Pacific Journal of Mathematics

Volume 338 No. 2 October 2025

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

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

The subscription price for 2025 is US $677/year for the electronic version, and $917/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber 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 1945-5844 electronic, 0030-8730 printed) 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. 338, No. 2, 2025
https://doi.org/10.2140/pjm.2025.338.209

EQUIVARIANT RIGIDITY OF RICHARDSON VARIETIES

ANDERS S. BUCH, PIERRE-EMMANUEL CHAPUT AND NICOLAS PERRIN

We prove that Schubert and Richardson varieties in flag manifolds are uniquely determined by their equivariant cohomology classes, as well as a stronger result that replaces Schubert varieties with closures of Białynicki-Birula cells under suitable conditions. This is used to prove a conjecture of Buch, Chaput, and Perrin, stating that any two-pointed curve neighborhood representing a quantum cohomology product with a Seidel class is a Schubert variety. We pose a stronger conjecture which implies a Seidel multiplication formula in equivariant quantum K-theory, and prove this conjecture for cominuscule flag varieties.

1. Introduction

A Schubert variety Ω in a flag manifold $X = G/P$ is called *rigid* if it is uniquely determined by its class $[\Omega]$ in the cohomology ring $H^*(X)$. More precisely, if $Z \subset X$ is any irreducible closed subvariety such that $[Z]$ is a multiple of $[\Omega]$ in $H^*(X)$, then Z is a G-translate of Ω. This problem has been studied in numerous papers; see, e.g., [Hong 2005; 2007; Coskun 2011; 2014; 2018; Robles and The 2012; Coskun and Robles 2013; Hong and Mok 2020; Liu et al. 2024]. In this paper we show that all Schubert varieties and Richardson varieties are *equivariantly rigid*. In other words, if $T \subset G$ is a maximal torus, $\Omega \subset X$ is a T-stable Richardson variety, and $Z \subset X$ is a (nonempty) T-stable closed subvariety such that the T-equivariant class $[Z] \in H_T^*(X)$ is a multiple of $[\Omega]$, then $Z = \Omega$.

More generally, let T be an algebraic torus over an algebraically closed field, let X be a nonsingular projective T-variety, and let $\Omega \subset X$ be a T-stable closed subvariety. Let Ω^T denote the set of T-fixed points in Ω. We will say that Ω is *T-convex* if, for any T-stable closed subvariety $Z \subset X$ satisfying $Z^T \subset \Omega$, we have $Z \subset \Omega$. A fixed point $p \in X^T$ is called *fully definite* if all T-weights of the Zariski tangent space T_pX belong to a strict half-space of the character lattice of T. We show that if

Buch was partially supported by NSF grant DMS-2152316. Perrin was partially supported by ANR project FanoHK, grant ANR-20-CE40-0023.
MSC2020: primary 14M15; secondary 14C25, 14L30, 14N35.
Keywords: rigidity, Schubert varieties, equivariant cohomology, Białynicki-Birula decomposition, curve neighborhoods, Seidel representation, quantum K-theory, horospherical varieties.

all T-fixed points in X are fully definite, then any irreducible T-convex subvariety of X is also T-equivariantly rigid. Here Ω is called T-equivariantly rigid if Ω is determined by its class in the T-equivariant Chow cohomology ring of X.

Let $\mathbb{G}_m \subset T$ be a one-parameter subgroup such that $X^T = X^{\mathbb{G}_m}$, and assume that this fixed point set is finite. The associated Białynicki-Birula decomposition of X is given by $X = \bigcup_{p \in X^T} X_p^+$, where $X_p^+ = \{x \in X : \lim_{t \to 0} t \cdot x = p\}$ is the Białynicki-Birula cell of points attracted to p by the action of $t \in \mathbb{G}_m$. This decomposition is called a *stratification* if each cell closure $\overline{X_p^+} \subset X$ is a union of smaller cells. In this case we show that all cell closures are T-convex. We arrive at the following result combining Theorem 4.3 and Proposition 5.3.

Theorem. *Let X be a nonsingular projective T-variety with finitely many T-fixed points, and choose $\mathbb{G}_m \subset T$ such that $X^T = X^{\mathbb{G}_m}$.*

(a) *Assume that the Białynicki-Birula decomposition of X is a stratification. Then each cell closure $\overline{X_p^+}$ is T-convex.*

(b) *Assume that all T-fixed points in X are fully definite. Then any irreducible T-convex subvariety of X is T-equivariantly rigid.*

This result applies to Schubert and Richardson varieties in flag varieties, as well as positroid varieties in Grassmannians, so these subvarieties are both T-convex and T-equivariantly rigid. However, projected Richardson varieties do not in general enjoy these properties; see Remark 6.5. Our theorem also covers a class of horospherical varieties, which includes all nonsingular horospherical varieties of Picard rank 1 [Pasquier 2009].

Our theorem has additional applications in quantum Schubert calculus. Let $X = G/P$ be a complex flag manifold. A Schubert class $[X^w]$ is called a *Seidel class* if the Weyl group element w is the minimal representative of a point in some cominuscule flag variety G/Q. Multiplication by Seidel classes in the quantum cohomology ring $\mathrm{QH}(X)$ is given by the identity $[X^w] \star [X^u] = q^{d(w,u)}[X^{wu}]$, where $d(w, u)$ is the unique minimal degree of a rational curve connecting the opposite Schubert varieties $X_{w_0 w}$ and X^u [Seidel 1997; Belkale 2004; Chaput et al. 2009]. This implies that $[X^{wu}]$ is equal to the class of the curve neighborhood $\Gamma_{d(w,u)}(X_{w_0 w}, X^u)$, defined as the union of all stable curves in X of degree $d(w, u)$ connecting $X_{w_0 w}$ to X^u. We conjectured in [Buch et al. 2023] that this curve neighborhood is in fact the translated Schubert variety

$$\Gamma_{d(w,u)}(X_{w_0 w}, X^u) = w^{-1} \cdot X^{wu}. \tag{1}$$

This has been proved in some cases when X is cominuscule, in all cases when X is a flag variety of type A [Li et al. 2025; Tarigradschi 2023], and for $X = \mathrm{SG}(2, 2n)$ [Benedetti et al. 2024]. Using that $\Gamma_{d(w,u)}(X_{w_0 w}, X^u)$ and $w^{-1} \cdot X^{wu}$ define the same class in $H_T^*(X)$ by an equivariant version of the Seidel multiplication formula from [Chaput et al. 2009; Chaput and Perrin 2023], the identity (1) follows from our result that Schubert varieties are equivariantly rigid.

In this paper we conjecture the more general identity

$$\Gamma_{d(w,u)+e}(X_{w_0w}, X^u) = \Gamma_e(w^{-1}(X^{wu})), \tag{2}$$

where the right-hand side is the union of all stable curves of degree e that pass through $w^{-1} \cdot X^{wu}$. This union is a Schubert variety [Buch et al. 2013] whose Weyl group element was determined in [Buch and Mihalcea 2015]. Denote by $M_{d(w,u)+e}(X_{w_0w}, X^u)$ the moduli space of three-pointed stable maps to X of degree $d(w,u)+e$ and genus zero, which send the first two marked points to X_{w_0w} and X^u, respectively. We further conjecture that the evaluation map $\mathrm{ev}_3 : M_{d(w,u)+e}(X_{w_0w}, X^u) \to \Gamma_e(w^{-1}(X^{wu}))$ is cohomologically trivial. This conjecture implies a Seidel multiplication formula in the equivariant quantum K-theory ring $\mathrm{QK}_T(X)$. We prove this conjecture when X is a cominuscule flag variety, thereby obtaining an equivariant generalization of our Seidel multiplication formula from [Buch et al. 2023]. This generalized Seidel multiplication formula has also been obtained for Grassmannians of type A in [Gorbounov et al. 2025] using different methods. Based on suggestions from Mihail Tarigradschi, we finally apply the methods of [Tarigradschi 2023] to prove the identity (2) when $X = \mathrm{GL}_n(\mathbb{C})/P$ is any flag manifold of Lie type A.

Our paper is organized as follows. In Section 2 we recall some basic facts and notation related to torus actions. In Section 3 we show that if all T-fixed points of X are fully definite, then the fixed point set Z^T of a T-stable subvariety $Z \subset X$ is determined by its equivariant class $[Z] \in H_T^*(X)$. This is used in Section 4 to prove part (b) of the above theorem. Section 5 proves part (a). Section 6 interprets our theorem for flag varieties, which is used in Section 7 to prove the conjecture about curve neighborhoods from [Buch et al. 2023]. Section 8 discusses the more general conjecture as well as its consequences in quantum K-theory. Finally, Section 9 interprets our theorem for certain horospherical varieties.

2. Torus actions

We work with varieties over a fixed algebraically closed field $\mathbb{K}$. Varieties are reduced but not necessarily irreducible. A point will always mean a closed point. The multiplicative group of $\mathbb{K}$ is denoted by $\mathbb{G}_m = \mathbb{K} \smallsetminus \{0\}$. An (algebraic) torus is a group variety isomorphic to $(\mathbb{G}_m)^r$ for some $r \in \mathbb{N}$.

Let $T = (\mathbb{G}_m)^r$ be an algebraic torus. Any rational representation V of T is a direct sum $V = \bigoplus_\lambda V_\lambda$ of weight spaces $V_\lambda = \{v \in V : t \cdot v = \lambda(t)v \ \forall t \in T\}$ defined by characters $\lambda : T \to \mathbb{G}_m$. The *weights* of V are the characters λ for which $V_\lambda \neq 0$. The group of all characters of T is called the *character lattice* and is isomorphic to $\mathbb{Z}^r$. Given a T-variety X, we let $X^T \subset X$ denote the closed subvariety of T-fixed points. A subvariety $Z \subset X$ is called *T-stable* if $t \cdot z \in Z$ for all $t \in T$ and $z \in Z$. In this case Z is itself a T-variety.

Definition 2.1. The T-fixed point $p \in X$ is *nondegenerate* in X if T acts with nonzero weights on the Zariski tangent space T_pX. The point p is *fully definite* if all T-weights of T_pX belong to a strict half-space of the character lattice of T.

Equivalently, $p \in X^T$ is fully definite in X if and only if there exists a cocharacter $\rho : \mathbb{G}_m \to T$ such that $\mathbb{G}_m$ acts with strictly positive weights on T_pX though ρ. For example, if $X = G/P$ is a flag variety and $T \subset G$ is a maximal torus, then all points of X^T are fully definite in X (see Section 6). Any nondegenerate T-fixed point must be isolated in X^T. Fully definite T-fixed points are called *attractive* in many sources, see, e.g., [Brion 1997]; here we follow the terminology from [Białynicki-Birula 1973].

Remark 2.2. If X is a normal quasiprojective T-variety, then $X^{\mathbb{G}_m} = X^T$ holds for all general cocharacters $\rho : \mathbb{G}_m \to T$. Here a cocharacter is called *general* if it avoids finitely many hyperplanes in the lattice of all cocharacters. This follows because X admits an equivariant embedding $X \subset \mathbb{P}(V)$, where V is a rational representation of T [Kambayashi 1966; Mumford 1965; Sumihiro 1974].

In the rest of this paper we let X be a nonsingular T-variety. The T-equivariant Chow cohomology ring of X will be denoted by $H_T^*(X)$; see [Fulton 1998; Anderson and Fulton 2024]. This is an algebra over the ring $H_T^*(\text{point})$, which may be identified with the symmetric algebra of the character lattice of T. Given a class $\sigma \in H_T^*(X)$ and a T-fixed point $p \in X^T$, we let $\sigma_p \in H_T^*(\text{point})$ denote the pullback of σ along the inclusion $\{p\} \to X$. When X is defined over $\mathbb{K} = \mathbb{C}$, Chow cohomology can be replaced with singular cohomology. In fact, our arguments will only depend on equivariant classes $[Z]_p \in H_T^*(\text{point})$ obtained by restricting the class of a T-stable closed subvariety $Z \subset X$ to a fixed point, and these restrictions are independent of the chosen cohomology theory. Similarly, we can use cohomology with coefficients in either $\mathbb{Z}$ or $\mathbb{Q}$.

3. Equivariant local classes

Let Z be a T-variety, fix $p \in Z^T$, and let $\mathfrak{m} \subset \mathcal{O}_{Z,p}$ be the maximal ideal in the local ring of p. Then the tangent cone $C_pZ = \operatorname{Spec}\big(\bigoplus \mathfrak{m}^i/\mathfrak{m}^{i+1}\big)$ is a T-stable closed subscheme of the Zariski tangent space $T_pZ = (\mathfrak{m}/\mathfrak{m}^2)^\vee = \operatorname{Spec}(\operatorname{Sym}(\mathfrak{m}/\mathfrak{m}^2))$. The *local class* of Z at p is defined by (see [Anderson and Fulton 2024, §17.4])

$$\eta_pZ = [C_pZ] \in H_T^*(T_pZ) = H_T^*(\text{point}). \tag{3}$$

When p is a nonsingular point of Z, we have $\eta_pZ = 1$.

Proposition 3.1. *Let Z be a T-variety and let $p \in Z^T$ be fully definite in Z. Then $\eta_p Z \neq 0$ in $H_T^*(\text{point})$.*

Proof. We may assume that p is a singular point of Z, so that $C_p Z$ has positive dimension. Choose $\mathbb{G}_m \subset T$ such that $\mathbb{G}_m$ acts with positive weights on $T_p Z$. It suffices to show that the class of $C_p Z$ is nonzero in $H^*_{\mathbb{G}_m}(T_p Z)$. Let $\{v_1, \dots, v_n\}$ be a basis of $T_p Z$ consisting of eigenvectors of $\mathbb{G}_m$. Then the action of $\mathbb{G}_m$ is given by $t \cdot v_i = t^{a_i} v_i$ for positive integers $a_1, \dots, a_n > 0$. Set $A = \prod_{i=1}^n a_i$, and let $\mathbb{G}_m$ act on $U = \mathbb{K}^n$ by $t \cdot u = t^A u$. Then the map $\phi : T_p Z \to U$ defined by

$$\phi(c_1 v_1 + \cdots + c_n v_n) = (c_1^{A/a_1}, \dots, c_n^{A/a_n})$$

is a finite $\mathbb{G}_m$-equivariant morphism. By [Edidin and Graham 1998, Theorem 4] we obtain

$$H^*_{\mathbb{G}_m}(U \smallsetminus \{0\}) \otimes \mathbb{Q} = H^*(\mathbb{P}U) \otimes \mathbb{Q},$$

where $\mathbb{P}U = (U \smallsetminus \{0\})/\mathbb{G}_m \cong \mathbb{P}^{n-1}$ is the projective space of lines in U, and

$$\phi_*[C_p Z]|_{U \smallsetminus \{0\}} = \deg(\phi)[\phi(C_p Z \smallsetminus \{0\})/\mathbb{G}_m] \in H^*(\mathbb{P}U) \otimes \mathbb{Q}.$$

The result now follows from the fact that every nonempty closed subvariety of projective space defines a nonzero Chow class. □

Corollary 3.2. *Let X be a nonsingular T-variety, $Z \subset X$ a T-stable closed subvariety, and $p \in Z^T$ a T-fixed point of Z. If p is nondegenerate in X and fully definite in Z, then $[Z]_p \neq 0 \in H_T^*(\text{point})$.*

Proof. By [Anderson and Fulton 2024, Proposition 17.4.1] we have

$$[Z]_p = c_m(T_p X / T_p Z) \cdot \eta_p Z,$$

where $m = \dim T_p X - \dim T_p Z$. The result therefore follows from Proposition 3.1, noting that T acts with nonzero weights on $T_p X / T_p Z$. □

The following example rules out some potential generalizations of Corollary 3.2.

Example 3.3. Let $\mathbb{G}_m$ act on $\mathbb{A}^4$ by

$$t.(a, b, c, d) = (ta,\ tb,\ t^{-1}c,\ t^{-1}d)\,.$$

Set $Z = V(ad - bc) \subset \mathbb{A}^4$, and let $p = (0, 0, 0, 0)$ be the origin in $\mathbb{A}^4$. Then

$$T_p Z = T_p \mathbb{A}^4 = \mathbb{A}^4 \quad \text{and} \quad C_p Z = Z.$$

Since $\mathbb{G}_m$ acts trivially on the equation $ad - bc$, we have $\eta_p Z = [Z] = 0$ in $H^*_{\mathbb{G}_m}(\mathbb{A}^4)$ (see [Anderson and Fulton 2024, §2.3]).

4. Rigidity of convex subvarieties

Let T be an algebraic torus and let X be a nonsingular T-variety. We will show in Section 6 that Schubert varieties and Richardson varieties in a flag variety X satisfy the following two definitions.

Definition 4.1. A T-stable closed subvariety $\Omega \subset X$ is *T-equivariantly rigid* if it is uniquely determined by its T-equivariant cohomology class up to a constant. More precisely, if $Z \subset X$ is any T-stable closed subvariety such that $[Z] = c\,[\Omega]$ holds in $H_T^*(X)$ for some $0 \neq c \in \mathbb{Q}$, then $Z = \Omega$.

Definition 4.2. A T-stable closed subvariety $\Omega \subset X$ is *T-convex* if, for any T-stable closed subvariety $Z \subset X$ satisfying $Z^T \subset \Omega$, we have $Z \subset \Omega$.

When the action of T is clear from the context, we frequently drop T from the notation and write simply *equivariantly rigid* and *convex*. Both notions are properties of the T-equivariant embedding $\Omega \subset X$; for example, any T-variety is convex as a subvariety of itself. Intersections of T-convex subvarieties are again T-convex (with the reduced scheme structure). Most of this paper concerns applications of the following observation.

Theorem 4.3. *Let X be a nonsingular projective T-variety such that all fixed points $p \in X^T$ are fully definite in X. Then any irreducible T-convex subvariety of X is T-equivariantly rigid.*

Proof. Let $\Omega \subset X$ be irreducible and convex, and let $Z \subset X$ be any T-stable closed subvariety such that $[Z] = c\,[\Omega]$ holds in $H_T^*(X)$, with $0 \neq c \in \mathbb{Q}$. Then Corollary 3.2 shows that $Z^T = \Omega^T = \{p \in X^T : [Z]_p \neq 0\}$. Since Ω is convex, we obtain $Z \subset \Omega$. Finally, the assumption $[Z] = c\,[\Omega]$ implies that Z and Ω have the same dimension, so we must have $Z = \Omega$. □

Example 4.4. Let X be a nonsingular projective T-variety, let $H_T^*(X)$ be the T-equivariant Chow cohomology ring, and let $\mathcal{L}$ be a T-equivariant line bundle. Given a section $f \in \Gamma(X, \mathcal{L})$, the associated divisor $D = Z(f)$ is T-stable if and only if f is semi-invariant, that is, $f \in \Gamma(X, \mathcal{L})_\lambda$ for some character λ. In this case f is an equivariant section of $\mathcal{L} \otimes \mathbb{K}_{-\lambda}$, and hence $[D] = c_1(\mathcal{L}) - c_1(\mathbb{K}_\lambda) \in H_T^*(X)$. Moreover, the T-stable effective Cartier divisors D' satisfying $[D'] = [D]$ are in bijective correspondence with $\mathbb{P}(\Gamma(X, \mathcal{L})_\lambda)$. It follows that if D is reduced and $\dim \Gamma(X, \mathcal{L}^{\otimes m})_{m\lambda} = 1$ for all $m \in \mathbb{N}$, then D is T-equivariantly rigid. This observation can be used to produce examples of equivariantly rigid subvarieties that are not convex. For example, if $T = (\mathbb{G}_m)^{n+1}$ acts on $\mathbb{P}^n$ through the standard action on $\mathbb{K}^{n+1}$, then any reduced T-stable divisor $D \subset \mathbb{P}^n$ is equivariantly rigid, but D is convex only if it is irreducible; see Theorem 6.3. We have not found an example of an irreducible T-stable subvariety that is equivariantly rigid but not convex.

5. Rigidity of Białynicki-Birula cells

The multiplicative group $\mathbb{G}_m$ is identified with the complement of the origin in $\mathbb{A}^1$. Given a morphism of varieties $f : \mathbb{G}_m \to X$, we write $\lim_{t\to 0} f(t) = p$ if f can be extended to a morphism $\bar{f} : \mathbb{A}^1 \to X$ such that $\bar{f}(0) = p$. This limit is unique when it exists, and it always exists when X is complete.

Let X be a nonsingular projective $\mathbb{G}_m$-variety such that $X^{\mathbb{G}_m}$ is finite. Then each fixed point $p \in X^{\mathbb{G}_m}$ defines the (positive) Białynicki-Birula cell

$$X_p^+ = \{x \in X : \lim_{t\to 0} t \cdot x = p\}.$$

A negative cell is similarly defined by $X_p^- = \{x \in X : \lim_{t\to 0} t^{-1} \cdot x = p\}$. By [Białynicki-Birula 1973, Theorem 4.4], these cells form a locally closed decomposition of X,

$$X = \bigcup_{p\in X^{\mathbb{G}_m}} X_p^+, \tag{4}$$

that is, a disjoint union of locally closed subsets. In addition, each cell X_p^+ is isomorphic to an affine space.

Lemma 5.1. *For any $\mathbb{G}_m$-stable closed subset $Z \subset X$, we have*

$$Z \subset \bigcup_{p\in Z^{\mathbb{G}_m}} X_p^+.$$

Proof. For any point $x \in Z$, we have $x \in X_p^+$, where $p = \lim_{t\to 0} t \cdot x \in Z^{\mathbb{G}_m}$. □

Definition 5.2. A locally closed decomposition $X = \bigcup X_i$ is called a *stratification* if each subset X_i is nonsingular and its closure $\overline{X_i}$ is a union of subsets X_j of the decomposition.

The Białynicki-Birula decomposition (4) typically fails to be a stratification, for example, when X is the blow-up of $\mathbb{P}^2$ at the point $[0, 1, 0]$, where $\mathbb{G}_m$ acts on $\mathbb{P}^2$ by $t \cdot [x, y, z] = [x, ty, t^2 z]$; see [Białynicki-Birula 1976, Example 1]. Lemma 5.1 shows that the Białynicki-Birula decomposition is a stratification if and only if $X_q^+ \subset \overline{X_p^+}$ holds for each fixed point $q \in (\overline{X_p^+})^{\mathbb{G}_m}$. It was proved in [Białynicki-Birula 1976, Theorem 5] that the decomposition is a stratification when each positive cell X_p^+ meets each negative cell X_q^- transversally. In particular, this holds when $X = G/P$ is a flag variety and $\mathbb{G}_m \subset G$ is a general one-parameter subgroup; see [McGovern 2002, Example 4.2] or Lemma 6.1. When both the positive and negative Białynicki-Birula decompositions are stratifications, all cells X_p^+ and X_q^- of complementary dimensions meet transversally, and hence the positive and negative cell closures form a pair of Poincaré dual bases of the cohomology ring $H^*(X)$; see [Benedetti and Perrin 2022, Lemma 3.11]. In this paper we utilize the following application, which is a consequence of Lemma 5.1.

Proposition 5.3. *Assume that the Białynicki-Birula decomposition of X is a stratification. Then each cell closure $\overline{X_p^+} \subset X$ is $\mathbb{G}_m$-convex.*

Corollary 5.4. *Let T be an algebraic torus and X a nonsingular projective T-variety such that all fixed points $p \in X^T$ are fully definite in X. Assume that $X^T = X^{\mathbb{G}_m}$ for some one-parameter subgroup $\mathbb{G}_m \subset T$, such that the associated Białynicki-Birula decomposition of X is a stratification. Then each cell closure $\overline{X_p^+}$ is T-convex and T-equivariantly rigid.*

Proof. The cell X_p^+ is T-stable because T is commutative and $p \in X^T$. The result now follows from Theorem 4.3 and Proposition 5.3. □

Question 5.5. We do not know whether Proposition 5.3 and Corollary 5.4 are true without the assumption that the Białynicki-Birula decomposition of X is a stratification. It would be very interesting to settle this question.

Example 5.6. Let X be a nonsingular projective toric variety, with torus $T \subset X$, and choose $\mathbb{G}_m \subset T$ such that $X^T = X^{\mathbb{G}_m}$. We show that the conclusion of Corollary 5.4 holds, even though the Białynicki-Birula decomposition is rarely a stratification. All fixed points $p \in X^T$ are fully definite in X, as the weights of T_pX form a basis of the character lattice of T. The T-orbits $O_\tau \subset X$ correspond to the cones τ of the fan defining X, and we have $O_\sigma \subset \overline{O_\tau}$ if and only if τ is a face of σ; see [Fulton 1993, §3.1]. In particular, the T-fixed points in X correspond to the maximal cones σ. Since X is complete, each cone τ is the intersection of the maximal cones σ corresponding to the T-fixed points in $\overline{O_\tau}$. Since all cell closures $\overline{X_p^+}$ are T-orbit closures, it suffices to show that each orbit closure $\overline{O_\tau}$ is T-convex. Let $Z \subset X$ be a T-stable closed subvariety such that $Z^T \subset \overline{O_\tau}$. We may assume that Z is irreducible, in which case $Z = \overline{O_\kappa}$ is also a T-orbit closure. Since κ is the intersection of the maximal cones given by the fixed points in Z^T, we obtain $\tau \subset \kappa$ and $\overline{O_\kappa} \subset \overline{O_\tau}$, as required. Now assume that X has dimension two. By [Białynicki-Birula 1973, Corollary 1 of Theorem 4.5], there is a unique repulsive fixed point $b \in X^{\mathbb{G}_m}$ with $X_b^+ = \{b\}$, and a unique attractive fixed point $a \in X^{\mathbb{G}_m}$ such that X_a^+ is a dense open subset of X. For all other fixed points $p \in X^{\mathbb{G}_m} \smallsetminus \{a, b\}$, the cell $X_p^+ \cong \mathbb{A}^1$ is a line. If the Białynicki-Birula decomposition of X is a stratification, then $b \in \overline{X_p^+}$ for all $p \in X^{\mathbb{G}_m}$. The T-fixed point b corresponds to a maximal cone σ, and b is connected to exactly two T-stable lines corresponding to the rays forming the boundary of this cone. We deduce that X contains at most four T-fixed points. Higher-dimensional toric varieties for which the Białynicki-Birula decomposition is not a stratification can be constructed by taking products. We do not know if the cell closures $\overline{X_p^+}$ are $\mathbb{G}_m$-convex when X is a toric variety.[1]

[1]Teddy Gonzales and Chayim Lowen [≥ 2025] have recently produced several examples showing that $\overline{X_p^+}$ may not be $\mathbb{G}_m$-convex when X is a nonsingular projective toric variety.

6. Rigidity of Richardson varieties

Let $X = G/P = \{g.P : g \in G\}$ be a flag variety defined by a connected reductive linear algebraic group G and a parabolic subgroup P. Fix a maximal torus T and a Borel subgroup B such that $T \subset B \subset P \subset G$. The opposite Borel subgroup $B^- \subset G$ is defined by $B^- \cap B = T$. Let Φ be the root system of nonzero weights of $T_1 G$, the tangent space of G at the identity element. The positive roots Φ^+ are the nonzero weights of $T_1 B$. Let $W = N_G(T)/T$ be the Weyl group of G, $W_P = N_P(T)/T$ the Weyl group of P, and let $W^P \subset W$ be the subset of minimal representatives of the cosets in W/W_P. The set of T-fixed points in X is given by $X^T = \{w \cdot P : w \in W\}$, where each point $w \cdot P$ depends only on the coset wW_P in W/W_P. Each fixed point $w \cdot P$ defines the *Schubert varieties* $X_w = \overline{Bw \cdot P}$ and $X^w = \overline{B^- w \cdot P}$. For $w \in W^P$ we have $\dim(X_w) = \operatorname{codim}(X^w, X) = \ell(w)$. The Bruhat order $\leq$ on W^P is defined by

$$u \leq w \iff X_u \subset X_w \iff X^u \supset X^w \iff X^u \cap X_w \neq \varnothing.$$

A *Richardson variety* is any nonempty intersection $X_w^u = X_w \cap X^u$ of opposite Schubert varieties in X. More generally, any G-translate of X_w^u will be called a Richardson variety. Any Richardson variety is reduced, irreducible, and rational; see [Deodhar 1977; Brion and Kumar 2005, §2].

Recall that a cocharacter $\rho : \mathbb{G}_m \to T$ is *strongly dominant* if $\langle \alpha, \rho \rangle > 0$ for all positive roots $\alpha \in \Phi^+$, where $\langle \alpha, \rho \rangle \in \mathbb{Z}$ is defined by $\alpha(\rho(t)) = t^{\langle \alpha, \rho \rangle}$ for $t \in \mathbb{G}_m$. The following lemma is well known; see, e.g., [McGovern 2002, Example 4.2] or [Benedetti and Perrin 2022, Corollary 3.14].

Lemma 6.1. *Let $\rho : \mathbb{G}_m \to T$ be a strongly dominant cocharacter. Then the associated Białynicki-Birula cells of X are given by $X_p^+ = B \cdot p$, for $p \in X^T$.*

Proof. Let $\mathbb{G}_m$ act on G by conjugation through ρ. The fixed point set for this action is [Springer 1998, (7.1.2), (7.6.4)]

$$T = \{g \in G : tgt^{-1} = g \ \forall t \in \mathbb{G}_m\},$$

and the corresponding Białynicki-Birula cell is [Springer 1998, (8.2.1)]

$$B = \{g \in G : \lim_{t \to 0} tgt^{-1} \in T\}.$$

This implies $B \cdot p \subset X_p^+$ for any fixed point $p \in X^{\mathbb{G}_m}$. We deduce from (4) that the positive Białynicki-Birula cells in X are the B-orbits. □

Lemma 6.2. *Let Y be any G-variety, and $\Omega \subset Y$ a T-stable closed subvariety. Any T-stable G-translate of Ω has the form $w \cdot \Omega$, with $w \in N_G(T)$.*

Proof. Let $\Omega' = g \cdot \Omega$ be a T-stable translate, and let $H \subset G$ be the stabilizer of Ω'. Since T and gTg^{-1} are maximal tori in H, we can choose $h \in H$ such that $T = hgTg^{-1}h^{-1}$. We obtain $hg \in N_G(T)$ and $\Omega' = h \cdot \Omega' = hg \cdot \Omega$, as required. □

Theorem 6.3. *Any T-stable Richardson variety in the flag variety $X = G/P$ is T-convex and T-equivariantly rigid.*

Proof. It follows from Proposition 5.3 and Lemma 6.1 that all Schubert varieties X_w and X^u are convex. This implies that every Richardson variety $X_w^u = X_w \cap X^u$ is convex; hence all T-stable Richardson varieties in X are convex by Lemma 6.2. The B-fixed point $p = 1 \cdot P$ is fully definite in X because the weights of T_pX are a subset of the negative roots of G. Since W acts transitively on X^T, this implies that all T-fixed points in X are fully definite. The result therefore follows from Theorem 4.3. □

Let $E = G/B$ denote the variety of complete flags, and let $\pi : E \to X$ be the natural projection. A *projected Richardson variety* in X is the image $\Pi_w^u(X) = \pi(E_w^u)$ of a Richardson variety in E. Projected Richardson varieties in the Grassmannian $X = \mathrm{Gr}(m, n)$ of type A, obtained as images of Richardson varieties in $\mathrm{Fl}(n)$, are also called *positroid varieties*.

Corollary 6.4. *Let $X = \mathrm{Gr}(m, n)$ be a Grassmannian of type A, and let $T = (\mathbb{G}_m)^n$ act on X through the diagonal action on $\mathbb{K}^n$. Then all positroid varieties in X are T-convex and T-equivariantly rigid.*

Proof. It was proved in [Knutson et al. 2013] that any positroid variety Ω is defined by Plucker equations. Equivalently, Ω is an intersection of T-stable Schubert divisors, so Ω is convex by Theorem 6.3 and equivariantly rigid by Theorem 4.3. □

Remark 6.5. Corollary 6.4 does not hold for projected Richardson varieties in arbitrary flag varieties $X = G/P$. Each simple root β defines a projected Richardson divisor $D_\beta = \Pi_{w_0^P}^{s_\beta}(X)$, where w_0^P denotes the longest element in W^P. It frequently happens that two distinct divisors $D_{\beta'}$ and $D_{\beta''}$ have the same T-equivariant cohomology and K-theory classes, which implies that these divisors are not equivariantly rigid. For example, this is the case for the quadric hypersurfaces of dimensions seven and eight, of Lie types B_4 and D_5, and the two-step flag variety $\mathrm{Fl}(1, 4; 5)$ of type A_4. For other flag varieties X, all projected Richardson varieties have distinct equivariant classes, but some projected Richardson divisor D_β contains all T-fixed points in X, which rules out that D_β is convex. For example, this is the case for the Lagrangian Grassmannian $\mathrm{LG}(2, 4)$ of type C_2 and the maximal orthogonal Grassmannian $\mathrm{OG}(4, 8)$ of type D_4. This is a special case of [Benedetti and Perrin 2022, Lemma 3.1], which can be used to produce many more examples.

Any element $u \in W$ has a unique factorization $u = u^P u_P$ for which $u^P \in W^P$ and $u_P \in W_P$, called the *parabolic factorization* with respect to P. This factorization is *reduced* in the sense that $\ell(u) = \ell(u^P) + \ell(u_P)$. The parabolic factorization of the longest element $w_0 \in W$ is $w_0 = w_0^P w_{0,P}$, where w_0^P and $w_{0,P}$ are the longest elements in W^P and W_P, respectively. Since w_0 and $w_{0,P}$ are self-inverse, we have $w_{0,P} = w_0 w_0^P$. As preparation for the next section, we prove the following identity of Schubert varieties.

Lemma 6.6. *Let $Q \subset G$ be a parabolic subgroup containing B and set $w = w_0^Q$. Then $w^{-1} \cdot X^w = X_{w_0 w}$.*

Proof. Since $X_{w_{0,Q}}$ is a Q-stable Schubert variety, we have $X_{w_{0,Q}} = w_{0,Q} \cdot X_{w_{0,Q}}$. By translating both sides by $w = w_0^Q$, we obtain $w \cdot X_{w_0 w} = w_0 \cdot X_{w_0 w} = X^w$. □

7. Seidel neighborhoods

In this section we prove a conjecture about curve neighborhoods from [Buch et al. 2023]. Since this conjecture and its proof relies on the moduli space of stable maps, we will restrict our attention to varieties defined over the field $\mathbb{K} = \mathbb{C}$ of complex numbers. As in Section 6, we let $X = G/P$ denote a flag variety.

For any effective degree $d \in H_2(X, \mathbb{Z})$, we let $M_d = \overline{\mathcal{M}}_{0,3}(X, d)$ denote the Kontsevich moduli space of three-pointed stable maps to X of degree d and genus zero; see [Fulton and Pandharipande 1997]. The evaluation map $\mathrm{ev}_i : M_d \to X$, defined for $1 \le i \le 3$, sends a stable map to the image of the i-th marked point in its domain. Given two opposite Schubert varieties X_v and X^u, the *Gromov–Witten variety* $M_d(X_v, X^u)$ is the variety of stable maps that send the first two marked points to X_v and X^u:

$$M_d(X_v, X^u) = \mathrm{ev}_1^{-1}(X_v) \cap X_2^{-1}(X^u) \subset M_d.$$

The *curve neighborhood* $\Gamma_d(X_v, X^u)$ is the union of all stable curves of degree d in X connecting X_v and X^u:

$$\Gamma_d(X_v, X^u) = \mathrm{ev}_3(M_d(X_v, X^u)) \subset X.$$

Let $\mathbb{Z}[q] = \mathrm{Span}_{\mathbb{Z}}\{q^d : d \in H_2(X, \mathbb{Z}) \text{ effective}\}$ be the semigroup ring defined by the effective curve classes on X. The equivariant quantum cohomology ring of X is an algebra over $H_T^*(\text{point}) \otimes_{\mathbb{Z}} \mathbb{Z}[q]$, which is defined by $\mathrm{QH}_T(X) = H_T^*(X) \otimes_{\mathbb{Z}} \mathbb{Z}[q]$ as a module. The *quantum product* of two opposite Schubert classes is given by

$$[X_v] \star [X^u] = \sum_{d \ge 0} q^d \, \mathrm{ev}_{3,*}[M_d(X_v, X^u)],$$

where the sum is over all effective degrees $d \in H_2(X; \mathbb{Z})$.

A simple root $\gamma \in \Phi^+$ is called *cominuscule* if, when the highest root is written in the basis of simple roots, the coefficient of γ is one. The flag variety G/Q is cominuscule if Q is a maximal parabolic subgroup corresponding to a cominuscule simple root γ, that is, s_γ is the unique simple reflection in W^Q. Let $W^{\text{comin}} \subset W$

be the subset of point representatives of cominuscule flag varieties of G, together with the identity element:

$$W^{\text{comin}} = \{w_0^Q : G/Q \text{ is cominuscule}\} \cup \{1\}.$$

This is a subgroup of W, which is isomorphic to the quotient of the coweight lattice of Φ modulo the coroot lattice [Bourbaki 1981, Proposition VI.2.6]. The isomorphism sends w_0^Q to the class of the fundamental coweight $\omega_\gamma^\vee$ corresponding to Q. Notice that γ is the unique simple root for which $w_0^Q \cdot \gamma < 0$. In the following we set $d(w_0^Q, u) = \omega_\gamma^\vee - u^{-1} \cdot \omega_\gamma^\vee \in H_2(X; \mathbb{Z})$ for any $u \in W$. Here we identify the group $H_2(X, \mathbb{Z})$ with a quotient of the coroot lattice, by mapping each simple coroot $\beta^\vee$ to the curve class $[X_{s_\beta}]$ if $s_\beta \in W^P$, and to zero otherwise.

The *Seidel representation* of W^{comin} on $\mathrm{QH}(X)/\langle q-1\rangle$ is defined by $w \cdot [X^u] = [X^w] \star [X^u]$ for $w \in W^{\text{comin}}$ and $u \in W$. In fact, we have [Seidel 1997; Belkale 2004; Chaput et al. 2009]

$$[X^w] \star [X^u] = q^{d(w,u)} [X^{wu}] \tag{5}$$

in the (nonequivariant) quantum ring $\mathrm{QH}(X)$. This implies that $d(w, u)$ is the unique minimal degree d for which $\Gamma_d(X_{w_0 w}, X^u)$ is not empty [Fulton and Woodward 2004; Buch et al. 2020]. More generally, it was proved in [Chaput et al. 2009; Chaput and Perrin 2023] that the identity

$$[X^w] \star [w \cdot X^u] = q^{d(w,u)} [X^{wu}] \tag{6}$$

holds in the equivariant quantum cohomology ring $\mathrm{QH}_T(X)$. We will discuss generalizations to quantum K-theory in Section 8.

It follows from (5) and the definition of the quantum product in $\mathrm{QH}(X)$ that $[\Gamma_{d(w,u)}(X_{w_0 w}, X^u)] = [X^{wu}]$ holds in $H^*(X)$. Conjecture 3.11 from [Buch et al. 2023] asserts that $\Gamma_{d(w,u)}(X_{w_0 w}, X^u)$ is in fact equal to the translated Schubert variety $w^{-1} \cdot X^{wu}$. This is proved below as a consequence of Theorem 6.3 and (6). This result was known when $X = G/P$ is cominuscule and $w = w_0^P$ [Buch et al. 2023], when X is a Grassmannian of type A and $[X^w]$ is a special Seidel class [Li et al. 2025, Corollary 4.6], when X is any flag variety of type A [Tarigradschi 2023], and when X is the symplectic Grassmannian $\mathrm{SG}(2, 2n)$ [Benedetti et al. 2024, Theorem 8.1].

Theorem 7.1. *Let $X = G/P$ be a complex flag variety. For $w \in W^{\text{comin}}$ and $u \in W$ we have $\Gamma_{d(w,u)}(X_{w_0 w}, X^u) = w^{-1} \cdot X^{wu}$.*

Proof. By applying w^{-1} to both sides of (6) and using Lemma 6.6, we obtain

$$[X_{w_0 w}] \star [X^u] = q^{d(w,u)} [w^{-1} \cdot X^{wu}]$$

in $\mathrm{QH}_T(X)$. By definition of the quantum product, this implies that

$$[w^{-1} \cdot X^{wu}] = \mathrm{ev}_{3,*}[M_{d(w,u)}(X_{w_0 w}, X^u)] = c\,[\Gamma_{d(w,u)}(X_{w_0 w}, X^u)]$$

holds in $H_T^*(X)$, where c is the degree of the map

$$\mathrm{ev}_3 : M_{d(w,u)}(X_{w_0 w}, X^u) \to \Gamma_{d(w,u)}(X_{w_0 w}, X^u).$$

The result therefore follows from Theorem 6.3. □

8. Seidel products in quantum K-theory

In this section we discuss a generalization of the Seidel multiplication formula to quantum K-theory. We start by briefly recalling the definition of quantum K-theory. A more detailed discussion can be found in [Buch et al. 2018a, §2].

Let $X = G/P$ be a flag variety defined over $\mathbb{K} = \mathbb{C}$. The equivariant K-theory ring $K^T(X)$ is an algebra over the representation ring $\Gamma = K^T(\text{point})$. The equivariant quantum K-theory ring $\mathrm{QK}_T(X)$ was originally constructed by Givental [2000] and Lee [2004]. This ring is an algebra over the formal power series ring $\Gamma[\![q]\!] = \Gamma[\![q_\beta : s_\beta \in W^P]\!]$, which has one variable q_β for each simple reflection s_β in W^P. As a module over $\Gamma[\![q]\!]$ we have $\mathrm{QK}_T(X) = K^T(X) \otimes_\Gamma \Gamma[\![q]\!]$. The *undeformed product* of two opposite Schubert classes in $\mathrm{QK}_T(X)$ is defined by

$$[\mathcal{O}_{X_v}] \odot [\mathcal{O}_{X^u}] = \sum_{d \geq 0} q^d \, \mathrm{ev}_{3,*}[\mathcal{O}_{M_d(X_v, X^u)}].$$

Let $\Psi : \mathrm{QK}_T(X) \to \mathrm{QK}_T(X)$ be the $\Gamma[\![q]\!]$-linear map defined by

$$\Psi([\mathcal{O}_{X^w}]) = \sum_{d \geq 0} q^d [\mathcal{O}_{\Gamma_d(X^w)}],$$

where the curve neighborhood $\Gamma_d(X^w) = \mathrm{ev}_2(\mathrm{ev}_1^{-1}(X^w))$ is defined using the evaluation maps from M_d. This curve neighborhood is a Schubert variety in X by [Buch et al. 2013, Proposition 3.2(b)], whose Weyl group element was determined in [Buch and Mihalcea 2015]. By [Buch et al. 2018a, Proposition 2.3], Givental's *quantum K-theory product* $\star$ is given by

$$[\mathcal{O}_{X_v}] \star [\mathcal{O}_{X^u}] = \Psi^{-1}([\mathcal{O}_{X_v}] \odot [\mathcal{O}_{X^u}]). \tag{7}$$

The following conjecture is the K-theoretic analogue of the Seidel multiplication formula (6) in $\mathrm{QH}_T(X)$ proved in [Chaput et al. 2009; Chaput and Perrin 2023].

Conjecture 8.1. *For $w \in W^{\mathrm{comin}}$ and $u \in W$ we have*

$$[\mathcal{O}_{X_{w_0 w}}] \star [\mathcal{O}_{X^u}] = q^{d(w,u)} [\mathcal{O}_{w^{-1}.X^{wu}}] \quad \textit{and} \quad [\mathcal{O}_{X^w}] \star [\mathcal{O}_{w.X^u}] = q^{d(w,u)} [\mathcal{O}_{X^{wu}}]$$

in $\mathrm{QK}_T(X)$.

The two identities in Conjecture 8.1 are equivalent by Lemma 6.6. The nonequivariant case of this conjecture was proved in [Buch et al. 2023, Corollary 3.7] when X is a cominuscule flag variety. When X is a Grassmannian of type A, the conjecture is equivalent to [Gorbounov et al. 2025, Corollary 10.4]. We will prove Conjecture 8.1 for cominuscule flag varieties below, based on the following conjectural generalization of Theorem 7.1. Recall that a morphism $\pi : Z \to Y$ is called *cohomologically trivial* if $\pi_* \mathcal{O}_Z = \mathcal{O}_Y$ and $R^j \pi_* \mathcal{O}_Z = 0$ for $j \geq 1$.

Conjecture 8.2. *Let $w \in W^{\text{comin}}$, $u \in W$, and let $e \in H_2(X, \mathbb{Z})$ be effective.*

(a) *We have $\Gamma_{d(w,u)+e}(X_{w_0 w}, X^u) = \Gamma_e(w^{-1} \cdot X^{wu})$.*

(b) *The evaluation map $\text{ev}_3 : M_{d(w,u)+e}(X_{w_0 w}, X^u) \to \Gamma_{d(w,u)+e}(X_{w_0 w}, X^u)$ is cohomologically trivial.*

Conjecture 8.2 is a variant of the quantum-equals-classical theorem for Gromov–Witten invariants as stated in [Buch et al. 2018b, Theorem 4.1]; see also [Xu 2024, Theorem 1.2]. The conjecture is true for $e = 0$; part (a) is equivalent to Theorem 7.1, and part (b) holds because the map $\text{ev}_3 : M_{d(w,u)}(X_{w_0 w}, X^u) \to \Gamma_{d(w,u)}(X_{w_0 w}, X^u)$ is birational by [Belkale 2004; Chaput et al. 2009], and $M_{d(w,u)}(X_{w_0 w}, X^u)$ has rational singularities by [Buch et al. 2013, Corollary 3.1]. For $e \geq 0$, Theorem 7.1 implies that

$$\Gamma_e(w^{-1}.X^{wu}) = \Gamma_e(\Gamma_{d(w,u)}(X_{w_0 w}, X^u)) \subset \Gamma_{d(w,u)+e}(X_{w_0 w}, X^u), \tag{8}$$

and $\Gamma_{d(w,u)+e}(X_{w_0 w}, X^u)$ is irreducible by Corollary 3.8 in [Buch et al. 2013]. Conjecture 8.2(a) is therefore true if and only if $\Gamma_{d(w,u)+e}(X_{w_0 w}, X^u)$ and $\Gamma_e(X^{wu})$ have the same dimension. We prove below that Conjecture 8.2(a) is true when $X = \text{GL}(n)/P$ is any flag variety of Lie type A. Conjecture 8.1 follows from Conjecture 8.2 by the following observation.

