordering, and moreover assume that each of G and H is nearly divisible. Then $\sigma(G) \cap \tau(H)$ is dense in $\mathbb{R}$.*

Proof. We use the characterization of order unit good traces on approximately divisible dimension groups; namely $\ker \phi$ has dense image in $\phi^{\vdash}$ [BeH 2014, Proposition 1.7].

Suppose the intersection is not dense; then there exists a real $\delta \geq 0$ such that $\sigma(G) \cap \tau(H) = \delta\mathbb{Z}$. We have that $\ker \phi$ has dense range in $\operatorname{Aff} S(K, (u, v)) = \operatorname{Aff} S(G, u) \times \operatorname{Aff} S(H, v)$. But

$$\ker \phi = \{(g, h) \in G \oplus H \mid \sigma(g) = -\tau(h)\}.$$

If $\delta = 0$, then $\ker \phi = \ker \sigma \oplus \ker \tau$ (since $\sigma(g) = -\tau(h)$ implies $\sigma(g) \in \tau(H) \cap \sigma(H)$, and hence is zero). The image of $\ker \phi$ is then contained in $\sigma^{\vdash} \times \tau^{\vdash}$, which is closed and of codimension two in $\operatorname{Aff} S(K, (u, v))$, and so $\ker \phi$ cannot be dense in $\phi^{\vdash}$ (which has codimension one), hence ϕ cannot be order unit good.

If $\delta \neq 0$, select g and h in G and H respectively such that $\sigma(g) = \delta = \tau(h)$. Then it is easy to see that $\ker \phi = (\ker \sigma \oplus \ker \tau) + (g, -h)\mathbb{Z}$, and then its range is contained in $(\sigma^{\vdash} \times \tau^{\vdash}) + (\hat{g}, -\hat{h})\mathbb{Z}$. However, the latter is closed (easy to see), and so the image of $\ker \phi$ is contained in a proper closed subspace (with a discrete direct summand) of $\phi^{\vdash}$; hence in this case as well, ϕ is not order unit good. □

Now we want to determine when $\sigma \oplus \tau$ is good or order unit good. Let $\pi_G : G \oplus H \to G$ and $\pi_H : G \oplus H \to H$ be the obvious projection maps. Unlike the inclusions $G, H \to G \oplus H$, these *are* order-preserving. First, consider $\sigma \circ \pi_G : \ker \phi \to \sigma(G) \cap \tau(H) \subseteq \mathbb{R}$. The kernel is exactly $\ker \sigma \oplus \ker \tau$; we also note that σ extends to a map $\Sigma : \phi^{\vdash} \to \mathbb{R}$ (sending (j, l) to $j(\sigma)$), the kernel of which is

$\sigma^{\vdash} \times \tau^{\vdash}$. Via the identification of $\mathrm{Aff}\, S(K, (u, v))$ with $\mathrm{Aff}\, S(G, u) \times \mathrm{Aff}\, S(H, v)$, we have the following diagram:

$$\begin{tikzcd}
0 \arrow[r] & \ker\sigma \oplus \ker\tau \arrow[r] \arrow[d] & \ker\phi \arrow[r, "\sigma\circ\pi_G"] \arrow[d] & \sigma(G)\cap\tau(H) \arrow[r] \arrow[d] & 0 \\
 & \overline{\widehat{\ker\sigma}} \times \overline{\widehat{\ker\tau}} \arrow[r] \arrow[d] & \overline{\widehat{\ker\phi}} \arrow[r, "\Sigma"] \arrow[d] & \mathbb{R} \arrow[d] & \\
0 \arrow[r] & \sigma^{\vdash} \times \tau^{\vdash} \arrow[r] & \phi^{\vdash} \arrow[r, "\Sigma"] & \mathbb{R} \arrow[r] & 0
\end{tikzcd}$$

The long horizontal overlines indicate closure, as subgroups of the affine function vector spaces; of course, there is no requirement that any of the three overlined groups be real vector spaces (they are norm-complete subgroups). The two leftmost top vertical arrows are just induced by the affine representations; the right one is the inclusion, compatible with Σ restricted to the image of $\ker\phi$. The two leftmost bottom vertical arrows are the obvious inclusions. The Σ in the middle row is an abuse of notation; it represents the restriction of Σ to $\overline{\widehat{\ker\phi}}$, the closed subgroup of $\phi^{\vdash}$, but the notation is already rather complex.

The middle row need not be exact at either end. For example, if $\ker\phi$ has dense image in $\phi^{\vdash}$ but one or both of $\ker\sigma$ or $\ker\tau$ does not have dense image in $\sigma^{\vdash}$ or respectively $\tau^{\vdash}$, then it is not left exact; if $\sigma(G)\cap\tau(H)$ is discrete, then the middle line is not right exact.

If $\ker\phi$ has dense image in $\phi^{\vdash}$, then $\sigma(G)\cap\tau(H)$ is a dense subgroup of $\mathbb{R}$: we simply note that density of the image of $\ker\phi$ in $\phi^{\vdash}$, the latter being a closed and therefore a norm-complete subspace of $\mathrm{Aff}\, S(K, (u, v))$, entails that for every bounded linear functional that is not zero on $\phi^{\vdash}$, its restriction to a dense subgroup must be not zero and have dense range in the reals. □

It also leads to a straightforward proof that if $\ker\sigma$ and $\ker\tau$ have dense images in $\sigma^{\vdash}$ and $\tau^{\vdash}$ respectively, and if $\sigma(G)\cap\tau(H)$ is a dense subgroup of $\mathbb{R}$, then $\ker\phi$ has dense image in $\phi^{\vdash}$. We have that

$$\sigma^{\vdash} \times \tau^{\vdash} = \overline{\widehat{\ker\sigma}} \times \overline{\widehat{\ker\tau}} \subseteq \overline{\widehat{\ker\phi}} \subseteq \phi^{\vdash}.$$

The left and right terms of these inclusions are vector spaces, and since $\sigma^{\vdash} \times \tau^{\vdash}$ is a closed codimension-two subspace of $\mathrm{Aff}\, S(K, (u, v))$ and $\phi^{\vdash}$ is codimension one, it follows that $\sigma^{\vdash} \times \tau^{\vdash}$ is a codimension-one subspace of $\phi^{\vdash}$. (The proof does not stop here — we do not know that $\overline{\widehat{\ker\phi}}$ is a real vector space.)

The map Σ induces a map from

$$\overline{\widehat{\ker\phi}}/(\overline{\widehat{\ker\sigma}} \oplus \overline{\widehat{\ker\tau}})$$

to a subgroup of the reals. However, this subgroup of the reals includes the dense subgroup $\sigma(G) \cap \tau(H)$, and as $\overline{\widehat{\ker\phi}}$ is a norm-complete abelian group, the image must be complete, and thus must be onto. In addition, since

$$\ker\Sigma = \sigma^{\vdash} \times \tau^{\vdash} = \overline{\widehat{\ker\sigma}} \times \overline{\widehat{\ker\tau}},$$

it follows that

$$\ker\Sigma \cap \overline{\widehat{\ker\phi}} = \overline{\widehat{\ker\sigma}} \times \overline{\widehat{\ker\tau}}.$$

We thus have

$$\ker\Sigma \subset \overline{\widehat{\ker\phi}} \subseteq \phi^{\vdash},$$

but Σ induces the equality

$$\overline{\widehat{\ker\phi}}/(\ker\Sigma \cap \overline{\widehat{\ker\phi}}) = \phi^{\vdash}/\ker\Sigma.$$

It follows immediately that $\overline{\widehat{\ker\phi}} = \phi^{\vdash}$.

Now we can show that if the closure of the images of $\ker\sigma$ and $\ker\tau$ are real vector spaces, and if $\ker\phi$ is order unit good, then σ and τ are order unit good.

We wish to show

$$\overline{\widehat{\ker\sigma}} \times \overline{\widehat{\ker\tau}} = \sigma^{\vdash} \times \tau^{\vdash},$$

as from this it follows trivially that

$$\overline{\widehat{\ker\sigma}} = \sigma^{\vdash} \quad \text{and} \quad \overline{\widehat{\ker\tau}} = \tau^{\vdash}.$$

Since the left thing is a vector space, and a complete normed abelian group (hence a closed vector subspace of $\operatorname{Aff} S(K, (u, v))$), if equality does not hold, there exists a bounded linear functional α on $\operatorname{Aff} S(K, (u, v))$ that kills $\overline{\widehat{\ker\sigma}} \times \overline{\widehat{\ker\tau}}$ but not $\sigma^{\vdash} \times \tau^{\vdash}$; in particular, α does not kill $\phi^{\vdash}$.

By composition with the affine representation, we "restrict" α to a real-valued bounded group homomorphism $\beta : G \oplus H \to \mathbb{R}$ (for a treatment of bounded group homomorphisms on dimension groups, see [Goodearl 1986]; their behaviour is just what you'd expect). Since α kills $\overline{\widehat{\ker\sigma}} \times \overline{\widehat{\ker\tau}}$, it follows that β kills $\ker\sigma \oplus \ker\tau$. We form the normed abelian group $\ker\phi/(\ker\sigma \oplus \ker\tau)$, which via σ, we know to be $\sigma(G) \cap \tau(H) \subset \mathbb{R}$. Thus β induces a bounded real-valued group homomorphism on $\ker\phi/(\ker\sigma \oplus \ker\tau)$; call it $\bar{\beta}$. We thus have two bounded group homomorphisms on the quotient, $\bar{\beta}$ and $\bar{\sigma}$, but as the quotient is isomorphic (as a normed abelian group) to a subgroup of the reals, there must be a positive real number λ such that $\bar{\beta} = \lambda\bar{\sigma}$.

This forces $\beta = \lambda \cdot \sigma \circ \pi_G$ (as bounded group homomorphisms on $\ker\phi$). Since $\ker\phi$ has dense image in its completion (!) which happens to be $\phi^{\vdash}$, we have that $\alpha|\phi^{\vdash} = \lambda\Sigma$. Thus α kills $\sigma^{\vdash} \times \tau^{\vdash}$, a contradiction. $\square$

To summarize, we have the following results.

Proposition 4.5. *Suppose (G, u, σ) and (H, v, τ) are nearly divisible dimension groups with order unit and distinguished trace, and form $K = G \oplus H$, and the trace $\phi = \sigma \oplus \tau : K \to \mathbb{R}$.*

(a) *If ϕ is order unit good (with respect to either the usual or the strict ordering on K), then $\sigma(G) \cap \tau(H)$ is a dense subgroup of the reals, and $\sigma \otimes 1_{\mathbb{Q}}$ and $\tau \otimes 1_{\mathbb{Q}}$ are order unit good as traces on $G \otimes \mathbb{Q}$ and $H \otimes \mathbb{Q}$ respectively.*

(b) *If the closure of the images of* $\ker \sigma$ *and* $\ker \tau$ *in $\sigma^{\vdash}$ and $\tau^{\vdash}$ respectively are real vector spaces, and if ϕ is order unit good, then both σ and τ are order unit good.*

(c) *If σ and τ are order unit good and $\sigma(G) \cap \tau(H)$ is dense in $\mathbb{R}$, then ϕ is order unit good.*

We can also rephrase this as follows.

Proposition 4.6. *Let (G, u, σ) and (H, v, τ) be nearly divisible dimension groups with order units and normalized traces. Consider the following properties:*

(1) *$\sigma \oplus \tau$ is an order unit good trace on $G \oplus_s H$.*

(2) *$\sigma(G) \cap \tau(H)$ is a dense subgroup of $\mathbb{R}$.*

(3) *The closures of the images of* $\ker \sigma$ *and* $\ker \tau$ *in their respective affine spaces are real vector spaces.*

(4) *Both σ and τ are order unit good traces.*

Then the following implications hold: (1) *implies* (2); (1) *and* (3) *jointly imply* (4); (4) *implies* (3); (4) *and* (2) *jointly imply* (1).

Remark. The implications are invariant under the transformation $(x) \mapsto (5 - x)$.

Examples exist (Example 4.8) where G and H are simple dimension groups that show that if ϕ is order unit good, then neither σ nor τ (or exactly one of them) need be order unit good.

This method suggests what to do with multiple traces. Let (G_i, u_i, σ_i), where $i = 1, 2, \dots, n$, be approximately divisible dimension groups, each with order unit and (unnormalized) trace. Form $K = \bigoplus G_i$ with the strict ordering, and $\phi = \sigma_1 \oplus \sigma_2 \oplus \cdots \oplus \sigma_n : K \to \mathbb{R}$, and the map $T : K \to \mathbb{R}^n$ defined by $\phi((g_i)) = \sum \sigma_i(g_i)$ and $T((g_i)) = (\sigma_1(g_1), \sigma_2(g_2), \dots, \sigma_n(g_n))$. Identify $\operatorname{Aff} S(K, ((u_i)))$ with the cartesian product $\operatorname{Aff} S(G_1, u_1) \times \cdots \times \operatorname{Aff} S(G_n, u_n)$.

If $(g_i) \in \ker \phi$, then $\sigma_n(g_n) = -\sum_{i=1}^{n-1} \sigma_i(g_i)$, and we can interchange n with any other integer less than n. In particular, $V := \sigma_n^{-1}(\sigma_n(G_n) \cap (\sum_{i=1}^{n} \sigma_i(G_i)))$ is independent of permutations and the range of T on $\ker \phi$ is $T(V)$.

Extend T to $\mathcal{T} : \operatorname{Aff} S(K, (u_i)) \to \mathbb{R}^n$ (sending (j_i) to $(j_i(\sigma_i))$). Restricted to $\phi^{\vdash}$, the range of $\mathcal{T}$ is exactly $(1, 1, 1, \dots, 1)^{\perp}$, i.e., the entries add to zero.

Now we can form the diagram analogous to the previous one.

$$\begin{tikzcd}
0 \arrow[r] & \ker\sigma_1 \oplus \cdots \oplus \ker\sigma_n \arrow[r] \arrow[d] & \ker\phi \arrow[r, "T"] \arrow[d] & T(V) \arrow[r] \arrow[d] & 0 \\
& \overline{\widehat{\ker\sigma_1}} \times \cdots \times \overline{\widehat{\ker\sigma_n}} \arrow[r] \arrow[d] & \overline{\widehat{\ker\phi}} \arrow[r, "\mathcal{T}"] \arrow[d] & \overline{T(V)} \arrow[d] & \\
0 \arrow[r] & \sigma_1^{\vdash} \times \cdots \times \sigma_n^{\vdash} \arrow[r] & \phi^{\vdash} \arrow[r, "\mathcal{T}"] & (1, 1, \ldots, 1)^{\perp} \arrow[r] & 0
\end{tikzcd}$$

We quickly see that density of $T(V)$ (a subgroup of $\mathbb{R}^n$ contained in $(1, \ldots, 1)^{\perp}$) in $(1, \ldots, 1)^{\perp}$ is necessary for ϕ to be order unit good; that is, it is necessary in order for $\ker\phi$ to have norm dense image in $\phi^{\vdash}$.

Suppose all the σ_i are order unit good and $T(V)$ is dense in $(1, \ldots, 1)^{\perp}$. Then

$$\overline{\widehat{\ker\sigma_1}} \times \cdots \times \overline{\widehat{\ker\sigma_n}} = \sigma_1^{\vdash} \times \cdots \times \sigma_n^{\vdash}$$

is a closed subspace of $\phi^{\vdash}$, and the middle line yields that the codimension of $\overline{\widehat{\ker\phi}}$ in $\operatorname{Aff} S(K)$ is $n - (n-1) = 1$, so being a closed subspace of the codimension-one space $\phi^{\vdash}$, $\overline{\widehat{\ker\phi}}$ must equal it, and therefore ϕ is order unit good.

Suppose that ϕ is order unit good (hence we have density of $T(V)$ in $(1, \ldots, 1)^{\perp}$) and each $\overline{\widehat{\ker\sigma_i}}$ is a vector space. To show that the σ_i are all order unit good, it is sufficient to show that the $\ker\sigma_i$ have dense image in $\sigma^{\vdash}$, and it is easy to show that $\overline{\widehat{\ker\sigma_1}} \times \cdots \times \overline{\widehat{\ker\sigma_n}}$ equals $\sigma_1^{\vdash} \times \cdots \times \sigma_n^{\vdash}$ is sufficient for this.

We note that the bounded real-valued group homomorphisms on $T(V)$, and of course on its closure, are linear combinations of the coordinate functions, which correspond to the σ_i, with lack of uniqueness arising from the relation that the sum of the coefficients is zero.

By assumption, each $\overline{\widehat{\ker\sigma_i}}$ is a vector space (and closed in $\operatorname{Aff} S(G_i, u_i)$), whence the whole batch $L := \overline{\widehat{\ker\sigma_1}} \times \cdots \times \overline{\widehat{\ker\sigma_n}}$ is a closed subspace of $M := \sigma_1^{\vdash} \times \cdots \times \sigma_n^{\vdash}$ (which is itself a closed codimension-n subspace of $\operatorname{Aff} S(K)$). If they are not equal, there exists a bounded linear functional α on $\operatorname{Aff} S(K, (u_i))$ that kills L but not M. This induces a bounded real-valued group homomorphism β on $\ker\phi$ which kills $\ker\sigma_1 \oplus \cdots \oplus \ker\sigma_n$. Hence it induces a bounded real-valued group homomorphism on the quotient, $T(V)$, $B : T(V) \to \mathbb{R}$.

Each σ_i induces Σ_i on $T(V)$, and these are the coordinate functions. Hence there exist real λ_i such that $B = \sum \lambda_i \Sigma_i$. Thus $\beta - \sum \lambda_i \sigma_i$ vanishes identically on $\ker\phi$, and by density, $\alpha = \sum \lambda_i \sigma_i$ (where σ_i is now interpreted as the map $(j_i) \mapsto j_i(\sigma_i)$ on $\operatorname{Aff} S(K)$). But this obviously kills $\sigma_1^{\vdash} \times \cdots \times \sigma_n^{\vdash}$, a contradiction. Hence each σ_i is order unit good.

To summarize what happens with multiple traces, we have the following:

Theorem 4.7. *Let (G_i, u_i, σ_i) be approximately divisible dimension groups with order unit (u_i) and (unnormalized) trace (σ_i). Form $K = \bigoplus G_i$ (with the strict ordering), and the trace $\phi = \bigoplus \sigma_i$ on K. Set $J = \sigma_n(G_n) \cap \left(\sum_{i \leq n-1} \sigma_i(G_i)\right)$, a subgroup of $\mathbb{R}$.*

(a) *If ϕ is order unit good, then $T(\sigma_n^{-1}(J))$ is dense in $(1, 1, \dots, 1)^{\perp}$.*

(b) *If the closure of the image of* $\ker \sigma_i$ *in $\sigma_i^{\vdash}$ is a real vector space for all i, and if ϕ is order unit good, then all σ_i are order unit good.*

(c) *If the image of* $\ker \sigma_i$ *is dense in $\sigma_i^{\vdash}$ for all i (that is, each σ_i is order unit good), and if $T(\sigma_n^{-1}(J))$ is dense in $(1, 1, \dots, 1)^{\perp}$, then ϕ is order unit good.*

The conditions for order unit goodness with n direct summands are slightly different, in that they involve the density of a subgroup of $\mathbb{R}^{n-1}$ (identified with $(1, \dots, 1)^{\perp}$), or simply that the closure of $T(V)$ is a vector space of dimension $n-1$ (in general, the closure need not be a vector space). To some extent, this explains some of the phenomena illustrated in the examples below, with direct sums of two not yielding an order unit good trace, but direct sums of three doing so. In fact, the argument in the example, $G_n = \mathbb{Z} + (\sqrt{3} + n\sqrt{2})\mathbb{Z}$, essentially boils down to showing the closure of $T(V)$ is a two-dimensional vector space. But actually calculating with $T(V)$ seems awkward.

However, computation is feasible in special cases. Suppose $G_1 = \mathbb{Z} + \sqrt{6}\mathbb{Z}$, $G_2 = \mathbb{Z} + \sqrt{15}\mathbb{Z}$, and $G_3 = \mathbb{Z} + \sqrt{10}\mathbb{Z}$. Then $T(V)$ is discrete, so $\sigma_1 \oplus \sigma_2 \oplus \sigma_3$ is not order unit good. However, if we add a fourth term, $G_4 = \mathbb{Z} + (\sqrt{6} + \sqrt{15} + \sqrt{10})\mathbb{Z}$, then with $\phi = \bigoplus_{i \leq 4} \sigma_i$,

$$\ker \phi = \left\{ \left(a + b\sqrt{6}, c + b\sqrt{15}, d + b\sqrt{10}, \right.\right.$$
$$\left.\left. -(a + c + d) - b\left(\sqrt{6} + \sqrt{15} + \sqrt{10}\right)\right) \;\middle|\; a, b, c, d \in \mathbb{Z} \right\}.$$

Let $v_1 = (1, 0, 0, -1)$, $v_2 = (0, 1, 0, -1)$, and $v_3 = (0, 0, 1, -1)$; then $\ker \phi$ is the $\mathbb{Z}$-span of

$$\{v_1, v_2, v_3, \sqrt{6}v_1 + \sqrt{15}v_2 + \sqrt{10}v_3\}.$$

The map from $\ker \phi$ to $\mathbb{R}^3$ given by $v_i \mapsto e_i$ (standard basis elements) has range equal to the free abelian group on $\{e_1, e_2, e_3, \sqrt{6}e_1 + \sqrt{15}e_2 + \sqrt{10}e_3\}$. Since $\{1, \sqrt{6}, \sqrt{10}, \sqrt{15}\}$ is linearly independent over the rationals, this group is dense. It is trivial that $\{v_i\}$ is a real basis for $\phi^{\vdash}$, so ϕ is good. In this example, all the $\ker \sigma_i$ are trivial, so $T(V)$ is all of $\ker \phi$.

On the other hand, if we omit any one or two of the G_i, the resulting trace is not order unit good, since the resulting $T(V)$ will be discrete.

We can similarly construct (G_i, σ_i) (the G_i subgroups of the reals), $i = 1, \dots, n$, such that $\bigoplus_{i=1}^n \sigma_i$ is order unit good, but for no proper subset J of $\{1, 2, \dots, n\}$ with $|J| > 1$ is $\bigoplus_{i \in J} \sigma_i$ order unit good: Let $\{p_i\}_{i=1}^n$ be distinct primes; set

$G_i = \mathbb{Z} + \sqrt{p_i}\,\mathbb{Z}$ for $1 \leq i \leq n-1$, and $G_n = \mathbb{Z} + \big(\sum_{i=1}^{n-1} \sqrt{p_i}\big)\mathbb{Z}$. The resulting $T(V)$ will be a critical group of rank n, so all subgroups of lesser rank are discrete.

Example 4.8. There exist simple dimension groups (G, σ) and (H, τ) with traces such that $\phi = \sigma \oplus \tau$ is (order unit) good on the strict direct sum $K = G \oplus H$, but σ is not good as a trace on G (and in one case, τ is good, in another case, it is not).

Proof. For simple dimension groups (as G, H, and K are), order unit goodness is equivalent to goodness. Begin with three subgroups of the reals,

$$G_1 = \mathbb{Z} + \sqrt{3}\mathbb{Z} + \sqrt{5}\mathbb{Z},$$
$$G_2 = \mathbb{Z} + \sqrt{2}\mathbb{Z} + \sqrt{5}\mathbb{Z},$$
$$G_3 = \mathbb{Z} + (\sqrt{3} + \sqrt{2})\mathbb{Z}.$$

Let τ_i denote the respective identifications of G_i with subgroups of the reals; these are traces on each of these three totally ordered dimension groups. Each τ_i is the unique (up to scalar multiple) trace, and thus is good. Now form $L = G_1 \oplus G_2$ with the strict order; since both subgroups contain $\mathbb{Z} + \sqrt{5}\mathbb{Z}$, which is dense, it follows from the criterion above that $\tau_1 \oplus \tau_2$ is a good trace thereon. Next, form $K = L \oplus G_3$, with the strict order; since the value group of $\tau_1 \oplus \tau_2(L)$ includes $\mathbb{Z} + (\sqrt{3} + \sqrt{2})\mathbb{Z}$ and the latter is dense, we can apply the criterion again, and so deduce that $\tau_1 \oplus \tau_2 \oplus \tau_3$ is good, as a trace on K.

However, we can obtain K by proceeding in a different order. Set $G = G_1 \oplus G_3$ with the strict order. Either by direct examination or by the necessity of the density condition, $\tau_1 \oplus \tau_3$ is *not* good — note in particular, that the intersection of the value groups is just $\mathbb{Z}$. Let $H = G_2$; then the obvious permutation-order isomorphism which takes K to $G \oplus H$ takes $\tau_1 \oplus \tau_2 \oplus \tau_3$ to $\tau_1 \oplus \tau_3 \oplus \tau_2$, hence the latter is good. But with $\sigma = \tau_1 \oplus \tau_3$ and $\tau = \tau_2$, we have that σ is not good, whereas $\sigma \oplus \tau$ (and τ) is good.

To obtain an example wherein neither σ nor τ is good, let G_4 be another copy of G_3, set $G = G_1 \oplus G_3$ and $H = G_2 \oplus G_4$ (with the strict orderings of course); $\tau = \tau_2 \oplus \tau_4$ is not good for the same reason as $\sigma = \tau_1 \oplus \tau_3$, but their direct sum is good. □

5. Good sets of traces

As in [BeH 2014], a compact convex subset Y of $S(G, u)$ is *order unit good* with respect to (G, u) if given $b \in G^+ \setminus \{0\}$ (b is an order unit of G) and $a \in G$ such that $0 \ll \hat{a}|K \ll \hat{b}|K$, there exists $a' \in G$ such that $\hat{a}|K = \hat{a}'|K$ and $0 \leq a' \leq b$. When Y is a face (it need not be; e.g., for any singleton subset of $S(G, u)$, $\{\tau\}$ is good if and only if the trace τ is good as defined for individual traces), Y is order unit good if and only if $\ker Y := \bigcap_{\tau \in K} \ker \tau$ has dense range in $Y^{\perp} = \{h \in \operatorname{Aff} S(G, u) \mid h|K \equiv 0\}$. When G is simple, this was defined as good in [BeH 2014]. When $G = \operatorname{Aff} K$ (where K is a Choquet simplex), equipped with the strict ordering, goodness of subsets of K is an interesting geometric property. In Appendix B, we show that at least when K is finite-dimensional, the good subsets of K are as conjectured in [BeH 2014, Section 7, Conjecture].

There is a problem in defining what a good subset Y should be in the nonsimple case. It should be consistent with what has been defined in the simple case (where good is the same as order unit good), and the singleton case (whence came the original definition of good); additionally, it would be desirable that if $Y = S(G, u)$, then Y should be good whenever G is a dimension group such that $\operatorname{Inf} G = \{0\}$.

We give a definition of good in more complicated situations, including for a set of traces; this extends some of the definitions in [BeH 2014]. For any partially ordered abelian group H and $h \in H^+$, recall the definition of the *interval generated by* h, denoted $[0, h]$ (possibly with a subscript H if necessary to avoid ambiguity about the choice of group), to be $\{j \in H \mid 0 \le j \le h\}$. Let (G, u) be a dimension group (at this stage, we really only require that it be a partially ordered unperforated group) with order unit. Let L be a subgroup of G; we say *L is a good subgroup of G* if

(i) L is convex (that is, if $a \le c \le b$ with $a, b \in L$ and $c \in G$, then $c \in L$), and G/L is unperforated;

(ii) using the quotient map $\pi : G \to G/L$, the latter equipped with the quotient ordering, for every $b \in G^+$, we have $\pi([0, b]) = [0, \pi(b)]$.

Convexity is required in order that the quotient positive cone be proper, that is, the only positive and negative elements are zero. Unperforation is often redundant; it may always be (in the presence of (ii); see the discussion concerning refinability in [BeH 2014]). The second property says that for all $b \in G^+$, and for all $a \in G$ such that $0 \le a + L \le b + L$ (or equivalently, $(a + L) \cap G^+$ and $(b - a + L) \cap G^+$ are both nonempty), there exists $a' \in G$ such that $a - a' \in L$ and $0 \le a' \le b$. This is a strong lifting property.

For example, if τ is a trace, set $L = \ker \tau$; this is convex, and is a good subgroup of G if and only if τ is good (as a trace); in this case, G/L is naturally isomorphic to a subgroup of the reals, so unperforation is automatic.

For a subset U of $S(G, u)$ define $\ker U = \bigcap_{\sigma \in U} \ker \sigma$; for a subset W of G, define $Z(W) = \{\sigma \in S(G, u) \mid \sigma(w) = 0\}$. For good sets, we may as well assume that $Y = Z(\ker Y)$ at the outset; in other words, $\sigma \in Y$ if and only if $\sigma(\ker K) = 0$, since in any reasonable definition for good or order unit good, the candidate set will satisfy $Y = Z(\ker Y)$. As explained in [BeH 2014], these form the collection of closed sets with respect to a Zariski-like topology, and also extend the definition of facial topology (relative to G) defined on $\partial_e S(G, u)$, to $S(G, u)$. If $Y \subset S(G, u)$, set $\widetilde{Y} = Z(\ker Y) = \{\sigma \in S(G, u) \mid \sigma(\ker Y) = \{0\}\}$; this is a closure operation, corresponding to the facial topology and analogous to the Zariski topology from algebraic geometry. In many cases, we just assume $Y = \widetilde{Y}$ already, since $\ker Y = \ker \widetilde{Y}$.

We say Y *is a good subset of* $S(G,u)$ *with respect to* (G,u) if $Y=\widetilde{Y}$ and $\ker Y$ is a good subgroup of G. If $Y=\{\tau\}$, and τ is merely an order unit good trace, then $\ker\tau$ has dense image in $\tau^{\vdash}$, and this implies $Y=\widetilde{Y}$.

If L is a subgroup of G, then we may form

$$Y\equiv Z(L)=\{\sigma\in S(G,u)\mid \sigma(L)=\{0\}\}.$$

Then Y satisfies $\widetilde{Y}=Y$. However, it can happen that L is a good subgroup of G, but $Z(L)$ is not a good subset of $S(G,u)$ with respect to G.

For example, let $(H,[\chi_X])$ be the ordered Čech cohomology group of any noncyclic primitive subshift of finite type. It is known not to be a dimension group, but is unperforated and has numerous other properties (see [Boyle and Handelman 1996] as well as unpublished results of Boyle and the author). There exists a dimension group (G,u) such that $H\cong G/\operatorname{Inf} G$ with the quotient ordering. Set $Y=S(G,u)$, so that $\ker Y=\operatorname{Inf} G$. Since the quotient $H=G/\operatorname{Inf} G$ is not a dimension group, it follows from results below that property (ii) fails. However, $L=\{0\}$ is clearly a good subgroup of G, and $Z(L)=Y$, but $\ker Y=\operatorname{Inf} G$. So Y is not a good subset of $S(G,u)$.

In the definition of a good subgroup, it may be that the relatively strong condition that G/L is unperforated can be replaced by the much weaker G/L is torsion-free, in the presence of (ii), the lifting property. This is the case in the situation described in [BeH 2014, Proposition 7.6], dealing with simple dimension groups and $L=\ker Y$. There are criteria for the quotient G/L to be unperforated [BeH 2014, Lemma B1], but these are not always easy to verify.

The following is implicit in [BeH 2014, Proposition 7.6].

Lemma 5.1. *Suppose* (G,u) *is a dimension group and* L *is a convex subgroup of* G *satisfying* (ii)*. Then* G/L *with the quotient ordering is an interpolation group, and its trace space is canonically affinely homeomorphic to* $L^{\vdash}$*. The latter is a Choquet simplex.*

Proof. If $0\le a+L\le (b+L)+(c+L)$ in G/L, first lift b and c separately to positive elements of G; it doesn't hurt to relabel them b and c. Applying (ii) to $0\le a+L\le (b+c)+L$, we can find $a'\in[0,b+c]$ such that $a-a'\in L$. Hence $0\le a'\le b+c$; by interpolation in G, we may find $a_1\in[0,b]$ and $a_2\in[0,c]$ such that $a'=a_1+a_2$. Then $a+L=a'+L=(a_1+L)+(a_2+L)$ and a_i+L are positive elements of G/L, and each is less than $b+L$, $c+L$ respectively. Thus G/L satisfies Riesz decomposition. The rest is standard. □

We may consider the simplest definition possible for goodness of a set; that is, Y is *better* (a facetious, but not inapt, name) for (G,u) if (i) $Y=Z(\ker Y)$ and (ii) whenever $a\in G$, $b\in G^+$ and $0\le \hat{a}|Y\le \hat{b}|Y$, there exists $a'\in G^+$ such that $\hat{a}'|Y=\hat{a}|Y$ and $a'\le b$. This turns out to be much too restrictive (although it is an interesting property); for example, if $Y = S(G, u)$, then Y is better implies $G/\operatorname{Inf} G$ (with the quotient ordering; this need not be a dimension group) is archimedean, which hardly ever occurs; and if G is simple, this is generally stronger than order unit good. If Y is a singleton, then better agrees with the original definition of good.

Lemma 5.2. *Let (G, u) be a dimension group with order unit u. If $Y \subseteq S(G, u)$ is good with respect to (G, u), then $G/\ker Y$ is a dimension group, with trace space canonically affinely homeomorphic to Y.*

Proof. As good implies order unit good, $\ker Y$ has dense image in $Y^{\vdash}$, and thus its closure is a vector space, so that by [BeH 2014, Corollary B2], $G/\ker Y$ is unperforated. Now suppose $0 \leq a + \ker Y \leq (b + \ker Y) + (b' + \ker Y)$, where the latter two terms are nonnegative. Hence we may assume $b, b' \geq 0$, and thus $0 \leq a + \ker Y \leq (b + b') + \ker Y$ implies there exists $a' \in G^+$ such that $a' + \ker Y = a + \ker Y$ and $a' \leq b + b'$. Riesz interpolation in G yields a decomposition $a' = a_1 + a_2$, where $0 \leq a_1, a_2$ and $a_1 \leq b$ and $a_2 \leq b'$. Hence $a + \ker Y = a' + \ker Y = (a_1 + \ker Y) + (a_2 + \ker Y)$, and $a_1 + \ker Y \leq b + \ker Y$, and $a_2 + \ker Y \leq b' + \ker Y$. Thus $G/\ker Y$ satisfies interpolation.

Any trace τ of $G/\ker Y$, normalized at $u + \ker Y$, induces a trace $\tilde{\tau}$ of (G, u) by composing with the quotient map. Conversely, if σ is a trace that kills $\ker Y$, then from the definition, $\sigma \in Y$. Hence the map $S(G/\ker Y, u + \ker Y) \to S(G, u)$ is one-to-one and onto, and it is easy to see that it is an affine homeomorphism to $\operatorname{Aff} Y$. □

Lemma 5.3. *If Y is a good subset of $S(G, u)$, then (I, w) is an order ideal of G with its own order unit, and for all $\sigma \in Y$, we have $\sigma|I \not\equiv 0$, then the map*

$$I/(I \cap \ker Y) \to G/\ker Y$$

is an order isomorphism.

Proof. First we show $I/(\ker Y \cap I)$ is unperforated, by showing the image of I is an order ideal in $G/\ker Y$ (which is unperforated, by the preceding). Select $0 \leq a + \ker Y \leq b + \ker Y$, where $b \in I$; we can write $b = b_1 - b_2$, where $b_i \in I^+$, and thus $0 \leq a + \ker Y \leq b_1 + \ker Y$, and now $b_1 \in I^+$. There thus exists $a' \in [0, b_1]$ such that $a - a_1 \in \ker Y$. As $0 \leq a' \leq b_1$ and $b_1 \in I$, it follows that $a_1 \in I^+$, so that $a_1 + \ker Y$ is in the image of I; the latter is thus a convex subgroup of $G/\ker Y$. Directedness of the image is trivial, so $I/(I \cap \ker Y)$ is an order ideal in $G/\ker Y$.

Any order ideal in an unperforated partially ordered group is itself unperforated, so $I/(\ker Y \cap I)$ is unperforated.

If $\sigma \in Y$ and $\sigma(w) = 0$, then $\sigma(I) = 0$, contradicting the property of Y; hence $\hat{w}|Y \gg \delta$ for some $\delta > 0$. Since $G/\ker Y$ is unperforated and its trace space is canonically identified with Y, it follows that $w + \ker Y$ is an order unit for $G/\ker Y$. Hence the order ideal generated by $w + \ker Y$ is all of $G/\ker Y$. Hence the image of I in $G/\ker Y$ is onto.

So far, the map $I/(I \cap \ker Y) \to G/\ker Y$ is one-to-one (by construction), order-preserving (by definition), and now we know that it is onto. To show it is an order-isomorphism, it suffices to show that the image of I^+ is all of the positive cone.

Select $b \in G^+$. Then $\hat{b}|Y \ll m$ for some integer m, so there exists an integer N such that $\hat{b} \ll N\hat{w}$, and thus $0 \le b + \ker Y \le Nw + \ker Y$ (the latter by unperforation, again). By goodness, there exists $a \in [0, Nw]$ such that $a - b \in \ker Y$; $0 \le a \le Nw$ implies $a \in I^+$, and it maps to $b + \ker Y$. □

The latter property is the analogue of $\tau(I) = \tau(G)$ for a single good trace τ of G. If we weaken the hypotheses to "$\ker Y$ does not contain I", then the result is unclear. We have similar problems with the following characterization when some points of Y are not faithful.

Lemma 5.4. *Let (I, w) be an order ideal of G with its own order unit, and suppose that every point of Y does not kill I. Then the map $\phi_I : Y \to S(I, w)$, given by $\sigma \mapsto \sigma/\sigma(w)|I$, is continuous. If Y is good with respect to (G, u), then $\phi_I(Y)$ is good with respect to (I, w).*

Proof. The restriction map on traces sends every point to a nonzero trace of I, and thus the map is continuous (as Y is compact, $\inf_{\sigma \in Y} \sigma(w) > 0$). Suppose ρ is a normalized trace on (I, w) such that $\rho|(I \cap \ker \rho)$ is identically zero. Then ρ induces a trace on $I/(I \cap \ker Y)$, hence is a trace on $G/\ker Y$, and therefore ρ is the restriction of a trace from G, necessarily killing Y. If r is the lifted trace, we must have $r \in Y$, and thus $\rho \in \phi_I(Y)$. In particular, relative to (I, w), we have $\phi_I(Y) = Z(\ker \phi_I(Y))$, and it follows immediately that $\phi_I(Y)$ is good with respect to (I, w). □

The condition on Y in the next result, that every point be faithful, is rather strong, but makes things easier to deal with. The much weaker faithfulness condition ($\ker Y \cap G^+ = \{0\}$) is innocuous, as we can factor out the maximal order ideal contained in $\ker Y$.

Lemma 5.5. *Let (G, u) be a dimension group, and Y a subset of $S(G, u)$ for normalized traces σ. Then $\sigma|\ker Y \equiv 0$ if and only if $\sigma \in Y$ and $\ker Y \cap G^+ = \{0\}$.*

(a) *The trace space of the quotient $G/\ker Y$ is canonically affinely homeomorphic to Y.*

(b) *If $G/\ker Y$ is unperforated and Y satisfies the additional condition that every element of Y is faithful, then $G/\ker Y$ is simple.*

Proof. Let ϕ be a normalized trace of $(G/\ker Y, u+\ker Y)$, and let $\pi : G \to G/\ker Y$ be the quotient map. Then $\sigma' := \sigma \circ \pi$ is a normalized trace of (G, u) satisfying $\sigma(\ker Y) = 0$, so $\sigma \in Y$. Thus the map $S(G/\ker Y, u + \ker Y) \to S(G, u)$ given by $\sigma \mapsto \sigma \circ \pi$ has image in Y, and is clearly onto Y.

(a) The map is obviously one-to-one, affine, and continuous, with continuous inverse $Y \to S(G/\ker Y, u + \ker Y)$, and so is an affine homeomorphism.

(b) Suppose $a+\ker Y$ is nonzero and $a+\ker Y \geq 0+\ker Y$; there thus exists $a' \geq 0$ such that $a'-a \in \ker Y$ (from the definition of the ordering on the quotient group). If $a+\ker Y$ is not an order unit, then there exists a normalized trace σ on $G/\ker Y$ such that $\sigma(a+\ker Y)=0$ (otherwise, $\hat{a}|Y$ is strictly positive, and as $G/\ker Y$ is unperforated, this would imply $a+\ker Y$ is an order unit in $G/\ker Y$). Then $\sigma'=\sigma\circ\pi$ belongs to Y and $\sigma'(a')=0$, contradicting $\ker\sigma'\cap G^+=\{0\}$.

Hence every nonzero element of $G/\ker Y$ is an order unit. □

If in part (b), we drop the unperforated hypothesis, then we can still say something. From $a+\ker Y \geq 0+\ker Y$, we have $0 \leq \hat{a}|Y$; if for all positive integers m, we have that $ma+\ker Y$ is not an order unit in $G/\ker Y$, then there must exist a trace ϕ on $G/\ker Y$ such that $\phi(a')=0$. As in the argument above, this leads to a contradiction. So in the perforated case, we obtain that there exists $m>0$ such that $m(a+\ker Y)$ is an order unit. If we define simple to mean no proper order ideals, then the quotient group is simple. (We normally deal with unperforated order groups, wherein the lack of order ideals is equivalent to every nonzero nonnegative element being an order unit.)

The following is a variant of [BeH 2014, Lemma 7.1].

Lemma 5.6. *Let (G,u) be an approximately divisible dimension group, and let L be a convex subgroup. If G/L is unperforated, then order units lift. (That is, given a such that $a+L$ is an order unit of G/L, there exists an order unit v of G such that $a-v \in L$.)*

Proof. The traces of G/L are the traces of G that kill L, $Z := Z(L) \subset S(G,u)$. As $a+L$ is an order unit, $\hat{a}|L \gg \delta$ for some $\delta>0$. As G is approximately divisible, there exists $w \in G$ such that $\delta/3 < \hat{w} < \delta/2$. Then $(\hat{a}-\hat{w})|Z \gg \delta/2$; as G/L is unperforated, $a-w+L$ is in $(G/L)^+$. From the definition of quotient ordering, there exists $c \in G^+$ such that $c+L=a-w+L$. Set $v=c+w$. Then $v+L=a-w+w+L=a+L$; since $v \geq w$ and w is an order unit, it follows that v is an order unit. □

If we drop approximate divisibility, we obtain that for all order units $a+L$ of G/L, there exists an integer N such that for all $n \geq N$, there exist order units v_n of G such that $v_n - na \in L$. (Instead of using a small order unit w, we take u or any other order unit we can find.)

The following gives a general result (without assuming $G/\ker Y$ is unperforated, but instead, that Y is a face) about lifting order units.

Lemma 5.7. *Suppose $Y=Z(\ker Y)$ is a face of $S(G,u)$ such that the image of* $\ker Y$ *is dense in $Y^\perp$. Let $a \in G$ satisfy $a+\ker Y \geq 0$ and $\hat{a}|Y \gg \delta$ for some $\delta>0$. Then there exists $a' \in G^{++}$ such that $a'+\ker Y = a+\ker Y$.*

Proof. From the quotient ordering, there exists $c \in G^+$ such that $c - a \in \ker Y$. Let $F = \{\tau \in S(G,u) \mid \tau(c) = 0\}$; because $c \in G^+$, F is a face, and is obviously closed. Since $\hat{c}|Y = \hat{a}|Y$, we must have $F \cap Y = \varnothing$. There thus exists $h \in \operatorname{Aff} S(G,u)^+$ such that $h|Y \equiv 0$ and $h|F \equiv 1$.

As $h \in Y^{\perp}$, there exist $g_n \in \ker Y$ such that $\hat{g}_n \to h$ uniformly. Hence $\widehat{g_n + c} \to h + \hat{c}$ uniformly. The latter however is strictly positive (since $\hat{c} \geq 0$ and $\hat{c}^{-1}(0) = F$). Hence there exists n such that $\widehat{g_n + c}$ is strictly positive; as G is unperforated, $a' := g_n + c$ is an order unit of G. Its image modulo $\ker Y$ is $c + \ker Y = a + \ker Y$. □

Proposition 5.8. *Suppose that (G,u) is a nearly divisible dimension group, and $Y = Z(\ker Y)$ is a subset of $S(G,u)$ such that for all $\sigma \in Y$, $\ker \sigma \cap G^+ = \{0\}$. Suppose that either Y is a face or $G/\ker Y$ is unperforated. Then Y is good (with respect to (G,u)) if and only if*

(a) *$\phi_I(Y)$ is order unit good for all order ideals I having their own order unit,*

(b) *for every nonzero order ideal I, we have $I + \ker Y = G + \ker Y$.*

Remark. Condition (b) is just a restatement of the map $I/(I \cap \ker Y) \to G/\ker Y$ being onto. It does not require the stronger property, that it be an order isomorphism.

Proof. Sufficiency of the conditions: Suppose $b \in G^+$ and $a \in G$ and in addition, $0 \leq a + \ker Y \leq b + \ker Y$. Let $I \equiv I(b)$ be the order ideal generated by b; that is, $I(b) = \{g \in G \mid \exists N \in \mathbb{N} \text{ such that } -Nb \leq g \leq Nb\}$. By (b), there exists $a_1 \in I(b)$ such that $a_1 + \ker Y = a + \ker Y$. Since $I/(I \cap \ker Y)$ is simple, $0 \leq a_1 + \ker Y \leq b + \ker Y$ entails either $a_1 + \ker Y = 0 + \ker Y$ or $a_1 + \ker Y$ is an order unit. In the former case, set $a' = 0$.

Otherwise, if Y is a face, there exists $a_2 \in I^{++}$ such that $a_2 + \ker Y = a_1 + \ker Y$. Similarly, either $b + \ker Y = a_1 + \ker Y$ (in which case, we take $a' = b$) or the difference $b + \ker Y - (a_2 + \ker Y)$ is an order unit in $I/(\ker Y \cap I)$.

If $G/\ker Y$ is unperforated, then $I/(I \cap \ker Y)$ is unperforated (this follows from I being an order ideal in G), and applying Lemma 5.6(b) to $\phi_I(Y)$, is simple with trace space canonically $\phi_I(Y)$. This means that the order-preserving one-to-one and onto map $I/(I \cap \ker Y) \to G/\ker Y$ induces an affine homeomorphism on their respective trace spaces; since the images in their affine function representations are the same, that of $I/(I \cap \ker Y)$ has dense range, and being simple (and $\phi_I(Y)$ being a simplex), the latter is a simple dimension group. A one-to-one order-preserving group homomorphism between simple dimension groups which induces an affine homeomorphism on the trace spaces is necessarily an order isomorphism.

Thus in either case, we have $0 \ll \hat{a}|Y \ll \hat{b}|Y$; now order unit goodness of $(I(b), b)$ yields $a' \in I^+$ such that $a' \leq b$.

Necessity of the conditions follows from the preceding results. □

Now we briefly examine examples in R_P. When R is a partially ordered commutative unperforated ring with 1 as an order unit, every closed face of $S(R, 1)$ is uniquely determined by its extreme points and these form a compact subset of $X = \partial_e S(R, 1)$ (and conversely, every closed subset of X yields a closed face in this way). Thus, as a preliminary question, we can ask when the closed face obtained from the closed subset Y of X is good (for R) or order unit good. We say Y *generates an (order unit) good face* when this occurs.

It is easy to see that Y generates an order unit good face for R if and only if for all pure traces $\sigma \notin Y$, we have $\sigma | \ker Y$ is not identically zero (we define $\ker Y = \bigcap_{\tau \in Y} \ker \tau$, as usual).

To verify this, if Y generates an order unit good face for R, then $\ker Y$ has dense image in $\operatorname{Ann} Y := \{f \in C(X, \mathbb{R}) \mid f|Y \equiv 0\}$. There exists $f \in \operatorname{Ann} Y$ such that $f(\sigma) \neq 0$, and there exist $a_n \in \ker Y$ such that $\hat{a}_n \to f$ uniformly, so there exists $a \in \{a_n\}$ such that $0 \neq \hat{a}(\sigma) = \sigma(a)$, hence $\sigma | \ker Y$ is not identically zero.

Conversely, suppose $\sigma(\ker Y) \neq \{0\}$ for every $\sigma \in X \setminus Y$. It is trivial that $\ker Y$ is an ideal of R (not generally an order ideal), so its closure in $C(X, \mathbb{R})$ is a closed ideal thereof, hence of the form $\operatorname{Ann} Z$ for some closed $Z \subset X$. Obviously $Y \subset X$, but if $\sigma \in Z \setminus Y$, there exists $a \in \ker Y$ such that $\sigma(a) \neq 0$, so that $\hat{a} \notin \operatorname{Ann} Z$, a contradiction. Hence $Z = Y$, so $\ker Y$ has dense image in $\operatorname{Ann} Z$, and thus Y is order unit good for R.

Lemma 5.9. *Let R be a partially ordered unperforated approximately divisible commutative ring, and let Y be a compact subset of the set of faithful pure traces. Let (I, v) be a nonzero order ideal with its own order unit.*

(a) *The set Y maps by normalized restriction to a compact set of pure faithful traces on (I, v), Y_I.*

(b) *If the closed face generated by Y is order unit good for R, then the closed face of $S(I, w)$ generated by Y_I is order unit good for (I, v).*

Goodness for R_P (several variables) of sets corresponding to faces (that is, closed subsets of the pure trace space) is dependent on the coefficients. For example, as we will see below, if V is the variety given by $f = (x-3)^2 + (y-3)^2 - 1$, the circle of radius one centred at (3, 3) and $P = c_0 + c_1 x + c_2 y$, then V (or its corresponding face in $S(R_P, 1)$ is order unit good, but not good, no matter what the choice of (positive) integers c_0, c_1, c_2. On the other hand, if $P_1 = P \cdot Q$ where $Q = c + xf + yg + xyh$, where f is a polynomial in x with no negative coefficients such that $(x-3)^2 + 8$ divides some power of $c + xf$ (such exist!), g is a polynomial in y such that $(y-3)^2 + 8$ divides some power of $c + yg$, and h is a polynomial in xy^{-1} such that $(1 + X^2)$ divides some power of $h(X)$, then V is a good subset for R_{P_1} (the conditions on the coefficients of monomials appearing in the faces of the Newton polytope described by the divisibility condition are necessary and sufficient for Proposition 5.8(b) to apply; however they are also extremely complicated).

Now we specialize to $R = R_P$ or $R_P \otimes \mathbb{Q}$, and to avoid severe complications, also assume that the compact Y consists of pure faithful traces (that is, Y is a compact subset of the positive orthant, $(\mathbb{R}^d)^{++}$, after identifying the pure faithful traces with points of the orthant). Then $\ker Y = \{f/P^k \in R_P \mid f|Y \equiv 0\}$. Recall that for $f \in \mathbb{Z}[x_1, \ldots, x_d]$, $f/P^k \in R_P$ means that there exists l such that $\operatorname{Log} f P^l \subseteq \operatorname{Log} P^{k+l}$; we can well assume $\operatorname{Log} f \subseteq \operatorname{Log} P^k$.

Hence Y is order unit good for R if and only if whenever σ is a pure trace not in Y, we have $\sigma|\ker Y \neq 0$. If we restrict σ to the faithful pure traces, then we deduce a necessary condition: if $Y \subset (\mathbb{R}^d)^{++}$ is compact, then Y is order unit good for R_P implies

$$ZI(Y) \cap (\mathbb{R}^d)^{++}) = Y.$$

That is, intersecting the Zariski closure of Y with $(\mathbb{R}^d)^{++}$ gives no new points. In the singleton case, we have seen that this condition, real isolation, is sufficient. However, for general compact Y, it is no longer sufficient.

In fact, examples to illustrate this are ubiquitous (when $d > 1$). The very simplest one I could think of is the following. Let $P = 1 + xy + x$ (the coefficients, here all ones, are not terribly important); then $\operatorname{Log} P$ is the triangle with vertices $\{(0, 0), (1, 1), (1, 0)\}$, and as rings $R_P \cong \mathbb{Z}[X, W]$ (the pure polynomial ring in two variables) via the transformation $X = x/P$ and $W = xy/P$. Let $f = (x-3)^2 + (y-3)^2 - 1$, so $Z(f) \cap \mathbb{R}^2$ is the circle of radius one centred at $(3, 3)$, and we set Y to be this circle, sitting inside the positive quadrant of $\mathbb{R}^2$. In particular, $\operatorname{Log} f = \{(0, 0), (1, 0), (2, 0), (0, 1), (0, 2)\}$. It is trivial that $ZI(Y) \cap (\mathbb{R}^2)^{++} = Y$. However, there exists $\sigma \in \partial_e S(R_P, 1) \setminus Y$ such that $\sigma|\ker Y = 0$.

Explicitly, σ is the pure trace corresponding to the extreme point of $\operatorname{cvx} \operatorname{Log} P$ given by $(0, 0)$; $\sigma(g/P^k) = (g, x^{0,0})/(P, x^{0,0})^k$. Suppose $a = h/P^k \in R_P$; we may assume $\operatorname{Log} h \subseteq \operatorname{Log} P^k$. If $r \in Y$ implies $h(r)/P^k(r) = 0$, that is, $\tau(a) = 0$ for all $\tau \in Y$, then $h|Y \equiv 0$ (since Y is in the positive quadrant, $P|Y$ vanishes nowhere). Hence there exists $e \in \mathbb{Q}[x, y]$ such that $h = e \cdot f$ (as $I_{\mathbb{Q}}(f) = f\mathbb{Q}[x, y]$); multiplying by a positive integer N, we may assume $Nh = e \cdot f$, where $e \in \mathbb{Z}[x, y]$.

We claim that this forces $h(0, 0) = 0$, that is, its constant term must be zero, from which it would follow that $\sigma(a) = 0$, showing that $\ker Y \subset \ker \sigma$, as desired. If $h(0, 0) \neq 0$, then as $\operatorname{Log} h \subseteq \operatorname{Log} P^k$, we would have to have $(0, 0) \in \operatorname{Log} h$, and in particular, this point is an extreme point of $\operatorname{cvx} \operatorname{Log} h$. Since $(0, 0)$ is also an extreme point of $\operatorname{cvx} \operatorname{Log} f$, it easily follows that $(0, 0)$ is an extreme point of $\operatorname{cvx} \operatorname{Log} e$ (we are working with Laurent polynomials as opposed to ordinary polynomials, hence this complicated argument). Now consider the coefficients of e and of f restricted to the line $x = 0$ (that is, throw away all the monomials with a power of x), e_0 and $f_0 = y^2 - 6y + 17$. The product is not zero, and cannot be a single monomial (since f_0 is not), hence there must be, in addition to the constant term, a term of the form y^j in the product. It is easy to check that this forces $(0, j) \in \operatorname{Log} e \cdot f = \operatorname{Log} h$. However, $\operatorname{Log} P^k$ is contained in the lattice cone generated by $\{(0,0), (1,1), (1,0)\}$, which does not contain $(0, j)$. This contradicts $\operatorname{Log} h \subseteq \operatorname{Log} P^k$.

This example does not depend on the coefficients in P, that is, we could just as well have taken $P = 2 + 3xy + 5x$ (which guarantees that R_P is approximately divisible), or whether we take R_P or $R_P \otimes \mathbb{Q}$.

In contrast, if we take the same f, but $P = 2 + 3x + 5y$ (or with any other positive coefficients), then $f/P^2 \in R_P$ and for all nonfaithful pure σ, we have $\sigma(f/P^2) > 0$; hence the same Y is now order unit good for R_P. This is part of a more general criterion.

Let h be a polynomial in d variables, and let S be a finite set of lattice points in $\mathbb{Z}^d$, and $K(S) = \operatorname{cvx} S$. Suppose F is a proper face of $K(S)$, and $\operatorname{Log} h \subseteq kS$ (the set of sums of k elements of S). We define $h_{F,k}$, the *facial polynomial of h relative to F and k,* by throwing away all the terms in x^w of h for which $w \notin kF$. In the case $S = \operatorname{Log} P$, we can form the element $h_{F,k}/(P_F)^k \in R_{P_F}$ (in fewer variables, the number being the dimension of F). This yields a positive homomorphism $R_P \to R_{P_F}$ as described in [Handelman 1985].

Let Y satisfy $ZI(Y) \cap (\mathbb{R}^d)^{++} = Y$, and form the ideal $I(Y)$ of $\mathbb{Z}[x_1, \dots, x_d]$. Let P be a projectively faithful polynomial in $\mathbb{Z}[x_1, \dots, x_d]$. We say that Y *can be fitted with respect to P* if there exists a polynomial $h \in I(Y)$ such that

(a) $\operatorname{Log} h \subseteq \operatorname{Log} P^k$ for some k,

(b) for every proper face F of $\operatorname{cvx} \operatorname{Log} P$, $h_{F,k}$ has no negative coefficients.

This depends on $\operatorname{Log} P$, but not so much on the coefficients P [Handelman 1987, Proposition II.5].

Condition (b) can be somewhat weakened, since we are permitted to multiply the numerator and denominator of h/P^k by powers of P, and apply eventual positivity criteria, e.g., [Handelman 1986]. The condition is equivalent to "for all pure σ that are not faithful, there exists $h \in I(Y)$ such that $\sigma(h/P_k) > 0$". For example, with $\operatorname{Log} P = \{(0,0), (0,1), (1,0)\}$ and Y the circle in $(\mathbb{R}^2)^{++}$ of radius 1 centred at $(3,3)$, Y is fitted with respect to P. Just observe that f has the three facial polynomials (corresponding to the three edges of $\operatorname{cvx} \operatorname{Log} P$ (the extreme points take care of themselves, so we need not worry about the zero-dimensional faces), $(x-3)^2 + 17$, $(y-3)^2 + 17$, $x^2 + y^2$. If we multiply the first two by a sufficiently high power, say N, of $1+x$ (respectively $(1+y)$), the outcome will have no negative coefficients. It follows that if $h = P^N f$, then h will be positively fitted with respect to P, with $k = N + 2$.

Now the following is practically tautological.

Proposition 5.10. *Let P be a faithfully projective element of $\mathbb{Z}[x_i]$, and Y a compact subset of $((\mathbb{R})^d)^{++}$. Then Y generates an order unit good face for R_P (and simultaneously for $R_P \otimes \mathbb{Q}$) if and only if*

(i) *$ZI(Y) \cap (\mathbb{R}^d)^{++} = Y$ and*

(ii) *Y can be fitted with respect to P.*

Conditions on Y to guarantee property (b) of Lemma 5.9 seem to be very difficult, involving divisibility of polynomials (and so depend on the coefficients). So goodness of subsets of $\partial_e S(R_P, 1)$ is still problematic.

Appendix A: Connections with zero-dimensional topological dynamics

The referee has observed that this paper uses methods almost entirely from partially ordered abelian groups and Choquet theory, and its results refer to the former. As a result, the connections with dynamics are invisible. This informal appendix is intended to outline some of the connections. We assume that the reader has some knowledge of Cantor dynamical systems.

Let (X, T) be a nonatomic zero-dimensional compact separable Hausdorff space (a *Cantor set*) together with a self-homeomorphism; we call this pair a *Cantor minimal system.* We may functorially attach a partially preordered abelian group, $K_0(X, T)$, to (X, T), in any of several equivalent ways, e.g., the preordered Grothendieck group of the crossed-product C*-algebra $C(X) \times_T \mathbb{Z}$, or directly by computing the preordered Čech cohomology, $C(X, \mathbb{Z})/(I - T)C(X, \mathbb{Z})$ (where T has its natural action on $C(X, \mathbb{Z})$), with the quotient preordering.

When T is minimal (an abbreviation for *the action of T on X is minimal*), not only is the preordering a genuine partial ordering, but $K_0(X, T)$ is a simple dimension group; moreover (in the minimal case), together with a distinguished order unit, it is a complete invariant for strong orbit equivalence, and a complete invariant for orbit equivalence is obtained from the simple dimension group $K_0(X, T)/\operatorname{Inf}(K_0(X, T))$ [Giordano et al. 1995; Herman et al. 1992].

When T is no longer minimal, chain recurrence (a weak condition) will guarantee that $K_0(X, T)$ is partially ordered [Boyle and Handelman 1996]. Unfortunately, even for rather natural systems, such as shifts of finite type, $K_0(X, T)$ need not be a dimension group [Kim et al. 2001], and more recent work suggests that being a dimension group is a relatively rare phenomenon — moreover, not all (countable) dimension groups can appear as a $K_0(X, T)$. This calls into question the usefulness of the results here, since we work almost exclusively with dimension groups.

Fortunately, many questions in nonminimal cases can be reduced to questions concerning dimension groups. This is because of the following result of Boyle and the author.

Theorem. *If T is chain recurrent, then there exists a dimension group G together with an onto, order-preserving group homomorphism $\phi : G \to K_0(X, T)$ such that*

(a) $\ker \phi \subseteq \operatorname{Inf} G$, *and*

(b) $\phi(G^+) = K_0(X, T)^+$.

Not only does this say that $K_0(X, T)$ is order-isomorphic to $G/\ker \phi$ with the quotient ordering on the latter, but it also implies that ϕ induces a natural affine homeomorphism between $S(G, u)$, the normalized trace space of G, and the normalized trace space of $K_0(X, T)$, which itself is just the space of invariant probability measures on X. The images of G and of $K_0(X, T)$ in their affine representation agree, and this means that some properties of measures/traces transfer between (X, T) and G (a dimension group). For example, order unit goodness is the same whether we take $K_0(X, T)$ or G.

This allows one to transfer problems about traces (or finite invariant measures) on $K_0(X, T)$ to the dimension group G. As a simple example (already a known consequence of Krieger's marker lemma), if T has no periodic points, then the image of $K_0(X, T)$ in its affine representation (taking as order unit $u = [\chi_X]$) is dense. In particular, almost divisibility transfers (near divisibility probably does not). More relevantly, order unit goodness and its refinable counterpart transfer completely between the two ordered groups, as does the purity criterion of [Goodearl and Handelman 1980]. Properties involving order ideals do not do so well, but very often there is a one-way implication. There is obviously more to be done.

Tensor products. Tensor products of dimension groups, or more generally, of partially ordered abelian groups, as discussed in Section 2, arise naturally. However, their translation to dynamical systems is not so clear. Nonetheless, there are examples — every minimal Cantor system can be realized as a continuous adic map on a Bratteli diagram, and the Cartesian product of the two Cantor sets admits an adic map compatible with the tensor product [BeH 2014, Appendix A].

There is a less tenuous interrelation. Let (X, T) and (Y, S) be Cantor dynamical systems (not necessarily minimal, although not much is known in the nonminimal case), which are at least chain recurrent. Form the product, $(X \times Y, T \times S)$ (meaning the $\mathbb{Z}$-action, not the $\mathbb{Z}^2$-action). There is a natural order-preserving group homomorphism, $\Phi : K_0(X \times Y, T \times S) \to K_0(X, T) \otimes K_0(Y, S)$ (with the usual positive cone on the tensor product). This is induced by the isomorphism $C(X \times Y, \mathbb{Z}) \cong C(X, \mathbb{Z}) \otimes C(Y, \mathbb{Z})$, the latter factoring onto $K_0(X, T) \otimes K_0(Y, S)$;

then $(\mathrm{I} - (T \times S))(C(X \times Y, \mathbb{Z})) \subseteq (\mathrm{I} - T)C(X \times Y, \mathbb{Z}) + (\mathrm{I} - S)(C(X \times Y, \mathbb{Z}))$ shows that Φ is well-defined.

Call the system constructed in [BeH 2014, Appendix A] realizing the tensor product, $(X \times Y, R)$. For the following observation, we don't really need the construction of R, merely that such a minimal system (realizing the tensor product of the dimension groups) exists, which is a consequence of [Giordano et al. 1995].

Observation. Suppose (X, T) and (Y, S) are minimal Cantor dynamical systems. In the following, any of (b), (c), or (d) implies that $T \times S$ is minimal. Moreover, (b), (c), and (d) are equivalent, and each implies (a); finally, if each of (X, T) and (Y, S) has only finitely many ergodic measures, then (a) implies (b).

(a) $T \times S$ is orbit equivalent to R.

(b) The kernel of the natural order-preserving homomorphism

$$\Phi : K_0(X \times Y, T \times S) \to K_0(X, T) \otimes K_0(Y, S)$$

consists of infinitesimals.

(c) Every invariant $(T \times S)$-ergodic measure is of the form $\mu \times \nu$, where μ is an invariant measure on (X, T) and ν is an invariant measure on (Y, S).

(d) For every continuous $f : X \to \mathbb{Z}$ of X and every coboundary $h = (\mathrm{I} - S)g$ of Y (where $g : Y \to \mathbb{Z}$ is continuous), $[f \cdot h]$ is an infinitesimal in $K_0(X \times Y, T \times S)$.

Remark. The apparent asymmetry in X and Y of (d) is illusory, as

$$(\mathrm{I} - T \times S)(f \cdot g) = f \cdot h + ((\mathrm{I} - T) f) \cdot (g \circ S).$$

Remark. Without the assumption that (X, T) and (Y, S) have only finitely many ergodic measures, it is still very likely true that (a) implies (b) anyway.

Proof. We show that (c) implies minimality of the product, and then that (b), (c), and (d) are equivalent in general. If $T \times S$ were not minimal, there would be a proper closed invariant subset $A \subset X \times Y$. Then any invariant ergodic measure supported on A cannot be a product measure, and of course, one exists.

(b) implies (c): Since the kernel of the map to the tensor product is onto and has only infinitesimals in its kernel, it induces a homeomorphism on the normalized trace spaces, and clearly product traces map to product traces under this; since all pure traces on the tensor product are of the form $\sigma \otimes \tau$, where σ and τ are pure traces on the two components respectively, it implies every pure trace on $K_0(X \times Y, T \times S)$ is a product trace, and this translates to product measure.

(c) implies (d): For any product measure on $X \times Y$, $d\rho = d\mu \, d\nu$, we have

$$\int_{X \times Y} f \cdot h \, d\rho = \left(\int_X f \, d\mu \right) \cdot \left(\int_Y h \, d\nu \right) = 0.$$

From (c), the closed convex hull of the product invariant measures is the set of all invariant measures, so $\int_{X\times Y} f\cdot h\, d\zeta = 0$ for every invariant measure ζ on $X \times Y$. Hence $[f\cdot h]$ vanishes at every trace on $K_0(X\times Y, T\times S)$, and thus is an infinitesimal.

(d) implies (b): The kernel of the map $K_0(X \times Y, T \times S) \to K_0(X, T) \otimes K_0(X, S)$ is spanned by the images (in K_0) of

$$(\mathrm{I}-T\times S)C(X\times Y, \mathbb{Z}), \quad C(X, \mathbb{Z})\cdot(\mathrm{I}-S)(C(Y, \mathbb{Z})), \quad (\mathrm{I}-T)(C(X, \mathbb{Z}))\cdot C(Y, \mathbb{Z}).$$

By the remark, every element of $(\mathrm{I} - T)(C(X, \mathbb{Z})) \cdot C(Y, \mathbb{Z})$ belongs to the abelian group generated by the other two, whose images obviously land in the infinitesimal subgroup.

(b) implies (a): Factoring out the infinitesimals from both groups yields a unital order isomorphism (since the original map sends the positive elements onto the positive elements), so $T \times S$ is orbit equivalent to R.

(a) implies (b) (if each of (X, T) and (Y, T) has only finitely many ergodic measures): Let m and n be the respective number of pure traces (ergodic measures). The number of pure traces on the tensor product is exactly mn, and orbit equivalence implies the same number of pure traces for $K_0(X \times Y, T \times S)$. The natural map to the tensor product induces a positive map between the corresponding affine function spaces (of the same dimension, mn), which must therefore be an isomorphism. Hence every trace on $K_0(X \times Y, T \times S)$ factors through the tensor product. Thus the kernel of the map must be contained in the kernel of all the traces, and hence is contained in the infinitesimal subgroup. □

Bernoulli measures and xerox actions. The results of Section 3 are reminiscent of those of [Akin et al. 2008], characterizing goodness of Bernoulli measures. Let $\{e_i\}$ be the standard basis for $\mathbb{Z}^n$, and set $x_i = x^{e_i}$ (in monomial notation), and $P = 1 + \sum_{i=1}^{n} x_i$ (for notational convenience, we sometimes write $1 = x^{e_0}$, where $e_0 = 0$). Form R_P, a very special case of the ordered rings discussed in this section. This particular one is ring isomorphic to the pure polynomial ring, $A_P = \mathbb{Z}[X_1, \ldots, X_n]$, under the assignment $X_i = x_i/P$, and the positive cone is generated multiplicatively and additively by $\{X_1, \ldots, X_n; 1 - \sum X_i\}$.

The pure traces of A_P are precisely the multiplicative ones, determined by $X_i \mapsto p_i$, where $0 \leq p_i$ and $\sum p_i \leq 1$. Let $p = (p_1, \ldots, p_n)$ denote the corresponding point in the standard simplex in $\mathbb{R}^n$. If $p_0 : 1 - \sum p_i > 0$ (which occurs at least when the corresponding pure trace is faithful, and in other cases as well), we can reconstruct the point in $(\mathbb{R}^n)^+$ (the faithful pure trace of R_P whence it came; explicitly, $x_i \mapsto p_i/p_0$), and when the measure is faithful (meaning all of p_i and p_0 are nonzero), these run over the entire open orthant, $(\mathbb{R}^n)^{++}$.

For all choices of (faithfully projective) P, there is a natural map from the pure trace space of $(R_P, 1)$ to the Newton polytope of P, given by the weighted moment map; because this particular choice for P is so pleasant, the weighted moment map is particularly explicit. (For generic P, R_P is not even a unique factorization domain — for example, if P is irreducible over the integers, faithfully projective, and R_P satisfies unique factorization, then up to the natural action of $\mathrm{AGL}(n, \mathbb{Z})$ on the exponents, $P = \sum_{i=0}^{n} a_i x_i$, where $x_0 = 1$ and a_i are positive integers.)

Now back to our specific choice of P; this R_P is not approximately divisible (for example, $X_i \mapsto 0$ yields a $\mathbb{Z}$-valued trace on A_P; we could fix this if we permitted nonmonic coefficients at all of the vertices). Nonetheless, we can analyze conditions (1) and (2) here.

For the point $r := (p_i/p_0) \in \mathbb{R}^n)^{++}$, τ_r will satisfy (1) and (2) precisely if r is really isolated and $p_0/p_i \in \mathbb{Z}[p_j/p_0] = \tau_r(R_P)$; in particular, p_0/p_i are units in $\tau_r(R_P)$ and are algebraic. Since $\sum_1^n p_i = 1 - p_0$, we quickly deduce that all of $p_0, p_1, \ldots, p_n$ must be algebraic and are units in $\mathbb{Z}[p_0, p_1, \ldots, p_n]$ (the image of A_P under the corresponding trace).

The density matrices that implement the pure traces on the fixed point C*-algebras are exactly those whose diagonals are p. In a sense, C*-algebra traces are the noncommutative analogue of Bernoulli measures.

Strict direct sums. There is no obvious connection between strict direct sums (even for simple dimension groups) and dynamical systems. In fact, although we know that a strict direct sum of dimension groups is again a dimension group, given realizations of the two components (as direct limits, that is, given Bratteli diagrams for each of them), there is no way known to construct the strict direct sum as such a direct limit from the two direct limits (that is, there does not seem to be a way of finding a Bratteli diagram based on the two given ones). There are a few (very few) ad hoc constructions in very special cases.

Appendix B: Order unit good traces on $\mathbb{Z}^k$

The criteria for goodness of traces on nearly divisible dimension groups depend on order unit goodness; and the usefulness of the former is a consequence of the relatively simple characterization of order unit good traces on approximately divisible dimension groups, namely density of the image of $\ker\tau$ in $\tau^{\vdash}$ via the affine representation of (G, u).

To obtain useful criteria for goodness on a larger class of dimension groups, it would be helpful to find an analogous characterization of order unit goodness in the presence of discrete traces. In this appendix, we consider the extreme dimension groups with discrete traces, namely the simplicial ones, $\mathbb{Z}^k$, with the usual ordering. It is already known that up to scalar multiple, the only good traces are given by left multiplication by a vector whose entries consist only of zeros and ones [Handelman 2013b, Lemma 6.2].

With the current definition of order unit good (really intended for approximately divisible groups), the order unit good traces on $\mathbb{Z}^k$ can be characterized, but the characterization makes it difficult to see how to obtain goodness criteria for more general dimension groups, as we did in the nearly divisible case.

Let $v \in (\mathbb{R}^{k\times 1})^+ \setminus \{\mathbf{0}\}$; then v induces a trace on $\mathbb{Z}^k$, via left multiplication, $\phi_v : \mathbb{Z}^k \to \mathbb{R}$ sending $w \mapsto vw$ (we think of $\mathbb{Z}^k$ as a set of columns, so matrix multiplication makes sense). Obviously we can replace v by any positive real multiple of itself without changing properties such as goodness or order unit goodness. In addition, we may apply any permutation to the entries, with the same lack of bad consequences. We may also discard any zeros (reducing the size of the vectors, that is, decreasing k).

Suppose v has only integer entries; then we may order the nonzero entries, so that

$$v = (n(1), n(2), \dots, n(r); 0, 0, \dots, 0), \quad \text{where } n(1) \le n(2) \le \cdots .$$

We may also assume that $\gcd\{n(i)\} = 1$.

Lemma B.1. *With this choice of v, we have that ϕ_v is order unit good if and only if $n(1) = 1$ and for all $r \ge j > 1$, we have $n(j) \le 1 + \sum_{i<j} n(i)$.*

Proof. Assume v is in the form indicated, and ϕ_v is order unit good. Since $\gcd\{n(i)\} = 1$, there exists a vector w such that $vw = 1$. Set $u = (1, 1, 1, \dots, 1)$; we have that u is an order unit, hence it is ϕ_v-order unit good. Since $vu > 1$ (unless $v = (1, 0, 0, \dots, 0)$ which is trivially good), there must exist $w_0 \in (\mathbb{Z}^d)^+$ such that $vw_0 = vw = 1 < vu$. Since the nonzero entries of v are increasing, this forces the smallest one, $n(1)$, to be 1. Hence $n(1) = 1$.

Since $vu = \sum n(i) := N$, and there exists $w \in \mathbb{Z}^k$ such that $vw = 1$, for each s with $1 < s < N$, there exists $w_s \in \{0, 1\}^k$ (as $0 \le w_0 \le u$) such that $vw_s = s$, by order unit goodness of u. Now suppose that for some j, we have $n(j) > 1 + \sum_{i<j} n(i)$. Then $n(j) - 1$ cannot be realized as a sum of $n(i)$s (using at most one for each choice of i), since $n(j) - 1 > \sum_{i<j} n(i)$, and $n(j) \le n(j')$ for all $j' > j$ (if there are any such j'). Hence no such w_0 can exist.

Thus, if u is ϕ_v-order unit good, then the constraint on growth must hold.

Conversely, suppose the inequalities hold. It is then an easy induction argument (on r, augmenting the vector by adjoining $n(k+1)$) to show that u is τ_v-order unit good, by realizing every integer in the interval $(0, N)$. Finally, to show that every order unit is ϕ_v-order unit good (u was the smallest choice), it suffices to show that if we add a single one to a ϕ_v-order unit good vector, the outcome is again ϕ_v-order unit good. □

In particular, the choices for v, $(1, 2, 4, 8, 16)$ and $(1, 1, 1, 4)$ yield order unit good traces, but $(1, 3)$ and $(1, 1, 1, 5)$ do not. This rather complicated set of conditions, when applied to order ideals in dimension groups that have a simplicial quotient by an order ideal, likely makes order unit goodness unusable for the purposes we had in mind.

Lemma B.2. *If* $\phi_v : \mathbb{Z}^k \to \mathbb{R}$ *is an order unit good trace*, *then up to scalar multiple*, $v \in (\mathbb{Z}^{k\times 1})^+$.

Proof. In $\mathbb{Z}^k$, all intervals of the form $[0, u]$ (where u is an order unit) are finite sets. If there were an irrational ratio among the nonzero entries of v, we would obtain $\phi_v(\mathbb{Z}^k) \cap [0, N]$ is infinite for any positive integer N. If order unit goodness held, this would be impossible. Hence all the ratios are rational, and it easily follows that after suitable scalar multiplication, we can convert v to an integer row. □

Proposition B.3. *Let* v *be an element of* $(\mathbb{R}^k)^+ \setminus \{\mathbf{\Gamma}\}$. *Then* ϕ_v *is an order unit good trace if and only if up to scalar multiple and after rearrangement so that* $v = (n(1), \dots, n(r); 0, 0, \dots)$ *with* $n(i-1) \le n(i)$, *we have* $n(i) \in \mathbb{N}$, $n(1) = 1$, *and for all* $1 < j \le r$,

$$n(j) \le 1 + \sum_{i<j} n(i).$$

Appendix C: Good simplices

In the finite-dimensional case, we verify a conjecture from [BeH 2014, Section 7] that good subsets of Choquet simplices are obtained as coproducts of faces with singleton subsets of disjoint faces.

Let K be a Choquet simplex. A nonempty subset J of K is said to be *good* (following [BeH 2014]) if it satisfies the following (redundant set of) properties:

(i) J is a (compact) Choquet simplex.

(ii) There exists a closed flat $\mathcal{L}$ such that $J = \mathcal{L} \cap K$.

(iii) If $a \in \mathrm{Aff}(J)^{++}$ and $b \in \mathrm{Aff}(K)^{++}$ are such that $a \ll b|J$, then there exists $a' \in \mathrm{Aff}(K)^{++}$ such that $a'|J = a$ and $a' \ll b$.

We denote this relationship between J and K by $J \Subset K$ (there is a G inside the inclusion sign). If F is a closed face of K, we denote it $F \triangleleft K$. A question arising out of [BeH 2014] is to characterize good subsets of Choquet simplices. For example, closed faces are good, and singleton sets are also good, and coproducts (within the category of simplices and good subsets) preserve these properties. A conjecture was made concerning the structure of good subsets; we verify this in the case that K is finite-dimensional.

Now (ii) is redundant, and only the compact convex part of (i) is necessary. This is based on the following simple construction.

If X is a subset of a real vector space, define the *affine span of* X, denoted Aspan X, as the set of finite sums $\left\{\sum r_i x_i \;\middle|\; r_i \in \mathbb{R},\ \sum r_i = 1,\ x_i \in X\right\}$.

If J is a singleton or a line segment, there is (almost) nothing to do. Define $\mathcal{L}_0 = \text{Aspan } J$. If there exists $v \in (K \cap \mathcal{L}_0) \setminus J$, we can write $v = \sum \alpha_i v_i - \beta_j w_j$, where $v_i, w_j \in J$, and $\alpha_i, \beta_j > 0$, and $\sum \alpha_i - \sum \beta_j = 1$. We can also arrange that $\text{cvx}\{v_i\} \cap \text{cvx}\{w_j\} = \varnothing$. Hence for any positive $\eta < 1$, there exists $a \in (\text{Aff } J)^{++}$ such that $1 - \eta < a|\,\text{cvx}\{w_j\} < 1$ and $a|\,\text{cvx}\{a_i\} < \eta$. Since a is continuous, it is bounded above, so (iii) applies with some constant $b \in \text{Aff } K$.

Hence there exists $a' \in (\text{Aff } K)^{++}$ such that $a = a'|J$. Evaluating the equation at a', we obtain

$$0 < a'(w) = \sum \alpha_i a(v_i) - \sum \beta_j a(w_j) < \eta \sum \alpha_i - (1 - \eta) \sum \beta_j.$$

This entails $\eta\big(\sum \alpha_i + \sum \beta_j\big) > \sum \beta_j$. Now $\sum \beta_j > 0$, since otherwise $v \in J$. Hence we can choose at the outset positive $\eta < \sum \beta_j \;/\; \big(\sum \alpha_i + \sum \beta_j\big)$, which yields a contradiction.

Thus $\mathcal{L}_0 \cap K = J$. If $x_n \in \mathcal{L}_0$ and $x_n \to x \in K$, but $x \notin J$, there exists a line segment joining x to an element of the relative interior of J; it must pass through at least two points in J, hence $x \in \mathcal{L}_0$. In other words, with $\mathcal{L}$ equalling the closure of $\mathcal{L}_0$, we have $J = \mathcal{L}_0 \cap K = \mathcal{L} \cap K$.

To check that the compact convex set J must be a simplex if (iii) is satisfied, observe that the quotient $\text{Aff } K / J^{\vdash}$ (with the strict ordering on $\text{Aff } K$, $J^{\vdash} = \{a \in \text{Aff } K \mid a|J \equiv 0\}$, and the quotient ordering) is order isomorphic to $\text{Aff } J$ (with the strict ordering). But goodness implies [BeH 2014] that it satisfies Riesz interpolation, which of course forces J to be a Choquet simplex.

Let K' and K'' be simplices (simplices mean Choquet simplices; but most of the time we will working in finite dimensions, so simplex means the usual simplex) sitting inside some common simplex K which in turn is contained in some topological vector space. Suppose that $\text{Aspan } K' \cap \text{Aspan } K'' = \varnothing$; we write this as $K' \wedge K'' = \varnothing$. Then the closure of $\text{cvx}(K', K'')$ is itself a simplex, and we refer to this as the coproduct, written $K' \dot{\vee} K''$ (this is more an internal coproduct, but we shall not distinguish internal from external). If K' and K'' are faces of K, sufficient for $K' \wedge K'' = \varnothing$ is that their intersection be empty (since K is a simplex); in this case, we say that K' and K'' are *disjoint.* If $\{K^i\}$ is a finite family of subsimplices, then disjointness of the set is defined inductively in the obvious way, so that $\dot{\bigvee}_i K^i$ makes sense and is a simplex.

We record elementary properties related to goodness.

Lemma C.1. (a) *Suppose* $J \Subset K$ *and* $K \Subset L$; *then* $J \Subset L$.

(b) *If* $F \lhd K$, *then* $F \Subset K$.

(c) *If* $J \Subset K$ *and* $F \lhd K$, *then* $J \cap F \lhd J$ *and* $J \cap F \Subset K$ *whenever* $J \cap F \neq \varnothing$.

(d) *If* $J_i \Subset K_i$ *for* $i = 1, 2$ *and* $K_1 \wedge K_2 = \varnothing$, *then* $J_1 \dot{\vee} J_2 \Subset K_1 \dot{\vee} K_2$.

The crucial result is the following. Its proof rests heavily on finite-dimensionality, but is a minor modification of the previous argument.

Lemma C.2. *Let K be a finite dimensional simplex, and suppose $J \Subset K$. Let J_1 and J_2 be disjoint faces of J. Set F_i $(i = 1, 2)$ to be the smallest face of K that contains J_i. Then F_1 and F_2 are disjoint.*

Proof. It suffices to show that $F_1 \cap F_2 = \varnothing$. If not, the intersection is a face, and hence contains a vertex (that is, extreme point) of K; call it v. We may suppose that $v \notin J_2$ (since $J_1 \cap J_2 = \varnothing$). Since J is itself a finite-dimensional simplex and J_i are disjoint faces, for any $\eta > 0$ (which we will specify later), we may find $a \in \mathrm{Aff}(J)^{++}$ such that $a|J_2 \ll 1 - \epsilon$, $a|J_1 \ll \eta$, and $a \ll 1$ (on all of J). Set b to be the constant function $\mathbf{1}$ on all of K, so that $0 \ll a \ll b|J$.

By goodness, there exists $a' \in \mathrm{Aff}(K)^{++}$ such that $a' \ll b$ and $a'|J = a$. It is now easy to show that for suitably small η (depending on the boundary measure of elements of $J_i \subset F_i$), this leads to a contradiction.

Since $v \notin J$ and F_2 is the smallest face containing J_2, there must exist $w \in J_2$ such that $w = \lambda v + \sum_s \lambda_s v_s$, where $v_s \in \partial_e F_2$, $\lambda > 0$, $\lambda_s \geq 0$ and $\lambda = 1 - \sum \lambda_s$. Evaluating at a', we obtain

$$\lambda a'(v) = a(w) - \sum \lambda_s a'(v_s) \geq 1 - \eta - (1 - \lambda)$$

(since $a'(v_s) \leq b(v_s) = 1$). Thus $a'(v) \geq 1 - \eta/\lambda$.

Now working within F_1, again since F_1 is the smallest face containing J_1, there must exist $y \in J_1$ such that $y = \mu v + \sum_t \mu_t y_t$, where $\{v, y_t\} \subseteq \partial_e F_1$, $\mu > 0$, $\mu_s \geq 0$, and $\mu = 1 - \sum \mu_s$. Applying a', we obtain $\mu a'(v) = a(y) - \sum \mu_t a'(y_t) < \eta$. Hence $a'(v) < \eta/\mu$.

Thus the two inequalities force $\eta/\mu + \eta/\lambda > 1$. We reach a contradiction if we choose $\eta < 1/(1/\mu + 1/\lambda)$. □

One obstruction (among several) to extending this to infinite-dimensional simplices is the fact that the representing measures of relative interior points might vanish on the intersection of the faces. We would also have to restrict to closed faces in this case (since otherwise it is not clear that the smallest face exists), and this presents problems.

Let $\{F_i\}$ be a disjoint collection of faces — that is, for all i, $F_i \wedge \left(\dot{\bigvee}_{j \neq i} F_j\right) = \varnothing$ — of the simplex K, and for each i, let v_i be a point in the relative interior of F_i; we also assume that the F_i are not themselves singletons. We may form $J_0 := \mathrm{cvx}\{v_i\}$ and $F_0 := \mathrm{cvx}\{F_i\}$; of course, this is the coproduct of $(\{v_i\}, F_i)$, and J_0 is thus a good subset of F_0 (since each $\{v_i\} \Subset F_i$). As in [BeH 2014], we call the (v_i, F_i), together with (F, F) (that is, the face $F \Subset F$) *building blocks.* It was conjectured (in the finite-dimensional case) that if $J \Subset K$, then there exists a face F of K, together with a disjoint face F_0 obtained as the coproduct, such that $J = F \dot{\vee} J_0$; in other words, that coproducts of the building blocks yield all good subsets; alternatively, that there is a maximal face F of K sitting inside J, and J is obtained by taking coproducts with respective singleton sets sitting inside pairwise disjoint faces. This now follows easily.

Corollary C.3. *Suppose K is a finite-dimensional simplex and $J \Subset K$. Then there exist a (possibly empty) face F of K together with a finite set of faces F_i of dimension at least one such that $\{F, F_1, \dots\}$ is disjoint, together with v_i in the relative interior of F_i such that $J = \mathrm{cvx}\{F, v_i\}$.*

Proof. We proceed by induction on the dimension of J. Let F be the convex hull of all the vertices of K that lie in J; these are automatically vertices of J. If this exhausts the vertices of J, then $F = J$ and F is a face (since K is a finite-dimensional simplex), and there is nothing to do. Of course, F can be empty.

Otherwise, there exists a vertex v_1 of J that is not in $\partial_e K$; necessarily this belongs to a proper face (it cannot be in the interior, in fact by property (ii), but this can also be proved using only (i) and (iii)) of K, and let F_1 be the smallest face of K containing v_1. Then v_1 is in the relative interior of F_1. Let J^1 be the complementary face to $\{v_1\}$ in J (that is, the convex hull of all the other vertices of J).

If J^1 is empty, then $J = J^1$ is already a singleton, and we are done.

If J^1 is not empty, then $J^1 \lhd J$, so $J^1 \Subset J$, and thus by transitivity, $J_1 \Subset K$. We can apply the previous lemma. Let F^1 be the smallest face of K containing J^1; then $F^1 \cap F_1 = \varnothing$, and thus J decomposes as the coproduct of J^1 and $\{v_1\}$ (using faces F^1 and F_1), so by induction on the dimension of J, and we are done. □

Acknowledgment

Discussions with my colleague Damien Roy concerning the material in Section 3 were very helpful.

References

[Akin 1999] E. Akin, "Measures on Cantor space", *Topology Proc.* **24**:Summer (1999), 1–34. MR 2002j:28002 Zbl 1025.28007

[Akin 2005] E. Akin, "Good measures on Cantor space", *Trans. Amer. Math. Soc.* **357**:7 (2005), 2681–2722. MR 2006e:37003 Zbl 1078.37004

[Akin et al. 2008] E. Akin, R. Dougherty, R. D. Mauldin, and A. Yingst, "Which Bernoulli measures are good measures?", *Colloq. Math.* **110**:2 (2008), 243–291. MR 2009f:37009 Zbl 1151.37011

[Bezuglyi and Handelman 2014] S. Bezuglyi and D. E. Handelman, "Measures on Cantor sets: the good, the ugly, the bad", *Trans. Amer. Math. Soc.* **366**:12 (2014), 6247–6311. MR 3267010 Zbl 06371207

[Boyle and Handelman 1996] M. Boyle and D. E. Handelman, "Orbit equivalence, flow equivalence and ordered cohomology", *Israel J. Math.* **95** (1996), 169–210. MR 98a:46082 Zbl 0871.58071

[Effros et al. 1980] E. G. Effros, D. E. Handelman, and C.-L. Shen, “Dimension groups and their affine representations”, *Amer. J. Math.* **102**:2 (1980), 385–407. MR 83g:46061 Zbl 0457.46047

[Frick and Ormes 2013] S. B. Frick and N. Ormes, “Dimension groups for polynomial odometers”, *Acta Appl. Math.* **126** (2013), 165–186. MR 3077947 Zbl 06210545

[Giordano et al. 1995] T. Giordano, I. F. Putnam, and C. F. Skau, “Topological orbit equivalence and C^*-crossed products”, *J. Reine Angew. Math.* **469** (1995), 51–111. MR 97g:46085 Zbl 0834.46053

[Giordano et al. $\geq$ 2016] T. Giordano, D. E. Handelman, and M. Hosseini, “Orbit equivalence of Cantor minimal systems and their continuous spectra”, In preparation.

[Goodearl 1986] K. R. Goodearl, *Partially ordered abelian groups with interpolation*, Mathematical Surveys and Monographs **20**, American Mathematical Society, Providence, RI, 1986. MR 88f:06013 Zbl 0589.06008

[Goodearl and Handelman 1980] K. R. Goodearl and D. E. Handelman, “Metric completions of partially ordered abelian groups”, *Indiana Univ. Math. J.* **29**:6 (1980), 861–895. MR 82b:06020 Zbl 0455.06012

[Goodearl and Handelman 1986] K. R. Goodearl and D. E. Handelman, “Tensor products of dimension groups and K_0 of unit-regular rings”, *Canad. J. Math.* **38**:3 (1986), 633–658. MR 87i:16043 Zbl 0608.16027

[Handelman 1982] D. E. Handelman, “Free rank $n+1$ dense subgroups of $\mathbb{R}^n$ and their endomorphisms”, *J. Funct. Anal.* **46**:1 (1982), 1–27. MR 84h:06019 Zbl 0489.20044

[Handelman 1985] D. E. Handelman, *Positive polynomials and product type actions of compact groups*, Mem. Amer. Math. Soc. **54**:320, American Mathematical Society, Providence, RI, 1985. MR 86h:46091 Zbl 0571.46045

[Handelman 1986] D. E. Handelman, “Deciding eventual positivity of polynomials”, *Ergodic Theory Dynam. Systems* **6**:1 (1986), 57–79. MR 87m:26013 Zbl 0606.22005

[Handelman 1987] D. E. Handelman, *Positive polynomials, convex integral polytopes, and a random walk problem*, Lecture Notes in Mathematics **1282**, Springer, Berlin, 1987. MR 89b:06014 Zbl 0654.46059

[Handelman 2009] D. E. Handelman, “Matrices of positive polynomials”, *Electron. J. Linear Algebra* **19** (2009), 2–89. MR 2011g:46105 Zbl 1195.37010

[Handelman 2013a] D. E. Handelman, “Invariants for critical dimension groups and permutation-Hermite equivalence”, preprint, 2013. arXiv 1309.7417

[Handelman 2013b] D. E. Handelman, “Realizing dimension groups, good measures, and Toeplitz factors”, *Illinois J. Math.* **57**:4 (2013), 1057–1109. MR 3285868 Zbl 1311.46060

[Handelman 2014] D. E. Handelman, “Limits of simple dimension groups and weakly initial objects”, *Indiana Univ. Math. J.* **63**:5 (2014), 1567–1600. MR 3283562 Zbl 06388780

[Herman et al. 1992] R. H. Herman, I. F. Putnam, and C. F. Skau, “Ordered Bratteli diagrams, dimension groups and topological dynamics”, *Internat. J. Math.* **3**:6 (1992), 827–864. MR 94f:46096 Zbl 0786.46053

[Kim et al. 2001] K. H. Kim, F. W. Roush, and S. G. Williams, “Duality and its consequences for ordered cohomology of finite type subshifts”, pp. 243–265 in *Combinatorial & computational mathematics* (Pohang, 2000), edited by S. Hong et al., World Scientific, River Edge, NJ, 2001. MR 2003a:37016 Zbl 1002.37006

[Medynets 2006] K. Medynets, “Cantor aperiodic systems and Bratteli diagrams”, *C. R. Math. Acad. Sci. Paris* **342**:1 (2006), 43–46. MR 2006g:37011 Zbl 1081.37004

[Petersen 2012] K. Petersen, "An adic dynamical system related to the Delannoy numbers", *Ergodic Theory Dynam. Systems* **32**:2 (2012), 809–823. MR 2901372 Zbl 1257.37006

[Putnam 1989] I. F. Putnam, "The C^*-algebras associated with minimal homeomorphisms of the Cantor set", *Pacific J. Math.* **136**:2 (1989), 329–353. MR 90a:46184 Zbl 0631.46068

[Renault 1980] J. Renault, *A groupoid approach to C^*-algebras*, Lecture Notes in Mathematics **793**, Springer, Berlin, 1980. MR 82h:46075 Zbl 0433.46049

Received September 4, 2014. Revised August 2, 2015.

DAVID HANDELMAN
MATHEMATICS DEPARTMENT
UNIVERSITY OF OTTAWA
585 KING EDWARD
OTTAWA K1N 6N5 ON
CANADA
dehsg@uottawa.ca

PACIFIC JOURNAL OF MATHEMATICS
Vol. 281, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.281.421

ON FOURIER COEFFICIENTS OF CERTAIN RESIDUAL REPRESENTATIONS OF SYMPLECTIC GROUPS

DIHUA JIANG AND BAIYING LIU

In the theory of automorphic descents developed by Ginzburg, Rallis, and Soudry in *The descent map from automorphic representations of* $\mathrm{GL}(n)$ *to classical groups* (World Scientific, 2011), the structure of Fourier coefficients of the residual representations of certain special Eisenstein series plays an essential role. In a series of papers starting with *Pacific J. Math.* 264:1 (2013), 83–123, we have looked for more general residual representations, which may yield a more general theory of automorphic descents. We continue this program here, investigating the structure of Fourier coefficients of certain residual representations of symplectic groups, associated with certain interesting families of global Arthur parameters. The results partially confirm a conjecture proposed by Jiang in *Contemp. Math.* 614 (2014), 179–242 on relations between the global Arthur parameters and the structure of Fourier coefficients of the automorphic representations in the associated global Arthur packets. The results of this paper can also be regarded as a first step towards more general automorphic descents for symplectic groups, which will be considered in our future work.

1. Introduction

Let Sp_{2n} be the symplectic group with symplectic form

$$\begin{pmatrix} 0 & v_n \\ -v_n & 0 \end{pmatrix},$$

where v_n is an $n \times n$ matrix with 1s on the second diagonal and 0s elsewhere. Fix a Borel subgroup $B = TU$ of Sp_{2n}, where the maximal torus T consists of elements of the form

$$\mathrm{diag}(t_1, \ldots, t_n; t_n^{-1}, \ldots, t_1^{-1})$$

The research of Jiang is supported in part by the NSF Grants DMS-1301567, and the research of Liu is supported in part by NSF Grants DMS-1302122, and in part by a postdoc research fund from Department of Mathematics, University of Utah.
MSC2010: primary 11F70, 22E55; secondary 11F30.
Keywords: Arthur parameters, Fourier coefficients, unipotent orbits, automorphic forms.

and the unipotent radical U consists of all upper unipotent matrices in Sp_{2n}. Let F be a number field and $\mathbb{A}$ be the ring of adeles of F.

The structure of Fourier coefficients for the residual representations of $\mathrm{Sp}_{4n}(\mathbb{A})$, with cuspidal support (GL_{2n}, τ), played an indispensable role in the theory of automorphic descent from GL_{2n} to the metaplectic double cover of Sp_{2n} by Ginzburg, Rallis, and Soudry in [Ginzburg et al. 2011]. As tested in a special case in our recent work joint with Xu and Zhang in [Jiang et al. 2015], we expected the residual representations investigated in [Jiang et al. 2013] may play important roles in extending the theory of automorphic descent in [Ginzburg et al. 2011] to a more general setting. In this paper, we take certain interesting families of residual representations of $\mathrm{Sp}_{2n}(\mathbb{A})$ obtained in [Jiang et al. 2013] and study the structure of their Fourier coefficients associated to nilpotent orbits as described in [Jiang 2014]. On one hand, the results of this paper partially confirm a conjecture proposed by the first named author in [loc. cit.] on relations between the global Arthur parameters and the structure of Fourier coefficients of the automorphic representations in the corresponding global Arthur packets. On the other hand, these results are preliminary steps towards the theory of more general automorphic descents for symplectic groups, which will be considered in our future work.

We first recall the global Arthur parameters for Sp_{2n} and the discrete spectrum, and the conjecture made in [loc. cit.]. Then we recall what has been proved about this conjecture before this current paper, in particular the results obtained in [Jiang and Liu 2015a]. Finally we describe more explicitly the objective of this paper. The main results will be precisely stated in Section 2.

1A. ***Arthur parameters and the discrete spectrum.*** Let F be a number field and $\mathbb{A}$ be the ring of adeles of F. Recall that the dual group of $G_n = \mathrm{Sp}_{2n}$ is $\mathrm{SO}_{2n+1}(\mathbb{C})$. The set of global Arthur parameters for the discrete spectrum of the space of all square-integrable automorphic functions on $\mathrm{Sp}_{2n}(\mathbb{A})$ is denoted by $\widetilde{\Psi}_2(\mathrm{Sp}_{2n})$, following the notation in [Arthur 2013]. The elements of $\widetilde{\Psi}_2(\mathrm{Sp}_{2n})$ are of the form

$$\psi := \psi_1 \boxplus \psi_2 \boxplus \cdots \boxplus \psi_r, \tag{1-1}$$

where ψ_i are pairwise distinct simple global Arthur parameters of orthogonal type. A simple global Arthur parameter is formally given by (τ, b) with an integer $b \geq 1$, and with $\tau \in \mathcal{A}_{\mathrm{cusp}}(a)$ being an irreducible unitary cuspidal automorphic representation of $\mathrm{GL}_a(\mathbb{A})$.

In (1-1), one has that $\psi_i = (\tau_i, b_i)$ with $\tau_i \in \mathcal{A}_{\mathrm{cusp}}(a_i)$, $2n+1 = \sum_{i=1}^r a_i b_i$, and $\prod_i \omega_{\tau_i}^{b_i} = 1$ (the condition on the central character of the parameter), following [Arthur 2013, Section 1.4]. In order for all the ψ_i to be of orthogonal type, the simple parameters $\psi_i = (\tau_i, b_i)$ for $i = 1, 2, \ldots, r$ satisfy the following parity condition: if τ_i is of symplectic type (i.e., $L(s, \tau_i, \bigwedge^2)$ has a pole at $s = 1$), then

b_i is even; and if τ_i is of orthogonal type (i.e., $L(s, \tau_i, \mathrm{Sym}^2)$ has a pole at $s = 1$), then b_i is odd. A global Arthur parameter $\psi = \boxplus_{i=1}^r (\tau_i, b_i)$ is called generic if $b_i = 1$ for all $1 \le i \le r$.

Theorem 1.1 [Arthur 2013, Theorem 1.5.2]. *For each global Arthur parameter $\psi \in \widetilde{\Psi}_2(\mathrm{Sp}_{2n})$, there exists a global Arthur packet $\widetilde{\Pi}_\psi$. The discrete spectrum of $\mathrm{Sp}_{2n}(\mathbb{A})$ has the following decomposition*

$$L^2_{\mathrm{disc}}(\mathrm{Sp}_{2n}(F) \setminus \mathrm{Sp}_{2n}(\mathbb{A})) \cong \bigoplus_{\substack{\psi \in \widetilde{\Psi}_2(\mathrm{Sp}_{2n}) \\ \pi \in \widetilde{\Pi}_\psi(\epsilon_\psi)}} \pi,$$

where $\widetilde{\Pi}_\psi(\epsilon_\psi)$ denotes the subset of $\widetilde{\Pi}_\psi$ consisting of members which occur in the discrete spectrum of $\mathrm{Sp}_{2n}(\mathbb{A})$.