Lemma 8.3. *Given $w \in W^{\text{comin}}$ and $u \in W$, the identity*

$$[\mathcal{O}_{X_{w_0 w}}] \star [\mathcal{O}_{X^u}] = q^{d(w,u)} [\mathcal{O}_{w^{-1}.X^{wu}}]$$

holds in $\text{QK}_T(X)$ if and only if

$$\text{ev}_{3,*}[\mathcal{O}_{M_{d(w,u)+e}(X_{w_0 w}, X^u)}] = [\mathcal{O}_{\Gamma_e(w^{-1}.X^{wu})}] \tag{9}$$

holds in $K_T(X)$ for all effective degrees $e \in H_2(X, \mathbb{Z})$.

Proof. Both assertions are equivalent to the identity

$$[\mathcal{O}_{X_{w_0 w}}] \odot [\mathcal{O}_{X^u}] = \sum_{e \geq 0} q^{d(w,u)+e} [\mathcal{O}_{\Gamma_e(w^{-1}.X^{wu})}]$$

by the definition (7) of the quantum product in $\text{QK}_T(X)$. □

Theorem 8.4. *Conjectures 8.1 and 8.2 hold when X is a cominuscule flag variety.*

Proof. Assume that X is cominuscule. Then Conjecture 8.2(b) is a special case of [Buch et al. 2018b, Theorem 4.1], and Conjecture 8.2(a) follows from Theorem 7.1 and [Buch et al. 2022, Corollary 8.24], noting that $q^{d(w,u)}$ is the maximal power of q occurring in the quantum cohomology product $[X_{w_0 w}]\star[X^u]$ by [Belkale 2004; Chaput et al. 2009]. This proves Conjecture 8.2, which implies Conjecture 8.1 by Lemma 8.3. □

We finish this section by proving that Conjecture 8.2(a) can be reduced to the case where X is a flag variety of Picard rank 1. In particular, Conjecture 8.2(a) follows from Theorem 8.4 in type A. These results were proved for $e=0$ in [Tarigradschi 2023]. We thank Mihail Tarigradschi for suggesting that his methods might apply to the general case of our conjecture.

Recall that $X=G/P$. Let $Q_1, Q_2\subset G$ be parabolic subgroups such that $P=Q_1\cap Q_2$. Set $Y_i=G/Q_i$ and let $\pi_i : X\to Y_i$ be the projection, for $i\in\{1,2\}$. Given a degree $d\in H_2(X,\mathbb{Z})$, we also let d denote the image $\pi_{i,*}(d)$ of this degree in $H_2(Y_i,\mathbb{Z})$. Let $\Gamma_d(Y_{i,v}, Y_i^u)\subset Y_i$ be the union of all stable curves of degree d in Y_i that connect the Schubert varieties $Y_{i,v}=\pi_i(X_v)$ and $Y_i^u=\pi_i(X^u)$, for $u, v\in W$. The next result generalizes [Björner and Brenti 2005, Theorem 2.6.1; Tarigradschi 2023, Lemma 4].

Lemma 8.5. *We have* $\Gamma_d(X^u)=\pi_1^{-1}(\Gamma_d(Y_1^u))\cap\pi_2^{-1}(\Gamma_d(Y_2^u))$.

Proof. Let $\operatorname{dist}_X(X_v, X^u)$ denote the unique minimal degree of a rational curve in X connecting X_v and X^u. It follows from [Buch et al. 2020, Theorem 5] that this degree is uniquely determined by $\pi_{i,*}(\operatorname{dist}_X(X_v, X^u))=\operatorname{dist}_{Y_i}(Y_{i,v}, Y_i^u)$ for $i\in\{1,2\}$. Using that $v\cdot P\in\Gamma_d(X^u)$ holds if and only if $d\geq\operatorname{dist}_X(X_v, X^u)$, we deduce that $\Gamma_d(X^u)$ and $\pi_1^{-1}(\Gamma_d(Y_1^u))\cap\pi_2^{-1}(\Gamma_d(Y_2^u))$ contain the same T-fixed points. The lemma follows from this, as both sets are B^--stable subvarieties of X. □

The following result implies that Conjecture 8.2(a) follows from the case where X has Picard rank 1. It was proved for $e=0$ in [Tarigradschi 2023, Theorem 3].

Theorem 8.6. *Let* $X=G/P$, $Y_1=G/Q_1$, *and* $Y_2=G/Q_2$ *be flag varieties such that* $P=Q_1\cap Q_2$. *Let* $w\in W^{\text{comin}}$, $u\in W$, *and let* $e\in H_2(X,\mathbb{Z})$ *be any effective degree. If* $\Gamma_{d(w,u)+e}(Y_{i,w_0w}, Y_i^u)=\Gamma_e(w^{-1}\cdot Y_i^{wu})$ *holds for* $i\in\{1,2\}$, *then* $\Gamma_{d(w,u)+e}(X_{w_0w}, X^u)=\Gamma_e(w^{-1}\cdot X^{wu})$.

Proof. The assumptions and Lemma 8.5 imply that

$$\begin{aligned}\Gamma_{d(w,u)+e}(X_{w_0w}, X^u)&\subset\pi_1^{-1}(\Gamma_{d(w,u)+e}(Y_{1,w_0w}, Y_1^u))\cap\pi_2^{-1}(\Gamma_{d(w,u)+e}(Y_{2,w_0w}, Y_2^u))\\&=\pi_1^{-1}(\Gamma_e(w^{-1}\cdot Y_1^{wu}))\cap\pi_2^{-1}(\Gamma_e(w^{-1}\cdot Y_2^{wu}))=\Gamma_e(w^{-1}\cdot X^{wu}),\end{aligned}$$

and the opposite inclusion holds by (8). □

Corollary 8.7. *Conjecture 8.2(a) is true when* $X=\mathrm{GL}(n)/P$ *has Lie type A.*

Proof. This follows from Theorems 8.4 and 8.6, noting that all flag varieties of type A with Picard rank 1 are Grassmannians, and therefore cominuscule. □

9. Horospherical varieties of Picard rank 1

In this section we interpret Theorem 4.3 and Proposition 5.3 for a class of horospherical varieties that includes all nonsingular projective horospherical varieties of Picard rank 1 (except flag varieties) by Pasquier's classification [2009]. Let G be a connected reductive linear algebraic group, $B \subset G$ a Borel subgroup, and $T \subset B$ a maximal torus. Let V_1 and V_2 be irreducible rational representations of G, and let $v_i \in V_i$ be a highest-weight vector of weight λ_i, for $i \in \{1, 2\}$. We assume that $\lambda_1 \neq \lambda_2$. Define

$$X = \overline{G \cdot [v_1 + v_2]} \subset \mathbb{P}(V_1 \oplus V_2).$$

If X is normal, then X is a horospherical variety of rank 1; see [Timashev 2011, Chapter 7]. We will assume that X is nonsingular and $\mathbb{K} = \mathbb{C}$, even though many claims hold more generally; this implies that X is fibered over a flag variety G/P_{12} with nonsingular horospherical fibers of Picard rank 1; see Remark 9.5. Any G-translate of a B-orbit closure in X will be called a *Schubert variety*. Our next result uses the action of $T \times \mathbb{G}_m$ on X defined by $(t, z) \cdot [u_1 + u_2] = t \cdot [u_1 + z u_2]$, for $u_i \in V_i$. We have $X^{T \times \mathbb{G}_m} = X^T$, and a Schubert variety is T-stable if and only if it is $T \times \mathbb{G}_m$-stable.

Theorem 9.1. *Any T-stable Schubert variety in X is $T \times \mathbb{G}_m$-convex and $T \times \mathbb{G}_m$-equivariantly rigid.*

Before proving Theorem 9.1, we sketch elementary proofs of some basic facts about X, which are also consequences of general results about spherical varieties; see [Timashev 2011; Perrin 2014; Pasquier 2009].

Given an element $[u_1 + u_2] \in \mathbb{P}(V_1 \oplus V_2)$, we will always assume $u_i \in V_i$, and i will always mean an element from $\{1, 2\}$. We consider $\mathbb{P}(V_i)$ as a subvariety of $\mathbb{P}(V_1 \oplus V_2)$. Let $\pi_i : \mathbb{P}(V_1 \oplus V_2) \smallsetminus \mathbb{P}(V_{3-i}) \to \mathbb{P}(V_i)$ denote the projection from V_{3-i}, defined by $\pi_i([u_1 + u_2]) = [u_i]$. Set $X_0 = G \cdot [v_1 + v_2] \subset \mathbb{P}(V_1 \oplus V_2)$, $X_i = G \cdot [v_i] \subset \mathbb{P}(V_i)$, and $X_{12} = G \cdot ([v_1], [v_2]) \subset \mathbb{P}(V_1) \times \mathbb{P}(V_2)$. Since v_i is a highest-weight vector, the stabilizer $P_i = G_{[v_i]}$ is a parabolic subgroup containing B. It follows that $X_i \cong G/P_i$ and $X_{12} \cong G/(P_1 \cap P_2)$ are flag varieties. In particular, X_i is closed in $\mathbb{P}(V_i)$, and X_{12} is closed in $\mathbb{P}(V_1) \times \mathbb{P}(V_2)$. Notice also that $X_0 \cong G/H$, where $H \subset P_1 \cap P_2$ is the kernel of the character $\lambda_1 - \lambda_2 : P_1 \cap P_2 \to \mathbb{G}_m$. This shows that X_0 is a $\mathbb{G}_m$-bundle over X_{12}, so X is a nonsingular projective horospherical variety of rank 1 (but not necessarily of Picard rank 1; see Remark 9.5).

Let W be the Weyl group of G, and recall the notation from Section 6.

Lemma 9.2. *We have $X = X_0 \cup X_1 \cup X_2$. The B-orbit closures in X are*

$$\overline{Bw \cdot [v_i]} = \bigcup_{w' \leq w} Bw' \cdot [v_i] \quad \text{for } w \in W^{P_i} \text{ and } i \in \{1, 2\},$$

$$\overline{Bw \cdot [v_1 + v_2]} = \bigcup_{w' \leq w} (Bw' \cdot [v_1 + v_2] \cup Bw' \cdot [v_1] \cup Bw' \cdot [v_2]) \quad \text{for } w \in W^{P_1 \cap P_2}.$$

Proof. Set $\mathbb{P}_0 = \mathbb{P}(V_1 \oplus V_2) \smallsetminus (\mathbb{P}(V_1) \cup \mathbb{P}(V_2))$. Since $\lambda_1 \neq \lambda_2$, it follows that $\overline{T \cdot [v_1 + v_2]}$ is the line through $[v_1]$ and $[v_2]$ in $\mathbb{P}(V_1 \oplus V_2)$. This implies $X_0 = (\pi_1 \times \pi_2)^{-1}(X_{12})$; hence X_0 is closed in $\mathbb{P}_0$ and $X_0 = X \cap \mathbb{P}_0$. We also have $X_i \subset X \cap \mathbb{P}(V_i) \subset \pi_i^{-1}(X_i) \cap \mathbb{P}(V_i) = X_i$, which proves the first claim. To finish the proof, it suffices to show $w' \cdot [v_i] \in \overline{Bw \cdot [v_1 + v_2]}$ if and only if $w' \leq w$ (when $w' \in W^{P_i}$). The implication 'if' holds because $w' \cdot [v_i] \in \overline{Tw' \cdot [v_1 + v_2]}$, and 'only if' holds because $\pi_i(\overline{Bw \cdot [v_1 + v_2]} \smallsetminus X_{3-i}) \subset \overline{Bw \cdot [v_i]}$. □

Define an alternative action of P_i on V_{3-i} by $p \bullet u = \lambda_i(p)^{-1} p \cdot u$, and use this action to form the space

$$G \times^{P_i} V_{3-i} = \{[g, u] : g \in G,\ u \in V_{3-i}\}/\{[gp, u] = [g, p \bullet u] : p \in P_i\}.$$

Define a morphism of varieties $\phi_i : G \times^{P_i} V_{3-i} \to \mathbb{P}(V_1 \oplus V_2)$ by

$$\phi_i([g, u]) = g \cdot [v_i + u].$$

This is well defined since $p \cdot (v_i + u) = \lambda_i(p)(v_i + p \bullet u)$ holds for $p \in P_i$ and $u \in V_{3-i}$. Set $E_i = (P_i \bullet v_{3-i}) \cup \{0\} \subset V_{3-i}$. Noting that E_i is the cone over $P_i \cdot [v_{3-i}] \cong P_i/(P_1 \cap P_2)$, it follows that E_i is closed in V_{3-i}.

Lemma 9.3. *The restricted map $\phi_i : G \times^{P_i} E_i \to X_0 \cup X_i$ is an isomorphism of varieties. In particular, $E_i \subset V_{3-i}$ is a linear subspace.*

Proof. Assume $\phi_i([g, u]) = \phi_i([g', u'])$, and set $p = g^{-1}g'$. We obtain $p \in P_i$ and $[v_i + u] = p \cdot [v_i + u'] = [v_i + p \bullet u']$ in $\mathbb{P}(V_1 \oplus V_2)$; hence

$$[g, u] = [g, p \bullet u'] = [gp, u'] = [g', u']$$

in $G \times^{P_i} V_{3-i}$. We deduce that $\phi_i : G \times^{P_i} E_i \to X_0 \cup X_i$ is bijective, so the lemma follows from Zariski's main theorem, using that $X_0 \cup X_i$ is nonsingular. □

Fix a strongly dominant cocharacter $\rho : \mathbb{G}_m \to T$. For $a \in \mathbb{Z}$, define the map $\rho_a : \mathbb{G}_m \to T \times \mathbb{G}_m$ by $\rho_a(z) = (\rho(z), z^a)$. The resulting action of $\mathbb{G}_m$ on X is given by $\rho_a(z) \cdot [u_1 + u_2] = \rho(z) \cdot [u_1 + z^a u_2]$.

Lemma 9.4. *All T-fixed points in X are fully definite for the action of $T \times \mathbb{G}_m$.*

Proof. Lemma 9.3 shows that $[v_1]$ has a $T \times \mathbb{G}_m$-stable open neighborhood in X isomorphic to $B^- \cdot [v_1] \times E_1$, where the action is given by $(t, z) \cdot (x, u) = (t \cdot x, t \bullet zu)$. If a is sufficiently negative, then $\mathbb{G}_m$ acts through ρ_a on $T_{[v_1]}X = T_{[v_1]}X_1 \oplus E_1$ with strictly negative weights; hence $[v_1]$ is fully definite in X for the action of $T \times \mathbb{G}_m$. A symmetric argument shows that $[v_2]$ is fully definite. The result follows from this, since all T-fixed points in X are obtained from $[v_1]$ or $[v_2]$ by the action of the Weyl group W. □

Proof of Theorem 9.1. For a sufficiently negative, it follows from Lemma 6.1 that the Białynicki-Birula cells of X defined by ρ_a are

$$X^+_{w\cdot[v_1]} = Bw\cdot[v_1] \quad\text{and}\quad X^+_{w\cdot[v_2]} = Bw\cdot[v_1+v_2]\cup Bw\cdot[v_2].$$

These cells form a stratification of X by Lemma 9.2, so Proposition 5.3 implies that $\overline{Bw\cdot[v_1]}$ and $\overline{Bw\cdot[v_1+v_2]}$ are $T\times\mathbb{G}_m$-convex for $w\in W$. A symmetric argument applies to $\overline{Bw\cdot[v_2]}$; hence all T-stable Schubert varieties in X are $T\times\mathbb{G}_m$-convex by Lemma 6.2. The result now follows from Theorem 4.3 and Lemma 9.4. □

Remark 9.5. The exact sequence of [Perrin 2014, Theorem 3.2.4] implies that $\mathrm{Pic}(X)$ is a free abelian group of rank equal to the rank of X (which is one) plus the number of B-stable prime divisors in X that do not contain a G-orbit. Any B-stable prime divisor meeting X_0 has the form $D=\overline{Bw_0s_\beta\cdot[v_1+v_2]}$, where β is a simple root, and Lemma 9.2 shows that D contains X_i if and only if β is a root of P_i. Let $P_{12}\subset G$ be the parabolic subgroup generated by P_1 and P_2. We obtain $\mathrm{Pic}(X)\cong\mathbb{Z}\oplus\mathrm{Pic}(G/P_{12})$. Let $\pi : X\to G/P_{12}$ be the map defined by $\pi(g\cdot[v_1+v_2])=\pi(g\cdot[v_i])=g\cdot P_{12}$. This is a G-equivariant morphism of varieties, as its restriction to $X_0\cup X_i$ is the composition of $\pi_i : X_0\cup X_i\to G/P_i$ with the projection $G/P_i\to G/P_{12}$. The fibers of π are translates of $\pi^{-1}(1\cdot P_{12}) = \overline{L\cdot[v_1+v_2]}\subset\mathbb{P}(V_1\oplus V_2)$, where L is the Levi subgroup of P_{12} containing T. Moreover, $\pi^{-1}(1\cdot P_{12})$ is a nonsingular projective horospherical variety of Picard rank 1, so it is either a flag variety or one of the nonhomogeneous spaces from Pasquier's classification [2009].

Question 9.6. Let X be any projective G-horospherical variety fibered over a flag variety G/P with nonsingular horospherical fibers of Picard rank 1. Is it true that X is isomorphic to an orbit closure $\overline{G\cdot[v_1+v_2]}\subset\mathbb{P}(V)$, where V is a rational representation of G, and $v_1, v_2\in V$ are highest-weight vectors?

Example 9.7. Let X be the blow-up of $\mathbb{P}^2$ at a point p, let $\pi : X\to\mathbb{P}^1$ be the morphism defined by projection from p, and set $G=\mathrm{SL}(2,\mathbb{C})$. Then X is G-horospherical and fibered over $\mathbb{P}^1$ with fiber $\mathbb{P}^1$. This variety X is isomorphic to $\overline{G\cdot[v_1+v_2]}\subset\mathbb{P}(V_1\oplus V_2)$, where v_1 is a highest-weight vector in $V_1=\mathbb{C}^2$, and v_2 is a highest-weight vector in $V_2=\mathrm{Sym}^2(\mathbb{C}^2)$.

Acknowledgements

We are particularly grateful to Mihail Tarigradschi, whose suggestions led to our proof of (2) in type A. We also thank Allen Knutson for suggesting the term *T-convex*, and we thank Leonardo Mihalcea for making us aware of [Gorbounov et al. 2025]. We finally thank an anonymous referee for a very careful reading of our manuscript and for several insightful suggestions.

References

[Anderson and Fulton 2024] D. Anderson and W. Fulton, *Equivariant cohomology in algebraic geometry*, Cambridge Studies in Advanced Mathematics **210**, Cambridge Univ. Press, 2024. MR

[Belkale 2004] P. Belkale, "Transformation formulas in quantum cohomology", *Compos. Math.* **140**:3 (2004), 778–792. MR

[Benedetti and Perrin 2022] V. Benedetti and N. Perrin, "Cohomology of hyperplane sections of (co)adjoint varieties", preprint, 2022. arXiv 2207.02089

[Benedetti et al. 2024] V. Benedetti, N. Perrin, and W. Xu, "Quantum K-theory of $\mathrm{IG}(2, 2n)$", *Int. Math. Res. Not.* **2024**:22 (2024), 14061–14093. MR

[Białynicki-Birula 1973] A. Białynicki-Birula, "Some theorems on actions of algebraic groups", *Ann. of Math.* (2) **98** (1973), 480–497. MR

[Białynicki-Birula 1976] A. Białynicki-Birula, "Some properties of the decompositions of algebraic varieties determined by actions of a torus", *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **24**:9 (1976), 667–674. MR

[Björner and Brenti 2005] A. Björner and F. Brenti, *Combinatorics of Coxeter groups*, Graduate Texts in Mathematics **231**, Springer, 2005. MR

[Bourbaki 1981] N. Bourbaki, *Éléments de mathématique*: *groupes et algèbres de Lie. Chapitres* 4, 5 *et* 6, Masson, Paris, 1981. Reprinted by Springer, Berlin, 2007. MR

[Brion 1997] M. Brion, "Equivariant Chow groups for torus actions", *Transform. Groups* **2**:3 (1997), 225–267. MR

[Brion and Kumar 2005] M. Brion and S. Kumar, *Frobenius splitting methods in geometry and representation theory*, Progress in Mathematics **231**, Birkhäuser, Boston, MA, 2005. MR

[Buch and Mihalcea 2015] A. S. Buch and L. C. Mihalcea, "Curve neighborhoods of Schubert varieties", *J. Differential Geom.* **99**:2 (2015), 255–283. MR

[Buch et al. 2013] A. S. Buch, P.-E. Chaput, L. C. Mihalcea, and N. Perrin, "Finiteness of cominuscule quantum K-theory", *Ann. Sci. Éc. Norm. Supér.* (4) **46**:3 (2013), 477–494. MR

[Buch et al. 2018a] A. S. Buch, P.-E. Chaput, L. C. Mihalcea, and N. Perrin, "A Chevalley formula for the equivariant quantum K-theory of cominuscule varieties", *Algebr. Geom.* **5**:5 (2018), 568–595. MR

[Buch et al. 2018b] A. S. Buch, P.-E. Chaput, L. C. Mihalcea, and N. Perrin, "Projected Gromov–Witten varieties in cominuscule spaces", *Proc. Amer. Math. Soc.* **146**:9 (2018), 3647–3660. MR

[Buch et al. 2020] A. S. Buch, S. Chung, C. Li, and L. C. Mihalcea, "Euler characteristics in the quantum K-theory of flag varieties", *Selecta Math.* (*N.S.*) **26**:2 (2020), art. id. 29. MR

[Buch et al. 2022] A. S. Buch, P.-E. Chaput, L. C. Mihalcea, and N. Perrin, "Positivity of minuscule quantum K-theory", preprint, 2022. arXiv 2205.08630

[Buch et al. 2023] A. S. Buch, P.-E. Chaput, and N. Perrin, "Seidel and Pieri products in cominuscule quantum K-theory", preprint, 2023. arXiv 2308.05307

[Chaput and Perrin 2023] P.-E. Chaput and N. Perrin, "Affine symmetries in quantum cohomology: corrections and new results", *Math. Res. Lett.* **30**:2 (2023), 341–374. MR

[Chaput et al. 2009] P.-E. Chaput, L. Manivel, and N. Perrin, "Affine symmetries of the equivariant quantum cohomology ring of rational homogeneous spaces", *Math. Res. Lett.* **16**:1 (2009), 7–21. MR

[Coskun 2011] I. Coskun, "Rigid and non-smoothable Schubert classes", *J. Differential Geom.* **87**:3 (2011), 493–514. MR

[Coskun 2014] I. Coskun, "Rigidity of Schubert classes in orthogonal Grassmannians", *Israel J. Math.* **200**:1 (2014), 85–126. MR

[Coskun 2018] I. Coskun, "Restriction varieties and the rigidity problem", pp. 49–95 in *Schubert varieties, equivariant cohomology and characteristic classes — IMPANGA 15*, edited by J. Buczyński et al., Eur. Math. Soc., Zürich, 2018. MR

[Coskun and Robles 2013] I. Coskun and C. Robles, "Flexibility of Schubert classes", *Differential Geom. Appl.* **31**:6 (2013), 759–774. MR

[Deodhar 1977] V. V. Deodhar, "Some characterizations of Bruhat ordering on a Coxeter group and determination of the relative Möbius function", *Invent. Math.* **39**:2 (1977), 187–198. MR

[Edidin and Graham 1998] D. Edidin and W. Graham, "Equivariant intersection theory", *Invent. Math.* **131**:3 (1998), 595–634. MR

[Fulton 1993] W. Fulton, *Introduction to toric varieties*, Annals of Mathematics Studies **131**, Princeton Univ. Press, 1993. MR

[Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Math. (3) **2**, Springer, 1998. MR

[Fulton and Pandharipande 1997] W. Fulton and R. Pandharipande, "Notes on stable maps and quantum cohomology", pp. 45–96 in *Algebraic geometry*, *II* (Santa Cruz, CA, 1995), edited by J. Kollár et al., Proc. Sympos. Pure Math. **62**, Amer. Math. Soc., Providence, RI, 1997. MR

[Fulton and Woodward 2004] W. Fulton and C. Woodward, "On the quantum product of Schubert classes", *J. Algebraic Geom.* **13**:4 (2004), 641–661. MR

[Givental 2000] A. Givental, "On the WDVV equation in quantum K-theory", *Michigan Math. J.* **48** (2000), 295–304. MR

[Gonzales and Lowen $\geq$ 2025] T. Gonzales and C. Lowen, "Counterexamples to $\mathbb{G}_m$-convexity for toric varieties", In preparation.

[Gorbounov et al. 2025] V. Gorbounov, C. Korff, and L. C. Mihalcea, "Quantum K-theory of Grassmannians from a Yang–Baxter algebra", preprint, 2025. arXiv 2503.08602

[Hong 2005] J. Hong, "Rigidity of singular Schubert varieties in $\mathrm{Gr}(m, n)$", *J. Differential Geom.* **71**:1 (2005), 1–22. MR

[Hong 2007] J. Hong, "Rigidity of smooth Schubert varieties in Hermitian symmetric spaces", *Trans. Amer. Math. Soc.* **359**:5 (2007), 2361–2381. MR

[Hong and Mok 2020] J. Hong and N. Mok, "Schur rigidity of Schubert varieties in rational homogeneous manifolds of Picard number one", *Selecta Math.* (*N.S.*) **26**:3 (2020), art. id. 41. MR

[Kambayashi 1966] T. Kambayashi, "Projective representation of algebraic linear groups of transformations", *Amer. J. Math.* **88** (1966), 199–205. MR

[Knutson et al. 2013] A. Knutson, T. Lam, and D. E. Speyer, "Positroid varieties: juggling and geometry", *Compos. Math.* **149**:10 (2013), 1710–1752. MR

[Lee 2004] Y.-P. Lee, "Quantum K-theory, I: Foundations", *Duke Math. J.* **121**:3 (2004), 389–424. MR

[Li et al. 2025] C. Li, Z. Liu, J. Song, and M. Yang, "On Seidel representation in quantum K-theory of Grassmannians", *Sci. China Math.* **68**:7 (2025), 1523–1548. MR

[Liu et al. 2024] Y. Liu, A. Sheshmani, and S.-T. Yau, "Multi-rigidity of Schubert classes in partial flag varieties", preprint, 2024. arXiv 2410.21726

[McGovern 2002] W. M. McGovern, "The adjoint representation and the adjoint action", pp. 159–238 in *Algebraic quotients*, *torus actions and cohomology*, *the adjoint representation and the adjoint action*, Encyclopaedia Math. Sci. **131**, Springer, 2002. MR

[Mumford 1965] D. Mumford, *Geometric invariant theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, (N.F.) **34**, Springer, 1965. MR

[Pasquier 2009] B. Pasquier, "On some smooth projective two-orbit varieties with Picard number 1", *Math. Ann.* **344**:4 (2009), 963–987. MR

[Perrin 2014] N. Perrin, "On the geometry of spherical varieties", *Transform. Groups* **19**:1 (2014), 171–223. MR

[Robles and The 2012] C. Robles and D. The, "Rigid Schubert varieties in compact Hermitian symmetric spaces", *Selecta Math.* (*N.S.*) **18**:3 (2012), 717–777. MR

[Seidel 1997] P. Seidel, "π_1 of symplectic automorphism groups and invertibles in quantum homology rings", *Geom. Funct. Anal.* **7**:6 (1997), 1046–1095. MR

[Springer 1998] T. A. Springer, *Linear algebraic groups*, 2nd ed., Progress in Mathematics **9**, Birkhäuser, Boston, MA, 1998. MR

[Sumihiro 1974] H. Sumihiro, "Equivariant completion", *J. Math. Kyoto Univ.* **14** (1974), 1–28. MR

[Tarigradschi 2023] M. Tarigradschi, "Curve neighborhoods of Seidel products in quantum cohomology", preprint, 2023. arXiv 2309.05985

[Timashev 2011] D. A. Timashev, *Homogeneous spaces and equivariant embeddings*, Encyclopaedia of Mathematical Sciences **138**, Springer, 2011. MR

[Xu 2024] W. Xu, "Quantum K-theory of incidence varieties", *Eur. J. Math.* **10**:2 (2024), art. id. 22. MR

Received October 10, 2024. Revised April 29, 2025.

ANDERS S. BUCH
DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
PISCATAWAY, NJ
UNITED STATES
asbuch@math.rutgers.edu

PIERRE-EMMANUEL CHAPUT
DOMAINE SCIENTIFIQUE VICTOR GRIGNARD
UNIVERSITÉ DE LORRAINE
NANCY
FRANCE
pierre-emmanuel.chaput@univ-lorraine.fr

NICOLAS PERRIN
CENTRE DE MATHÉMATIQUES LAURENT SCHWARTZ
ECOLE POLYTECHNIQUE
PALAISEAU
FRANCE
nicolas.perrin.cmls@polytechnique.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 338, No. 2, 2025
https://doi.org/10.2140/pjm.2025.338.231

GENUS THREE GOERITZ GROUPS OF CONNECTED SUMS OF TWO LENS SPACES

HAO CHEN AND YANQING ZOU

We prove that the mapping class groups of the genus three Heegaard splittings of the connected sums of two lens spaces are finitely generated, and the corresponding reducing sphere complexes are all connected.

1. Introduction

It is well known that every closed orientable 3-manifold N admits a *Heegaard splitting* $V \cup_\Sigma W$, which is a decomposition of N into two handlebodies V and W of the same genus. Their common boundary surface Σ is called *Heegaard surface*, and the genus of Σ is called the genus of the Heegaard splitting. Two Heegaard splittings of N are said to be isotopic if their corresponding Heegaard surfaces are ambient isotopic. The *Goeritz group* $\mathcal{G}(N, \Sigma)$, first introduced by Goeritz [1933], is the group of isotopy classes of orientation-preserving diffeomorphisms of N that preserve these two handlebodies of the splitting setwise. As a subgroup of the mapping class group of Σ, there is an open question about it in [Gordon 2007].

Question. Is the Goeritz group finite or finitely generated?

By the works of Johnson [2010], Namazi [2007], as well as Zou and Qiu [2020], the Goeritz groups of almost all Heegaard splittings of distance at least 2 are finite. However, if $V \cup_\Sigma W$ is weakly reducible, or equivalently, of distance at most 1, $\mathcal{G}(N, \Sigma)$ is infinite as shown in Namazi's construction. We focus on studying the finite generation problem for the Goeritz group of reducible Heegaard splittings.

A Heegaard splitting $V \cup_\Sigma W$ is reducible if there is a 2-sphere $S \subset N$ that intersects Σ transversely in one essential simple closed curve. Such a 2-sphere S is called a reducing sphere for Σ. When $V \cup_\Sigma W$ is reducible, we can decompose it as a connected sum of two Heegaard splittings of smaller genus, denoted by $V \cup_\Sigma W = N_1 \sharp N_2$, where $N_1 = V_1 \cup_{\Sigma_1} W_1$ and $N_2 = V_2 \cup_{\Sigma_2} W_2$. A natural question arises: If both $\mathcal{G}(N_1, \Sigma_1)$ and $\mathcal{G}(N_2, \Sigma_2)$ are finitely generated, is $\mathcal{G}(N, \Sigma)$ also finitely generated?

MSC2020: 57K30, 57K31.
Keywords: reducing sphere complex, Goeritz group, Heegaard splitting.

If $g(\Sigma_1) + g(\Sigma_2) = 2$, Cho and Koda [2019] proved that $\mathcal{G}(N, \Sigma)$ is finitely presented. Here we study the case $g(\Sigma_1) + g(\Sigma_2) = 3$ and give an answer to the Question as follows.

Theorem 1.1. *If $N = V \cup_\Sigma W$ is a genus three Heegaard splitting*[1] *for a connected sum of two lens spaces ($\neq S^3$, $S^1 \times S^2$), then $\mathcal{G}(N, \Sigma)$ is finitely generated.*

Furthermore, we study the reducing sphere complex $\mathscr{R}$ for $V \cup_\Sigma W$, which is a subcomplex of the curve complex spanned by those curves that bound disks in both handlebodies. As a corollary, we have the following.

Corollary 1.2. *Under the same condition as in Theorem 1.1, $\mathscr{R}$ is connected.*

For any reducible Heegaard splitting $V \cup_\Sigma W = (V_1 \cup_{\Sigma_1} W_1) \sharp (V_2 \cup_{\Sigma_2} W_2)$, let $\mu = S \cap \Sigma$ be the intersection of a reducing sphere S and the Heegaard surface Σ. Although the method in the proof of Theorem 1.1 does not apply in general, it provides insight into the widely studied subgroup $G_\mu \leqslant \mathcal{G}(N, \Sigma)$, the stabilizer of μ, which is a key subgroup of $\mathcal{G}(N, \Sigma)$. By its definition, it is not hard to see that there is a natural homomorphism from G_μ to $\mathcal{G}(N_i, \Sigma_i)$ for each i. Thus, it is of interest to determine whether G_μ is finitely generated (or finitely presented) when both of those two Goeritz groups are finitely generated (or finitely presented). Using standard combinatorial techniques, we obtain the following result.

Theorem 1.3. *If $\mathcal{G}(N_1, \Sigma_1)$ and $\mathcal{G}(N_2, \Sigma_2)$ are both finitely generated* (*or finitely presented*), *then so is G_μ.*[2]

Overview of the proof. We first show that the Goeritz group under consideration can be generated by three stabilizers of reducing curves, as shown in Theorem 4.8. As a corollary, the corresponding reducing sphere complex is connected. Next, we carefully study the stabilizer of a reducing sphere and give a proof of Theorem 1.3. Finally, by the previous work [Cho and Koda 2019] on genus two reducible Heegaard splittings, we arrive at the finite generation of each stabilizer.

This paper is organized as follows. We introduce some notations in Section 2 and study two classes of automorphisms, *eyeglass twist* and *visional bubble move*, in Section 3. Next, we carefully study the properties of three stabilizers in Section 4. After all preparations have been done, we complete the proof of Theorem 1.1 and Theorem 1.3 in Section 5.

Notations. We respectively denote the isotopy class of a curve μ in a surface and of a diffeomorphism h of a manifold by $\bar{\mu}$ and $\bar{h}$. When μ is endowed with an orientation, we denote it by $\vec{\mu}$. For an oriented curve $\vec{\mu}$, we denote its isotopy by $[\vec{\mu}]$.

[1] By Haken's lemma, the Heegaard splitting is reducible.

[2] Here, N_i $(i = 1, 2)$ is not necessarily a lens space.

2. Preliminaries

Throughout the paper, we respectively denote the isotopy class of a curve μ and of a diffeomorphism f by $\bar{\mu}$ and $\bar{f}$. From now on, we assume that N is the connected sum of two lens spaces unless otherwise specified, and $V \cup_\Sigma W$ is a genus three Heegaard splitting of N. Let $\mathrm{Diff}^+(N, \Sigma)$ be a subgroup of $\mathrm{Diff}^+(N)$ defined as

$$\mathrm{Diff}^+(N, \Sigma) \overset{\text{def}}{=} \{ f \in \mathrm{Diff}^+(N) : f(\Sigma) = \Sigma \text{ and } f \text{ preserves the orientation of } \Sigma \}.$$

It is clear that if an orientation-preserving diffeomorphism of N preserves the Heegaard splitting of N, it must preserve the orientation of Σ. Hence, the natural homomorphism $\rho_1 : \mathrm{Diff}^+(N, \Sigma) \to \mathcal{G}(N, \Sigma)$ is an epimorphism.

Definition 2.1. Two reducing spheres S_1, S_2 (for Σ) are isotopic if there is an isotopy

$$H_t : (N, \Sigma) \to (N, \Sigma), \quad 0 \le t \le 1,$$

such that $H_0 = \mathrm{id}$ and $H_1(S_1) = S_2$.

Definition 2.2. A triplet $\mathcal{T} = (S_1, S_2, S_3)$ of pairwise nonisotopic reducing spheres for Σ is called a *sphere triplet* (for Σ), and spheres S_i $(i = 1, 2, 3)$ are called the components of $\mathcal{T}$. We say the triplet is *complete* if the reducing spheres are pairwise disjoint (i.e., its three components span a 2-simplex in the corresponding reducing sphere complex).

Definition 2.3. Two sphere triplets $\mathcal{T}_1$, $\mathcal{T}_2$ are isotopic if there is an isotopy

$$H_t : (N, \Sigma) \to (N, \Sigma), \quad 0 \le t \le 1,$$

such that $H_0 = \mathrm{id}$ and $H_1(\mathcal{T}_1) = \mathcal{T}_2$.

Note 2.4. We usually make no notational distinction between triplets and their isotopy classes when the context is clear.

Definition 2.5. Two sphere triplets $\mathcal{T}_1$, $\mathcal{T}_2$ are *congruent* if they differ by a permutation. For instance, (S_1, S_2, S_3) is congruent with (S_3, S_1, S_2).

We designate a complete sphere triplet $\mathcal{T} = (S_1, S_2, S_3)$ for Σ, as depicted in Figure 1, such that (1) S_i $(i = 1, 2)$ cuts off a genus one Heegaard splitting of $M_i \setminus B^3$; (2) S_3 cuts off a genus one Heegaard splitting of a 3-ball. Clearly, S_1 and S_2 are two reducible 2-spheres and cobound $S^2 \times I$ in N. We also write $\mu_i = S_i \cap \Sigma$, for $i = 1, 2, 3$. Throughout the remainder of this paper, we fix the notations S_i and μ_i for these designated reducing spheres in Figure 1 and corresponding reducing curves respectively.

Lemma 2.6. *Up to congruences, $\mathcal{G}(N, \Sigma)$ acts transitively on the set of isotopy classes of complete sphere triplets for Σ.*

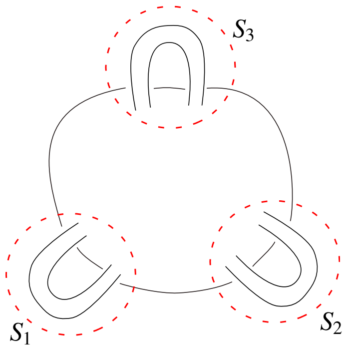

Figure 1. Heegaard surface Σ and triplet $\mathcal{T} = (S_1, S_2, S_3)$.

Proof. It suffices to show that for any given triplet $\mathcal{T}' = (S'_1, S'_2, S'_3)$, there exists a diffeomorphism $h \in \operatorname{Diff}^+(N, \Sigma)$ such that $h(\mathcal{T})$ is congruent to $\mathcal{T}'$. Firstly, each reducing sphere S'_i cuts off a genus one summand of Σ. Then, by uniqueness of the prime decomposition of 3-manifolds, one of the three reducing spheres bounds a 3-ball $\mathcal{B} \subset N$. By assumption, $N \setminus \operatorname{int}(\mathcal{B})$ does not admit a genus one Heegaard splitting. This implies that $\Sigma \cap \mathcal{B}$ is a torus with a open disk removed. Thus, the other two reducing spheres are isotopic in N and each cuts off a genus one Heegaard splitting of a once-punctured lens space. It follows that $\bigcup_{i=1}^{3} S'_i$ divides N into four parts, a 3-ball, a thrice-punctured 3-sphere, and two once-punctured lens spaces. Since $\bigcup_{i=1}^{3} S'_i$ divides N into four parts of the same diffeomorphism type as those divided by $\bigcup_{i=1}^{3} S_i$, we glue all diffeomorphisms of these four parts along spheres to obtain the desired h. □

Using similar arguments, we can prove the following lemma.

Lemma 2.7. *If S is a common component of these two complete sphere triplets $\mathcal{T}$ and $\mathcal{T}'$, then there exists a diffeomorphism $h \in \operatorname{Diff}^+(N, \Sigma)$ such that $h(\mathcal{T}) = \mathcal{T}'$ and $h(S) = S$.*

3. Eyeglass twist and visional bubble move

A Heegaard splitting $N = A \cup_\Sigma B$ is weakly reducible if there are two properly embedded disjoint essential disks, $a \subset A$ and $b \subset B$. We call (a, b) a weakly reducing pair for Σ. An eyeglass is a triple (a, b, λ), where (a, b) is a weakly reducing pair for Σ and $\lambda \subset \Sigma$ is an arc connecting a and b with its interior disjoint from them. For an eyeglass $\eta = (a, b, \lambda)$, we refer to (a, b) as the lenses of η and λ as the bridge of η. Given a normal direction $\vec{n}$ pointing toward the interior of B, we can push the 1-handle $a \times I$ around the circumference of the disk b in a counterclockwise direction as in Figure 2 (left). In fact, it is exactly an excursion

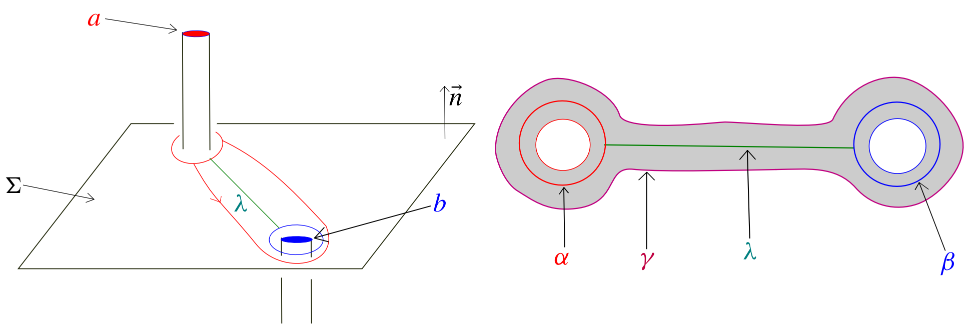

Figure 2. Left: eyeglass twist. Right: regular neighborhood Δ.

of the handlebody A that ends at the initial position. More formally, an eyeglass η defines a natural automorphism $T_\eta : (N, \Sigma) \to (N, \Sigma)$, known as the *positive eyeglass twist*. The inverse of this operation, which involves a clockwise excursion of A, is called *negative eyeglass twist* and denoted by $T_{\bar{\eta}}$. For an eyeglass twist T_η, the eyeglass η is referred to as its base eyeglass. It is not hard to see that an eyeglass twist preserves the isotopy classes of its lenses.

Note 3.1. The above definition does not depend on the order of the two lenses of the eyeglass. In other words, if $\eta = (a, b, \lambda)$ and $\eta' = (b, a, \lambda)$, then we have $T_\eta = T_{\eta'}$.

Let $\eta = (a, b, \lambda)$ be an eyeglass, where $\alpha \overset{\text{def}}{=} \partial a$, $\beta \overset{\text{def}}{=} \partial b$, and let Δ be a regular neighborhood of $\alpha \cup \lambda \cup \beta$ in surface Σ. Denote by γ the component (as in Figure 2 (right)) of $\partial\Delta$, which is isotopic to neither α nor β. We can now describe the above situation as

$$T_\eta = \tau_\alpha \cdot \tau_\beta \cdot \tau_\gamma^{-1},$$

where $\tau_{[\,\cdot\,]}$ denotes the *left-handed Dehn twist*. See more details in [Zupan 2020, Lemma 2.5].

Remark. Although different choices of regular neighborhoods of the eyeglass η yield different eyeglass twists, they are all equivalent up to isotopy. Therefore, for the eyeglass η, we obtain two eyeglass twists $T_\eta, T_{\bar{\eta}} \in \mathcal{G}(N, \Sigma)$.

Definition 3.2. Suppose η_1 and η_2 are two eyeglasses in N. They are isotopic if there is an isotopy $H_t : (N, \Sigma) \to (N, \Sigma)$, $0 \le t \le 1$, such that $H_0 = \mathrm{id}$ and $H_1(\eta_1) = \eta_2$. Furthermore, the isotopy class of an eyeglass η is denoted by $[\eta]$.

It is not hard to see that the eyeglass twist T_η depends only on the isotopy class of η. In analogy to the case for Dehn twists, we have the following lemma.

Lemma 3.3. *Given any* $\varphi \in \mathcal{G}(N, \Sigma)$, *we have* $T_{\varphi(\eta)} = \varphi \cdot T_\eta \cdot \varphi^{-1}$.

When a lens of an eyeglass η is decomposed into two disks, the corresponding eyeglass twist T_η can be expressed as the composition of two new eyeglass twists. We write it as the following lemma.

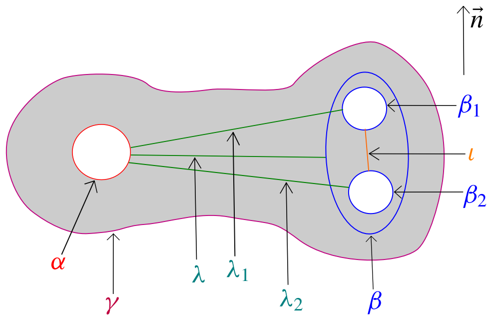

Figure 3. Composition of eyeglass twists.