1B. ***A conjecture on the Fourier coefficients.*** We will use the notation in [Jiang and Liu 2015c; 2015a] freely. Following [Jiang and Liu 2015c, Section 2], for a symplectic partition $\underline{p}$ of $2n$, or equivalently each F-stable unipotent orbit $\mathcal{O}_{\underline{p}}$, via the standard $\mathfrak{sl}_2(F)$-triple, one may construct an F-unipotent subgroup $V_{\underline{p},2}$. In this case, the F-rational unipotent orbits in the F-stable unipotent orbit $\mathcal{O}_{\underline{p}}$ are parametrized by a datum $\underline{a}$ (see [loc. cit.] for details), which defines a character $\psi_{\underline{p},\underline{a}}$ of $V_{\underline{p},2}(\mathbb{A})$. This character $\psi_{\underline{p},\underline{a}}$ is automorphic in the sense that it is trivial on $V_{\underline{p},2}(F)$. The $\psi_{\underline{p},\underline{a}}$-Fourier coefficient of an automorphic form φ on $\mathrm{Sp}_{2n}(\mathbb{A})$ is defined by

$$\varphi^{\psi_{\underline{p},\underline{a}}}(g) := \int_{V_{\underline{p},2}(F)\setminus V_{\underline{p},2}(\mathbb{A})} \varphi(vg)\, \psi_{\underline{p},\underline{a}}(v)^{-1}\, dv. \tag{1-2}$$

We say that an irreducible automorphic representation π of $\mathrm{Sp}_{2n}(\mathbb{A})$ has a nonzero $\psi_{\underline{p},\underline{a}}$-Fourier coefficient or a nonzero Fourier coefficient attached to a (symplectic) partition $\underline{p}$ if there exists an automorphic form φ in the space of π with a nonzero $\psi_{\underline{p},\underline{a}}$-Fourier coefficient $\varphi^{\psi_{\underline{p},\underline{a}}}(g)$, for some choice of $\underline{a}$. For any irreducible automorphic representation π of $\mathrm{Sp}_{2n}(\mathbb{A})$, as in [Jiang 2014], we define $\mathfrak{p}^m(\pi)$ (which corresponds to $\mathfrak{n}^m(\pi)$ in the notation of [loc. cit.]) to be the set of all symplectic partitions $\underline{p}$ with the properties that π has a nonzero $\psi_{\underline{p},\underline{a}}$-Fourier coefficient for some choice of $\underline{a}$, but for any $\underline{p}' > \underline{p}$ (with the natural ordering of partitions), π has no nonzero Fourier coefficients attached to $\underline{p}'$. It is generally believed (and may be called a conjecture) that the set $\mathfrak{p}^m(\pi)$ contains only one partition for any irreducible automorphic representation π (or locally for any irreducible admissible representation π). We refer to [Jiang and Liu 2015b, Section 3], in particular Conjecture 3.1, for more detailed discussions on this issue.

As in [Jiang 2014], $\widetilde{\Pi}_\psi(\epsilon_\psi)$ is called the automorphic L^2-packet attached to the global Arthur parameter ψ. For each ψ of the form in (1-1), let $\underline{p}(\psi) = [(b_1)^{(a_1)} \cdots (b_r)^{(a_r)}]$ be a partition of $2n+1$ attached to the global Arthur parameter

ψ, following the discussion in [op. cit., Section 4]. For $\pi \in \widetilde{\Pi}_{\psi}(\epsilon_{\psi})$, the structure of the global Arthur parameter ψ deduces constraints on the structure of $\mathfrak{p}^m(\pi)$, which are given by the following conjecture.

Conjecture 1.2 [Jiang 2014, Conjecture 4.2]. For any $\psi \in \widetilde{\Psi}_2(\mathrm{Sp}_{2n})$, let $\widetilde{\Pi}_{\psi}(\epsilon_{\psi})$ be the automorphic L^2-packet attached to ψ. Then the following hold.

(1) Any symplectic partition $\underline{p}$ of $2n$ satisfying $\underline{p} > \eta_{\mathfrak{g}^\vee,\mathfrak{g}}(\underline{p}(\psi))$ does not belong to $\mathfrak{p}^m(\pi)$ for any $\pi \in \widetilde{\Pi}_{\psi}(\epsilon_{\psi})$.

(2) For every $\pi \in \widetilde{\Pi}_{\psi}(\epsilon_{\psi})$, every partition $\underline{p} \in \mathfrak{p}^m(\pi)$ has the property that $\underline{p} \leq \eta_{\mathfrak{g}^\vee,\mathfrak{g}}(\underline{p}(\psi))$.

(3) There exists at least one member $\pi \in \widetilde{\Pi}_{\psi}(\epsilon_{\psi})$ having the property that $\eta_{\mathfrak{g}^\vee,\mathfrak{g}}(\underline{p}(\psi)) \in \mathfrak{p}^m(\pi)$.

Here $\eta_{\mathfrak{g}^\vee,\mathfrak{g}}$ denotes the Barbasch–Vogan duality map (see Definition 2.2) from the partitions for $\mathfrak{so}_{2n+1}(\mathbb{C})$ to the partitions for $\mathfrak{sp}_{2n}(\mathbb{C})$.

We remark that part (2) is stronger than part (1) in Conjecture 1.2. More related discussions can be found in [Jiang and Liu 2015b].

There has been progress toward the proof of Conjecture 1.2. When the global Arthur parameter $\psi = \boxplus_{i=1}^r(\tau_i, 1)$ is generic, in Conjecture 1.2, part (1) is trivial, part (2) is automatic, and part (3) of Conjecture 1.2 can be viewed as the global version of the Shahidi conjecture, namely, any global tempered L-packet has a generic member. This can be proved following the theory of automorphic descent developed by Ginzburg, Rallis, and Soudry [Ginzburg et al. 2011] and the endoscopy classification of Arthur [2013]. We refer to [Jiang and Liu 2015b, Section 3.1], in particular Theorem 3.3, for more precise discussion on this issue. Hence Conjecture 1.2 holds for all generic global Arthur parameters, and those π satisfying part (3) are generic cuspidal representations.

For Arthur parameters of form $\psi = (\tau, b) \boxplus (1_{\mathrm{GL}_1(\mathbb{A})}, 1)$, where τ is an irreducible cuspidal representation of $\mathrm{GL}_{2k}(\mathbb{A})$ and is of symplectic type, and b is even, one has that $\underline{p}(\psi) = [b^{(2k)}1]$. In this case, part (3) of Conjecture 1.2 has been proved by Liu in [2013a], where it is also shown that $\mathfrak{p}^m(\pi)$ contains only one partition in this particular case.

For a general global Arthur parameter ψ, part (1) of Conjecture 1.2 is completely proved in [Jiang and Liu 2015a]. We remark that if we assume that $\mathfrak{p}^m(\pi)$ contains only one partition, then part (2) of Conjecture 1.2 essentially follows from parts (1) and (3) of Conjecture 1.2 plus certain local constraints at unramified local places as discussed in [loc. cit.]. We omit the details here. However, without knowing that the set $\mathfrak{p}^m(\pi)$ contains only one partition, part (2) of Conjecture 1.2 is also settled in [loc. cit.] partially; namely, any symplectic partition $\underline{p}$ of $2n$, for which $\underline{p} > \eta_{\mathfrak{g}^\vee,\mathfrak{g}}(\underline{p}(\psi))$ under the lexicographical ordering, does not belong to $\mathfrak{p}^m(\pi)$

for any $\pi \in \widetilde{\Pi}_\psi(\epsilon_\psi)$. We refer to [Jiang 2014, Section 4] and also [Jiang and Liu 2015b] for more discussion on this conjecture and related topics.

1C. ***The objective of this paper.*** In this section, we begin to investigate part (3) of Conjecture 1.2. This means that we have to construct or determine a particular member in a given automorphic L^2-packet $\widetilde{\Pi}_\psi(\epsilon_\psi)$ attached to a general global Arthur parameter ψ, whose Fourier coefficients achieve the partition $\eta_{\mathfrak{g}^\vee,\mathfrak{g}}(\underline{p}(\psi))$. Such members should be the distinguished members in $\widetilde{\Pi}_\psi(\epsilon_\psi)$, following the Whittaker normalization in the sense of Arthur [2013] for global generic Arthur parameters. For general nongeneric global Arthur parameters, the distinguished members in $\widetilde{\Pi}_\psi(\epsilon_\psi)$ can be certain residual representations determined by ψ as conjectured by Mœglin [2008; 2011], or certain cuspidal automorphic representations, which may be explicitly constructed through the framework of endoscopy correspondences as outlined in [Jiang 2014]. Due to the different nature of the two construction methods, we are going to treat them separately, in order to prove part (3) of Conjecture 1.2.

As explained in [Jiang and Liu 2015b], when the distinguished members π in a given $\widetilde{\Pi}_\psi(\epsilon_\psi)$ are residual representations, they can be constructed explicitly from the given cuspidal data. In this case, our method is to establish the nonvanishing of the Fourier coefficients of those π associated to the partition $\eta_{\mathfrak{g}^\vee,\mathfrak{g}}(\underline{p}(\psi))$, in terms of the nonvanishing condition (Fourier coefficients or periods) on the construction data that is also defined by the given nongeneric global Arthur parameter ψ. Hence, such a method can be regarded as a natural extension of the well-known Langlands–Shahidi method from generic Eisenstein series [Shahidi 2010] to nongeneric Eisenstein series, and in particular to the singularity of Eisenstein series, i.e., the residues of Eisenstein series. On the other hand, this method can also be regarded as an extension of the automorphic descent method of Ginzburg–Rallis–Soudry for particular residual representations [Ginzburg et al. 2011] to general residual representations.

In this paper, we are going to test our method for these nongeneric global Arthur parameters ψ, whose automorphic L^2-packets $\widetilde{\Pi}_\psi(\epsilon_\psi)$ contain the residual representations that are completely determined in our previous work joint with Zhang [Jiang et al. 2013]. Those nongeneric global Arthur parameters of $\mathrm{Sp}_{2n}(\mathbb{A})$ are of the following form

$$\psi = (\tau_1, b_1) \boxplus \boxplus_{i=2}^{r} (\tau_i, 1), \quad \text{with } b_1 > 1,$$

which has three cases, depending on the symmetry of τ_1 and the relationship between τ_1 and τ_i for $i = 2, 3, \ldots, r$. In each case, $b \geq 1$.

Case I: $\psi = (\tau, 2b+1) \boxplus \boxplus_{i=2}^{r} (\tau_i, 1)$, where $\tau \not\cong \tau_i$ for any $2 \leq i \leq r$.

Case II: $\psi = (\tau, 2b+1) \boxplus (\tau, 1) \boxplus \boxplus_{i=3}^{r} (\tau_i, 1)$, where $\tau \ncong \tau_i$ for any $3 \leq i \leq r$.

Case III: $\psi = (\tau, 2b) \boxplus \boxplus_{i=2}^{r} (\tau_i, 1)$.

For $\psi \in \widetilde{\Psi}_2(\mathrm{Sp}_{2n})$, $\tau \in \mathcal{A}_{\mathrm{cusp}}(\mathrm{GL}_a)$ is of orthogonal type in Case I and Case II, and of symplectic type in Case III. Of course, the remaining τ_i are of orthogonal type in all three cases.

When τ is of orthogonal type, i.e., in both Case I and Case II, the corresponding residual representations given in [Jiang et al. 2013] must be nonzero. In this paper, we prove part (3) of Conjecture 1.2 in those two cases, and refer to Section 2 for more details.

When τ is of symplectic type and $r \geq 2$, the relation between τ and τ_i, for $i = 2, 3, \ldots, r$, is governed by the corresponding Gan–Gross–Prasad conjecture [Gan et al. 2012], which controls the structure of the automorphic L^2-packet $\widetilde{\Pi}_{\psi}(\epsilon_{\psi})$. We prove part (3) of Conjecture 1.2 for Case III when $\widetilde{\Pi}_{\psi}(\epsilon_{\psi})$ contains residual representations. While the automorphic L^2-packet $\widetilde{\Pi}_{\psi}(\epsilon_{\psi})$ does not contain any residual representation, the situation is more involved, and will be left for a separate treatment in our future work. We discuss with more details in Section 2.

We will state the main results more explicitly in Section 2. After recalling a technical lemma from [Jiang and Liu 2015b] in Section 3, we are ready to treat Case I in both Sections 4 and 5. Case II is treated in Section 6. The final section is devoted to Case III. One may find more detailed description of the arguments and methods used in the proof of those cases in each relevant section.

2. The main results

After introducing more notation and basic facts about the discrete spectrum and Fourier coefficients attached to partitions, we will state the main results explicitly for each case.

Throughout the paper, we let $P_r^{2n} = M_r^{2n} N_r^{2n}$ (with $1 \leq r \leq n$) be the standard parabolic subgroup of Sp_{2n} with Levi part M_r^{2n} isomorphic to $\mathrm{GL}_r \times \mathrm{Sp}_{2n-2r}$ and unipotent radical N_r^{2n}. Also let $\tilde{P}_r^{2n}(\mathbb{A}) = \widetilde{M}_r^{2n}(\mathbb{A}) N_r^{2n}(\mathbb{A})$ be the preimage of $P_r^{2n}(\mathbb{A})$ in $\widetilde{\mathrm{Sp}}_{2n}(\mathbb{A})$ (the superscript $2n$ may be dropped when there is no confusion). The description of the three cases was briefly given in [Jiang and Liu 2015b]. Here are the details.

2A. ***Case I.*** $\psi \in \widetilde{\Psi}_2(\mathrm{Sp}_{2n})$ is written as

$$\psi = (\tau, 2b+1) \boxplus \boxplus_{i=2}^{r} (\tau_i, 1), \tag{2-1}$$

where $b \geq 1$ and $\tau \ncong \tau_i$ for any $2 \leq i \leq r$. Assume $\tau \in \mathcal{A}_{\mathrm{cusp}}(\mathrm{GL}_a)$ has central character ω_τ, and $\tau_i \in \mathcal{A}_{\mathrm{cusp}}(\mathrm{GL}_{a_i})$ has central character ω_{τ_i} for $2 \leq i \leq r$. Following the definition of $\widetilde{\Psi}_2(\mathrm{Sp}_{2n})$, one must have that $2n+1 = a(2b+1)+\sum_{i=2}^r a_i$, and $\omega_\tau^{2b+1}\cdot\prod_{i=2}^r \omega_{\tau_i} = 1$. Consider the isobaric representation $\pi = \tau \boxplus \tau_2 \boxplus \cdots \boxplus \tau_r$ of $\mathrm{GL}_{2m+1}(\mathbb{A})$, where $2m+1 = a+\sum_{i=2}^r a_i = 2n+1-2ab$. It follows that π has central character $\omega_\pi = \omega_\tau \cdot \prod_{i=2}^r \omega_{\tau_i} = 1$ and $a \le 2m+1 = 2n+1-2ab$.

By [Ginzburg et al. 2011, Theorem 3.1], π descends to an irreducible generic cuspidal representation σ of $\mathrm{Sp}_{2n-2ab}(\mathbb{A})$, which has the functorial transfer back to π. As remarked before, this is part (3) of Conjecture 1.2 for the generic global Arthur parameter

$$\psi_\pi = (\tau, 1) \boxplus (\tau_2, 1) \boxplus \cdots \boxplus (\tau_r, 1).$$

Hence $L(s, \tau \times \sigma)$ has a (simple) pole at $s = 1$.

Let $\Delta(\tau, b)$ be the Speh residual representation in the discrete spectrum of $\mathrm{GL}_{ab}(\mathbb{A})$; see [Mœglin and Waldspurger 1989], or [Jiang et al. 2013, Section 1.2]. For any automorphic form

$$\phi \in \mathcal{A}\big(N_{ab}(\mathbb{A})M_{ab}(F) \setminus \mathrm{Sp}_{2ab+2m}(\mathbb{A})\big)_{\Delta(\tau,b)\otimes\sigma},$$

following [Langlands 1976; Mœglin and Waldspurger 1995], one has a residual Eisenstein series

$$E(\phi, s)(g) = E(g, \phi_{\Delta(\tau,b)\otimes\sigma}, s).$$

We refer to [Jiang et al. 2013] for particular details about this family of Eisenstein series. In particular, it is proved in [Jiang et al. 2013] that $E(\phi, s)(g)$ has a simple pole at $(b+1)/2$, which is the right-most one. We denote by $\mathcal{E}(g, \phi)$ the residue, which is square-integrable. They generate the residual representation $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ of $\mathrm{Sp}_{2n}(\mathbb{A})$. Following [Jiang et al. 2013, Section 6.2], the global Arthur parameter of this nonzero square-integrable automorphic representation $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ is exactly $\psi = (\tau, 2b+1) \boxplus \boxplus_{i=2}^r (\tau_i, 1)$ as in (2-1). We prove part (3) of Conjecture 1.2 for Case I.

Theorem 2.1. *For any global Arthur parameter of the form*

$$\psi = (\tau, 2b+1) \boxplus \mathop{\boxplus}_{i=2}^{r} (\tau_i, 1)$$

with $b \ge 1$ and $\tau \not\cong \tau_i$ for any $2 \le i \le r$, the residual representation $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the Barbasch–Vogan duality

$$\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}(\underline{p}(\psi))$$

of the partition $\underline{p}(\psi)$ associated to $(\psi, \mathrm{SO}_{2n+1}(\mathbb{C}))$.

In order to prove Theorem 2.1, we have to precisely figure out the partition $\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}(\underline{p}(\psi))$. We recall

Definition 2.2. Given any partition $\underline{pq} = [q_1 q_2 \cdots q_r]$ for $\mathfrak{so}_{2n+1}(\mathbb{C})$ satisfying $q_1 \geq q_2 \geq \cdots \geq q_r > 0$, whose even parts occur with even multiplicity, let $\underline{q}^- = [q_1 q_2 \cdots q_{r-1}(q_r - 1)]$. Then the Barbasch–Vogan duality $\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}$, following [Barbasch and Vogan 1985, Definition A1; Achar 2003, Section 3.5], is defined by

$$\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}(\underline{q}) := ((\underline{q}^-)_{\mathrm{Sp}_{2n}})^t,$$

where $(\underline{q}^-)_{\mathrm{Sp}_{2n}}$ is the Sp_{2n}-collapse of $\underline{q}^-$, which is the biggest special symplectic partition which is smaller than $\underline{q}^-$.

Following [Jiang 2014, Section 4], $\underline{p}(\psi) = [(2b+1)^a(1)^{2m+1-a}]$. As calculated in [Jiang and Liu 2015b], when $a = 2m+1$, by Definition 2.2,

$$\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}(\underline{p}(\psi)) = [(a)^{2b}(2m)];$$

when $a \leq 2m$ and a is even,

$$\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}(\underline{p}(\psi)) = [(2m)(a)^{2b}];$$

and finally, when $a \leq 2m$ and a is odd,

$$\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}(\underline{p}(\psi)) = [(2m)(a+1)(a)^{2b-2}(a-1)].$$

The proof of Theorem 2.1 goes as follows. Given a symplectic partition $\underline{p}$ of $2n$ (that is, where odd parts occur with even multiplicities), denote by $\underline{p}^{\mathrm{Sp}_{2n}}$ the Sp_{2n}-expansion of $\underline{p}$, which is the smallest special symplectic partition that is bigger than $\underline{p}$. In [Jiang and Liu 2015c], we proved the following theorem which provides a crucial reduction in the proof of Theorem 2.1.

Theorem 2.3 [Jiang and Liu 2015c, Theorem 4.1]. *Let π be an irreducible automorphic representation of $\mathrm{Sp}_{2n}(\mathbb{A})$. If π has a nonzero Fourier coefficient attached to a nonspecial symplectic partition $\underline{p}$ of $2n$, then π must have a nonzero Fourier coefficient attached to $\underline{p}^{\mathrm{Sp}_{2n}}$, the Sp_{2n}-expansion of the partition $\underline{p}$.*

If $a \leq 2m$ and a is odd, by [Collingwood and McGovern 1993, Lemma 6.3.9],

$$[(2m)(a+1)(a)^{2b-2}(a-1)] = [(2m)(a)^{2b}]^{\mathrm{Sp}_{2n}}.$$

Hence it suffices to prove the following theorem.

Theorem 2.4. *With notation above, the following hold.*

(1) *If $a = 2m+1$, then $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to $[(a)^{2b}(2m)]$.*

(2) *If $a \leq 2m$, then $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to $[(2m)(a)^{2b}]$.*

Parts (1) and (2) of Theorem 2.4 will be proved in Sections 4 and 5, respectively.

2B. *Case II.* $\psi \in \widetilde{\Psi}_2(\mathrm{Sp}_{2n})$ is written as

$$\psi = (\tau, 2b+1) \boxplus (\tau, 1) \boxplus \boxplus_{i=3}^{r} (\tau_i, 1), \tag{2-2}$$

where $b \geq 1$ and $\tau \ncong \tau_i$ for any $3 \leq i \leq r$. Assume that $\tau \in \mathcal{A}_{\mathrm{cusp}}(\mathrm{GL}_a)$ has central character ω_τ, and $\tau_i \in \mathcal{A}_{\mathrm{cusp}}(\mathrm{GL}_{a_i})$ has central character ω_{τ_i} for $3 \leq i \leq r$. Then $2n+1 = a(2b+1) + a + \sum_{i=3}^{r} a_i$ and $\omega_\tau^{2b+1} \cdot \omega_\tau \cdot \prod_{i=3}^{r} \omega_{\tau_i} = 1$. Consider the isobaric representation $\pi = \tau_3 \boxplus \cdots \boxplus \tau_r$ of $\mathrm{GL}_{2m+1}(\mathbb{A})$, where $2m+1 = \sum_{i=3}^{r} a_i = 2n+1-a(2b+2)$. Then π has central character $\omega_\pi = \prod_{i=3}^{r} \omega_{\tau_i} = 1$.

By [Ginzburg et al. 2011, Theorem 3.1], there is a generic $\sigma \in \mathcal{A}_{\mathrm{cusp}}(\mathrm{Sp}_{2m})$ such that σ has the functorial transfer π and hence $L(s, \tau \times \sigma)$ is holomorphic at $s = 1$ in this case. For any automorphic form

$$\phi \in \mathcal{A}\big(N_{a(b+1)}(\mathbb{A}) M_{a(b+1)}(F) \setminus \mathrm{Sp}_{2a(b+1)+2m}(\mathbb{A})\big)_{\Delta(\tau, b+1) \otimes \sigma},$$

one defines a residual Eisenstein series as in Case I

$$E(\phi, s)(g) = E(g, \phi_{\Delta(\tau, b+1) \otimes \sigma}, s).$$

By [Jiang et al. 2013], this Eisenstein series has a simple pole at $b/2$, which is the right-most one. Denote the representation generated by these residues at $s = b/2$ by $\mathcal{E}_{\Delta(\tau, b+1) \otimes \sigma}$, which is square-integrable. Following [Jiang et al. 2013] and [Shahidi 2010, Theorem 7.1.2], this residual representation $\mathcal{E}_{\Delta(\tau, b+1) \otimes \sigma}$ is nonzero. In particular, by Section 6.2 of [Jiang et al. 2013], the global Arthur parameter of $\mathcal{E}_{\Delta(\tau, b+1) \otimes \sigma}$ is exactly $\psi = (\tau, 2b+1) \boxplus (\tau, 1) \boxplus \boxplus_{i=3}^{r} (\tau_i, 1)$ as in Case II. In this case, we prove

Theorem 2.5. *For any global Arthur parameter of the form*

$$\psi = (\tau, 2b+1) \boxplus (\tau, 1) \boxplus \boxplus_{i=3}^{r} (\tau_i, 1)$$

with $b \geq 1$ and $\tau \ncong \tau_i$ for any $3 \leq i \leq r$, the residual representation $\mathcal{E}_{\Delta(\tau, b+1) \otimes \sigma}$ has a nonzero Fourier coefficient attached to the Barbasch–Vogan duality

$$\eta_{\mathfrak{so}_{2n+1}, \mathfrak{sp}_{2n}}(\underline{p}(\psi))$$

of the partition $\underline{p}(\psi)$ associated to $(\psi, \mathrm{SO}_{2n+1}(\mathbb{C}))$.

Following [Jiang 2014, Section 4], $\underline{p}(\psi) = [(2b+1)^a (1)^a (1)^{2m+1}]$. Now by Definition 2.2, we may calculate the partition $\eta_{\mathfrak{so}_{2n+1}, \mathfrak{sp}_{2n}}(\underline{p}(\psi))$ explicitly as follows. When a is even,

$$\begin{aligned}\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}(\underline{p}(\psi)) &= \eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}([(2b+1)^a(1)^{2m+1+a}]) \\ &= [(2b+1)^a(1)^{2m+a}]^t \\ &= [(a)^{2b+1}]+[(2m+a)] \\ &= [(2m+2a)(a)^{2b}].\end{aligned}$$

When a is odd,

$$\begin{aligned}\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}(\underline{p}(\psi)) &= \eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}([(2b+1)^a(1)^{2m+1+a}]) \\ &= ([(2b+1)^a(1)^{2m+a}]_{\mathrm{Sp}_{2n}})^t \\ &= [(2b+1)^{a-1}(2b)(2)(1)^{2m+a-1}]^t \\ &= [(a-1)^{2b+1}]+[(1)^{2b}]+[(1)^2]+[(2m+a-1)] \\ &= [(2m+2a)(a+1)(a)^{2b-2}(a-1)].\end{aligned}$$

As before, if a is odd, then, by the recipe for obtaining the Sp_{2n}-expansion of a symplectic partition $\underline{p}$ given in [Collingwood and McGovern 1993, Lemma 6.3.9],

$$[(2m+2a)(a+1)(a)^{2b-2}(a-1)] = [(2m+2a)(a)^{2b}]^{\mathrm{Sp}_{2n}}.$$

Hence it suffices to prove the following theorem.

Theorem 2.6. *The residual representation $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$ has a nonzero Fourier coefficient attached to* $[(2m+2a)(a)^{2b}]$.

The proof of Theorem 2.6 is given in Section 6, using induction on the integer b. We note that when $b=0$, the Arthur parameter is

$$\psi = 2(\tau,1)\boxplus\boxplus_{i=3}^{r}(\tau_i,1),$$

which does not parametrize automorphic representations in the discrete spectrum. Indeed, in this case, the corresponding automorphic representation constructed from the Eisenstein series is the value at $s=0$, which we still denote by $\mathcal{E}_{\Delta(\tau,1)\otimes\sigma}=\mathcal{E}_{\tau\otimes\sigma}$. It is clear that in this case, the partition $\underline{p}(\psi)$ is the trivial partition. On the other hand, following [Shahidi 2010, Theorem 7.1.3], the representation $\mathcal{E}_{\Delta(\tau,1)\otimes\sigma}$ has a nonzero Whittaker–Fourier coefficient. In other words, Theorem 2.6 still holds for $b=0$. As we proceed in Section 6, the case of $b=0$ will serve as the base of the induction argument.

2C. ***Case III.*** $\psi\in\widetilde{\Psi}_2(\mathrm{Sp}_{2n})$ is written as

$$\psi = (\tau,2b)\boxplus\boxplus_{i=2}^{r}(\tau_i,1), \tag{2-3}$$

where $b \geq 1$. In this case, τ is of symplectic type (and hence $a = 2k$ is even), while τ_i for all $2 \leq i \leq r$ are of orthogonal type. Assume that $\tau \in \mathcal{A}_{\text{cusp}}(\text{GL}_a)$ has central character ω_τ, and $\tau_i \in \mathcal{A}_{\text{cusp}}(\text{GL}_{a_i})$ has central character ω_{τ_i} for $2 \leq i \leq r$. By the definition of Arthur parameters, one has that $2n+1 = 2ab + \sum_{i=2}^r a_i$, and $\prod_{i=2}^r \omega_{\tau_i} = 1$. Consider the isobaric representation $\pi = \tau_2 \boxplus \cdots \boxplus \tau_r$ of $\text{GL}_{2m+1}(\mathbb{A})$, where $2m+1 = \sum_{i=2}^r a_i = 2n+1-2ab$. Hence π has central character $\omega_\pi = \prod_{i=2}^r \omega_{\tau_i} = 1$.

By [Ginzburg et al. 2011, Theorem 3.1], there is a generic $\sigma \in \mathcal{A}_{\text{cusp}}(\text{Sp}_{2m})$ that has the functorial transfer π. Then we define a residual Eisenstein series

$$E(\phi, s)(g) = E(g, \phi_{\Delta(\tau,b)\otimes\sigma}, s)$$

associated to any automorphic form

$$\phi \in \mathcal{A}\big(N_{ab}(\mathbb{A})\, M_{ab}(F) \setminus \text{Sp}_{2ab+2m}(\mathbb{A})\big)_{\Delta(\tau,b)\otimes\sigma}.$$

By [Jiang et al. 2013], this Eisenstein series may have a simple pole at $b/2$, which is the right-most one. Denote the representation generated by these residues at $s = b/2$ by $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$. This residual representation is square-integrable. If $L\big(\frac{1}{2}, \tau \times \sigma\big) \neq 0$, the residual representation $\mathcal{E}_{\tau\otimes\sigma}$ is nonzero, and hence by the induction argument in [Jiang et al. 2013], the residual representation $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ is also nonzero. Finally, following [op. cit., Section 6.2], we see that the global Arthur parameter of $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ is exactly $\psi = (\tau, 2b) \boxplus \boxplus_{i=2}^r (\tau_i, 1)$ as in (2-3).

Theorem 2.7. *Assume that $a = 2k$ and $L\big(\frac{1}{2}, \tau \times \sigma\big) \neq 0$. If the residual representation $\mathcal{E}_{\tau\otimes\sigma}$ of $\text{Sp}_{4k+2m}(\mathbb{A})$, with $\sigma \ncong 1_{\text{Sp}_0(\mathbb{A})}$, has a nonzero Fourier coefficient attached to the partition $[(2k+2m)(2k)]$, then, for any $b \geq 1$, the residual representation $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2k+2m)(2k)^{2b-1}]$.*

We remark that if $\sigma \cong 1_{\text{Sp}_0(\mathbb{A})}$, $\big(\frac{1}{2}, \big) = L\big(\frac{1}{2}, \tau \times \sigma\big) \neq 0$. In this case, [Liu 2013a, Theorem 4.2.2] shows that $\mathfrak{p}^m(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}) = \{[(2k)^{2b}]\}$.

In fact, the assumption that the residual representation $\mathcal{E}_{\tau\otimes\sigma}$ of $\text{Sp}_{4k+2m}(\mathbb{A})$, with $\sigma \ncong 1_{\text{Sp}_0(\mathbb{A})}$, has a nonzero Fourier coefficient attached to the partition $[(2k+2m)(2k)]$ is exactly [Ginzburg et al. 2004, Conjecture 6.1], and hence Theorem 2.7 has a close connection to the Gan–Gross–Prasad conjecture [Gan et al. 2012]. We will come back to this issue in our future work.

In this case, $\underline{p}(\psi) = [(2b)^a (1)^{2m+1}]$, and following the calculation in [Jiang and Liu 2015b],

$$\eta_{\mathfrak{so}_{2n+1},\mathfrak{sp}_{2n}}(\underline{p}(\psi)) = [(a+2m)(a)^{2b-1}],$$

where $a = 2k$ is even. The proof of Theorem 2.7 is given in Section 7.

When $L\left(\frac{1}{2}, \tau \times \sigma\right)$ is zero for the Arthur parameter in (2-3), the corresponding automorphic L^2-packet $\widetilde{\Pi}_\psi(\epsilon_\psi)$ are expected to contain all cuspidal automorphic representations if it is not empty. We are going to apply the construction of endoscopy correspondences outlined in [Jiang 2014] to construct the distinguished cuspidal members in $\widetilde{\Pi}_\psi(\epsilon_\psi)$. The details for this case will be considered in our future work. See [Jiang and Liu 2015b] for a brief discussion in this aspect.

3. A basic lemma

We recall a basic lemma from [Jiang and Liu 2015b], which will be a technical key step in the proofs of this paper. Let H be a reductive group defined over F. We first recall [Jiang and Liu 2013, Lemma 5.2], which is also formulated in a slightly different version in [Ginzburg et al. 2011, Corollary 7.1]. Note that the proof of [Jiang and Liu 2013, Lemma 5.2] is valid for $H(\mathbb{A})$.

Let C be an F-subgroup of a maximal unipotent subgroup of H, and let ψ_C be a nontrivial character of $[C] = C(F) \setminus C(\mathbb{A})$. Suppose that $\widetilde{X}, \widetilde{Y}$ are two unipotent F-subgroups, satisfying the following conditions:

(1) $\widetilde{X}$ and $\widetilde{Y}$ normalize C;

(2) $\widetilde{X} \cap C$ and $\widetilde{Y} \cap C$ are normal in $\widetilde{X}$ and $\widetilde{Y}$, respectively, $(\widetilde{X} \cap C) \setminus \widetilde{X}$ and $(\widetilde{Y} \cap C) \setminus \widetilde{Y}$ are abelian;

(3) $\widetilde{X}(\mathbb{A})$ and $\widetilde{Y}(\mathbb{A})$ preserve ψ_C;

(4) ψ_C is trivial on $(\widetilde{X} \cap C)(\mathbb{A})$ and $(\widetilde{Y} \cap C)(\mathbb{A})$;

(5) $[\widetilde{X}, \widetilde{Y}] \subset C$;

(6) there is a nondegenerate pairing $(\widetilde{X} \cap C)(\mathbb{A}) \times (\widetilde{Y} \cap C)(\mathbb{A}) \to \mathbb{C}^*$, given by $(x, y) \mapsto \psi_C([x, y])$, which is multiplicative in each coordinate, and identifies $(\widetilde{Y} \cap C)(F) \setminus \widetilde{Y}(F)$ and $(\widetilde{X} \cap C)(F) \setminus \widetilde{X}(F)$ with the duals of the subgroups $\widetilde{X}(F)(\widetilde{X} \cap C)(\mathbb{A}) \setminus \widetilde{X}(\mathbb{A})$ and $\widetilde{Y}(F)(\widetilde{Y} \cap C)(\mathbb{A}) \setminus \widetilde{Y}(\mathbb{A})$, respectively.

Let $B = C\widetilde{Y}$ and $D = C\widetilde{X}$, and extend ψ_C trivially to characters of $[B] = B(F) \setminus B(\mathbb{A})$ and $[D] = D(F) \setminus D(\mathbb{A})$, which will be denoted by ψ_B and ψ_D, respectively.

Lemma 3.1 [Jiang and Liu 2013, Lemma 5.2]. *Assume that $(C, \psi_C, \widetilde{X}, \widetilde{Y})$ satisfies all the above conditions. Let f be an automorphic form on $H(\mathbb{A})$. Then*

$$\int_{[C]} f(cg)\psi_C^{-1}(c)\, dc \equiv 0, \quad \textit{for all } g \in H(\mathbb{A}),$$

if and only if

$$\int_{[D]} f(ug)\psi_D^{-1}(u)\, du \equiv 0, \quad \textit{for all } g \in H(\mathbb{A}),$$

if and only if

$$\int_{[B]} f(vg)\psi_B^{-1}(v)\,dv \equiv 0, \quad \textit{for all } g \in H(\mathbb{A}).$$

For simplicity, we always use ψ_C to denote its extensions ψ_B and ψ_D when we apply Lemma 3.1 to various circumstances. Lemma 3.1 can be extended as follows and will be a technical key in this paper.

Lemma 3.2 [Jiang and Liu 2015b, Lemma 6.2]. *Assume that* $(C, \psi_C, \widetilde{X}, \widetilde{Y})$ *satisfies the following conditions*: $\widetilde{X} = \{\widetilde{X}_i\}_{i=1}^r$, $\widetilde{Y} = \{\widetilde{Y}_i\}_{i=1}^r$, *and for* $1 \le i \le r$, *each quadruple*

$$(\widetilde{X}_{i-1} \cdots \widetilde{X}_1 C \widetilde{Y}_r \cdots \widetilde{Y}_{i+1}, \psi_C, \widetilde{X}_i, \widetilde{Y}_i)$$

satisfies all the conditions of Lemma 3.1. Let f *be an automorphic form on* $H(\mathbb{A})$. *Then*

$$\int_{[\widetilde{X}_r \cdots \widetilde{X}_1 C]} f(xcg)\psi_C^{-1}(c)\,dc\,dx \equiv 0, \quad \textit{for all } g \in H(\mathbb{A}),$$

if and only if

$$\int_{[C\widetilde{Y}_r \cdots \widetilde{Y}_1]} f(cyg)\psi_C^{-1}(c)\,dy\,dc \equiv 0, \quad \textit{for all } g \in H(\mathbb{A}).$$

The proof of this lemma is carried out by using Lemma 3.1 inductively, and was given with full details in [loc. cit.].

4. Proof of part (1) of Theorem 2.4

In this section, we assume that $a = 2m+1$ and show that $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to $\underline{p} := [(2m+1)^{2b}(2m)]$.

Proof of part (1) of Theorem 2.4. We will prove the theorem by induction on b. Note that when $b = 0$, $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma} \cong \sigma$ which has a nonzero Fourier coefficient attached to $[(2m)]$ since σ is generic. Now assume that $\mathcal{E}_{\Delta(\tau,b-1)\otimes\sigma}$ has a nonzero $\psi_{[(2m+1)^{2b-2}(2m)],\alpha}$-Fourier coefficient attached to $[(2m+1)^{2b-2}(2m)]$, for some $\alpha \in F^*/(F^*)^2$.

Take any $\varphi \in \mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ and consider its $\psi_{\underline{p},\alpha}$-Fourier coefficients attached to $\underline{p}$:

$$\varphi^{\psi_{\underline{p},\alpha}}(g) = \int_{[V_{\underline{p},2}]} \varphi(vg)\psi_{\underline{p},\alpha}^{-1}(v)\,dv. \tag{4-1}$$

For definitions of the unipotent group $V_{\underline{p},2}$ and its character $\psi_{\underline{p},\alpha}$, see [Jiang and Liu 2015c, Section 2]. By [op. cit., Corollary 2.4], the integral in (4-1) is nonvanishing if and only if the following integral is nonvanishing:

$$\int_{[Y_1 V_{\underline{p},2}]} \varphi(vg)\psi_{\underline{p},\alpha}^{-1}(v)\,dv, \tag{4-2}$$

where Y_1 is defined in [Jiang and Liu 2015c, (2.5)] corresponding to the partition $[(2m+1)^{2b}(2m)]$ and the character $\psi_{\underline{p},\alpha}$ extends to $Y_1 V_{\underline{p},2}$ trivially.

Assume that T is the maximal split torus in $\mathrm{Sp}_{2b(2m+1)+2m}$, consisting of elements

$$\mathrm{diag}(t_1, t_2, \dots, t_{b(2m+1)+m}, t_{b(2m+1)+m}^{-1}, \dots, t_2^{-1}, t_1^{-1}).$$

Let ω_1 be the Weyl element of $\mathrm{Sp}_{2b(2m+1)+2m}$, sending elements $t \in T$ to the torus elements

$$t' = \mathrm{diag}(t^{(0)}, t^{(1)}, t^{(2)}, \dots, t^{(m)}, t^{(m+1)}, t^{(m),*}, \dots, t^{(2),*}, t^{(1),*}, t^{(0),*}), \tag{4-3}$$

where $t^{(0)} = \mathrm{diag}(t_1, t_2, \dots, t_{2m+1})$, and with $e = 2m+1$,

$$t^{(m+1)} = \mathrm{diag}(t_{e+m+1}, \dots, t_{(b-1)e+m+1}, t_{be-m}^{-1}, \dots, t_{2e-m}^{-1})$$

and

$$t^{(j)} = \mathrm{diag}(t_{e+j}, \dots, t_{(b-1)e+j}, t_{be-j+1}^{-1}, \dots, t_{2e-j+1}^{-1}, t_{be+j}),$$

for $1 \le j \le m$.

Now identify $\mathrm{Sp}_{(2b-1)(2m+1)+2m}$ with its image in $\mathrm{Sp}_{2b(2m+1)+2m}$ under the embedding $g \mapsto \mathrm{diag}(I_{2m+1}, g, I_{2m+1})$, and denote the restriction of ω_1 to $\mathrm{Sp}_{(2b-1)(2m+1)+2m}$ by ω_1'. We conjugate cross the integration variables by ω_1 from the left; then the integral in (4-2) becomes

$$\int_{[U_{\underline{p},2}]} \varphi(u\omega_1 g)\psi_{\underline{p},\alpha}^{\omega_1}(u)^{-1}\, du, \tag{4-4}$$

where $U_{\underline{p},2} = \omega_1 Y_1 V_{\underline{p},2}\omega_1^{-1}$, and $\psi_{\underline{p},\alpha}^{\omega_1}(u) = \psi_{\underline{p},\alpha}(\omega_1^{-1}u\omega_1)$.

Now, we describe the structure of elements in $U_{\underline{p},2}$, each of which has the form

$$u = \begin{pmatrix} z_{2m+1} & q_1 & q_2 \\ 0 & u' & q_1^* \\ 0 & 0 & z_{2m+1}^* \end{pmatrix} \begin{pmatrix} I_{2m+1} & 0 & 0 \\ p_1 & I_{(2b-2)(2m+1)+2m} & 0 \\ p_2 & p_1^* & I_{2m+1} \end{pmatrix}, \tag{4-5}$$

where $z_{2m+1} \in V_{2m+1}$, the standard maximal unipotent subgroup of GL_{2m+1}; $u' \in U_{[(2m+1)^{2b-2}(2m)],2} := \omega_1' Y_2 V_{[(2m+1)^{2b-2}(2m)],2}\omega_1'^{-1}$ with Y_2 as in [Jiang and Liu 2015c, (2.5)] corresponding to the partition $[(2m+1)^{2b-2}(2m)]$; and p_i, q_i, $1 \le i \le 2$, are described as follows:

- $q_1 \in M_{(2m+1)\times((2b-2)(2m+1)+2m)}$, such that $q_1(i,j) = 0$ for $1 \le i \le 2m+1$ and $1 \le j \le (2b-2) + (2b-1)(i-1)$.
- $p_1 \in M_{((2b-2)(2m+1)+2m)\times(2m+1)}$, such that $p_1(i,j) = 0$ for $1 \le j \le 2m+1$ and $(2b-2) + (2b-1)(i-1) + 1 \le i \le (2b-2)(2m+1) + 2m$.
- $q_2 \in M_{(2m+1)\times(2m+1)}$, symmetric with respect to the secondary diagonal, such that $q_2(i,j) = 0$ for $1 \le i \le 2m+1$ and $1 \le j \le i$.

- $p_2 \in M_{(2m+1)\times(2m+1)}$, symmetric with respect to the secondary diagonal, such that $p_2(i,j) = 0$ for $1 \le i \le 2m+1$ and $1 \le j \le i$.

Note that

$$\psi^{\omega_1}_{\underline{p},\alpha}\begin{pmatrix} z_{2m+1} & q_1 & q_2 \\ 0 & I_{(2b-2)(2m+1)+2m} & q_1^* \\ 0 & 0 & z^*_{2m+1} \end{pmatrix} = \psi\Bigl(\sum_{i=1}^{2m} z_{2m+1}(i,i+1)\Bigr).$$

Next, we apply Lemma 3.2 to fill the zero entries in q_1, q_2 using the nonzero entries in p_1, p_2. To proceed, we need to define a sequence of one-dimensional root subgroups and put them in a correct order.

Let X_j, with $1 \le j \le (2b-2)+1$, be the one-dimensional subgroups corresponding to the roots such that the corresponding entries are in the first row of q_1 or q_2 and are identically zero, from right to left. For $1 < i \le m$, let X_j, with

$$\Bigl(\sum_{k=1}^{i-1}[(2b-2)+(2b-1)(k-1)+k]\Bigr)+1 \le j \le \sum_{k=1}^{i}[(2b-2)+(2b-1)(k-1)+k],$$

be the one-dimensional subgroups corresponding to the roots such that the corresponding entries are in the i-th row of q_1 or q_2 and are identically zero, from right to left.

Let Y_j, with $1 \le j \le (2b-2)+1$, be the one-dimensional subgroups corresponding to the roots such that the corresponding entries are in the second column of p_1 or p_2 and are not identically zero, from bottom to top. For $1 < i \le m$, let Y_j, with

$$1+\sum_{k=1}^{i-1}[(2b-2)+(2b-1)(k-1)+k] \le j \le \sum_{k=1}^{i}[(2b-2)+(2b-1)(k-1)+k],$$

be the one-dimensional subgroups corresponding to the roots such that the corresponding entries are in the $(i+1)$-th column of p_1 or p_2 and are not identically zero, from bottom to top.

Let W_1 be the subgroup of $U_{\underline{p},2}$ such that the entries corresponding to the one-dimensional subgroups Y_j above, with

$$1 \le j \le \ell := \sum_{k=1}^{m}[(2b-2)+(2b-1)(k-1)+k],$$

are all identically zero. And let $\psi_{W_1} = \psi^{\omega_1}_{\underline{p},\alpha}|_{W_1}$. Then $(W_1, \psi_{W_1}, \{X_j\}_j^\ell, \{Y_j\}_j^\ell)$ satisfies all the conditions for Lemma 3.2. Hence, by that lemma, the integral in (4-4) is nonvanishing if and only if the following integral is nonvanishing:

$$\int_{[W_2]} \varphi(w\omega_1 g)\,\psi_{W_2}(w)^{-1}\,dw, \tag{4-6}$$

where $W_2 := \prod_{j=1}^{\ell} X_j W_1$ and ψ_{W_2} is the character on W_2 extended trivially from ψ_{W_1}.

Now we consider the i-th row of q_1 and q_2, with $m+1 \leq i \leq 2m$. We will continue to apply Lemma 3.2 to fill the zero entries in q_1 and q_2, row by row, from the $(m+1)$-th row to $2m$-th row. But for each $m+1 \leq i \leq 2m$, before we apply Lemma 3.2 as above, we need to take the Fourier expansion along the one-dimensional root subgroup X_{2e_i}. For example, for $i = m+1$, we first take the Fourier expansion of the integral in (4-6) along the one-dimensional root subgroup $X_{2e_{m+1}}$. We will get two kinds of Fourier coefficients corresponding to the orbits of the dual of $[X_{2e_{m+1}}] := X_{2e_{m+1}}(F) \setminus X_{2e_{m+1}}(\mathbb{A})$: the trivial orbit and the nontrivial one. For the Fourier coefficients attached to the nontrivial orbit, we can see that there is an inner integral

$$\varphi^{\psi_{[(2m+2)1^{2b(2m+1)-2}],\beta}}, \quad \beta \in F^*,$$

which is identically zero by [Jiang and Liu 2015a, Proposition 6.4]. Therefore only the Fourier coefficient attached to the trivial orbit, which actually equals to the integral in (4-6), survives. Then, we can apply the Lemma 3.2 to the $(m+1)$-th row of q_1 and q_2 similarly as above.

After considering all the i-th row of q_1 and q_2, $m+1 \leq i \leq 2m$ as above, we get that the integral in (4-6) is nonvanishing if and only if the following integral is nonvanishing:

$$\int_{[W_3]} \varphi(w\omega_1 g)\, \psi_{W_3}(w)^{-1}\, dw, \tag{4-7}$$

where W_3 has elements of the following form:

$$w = \begin{pmatrix} z_{2m+1} & q_1 & q_2 \\ 0 & u' & q_1^* \\ 0 & 0 & z_{2m+1}^* \end{pmatrix}, \tag{4-8}$$

where $z_{2m+1} \in V_{2m+1}$, the standard maximal unipotent subgroup of GL_{2m+1};

$$u' \in U_{[(2m+1)^{2b-2}(2m)],2} := \omega_1' Y_2 V_{[(2m+1)^{2b-2}(2m)],2} \omega_1'^{-1}$$

with Y_2 as in [op. cit., (2.5)] corresponding to the partition $[(2m+1)^{2b-2}(2m)]$;

$$q_1 \in M_{(2m+1)\times((2b-2)(2m+1)+2m)},$$

such that $q_1(2m+1, j) = 0$ for $1 \leq j \leq (2b-2)(2m+1)+2m$;

$$q_2 \in M_{(2m+1)\times(2m+1)},$$

symmetric with respect to the secondary diagonal, such that $q_2(2m+1,1)=0$. Also,

$$\psi_{W_3}\begin{pmatrix} z_{2m+1} & q_1 & q_2 \\ 0 & I_{(2b-2)(2m+1)+2m} & q_1^* \\ 0 & 0 & z_{2m+1}^* \end{pmatrix} = \psi\Big(\sum_{i=1}^{2m} z_{2m+1}(i,i+1)\Big).$$

Now consider the Fourier expansion of the integral in (4-7) along the one-dimensional root subgroup $X_{2e_{2m+1}}$. By the same reason as above, only the Fourier coefficient corresponding to the trivial orbit of the dual of $[X_{2e_{2m+1}}]$ survives, which is actually equal to the integral in (4-7):

$$\int_{[W_4]} \varphi(w\omega_1 g)\,\psi_{W_4}(w)^{-1}\,dw, \tag{4-9}$$

where elements in W_4 have the same structure as in (4-8), except that $q_2(2m+1,1)$ is not identically zero.

It is easy to see that the integral in (4-9) has an inner integral which is exactly $\varphi^{\psi_{N_{1^{2m}}}}$, using notation in Lemma 4.2 below. On the other hand, we know that by Lemma 4.2 below, $\varphi^{\psi_{N_{1^{2m}}}} = \varphi^{\tilde{\psi}_{N_{1^{2m+1}}}}$. Therefore, the integral in (4-9) becomes

$$\int_{[W_5]} \varphi(w\omega_1 g)\psi_{W_5}(w)^{-1}\,dw, \tag{4-10}$$

where elements in W_5 are of the form:

$$w = w(z_{2m+1}, u', q_1, q_2) = \begin{pmatrix} z_{2m+1} & q_1 & q_2 \\ 0 & u' & q_1^* \\ 0 & 0 & z_{2m+1}^* \end{pmatrix},$$

where $z_{2m+1} \in V_{2m+1}$, the standard maximal unipotent subgroup of GL_{2m+1};

$$u' \in U_{[(2m+1)^{2b-2}(2m)],2} := \omega_1' Y_2 V_{[(2m+1)^{2b-2}(2m)],2} \omega_1'^{-1}$$

with Y_2 as in [loc. cit.] corresponding to the partition $[(2m+1)^{2b-2}(2m)]$;

$$q_1 \in M_{(2m+1)\times((2b-2)(2m+1)+2m)},$$

and $q_2 \in M_{(2m+1)\times(2m+1)}$, symmetric with respect to the secondary diagonal. And

$$\psi_{W_5}\begin{pmatrix} z_{2m+1} & q_1 & q_2 \\ 0 & I_{(2b-2)(2m+1)+2m} & q_1^* \\ 0 & 0 & z_{2m+1}^* \end{pmatrix} = \psi\Big(\sum_{i=1}^{2m} z_{2m+1}(i,i+1)\Big).$$

Hence, the integral in (4-10) can be written as

$$\int_{W_6} \varphi_{P_{2m+1}}(w\omega_1 g)\,\psi_{W_6}(w)^{-1}\,dw, \tag{4-11}$$

where W_6 is a subgroup of W_5 consisting of elements of the form $w(z_{2m+1}, u', 0, 0)$, $\psi_{W_6} = \psi_{W_5}|_{W_6}$, and $\varphi_{P_{2m+1}}$ is the constant term of φ along the parabolic subgroup

$P_{2m+1} = M_{2m+1}N_{2m+1}$ of $\mathrm{Sp}_{2b(2m+1)+2m}$ with the Levi subgroup isomorphic to $\mathrm{GL}_{2m+1} \times \mathrm{Sp}_{(2b-2)(2m+1)+2m}$.

By Lemma 4.1 below, $\varphi_{P_{2m+1}}(w\omega_1 g)$ is an automorphic form in $\tau|\cdot|^{-b} \otimes \mathcal{E}_{\Delta(\tau,b-1)\otimes\sigma}$ when restricted to the Levi subgroup. Note that the restriction of ψ_{W_5} to the z_{2m+1}-part gives a Whittaker coefficient of τ, and the restriction to the u'-part gives a $\psi_{[(2m+1)^{2b-2}(2m)],\alpha}$-Fourier coefficient of $\mathcal{E}_{\Delta(\tau,b-1)\otimes\sigma}$ up to the conjugation of the Weyl element ω_1'. On the other hand, τ is generic, and by induction assumption, $\mathcal{E}_{\Delta(\tau,b-1)\otimes\sigma}$ has a nonzero $\psi_{[(2m+1)^{2b-2}(2m)],\alpha}$-Fourier coefficient. Therefore, we conclude that $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero $\psi_{\underline{p},\alpha}$-Fourier coefficient attached to the partition $\underline{p} = [(2m+1)^{2b}(2m)]$. This completes the proof of part (1) of Theorem 2.4, up to Lemmas 4.1 and 4.2, which are stated below. $\square$

Note that Lemmas 4.1 and 4.2 are analogs of [Liu 2013a, Lemmas 4.2.4 and 4.2.6], with similar arguments, and hence we state them without proofs.

Lemma 4.1. *Let $P_{ai} = M_{ai}N_{ai}$, with $1 \le i \le b$ and $a \le 2m+1$, be the parabolic subgroup of Sp_{2ab+2m} with Levi part*

$$M_{ai} \cong \mathrm{GL}_{ai} \times \mathrm{Sp}_{a(2b-2i)+2m}\,.$$

Let φ be an arbitrary automorphic form in $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$. Denote by $\varphi_{P_{ai}}(g)$ the constant term of φ along P_{ai}. Then, for $1 \le i \le b$,

$$\varphi_{P_{ai}} \in \mathcal{A}\big(N_{ai}(\mathbb{A})M_{ai}(F)\setminus \mathrm{Sp}_{2ab+2m}(\mathbb{A})\big)_{\Delta(\tau,i)|\cdot|^{-(2b+1-i)/2}\otimes\mathcal{E}_{\Delta(\tau,b-i)\otimes\sigma}}.$$

Note that when $b = i$, $\mathcal{E}_{\Delta(\tau,b-i)\otimes\sigma} = \sigma$.

Lemma 4.2. *Let N_{1^p} be the unipotent radical of the parabolic subgroup P_{1^p} of $\mathrm{Sp}_{2b(2m+1)+2m}$ with the Levi part being $\mathrm{GL}_1^{\times p} \times \mathrm{Sp}_{2b(2m+1)+2m-2p}$. Let*

$$\psi_{N_{1^p}}(n) := \psi(n_{1,2}+\cdots+n_{p,p+1}) \quad \textit{and} \quad \tilde{\psi}_{N_{1^p}}(n) := \psi(n_{1,2}+\cdots+n_{p-1,p})$$

be two characters of N_{1^p}. For any automorphic form $\varphi \in \mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$, define $\psi_{N_{1^p}}$ and $\tilde{\psi}_{N_{1^p}}$-Fourier coefficients as follows:

$$\varphi^{\psi_{N_{1^p}}}(g) := \int_{[N_{1^p}]} \varphi(ng)\,\psi_{N_{1^p}}(n)^{-1}\,dn \tag{4-12}$$

and

$$\varphi^{\tilde{\psi}_{N_{1^p}}}(g) := \int_{[N_{1^p}]} \varphi(ng)\,\tilde{\psi}_{N_{1^p}}(n)^{-1}\,du. \tag{4-13}$$

Then $\varphi^{\psi_{N_{1^p}}} \equiv 0$ for all $p \ge 2m+1$, and $\varphi^{\psi_{N_{1^{2m}}}} = \varphi^{\tilde{\psi}_{N_{1^{2m+1}}}}$.

5. Proof of part (2) of Theorem 2.4

In this section, we assume that $a \le 2m$ and σ is ψ^α-generic for $\alpha \in F^*/(F^*)^2$, and show that $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to $[(2m)(a)^{2b}]$.

First, we construct a residual representation of $\widetilde{\mathrm{Sp}}_{2ab}(\mathbb{A})$ as follows. For any $\tilde{\phi} \in \mathcal{A}\big(N_{ab}(\mathbb{A})\widetilde{M}_{ab}(F)\setminus\widetilde{\mathrm{Sp}}_{2ab}(\mathbb{A})\big)_{\gamma_{\psi^{-\alpha}}\Delta(\tau,b)}$, following [Mœglin and Waldspurger 1995], an residual Eisenstein series can be defined by

$$\widetilde{E}(\tilde{\phi},s)(g) = \sum_{\gamma \in P_{ab}(F)\setminus \mathrm{Sp}_{2ab}(F)} \lambda_s \tilde{\phi}(\gamma g).$$

It converges absolutely for real part of s large and has meromorphic continuation to the whole complex plane $\mathbb{C}$. By similar argument as that in [Jiang et al. 2013], this Eisenstein series has a simple pole at $b/2$, which is the right-most one. Denote the representation generated by these residues at $s = b/2$ by $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$. This residual representation is square-integrable.

We separate the proof of part (2) of Theorem 2.4 into *three steps*:

***Step* (1)** $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2m)1^{2ab}]$ with respect to the character $\psi_{[(2m)1^{2ab}],\alpha}$ (for definition, see [Jiang and Liu 2015c, Section 2]).

***Step* (2)** $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ is irreducible. Let $\mathcal{D}^{2ab+2m}_{2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$ be the ψ^α-descent of $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ [Ginzburg et al. 2011, Section 3.2]. Then, as a representation of $\widetilde{\mathrm{Sp}}_{2ab}(\mathbb{A})$, it is square-integrable and contains the whole space of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$.

***Step* (3)** $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ has a nonzero Fourier coefficient attached to the symplectic partition $[(a)^{2b}]$.

Proof of part (2) of Theorem 2.4. From the results in steps (1)–(3) above, we can see that $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the composite partition $[(2m)1^{2ab}] \circ [(a)^{2b}]$ (for the definition of composite partitions and the attached Fourier coefficients, we refer to [Ginzburg et al. 2003, Section 1]). Therefore, by [Jiang and Liu 2015c, Lemma 3.1] or [Ginzburg et al. 2003, Lemma 2.6], $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to $[(2m)(a)^{2b}]$, which completes the proof of the part (2) of Theorem 2.4. □

5A. ***Proof of step (1).*** Note that by [Ginzburg et al. 2003, Lemma 1.1], $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2m)1^{2ab}]$ with respect to the character $\psi_{[(2m)1^{2ab}],\alpha}$ if and only if the ψ^α-descent $\mathcal{D}^{2ab+2m}_{2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$ of $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ is not identically zero as a representation of $\widetilde{\mathrm{Sp}}_{2ab}(\mathbb{A})$.

Recall that $P_r^{2l} = M_r^{2l}N_r^{2l}$ (with $1 \le r \le l$) is the standard parabolic subgroup of Sp_{2l} with Levi part M_r^{2l} isomorphic to $\mathrm{GL}_r \times \mathrm{Sp}_{2l-2r}$ and N_r^{2l} the unipotent radical. $\tilde{P}_r^{2l}(\mathbb{A})$ is the preimage of $P_r^{2l}(\mathbb{A}) = \widetilde{M}_r^{2l}(\mathbb{A})N_r^{2l}(\mathbb{A})$ in $\widetilde{\mathrm{Sp}}_{2l}(\mathbb{A})$.

Take any $\xi \in \mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$; we will calculate the constant term of the Fourier–Jacobi coefficient $\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)$ along P^{2ab}_r, which is denoted by $\mathcal{C}_{N^{2ab}_r}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)\big)$, where $1 \le r \le ab$.

By [Ginzburg et al. 2011, Theorem 7.8],

$$\tag{5-1} \mathcal{C}_{N^{2ab}_r}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)\big) = \sum_{\substack{0\le k\le r\\ \gamma\in P^1_{r-k,1^k}(F)\backslash \mathrm{GL}_r(F)}} \int_{L(\mathbb{A})} \phi_1(i(\lambda))\,\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1+k}}(\mathcal{C}_{N^{2ab+2m}_{r-k}}\big(\xi\big))(\hat{\gamma}\lambda\beta)\,d\lambda.$$

We explain the notation used in (5-1) as follows: N^{2ab+2m}_{r-k} denotes the unipotent radical of the parabolic subgroup P^{2ab+2m}_{r-k} of Sp_{2ab+2m} with the Levi subgroup $\mathrm{GL}_{r-k}\times \mathrm{Sp}_{2ab+2m-2r+2k}$, and $P^1_{r-k,1^k}$ is a subgroup of GL_r consisting of matrices of the form

$$\begin{pmatrix} g & x\\ 0 & z\end{pmatrix},$$

with $z \in U_k$, the standard maximal unipotent subgroup of GL_k. For $g \in \mathrm{GL}_j$, with $j \le ab+m$, $\hat{g} = \mathrm{diag}(g, I_{2ab+2m-2j}, g^*)$, and L is a unipotent subgroup, consisting of matrices of the form

$$\lambda = \begin{pmatrix} I_r & 0\\ x & I_m\end{pmatrix}^{\wedge}$$

with $i(\lambda)$ in the last row of x, and

$$\beta = \begin{pmatrix} 0 & I_r\\ I_m & 0\end{pmatrix}^{\wedge}.$$

We assume that $\phi = \phi_1\otimes\phi_2$, with $\phi_1 \in \mathcal{S}(\mathbb{A}^r)$ and $\phi_2 \in \mathcal{S}(\mathbb{A}^{ab-r})$. Finally, the Fourier–Jacobi coefficients satisfy the identity

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m-1+k}}\big(\mathcal{C}_{N^{2ab+2m}_{r-k}}(\xi)\big)(\hat{\gamma}\lambda\beta) := \mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m-1+k}}\big(\mathcal{C}_{N^{2ab+2m}_{r-k}}(\rho(\hat{\gamma}\lambda\beta)\xi)\big)(I),$$

with $\rho(\hat{\gamma}\lambda\beta)$ denoting the right translation by $\hat{\gamma}\lambda\beta$, where the function is regarded as taking first the constant term $\mathcal{C}_{N^{2ab+2m}_{r-k}}(\rho(\hat{\gamma}\lambda\beta)\xi)$, and then after restricted to $\mathrm{Sp}_{2ab+2m-2r+2k}(\mathbb{A})$, taking the Fourier–Jacobi coefficient

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m-1+k}},$$

which is a map taking automorphic forms on $\mathrm{Sp}_{2ab+2m-2r+2k}(\mathbb{A})$ to those on $\widetilde{\mathrm{Sp}}_{2ab-2r}(\mathbb{A})$.

By the cuspidal support of ξ, $\mathcal{C}_{N^{2ab+2m}_{r-k}}(\xi)$ is identically zero, unless $k = r$ or $r-k = la$ with $1 \le l \le b$. When $k = r$, since $[(2m+2r)1^{2ab-2r}]$ is bigger than $\eta_{\mathfrak{so}_{2n+1}(\mathbb{C}),\mathfrak{sp}_{2n}(\mathbb{C})}(\underline{p}(\psi))$ under the lexicographical ordering, by [Jiang and Liu 2015a, Proposition 6.4; Ginzburg et al. 2003, Lemma 1.1], $\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m-1+r}}(\xi)$ is identically zero, and hence the corresponding term is zero. When $r-k=la$, with $1\le l\le b$ and $1\le k\le r$, then by Lemma 4.1, after restricting to $\mathrm{Sp}_{2a(b-l)+2m}(\mathbb{A})$, $\mathcal{C}_{N^{2ab+2m}_{r-k}}(\rho(\hat{\gamma}\lambda\beta)\xi)$ becomes a form in $\mathcal{E}_{\Delta(\tau,b-l)\otimes\sigma}$ whose Arthur parameter is

$$\psi'=(\tau,2b-2l+1)\boxplus\boxplus_{i=2}^{r}(\tau_i,1).$$

Since $[(2m+2k)1^{2a(b-l)-2k}]$ is bigger than $\eta_{\mathfrak{so}_{2n'+1}(\mathbb{C}),\mathfrak{sp}_{2n'}(\mathbb{C})}(\underline{p}(\psi'))$ under the lexicographical ordering, where $2n'=2a(b-l)+2m$, by [Jiang and Liu 2015a, Proposition 6.4; Ginzburg et al. 2003, Lemma 1.1], it follows that

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m-1+k}}\big(\mathcal{C}_{N^{2ab+2m}_{r-k}}(\rho(\hat{\gamma}\lambda\beta)\xi)\big)$$

is also identically zero, and hence the corresponding term is also zero. Therefore, the only possibilities that

$$\mathcal{C}_{N^{2ab}_{r}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)\big)\neq 0$$

are $r=la$ with $1\le l\le b$, and $k=0$. To prove that $\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)$ is not identically zero, we just have to show that

$$\mathcal{C}_{N^{2ab}_{r}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)\big)\neq 0\quad\text{for some } r.$$

Let $r=ab$; then

$$\mathcal{C}_{N^{2ab}_{ab}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)\big)=\int_{L(\mathbb{A})}\phi_1(i(\lambda))\,\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m-1}}\big(\mathcal{C}_{N^{2ab+2m}_{ab}}(\xi)\big)(\lambda\beta)\,d\lambda. \tag{5-2}$$

By Lemma 4.1, when restricted to $\mathrm{GL}_{2ab}(\mathbb{A})\times\mathrm{Sp}_{2m}(\mathbb{A})$,

$$\mathcal{C}_{N^{2ab+2m}_{ab}}(\xi)\in\delta^{1/2}_{P^{2ab+2m}_{ab}}|\det|^{-\frac{b+1}{2}}\Delta(\tau,b)\otimes\sigma.$$

Clearly, the integral in (5-2) is not identically zero if and only if σ is ψ^{α}-generic. By assumption, σ is ψ^{α}-generic, and hence

$$\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)$$

is not identically zero. Therefore, $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2m)1^{2ab}]$ with respect to the character $\psi_{[(2m)1^{2ab}],\alpha}$. This completes the proof of step (1).

5B. ***Proof of step (2).*** The proof of irreducibility of $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ is similar to that of $\tilde{\mathcal{E}}_{\Delta(\tau,1)}$ which is given in the proof of [Ginzburg et al. 2011, Theorem 2.1]. To show the square-integrable residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ is irreducible, it suffices to show that at each local place v,

$$\mathrm{Ind}^{\widetilde{\mathrm{Sp}}_{2ab}(F_v)}_{\tilde{P}_{ab}(F_v)}\,\mu_{\psi_v^{-\alpha}}\Delta(\tau_v,b)|\cdot|^{\frac{b}{2}} \tag{5-3}$$

has a unique irreducible quotient, where we assume that $\psi \cong \otimes_v \psi_v$, P_{ab} is the parabolic subgroup of Sp_{2ab} with Levi subgroup isomorphic to GL_{ab}, and $\tilde{P}_{ab}(F_v)$ is the preimage of $P_{ab}(F_v)$ in $\widetilde{\mathrm{Sp}}_{2ab}(F_v)$. Note that $\Delta(\tau_v, b)$ is the unique irreducible quotient of the following induced representation

$$\mathrm{Ind}_{Q_{a^b}(F_v)}^{\mathrm{GL}_{ab}(F_v)} \tau_v|\cdot|^{\frac{b-1}{2}} \otimes \tau_v|\cdot|^{\frac{b-3}{2}} \otimes \cdots \otimes \tau_v|\cdot|^{\frac{1-b}{2}},$$

where Q_{a^b} is the parabolic subgroup of GL_{ab} with Levi subgroup isomorphic to $\mathrm{GL}_a^{\times b}$. Let P_{a^b} be the parabolic subgroup of Sp_{2ab} with Levi subgroup isomorphic to $\mathrm{GL}_a^{\times b}$, and $\tilde{P}_{a^b}(F_v)$ is the preimage of $P_{a^b}(F_v)$ in $\widetilde{\mathrm{Sp}}_{2ab}(F_v)$. We just have to show that the following induced representation has a unique irreducible quotient

$$\mathrm{Ind}_{\tilde{P}_{a^b}(F_v)}^{\widetilde{\mathrm{Sp}}_{2ab}(F_v)} \mu_{\psi_v^{-\alpha}} \tau_v|\cdot|^{\frac{2b-1}{2}} \otimes \tau_v|\cdot|^{\frac{2b-3}{2}} \otimes \cdots \otimes \tau_v|\cdot|^{\frac{1}{2}}. \tag{5-4}$$

Since τ_v is generic and unitary, by [Tadić 1986; Vogan 1986], τ_v is fully parabolic, induced from its Langlands data with exponents in the open interval $\left(-\frac{1}{2}, \frac{1}{2}\right)$. Explicitly, we can assume that

$$\tau_v \cong \rho_1|\cdot|^{\alpha_1} \times \rho_2|\cdot|^{\alpha_2} \times \cdots \times \rho_r|\cdot|^{\alpha_r},$$

where the ρ_i are tempered representations, $\alpha_i \in \mathbb{R}$, and $\frac{1}{2} > \alpha_1 > \alpha_2 > \cdots > \alpha_r > -\frac{1}{2}$. Therefore, the induced representation in (5-4) can be written as

$$\begin{aligned}\mu_{\psi_v^{-\alpha}} \rho_1|\cdot|^{\frac{2b-1}{2}+\alpha_1} \times \rho_2|\cdot|^{\frac{2b-1}{2}+\alpha_2} \times \cdots \times \rho_r|\cdot|^{\frac{2b-1}{2}+\alpha_r} \\ \times \rho_1|\cdot|^{\frac{2b-3}{2}+\alpha_1} \times \rho_2|\cdot|^{\frac{2b-3}{2}+\alpha_2} \times \cdots \times \rho_r|\cdot|^{\frac{2b-3}{2}+\alpha_r} \\ \times \cdots \times \rho_1|\cdot|^{\frac{1}{2}+\alpha_1} \times \rho_2|\cdot|^{\frac{1}{2}+\alpha_2} \times \cdots \times \rho_r|\cdot|^{\frac{1}{2}+\alpha_r} \rtimes 1_{\widetilde{\mathrm{Sp}}_0(F_v)}.\end{aligned}$$

Since $\alpha_i \in \mathbb{R}$ and $\frac{1}{2} > \alpha_1 > \alpha_2 > \cdots > \alpha_r > -\frac{1}{2}$, we can easily see that the exponents satisfy

$$\begin{aligned}\frac{2b-1}{2}+\alpha_1 > \frac{2b-1}{2}+\alpha_2 > \cdots > \frac{2b-1}{2}+\alpha_r \\ > \frac{2b-3}{2}+\alpha_1 > \frac{2b-3}{2}+\alpha_2 > \cdots > \frac{2b-3}{2}+\alpha_r \\ > \cdots > \frac{1}{2}+\alpha_1 > \frac{1}{2}+\alpha_2 > \cdots > \frac{1}{2}+\alpha_r > 0.\end{aligned}$$

By Langlands classification of metaplectic groups (see [Borel and Wallach 2000; Ban and Jantzen 2013]), one can see that the induced representation in (5-4) has a unique irreducible quotient which is the Langlands quotient. This completes the proof of irreducibility of $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$.

To prove the square-integrability of $\mathcal{D}_{2m,\psi^\alpha}^{2ab+2m}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$, we need to calculate the automorphic exponent attached to the nontrivial constant term considered in step (1); $r = ab$, and for definition of automorphic exponent see [Mœglin and Waldspurger 1995, I.3.3]. For this, we need to consider the action of

$$\overline{g} = \mathrm{diag}(g, g^*) \in \mathrm{GL}_{ab}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_0(\mathbb{A}).$$

Since $r = ab$,

$$\beta = \begin{pmatrix} 0 & I_{ab} \\ I_m & 0 \end{pmatrix}^{\wedge}.$$

Let

$$\tilde{g} := \beta\, \mathrm{diag}(I_m, \bar{g}, I_m)\beta^{-1} = \mathrm{diag}(g, I_{2m}, g^*).$$

Then changing variables in (5-2) via $\lambda \mapsto \tilde{g}\lambda\tilde{g}^{-1}$ will give a Jacobian $|\det g|^{-m}$. On the other hand, by [Ginzburg et al. 2011, Formula (1.4)], the action of $\bar{g}$ on ϕ_1 gives $\gamma_{\psi^{-\alpha}}(\det g)|\det g|^{1/2}$. Therefore, $\bar{g}$ acts by $\Delta(\tau, b)(g)$ with character

$$\delta^{1/2}_{P^{2ab+2m}_{ab}}(\tilde{g})\,|\det g|^{-\frac{b+1}{2}}\,|\det g|^{-m}\,\gamma_{\psi^{-\alpha}}(\det g)|\det g|^{\frac{1}{2}} = \gamma_{\psi^{-\alpha}}(\det g)\,\delta^{1/2}_{P^{2ab}_{ab}}(\bar{g})\,|\det g|^{-\frac{b}{2}}.$$

Therefore, as a function on $\mathrm{GL}_{ab}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_0(\mathbb{A})$,

$$\mathcal{C}_{N^{2ab}_{ab}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)\big) \in \gamma_{\psi^{-\alpha}}\delta^{1/2}_{P^{2ab}_{ab}}|\det(\cdot)|^{-\frac{b}{2}}\,\Delta(\tau, b) \otimes 1_{\widetilde{\mathrm{Sp}}_0(\mathbb{A})}. \tag{5-5}$$

Since, the cuspidal exponent of $\Delta(\tau, b)$ is

$$\Big\{\Big(\frac{1-b}{2}, \frac{3-b}{2}, \dots, \frac{b-1}{2}\Big)\Big\},$$

the cuspidal exponent of $\mathcal{C}_{N^{2ab}_{ab}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m-1}}(\xi)\big)$ is

$$\Big\{\Big(\frac{1-2b}{2}, \frac{3-2b}{2}, \dots, -\frac{1}{2}\Big)\Big\}.$$

Hence, by Langlands square-integrability criterion [Mœglin and Waldspurger 1995, Lemma I.4.11], the automorphic representation $\mathcal{D}^{2ab+2m}_{2m,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$ is square-integrable.

From (5-5), it is easy to see that as a representation of $\mathrm{GL}_{ab}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_0(\mathbb{A})$,

$$\mathcal{C}_{N^{2ab}_{ab}}\big(\mathcal{D}^{2ab+2m}_{2m,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})\big) = \gamma_{\psi^{-\alpha}}\delta^{1/2}_{P^{2ab}_{ab}}|\det(\cdot)|^{-\frac{b}{2}}\,\Delta(\tau, b) \otimes 1_{\widetilde{\mathrm{Sp}}_0(\mathbb{A})}. \tag{5-6}$$

From the cuspidal support of the Speh residual representation $\Delta(\tau, b)$ of $\mathrm{GL}_{ab}(\mathbb{A})$, one can now easily see that

$$\mathcal{C}_{N^{2ab}_{a^b}}\big(\mathcal{D}^{2ab+2m}_{2m,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})\big) = \gamma_{\psi^{-\alpha}}\delta^{1/2}_{P^{2ab}_{a^b}}\tau|\cdot|^{\frac{1-2b}{2}} \otimes \tau|\cdot|^{\frac{3-2b}{2}} \otimes \cdots \otimes \tau|\cdot|^{-\frac{1}{2}} \otimes 1_{\widetilde{\mathrm{Sp}}_0(\mathbb{A})},$$

where $N^{2ab}_{a^b}$ is the unipotent radical of the parabolic subgroup $P^{2ab}_{a^b}$ with Levi isomorphic to $\mathrm{GL}_a^{\times b}$. By [op. cit., Corollary 3.14(ii)], any noncuspidal irreducible summand of $\mathcal{D}^{2ab+2m}_{2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$ must be contained in the space $\tilde{\mathcal{E}}_{\tau^{\otimes b},\Lambda}$, which is the residual representation generated by residues of the Eisenstein series associated to the induced representation

$$\mathrm{Ind}^{\widetilde{\mathrm{Sp}}_{2ab}(\mathbb{A})}_{\tilde{P}^{2ab}_{a^b}(\mathbb{A})}\gamma_{\psi^{-\alpha}}\tau|\cdot|^{s_1}\otimes\tau|\cdot|^{s_2}\otimes\cdots\otimes\tau|\cdot|^{s_b},$$

at the point

$$\Lambda=\left\{\frac{1-2b}{2},\frac{3-2b}{2},\dots,\frac{-1}{2}\right\}.$$

Since the Speh residual representation $\Delta(\tau,b)$ of $\mathrm{GL}_{ab}(\mathbb{A})$ is irreducible, by taking residues in stages, one can easily see that the space of the residual representation $\tilde{\mathcal{E}}_{\tau^{\otimes b},\Lambda}$ is exactly identical to that of $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$. Therefore, any noncuspidal irreducible summand of $\mathcal{D}^{2ab+2m}_{2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$ must be contained in the space $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$. Hence, the descent representation $\mathcal{D}^{2ab+2m}_{2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$ has a nontrivial intersection with the space of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$. Since we have seen that $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ is irreducible, $\mathcal{D}^{2ab+2m}_{2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$ must contain the whole space of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$. This completes the proof of step (2).

5C. ***Proof of step (3).*** The proof of the fact that $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ has a nonzero Fourier coefficient attached to the symplectic partition $[(a)^{2b}]$ is very similar to the proof of [Liu 2013a, Theorem 4.2.2], if a is even. The idea is to apply Lemma 3.2 repeatedly and use induction on b. Note that the case of $\tilde{\mathcal{E}}_{\Delta(\tau,1)}$ has already been proved in [Ginzburg et al. 2011, Theorem 8.1]. We omit the details here for this case.

In the following, we assume that $a=2k+1$ and prove $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ has a nonzero Fourier coefficient attached to the symplectic partition $\underline{p}:=[(2k+1)^{2b}]$ by induction on b. When $b=1$, it has been proved in [op. cit., Theorem 8.2], we will use similar idea here. Assume that $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)}$ has a nonzero Fourier coefficient attached to the symplectic partition $[(2k+1)^{2b-2}]$.

Take any $\varphi\in\tilde{\mathcal{E}}_{\Delta(\tau,b)}$; its Fourier coefficients attached to $\underline{p}$ are of the following form

$$\varphi^{\psi_{\underline{p}}}(g)=\int_{[V_{\underline{p},2}]}\varphi(vg)\psi_{\underline{p}}^{-1}(v)\,dv. \tag{5-7}$$

For definitions of the unipotent group $V_{\underline{p},2}$ and its character $\psi_{\underline{p}}$, see [Jiang and Liu 2015c, Section 2].

Note that the one-dimensional torus $\mathcal{H}_{\underline{p}}$ defined in [op. cit, (2.1)] has elements of the form

$$\mathcal{H}_{\underline{p}}(t)=\mathrm{diag}(A(t),A(t),\dots,A(t)),\quad \text{where } A(t)=\mathrm{diag}(t^{2k},t^{2k-2},\dots,t^{-2k}),$$

and there are $2b$ copies of $A(t)$. Also note that the group $L_{\underline{p}}(\mathbb{A})$ defined in [op. cit, Section 2] is isomorphic to $\mathrm{GL}^{2k+1}_{2b}(\mathbb{A})$, and the stabilizer of the character $\psi_{\underline{p}}$

in $L_{\underline{p}}$ is isomorphic to the diagonal embedding $\widetilde{\mathrm{Sp}}_{2b}^{\Delta}(\mathbb{A})$. Let ι be this diagonal embedding. Let

$$N = \left\{ n(x) := \begin{pmatrix} 1 & 0 & x \\ 0 & I_{2b-2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}.$$

Then

$$\iota(N) = \left\{ \iota(n(x)) = \begin{pmatrix} I_{2k+1} & 0 & xI_{2k+1} \\ 0 & I_{(2k+1)(2b-2)} & 0 \\ 0 & 0 & I_{2k+1} \end{pmatrix} \right\}. \tag{5-8}$$

To show the integral in (5-7) is nonvanishing, it suffices to show that the following integral is nonvanishing:

$$\int_{F\backslash\mathbb{A}} \int_{[V_{\underline{p},2}]} \varphi(vn(x)g)\,\psi_{\underline{p}}^{-1}(v)\,dv\,dx. \tag{5-9}$$

Let ω be a Weyl element which sends $\mathcal{H}_{\underline{p}}(t)$ to the torus element

$$\mathrm{diag}\big(A(t), t^{2k} I_{2b-2}, t^{2k-2} I_{2b-2}, \dots, t^{-2k} I_{2b-2}, A(t)\big).$$

Then ω has the form $\mathrm{diag}(I_{2k+1}, \omega_1, I_{2k+1})$. Conjugating from left by ω, the integral in (5-9) becomes

$$\int_{[W]} \varphi(w\omega g)\,\psi_W^{-1}(w)\,dw, \tag{5-10}$$

where $W = \omega V_{\underline{p},2}\iota(N)\omega^{-1}$ and $\psi_W(w) = \psi_{\underline{p}}(\omega^{-1}w\omega)$. Then elements of W have the form

$$w = \begin{pmatrix} z_{2k+1} & q_1 & q_2 \\ 0 & w' & q_1^* \\ 0 & 0 & z_{2k+1}^* \end{pmatrix} \begin{pmatrix} I_{2k+1} & 0 & 0 \\ p_1 & I_{(2b-2)(2k+1)} & 0 \\ p_2 & p_1^* & I_{2k+1} \end{pmatrix}, \tag{5-11}$$

where $z_{2k+1} \in V_{2k+1}$, the standard maximal unipotent subgroup of GL_{2k+1}; $w' \in \omega_1 V_{[(2k+1)^{2b-2}],2}\omega_1^{-1}$; $q_1 \in M_{(2k+1)\times((2b-2)(2k+1))}$ with certain conditions; $p_1 \in M_{((2b-2)(2m+1))\times(2m+1)}$ with certain conditions; $q_2 \in M_{(2k+1)\times(2k+1)}$, symmetric with respect to the secondary diagonal, such that $q_2(i,j) = 0$ for $1 \le j < i \le 2k+1$, and $q_2(1,1) = q_2(2,2) = \cdots = q_2(2k+1, 2k+1)$; $p_2 \in M_{(2k+1)\times(2k+1)}$, symmetric with respect to the secondary diagonal, such that $p_2(i,j) = 0$ for $1 \le j \le i \le 2k+1$.

Next, as in the proof of Section 4, we apply Lemma 3.2 to fill the zero entries in q_1, q_2 using the nonzero entries in p_1, p_2. Similarly, to proceed, we need to define a sequence of one-dimensional root subgroups and put them in a correct order:

$$\alpha_j^i = \begin{cases} e_i + e_{2k+1-i+j} \\ \quad \text{if } 1 \le i \le k \text{ and } 1 \le j \le i, \\ e_i - e_{(2k+1)+(2b-2)i-(j-1)} \\ \quad \text{if } 1 \le i \le k \text{ and } i+1 \le j \le i+(2b-2)i, \\ e_i + e_{i+j} \\ \quad \text{if } k+1 \le i \le 2k \text{ and } 1 \le j \le 2k+1-i, \\ ve_i + e_{(2k+1)b-(b-1)-(2b-2)(i-k-1)+j} \\ \quad \text{if } k+1 \le i \le 2k \text{ and } (2k+1-i)+1 \le j \\ \qquad\qquad \le (2k+1-i)+((b-1)+(2b-2)(i-k-1)), \\ e_i - e_{(2k+1)b-(j-1)} \\ \quad \text{if } k+1 \le i \le 2k \text{ and } (2k+1-i)+((b-1)+(2b-2)(i-k-1))+1 \le j \\ \qquad\qquad \le (2k+1-i)+(2b-2)i. \end{cases}$$

For the above roots α_j^i, let $X_{\alpha_j^i}$ be the corresponding one-dimensional root subgroup.

For $1 \le i \le k$ and $1 \le j \le i$, let $\beta_j^i = -e_{2k+1-i+j} - e_{i+1}$. For $1 \le i \le k$ and $i+1 \le j \le i+(2b-2)i$, let $\beta_j^i = e_{(2k+1)+(2b-2)i-(j-1)} - e_{i+1}$. For $k+1 \le i \le 2k$ and $1 \le j \le 2k+1-i$, let $\beta_j^i = -e_{i+j} - e_{i+1}$. For $k+1 \le i \le 2k$ and $(2k+1-i)+1 \le j \le (2k+1-i)+((b-1)+(2b-2)(i-k-1))$, let $\beta_j^i = -e_{(2k+1)b-(b-1)-(2b-2)(i-k-1)+j} - e_{i+1}$. Finally, for $k+1 \le i \le 2k$ and

$$(2k+1-i)+((b-1)+(2b-2)(i-k-1))+1 \le j \le (2k+1-i)+(2b-2)i,$$

let $\beta_j^i = e_{(2k+1)b-(j-1)} - e_{i+1}$. For the above roots β_j^i, let $X_{\beta_j^i}$ be the corresponding one-dimensional root subgroup.

Let

$$m_i = \begin{cases} i+(2b-2)i & \text{if } 1 \le i \le k, \\ (2k+1-i)+(2b-2)i & \text{if } k+1 \le i \le 2k. \end{cases}$$

Let $\widetilde{W}$ be the subgroup of W with elements of the form as in (5-11), but with the p_1 and p_2 parts zero. Let $\psi_{\widetilde{W}} = \psi_W|_{\widetilde{W}}$. For any subgroup of W containing $\widetilde{W}$, we automatically extend $\psi_{\widetilde{W}}$ trivially to this subgroup and still denote the character by $\psi_{\widetilde{W}}$.

Next, we will apply Lemma 3.2 to a sequence of quadruples. For any i such that $1 \le i \le k+1$, one can see that the following quadruple satisfies all the conditions for Lemma 3.2:

$$\big(\widetilde{W}_i, \psi_{\widetilde{W}}, \{X_{\alpha_j^i}\}_{j=1}^{m_i}, \{X_{\beta_j^i}\}_{j=1}^{m_i}\big),$$

where

$$\widetilde{W}_i = \prod_{s=1}^{i-1} \prod_{j=1}^{m_s} X_{\alpha_j^s} \widetilde{W} \prod_{l=i+1}^{2k} \prod_{j=1}^{m_l} X_{\beta_j^l}.$$

Applying Lemma 3.2, one can see that the integral in (5-10) is nonvanishing if and only if the following integral is nonvanishing:

$$\int_{[\widetilde{W}_i']} \varphi(w\omega g)\,\psi^{-1}_{\widetilde{W}_i'}(w)\,dw, \tag{5-12}$$

where

$$\widetilde{W}_i' = \prod_{s=1}^{i}\prod_{j=1}^{m_s} X_{\alpha_j^s}\,\widetilde{W}\prod_{l=i+1}^{2k}\prod_{j=1}^{m_l} X_{\beta_j^l}, \tag{5-13}$$

and $\psi_{\widetilde{W}_i'}$ is extended from $\psi_{\widetilde{W}}$ trivially.

For any i such that $k+2 \le i \le 2k$, before applying Lemma 3.2 repeatedly to certain sequence of quadruples as above, we need to take the Fourier expansion of the resulting integral at the end of the step $i-1$ along $X_{e_i+e_i}$ (at the end of step $k+1$, one gets the integral in (5-12) with $i = k+1$ there, at the end of step s, $k+2 \le s \le 2k-1$, one would get the integral in (5-14)). Under the action of GL_1, we get two kinds of Fourier coefficients corresponding to the two orbits of the dual of $[X_{e_i+e_i}]$: the trivial one and the nontrivial one. It turns out that any Fourier coefficient corresponding to the nontrivial orbit contains an inner integral which is exactly the Fourier coefficients attached to the partition $[(2i)1^{(2k+1)(2b)-2i}]$, which is identically zero by [Jiang and Liu 2015a, Proposition 6.4], since $i \ge k+2$. Therefore only the Fourier coefficient attached to the trivial orbit survives.

After taking Fourier expansion of the resulting integral at the end of step $i-1$ along $X_{e_i+e_i}$ as above, one can see that the following quadruple satisfies all the conditions for Lemma 3.2:

$$\big(X_{e_i+e_i}\widetilde{W}_i,\ \psi_{\widetilde{W}},\ \{X_{\alpha_j^i}\}_{j=1}^{m_i},\ \{X_{\beta_1^i}\}_{j=1}^{m_i}\big),$$

where

$$\widetilde{W}_i = \prod_{s=1}^{i-1}\prod_{j=1}^{m_s} X_{\alpha_j^s}\prod_{t=k+2}^{i-1} X_{e_t+e_t}\,\widetilde{W}\prod_{l=i+1}^{2k}\prod_{j=1}^{m_l} X_{\beta_j^l}.$$

Applying Lemma 3.2, we can see that the resulting integral at the end of step $i-1$ is nonvanishing if and only if the following integral is nonvanishing:

$$\int_{[\widetilde{W}_i']} \varphi(w\omega g)\,\psi^{-1}_{\widetilde{W}_i'}(w)\,dw, \tag{5-14}$$

where

$$\widetilde{W}_i' = \prod_{s=1}^{i}\prod_{j=1}^{m_s} X_{\alpha_j^s}\prod_{t=k+2}^{i} X_{e_t+e_t}\,\widetilde{W}\prod_{l=i+1}^{2k}\prod_{j=1}^{m_l} X_{\beta_j^l}, \tag{5-15}$$

and $\psi_{\widetilde{W}_i'}$ is the trivial extension of $\psi_{\widetilde{W}}$.

One can see that elements of $\widetilde{W}'_{2k}$ have the following form:

$$w = \begin{pmatrix} z_{2k+1} & q_1 & q_2 \\ 0 & w' & q_1^* \\ 0 & 0 & z_{2k+1}^* \end{pmatrix}, \tag{5-16}$$

where $z_{2k+1} \in V_{2k+1}$, which is the standard maximal unipotent subgroup of GL_{2k+1};

$$w' \in \omega_1 V_{[(2k+1)^{2b-2}],2}\,\omega_1^{-1};$$

$q_1 \in \mathrm{Mat}_{(2k+1)\times(2k+1)(2b-2)}$ with $q_1(2k+1, j) = 0$ for $1 \le j \le (2k+1)(2b-2)$; $q_2 \in \mathrm{Mat}_{(2k+1)\times(2k+1)}$, symmetric with respect to the secondary diagonal, with $q_2(2k+1,1) = 0$. For $w \in \widetilde{W}'_{2k}$ of form in (5-16),

$$\psi_{\widetilde{W}'_{2k}}(w) = \psi\left(\sum_{i=1}^{2k} z_{i,i+1}\right)\psi_{[(2k+1)^{2b-2}]}(\omega_1^{-1} w' \omega_1).$$

Now consider the Fourier expansion of the integral in (5-14) along the one-dimensional root subgroup $X_{2e_{2k+1}}$. By the same reason as above, only the Fourier coefficient corresponding to the trivial orbit of the dual of $[X_{2e_{2k+1}}]$ survives, which is actually equal to the integral in (5-14) (with $i = 2k$ there):

$$\int_{[W_{2k+1}]} \varphi(w\omega g)\,\psi_{W_{2k+1}}(w)^{-1}\,dw, \tag{5-17}$$

where elements in W_{2k+1} have the same structure as in (5-16), except that the element $q_2(2k+1,1)$ is not identically zero.

One can see that the integral in (5-17) has an inner integral which is exactly $\varphi^{\psi_{N_{12k}}}$, using notation in Lemma 5.2 below. On the other hand, we know that by Lemma 5.2 below, $\varphi^{\psi_{N_{12k}}} = \varphi^{\widetilde{\psi}_{N_{12k+1}}}$. Therefore, the integral in (5-17) becomes

$$\int_{[W'_{2k+1}]} \varphi(w\omega g)\,\psi_{W'_{2k+1}}(w)^{-1}\,dw, \tag{5-18}$$

where any element in W'_{2k+1} has the following form:

$$w = w(z_{2k+1}, w', q_1, q_2) = \begin{pmatrix} z_{2k+1} & q_1 & q_2 \\ 0 & w' & q_1^* \\ 0 & 0 & z_{2k+1}^* \end{pmatrix},$$

where $z_{2k+1} \in V_{2k+1}$, the standard maximal unipotent subgroup of GL_{2k+1}; $w' \in \omega_1 V_{[(2k+1)^{2b-2}],2}\omega_1^{-1}$; $q_1 \in \mathrm{Mat}_{(2k+1)\times(2k+1)(2b-2)}$; $q_2 \in \mathrm{Mat}_{(2k+1)\times(2k+1)}$, symmetric with respect to the secondary diagonal. For $w \in W'_{2k+1}$ as above,

$$\psi_{W'_{2k+1}}(w) = \psi\left(\sum_{i=1}^{2k} z_{i,i+1}\right)\psi_{[(2k+1)^{2b-2}]}(\omega_1^{-1} w' \omega_1).$$

Hence, the integral in (5-18) can be written as

$$\int_{W''_{2k+1}} \varphi_{P_{2k+1}}(w\omega g)\,\psi_{W''_{2k+1}}(w)^{-1}\,dw, \tag{5-19}$$

where W''_{2k+1} is a subgroup of W'_{2k+1} consisting only of elements of the form $w(z_{2k+1}, w', 0, 0)$,

$$\psi_{W''_{2k+1}} = \psi_{W'_{2k+1}}\big|_{W''_{2k+1}},$$

and $\varphi_{P_{2m+1}}$ is the constant term of φ along the parabolic subgroup $\tilde{P}_{2k+1}(\mathbb{A}) = \widetilde{M}_{2k+1}(\mathbb{A})N_{2k+1}(\mathbb{A})$ of $\widetilde{\mathrm{Sp}}_{2b(2k+1)}(\mathbb{A})$ with the Levi subgroup isomorphic to $\mathrm{GL}_{2k+1}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_{(2b-2)(2k+1)}(\mathbb{A})$.

By Lemma 5.1 below, $\varphi(w\omega g)_{\tilde{P}_{2k+1}(\mathbb{A})}$ is an automorphic form in

$$\gamma_{\psi^{-\alpha}}\,\tau|\cdot|^{-\frac{2b-1}{2}} \otimes \tilde{\mathcal{E}}_{\Delta(\tau,b-1)}$$

when restricted to the Levi subgroup. Note that the restriction of $\psi_{W'_{2k+1}}$ to the z_{2k+1}-part gives us a Whittaker coefficient of τ, and the restriction to the w'-part gives a Fourier coefficient of $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)}$ attached to the partition $[(2k+1)^{2b-2}]$, up to the conjugation of the Weyl element ω_1. On the other hand, τ is generic, and by induction assumption, $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)}$ has a nonzero Fourier coefficient attached to the partition $[(2k+1)^{2b-2}]$. Therefore, we can conclude that $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ has a nonzero $\psi_{\underline{p}}$-Fourier coefficient attached to the partition $[(2k+1)^{2b}]$. This completes the proof of step (3), up to Lemmas 5.1 and 5.2, which are stated below.

We remark that as Lemmas 4.1 and 4.2, Lemmas 5.1 and 5.2 below are also analogues of [Liu 2013a, Lemmas 4.2.4 and 4.2.6], with similar arguments, and hence we again only state them without proofs.

Lemma 5.1. *Let $\tilde{P}_{(2k+1)i}(\mathbb{A}) = \widetilde{M}_{(2k+1)i}(\mathbb{A})N_{(2k+1)i}(\mathbb{A})$, with $1 \le i \le b$, be the parabolic subgroup of $\widetilde{\mathrm{Sp}}_{2b(2k+1)}(\mathbb{A})$ with Levi part*

$$\widetilde{M}_{(2k+1)i}(\mathbb{A}) \cong \mathrm{GL}_{(2k+1)i}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_{(2k+1)(2b-2)}(\mathbb{A}).$$

Let φ be an arbitrary automorphic form in $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$. Denote by $\varphi_{P_{(2k+1)i}}$ the constant term of φ along $P_{(2k+1)i}$. Then, for $1 \le i \le b$, $\varphi_{P_{(2k+1)i}}$ belongs to

$$\mathcal{A}\big(N_{(2k+1)i}(\mathbb{A})\,\widetilde{M}_{(2k+1)i}(F) \setminus \widetilde{\mathrm{Sp}}_{2b(2k+1)}(\mathbb{A})\big)_{\gamma_{\psi^{-\alpha}}\Delta(\tau,i)|\cdot|^{-(2b-i)/2}\otimes\tilde{\mathcal{E}}_{\Delta(\tau,b-i)}}.$$

Lemma 5.2. *Let $N_{1^p}(\mathbb{A})$ be the unipotent radical of the parabolic subgroup $\tilde{P}_{1^p}(\mathbb{A})$ of $\widetilde{\mathrm{Sp}}_{2b(2k+1)}(\mathbb{A})$ with Levi part isomorphic to*

$$\mathrm{GL}_1^{\times p}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_{2b(2k+1)-2p}(\mathbb{A}).$$

Let

$$\psi_{N_{1^p}}(n) := \psi(n_{1,2} + \cdots + n_{p,p+1}) \quad \text{and} \quad \tilde{\psi}_{N_{1^p}}(n) := \psi(n_{1,2} + \cdots + n_{p-1,p}),$$

be two characters of $N_{1^p}(\mathbb{A})$. *For any automorphic form* $\varphi \in \tilde{\mathcal{E}}_{\Delta(\tau,b)}$, *define* $\psi_{N_{1^p}}$ *and* $\tilde{\psi}_{N_{1^p}}$-*Fourier coefficients by*:

$$\varphi^{\psi_{N_{1^p}}}(g) := \int_{[N_{1^p}]} \varphi(ng)\,\psi_{N_{1^p}}(n)^{-1}\,dn, \tag{5-20}$$

$$\varphi^{\tilde{\psi}_{N_{1^p}}}(g) := \int_{[N_{1^p}]} \varphi(ng)\,\tilde{\psi}_{N_{1^p}}(n)^{-1}\,du. \tag{5-21}$$

Then $\varphi^{\psi_{N_{1^p}}} \equiv 0$ *for all* $p \geq 2k+1$, *and* $\varphi^{\psi_{N_{1^{2k}}}} = \varphi^{\tilde{\psi}_{N_{1^{2k+1}}}}$.

6. Proof of Theorem 2.6

In this section, we prove that $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$ has a nonzero Fourier coefficient attached to $[(2m+2a)(a)^{2b}]$. Assume that σ is ψ^α-generic with $\alpha \in F^*/(F^*)^2$.

As in the proof of part (2) of Theorem 2.4 in Section 5 we separate the proof of Theorem 2.6 into two steps:

***Step* (1)** $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2m+2a)1^{2ab}]$ with respect to the character $\psi_{[(2m+2a)1^{2ab}],\alpha}$ (for the definition, see [Jiang and Liu 2015c, Section 2]).

***Step* (2)** Let

$$\mathcal{D}^{2a(b+1)+2m}_{2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma})$$

be the ψ^α-descent from the representation $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$ of $\mathrm{Sp}_{2a(b+1)+2m}(\mathbb{A})$ to a representation of $\widetilde{\mathrm{Sp}}_{2ab}(\mathbb{A})$. Then it is square-integrable and contains the whole space of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ which is irreducible and constructed at the beginning of Section 5.

Proof of Theorem 2.6. First, recall from the step (3) in the proof of part (2) of Theorem 2.4 that $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ has a nonzero Fourier coefficient attached to the symplectic partition $[(a)^{2b}]$. From the results in steps (1) and (2) above, we can see that $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the composite partition $[(2m+2a)1^{2ab}]\circ[(a)^{2b}]$ (for the definition of composite partitions and the attached Fourier coefficients, we refer to [Ginzburg et al. 2003, Section 1]). Therefore, by [Jiang and Liu 2015c, Lemma 3.1] or [Ginzburg et al. 2003, Lemma 2.6], $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$ has a nonzero Fourier coefficient attached to $[(2m+2a)(a)^{2b}]$. $\square$

Before proving the above two steps, we record the following lemma which is analogous to Lemma 4.1, whose proof will be omitted.

Lemma 6.1. *Let* $P_{ai} = M_{ai}N_{ai}$, *with* $1 \leq i \leq b+1$, *be the parabolic subgroup of* $\mathrm{Sp}_{2a(b+1)+2m}$ *whose Levi part* $M_{ai} \cong \mathrm{GL}_{ai} \times \mathrm{Sp}_{a(2b+2-2i)+2m}$. *Let* φ *be an arbitrary automorphic form in* $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$. *Denote by* $\varphi_{P_{ai}}(g)$ *the constant term* *of φ along P_{ai}. Then, for $1 \le i \le b+1$,*

$$\varphi_{P_{ai}} \in \mathcal{A}\big(N_{ai}(\mathbb{A})\,M_{ai}(F)\setminus \mathrm{Sp}_{2a(b+1)+2m}(\mathbb{A})\big)_{\Delta(\tau,i)|\cdot|^{-(2b+1-i)/2}\otimes\mathcal{E}_{\Delta(\tau,b+1-i)\otimes\sigma}}.$$

Note that when $i=b$, $\mathcal{E}_{\Delta(\tau,b+1-i)\otimes\sigma}=\mathcal{E}_{\tau\otimes\sigma}$, which is not a residual representation as explained at the end of Section 2B, is nonzero and generic by [Shahidi 2010, Theorem 7.1.3]; *and when $i=b+1$, $\mathcal{E}_{\Delta(\tau,b+1-i)\otimes\sigma}=\sigma$.*

6A. ***Proof of step (1).*** By [Ginzburg et al. 2003, Lemma 1.1], $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2m+2a)1^{2ab}]$ with respect to the character $\psi_{[(2m+2a)1^{2ab}],\alpha}$ if and only if the ψ^{α}-descent

$$\mathcal{D}^{2a(b+1)+2m}_{2m+2a,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma})$$

of $\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$, which is a representation of $\widetilde{\mathrm{Sp}}_{2ab}(\mathbb{A})$, is not identically zero.