Lemma 3.4 [Freedman and Scharlemann 2018, Figure 8]. *Let $\eta = (a, b, \lambda)$ be an eyeglass in N. If b is the band sum of two disks b_1 and b_2 along arc ι such that $\beta_i = \partial b_i$ is disjoint from both the bridge λ and the disk a, then for $i = 1, 2$, we can choose a proper arc λ_i which connects α and β_i, in the planar surface bounded by α, β_1, β_2 and γ. Given such a choice, we can obtain two new eyeglasses, η_1 $(= (a, b_1, \lambda_1))$ and η_2 $(= (a, b_2, \lambda_2))$. Moreover, there exists a suitable choice $\{\lambda_i\}$ such that T_η is a composition of T_{η_1} and T_{η_2}.*

Proof. Let $P \subset \Sigma$ be the pair of pants bounded by β_1, β_2, and β, and let $p = \lambda \cap \beta$. Let $\ell \subset P$ be an embedded arc that connects p and another point in β. Clearly, ℓ divides P into two annuli A_1 and A_2 that contain β_1 and β_2, respectively. Next, we choose for each $i \in \{1, 2\}$ an embedded arc $\lambda_i' \subset A_i$ that connects p and β_i. Finally, let $\lambda_i = \lambda \cup \lambda_i'$. We can verify that $\{\lambda_i\}$ is a desired choice. We also provide a specific choice in Figure 3, in which $T_\eta = T_{\eta_1} \cdot T_{\eta_2}$. □

Definition 3.5. Suppose η is an eyeglass in N, and S a reducing sphere for Σ. We say S separates η if these two lenses of η are disjoint from S and lie in different components of $N \setminus S$.

Definition 3.6. Suppose S is a reducing sphere for Σ, $\mu_s = S \cap \Sigma$, and $\eta \subset N$ (with $\partial\eta = \alpha \cup \beta \cup \lambda$)) an eyeglass with lenses disjoint from S. The *geometric intersection number* between S and η is defined as $I(\eta, S) = \tilde{I}(\lambda, \mu_s)$, where $\tilde{I}(\cdot, \cdot)$ is the geometric intersection number up to isotopies (of Σ) that leave α and β invariant.

Definition 3.7. For any separating reducing sphere S for Σ, we associate it with a subgroup $\mathcal{E}_k(S)$ of $\mathcal{G}(N, \Sigma)$ for each $k \in \mathbb{N}_+$ defined as

$$\mathcal{E}_k(S) = \langle E_k(S) \rangle,$$

where

$$E_k(S) = \{T_\eta \in \mathcal{G}(N, \Sigma) : S \text{ separates } \eta \text{ and } I(\eta, S) \leq k\}.$$

By definition, we have $E_k(S) \subset E_{k+1}(S)$. Then we have the ascending sequence $\mathcal{E}_1(S) \leq \mathcal{E}_2(S) \leq \mathcal{E}_3(S) \leq \cdots$.

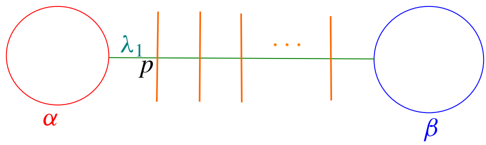

Figure 4. Intersection pattern.

Let us return to the setting from the previous section, in which $N = V \cup_\Sigma W$ is a genus three Heegaard splitting for a connected sum of two lens spaces and each S_i ($i = 1, 2, 3$) is a fixed separating reducing sphere for Σ, as in the previous section. Then we have the following lemma.

Lemma 3.8. *For $i = 1, 2, 3$ and $k \in \mathbb{N}_+$, we have $\mathcal{E}_{k+1}(S_i) \leqslant \mathcal{E}_k(S_i)$.*

Proof. It is sufficient to prove that $E_{k+1}(S_i) \subset \mathcal{E}_k(S_i)$. Consider an eyeglass twist T_η (suppose $\eta = (a, b, \lambda)$ and $\partial\eta = \alpha \cup \beta \cup \lambda$) representing an element of $E_{k+1}(S_i)$. We aim to show that $T_\eta \in E_k(S_i)$. Without loss of generality, we assume that β lies in the genus one component of $\Sigma \setminus S_i$, and the bridge λ intersects $\mu_i (= S_i \cap \Sigma)$ at $k+1$ points. Let one of these points, say p, be closest to α, as depicted in Figure 4. The point p divides λ into two segments, λ_1 and λ_2, where λ_1 denotes the one that is disjoint from μ_i.

Next, we choose an arc λ_3 in the pair of pants $\Sigma \setminus (S_i \cup \beta)$ that connects the point p and β, and whose interior does not intersect β (see Figure 5 (left)). Then we obtain a new eyeglass $\eta' = (a, b, \lambda_1 \cup \lambda_3)$, with the corresponding two eyeglass twists φ_1 and φ_2 ($\varphi_1 = T_{\eta'}$, $\varphi_2 = T_{\bar{\eta}'}$). Let $\gamma \subset \Sigma$ be a curve such that α, β, and γ cobound a pair of pants $\Delta \subset \Sigma$ containing η', as shown in Figure 6. Notice that $I(\eta', S_i) = 1$, which implies that $\varphi_1, \varphi_2 \in \mathcal{E}_k(S_i)$.

Pushing the 1-handle $a \times I$ along the path ι, as illustrated by the green line in Figure 5 (left), produces an eyeglass twist. This twist is exactly φ_2 $(= \tau_\alpha^{-1} \cdot \tau_\beta^{-1} \cdot \tau_\gamma)$. After the excursion of the 1-handle $a \times I$ along path ι, the intersection points p and p' are eliminated, as shown in Figure 5 (right). To see this, we assume that φ_2 is supported in Δ. Thus, we only need to observe where the arcs $\lambda \cap \Delta$ are sent by φ_2. The precise picture is illustrated in Figure 5 (middle). Let $\ell \subset \lambda \cap \Delta$ be the component that is closest to α (i.e., $\{p, p'\} \subset \ell$). After having removed all bigons, we can see that $\varphi_2(\ell)$ is disjoint from S_i (See details in Figure 5 (right).) On the other hand, for any other component $\ell' \subset \lambda \cap \Delta$, $|\varphi_2(\ell') \cap S_i| = |\ell' \cap S_i|$. This implies that $I(\varphi_2(\eta), S_i) < k+1$. By Lemma 3.3, we then have $T_\eta = \varphi_2^{-1} \cdot T_{\varphi_2(\eta)} \cdot \varphi_2$. Since both $T_{\varphi_2(\eta)}$ and φ_2 belong to $\mathcal{E}_k(S_i)$, it follows that $T_\eta \in \mathcal{E}_k(S_i)$. □

In summary, we have the equation

$$\mathcal{E}_1(S_i) = \mathcal{E}_2(S_i) = \mathcal{E}_3(S_i) = \cdots$$

for $i = 1, 2, 3$.

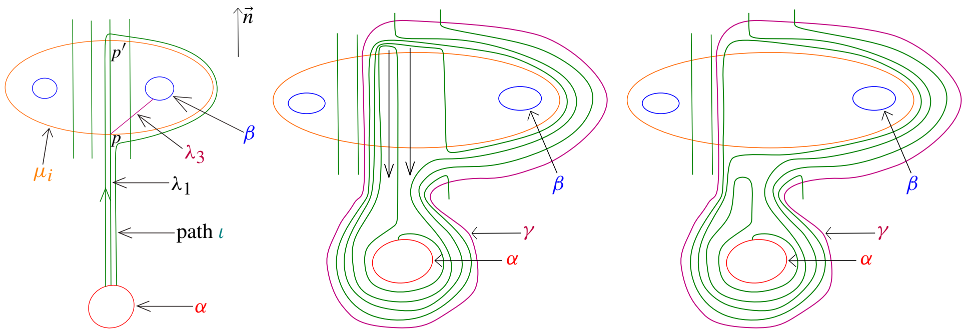

Figure 5. Left: new eyeglass. Middle: new intersection pattern. Right: new pattern with bigons removed.

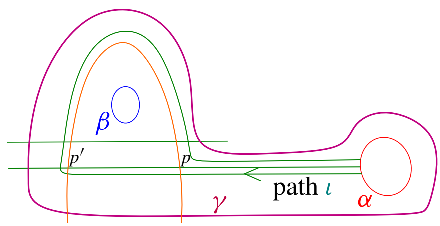

Figure 6. Specific neighborhood Δ of η.

Scharlemann [2022] defines a class of automorphisms of a Heegaard splitting, called bubble moves, which generalize one of the five automorphisms proposed by Powell [1980]. In this work, we extend the concept to a more general setting.

Definition 3.9. For a 3-manifold N with a Heegaard splitting $N = A \cup_\Sigma B$, a bubble is a 3-submanifold $\mathcal{B}$ of N, whose boundary is a 2-sphere. If the boundary $\partial\mathcal{B}$ is a reducing sphere for Σ, we call $\mathcal{B}$ a *bubble* for Σ. The genus of $\mathcal{B} \cap \Sigma$ is referred to as the *genus* of $\mathcal{B}$. In addition, a bubble is called *trivial* if it is a 3-ball, otherwise it is called *essential*.

Definition 3.10 [Scharlemann 2022, Section 2]. Let $N = A \cup_\Sigma B$ be a Heegaard splitting, and $\mathcal{B}$ a trivial bubble for Σ. A *bubble move* is an isotopy of $\mathcal{B}$ along a closed path in $\Sigma \setminus \operatorname{int}(\mathcal{B})$ that starts and ends at $\mathcal{B}$, returning $(\mathcal{B}, \mathcal{B} \cap \Sigma)$ to itself. See Figure 7 (left).

For essential bubbles, we introduce a new class of automorphisms defined as follows.

Definition 3.11 (visional bubble move). Let $N = A \cup_\Sigma B$ be a Heegaard splitting for a closed 3-manifold N, and $\mathcal{B}$ a bubble for Σ with S as its boundary. The submanifold $N \setminus \operatorname{int}(\mathcal{B})$ is also a bubble for Σ, which we refer to as the *dual bubble* of $\mathcal{B}$ and denote by $\mathcal{B}'$. By capping off the sphere boundary of $\mathcal{B}'$ with a 3-ball, we obtain a new manifold $N(\mathcal{B})$. The manifold $N(\mathcal{B})$ inherits a Heegaard splitting,

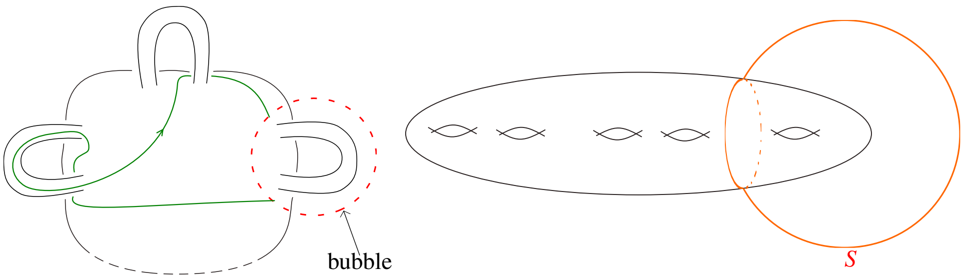

Figure 7. Left: bubble move. Right: Heegaard surface Σ'.

with its Heegaard surface Σ' being the boundary sum of $\Sigma \setminus \operatorname{int}(\mathcal{B})$ and a bordered torus (torus with an open disk removed), as illustrated in Figure 7 (right). Clearly, $N(\mathcal{B}) \setminus \operatorname{int}(\mathcal{B}')$ is a trivial bubble in $N(\mathcal{B})$, bounded by the sphere S, denoted by $\mathcal{B}^3$. As in the case of trivial bubbles, a bubble ($\mathcal{B}^3$) move induces a diffeomorphism $h : N(\mathcal{B}) \to N(\mathcal{B})$ such that $h|_{\mathcal{B}^3} = \mathrm{id}$. We then glue the two diffeomorphisms $h|_{\mathcal{B}'} : \mathcal{B}' \to \mathcal{B}'$ and $\mathrm{id} : \mathcal{B} \to \mathcal{B}$ along the sphere S to obtain a diffeomorphism $\tilde{h} : (N, \Sigma) \to (N, \Sigma)$, which we refer to as a *visional bubble* ($\mathcal{B}$) *move*.

In our setting, each sphere S_i ($i \in \{1, 2, 3\}$) bounds a genus one bubble, denoted by $\mathcal{B}_i$. It is easy to see that a visional $\mathcal{B}_i$ move fixes the sphere S_i. In other words, any visional $\mathcal{B}_i$ move lies in the stabilizer $\mathcal{H}_i \leqslant \mathcal{G}(N, \Sigma)$ of the isotopy class of the curve $\mu_i = S_i \cap \Sigma$. Furthermore, let $\mathcal{H}$ be the subgroup of $\mathcal{G}(N, \Sigma)$ generated by $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$; i.e., $\mathcal{H} = \langle \mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3 \rangle$.

4. Stabilizers of reducing 2-spheres

To prove the main theorem, we first show that $\mathcal{G}(N, \Sigma) = \mathcal{H}$ in Theorem 4.8. Then we will prove that $\mathcal{H}$ is finitely generated in the next section. As a corollary, the reducing sphere complex $\mathscr{R}$ is connected. Before proving Theorem 4.8, we introduce the following two lemmas.

Lemma 4.1. *$\mathcal{E}_1(S_i) < \mathcal{H}$ for $i = 1, 2, 3$.*

Proof. It is sufficient to prove that all generators of $\mathcal{E}_1(S_i)$ belong to the subgroup $\mathcal{H}$; i.e., $E_1(S_i) \subseteq \mathcal{H}$ for $i = 1, 2, 3$. Without loss of generality, we assume that $i = 1$. For any element $T_\eta \in E_1(S_1)$, $\eta = (a, b, \lambda)$ is an eyeglass such that S_1 separates η and the bridge λ intersects S_1 transversely at one point p_1. In this case, both $\alpha = \partial a$ and $\beta = \partial b$ are disjoint from $S_1 \cap \Sigma$. We assume the following conditions:

(1) a and b are, respectively, a disk in V and W.

(2) α and β lie in the genus two and the genus one component of $\Sigma \setminus S_1$ respectively.

We prove by induction on the geometric intersection number $I(\alpha, \mu_3)$ that $T_\eta \in \mathcal{H}$. The base case $\alpha \cap \mu_3 = \varnothing$ is divided into the following Case 1 and Case 2.

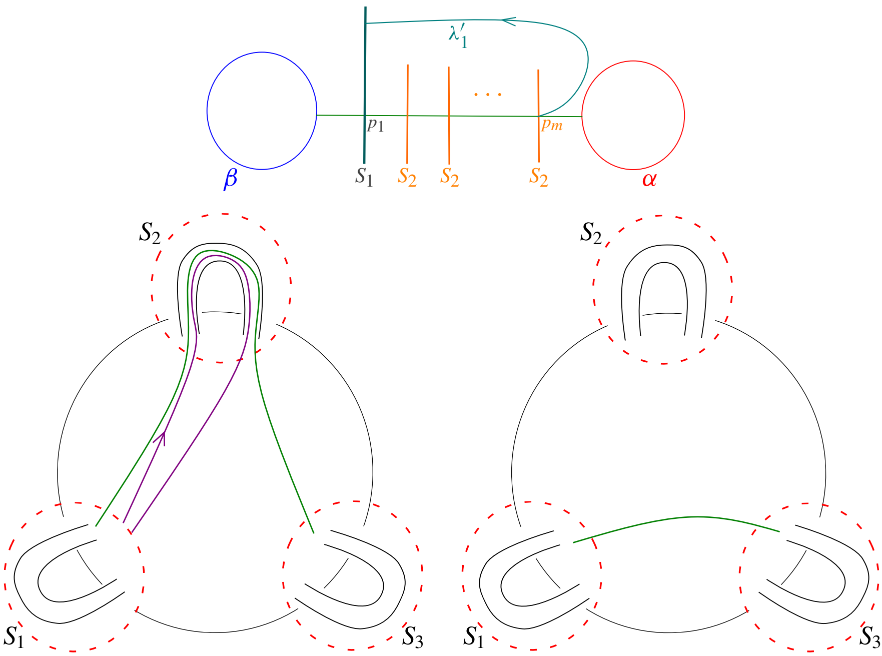

Figure 8. Top: finding a new arc λ_1'. Bottom left: a visional $\mathcal{B}_1$ move along the pink path. Bottom right: the reduction.

Case 1. $\alpha \cap \mu_3 = \varnothing$ and α lies in the genus one component of $\Sigma \setminus S_3$. Let $p_m \in \lambda \cap S_2$ be the point that is closest to α and $\lambda_1 \subset \lambda$ be the subarc bounded by p_0 and p_m, as shown in Figure 8 (top row). Then we choose an arc $\lambda_1' \subset \Sigma$ connecting p_m and a point of S_1, with its interior disjoint from S_1, S_2, and α. Let $\lambda' = \lambda_1 \cup \lambda_1'$. We endow λ' with an orientation as illustrated in Figure 8 (top row) and write $\vec{\lambda}'$ for the resulting oriented arc. Notice that the visional $\mathcal{B}_1$ move along $\vec{\lambda}'$, denoted by ψ_1, can reduce the intersection number $I(\eta, S_2)$. To be precise, $\psi_1 \in \mathcal{H}_1$ satisfies

$$I(\psi_1(\eta), S_2) = 0, \quad \psi_1(\alpha) = \alpha, \quad \psi_1(\beta) = \beta.$$

Figure 8 (bottom row) illustrates how a visional bubble move reduces the intersection.

Note that $\psi_1(\eta)$ is disjoint from S_2. It means that $T_{\psi_1(\eta)} \in \mathcal{H}_2$. Then we have

$$T_\eta = \psi_1^{-1} \cdot T_{\psi_1(\eta)} \cdot \psi_1 \in \mathcal{H}.$$

Case 2. $\alpha \cap \mu_3 = \varnothing$ and α lies in the genus two component of $\Sigma \setminus S_3$. Similarly, we can find a visional $\mathcal{B}_1$ move $\psi_2 \in \mathcal{H}_1$ such that

$$I(\psi_2(\eta), S_3) = 0, \quad \psi_2(\alpha) = \alpha, \quad \psi_2(\beta) = \beta.$$

It follows that $T_{\psi_2(\eta)} \in \mathcal{H}_3$ and $T_\eta = \psi_2^{-1} \cdot T_{\psi_2(\eta)} \cdot \psi_2 \in \mathcal{H}$.

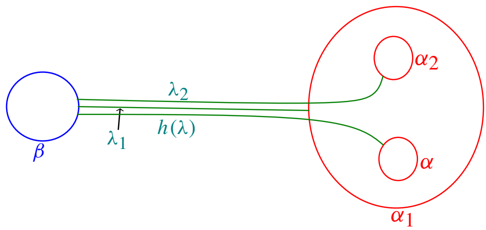

Figure 9. New eyeglasses.

Assume the statement is true for $I(\alpha, \mu_3) \leq 2n$. We consider the case that $I(\alpha, \mu_3) = 2n+2$. We first isotope η so that all the endpoints of the bridge λ lie in the genus two component of $\Sigma \setminus S_3$, while keeping the curves α and β invariant during the isotopy. Next, we apply a visional $\mathcal{B}_1$ move $\bar{h} \in \mathcal{H}_1$ such that $I(h(\alpha), \mu_3) \leqslant I(\alpha, \mu_3)$ and $h(\lambda)$ is disjoint from μ_3.

Let D and D' be, respectively, the disks bounded by $h(\alpha)$ and μ_3 in V. Without loss of generality, we assume that $|D \cap D'|$ is minimal. Let $D'' \subset D'$ be an outermost subdisk cut off by D. Doing a compression on D along D'' results in two essential disks D_1 and D_2, with boundaries α_1 and α_2. Since $h(\lambda) \cap D'' \subset h(\lambda) \cap \mu_3 = \emptyset$, $\alpha_1 \cup \alpha_2$ intersects $h(\lambda)$ at most once. Thus, we consider the following two subcases:

Case 3.1. $(\alpha_1 \cup \alpha_2) \cap h(\lambda) = \emptyset$. By Lemma 3.4, $T_{h(\eta)}$ is a composition of two eyeglass twists whose bases have a smaller intersection with μ_3. Since the intersection number is smaller, the induction hypothesis applies. Therefore, by the induction assumption, we have $T_{h(\eta)} \in \mathcal{H}$.

Case 3.2. $|(\alpha_1 \cup \alpha_2) \cap h(\lambda)| = 1$. Without loss of generality, we assume that α_1 intersects $h(\lambda)$ at one point. We now construct two new eyeglasses η_1 $(= (D_1, b, \lambda_1))$ and η_2 $(= (D_2, b, \lambda_2))$, as depicted in Figure 9. By Lemma 3.4, we know that $T_{\eta_1} = T_{\eta_2} \cdot T_{h(\eta)}$. From the induction assumption, it follows that $T_{h(\eta)} \in \mathcal{H}$. □

Lemma 4.2. *For any eyeglass $\eta = (a, b, \lambda)$ (with $\partial\eta = (\alpha, \beta, \lambda)$) satisfying that both α and β lie in the genus two component of $\Sigma \setminus S_i$ ($i = 1$ or 2), we have $T_\eta \in \mathcal{H}$.*

Proof. Without loss of generality, we assume that both α and β lie in the genus two component of $\Sigma \setminus S_1$. The other case is similar. So we omit it. If λ is disjoint from $\mu_1 = S_1 \cap \Sigma$, then T_η fixes μ_1. This means that $T_\eta \in \mathcal{H}$. So we assume that neither α nor β is isotopic to μ_1, and $\lambda \cap \mu_1 \neq \emptyset$. After performing three compressions on Σ along a, b, and the disk $D \subset V$ bounded by μ_1, we obtain a 2-sphere S.

Claim 4.3. *The sphere S contains a scar[3] of D.*

[3]A disk compression will produce two copies of the surgery disk in the resulting surface, which we refer to as scars of the surgery disk.

Proof. Suppose that the conclusion is false. In this case, one of α and β is separating in Σ. Suppose it is α. Then β lies in the genus one component of $\Sigma \setminus \alpha$. Since α is disjoint from μ_1, it means that λ is disjoint from μ_1, a contradiction. □

Hence, S contains a scar of D and the scars of a and b. Then we choose two disjoint simple closed curves $\ell_1, \ell_2 \subset S$ such that ℓ_1 cuts off a disk that contains only the scars of a, while ℓ_2 cuts off a disk containing only the scars of b. Since ℓ_1 (resp. ℓ_2) is a band sum of two scars of a (resp. b), ℓ_1 (resp. ℓ_2) bounds an essential disk in V (resp. W). Moreover, ℓ_1, ℓ_2, and the reducing curve μ_1 $(= \partial D)$ cobound a pair of pants in Σ. Then ℓ_1 is the band sum of μ_1 and ℓ_2. Hence ℓ_1 bounds a disk in W. So ℓ_1 is a reducing curve. Similarly, ℓ_2 is also a reducing curve. Then, there are two disjoint reducing spheres S_{ℓ_1} and S_{ℓ_2} such that $S_{\ell_i} \cap \Sigma = \ell_i$. By definition, these three spheres (S_{ℓ_1}, S_{ℓ_2} and S_1) constitute a complete sphere triplet, denoted by $\mathcal{T}'$. Note that S_1 is a common 2-sphere of $\mathcal{T}$ $(= (S_1, S_2, S_3))$ and $\mathcal{T}'$. By Lemma 2.7, there exists an element $\phi \in \mathcal{H}_i$ such that $\phi(\mathcal{T}') = \mathcal{T}$. To prove $T_\eta \in \mathcal{H}$, it suffices to prove $T_{\phi(\eta)} \in \mathcal{H}$.

If α is isotopic to ℓ_1, then $\phi(\alpha)$ is isotopic to one of μ_1, μ_2, and μ_3. So $T_{\phi(\eta)} \in \mathcal{H}$. Otherwise, S_{ℓ_1} separates η. Then $\phi(S_{\ell_1})$ separates $\phi(\eta)$. By Lemma 3.8, we know that $\mathcal{E}_{I(\phi(\eta),\phi(S_{\ell_1}))}(\phi(S_{\ell_1})) = \mathcal{E}_1(\phi(S_{\ell_1}))$. On the other hand, $\phi(\mathcal{T}') = \mathcal{T}$ implies that $\phi(S_{\ell_1})$ is exactly one component of $\mathcal{T}$. Further, by Lemma 4.1, it follows that $\mathcal{E}_1(\phi(S_{\ell_1})) < \mathcal{H}$. In summary, we have

$$T_{\phi(\eta)} \in E_{I(\phi(\eta),\phi(S_{\ell_1}))}(\phi(S_{\ell_1})) \subseteq \mathcal{E}_{I(\phi(\eta),\phi(S_{\ell_1}))}(\phi(S_{\ell_1})) = \mathcal{E}_1(\phi(S_{\ell_1})) < \mathcal{H}. \quad \square$$

Recently, the classical "Haken's lemma" has been strengthened into the "strong Haken's lemma" by several authors in various ways [Scharlemann 2024; Hensel and Schultens 2024; Taylor 2025], which says that a sphere set in a 3-manifold can be isotoped to be aligned with the given Heegaard surface (See the following definition for "aligned".)

Definition 4.4 [Freedman and Scharlemann 2024, Section 1]. A sphere set $E \subset N$ is a compact properly embedded surface in N such that each component of E is a sphere. Then, a Heegaard surface Σ and a sphere set E in N $(= A \cup_\Sigma B)$ are *aligned* if they are transverse, and each component of E intersects Σ in at most one circle. In addition, each disk component of $E \setminus \Sigma$ is essential in either A or B.

Definition 4.5 (bubble-sum). Suppose $R_1 \subset N$ is a reducing sphere for Σ and $\mathcal{B} \subset N$ a trivial bubble disjoint from R_1. We will use the notation $R_2 = \partial\mathcal{B}$, $\gamma_1 = R_1 \cap \Sigma$, and $\gamma_2 = R_2 \cap \Sigma$. Let $\lambda \subset \Sigma$ be an embedded arc that connects γ_1 and γ_2, with its interior disjoint from γ_1 and γ_2. Then for a good closed neighborhood $N(R_1 \cup R_2 \cup \lambda)$, $\partial N(R_1 \cup R_2 \cup \lambda)$ consists of copies of $R_1 \cup R_2$ and a new reducing sphere R_3. The reducing sphere R_3 is called a bubble-sum of R_1 and R_2 along λ.

Definition 4.6 [Freedman and Scharlemann 2024, Definition 1.4]. Let E_0 and E_1 be two sphere sets aligned with Σ in N. E_0 and E_1 are *equivalent* if there is an isotopy $H : N \times I \to N$ with $H_s : \Sigma \times \{s\} \to \Sigma$ for $0 \le s \le 1$ such that H_1 maps E_0 to E_1.

Freedman and Scharlemann [2024] prove that any two alignments of a sphere set are related by a sequence of *bubble-sums*, which are called "bubble moves" in their paper, and *eyeglass twists*.

Theorem 4.7 [Freedman and Scharlemann 2024, Theorem 1.6]. *If E_0 and E_1 are two sphere sets that are properly isotopic in N and each aligns with Σ, then up to equivalence, E_1 can be obtained from E_0 by a sequence of bubble-sums and eyeglass twists.*[4]

There is a bijection between the isotopy classes of reducing curves and the isotopy classes of reducing spheres. We identify a reducing sphere with its corresponding reducing curve. Therefore, the reducing sphere complex can be treated as a subcomplex of the curve complex of Σ. In subsequent arguments, we will use the symbol $\mathscr{R}$ to represent the reducing sphere complex for the Heegaard splitting $N = V \cup_\Sigma W$.

Theorem 4.8. $\mathcal{G}(N, \Sigma) = \mathcal{H}$.

Proof. We divide the proof into two cases: (a) $N_1 \neq N_2$; (b) $N_1 = N_2$.

Case (a). Let $\mathscr{R}^0$ be the 0-skeleton of $\mathscr{R}$. As both $\mathcal{G}(N, \Sigma)$ and $\mathcal{H}$ naturally act on $\mathscr{R}^0$, we denote the corresponding orbit containing the isotopy class $\bar{\mu}_i$ by $\mathscr{O}_i$ and $\mathscr{O}_i'$ respectively. Since we have assumed that $N_1 \neq N_2$, we know that $\mathscr{O}_1 \cap \mathscr{O}_2 = \varnothing$. To prove the theorem, it suffices[5] to show that $\mathscr{O}_1 = \mathscr{O}_1'$.

Given any reducing sphere S for Σ with the intersection curve μ $(= S \cap \Sigma)$ representing an isotopy class $\bar{\mu}$ of $\mathscr{O}_1$, we will prove that $\bar{\mu} \in \mathscr{O}_1'$. If it does, we immediately obtain the desired result $\mathscr{O}_1 = \mathscr{O}_1'$.

By an innermost curve argument, these two essential spheres S and S_1 are isotopic. Both S and S_1 are aligned with Σ. By Theorem 4.7, S is related to S_1 by a sequence of bubble-sums and eyeglass twists. Thus, there is a sequence of reducing curves λ_i in Σ such that λ_{i+1} can be obtained from λ_i by a bubble-sum or an eyeglass twist

$$\mu_1 = \Sigma \cap S_1 = \lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n = \Sigma \cap S = \mu. \tag{1}$$

We first prove by induction that $\bar{\lambda}_i \in \mathscr{O}_1' \cup \mathscr{O}_2'$. The base case $\bar{\lambda}_1 \in \mathscr{O}_1' \cup \mathscr{O}_2'$ clearly holds. We assume that $\bar{\lambda}_i \in \mathscr{O}_1' \cup \mathscr{O}_2'$. Then we prove that $\bar{\lambda}_{i+1} \in \mathscr{O}_1' \cup \mathscr{O}_2'$. Without loss of generality, we assume that $\bar{\lambda}_i \in \mathscr{O}_1'$. It means that there is an element $\bar{g} \in \mathcal{H}$ such that $\bar{g}(\bar{\lambda}_i) = \bar{\mu}_1$. Let $g \in \mathrm{Diff}^+(N, \Sigma)$ be a representative of $\bar{g}$ such that $g(\lambda_i) = \mu_1$. Then there are two cases as follows.

[4]Each involved eyeglass twist must have its base lenses disjoint from the sphere set on which it acts. See details in [Freedman and Scharlemann 2024, Section 1].

[5]Assume that $\mathscr{O}_1 = \mathscr{O}_1'$. Then, for any $\phi \in \mathcal{G}(N, \Sigma)$, we can find an element $\varphi \in \mathcal{H}$ such that $\phi(\bar{\mu}_1) = \varphi(\bar{\mu}_1)$. So $\varphi^{-1} \cdot \phi \in \mathcal{H}_1 \leqslant \mathcal{H}$. It follows that $\phi \in \mathcal{H}$.

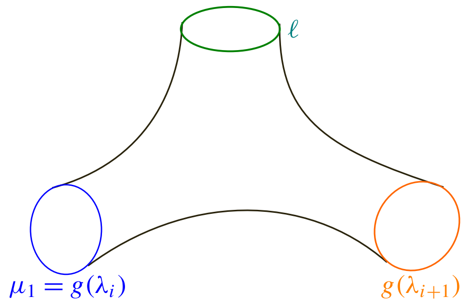

Figure 10. A pair of pants bounded by $g(\lambda_{i+1})$, ℓ, and μ_1.

Case 1. λ_{i+1} is obtained from λ_i by a bubble-sum. It implies that $g(\lambda_{i+1})$ can be obtained from $g(\lambda_i)$ by a bubble-sum. In this bubble-sum, the involved trivial bubble has sphere boundary S_ℓ, where $S_\ell \cap \Sigma = \ell$. By Definition 4.5, the three curves $g(\lambda_{i+1})$, ℓ, and μ_1 $(= g(\lambda_i))$ cobound a pair of pants illustrated as in Figure 10. Denote by $S_{g(\lambda_{i+1})}$ the reducing 2-sphere intersecting transversely with Σ in $g(\lambda_{i+1})$. Note that $S_1 \cup S_\ell \cup S_{g(\lambda_{i+1})}$ divides N into four submanifolds of the same diffeomorphism type as those divided by $S_1 \cup S_3 \cup S_2$. It follows that there is a diffeomorphism $h \in \mathrm{Diff}^+(N, \Sigma)$ such that $h(S_1, S_\ell, S_{g(\lambda_{i+1})}) = (S_1, S_3, S_2)$. Thus, we have $h \cdot g(\lambda_{i+1}) = \mu_2$. Furthermore, $\bar{h} \cdot \bar{g}(\bar{\lambda}_{i+1}) = \bar{\mu}_2$, where $\bar{g} \in \mathcal{H}$ and $\bar{h} \in \mathcal{H}_1$. So $\bar{\lambda}_{i+1} \in \mathcal{O}_2'$.

Case 2. λ_{i+1} is obtained from λ_i by an eyeglass twist $T_{\eta_i'}$. It means that $g(\lambda_{i+1})$ can be obtained from $g(\lambda_i)$ $(= \mu_1)$ by the eyeglass twist $T_{g(\eta_i')}$. Write $\eta_i = g(\eta_i')$. From the hypotheses of Theorem 4.7, we know that the lenses of η_i' are disjoint from λ_i. This implies the lenses of η_i are also disjoint from μ_1. Accordingly, there are two subcases as follows.

Subcase 2.1. S_1 separates η_i. By Lemma 3.8 and Lemma 4.1, we have $T_{\eta_i} \in \mathcal{H}$. Since $T_{\eta_i}^{-1} \cdot g(\lambda_{i+1}) = \mu_1$, we have that $\bar{\lambda}_{i+1} \in \mathcal{O}_1'$.

Subcase 2.2. S_1 does not separate η_i. Then these two lenses of η_i both lie in the component of $N \setminus S_1$ which contains the genus two component of $\Sigma \setminus S_1$. By Lemma 4.2, $T_{\eta_i} \in \mathcal{H}$. So we have $\bar{\lambda}_{i+1} \in \mathcal{O}_1'$.

The above argument completes the proof for the statement that $\bar{\lambda}_i \in \mathcal{O}_1' \cup \mathcal{O}_2'$ for all $i \leqslant n$. In particular, $\bar{\mu} = \bar{\lambda}_n \in \mathcal{O}_1' \cup \mathcal{O}_2'$. However, $\bar{\mu} \in \mathcal{O}_1$, $\mathcal{O}_1 \cap \mathcal{O}_2 = \varnothing$, and $\mathcal{O}_i' \subset \mathcal{O}_i$. This implies that $\bar{\mu} \in \mathcal{O}_1'$. This completes the proof of the case (a).

Case (b). Since any lens space has a unique genus one Heegaard splitting up to diffeomorphism, we can construct a diffeomorphism $f \in \mathrm{Diff}^+(N, \Sigma)$ such that $f(S_1, S_2, S_3) = (S_2, S_1, S_3)$, as in the proof of Lemma 2.6. This means that $\mathcal{O}_1' = \mathcal{O}_2'$. Then we can prove by induction on the sequence (1), as in Case (a), that $\bar{\lambda}_i \in \mathcal{O}_1'$. It follows that $\mathcal{O}_1' = \mathcal{O}_1$. Overall, $\mathcal{G}(N, \Sigma) = \mathcal{H}$. □

We use the above results to prove the connectedness of the reducing sphere complexes $\mathscr{R}$.

Proof of Corollary 1.2. These three reducing 2-spheres S_1, S_2 and S_3 are contained in a same component of $\mathscr{R}$, say $\mathscr{R}'$. To prove the connectedness of $\mathscr{R}$, it suffices to show that $\mathscr{R} = \mathscr{R}'$.

Given any reducing sphere $S \in \mathscr{R}$, there are two other reducing spheres S' and S'' in $\mathscr{R}$ such that the sphere triplet $\mathcal{T}' = (S, S', S'')$ is complete. By Lemma 2.6, there is an element $\nu \in \mathcal{G}(N, \Sigma)$ such that $\nu(\mathcal{T})$ is congruent to $\mathcal{T}'$. By Theorem 4.8,

$$\nu = \theta_n \cdot \theta_{n-1} \cdots \theta_2 \cdot \theta_1,$$

where $\theta_i \in \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3$.

Let $\nu_i = \theta_i \cdot \theta_{i-1} \cdots \theta_2 \cdot \theta_1$ $(0 \leqslant i \leqslant n)$ and $\mathcal{T}_i = \nu_i(\mathcal{T})$. Then we obtain a sequence of triplets

$$\mathcal{T} = \mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_{n-1}, \mathcal{T}_n = \mathcal{T}'.$$

We identify a complete sphere triplet with a 2-simplex of $\mathscr{R}$. Then we prove by induction that $\mathcal{T}_i \subset \mathscr{R}'$. The base case $\mathcal{T}_0 \subset \mathscr{R}'$ clearly holds. We assume that $\mathcal{T}_k \subset \mathscr{R}'$. Without loss of generality, we assume that $\theta_{k+1} \in \mathcal{H}_1$. By the induction assumption, $\mathcal{T}_k$ and S_1 are contained in the same component $\mathscr{R}'$. So $\theta_{k+1}(\mathcal{T}_k)$ and $\theta_{k+1}(S_1)$ are also in the same component of $\mathscr{R}$. Since $\theta_{k+1}(S_1) = S_1 \in \mathscr{R}'$, we have that $\mathcal{T}_{k+1} = \theta_{k+1}(\mathcal{T}_k) \subset \mathscr{R}'$. Therefore, $S \in \mathcal{T}' = \mathcal{T}_n \subset \mathscr{R}'$. □

5. Finitely many generators

In this section, we prove that each $\mathcal{H}_i$ is finitely generated. Then, by Theorem 4.8, $\mathcal{G}(N, \Sigma)$ $(= \langle \mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3 \rangle)$ is also finitely generated. Before that, we first introduce the following definition.

Definition 5.1. Let $N = A \cup_\Sigma B$ be a Heegaard splitting and D a union of finitely many disjoint *marked*[6] disks in Σ. In this case, we define the Goeritz group $\mathcal{G}(N, \Sigma, D)$ to be the group of diffeomorphisms of N that preserve the Heegaard splitting and fix D pointwise, modulo isotopies[7] that leave Σ and D invariant.

Note 5.2. Unless otherwise specified, all manifolds considered in this section are not assumed to be the connected sum of lens spaces.

Suppose $\mathcal{B}$ is a bubble for Σ, with the boundary sphere $S = \partial\mathcal{B}$ intersecting Σ in an essential simple closed curve μ. Let D_a be the disk $S \cap A$ and $\Sigma_{\mathcal{B}}$ be the bordered surface $\Sigma \setminus \mathrm{int}(\mathcal{B})$. Denote by $\Sigma(\mathcal{B})$ the closed surface $\Sigma_{\mathcal{B}} \cup D_a$. Since S is separating

[6] In classical surface mapping class group theory, the surfaces with marked points are frequently considered. Here, we simply follow this tradition by discussing mapping class groups of surfaces with marked disks.

[7] Precisely, such an isotopy $H_t : N \to N$, $0 \leq t \leq 1$, is required to satisfy that $H_t(\Sigma, D) = (\Sigma, D)$ and $H_0|_D = H_1|_D = \mathrm{id}$.

in N, we cap off the sphere boundary S of the dual bubble $\mathcal{B}'$ $(= N \setminus \mathrm{int}(\mathcal{B}))$ with a 3-ball to obtain a new closed orientable 3-manifold $N(\mathcal{B})$. It is not hard to see that $\Sigma(\mathcal{B})$ is a Heegaard surface for $N(\mathcal{B})$. Putting an orientation for the curve μ, we get the oriented curve $\vec{\mu}$ and its isotopy class $[\vec{\mu}]$. Denote by $G_{\vec{\mu}}$ the stabilizer of $[\vec{\mu}]$:

$$G_{\vec{\mu}} \overset{\text{def}}{=} \{\phi \in \mathcal{G}(N, \Sigma) : \phi([\vec{\mu}]) = [\vec{\mu}]\}.$$

We define a subgroup of $\mathrm{Diff}^+(N, \Sigma)$ by

$$\mathrm{Diff}^+(N, \Sigma, S) \overset{\text{def}}{=} \{f \in \mathrm{Diff}^+(N, \Sigma) : f|_S = \mathrm{id}\}.$$

By the definition, if $f \in \mathrm{Diff}^+(N, \Sigma, S)$, then $f(\mathcal{B}') = \mathcal{B}'$ $(\mathcal{B}' = \overline{N \setminus \mathcal{B}})$. We associate it with an element of the Goeritz group $\mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}), D_a)$ as follows. Since $N(\mathcal{B})$ is the union of B' and a 3-ball, $f|_{\mathcal{B}'}$ can be naturally extended into a diffeomorphism $\hat{f} : (N(\mathcal{B}), \Sigma(\mathcal{B}), D_a) \to (N(\mathcal{B}), \Sigma(\mathcal{B}), D_a)$. All different extensions of f are pairwise isotopic. Subsequently, we obtain a map $\rho_2 : \mathrm{Diff}^+(N, \Sigma, S) \to \mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}), D_a)$. It is not hard to verify that ρ_2 is also an epimorphism.

Restricting the natural homomorphism $\rho_1 : \mathrm{Diff}^+(N, \Sigma) \to \mathcal{G}(N, \Sigma)$ to the subgroup $\mathrm{Diff}^+(N, \Sigma, S)$ results in a restriction map, which we still denote by ρ_1. It is easy to see that $\rho_1(\mathrm{Diff}^+(N, \Sigma, S)) = G_{\vec{\mu}}$.

We want to define a homomorphism $\rho : G_{\vec{\mu}} \to \mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}), D_a)$ so that the following diagram commutes:

$$\begin{array}{ccc} \mathrm{Diff}^+(N, \Sigma, S) & \xrightarrow{\ \rho_1\ } & G_{\vec{\mu}} \\ & \searrow^{\rho_2} & \downarrow{\scriptstyle\rho} \\ & & \mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}), D_a) \end{array}$$

If such a ρ exists, it is uniquely determined by the requirement. Its existence is guaranteed by the following lemma.

Lemma 5.3. *For any diffeomorphism* $f \in \mathrm{Diff}^+(N, \Sigma, S)$, *if* $\rho_1(f) = \mathrm{id}$, *then* $\rho_2(f) = \mathrm{id}$.

Proof. Assume that $\vec{\alpha}$ is an oriented essential simple closed curve in $\Sigma_{\mathcal{B}}$. Since $\rho_1(f) = \mathrm{id}$, $f(\vec{\alpha})$ is isotopic to $\vec{\alpha}$ in Σ. By [Farb and Margalit 2012, Lemma 3.16], we know that $f(\vec{\alpha})$ is also isotopic to $\vec{\alpha}$ in $\Sigma_{\mathcal{B}}$. In other words, f preserves the isotopy classes of all oriented essential simple closed curves in $\Sigma_{\mathcal{B}}$. We can prove by the *Alexander method*[8] that such a diffeomorphism is isotopic to a power of the Dehn twist τ_μ. This means that $\rho_2(f) = \mathrm{id}$. □

[8]Here, we first choose a collection $\{\gamma_i\}$ of essential simple closed curves in $\Sigma_{\mathcal{B}}$, as shown in Figure 11. By the Alexander method [Farb and Margalit 2012, Proposition 2.8], we know that $f\left(\bigcup \gamma_i\right)$ is isotopic to $\bigcup \gamma_i$ relative to μ $(= \partial \Sigma_{\mathcal{B}})$. So we assume that $f\left(\bigcup \gamma_i\right) = \bigcup \gamma_i$. In addition, f also

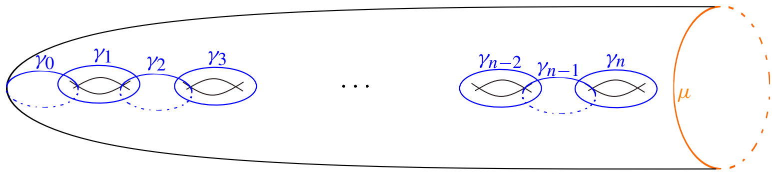

Figure 11. The surface $\Sigma_{\mathcal{B}}$.

It follows that such ρ exists and is surjective, and the kernel of ρ is denoted by $\mathcal{I}(\rho)$. Then we have the exact sequence

$$1 \to \mathcal{I}(\rho) \xrightarrow{i} G_{\vec{\mu}} \xrightarrow{\rho} \mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}), D_a) \to 1. \tag{2}$$

Similarly, for the dual bubble $\mathcal{B}'$, we also have a dual exact sequence

$$1 \to \mathcal{I}(\rho') \xrightarrow{i'} G_{\vec{\mu}} \xrightarrow{\rho'} \mathcal{G}(N(\mathcal{B}'), \Sigma(\mathcal{B}'), D_a) \to 1. \tag{3}$$

Lemma 5.4. *The composite homomorphism $\rho' \cdot i$ is an epimorphism.*

Proof. By a similar argument for the dual bubble $\mathcal{B}'$, we have the following commutative diagram:

$$\begin{array}{ccc} \mathrm{Diff}^+(N, \Sigma, S) & \xrightarrow{\quad\rho_1\quad} & G_{\vec{\mu}} \\ & \searrow^{\rho_2'} & \downarrow{\scriptstyle \rho'} \\ & & \mathcal{G}(N(\mathcal{B}'), \Sigma(\mathcal{B}'), D_a) \end{array}$$

Since ρ_2 is surjective, for any element $\phi \in \mathcal{G}(N(\mathcal{B}'), \Sigma(\mathcal{B}'), D_a)$, we can find a diffeomorphism $f \in \mathrm{Diff}^+(N, \Sigma, S)$ such that $\rho_2'(f) = \phi$. We extend $f|_{\mathcal{B}}$ by the identity to obtain a diffeomorphism $\hat{f} \in \mathrm{Diff}^+(N, \Sigma, S)$. It is not hard to see that $\rho' \cdot \rho_1(\hat{f}) = \rho_2'(\hat{f}) = \rho_2'(f) = \phi$ and $\rho_1(\hat{f}) \in \mathcal{I}(\rho)$. The lemma follows immediately. □

Subsequently, we have the exact sequence

$$1 \to \mathcal{I}(\rho') \cap \mathcal{I}(\rho) \xrightarrow{i'''} \mathcal{I}(\rho) \xrightarrow{\rho' \cdot i} \mathcal{G}(N(\mathcal{B}'), \Sigma(\mathcal{B}'), D_a) \to 1. \tag{4}$$

Since the marked disk D_a can be treated as a marked point in $\Sigma(\mathcal{B})$, we apply the description of the kernel of the *capping homomorphism*[9] [Farb and Margalit 2012, Proposition 3.19] to obtain $\mathcal{I}(\rho') \cap \mathcal{I}(\rho) = \langle \tilde{\tau}_\mu \rangle$, where $\tilde{\tau}_\mu$ is the extension of the Dehn twist τ_μ to the whole manifold N.