Take any $\xi\in\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}$, we will calculate the constant term of

$$\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)$$

along the parabolic subgroup $\tilde{P}^{2ab}_r(\mathbb{A})=\widetilde{M}^{2ab}_r(\mathbb{A})\,N^{2ab}_r(\mathbb{A})$ of $\widetilde{\mathrm{Sp}}_{2ab}(\mathbb{A})$ with Levi subgroup isomorphic to $\mathrm{GL}_r(\mathbb{A})\times\widetilde{\mathrm{Sp}}_{2ab-2r}(\mathbb{A})$, $1\le r\le ab$, which is denoted by

$$\mathcal{C}_{N^{2ab}_r}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)\big).$$

By [Ginzburg et al. 2011, Theorem 7.8],

$$\begin{aligned}\mathcal{C}_{N^{2ab}_r}&\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)\big)\\ &=\sum_{\substack{0\le k\le r\\ \gamma\in P^1_{r-k,1^k}(F)\backslash \mathrm{GL}_r(F)}}\int_{L(\mathbb{A})}\phi_1(i(\lambda))\,\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m+a-1+k}}\big(\mathcal{C}_{N^{2a(b+1)+2m}_{r-k}}(\xi)\big)(\hat{\gamma}\lambda\beta)\,d\lambda,\end{aligned}\tag{6-1}$$

The notation in (6-1) is explained in order: $N^{2a(b+1)+2m}_{r-k}$ is the unipotent radical of the parabolic subgroup $P^{2a(b+1)+2m}_{r-k}$ of $\mathrm{Sp}_{2a(b+1)+2m}$; $P^1_{r-k,1^k}$ is a subgroup of GL_r consisting of matrices of the form

$$\begin{pmatrix} g & x\\ 0 & z\end{pmatrix},$$

with $z\in U_k$, the standard maximal unipotent subgroup of GL_k. For $g\in\mathrm{GL}_j$, $j\le a(b+1)+m$, $\hat{g}=\mathrm{diag}(g, I_{2a(b+1)+2m-2j}, g^*)$; L is a unipotent subgroup, consisting of matrices of the form

$$\lambda=\begin{pmatrix} I_r & 0\\ x & I_{m+a}\end{pmatrix}^{\wedge},$$

and $i(\lambda)$ is the last row of x, and

$$\beta = \begin{pmatrix} 0 & I_r \\ I_{m+a} & 0 \end{pmatrix}^{\wedge}.$$

Finally, the Schwartz function $\phi = \phi_1 \otimes \phi_2$ with $\phi_1 \in \mathcal{S}(\mathbb{A}^r)$ and $\phi_2 \in \mathcal{S}(\mathbb{A}^{ab-r})$, and the function

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m+a-1+k}}\big(\mathcal{C}_{N^{2a(b+1)+2m}_{r-k}}(\xi)\big)(\hat{\gamma}\lambda\beta) \\ := \mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m+a-1+k}}\big(\mathcal{C}_{N^{2a(b+1)+2m}_{r-k}}(\rho(\hat{\gamma}\lambda\beta)\xi)\big)(I),$$

with $\rho(\hat{\gamma}\lambda\beta)$ denoting the right translation by $\hat{\gamma}\lambda\beta$, is a composition of the restriction to $\mathrm{Sp}_{2a(b+1)+2m-2r+2k}(\mathbb{A})$ of $\mathcal{C}_{N^{2a(b+1)+2m}_{r-k}}(\rho(\hat{\gamma}\lambda\beta)\xi)$ with the Fourier–Jacobi coefficient

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m+a-1+k}},$$

which takes automorphic forms on $\mathrm{Sp}_{2a(b+1)+2m-2r+2k}(\mathbb{A})$ to those forms on $\widetilde{\mathrm{Sp}}_{2ab-2r}(\mathbb{A})$.

By the cuspidal support of ξ,

$$\mathcal{C}_{N^{2a(b+1)+2m}_{r-k}}(\xi)$$

is identically zero, unless $k = r$ or $r - k = la$ with $1 \le l \le b+1$. When $k = r$, since $[(2m+2a+2r)1^{2ab-2r}]$ is bigger than $\eta_{\mathfrak{so}_{2n+1}(\mathbb{C}),\mathfrak{sp}_{2n}(\mathbb{C})}(\underline{p}(\psi))$ under the lexicographical ordering, by [Jiang and Liu 2015a, Proposition 6.4; Ginzburg et al. 2003, Lemma 1.1],

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m+a-1+r}}(\xi)$$

is identically zero, hence the corresponding term is zero. When $r - k = la$, $1 \le l \le b+1$ and $1 \le k \le r$, by Lemma 6.1, after restricting to $\mathrm{Sp}_{2a(b+1-l)+2m}(\mathbb{A})$, $\mathcal{C}_{N^{2a(b+1)+2m}_{r-k}}(\rho(\hat{\gamma}\lambda\beta)\xi)$ becomes a form in $\mathcal{E}_{\Delta(\tau,b+1-l)\otimes\sigma}$. Note that the Arthur parameter of $\mathcal{E}_{\Delta(\tau,b+1-l)\otimes\sigma}$ is

$$\psi' = \begin{cases} (\tau, 2b-2l+1) \boxplus (\tau,1) \boxplus_{i=3}^{r}(\tau_i,1) & \text{if } 1 \le l \le b, \\ \boxplus_{i=3}^{r}(\tau_i,1) & \text{if } l = b+1. \end{cases}$$

Since $[(2m+2k)1^{2a(b+1-l)-2k}]$ is bigger than $\eta_{\mathfrak{so}_{2n'+1}(\mathbb{C}),\mathfrak{sp}_{2n'}(\mathbb{C})}(\underline{p}(\psi'))$ under the lexicographical ordering, where $2n' = 2a(b+1-l)+2m$, by [Jiang and Liu 2015a, Proposition 6.4; Ginzburg et al. 2003, Lemma 1.1],

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m+a-1+k}}\big(\mathcal{C}_{N^{2a(b+1)+2m}_{r-k}}(\rho(\hat{\gamma}\lambda\beta)\xi)\big)$$

is also identically zero. Hence the corresponding term is also zero.

It follows that the only possibilities for which

$$\mathcal{C}_{N^{2a(b+1)+2m}_{r}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)\big) \neq 0$$

are $r = la$ with $1 \le l \le b+1$, and $k = 0$. To prove that $\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)$ is not identically zero, we just have to show that

$$\mathcal{C}_{N_r^{2ab}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)\big) \neq 0$$

for some r.

Take $r = ab$. Then we have

$$\mathcal{C}_{N_{ab}^{2ab}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)\big) = \int_{L(\mathbb{A})} \phi_1(i(\lambda))\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{m+a-1}}\big(\mathcal{C}_{N_{ab}^{2a(b+1)+2m}}(\xi)\big)(\lambda\beta)\,d\lambda. \tag{6-2}$$

By Lemma 6.1, when restricted to $\mathrm{GL}_{2ab}(\mathbb{A}) \times \mathrm{Sp}_{2m+2a}(\mathbb{A})$,

$$\mathcal{C}_{N_{ab}^{2a(b+1)+2m}}(\xi) \in \delta^{1/2}_{P_{ab}^{2a(b+1)+2m}}|\cdot|^{-\frac{b+1}{2}}\Delta(\tau, b) \otimes (\mathcal{E}_{\tau\otimes\sigma}),$$

where $\mathcal{E}_{\tau\otimes\sigma}$ is not a residual representation as explained at the end of Section 2B.

Clearly, the integral in (6-2) is not identically zero if and only if $\mathcal{E}_{\tau\otimes\sigma}$ is ψ^{α}-generic. Since by assumption, σ is ψ^{α}-generic, we have that $\mathcal{E}_{\tau\otimes\sigma}$ is also ψ^{α}-generic by [Shahidi 2010, Theorem 7.1.3]. Hence,

$$\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)$$

is not identically zero. Therefore, $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2m+2a)1^{2ab}]$ with respect to the character

$$\psi_{[(2m+2a)1^{2ab}],\alpha}.$$

This completes the proof of step (1).

6B. ***Proof of step (2).*** To prove the square-integrability of the descent representation

$$\mathcal{D}^{2a(b+1)+2m}_{2m+2a,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma}),$$

as in Section 5B, we need to calculate the automorphic exponent attached to the nontrivial constant term considered in step (1) ($r = ab$). For this, we need to consider the action of

$$\bar{g} = \mathrm{diag}(g, g^*) \in \mathrm{GL}_{ab}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_0(\mathbb{A}).$$

Since $r = ab$, we have that $\beta = \begin{pmatrix} 0 & I_{ab} \\ I_{m+a} & 0 \end{pmatrix}^{\wedge}$. Let

$$\tilde{g} := \beta\,\mathrm{diag}(I_{m+a}, \bar{g}, I_{m+a})\beta^{-1} = \mathrm{diag}(g, I_{2m+2a}, g^*).$$

Then changing variables in (5-2), $\lambda \mapsto \tilde{g}\lambda\tilde{g}^{-1}$ will give a Jacobian $|\det g|^{-m-a}$. On the other hand, by [Ginzburg et al. 2011, Formula (1.4)], the action of $\bar{g}$ on ϕ_1

gives $\gamma_{\psi^{-\alpha}}(\det g)\,|\det g|^{\frac{1}{2}}$. Therefore, $\bar{g}$ acts by $\Delta(\tau, b)(g)$ with character

$$\delta^{1/2}_{P^{2a(b+1)+2m}_{ab}}(\tilde{g})\,|\det g|^{-\frac{b+1}{2}}\,|\det g|^{-m-a}\,\gamma_{\psi^{-\alpha}}(\det g)|\det g|^{\frac{1}{2}} \\ = \gamma_{\psi^{-\alpha}}(\det g)\delta^{1/2}_{P^{2ab}_{ab}}(\bar{g})\,|\det g|^{-\frac{b}{2}}.$$

Therefore, as a function on $\mathrm{GL}_{ab}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_0(\mathbb{A})$,

$$\mathcal{C}_{N^{2ab}_{ab}}\big(\mathcal{F}\mathcal{J}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)\big) \in \gamma_{\psi^{-\alpha}}\delta^{1/2}_{P^{2ab}_{ab}}\,|\det(\cdot)|^{-\frac{b}{2}}\,\Delta(\tau, b) \otimes 1_{\widetilde{\mathrm{Sp}}_0(\mathbb{A})}. \tag{6-3}$$

Since the cuspidal exponent of $\Delta(\tau, b)$ is

$$\Big\{\Big(\frac{1-b}{2}, \frac{3-b}{2}, \dots, \frac{b-1}{2}\Big)\Big\},$$

the cuspidal exponent of $\mathcal{C}_{N^{2ab}_{ab}}\big(\mathcal{F}\mathcal{J}^{\phi}_{\psi^{\alpha}_{m+a-1}}(\xi)\big)$ is

$$\Big\{\Big(\frac{1-2b}{2}, \frac{3-2b}{2}, \dots, -\frac{1}{2}\Big)\Big\}.$$

Hence, by the Langlands square-integrability criterion ([Mœglin and Waldspurger 1995, Lemma I.4.11]), the automorphic representation

$$\mathcal{D}^{2a(b+1)+2m}_{2m+2a,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma})$$

is square integrable.

From (6-3), it follows that as a representation of $\mathrm{GL}_{ab}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_0(\mathbb{A})$,

$$\mathcal{C}_{N^{2ab}_{ab}}\big(\mathcal{D}^{2a(b+1)+2m}_{2m+2a,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma})\big) \\ = \gamma_{\psi^{-\alpha}}\delta^{1/2}_{P^{2ab}_{ab}}\,|\det(\cdot)|^{-\frac{b}{2}}\,\Delta(\tau, b) \otimes 1_{\widetilde{\mathrm{Sp}}_0(\mathbb{A})}. \tag{6-4}$$

Therefore, a similar argument as in Section 5B implies that any noncuspidal summand of

$$\mathcal{D}^{2a(b+1)+2m}_{2m+2a,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma})$$

must be an irreducible subrepresentation of $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$. Hence,

$$\mathcal{D}^{2a(b+1)+2m}_{2m+2a,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma})$$

has a nontrivial intersection with the space of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$. Since $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$ is irreducible,

$$\mathcal{D}^{2a(b+1)+2m}_{2m+2a,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b+1)\otimes\sigma})$$

must contain the whole space of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b)}$. This completes the proof of step (2).

7. Proof of Theorem 2.7

In this section, assuming that $a = 2k$, $L(\frac{1}{2}, \tau \times \sigma) \neq 0$, $\sigma \ncong 1_{\mathrm{Sp}_0(\mathbb{A})}$, and $\mathcal{E}_{\tau\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2k+2m)(2k)]$, we prove that $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2k+2m)(2k)^{2b-1}]$, for any $b \geq 1$.

Without loss of generality, by [Jiang and Liu 2015c, Lemma 3.1] or [Ginzburg et al. 2003, Lemma 2.6], we may assume that $\mathcal{E}_{\tau\otimes\sigma}$ has a nonzero Fourier coefficient corresponding to the partition $[(2k+2m)1^{2k}]$ with respect to the character $\psi_{[(2k+2m)1^{2k}],\alpha}$ for some $\alpha \in F^*/(F^*)^2$. Then the ψ^α-descent of $\mathcal{E}_{\tau\otimes\sigma}$ is a generic representation of $\widetilde{\mathrm{Sp}}_{2k}(\mathbb{A})$. Note that by the constant formula in [Ginzburg et al. 2011, Theorem 7.8], one can easily see that this descent is also a cuspidal representation of $\widetilde{\mathrm{Sp}}_{2k}(\mathbb{A})$

Similarly as in previous sections, we separate the proof of Theorem 2.7 into three steps:

Step (1) $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero $\psi_{[(2k+2m)1^{2k(2b-1)}],\alpha}$-Fourier coefficient attached to the partition $[(2k+2m)1^{2k(2b-1)}]$ (for definition, see [Jiang and Liu 2015c, Section 2]).

Step (2) Let $\tilde{\sigma}$ be any irreducible subrepresentation of the ψ^α-descent of $\mathcal{E}_{\tau\otimes\sigma}$. Then it is a generic cuspidal representation of $\widetilde{\mathrm{Sp}}_{2k}(\mathbb{A})$ which is weakly lifting to τ. Using the theory of theta correspondence and the strong lifting from generic cuspidal representations of $\mathrm{SO}_{2n+1}(\mathbb{A})$ to automorphic representations of $\mathrm{GL}_{2n}(\mathbb{A})$, proved in [Jiang and Soudry 2003] (see also [Cogdell et al. 2004]), τ is also a strong lifting of $\tilde{\sigma}$.

Define a residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ as follows: for any

$$\tilde{\phi} \in \mathcal{A}\big(N_{k(2b-1)}(\mathbb{A})\,\widetilde{M}_{k(2b-1)}(F) \setminus \widetilde{\mathrm{Sp}}_{2k(2b-1)}(\mathbb{A})\big)_{\gamma_{\psi^{-\alpha}}\Delta(\tau,b-1)\otimes\tilde{\sigma}}$$

one defines as in [Mœglin and Waldspurger 1995]) the residual Eisenstein series

$$\widetilde{E}(\tilde{\phi},s)(g) = \sum_{\gamma\in P_{k(2b-1)}(F)\setminus \mathrm{Sp}_{2k(2b-1)}(F)} \lambda_s\tilde{\phi}(\gamma g).$$

It converges absolutely for real part of s large and has meromorphic continuation to the whole complex plane $\mathbb{C}$. By similar argument as that in [Jiang et al. 2013], this Eisenstein series has a simple pole at $b/2$, which is the right-most one. Denote the representation generated by these residues at $s = b/2$ by $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$. This residual representation is square-integrable. Since τ is also a strong lifting of $\tilde{\sigma}$, the same argument as in Section 5B implies that $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ is also irreducible (details will be omitted).

Let

$$\mathcal{D}^{4kb+2m}_{2k+2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$$

be the ψ^{α}-descent of $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$. Then as a representation of $\widetilde{\mathrm{Sp}}_{2k(2b-1)}(\mathbb{A})$, it is square-integrable and contains the whole space of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$, where $\tilde{\sigma}$ is an irreducible subrepresentation of the ψ^{α}-descent of $\mathcal{E}_{\tau\otimes\sigma}$.

Step (3) Let $\tilde{\sigma}$ be any irreducible subrepresentation of the ψ^{α}-descent of $\mathcal{E}_{\tau\otimes\sigma}$. $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ has a nonzero Fourier coefficient attached to the partition $[(2k)^{2b-1}]$.

Proof of Theorem 2.7. From the results in steps (1)–(3) above, we can see that $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the composite partition $[(2k+2m)1^{2k(2b-1)}]\circ[(2k)^{2b-1}]$ (for the definition of composite partitions and the attached Fourier coefficients, we refer to [Ginzburg et al. 2003, Section 1]). Therefore, by [Jiang and Liu 2015c, Lemma 3.1] or [Ginzburg et al. 2003, Lemma 2.6], $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to $[(2k+2m)(2k)^{2b-1}]$. $\square$

Before proving the above three steps, we record the following lemma which is analogous to Lemmas 4.1 and 6.1.

Lemma 7.1. *Let $P_{ai} = M_{ai}N_{ai}$, with $1 \le i \le b$, be the parabolic subgroup of Sp_{2ab+2m} with Levi part $M_{ai} \cong \mathrm{GL}_{ai} \times \mathrm{Sp}_{a(2b-2i)+2m}$. Let φ be an arbitrary automorphic form in $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$. Denote by $\varphi_{P_{ai}}(g)$ the constant term of φ along P_{ai}. Then, for $1 \le i \le b$,*

$$\varphi_{P_{ai}} \in \mathcal{A}\big(N_{ai}(\mathbb{A})M_{ai}(F)\setminus \mathrm{Sp}_{2ab+2m}(\mathbb{A})\big)_{\Delta(\tau,i)|\cdot|^{-(2b-i)/2}\otimes\mathcal{E}_{\Delta(\tau,b-i)\otimes\sigma}}.$$

Note that when $i = b$, $\mathcal{E}_{\Delta(\tau,b-i)\otimes\sigma} = \sigma$.

7A. *Proof of step (1).* By [Ginzburg et al. 2003, Lemma 1.1], $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2k+2m)1^{2ab}]$ with respect to the character $\psi_{[(2k+2m)1^{2ab}],\alpha}$ if and only if the ψ^{α}-descent

$$\mathcal{D}^{4kb+2m}_{2k+2m,\psi^{\alpha}}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$$

of $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ is not identically zero, as a representation of $\widetilde{\mathrm{Sp}}_{2k(2b-1)}(\mathbb{A})$.

We calculate the constant term of

$$\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi),$$

for $\xi \in \mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$, along the parabolic subgroup

$$\tilde{P}^{2k(2b-1)}_r(\mathbb{A}) = \widetilde{M}^{2k(2b-1)}_r(\mathbb{A})\,N^{2k(2b-1)}_r(\mathbb{A})$$

of $\widetilde{\mathrm{Sp}}_{2k(2b-1)}(\mathbb{A})$ with Levi isomorphic to $\mathrm{GL}_r(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_{2k(2b-1)-2r}(\mathbb{A})$, which is denoted by $\mathcal{C}_{N_r^{2k(2b-1)}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi)\big)$, where $1 \le r \le k(2b-1)$.

By [Ginzburg et al. 2011, Theorem 7.8],

$$\mathcal{C}_{N_r^{2k(2b-1)}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi)\big) = \sum_{\substack{0\le s\le r\\ \gamma\in P^1_{r-s,1^s}(F)\backslash \mathrm{GL}_r(F)}} \int_{L(\mathbb{A})} \phi_1(i(\lambda))\,\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{k+m-1+s}}\big(\mathcal{C}_{N^{4kb+2m}_{r-s}}(\xi)\big)(\hat\gamma\lambda\beta)\,d\lambda. \tag{7-1}$$

The notation in this formula is as follows: N^{4kb+2m}_{r-s} is the unipotent radical of the parabolic subgroup P^{4kb+2m}_{r-s} of Sp_{4kb+2m} with Levi isomorphic to

$$\mathrm{GL}_{r-s}\times \mathrm{Sp}_{4kb+2m-2r+2s},$$

and $P^1_{r-s,1^s}$ is a subgroup of GL_r consisting of matrices of the form

$$\begin{pmatrix} g & x\\ 0 & z\end{pmatrix},$$

with $z\in U_s$, the standard maximal unipotent subgroup of GL_s. For $g\in \mathrm{GL}_j$, $j\le 2kb+m$, $\hat g=\mathrm{diag}(g, I_{4kb+2m-2j}, g^*)$; L is a unipotent subgroup, consisting of matrices of the form

$$\lambda=\begin{pmatrix} I_r & 0\\ x & I_{k+m}\end{pmatrix}^{\wedge},$$

$i(\lambda)$ is the last row of x, and

$$\beta=\begin{pmatrix} 0 & I_r\\ I_{k+m} & 0\end{pmatrix}^{\wedge}.$$

The Schwartz function $\phi=\phi_1\otimes\phi_2$ with $\phi_1\in\mathcal{S}(\mathbb{A}^r)$ and $\phi_2\in\mathcal{S}(\mathbb{A}^{k(2b-1)-r})$, and the function

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{k+m-1+s}}\big(\mathcal{C}_{N^{4kb+2m}_{r-s}}(\xi)\big)(\hat\gamma\lambda\beta) := \mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{k+m-1+s}}\big(\mathcal{C}_{N^{4kb+2m}_{r-s}}(\rho(\hat\gamma\lambda\beta)\xi)\big)(I),$$

with $\rho(\hat\gamma\lambda\beta)$ denoting the right translation by $\hat\gamma\lambda\beta$, is a composition of the restriction of $\mathcal{C}_{N^{2ab+2m}_{r-k}}(\rho(\hat\gamma\lambda\beta)\xi)$ to $\mathrm{Sp}_{4kb+2m-2r+2s}(\mathbb{A})$ with the Fourier–Jacobi coefficient

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{k+m-1+s}},$$

taking automorphic forms on $\mathrm{Sp}_{4kb+2m-2r+2s}(\mathbb{A})$ to those on $\widetilde{\mathrm{Sp}}_{4kb-2k-2r}(\mathbb{A})$.

By the cuspidal support of ξ, $\mathcal{C}_{N^{4kb+2m}_{r-s}}(\xi)$ is identically zero, unless $s=r$ or $r-s=2kl$ with $1\le l\le b$. When $s=r$, since $[(2k+2m+2r)1^{4kb-2k-2r}]$ is bigger than $\eta_{\mathfrak{so}_{2n+1}(\mathbb{C}),\mathfrak{sp}_{2n}(\mathbb{C})}(\underline{p}(\psi))$ under the lexicographical ordering, by [Jiang and Liu 2015a, Proposition 6.4; Ginzburg et al. 2003, Lemma 1.1],

$$\mathcal{FJ}^{\phi_2}_{\psi^{\alpha}_{k+m-1+r}}(\xi)$$

is identically zero, and hence the corresponding term is zero. When $r-s=la$, $1\le l\le b$ and $1\le s\le r$, by Lemma 7.1, after restricting to $\mathrm{Sp}_{4k(b-l)+2m}(\mathbb{A})$,

$\mathcal{C}_{N_{r-s}^{4kb+2m}}(\rho(\hat{\gamma}\lambda\beta)\xi)$ becomes a form in $\mathcal{E}_{\Delta(\tau,b-l)\otimes\sigma}$. The Arthur parameter of $\mathcal{E}_{\Delta(\tau,b-l)\otimes\sigma}$ is

$$\psi' = (\tau, 2b-2l) \boxplus \boxplus_{i=2}^{r} (\tau_i, 1).$$

Since $[(2k+2m+2s)1^{4k(b-l)-2k-2s}]$ is bigger than $\eta_{\mathfrak{so}_{2n'+1}(\mathbb{C}),\mathfrak{sp}_{2n'}(\mathbb{C})}(\underline{p}(\psi'))$ under the lexicographical ordering, where $2n' = 4k(b-l)+2m$, by [Jiang and Liu 2015a, Proposition 6.4; Ginzburg et al. 2003, Lemma 1.1],

$$\mathcal{F}\mathcal{J}^{\phi_2}_{\psi^{\alpha}_{k+m-1+s}}\big(\mathcal{C}_{N_{r-s}^{4kb+2m}}(\rho(\hat{\gamma}\lambda\beta)\xi)\big)$$

is also identically zero, and hence the corresponding term is also zero. Therefore, the only possibilities that

$$\mathcal{C}_{N_r^{2k(2b-1)}}\big(\mathcal{F}\mathcal{J}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi)\big) \neq 0$$

are $r = 2kl$, $1 \le l \le b$, and $s = 0$. To prove that $\mathcal{F}\mathcal{J}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi)$ is not identically zero, we just have to show that

$$\mathcal{C}_{N_r^{2k(2b-1)}}\big(\mathcal{F}\mathcal{J}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi)\big) \neq 0$$

for some r.

Taking $r = 2k(b-1)$, we have

$$\begin{aligned}(7\text{-}2)\quad \mathcal{C}_{N_{2k(b-1)}^{2k(2b-1)}}\big(\mathcal{F}\mathcal{J}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi)\big) \\ = \int_{L(\mathbb{A})} \phi_1(i(\lambda))\,\mathcal{F}\mathcal{J}^{\phi_2}_{\psi^{\alpha}_{k+m-1}}\big(\mathcal{C}_{N_{2k(b-1)}^{4kb+2m}}(\xi)\big)(\lambda\beta)\,d\lambda.\end{aligned}$$

By Lemma 7.1, when restricted to $\mathrm{GL}_{2k(2b-2)}(\mathbb{A}) \times \mathrm{Sp}_{4k+2m}(\mathbb{A})$,

$$\mathcal{C}_{N_{2k(b-1)}^{4kb+2m}}(\xi) \in \delta^{1/2}_{P_{2k(b-1)}^{4kb+2m}}|\det|^{-\frac{b+1}{2}}\Delta(\tau, b-1) \otimes \mathcal{E}_{\tau\otimes\sigma}.$$

It follows that the integral in (7-2) is not identically zero if and only if $\mathcal{E}_{\tau\otimes\sigma}$ has a nonzero Fourier coefficient corresponding to the partition $[(2k+2m)1^{2k}]$ with respect to the character $\psi_{[(2k+2m)1^{2k}],\alpha}$. Hence, by assumption,

$$\mathcal{F}\mathcal{J}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi)$$

is not identically zero. Therefore, $\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}$ has a nonzero Fourier coefficient attached to the partition $[(2k+2m)1^{2k(2b-1)}]$ with respect to the character

$$\psi_{[(2k+2m)1^{2k(2b-1)}],\alpha}.$$

This completes the proof of step (1).

7B. ***Proof of step (2).*** In order to prove the square-integrability of the descent representation

$$\mathcal{D}^{4kb+2m}_{2k+2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma}),$$

we need to calculate the automorphic exponent attached to the nontrivial constant term considered in step (1) with $r = 2k(b-1)$ (for the definition of automorphic exponent, see [Mœglin and Waldspurger 1995, I.3.3]). For this, we need to consider the action of

$$\bar{g} = \mathrm{diag}(g, I_{2k}, g^*) \in \mathrm{GL}_{2k(b-1)}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_{2k}(\mathbb{A}).$$

Since $r = 2k(b-1)$, $\beta = \begin{pmatrix} 0 & I_{2k(b-1)} \\ I_{k+m} & 0 \end{pmatrix}^{\wedge}$. Let

$$\tilde{g} := \beta\,\mathrm{diag}(I_{k+m}, \bar{g}, I_{k+m})\beta^{-1} = \mathrm{diag}(g, I_{4k+2m}, g^*).$$

Then changing variables in (5-2) via $\lambda \mapsto \tilde{g}\lambda\tilde{g}^{-1}$ will give a Jacobian $|\det g|^{-k-m}$. On the other hand, by [Ginzburg et al. 2011, Formula (1.4)], the action of $\bar{g}$ on ϕ_1 gives $\gamma_{\psi^{-\alpha}}(\det g)|\det g|^{1/2}$. Therefore, $\bar{g}$ acts by $\Delta(\tau, b-1)(g)$ with character

$$\delta^{1/2}_{P^{4kb+2m}_{2k(b-1)}} |\det g|^{-\frac{b+1}{2}}\, |\det g|^{-k-m}\, \gamma_{\psi^{-\alpha}}(\det g)\, |\det g|^{\frac{1}{2}} = \gamma_{\psi^{-\alpha}}(\det g)\,\delta^{1/2}_{P^{2k(2b-1)}_{2k(b-1)}}(\bar{g})\,|\det g|^{-\frac{b}{2}}.$$

Thus, combined with the calculation in step (1), as a function on $\mathrm{GL}_{2k(b-1)}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_{2k}(\mathbb{A})$,

$$\mathcal{C}_{N^{2k(2b-1)}_{2k(b-1)}}\big(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi)\big) \in \gamma_{\psi^{-\alpha}}\,\delta^{1/2}_{P^{2k(2b-1)}_{2k(b-1)}}\,|\det(\cdot)|^{-\frac{b}{2}}\,\Delta(\tau, b-1) \otimes \mathcal{D}^{4k+2m}_{2k}(\mathcal{E}_{\tau\otimes\sigma}). \tag{7-3}$$

Note that by the constant formula in [op. cit., Theorem 7.8], one can easily see that

$$\mathcal{D}^{4k+2m}_{2k}(\mathcal{E}_{\tau\otimes\sigma})$$

is a cuspidal representation of $\widetilde{\mathrm{Sp}}_{2k}(\mathbb{A})$. Since the cuspidal exponent of $\Delta(\tau, b-1)$ is

$$\Big\{\Big(\frac{2-b}{2}, \frac{4-b}{2}, \dots, \frac{b-2}{2}\Big)\Big\},$$

the cuspidal exponent of $\mathcal{C}_{N^{2k(2b-1)}_{2k(b-1)}}(\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{k+m-1}}(\xi))$ is

$$\Big\{\Big(\frac{2-2b}{2}, \frac{4-2b}{2}, \dots, -1\Big)\Big\}.$$

Hence, by the Langlands square-integrability criterion [Mœglin and Waldspurger 1995, Lemma I.4.11], the automorphic representation $\mathcal{D}^{4kb+2m}_{2k+2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$ is square-integrable.

From (7-3), as a representation of $\mathrm{GL}_{2k(b-1)}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_{2k}(\mathbb{A})$, we have

$$\mathcal{C}_{N^{2k(2b-1)}_{2k(b-1)}}\big(\mathcal{D}^{4kb+2m}_{2k+2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})\big) = \gamma_{\psi^{-\alpha}}\,\delta^{1/2}_{P^{2k(2b-1)}_{2k(b-1)}}\,|\det(\cdot)|^{-\frac{b}{2}}\,\Delta(\tau,b-1)\otimes\mathcal{D}^{4k+2m}_{2k,\psi^\alpha}(\mathcal{E}_{\tau\otimes\sigma}). \tag{7-4}$$

Therefore, using a similar argument as in Section 5B, one can see that

$$\mathcal{D}^{4kb+2m}_{2k+2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$$

contains an irreducible subrepresentation of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$, where $\tilde{\sigma}$ is an irreducible generic cuspidal representation of $\widetilde{\mathrm{Sp}}_{2k}(\mathbb{A})$ which is a subrepresentation of the ψ^α-descent of $\mathcal{E}_{\tau\otimes\sigma}$, and is weakly lifting to τ. Since τ is also a strong lifting of $\tilde{\sigma}$, a similar argument as in Section 5B implies that $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ is irreducible. Hence $\mathcal{D}^{4kb+2m}_{2k+2m,\psi^\alpha}(\mathcal{E}_{\Delta(\tau,b)\otimes\sigma})$ must contain the whole space of residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$. This completes the proof of step (2).

7C. ***Proof of step (3).*** Let $\tilde{\sigma}$ be any irreducible subrepresentation of the ψ^α-descent of $\mathcal{E}_{\tau\otimes\sigma}$, then it is a generic cuspidal representation of $\widetilde{\mathrm{Sp}}_{2k}(\mathbb{A})$. Assume that $\tilde{\sigma}$ is ψ^β-generic for some $\beta \in F^*/(F^*)^2$.

As in previous sections, we need to record the following lemma which is analogous to Lemma 5.1.

Lemma 7.2. *Let $\tilde{P}_{ai}(\mathbb{A}) = \widetilde{M}_{ai}(\mathbb{A})N_{ai}(\mathbb{A})$ with $1 \le i \le b-1$ be the parabolic subgroup of $\widetilde{\mathrm{Sp}}_{2k(2b-1)}(\mathbb{A})$ with Levi part*

$$\widetilde{M}_{ai}(\mathbb{A}) \cong \mathrm{GL}_{ai}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_{2k(2b-1-2i)}(\mathbb{A}).$$

Let φ be an arbitrary automorphic form in $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$. Denote by $\varphi_{P_{ai}}(g)$ the constant term of φ along P_{ai}. Then, for $1 \le i \le b-1$,

$$\varphi_{P_{ai}} \in \mathcal{A}\big(N_{ai}(\mathbb{A})\widetilde{M}_{ai}(F)\backslash\widetilde{\mathrm{Sp}}_{2k(2b-1)}(\mathbb{A})\big)_{\gamma_{\psi^{-\alpha}}\Delta(\tau,i)|\cdot|^{-(2b-1-i)/2}\otimes\tilde{\mathcal{E}}_{\Delta(\tau,b-1-i)\otimes\tilde{\sigma}}}.$$

Note that when $i = b-1$, $\tilde{\mathcal{E}}_{\Delta(\tau,b-1-i)\otimes\tilde{\sigma}} = \tilde{\sigma}$.

First, we show that $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ has a nonzero Fourier coefficient attached to the partition $[(2k)1^{2k(2b-2)}]$ with respect to the character $\psi_{[(2k)1^{2k(2b-2)}],\beta}$. By [Ginzburg et al. 2003, Lemma 1.1], we know that $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ has a nonzero $\psi_{[(2k)1^{2k(2b-2)}],\beta}$-Fourier coefficient attached to the partition $[(2k)1^{2k(2b-2)}]$ if and only if the ψ^β-descent

$$\widetilde{\mathcal{D}}^{2k(2b-1)}_{2k,\psi^\beta}(\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}})$$

of $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ is not identically zero, as a representation of $\mathrm{Sp}_{2k(2b-2)}(\mathbb{A})$.

Take any $\xi \in \tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$; we will calculate the constant term of

$$\mathcal{F}\mathcal{J}^{\phi}_{\psi^\beta_{k-1}}(\xi)$$

along $P_r^{2k(2b-2)}$, which is denoted by

$$\mathcal{C}_{N_r^{2k(2b-2)}}\big(\mathcal{FJ}^{\phi}_{\psi^{\beta}_{k-1}}(\xi)\big),$$

where $1 \le r \le k(2b-2)$. Recall that $P_r^{2k(2b-2)} = M_r^{2k(2b-2)} N_r^{2k(2b-2)}$ is the parabolic subgroup of $\mathrm{Sp}_{2k(2b-2)}$ with Levi subgroup isomorphic to

$$\mathrm{GL}_r \times \mathrm{Sp}_{2k(2b-2)-2r}\,.$$

By [Ginzburg et al. 2011, Theorem 7.8],

$$\begin{aligned}(7\text{-}5)\quad &\mathcal{C}_{N_r^{2k(2b-2)}}(\mathcal{FJ}^{\phi}_{\psi^{\beta}_{k-1}}(\xi))\\ &\qquad = \sum_{\substack{0\le s\le r\\ \gamma\in P^1_{r-s,1^s}(F)\backslash \mathrm{GL}_r(F)}} \int_{L(\mathbb{A})} \phi_1(i(\lambda))\,\mathcal{FJ}^{\phi_2}_{\psi^{\beta}_{k-1+s}}\big(\mathcal{C}_{N_{r-s}^{2k(2b-1)}}(\xi)\big)(\hat{\gamma}\lambda\eta)\,d\lambda.\end{aligned}$$

Here is the notation in the formula: $N_{r-s}^{2k(2b-1)}(\mathbb{A})$ is the unipotent radical of the parabolic subgroup $\tilde{P}_{r-s}^{2k(2b-1)}(\mathbb{A})$ of $\widetilde{\mathrm{Sp}}_{2k(2b-1)}(\mathbb{A})$ with Levi subgroup isomorphic to $\mathrm{GL}_{r-s}(\mathbb{A}) \times \widetilde{\mathrm{Sp}}_{2k(2b-1)-2r+2s}(\mathbb{A})$, $P^1_{r-s,1^s}$ is a subgroup of GL_r consisting of matrices of the form

$$\begin{pmatrix} g & x\\ 0 & z\end{pmatrix},$$

with $z \in U_s$, the standard maximal unipotent subgroup of GL_s. For $g \in \mathrm{GL}_j$, with $j \le k(2b-1)$, $\hat{g} = \mathrm{diag}(g, I_{2k(2b-1)-2j}, g^*)$, L is a unipotent subgroup, consisting of matrices of the form

$$\lambda = \begin{pmatrix} I_r & 0\\ x & I_k\end{pmatrix}^{\wedge},$$

$i(\lambda)$ is the last row of x, and

$$\eta = \begin{pmatrix} 0 & I_r\\ I_k & 0\end{pmatrix}^{\wedge}.$$

The Schwartz function $\phi = \phi_1 \otimes \phi_2$ with $\phi_1 \in \mathcal{S}(\mathbb{A}^r)$ and $\phi_2 \in \mathcal{S}(\mathbb{A}^{k(2b-2)-r})$, and the function

$$\mathcal{FJ}^{\phi_2}_{\psi^{\beta}_{k-1+s}}\big(\mathcal{C}_{N_{r-s}^{2k(2b-1)}}(\xi)\big)(\hat{\gamma}\lambda\eta) := \mathcal{FJ}^{\phi_2}_{\psi^{\beta}_{k-1+s}}\big(\mathcal{C}_{N_{r-s}^{2k(2b-1)}}(\rho(\hat{\gamma}\lambda\eta)\xi)\big)(I),$$

with $\rho(\hat{\gamma}\lambda\eta)$ denoting the right translation by $\hat{\gamma}\lambda\eta$, is a composition of the restriction to $\widetilde{\mathrm{Sp}}_{2k(2b-1)-2r+2s}(\mathbb{A})$ of $\mathcal{C}_{N_{r-s}^{2ab+2m}}(\rho(\hat{\gamma}\lambda\eta)\xi)$ with Fourier–Jacobi coefficient

$$\mathcal{FJ}^{\phi_2}_{\psi^{\beta}_{k-1+s}},$$

taking automorphic forms on $\widetilde{\mathrm{Sp}}_{2k(2b-1)-2r+2s}(\mathbb{A})$ to those on $\mathrm{Sp}_{2k(2b-2)-2r}(\mathbb{A})$.

By the cuspidal support of ξ, $\mathcal{C}_{N_{r-s}^{2k(2b-1)}}(\xi)$ is identically zero, unless $s = r$ or $r - s = 2kl$ with $1 \le l \le b-1$. When $s = r$, from the structure of the unramified components of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$, by [Jiang and Liu 2015a, Lemma 3.2],

$$\mathcal{FJ}^{\phi_2}_{\psi^{\beta}_{k-1+r}}(\xi)$$

is identically zero, and hence the corresponding term is zero. When $r-s=2kl$, $1\le l\le b-1$ and $1\le s\le r$, then by Lemma 7.2, after restricting to $\widetilde{\mathrm{Sp}}_{2k(2b-1-2l)}(\mathbb{A})$, $\mathcal{C}_{N^{2k(2b-1)}_{r-s}}(\rho(\hat{\gamma}\lambda\eta)\xi)$ becomes a form in $\tilde{\mathcal{E}}_{\Delta(\tau,b-1-l)\otimes\tilde{\sigma}}$. From the structure of the unramified components of the residual representation $\tilde{\mathcal{E}}_{\Delta(\tau,b-1-l)\otimes\tilde{\sigma}}$, by [loc. cit.],

$$\mathcal{FJ}^{\phi_2}_{\psi^{\beta}_{k-1+s}}\big(\mathcal{C}_{N^{2k(2b-1)}_{r-s}}(\rho(\hat{\gamma}\lambda\eta)\xi)\big)$$

is also identically zero, and hence the corresponding term is also zero. Therefore, the only possibilities that

$$\mathcal{C}_{N^{2k(2b-2)}_{r}}\big(\mathcal{FJ}^{\phi}_{\psi^{\beta}_{k-1}}(\xi)\big)\neq 0$$

are $r=2kl$, $1\le l\le b-1$, and $s=0$. To prove that $\mathcal{FJ}^{\phi}_{\psi^{\beta}_{k-1}}(\xi)$ is not identically zero, we just have to show

$$\mathcal{C}_{N^{2k(2b-2)}_{r}}\big(\mathcal{FJ}^{\phi}_{\psi^{\beta}_{k-1}}(\xi)\big)\neq 0$$

for some r.

Taking $r=2k(b-1)$, we have

$$\mathcal{C}_{N^{2k(2b-2)}_{r}}\big(\mathcal{FJ}^{\phi}_{\psi^{\beta}_{k-1}}(\xi)\big)=\int_{L(\mathbb{A})}\phi_1(i(\lambda))\,\mathcal{FJ}^{\phi_2}_{\psi^{\beta}_{k-1}}\big(\mathcal{C}_{N^{2k(2b-1)}_{2k(b-1)}}(\xi)\big)(\lambda\eta)\,d\lambda. \tag{7-6}$$

By Lemma 7.2, when restricted to $\mathrm{GL}_{2k(b-1)}(\mathbb{A})\times\widetilde{\mathrm{Sp}}_{2k}(\mathbb{A})$,

$$\mathcal{C}_{N^{2k(2b-1)}_{2k(b-1)}}(\xi)\in\delta^{1/2}_{P^{2k(2b-1)}_{2k(b-1)}}\,|\mathrm{det}|^{-\frac{b}{2}}\,\gamma_{\psi^{-\alpha}}\,\Delta(\tau,b-1)\otimes\tilde{\sigma}.$$

It is clear that the integral in (7-6) is not identically zero if and only if $\tilde{\sigma}$ is ψ^{β}-generic. Hence, by assumption,

$$\mathcal{FJ}^{\phi}_{\psi^{\alpha}_{k-1}}(\xi)$$

is not identically zero. Thus, $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ has a nonzero Fourier coefficient attached to the partition $[(2k)1^{2k(2b-2)}]$ with respect to the character $\psi_{[(2k)1^{2k(2b-2)}],\beta}$.

Next, we show that the ψ^{β}-descent

$$\widetilde{\mathcal{D}}^{2k(2b-1)}_{2k,\psi^{\beta}}(\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}})$$

of $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ is square-integrable and contains the whole space of the residual representation $\mathcal{E}_{\Delta(\tau,b-1)}$ which is irreducible, as shown in [Liu 2013b, Theorem 7.1].

To prove the square-integrability of

$$\widetilde{\mathcal{D}}^{2k(2b-1)}_{2k,\psi^{\beta}}(\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}),$$

we need to calculate the automorphic exponent attached to the nontrivial constant term considered above ($r = 2k(b-1)$). For this, we need to consider the action of

$$\bar{g} = \mathrm{diag}(g, g^*) \in \mathrm{GL}_{2k(b-1)}(\mathbb{A}) \times \mathrm{Sp}_0(\mathbb{A}).$$

Since $r = 2k(b-1)$, $\eta = \begin{pmatrix} 0 & I_{2k(b-1)} \\ I_k & 0 \end{pmatrix}^{\wedge}$. Let

$$\tilde{g} := \eta\, \mathrm{diag}(I_k, \bar{g}, I_k)\eta^{-1} = \mathrm{diag}(g, I_{2k}, g^*).$$

Then changing variables in (7-6) via $\lambda \mapsto \tilde{g}\lambda\tilde{g}^{-1}$ will give a Jacobian $|\det g|^{-k}$. On the other hand, by [Ginzburg et al. 2011, Formula (1.4)], the action of $\bar{g}$ on ϕ_1 gives $|\det g|^{1/2}$. Therefore, $\bar{g}$ acts by $\Delta(\tau, b-1)(g)$ with character

$$\delta^{1/2}_{P^{2k(2b-1)}_{2k(b-1)}} |\det g|^{-\frac{b}{2}} |\det g|^{-k} \gamma_{\psi^{-\beta}}(\det g) |\det g|^{\frac{1}{2}} = \delta^{1/2}_{P^{2k(2b-2)}_{2k(b-1)}}(\bar{g}) |\det g|^{-\frac{b-1}{2}}.$$

Therefore, as a function on $\mathrm{GL}_{2k(b-1)}(\mathbb{A}) \times \mathrm{Sp}_0(\mathbb{A})$,

$$\mathcal{C}_{N^{2k(2b-2)}_{2k(b-1)}}\big(\mathcal{FJ}^{\phi}_{\psi^{\beta}_{k-1}}(\xi)\big) \in \delta^{1/2}_{P^{2k(2b-2)}_{2k(b-1)}} |\det(\cdot)|^{-\frac{b-1}{2}} \Delta(\tau, b-1) \otimes 1_{\mathrm{Sp}_0(\mathbb{A})}. \tag{7-7}$$

Since the cuspidal exponent of $\Delta(\tau, b-1)$ is

$$\Big\{\Big(\frac{2-b}{2}, \frac{4-b}{2}, \dots, \frac{b-2}{2}\Big)\Big\},$$

the cuspidal exponent of $\mathcal{C}_{N^{2k(2b-2)}_{2k(b-1)}}\big(\mathcal{FJ}^{\phi}_{\psi^{\beta}_{k-1}}(\xi)\big)$ is

$$\Big\{\Big(\frac{3-2b}{2}, \frac{5-2b}{2}, \dots, -\frac{1}{2}\Big)\Big\}.$$

By the Langlands square-integrability criterion ([Mœglin and Waldspurger 1995, Lemma I.4.11]), the automorphic representation

$$\widetilde{\mathcal{D}}^{2k(2b-1)}_{2k,\psi^{\beta}}(\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}})$$

is square integrable.

From (7-7), it is easy to see that as a representation of $\mathrm{GL}_{2k(b-1)}(\mathbb{A}) \times \mathrm{Sp}_0(\mathbb{A})$,

$$\mathcal{C}_{N^{2k(2b-2)}_{2k(b-1)}}\big(\widetilde{\mathcal{D}}^{2k(2b-1)}_{2k,\psi^{\beta}}(\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}})\big) = \delta^{1/2}_{P^{2k(2b-2)}_{2k(b-1)}} |\det(\cdot)|^{-\frac{b-1}{2}} \Delta(\tau, b-1) \otimes 1_{\mathrm{Sp}_0(\mathbb{A})}. \tag{7-8}$$

It follows that

$$\widetilde{\mathcal{D}}^{2k(2b-1)}_{2k,\psi^{\beta}}(\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}})$$

has a nontrivial intersection with the space of the residual representation $\mathcal{E}_{\Delta(\tau,b-1)}$. Since by [Liu 2013b, Theorem 7.1, part (2)], $\mathcal{E}_{\Delta(\tau,b-1)}$ is irreducible,

$$\widetilde{\mathcal{D}}^{2k(2b-1)}_{2k,\psi^{\beta}}(\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}})$$

must contains the whole space of the residual representation $\mathcal{E}_{\Delta(\tau,b-1)}$. By [op. cit., Theorem 7.1, part (3)], the descent

$$\widetilde{\mathcal{D}}^{2k(2b-1)}_{2k,\psi^\beta}(\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}})$$

is actually irreducible and equals the residual representation $\mathcal{E}_{\Delta(\tau,b-1)}$ identically.

By [op. cit., Theorem 4.2.2], we know that $\mathfrak{p}^m(\mathcal{E}_{\Delta(\tau,b-1)}) = \{[(2k)^{2b-2}]\}$. Therefore, by [Jiang and Liu 2015c, Lemma 3.1] or [Ginzburg et al. 2003, Lemma 2.6], $\tilde{\mathcal{E}}_{\Delta(\tau,b-1)\otimes\tilde{\sigma}}$ has a nonzero Fourier coefficient attached to the partition $[(2k)^{2b-1}]$. This completes the proof of step (3).