---

preserves the orientation of γ_i. This implies that f fixes each vertex and each edge of $\bigcup \gamma_i$ (regard $\bigcup \gamma_i$ as a graph). Without loss of generality, we may further assume that $f|_{\bigcup \gamma_i} = \mathrm{id}$. On the other hand, $\Sigma_{\mathcal{B}} \setminus (\bigcup \gamma_i)$ is an annulus. It follows that f is isotopic to a power of the *Dehn twist* τ_μ.

[9]Let $f \in \mathrm{Diff}^+(N, \Sigma, S)$ be a diffeomorphism that represents an element of $\mathcal{I}(\rho') \cap \mathcal{I}(\rho)$. Then, with D_a identified with a marked point in $\Sigma(\mathcal{B})$, $f|_{\Sigma_{\mathcal{B}}}$ represents an element of the kernel of the capping homomorphism $\mathrm{Cap} : \mathrm{Mod}(\Sigma_{\mathcal{B}}, \partial \Sigma_{\mathcal{B}}) \to \mathrm{Mod}(\Sigma(\mathcal{B}), D_a)$. By [Farb and Margalit 2012, Proposition 3.19], $f|_{\Sigma_{\mathcal{B}}}$ is isotopic to a power of the Dehn twist along μ $(= \partial \Sigma_{\mathcal{B}})$, and so is $f|_{\Sigma_{\mathcal{B}'}}$. This implies that f is isotopic to a power of the Dehn twist along μ.

Lemma 5.5. *If both $\mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}))$ and $\mathcal{G}(N(\mathcal{B}'), \Sigma(\mathcal{B}'))$ are finitely generated (or finitely presented), then $G_{\vec{\mu}}$ is finitely generated (or finitely presented) as well.*

Proof. We first prove that $\mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}), D_a)$ is finitely generated (or finitely presented). If the genus $g(\Sigma(\mathcal{B}))$ is 1, we have $\mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}), D_a) = \mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}))$. Then there is nothing to prove. If $g(\Sigma(\mathcal{B})) \geqslant 2$, we apply the *Birman exact sequence* [Birman 1969] for the pair $(\Sigma(\mathcal{B}), D_a)$ to obtain the commutative diagram

$$\begin{array}{ccccccccc} 1 & \longrightarrow & K & \xrightarrow{\text{push}} & \mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}), D_a) & \xrightarrow{\text{forget}} & \mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B})) & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle i} & & \downarrow{\scriptstyle i} & & \downarrow{\scriptstyle i} & & \\ 1 & \longrightarrow & \pi_1(\Sigma(\mathcal{B})) & \xrightarrow{\text{push}} & \mathrm{Mod}(\Sigma(\mathcal{B}), D_a) & \xrightarrow{\text{forget}} & \mathrm{Mod}(\Sigma(\mathcal{B})) & \longrightarrow & 1 \end{array}$$

where i denotes the inclusion map. The Birman exact sequence provides a description for the kernel of the *forget* map, which asserts that the kernel is generated by the isotopies (of $\Sigma(\mathcal{B})$) that push D_a along a closed path (that begins and ends at D_a) in $\Sigma(\mathcal{B})$. Note that all such isotopies can be extended to the whole manifold $N(\mathcal{B})$. It follows that $K = \pi_1(\Sigma(\mathcal{B}))$. Since both $\pi_1(\Sigma(\mathcal{B}))$ and $\mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}))$ are finitely generated (or finitely presented), so is $\mathcal{G}(N(\mathcal{B}), \Sigma(\mathcal{B}), D_a)$. Similarly, $\mathcal{G}(N(\mathcal{B}'), \Sigma(\mathcal{B}'), D_a)$ is also finitely generated (or finitely presented).

By the exact sequence (4), $\mathcal{I}(\rho)$ is finitely generated (or finitely presented). Then by the exact sequence (2), $G_{\vec{\mu}}$ is also finitely generated (or finitely presented). □

With the above preparations completed, we can present the proof of Theorem 1.3.

Proof of Theorem 1.3. By Lemma 5.5, we know that $G_{\vec{\mu}}$ is finitely generated (or finitely presented). Since $G_{\vec{\mu}}$ is a subgroup of G_μ with index at most two, G_μ is also finitely generated (or finitely presented). □

The genus at most two Goeritz groups for lens spaces or their connected sum have been shown to be finitely generated in [Cho 2013; Cho and Koda 2016; 2019]. Then by Theorem 1.3, the stabilizer G_{μ_i}, which is exactly the group $\mathcal{H}_i$, is finitely generated. By Theorem 4.8, it follows that $\mathcal{G}(N, \Sigma)$ is finitely generated. So we complete the proof of Theorem 1.1.

Acknowledgements

We would like to thank Professor Ruifeng Qiu and Chao Wang for many helpful suggestions. This work was partially supported by the NSFC (grant nos. 12131009, 12471065 and 12326601) and the Science and Technology Commission of Shanghai Municipality (grant no. 22DZ2229014).

References

[Birman 1969] J. S. Birman, "Mapping class groups and their relationship to braid groups", *Comm. Pure Appl. Math.* **22** (1969), 213–238. MR

[Cho 2013] S. Cho, "Genus-two Goeritz groups of lens spaces", *Pacific J. Math.* **265**:1 (2013), 1–16. MR

[Cho and Koda 2016] S. Cho and Y. Koda, "Connected primitive disk complexes and genus two Goeritz groups of lens spaces", *Int. Math. Res. Not.* **2016**:23 (2016), 7302–7340. MR

[Cho and Koda 2019] S. Cho and Y. Koda, "The mapping class groups of reducible Heegaard splittings of genus two", *Trans. Amer. Math. Soc.* **371**:4 (2019), 2473–2502. MR

[Farb and Margalit 2012] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton Mathematical Series **49**, Princeton Univ. Press, 2012. MR

[Freedman and Scharlemann 2018] M. Freedman and M. Scharlemann, "Powell moves and the Goeritz group", preprint, 2018. arXiv 1804.05909

[Freedman and Scharlemann 2024] M. Freedman and M. Scharlemann, "Uniqueness in Haken's theorem", *Michigan Math. J.* **74**:1 (2024), 119–142. MR

[Goeritz 1933] L. Goeritz, "Die Abbildungen der Brezelfläche und der Vollbrezel vom Geschlecht 2", *Abh. Math. Sem. Univ. Hamburg* **9**:1 (1933), 244–259. MR

[Gordon 2007] C. M. Gordon, "Problems", pp. 401–411 in *Workshop on Heegaard splittings*, edited by C. M. Gordon and Y. Moriah, Geom. Topol. Monogr. **12**, Geom. Topol. Publ., Coventry, 2007. MR

[Hensel and Schultens 2024] S. Hensel and J. Schultens, "The strong Haken theorem via sphere complexes", *Algebr. Geom. Topol.* **24**:5 (2024), 2707–2719. MR

[Johnson 2010] J. Johnson, "Mapping class groups of medium distance Heegaard splittings", *Proc. Amer. Math. Soc.* **138**:12 (2010), 4529–4535. MR

[Namazi 2007] H. Namazi, "Big Heegaard distance implies finite mapping class group", *Topology Appl.* **154**:16 (2007), 2939–2949. MR

[Powell 1980] J. Powell, "Homeomorphisms of S^3 leaving a Heegaard surface invariant", *Trans. Amer. Math. Soc.* **257**:1 (1980), 193–216. MR

[Scharlemann 2022] M. Scharlemann, "Powell's conjecture on the Goeritz group of S^3 is stably true", preprint, 2022. arXiv 2210.13629

[Scharlemann 2024] M. Scharlemann, "A strong Haken theorem", *Algebr. Geom. Topol.* **24**:2 (2024), 717–753. MR

[Taylor 2025] S. A. Taylor, "Strong Haken via thin position", *Bol. Soc. Mat. Mex.* (3) **31**:1 (2025), art. id. 14. MR

[Zou and Qiu 2020] Y. Zou and R. Qiu, "Finiteness of mapping class groups: locally large strongly irreducible Heegaard splittings", *Groups Geom. Dyn.* **14**:2 (2020), 591–605. MR

[Zupan 2020] A. Zupan, "The Powell conjecture and reducing sphere complexes", *J. Lond. Math. Soc.* (2) **101**:1 (2020), 328–348. MR

Received December 25, 2024. Revised August 1, 2025.

HAO CHEN
SCHOOL OF MATHEMATICAL SCIENCES
EAST CHINA NORMAL UNIVERSITY
SHANGHAI
CHINA
hchen@stu.ecnu.edu.cn

YANQING ZOU
SCHOOL OF MATHEMATICAL SCIENCES
KEY LABORATORY OF MEA (MINISTRY OF EDUCATION) AND SHANGHAI KEY LABORATORY OF PMMP
EAST CHINA NORMAL UNIVERSITY
SHANGHAI
CHINA
yqzou@math.ecnu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 338, No. 2, 2025
https://doi.org/10.2140/pjm.2025.338.251

COMPLETE MINIMAL HYPERSURFACES IN A HYPERBOLIC SPACE $H^4(-1)$

QING-MING CHENG AND YEJUAN PENG

We study n-dimensional complete minimal hypersurfaces in the hyperbolic space $H^{n+1}(-1)$ of constant curvature -1. We prove that a 3-dimensional complete minimal hypersurface with constant scalar curvature in $H^4(-1)$ satisfies $S \leq \frac{21}{29}$ by making use of the generalized maximum principle, where S denotes the squared norm of the second fundamental form of the hypersurface.

1. Introduction

Let M^n be an n-dimensional minimal hypersurface in the hyperbolic space $H^{n+1}(-1)$ of constant curvature -1. A very important subject of study is the rigidity of complete minimal hypersurfaces in the hyperbolic space $H^{n+1}(-1)$. It is well known that there are many important results on the rigidity of compact minimal hypersurfaces in the unit sphere $S^{n+1}(1)$. For example, Simons [7], Chern, do Carmo and Kobayashi [3] and Lawson [4] prove that an n-dimensional compact minimal hypersurface in the unit sphere $S^{n+1}(1)$ is isometric to a totally geodesic sphere or a Clifford torus if the squared norm S of its second fundamental form satisfies $S \leq n$. In particular, for $n = 3$, it is known that a 3-dimensional compact minimal hypersurface in the unit sphere $S^4(1)$ with constant scalar curvature is isometric to a totally geodesic sphere or a Clifford torus or the Cartan minimal isoparametric hypersurface (see [1; 6]). On the other hand, Cheng and Wan [2] proved complete minimal hypersurfaces with constant scalar curvature in the Euclidean space $\mathbb{R}^4$ are isometric to the hyperplane $\mathbb{R}^3$. But for complete minimal hypersurfaces in the hyperbolic space $H^{n+1}(-1)$, there are only few results on rigidity of complete minimal hypersurfaces. It is our main purpose to study the following conjecture:

Conjecture. A complete minimal hypersurface with constant scalar curvature in the hyperbolic space $H^4(-1)$ is isometric to the hyperbolic space $H^3(-1)$.

We will prove the following:

Peng is the corresponding author. This work was supported by JSPS Grant-in-Aid for Scientific Research No: 25K06992.

MSC2020: 53C40, 53C42.

Keywords: minimal hypersurfaces, hyperbolic space, constant scalar curvature, the generalized maximum principle.

Theorem 1.1. *A complete minimal hypersurface with constant scalar curvature in the hyperbolic space $H^4(-1)$ satisfies $S \le \frac{21}{29}$, where S denotes the squared norm of the second fundamental form of the hypersurface.*

2. Basic formulas

Let M^n be an n-dimensional hypersurface in an $(n+1)$-dimensional hyperbolic space $H^{n+1}(-1)$. At each point p in $H^{n+1}(-1)$, we choose a local orthonormal frame field $\{e_1, e_2, \dots, e_{n+1}\}$ and the dual coframe $\{\omega^1, \omega^2, \dots, \omega^{n+1}\}$ such that, restricted to M^n, $\{e_1, e_2, \dots, e_n\}$ is tangent to M^n. Structure equations of $H^{n+1}(-1)$ are given by

$$\begin{gathered} d\omega_A = -\sum_B \omega_{AB} \wedge \omega_B, \quad \omega_{AB} + \omega_{BA} = 0, \\ d\omega_{AB} + \sum_C \omega_{AC} \wedge \omega_{CB} = \tfrac{1}{2} \sum_{C,D} K_{ABCD}\, \omega_C \wedge \omega_D, \end{gathered} \tag{2-1}$$

with

$$K_{ABCD} = -(\delta_{AC}\delta_{BD} - \delta_{AD}\delta_{BC}). \tag{2-2}$$

If we restrict these forms to M^n, then $\omega^{n+1} = 0$. We have

$$\omega_{i,n+1} = \sum_j h_{ij}\omega_j, \quad h_{ij} = h_{ji}. \tag{2-3}$$

One calls

$$H = \frac{1}{n} \sum_i h_{ii}, \quad h = \sum_{i,j} h_{ij}\omega_i \otimes \omega_j \tag{2-4}$$

the mean curvature and the second fundamental form of M^n, respectively. If H is identically zero, M^n is called minimal. The structure equations of M^n are given by

$$\begin{gathered} d\omega_i = -\sum_j \omega_{ij} \wedge \omega_j, \quad \omega_{ij} + \omega_{ji} = 0, \\ d\omega_{ij} + \sum_k \omega_{ik} \wedge \omega_{kj} = \tfrac{1}{2} \sum_{k,l} R_{ijkl}\, \omega_k \wedge \omega_l, \end{gathered} \tag{2-5}$$

where

$$R_{ijkl} = -(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}) + (h_{ik}h_{jl} - h_{il}h_{jk}). \tag{2-6}$$

For minimal hypersurfaces in $H^{n+1}(-1)$, we obtain

$$R = -n(n-1) - S,$$

where R and S denote the scalar curvature and the squared norm of the second fundamental form of M^n, respectively. From the structure equations of M^n, Codazzi equations and Ricci formulas are given by

$$h_{ijk} = h_{ikj}, \quad h_{ijkl} - h_{ijlk} = \sum_m h_{im} R_{mjkl} + \sum_m h_{mj} R_{mikl},$$

where $h_{ijk} = \nabla_k h_{ij}$ and $h_{ijkl} = \nabla_l \nabla_k h_{ij}$, respectively. Define f_3 and f_4 by

$$f_3 = \sum_{i,j,k=1}^{n} h_{ij} h_{jk} h_{ki} \quad \text{and} \quad f_4 = \sum_{i,j,k,l=1}^{n} h_{ij} h_{jk} h_{kl} h_{li},$$

respectively. We have, for minimal hypersurfaces,

$$\begin{aligned} \tfrac{1}{3}\Delta f_3 &= -(n+S) f_3 + 2C, \\ \tfrac{1}{4}\Delta f_4 &= -(n+S) f_4 + (2A+B), \end{aligned} \tag{2-7}$$

where

$$C = \sum_{i,j,k} \lambda_i h_{ijk}^2, \quad A = \sum_{i,j,k} \lambda_i^2 h_{ijk}^2, \quad B = \sum_{i,j,k} \lambda_i \lambda_j h_{ijk}^2$$

and λ_i's are principal curvatures of M^n, that is,

$$\sum_i h_{ii} = \sum_i \lambda_i = 0, \quad S = \sum_{i,j} h_{ij}^2 = \sum_i \lambda_i^2,$$

$$h_{ijij} - h_{jiji} = (\lambda_i - \lambda_j)(-1 + \lambda_i \lambda_j).$$

By a direct computation, we have

$$S = n(1-n) - R, \quad \Delta h_{ij} = -(S+n) h_{ij}, \quad \tfrac{1}{2}\Delta S = -S(S+n) + \sum_{i,j,k} h_{ijk}^2.$$

If the squared norm S of the second fundamental form is constant, we have

$$\sum_{i,j,k} h_{ijk}^2 = S(S+n), \quad \sum_{i,j,k,l} h_{ijkl}^2 = S(S+n)(2n+3+S) + 3(A-2B).$$

The following generalized maximum principle due to Omori [5] (see Yau [8]) will play an important role in this paper.

Theorem 2.1. *Let M^n be a complete Riemannian manifold with sectional curvature bounded from below. If a C^2-function f is bounded from above in M^n, then there exists a sequence $\{p_k\}_{k=1}^{\infty} \subset M^n$ such that*

(1) $\lim_{k\to\infty} f(p_k) = \sup_{M^n} f$,

(2) $\lim_{k\to\infty} |\nabla f(p_k)| = 0$,

(3) $\lim_{k\to\infty} \sup \nabla_l \nabla_l f(p_k) \le 0$, *for* $l = 1, 2, \ldots, n$.

3. Minimal hypersurfaces with two distinct principal curvatures

Theorem 3.1. *Let M^3 be a minimal hypersurface in $H^4(-1)$ with constant scalar curvature. If M^3 has two principal curvatures somewhere, we have $S \leq \frac{21}{29}$.*

Proof. We assume, at $p \in M^3$, that M^3 has two distinct principal curvatures. At p, we may choose an orthonormal frame e_1, e_2, e_3 such that $h_{ij} = \lambda_i \delta_{ij}$. We can assume

$$\lambda_1 = \lambda_2 = \lambda.$$

Since M^3 is minimal, we have

$$\lambda_3 = -2\lambda, \quad \lambda^2 = \tfrac{1}{6}S.$$

Because $\sum_i h_{ii} = 0$ and S is constant, we have

$$h_{11k} + h_{22k} + h_{33k} = 0, \quad h_{11k} + h_{22k} - 2h_{33k} = 0.$$

We obtain

$$h_{11k} + h_{22k} = 0, \quad h_{33k} = 0, \qquad k = 1, 2, 3.$$

We can choose e_1, e_2 such that $h_{123}(p) = 0$ at p. In fact, if necessary, we make a rotation of e_1, e_2 with angle θ, which satisfies

$$\cos(-2\theta) = \frac{h_{223}(p)}{\sqrt{h_{223}^2(p) + h_{123}^2(p)}}, \quad \sin(-2\theta) = \frac{h_{123}(p)}{\sqrt{h_{223}^2(p) + h_{123}^2(p)}}.$$

Letting

$$a = h_{113}^2, \quad b = h_{111}^2 + h_{112}^2,$$

in view of

$$\begin{aligned} S(S+3) = \sum_{i,j,k} h_{ijk}^2 &= 3(h_{112}^2 + h_{113}^2 + h_{221}^2 + h_{223}^2) + (h_{111}^2 + h_{222}^2) \\ &= 6h_{113}^2 + 4(h_{111}^2 + h_{112}^2), \end{aligned}$$

we have

$$6a + 4b = S(S+3).$$

Since $n = 3$, we have

$$f_4 = \tfrac{1}{2}S^2, \quad 2A + B = \tfrac{1}{2}S^2(S+3).$$

Lemma 3.1. *h_{ijkl} are symmetric in i, j, k, l if i, j, k, l are not $\{1, 1, 3, 3\}$, $\{2, 2, 3, 3\}$ and*

$$h_{3311} = h_{3322} = \frac{2}{3\lambda}(a+b), \quad h_{3333} = \frac{2a}{3\lambda}, \quad h_{3312} = 0, \quad h_{3313} = \frac{2}{3\lambda}h_{111}h_{113},$$
$$h_{3323} = \frac{2}{3\lambda}h_{112}h_{113}, \quad h_{1111} = h_{2222}, \quad h_{1133} = h_{2233} = -\frac{a}{3\lambda}.$$

Proof. According to the Ricci formula,

$$\begin{aligned} h_{ijkl} - h_{ijlk} &= \sum_m h_{mj} R_{mikl} + \sum_m h_{im} R_{mjkl} \\ &= (\lambda_i - \lambda_j) R_{ijkl} \\ &= (\lambda_i - \lambda_j)(-1 + \lambda_i \lambda_j)(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}). \end{aligned}$$

Also $S = \sum_{i,j} h_{ij}^2$ is constant. We have

$$0 = \sum_{i,j} (h_{ij}^2)_{kl} = 2\biggl(\sum_{i,j} h_{ijk} h_{ijl} + \sum_{i,j} h_{ij} h_{ijkl} \biggr) = 2\biggl(\sum_{i,j} h_{ijk} h_{ijl} - 3\lambda h_{33kl} \biggr). \quad \square$$

Lemma 3.2. *We have*

$$x + 2y = \tfrac{26}{9} a^2 + \tfrac{7}{18} ab - b^2 + \tfrac{5}{4} Sb, \tag{3-1}$$

where

$$x = \lambda^2 [3(h_{1123}^2 + h_{2213}^2) + h_{1113}^2 + h_{2223}^2], \quad y = \lambda^2 (h_{1111}^2 + h_{1112}^2) + (a+b)\lambda h_{1111}.$$

Proof. We have

$$\begin{aligned} \sum_{i,j,k,l} h_{ijkl}^2 &= S(S+3)(S+9) + 3(A - 2B) \\ &= S(S+3)(S+9) + 4(2A + B) - 5(A + 2B) \\ &= S(S+3)(S+9) + 2S^2(S+3) - 5\biggl(\sum_{i,j,k} h_{ijk}^2 \lambda_i^2 + 2 \sum_{i,j,k} h_{ijk}^2 \lambda_i \lambda_j \biggr) \\ &= 3S(S+3)^2 - \tfrac{5}{3} \sum_{i,j,k} h_{ijk}^2 (\lambda_i + \lambda_j + \lambda_k)^2, \end{aligned}$$

where

$$\sum_{i,j,k} h_{ijk}^2 (\lambda_i + \lambda_j + \lambda_k)^2 = 3 \sum_{i \neq k} h_{iik}^2 (2\lambda_i + \lambda_k)^2 + 9 \sum_i h_{iii}^2 \lambda_i^2 = 36\lambda^2 b.$$

We have

$$\sum_{i,j,k,l} h_{ijkl}^2 = 3S(S+3)^2 - 60\lambda^2 b = 3S(S+3)^2 - 10Sb,$$

$$\begin{aligned} \sum_{i,j,k} h_{ijk1}^2 &= \sum_{i \neq j \neq k} h_{ijk1}^2 + 3 \sum_{i \neq k} h_{iik1}^2 + \sum_i h_{iii1}^2 \\ &= 6h_{1231}^2 + 3(h_{1121}^2 + h_{1131}^2 + h_{2211}^2 + h_{2231}^2 + h_{3311}^2) + (h_{1111}^2 + h_{2221}^2 + h_{3331}^2) \\ &= 3(2h_{1123}^2 + h_{1113}^2 + h_{2213}^2) + (h_{1111}^2 + 3h_{1112}^2) + h_{3331}^2 + (h_{2221}^2 + 3h_{2211}^2 + 3h_{3311}^2) \\ &= 3(2h_{1123}^2 + h_{1113}^2 + h_{2213}^2) + 4(h_{1111}^2 + h_{1112}^2) + h_{3331}^2 + 6(h_{1111} h_{3311} + h_{3311}^2). \end{aligned}$$

In the same way, we have

$$\begin{aligned}\sum_{i,j,k} h_{ijk2}^2 &= \sum_{i\neq j\neq k} h_{ijk2}^2 + 3\sum_{i\neq k} h_{iik2}^2 + \sum_i h_{iii2}^2 \\ &= 3(2h_{2213}^2 + h_{1123}^2 + h_{2223}^2) + 4(h_{1111}^2 + h_{1112}^2) \\ &\qquad + h_{3332}^2 + 3(2h_{1111}h_{3322} + h_{3311}^2 + h_{3322}^2),\end{aligned}$$

$$\begin{aligned}\sum_{i,j,k} h_{ijk3}^2 &= \sum_{i\neq j\neq k} h_{ijk3}^2 + 3\sum_{i\neq k} h_{iik3}^2 + \sum_i h_{iii3}^2 \\ &= 3(h_{1123}^2 + h_{2213}^2) + h_{1113}^2 + h_{2223}^2 \\ &\qquad + 3(h_{1133}^2 + h_{2233}^2 + h_{3313}^2 + h_{3323}^2) + h_{3333}^2 \\ &= 3(h_{1123}^2 + h_{2213}^2) + h_{1113}^2 + h_{2223}^2 + \left(\frac{2a^2}{3\lambda^2} + \frac{4}{3\lambda^2}ab\right) + \frac{4a^2}{9\lambda^2} \\ &= 3(h_{1123}^2 + h_{2213}^2) + h_{1113}^2 + h_{2223}^2 + \frac{10a^2 + 12ab}{9\lambda^2}.\end{aligned}$$

Hence we obtain

$$\begin{aligned}\sum_{i,j,k,l} h_{ijkl}^2 &= \sum_{i,j,k} h_{ijk1}^2 + \sum_{i,j,k} h_{ijk2}^2 + \sum_{i,j,k} h_{ijk3}^2 \\ &= 12(h_{1123}^2 + h_{2213}^2) + 4(h_{1113}^2 + h_{2223}^2) + 8(h_{1111}^2 + h_{1112}^2) \\ &\qquad + 12h_{1111}h_{3311} + (h_{3331}^2 + h_{3332}^2) + 12h_{3311}^2 + \frac{10a^2 + 12ab}{9\lambda^2} \\ &= [12(h_{1123}^2 + h_{2213}^2) + 4(h_{1113}^2 + h_{2223}^2)] \\ &\qquad + \left[8(h_{1111}^2 + h_{1112}^2) + \frac{8}{\lambda}h_{1111}(a+b)\right] \\ &\qquad + \frac{4}{9\lambda^2}ab + 12\left[\frac{2}{3\lambda}(a+b)\right]^2 + \frac{10a^2 + 12ab}{9\lambda^2}.\end{aligned}$$

We infer, from the above formulas,

$$\frac{4}{\lambda^2}x + \frac{8}{\lambda^2}y + \frac{4ab + 48(a+b)^2 + 10a^2 + 12ab}{9\lambda^2} = 3S(S+3)^2 - 10Sb,$$

that is,

$$\begin{aligned}x + 2y &= \tfrac{26}{9}a^2 + \tfrac{26}{9}ab + \tfrac{2}{3}b^2 - \tfrac{5}{12}S^2 b \\ &= \tfrac{26}{9}a^2 + \tfrac{7}{18}ab - b^2 + \tfrac{5}{4}Sb.\end{aligned}$$

□

Lemma 3.3. *We have*

$$x + 4a\lambda h_{1111} = -\tfrac{34}{9}a^2 - \tfrac{4}{3}ab + \tfrac{4}{3}b^2 + \lambda^2(72\lambda^2 + 18)a + \lambda^2(40\lambda^2 + 8)b. \tag{3-2}$$

Proof. Since $S=\sum_{i,j} h_{ij}^2$ is constant, we get, for any k, l, m,

$$0=\Big(\sum_{i,j} h_{ij}^2\Big)_{klm}=2\sum_{i,j}(h_{ij}h_{ijklm}+h_{ijm}h_{ijkl}+h_{ijk}h_{ijlm}+h_{ijl}h_{ijkm}).$$

Since

$$\sum_{i,j} h_{ij}h_{ijklm}=-3\lambda h_{33klm},$$

we have

$$3\lambda h_{33klm}=\sum_{i,j} h_{ijm}h_{ijkl}+\sum_{i,j} h_{ijk}h_{ijlm}+\sum_{i,j} h_{ijl}h_{ijkm}.$$

Hence,

$$\sum_{k,l,m} h_{klm}h_{33klm}=\frac{1}{\lambda}\sum_{i,j,k,l,m} h_{ijk}h_{klm}h_{ijlm}.$$

On the other hand, we have

$$0=\Big(\sum_{i,j,k} h_{ijk}^2\Big)_{33}=2\sum_{i,j,k}(h_{ijk}h_{ijk33}+h_{ijk3}^2).$$

$$\begin{aligned}\sum_{i,j,k} h_{ijk}(h_{33ijk}-h_{ijk33})&=\sum_{i,j,k} h_{ijk}h_{33ijk}+\sum_{i,j,k} h_{ijk3}^2\\ &=\frac{1}{\lambda}\sum_{i,j,k,l,m} h_{ijk}h_{klm}h_{ijlm}+\frac{x}{\lambda^2}+\frac{10a^2+12ab}{9\lambda^2}.\end{aligned}$$

Since

$$\begin{aligned}&\sum_{i,j,k} h_{ijk}(h_{33ijk}-h_{ijk33})\\ &=\sum_{i,j,k} h_{ijk}[h_{3i3jk}-h_{ijk33}]\\ &=\sum_{i,j,k} h_{ijk}\Big[\Big(h_{3ij3}+\sum_m h_{mi}R_{m33j}+\sum_m h_{3m}R_{mi3j}\Big)_k-\Big(h_{ij3k}+2\sum_m h_{mj}R_{mik3}\Big)_3\Big]\\ &=\sum_{i,j,k} h_{ijk}\Big[h_{3ij3k}-h_{ij3k3}+\sum_m h_{mik}R_{m33j}+\sum_m h_{3mk}R_{mi3j}-2\sum_m h_{mj3}R_{mik3}\Big]\\ &\qquad+\sum_{i,j,k,m} h_{ijk}h_{mi}(h_{m3}h_{3j}-h_{mj}h_{33})_k+\sum_{i,j,k,m} h_{ijk}h_{3m}(h_{m3}h_{ij}-h_{mj}h_{i3})_k\\ &\qquad-2\sum_{i,j,k,m} h_{ijk}h_{mj}(h_{mk}h_{i3}-h_{m3}h_{ik})_3\\ &=\sum_{i,j,k} h_{ijk}\Big[2\sum_m h_{mij}R_{m33k}+5\sum_m h_{3mj}R_{mi3k}\Big]\\ &\qquad+\sum_{i,j,k,m} h_{ijk}h_{mi}(h_{m3k}h_{3j}+h_{m3}h_{3jk}-h_{mjk}h_{33})\\ &\qquad+\sum_{i,j,k,m} h_{ijk}h_{3m}(h_{m3}h_{ijk}-h_{mjk}h_{i3}-h_{mj}h_{i3k}+h_{m3k}h_{ij})\\ &\qquad-2\sum_{i,j,k,m} h_{ijk}h_{mj}(h_{mk3}h_{i3}-h_{m3}h_{ik3})\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i,j,k,m} \Big[2h_{ijk}h_{mij}(-1+\lambda_k\lambda_3)(\delta_{k3}\delta_{3m}-\delta_{mk}\delta_{33}) \\
&\qquad\qquad\qquad\qquad + 5h_{ijk}h_{3mj}(-1+\lambda_i\lambda_k)(\delta_{m3}\delta_{ik}-\delta_{mk}\delta_{i3})\Big] \\
&\qquad + \sum_{i,k}\lambda_3\lambda_i h_{i3k}^2 - \sum_{i,j,k}\lambda_3\lambda_i h_{ijk}^2 \\
&\qquad + \sum_{j,k}\lambda_3^2 h_{3jk}^2 + \sum_{i,j,k} h_{ijk}^2\lambda_3^2 - \sum_{j,k} h_{3jk}^2\lambda_3^2 \\
&\qquad - \sum_{i,k}\lambda_3^2 h_{i3k}^2 - 2\sum_{j,k}\lambda_3\lambda_j h_{3jk}^2 + 2\sum_{i,k}\lambda_3^2 h_{i3k}^2 \\
&= \Big[2\sum_{i,j} h_{ij3}^2(-1+\lambda_3^2) - 2\sum_{i,j,k} h_{ijk}^2(-1+\lambda_k\lambda_3) - 5\sum_{j,k} h_{3jk}^2(-1+\lambda_3\lambda_k)\Big] \\
&\qquad\qquad - \sum_{i,j}\lambda_3\lambda_i h_{ij3}^2 - \sum_{i,j,k}\lambda_3\lambda_i h_{ijk}^2 + \sum_{i,j,k}\lambda_3^2 h_{ijk}^2 + \sum_{i,j}\lambda_3^2 h_{ij3}^2 \\
&= \big[2(-1+4\lambda^2)(2a) - 2(-(6a+4b)-2\lambda(4\lambda b)) + 5(1+2\lambda^2)(2a)\big] \\
&\qquad\qquad + 4a\lambda^2 + 2\lambda(4\lambda b) + 4\lambda^2(6a+4b) + 8\lambda^2 a \\
&= (72\lambda^2+18)a + (40\lambda^2+8)b.
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{i,j,k,l,m} h_{ijk}h_{klm}h_{ijlm} \\
&= \sum_{k,l,m} h_{klm}(h_{11k}h_{11lm} + h_{22k}h_{22lm} + 2h_{12k}h_{12lm} + 2h_{13k}h_{13lm} + 2h_{23k}h_{23lm}) \\
&= \sum_{k,l,m} h_{11k}h_{klm}(h_{11lm} - h_{22lm}) + \sum_{l,m} 2(h_{112}h_{1lm} - h_{111}h_{2lm})h_{12lm} \\
&\qquad\qquad + \sum_{l,m} 2h_{113}h_{1lm}h_{13lm} - \sum_{l,m} 2h_{113}h_{2lm}h_{23lm} \\
&= \sum_k h_{11k}\big[h_{k11}(h_{1111} - h_{2211}) + h_{k22}(h_{1122} - h_{2222}) + 2h_{k13}(h_{1113} - h_{2213}) \\
&\qquad + 2h_{k23}(h_{1123} - h_{2223}) + 2h_{k12}(h_{1112} - h_{2212})\big] \\
&\qquad + 2h_{112}(2h_{112}h_{1212} + 2h_{113}h_{1213} + h_{111}h_{1211} + h_{122}h_{1222}) \\
&\qquad - 2h_{111}(2h_{212}h_{1212} + h_{211}h_{1211} + h_{222}h_{1222} + 2h_{223}h_{1223}) \\
&\qquad + 2h_{113}\big[h_{111}h_{1311} + h_{113}(h_{1313} + h_{1331}) + 2h_{112}h_{1312} + h_{122}h_{1322}\big] \\
&\qquad - 2h_{113}\big[h_{222}h_{2322} + 2h_{212}h_{2312} + h_{223}(h_{2323} + h_{2332}) + h_{211}h_{2311}\big] \\
&= (a+b)(h_{1111} - h_{2211}) + \sum_k h_{11k}^2(h_{1111} - h_{2211}) \\
&\qquad + 4bh_{1122} + 2a(h_{1133} + h_{3311} + h_{2233} + h_{3322}) \\
&\qquad + 4h_{111}h_{113}h_{1113} - 4h_{112}h_{223}h_{2223} + 4h_{112}h_{113}h_{1123} - 4h_{113}h_{221}h_{2213} \\
&= 2(a+b)(h_{1111} - h_{2211}) + 4bh_{1122} \\
&\qquad + 4h_{113}(h_{111}h_{1113} + h_{112}h_{2223} + h_{112}h_{1123} + h_{111}h_{2213}) \\
&\qquad + 2a(h_{1133} + h_{3311} + h_{2233} + h_{3322})
\end{aligned}
$$

$$
\begin{aligned}
&= 2(a+b)(h_{1111} - h_{2211}) + 4bh_{1122} \\
&\qquad - 4h_{113}(h_{111}h_{3313} + h_{112}h_{3323}) + 2a(h_{1133} + h_{3311} + h_{2233} + h_{3322}) \\
&= 2(a+b)(2h_{1111} + h_{3311}) - 4b(h_{1111} + h_{3311}) \\
&\qquad - 4h_{113}\Big(h_{111} \cdot \frac{2}{3\lambda} h_{111}h_{113} + h_{112} \cdot \frac{2}{3\lambda} h_{112}h_{113}\Big) \\
&\qquad + 2a\Big(-\frac{a}{3\lambda} + \frac{2}{3\lambda}(a+b) - \frac{a}{3\lambda} + \frac{2}{3\lambda}(a+b)\Big) \\
&= 4ah_{1111} + 2(a-b) \cdot \frac{2}{3\lambda}(a+b) - \frac{8}{3\lambda} a(h_{111}^2 + h_{112}^2) - \frac{4a^2}{3\lambda} + \frac{8a}{3\lambda}(a+b) \\
&= 4ah_{1111} + \frac{8a^2 - 4b^2}{3\lambda}.
\end{aligned}
$$

Hence, we have

$$
(72\lambda^2 + 18)a + (40\lambda^2 + 8)b = \frac{4ah_{1111}}{\lambda} + \frac{8a^2 - 4b^2}{3\lambda^2} + \frac{x}{\lambda^2} + \frac{10a^2 + 12ab}{9\lambda^2},
$$

$$
\begin{aligned}
x + 4a\lambda h_{1111} &= \lambda^2(72\lambda^2 + 18)a + \lambda^2(40\lambda^2 + 8)b - \tfrac{1}{9}(34a^2 + 12ab - 12b^2) \\
&= -\tfrac{34}{9}a^2 - \tfrac{4}{3}ab + \tfrac{4}{3}b^2 + \lambda^2(72\lambda^2 + 18)a + \lambda^2(40\lambda^2 + 8)b. \qquad \square
\end{aligned}
$$

In view of (3-1) and (3-2), we have from $6a + 4b = S(S+3)$, $6\lambda^2 = S$,

$$
\begin{aligned}
\text{(3-3)} \quad & 2\lambda^2(h_{1111}^2 + h_{1112}^2) - 2(a-b)\lambda h_{1111} \\
&= \tfrac{60}{9}a^2 + \tfrac{31}{18}ab - \tfrac{7}{3}b^2 + \tfrac{5}{4}Sb - \lambda^2(72\lambda^2 + 18)a - \lambda^2(40\lambda^2 + 8)b \\
&= \tfrac{20}{3}a^2 + \tfrac{31}{18}ab - \tfrac{7}{3}b^2 + \tfrac{5}{4}Sb - \tfrac{1}{6}S(12S + 18)a - \tfrac{1}{6}S\big(\tfrac{20}{3}S + 8\big)b \\
&= \tfrac{20}{3}a^2 + \tfrac{31}{18}ab - \tfrac{7}{3}b^2 - 2S(S+3)a + 3Sa - \tfrac{10}{9}S\big(S + \tfrac{3}{40}\big)b \\
&= \tfrac{20}{3}a^2 + \tfrac{31}{18}ab - \tfrac{7}{3}b^2 - 2(6a + 4b)a + 3Sa - \tfrac{10}{9}S\big(S + 3 - \tfrac{117}{40}\big)b \\
&= -\tfrac{16}{3}a^2 - \tfrac{233}{18}ab - \tfrac{61}{9}b^2 + S\big(3a + \tfrac{13}{4}b\big).
\end{aligned}
$$

According to

$$
2\lambda^2(h_{1111}^2 + h_{1112}^2) - 2(a-b)\lambda h_{1111} \geq -\tfrac{1}{2}(a-b)^2,
$$

we obtain

$$
\text{(3-4)} \qquad -\tfrac{29}{6}a^2 - \tfrac{251}{18}ab - \tfrac{113}{18}b^2 + S\big(3a + \tfrac{13}{4}b\big) \geq 0.
$$

Since

$$
\begin{aligned}
&-\tfrac{29}{6}a^2 - \tfrac{58}{18}ab - \tfrac{13}{2}ab - \tfrac{13}{3}b^2 \\
&\qquad = -\tfrac{29}{36}a(6a + 4b) - \tfrac{13}{12}(4b + 6a)b = -\tfrac{29}{36}S(S+3)a - \tfrac{13}{12}S(S+3)b,
\end{aligned}
$$

we have from (3-4)

$$
\big(\tfrac{21}{36} - \tfrac{29}{36}S\big)Sa - \tfrac{76}{18}ab - \tfrac{35}{18}b^2 - \tfrac{13}{12}S^2b \geq 0.
$$

Hence we have $S \leq \frac{21}{29}$. $\square$

4. Proof of Theorem 1.1

In this section, we will give a proof of the Theorem 1.1.

Proof of Theorem 1.1. We choose a local frame field $\{e_1, e_2, e_3, e_4\}$ such that at any point p,

$$h_{ij} = \lambda_i \delta_{ij}.$$

Since S is constant, we notice that the sectional curvature is bounded from below from Gauss equations. By making using of the generalized maximum principle due to Omori [5], there exists a sequence $\{p_k\}_{k=1}^{\infty} \subset M^3$ such that

$$\lim_{k\to\infty} f_3(p_k) = \sup_{M^3} f_3, \quad \lim_{k\to\infty} |\nabla f_3(p_k)| = 0, \quad \lim_{k\to\infty} \sup \nabla_l \nabla_l f_3(p_k) \le 0 \text{ for } l = 1, 2, 3.$$

Since S is constant,

$$\sum_{i,j,k} h_{ijk}^2 = S(S+3), \qquad \sum_{i,j,k,l} h_{ijkl}^2 = S(S+3)(S+9) + 3(A-2B),$$

we know that, for any i, j, k, l, $\{\lambda_i(p_k)\}$, $\{h_{ijk}(p_k)\}$ and $\{h_{ijkl}(p_k)\}$ are bounded sequences, respectively. Thus, we can assume, if necessary, by taking a subsequences of $\{p_m\}$,

$$\lim_{m\to\infty} \lambda_i(p_m) = \hat{\lambda}_i, \quad \lim_{m\to\infty} h_{ijk}(p_m) = \hat{h}_{ijk}, \quad \lim_{m\to\infty} h_{ijkl}(p_m) = \hat{h}_{ijkl} \quad \text{for all } i, j, k, l.$$

From now on, all the computations are considered for $\hat{\lambda}_i, \hat{h}_{ijk}$ and $\hat{h}_{ijkl}$. For simplicity, we omit $\hat{}$.

If the principal curvatures are the same, $S \equiv 0$ since M^3 is minimal. We only consider the following two cases.

Case 1. The number of distinct principal curvatures is two. By the same proof as in the Section 3, we get

$$S \le \tfrac{21}{29}.$$

Case 2. All three principal curvatures are distinct. If f_3 is constant, M^3 is isoparametric and $S \equiv 0$. This is impossible. From now on, we suppose that f_3 is not constant. We will derive a contradiction. Without loss of the generality, we assume that $\lambda_1 < \lambda_2 < \lambda_3$. We also assume $\sup f_3 \neq 0$; otherwise we use $\inf f_3 \neq 0$.

Lemma 4.1. *We have*

$$h_{iik} = 0 \quad \textit{for any } i, k \qquad \textit{and} \qquad h_{123}^2 = \tfrac{1}{6} S(S+3).$$

Proof. Since $\sum_i h_{ii} = 0$ and $S = \sum_{i,j} h_{ij}^2$ is constant, we have

$$\sum_i h_{iik} = 0, \quad \sum_i h_{iik}\lambda_i = 0.$$

Since $\lim_{k\to\infty} |\nabla f_3(p_k)| = 0$, we have

$$\sum_i h_{iik}\lambda_i^2 = 0.$$

Since $\lambda_i \neq \lambda_j$ for $i \neq j$, we have $h_{iik} = 0$ for any i, k. From

$$S(S+3) = \sum_{i,j,k} h_{ijk}^2 = 6h_{123}^2,$$

we obtain

$$h_{123}^2 = \tfrac{1}{6}S(S+3). \qquad \square$$

Lemma 4.2. *We have*

$$h_{iijk} = h_{iiik} = h_{kiii} = 0 \quad \textit{for } i \neq j \neq k.$$

Proof. Since $\sum_{i,j,k} h_{ijk}^2 = S(S+3)$, we have $h_{123l} = 0$ for any l, i.e.,

$$h_{iijk} = 0 \quad \text{for } i \neq j \neq k. \tag{4-1}$$

Since $\sum_i h_{ii} = 0$ and $S = \sum_{i,j} h_{ij}^2$ is constant, we have

$$\sum_i h_{iijk} = 0, \qquad \sum_i h_{iijk}\lambda_i = 0.$$

For $j \neq k$, using (4-1), we have

$$h_{jjjk} + h_{kkjk} = 0, \quad \sum_i h_{iijk}\lambda_i = 0 \qquad \text{for } j \neq k. \tag{4-2}$$

From (4-2), we have $h_{jjjk} = h_{kkjk} = 0$ for $j \neq k$. $\square$

Lemma 4.3. *We have*

$$\sum_{i,k} h_{iikk}^2 + 2\sum_{i\neq k} h_{iikk}^2 = 3S(S+3)^2.$$

Proof. From

$$\sum_{i,j,k,l} h_{ijkl}^2 = S(S+3)(S+9) + 3(A-2B),$$

$$3(A-2B) = 6h_{123}^2(\lambda_1^2+\lambda_2^2+\lambda_3^2-2\lambda_1\lambda_2-2\lambda_2\lambda_3-2\lambda_3\lambda_1) = 2S^2(S+3),$$

$$\begin{aligned}\sum_{i,j,k,l} h_{ijkl}^2 &= \sum_{\substack{i\neq j\neq k\\ l}} h_{ijkl}^2 + 3\sum_{\substack{i\neq k\\ l}} h_{iikl}^2 + \sum_{i,l} h_{iiil}^2 \\ &= 3\sum_{i\neq k} h_{iikk}^2 + \sum_i h_{iiii}^2 = \sum_{i,k} h_{iikk}^2 + 2\sum_{i\neq k} h_{iikk}^2,\end{aligned}$$

we have

$$\sum_{i,k} h_{iikk}^2 + 2\sum_{i\neq k} h_{iikk}^2 = S(S+3)(S+9) + 2S^2(S+3) = 3S(S+3)^2. \qquad \square$$

Lemma 4.4. *We have*

$$\sup f_3 > 0, \quad -\sqrt{\tfrac{1}{2}S} < \lambda_1 < -\sqrt{\tfrac{1}{6}S}, \quad -\sqrt{\tfrac{1}{6}S} < \lambda_2 < 0.$$

Proof. Since $\lim_{k\to\infty} \sup \Delta f_3(p_k) \le 0$ and $\frac{1}{3}\Delta f_3 = -(S+3) f_3$, we have

$$\begin{aligned} 0 &\ge -(S+3) \lim_{k\to\infty} \sup f_3(p_k) \\ &= -(S+3) \sup_{M^3} f_3. \end{aligned}$$

We get $\sup f_3 > 0$. We also notice that $\lambda_1 < 0$ and $\lambda_3 > 0$. By a direct computation,

$$\sup f_3 = \lambda_1^3 + \lambda_2^3 + \lambda_3^3 = 3\lambda_1\lambda_2\lambda_3 = 3\lambda_i\big(\lambda_i^2 - \tfrac{1}{2}S\big) \quad \text{for all } i.$$

We obtain $\lambda_1^2 < \frac{1}{2}S$ and $\lambda_3^2 > \frac{1}{2}S$, $\lambda_2 < 0$ and

$$\lambda_1^2 + \lambda_2^2 < \tfrac{1}{2}S. \tag{4-3}$$

Because of $\lambda_1^2 + \lambda_2^2 + \lambda_1\lambda_2 = \frac{1}{2}S$ and $\lambda_1 < \lambda_2 < 0$, we have

$$\lambda_1^2 > \tfrac{1}{6}S, \quad \lambda_2^2 < \tfrac{1}{6}S.$$

and

$$\tfrac{1}{6}S < \lambda_1^2 < \tfrac{1}{2}S, \quad 0 < \lambda_2^2 < \tfrac{1}{6}S, \quad \lambda_1 < \lambda_2 < 0. \qquad \square$$

For simplicity, we use f_3 in place of $\sup f_3$ in the following.

Lemma 4.5. *We have*

$$h_{iikk} = -\tfrac{1}{3}(S+3)\lambda_i + g_i\lambda_k + w g_i g_k,$$

where

$$g_i = \lambda_i^2 - \frac{f_3}{S}\lambda_i - \tfrac{1}{3}S.$$

Proof. Taking derivatives of $\sum_i h_{ii} = 0$ and $\sum_{i,j} h_{ij}^2 = S$, we have

$$\sum_i h_{iikk} = 0, \quad \sum_i h_{iikk}\lambda_i = -\tfrac{1}{3}S(S+3).$$

We solve this rank-5 linear system of six equations with six unknowns h_{iikk}, $i \le k$, with $h_{iijj} = h_{jjii} + (\lambda_i - \lambda_j)(-1 + \lambda_i\lambda_j)$. $\square$

Lemma 4.6. *We have*

$$f_5 = \tfrac{5}{6}S f_3, \quad f_6 = \tfrac{1}{3}f_3^2 + \tfrac{1}{4}S^3,$$

$$\sum_i g_i^2 = \sum_i g_i\lambda_i^2 = \tfrac{1}{6}S^2 - \frac{f_3^2}{S}, \quad \sum_i g_i^4 = \tfrac{1}{2}\Big(\tfrac{1}{6}S^2 - \frac{f_3^2}{S}\Big)^2,$$

$$\sum_i g_i^2\lambda_i = \frac{f_3^3}{S^2} - \tfrac{1}{6}S f_3, \quad \sum_i g_i^2\lambda_i^2 = \tfrac{1}{36}S^3 - \tfrac{1}{6}f_3^2, \quad \sum_i g_i^3\lambda_i = 0.$$

Proof. From $f_3 = 3\lambda_i\big(\lambda_i^2 - \frac{1}{2}S\big)$, for $i = 1, 2, 3$, we have

$$f_5 = \tfrac{5}{6}Sf_3, \quad f_6 = \tfrac{1}{3}f_3^2 + \tfrac{1}{4}S^3.$$

According to $g_i = \lambda_i^2 - \frac{f_3}{S}\lambda_i - \frac{1}{3}S$, we infer

$$\sum_i g_i^2 = \tfrac{1}{6}S^2 - \frac{f_3^2}{S}, \quad \sum_i g_i^4 = \tfrac{1}{2}\Big(\tfrac{1}{6}S^2 - \frac{f_3^2}{S}\Big)^2,$$

$$\sum_i g_i^2\lambda_i = \frac{f_3^3}{S^2} - \tfrac{1}{6}Sf_3, \quad \sum_i g_i^2\lambda_i^2 = \tfrac{1}{36}S^3 - \tfrac{1}{6}f_3^2, \quad \sum_i g_i\lambda_i^2 = \tfrac{1}{6}S^2 - \frac{f_3^2}{S},$$

Because of $F_3 = 3g_i\big(g_i^2 - \frac{1}{2}F_2\big)$, for $i = 1, 2, 3$, we have

$$\sum_i g_i^3\lambda_i = 0,$$

where $F_k = \sum_i g_i^k$. □

Lemma 4.7. *We have*

$$y = \Big(\tfrac{1}{3} + \frac{1}{S}\Big)f_3 \pm \Big[\frac{f_3^2}{S^2}\big(\tfrac{19}{9}S^2 + \tfrac{8}{3}S + 1\big) + \tfrac{7}{9}S(S+6)\big(S + \tfrac{15}{7}\big)\Big]^{\frac{1}{2}},$$

where

$$y = \Big(\tfrac{1}{6}S^2 - \frac{f_3^2}{S}\Big)w.$$

Proof. By using the Lemmas 4.5 and 4.6, we have

$$\sum_{i,k} h_{iikk}^2 = \sum_{i,k}\big(-\tfrac{1}{3}(S+3)\lambda_i + g_i\lambda_k + wg_ig_k\big)^2 = \tfrac{1}{3}S(S+3)^2 + S\Big(\tfrac{1}{6}S^2 - \frac{f_3^2}{S}\Big) + w^2\Big(\tfrac{1}{6}S^2 - \frac{f_3^2}{S}\Big)^2, \tag{4-4}$$

$$\begin{aligned}\sum_i h_{iiii}^2 &= \sum_i\big(-\tfrac{1}{3}(S+3)\lambda_i + g_i\lambda_i + wg_i^2\big)^2\\ &= \tfrac{1}{9}(S+3)^2S + \sum_i g_i^2\lambda_i^2 + w^2\sum_i g_i^4 + 2w\sum_i \lambda_i g_i^3\\ &\qquad - \tfrac{2}{3}(S+3)\sum_i g_i\lambda_i^2 - \tfrac{2}{3}(S+3)w\sum_i g_i^2\lambda_i\\ &= \tfrac{1}{9}S(S+3)^2 + \tfrac{1}{6}S\Big(\tfrac{1}{6}S^2 - \frac{f_3^2}{S}\Big) + w^2\tfrac{1}{2}\Big(\tfrac{1}{6}S^2 - \frac{f^2}{S}\Big)^2\\ &\qquad - \tfrac{2}{3}(S+3)\Big(\tfrac{1}{6}S^2 - \frac{f_3^2}{S}\Big) - \tfrac{2}{3}(S+3)w\Big(\frac{f_3^3}{S^2} - \tfrac{1}{6}Sf_3\Big),\end{aligned} \tag{4-5}$$

$$\begin{aligned}\sum_{i\neq k} h_{iikk}^2 &= \sum_{i,k} h_{iikk}^2 - \sum_i h_{iiii}^2\\ &= \tfrac{2}{9}S(S+3)^2 + \big[\tfrac{5}{36}S^3 - \tfrac{5}{6}f_3^2\big] + \tfrac{2}{3}(S+3)\Big(\tfrac{1}{6}S^2 - \frac{f_3^2}{S}\Big)\\ &\qquad + \tfrac{2}{3}w(S+3)\Big(-\tfrac{1}{6}Sf_3 + \frac{f_3^3}{S^2}\Big) + \tfrac{1}{2}w^2\Big[\tfrac{1}{36}S^4 - \tfrac{1}{3}Sf_3^2 + \frac{f_3^4}{S^2}\Big].\end{aligned} \tag{4-6}$$

Substituting (4-4) and (4-6) into the Lemma 4.3 completes the proof. □

Lemma 4.8. *We have*

$$-\frac{3}{S}y\big(\lambda_l^2-\tfrac{1}{6}S\big)\big(\lambda_l^2-\tfrac{2}{3}S\big)\le\lambda_l\big(\lambda_l^2-\tfrac{1}{6}S\big)\Big(\frac{9}{S}\lambda_l^4-\tfrac{15}{2}\lambda_l^2+2S+3\Big).$$

Proof. Since

$$\tfrac{1}{3}(f_3)_{ll}=\sum_i h_{iill}\lambda_i^2+2\sum_{i,j}h_{ijl}^2\lambda_i,$$

we have

$$0\ge\tfrac{1}{3}\lim_{k\to\infty}\sup(f_3)_{ll}=\tfrac{1}{3}\lim_{k\to\infty}(f_3)_{ll}=\sum_i h_{iill}\lambda_i^2+2\sum_{i,j}h_{ijl}^2\lambda_i. \tag{4-7}$$

By a direct computation, we infer

$$\sum_i h_{iill}\lambda_i^2=-\lambda_l(S+3)\big(\lambda_l^2-\tfrac{1}{2}S\big)+\lambda_l\Big(\tfrac{1}{6}S^2-\frac{f_3^2}{S}\Big)+\Big(\lambda_l^2-\frac{f_3}{S}\lambda_l-\tfrac{1}{3}S\Big)y, \tag{4-8}$$

$$2\sum_{i,j}h_{ijl}^2\lambda_i=-\tfrac{1}{3}S(S+3)\lambda_l. \tag{4-9}$$

By substituting (4-8) and (4-9) into (4-7), we have

$$\Big(\lambda_l^2-\frac{f_3}{S}\lambda_l-\tfrac{1}{3}S\Big)y\le\lambda_l\Big[(S+3)\big(\lambda_l^2-\tfrac{1}{2}S\big)-\Big(\tfrac{1}{6}S^2-\frac{f_3^2}{S}\Big)+\tfrac{1}{3}S(S+3)\Big].\qquad\square$$

If $y\ge 0$, by using the Lemma 4.6, we have

$$\begin{aligned} y&=\Big(\tfrac{1}{3}+\frac{1}{S}\Big)f_3+\Big[\frac{f_3^2}{S^2}\big(\tfrac{19}{9}S^2+\tfrac{8}{3}S+1\big)+\tfrac{7}{9}S(S+6)\big(S+\tfrac{15}{7}\big)\Big]^{\frac{1}{2}},\\ &>\Big(\tfrac{1}{3}+\frac{1}{S}\Big)f_3+\Big[\frac{f_3^2}{S^2}\big(\tfrac{4}{3}S+1\big)^2\Big]^{\frac{1}{2}}\\ &=\Big(\tfrac{5}{3}+\frac{2}{S}\Big)f_3.\end{aligned} \tag{4-10}$$

By substituting (4-10) into the Lemma 4.8 with $l=1$, we have

$$\Big(\frac{24}{S}+\frac{18}{S^2}\Big)\lambda_1^4-\Big(25+\frac{21}{S}\Big)\lambda_1^2+7S+9<0. \tag{4-11}$$

We notice that the left-hand side of (4-11) is an increasing function of λ_1^2 for $\lambda_1^2>\frac{1}{6}S$. Substituting $\lambda_1^2=\frac{1}{6}S$ into (4-11), we have

$$S<-\tfrac{12}{7}.$$

It is a contradiction.

If $y<0$, by taking $l=2$ in the Lemma 4.8, we have

$$\frac{3}{S}y\big(\lambda_2^2-\tfrac{2}{3}S\big)+\lambda_2\Big(\frac{9}{S}\lambda_2^4-\tfrac{15}{2}\lambda_2^2+2S+3\Big)\le 0. \tag{4-12}$$

Because of

$$y=\Big(\tfrac{1}{3}+\frac{1}{S}\Big)f_3-\Big[\frac{f_3^2}{S^2}\big(\tfrac{19}{9}S^2+\tfrac{8}{3}S+1\big)+\tfrac{7}{9}S(S+6)\big(S+\tfrac{15}{7}\big)\Big]^{\frac{1}{2}},$$

and since the left-hand side of (4-12) is an increasing function of λ_2 for $0>\lambda_2>-\sqrt{\frac{1}{6}S}$, substituting $\lambda_2=-\sqrt{\frac{1}{6}S}$ into (4-12), we have

$$\text{LHS of (4-12)} > 0.$$

It is a contradiction. □

References

[1] S. Chang, "On minimal hypersurfaces with constant scalar curvatures in S^4", *J. Differential Geom.* **37**:3 (1993), 523–534. MR

[2] Q. M. Cheng and Q. R. Wan, "Complete hypersurfaces of $\mathbb{R}^4$ with constant mean curvature", *Monatsh. Math.* **118**:3-4 (1994), 171–204. MR

[3] S. S. Chern, M. do Carmo, and S. Kobayashi, "Minimal submanifolds of a sphere with second fundamental form of constant length", pp. 59–75 in *Functional Analysis and Related Fields* (Chicago, IL, 1968), edited by F. E. Browder and M. Stone, Springer, 1970. MR

[4] H. B. Lawson, Jr., "Local rigidity theorems for minimal hypersurfaces", *Ann. of Math.* (2) **89** (1969), 187–197. MR

[5] H. Omori, "Isometric immersions of Riemannian manifolds", *J. Math. Soc. Japan* **19** (1967), 205–214. MR

[6] C.-K. Peng and C.-L. Terng, "Minimal hypersurfaces of spheres with constant scalar curvature", pp. 177–198 in *Seminar on minimal submanifolds*, edited by E. Bombieri, Ann. of Math. Stud. **103**, Princeton Univ. Press, 1983. MR

[7] J. Simons, "Minimal varieties in riemannian manifolds", *Ann. of Math.* (2) **88** (1968), 62–105. MR

[8] S. T. Yau, "Harmonic functions on complete Riemannian manifolds", *Comm. Pure Appl. Math.* **28** (1975), 201–228. MR

Received August 4, 2024. Revised April 23, 2025.

QING-MING CHENG
MATHEMATICAL SCIENCE RESEARCH CENTER
CHONGQING UNIVERSITY OF TECHNOLOGY
CHONGQING
CHINA
and
OSAKA CENTER ADVANCED MATHEMATICAL INSTITUTE
OSAKA METROPOLITAN UNIVERSITY
OSAKA
JAPAN
chengqingming@cqut.edu.cn

YEJUAN PENG
SCHOOL OF MATHEMATICS AND STATISTICS
HENAN NORMAL UNIVERSITY
XINXIANG
CHINA
yejuan666@126.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 338, No. 2, 2025
https://doi.org/10.2140/pjm.2025.338.267

THE RECIPROCAL COMPLEMENT OF A POLYNOMIAL RING IN SEVERAL VARIABLES OVER A FIELD

NEIL EPSTEIN, LORENZO GUERRIERI AND K. ALAN LOPER

The *reciprocal complement* $R(D)$ of an integral domain D is the subring of its fraction field generated by the reciprocals of its nonzero elements. Many properties of $R(D)$ are determined when D is a polynomial ring in $n \geq 2$ variables over a field. In particular, $R(D)$ is an n-dimensional, local, non-Noetherian, non-integrally closed, non-factorial, atomic G-domain, with infinitely many prime ideals at each height other than 0 and n.

1. Introduction

Let D be an integral domain with fraction field F. What can we say about the subring $R(D)$ of F generated by the reciprocals of all the nonzero elements of D (called the *reciprocal complement*, or *ring of reciprocals*, of D)?

Simple as the above question is, it appears to be a new one, and as we will see in this paper, the answer can be both surprising and satisfying. The question arises naturally in the study of *Egyptian domains*, which extends the notion of Egyptian fractions from the integers to arbitrary integral domains. This study was initiated in [Guerrieri et al. 2024] and continued in [Epstein 2024a]. Recall [Guerrieri et al. 2024] that an integral domain D is *Egyptian* if every element of its fraction field F can be written as a sum of (resp., of distinct) reciprocals of elements of D (in general, such an element of F is called *Egyptian* or *D-Egyptian*). From the viewpoint of the above question then, D is Egyptian if and only if $F = R(D)$. So the distinction between $R(D)$ and F can be seen as a measure of how far an integral domain is from being Egyptian.

The idea of Egyptian domains ultimately comes from the way the ancient Egyptians represented fractions. Namely, they represented an element of $\mathbb{Q} \cap (0, 1)$ as a sum of reciprocals of (distinct) positive integers (so-called *unit fractions*). More than eight centuries ago, Fibonacci [Dunton and Grimm 1966] showed that this is always

Lorenzo Guerrieri is supported by the grant MAESTRO NCN-UMO-2019/34/A/ST1/00263 (Research in commutative algebra and representation theory) and the grant NAWA POWROTY-PPN/PPO/2018/1/00013/U/00001 (Applications of lie algebras to commutative algebra).
MSC2020: primary 13F99; secondary 13A18, 13F20, 13G05.
Keywords: reciprocal complement, Egyptian fraction, polynomial rings.

possible. However, it is far from unique. There are always infinitely many ways to represent a positive rational number as a sum of distinct unit fractions. For the past century or so, number theorists have taken up questions of the diversity of ways to represent fractions as unit fractions. Indeed, such questions can always be rephrased as diophantine equations. Most prominently, the Erdős–Straus conjecture posits that for any $n \geq 5$, the number $4/n$ can be written as a sum of at most three unit fractions.

In addition to its connection with Egyptian fractions, the setting of Egyptian domains and reciprocal complements also isolates natural properties of affine semigroups, rings, and even of algebraic varieties. For instance, it can distinguish whether a subsemigroup Λ of $\mathbb{Q}^n$ is a group, in the following sense. Let D be an Egyptian domain (e.g., $\mathbb{Z}$, or any field). Then $D[\Lambda]$ is Egyptian if and only if Λ is a group (see [Guerrieri et al. 2024, Proposition 3] for "if" and [Epstein 2024a, Theorem 2.6] for "only if"). On the other hand, any *local* domain is Egyptian [Guerrieri et al. 2024, Example 3] and in an affine-local sense, any domain that is finitely generated over a field is *locally* Egyptian [Epstein 2024a, Corollary 3.11], even though $k[x]$ is not Egyptian. Thus, the Egyptian property is an essentially global property that cannot be checked locally, unlike many ring-theoretic properties. Passing to algebraic varieties, Dario Spirito [2025, Theorem 2.1] has shown that when k is an algebraically closed field, and D is a one-dimensional finitely generated k-algebra that is a domain, then D is Egyptian if and only if there is some realization $X \subseteq \mathbb{A}^n_k$ of D that is regular at ∞ such that $|\overline{X} \setminus X| \neq 1$, where $\overline{X}$ is the projective closure of X in $\mathbb{P}^n_k$. Otherwise, if $\{p\} = \overline{X} \setminus X$, he shows that the reciprocal complement of D is isomorphic to $\mathcal{O}_{\overline{X},\{p\}}$.

The first-named author called an integral domain *Bonaccian* if for any nonzero $f \in F$, either f or $1/f$ can be written as a sum of reciprocals from D. Equivalently, $R(D)$ is a valuation domain. He then showed that a Euclidean domain is always Bonaccian [Epstein 2024c], and indeed $R(D)$ is either a DVR or a field. In particular, he [Epstein 2024b] showed that the reciprocal complement of $K[X]$ (K a field, X an indeterminate) is $K[T]_{(T)}$, where $T = 1/X$.

In the current paper, we show that no such thing is true for the reciprocal complement of a polynomial ring in two or more variables over a field. Indeed, let $D = K[X_1, \dots, X_n]$ and $R = R(D)$. Then R has many properties like those of D, but also many interesting, even exotic features. Our main results include the following:

- Any prime ideal of R is generated by elements of the form $1/f$, where $f \in D \setminus K$. This is a special case of a result that holds in any reciprocal complement (see Proposition 2.8).
- R is a local ring whose unique maximal ideal is generated by all elements of the form $1/f$ for $f \in D \setminus K$ (see Theorem 3.3). This is a special case of a result that holds in any reciprocal complement (see Theorem 2.4).

- For every $1 \le i \le n-1$, R has infinitely many primes of height i (see Theorem 6.6).
- $\dim R = n$ (see Theorem 4.4).
- R is atomic (see Theorem 3.11).
- For every $j \le n$, there is a prime ideal $\mathfrak{p} \in \operatorname{Spec} R$ such that $R_{\mathfrak{p}}$ is isomorphic to the reciprocal complement of a polynomial ring in j variables over a field (see Proposition 4.2).
- R is a *G-domain*. In fact, $R\left[\prod_{i=1}^{n} X_i\right] = \operatorname{Frac} R$ (see Proposition 2.9).
- For certain height-one primes $\mathfrak{p}$, we have that $R_{\mathfrak{p}}$ is a DVR (see Lemma 3.6).
- If $n \ge 2$, R is not Noetherian. In fact it is not even coherent (see Corollary 5.7).
- If $n \ge 2$, R is not integrally closed (see Theorem 5.8).
- When $n \le 2$ and $\operatorname{ht} \mathfrak{p} = 1$, $R_{\mathfrak{p}}$ is always Noetherian (see Theorem 7.3).
- When $n = 2$, any finitely generated ideal is contained in all but finitely many prime ideals (see Theorem 7.5). Since all but two prime ideals of R have height one, this behavior can be seen as an extreme version of Krull's principal ideal theorem.

A key to our results has been a change in perspective, wherein one uses the K-automorphism σ of the field $K(X_1, \dots, X_n)$ that sends $X_i \mapsto 1/X_i$. Then $R^* := \sigma(R(D))$ contains D as a subring, even though R^* and R are isomorphic as rings. The effect of this on reciprocals of explicit polynomials is captured in Lemma 3.5. We will sometimes use the R^* point of view to analyze R, and sometimes the R point of view.

We also use valuations in a variety of ways to control the behavior of prime ideals in R.

Prior to this paper, there were four standard constructions that generally lead to non-Noetherian rings in ways that are essentially different from one another:

- polynomial rings in infinitely many variables over a field or $\mathbb{Z}$ (and their quotients),
- putting a valuation on a field with value group not isomorphic to $\mathbb{Z}$ and extracting the valuation ring,
- pullbacks, and
- rings of integer-valued polynomials.

Now we know there is a fifth such construction: the reciprocal complement. The fact that the reciprocal complement of such a well-behaved ring as $K[X_1, \dots, X_d]$ for $d > 1$ is non-Noetherian, not integrally closed, and so forth (see above) indicates that these properties probably also fail for reciprocal complements of many other otherwise well-behaved integral domains. This provides a fertile ground for and source of problems for factorization theory and other investigations in non-Noetherian commutative algebra, being so different (as seen in the list of properties above) from valuation rings, pullbacks, integer-valued polynomial rings, and infinite-dimensional polynomial rings.

2. General properties

In this section, we determine some properties of the reciprocal complement $R = R(T)$ of any integral domain T. In particular, we show that R is always local (see Theorem 2.4), that the prime ideals of R are generated by reciprocals of elements of T (see Proposition 2.8), and that if T is finitely generated over a field, then the fraction field of R is a finitely generated R-algebra (see Proposition 2.9).

Definition 2.1. For any integral domain T, we let $R(T)$ be the *reciprocal complement* of T. That is, if $F(T)$ is the fraction field of T, then $R(T)$ is the subring of $F(T)$ generated by all terms of the form $1/f$, where $f \in T \setminus \{0\}$. Equivalently, $R(T)$ is the set of all finite sums $1/f_1 + \cdots + 1/f_t$, where $t \geq 0$ and each f_i is an element of $T \setminus \{0\}$.

Let T be an integral domain. Let E be the set of Egyptian elements of T, and set $G := E \cup \{0\}$. Recall [Guerrieri et al. 2024, Proposition 8(3), where G is called E_3] that G is a subring of T. Since it must then be an integral domain, $E = G \setminus \{0\}$ is a multiplicatively closed subset of T.

Proposition 2.2. *Let T, E, G be as above. Then*

(1) $R(E^{-1}T) = R(T)$, *and*

(2) *the set of Egyptian elements of $E^{-1}T$, along with* 0, *coincides with the fraction field of G.*

Proof. To prove (1), note that taking the reciprocal complement preserves inclusion; hence, $R(E^{-1}T) \supseteq R(T)$. Conversely, let $y \in R(E^{-1}T)$. By definition there exists $d_1, \ldots, d_n \in T$ and $e_1, \ldots, e_n \in E$ such that

$$y = \frac{e_1}{d_1} + \cdots + \frac{e_n}{d_n}.$$

But each e_i is an Egyptian element of T, and thus we can write $e_i = \sum_{j=1}^{n_i}(1/d_{i_j})$. It follows that

$$y = \sum_{i=1}^{n} \frac{1}{d_i}\left(\sum_{j=1}^{n_i} \frac{1}{d_{i_j}}\right) \in R(T).$$

To prove (2), let K the fraction field of G. Pick an $(E^{-1}T)$-Egyptian element x of $E^{-1}T$. Hence, $x \in R(E^{-1}T) = R(T)$. We can write $x = d/e$ with $d \in T$ and

$e \in E = G \setminus \{0\}$. But $e \in R(T)$ and therefore $d = ex \in R(T)$. It follows that $d \in R(T) \cap T = G$. Thus $x = d/e \in K$.

Conversely, let $0 \neq x \in K$. Then $x = g/h$, where $g, h \in E$. Write $g = \sum_{i=1}^{s}(1/d_i)$, $d_i \in T$. Then $x = g/h = \sum_{i=1}^{s} 1/(d_i/g)$. Since each d_i/g is an element of $E^{-1}T$, we have $x \in R(E^{-1}T)$. Since $g \in T$ and $h \in E$, we have $x \in E^{-1}T$. Hence x is an Egyptian element of $E^{-1}T$. □

Lemma 2.3. *Let K be a field. Let T be a K-algebra all of whose Egyptian elements are in K. Let $x_1, \dots, x_n \in T \setminus K$ and $u \in K \setminus \{0\}$. Then*

$$y = u + \frac{1}{x_1} + \cdots + \frac{1}{x_n}$$

is a unit in $R := R(T)$.

Proof. We can reduce to the case $u = 1$ by dividing all the x_i's by u.

If $n = 0$, the statement is vacuously true. So we may assume $n \geq 1$ and work by induction on n. Set $\alpha_i := x_i^{-1}$ for each $1 \leq i \leq n$.

If $y = 0$, then $\alpha_n = -\left(1 + \sum_{i=1}^{n-1} \alpha_i\right) \in U(R)$ by the inductive hypothesis. Thus, $x_n = \alpha_n^{-1} \in R$, so x_n is an Egyptian element of T, whence $x_n \in K$ by the assumptions on T. But that contradicts the assumption on x_n that $x_n \notin K$. Hence, $y \neq 0$.

Next, notice that

$$H_n := \frac{\prod_{i=1}^{n} \alpha_i}{1 + \alpha_1 + \cdots + \alpha_n} = \frac{1}{\prod_{i=1}^{n} x_i + \sum_{i=1}^{n}\left(\prod_{j \neq i} x_i\right)} \in R(T).$$

Starting from this fact we prove by reverse induction that

$$H_k := \frac{\prod_{i=1}^{k} \alpha_i}{1 + \alpha_1 + \cdots + \alpha_n} \in R(T)$$

also for every $0 \leq k \leq n$. Suppose that $H_{k+1} \in R(T)$. Then notice that

$$\prod_{i=1}^{k} \alpha_i - H_{k+1} = \frac{\left(\prod_{i=1}^{k} \alpha_i\right)\left(1 + \sum_{i \neq k+1} \alpha_i\right)}{1 + \alpha_1 + \cdots + \alpha_n} = \left(1 + \sum_{i \neq k+1} \alpha_i\right) H_k.$$

Since by the original inductive hypothesis $1 + \sum_{i \neq k+1} \alpha_i$ is a unit in $R(T)$, we get $H_k \in R(T)$. In particular, $H_0 \in R(T)$, as was to be shown. □

Theorem 2.4. *Let T be an integral domain. Then $R(T)$ is a local ring, with maximal ideal generated by all elements of the form $1/x$, with $0 \neq x \in T$ not an Egyptian element.*

Proof. First assume that all the Egyptian elements of T are in a subring K of T that is a field. Let $\mathfrak{m}$ be the set of finite sums of elements of the form $1/x$, where $x \in T \setminus K$. Since any multiple of a nonunit of T is a nonunit of T, and all units of T are in K, it follows that $\mathfrak{m}$ is an ideal of $R(T)$. Moreover, any element of $R(T) \setminus \mathfrak{m}$ is a unit of $R(T)$ by Lemma 2.3. Thus it suffices to show that $\mathfrak{m}$ does not contain a unit of $R(T)$.

Let $\alpha \in \mathfrak{m}$. Write $\alpha = 1/x_1 + \cdots + 1/x_n$, with each x_i in $T \setminus K$. We proceed by induction on n to show that α is not a unit. If $n = 0$ (so $\alpha = 0$) the claim is vacuously true. If $n > 0$ and α is a unit of $R(T)$, then $\alpha^{-1} \in R(T)$. We have

$$x_1 = (1/x_1)^{-1} = \frac{1}{\alpha - \sum_{i=2}^{n}(1/x_i)} = \frac{\alpha^{-1}}{1 - \alpha^{-1}\sum_{i=2}^{n}(1/x_i)}. \tag{1}$$

But since $\alpha^{-1} \in R(T)$, $\sum_{i=2}^{n}(1/x_i) \in \mathfrak{m}$ by the inductive hypothesis, and $\mathfrak{m}$ is an ideal, we have $-\alpha^{-1}\sum_{i=2}^{n}(1/x_i) \in \mathfrak{m}$. By Lemma 2.3, it follows that the denominator of (1) is a unit. Hence, $x_1 \in R(T)$, so that $1/x_1 \in U(R(T)) \subseteq K$, contradicting the fact that $x_1 \notin K$.

Finally, we drop the assumption on T. Let E be the set of Egyptian elements of T. Then by Proposition 2.2, $R(T) = R(E^{-1}T)$, and $E^{-1}T$ has all its Egyptian elements in the subfield $E^{-1}G$, where G is the subring $E \cup \{0\}$ of T. Then by the first part of the proof, $R(T)$ is a local ring whose maximal ideal $\mathfrak{m}$ is generated by all elements of the form $1/(x/e)$, where $x \in T \setminus \{0\}$, $e \in E$, and $x/e \notin E^{-1}G$. First note that since every $e \in E$ is a unit of $R(E)$, it follows that $\mathfrak{m}$ is generated by those elements $1/x$ where $x \in T \setminus \{0\}$ and $x \notin E^{-1}G$. But since $E^{-1}G \cap T = G$, the result follows. □

We will see shortly that every prime ideal of $R(T)$ shares the property with $\mathfrak{m}$ that it is generated by reciprocals of elements of T. First, we need the following notion of length.

Definition 2.5. For $\alpha \in R(T)$, the *T-length* of α, denoted by $\ell_T(\alpha)$ (or the *length of* α, denoted by $\ell(\alpha)$, if the ring is understood) is the minimum number t such that there exist $f_1, \ldots, f_t \in T$ such that $\alpha = 1/f_1 + \cdots + 1/f_t$. For α in the fraction field of T but not in $R(T)$, we write $\ell_T(\alpha) = \infty$.

Lemma 2.6. *Let T be an integral domain and $0 \neq \alpha \in R(T)$. Write $\alpha = \sum_{i=1}^{t}(1/f_i)$, where $t = \ell_T(\alpha)$ and each f_i is an element of $T \setminus \{0\}$. Then, in $R(T)$, α is a factor of the product of all the elements $1/f_i$.*

Proof. Set $F := \prod_{i=1}^{t} f_i$. Since $F/f_i \in T$ for each i, we have $F\alpha \in T$, and since both F and α are nonzero elements of the fraction field of T, we have $F\alpha \neq 0$. Hence $1/(F\alpha) \in R(T)$. Then the equation $1/F = 1/(F\alpha) \cdot \alpha$ finishes the proof. □

Lemma 2.7. *Let T be an integral domain and $0 \neq \alpha \in R(T)$. Let $t = \ell_T(\alpha)$ and write $\alpha = \sum_{i=1}^{t} 1/f_i$ with $f_i \in T \setminus \{0\}$. Then $\ell_T(\alpha - (1/f_t)) = t - 1$.*

Proof. Write $\beta = \alpha - 1/f_t$. Since $\beta = \sum_{i=1}^{t-1}(1/f_i)$, we have $\ell_T(\beta) \leq t - 1$. Write $\beta = \sum_{j=1}^{s}(1/g_j)$, where $s = \ell_T(\beta)$ and each g_j is an element of $T \setminus \{0\}$. Then $\alpha = (1/f_t) + \sum_{j=1}^{s}(1/g_j)$. Thus,

$$t = \ell_T(\alpha) \leq s + 1 \leq (t-1) + 1 = t.$$

Hence, $s + 1 = t$, as was to be shown. □

As a consequence of the above two lemmas, we obtain the following result about the generators of any prime ideal of $R(T)$, which recapitulates the fact about the maximal ideal of $R(T)$ given in Theorem 2.4.

Proposition 2.8. *Any prime ideal of $R(T)$ is generated by elements of the form $1/f$, where $f \in T$.*

Proof. Let $0 \neq \alpha \in \mathfrak{p}$ and $t = \ell_T(\alpha)$. We proceed by induction on t to show that α is a sum of elements of $\mathfrak{p}$ of the form $1/f$, where $f \in T$.

When $t = 1$, it is clear. Suppose $t > 1$. Write $\alpha = \sum_{i=1}^{t}(1/f_i)$, where $f_i \in T$. By Lemma 2.6, $\prod_{i=1}^{t}(1/f_i)$ is a multiple of $\alpha \in R(D)$. Hence, $\prod_{i=1}^{t}(1/f_i) \in \mathfrak{p}$. Since $\mathfrak{p}$ is prime, it follows that $1/f_i \in \mathfrak{p}$ for some $1 \leq i \leq t$. Let $\beta = \alpha - 1/f_i$. Clearly $\beta \in \mathfrak{p}$, and by Lemma 2.7, $\ell_T(\beta) = t - 1$, so by the inductive hypothesis, β is a sum of elements of the form $1/g \in \mathfrak{p}$, with $g \in T$. Thus, $\alpha = \beta + 1/f_i$ is also such a sum. □

We culminate this section with a result on reciprocal complements of finitely generated K-algebras.

Proposition 2.9. *Let L/K be a field extension, let $f_1, \ldots, f_n \in L$, and let $T = K[f_1, \ldots, f_n]$. Then $R(T)\big[\prod_{i=1}^{n} f_i\big] = \operatorname{Frac} T$. Hence $1/\prod_{i=1}^{n} f_i \in \mathfrak{p}$ for every nonzero prime ideal $\mathfrak{p}$ of $R(T)$.*

Proof. Write $g = \prod_{i=1}^{n} f_i$. First, note that since $g \in T$, we have $1/g \in R(T)$.

We have

$$f_1 = \frac{g}{\prod_{i=2}^{n} f_i} = g \cdot \frac{1}{\prod_{i=2}^{n} f_i} \in R(T)[g],$$

since $1/\big(\prod_{i=2}^{n} f_i\big) \in R(T)$. By symmetry, we have $f_1, \ldots, f_n \in R(T)[g]$. Obviously $K \subseteq R(T)$ as well, so $T = K[f_1, \ldots, f_n] \subseteq R(T)[g]$. Let $\alpha \in \operatorname{Frac} T$. We may write $\alpha = u/v$ with $u, v \in T$ and $v \neq 0$. Then, since $u \in T \subseteq R(T)[g]$ and $1/v \in R(T) \subseteq R(T)[g]$, we have $\alpha = u \cdot (1/v) \in R(T)[g]$. Thus, $\operatorname{Frac} T \subseteq R(T)[g]$, but the reverse containment is obvious, so $R(T)[g] = \operatorname{Frac} T$.

The final statement follows from [Kaplansky 1970, Theorem 19]. □

Remark 2.10. Recall that a *G-domain* is an integral domain whose fraction field is a finitely generated algebra over it [Kaplansky 1970, Definition following Theorem 18]. Hence, Proposition 2.9 implies that for any integral domain T that is finitely generated over a field, $R(T)$ is a G-domain.

3. Properties and bounds on the ring of polynomial reciprocals

In this section, we give bounds on the reciprocal complement of a polynomial ring in n variables. That is, we exhibit rings that it is contained in and rings that it contains. We also show it is atomic, but fails unique factorization. A main tool is the map σ, an involution on $K(X_1, \ldots, X_n)$, which makes our ring isomorphic to an overring of $K[X_1, \ldots, X_n]$.

Notation 3.1. Let $D = D_n = K[X_1, \ldots, X_n]$, the polynomial ring in n variables over a field K, where $n \geq 1$. Let $F = F_n$ the fraction field of D_n. That is, $F_n = K(X_1, \ldots, X_n)$. We set $R := R_n = R(D_n)$.

We let $\sigma = \sigma_n : F_n \to F_n$ be the unique K-algebra homomorphism that sends $X_i \mapsto 1/X_i$ for $1 \leq i \leq n$. Note that $\sigma \circ \sigma = 1_F$; hence σ is an *involution*, whence a K-automorphism of F. For any subring T of F, we set $T^* := \sigma(T)$.

We define $2n$ functions $t_i, a_i : D \setminus \{0\} \to \mathbb{N}_0$ for $1 \leq i \leq n$ as follows. We set $t_i(f) = c$ if $X_i^c \mid f$ but $X_i^{c+1} \nmid f$, and we let $a_i(f) = \deg_{X_i}(f) - t_i(f)$, where $\deg_{X_i}(f)$ is equal to the degree of f as a polynomial in X_i with coefficients in $K[X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n]$. Then, for any $f \in D \setminus \{0\}$, we write $f = \left(\prod_{i=1}^n X_i^{t_i(f)}\right) f_0$, where $f_0 \in D \setminus \bigcup_{i=1}^n X_i D$.

For any n-tuple $(u_1, \ldots, u_n) \in \mathbb{Z}^n$, write $\boldsymbol{u} := (u_1, \ldots, u_n)$ and $\boldsymbol{X}^{\boldsymbol{u}} := \prod_{i=1}^n X_i^{u_i}$.

Our first goal will be to prove that the maximal ideal is generated by the reciprocals of the nonconstant polynomials.

Lemma 3.2. *Let T be an $\mathbb{N}$-graded integral domain. Then all the Egyptian elements of T are in T_0, its 0-th graded component.*

Proof. For $f \in T$, we let its *degree* be the degree of its largest nonzero graded component. Note that degree is then additive in T; that is, if $f, g \in T \setminus \{0\}$, then $\deg(fg) = \deg(f) + \deg(g)$. Also, if $f + g \neq 0$, then $\deg(f + g) \leq \max\{\deg f, \deg g\}$.

With this in mind, let $f \in T$ be Egyptian. Write

$$f = \frac{1}{f_1} + \cdots + \frac{1}{f_s}$$

with $f_j \in T$ for each j. Clearing denominators by multiplying through by $\prod_{i=1}^s f_i$, we have

$$f f_1 \cdots f_s = \sum_{i=1}^{s} \prod_{j \neq i} f_j.$$

By equating degrees, it follows that

$$\deg(f) + \sum_{i=1}^{s} \deg(f_i) \leq \max_i \sum_{j \neq i} \deg(f_j).$$

Since all degrees are nonnegative, it follows that $\deg f = 0$, so that $f \in T_0$. □

Theorem 3.3. *R (as in Notation 3.1) is a local ring, with maximal ideal generated by all elements of the form $1/f$, with f a nonconstant polynomial.*

Proof. By Theorem 2.4, $R(D)$ is a local ring with maximal ideal $\mathfrak{m}$ generated by all elements of the form $1/f$ with $0 \neq f \in D$ non-Egyptian. However, by Lemma 3.2 (and using the standard grading on the polynomial ring), no nonconstant polynomial can be Egyptian. Since the nonzero constant polynomials are units and hence Egyptian, the result follows. □

Proposition 3.4. *Let $0 \le j < n$ be integers. Then $R_n \cap F_j = R_j$ and $R_n^* \cap F_j = R_j^*$.*

Proof. We need only prove the first statement, since it then follows that

$$R_n^* \cap F_j = \sigma(R_n) \cap F_j = \sigma(R_n \cap F_j) = \sigma(R_j) = R_j^*.$$

Moreover, by an easy induction, we may assume $j = n-1$.

It is clear that $R_{n-1} \subseteq R_n \cap F_{n-1}$. So let $\alpha \in R_n \cap F_{n-1}$. Then $\alpha = \sum_{i=1}^t (1/f_i)$, where each f_i is an element of $D_n \setminus \{0\}$. Reorder the f_i such that $f_1, \ldots, f_s \in D_{n-1}$ and $f_{s+1}, \ldots, f_t \in D_n \setminus D_{n-1}$. Then for $1 \le i \le s$, we have $1/f_i \in R_{n-1} \subseteq R_n \cap F_{n-1}$. Let $\beta = \alpha - \sum_{i=1}^s (1/f_i)$; then we have

$$\beta = \sum_{i=s+1}^{t} \frac{1}{f_i}. \tag{2}$$

Assume $\beta \neq 0$. Then multiplying (2) by $\prod_{i=s+1}^t f_i$, we have

$$\beta f_{s+1} \cdots f_t = \sum_{i=s+1}^{t} \prod_{\substack{j \neq i \\ j > s}} f_j.$$

With respect to the polynomial ring $F_{n-1}[X_n]$, note that $\beta \in F_{n-1}$ and each f_i for $i > s$ is a nonconstant polynomial. Say $\deg f_i = d_i > 0$ for each $i > s$. Then the left hand side above has degree $\sum_{i=s+1}^t d_i$, whereas the right hand side has degree $\le \max_i \left\{\sum_{j \neq i,\ j > s} d_j\right\} < \sum_{s=i+1}^t d_i$, a contradiction. Hence $\beta = 0$. That is, $\alpha = \sum_{i=1}^t (1/f_i) \in R_{n-1}$, since each f_i is an element of D_{n-1}. □

Lemma 3.5. *Let $f \in D \setminus \{0\}$. Then*

$$\sigma\left(\frac{1}{f}\right) = \frac{\boldsymbol{X}^{\boldsymbol{a}(f)+\boldsymbol{t}(f)}}{f^*},$$

where $f^* \in D \setminus \bigcup_{i=1}^n X_i D$. Moreover, $\boldsymbol{a}(f) = \boldsymbol{a}(f^*)$ and $f = \boldsymbol{X}^{\boldsymbol{t}(f)} f^{**}$.

Proof. First suppose $f \in D \setminus \bigcup_{i=1}^n X_i D$, so that $\boldsymbol{t}(f) = \boldsymbol{0}$. Write $f = \sum_{\boldsymbol{j} \in \mathbb{N}_0^n} u_{\boldsymbol{j}} \boldsymbol{X}^{\boldsymbol{j}}$, where each $u_{\boldsymbol{j}}$ is in K and $u_{\boldsymbol{j}} = 0$ for all but finitely many n-tuples $\boldsymbol{j}$. Then $a_i(f) = \deg_{X_i}(f) = \max\{c \mid \exists \boldsymbol{j} \text{ with } j_i = c \text{ and } u_{\boldsymbol{j}} \neq 0\}$ for each $1 \le i \le n$. We have

$$\sigma\left(\frac{1}{f}\right) = \frac{1}{\sum_{\boldsymbol{j}} u_{\boldsymbol{j}} / \boldsymbol{X}^{\boldsymbol{j}}} = \frac{\boldsymbol{X}^{\boldsymbol{a}(f)}}{\sum_{\boldsymbol{j}} u_{\boldsymbol{j}} \boldsymbol{X}^{\boldsymbol{a}(f)-\boldsymbol{j}}}.$$

Let f^* denote the expression in the denominator above. Note that $\boldsymbol{a}(f) - \boldsymbol{j} \in \mathbb{N}_0^n$ whenever $u_{\boldsymbol{j}} \neq 0$, since for each such $\boldsymbol{j}$ we have $j_i \le a_i(f)$ for all $1 \le i \le n$. Hence f^* is a true polynomial. Moreover, for each i, since $X_i \nmid f$, there is some $\boldsymbol{j}$ with $u_{\boldsymbol{j}} \neq 0$ and $j_i = 0$, and hence $a_i(f) - j_i = a_i(f)$. Thus, $a_i(f^*) = a_i(f)$. Finally, for each i, there is some $\boldsymbol{j}$ with $u_{\boldsymbol{j}} \neq 0$ and $j_i = a_i(f)$. Hence $a_i(f) - j_i = 0$, so $X_i \nmid f^*$, whence $f_i^* \in D \setminus \bigcup_{i=1}^n X_i D$.

For the final claim, we have

$$\sigma\left(\frac{1}{f^*}\right)=\frac{\boldsymbol{X}^{\boldsymbol{a}(f^*)}}{\sum_{\boldsymbol{j}} u_{\boldsymbol{j}} \boldsymbol{X}^{\boldsymbol{a}(f^*)-(\boldsymbol{a}(f)-\boldsymbol{j})}}=\frac{\boldsymbol{X}^{\boldsymbol{a}(f^*)}}{\sum_{\boldsymbol{j}} u_{\boldsymbol{j}} \boldsymbol{X}^{\boldsymbol{j}}}=\frac{\boldsymbol{X}^{\boldsymbol{a}(f^*)}}{f}.$$

Now we go to the general case, where $\boldsymbol{t}(f)$ is not necessarily the zero vector. We have $f=\boldsymbol{X}^{\boldsymbol{t}(f)} f_0$, where $f_0 \in D \setminus \bigcup_{i=1}^n X_i D$. Then

$$\sigma\left(\frac{1}{f}\right)=\left(\prod_{i=1}^n \sigma\left(\frac{1}{X_i}\right)^{t_i(f)}\right) \sigma\left(\frac{1}{f_0}\right)=\boldsymbol{X}^{\boldsymbol{t}(f)} \sigma\left(\frac{1}{f_0}\right)=\frac{\boldsymbol{X}^{\boldsymbol{a}(f_0)+\boldsymbol{t}(f)}}{f_0^*}.$$

Moreover, $a_i(f)=\deg_{X_i}(f)-t_i(f)=\deg_{X_i}(f_0)=a_i(f_0)$ for each $1 \leq i \leq n$, so $\boldsymbol{a}(f)=\boldsymbol{a}(f_0)$. Setting $f^*=f_0^*$, we have $f^{**}=f_0^{**}=f_0$, so that $f=\boldsymbol{X}^{\boldsymbol{t}(f)} f_0= \boldsymbol{X}^{\boldsymbol{t}(f)} f_0^{**}$, completing the proof. □

Lemma 3.6. *We have $R^* \subseteq D_{(X_i)}$ for each $1 \leq i \leq n$. In particular there are n distinct height-one prime ideals $\mathfrak{p}_i$ of R^* obtained as centers of the X_i-adic valuations of D, and $R^*_{\mathfrak{p}_i}=D_{(X_i)}$.*

Proof. Choose i with $1 \leq i \leq n$. Let $f \in D$. Let v_i be the X_i-adic valuation function. Then by Lemma 3.5, $v_i(\sigma(1/f))=t_i(f)+a_i(f)-v_i(f^*) \geq 0$, as $a_i(f)=\deg_{X_i}(f^*)$, and $v_i(f^*)$ cannot exceed the X_i-degree of f^*. Since every nonzero element $\alpha \in R^*$ is a sum of terms of the form $\sigma(1/f)$, it follows that $v_i(R^*) \geq 0$.