Acknowledgments

We would like to thank David Soudry for helpful discussion on related topics. We also would like to thank the referee for careful reading of the paper and helpful comments. This material is based upon work supported by the National Science Foundation under agreement No. DMS-1128155. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[Achar 2003] P. N. Achar, "An order-reversing duality map for conjugacy classes in Lusztig's canonical quotient", *Transform. Groups* **8**:2 (2003), 107–145. MR 2004c:20073 Zbl 1021.22003

[Arthur 2013] J. Arthur, *The endoscopic classification of representations: Orthogonal and symplectic groups*, American Mathematical Society Colloquium Publications **61**, Amer. Math. Soc., Providence, RI, 2013. MR 3135650 Zbl 1310.22014

[Ban and Jantzen 2013] D. Ban and C. Jantzen, "The Langlands quotient theorem for finite central extensions of p-adic groups", *Glas. Mat. Ser. III* **48(68)**:2 (2013), 313–334. MR 3151110 Zbl 1304.22019

[Barbasch and Vogan 1985] D. Barbasch and D. A. Vogan, Jr., "Unipotent representations of complex semisimple groups", *Ann. of Math.* (2) **121**:1 (1985), 41–110. MR 86i:22031 Zbl 0582.22007

[Borel and Wallach 2000] A. Borel and N. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, 2nd ed., Mathematical Surveys and Monographs **67**, Amer. Math. Soc., Providence, RI, 2000. MR 2000j:22015 Zbl 0980.22015

[Cogdell et al. 2004] J. W. Cogdell, H. H. Kim, I. I. Piatetski-Shapiro, and F. Shahidi, "Functoriality for the classical groups", *Publ. Math. Inst. Hautes Études Sci.* **99**:1 (2004), 163–233. MR 2006a:22010 Zbl 1090.22010

[Collingwood and McGovern 1993] D. H. Collingwood and W. M. McGovern, *Nilpotent orbits in semisimple Lie algebras*, Van Nostrand Reinhold, New York, 1993. MR 94j:17001 Zbl 0972.17008

[Gan et al. 2012] W. T. Gan, B. H. Gross, and D. Prasad, "Symplectic local root numbers, central critical L values, and restriction problems in the representation theory of classical groups", pp. 1–109 in *Sur les conjectures de Gross et Prasad, I*, Astérisque **346**, Société Mathématique de France, Paris, 2012. MR 3202556 Zbl 1280.22019

[Ginzburg et al. 2003] D. Ginzburg, S. Rallis, and D. Soudry, "On Fourier coefficients of automorphic forms of symplectic groups", *Manuscripta Math.* **111**:1 (2003), 1–16. MR 2004g:11033 Zbl 1027.11034

[Ginzburg et al. 2004] D. Ginzburg, D. Jiang, and S. Rallis, "On the nonvanishing of the central value of the Rankin–Selberg L-functions", *J. Amer. Math. Soc.* **17**:3 (2004), 679–722. MR 2005g:11078 Zbl 1057.11029

[Ginzburg et al. 2011] D. Ginzburg, S. Rallis, and D. Soudry, *The descent map from automorphic representations of* $\mathrm{GL}(n)$ *to classical groups*, World Scientific, Hackensack, NJ, 2011. MR 2012g:22020 Zbl 1233.11056

[Jiang 2014] D. Jiang, "Automorphic integral transforms for classical groups, I: Endoscopy correspondences", pp. 179–242 in *Automorphic forms and related geometry: Assessing the legacy of I. I. Piatetski–Shapiro*, edited by J. W. Cogdell et al., Contemp. Math. **614**, Amer. Math. Soc., Providence, RI, 2014. MR 3220929 Zbl 1315.11037

[Jiang and Liu 2013] D. Jiang and B. Liu, "On Fourier coefficients of automorphic forms of $\mathrm{GL}(n)$", *Int. Math. Res. Not.* **2013**:17 (2013), 4029–4071. MR 3096918 Zbl 06438742

[Jiang and Liu 2015a] D. Jiang and B. Liu, "Arthur parameters and Fourier coefficients for automorphic forms on symplectic groups", 2015. To appear in *Annales de l'Institut Fourier*. arXiv 1309.6239

[Jiang and Liu 2015b] D. Jiang and B. Liu, "Fourier coefficients for automorphic forms on quasisplit classical groups", 2015. To appear in the special volume in honor of J. Cogdell. arXiv 1412.7553

[Jiang and Liu 2015c] D. Jiang and B. Liu, "On special unipotent orbits and Fourier coefficients for automorphic forms on symplectic groups", *J. Number Theory* **146** (2015), 343–389. MR 3267118 Zbl 06359741

[Jiang and Soudry 2003] D. Jiang and D. Soudry, "The local converse theorem for $\mathrm{SO}(2n+1)$ and applications", *Ann. of Math.* (2) **157**:3 (2003), 743–806. MR 2005b:11193 Zbl 1049.11055

[Jiang et al. 2013] D. Jiang, B. Liu, and L. Zhang, "Poles of certain residual Eisenstein series of classical groups", *Pacific J. Math.* **264**:1 (2013), 83–123. MR 3079762 Zbl 06203663

[Jiang et al. 2015] D. Jiang, B. Liu, B. Xu, and L. Zhang, "The Jacquet–Langlands correspondence via twisted descent", *Int. Math. Res. Not.* (online publication October 2015). arXiv 1501.00506

[Langlands 1976] R. P. Langlands, *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Mathematics **544**, Springer, Berlin, 1976. MR 58 #28319 Zbl 0332.10018

[Liu 2013a] B. Liu, *Fourier coefficients of automorphic forms and Arthur classification*, Ph.D. thesis, University of Minnesota, 2013, Available at http://search.proquest.com/docview/1419458650.

[Liu 2013b] B. Liu, "On extension of Ginzburg–Jiang–Soudry correspondence to certain automorphic forms on $\mathrm{Sp}_{4mn}(\mathbb{A})$ and $\widetilde{\mathrm{Sp}}_{4mn\pm 2n}(\mathbb{A})$", 2013. submitted. arXiv 1309.6240

[Mœglin 2008] C. Mœglin, "Formes automorphes de carré intégrable non cuspidales", *Manuscripta Math.* **127**:4 (2008), 411–467. MR 2010i:11071 Zbl 1306.11041

[Mœglin 2011] C. Mœglin, "Image des opérateurs d'entrelacements normalisés et pôles des séries d'Eisenstein", *Adv. Math.* **228**:2 (2011), 1068–1134. MR 2822218 Zbl 1225.22016

[Mœglin and Waldspurger 1989] C. Mœglin and J.-L. Waldspurger, "Le spectre résiduel de $\mathrm{GL}(n)$", *Ann. Sci. École Norm. Sup.* (4) **22**:4 (1989), 605–674. MR 91b:22028 Zbl 0696.10023

[Mœglin and Waldspurger 1995] C. Mœglin and J.-L. Waldspurger, *Spectral decomposition and Eisenstein series: A paraphrase of the scriptures*, Cambridge Tracts in Mathematics **113**, Cambridge Univ. Press, 1995. MR 97d:11083 Zbl 0846.11032

[Shahidi 2010] F. Shahidi, *Eisenstein series and automorphic L-functions*, American Mathematical Society Colloquium Publications **58**, Amer. Math. Sci., Providence, RI, 2010. MR 2012d:11119 Zbl 1215.11054

[Tadić 1986] M. Tadić, "Classification of unitary representations in irreducible representations of general linear group (non–Archimedean case)", *Ann. Sci. École Norm. Sup.* (4) **19**:3 (1986), 335–382. MR 88b:22021 Zbl 0614.22005

[Vogan 1986] D. A. Vogan, Jr., "The unitary dual of $\mathrm{GL}(n)$ over an Archimedean field", *Invent. Math.* **83**:3 (1986), 449–505. MR 87i:22042 Zbl 0598.22008

Received October 21, 2014. Revised June 12, 2015.

DIHUA JIANG
UNIVERSITY OF MINNESOTA
127 VINCENT HALL
206 CHURCH ST. SE
MINNEAPOLIS, MN 55455
UNITED STATES
dhjiang@math.umn.edu

BAIYING LIU
SCHOOL OF MATHEMATICS
INSTITUTE FOR ADVANCED STUDY
EINSTEIN DRIVE
PRINCETON, NJ 08540
UNITED STATES
liu@ias.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 281, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.281.467

ON THE EXISTENCE OF CENTRAL FANS OF CAPILLARY SURFACES

AMMAR KHANFER

We prove that under some conditions every nonparametric capillary surface which has a central fan (of radial limits at a point $\mathcal{O}$) can be perturbed with respect to the contact angle and the perturbed surfaces continue to have central fans. In particular, any nonparametric capillary surface which is symmetric with respect to a vertical plane through $\mathcal{O}$ and has a central fan may be perturbed (with respect to the contact angle) in a nonsymmetric manner and the resulting capillary surfaces will not be symmetric with respect to the vertical plane but will continue to have central fans.

1. Introduction

Let Ω be a bounded open set in $\mathbb{R}^2$ with locally Lipschitz boundary $\partial\Omega$ such that a point $\mathcal{O}$ lies on $\partial\Omega$, $\partial\Omega\setminus\{\mathcal{O}\}$ is a C^2 curve and there exist distinct rays $l^{\pm}$ starting at $\mathcal{O}$ such that $\partial\Omega$ is tangent to $l^{+}\cup l^{-}$ at $\mathcal{O}$. By rotating and translating the domain, we may assume $\mathcal{O}=(0,0)$, $l^{+}=\{r(\cos\alpha,\sin\alpha):r\geq 0\}$, $l^{-}=\{r(\cos\alpha,-\sin\alpha):r\geq 0\}$ and

$$\Omega\cap B(\mathcal{O},\delta)=\{r(\cos\theta,\sin\theta):0<r<\delta,\ \theta^{-}(r)<\theta<\theta^{+}(r)\}$$

for some $\alpha\in(0,\pi)$, $\delta>0$ and functions $\theta^{\pm}\in C^0([0,\delta))$ which satisfy $\theta^{-}<\theta^{+}$, $\theta^{-}(0)=-\alpha$ and $\theta^{+}(0)=\alpha$; here $B(\mathcal{O},\delta)$ is the open ball in $\mathbb{R}^2$ centered at $\mathcal{O}$ of radius δ. We will assume this description of Ω holds throughout this paper.

Let γ be a measurable function mapping $\partial\Omega$ into $[0,\pi]$ and $f\in C^2(\Omega)\cap L^{\infty}(\Omega)$ be a (bounded) variational solution of the nonparametric capillary surface problem of finding a function $u\in C^2(\Omega)$ such that

$$\operatorname{div}(Tu)=\kappa u+\lambda \quad \text{in } \Omega, \tag{1}$$

$$Tu\cdot\nu=\cos\gamma \quad \text{a.e. on } \partial\Omega, \tag{2}$$

where

$$Tu=\left(\frac{D_1u}{\sqrt{1+|Du|^2}},\frac{D_2u}{\sqrt{1+|Du|^2}}\right),$$

MSC2010: primary 35J93, 76B45; secondary 35J62, 53A10.
Keywords: capillary surfaces, nonconvex corner, central fans.

κ and λ are constants and ν is the outer unit normal to $\partial\Omega$. We will assume $\kappa > 0$ and therefore, by vertical translation, assume $\lambda = 0$. (Since $\kappa > 0$, f is unique.)

Lancaster and Siegel [1996] proved that if γ is bounded away from 0 and π near $\mathcal{O}$, then the radial limit of f at $\mathcal{O}$ in the direction θ,

$$Rf(\theta) \stackrel{\text{def}}{=} \lim_{r\downarrow 0} f(r\cos\theta, r\sin\theta),$$

exists for each $\theta \in [-\alpha, \alpha]$, Rf belongs to $C^0([-\alpha, \alpha])$, $Rf(-\alpha)$ is the limiting height at $\mathcal{O}$ of the trace of f on $\partial^-\Omega = \partial\Omega \cap \{y < 0\}$ and $Rf(\alpha)$ is the limiting height at $\mathcal{O}$ of the trace of f on $\partial^+\Omega = \partial\Omega \cap \{y > 0\}$. In particular, when $\alpha > \frac{\pi}{2}$, so that $\partial\Omega$ has a nonconvex (or reentrant) corner at $\mathcal{O}$, and f is discontinuous at $\mathcal{O}$, the conclusion of Theorem 1 of [Lancaster and Siegel 1996] is that the radial limits of f behave in one of the following ways:

(i) There exist α_1 and α_2 so that $-\alpha \le \alpha_1 < \alpha_2 \le \alpha$ and Rf is constant on $[-\alpha, \alpha_1]$ and $[\alpha_2, \alpha]$ and is strictly increasing or strictly decreasing on $[\alpha_1, \alpha_2]$. Label these case (I) and case (D), respectively.

(ii) There exist $\alpha_1, \alpha_L, \alpha_R, \alpha_2$ so that $-\alpha \le \alpha_1 < \alpha_L < \alpha_R < \alpha_2 \le \alpha$, $\alpha_R = \alpha_L + \pi$, and Rf is constant on $[-\alpha, \alpha_1]$, $[\alpha_L, \alpha_R]$, and $[\alpha_2, \alpha]$ and is either strictly increasing on $[\alpha_1, \alpha_L]$ and strictly decreasing on $[\alpha_R, \alpha_2]$ or strictly decreasing on $[\alpha_1, \alpha_L]$ and strictly increasing on $[\alpha_R, \alpha_2]$. Label these case (ID) and case (DI), respectively.

In addition, if the limits

$$\gamma_1 = \lim_{\partial^+\Omega \ni (x,y) \to \mathcal{O}} \gamma(x, y) \quad \text{and} \quad \gamma_2 = \lim_{\partial^-\Omega \ni (x,y) \to \mathcal{O}} \gamma(x, y) \tag{3}$$

both exist, then [Lancaster 2010; 2012; Lancaster and Siegel 1996] imply that α_2 equals $\alpha - \gamma_1$ in cases (I) and (DI) and $\alpha + \gamma_1 - \pi$ in cases (D) and (ID) while α_1 equals $-\alpha + \gamma_2$ in cases (D) and (DI) and $\pi - \alpha - \gamma_2$ in cases (I) and (ID).

The intervals in $[-\alpha, \alpha]$ on which Rf is constant are called "fans" in, for example, [Lancaster 1985]; specifically, $[-\alpha, \alpha_1]$ and $[\alpha_2, \alpha]$ are called "side fans" and, if it exists, $[\alpha_L, \alpha_L + \pi]$ is called a "central fan". When Ω and γ are symmetric with respect to the x-axis, we have $Rf(\alpha) = Rf(-\alpha)$ and, if $\alpha > \frac{\pi}{2}$, $\alpha_L = -\frac{\pi}{2}$ and $\alpha_R = \frac{\pi}{2}$. (If $\kappa < 0$ in (1), we would need to explicitly assume $f(x, y) = f(x, -y)$ for $(x, y) \in \Omega$.) If the fans touch or overlap (e.g., $\gamma_1 + \gamma_2 \ge 2\alpha - \pi$ in a situation where case (DI) would hold), then f is continuous at $\mathcal{O}$.

Let Ω be a bounded domain in $\mathbb{R}^2$ which is symmetric with respect to the x-axis and has a reentrant corner of size $2\alpha > \pi$ at the origin $\mathcal{O}$. Let $\gamma : \partial\Omega \to (0, \pi)$ also be symmetric with respect to the x-axis such that the limits in (3) exist and $\gamma_1 = \gamma_2 < \frac{\pi}{2}$. As in Example 2 of [Lancaster and Siegel 1996], it follows that the solution f of (1)–(2) with the domain Ω and contact angle γ above is continuous at $\mathcal{O}$ if and only if $\gamma_1 \geq \frac{\pi}{2} - \alpha$ and the radial limits $Rf(\theta)$ of f at $\mathcal{O}$ have a central fan if $\gamma_1 < \frac{\pi}{2} - \alpha$. Danzhu Shi and Robert Finn [2004] considered the borderline case in which $\gamma_1 = \alpha - \frac{\pi}{2}$, so that f is continuous at $\mathcal{O}$. By perturbing the domain (using "an asymmetric domain perturbation that is in an asymptotic sense arbitrarily small"), they showed that the solution of the perturbed capillary problem is discontinuous at $\mathcal{O}$. (They convert the behavior of the radial limit function from a constant in Example 2 to case (I) in the perturbed problem.)

Consider a similar (symmetric) situation with a constant contact angle γ which satisfies $\gamma < \alpha - \frac{\pi}{2}$, so that the solution f of (1)–(2) with the (symmetric) domain Ω and contact angle γ is discontinuous at $\mathcal{O}$, the radial limits $Rf(\theta)$ of f at $\mathcal{O}$ have a central fan and case (DI) holds. Applying the procedure of Finn and Shi, one makes a suitable, nonsymmetric (with respect to the x-axis) perturbation of Ω outside a neighborhood of $\mathcal{O}$ and obtains a new solution $\tilde{f}$ of (1)–(2) in the perturbed domain $\widetilde{\Omega}$, and one then shows that $\tilde{f}$ is discontinuous at $\mathcal{O}$ and the radial limits $R\tilde{f}(\theta)$ have no central fan (i.e., case (I) holds); the size of the domain perturbation required to achieve this depends on the size of $\alpha - \frac{\pi}{2} - \gamma$.

We might view their example and procedure as a perturbation of the contact angle in a fixed domain $\widehat{\Omega}$ as follows. Let $\widehat{\Omega}$ be the largest open subset of $\Omega \cap \widetilde{\Omega}$ which is symmetric with respect to the x-axis. Let $\hat{\nu}$ denote the exterior unit normal to $\widehat{\Omega}$ at points of $\partial\widehat{\Omega}$ where it exists. Define (variable) contact angles $\lambda, \tilde{\lambda} : \partial\widehat{\Omega} \to [0, \pi]$ as follows:

- On $\partial\widehat{\Omega} \cap \partial\Omega$, set $\lambda = \gamma$.
- On $\partial\widehat{\Omega} \cap \partial\widetilde{\Omega}$, set $\tilde{\lambda} = \gamma$.
- On $\partial\widehat{\Omega} \cap \Omega$, set $\lambda = Tf \cdot \hat{\nu}$ when $\hat{\nu}$ is defined; recall that $f \in C^2(\Omega)$.
- On $\partial\widehat{\Omega} \cap \widetilde{\Omega}$, set $\tilde{\lambda} = T\tilde{f} \cdot \hat{\nu}$ when $\hat{\nu}$ is defined; recall that $\tilde{f} \in C^2(\widetilde{\Omega})$.

Using the procedure given in [Shi and Finn 2004], notice that $\hat{\nu}$ exists at all but a finite number of points and so λ and $\tilde{\lambda}$ are defined almost everywhere on $\partial\widehat{\Omega}$. From Theorem 5.1 of [Finn 1986], we see that f and $\tilde{f}$ are the solutions of (1)–(2) with domain $\widehat{\Omega}$ and contact angles λ and $\tilde{\lambda}$ respectively. We may therefore view $\tilde{\lambda}$ as a perturbation of the (symmetric) contact angle λ and, when $\gamma < \alpha - \frac{\pi}{2}$, this perturbation $\tilde{\lambda}$ destroys the central fan. In this paper, we establish the stability of central fans with respect to sufficiently small perturbations of the contact angle γ, leaving the domain Ω fixed; this implies that $\tilde{\lambda}$ is a "large" perturbation of λ. We shall prove the following result.

Theorem 1. *Suppose Ω is a bounded open domain in $\mathbb{R}^2$ which has a reentrant corner at $\mathcal{O}$ of size 2α with $\alpha \in \left(\frac{\pi}{2}, \pi\right)$ as described above. Suppose also that there is a finite set $A = \{P_1, \ldots, P_m\} \subset \partial\Omega$ with $m \geq 1$ and $P_1 = \mathcal{O}$ such that $\partial\Omega \setminus A$ is a C^4 curve (if $m = 1$) or a finite disjoint union of C^4 curves (if $m > 1$). Let*

$\gamma \in C^{1,\beta}(\partial\Omega \setminus A)$, for some $\beta \in (0,1)$, satisfy $\delta_0 \le \gamma \le \pi - \delta_0$ for some $\delta_0 > 0$ such that the limits

$$\gamma_1 = \lim_{\partial^+\Omega \ni (x,y) \to \mathcal{O}} \gamma(x,y) \quad \text{and} \quad \gamma_2 = \lim_{\partial^-\Omega \ni (x,y) \to \mathcal{O}} \gamma(x,y)$$

both exist. Suppose there exists $f \in C^2(\Omega) \cap L^\infty(\Omega)$ which satisfies (1)–(2) and is discontinuous at $\mathcal{O}$ and the radial limit function of f at $\mathcal{O}$, $Rf(\,\cdot\,)$, behaves as in case (ID) or case (DI).

Then there exist functions $\omega^\pm : \partial\Omega \to [0,\pi]$ with $0 \le \omega^+ \le \gamma \le \omega^- \le \pi$ on $\partial\Omega$ and $\omega^+ < \gamma < \omega^-$ on $\partial\Omega \setminus A$ such that if $\sigma : \partial\Omega \to (0,\pi)$ with $\omega^+ \le \sigma \le \omega^-$ a.e. on $\partial\Omega$ and $\delta_1 \le \sigma \le \pi - \delta_1$ for some $\delta_1 \in (0,\delta_0)$, then the radial limit function Rh of the solution $h \in C^2(\Omega)$ of (1)–(2) with γ replaced by σ in (2) has the same type of behavior (i.e., case (ID) or case (DI) holds) as does Rf. In particular, the radial limits of h have a central fan.

The following corollary shows that Example 2 of [Lancaster and Siegel 1996] can be perturbed (with respect to the contact angle) and that the resulting nonsymmetric nonparametric capillary surfaces will have central fans.

Corollary 2. *Let Ω be an open, connected, bounded Lipschitz domain which is symmetric with respect to the x-axis such that $\mathcal{O} = (0,0) \in \partial\Omega$, $\partial\Omega \setminus \{\mathcal{O}\}$ is a C^4 curve and Ω has a corner at $\mathcal{O}$ with opening angle $2\alpha > \pi$. Suppose $\gamma : \partial\Omega \setminus \{\mathcal{O}\} \to (0,\pi)$ is a $C^{1,\beta}$ map which satisfies $\gamma(x,-y) = \gamma(x,y)$ for $(x,y) \in \partial\Omega$ and for which the limit*

$$\lim_{\partial\Omega \ni (x,y) \to \mathcal{O}} \gamma(x,y) = \gamma_0,$$

exists and $0 < \gamma_0 < \alpha - \frac{\pi}{2}$. Let $f \in C^2(\Omega) \cap C^{1,\beta}(\overline{\Omega} \setminus \{\mathcal{O}\})$ of (1)–(2). Then f is discontinuous at $\mathcal{O}$, the radial limit function Rf behaves as in case (DI) and there exist functions $\omega^\pm : \partial\Omega \to [0,\pi]$ with $0 \le \omega^+ \le \gamma \le \omega^- \le \pi$ on $\partial\Omega$ and $\omega^+ < \gamma < \omega^-$ on $\partial\Omega \setminus A$ such that if $\sigma : \partial\Omega \to (0,\pi)$ with $\omega^+ \le \sigma \le \omega^-$ a.e. on $\partial\Omega$ and $\delta_1 \le \sigma \le \pi - \delta_1$ for some $\delta_1 \in (0,\delta_0)$, then the radial limit function Rh of the solution $h \in C^2(\Omega)$ of (1)–(2) with γ replaced by σ in (2) is discontinuous at $\mathcal{O}$ and behaves as in case (DI).

We do not address the stability of the continuity at $\mathcal{O}$ of a solution f of (1)–(2) but we note that the procedure in [Shi and Finn 2004], as stated, would not establish the discontinuity at $\mathcal{O}$ of f for arbitrarily small perturbations of the domain (in the asymptotic sense of Shi and Finn) when $\gamma_1 = \gamma_2 > \alpha - \frac{\pi}{2}$.

2. Some lemmas

Lemma 3. *Let Ω be a bounded open domain in $\mathbb{R}^2$ with Lipschitz boundary and let Γ be an open subset of $\partial\Omega$ which is a $C^{2,\beta}$ curve for some $\beta \in (0,1)$. Let*

$\gamma \in L^\infty(\partial\Omega)$ satisfy $\delta \le \gamma \le \pi - \delta$ a.e. on $\partial\Omega$ for some $\delta > 0$ and $\gamma \in C^{1,\beta}(\Gamma)$. Suppose there exists $f \in C^2(\Omega) \cap L^\infty(\Omega)$ which satisfies

$$\operatorname{div}(Tu) = \kappa u \quad \text{in } \Omega \tag{4}$$

and

$$Tu \cdot \nu = \cos\gamma \quad \text{on } \Gamma. \tag{5}$$

Then $f \in C^{2,\beta}(\Omega \cup \Gamma)$.

See [Finn 1986, p. 210, Note 5], or [Finn 1988], or the introduction of [Korevaar and Simon 1996], which references [Simon and Spruck 1976; Taylor 1977].

The next result uses the notation of [Korevaar and Simon 1996, Theorem 2]; in particular,

$$\frac{(\nabla g(x), -1)}{\sqrt{1 + |\nabla g(x)|^2}}$$

denotes the continuous extension of the (downward) unit normal to the graph of g when considered as a function on this graph.

Lemma 4. *Let Ω be a bounded open domain in $\mathbb{R}^2$ with Lipschitz boundary and let Γ be an open subset of $\partial\Omega$ which is a C^3 curve. Let $\phi \in L^\infty(\partial\Omega)$ be in $C^{1,\beta}(\Gamma)$ for some $\beta \in (0, 1)$. Suppose $g \in C^2(\Omega) \cap L^\infty(\Omega)$ is the variational solution of*

$$\begin{aligned} \operatorname{div}(Tu) &= \kappa u && \text{in } \Omega, \\ u &= \phi && \text{on } \partial\Omega; \end{aligned}$$

that is, g minimizes $J(\cdot)$ over $\mathrm{BV}(\Omega)$, where

$$J(u) = \int_\Omega \sqrt{1 + |Du|^2} + \int_\Omega \int_0^u \kappa t \, dt \, dx + \int_{\partial\Omega} |u - \phi| \, ds, \quad u \in \mathrm{BV}(\Omega).$$

Set

$$Q = \{(x, t) \in \Gamma \times \mathbb{R} : \min\{\phi(x), g(x)\} \le t \le \max\{\phi(x), g(x)\}\}$$

and $Q_0 = Q \setminus T$, where $T \subset \partial\Omega \times \mathbb{R}$ is the graph of ϕ, and let G be the graph of g over Ω. Then for each $x_0 \in \Gamma$, there exists a $\delta > 0$ such that $\{x \in \partial\Omega : |x - x_0| \le \delta\} \subset \Gamma$ and the following conclusions hold:

(a) *$\Pi = \{(x, t) \in Q \cup G : |x - x_0| \le \delta\}$ is a $C^{1,\sigma}$ manifold with boundary whose boundary is the union of $\{(x, \phi(x)) \in T : x \in \Gamma,\ |x - x_0| \le \delta\}$ and $\{(x, g(x)) : x \in \Omega,\ |x - x_0| = \delta\}$ for some $\sigma \in (0, 1)$.*

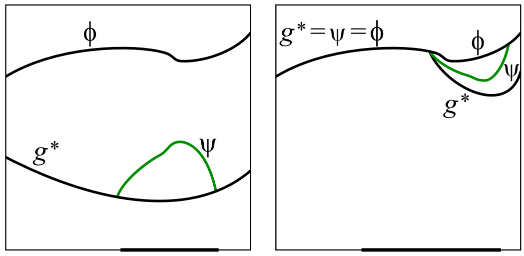
(b) *The (downward) unit normal $\vec{N}$ to Π is a continuous function and*

$$\vec{N}(x,t)=\begin{cases}\dfrac{(\nabla g(x),-1)}{\sqrt{1+|\nabla g(x)|^2}} & \text{if } x\in\Omega\cup\Gamma \text{ and } t=g(x),\\ (\nu(x),0) & \text{if } (x,t)\in Q \text{ and } g(x)\le t<\phi(x),\\ (-\nu(x),0) & \text{if } (x,t)\in Q \text{ and } \phi(x)<t\le g(x),\end{cases}$$

where ν denotes the outward unit normal to $\partial\Omega$.

Proof. Let $A=\{x\in\Gamma: g^*(x)=\phi(x)\}$, $B=\{x\in\Gamma: g^*(x)\neq\phi(x)\}$, and A_0 be the interior (in Γ) of A, where g^* is the trace of g on $\partial\Omega$; let us define $g^*(x)$ to be $\phi(x)$ if $x\in\Gamma$ and $g^*(x)$ is not otherwise defined. Using the arguments in [Elcrat and Lancaster 1986], we see that if $x_0\in A_0$, then there exists a $\delta>0$ such that $\{x\in\partial\Omega: |x-x_0|\le\delta\}\subset A_0$ and (a) and (b) hold.

Suppose $x_0\in B$ such that $g^*(x_0)=z_0<\phi(x_0)$ and so (x_0,z_0) is an interior point of $\mathcal{Q}\cup G$. Standard results on the regularity of solutions of obstacle problems at interior points imply g is continuous on $(\overline{\Omega}\times\mathbb{R})\cap U$, where U is a neighborhood in $\mathbb{R}^3$ of (x_0,z_0), and, considered as a parametric surface, $(\mathcal{Q}\cup G)\cap U$ is a $C^{1,\alpha}$ surface for some $\alpha\in(0,1)$. (For example, this follows from [Simon and Spruck 1976] or [Taylor 1977], since the contact angle is zero at these interior points. Another argument follows from [Hildebrandt 1973]; by "blowing up" or dilating $\mathbb{R}^3$ about (x_0,z_0), we may assume the function $f: E\times\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}^3\to\mathbb{R}$ given by $f(w,X,p,q)=|p|^2+|q|^2+\frac{1}{2}\kappa(X_3-z_0)((X-(x_0,z_0))\cdot(p\times q))$ satisfies conditions A and B of that paper in a neighborhood $U\times I$ of (x_0,z_0), where I is an open interval containing z_0, and so, for smooth Dirichlet data ψ slightly larger than z_0 near x_0 and equal to g on $\Omega\cap\partial U$, a theorem in [Hildebrandt 1973] shows there is a parametric minimizer of $\int_E f(w,z,\nabla z)\,du\,dv$ that is smooth in the interior of its domain $E=\{(u,v): u^2+v^2<1\}$ and, from [Miranda 1964] (or [Finn 1986, p. 16, Note 10]), we see that this parametric solution is a graph $z=h(x,y)$. Since $g\le h$ on ∂U by the choice of ψ and $h\le g$ on ∂U since $\psi<\phi$ on $U\cap\partial\Omega$, we see that $g=h$ in U. In particular, g is continuous at each point of $\Gamma\cap B$ and the points of $A\setminus A_0$ are isolated.) An application of [Bourni 2011] shows that (a) and (b) hold; that is, we may choose a domain $\mathcal{V}\subset\Omega$ such that $\partial\mathcal{V}$ is a $C^{1,\alpha}$ curve in $\mathbb{R}^2$, $x_0\in\partial\mathcal{V}\subset\Gamma\cup\Omega$, $\partial\Omega\cap\overline{(\Omega\cap\partial\mathcal{V})}=\{x^{(j)}: j=1,2\}$ with $x^{(j)}\in B$ $(j=1,2)$, the closure in $\mathbb{R}^3$ of $\{(x,g(x))\in\Omega\times\mathbb{R}: x\in\partial\mathcal{V}\}$ is a $C^{1,\alpha}$ curve in $\mathbb{R}^3$ which meets $\Gamma\times\mathbb{R}$ tangentially at $(x^{(j)},g^*(x^{(j)}))$, $j=1,2$, and we can find a function $\psi:\partial\mathcal{V}\to\mathbb{R}$ whose graph is a $C^{1,\alpha}$ curve in $\mathbb{R}^3$ such that $g^*\le\psi\le\phi$ on $\partial\Omega\cap\partial\mathcal{V}$ (see Figure 1(a)) and apply the conclusion of [Bourni 2011] to see that (a) and (b) hold in a neighborhood of x_0. If $x_0\in B$ such that $g^*(x_0)=z_0>\phi(x_0)$, apply the argument above to $-g$ (with $-\phi$ as Dirichlet data).



(a) $\partial\nu\cap\partial\Omega$ (b) $\partial\nu\cap\partial\Omega$

Figure 1. The traces of g, ψ and ϕ.

Suppose $x_0 \in A \setminus A_0$. Notice, from the arguments above, that A_0 and B are open. There exist a domain $\mathcal{V} \subset \Omega$ and a function $\psi : \partial\mathcal{V} \to \mathbb{R}$ as above such that $x_0 \in \partial\mathcal{V} \subset \Gamma \cup \Omega$ and $\partial\Omega \cap \overline{(\Omega \cap \partial\mathcal{V})} = \{x^{(j)} : j = 1, 2\}$ with $x^{(j)} \in A_0 \cup B$ $(j = 1, 2)$; we argue as above (see, for example, Figure 1(b)). □

Lemma 5. *Let Ω be a bounded open domain in $\mathbb{R}^2$ with Lipschitz boundary and let Γ be an open subset of $\partial\Omega$ which is a C^4 curve or a finite disjoint union of C^4 curves. Let $\gamma \in L^\infty(\partial\Omega)$ satisfy $\delta \le \gamma \le \pi - \delta$ a.e. on $\partial\Omega$ for some $\delta > 0$ and $\gamma \in C^{1,\beta}(\Gamma)$ for some $\beta \in (0, 1)$. Suppose there exists $f \in C^2(\Omega) \cap L^\infty(\Omega)$ which satisfies* (4) *and* (5). *Let $\epsilon > 0$. Define $g = g_\epsilon \in \mathrm{BV}(\Omega)$ to be the minimizer over $\mathrm{BV}(\Omega)$ of $J_\epsilon(\,\cdot\,)$, where*

$$J_\epsilon(u) = \int_\Omega \sqrt{1 + |Du|^2} + \int_\Omega \int_0^u kt\,dt\,dx + \int_{\partial\Omega} |u - (f + \epsilon)|\,ds$$

for $u \in \mathrm{BV}(\Omega)$. We have:

(i) *$g \in C^2(\Omega)$ and satisfies* (4).

(ii) *The unit normal $\vec{N}$ to Π is in $C^{0,\beta}(\Omega \cup E)$ for each compact subset E of Γ and hence the contact angle*

$$\gamma_g \stackrel{\text{def}}{=} \arccos(Tg \cdot \nu) \in [0, \pi] \tag{6}$$

is well defined and continuous on Γ, where ν denotes the outward unit normal to $\partial\Omega$. In particular, $\gamma_g = 0$ on $\{x \in \Gamma : g(x) < f(x) + \epsilon\}$.

(iii) *Suppose there is a finite set $A = \{x_1, \dots, x_m\} \subset \partial\Omega$ such that $\Gamma = \partial\Omega \setminus A$. Then $f \le g \le f + \epsilon$ in Ω.*

(iv) *Suppose there is a finite set $A = \{x_1, \dots, x_m\} \subset \partial\Omega$ such that $\Gamma = \partial\Omega \setminus A$. Then $\gamma_g < \gamma$ on Γ.*

Proof. (i) The existence of g follows from Theorem 5 of [Gerhardt 1974], or Theorem 2.1 of [Giusti 1976]. The interior regularity of g follows from Theorem 3.1

of [Giusti 1976] (see also [Gerhardt 1974, p. 174; Williams 1978, Theorem 3]). The fact that g satisfies (4) is standard (e.g., [Gerhardt 1974, p. 174]).

(ii) The boundary regularity of g follows from Lemma 4. On $\{x \in \Gamma : g(x) < f(x)+\epsilon\}$, we have $\vec{N}(x, g(x)) = (\nu(x), 0)$, $Tg(x) = \nu(x)$, and so

$$\gamma_g(x) = \arccos(\nu(x) \cdot \nu(x)) = \arccos(1) = 0.$$

(iii) Notice that $f, g \in C^2(\Omega) \cap C^0(\Omega \cup \Gamma)$. Set $M = \{x \in \Omega : f(x) > g(x)\}$. On $\partial M \cap \Gamma$, $g < f + \epsilon$ and so by (ii) and Lemma 4, with $\Pi = Q \cup G$, where $Q = \{(x, z) : (x, z) \in E,\ g(x) \leq z < f(x)\} \in M \times \mathbb{R}$, implies that $\sup \cos \gamma_g = Tg \cdot \nu = 1$; hence $\gamma_g = 0$ on $\partial M \cap \Gamma$. Thus $f = g$ on $\Omega \cap \partial M$ and $\gamma_g = 0$ almost everywhere on $\partial\Omega \cap \partial M$ and so the general comparison principle (e.g., [Finn 1986, Theorem 5.1]) implies $f \leq g$ in M; hence $M = \varnothing$.

Now let $\tau > 0$ and set $N = \{x \in \Omega : g(x) > f(x)+\epsilon+\tau\}$. Then $g = f+\epsilon+\tau$ on $\Omega \cap \partial N$ and $g > f+\epsilon$ on $\partial N \cap \Gamma$ and so Lemma 4 implies $\gamma_g = \pi$ almost everywhere on $\partial\Omega \cap \partial N$. The general comparison principle then implies $g \leq f + \epsilon + \tau$ and so $N = \varnothing$. Therefore $g \leq f + \epsilon + \tau$ in Ω for each $\tau > 0$ and so $g \leq f + \epsilon$ in Ω.

(iv) Suppose first $x \in \Gamma$ and there is a sequence $\{y_j\}$ in Γ such that $x = \lim_{j\to\infty} y_j$ and $g(y_j) < f(y_j) + \epsilon$ for each j. Then (ii) implies $\gamma_g(y_j) = 0$ for each j and so $\gamma_g(x) = 0$. Since $\gamma \in (0, \pi)$, we see that $\gamma_g(x) = 0 < \gamma(x)$.

Suppose next that $x \in \Gamma$ and $g \geq f + \epsilon$ in $\mathcal{P} \cap \Gamma$, where $\mathcal{P}$ is a neighborhood of x in $\mathbb{R}^2$. From (iii), we see that $g = f + \epsilon$ in $\mathcal{P} \cap \Gamma$. If $\gamma_g(x) > \gamma(x)$, then $g(x - t\nu(x)) > f(x - t\nu(x)) + \epsilon$ for $t > 0$ small and this contradicts (iii). (Recall that $\nu(x)$ is the exterior unit normal to $\partial\Omega$ at x.) Thus $\gamma_g \leq \gamma$ on Γ.

Finally, suppose $x \in \Gamma$, $\gamma_g(x) = \gamma(x)$ and $g = f + \epsilon$ in $\mathcal{P} \cap \partial\Omega$, where $\mathcal{P}$ is a neighborhood of x in $\mathbb{R}^2$; notice that [Heinz 1970] and Lemma 3 imply $g \in C^{2,\beta}(\mathcal{P} \cap \overline{\Omega})$. Since $g \leq f + \epsilon$ in Ω and $\gamma_g(x) = \gamma(x)$, the tangent plane Π_g to $z = g$ at $(x, g(x))$ and the tangent plane Π to $z = f+\epsilon$ at $(x, g(x)) = (x, f(x)+\epsilon)$ must coincide. Now the mean curvature H_g of $z = g$ at $(x, g(x))$ is $\frac{1}{2}\kappa g(x)$ and the mean curvature H_f of $z = f + \epsilon$ at $(x, g(x))$ is $\frac{1}{2}\kappa f(x) = \frac{1}{2}(\kappa g(x) - \kappa\epsilon)$. Since $g = f + \epsilon$ in $\mathcal{P} \cap \Gamma$, the (signed) curvature of the curve $z = f(x - t\nu(x)) + \epsilon$ must be strictly less than the (signed) curvature of the curve $z = g(x - t\nu(x))$ for $t > 0$ small and so $g(x - t\nu(x)) > f(x - t\nu(x)) + \epsilon$ for $t > 0$ small, contradicting (iii). □

3. Stability of central fans

We will begin by establishing the stability of the central fans with respect to "one-sided" perturbations of γ. (See Figures 2 and 3.)

Theorem 6. *Let Ω be an open, connected, bounded Lipschitz domain such that $\mathcal{O} = (0, 0) \in \partial\Omega$, Ω has a corner at $\mathcal{O}$ with opening angle $2\alpha > \pi$ and there is a finite set $A = \{P_1, \ldots, P_m\} \subset \partial\Omega$ with $m \geq 1$ and $P_1 = \mathcal{O}$ such that $\partial\Omega \setminus A$ is*

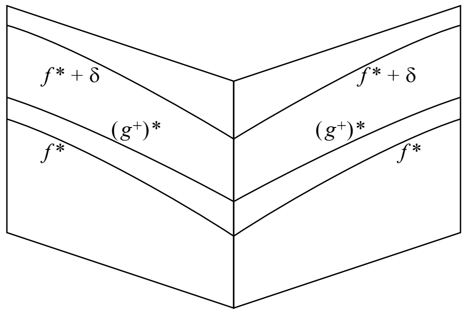


Figure 2. The traces of f, g^+ and $f + \delta$.

a C^4 curve (if $m = 1$) or a finite disjoint union of C^4 curves (if $m > 1$). Suppose $\gamma : \partial\Omega \setminus \mathcal{O} \to (0, \pi)$ is a $C^{1,\beta}$ map for which the limits

$$\lim_{\partial^+\Omega \ni (x,y) \to \mathcal{O}} \gamma(x, y) = \gamma_1, \qquad \lim_{\partial^-\Omega \ni (x,y) \to \mathcal{O}} \gamma(x, y) = \gamma_2$$

exist with $\gamma_i \in (0, \pi)$, $i = 1, 2$, and $f \in C^2(\Omega) \cap C^{1,\beta}(\overline{\Omega} \setminus \{\mathcal{O}\})$ satisfies

$$\begin{aligned} \operatorname{div}(Tf) &= \kappa f && \text{in } \Omega, \\ Tf \cdot \nu &= \cos\gamma && \text{on } \partial\Omega \setminus \{\mathcal{O}\} \end{aligned}$$

such that f is discontinuous at $\mathcal{O}$ and the radial limits $Rf(\cdot)$ of f at $\mathcal{O}$ have a central fan.

There exists a $\delta > 0$ such that if $g^+ \in \mathrm{BV}(\Omega) \cap C^2(\Omega)$ is the variational solution of the Dirichlet problem

$$\operatorname{div}(Tg) = \kappa g \quad \text{in } \Omega, \tag{7}$$

$$g = f + \delta \quad \text{on } \partial\Omega \setminus A, \tag{8}$$

and if $\omega^+ \stackrel{\text{def}}{=} \arccos(Tg^+ \cdot \nu)$ on $\partial\Omega \setminus A$, then for any function $\sigma \in L^\infty(\partial\Omega)$ satisfying

$$\omega^+ \le \sigma \le \gamma \quad \text{a.e. on } \partial\Omega \tag{9}$$

and $\delta_1 \le \sigma \le \pi - \delta_1$ for some $\delta_1 > 0$, the variational solution $h \in \mathrm{BV}(\Omega) \cap C^2(\Omega)$ of the capillary problem

$$\operatorname{div}(Th) = \kappa h \ \text{ in } \Omega, \qquad Th \cdot \nu = \cos\sigma \ \text{ on } \partial\Omega \setminus A \tag{10}$$

is discontinuous at $\mathcal{O}$, the radial limits $Rh(\cdot)$ of h at $\mathcal{O}$ have a central fan and they have the same type of behavior (i.e., case (DI) or (ID)) as $Rf(\cdot)$.

Proof. Suppose first that Rf behaves as in case (DI) and so

$$Rf(0) < \min\{Rf(\alpha), Rf(-\alpha)\}.$$

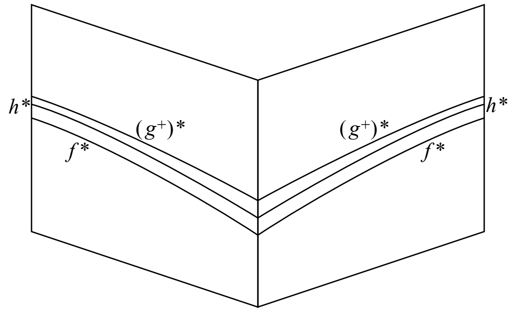


Figure 3. The traces of f, h and g^+.

Let $\delta < \min\{Rf(\alpha) - Rf(0),\ Rf(-\alpha) - Rf(0)\}$ and let g^+ be the variational solution of the Dirichlet problem (7)–(8) for this choice of δ (see Figure 2). From Lemma 5(iv), we see that $\omega^+ < \gamma$ on $\partial\Omega \setminus A$ and therefore there exist $\sigma \in L^\infty(\partial\Omega)$ which satisfy (9); let us select σ and h as in the theorem (see Figure 3). From Lemma 5(iii) and the general comparison principle, we see that $f \le h \le g^+ \le f + \delta$ in Ω and hence

$$Rf(\theta) \le Rh(\theta) \le Rg^+(\theta) \le Rf(\theta) + \delta \quad \text{for } \theta \in [-\alpha, \alpha]; \tag{11}$$

thus

$$Rh(\alpha) - Rh(0) \ge Rf(\alpha) - (Rf(0) + \delta) = Rf(\alpha) - Rf(0) - \delta > 0$$

and

$$Rh(-\alpha) - Rh(0) \ge Rf(-\alpha) - (Rf(0) + \delta) = Rf(-\alpha) - Rf(0) - \delta > 0.$$

Now we know that the radial limits of h at $\mathcal{O}$ exist and behave as in [Lancaster and Siegel 1996] (i.e., one of case (I), (D), (ID) or (DI) must hold; if, for example, case (I) held, one would have $Rf(-\alpha) < Rf(0) < Rf(\alpha)$). The calculations above show that $Rf(-\alpha) > Rf(0)$ and $Rf(\alpha) > Rf(0)$ and therefore $Rh(\cdot)$ must behave as in case (DI); hence $Rh(\cdot)$ has a central fan (see Figure 4).

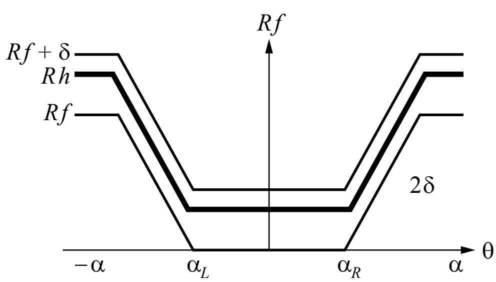


Figure 4. The radial limits of f, h and $f + \delta$.

Suppose next that Rf behaves as in case (ID) and so

$$Rf(0) > \min\{Rf(\alpha),\, Rf(-\alpha)\}.$$

If we let $\delta < \min\{Rf(0) - Rf(\alpha),\ Rf(0) - Rf(-\alpha)\}$, we may repeat the same argument as above and obtain $Rh(0) > Rh(-\alpha)$ and $Rh(0) > Rh(\alpha)$; hence $Rh(\,\cdot\,)$ must behave as in case (ID) and therefore has a central fan. □

Remark 7. The corresponding theorem with (8) replaced by

$$g^- = f - \delta \quad \text{on } \partial\Omega \setminus A \tag{12}$$

and with $\omega^- \overset{\text{def}}{=} \arccos(Tg^- \cdot \nu)$ on $\partial\Omega \setminus A$, $\omega^- \ge \sigma \ge \gamma$ on $\partial\Omega \setminus A$, $\delta_1 \le \sigma \le \pi - \delta_1$ for some $\delta_1 > 0$ and h a solution of (10) yields $f - \delta \le g^- \le h \le f$ in Ω and

$$Rf(\theta) \le Rh(\theta) \le Rf(\theta) + \delta \quad \text{for } \theta \in [-\alpha, \alpha]; \tag{13}$$

hence h is discontinuous at $\mathcal{O}$ and the radial limits $Rh(\,\cdot\,)$ of h at $\mathcal{O}$ have the same type of behavior (i.e., case (DI) or (ID)) as $Rf(\,\cdot\,)$.

Proof of Theorem 1. Suppose Rf behaves as in case (DI) or case (ID) and define $\delta = \frac{1}{2}\min\{|Rf(\alpha) - Rf(0)|, |Rf(-\alpha) - Rf(0)|\}$. Combining the arguments in Theorem 6 and Remark 7, we obtain

$$Rf(\alpha) - Rf(0) - \delta \le Rh(\alpha) - Rh(0) \le Rf(\alpha) - Rf(0) + \delta$$

and

$$Rf(-\alpha) - Rf(0) - \delta \le Rh(\alpha) - Rh(0) \le Rf(-\alpha) - Rf(0) + \delta.$$

If Rf behaves as in case (DI), we have $0 < Rh(\alpha) - Rh(0)$ and $0 < Rh(-\alpha) - Rh(0)$ and therefore Rh behaves as in case (DI). If Rf behaves as in case (ID), we have $Rh(\alpha) - Rh(0) < 0$ and $Rh(-\alpha) - Rh(0) < 0$ and therefore Rh behaves as in case (ID). □

Proof of Corollary 2. Since $\gamma_0 < \alpha - \frac{\pi}{2}$, we see that $|2\gamma - \pi| > 2\pi - 2\alpha$. It follows from [Lancaster 2012] that f is discontinuous at $\mathcal{O}$. Since $f(x, y) = f(x, -y)$ for each $(x, y) \in \Omega$, the radial limits of f cannot behave as in cases (I) or (D) of Theorem 1 of [Lancaster and Siegel 1996] and therefore they have a central fan. That case (DI) holds for $Rf(\,\cdot\,)$ follows from [Lancaster 2012] or directly from the fact that $(\pi - \gamma_0) + (\pi - \gamma_0) + \pi > 4\pi - 2\alpha > 2\alpha$ means case (ID) cannot hold. The claim follows from Theorem 1. □

Remark 8. It should be emphasized that the conclusion of Theorem 1 is not that "there exists a $\delta > 0$ such that if $\sigma : \partial\Omega \to [0, \pi]$ satisfies $\gamma - \delta \le \sigma \le \gamma + \delta$ a.e. on $\partial\Omega$, then the radial limit function Rh of the solution $h \in C^2(\Omega)$ of (1)–(2) with γ replaced by σ in (2) has the same type of behavior (i.e., case (ID) or case (DI) holds) as does Rf". The validity of such a conclusion is an interesting question which might spur further investigation.

Acknowledgements

The author wishes to thank Professor K. Lancaster for his valuable comments and suggestions throughout this research, and also to thank the editor and referee for their patience and to acknowledge the contributions of the referee, who provided many helpful comments and Figure 4.

References

[Bourni 2011] T. Bourni, "$C^{1,\alpha}$ theory for the prescribed mean curvature equation with Dirichlet data", *J. Geom. Anal.* **21**:4 (2011), 982–1035. MR 2836589 Zbl 1232.53009

[Elcrat and Lancaster 1986] A. R. Elcrat and K. E. Lancaster, "Boundary behavior of a nonparametric surface of prescribed mean curvature near a reentrant corner", *Trans. Amer. Math. Soc.* **297**:2 (1986), 645–650. MR 87h:35098 Zbl 0602.35042

[Finn 1986] R. Finn, *Equilibrium capillary surfaces*, Grundlehren der Mathematischen Wissenschaften **284**, Springer, New York, 1986. MR 88f:49001 Zbl 0583.35002

[Finn 1988] R. Finn, "Moon surfaces, and boundary behaviour of capillary surfaces for perfect wetting and nonwetting", *Proc. London Math. Soc.* (3) **57**:3 (1988), 542–576. MR 89m:49076 Zbl 0668.76019

[Gerhardt 1974] C. Gerhardt, "Existence, regularity, and boundary behaviour of generalized surfaces of prescribed mean curvature", *Math. Z.* **139** (1974), 173–198. MR 55 #10846 Zbl 0316.49005

[Giusti 1976] E. Giusti, "Boundary value problems for non-parametric surfaces of prescribed mean curvature", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **3**:3 (1976), 501–548. MR 58 #2572 Zbl 0344.35036

[Heinz 1970] E. Heinz, "Über das Randverhalten quasilinearer elliptischer Systeme mit isothermen Parametern", *Math. Z.* **113** (1970), 99–105. MR 41 #7288 Zbl 0176.41004

[Hildebrandt 1973] S. Hildebrandt, "Interior $C^{1+\alpha}$-regularity of solutions of two-dimensional variational problems with obstacles", *Math. Z.* **131** (1973), 233–240. MR 51 #1568 Zbl 0275.49007

[Korevaar and Simon 1996] N. Korevaar and L. Simon, "Equations of mean curvature type with contact angle boundary conditions", pp. 175–201 in *Geometric analysis and the calculus of variations*, edited by J. Jost, International Press, Cambridge, MA, 1996. MR 98e:35071 Zbl 0932.35091

[Lancaster 1985] K. E. Lancaster, "Boundary behavior of a nonparametric minimal surface in $\mathbb{R}^3$ at a non-convex point", *Analysis* **5**:1-2 (1985), 61–69. Corrigendum in **6**:4 (1986), 413. MR 86m:49053 Zbl 0601.35035

[Lancaster 2010] K. E. Lancaster, "A proof of the Concus–Finn conjecture", *Pacific J. Math.* **247**:1 (2010), 75–108. MR 2011m:53011 Zbl 1256.53009

[Lancaster 2012] K. E. Lancaster, "Remarks on the behavior of nonparametric capillary surfaces at corners", *Pacific J. Math.* **258**:2 (2012), 369–392. MR 2981959 Zbl 1253.35069

[Lancaster and Siegel 1996] K. E. Lancaster and D. Siegel, "Existence and behavior of the radial limits of a bounded capillary surface at a corner", *Pacific J. Math.* **176**:1 (1996), 165–194. MR 98g:58030a Zbl 0866.76018

[Miranda 1964] M. Miranda, “Superfici Cartesiane generalizzate ed insiemi di perimetro localmente finito sui prodotti Cartesiani”, *Ann. Scuola Norm. Sup. Pisa* (3) **18** (1964), 515–542. MR 30 #4906 Zbl 0152.24402

[Shi and Finn 2004] D. Shi and R. Finn, “On a theorem of Lancaster and Siegel”, *Pacific J. Math.* **213**:1 (2004), 111–119. MR 2004m:76038 Zbl 1156.76361

[Simon and Spruck 1976] L. Simon and J. Spruck, “Existence and regularity of a capillary surface with prescribed contact angle”, *Arch. Rational Mech. Anal.* **61**:1 (1976), 19–34. MR 58 #7339 Zbl 0361.35014

[Taylor 1977] J. E. Taylor, “Boundary regularity for solutions to various capillarity and free boundary problems”, *Comm. Partial Differential Equ.* **2**:4 (1977), 323–357. MR 58 #7336 Zbl 0357.35010

[Williams 1978] G. H. Williams, “Surfaces of prescribed mean curvature with inequalities on the boundary”, *Math. Z.* **164**:1 (1978), 31–51. MR 80d:53006 Zbl 0378.49029

Received August 30, 2013. Revised August 11, 2015.