Now let $1 \leq i<j \leq n$, and let $\mathfrak{p}_i$ and $\mathfrak{p}_j$ be the centers of the X_i- and X_j-adic valuations on D in R^*, respectively. Since $v_i(X_i)=1$ but $v_j(X_i)=0$, we have $X_i \in \mathfrak{p}_i \setminus \mathfrak{p}_j$. Similarly, $v_i(X_j)=0$ and $v_j(X_j)=1$, so $X_j \in \mathfrak{p}_j \setminus \mathfrak{p}_i$.

For the final claim, let $\alpha \in D_{(X_i)}$. Then $\alpha=f/g$ for some $f \in D$ and $g \in D \setminus X_i D$. If $g \in \mathfrak{p}_i$, then $g \in X_i D_{(X_i)} \cap D=X_i D$, which is a contradiction. Hence, $f \in R^*$ and $g \in R^* \setminus \mathfrak{p}_i$, so $f/g \in R^*_{\mathfrak{p}_i}$. Thus, $R^*_{\mathfrak{p}_i}=D_{(X_i)}$, whence $\operatorname{ht} \mathfrak{p}_i=1$. □

Lemma 3.7. *Let $\mathfrak{p}$ be a nonzero prime ideal of R. Then $1/X_i \in \mathfrak{p}$ for some $1 \leq i \leq n$.*

Proof. By Proposition 2.9, $1/\left(\prod_{i=1}^n X_i\right)=\prod_{i=1}^n(1/X_i) \in \mathfrak{p}$. Since $\mathfrak{p}$ is prime, some $1/X_i$ is in $\mathfrak{p}$. □

Lemma 3.8. *Let $f \in D$. If $f(\mathbf{0}) \neq 0$ then f is a unit in R^*. In particular,*

$$K[X_1, \ldots, X_n]_{(X_1, \ldots, X_n)} \subseteq R^*.$$

Proof. We prove this by induction on n. If $n=0$, then the result is vacuous. Thus, let $n \geq 1$ and assume the result true for smaller n.

By way of contradiction suppose that $f \in \mathfrak{p}$ for a prime ideal $\mathfrak{p}$ of R^*. By Lemma 3.7, some X_i is in $\mathfrak{p}$; without loss of generality assume $i=n$, so that $X_n \in \mathfrak{p}$. Then $f=X_n g+h$ for some $g \in D_n$ and $h \in D_{n-1} \setminus (X_1, \ldots, X_{n-1}) D_{n-1}$. Then $h \in \mathfrak{p} \cap D_{n-1} \setminus (X_1, \ldots, X_{n-1}) D_{n-1}$, so that, by the inductive hypothesis,

$1/h \in K[X_1, \dots, X_{n-1}]_{(X_1,\dots,X_{n-1})} \subseteq R^*_{n-1}$. Hence also $1/h \in R^*$. But then $1 = (1/h)(f - X_n g) \in \mathfrak{p}$, a contradiction. □

Lemma 3.9. *Let w be the **order valuation** of $D_{(X_1,\dots,X_n)}$ on F, i.e., the unique valuation on F such that for any nonzero $g \in D$, $w(g) = \max\{j \mid g \in (X_1, \dots, X_n)^j\}$. Let $(W, \mathfrak{m}_W)$ be the corresponding DVR. Then $R^* \subseteq W$, and $\mathfrak{m}_W \cap R^*$ is the maximal ideal $\mathfrak{m}$ of R^*.*

Proof. By construction $D \subseteq W$ and $X_i \in \mathfrak{m}_W$ for all i, so that $w(X_i) > 0$. We first show that for any nonconstant polynomial f not divisible by any of the variables, $w(\sigma(1/f)) \geq 1$. Under these assumptions, $\sigma(1/f) = \boldsymbol{X}^{\boldsymbol{a}(f)}/f^*$ by Lemma 3.5. Note that there is some i and some monomial m in f^* such that $X_i^{a_i(f)} \nmid m$. Thus, $w(f^*) \leq w(m) \leq (a_i(f) - 1) + \sum_{j \neq i} a_j(f) < \sum_j a_j(f)$. Thus, $w(\sigma(1/f)) = \sum_j a_j(f) - w(f^*) > 0$.

For a general nonconstant polynomial f, we have $f = \boldsymbol{X}^{\boldsymbol{t}(f)} f_0$, where f_0 is not a multiple of any of the X_i, and if f_0 is constant then some $t_i(f) > 0$. Hence, $w(\sigma(1/f)) = \sum_{i=1}^n t_i(f) w(X_i) + w(\sigma(1/f_0)) > 0$.

Now let $\alpha \in R^*$. Write $\alpha = u + \sum_{j=1}^t \sigma(1/f_j)$, where $u \in K$ and each f_j is in $D \setminus K$. Since $K \subseteq D \subseteq W$, by the above we have $\alpha \in W$, whence $R^* \subseteq W$. If $u = 0$ then $w(\alpha) \geq \min\{w(\sigma(1/f_i) \mid 1 \leq i \leq t\} > 0$, so that $\alpha \in \mathfrak{m}_W$. Thus by Theorem 3.3, $\mathfrak{m} \subseteq \mathfrak{m}_W \cap R^*$, but then since $\mathfrak{m}$ is maximal, the result follows. □

The following result must be well known but we provide a proof for the convenience of the reader.

Lemma 3.10. *Let $(T, \mathfrak{m})$ be a local integral domain with fraction field F, and let $(V, \mathfrak{n})$ be a discrete rank-one valuation ring such that $T \subseteq V \subseteq F$ and $\mathfrak{n} \cap T = \mathfrak{m}$. Then T is atomic, and any $x \in \mathfrak{m} \setminus \mathfrak{n}^2$ is an irreducible element of T.*

Proof. We begin with the second statement. Let $x \in \mathfrak{m} \setminus \mathfrak{n}^2$. Write $x = st$ with $s, t \in T$ and s not a unit. Then if v is the valuation function of v, we have $v(s) \geq 1$, so that $1 = v(x) = v(st) = v(s) + v(t) \geq 1 + v(t)$, whence $v(t) = 0$, so that $t \in T \setminus \mathfrak{n} = T \setminus \mathfrak{m}$ and is thus a unit. Thus, x is irreducible.

For the first statement, let $A = \{x \in \mathfrak{m} \mid x$ cannot be written as a product of irreducible elements$\}$. If $A \neq \varnothing$, choose $a \in A$ such that $v(a) \leq v(b)$ for all $b \in A$. Since a is not irreducible, we may write $a = bc$ for some nonunits $b, c \in T$; hence $b, c \in \mathfrak{m}$. But then $v(c) \geq 1$, so $v(b) = v(a) - v(c) < v(a)$. Thus, $b \notin A$, so b can be written as a product of irreducible elements. By the same argument, the same holds for c. Hence, $a = bc$ is also a product of irreducible elements, which is a contradiction. Thus, $A = \varnothing$, so every element of T is either a unit or a product of irreducibles. That is, T is atomic. □

Theorem 3.11. *The ring R is atomic. That is, every nonzero nonunit element factors into a product of irreducible elements.*

Proof. Since $R \cong R^*$, we may work with R^*. By Lemma 3.9, the maximal ideal of R^* is the center of a rank-one discrete valuation. The result then follows from Lemma 3.10. □

Remark 3.12. However, R is not a UFD provided $n \geq 2$. To see this (and working in R^*), use the labels $X := X_1$ and $Y := X_2$, and first note that $s := \sigma(1/(X+Y)) = XY/(X+Y) \in R^*$, and also that $t := X^2/(X+Y) = X - s$ and $u := Y^2/(X+Y) = Y - s \in R^*$. But each of s, t, u has value 1 in the order valuation w from Lemma 3.9, hence must be irreducible elements of R^* by Lemma 3.10.

Then we have $X^2Y^2/(X+Y)^2 = s^2 = tu$, so if R^* were a UFD, s would be associate to either t or u. But if s is an associate of t, then $s \in tR^*$, which implies that $Y/X \in R^*$. And if s is an associate of u, then $u \in sR^*$, which again implies that $Y/X \in R^*$. So in either case we obtain $X/Y = \sigma(Y/X) \in R = R(D) \subseteq R(K(X_3, \ldots, X_n)[X, Y])$, contradicting [Epstein 2024c, Example 2.9].

4. Dimension

In this section, we show that R has the same Krull dimension as D (see Theorem 4.4).

Lemma 4.1. *For any $1 \leq i \leq n$, we have*

$$R[X_i] = R(K(X_i)[X_1, \ldots, \widehat{X}_i, \ldots, X_n]).$$

Hence, $R^*[X_i^{-1}]$ is isomorphic to the reciprocal complement of the polynomial ring in $n-1$ variables over a field.

Proof. Without loss of generality set $i = n$. Set $S := R(K(X_n)[X_1, \ldots, X_{n-1}])$.

For the forward containment, first note that $X_n = 1/(1/X_n) \in S$ since $1/X_n \in K(X_n)[X_1, \ldots, X_{n-1}]$. Moreover, since $D_n \subseteq K(X_n)[X_1, \ldots, X_{n-1}]$ and $R(-)$ preserves containment, we have $R \subseteq S$. Thus, $R[X_n] \subseteq S$.

For the reverse, let $0 \neq f \in K(X_n)[X_1, \ldots, X_{n-1}]$. By finding a common denominator to the $K(X_n)$-coefficients of the monomials in $X_1, \ldots, X_{n-1}$, we may write $f = g/h$, where $g \in D_n$ and $h \in K[X_n]$. Write $h = \sum_{i=0}^{t} c_i X_n^i$ with all $c_i \in K$. Then

$$\frac{1}{f} = \frac{h}{g} = \sum_{i=0}^{t} \frac{c_i}{g} \cdot X_n^i.$$

But for each i, if $c_i = 0$, then $c_i/g = 0 \in R_n$; otherwise $c_i/g = (1/g/c_i) \in R$. Thus, $1/f = \sum_{i=0}^{t} (c_i/g) \cdot X_n^i \in R[X_n]$. Hence, $S \subseteq R[X_n]$, completing the proof. □

Proposition 4.2. *Let j and n be integers with $0 \leq j \leq n$. Then there is a unique prime ideal P of R such that R_P is the reciprocal complement of $L[X_1, \ldots, X_j]$, where $L = K(X_{j+1}, \ldots, X_n)$.*

Proof. By applying induction to Lemma 4.1, we have

$$R[X_{j+1}, \ldots, X_n] = R(L[X_1, \ldots, X_j]) =: S.$$

But, since S is a local ring by Theorem 2.4, and since each X_i^{-1} is an element of R, there is a unique prime ideal P of R maximal with respect to avoiding all of X_i^{-1} for $j+1 \le i \le n$ such that $R_P = S$. □

Lemma 4.3. *For each i, there is a unique prime ideal $\mathfrak{q}_i$ of R maximal with respect to not containing $1/X_i$. We have $R_{\mathfrak{q}_i} = R[X_i]$.*

Proof. We may assume $i = n \ge 1$. By Lemma 4.1, $R[X_n]$ is the reciprocal complement of the polynomial ring in $n-1$ variables over a field, which by Theorem 2.4 is local. Hence by elementary localization theory, there is a unique prime ideal $\mathfrak{q}_n$ of R maximal with respect to avoiding X_n^{-1}, and $R_{\mathfrak{q}_n} = R[X_n] = R(K(X_n)[X_1, \dots, X_{n-1}])$. □

Theorem 4.4. $\dim R = n$.

Proof. We proceed by induction on n. Of course $R_0 = K$, which has dimension 0, so we may assume $n > 0$. By Theorem 3.3, R is local, and its maximal ideal $\mathfrak{m}$ contains the reciprocals of all the variables. Let $\mathfrak{q} = \mathfrak{q}_n$ be as in Lemma 4.3, so that $1/X_n \notin \mathfrak{q}_n$ and $R_\mathfrak{q}$ is isomorphic to the reciprocal complement of a polynomial ring in $n-1$ variables over a field by Lemma 4.1. Then by the inductive hypothesis, $\operatorname{ht}\mathfrak{q} = \dim R_\mathfrak{q} = n-1$. Since $1/X_n \in \mathfrak{m} \setminus \mathfrak{q}$, we have $\operatorname{ht}\mathfrak{m} > \operatorname{ht}\mathfrak{q} = n-1$, whence $\dim R = \operatorname{ht}\mathfrak{m} \ge n$.

On the other hand, since R^* is an overring of the n-dimensional Noetherian domain D, we have $\dim R = \dim R^* \le n$; see [Anderson et al. 1988]. Hence $\dim R = n$. □

5. Exotic properties of R

For most of this section, we work in two variables, so that $D = K[X, Y]$, where $X = X_1$ and $Y = X_2$ for short. Then the notation D_n, R_n, etc. when $n \ne 2$ will stand for the corresponding rings of other dimensions. We show that R_n^* is not integrally closed when $n \ge 2$. We also show that it is not a finite conductor domain, hence not coherent, and is thus also non-Noetherian.

Before we begin, recall the following presumably well-known result:

Lemma 5.1. *Let $A \subseteq B$ be integral domains such that B is free as an A-module. Then $B \cap \operatorname{Frac}(A) = A$.*

Proof. Let $b \in B \cap \operatorname{Frac} A$. Write $b = x/y$ with $x, y \in A$ and $y \ne 0$. Then $x = yb \in yB \cap A = (yA)B \cap A = yA$, where the latter equation holds by freeness. Thus, $x = ya$ for some $a \in A$, whence $yb = ya$, so by cancellation, $b = a \in A$. □

Our methodology here is to construct a family of valuation rings that contain R^*, which serve as a tool to analyze the elements and prime ideals of our ring. We must start with notation that will be useful:

Notation 5.2. Choose two relatively prime positive integers p and q with $p < q$, such that neither p nor q is a multiple of char K. Let K' be the smallest field extension of K that contains all the primitive p-th and q-th roots of 1. Let $L/K'(X, Y)$ be generated by elements s and t such that $s^p = X$ and $t^q = Y$. Note that $K'[s, t]$ is free as a $K'[X, Y]$-module on the basis $\{s^i t^j \mid 0 \le i < p,\ 0 \le j < q\}$.

Lemma 5.3. *Let $g \in K[X, Y]$. Then $s - t \mid g$ in $K'[s, t]$ if and only if $X^q - Y^p \mid g$ in $K[X, Y]$.*

Proof. Since $K'[X, Y] \cap K(X, Y) = K[X, Y]$ by Lemma 5.1, we may assume $K = K'$. Thus, we may let ξ_p (resp. ξ_q) be a primitive p-th (resp. q-th) root of unity in K. Then

$$\prod_{i=0}^{p-1}\prod_{i=0}^{q-1}(\xi_p^i s - \xi_q^j t) = \prod_{i=0}^{p-1}((\xi_p^i s)^q - t^q) = \prod_{k=0}^{p-1}(\xi_p^k s^q - t^q) = (-1)^p \prod_{k=0}^{p-1}(t^q - \xi_p^k s^q)$$
$$= (-1)^p(t^{pq} - s^{pq}) = (-1)^p(Y^p - X^q).$$

The second equality above holds because p and q are relatively prime, so that the order of $q + p\mathbb{Z}$ in $\mathbb{Z}/p\mathbb{Z}$ must be p.

Also note that for each pair (i, j) of integers with $0 \le i < p$ and $0 \le j < q$, there is a unique $\tau_{ij} \in \operatorname{Aut}_{K(X,Y)} L$ such that $\tau_{ij}(s) = \xi_p^i s$ and $\tau_{ij}(t) = \xi_q^j t$. Thus, if $s - t \mid g$ in $K[s, t]$, then for each i and j, we have $\tau_{ij}(s - t) = \xi_p^i s - \xi_q^j t \mid \tau_{ij}(g) = g$. Since the $\xi_p^i s - \xi_q^j t$ are mutually nonassociate irreducible elements of $K[s, t]$, a UFD, it follows that $X^q - Y^p = \pm \prod_{i,j}(\xi_p^i s - \xi_q^j t) \mid g$ in $K[s, t]$. Hence,

$$\frac{g}{X^q - Y^p} \in K(X, Y) \cap K[s, t] = K[X, Y],$$

again by Lemma 5.1, which implies that $X^q - Y^p \mid g$ in $K[X, Y]$.

For the converse, simply note that if $X^q - Y^p \mid g$ in $K[X, Y]$, then as $s - t \mid X^q - Y^p$ in $K[s, t]$ and $K[X, Y] \subset K[s, t]$, it follows by transitivity of divisibility that $s - t \mid g$ in $K[s, t]$. □

Notation 5.4. Let $u := s - t$. Then s and u are algebraically independent over K', and $K'[s, t] = K'[s, u]$. We define a valuation $w := w_h := w_{p,q,h}$ on $K'[s, u]$ by setting $w(s) = 1$ and $w(u) = h$ for some integer $h \ge 1$, and for any nonzero $f = \sum_{i,j} c_{ij} s^i u^j$ in $K'[s, t]$, where $c_{ij} \in K'$, we set

$$w(f) = \min\{w(s^i u^j) \mid c_{ij} \ne 0\} = \min\{i + hj \mid c_{ij} \ne 0\}.$$

Then we let $W := W_h := W_{p,q,h}$ be the corresponding valuation ring in the field L. Clearly $K'[s, u] \subseteq W$. Set $V := W \cap F$ (denoted by V_h or $V_{p,q,h}$ if needed) and let $v = v_h = v_{p,q,h}$ be the corresponding valuation on F.

Lemma 5.5. *The valuation ring V is an overring of $R^* := R_2^*$ if and only if $h \le pq + 1$. If $h < pq + 1$, then $\sigma(1/f) \in \mathfrak{m}_V$ for all $f \in D \setminus K$.*

Suppose on the other hand that $h = pq + 1$. *Then for an irreducible polynomial* $f \in K[X, Y]$, *with* $\alpha = \sigma(1/f^*)$ *and* $\theta = \sigma(1/(X^q - Y^p)) = X^q Y^p/(Y^p - X^q)$, *we have* $v(\alpha) = 0$ *if and only if* $\alpha = \theta^m \delta^{-1}$ *for some* $m \geq 1$ *and some element* $\delta \in K[\theta] \setminus (\theta)K[\theta]$. *Otherwise* $v(\alpha) \geq p$.

Proof. Since $X = X^*$ and $Y = Y^*$, we have $v(\sigma(1/X^*)) = v(X) = w(s^p) = pw(s) = p$, and $w(t) = w(s - u) = 1$, so $v(\sigma(1/Y^*)) = v(Y) = w(t^q) = qw(t) = q$. Hence, V is an overring of $K[X, Y]$, with $(X, Y) \subseteq \mathfrak{m}_V$.

Now suppose (i, j) is a pair of integers with $0 \leq i < p$, $0 \leq j < q$, and $(i, j) \neq (0, 0)$. Then $\xi_p^i s - \xi_q^j t = (\xi_p^i - \xi_q^j)s + \xi_p^j u$, so since $\xi_p^i - \xi_q^j \in K \setminus \{0\}$, we have $w(\xi_p^i s - \xi_q^j t) = 1$. Therefore,

$$\begin{aligned} v(X^q - Y^p) &= w\left(\pm \prod_{i=0}^{p-1} \prod_{i=0}^{q-1} (\xi_p^i s - \xi_q^j t) \right) = w(s - t) + \sum_{i,j}^{(i,j)\neq(0,0)} w(\xi_p^i s - \xi_q^j t) \\ &= h + pq - 1. \end{aligned}$$

It follows that

$$v(\theta) = v\left(\frac{X^q Y^p}{X^q - Y^p} \right) = qp + pq - (h + pq - 1) = pq + 1 - h.$$

Thus, if $h > pq + 1$, we have $v(\theta) < 0$, so that $\theta \notin V$ and $R^* \not\subseteq V$. But as long as $h \leq pq + 1$ we have $v(\theta) \geq 0$, with $v(\theta) > 0 \iff h < pq + 1$. From now on we assume $h \leq pq + 1$.

Now let $f \in K[X, Y]$ be nonconstant, irreducible, and not associate to any of X, Y, $X^q - Y^p$. Then for some $c_{ij} \in K$, we have

$$\begin{aligned} f &= \sum_{i,j} c_{ij} X^i Y^j = \sum_{ij} c_{ij} s^{pi} (s - u)^{qj} \\ &= \sum_{i,j} c_{ij} s^{pi} \sum_{k=0}^{qj} (-1)^k \binom{qj}{k} s^{qj-k} u^k \\ &= \left(\sum_{i,j} c_{ij} s^{pi+qj} \right) + u \cdot \sum_{i,j} c_{ij} s^{pi} \sum_{k=1}^{qj} (-1)^k \binom{qj}{k} s^{qj-k} u^{k-1}. \end{aligned} \tag{3}$$

Then since f is not associate to (hence not divisible by) $X^q - Y^p$ in $K[X, Y]$, it follows from Lemma 5.3 that $u \nmid f$ in $K'[s, u]$. Therefore, $f = f_1 + uf_2$ with $0 \neq f_1 \in K'[s]$ and $f_2 \in K'[s, u]$. in particular, $f_1 = \sum_{i,j} c_{ij} s^{pi+qj}$. Then $v(f) = w(f) \leq w(f_1) = \min\{pi + qj \mid c_{ij} \neq 0\}$. As usual, recalling the notation of Lemma 3.5, write $\alpha = \sigma(1/f^*) = X^a Y^b / f$, where $(a, b) = (a_1(f), a_2(f))$, $a \geq 1$, and $b \geq 1$. Thus in the sums in (3) above, we have $i \leq a$ and $j \leq b$ for all pairs (i, j) such that $c_{ij} \neq 0$. Hence, $w(f_1)$ takes the form $pi + qj$ for some $i \leq a$ and $j \leq b$.

Therefore $v(\alpha) = pa + qb - v(f) \geq p(a-i) + q(b-j) \geq 0$, and if it is nonzero it must be at least p. Since R^* is generated as a K-algebra by all such terms $\sigma(1/f^*)$, it follows that $R^* \subseteq V$.

Now, suppose $\delta \in K[\theta] \setminus (\theta)K[\theta]$. Then by Theorem 3.3, δ is a unit of R^*, hence also in V, so $v(\delta) = 0$. Hence for any nonnegative integer m, we have $v(\theta^m/\delta) = m \cdot (pq + 1 - h)$. Thus, it has value 0 if and only if $h = pq + 1$, and is otherwise positive.

It remains to show that if $h = pq + 1$ and $v(\alpha) = 0$, then there exist some $m \geq 0$ and some $\delta \in K[\theta] \setminus (\theta)K[\theta]$ with $\alpha = \theta^m/\delta$, whereas if $h < pq + 1$ then $v(\alpha) > 0$. To prove this, let f, α, f_1, f_2 be as above. We proceed by induction on the number $\ell = a + b = a_1(f) + a_2(f)$, noting that the statement is vacuously true for $\ell = 0, 1$.

We first dispense with the case that some monomial $c_{ij}s^{pi+qj}$ appearing in f_1 satisfies either $i < a$ or $j < b$. Then $v(f) \leq w(f_1) \leq pi + qj \leq pa + qb - p$, so that $v(\alpha) = pa + qb - v(f) \geq p$.

Thus, we may assume that $f_1 = c_{ab}s^{pa+qb}$, so that $c_{ab} \neq 0$. Set $g := f - c_{ab}X^aY^b$. Then rewriting g as an element of $K'[s, u]$, we have $g = g_1 + ug_2$, where $g_1 = f_1 - c_{ab}s^{pa+qb} = 0$. Hence $u \mid g$ in $K'[s, t]$, whence $X^q - Y^p \mid g$ in $K[X, Y]$ by Lemma 5.3. That is, we have $f = c_{ab}X^aY^b + (X^q - Y^p)^m H$, where $m \geq 1$ and $H \in K[X, Y]$ is relatively prime to each of X, Y, and $X^q - Y^p$. Thus $a \geq p$ and $b \geq q$. Also note that $a_1(H) \leq a - qm$ and $a_2(H) \leq b - pm$.

Set $\alpha' := X^aY^b/((Y^p - X^q)^m H)$. Then there are nonnegative integers e_1, e_2 with $\alpha' = X^{e_1}Y^{e_2}\theta^m\sigma(1/H^*)$. In particular, $e_1 = a - mp - a_1(H)$ and $e_2 = b - mq - a_2(H)$. Thus, $c_{ab}\alpha' + 1$ is a unit of R^*, so since $\alpha = \alpha'/(c_{ab}\alpha' + 1)$, we have $v(\alpha') = v(\alpha)$, which we assume to be 0. But $v(\alpha') = e_1 p + e_2 q + m(pq + 1 - h) + v(\sigma(1/H^*))$, whence, since $v(\alpha') = 0$, we have $e_1 = e_2 = 0$, and every irreducible factor τ of H satisfies $v(\sigma(1/\tau^*)) = 0$. Moreover if $h < pq + 1$ it further follows that $m = 0$, so that $f = c_{ab}X^aY^b$, contradicting the fact that f is relatively prime to X and Y, finishing this case.

Then in the remaining case (where $h = pq + 1$), by the inductive hypothesis each such τ satisfies $\sigma(1/\tau^*) = \theta^{m(\tau)}/\delta(\tau)$. As these terms are multiplicative, there is some $k \in \mathbb{N}$ and $\epsilon \in K[\theta] \setminus (\theta)K[\theta]$ with $\sigma(1/H^*) = \theta^k/\epsilon$ Thus, we have

$$\alpha = \frac{\alpha'}{c_{ab}\alpha' + 1} = \frac{\theta^{m+k}/\epsilon}{c_{ab}(\theta^{m+k}/\epsilon) + 1} = \frac{\theta^{m+k}}{c_{ab}\theta^{m+k} + \epsilon}.$$

Since $c_{ab}\theta^{m+k} + \epsilon \in K[\theta] \setminus (\theta)K[\theta]$, we are done. □

Recall (see [Zafrullah 1978]) that an integral domain is a *finite conductor domain* if the intersection of any pair of principal ideals is finitely generated.

Theorem 5.6. *For any $n \geq 2$, the ideal $(1/X_1)R_n \cap (1/X_2)R_n$ is not finitely generated. Hence R_n is not a finite conductor domain.*

Proof. Let $n \geq 3$ and suppose $(1/X_1)R_n \cap (1/X_2)R_n = (\alpha_1, \ldots, \alpha_t)R_n$ for some $\alpha_1, \ldots, \alpha_t \in R_n$. Let $S = R_n[X_3, \ldots, X_n]$. Let $L = K(X_3, \ldots, X_n)$. By Lemma 4.1, we have $S = R(L[X_1, X_2])$. Let $(-)'$ denote the image of an element of R_n in S. Then $\alpha'_j \in (1/X_1)S \cap (1/X_2)S$ for all j, so $(\alpha'_1, \ldots, \alpha'_t) \subseteq (1/X_1)S \cap (1/X_2)S$. Conversely let $u \in (1/X_1)S \cap (1/X_2)S$. Then by clearing denominators, there is some positive integer d such that $(X_3 \cdots X_n)^{-d}u \in (1/X_1)R_n \cap (1/X_2)R_n = (\alpha_1, \ldots, \alpha_t)R_n$. Since $S = R_n[X_3, \ldots, X_n]$, it follows that $u \in (\alpha'_1, \ldots, \alpha'_t)S$. Thus, $(1/X_1)S \cap (1/X_2)S$ is a finitely generated ideal, and we have reduced to the 2-dimensional case. So from now on we assume $n = 2$ and we rewrite $X = X_1$ and $Y = X_2$. For the rest of the proof, we pass to the R^* notation.

Suppose $XR^* \cap YR^* = (\alpha_1, \ldots, \alpha_t)$ for some finite list of nonzero $\alpha_i \in R^*$; a contradiction will complete the proof. Then there exist $\beta_i, \gamma_i \in R^*$ with $\alpha_i = X\beta_i = Y\gamma_i$ for all i. Write $\gamma_i = c_i + \sum_{j=1}^{m_i} \sigma(1/f_{ij})$, where $c_i \in K$, $m_i \geq 0$, and each f_{ij} is in $D \setminus K$. If some $c_i \neq 0$, then γ_i is a unit by Lemma 2.3, so $Y/X = \gamma_i^{-1}\beta_i \in R^*$, which is false by [Epstein 2024c, Example 2.9]. Hence $m_i \geq 1$ and $\gamma_i = \sum_{j=1}^{m_i} \sigma(1/f_{ij})$. Choose some positive integer q that is not a multiple of $\operatorname{char} K$ and such that $q > \max\{\deg_X f_{ij} \mid 1 \leq i \leq t,\ 1 \leq j \leq m_i\}$. Set $v := v_{1,q,q+1}$ and $\theta = \theta_{1,q} = X^qY/(Y - X^q)$ as in Lemma 5.5.

Then $Y\theta = X^q \cdot (\theta + Y) \in XR^*$, and therefore $Y\theta \in XR^* \cap YR^*$. It follows that $\theta \in (\gamma_1, \ldots, \gamma_t)$. Since $v(\theta) = 0$, it follows that for some pair (i, j), we have $v(\sigma(1/f_{ij})) = 0$. By Lemma 5.5, there exists some positive integer m and some element $\delta \in K[\theta] \setminus (\theta)K[\theta]$ such that $\sigma(1/f_{ij}) = \theta^m\delta^{-1}$. Write $f = f_{ij}$.

Let $d = \deg_X(f)$ and $e = \deg_Y(f)$. Then by Lemma 3.5, we have $\sigma(1/f) = X^dY^e/f^*$, where $\deg_X(f^*) \leq d$ and $\deg_Y(f^*) \leq e$.

Write $\delta = c_0 + \sum_{i=1}^{s} c_i\theta^i$, where each c_i is in K and $c_0 \neq 0$. Then

$$\frac{X^dY^e}{f^*} = \sigma(1/f) = \theta^m\delta^{-1} = \frac{(X^qY)^m/(Y - X^q)^m}{c_0 + \sum_{i=1}^{s} c_i(X^qY)^i/(Y - X^q)^i}.$$

If $m \geq s$, then the latter equation simplifies to an equation where both the numerator and denominator of each fraction is a polynomial, as follows:

$$\frac{X^dY^e}{f^*} = \frac{X^{qm}Y^m}{c_0(Y - X^q)^m + \sum_{i=1}^{s} c_i(X^qY)^i(Y - X^q)^{m-i}}.$$

Since $q > d$ by the choice of q, we have $qm > d$. It follows that $X \mid c_0(Y - X^q)^m$, which contradicts the fact that $c_0 \in K^\times$.

On the other hand if $m < s$, then the equation simplifies with numerators and denominators being polynomials, as follows:

$$\frac{X^dY^e}{f^*} = \frac{X^{qm}Y^m(Y - X^q)^s}{c_0(Y - X^q)^s + \sum_{i=1}^{s} c_i(X^qY)^i(Y - X^q)^{s-i}}.$$

Cross-multiplying, we have

$$X^{qm}Y^m(Y-X^q)^s f^* = X^dY^e\cdot\left(c_0(Y-X^q)^s+\sum_{i=1}^{s}c_i(X^qY)^i(Y-X^q)^{s-i}\right).$$

Since $qm > d$, it follows that $X \mid c_0(Y-X^q)^s$, which again contradicts the fact that $c_0 \in K^\times$. □

Recall that a ring is *coherent* if every finitely generated ideal is finitely presented. The coherent rings include the Noetherian rings and also all valuation domains (see [Bourbaki 1972, Chapter I, §2, Exercise 12 and Chapter VI, §1, Exercise 3]).

Corollary 5.7. *For any $n \geq 2$, the ring R_n is not coherent. Hence it is non-Noetherian.*

Proof. This follows from Theorem 5.6 and [Chase 1960, Theorem 2.2]. □

The next result is notably unlike the behavior of localized polynomial rings.

Theorem 5.8. *For any $n \geq 2$, R_n is not integrally closed.*

Proof. We first consider the 2-dimensional case. Let p and q be relatively prime integers with $1 < p < q$, such that neither p nor q is a multiple of char K. By elementary number theory, there is a unique pair of integers c and d with $qd-pc=1$, $0<c<q$, and $0 < d < p$. Consider the element $\beta := \beta_{p,q} := (X^{2q-c}Y^d)/(X^q-Y^p) \in F$. We claim that β is integral over R^* — in fact, $\beta^p \in R^*$ — but $\beta \notin R^*$.

To see that $\beta^p \in R^*$, simply note the following:

$$\beta^p = \frac{X^{(2q-c)p}Y^{dp}}{(X^q-Y^p)^p} = \left(\frac{X^qY^p}{X^q-Y^p}\right)^d\cdot\left(\frac{X^qY^p}{X^q-Y^p}+X^q\right)^{p-d}\cdot X,$$

which is in R^* since $X \in R^*$ and $\sigma(1/(Y^p-X^q)) = X^qY^p/(X^q-Y^p) \in R^*$.

On the other hand, let $v = v_{p,q,pq+1}$. Then

$$v(\beta) = v\left(\frac{X^{2q-c}Y^d}{X^q-Y^p}\right) = (2q-c)p+qd-2pq = qd-pc = 1.$$

Suppose $\beta \in R^*$. Since $v(\beta) > 0$, it follows that $\beta \in \mathfrak{m}_V \cap R^* \subseteq \mathfrak{m}$, the maximal ideal of R^*. So by Theorem 3.3, we have $\beta = \sum_{i=1}^{t}\sigma(1/f_i^*)$ for nonconstant polynomials $f_i \in K[X,Y]$. By reordering, let $f_1,\ldots,f_s$ be the polynomials whose only irreducible factor is X^q-Y^p up to associate and multiplicity, whereas each of $f_{s+1},\ldots,f_t$ has an irreducible factor not associate to X^q-Y^p. Set $\gamma := \sum_{i=1}^{s}\sigma(1/f_i^*)$ and $\delta := \sum_{i=s+1}^{t}\sigma(1/f_i^*)$, so that $\beta = \gamma+\delta$. By Lemma 5.5, we have $v(\delta) \geq p$, so that since $v(\beta) = 1 < p$, we have $v(\gamma) = 1$.

On the other hand, for $1 \leq i \leq s$, there exist $\lambda_i \in K$ and $\ell_i \in \mathbb{N}_0$ with $f_i = \lambda_i(X^q-Y^p)^{\ell_i}$. Thus, $\sigma(1/f^*) = \lambda_i\theta^{\ell_i}$, so that by Lemma 5.5 we have $v(\sigma(1/f_i^*)) =$

$\ell_i v(\theta) = 0$. Thus, either $\gamma = 0$ or $v(\gamma) = 0$, either of which is a contradiction. Hence, $\beta \notin R^*$.

Finally, we pass to the n-dimensional case. We have $\beta^p \in R_2^* \subseteq R_n^*$. Since $R_2^* = R_n^* \cap K(X, Y)$ by Proposition 3.4, and $\beta \in K(X, Y) \setminus R_2^*$, it follows that $\beta \notin R_n^*$. □

6. The abundance of prime ideals in *R*

In this section, we show that R_n, which as we have seen is far from Noetherian when $n > 1$ (see Corollary 5.7), does have infinitely many prime ideals of each height other than 0 and n, a property enjoyed by any n-dimensional Noetherian ring, but not by some non-Noetherian rings (e.g., any valuation domain of dimension at least 2). We start with the following result to bootstrap our efforts.

Proposition 6.1. *For any $n \geq 2$, R_n has infinitely many height-one prime ideals.*

Proof. In this proof, we use R^* notation.

First suppose $n = 2$. For any relatively prime pair (p, q) of positive integers with $p < q$, let V, v, θ, and h be as in Lemma 5.5, with $h = pq + 1$. Let $\mathfrak{p} = \mathfrak{p}_{p,q}$ be the contraction of $\mathfrak{m}_V$ to R^*. Then since $v(\theta) = 0$, we have $\theta \notin \mathfrak{p}$. Since $\mathfrak{p}$ is a nonzero prime but not the maximal ideal of R^* (as $\theta \in \mathfrak{m}$), it follows that $\mathfrak{p}$ is a height-one prime.

On the other hand, let (r, s) be a different pair of relatively prime positive integers with $r < s$. We claim that $v(X^s - Y^r) = \min\{v(X^s), v(Y^r)\} = \min\{ps, qr\}$. Otherwise we would have $v(X^s) = v(Y^r)$, whence $ps = qr$. But then by assumption of relatively prime pairs, we would have $p = r$ and $q = s$, contradicting the assumption of distinctness. Therefore, $v(\theta_{r,s}) = v(X^s Y^r) - v(X^s - Y^r) = \max\{qr, ps\}$. Thus, $\theta_{r,s} \in \mathfrak{p}_{p,q}$. But by the proof of Lemma 5.5, $\theta_{r,s} \notin \mathfrak{p}_{r,s}$. Hence, $\mathfrak{p}_{p,q} \neq \mathfrak{p}_{r,s}$. Since there are infinitely many such pairs of integers, it follows that R^* has infinitely many height-one primes.

Finally, we drop the assumption that $n = 2$. By Proposition 4.2, there is a prime ideal Q of R^* such that R^*_Q is isomorphic to the reciprocal complement of $L[X, Y]$ for some field L. But then by the dimension 2 part of the proof above, R^*_Q has infinitely many height-one primes. Thus, there are infinitely many height-one primes of R^* that are contained in Q. □

Notation 6.2. Recall that given a valuation ring with fraction field K and an indeterminate t over K, the ring $V(t)$ is a valuation ring of $K(t)$ called the trivial extension of V. Given $\varphi = \sum_{j=0}^{e} f_j t^j$ with $f_j \in K$, then the value of φ with respect to $V(t)$ is $\min_j\{v(f_j)\}$ (see [Gilmer 1972, p. 218]).

By Lemma 4.3, there is a prime ideal $Q \in \operatorname{Spec} R_n$ such that

$$(R_n)_Q = R(K(X_n)[X_1, \ldots, X_{n-1}]).$$

Fix this prime for the next two lemmas.

Lemma 6.3. *Let V be a valuation overring of R_{n-1}; then the trivial extension $V(X_n)$ is an overring of $(R_n)_Q$, where Q is as in Notation 6.2.*

Proof. By the comment before the Lemma, it suffices to show that $1/\varphi \in V(X_n)$ for every $\varphi \in K(X_n)[X_1, \dots, X_{n-1}]$. Since $K(X_n) \subseteq V(X_n)$, we may assume $\varphi \in K[X_1, \dots, X_n]$. Let v^* be the valuation for $V(X_n)$; write $\varphi = \sum_{j=0}^{e} f_j X_n^e$ with $f_j \in K[X_1, \dots, X_{n-1}]$. Since V is an overring of R_{n-1}, we have that $v(f_j) \le 0$ whenever $f_j \ne 0$. Hence, $v^*(1/\varphi) = -\min\{v(f_j) \mid 0 \le j \le e \text{ and } f_j \ne 0\} \ge 0$. □

Lemma 6.4. *Let $\mathfrak{p} \in \operatorname{Spec} R_{n-1}$, and let V be a valuation overring of R_{n-1} centered on $\mathfrak{p}$. Let $\mathfrak{p}'$ be the center of $V(X_n)$ in R_n. Then $\operatorname{ht}\mathfrak{p}' \ge \operatorname{ht}\mathfrak{p}$, with equality if $\operatorname{ht}\mathfrak{p} \in \{0, n-2, n-1\}$.*

Proof. Let $i = \operatorname{ht}\mathfrak{p}$. If $i = 0$, then $\mathfrak{p} = (0)$, so that $V = \operatorname{Frac} R_{n-1} = K(X_1, \dots, X_{n-1})$, whence $\mathfrak{p}' = (0)$. Assume by induction that $i \ge 1$ and the inequality holds for all primes with smaller height. Note that $\mathfrak{p}' \cap R_{n-1} = \mathfrak{p}$.

Let $\mathfrak{q} \subsetneq \mathfrak{p}$ with $\mathfrak{q} \in \operatorname{Spec} R_{n-1}$ and $\operatorname{ht}\mathfrak{q} = i-1$. Let W be a valuation overring of R_{n-1} centered on $\mathfrak{q}$; let $\mathfrak{q}'$ be the center of $W(X_n)$ in R_n. To show that $\mathfrak{q}' \subseteq \mathfrak{p}'$, it suffices by Proposition 2.8 to show that for any $\varphi \in K[X_1, \dots, X_n]$ with $1/\varphi \in \mathfrak{q}'$, we have $1/\varphi \in \mathfrak{p}'$. Write $\varphi = \sum_{j=0}^{e} f_j X_n^j$, where $f_j \in K[X_1, \dots, X_{n-1}$. Since $1/\varphi \in \mathfrak{q}'$, there is some $0 \le k \le e$ with $1/f_k \in \mathfrak{q}$, by the way the valuation on $W(X_n)$ is defined. Thus, $1/f_k \in \mathfrak{p}$, so $v^*(1/\varphi) = -\min\{v(f_j) \mid 0 \le j \le e\} \ge -v(f_k) > 0$. Hence $\mathfrak{q}' \subseteq \mathfrak{p}'$. On the other hand $\mathfrak{q}' \ne \mathfrak{p}'$, since for any $\alpha \in \mathfrak{p} \setminus \mathfrak{q}$, we have $\alpha \in \mathfrak{p}' \setminus \mathfrak{q}'$. Thus, $\mathfrak{q}' \subsetneq \mathfrak{p}'$, so that

$$\operatorname{ht}\mathfrak{p}' \ge 1 + \operatorname{ht}\mathfrak{q}' \ge 1 + (i-1) = i,$$

with the second inequality by the inductive hypothesis.

Suppose $i = n-1$. Since $X_n \notin \mathfrak{p}'$, we have that $\mathfrak{p}'$ is not the maximal ideal of R_n, so that $\operatorname{ht}\mathfrak{p}' \le n-1$. But also $\operatorname{ht}\mathfrak{p}' \ge \operatorname{ht}\mathfrak{p} = n-1$, so that $\operatorname{ht}\mathfrak{p}' = n-1$.

Finally, suppose $i = n-2$. Let $\mathfrak{m}$ be the maximal ideal of R_{n-1}. Since $\mathfrak{m}$ contains all nonunits of R_{n-1} and $\operatorname{ht}\mathfrak{m} = n-1$, there is some $\alpha \in \mathfrak{m} \setminus \mathfrak{p}$. Thus $\alpha \notin \mathfrak{p}'$. But since $V(X_n) \supseteq (R_n)_Q$ by Lemma 6.3, we have $\mathfrak{p}' \subseteq Q$. Moreover, the containment must be strict, since $\alpha \in \mathfrak{m} \subseteq Q$ but $\alpha \notin \mathfrak{p}'$. Thus, $\operatorname{ht}\mathfrak{p}' \le n-2$, so $\operatorname{ht}\mathfrak{p}' = \operatorname{ht}\mathfrak{p} = n-2$. □

Lemma 6.5. *Let $\mathfrak{p}$ be a prime ideal of R_{n-1} of height $n-2$. Let V a valuation ring of R_{n-1} centered on $\mathfrak{p}$. Let κ_V be the residue field of V and let $\pi : V(X_n) \twoheadrightarrow \kappa_V(X_n)$ be the canonical surjection. Let $W := \pi^{-1}(\kappa_V[X_n^{-1}]_{(X_n^{-1})})$. Then W is a valuation overring of R_n centered on a prime ideal $\mathfrak{a}$ with $\operatorname{ht}\mathfrak{a} = n-1$, such that $\mathfrak{a} \cap R_{n-1} = \mathfrak{p}$.*

Proof. We have that W is a valuation ring with quotient field $K(X_1, \dots, X_n)$ by [Bastida and Gilmer 1973, Theorem 2.1(h)]. If G is the value group of V, then $G \oplus \mathbb{Z}$, ordered lexicographically, is the value group of W. In particular, given $\varphi =$

$\sum_{j=0}^{e} f_j X_n^j \in K[X_1, \ldots, X_n]$, with each f_j in $K[X_1, \ldots, X_{n-1}]$, the valuation w is given by $w(\varphi) = (v(f_k), -k)$, where k is the largest index i with $0 \le i \le e$ such that $v(f_i) \le v(f_j)$ for all $0 \le j \le e$. Then $w(1/\varphi) = (-v(f_k), k) \ge (0, 0)$ since $1/f_k \in R_{n-1} \subseteq V$, whence $v(f_k) \le 0$.