AMMAR KHANFER
DEPARTMENT OF MATHEMATICS AND STATISTICS
AL-IMAM MOHAMMAD IBN SAUD ISLAMIC UNIVERSITY
P.O. BOX 90950
AL-RIYADH 11623
SAUDI ARABIA
a.khanfer@yahoo.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 281, No. 2, 2016
dx.doi.org/10.2140/pjm.2016.281.481

SURFACES OF PRESCRIBED MEAN CURVATURE $H(x, y, z)$ WITH ONE-TO-ONE CENTRAL PROJECTION ONTO A PLANE

FRIEDRICH SAUVIGNY

Dedicated to the memory of Professor Stefan Hildebrandt in gratitude

When we consider surfaces of prescribed mean curvature H with a one-to-one orthogonal projection onto a plane, we have to study the nonparametric H-surface equation. Now the H-surfaces with a one-to-one central projection onto a plane lead to an interesting elliptic differential equation, which is derived in Section 2; in the case $H = 0$ this PDE was invented by T. Radó. We establish the uniqueness of the Dirichlet problem for this H-surface equation in central projection in Section 3, and develop an estimate for the maximal deviation of large H-surfaces from their boundary values, resembling an inequality by J. Serrin. In Section 4 we provide a Bernstein-type result for the case $H = 0$ and classify the entire solutions of the minimal surface equation in central projection. We also solve the Dirichlet problem for $H = 0$ by a variational method. In Section 5 we solve the Dirichlet problem for nonvanishing H with compact support via a nonlinear continuity method, and we construct large H-surfaces bounding extreme contours by an approximation. Finally, in Section 6 we solve the Dirichlet problem on discs for the nonparametric H-surface equation in central projection under certain restrictions for the mean curvature.

1. Introduction

In Plateau's problem for variable $H = H(x, y, z)$, one constructs branched immersions of prescribed mean curvature $H(x, y, z)$ bounding a given Jordan contour Γ in $\mathbb{R}^3$ by minimizing an energy functional (see [Dierkes et al. 2010a, Part II]). This parametric H-surface

$$X = X(u, v) = (x(u, v), y(u, v), z(u, v))$$

satisfies Rellich's nonlinear elliptic system (3-31) and is given in conformal parameters — apart from the isolated branch points. In [Sauvigny 1982], this parametric

MSC2010: 35J60, 53A10.
Keywords: surfaces of prescribed mean curvature, solution of the Dirichlet problem, H-surface equation in central projection, large H-surfaces for Plateau's problem.

surface X is shown to be a graph $z = \zeta(x, y)$ above the (x, y)-plane for certain contours Γ and solves the Dirichlet problem for the nonparametric H-surface equation to the given boundary values.

In the present paper we solve the Dirichlet problem for H-surfaces in the representation (2-3) with a one-to-one central projection by a nonlinear continuity method (compare Theorem 5.1) and an approximation (see Theorem 6.1). We start with a solution of Plateau's problem for $H = 0$ which possesses a one-to-one central projection (see Theorem 4.1). Having answered the uniqueness question (compare Theorem 3.1), we study intensively the stability and the compactness of this boundary value problem with the aid of [Sauvigny 1982, Satz 1]. In the minimal surface case, the relevant PDEs (2-17) and (2-24) already appear in a paper by T. Radó [1932] — but the inhomogeneous equations seem to be investigated here for the first time.

We can determine the set of entire solutions for the nonparametric minimal surface equation in central projection (compare Theorem 4.2). While minimal surfaces remain in the convex hull of their bounding contour, this is not the case for surfaces of prescribed mean curvature. However, we can estimate the deviation of our solution from their boundary values by comparison with large spherical caps (see Theorems 2.1 and 3.2). These surfaces do not belong to the family of graphs; however, they possess a one-to-one central projection and can be used here. Moreover, we can construct large solutions of Plateau's problem by a continuity and approximation method (compare Theorem 5.2).

2. The *H*-surface equation in central projection

It is well-known that the set of surfaces of constant mean curvature $H \in \mathbb{R}$ is invariant under translations and rotations. When we consider these H-surfaces with one-to-one central projection onto a plane $\mathcal{E}$, we can assume by translation that the origin $(0, 0, 0) \in \mathbb{R}^3$ represents the center of projection. Furthermore, we can attain by rotation that this plane $\mathcal{E}$ is parallel to the xy-plane. Now H-surfaces are transformed into $(a^{-1} \cdot H)$-surfaces after a dilation by the factor $a \in \mathbb{R} \setminus \{0\}$. Therefore, we can select the set

$$\mathcal{E} := \{(x, y, 1) \in \mathbb{R}^3 \mid x, y \in \mathbb{R}\} \tag{2-1}$$

without loss of generality as our *projection plane* after a suitable dilation. For the general study of H-surfaces with prescribed mean curvature $H = H(x, y, z)$ on domains in the Euclidean space, we refer our readers to Chapter 5 of the treatise [Dierkes et al. 2010a] by U. Dierkes, S. Hildebrandt, and F. Sauvigny.

Choose an arbitrary domain $\Omega \subset \mathbb{R}^2$ in the plane; with the real-valued function

$$\varrho = \varrho(x, y) \in C^2(\Omega, \mathbb{R}) \tag{2-2}$$

we associate the vector-valued function

$$X = X(x, y) := \varrho(x, y) \cdot (x, y, 1), \quad (x, y) \in \Omega. \tag{2-3}$$

At all points $(x, y) \in \Omega$ with $\varrho(x, y) \neq 0$, we obtain in (2-3) a differential-geometrically *regular surface with one-to-one central projection onto the plane* $\mathcal{E}$.

In this context, let us calculate the first derivatives of the surface X, namely

$$\begin{aligned} X_x(x, y) &= \varrho_x(x, y) \cdot (x, y, 1) + \varrho(x, y) \cdot (1, 0, 0), \\ X_y(x, y) &= \varrho_y(x, y) \cdot (x, y, 1) + \varrho(x, y) \cdot (0, 1, 0), \end{aligned} \tag{2-4}$$

and the *coefficients of its first fundamental form*, which are

$$\begin{aligned} X_x^2(x, y) &= X_x \cdot X_x(x, y) \\ &= \varrho_x^2(x, y) \cdot (x^2 + y^2 + 1) + 2\varrho(x, y) x \varrho_x(x, y) + \varrho^2(x, y), \\ X_x \cdot X_y &= \varrho_x(x, y)\varrho_y(x, y)(x^2+y^2+1) + \varrho(x, y)[y\varrho_x(x, y) + x\varrho_y(x, y)], \\ X_y^2(x, y) &= \varrho_y^2(x, y) \cdot (x^2 + y^2 + 1) + 2\varrho(x, y) y \varrho_y(x, y) + \varrho^2(x, y), \end{aligned} \tag{2-5}$$

for $(x, y) \in \Omega$. Furthermore, we determine the exterior product of the vectors (2-4) as follows:

$$\begin{aligned} X_x \wedge X_y(x, y) \\ &= \bigl(\varrho_x \cdot (x, y, 1) + \varrho \cdot (1, 0, 0)\bigr) \wedge \bigl(\varrho_y \cdot (x, y, 1) + \varrho \cdot (0, 1, 0)\bigr) \\ &= \varrho\varrho_x(x, y) \cdot (-1, 0, x) + \varrho\varrho_y(x, y) \cdot (0, -1, y) + \varrho^2(x, y) \cdot (0, 0, 1) \\ &= \varrho(x, y) \cdot \bigl(-\varrho_x(x, y), -\varrho_y(x, y), \varrho(x, y) + x\varrho_x(x, y) + y\varrho_y(x, y)\bigr). \end{aligned} \tag{2-6}$$

The *surface element* $W(x, y)$ is given by

$$\begin{aligned} W(x, y)^2 := |X_x \wedge X_y(x, y)|^2 \\ &= \varrho^2(x, y) \cdot \bigl(|\nabla\varrho(x, y)|^2 + [\varrho(x, y) + x\varrho_x + y\varrho_y]^2\bigr) \\ &= \varrho^2(x, y) \cdot \bigl(\varrho^2(x, y) + (1 + x^2)\varrho_x^2 + (1 + y^2)\varrho_y^2 \\ &\qquad + 2xy\varrho_x\varrho_y + 2x\varrho\varrho_x + 2y\varrho\varrho_y\bigr). \end{aligned} \tag{2-7}$$

Therefore, the equivalence

$$X_x \wedge X_y(x, y) \neq 0 \quad \text{if and only if} \quad \varrho(x, y) \neq 0, \tag{2-8}$$

which we have already used above, holds true.

We determine the second derivatives of our surface (2-3) via (2-4) and obtain

$$\begin{aligned} X_{xx}(x, y) &= \varrho_{xx}(x, y) \cdot (x, y, 1) + 2\varrho_x(x, y) \cdot (1, 0, 0), \\ X_{xy}(x, y) &= \varrho_{xy}(x, y) \cdot (x, y, 1) + (\varrho_y(x, y), \varrho_x(x, y), 0), \\ X_{yy}(x, y) &= \varrho_{yy}(x, y) \cdot (x, y, 1) + 2\varrho_y(x, y) \cdot (0, 1, 0). \end{aligned} \tag{2-9}$$

With the aid of (2-6) and (2-9), we determine the *coefficients of its second fundamental form* — multiplied by $W(x, y)$ — in the following triple products:

$$
\begin{aligned}
(X_x, X_y, X_{xx})|_{(x,y)} &= \varrho^2 \varrho_{xx}(x, y) - 2\varrho\varrho_x^2(x, y),\\
(X_x, X_y, X_{xy})|_{(x,y)} &= \varrho^2 \varrho_{xy}(x, y) - 2\varrho\varrho_x\varrho_y(x, y),\\
(X_x, X_y, X_{yy})|_{(x,y)} &= \varrho^2 \varrho_{yy}(x, y) - 2\varrho\varrho_y^2(x, y).
\end{aligned}
\tag{2-10}
$$

For an adequate geometric formulation we need some definitions.

Definition 2.1. With each domain $\Omega \subset \mathbb{R}^2$ we associate the *cone*

$$
\mathcal{C}(\Omega) := \{(\xi, \eta, \zeta) \in \mathbb{R}^3 \mid \xi = rx,\ \eta = ry,\ \zeta = r,\ (x, y) \in \Omega,\ 0 < r < +\infty\},
\tag{2-11}
$$

where $\Omega \times \{1\} \subset \mathbb{R}^3$ represents its *base* and $(0, 0, 0)$ its *vertex*. The cone $\mathcal{C}(\Omega)$ consists of the *generating lines*

$$
L_{(x,y)} := \{(rx, ry, r) \in \mathbb{R}^3 \mid 0 < r < +\infty\} \quad \text{for all } (x, y) \in \Omega.
$$

The boundary of our cone $\partial\mathcal{C}(\Omega)$ is composed of the generating lines $L_{(x,y)}$, $(x, y) \in \partial\Omega$.

Definition 2.2. At first, we define the *logarithmic mean curvature* on the cylinder $\Omega \times \mathbb{R}$ by the continuous function

$$
D = D(x, y, z) : \Omega \times \mathbb{R} \to \mathbb{R} \in C^0(\Omega \times \mathbb{R}).
\tag{2-12}
$$

Then we prescribe the *associate mean curvature* on the cone $\mathcal{C}(\Omega)$ by setting

$$
\begin{gathered}
H(\xi, \eta, \zeta) := \frac{D(x, y, \ln r)}{r},\\
(\xi, \eta, \zeta) = (rx, ry, r) \in \mathcal{C}(\Omega), \quad (x, y) \in \Omega, \quad 0 < r < +\infty.
\end{gathered}
\tag{2-13}
$$

At all points with $\varrho(x, y) > 0$, the representation (2-3) yields a *surface of prescribed mean curvature H* from (2-12) and (2-13), or briefly an *H-surface*, if and only if the following partial differential equation (PDE) holds true:

$$
\begin{aligned}
&2D(x, y, \ln\varrho)\varrho^2\\
&\quad\times \bigl(\varrho^2 + (1+x^2)\varrho_x^2 + (1+y^2)\varrho_y^2 + 2xy\varrho_x\varrho_y + 2x\varrho\varrho_x + 2y\varrho\varrho_y\bigr)^{3/2}\\
&\quad = 2H(\varrho x, \varrho y, \varrho)\varrho^3\\
&\qquad\times \bigl(\varrho^2 + (1+x^2)\varrho_x^2 + (1+y^2)\varrho_y^2 + 2xy\varrho_x\varrho_y + 2x\varrho\varrho_x + 2y\varrho\varrho_y\bigr)^{3/2}\\
&\quad = 2H(\varrho x, \varrho y, \varrho)W^3(x, y)\\
&\quad = X_y^2(X_x, X_y, X_{xx}) - 2(X_x \cdot X_y)(X_x, X_y, X_{xy}) + X_x^2(X_x, X_y, X_{yy})\\
&\quad = \varrho^2\bigl(X_y^2\varrho_{xx}(x, y) - 2(X_x \cdot X_y)\varrho_{xy}(x, y) + X_x^2\varrho_{yy}(x, y)\bigr)\\
&\qquad - 2\varrho(x, y)\bigl(X_y^2\varrho_x^2(x, y) - 2(X_x \cdot X_y)\varrho_x\varrho_y(x, y) + X_x^2\varrho_y^2(x, y)\bigr).
\end{aligned}
\tag{2-14}
$$

Besides the prescription (2-13) for the mean curvature, we have used the identity (2-7) for the surface element and the equations (2-10) for the triple products. With the aid of the relations (2-5) we immediately calculate

$$X_y^2\varrho_x^2(x,y) - 2(X_x\cdot X_y)\varrho_x\varrho_y(x,y) + X_x^2\varrho_y^2(x,y) = \varrho^2(x,y)|\nabla\varrho(x,y)|^2. \tag{2-15}$$

When we insert the identity (2-15) into the equation (2-14), we arrive at the PDE

$$\begin{aligned}
&2D(x,y,\ln\varrho)\\
&\quad\times\big(\varrho^2+(1+x^2)\varrho_x^2+(1+y^2)\varrho_y^2+2xy\varrho_x\varrho_y+2x\varrho\varrho_x+2y\varrho\varrho_y\big)^{3/2}\\
&\quad\quad= X_y^2\varrho_{xx}(x,y)-2(X_x\cdot X_y)\varrho_{xy}(x,y)+X_x^2\varrho_{yy}(x,y)\\
&\quad\quad\quad-2\varrho(x,y)|\nabla\varrho(x,y)|^2.
\end{aligned}\tag{2-16}$$

Taking the coefficients of the first fundamental form (2-5) into account, we obtain the PDE

$$\begin{aligned}
&2D(x,y,\ln\varrho)\\
&\quad\times\big(\varrho^2+(1+x^2)\varrho_x^2+(1+y^2)\varrho_y^2+2xy\varrho_x\varrho_y+2x\varrho\varrho_x+2y\varrho\varrho_y\big)^{3/2}\\
&\quad\quad=\big(\varrho_y^2\cdot(x^2+y^2+1)+2\varrho y\varrho_y+\varrho^2\big)\varrho_{xx}\\
&\quad\quad\quad-2\big(\varrho_x\varrho_y\cdot(x^2+y^2+1)+\varrho[y\varrho_x+x\varrho_y]\big)\varrho_{xy}\\
&\quad\quad\quad+\big(\varrho_x^2\cdot(x^2+y^2+1)+2\varrho x\varrho_x+\varrho^2\big)\varrho_{yy}(x,y)-2\varrho|\nabla\varrho(x,y)|^2,
\end{aligned}\tag{2-17}$$

for $(x,y)\in\Omega$.

Since our surface X is regular in Ω, we can assume the property

$$\varrho(x,y)>0,\quad (x,y)\in\Omega,\tag{2-18}$$

after an eventual reflection. Now we use the *logarithmic representation*

$$\sigma(x,y):=\ln\varrho(x,y),\quad (x,y)\in\Omega.\tag{2-19}$$

Then we determine their first derivatives

$$\sigma_x(x,y)=\frac{\varrho_x}{\varrho}(x,y),\quad \sigma_y(x,y)=\frac{\varrho_y}{\varrho}(x,y),\quad (x,y)\in\Omega,\tag{2-20}$$

as well as their second derivatives

$$\begin{aligned}
(\sigma_{xx}+\sigma_x^2)|_{(x,y)}&=\frac{\varrho_{xx}}{\varrho}(x,y),\\
(\sigma_{xy}+\sigma_x\sigma_y)|_{(x,y)}&=\frac{\varrho_{xy}}{\varrho}(x,y),\\
(\sigma_{yy}+\sigma_y^2)|_{(x,y)}&=\frac{\varrho_{yy}}{\varrho}(x,y).
\end{aligned}\tag{2-21}$$

From the identity (2-15) we deduce the relation

$$X_y^2\sigma_x^2(x,y) - 2(X_x\cdot X_y)\sigma_x\sigma_y(x,y) + X_x^2\sigma_y^2(x,y) = |\nabla\varrho(x,y)|^2 = \varrho^2|\nabla\sigma(x,y)|^2. \tag{2-22}$$

Into the equation (2-17) we insert the second derivatives (2-21) and observe (2-22) to obtain

$$\begin{aligned}
&2D(x,y,\ln\varrho)\varrho^2 \\
&\quad\times\left(1+(1+x^2)\left(\frac{\varrho_x}{\varrho}\right)^2+(1+y^2)\left(\frac{\varrho_y}{\varrho}\right)^2+2xy\frac{\varrho_x}{\varrho}\frac{\varrho_y}{\varrho}+2x\frac{\varrho_x}{\varrho}+2y\frac{\varrho_y}{\varrho}\right)^{3/2} \\
&\quad= \big(\varrho_y^2\cdot(x^2+y^2+1)+2\varrho y\varrho_y+\varrho^2\big)\frac{\varrho_{xx}}{\varrho} \\
&\qquad -2\big(\varrho_x\varrho_y\cdot(x^2+y^2+1)+\varrho[y\varrho_x+x\varrho_y]\big)\frac{\varrho_{xy}}{\varrho} \\
&\qquad +\big(\varrho_x^2\cdot(x^2+y^2+1)+2\varrho x\varrho_x+\varrho^2\big)\frac{\varrho_{yy}}{\varrho}(x,y) \\
&\qquad -2|\nabla\varrho(x,y)|^2 \\
&\quad= \big(\varrho_y^2\cdot(x^2+y^2+1)+2\varrho y\varrho_y+\varrho^2\big)\sigma_{xx} \\
&\qquad -2\big(\varrho_x\varrho_y\cdot(x^2+y^2+1)+\varrho[y\varrho_x+x\varrho_y]\big)\sigma_{xy} \\
&\qquad +\big(\varrho_x^2\cdot(x^2+y^2+1)+2\varrho x\varrho_x+\varrho^2\big)\sigma_{yy}(x,y) \\
&\qquad -\varrho^2(x,y)|\nabla\sigma(x,y)|^2,
\end{aligned} \tag{2-23}$$

for $(x,y)\in\Omega$. Now we use formulae (2-20) for the first derivatives and arrive at the PDE

$$\begin{aligned}
&2D(x,y,\sigma)\big(1+(1+x^2)\sigma_x^2+(1+y^2)\sigma_y^2+2xy\sigma_x\sigma_y+2x\sigma_x+2y\sigma_y\big)^{3/2} \\
&\quad= \big(\sigma_y^2\cdot(x^2+y^2+1)+2y\sigma_y+1\big)\sigma_{xx} \\
&\qquad -2\big(\sigma_x\sigma_y\cdot(x^2+y^2+1)+[y\sigma_x+x\sigma_y]\big)\sigma_{xy} \\
&\qquad +\big(\sigma_x^2\cdot(x^2+y^2+1)+2x\sigma_x+1\big)\sigma_{yy}(x,y)-|\nabla\sigma(x,y)|^2,
\end{aligned} \tag{2-24}$$

for $(x,y)\in\Omega$.

Definition 2.3. Let us address the PDE (2-17) as the *H-surface equation in central projection* and the PDE (2-24) as the *logarithmic H-surface equation*. In the special case $D\equiv 0\equiv H$, we speak of the PDE (2-17) as the *minimal surface equation in central projection* and of the PDE (2-24) as the *logarithmic minimal surface equation*.

In the case $D=D(x,y):\Omega\to\mathbb{R}$, where the logarithmic mean curvature does not depend on the z-variable, we prescribe the associate mean curvature H from (2-13) on the base of the cone $\mathcal{C}(\Omega)$. The mean curvature is positive-homogeneously continued on each generating line $L_{(x,y)}$, $(x, y) \in \Omega$, of the degree -1. Then our PDE (2-17) is *positive-homogeneous* in the following sense: for any positive solution $\varrho(x, y)$ of (2-17) and all parameters $a > 0$, the function $a \cdot \varrho(x, y)$ solves this differential equation as well. In the special case that the logarithmic mean curvature

$$\widehat{D}(x, y) := \frac{-2}{1+x^2+y^2}, \quad (x, y) \in \mathbb{R}^2, \tag{2-25}$$

is prescribed on the base of our cone, we can explicitly solve the PDE (2-17) as follows.

Theorem 2.1. *Let the right-hand side* $\widehat{D}$ *from* (2-25) *with its homogeneous continuation* $\widehat{H}$ *of* (2-13) *be given on the cone* $\mathcal{C}(\Omega)$ *for* $\Omega = \mathbb{R}^2$. *Then the functions*

$$\hat{\varrho}(x, y) := \frac{a}{1+x^2+y^2}, \quad (x, y) \in \mathbb{R}^2, \quad \textit{for arbitrary parameters } a > 0, \tag{2-26}$$

solve the $\widehat{H}$*-surface equation* (2-17) *in central projection.*

Proof. Equivalently to the PDE (2-17) for the function $\hat{\varrho}$, we consider the PDE (2-24) for its logarithmic representation $\hat{\sigma}(x, y) := \ln \hat{\varrho}(x, y)$, $(x, y) \in \mathbb{R}^2$, and obtain

$$\begin{aligned} \hat{\sigma}(x, y) &= \ln a - \ln(1+x^2+y^2), \\ \hat{\sigma}_x(x, y) = \frac{-2x}{1+x^2+y^2}, &\quad \hat{\sigma}_y(x, y) = \frac{-2y}{1+x^2+y^2}. \end{aligned} \tag{2-27}$$

We easily determine the expressions

$$\begin{aligned} (1+x^2)\hat{\sigma}_x^2 + (1+y^2)\hat{\sigma}_y^2 + 2xy\hat{\sigma}_x\hat{\sigma}_y + 2x\hat{\sigma}_x + 2y\hat{\sigma}_y &= 0, \\ \hat{\sigma}_y^2 \cdot (x^2+y^2+1) + 2y\hat{\sigma}_y + 1 &= 1, \\ \hat{\sigma}_x\hat{\sigma}_y \cdot (x^2+y^2+1) + [y\hat{\sigma}_x + x\hat{\sigma}_y] &= 0, \\ \hat{\sigma}_x^2 \cdot (x^2+y^2+1) + 2x\hat{\sigma}_x + 1 &= 1, \end{aligned} \tag{2-28}$$

for $(x, y) \in \Omega$. Therefore, the PDE (2-24) is reduced to the equation

$$\begin{aligned} \Delta\hat{\sigma}(x, y) - |\nabla\hat{\sigma}(x, y)|^2 &= \frac{-4}{1+x^2+y^2} + \frac{4x^2+4y^2}{(1+x^2+y^2)^2} - \frac{4x^2+4y^2}{(1+x^2+y^2)^2} \\ &= \frac{-4}{1+x^2+y^2} = 2\widehat{D}(x, y), \end{aligned}$$

for $(x, y) \in \mathbb{R}^2$. Consequently, the PDE (2-17) with the right-hand side from (2-25) is satisfied for the functions (2-26). $\square$

Remark. For each $a > 0$, the surface (2-3) on the domain $\Omega = \mathbb{R}^2$ with the function (2-26) represents a surface of constant mean curvature $-2/a$ with one-to-one central projection onto the plane $\mathcal{E}$. More precisely, we obtain a sphere of radius $a/2$ about the center $(0, 0, a/2)$, where its south pole $(0, 0, 0)$ has been exempted. We shall use these solutions, which represent a *foliation of spheres*, as comparison surfaces in the next section.

Proof. In the case $a = 1$, the equations (2-3) and (2-26) represent the *stereographic projection* of this sphere onto the plane $\mathcal{E}$. Here we employ a theorem of Euclid on right triangles: *the square of a small side equals its projection on the hypotenuse times the hypotenuse*. Based on a simple diagram with a right triangle, we thus obtain $1^2 = \varrho(x, y)\sqrt{1 + x^2 + y^2} \cdot \sqrt{1 + x^2 + y^2}$, which yields

$$\varrho(x, y) = \frac{1}{1 + x^2 + y^2}.$$

By a dilation with the factor a, we can easily inspect the general case $a > 0$. $\square$

3. Uniqueness of Dirichlet's problem and estimates

Definition 3.1. Let the logarithmic mean curvature $D(x, y, z) : \Omega \times \mathbb{R} \to \mathbb{R}$ be given on the cylinder adjoint to the bounded Jordan domain $\Omega \subset \mathbb{R}^2$, with its associate mean curvature $H(\xi, \eta, \zeta) : \mathcal{C}(\Omega) \to \mathbb{R}$ from (2-13) on the cone $\mathcal{C}(\Omega)$. On the Jordan contour $\partial\Omega$ let the positive continuous boundary distribution $\phi : \partial\Omega \to (0, +\infty)$ be prescribed. Then the positive solution

$$\varrho = \varrho(x, y) : \bar{\Omega} \to (0, +\infty) \in C^2(\Omega) \cap C^0(\bar{\Omega})$$

of the PDE (2-17) under the *Dirichlet boundary condition*

$$\varrho(x, y) = \phi(x, y) \quad \text{for all } (x, y) \in \partial\Omega \tag{3-1}$$

is named the *solution of the Dirichlet problem* $\boldsymbol{P}(\Omega, \phi, H)$ *for the H-surface equation in central projection.*

Definition 3.2. The logarithmic mean curvature $D(x, y, z) : \Omega \times \mathbb{R} \to \mathbb{R}$ satisfies the *monotonicity condition* if, for each point (x, y) in the domain Ω, the function d defined by

$$d(z) := D(x, y, z) \quad \text{for } z \in \mathbb{R}$$

is of class C^1 and satisfies

$$d'(z) = \frac{\partial}{\partial z} D(x, y, z) \geq 0 \quad \text{for } z \in \mathbb{R}. \tag{3-2}$$

The maximum principle for elliptic equations implies the following.

Theorem 3.1 (uniqueness of $\boldsymbol{P}(\Omega, \phi, H)$). *Let* $\varrho^{(j)} = \varrho^{(j)}(x, y)$, $j = 1, 2$, *denote two solutions of the Dirichlet problem* $\boldsymbol{P}(\Omega, \phi, H)$ *in the Jordan domain* $\Omega \subset \mathbb{R}^2$, *where the logarithmic mean curvature satisfies the monotonicity condition. Then*

$$\varrho^{(1)}(x, y) = \varrho^{(2)}(x, y) \quad \textit{for } (x, y) \in \bar{\Omega}.$$

Proof. Let us consider two solutions

$$\varrho^{(j)} = \varrho^{(j)}(x, y) : \mathbb{R}^2 \to (0, +\infty) \in C^2(\Omega) \cap C^0(\bar{\Omega}), \quad \text{with } j = 1, 2, \tag{3-3}$$

of the Dirichlet problem $\boldsymbol{P}(\Omega, \phi, H)$. We can apply the maximum principle to the difference of their logarithmic representations,

$$\sigma^{(j)}(x, y) := \ln \varrho^{(j)}(x, y), \quad (x, y) \in \bar{\Omega}, \quad \text{with } j = 1, 2,$$

since the associate PDE (2-24) is quasilinear. The elliptic differential operator for the difference function possesses a nonpositive coefficient of 0 order, due to the monotonicity condition (compare [Sauvigny 2012a, Chapter 6, §2]). Therefore, we obtain

$$\varrho^{(1)}(x, y) = \varrho^{(2)}(x, y), \quad (x, y) \in \bar{\Omega},$$

and the Dirichlet problem $\boldsymbol{P}(\Omega, \phi, H)$ is uniquely determined. □

Furthermore, we prove the following interesting theorem.

Theorem 3.2 (geometric maximum principle). *Let* $\varrho = \varrho(x, y)$ *denote a solution of the Dirichlet problem* $\boldsymbol{P}(\Omega, \phi, H)$ *in the Jordan domain* $\Omega \subset \mathbb{R}^2$, *which is contained in the* disc $\Omega_b := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < b^2\}$ of radius $0 < b < +\infty$. *Furthermore*, *let the logarithmic mean curvature satisfy the monotonicity condition and the inequalities*

$$\widehat{D}(x, y) \le D(x, y, z) \le 0 \quad \textit{for all } (x, y, z) \in \Omega \times \mathbb{R}. \tag{3-4}$$

Then we have the estimate

$$0 < \min_{(\xi,\eta)\in\partial\Omega} \phi(\xi, \eta) \le \varrho(x, y) \le (1 + b^2) \cdot \max_{(\xi,\eta)\in\partial\Omega} \phi(\xi, \eta), \tag{3-5}$$

for all points $(x, y) \in \bar{\Omega}$.

Proof. (1) From (2-17) and (3-4) we infer the elliptic differential inequality

$$\begin{aligned} &\big(\varrho_y^2 \cdot (x^2 + y^2 + 1) + 2\varrho y \varrho_y + \varrho^2\big)\varrho_{xx} \\ &\quad - 2\big(\varrho_x \varrho_y \cdot (x^2 + y^2 + 1) + \varrho[y\varrho_x + x\varrho_y]\big)\varrho_{xy} \\ &\qquad + \big(\varrho_x^2 \cdot (x^2 + y^2 + 1) + 2\varrho x \varrho_x + \varrho^2\big)\varrho_{yy}(x, y) - 2\varrho|\nabla\varrho(x, y)|^2 \le 0, \end{aligned} \tag{3-6}$$

for $(x, y) \in \Omega$. Within the domain Ω our function ϱ cannot attain a strict minimum, and the estimate on the left-hand side of (3-5) is established.

(2) We compare the solution ϱ with the spherical solution of Theorem 2.1,

$$\hat{\varrho}(x, y) := \frac{a}{1 + x^2 + y^2}, \quad (x, y) \in \mathbb{R}^2, \quad \text{where } a := (1+b^2) \max_{(\xi,\eta)\in\partial\Omega} \phi(\xi, \eta). \tag{3-7}$$

By construction we have the inequality

$$\hat{\varrho}(x, y) \ge \varrho(x, y) \quad \text{for all } (x, y) \in \partial\Omega \tag{3-8}$$

on the boundary. From the condition (3-4) and the PDE (2-24) for the logarithmic representation

$$\hat{\sigma}(x, y) := \ln \hat{\varrho}(x, y), \quad (x, y) \in \bar{\Omega},$$

we deduce the differential inequality

$$\begin{aligned}(3\text{-}9)\quad & \big(\hat{\sigma}_y^2 \cdot (x^2+y^2+1)+2y\hat{\sigma}_y+1\big)\hat{\sigma}_{xx} - 2\big(\hat{\sigma}_x\hat{\sigma}_y \cdot (x^2+y^2+1)+[y\hat{\sigma}_x+x\hat{\sigma}_y]\big)\hat{\sigma}_{xy} \\ & + \big(\hat{\sigma}_x^2 \cdot (x^2+y^2+1)+2x\hat{\sigma}_x+1\big)\hat{\sigma}_{yy}(x, y) - |\nabla\hat{\sigma}(x, y)|^2 \\ & = 2\widehat{D}(x, y)\big(1+(1+x^2)\hat{\sigma}_x^2+(1+y^2)\hat{\sigma}_y^2+2xy\hat{\sigma}_x\hat{\sigma}_y+2x\hat{\sigma}_x+2y\hat{\sigma}_y\big)^{3/2} \\ & \le 2D(x, y, \hat{\sigma}) \\ & \qquad \times\big(1+(1+x^2)\hat{\sigma}_x^2+(1+y^2)\hat{\sigma}_y^2+2xy\hat{\sigma}_x\hat{\sigma}_y+2x\hat{\sigma}_x+2y\hat{\sigma}_y\big)^{3/2},\end{aligned}$$

at all points $(x, y) \in \Omega$. The logarithmic representation

$$\sigma(x, y) := \ln \varrho(x, y), \quad (x, y) \in \bar{\Omega},$$

of the function ϱ satisfies the quasilinear PDE (2-24). Together with (3-9) the difference function

$$\tau(x, y) := \sigma(x, y) - \hat{\sigma}(x, y), \quad (x, y) \in \bar{\Omega},$$

is subject to the differential inequality

$$\mathcal{L}\tau(x, y) \ge 0, \quad (x, y) \in \Omega,$$

for an elliptic differential operator $\mathcal{L}$ (see [Sauvigny 2012a, Chapter 6, §2]). Due to the monotonicity condition, the coefficient for the 0-order term of $\mathcal{L}$ is nonpositive. Because of (3-8) we have $\tau(x, y) \le 0$ for $(x, y) \in \partial\Omega$ and hence, by the maximum principle for elliptic operators, $\tau(x, y) \le 0$ for $(x, y) \in \bar{\Omega}$. With the aid of (3-7) we arrive at the estimate

$$(3\text{-}10)\qquad \varrho(x, y) \le \hat{\varrho}(x, y) \le a = (1+b^2) \cdot \max_{(\xi,\eta)\in\partial\Omega} \phi(\xi, \eta), \quad (x, y) \in \bar{\Omega}.$$

This shows the right-hand side of the statement (3-5) above. □

Remark. For embedded surfaces of constant mean curvature, J. Serrin [1969] established a maximum-estimate by the bounding contour; there the reflection method of A. D. Alexandroff was employed. Our method above is based on a foliation of H-surfaces with variable mean curvature.

Already at the beginning of the last century, A. Korn and C. H. Müntz solved the boundary value problem of the minimal surface equation for contours deviating only a little from planar curves. From §§ 413–415 of J. C. C. Nitsche's treatise [1975] we learn that Plateau's problem for parametric minimal surfaces with positive second variation is stable with respect to small perturbations of the bounding contour.

The approximate solutions are obtained by solving a peculiar nonlinear elliptic PDE which is accessible to Banach's fixed point theorem within the Hölder space $C^{2+\alpha}(\bar{B})$ (compare [Dierkes et al. 2010a, §5.6, Proposition 1]). We have used this method in Proposition 1.8 of [Sauvigny 2012b, Chapter 13], in order to establish the stability of the nonparametric H-surface equation under small perturbations with respect to the $C^{2+\alpha}$-norm of the boundary data. Since the logarithmic H-surface equation (2-24) has a similar structure, we can prove the stability for our Dirichlet problem $\boldsymbol{P}(\Omega, \phi, H)$ under homothetic transformations — near the identity — of the boundary values.

Lemma 3.1 (perturbation result). *Let $\Omega \subset \mathbb{R}^2$ denote a convex $C^{2+\alpha}$-Jordan domain with $0 < \alpha < 1$, such that the logarithmic mean curvature*

$$D = D(x, y, z) \in C^{1+\alpha}(\bar{\Omega} \times \mathbb{R}, \mathbb{R})$$

satisfies the monotonicity condition. For the associate mean curvature H due to (2-13) on the cone $\mathcal{C}(\Omega)$ and the positive boundary distribution $\phi : \partial\Omega \to (0, +\infty)$ of class $C^{2+\alpha}$, the problem $\boldsymbol{P}(\Omega, \phi, H)$ possesses a solution. Then the Dirichlet problem $\boldsymbol{P}(\Omega, \lambda\phi, H)$ is solvable for all parameters $1 - \varepsilon \le \lambda \le 1 + \varepsilon$, where the quantity $\varepsilon > 0$ is sufficiently small.

Proof. (1) Instead of the PDE (2-17) we use the equivalent equation (2-24) and start with a solution $\sigma = \sigma(x, y) \in C^{2+\alpha}(\bar{\Omega})$ of this logarithmic H-surface equation. Then we consider the *perturbation*

$$\sigma(x, y) + t + \tau(x, y), \quad (x, y) \in \bar{\Omega}, \tag{3-11}$$

with a parameter $-\epsilon \le t \le \epsilon$ and a function τ in the Banach space

$$\mathcal{B} := \{\tau = \tau(x, y) \in C^{2+\alpha}(\bar{\Omega}) \mid \tau(x, y) = 0 \text{ for all } (x, y) \in \partial\Omega\}.$$

We insert (3-11) into (2-24) and observe that our perturbed function (3-11) satisfies the logarithmic H-surface equation (2-24) if and only if the function $\tau \in \mathcal{B}$ fulfills the PDE

$$\begin{aligned}
&\big((\sigma_y + \tau_y)^2 \cdot (x^2 + y^2 + 1) + 2y(\sigma_y + \tau_y) + 1\big)(\sigma_{xx} + \tau_{xx}) \\
&\quad -2\big((\sigma_x+\tau_x)(\sigma_y+\tau_y)\cdot(x^2+y^2+1)+[y(\sigma_x+\tau_x)+x(\sigma_y+\tau_y)]\big)(\sigma_{xy}+\tau_{xy}) \\
&\quad + \big((\sigma_x + \tau_x)^2 \cdot (x^2 + y^2 + 1) + 2x(\sigma_x + \tau_x) + 1\big)(\sigma_{yy} + \tau_{yy}) \\
&\quad - |(\nabla\sigma + \nabla\tau)|_{(x,y)}|^2 \\
&\qquad = 2D(x, y, \sigma + t + \tau)\big(1 + (1 + x^2)(\sigma_x + \tau_x)^2 + (1 + y^2)(\sigma_y + \tau_y)^2 \\
&\qquad\qquad + 2xy(\sigma_x + \tau_x)(\sigma_y + \tau_y) + 2x(\sigma_x + \tau_x) + 2y(\sigma_y + \tau_y)\big)^{3/2},
\end{aligned} \tag{3-12}$$

for $(x, y) \in \Omega$.

(2) On the left-hand side of (3-12) we collect all those terms, where the factors of the set

$$\mathcal{F} := \{\tau_x, \tau_y, \tau_{xx}, \tau_{xy}, \tau_{yy}\}$$

appear in the same order, to the following differential operators: the terms of order 0 result in the expression

$$2D(x, y, \sigma)\big(1 + (1 + x^2)\sigma_x^2 + (1 + y^2)\sigma_y^2 + 2xy\sigma_x\sigma_y + 2x\sigma_x + 2y\sigma_y\big)^{3/2}, \tag{3-13}$$

since σ solves the PDE (2-24). Then we collect all terms of order 1 to the linear elliptic differential operator

$$\begin{aligned} \mathcal{L}_0\tau := &\big(\sigma_y^2 \cdot (x^2 + y^2 + 1) + 2y\sigma_y + 1\big)\tau_{xx} \\ &- 2\big(\sigma_x\sigma_y \cdot (x^2 + y^2 + 1) + [y\sigma_x + x\sigma_y]\big)\tau_{xy} \\ &+ \big(\sigma_x^2 \cdot (x^2 + y^2 + 1) + 2x\sigma_x + 1\big)\tau_{yy}(x, y) \\ &+ a(x, y)\tau_x + b(x, y)\tau_y \end{aligned} \tag{3-14}$$

with coefficients $a(x, y)$, $b(x, y)$ of the class $C^\alpha(\bar{\Omega})$ depending on the solution σ. The remaining terms of order 2 and 3 are assembled to the nonlinear operator $\mathcal{Q} = \mathcal{Q}(\tau) : \mathcal{B} \to C^\alpha(\bar{\Omega})$. On the ball $\mathcal{B}_\delta := \{\tau = \tau(x, y) \in \mathcal{B} \mid \|\tau\|_{C^{2+\alpha}(\bar{\Omega})} \le \delta\}$ of radius $\delta > 0$ the estimates

$$\begin{aligned} \|\mathcal{Q}(\tau)\|_{C^\alpha(\bar{\Omega})} &\le L_1(\delta)\|\tau\|_{C^{2+\alpha}(\bar{\Omega})} \quad \text{for all } \tau \in \mathcal{B}_\delta, \\ \|\mathcal{Q}(\tilde{\tau}) - \mathcal{Q}(\hat{\tau})\|_{C^\alpha(\bar{\Omega})} &\le L_2(\delta)\|\tilde{\tau} - \hat{\tau}\|_{C^{2+\alpha}(\bar{\Omega})} \quad \text{for all } \tilde{\tau}, \hat{\tau} \in \mathcal{B}_\delta \end{aligned} \tag{3-15}$$

hold true. Here as well as in (3-18), (3-19), (3-26) below, the constants $L_j(\delta) > 0$ satisfy $\lim_{\delta\to 0+} L_j(\delta) = 0$ for $j = 1, 2, 3, 4$. When we respect that all terms in $\mathcal{Q}$ are either quadratic or cubic in $\mathcal{F}$ and control their Hölder-norms, we immediately see the assertions (3-15) above, where the upper inequality implies that the *operator* $\mathcal{Q}$ is *superlinear*.

Now the equation (3-12) appears in the equivalent form

$$\begin{aligned} &\mathcal{L}_0\tau + \mathcal{Q}(\tau) \\ &\quad = 2D(x, y, \sigma + t + \tau)\big(1 + (1 + x^2)(\sigma_x + \tau_x)^2 + (1 + y^2)(\sigma_y + \tau_y)^2 \\ &\qquad\quad + 2xy(\sigma_x + \tau_x)(\sigma_y + \tau_y) + 2x(\sigma_x + \tau_x) + 2y(\sigma_y + \tau_y)\big)^{3/2} \\ &\qquad - 2D(x, y, \sigma)\big(1 + (1 + x^2)\sigma_x^2 + (1 + y^2)\sigma_y^2 \\ &\qquad\quad + 2xy\sigma_x\sigma_y + 2x\sigma_x + 2y\sigma_y\big)^{3/2}. \end{aligned} \tag{3-16}$$

(3) We introduce the nonlinear operator

$$\begin{aligned} \mathcal{N}(\tau) := \big(&1 + (1 + x^2)(\sigma_x + \tau_x)^2 + (1 + y^2)(\sigma_y + \tau_y)^2 \\ &+ 2xy(\sigma_x + \tau_x)(\sigma_y + \tau_y) + 2x(\sigma_x + \tau_x) + 2y(\sigma_y + \tau_y)\big)^{3/2}, \end{aligned} \tag{3-17}$$

for $\tau \in \mathcal{B}$, where the power $3/2$ is larger than 1. Therefore, we obtain the estimate

$$\|\mathcal{N}(\tilde{\tau}) - \mathcal{N}(\hat{\tau})\|_{C^\alpha(\bar{\Omega})} \le L_3(\delta)\|\tilde{\tau} - \hat{\tau}\|_{C^{2+\alpha}(\bar{\Omega})} \quad \text{for all } \tilde{\tau}, \hat{\tau} \in \mathcal{B}_\delta. \tag{3-18}$$

Thus we receive the superlinear operator

$$\begin{gathered} \mathcal{R}(\tau) := \mathcal{N}(\tau) - \mathcal{N}(0), \quad \tau \in \mathcal{B}, \\ \text{satisfying } \|\mathcal{R}(\tau)\|_{C^\alpha(\bar{\Omega})} \le L_3(\delta)\|\tau\|_{C^{2+\alpha}(\bar{\Omega})}, \quad \tau \in \mathcal{B}_\delta. \end{gathered} \tag{3-19}$$

Now we rewrite (3-16) into the equivalent form

$$\begin{aligned} \mathcal{L}_0\tau + \mathcal{Q}(\tau) &= 2D(x, y, \sigma + t + \tau)\mathcal{N}(\tau) - 2D(x, y, \sigma)\mathcal{N}(0) \\ &= 2D(x, y, \sigma + t + \tau)(\mathcal{N}(0) + \mathcal{R}(\tau)) - 2D(x, y, \sigma)\mathcal{N}(0) \\ &= 2(D(x, y, \sigma + t + \tau) - D(x, y, \sigma))\mathcal{N}(0) \\ &\qquad + 2D(x, y, \sigma + t + \tau)\mathcal{R}(\tau). \end{aligned} \tag{3-20}$$

(4) Let us determine

$$\begin{aligned} &D(x, y, \sigma(x, y) + t + \tau(x, y)) - D(x, y, \sigma(x, y)) \\ &\quad = \int_0^1 \tfrac{d}{ds} D(x, y, s[\tau(x, y) + t] + \sigma(x, y))\, ds = c_0(x, y)[\tau(x, y) + t], \end{aligned} \tag{3-21}$$

for $(x, y) \in \Omega$, with the nonnegative function

$$c_0(x, y) := \int_0^1 D_z(x, y, s[\tau(x, y) + t] + \sigma(x, y))\, ds,$$

due to the monotonicity condition. We insert (3-21) into the PDE (3-20) and arrive at

$$\mathcal{L}_0\tau + \mathcal{Q}(\tau) = 2c_0(x, y)[\tau(x, y) + t]\mathcal{N}(0) + 2D(x, y, \sigma + t + \tau)\mathcal{R}(\tau). \tag{3-22}$$

Introducing the coefficient function $c(x, y) := -2c_0(x, y)\mathcal{N}(0) \le 0$, $(x, y) \in \Omega$, and the linear elliptic operator $\mathcal{L}\tau := \mathcal{L}_0\tau + c(x, y)\tau$, $\tau \in \mathcal{B}$, we obtain the PDE

$$\mathcal{L}\tau = 2tc_0(x, y)\mathcal{N}(0) - \mathcal{Q}(\tau) + 2D(x, y, \sigma + t + \tau)\mathcal{R}(\tau) =: \mathcal{M}_t(\tau) \tag{3-23}$$

with the nonlinear operator $\mathcal{M}_t : \mathcal{B} \to C^\alpha(\bar{\Omega})$ on the right-hand side.

For each $\theta > 0$, we can find quantities $\delta = \delta(\theta) > 0$ and $\varepsilon = \varepsilon(\theta) > 0$ such that

$$\|\mathcal{M}_t(\tau)\|_{C^\alpha(\bar{\Omega})} \le \theta^{-1}\delta \quad \text{for all } \tau \in \mathcal{B}_\delta \text{ and all } -\varepsilon \le t \le +\varepsilon. \tag{3-24}$$

This follows from the structure of the operator $\mathcal{M}_t$ in (3-23), since the operators $\mathcal{Q}$ and $\mathcal{R}$ are superlinear as in (3-15) and (3-19) and the expression

$$\sup_{\tau \in \mathcal{B}_\delta, t \in [-1, +1]} \|D(x, y, \sigma + t + \tau)\|_{C^\alpha(\bar{\Omega})}$$

is finite, due to the regularity of D on the convex domain Ω.