Set $\mathfrak{a} := \mathfrak{m}_W \cap R_n$ and $\mathfrak{p}' := \mathfrak{m}_{V(X_n)} \cap R_n$. By standard pullback results, the maximal ideal of $V(X_n)$ is a nonmaximal prime of W; thus $\mathfrak{p}' \subseteq \mathfrak{a}$. On the other hand, $X_n^{-1} \in \mathfrak{a} \setminus \mathfrak{p}'$, so that $\mathfrak{p}' \subsetneq \mathfrak{a}$. Since $\mathfrak{p}$ is a nonmaximal ideal of R_{n-1}, there is some nonunit α of R_{n-1} (hence also of R_n) that avoids $\mathfrak{p}$. We have $w(\alpha) = (v(\alpha), 0) = (0, 0)$, so that $\alpha \notin \mathfrak{a}$. Thus, $\mathfrak{p}' \subsetneq \mathfrak{a} \subsetneq \mathfrak{m}_{R_n}$, so that since $\operatorname{ht} \mathfrak{p}' = n - 2$ by Lemma 6.4, we have $\operatorname{ht} \mathfrak{a} = n - 1$.

Now, $\mathfrak{p} = \mathfrak{p}' \cap R_{n-1} \subseteq \mathfrak{a} \cap R_{n-1}$. Hence, $\operatorname{ht}(\mathfrak{a} \cap R_{n-1}) \ge \operatorname{ht} \mathfrak{p} = n - 2$. But $\alpha \in \mathfrak{m}_{R_{n-1}} \setminus \mathfrak{a}$, so $\operatorname{ht}(\mathfrak{a} \cap R_{n-1}) = n - 2$, whence $\mathfrak{a} \cap R_{n-1} = \mathfrak{p}$. □

Theorem 6.6. *For every* $1 \le i \le n - 1$, *there exist infinitely many primes of* R_n *of height* i.

Proof. When $n = 0, 1$, the statement is vacuous. Moreover, since when $n \ge 2$ we know that R_n has infinitely many height-one primes by Proposition 6.1, the result holds for $n = 2$. Thus, we assume inductively that $n > 2$ and the result holds for smaller n. Since $(R_n)_Q = R(K(X_n)[X_1, \ldots, X_{n-1}])$ (see Notation 6.2), it has infinitely many primes of height i for $1 \le i \le n - 2$, which then restrict to distinct primes of these heights in R_n via the localization map. So we need only show that R_n has infinitely many primes of height $n - 1$.

Let $\mathfrak{p}$ and $\mathfrak{q}$ be distinct prime ideals of height $n - 2$ in R_{n-1}. By Lemma 6.5, there are valuation overrings W_1 and W_2 of R_n whose centers in R_n are height $n - 1$ primes $\mathfrak{p}'$ and $\mathfrak{q}'$ such that $\mathfrak{p}' \cap R_{n-1} = \mathfrak{p}$ and $\mathfrak{q}' \cap R_{n-1} = \mathfrak{q}$. Since $\mathfrak{p} \ne \mathfrak{q}$, it follows that $\mathfrak{p}' \ne \mathfrak{q}'$. Since there are infinitely many primes of height $n - 2$ in R_{n-1} by the inductive hypothesis, it follows that there are infinitely many primes of height $n - 1$ in R_n. □

7. The dimension 2 case

In this section, we work in two variables, so that $D = K[X, Y]$ where $X = X_1$, $Y = X_2$ for short, $R = R(K[X, Y])$, $F = K(X, Y)$, etc. We have done likewise in many results earlier in the paper in service of extending the results to higher dimensions. However, for each of the results in this section, either we do not know how to extend it into higher dimension, or else we know it to be false in higher dimension. In the dimension 2 case, we will show that R has an integral overring that is not finitely generated, that localizing R at height-one primes always yields Noetherian domains, and that any finitely generated proper ideal lives in almost all height-one primes.

We start by expanding Theorem 5.8 to show that the integral closure if R^* is quite a bit larger than R^* itself:

Proposition 7.1. *There is an overring S of $R = R_2$, that is integral over R but not finitely generated over it.*

Proof. As usual, we will work with R^* instead of R.

Let $\Sigma := \{\beta_{p,q}\}$ as in the proof of Theorem 5.8, where the pairs (p, q) range over all relatively prime pairs of integers $1 < p < q$ such that neither p nor q is a multiple of $\operatorname{char} K$. Let $S := R^*[\Sigma]$. Then, as seen in the proof of Theorem 5.8, each $\beta_{p,q}$ is in the integral closure of R^*. Hence, S is integral over R^*.

Suppose that S is finitely generated as an R^*-algebra. Then there is a finite list of such pairs $\{(p_i, q_i)\}_{1\le i\le s}$ such that $S = R^*[\beta_{p_1,q_1}, \dots, \beta_{p_s,q_s}]$. Choose $r > \max\{p_i \mid 1 \le i \le s\}$ such that $r > 1$. Then $(r, r+1)$ is such a pair, so $\beta_{r,r+1} \in S$. Let $v = v_{r,r+1,r^2+r+1}$.

Then for any relatively prime pair (p, q) with $p < r$, we claim that $v(X^q - Y^p) = \min\{v(X^q), v(Y^p)\}$. If this were not the case, we would have $rq = v(X^q) = v(Y^p) = (r+1)p$, so that $(q - p)r = p$, contradicting the facts that $q - p \ge 1$ and $r > p$. Thus, $v(X^q - Y^p) = \min\{rq, (r+1)p\}$.

Now choose any $(p, q) = (p_i, q_i)$ with $1 \le i \le s$. Let (c, d) be the unique pair of integers with $0 < c < q,\ 0 < d < p$, and $qd = pc + 1$. Then

$$\begin{aligned} v(\beta_{p,q}) = v\left(\frac{X^{2q-c}Y^d}{X^q - Y^p}\right) &= (2q-c)r + d(r+1) - \min\{rq, (r+1)p\} \\ &\ge (2q-c)r + d(r+1) - rq = (q-c)r + (r+1)d \ge 2r+1. \end{aligned}$$

On the other hand, by the proof of Theorem 5.8, we have $v(\beta_{r,r+1}) = 1$.

Let $Z_1, \dots, Z_s$ be algebraically independent indeterminates over R^*. It follows that if $g \in R^*[Z_1, \dots, Z_s]$ such that $\beta_{r,r+1} = g(\beta_{p_1,q_1}, \dots, \beta_{p_s,q_s})$, then g has a nonzero constant term c, and $1 = v(\beta_{r,r+1}) = v(c)$. But this contradicts the fact (see Lemma 5.5) that every element of R^* has value either 0 or $\ge r$ under v. Thus, S is not finitely generated as an R^*-algebra. □

Our main result for dimension 2 shows that the localizations at height-one primes are surprisingly well behaved. First, though, we need the following lemma.

Lemma 7.2. *Let $\mathfrak{p} \in \operatorname{Spec} R_n$ and $1 \le i \le n$ such that $\mathfrak{p} \not\subseteq \mathfrak{q}_i$, where $\mathfrak{q}_i$ is as in Lemma 4.3. Then $1/X_i \in \mathfrak{p}$.*

The $\mathfrak{q}_i$ are mutually incomparable, and for each i, $1/X_j \in \mathfrak{q}_i$ for each $j \ne i$.

If $n = 2$, then any nonzero prime distinct from $\mathfrak{q}_1$ and $\mathfrak{q}_2$ contains $(1/X_1, 1/X_2)R$. Hence with the notation of Lemma 3.6, $\mathfrak{p}_1 = \mathfrak{q}_2$ and $\mathfrak{p}_2 = \mathfrak{q}_1$.

Proof. By Lemma 4.3, any prime ideal avoiding X_i^{-1} must be contained in $\mathfrak{q}_i$. Thus $1/X_i \in \mathfrak{p}$.

Now let i and j be distinct integers between 1 and n. Combining Lemmas 4.1 and 4.3 and Theorem 4.4 yields $\operatorname{ht}\mathfrak{q}_i = \operatorname{ht}\mathfrak{q}_j = n-1$. Thus, $\mathfrak{q}_i$ and $\mathfrak{q}_j$ must be incomparable. Since $\mathfrak{q}_i \not\subseteq \mathfrak{q}_j$, we have $X_j \in \mathfrak{q}_i$ by the first paragraph.

In the $n = 2$ case, by Theorem 4.4 we have that any nonmaximal nonzero prime has height one. In particular, if $\mathfrak{p}$ is distinct from $\mathfrak{q}_1$ and $\mathfrak{q}_2$, then since $\operatorname{ht}\mathfrak{p} = 1 = \operatorname{ht}\mathfrak{q}_1 = \operatorname{ht}\mathfrak{q}_2$, we have $\mathfrak{p} \nsubseteq \mathfrak{q}_i$ for $i = 1, 2$. Since $1/X_i \in \mathfrak{p}_i \setminus \mathfrak{q}_i$ for $i = 1, 2$, the final claim follows. □

Theorem 7.3. *Let $\mathfrak{p}$ be a height-one prime ideal of $R = R_2$. Then $R_\mathfrak{p}$ is a Noetherian one-dimensional local domain.*

Proof. By Lemma 3.6, we may assume $\mathfrak{p} \neq \mathfrak{p}_1, \mathfrak{p}_2$. We work in R^* and use the notation $D = K[X, Y]$ and $R^* = \sigma(D)$. Since $\mathfrak{p}$ is not maximal there must exist $f \in (X, Y)D$ irreducible and not associate to either X or Y in D, such that $\alpha := \sigma(1/f^*) \notin \mathfrak{p}$. As neither α nor α^{-1} are in $S = D_{(X,Y)}$ it follows from [Seidenberg 1953, Theorem 7] that $S[\alpha]$ is a two-dimensional Noetherian ring having a height-one prime ideal $\mathfrak{q}$ generated by X and Y. Hence $S[\alpha]_\mathfrak{q} = D[\alpha]_{(X,Y)}$ is a one-dimensional local Noetherian domain.

Moreover, since $S \subseteq R^*$, we get $S[\alpha] \subseteq R^*$. Let us show that $\mathfrak{p} \cap S[\alpha] = \mathfrak{q}$. By Lemma 7.2, $X, Y \in \mathfrak{p}$, so that $\mathfrak{q} \subseteq \mathfrak{p} \cap S[\alpha]$. For the reverse containment, it will be enough to show that $\mathfrak{p} \cap S[\alpha]$ is a height-one prime, for which it will suffice to show it is not a maximal ideal.

Every maximal ideal of $S[\alpha]$ that contains $\mathfrak{q}$ is of the form $(\mathfrak{q}, h(\alpha))$ where $h \in K[T]$ is an irreducible monic polynomial. This is because by [Seidenberg 1953, Theorem 7], we have $S[\alpha]/\mathfrak{q} \cong K[T]$, where T is an indeterminate over K and $\overline{2}2\alpha \mapsto T$ in the isomorphism. We know that $\alpha \notin \mathfrak{p}$. If $h(T) \neq T$, then it has a nonzero constant term, so by Theorem 3.3, $h(\alpha)$ is a unit in R^*, whence $h(\alpha) \notin \mathfrak{p}$. On the other hand, $\alpha \notin \mathfrak{p}$ by assumption.

Thus $\mathfrak{p} \cap S[\alpha] = \mathfrak{q}$, so that $S[\alpha]_\mathfrak{q} \subseteq R^*_\mathfrak{p}$. By the Krull–Akizuki theorem [Matsumura 1986, Theorem 11.7], $R^*_\mathfrak{p}$ is Noetherian, finishing the proof. □

Remark 7.4. Suppose, in the setting of Theorem 7.3, that there exists $f \in (X, Y)D \setminus (X, Y)^2 D$ such that $1/f^* \notin \mathfrak{p}$. Then $R_\mathfrak{p}$ is not merely one-dimensional and Noetherian, but a DVR.

To see this, and continuing the notation in the proof above, let Z be an indeterminate over D and let $\pi : K[X, Y, Z] \to D[\alpha]$ be the unique K-algebra homomorphism that fixes D and sends $Z \mapsto \alpha$. Note that π is surjective. Since $D[\alpha]$ has dimension 2, the kernel of π is a height-one prime of $K[X, Y, Z]$. It is clear that the polynomial $h = fZ - X^{a_1(f)} Y^{a_2(f)}$ is irreducible in $K[X, Y, Z]$ and contained in the kernel of π. It follows that

$$D[\alpha] \cong \frac{K[X, Y, Z]}{(h)}.$$

The ring

$$S[\alpha]_\mathfrak{q} \cong \left(\frac{K[X, Y, Z]}{(h)}\right)_{(X,Y)}$$

is a DVR if and only if h is a regular parameter in $K[X, Y, Z]_{(X,Y)}$. This happens if and only if $f \in (X, Y)D \setminus (X, Y)^2 D$. In case $S[\alpha]_{\mathfrak{q}}$ is a DVR, we clearly have $S[\alpha]_{\mathfrak{q}} = R^*_{\mathfrak{p}}$ since a DVR has no proper overring other than its fraction field.

Next we show a "cofinite character" result that is somewhat dual to the finite character property of Krull domains.

Theorem 7.5. *Every finitely generated proper ideal of $R = R_2$ is contained in all but finitely many prime ideals.*

Proof. Every nonunit element $\varphi \in R^*$ can be written as a finite sum $\varphi = \varphi_1 + \cdots + \varphi_t$ such that any φ_j is a finite product of elements of the form $\sigma(1/f)$ with f irreducible in $K[X, Y]$. Thus it is sufficient to prove that any $\sigma(1/f)$ with f irreducible is contained in all but finitely many primes of R^*. Since we already know this fact for X and Y by Lemma 3.7, we can assume f is not an associate of X nor Y. Set $\alpha = \sigma(1/f)$. We know that α is in the maximal ideal of R^*. By the proof of Theorem 7.3, we get that if a prime $\mathfrak{p}$ of R^* does not contain α, then $R^*_{\mathfrak{p}}$ contains the one-dimensional Noetherian local domain $D[\alpha]_{(X,Y)}$. Suppose there exist two distinct nonzero prime ideals $\mathfrak{p}$ and $\mathfrak{q}$ of R^* with $\alpha \notin \mathfrak{p} \cup \mathfrak{q}$. Necessarily $\mathfrak{p}$ and $\mathfrak{q}$ have height one. We show that $R^*_{\mathfrak{p}}$ and $R^*_{\mathfrak{q}}$ cannot be contained in a common valuation overring. Suppose by way of contradiction that $R^*_{\mathfrak{p}} \cup R^*_{\mathfrak{q}} \subseteq V$ for some valuation ring V contained in $K(X, Y)$. Then V is an overring of $D[\alpha]_{(X,Y)}$, hence a DVR. Since $R^*_{\mathfrak{p}}$ and $R^*_{\mathfrak{q}}$ are one-dimensional, the maximal ideal $\mathfrak{m}_V$ of V contains both the maximal ideals of $R^*_{\mathfrak{p}}$ and $R^*_{\mathfrak{q}}$. Therefore the intersection $R^*_{\mathfrak{p}} \cap R^*_{\mathfrak{q}}$ is local. For this observe that given two nonunits $\beta, \theta \in R^*_{\mathfrak{p}} \cap R^*_{\mathfrak{q}}$, we have $\beta, \theta \in \mathfrak{p}R^*_{\mathfrak{p}} \cup \mathfrak{q}R^*_{\mathfrak{q}} \subseteq \mathfrak{m}_V$ and hence their sum $\beta + \theta \in \mathfrak{m}_V \cap R^*_{\mathfrak{p}} \cap R^*_{\mathfrak{q}} \subseteq \mathfrak{p}R^*_{\mathfrak{p}} \cap \mathfrak{q}R^*_{\mathfrak{q}}$ is a nonunit in $R^*_{\mathfrak{p}} \cap R^*_{\mathfrak{q}}$. Moreover, we have $R^* \subseteq R^*_{\mathfrak{p}} \cap R^*_{\mathfrak{q}} \subseteq R^*_{\mathfrak{p}}$. Because $\alpha^{-1} \in (R^*_{\mathfrak{p}} \cap R^*_{\mathfrak{q}}) \setminus R^*$, the maximal ideal of $R^*_{\mathfrak{p}} \cap R^*_{\mathfrak{q}}$ must contract to a nonzero prime ideal of R^* not containing α. But any such ideal has height one, so that $R^*_{\mathfrak{p}} \cap R^*_{\mathfrak{q}}$ contains the localization of R^* at some height-one prime. This is a contradiction since two localizations at distinct height-one primes cannot be comparable with respect to inclusion. Hence, $R^*_{\mathfrak{p}}$ and $R^*_{\mathfrak{q}}$ cannot have a common valuation overring.

Suppose there are infinitely many distinct primes of R^* not containing α. Then the above implies that $D[\alpha]_{(X,Y)}$ has infinitely many valuation overrings, but it is clearly a contradiction since $D[\alpha]_{(X,Y)}$ is local, Noetherian and one-dimensional. □

Recall that in a domain S, with fraction field F, for any S-submodule I of F we can write $I^{-1} := \{x \in F \mid xI \subseteq S\}$. Then for any ideal I of S, we set $I_v := (I^{-1})^{-1}$ and $I_t := \bigcup\{J_v \mid J \subseteq I \text{ and } J \text{ is finitely generated}\}$. A *t-ideal* is then an ideal I such that $I = I_t$. If there is a unique maximal element among the proper t-ideals of S, we say S is *t-local*. See [Fontana and Zafrullah 2019].

Corollary 7.6. *The ring $R = R_2$ is t-local.*

Proof. In fact the unique maximal ideal $\mathfrak{m}$ of R is a t-ideal. To see this, let I be a finitely generated proper ideal of R. Then by Theorem 7.5, I is contained in some height-one prime ideal $\mathfrak{p}$. But $\mathfrak{p}$ is a t-ideal by [Elliott 2019, top of p. 23]. Hence, $I_t \subseteq \mathfrak{p}_t = \mathfrak{p} \subset \mathfrak{m}$. Thus, $\mathfrak{m}_t = \mathfrak{m}$. □

8. Questions

The study of reciprocal complements is an entirely new field of inquiry. There are many interesting questions one could ask about this particular R, or about reciprocal complements in general. The following are just some questions that occurred to these authors, but such questions are easy to generate. As seen below, some of these questions have had progress on them since they were first proposed in an earlier draft of this paper.

Question 1. What can be said about the integral closure of R, where R is the reciprocal complement of a polynomial ring in two or more variables over a field? In particular:

(a) Is the integral closure of R infinitely generated over R? (We can't conclude this from Proposition 7.1 since R is not a Noetherian ring, so finitely generated R-modules can have infinitely generated submodules.)

(b) Is the integral closure of R local? If not, is it at least semilocal?

(c) Is the integral closure of R completely integrally closed?

Question 2. Let R be the reciprocal complement of a polynomial ring in finitely many variables over a field. Is R a *strong Bézout intersection domain* (SBID) (see [Guerrieri and Loper 2021])? That is, is it true that every finite intersection of non-comparable principal ideals fails to be finitely generated?

Some positive evidence is given by Theorem 5.6, and also by Corollary 7.6, since any SBID is t-local.

Question 3. Let D be a Noetherian domain. Is $R(D)$ a G-domain?

For some evidence of this, see Proposition 2.9. More generally, by [Guerrieri 2025, Theorem 2.12], the above holds whenever $\dim R(D) < \infty$.

Question 4. Let D be an integral domain of dimension ≥ 2. Assume that any nonzero Egyptian element of D is a unit. Must $R(D)$ be non-Noetherian?

We have seen in Corollary 5.7 that $D = D_n$ is an example of the above phenomenon when $n \geq 2$. More generally, by [Guerrieri 2025, Corollary 2.9], the answer is yes whenever $\dim R(D) \geq 2$.

Question 5. For any integral domain D, must we have $\dim R(D) \leq \dim(D)$?

Note that there is no hope for *equality* in the above, as we have for $D = D_n$ by Theorem 4.4. Indeed, the quantity $\varphi(D) := \dim(D) - \dim R(D)$ can be any nonnegative value ω, by letting A be a Jaffard (e.g., Noetherian) Egyptian domain of dimension ω and $D = A[X_1, \dots, X_d]$; then by Proposition 2.2 and Theorem 4.4, $\dim R(D) = d$, but $\dim D = d + \dim A$, so $\varphi(D) = \dim A = \omega$. Moreover, one can make A have any dimension ω by letting $G = \mathbb{Z}^{\oplus\omega}$ and $A = K[G]$ for any field K, which is Egyptian by [Guerrieri et al. 2024, Proposition 3].

By [Guerrieri 2025, Theorem 5.5], for any nonnegative integer c, one can construct integral domains D all of whose Egyptian elements are units, such that $\varphi(D) = c$.

The above question has a positive answer when D is finitely generated over a field or falls into certain classes of semigroup algebras [Guerrieri 2025, Theorem 3.2, Remark 4.11].

Question 6. Let D be a Noetherian integral domain with $\dim D = n \geq 2$. Are there infinitely many prime ideals of $R(D)$ of height i for each $1 \leq i \leq n-1$?

We see an example of this phenomenon when $D = D_n$ by Theorem 6.6. If D is not restricted to be Noetherian, however, there are counterexamples [Guerrieri 2025, Theorem 4.2].

Acknowledgement

The authors would like to thank the referees for their thorough work on our paper, which reads much better and has clearer focus as a result.

References

[Anderson et al. 1988] D. F. Anderson, A. Bouvier, D. E. Dobbs, M. Fontana, and S. Kabbaj, “On Jaffard domains”, *Exposition. Math.* **6**:2 (1988), 145–175. MR

[Bastida and Gilmer 1973] E. Bastida and R. Gilmer, “Overrings and divisorial ideals of rings of the form $D + M$”, *Michigan Math. J.* **20** (1973), 79–95. MR

[Bourbaki 1972] N. Bourbaki, *Elements of mathematics*: *Commutative algebra*, Hermann, Paris, 1972. MR

[Chase 1960] S. U. Chase, “Direct products of modules”, *Trans. Amer. Math. Soc.* **97** (1960), 457–473. MR

[Dunton and Grimm 1966] M. Dunton and R. E. Grimm, “Fibonacci on Egyptian fractions”, *Fibonacci Quart.* **4** (1966), 339–354. MR

[Elliott 2019] J. Elliott, *Rings, modules, and closure operations*, Springer, 2019. MR

[Epstein 2024a] N. Epstein, “Egyptian integral domains”, *Ric. Mat.* **73**:5 (2024), 2901–2910. MR

[Epstein 2024b] N. Epstein, “Rational functions as sums of reciprocals of polynomials”, *Amer. Math. Monthly* **131**:9 (2024), 794–801. MR

[Epstein 2024c] N. Epstein, “The unit fractions from a Euclidean domain generate a DVR”, *Ric. Mat.* (2024), 7 pages.

[Fontana and Zafrullah 2019] M. Fontana and M. Zafrullah, "t-local domains and valuation domains", pp. 33–62 in *Advances in commutative algebra*, edited by A. Badawi and J. Coykendall, Springer, 2019. MR

[Gilmer 1972] R. Gilmer, *Multiplicative ideal theory*, Pure and Applied Mathematics **12**, Marcel Dekker, New York, 1972. MR

[Guerrieri 2025] L. Guerrieri, "The reciprocal complements of classes of integral domains", *J. Algebra* **682** (2025), 188–214. MR

[Guerrieri and Loper 2021] L. Guerrieri and K. A. Loper, "On the integral domains characterized by a Bezout property on intersections of principal ideals", *J. Algebra* **586** (2021), 208–231. MR

[Guerrieri et al. 2024] L. Guerrieri, K. A. Loper, and G. Oman, "From ancient Egyptian fractions to modern algebra", *J. Algebra Appl.* (2024), 14 pages.

[Kaplansky 1970] I. Kaplansky, *Commutative rings*, Allyn and Bacon, Boston, MA, 1970. MR

[Matsumura 1986] H. Matsumura, *Commutative ring theory*, Cambridge Studies in Advanced Mathematics **8**, Cambridge Univ. Press, 1986. MR

[Seidenberg 1953] A. Seidenberg, "A note on the dimension theory of rings", *Pacific J. Math.* **3** (1953), 505–512. MR

[Spirito 2025] D. Spirito, "The reciprocal complement of a curve", 2025. arXiv 2501.10094

[Zafrullah 1978] M. Zafrullah, "On finite conductor domains", *Manuscripta Math.* **24**:2 (1978), 191–204. MR

Received July 31, 2024. Revised April 18, 2025.

NEIL EPSTEIN
DEPARTMENT OF MATHEMATICAL SCIENCES
GEORGE MASON UNIVERSITY
FAIRFAX, VA
UNITED STATES
nepstei2@gmu.edu

LORENZO GUERRIERI
INSTYTUT MATEMATYKI
JAGIELLONIAN UNIVERSITY
KRAKÓW
POLAND
lorenzo.guerrieri@uj.edu.pl

K. ALAN LOPER
DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY
NEWARK
UNITED STATES
lopera@math.ohio-state.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 338, No. 2, 2025
https://doi.org/10.2140/pjm.2025.338.295

ON A-PACKETS CONTAINING UNITARY LOWEST-WEIGHT REPRESENTATIONS OF $U(p,q)$

SHUJI HORINAGA

We determine all the Arthur packets containing an irreducible unitary lowest-weight representation π of a real unitary group $G = U(p,q)$, including nonscalar cases. Our methods are the Barbasch–Vogan parametrization of representations of G and Trapa's algorithm to calculate the cohomological inductions. In particular, we show that an Arthur packet has at most one irreducible unitary lowest-weight representation of G. As a consequence, if an irreducible unitary lowest-weight representation π exists in the Arthur packet of ψ, we give an explicit formula for the lowest K-type of π.

1. Introduction

One of the fundamental problems in number theory is to investigate arithmetic properties of holomorphic cusp forms on Hermitian symmetric spaces and the geometry of Shimura varieties. For instance, in the case of Siegel modular forms, the detailed analysis of Fourier coefficients and zeta integrals leads to arithmeticity of standard L-values and the cohomology of Siegel modular varieties. In contrast, holomorphic cusp forms associated with unitary groups remain less understood despite their importance.

Recently, there has been significant progress in Arthur's endoscopic classification of automorphic representations on classical groups, and now it is possible to study the automorphic forms systematically. When we try to apply these advances to the study of holomorphic cusp forms on Hermitian symmetric spaces, we face problems in the local representation theory. A key local question is how to classify the local Arthur packets containing a given unitary lowest-weight representation. In this paper, we completely determine the local Arthur packets containing a given unitary lowest-weight representation for unitary groups using the Barbasch–Vogan parametrization of irreducible admissible representations.

To state the main theorem, we recall lowest-weight representations and Mœglin and Renard's description of Arthur packets. Let $G = U(p,q)$ with $N = p+q$ and K be the maximal compact subgroup of G. With the usual choice of positive roots, the

MSC2020: 11F55, 11F70, 17B10, 22E50.

Keywords: unitary groups, lowest-weight representations, Arthur classification.

highest weights of irreducible representations of K are $(\lambda_1,\dots,\lambda_p,\lambda_{p+1},\dots,\lambda_N)\in\mathbb{Z}^N$ satisfying

$$\lambda_1\geq\cdots\geq\lambda_p,\quad \lambda_{p+1}\geq\cdots\geq\lambda_N.$$

We denote by λ the irreducible representation of K with highest-weight λ for short. For each irreducible representation λ of K, there exists a unique irreducible lowest-weight representation π_λ with the lowest K-type λ. For an irreducible representation π, let χ_π be the infinitesimal character of π.

Let us recall the Arthur classification for G. The Arthur classification associates the A-parameters ψ with a finite set $\Pi(\psi)$, called the Arthur packet (or A-packet) for ψ, consisting of unitary representations of finite length of G. The A-parameters are equivalence classes of N-dimensional representations of $\mathbb{C}^\times\times\mathrm{SL}_2(\mathbb{C})$ with suitable properties. The A-packet satisfies several properties that can be easily investigated by ψ. For example, all the representations in $\Pi(\psi)$ have the same infinitesimal characters χ_ψ, and their Harish-Chandra parameter is given by the exponents in the representation $\mathbb{C}^\times\to\mathrm{GL}_N(\mathbb{C})$ defined by $z\mapsto\psi\big(z,\mathrm{diag}((z/\bar{z})^{1/2},(z/\bar{z})^{-1/2})\big)$. When the corresponding Harish-Chandra parameter for ψ is integral, the parameter ψ is called good or good parity. The good A-parameters ψ of G can be viewed as a formal sum

$$\psi=\bigoplus_{i=1}^{r}\chi_{t_i}\otimes S_{a_i}$$

such that $\sum_i a_i=N$, S_m is the m-dimensional irreducible representation of $\mathrm{SL}_2(\mathbb{C})$ and $t_i+a_i+N\in2\mathbb{Z}$, where χ_t is the character of $\mathbb{C}^\times$ defined by $\chi_t(z)=z^{t/2}\bar{z}^{-t/2}$. Suppose $t_i\geq t_{i+1}$ and $a_i\geq a_{i+1}$ if $t_i=t_{i+1}$. Put

$$\mathcal{D}(\psi)=\Big\{(p_i,q_i)\in(\mathbb{Z}_{\geq0}\times\mathbb{Z}_{\geq0})^r\mid p_i+q_i=a_i,\ \sum_{i=1}^{r}p_i=p,\ \sum_{i=1}^{r}q_i=q\Big\}.$$

For each $\underline{d}\in\mathcal{D}(\psi)$, we will attach a cohomological induction $\mathscr{A}_{\underline{d}}(\psi)=A_{\mathfrak{q}(x_{\underline{d}})}(\mu_{\underline{d}})$; see (4-3). By [Mœglin and Renard 2019, Théorème 1.1], the A-packet $\Pi(\psi)$ is equal to the set

$$\Pi(\psi)=\{\mathscr{A}_{\underline{d}}(\psi)\mid\underline{d}\in\mathcal{D}(\psi)\}.$$

For a good A-parameter $\psi=\bigoplus_{i=1}^{r}\chi_{t_i}\otimes S_{a_i}$, let $j=j(\psi)$ be the minimal number i so that $\sum_{\ell=1}^{i}a_\ell\geq p$. Put $a_{<i}=\sum_{\ell<i}a_\ell$ and $a_{>i}=\sum_{\ell>i}a_\ell$. Define $\underline{d}_0\in\mathcal{D}(\psi)$ by

$$\underline{d}_0=\underline{d}_0(\psi)=\{(a_1,0),\dots,(a_{j-1},0),(p_j,q_j),(0,a_{j+1}),\dots,(0,a_r)\}$$

where $p_j=p-a_{<j}$, $q_j=q-a_{>j}$. Let ν_i be the segment $\big[\frac{1}{2}(t_i-a_i+1),\frac{1}{2}(t_i+a_i-1)\big]$ and

$$\nu_{<i}=\bigsqcup_{k<i}\nu_k,\quad \nu_{>i}=\bigsqcup_{i<k}\nu_k.$$

Here, we consider the union as multisets. The multisets $\nu_{\leq i}$ and $\nu_{\geq i}$ are defined similarly.

For an irreducible representation $\lambda = (\lambda_1, \dots, \lambda_N)$ of K, put

$$p' = p'(\lambda) = \#\{i \mid \lambda_p = \lambda_i,\ 1 \leq i \leq p\}, \quad q' = q'(\lambda) = \#\{i \mid \lambda_i = \lambda_{p+1},\ p+1 \leq i \leq N\}.$$

Set

$$P = P(\lambda) = \left\{\lambda_p - \tfrac{1}{2}(N-1), \lambda_{p-1} - \tfrac{1}{2}(N-1) + 1 \dots, \lambda_1 + \tfrac{1}{2}(p-q-1)\right\},$$
$$Q = Q(\lambda) = \left\{\lambda_N + \tfrac{1}{2}(p-q+1), \lambda_{N-1} + \tfrac{1}{2}(p-q+3), \dots, \lambda_{p+1} + \tfrac{1}{2}(N-1)\right\}.$$

The multiset $P \sqcup Q$ can be identified with the infinitesimal character of the lowest-weight representation π_λ. We define the segments P' and Q' by

$$P' = \left[\lambda_p - \tfrac{1}{2}(N-1), \lambda_p - \tfrac{1}{2}(N+1) + p'\right],$$
$$Q' = \left[\lambda_{p+1} + \tfrac{1}{2}(N+1) - q', \lambda_{p+1} + \tfrac{1}{2}(N-1)\right].$$

Put $I = P' \cap Q'$.

Lemma 1.1 (Lemma 4.6). *Let $\psi = \bigoplus_{i=1}^r \chi_{t_i,s} \otimes S_{a_i}$ be an A-parameter. If $\Pi(\psi)$ contains an irreducible lowest-weight representation π, the parameter ψ is good and $\chi_\psi = \chi_\pi$. Moreover, if $\mathscr{A}_{\underline{d}}(\psi) \in \Pi(\psi)$ is nonzero and lowest weight, there exists j such that $q_i = 0$ for any $i < j$ and $p_\ell = 0$ for any $\ell > j$, i.e., $\underline{d} = \underline{d}_0$.*

As a consequence of this lemma, there exists at most one unitary lowest-weight representation in $\Pi(\psi)$. We now state the main theorem of the present paper.

Theorem 1.2 (Theorem 4.7). *Let $\lambda = (\lambda_1, \dots, \lambda_N)$ be an irreducible representation of K and π_λ be the irreducible lowest-weight representation with lowest K-type λ. Suppose that ψ is a good A-parameter such that $\chi_\psi = \chi_{\pi_\lambda}$ and $\mathscr{A}_{\underline{d}_0}(\psi)$ is nonzero.*

(1) *If $N - p' \leq \lambda_p - \lambda_{p+1} < N - q'$, the packet $\Pi(\psi)$ contains π_λ if and only if $\left[\lambda_p - \tfrac{1}{2}(N-1), \lambda_{p+1} + \tfrac{1}{2}(N-1)\right] \subset \nu_j \subset P'$.*

(2) *If $N - q' \leq \lambda_p - \lambda_{p+1} < N - p'$, the packet $\Pi(\psi)$ contains π_λ if and only if either*

- *$\nu_{\leq j} = P$, or*
- *$\left[\lambda_p - \tfrac{1}{2}(N-1), \lambda_{p+1} + \tfrac{1}{2}(N-1)\right] \subset \nu_j \subset Q'$.*

(3) *If $N - p', N - q' \leq \lambda_p - \lambda_{p+1}$, the packet $\Pi(\psi)$ contains π_λ if and only if either*

- *$P \subset \nu_{\leq j} \subset P \sqcup I$, or*
- *$I \subset \nu_j \subset Q'$.*

(4) *If $\lambda_p - \lambda_{p+1} < N - p', N - q'$, the packet $\Pi(\psi)$ contains π_λ if and only if $\left[\lambda_p - \tfrac{1}{2}(N-1), \lambda_{p+1} + \tfrac{1}{2}(N-1)\right] = \nu_j$.*

Conversely, we have the following:

Theorem 1.3 (Corollary 4.8). *Let $\psi = \bigoplus_{i=1}^{r} \chi_{t_i} \otimes S_{a_i}$ be a good A-parameter. The packet $\Pi(\psi)$ contains a nonzero unitary lowest-weight representation if and only if both*

- *$\nu_{<j}$ and $\nu_{>j}$ are multiplicity free, and*
- *$\#(\nu_j \cap \nu_{>j}) \leq p_j$ and $\#(\nu_j \cap \nu_{<j}) \leq q_j$.*

When $\Pi(\psi)$ contains a nonzero unitary lowest-weight representation π in $\Pi(\psi)$, the lowest K-type λ of π is given as follows:

(1) *When $q_j = 0$, the lowest K-type λ of π satisfies $P(\lambda) = \nu_{\leq j}$ and $Q(\lambda) = \nu_{>j}$.*

(2) *When $p_j = \#(\nu_j \cap \nu_{>j})$ and $q_j \neq 0$, the lowest K-type λ of π satisfies $P(\lambda) = \nu_{<j} \sqcup (\nu_j \cap \nu_{>j})$ and $Q(\lambda) = \nu_{\geq j} \setminus (\nu_j \cap \nu_{>j})$.*

(3) *When $q_j = \#(\nu_j \cap \nu_{<j}) \neq 0$, the lowest K-type λ of π satisfies*

$$P(\lambda) = \nu_{\leq j} \setminus (\nu_{<j} \cap \nu_j) \quad \text{and} \quad Q(\lambda) = (\nu_j \cap \nu_{<j}) \sqcup \nu_{>j}.$$

(4) *When $p_j \neq \#(\nu_j \cap \nu_{>j})$ and $q_j \neq \#(\nu_j \cap \nu_{<j})$, put $\nu_{<j} \sqcup \nu_{>j} = \{\sigma_1, \ldots, \sigma_{N-\#(\nu_j)}\}$. Let i_0 be the minimal integer such that*

$$1 \leq i_0 \leq N - \#(\nu_j) \quad \text{and} \quad \#(\nu_j) - i_0 + 1 + \#\{x \in \nu_{<j} \sqcup \nu_{>j} \mid x > \nu_{j,i_0}\} = p.$$

Then, the lowest K-type $\lambda = (\lambda_1, \ldots, \lambda_N)$ of π is given by

$$\lambda_i = \begin{cases} \sigma_i - \frac{1}{2}(p - q + 1) + i & \text{if } i < p - \#(\nu_j) + i_0, \\ \nu_{j,1} + \frac{1}{2}(N+1) - \#(\nu_j) & \text{if } p - \#(\nu_j) + i_0 \leq i \leq p, \\ \nu_{j,1} - \frac{1}{2}(N-1) & \text{if } p+1 \leq i \leq p + i_0 - 1, \\ \sigma_{i-\#(\nu_j)} - \frac{1}{2}(N+1) - p + i & \text{if } p + i_0 \leq i. \end{cases}$$

To conclude the introduction, we give some remarks and one application of the present paper. In the proof of the main theorem, we do not calculate the K-types of cohomological inductions except for special cases. Our proof is based on the Barbasch–Vogan parametrization of representations of G. This parametrization says that for any irreducible representation π, the map $\pi \mapsto (\mathrm{Ann}(\pi), \mathrm{AS}(\pi))$ is injective, where $\mathrm{Ann}(\pi)$ is the annihilator and $\mathrm{AS}(\pi)$ is the asymptotic support. The invariants $\mathrm{Ann}(\pi)$ and $\mathrm{AS}(\pi)$ can be described as certain tableaux in our case. Trapa [2001] gave an algorithm to compute such invariants for cohomological inductions $A_{\mathfrak{q}}(\mu)$. We calculate the tableaux and investigate the conditions where the cohomological inductions $A_{\mathfrak{q}}(\mu)$ are isomorphic to a given π_λ. When G is not a unitary group, the problem becomes complicated, and there are at least two difficulties that do not occur in the unitary group case: one, the description of unipotent A-parameters and two, the reducibility of $A_{\mathfrak{q}}(\mu)$ in the weakly fair range.

Finally, the results of this paper have been applied to the birational geometry of Shimura varieties of $U(1, n)$ in [Horinaga et al. 2025]. In the study of the Kodaira dimension of Shimura varieties, it is crucial to show that the existence of low-weight cusp forms, i.e., cusp forms whose weights are less than that of the discrete series. The method is based on Arthur's multiplicity formula, and the result in the present paper plays a vital role.

2. Unitary groups and representations

Here, we review definitions of unitary groups and representations. We also recall the Barbasch–Vogan parametrization of the representations, which is key to our study.

2.1. ***Tableau notation.*** By a segment, we mean a set of the form $\{a, a+1, \dots, a+n\}$, say $[a, a+n]$, for a real number a and $n \in \mathbb{Z}_{\geq 0}$. In this paper, the segments are always regarded as multisets, that is, sets with multiplicities. For segments $\nu_1 = [a, b]$ and $\nu_2 = [c, d]$, we say that ν_1 and ν_2 are linked (resp. $\nu_1 \leq \nu_2$) if either $c = b+1$ or $a = d+1$ (resp. $a \leq c$ and $b \leq d$) holds.

For a partition $n = n_1 + \cdots + n_\ell$ with $n_1 \geq \cdots \geq n_\ell$, we have a diagram with n_i boxes in the i-th row. This diagram is called a Young diagram of size $n = n_1 + \cdots + n_\ell$. If $\nu = (\nu_1, \dots, \nu_n)$ is an n-tuple of real numbers, a ν-quasitableau is defined as a tableau such that the shape is a Young diagram of size n and the entries are an arrangement of $\nu_1, \dots, \nu_n$. For a ν-quasitableau T, we say that T is ν-antitableau if entries strictly decrease down along each column and weakly decrease along each row. This definition is the same as the definition of the semistandard tableau by replacing "decreasing" with "increasing."

We define a (p, q)-signed tableau as an equivalence class of Young diagrams whose boxes are p plus boxes and q minus boxes so that the signs alternate across the row. Here, we say that two signed tableaux are equivalent if the signatures (p, q) are the same and coincide by interchanging rows of the same length. For a tableau T with entries, we say that the (a, b)-th entry of T is the entry in the a-th row and the b-th column.

The definitions above are those in the previous works [Huang 2025; Chengyu 2025; Trapa 2001] on the nonvanishing of cohomological inductions. See these references for examples of tableaux.

2.2. ***Unitary groups.*** For a Lie group H, we denote by $\mathfrak{h}$ (resp. $\mathfrak{h}_{\mathbb{C}}$) the Lie algebra of H (resp. complexification of $\mathfrak{h}$) and by $\mathcal{U}(\mathfrak{h}_{\mathbb{C}})$ the universal enveloping algebra of $\mathfrak{h}_{\mathbb{C}}$. Fix a positive integer N with a partition $N = p + q$ and $p, q \geq 0$. We define the unitary group $G = U(p, q)$ by

$$G = U(p,q) = \{g \in \mathrm{GL}_N(\mathbb{C}) \mid {}^t\bar{g} I_{p,q} g = I_{p,q}\}, \quad I_{p,q} = \begin{pmatrix} \mathbb{1}_p & \\ & -\mathbb{1}_q \end{pmatrix}.$$

Here, $\bar{g}$ is the complex conjugate of g. For a Cartan involution $\theta : g \mapsto \mathrm{Ad}(I_{p,q})({}^t\bar{g}^{-1})$, let K be the group of the fixed points of θ. Then, K is a maximal compact subgroup of G, which is isomorphic to $U(p)\times U(q)$. The Cartan involution θ induces an involution θ on $\mathfrak{g}$ and a decomposition

$$\mathfrak{g}=\mathfrak{k}\oplus\mathfrak{p},\quad \mathfrak{g}_{\mathbb{C}}=\mathfrak{k}_{\mathbb{C}}\oplus\mathfrak{p}_{+}\oplus\mathfrak{p}_{-}.$$

Here, $\mathfrak{p}$ is the (-1)-eigenspace of θ on $\mathfrak{g}$, and $\mathfrak{p}_+$ (resp. $\mathfrak{p}_-$) corresponds to the holomorphic (resp. antiholomorphic) tangent space of a Hermitian symmetric space G/K. Let T be the diagonal subgroup of G. Then, T is a Cartan subgroup of G and K. Define $e_i\in\mathfrak{t}_{\mathbb{C}}^{*}$ by $e_i(\mathrm{diag}(t_1,\dots,t_N))=t_i$. We regard $\mathfrak{t}_{\mathbb{C}}^{*}$ as $\mathbb{C}^N$ by the basis $\{e_1,\dots,e_N\}$. Then, the root system $\triangle$ of $\mathfrak{g}_{\mathbb{C}}$ with respect to $\mathfrak{t}_{\mathbb{C}}$ is equal to

$$\triangle=\{\pm(e_i-e_j)\mid 1\le i<j\le N\}.$$

We choose a positive system $\triangle^+$ as

$$\triangle^+=\{e_i-e_j\mid 1\le i<j\le N\}.$$

Let $\triangle_c$ (resp. $\triangle_n$) be the compact (resp. noncompact) root system with the positive system $\triangle_c^+=\triangle_c\cap\triangle^+$ (resp. $\triangle_n^+=\triangle_n\cap\triangle^+$), explicitly,

$$\triangle_c=\{\pm(e_i-e_j)\mid 1\le i<j\le p \text{ or } p+1\le i<j\le N\}$$

and

$$\triangle_n=\{\pm(e_i-e_j)\mid 1\le i\le p<j\le N\}.$$

Then, the root system of $\mathfrak{p}_+$ associated with $\mathfrak{t}_{\mathbb{C}}$ is $\triangle_n^+$. Let $\mathfrak{b}$ be the Borel subalgebra of $\mathfrak{g}_{\mathbb{C}}$ associated with $\triangle^+$ and $\mathfrak{b}^-$ be the opposite of $\mathfrak{b}$. For a Lie subalgebra $\mathfrak{u}$ of $\mathfrak{g}_{\mathbb{C}}$ stable under the adjoint action of $\mathfrak{t}_{\mathbb{C}}$, let $\rho(\mathfrak{u})$ be half the sum of roots in $\mathfrak{u}$. Put $\rho=\rho(\mathfrak{b})$.

For $\lambda=(\lambda_1,\dots,\lambda_N)\in\mathbb{Z}^N$ with $\lambda_1\ge\cdots\ge\lambda_p$ and $\lambda_{p+1}\ge\cdots\ge\lambda_N$, let $F(\lambda)$ be an irreducible representation of K with the highest-weight λ. We often write $F(\lambda)$ as λ, for short. The restriction of an admissible $(\mathfrak{g},K)$-module π to K is decomposed as a direct sum

$$\pi|_K=\bigoplus_{\lambda}F(\lambda)^{\oplus m_\pi(\lambda)},\quad m_\pi(\lambda)\in\mathbb{Z}_{\ge 0},$$

where λ runs over all $\triangle_c^+$-dominant integral weights. The nonnegative integer $m_\pi(\lambda)$ is the multiplicity of $F(\lambda)$ in π. By a K-type of π, we mean $F(\lambda)$ with $m_\pi(\lambda)\neq 0$.