Furthermore, with $\delta = \delta(\theta) > 0$ *and* $\varepsilon = \varepsilon(\theta) > 0$ *we realize the estimate*

$$\|\mathcal{M}_t(\tilde{\tau}) - \mathcal{M}_t(\hat{\tau})\|_{C^\alpha(\bar{\Omega})} \le \frac{1}{2\theta}\|\tilde{\tau} - \hat{\tau}\|_{C^{2+\alpha}(\bar{\Omega})} \quad \text{for all } \tilde{\tau}, \hat{\tau} \in \mathcal{B}_\delta \text{ and } -\varepsilon \le t \le +\varepsilon. \tag{3-25}$$

Here the structure of the operator $\mathcal{M}_t$ in (3-23) is combined with the inequalities (3-15), (3-18) and the following estimate, which is based on the regularity of D in the convex domain Ω:

$$\|D(\cdot,\cdot,\sigma + t + \tilde{\tau}) - D(\cdot,\cdot,\sigma + t + \hat{\tau})\|_{C^\alpha(\bar{\Omega})} \le L_4(\delta)\|\tilde{\tau} - \hat{\tau}\|_{C^{2+\alpha}(\bar{\Omega})} \quad \text{for all } \tilde{\tau}, \hat{\tau} \in \mathcal{B}_\delta \text{ and } t \in [-1, +1]. \tag{3-26}$$

(5) Due to Theorem 5.2 in [Sauvigny 2012b, Chapter 9], the linear elliptic operator $\mathcal{L} : \mathcal{B} \to C^\alpha(\bar{\Omega})$ satisfies the *Schauder estimate*

$$\|\tau\|_{C^{2+\alpha}(\bar{\Omega})} \le \theta\|\mathcal{L}\tau\|_{C^\alpha(\bar{\Omega})} \quad \text{for all } \tau \in \mathcal{B}, \tag{3-27}$$

where $\theta > 0$ represents the *Schauder constant*. Consequently, $\mathcal{L}$ possesses an inverse $\mathcal{L}^{-1}$ bounded with respect to the respective Hölder norms. With $\delta = \delta(\theta)$ the set $\mathcal{L}(\mathcal{B}_\delta)$ contains a ball of radius $\theta^{-1}\delta$ within the Banach space $C^\alpha(\bar{\Omega})$. When we remember (3-24) with $\varepsilon = \varepsilon(\theta)$, we can transform (3-23) into the fixed point equation

$$\tau = \mathcal{L}^{-1} \circ \mathcal{M}_t(\tau), \quad \text{with } \tau \in \mathcal{B}_\delta \text{ for all } -\varepsilon \le t \le +\varepsilon. \tag{3-28}$$

The nonlinear operator $\mathcal{L}^{-1} \circ \mathcal{M}_t : \mathcal{B}_\delta \to \mathcal{B}_\delta$ yields a contraction

$$\|\mathcal{L}^{-1} \circ \mathcal{M}_t(\tilde{\tau}) - \mathcal{L}^{-1} \circ \mathcal{M}_t(\hat{\tau})\|_{C^{2+\alpha}(\bar{\Omega})} \le \tfrac{1}{2}\|\tilde{\tau} - \hat{\tau}\|_{C^{2+\alpha}(\bar{\Omega})} \quad \text{for all } \tilde{\tau}, \hat{\tau} \in \mathcal{B}_\delta \text{ and } -\varepsilon \le t \le +\varepsilon \tag{3-29}$$

due to (3-25) and (3-27). Banach's fixed point theorem furnishes a unique solution $\tau \in \mathcal{B}_\delta$ of the equation (3-28) for all $-\varepsilon \le t \le +\varepsilon$. □

With the aid of the uniformization method, we shall estimate the area of the solutions for our Dirichlet problem. Let $\Omega \subset \mathbb{R}^2$ denote a $C^{2+\alpha}$-Jordan domain with the positive boundary distribution $\phi : \partial\Omega \to (0, +\infty)$ of the class $C^{2+\alpha}$. We define the Jordan contour

$$\Gamma(\phi) := \{(x\phi(x,y), y\phi(x,y), \phi(x,y)) \in \mathbb{R}^3 \mid (x,y) \in \partial\Omega\}$$

and the *area* $M(\phi) > 0$ *of the conical surface*

$$S(\phi) := \{(\lambda x\phi(x,y), \lambda y\phi(x,y), \lambda\phi(x,y)) \in \mathbb{R}^3 \mid (x,y) \in \partial\Omega,\ \lambda \in (0,1)\}.$$

With the logarithmic mean curvature (2-12) let us define the mean curvature H due to (2-13) on the cone $\mathcal{C}(\Omega)$. We consider a solution $\varrho = \varrho(x,y) \in \boldsymbol{P}(\Omega, \phi, H)$.

The associate surface (2-3) possesses the *area* $A(\varrho)$ and the *volume* $V(\varrho)$ *of the conical domain*

$$G(\varrho) := \{(\lambda x\varrho(x, y), \lambda y\varrho(x, y), \lambda\varrho(x, y)) \in \mathbb{R}^3 \mid (x, y) \in \Omega,\ \lambda \in (0, 1)\}.$$

This H-surface has the *minimal mean curvature*

$$m(\varrho) := \inf_{(x,y)\in\Omega} H(x\varrho(x, y), y\varrho(x, y), \varrho(x, y)) \in \mathbb{R}. \tag{3-30}$$

Lemma 3.2 (area estimate). *We can estimate the area $A(\varrho)$ of a solution ϱ for the Dirichlet problem $\boldsymbol{P}(\Omega, \phi, H)$ in*

$$A(\varrho) \le -3m(\varrho)V(\varrho) + M(\phi)$$

by the minimal mean curvature $m(\varrho)$ as well as the volume $V(\varrho)$ of the conical domain $G(\varrho)$ for the solution, and by the area $M(\phi)$ of the given conical surface $S(\phi)$.

Proof. Let us introduce conformal parameters into the surface (2-3). Then we obtain a *parametric H-surface* $X(u, v) = (x(u, v), y(u, v), z(u, v)) : \bar{B} \to \mathbb{R}^3 \in C^{2+\alpha}(\bar{B})$ which is regular on the closure of the *unit disc* $B := \{(u, v) \in \mathbb{R}^2 \mid u^2 + v^2 < 1\}$ in the differential-geometric sense. More precisely, we have the following conditions:

$$\begin{gathered} \Delta X(u, v) = 2H(X(u, v))X_u \wedge X_v, \quad |X_u|^2 = |X_v|^2 > 0, \\ X_u \cdot X_v = 0, \quad X \cdot X_u \wedge X_v > 0 \quad \text{on } \bar{B}. \end{gathered} \tag{3-31}$$

With the aid of triple products $(\cdot\,,\cdot\,,\cdot)$ we calculate

$$2H(X)(X, X_u, X_v) = X \cdot \Delta X = (X \cdot X_u)_u + (X \cdot X_v)_v - |\nabla X|^2 \quad \text{on } \bar{B} \tag{3-32}$$

and obtain

$$\tfrac{1}{2}|\nabla X|^2 = -H(X)(X, X_u, X_v) + \tfrac{1}{2}\{(X \cdot X_u)_u + (X \cdot X_v)_v\} \quad \text{on } \bar{B}. \tag{3-33}$$

Let us denote the exterior normal to the unit disc by $\nu : \partial B \to S^1$. Furthermore, we use the arc length σ and the line element $d\sigma$ on ∂B. Then we integrate (3-33) as follows:

$$\begin{aligned} A(\varrho) &= \frac{1}{2}\iint_B |\nabla X|^2 \, du\, dv \\ &= \iint_B -H(X)(X, X_u, X_v)\, du\, dv + \frac{1}{2}\int_{\partial B} (X \cdot X_\nu)\, d\sigma. \end{aligned} \tag{3-34}$$

We use the positive orientation of the conformal parameters as well as the minimal mean curvature (3-30) in order to estimate the two-dimensional integral on the right-hand side. Denoting by $N(u, v)$ the unit normal, the conformal parametrization yields the identity

$$X_\sigma \wedge N = X_\nu \quad \text{on } \partial B,$$

which we use in (3-34) for the one-dimensional integral. Thus we obtain

$$\begin{aligned} A(\varrho) &\le -m(\varrho)\iint_B (X, X_u, X_v)\,du\,dv + \frac{1}{2}\int_{\partial B}(X, X_\sigma, N)\,d\sigma \\ &\le -m(\varrho)\iint_B (X, X_u, X_v)\,du\,dv + \frac{1}{2}\int_{\partial B}|X \wedge X_\sigma|\,d\sigma \\ &= -m(\varrho)\iint_B (X, X_u, X_v)\,du\,dv + M(\phi). \end{aligned} \tag{3-35}$$

Then we apply the Gaussian integral theorem to the conical domain $G(\varrho)$ and the vector field

$$W(x, y, z) := (x, y, z), \quad (x, y, z) \in G(\varrho),$$

which is tangential on the conical boundary $\overline{S(\phi)} = \partial G(\varrho) \cap \partial \mathcal{C}(\Omega)$. Therefore, we receive the expression

$$\iint_B (X, X_u, X_v)\,du\,dv = \iiint_{G(\varrho)} \operatorname{div} W(x, y, z)\,dx\,dy\,dz = 3V(\varrho). \tag{3-36}$$

We insert (3-36) into (3-35) and obtain with

$$A(\varrho) \le -3m(\varrho)V(\varrho) + M(\phi) \tag{3-37}$$

the final estimate. □

Remark. Originally, R. Finn [1954] established a priori estimates of the area for graphs of minimal surface type. E. Heinz [1971] proved such an estimate for graphs of prescribed mean curvature. Here we refer our readers to Proposition 1.2 in [Sauvigny 2012b, Chapter 13].

4. Some results on Radó's minimal surface equation

In this section, we consider the special case $H \equiv 0 \equiv D$. With the aid of Plateau's problem, we can solve Dirichlet's problem for the PDE (2-17) with vanishing right-hand side. This has already been proposed by Radó [1932] (compare [Nitsche 1975, §402]). However, we shall apply alternative methods from my dissertation [Sauvigny 1982] and book with Dierkes and Hildebrandt [Dierkes et al. 2010a, §§5.1–5.3], in order to realize that the central projection is one-to-one. The n-dimensional situation has been studied by E. Tausch [1981] using nonparametric methods.

By variational methods we establish the following theorem.

Theorem 4.1 (solution of $\boldsymbol{P}(\Omega, \phi, 0)$). *Let $\Omega \subset \mathbb{R}^2$ denote a convex $C^{2+\alpha}$-Jordan domain and let $\phi : \partial\Omega \to (0, +\infty)$ denote a positive $C^{2+\alpha}$-boundary distribution with $0 < \alpha < 1$. Then the Dirichlet problem $\boldsymbol{P}(\Omega, \phi, 0)$ possesses exactly one solution $\varrho(x, y)$, $(x, y) \in \overline{\Omega}$.*

Proof. (1) We solve Plateau's problem for the regular $C^{2+\alpha}$-Jordan contour

$$\Gamma := \{(x\phi(x, y), y\phi(x, y), \phi(x, y)) \in \mathbb{R}^3 \mid (x, y) \in \partial\Omega\}$$

with a parametric minimal surface $X(u, v) = (x(u, v), y(u, v), z(u, v)) : \bar{B} \to \mathbb{R}^3$. This surface is defined on the closure of the unit disc $B := \{(u, v) \in \mathbb{R}^2 \mid u^2 + v^2 < 1\}$ and satisfies the Laplace equation

$$\Delta X(u, v) = 0, \quad (u, v) \in B.$$

Its isothermal first fundamental form

$$ds^2 = X_u^2\, du^2 + 2X_u \cdot X_v\, du\, dv + X_v^2\, dv^2 = E(u, v)(du^2 + dv^2), \quad (u, v) \in \bar{B},$$

might only degenerate at isolated *branch points* of X, and its unit normal $N(u, v)$, $(u, v) \in \bar{B}$, exists within the class $C^2(B) \cap C^1(\bar{B})$ subject to the *Schwarzian differential equation*

$$\Delta N(u, v) - 2E(u, v)K(u, v)N(u, v) = 0, \quad (u, v) \in B. \tag{4-1}$$

Here $K(u, v) \leq 0$ denotes the Gaussian curvature of the metric ds^2 at regular points (u, v). The differential equation (4-1) can be found in Hilfssatz 1 and Satz 1 of [Sauvigny 1982] (see [Dierkes et al. 2010a, §5.1, Theorem 1] as well). The necessary investigations on the regularity and branch points of H-surfaces are contained in Chapter 2 of the treatise [Dierkes et al. 2010b] by Dierkes, Hildebrandt, and A. Tromba.

(2) The minimal surface $X(\bar{B})$ lies in the convex hull of its bounding contour Γ, where the latter is situated on the boundary of the convex cone $\mathcal{C}(\Omega)$, outside its vertex. This implies the inclusions

$$X(B) \subset \mathcal{C}(\Omega) \quad \text{and} \quad X(\partial B) \subset \partial\mathcal{C}(\Omega) \setminus \{(0, 0, 0)\}. \tag{4-2}$$

The arguments from [Sauvigny 1982, §2] show that the minimal surface approaches the bounding cone $\partial\mathcal{C}(\Omega)$ transversally and does not possess boundary branch points. When we consider the auxiliary function

$$\theta(u, v) := N(u, v) \cdot X(u, v), \quad (u, v) \in \bar{B},$$

we obtain the boundary condition

$$\theta(u, v) > 0, \quad (u, v) \in \partial B. \tag{4-3}$$

With the aid of (4-1), we derive the PDE for our auxiliary function

$$\begin{aligned} \Delta\theta(u, v) &\\ &= (\Delta N(u, v)) \cdot X(u, v) + 2\nabla N(u, v) \cdot \nabla X(u, v) + N(u, v) \cdot (\Delta X(u, v)) \\ &= X(u, v) \cdot \Delta N(u, v) - N(u, v) \cdot \Delta X(u, v) = 2E(u, v)K(u, v)\theta(u, v), \end{aligned} \tag{4-4}$$

for $(u, v) \in B$. Before we have fixed three points on the contour Γ, such that the boundary representation is positive-oriented with respect to the projection plane $\mathcal{E}$.

(3) Now the metric ds^2 is stable in the following sense:

$$\text{(4-5)} \quad \iint_B |\nabla \psi(u, v)|^2 \, du \, dv \\ \geq -2 \iint_B E(u, v) K(u, v) \psi(u, v)^2 \, du \, dv \quad \text{for all } \psi \in C_0^1(B).$$

This stability condition has been established in [Sauvigny 1982, §3] by the area-minimizing property for the solutions of Plateau's problem. Due to Hilfssatz 6 of [Sauvigny 1982] (compare [Dierkes et al. 2010a, §5.3, Proposition 1]), we obtain

$$\text{(4-6)} \qquad \theta(u, v) > 0, \quad (u, v) \in \bar{B}.$$

On the basis of the property (4-6), we can exclude interior branch points for our minimal surface by a winding number argument. Therefore, the surface

$$X : \bar{B} \to \mathcal{C}(\bar{\Omega})$$

represents a minimal embedding, with one-to-one central projection onto the plane $\mathcal{E}$, which bounds the contour Γ. Thus we have solved the Dirichlet problem $\boldsymbol{P}(\Omega, \phi, 0)$. $\square$

Finally, we classify the entire solutions of Radó's partial differential equation.

Theorem 4.2 (Bernstein-type result). *Let $\varrho = \varrho(x, y) \in C^2(\mathbb{R}^2, (0, +\infty))$ represent a positive solution of the minimal surface equation in central projection*

$$\text{(4-7)} \quad \big(\varrho_y^2 \cdot (x^2 + y^2 + 1) + 2\varrho y \varrho_y + \varrho^2\big)\varrho_{xx} \\ - 2\big(\varrho_x \varrho_y \cdot (x^2 + y^2 + 1) + \varrho[y\varrho_x + x\varrho_y]\big)\varrho_{xy} \\ + \big(\varrho_x^2 \cdot (x^2 + y^2 + 1) + 2\varrho x \varrho_x + \varrho^2\big)\varrho_{yy}(x, y) - 2\varrho|\nabla\varrho(x, y)|^2 = 0,$$

for $(x, y) \in \mathbb{R}^2$. Then it follows that $\varrho(x, y) = c$ for all $(x, y) \in \mathbb{R}^2$, with a positive constant $0 < c < \infty$.

Proof. (1) We define the complete minimal embedding

$$\text{(4-8)} \qquad X(x, y) := (x\varrho(x, y), y\varrho(x, y), \varrho(x, y)), \quad (x, y) \in \mathbb{R}^2,$$

in the Euclidean space $\mathbb{R}^3$. For an arbitrary radius $0 < R < +\infty$, we consider the geodesic disc about the center $X_0 := X(0, 0)$ parametrized over the domain $(0, 0) \in D_R \subset \mathbb{R}^2$. Into this minimal disc $X(x, y)$, $(x, y) \in D_R$, we introduce conformal parameters and obtain the parametric minimal surface $X(u, v)$, $(u, v) \in B$ with its unit normal $N(u, v)$, $(u, v) \in B$.

(2) The auxiliary function

$$\theta(u,v) := N(u,v)\cdot X(u,v), \quad (u,v)\in B,$$

satisfies the conditions

$$\text{(4-9)}\qquad \begin{aligned} \Delta\theta(u,v) - 2E(u,v)K(u,v)\theta(u,v) = 0, &\quad (u,v)\in B,\\ \theta(u,v) > 0, &\quad (u,v)\in \bar{B}. \end{aligned}$$

The arguments of Theorem 1 in [Dierkes et al. 2010a, §5.4] show that our minimal surface $X(u,v)$, $(u,v)\in B$, is stable in the sense of the inequality (4-5). This property holds true for the discs of all radii $0 < R < +\infty$ about X_0.

(3) Theorem 3 in [Dierkes et al. 2010a, §5.5] shows that the surface (4-8) represents a plane within the half-space $z > 0$. It follows that $\varrho(x,y) = c$ for all $(x,y)\in\mathbb{R}^2$, with a positive constant $0 < c < +\infty$. □

5. Large *H*-surfaces bounding extreme contours

We return to surfaces of prescribed mean curvature with compact support and solve the associate Dirichlet problem on convex domains. We use the deformation method, which is presented in [Sauvigny 2012b, Chapter 13, §1] for the nonparametric H-surface equation.

Lemma 5.1 (nondegeneracy result). *Suppose $X(u,v) := (x(u,v), y(u,v), z(u,v))$, $(u,v)\in B$, is a parametric H-surface, with unit normal $N(u,v)$, $(u,v)\in B$, and suppose its prescribed mean curvature* (2-12) *in the class $C^{1+\alpha}(\Omega\times\mathbb{R})$ with* (2-13) *satisfies the monotonicity condition. Furthermore, let the auxiliary function $\theta(u,v) := N(u,v)\cdot X(u,v) \geq 0$, $(u,v)\in B$, possess a zero $(u_0,v_0)\in B$ within this disc. Then the identity $\theta(u,v)\equiv 0$, $(u,v)\in B$, follows.*

Proof. For our prescribed mean curvature H, we easily determine the equation

$$\text{(5-1)}\quad \begin{aligned} \nabla H(X)\cdot X &= r\cdot\frac{d}{dr}\left\{\frac{D(x,y,\ln r)}{r}\right\} = r\cdot\left\{\frac{D_z(x,y,\ln r)}{r^2} - \frac{D(x,y,\ln r)}{r^2}\right\}\\ &= \frac{D_z(x,y,\ln r)}{r} - \frac{D(x,y,\ln r)}{r} \geq -\frac{D(x,y,\ln r)}{r}\\ &= -H(\xi,\eta,\zeta) = -H(X), \end{aligned}$$

for all points $X = (\xi,\eta,\zeta) = (rx, ry, r)\in\mathcal{C}(\Omega)$, $(x,y)\in\Omega$, $0 < r < +\infty$. With the aid of Hilfssatz 1 and Satz 1 in [Sauvigny 1982] (see also [Dierkes et al. 2010a, §5.1, Theorem 1]) and the estimate (5-1), we obtain the following differential inequality on the unit disc B for our auxiliary function:

$$
\begin{aligned}
(5\text{-}2)\quad \Delta\theta(u,v) &\\
&= (\Delta N(u,v))\cdot X(u,v) + 2\nabla N(u,v)\cdot\nabla X(u,v) + N(u,v)\cdot(\Delta X(u,v))\\
&= X(u,v)\cdot\Delta N(u,v) - N(u,v)\cdot\Delta X(u,v)\\
&= X(u,v)\cdot\Delta N(u,v) - 2E(u,v)H|_{X(u,v)}\\
&= -q(u,v)\theta(u,v) - 2E(u,v)\nabla H|_{X(u,v)}\cdot X(u,v) - 2E(u,v)H|_{X(u,v)}\\
&\le -q(u,v)\theta(u,v).
\end{aligned}
$$

Here we have used the potential

$$(5\text{-}3)\quad q(u,v) := 2\big(2E(u,v)H|^2_{X(u,v)} - EK(u,v) - E(u,v)\nabla H|_{X(u,v)}\cdot N(u,v)\big),$$

for $(u,v)\in B$. Since a point $(u_0,v_0)\in B$ with $\theta(u_0,v_0)=0$ exists, the nonnegative function θ solving (5-2) has to vanish in B identically, due to Hilfssatz 5 in [Sauvigny 1982]. □

By a nonlinear continuity method we prove the following theorem.

Theorem 5.1 (solution of $\boldsymbol{P}(\Omega,\phi,H^*)$). *On the convex $C^{2+\alpha}$-Jordan domain Ω we prescribe the logarithmic mean curvature*

$$D(x,y) = D^*(x,y) : \Omega\to\mathbb{R}\in C_0^{1+\alpha}(\Omega)$$

with compact support, subject to the inequalities

$$(5\text{-}4)\qquad \widehat{D}(x,y)\le D(x,y)\le 0\quad \textit{for all points } (x,y)\in\Omega.$$

Denote by H^* its homogeneous continuation onto the cone $\mathcal{C}(\Omega)$ due to (2-13). *Then the Dirichlet problem $\boldsymbol{P}(\Omega,\phi,H^*)$ of the H^*-surface equation in central projection possesses a solution $\varrho=\varrho(x,y)\in C^{2+\alpha}(\bar{\Omega})$ for all $C^{2+\alpha}$ functions $\phi:\partial\Omega\to(0,+\infty)$.*

Proof. (1) We introduce the positive quantity

$$(5\text{-}5)\qquad r^* := (1+b^2)\cdot\max_{(\xi,\eta)\in\partial\Omega}\phi(\xi,\eta).$$

Then we choose a weakly monotonically decreasing function $\chi=\chi(z)\in C^1(\mathbb{R},[0,1])$ with the properties

$$(5\text{-}6)\quad \chi(z)=1\ \text{ for all } z\in(-\infty,\ln r^*],\qquad \chi(z)=0\ \text{ for all } z\in[\ln(r^*+1),+\infty).$$

Thus we prescribe the logarithmic mean curvature

$$(5\text{-}7)\qquad D=D(x,y,z) := D^*(x,y)\cdot\chi(z),\qquad (x,y,z)\in\Omega\times\mathbb{R},$$

on the cylinder $\Omega \times \mathbb{R}$. Due to (2-13), we observe the associate mean curvature H on the cone $\mathcal{C}(\Omega)$ as follows:

$$H(\xi,\eta,\zeta) = \frac{D^*(x,y)}{r} = H^*(\xi,\eta,\zeta), \quad (\xi,\eta,\zeta) = (rx, ry, r), \quad \text{with } (x,y) \in \Omega,\ 0 < r \le r^*, \tag{5-8}$$

and

$$H(\xi,\eta,\zeta) = 0, \quad (\xi,\eta,\zeta) = (rx, ry, r), \quad \text{with } (x,y) \in \Omega,\ r^* + 1 \le r < +\infty. \tag{5-9}$$

(2) For the parameters $1 \le \lambda < +\infty$ let us consider the Dirichlet problems $\boldsymbol{P}(\Omega, \lambda\phi, H)$ with the mean curvature H being prescribed. These H-surfaces with one-to-one central projection bound the contours

$$\Gamma_\lambda := \{(x\lambda\phi(x,y), y\lambda\phi(x,y), \lambda\phi(x,y)) \in \mathbb{R}^3 \mid (x,y) \in \partial\Omega\} \tag{5-10}$$

situated on the boundary of our cone. On account of (5-9) we can choose a parameter $1 < \lambda^* < +\infty$ large enough that the mean curvature H vanishes within the convex hull of the contour Γ_{λ^*}. Therefore, we can solve the Dirichlet problem $\boldsymbol{P}(\Omega, \lambda^*\phi, H)$ with the aid of Theorem 4.1.

(3) Now the set

$$\Lambda := \{\lambda \in [1, \lambda^*] \mid \boldsymbol{P}(\Omega, \lambda\phi, H) \text{ possesses a solution}\} \tag{5-11}$$

is open, since the solutions are stable with respect to small homothetic perturbations of the bounding contour due to Lemma 3.1. Here we use that the logarithmic mean curvature from (5-6) and (5-7) satisfies the monotonicity condition in Definition 3.2. Moreover, the set Λ is closed since the property of *one-to-one central projection* remains valid in the limit. This follows from Lemma 5.1 which requires the monotonicity condition again. Since the cone is convex and the mean curvature vanishes near the boundary, the surfaces approach the conical boundary $\partial\mathcal{C}(\Omega)$ transversally.

(4) In order to establish the compactness of our solutions, we need a joint bound on the area of the surfaces. Here we use the area estimate from Lemma 3.2 as follows: on account of Theorem 3.2 we obtain a bound from below for the mean curvature of our surfaces. Furthermore, this Theorem 3.2 yields a uniform estimate for the volumes appearing in Lemma 3.2.

With the aid of the Courant–Lebesgue lemma and the geometric maximum principle in a local version, we can easily derive a modulus of continuity for our parametric solutions in order to establish the equicontinuity of our functions on the closed disc $\bar{B}$. In this context, we use at each point $(\xi_0, \eta_0, \zeta_0) = (r_0 x_0, r_0 y_0, r_0) \in \mathcal{C}(\bar{\Omega})$, with $(x_0, y_0) \in \bar{\Omega}$ and $r_0 > 0$, the *conical ε-neighborhood*

$$
\begin{aligned}
\text{(5-12)}\quad & U_\varepsilon(\xi_0, \eta_0, \zeta_0) \\
&:= \big\{(rx, ry, r) \in \mathbb{R}^3 \mid (x, y) \in \bar{\Omega} \text{ with } (x - x_0)^2 + (y - y_0)^2 \le \varepsilon^2 \text{ and} \\
&\qquad r \in \mathbb{R} \text{ with } (1-\varepsilon) r_0 \le r \le (1+\varepsilon) r_0 \mu(x_0, y_0, \varepsilon)/(1 + x^2 + y^2)\big\},
\end{aligned}
$$

where we need the function

$$\mu(x_0, y_0, \varepsilon) := \sup\{1 + x^2 + y^2 \mid (x, y) \in \Omega \text{ with } (x - x_0)^2 + (y - y_0)^2 < \varepsilon^2\}.$$

The parametric representation $X(u, v)$, $(u, v) \in \bar{B}$, of a solution $\varrho = \varrho(x, y)$, $(x, y) \in \bar{\Omega}$, for our Dirichlet problem under condition (5-4) is subject to the *inclusion principle*

$$X(\partial\Theta) \subset U_\varepsilon(\xi_0, \eta_0, \zeta_0) \quad \Rightarrow \quad X(\bar{\Theta}) \subset U_\varepsilon(\xi_0, \eta_0, \zeta_0) \quad \text{for all domains } \Theta \subset \Omega,$$

where we use the proof of Theorem 3.2. Then we can adapt the proof of Theorem 2(iii) in [Dierkes et al. 2010a, §7.1] and especially Lemma 4 to obtain the desired equicontinuity. Alternatively, we can modify the proofs of Satz 5 and Hilfssatz 10 in [Sauvigny 1982] by using the inclusion principle above.

(5) Now we combine from [Sauvigny 2012b, Chapter 12] the gradient estimate Theorem 2.6 by Heinz and the inner $C^{1+\alpha}$-estimate Theorem 2.7 for the H-surface system (3-31), which both require a smallness condition, with the modulus of continuity as in proof of Theorem 5.4(2). Thus we obtain an inner $C^{1+\alpha}$-estimate for our solutions, which implies an interior $C^{2+\alpha}$-estimate via Theorem 4.4 in [Sauvigny 2012b, Chapter 9].

Therefore, we can extract a uniformly convergent subsequence on $\bar{B}$ which converges in $C^{2+\alpha}_{\text{loc}}(B) \cap C^0(\bar{B})$ to a solution of Plateau's problem. We invoke the boundary regularity result proved by Heinz [1970] and Hildebrandt (see [Dierkes et al. 2010b, §2.3, Theorem 2]). Thus our limit surface belongs to the Banach space $C^{2+\alpha}(\bar{B})$.

Alternatively, we could control the convergence within the Banach space $C^{2+\alpha}(\bar{B})$ with the aid of [Dierkes et al. 2010b, §2.1, Proposition 2 and Lemma 7; §2.2, Theorem 2], together with Theorems 4.6 and 5.2 from [Sauvigny 2012b, Chapter 9].

(6) Since our set Λ is nonempty, we obtain the identity $\Lambda = [1, \lambda^*]$. Therefore, the problem $\boldsymbol{P}(\Omega, \phi, H)$ possesses a solution. Finally, we use Theorem 3.2 and remember (5-5). Consequently, the solution of $\boldsymbol{P}(\Omega, \phi, H)$ lies within the conical section described in (5-8), where the curvatures H and H^* coincide. Thus we have found a solution of the Dirichlet problem $\boldsymbol{P}(\Omega, \phi, H^*)$. □

Definition 5.1. To each Jordan domain $\Omega \subset \mathbb{R}^2$ and positive boundary distribution $\phi : \partial\Omega \to (0, +\infty) \in C^0(\partial\Omega, \mathbb{R})$ we associate the Jordan contour

$$\Gamma = \Gamma(\Omega, \phi) := \{(x\phi(x, y), y\phi(x, y), \phi(x, y)) \in \mathbb{R}^3 \mid (x, y) \in \partial\Omega\} \tag{5-13}$$

with one-to-one central projection onto the curve $\partial\Omega$. We name Γ an *extreme contour* if each point $X_0 = (\xi_0, \eta_0, \zeta_0) \in \Gamma$ admits a real number $a = a(X_0) > 0$ such that

$$\Gamma \subset \mathcal{K}(a) \quad \text{and} \quad X_0 \in \Gamma \cap \partial\mathcal{K}(a). \tag{5-14}$$

Here we use the *balls of support*

$$\mathcal{K}(a) := \{(\xi, \eta, \zeta) \in \mathbb{R}^3 \mid \xi^2 + \eta^2 + (\zeta - a/2)^2 \le a^2/4\} \tag{5-15}$$

of radius $a/2 > 0$ about the center $(0, 0, a/2)$.

Theorem 5.2 (large embedded solutions for Plateau's problem). *We have a function $\phi = \phi(x, y) : \partial\Omega \to (0, +\infty) \in C^{2+\alpha}$, on the boundary of a convex $C^{2+\alpha}$-Jordan domain $\Omega \subset \mathbb{R}^2$, such that $\Gamma = \Gamma(\Omega, \phi)$ represents an extreme contour. Furthermore, let us prescribe with $D = D(x, y) : \bar{\Omega} \to \mathbb{R} \in C^{1+\alpha}(\bar{\Omega})$ the logarithmic mean curvature subject to the restriction* (5-4) *and denote by H its homogeneous continuation onto the cone $\mathcal{C}(\Omega)$ due to* (2-13). *Then we can solve the Dirichlet problem $\boldsymbol{P}(\Omega, \phi, H)$ by a function $\varrho = \varrho(x, y) \in C^{2+\alpha}(\bar{\Omega})$.*

Proof. (1) Approximate $D = D(x, y)$ by functions $D^k = D^k(x, y) : \Omega \to \mathbb{R} \in C_0^{1+\alpha}(\Omega)$ for $k = 1, 2, \ldots$ within $C^{\alpha}_{\text{loc}}(\Omega)$ which are dominated as follows:

$$D(x, y) \le D^k(x, y) \le 0, \quad (x, y) \in \Omega, \quad \text{for all } k \in \mathbb{N}. \tag{5-16}$$

Therefore, these functions D^k are subject to the restrictions (5-4), and the convergence in Ω is compactly uniform. Then we denote by H^k their continuation onto the cone $\mathcal{C}(\Omega)$ due to (2-13). With the aid of Theorem 5.1 we solve the Dirichlet problems $\boldsymbol{P}(\Omega, \phi, H^k)$ by the functions $\varrho^k = \varrho^k(x, y) \in C^{2+\alpha}(\bar{\Omega})$ for all $k \in \mathbb{N}$.

(2) We choose the solution $\tilde{\varrho}(x, y)$, $(x, y) \in \bar{\Omega}$, of the problem $\boldsymbol{P}(\Omega, \phi, 0)$ in Theorem 4.1 as a *lower barrier function*. Now we observe that Γ represents an extreme contour. For each $(x_0, y_0) \in \partial\Omega$ with the associate point

$$X_0 := (x_0\phi(x_0, y_0), y_0\phi(x_0, y_0), \phi(x_0, y_0)) \in \Gamma$$

we can find a real number $a = a(x_0, y_0) > 0$ such that the ball $\mathcal{K}(a)$ from (5-15) satisfies the conditions (5-14). Together with the solution

$$\hat{\varrho}(x, y) := a(1 + x^2 + y^2)^{-1}, \quad (x, y) \in \bar{\Omega},$$

from Theorem 2.1 as an *upper barrier function*, we obtain the estimates

$$(5\text{-}17)\qquad \tilde{\varrho}(x,y) \le \varrho^k(x,y) \le \hat{\varrho}(x,y) \quad \text{for all } (x,y) \in \bar{\Omega} \text{ and all } k \in \mathbb{N}$$

via the method of proof in Theorem 3.2.

(3) The parametric representations of these solutions yield H^k-surfaces $X^k(u,v)$ of equibounded mean curvature and belonging to $C^{2+\alpha}(\bar{B})$. Now we use the arguments in parts (4) and (5) of the proof for Theorem 5.1 and establish the equicontinuity of $\{X^k\}_{k=1,2,3,\dots}$ on the closed disc by the inclusion principle. With the aid of [Dierkes et al. 2010b, §2.2, Theorem 2 and §2.3, Theorem 2] we see that these functions converge to a function $X \in C^{1+\alpha}(\bar{B}, \mathcal{T})$ in isothermal parameters which is situated in the *spherical solid*

$$(5\text{-}18)\qquad \mathcal{T} := \{Y = (rx, ry, r) \in \mathbb{R}^3 \mid (x,y) \in \bar{\Omega},\ \tilde{\varrho}(x,y) \le r \le \hat{\varrho}(x,y)\}.$$

At the point $X_0 \in \partial\mathcal{T}$ the surfaces associated with the lower and upper barrier functions form an angle $\omega = \omega(x_0, y_0) \in (0, \pi)$.

The inclusion $X(\bar{B}) \subset \mathcal{T}$ and the representation $X(u_0, v_0) = X_0 \in \partial\mathcal{T}$ with $(u_0, v_0) \in \partial B$ imply that the point (u_0, v_0) does not constitute a branch point of X. Otherwise the local expansion there would imply that the surface X protrudes from $\mathcal{T}$ — an evident contradiction.

Furthermore, the inclusion $X(B) \subset \mathcal{C}(\Omega)$ holds true. If $X(u_0, v_0) = X_0 \in \partial\mathcal{T}$ were true for a point $(u_0, v_0) \in B$, the local expansion of X would force the surface to protrude from $\mathcal{T}$ — which is impossible. Since the boundary point X_0 can be chosen arbitrarily on Γ, the inclusion above and the exclusion of branch points on ∂B is established.

For the local expansions, we refer our readers to Theorem 2 and Corollary 2 in [Dierkes et al. 2010b, §3.1] and to the original paper by Heinz [1970].

(4) The functions $X^k(u,v) \in C^{2+\alpha}(\bar{B})$ satisfy the nonlinear elliptic systems

$$(5\text{-}19)\qquad \Delta X^k(u,v) = 2H^k(X^k(u,v))X^k_u \wedge X^k_v \quad \text{on } \bar{B} \text{ for } k = 1, 2, \dots.$$

Since the mean curvatures $H^k = H^k(\xi, \eta, \zeta) : \mathcal{C}(\Omega) \to \mathbb{R}$ converge compactly uniformly in the open cone $\mathcal{C}(\Omega)$ and the surfaces $X^k = X^k(u,v) : \bar{B} \to \overline{\mathcal{C}(\Omega)}$ converge due to (3) uniformly on $\bar{B}$ to the continuous function $X : \bar{B} \to \overline{\mathcal{C}(\Omega)}$ with the property $X(B) \subset \mathcal{C}(\Omega)$, we see the limit relation

$$(5\text{-}20)\qquad \lim_{k\to\infty} H^k(X^k(u,v)) = H(X(u,v)) \quad \text{for all } (u,v) \in B.$$

Since the relation (5-20) occurs within $C^{\alpha}_{\mathrm{loc}}(B)$ and a modulus of continuity in (3) has been established, we can use the arguments from part (5) in the proof of Theorem 5.1. Consequently, the functions X^k converge within the space $C^{2+\alpha}_{\mathrm{loc}}(B)$

to the H-surface

$$X = X(u, v) \in C^{2+\alpha}(B) \cap C^0(\bar{B})$$

bounding the regular $C^{2+\alpha}$-contour Γ.

(5) We invoke Theorem 2 in [Dierkes et al. 2010b, §2.3] again and see that $X = X(u, v) \in C^{2+\alpha}(\bar{B})$. Furthermore, Lemma 5.1 guarantees that our limit surface satisfies

$$X(u, v) \cdot N(u, v) > 0, \quad (u, v) \in \bar{B}, \tag{5-21}$$

where we use the inclusion $X(\bar{B}) \subset \mathcal{T}$ at the boundary. By a winding number argument, we can easily exclude the interior branch points, and the H-surface $X : \bar{B} \to \mathbb{R}^3$ is a differential-geometrically regular surface. Finally, the nonparametric representation of this surface $\varrho = \varrho(x, y) \in C^{2+\alpha}(\bar{\Omega})$ solves the Dirichlet problem $\boldsymbol{P}(\Omega, \phi, H)$. □

Remark. For arbitrary $a > 0$, let us consider a regular $C^{2+\alpha}$-Jordan contour on the boundary of the ball $\mathcal{K}(a)$ with a one-to-one and convex central projection onto the plane $\mathcal{E}$. Due to Theorem 5.2 above, we can construct for all nonpositive curvatures H greater or equal to the mean curvature of the upper hemisphere an H-surface bounding the contour Γ with one-to-one projection onto $\mathcal{E}$. Since these surfaces include the large spherical caps, we receive *large embedded solutions of Plateau's problem*. We have to distinguish our considerations from the investigations of H. Brézis and M. Coron [1984] or independently of M. Struwe [1985]. For constant H they construct two solutions of Rellich's H-surface system by variational methods and obtain two not necessarily immersed H-surfaces which solve Plateau's problem for the same contour.

6. The Dirichlet problem $\boldsymbol{P}(\Omega_b, \phi, H)$ on discs

In this section we concentrate on *circular cones* $\mathcal{C}(\Omega_b)$ associated with the discs

$$\Omega_b := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < b^2\}$$

of radius $0 < b < +\infty$ about the origin. One can easily prove the following lemma.

Lemma 6.1 (boundary curvature). *For arbitrary radii $0 < b < +\infty$ we parametrize the circular cones $\partial\mathcal{C}(\Omega_b)$ by*

$$Y = Y(r, t) := (rb \cos t, rb \sin t, r), \quad 0 < r < +\infty, \quad 0 \le t < 2\pi. \tag{6-1}$$

Their mean curvature with respect to the interior normal is given by

$$H_b(r, t) = \frac{1}{2rb\sqrt{1 + b^2}}, \quad 0 < r < +\infty, \quad 0 \le t < 2\pi. \tag{6-2}$$

Theorem 6.1 (solution of $\boldsymbol{P}(\Omega_b, \phi, H)$). *On the closed disc $\bar{\Omega}_b$ of radius $0 < b < \infty$, let us prescribe the logarithmic mean curvature $D(x, y) \in C^{1+\alpha}(\bar{\Omega}_b)$ subject to the inequalities* (5-4) *and the estimate*

$$-\frac{1}{2b\sqrt{1+b^2}} < D(x, y) \leq 0 \quad \textit{for all points } (x, y) \in \partial\Omega_b. \tag{6-3}$$

We denote by H its homogeneous continuation onto the circular cone $\mathcal{C}(\Omega_b)$ due to (2-13). *Then the Dirichlet problem $\boldsymbol{P}(\Omega_b, \phi, H)$ possesses a solution*

$$\varrho = \varrho(x, y) \in C^{2+\alpha}(\Omega_b) \cap C^0(\bar{\Omega}_b)$$

for all continuous boundary distributions $\phi : \partial\Omega_b \to (0, +\infty)$.

Proof. (1) As in the proof of Theorem 5.2, let us approximate $D = D(x, y)$ by the functions

$$D^k = D^k(x, y) : \Omega_b \to \mathbb{R} \in C_0^{1+\alpha}(\Omega_b) \quad \text{for } k = 1, 2, \ldots$$

within $C^{\alpha}_{\text{loc}}(\Omega_b)$, dominated due to (5-16) when we replace the domain Ω with Ω_b. We denote by H^k their continuation onto the circular cone $\mathcal{C}(\Omega_b)$ due to (2-13) and approximate the continuous boundary distribution $\phi : \partial\Omega_b \to (0, +\infty) \in C^0$ uniformly by the sequence $\phi^k : \partial\Omega_b \to (0, +\infty) \in C^{2+\alpha}$ for $k = 1, 2, \ldots$.

With the aid of Theorem 5.1, we solve the Dirichlet problems $\boldsymbol{P}(\Omega_b, \phi^k, H^k)$ by the functions $\varrho^k = \varrho^k(x, y) \in C^{2+\alpha}(\bar{\Omega}_b)$ for all $k \in \mathbb{N}$. In the parametric form we receive H^k-surfaces $X^k(u, v) \in C^{2+\alpha}(\bar{B})$ bounding the Jordan contours $\Gamma^k := \Gamma(\Omega_b, \phi_k)$ from (5-13). By an area estimate as in part (4) of the proof of Theorem 5.1, we select a subsequence of $\{X^k(u, v)\}_{k=1,2,\ldots}$ which is uniformly convergent on $\bar{B}$ to the limit

$$X = X(u, v) \in C^0(\bar{B}, \mathcal{C}(\bar{\Omega}_b)).$$

(2) Let us take a point $(x, y) = (b\cos t, b\sin t) \in \partial\Omega_b$ with an appropriate $0 \leq t < 2\pi$ and a number $r > 0$, such that we obtain the boundary point $(rx, ry, r) \in \partial\mathcal{C}(\Omega_b)$ of the cone. We use the balls

$$K_\delta(rx, ry, r) := \{(\xi, \eta, \zeta) \in \mathbb{R}^3 \mid (\xi - rx)^2 + (\eta - ry)^2 + (\zeta - r)^2 < \delta^2\}$$

about this point of radius $\delta = \delta(rx, ry, r) > 0$, which we shall choose sufficiently small. Now we need *circular cylinders of curvature $h > 0$* which are generated as images of the *standard cylinder*

$$\mathcal{S}_h := \{(\xi, \eta, \zeta) \in \mathbb{R}^3 \mid \xi^2 + \eta^2 < 1/4h^2\} \tag{6-4}$$

of curvature h under an appropriate rotation and translation.

At the point (rx, ry, r) we use as the *cylinder of support* $\mathcal{S}_h(rx, ry, r)$ the circular cylinder of curvature

$$h := \frac{1}{2r}\left(\frac{1}{2b\sqrt{1+b^2}} - D(x, y)\right) \in \left(-\frac{1}{r}D(x, y), H_b(r, t)\right) \tag{6-5}$$

with the properties

$$\begin{gathered}\mathcal{C}(\Omega_b) \cap K_\delta(rx, ry, r) \subset \mathcal{S}_h(rx, ry, r) \cap K_\delta(rx, ry, r),\\ \partial\mathcal{C}(\Omega_b) \cap \partial\mathcal{S}_h(rx, ry, r) \cap K_\delta(rx, ry, r) = L_{(x,y)} \cap K_\delta(rx, ry, r).\end{gathered} \tag{6-6}$$

Here we have to apply our Lemma 6.1 from above. For the prescribed mean curvature, the estimate

$$|H(\xi, \eta, \zeta)| \le h \quad \text{at all points } (\xi, \eta, \zeta) \in \mathcal{C}(\bar{\Omega}_b) \cap K_\delta(rx, ry, r) \tag{6-7}$$

holds true.

(3) The limit surface $X(u, v)$ from (1) cannot touch the cone $\partial\mathcal{C}(\Omega_b)$ at an interior point. If this happened, we could find a point $(u_0, v_0) \in B$ such that $X(u_0, v_0) = (rx, ry, r)$ holds true for a boundary point $(rx, ry, r) \in \partial\mathcal{C}(\Omega_b)$ considered in (2) above. Now we use Hildebrandt's *geometric maximum principle for H-surfaces in circular cylinders*, presented in Hilfssatz 3 in [Sauvigny 1982] or Proposition 1.6 in [Sauvigny 2012b, Chapter 13].

Transforming the setting into the standard cylinder (6-4) by rotation and translation, we show in (4) that the continuous *auxiliary function*

$$\begin{gathered}\Psi(u, v) := x(u, v)^2 + y(u, v)^2, \quad (u, v) \in B,\\ \text{with } (u-u_0)^2 + (v-v_0)^2 < \epsilon^2,\end{gathered} \tag{6-8}$$

is subharmonic in the sense of mean values, where $\epsilon > 0$ is sufficiently small. Due to Theorem 2.9 in [Sauvigny 2012a, Chapter 5], the function Ψ is subject to the maximum principle. Therefore, the surface X would locally coincide with the bounding cylinder and protrude from $\mathcal{C}(\bar{\Omega}_b)$, which is impossible. Consequently, we have

$$X(B) \subset \mathcal{C}(\Omega_b). \tag{6-9}$$

(4) Now we prove that the function Ψ is subharmonic: for $k = 1, 2, \dots$ we consider the *approximate auxiliary functions*

$$\begin{gathered}\Psi^k(u, v) := x^k(u, v)^2 + y^k(u, v)^2, \quad (u, v) \in B,\\ \text{with } (u-u_0)^2 + (v-v_0)^2 < \epsilon^2,\end{gathered} \tag{6-10}$$

associated with the solutions $X^k \in C^{2+\alpha}(\bar{B}, \mathcal{C}(\bar{\Omega}_b))$ of the H-surface system (5-19). Since their mean curvatures H^k are equally bounded as in (6-7) due to (5-16), these functions satisfy

$$\Delta\Psi^k(u,v) \geq 0, \quad (u,v) \in B, \quad \text{with } (u-u_0)^2+(v-v_0)^2<\epsilon^2, \tag{6-11}$$

by Hildebrandt's maximum principle cited above. The functions X^k and Ψ^k converge uniformly, and consequently the subharmonic property (6-11) for Ψ^k — in the mean-value sense — is transferred to the limit function Ψ.

(5) As we have seen in part (4) of the proof of Theorem 5.2, the inclusion (6-9) implies that the convergence of our sequence $X^k(u,v) \in C^{2+\alpha}(\bar{B}),\ k=1,2,\dots,$ occurs in the space $C^{2+\alpha}_{\text{loc}}(B) \cap C^0(\bar{B})$ to the limit surface

$$X = X(u,v) \in C^{2+\alpha}(B) \cap C^0(\bar{B}). \tag{6-12}$$

Lemma 5.1 guarantees that our surface satisfies

$$X(u,v)\cdot N(u,v) > 0, \quad (u,v) \in B. \tag{6-13}$$

Therefore, the surface X has a one-to-one central projection onto the plane $\mathcal{E}$ and possesses the nonparametric representation

$$\varrho = \varrho(x,y) \in C^{2+\alpha}(\Omega_b) \cap C^0(\bar{\Omega}_b) \tag{6-14}$$

solving the Dirichlet problem $\boldsymbol{P}(\Omega_b, \phi, H)$. □

Acknowledgements

I cordially thank Professor Dr. Dr.h.c. Robert Finn (Stanford University) for his interest in my investigations. I am also very grateful to Privatdozent Dr. Ruben Jakob (Eberhard Karls Universität Tübingen) for many valuable suggestions. His detailed comments rendered my paper more readable and the proofs of Lemma 3.1 and Theorems 5.1 and 5.2 more rigorous. Finally I would like to thank the production editor Silvio Levy and his staff for their excellent work.

References

[Brézis and Coron 1984] H. Brézis and J.-M. Coron, "Multiple solutions of H-systems and Rellich's conjecture", *Comm. Pure Appl. Math.* **37**:2 (1984), 149–187. MR 85i:53010 Zbl 0537.49022

[Dierkes et al. 2010a] U. Dierkes, S. Hildebrandt, and F. Sauvigny, *Minimal surfaces*, 2nd ed., Grundlehren der Mathematischen Wissenschaften **339**, Springer, Heidelberg, 2010. MR 2012b:49002 Zbl 1213.53002

[Dierkes et al. 2010b] U. Dierkes, S. Hildebrandt, and A. J. Tromba, *Regularity of minimal surfaces*, 2nd ed., Grundlehren der Mathematischen Wissenschaften **340**, Springer, Heidelberg, 2010. MR 2012b:49003 Zbl 1213.53003

[Finn 1954] R. Finn, "On equations of minimal surface type", *Ann. of Math.* (2) **60** (1954), 397–416. MR 16,592b Zbl 0058.32501

[Heinz 1970] E. Heinz, "Über das Randverhalten quasilinearer elliptischer Systeme mit isothermen Parametern", *Math. Z.* **113** (1970), 99–105. MR 41 #7288 Zbl 0176.41004

[Heinz 1971] E. Heinz, "Interior gradient estimates for surface $z = f(x, y)$ with prescribed mean curvature", *J. Differential Geometry* **5**:1–2 (1971), 149–157. MR 44 #7125 Zbl 0212.44001

[Nitsche 1975] J. C. C. Nitsche, *Vorlesungen über Minimalflächen*, Grundlehren der mathematischen Wissenschaften **199**, Springer, Berlin, 1975. MR 56 #6533 Zbl 0319.53003

[Radó 1932] T. Radó, "Contributions to the theory of minimal surfaces", *Acta Sci. Math.* (*Szeged*) **6**:1 (1932), 1–20. Zbl 0005.17901

[Sauvigny 1982] F. Sauvigny, "Flächen vorgeschriebener mittlerer Krümmung mit eineindeutiger Projektion auf eine Ebene", *Math. Z.* **180**:1 (1982), 41–67. MR 83j:53004 Zbl 0465.53003

[Sauvigny 2012a] F. Sauvigny, *Partial differential equations, 1: Foundations and integral representations*, 2nd ed., Springer, Berlin, 2012. MR 2007c:35001 Zbl 1246.35001

[Sauvigny 2012b] F. Sauvigny, *Partial differential equations, 2: Functional analytic methods*, 2nd ed., Springer, London, 2012. MR 2012m:35001 Zbl 1246.35002

[Serrin 1969] J. Serrin, "On surfaces of constant mean curvature which span a given space curve", *Math. Z.* **112** (1969), 77–88. MR 40 #3447 Zbl 0182.24001

[Struwe 1985] M. Struwe, "Large H-surfaces via the mountain-pass-lemma", *Math. Ann.* **270**:3 (1985), 441–459. MR 86d:58024 Zbl 0582.58010

[Tausch 1981] E. Tausch, "The n-dimensional least area problem for boundaries on a convex cone", *Arch. Rational Mech. Anal.* **75**:4 (1981), 407–416. MR 82e:53014 Zbl 0477.53006

Received August 6, 2015. Revised August 11, 2015.

FRIEDRICH SAUVIGNY
MATHEMATISCHES INSTITUT
BRANDENBURGISCHE TECHNISCHE UNIVERSITÄT COTTBUS - SENFTENBERG
PLATZ DER DEUTSCHEN EINHEIT 1
D-03046 COTTBUS
GERMANY
sauvigny@b-tu.de

CONTENTS

Volume 281, no. 1 and no. 2

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

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.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 281 No. 2 April 2016

0030-8730(2016)281:2;1-4