The restriction of π to $\mathcal{Z}(\mathfrak{g}_{\mathbb{C}})$ defines a character χ_π of $\mathcal{Z}(\mathfrak{g}_{\mathbb{C}})$. The character χ_π is called the infinitesimal character of π. By the Harish-Chandra isomorphism, infinitesimal characters are parametrized by $\mathfrak{t}_{\mathbb{C}}^{*}/W$, where $W=W(G;T)$ is the Weyl group of G for T. For $\lambda\in\mathfrak{t}_{\mathbb{C}}^{*}/W$, let χ_λ (or λ for short) be the corresponding character of $\mathcal{Z}(\mathfrak{g}_{\mathbb{C}})$. We say that an infinitesimal character ν is integral if ν is in the image of $\mathbb{Z}^N+\rho$. In this paper, we regard integral infinitesimal characters as multisets with N-elements or an element $(x_1,\dots,x_N)$ in $\mathbb{Z}^N+\frac{1}{2}(N-1)$ with $x_1\ge\cdots\ge x_N$.

2.3. ***Unitary lowest-weight representations.*** For a $(\mathfrak{g}, K)$-module π, we say that π is lowest-weight if there exists $v \in \pi$ such that v generates π and v is annihilated by $\mathfrak{b}^-$. Let $\lambda = (\lambda_1, \dots, \lambda_N) \in \mathbb{Z}^N \subset \mathfrak{t}_{\mathbb{C}}^*$ be a Δ_c^+-dominant integral weight. We regard the irreducible representation $F(\lambda)$ of K as an irreducible $\mathfrak{p}_- \oplus \mathfrak{k}_{\mathbb{C}}$ module by letting $\mathfrak{p}_-$ act trivially. Set

$$N(\lambda) = \mathcal{U}(\mathfrak{g}_{\mathbb{C}}) \otimes_{\mathcal{U}(\mathfrak{p}_- \oplus \mathfrak{k}_{\mathbb{C}})} F(\lambda).$$

The module $N(\lambda)$ is called the parabolic Verma module and has the unique irreducible quotient $L(\lambda)$ by [Humphreys 2008, §9.4]. By [Enright et al. 1983, Theorem 2.4] the module $L(\lambda)$ is unitarizable if $\lambda_p - \lambda_{p+1} \geq N - p' - q'$ where $p' = p'(\lambda) = \#\{i \mid \lambda_i = \lambda_p,\ i \leq p\}$ and $q' = q'(\lambda) = \#\{i \mid \lambda_i = \lambda_{p+1},\ p+1 \leq i \leq N\}$. In particular, $L(\lambda)$ is a discrete series representation if $\lambda_p - \lambda_{p+1} > N - 1$, and is limits of discrete series if $\lambda_p - \lambda_{p+1} = N - 1$. The infinitesimal character of π_λ equals

$$\big(\lambda_1 + \tfrac{1}{2}(p-q-1), \dots, \lambda_p - \tfrac{1}{2}(N-1), \lambda_{p+1} + \tfrac{1}{2}(N-1), \dots, \lambda_N + \tfrac{1}{2}(p-q+1)\big).$$

Note that the infinitesimal character χ_{π_λ} is integral by $\lambda \in \mathbb{Z}^N$.

2.4. ***Barbasch–Vogan parametrization of representations of G.*** For the details of this subsection, we refer to [Trapa 2001, §4,5; Barbasch and Vogan 1983]. We introduce two invariants $\mathrm{Ann}(\pi)$ and $\mathrm{AS}(\pi)$ associated to $(\mathfrak{g}, K)$-modules π. For a $(\mathfrak{g}, K)$-module π, let $\mathrm{Ann}(\pi)$ be the annihilator of π in $\mathcal{U}(\mathfrak{g}_{\mathbb{C}})$. When π is irreducible, the ideal $\mathrm{Ann}(\pi)$ is called a primitive ideal. Primitive ideals for an irreducible representation with infinitesimal character ν are parametrized by ν-antitableau. The asymptotic support $\mathrm{AS}(\pi)$ is defined via the local behavior of the character of π. By definition, $\mathrm{AS}(\pi)$ is a union of nilpotent orbits of $\mathfrak{g}_{\mathbb{C}}$.

Theorem 2.1 [Barbasch and Vogan 1983, Theorem 4.2; Trapa 2001, Theorem 6.1]. *For irreducible $(\mathfrak{g}, K)$-modules π and π' with integral infinitesimal characters, the representation π is isomorphic to π' if and only if*

$$(\mathrm{Ann}(\pi), \mathrm{AS}(\pi)) = (\mathrm{Ann}(\pi'), \mathrm{AS}(\pi')).$$

Trapa and Vogan [Trapa 2001, Conjecture 1.1] conjectured that the cohomological inductions in the weakly fair range exhaust the unitary $(\mathfrak{g}, K)$-modules with integral infinitesimal characters. As far as the author knows, the conjecture has been proven only for specific cases such as $U(n, 1)$ and $U(n, 2)$ (see [Wong and Zhang 2024] for details).

3. Cohomological inductions

In this section, we introduce the cohomological inductions $A_{\mathfrak{q}}(\mu)$, recall their basic properties, review Trapa's algorithm to determine the tableaux for $A_{\mathfrak{q}}(\mu)$, and state a nonvanishing criterion for certain $A_{\mathfrak{q}}(\mu)$.

3.1. ***θ*-stable parabolic subalgebras and cohomological induction.** Take $x \in \sqrt{-1}\mathfrak{k}$. Since the action $\mathrm{ad}(x)$ on $\mathfrak{g}_{\mathbb{C}}$ is diagonalizable with real eigenvalues, we define the subalgebras of $\mathfrak{g}_{\mathbb{C}}$ by

$$\begin{aligned}\mathfrak{q} = \mathfrak{q}(x) &= \text{sum of root vectors with nonnegative eigenvalues},\\ \mathfrak{u} = \mathfrak{u}(x) &= \text{sum of root vectors with positive eigenvalues},\\ \mathfrak{l} = \mathfrak{l}(x) &= \text{sum of root vectors with zero eigenvalues}.\end{aligned}$$

We call parabolic subalgebras $\mathfrak{q}$ of $\mathfrak{g}_{\mathbb{C}}$ obtained in the above way θ-stable parabolic subalgebras of $\mathfrak{g}_{\mathbb{C}}$. By a conjugate of an element in K, we may assume $x \in \sqrt{-1}\,\mathfrak{t}$. For $\mu \in \mathfrak{t}_{\mathbb{C}}^*$, let $\mathbb{C}_\mu$ denote the character of $\mathfrak{l}$ with $\mathbb{C}_\mu|_{\mathfrak{t}_{\mathbb{C}}} = \mu$, if it exists. We then obtain the cohomological induction $A_{\mathfrak{q}}(\mu)$ by the induction of $\mathbb{C}_\mu$ as in [Knapp and Vogan 1995, (5.6)].

For $G = U(p,q)$, θ-stable parabolic subalgebras $\mathfrak{q}(x)$ arise from $\underline{d} = \{(p_i, q_i)_{1 \le i \le r}\}$ and an element $x_{\underline{d}}$ in $\sqrt{-1}\,\mathfrak{t} \cong \mathbb{R}^N$ of the form

$$(3\text{-}1)\quad x_{\underline{d}} = (\underbrace{x_1, \dots, x_1}_{p_1}, \dots, \underbrace{x_r, \dots, x_r}_{p_r}, \underbrace{x_1, \dots, x_1}_{q_1}, \dots, \underbrace{x_r, \dots, x_r}_{q_r}), \quad x_1 > \cdots > x_r,$$

such that $(p_i, q_i) \in (\mathbb{Z}_{\ge 0})^2$ with $(p_i, q_i) \neq (0,0)$ for any i and $p_1 + \cdots + p_r = p$, $q_1 + \cdots + q_r = q$. Set $\mathfrak{q}_{\underline{d}} = \mathfrak{q}(x_{\underline{d}})$. The centralizer $C_G(\mathfrak{q}_{\underline{d}})$ is a connected reductive group $L_{\underline{d}}$ isomorphic to

$$L_{\underline{d}} \cong U(p_1, q_1) \times \cdots \times U(p_r, q_r)$$

such that $\mathfrak{l}_{\underline{d}} = \mathfrak{l}(x_{\underline{d}})_{\mathbb{C}} = \mathrm{Lie}(L_{\underline{d}}) \otimes \mathbb{C}$. Note that our choice of $\mathfrak{q}_{\underline{d}}$ or the hermitian form $I_{p,q}$ is different from that of [Mœglin and Renard 2019; Trapa 2001; Vogan 1997], but the same as [Ichino 2022]. We say that a θ-stable parabolic subalgebra $\mathfrak{q}$ is holomorphic if there exists j such that $q_i = 0$ for any $i < j$ and $p_\ell = 0$ for any $\ell > j$, in other words, $\mathfrak{u} \cap \mathfrak{p}_{\mathbb{C}} \subset \mathfrak{p}_+$. The choices for a holomorphic θ-stable parabolic subalgebra are the same as [Trapa 2001; Vogan 1997].

3.2. ***Properties of $A_{\mathfrak{q}}(\mu)$.*** Here, we review the basic properties of $A_{\mathfrak{q}}(\mu)$. Let $\underline{d} = \{(p_i, q_i)_{1 \le i \le r}\}$ be a set of pairs of nonnegative integers with $(p_i, q_i) \neq 0$, $\sum_i p_i = p$, and $\sum_i q_i = q$. Take $\mu \in \mathbb{Z}^N$ such that

$$\mu = (\underbrace{\mu_1, \dots, \mu_1}_{a_1}, \dots, \underbrace{\mu_r, \dots, \mu_r}_{a_r}), \quad a_i = p_i + q_i.$$

We define the segments ν_i associated with $\underline{d}$ and μ by

$$(3\text{-}2)\qquad \nu_i = \nu_i(\underline{d}, \mu) = \left[\mu_i + \tfrac{1}{2}(N+1) - a_{\le i},\, \mu_i + \tfrac{1}{2}(N-1) - a_{<i}\right].$$

Here, $a_{<i} = \sum_{\ell < i} a_\ell$ and $a_{\le i} = \sum_{\ell \le i} a_\ell$. We denote the cohomological induction $A_{\mathfrak{q}_{\underline{d}}}(\mu)$ by $A(\mathfrak{q}_{\underline{d}}, \nu_1, \dots, \nu_r) = A(\underline{d}, \nu_1, \dots, \nu_r)$ for short. We say that the cohomological induction $A_{\mathfrak{q}_{\underline{d}}}(\mu)$ is in the weakly fair range (resp. mediocre range) if $\mu_i - \mu_{i+1} \geq -\frac{1}{2}(a_i + a_{i+1})$ for any i (resp. $\mu_i - \mu_j \geq -\max\{a_i, a_j\} - \sum_{i<k<j} a_k$ for any $i < j$). We also say that $\nu = \mu + \rho = \bigsqcup_i \nu_i$ is in the weakly fair range (resp. mediocre range) if $A_{\mathfrak{q}_{\underline{d}}}(\mu)$ is so.

The cohomological induction $A_{\mathfrak{q}_{\underline{d}}}(\mu)$ is in the weakly fair range (resp. mediocre range) if and only if the mean value in ν_i is greater than or equal to the mean value in ν_{i+1} (resp. $\nu_i \not\succeq \nu_j$ for any $i < j$) by the explicit calculation (see [Chengyu 2025, Lemma 2.4]).

The following statements are well known (for example, see [Knapp and Vogan 1995; Adams 1987, Lemma 4.2; Huang and Pandžić 2006, Theorem 6.4.4; Trapa 2001, Theorem 3.1 (iv)]):

- In the weakly fair range, $A_{\mathfrak{q}}(\mu)$ is unitarizable. Moreover, it is zero or irreducible for $G = U(p,q)$.
- For $A_{\mathfrak{q}}(\mu)$ in the mediocre range,

$$\dim \operatorname{Hom}_K(F(\lambda), A_{\mathfrak{q}}(\mu)) = \sum_{w \in W^1} \operatorname{sgn}(w) P_{\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}}}\big(w(\lambda+\rho_c) - (\mu + 2\rho(\mathfrak{u}\cap\mathfrak{p}) + \rho_c)\big)$$

holds. Here, W^1 is the subgroup of the Weyl group of K with respect to T consisting of w for which $\alpha \in \Delta_c^+$, $w^{-1}\alpha < 0$ implies $\alpha \in \Delta(\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}})$, and $P_{\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}}}$ is the partition function with respect to $\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}}$, i.e., $P_{\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}}}(x)$ is the multiplicity of weight x in the symmetric algebra $S(\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}})^{\mathfrak{l}\cap\mathfrak{k}\cap\mathfrak{n}}$.

- If $\mu + 2\rho(\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}})$ is Δ_c^+-dominant, $\mu + 2\rho(\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}})$ occurs in $A_{\mathfrak{q}}(\mu)$ and K-types in $A_{\mathfrak{q}}(\mu)$ are of the form

$$\mu + 2\rho(\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}}) + \sum_{\alpha\in\Delta(\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}})} n_\alpha \alpha, \quad n_\alpha \geq 0. \tag{3-3}$$

Calculating the K-type formula directly seems complicated, even when $\mathfrak{q}$ is a holomorphic θ-stable parabolic subalgebra. To avoid such complexity, we will calculate the tableaux described below instead.

3.3. *Tableaux associated with $A_{\mathfrak{q}}(\mu)$.* Let π be a cohomological induction in the weakly fair range. We associate two invariants $\operatorname{Ann}(\pi)$ and $\operatorname{AS}(\pi)$ to π in Section 2.4. The annihilator $\operatorname{Ann}(\pi)$ can be regarded as a ν-antitableau, but $\operatorname{AS}(\pi)$ is a union of unipotent orbits. If π is isomorphic to a cohomological induction $A_{\mathfrak{q}}(\mu)$ in the weakly fair range, the asymptotic support $\operatorname{AS}(\pi)$ is a single unipotent orbit by [Trapa 2001, Proposition 5.4]. Indeed, for $A_{\mathfrak{q}}(\mu)$ in the good range, its asymptotic support is a single unipotent orbit. Since we can obtain π as a translation of cohomological inductions in the good range, one can show that the asymptotic supports coincide. Hence, $\operatorname{AS}(A_{\mathfrak{q}}(\mu))$ is a single unipotent orbit. Thus, we may associate a ν-antitableau and a (p,q)-signed tableau for a cohomological induction in the weakly fair range by [Collingwood and McGovern 1993, Theorem 9.3.3]. The examples of tableaux associated with $A_{\mathfrak{q}}(\mu)$ are available in [Huang 2025; Chengyu 2025; Trapa 2001].

In the following, we associate tableaux for $A_{\mathfrak{q}}(\mu)$. Let $\mathfrak{q}$ be the θ-stable parabolic subalgebra corresponding to $\{(p_i, q_i)_{1\le i\le r}\}$ and μ be a $\triangle_c^+$-dominant integral weight such that $A_{\mathfrak{q}}(\mu)$ is in the mediocre range. We first construct the (p, q)-signed tableau inductively. Let S_1 be the Young diagram of size $1+\cdots+1$ with p_1+q_1 boxes filled with p_1 plus boxes and q_1 minus boxes. Suppose that the $\big(\sum_{i<k} p_i, \sum_{i<k} q_i\big)$-signed tableau $\bigsqcup_{i<k} S_i$ for $k\le r$ is defined. We then construct the signed tableau $\bigsqcup_{i\le k} S_i$ by adding p_k boxes filled with $+$ and q_k boxes filled with $-$, from top to bottom, to each row-ends of $\bigsqcup_{i<k} S_i$ such that

- at most one box is added in each row-end, and
- the signs in $\bigsqcup_{i\le k} S_i$ are alternating across the row.

Then, $\bigsqcup_{i\le k} S_i$ is defined as a Young diagram with decreasing rows by rearranging the rows. The resulting tableau $S=\bigsqcup_{i\le r} S_i$ is the asymptotic support of $A_{\mathfrak{q}}(\mu)$.

We next construct the ν-antitableau. For $A_{\mathfrak{q}}(\mu)$, the shape of (p, q)-signed tableau and the ν-antitableau are the same. The shape S of the (p, q)-signed tableau is partitioned into $\bigsqcup_{1\le i\le r} S_i$, which is the same as in the definition of the (p, q)-signed tableau. For each S_i, we fill $\nu_{i,1}, \dots, \nu_{i,a_i}$ from top to bottom, where $\nu_i=\nu_i(\underline{d}, \mu)=\{\nu_{i,1}, \dots, \nu_{i,a_i}\}$ with $\nu_{i,1}>\cdots>\nu_{i,a_i}$. Then, S is a ν-quasitableau. When the S is a ν-antitableau, let $\mathrm{Ann}(A_{\mathfrak{q}}(\mu))=S$, which is possibly equivalent to the formal zero tableau explained below. We introduce the two invariants $\mathrm{overlap}(S_i, S_{i+1})$ and $\mathrm{sing}(S_i, S_{i+1})$ associated with cohomological inductions. Set

$$\mathrm{sing}(S_i, S_{i+1})=\#(\nu_i\cap\nu_{i+1}).$$

For S_i and S_{i+1} as in the definition of (p, q)-signed tableau $S=\bigsqcup_i S_i$, let m be the largest integer such that for any i with $i\le m$, the (a_i-m+i)-th box of S_i is strictly left to the i-th box of S_{i+1}. If such an integer m does not exist, set $m=0$. We then define $\mathrm{overlap}(S_i, S_{i+1})$ by the nonnegative integer m. In the following, we give an algorithm [Trapa 2001, Procedure 7.5] to obtain a ν-antitableau or the formal zero tableau from the tableau S. By [Trapa 2001, Theorem 7.9], the cohomological induction $A_{\mathfrak{q}}(\mu)$ is zero if and only if the tableau S is equivalent to the formal zero tableau. More precisely, let $S'=\bigsqcup_i S_i'$ be the tableau after Trapa's algorithm. Let ν_i' be the segment consisting of the entries of S_i'. The cohomological induction $A_{\mathfrak{q}}(\mu)$ is nonzero if and only if the resulting tableau $S'=\bigsqcup_i S_i'$ satisfies

- $\nu_i\ge\nu_{i+1}$, and
- $\mathrm{overlap}(S_i', S_{i+1}')\ge\mathrm{sing}(S_i', S_{i+1}')$ for any i.

We give Trapa's algorithm to transform the tableau to ν-antitableau or the formal zero tableau. For $S=\bigsqcup_i S_i$, the algorithm is generated by replacing adjacent skew columns $S_i \sqcup S_{i+1}$ with $S'_i \sqcup S'_{i+1}$. Set $R = S_i \sqcup S_{i+1}$ and $R' = S'_i \sqcup S'_{i+1}$. When R' is the formal zero tableau, we understand that the tableau S is equivalent to the formal zero tableau. We review the construction of R', which is an arrangement of $\nu_i \sqcup \nu_{i+1}$ with the shape R.

(1) If $\mathrm{overlap}(S_i, S_{i+1}) = \mathrm{sing}(S_i, S_{i+1}) < \min\{a_i, a_{i+1}\}$ or $\mathrm{overlap}(S_i, S_{i+1}) > \mathrm{sing}(S_i, S_{i+1})$, set $R' = R$. Here, $a_i = \#(\nu_i)$.

(2) If $\mathrm{overlap}(S_i, S_{i+1}) < \mathrm{sing}(S_i, S_{i+1})$, then R' is the formal zero tableau.

(3) Assume $\mathrm{overlap}(S_i, S_{i+1}) = \mathrm{sing}(S_i, S_{i+1}) = a_{i+1}$. In this case, $\nu_{i+1} \subset \nu_i$. We define R' by induction on $m = \nu_{i+1,a_{i+1}} - \nu_{i,a_i}$. When $m = 0$, set $R' = R$. When $m > 0$, set $\nu_{i+1}(-) = \{\nu_{i+1,1} - 1, \ldots, \nu_{i+1,a_{i+1}} - 1\}$. We define $S_{i+1}(-)$ by the tableau with the shape S_{i+1} filled with $\nu_{i+1}(-)$. Set $R(-) = S_i \sqcup S_{i+1}(-)$. By the induction hypothesis, $R(-)'$ is defined. Then, in $R(-)'$, there exists at most one box B filled with $\nu_{i+1,1} - 1$ and strictly to the right of the unique box filled with $\nu_{i+1,1}$ in $R(-)'$. If the box B exists, add one to the entry in B. If no such box exists, add one to the entry in the left-most box filled with $\nu_{i+1,1} - 1$ in $R(-)'$. We denote the resulting tableau by $R(-)'_1$. Now construct $R(-)'_2$ by the same procedure applied to $R(-)'_1$, but instead considering the entries $\nu_{i+1,2}$ and $\nu_{i+1,2} - 2$. By the same procedure, we get $R(-)'_{a_{i+1}}$. Set $R' = R(-)'_{a_{i+1}}$.

(4) Assume $\mathrm{overlap}(S_i, S_{i+1}) = \mathrm{sing}(S_i, S_{i+1}) = a_i$. In this case $\nu_i \subset \nu_{i+1}$. We define R' by induction on $m = \nu_{i+1,1} - \nu_{i,1}$. When $m = 0$, set $R' = R$. When $m > 0$, set $\nu_i(+) = \{\nu_{i,1} + 1, \ldots, \nu_{i,a_i} + 1\}$. We define the tableau $S_i(+)$ by the tableau with the shape S_i filled with $\nu_i(+)$. Set $R(+) = S_i(+) \sqcup S_{i+1}$. By the induction hypothesis, $R(+)'$ is defined. Then, in $R(+)'$, there exists at most one box B filled with $\nu_{i,a_i} + 1$ and strictly to the left of the unique box filled with ν_{i,a_i} in $R(+)'$. If the box B exists, subtract one from the entry in the box B. If no such box exists, subtract one from the entry in the right-most box filled with $\nu_{i,a_i} + 1$ in $R(+)'$. We denote the resulting tableau by $R(+)'_1$. Now construct $R(+)'_2$ by the same procedure applied to $R(+)'_1$, by the same procedure again, we get $R(+)'_{a_i}$. Set $R' = R(+)'_{a_i}$.

We still need to consider a partition of the resulting tableau R' into $R' = S'_i \sqcup S'_{i+1}$. The last box in R' is the right-most box filled with $\nu'_{i+1,a_{i+1}} = \min\{\nu_{i,a_i}, \nu_{i+1,a_{i+1}}\}$. The box next to the last box is the right-most box filled with $\nu_{i+1,a_{i+1}} + 1$. This procedure stops when the entry of the box reaches $\min\{\nu_{k,1}, \nu_{k+1,1}\}$. Let S'_i be the tableau of the remaining boxes. We then obtain the partition $R' = S'_i \sqcup S'_{i+1}$.

If interested, the author recommends to calculate examples and check the well-definedness of the above definition. The explicit formula for the overlap is investigated in [Chengyu 2025, Theorem 3.6; Huang 2025, Lemma 5.4]. For the explicit description of entries in the tableau R', see [Huang 2025, §4.5].

3.4. ***Unitary lowest-weight representations as cohomological inductions.*** In this subsection, we determine a cohomological induction that is isomorphic to a given lowest-weight representation π_λ. The following is one of the easiest cases in which to calculate the K-types.

Lemma 3.1. *Let $\mathfrak{q}_{\underline{d}}$ be a holomorphic θ-stable parabolic subalgebra corresponding to $\underline{d} = \{(p_i, q_i)_i\}$ such that $p_i q_i = 0$ for any i and $\pi = A_{\mathfrak{q}_{\underline{d}}}(\mu)$ be a cohomological induction in the mediocre range. Put $j = \max\{i \mid q_i = 0\}$. Then, π is nonzero if and only if the multisets $\nu_{\le j}$ and $\nu_{>j}$ are multiplicity free. If π is nonzero, then π is a unitary lowest-weight representation with the lowest K-type*

$$\mu + 2\rho(\mathfrak{u} \cap \mathfrak{p}_{\mathbb{C}}) = \mu + (\underbrace{q, \dots, q}_{p}, \underbrace{-p, \dots, -p}_{q}).$$

Moreover, let

$$\sigma_i = \begin{cases} \mu_i + \frac{1}{2}(p - q + 1) - i & \text{if } 1 \le i \le p, \\ \mu_i + \frac{1}{2}(N+1) - (i-p) & \text{if } p+1 \le i \le N. \end{cases}$$

We denote by i_0 the maximal positive integer such that $\sigma_{p+\min\{p,q\}+1-i_0} \ge \sigma_{p+1-i_0}$, if it exists. If there is no such integer, put $i_0 = 0$. Then, the first column of the tableau $\mathrm{Ann}(\pi)$ consists of

$$\sigma_1, \sigma_2, \dots, \sigma_{p-i_0}, \underbrace{\sigma_{p+\min\{p,q\}+1-i_0}, \dots, \sigma_{p+\min\{p,q\}}}_{i_0}, \sigma_{p+\min\{p,q\}+1}, \dots, \sigma_N$$

and the second column consists of

$$\sigma_{p+1}, \dots, \sigma_{p+\min\{p,q\}-i_0}, \underbrace{\sigma_{p+1-i_0}, \dots, \sigma_p}_{i_0}.$$

Proof. Before applying the algorithm, the first column of the tableau $\mathrm{Ann}(\pi)$ consists of

$$\underbrace{\sigma_1, \dots, \sigma_p}_{p}, \underbrace{\sigma_{p+\min\{p,q\}+1}, \dots, \sigma_N}_{q-\min\{p,q\}}$$

and the second column consists of

$$\sigma_{p+1}, \dots, \sigma_{p+\min\{p,q\}}$$

from top to bottom. To apply Trapa's algorithm, we use the partition $\mathrm{AS}(\pi) = \bigsqcup_i S_i$. Let ν_i be the segment defined in (3-2). Then, $\nu_i = [\sigma_{a_{\le i}}, \sigma_{a_{<i}+1}]$ and the tableaux S_i are filled with ν_i. By the algorithm for the partition $\bigsqcup_{i \le j} S_i$ and $\bigsqcup_{i>j} S_i$, the representation π is zero if $\nu_{\le j}$ or $\nu_{>j}$ is not multiplicity free.

Suppose that $\nu_{\le j}$ and $\nu_{>j}$ are multiplicity free. By the algorithm, the diagram $\bigsqcup_i S_i$ is invariant under the algorithm if $\nu_j \not\subset \nu_{j+1}$ and $\nu_j \not\supset \nu_{j+1}$. Consider the partition $S_j \sqcup S_{j+1}$. When $\nu_j \supset \nu_{j+1}$, the tableau S_j consists only of the first column with the entries $\sigma_{p+1-a_j}, \dots, \sigma_p$ and the tableau S_{j+1} consists only of the second column with the entries $\sigma_{p+1}, \dots, \sigma_{p+a_{j+1}}$. Hence, the tableau $S_j \sqcup S_{j+1}$ is invariant under the algorithm, but the partition $S'_j \sqcup S'_{j+1}$ is different. The tableau S'_j consists only of the first column with entries $\sigma_{p+1-a_j}, \dots, \sigma_{p+a_{j+1}}$ and the tableau S'_{j+1} consists of the remaining boxes. Then, if S'_{j+1} has a box in the first column, the maximal entry is $\sigma_{p+a_{j+1}} - 1$. The tableau $\left(\bigsqcup_{i \le j-1} S_i\right) \sqcup S'_j$ is invariant under the algorithm by definition. Since the entry $\sigma_{p+a_{j+1}} - 1$ is greater than or equal to the maximal entry in S_{j+2}, the tableau $S'_{j+1} \sqcup S_{j+2}$ is invariant under the algorithm. Continuing the same procedure, one obtains the resulting tableau $\mathrm{Ann}(\pi)$. This shows that $i_0 = 0$ and the representation π is nonzero. The lowest K-type of π is given as (3-3), since $\mu + 2\rho(\mathfrak{u} \cap \mathfrak{p}_{\mathbb{C}})$ is Δ_c^+-dominant.

When $\nu_j \subset \nu_{j+1}$, we first assume that $a_{j+1} > \min\{p, q\}$. By the same procedure as above, if $\sigma_p > \sigma_{p+\min\{p,q\}+1}$, the tableau $S'_j \sqcup S'_{j+1}$ is the same as $S_j \sqcup S_{j+1}$ and $i_0 = 0$. The first column of the tableau S'_j consists of $\sigma_{p+1-a_j}, \dots, \sigma_p$ and the second column consists of $\sigma_{p+1}, \dots, \sigma_{p+1-a_j} + 1$. Since $\nu_{\le j}$ is multiplicity free, we have $\sigma_{p-a_j} \ge \sigma_{p+1-a_j}$. The tableau $S_{j-1} \sqcup S'_j$ is stable under the algorithm. Similarly, $S'_{j+1} \sqcup S_{j+2}$ is stable. Hence, the tableau $\bigsqcup_i S_i$ is stable under the algorithm. If $\sigma_p \le \sigma_{p+\min\{p,q\}+1}$, the first column of the tableau $S'_j \sqcup S'_{j+1}$ consists of

$$\underbrace{\sigma_{p+\min\{p,q\}-a_j+1}, \dots, \sigma_{p+\min\{p,q\}}}_{a_j}, \underbrace{\sigma_{p+\min\{p,q\}+1}, \dots, \sigma_{p+a_{j+1}}}_{a_{j+1}-\min\{p,q\}}$$

and the second column consists of

$$\underbrace{\sigma_{p+1}, \dots, \sigma_{p+\min\{p,q\}-a_j}}_{\min\{p,q\}-a_j}, \underbrace{\sigma_{p+1-a_j}, \dots, \sigma_p}_{a_j}.$$

The first column of S'_j consists of

$$\sigma_{p+\min\{p,q\}+1-a_j}, \dots, \sigma_p$$

and the second column consists of

$$\sigma_{p+1}, \dots, \sigma_{p+\min\{p,q\}-a_j}.$$

By the same procedure to the end, the statement follows.

We suppose $\nu_j \subset \nu_{j+1}$ and $a_{j+1} \le \min\{p, q\}$. Then, $i_0 = 0$. Since $\nu_{>j}$ is multiplicity free, σ_p is greater than or equal to $\sigma_{p+\min\{p,q\}}$. In this case, the tableau $S'_j \sqcup S'_{j+1}$ is the same as $S_j \sqcup S'_{j+1}$, but the tableau S'_{j+1} consists only of the second column with entries $\sigma_p + 1, \dots, \sigma_{p+\min\{p,q\}}$. By the routine discussion, the tableau $\bigsqcup_i S_i$ is invariant under the algorithm. □

We explicitly describe π_λ in terms of cohomological induction as follows:

Lemma 3.2. *Let π_λ be a unitary lowest-weight representation with lowest K-type $\lambda = (\lambda_1, \dots \lambda_N)$.*

(1) *When $\lambda_p - \lambda_{p+1} < N - p', N - q'$, let $\mathfrak{q}$ be the θ-stable parabolic subalgebra corresponding to*

$$\Big\{\underbrace{(1,0), \dots, (1,0)}_{p-p'}, (p', N-(\lambda_p - \lambda_{p+1}) - p'), \underbrace{(0,1), \dots, (0,1)}_{\lambda_p - \lambda_{p+1} - p + p'}\Big\}$$

and

$$\mu = (\mu_1, \dots, \mu_N)$$

be an element in $\mathbb{Z}^N$ defined by

$$\mu_i = \begin{cases} \lambda_i - q & \text{if } i \leq p - p', \\ \lambda_{p+1} + p - p' & \text{if } p - p' < i \leq N - (\lambda_p - \lambda_{p+1}) - p', \\ \lambda_i + p & \text{if } N - (\lambda_p - \lambda_{p+1}) - p' < i. \end{cases}$$

Then, $A_\mathfrak{q}(\mu) \cong \pi_\lambda$.

(2) *When $\min\{N - p', N - q'\} \leq \lambda_p - \lambda_{p+1}$, let $\mathfrak{q}$ be the θ-stable parabolic subalgebra corresponding to*

$$\{\underbrace{(1,0), \dots, (1,0)}_{p-p'}, (p', 0), (0, q'), \underbrace{(0,1), \dots, (0,1)}_{q-q'}\}$$

and μ be an element in $\mathbb{Z}^N$ defined by

$$\mu = (\underbrace{\lambda_1 - q, \dots, \lambda_p - q}_{p}, \underbrace{\lambda_{p+1} + p, \dots, \lambda_N + p}_{q}).$$

Then, $A_\mathfrak{q}(\mu) \cong \pi_\lambda$.

(3) *Let $\mathfrak{q}$ be the θ-stable parabolic subalgebra corresponding to*

$$\{\underbrace{(1,0), \dots, (1,0)}_{p-p'}, (p', 0), (0, q'), \underbrace{(0,1), \dots, (0,1)}_{q-q'}\}$$

and μ be an element in $\mathbb{Z}^N$ defined by $\mu = (\mu_1, \dots, \mu_N)$ such that $\mu_1 \geq \cdots \geq \mu_p$, $\mu_{p+1} \geq \cdots \geq \mu_N$, $\mu_p = \mu_{p-1} = \cdots = \mu_{p-p'+1}$, and $\mu_{p+1} = \mu_{p+2} = \cdots = \mu_{p+q'}$. Suppose that $A_\mathfrak{q}(\mu)$ is in the mediocre range. Then, $A_\mathfrak{q}(\mu)$ is a nonzero lowest-weight representation and the lowest K-type $\lambda = (\lambda_1, \dots, \lambda_N)$ of $A_\mathfrak{q}(\mu)$ satisfies

$$\min\{N - p'(\lambda), N - q'(\lambda)\} \leq \lambda_p - \lambda_{p+1},$$

where $p'(\lambda) = \#\{i \mid 1 \leq i \leq p,\ \lambda_i = \lambda_p\}$ and $q'(\lambda) = \#\{i \mid p+1 \leq i \leq N,\ \lambda_i = \lambda_{p+1}\}$.

Proof. We calculate the K-types of $A_{\mathfrak{q}}(\mu)$. For (1), we have

$$2\rho(\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}}) = (\underbrace{q,\dots,q}_{p-p'}, \underbrace{\lambda_p-\lambda_{p+1}-p+p',\dots,\lambda_p-\lambda_{p+1}-p+p'}_{p'}, \underbrace{-p+p',\dots,-p+p'}_{N-(\lambda_p-\lambda_{p+1})-p')}, \underbrace{-p,\dots,-p}_{\lambda_p-\lambda_{p+1}-p+p'})$$

and

$$\mu+2\rho(\mathfrak{u}\cap\mathfrak{p}) = (\lambda_1,\dots,\lambda_p,\lambda_{p+1},\dots,\lambda_N).$$

Since $\mu+2\rho(\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}})$ is $\triangle_c^+$-dominant, the representation $A_{\mathfrak{q}}(\mu)$ is isomorphic to π_λ by (3-3). The statement (2) is a restatement of Lemma 3.1. We prove (3). Since $\mathfrak{q}$ is holomorphic, $A_{\mathfrak{q}}(\mu)$ is a lowest-weight representation by Lemma 3.1. Suppose $A_{\mathfrak{q}}(\mu)\cong\pi_\lambda$. In this case, we have

$$\nu_{p-p'+1} = \left[\mu_p-\tfrac{1}{2}(p-q-1),\,\mu_p-\tfrac{1}{2}(p-q+1)+p'\right]$$

and

$$\nu_{p-p'+2} = \left[\mu_{p+1}-\tfrac{1}{2}(p-q-1)-q',\,\mu_{p+1}-\tfrac{1}{2}(p-q+1)\right].$$

Here, ν_i is the segment defined in (3-2). Since it is in the mediocre range, either

$$\mu_p-\tfrac{1}{2}(p-q+1)+p' \geq \mu_{p+1}-\tfrac{1}{2}(p-q+1)$$

or

$$\mu_p-\tfrac{1}{2}(p-q-1) \geq \mu_{p+1}-\tfrac{1}{2}(p-q-1)-q'$$

holds. This is equivalent to

$$\mu_p-\mu_{p+1}\geq -p' \quad \text{or} \quad \mu_p-\mu_{p+1}\geq -q'.$$

By (2), we have

$$\mu_p=\lambda_p-q \quad \text{and} \quad \mu_{p+1}=\lambda_{p+1}+p.$$

Then, the statement (3) follows from $p'(\lambda)\geq p'$ and $q'(\lambda)\geq q'$. □

The signed tableaux for unitary lowest-weight representations are as follows:

Corollary 3.3. *Let π be a unitary lowest-weight representation. Then the signed tableau* $\mathrm{AS}(\pi)$ *has at most two columns. If the tableau* $\mathrm{AS}(\pi)$ *has only one column, π is a character. If* $\mathrm{AS}(\pi)$ *has two columns, the signs are arranged in the order of $+$ and $-$ in rows with two boxes. In particular, the tableau* $\mathrm{AS}(\pi)$ *is uniquely determined by its shape.*

Proof. This follows from Lemma 3.2 and the definition of (p,q)-signed tableau for $A_{\mathfrak{q}}(\mu)$. □

3.5. ***Nonvanishing criterion for $A_{\mathfrak{q}}(\mu)$.*** For a θ-stable maximal parabolic subalgebra $\mathfrak{q}$, Trapa's algorithm gives the following nonvanishing criterion for $A_{\mathfrak{q}}(\mu)$.

Lemma 3.4 [Huang 2025, Lemma 2.6]. *Let $\mathfrak{q}$ be a θ-stable parabolic subalgebra corresponding to $\underline{d} = \{(p_1, q_1), (p_2, q_2)\}$ and μ be a Δ_c^+-dominant integral weight such that $A_{\mathfrak{q}}(\mu)$ is in the mediocre range. Then, $A_{\mathfrak{q}}(\mu)$ is nonzero if and only if*

$$\min\{p_1, q_2\} + \min\{q_1, p_2\} \geq \#(\nu_1 \cap \nu_2).$$

Take a holomorphic θ-stable parabolic subalgebra associated to $\underline{d}=\{(p_i, q_i)_{1\leq i\leq r}\}$. Let $j = j(\underline{d})$ be the minimal integer i such that $a_{\leq i} \geq p$. The cohomological inductions $A_{\mathfrak{q}}(\mu)$ for holomorphic θ-stable parabolic subalgebras $\mathfrak{q}$ are lowest weight by [Adams 1987, Lemma 1.7]. We will show that the converse holds in Lemma 4.6. We can calculate the nonvanishing conditions and tableaux of such cohomological induction as follows:

Lemma 3.5. *Let $\mathfrak{q}_{\underline{d}}$ be a holomorphic θ-stable parabolic subalgebra corresponding to $\underline{d} = \{(p_i, q_i)_{1\leq i\leq r}\}$ with $a_i = p_i + q_i$ and π be a cohomological induction $A_{\mathfrak{q}_{\underline{d}}}(\mu)$ in the mediocre range. Then, π is nonzero if and only if*

- *$\nu_{<j}$ and $\nu_{>j}$ are multiplicity free, and*
- *$\#(\nu_j \cap \nu_{<j}) \leq q_j$ and $\#(\nu_j \cap \nu_{>j}) \leq p_j$.*

When π is nonzero and $q_j \neq 0$, we have $\nu_{<j} \cap \nu_{>j} = \varnothing$. Write

$$\nu_j = \{\nu_{j,1}, \dots, \nu_{j,a_j}\} \quad \text{and} \quad \nu_{<j} \sqcup \nu_{>j} = \{\sigma_1, \dots, \sigma_t\}$$

such that $\nu_{j,1} > \dots > \nu_{j,a_j}$ and $\sigma_1 > \dots > \sigma_t$. Put $\nu_j \cap (\nu_{<j} \sqcup \nu_{>j}) = \{\sigma_{f+1}, \dots, \sigma_g\}$. Let $m = \min\{f, q_j - \#(\nu_j \cap \nu_{<j})\}$ and let i_0 be the maximal integer such that $1 \leq i_0 \leq g - f$ and $\sigma_{f+i_0} \geq \nu_{j,m+i_0}$. The first column of the tableau $\mathrm{Ann}(\pi)$ consists of

$$\sigma_1, \sigma_2, \dots, \sigma_f, \underbrace{\sigma_{f+1}, \sigma_{f+2}, \dots, \sigma_{f+i_0}}_{i_0},$$
$$\nu_{j,m+i_0+1}, \nu_{j,m+i_0+2}, \dots, \nu_{j,g}, \nu_{j,g+1}, \dots, \nu_{j,a_j}, \sigma_{f+p-m+1}, \dots, \sigma_t$$

and the second column consists of

$$\underbrace{\nu_{j,1}, \nu_{j,2} \dots, \nu_{j,m}}_{m}, \underbrace{\nu_{j,m+1}, \nu_{j,m+2}, \dots, \nu_{j,m+i_0}}_{i_0},$$
$$\underbrace{\sigma_{f+i_0+1}, \sigma_{f+i_0+2}, \dots, \sigma_g}_{g-i_0}, \sigma_{g+1}, \dots, \sigma_{\min\{t, f+p-m\}}$$

from top to bottom. Here, we understand that there is no box next to the box filled with ν_{j,a_j} if $t \leq f + p - m$ in the first column. In particular, if $m \neq 0$, then the $(1, 2)$-th entry of $\mathrm{Ann}(\pi)$ is $\nu_{j,1}$.

Proof. The proof is the same as Lemma 3.1. The case $q_j=0$ follows from Lemma 3.1. When $q_j\neq 0$, we compute the tableau $\mathrm{Ann}(\pi)$. If $\nu_{<j}\cap\nu_{>j}\neq\varnothing$, take $x\in\nu_{<j}\cap\nu_{>j}$. Since any element in ν has multiplicity at most two, any element in ν_j is greater than or less than x. If x is less than any element in ν, the segment ν_j is contained in ν_{j-1}, since π is in the mediocre range. Applying the algorithm to the tableau $S_{j-1}\sqcup S_j$, the representation $A_{\mathfrak{q}_{\underline{d}}}(\mu)$ is zero by $p_j\neq 0$. If x is greater than any element in ν_j, the representation is zero by the same method as $q_j\neq 0$. Hence, $\nu_{<j}\cap\nu_{>j}=\varnothing$.

If $\nu_{<j}$ or $\nu_{>j}$ is not multiplicity one, we have $\pi=0$. Hence, we may assume that $\nu_{<j}$ and $\nu_{>j}$ are multiplicity free. Let h be the integer such that $\nu_{<j}=\{\sigma_1,\dots,\sigma_h\}$ with $h=p-p_j=f+\#(\nu_j\cap\nu_{<j})$. Set $m'=\min\{h,q_j\}$. Before Trapa's algorithm, the first column of the tableau $\mathrm{Ann}(\pi)$ consists of

$$\sigma_1,\dots,\sigma_h,\nu_{j,m'+1},\dots,\nu_{j,a_j},\dots$$

and the second column consists of

$$\nu_{j,1},\nu_{j,2},\dots,\nu_{j,m'},\sigma_{h+1},\dots$$

from top to bottom. Let $\nu_{j-1}=\{\sigma_k,\dots,\sigma_h\}$ and S_i be the tableau consisting of entries with elements in ν_i defined in Trapa's algorithm. Consider $R=S_{j-1}\sqcup S_j$. By Trapa's algorithm, we have $R=R'$ if $\sigma_h\geq\nu_{j,m'}$ or $\nu_{j-1}\not\subset\nu_j$. In this case, R is equivalent to zero when $q_j<\#(\nu_{j-1}\cap\nu_j)$. If not, the first column of R' consists of

$$\nu_{j,m'-h+k},\dots,\nu_{j,m'},\nu_{j,m'+1},\dots$$

and the second column consists of

$$\nu_{j,1},\dots,\nu_{j,m'-h+k-1},\sigma_k,\dots,\sigma_h,\sigma_{h+1},\dots$$

from top to bottom. If $\sigma_h<\nu_{j,m'}$, then $\sigma_k<\nu_{j,m'-h+k}$. Trapa's algorithm for $R=S'_{j-2}\sqcup S'_{j-1}$ is similar and R is equivalent to zero if $q_j-\#(\nu_{j-1})<\#(\nu_{j-2}\cap\nu'_{j-1})$. Also, the algorithm for $S'_j\sqcup S_{j+1}$ is similar. Continuing this procedure to the end, one obtains the lemma. Note that $f-m=h-m'$ by elementary computations. □

The following statements, which are helpful to compute $A_{\mathfrak{q}}(\mu)$, show that when the associated segments are linked, we may replace the linked segments or, conversely, partition them into smaller segments.

Corollary 3.6. *Under the same notation as in Lemma 3.5, suppose $A_{\mathfrak{q}}(\mu)$ is nonzero. Let $\nu'_1\sqcup\cdots\sqcup\nu'_k=\nu_{<j}$ be a partition of $\nu_{<j}$ into segments with $\nu'_1>\cdots>\nu'_k$ and $\nu''_1\sqcup\cdots\sqcup\nu''_\ell=\nu_{>j}$ be a partition of $\nu_{>j}$ into segments with $\nu''_1>\cdots>\nu''_\ell$. Let $\mathfrak{q}'$ be the θ-stable parabolic subalgebra associated with*

$$\big\{(\#(\nu'_1),0),\dots,(\#(\nu'_k),0),(p-\#(\nu_{<j}),q-\#(\nu_{>j})),(0,\#(\nu''_1)),\dots,(0,\#(\nu''_\ell))\big\}.$$

Put $\pi=A(\mathfrak{q}',\nu'_1\sqcup\cdots\sqcup\nu'_k\sqcup\nu_j\sqcup\nu''_1\sqcup\cdots\sqcup\nu''_\ell)$. Then, if π is in the mediocre range, we have $A_{\mathfrak{q}}(\mu)\cong\pi$.

Proof. The statement follows from Lemma 3.5, since the tableau $A_{\mathfrak{q}}(\mu)$ does not depend on a partition of $\nu_{<j}$ and $\nu_{>j}$. □

Corollary 3.7. *Under the same notation as in Lemma 3.5, suppose $A_{\mathfrak{q}}(\mu)$ is nonzero. When $\nu_{j-1} \subset \nu_j$ (resp. $\nu_{j+1} \subset \nu_j$), let $\mathfrak{q}'$ be the θ-stable parabolic subalgebra corresponding to*

$$\big\{(p_1, q_1), \dots, (p_{j-2}, q_{j-2}), (p_j + p_{j-1}, q_j - p_{j-1}), \\ (0, p_{j-1}), (p_{j+1}, q_{j+1}), \dots, (p_r, q_r)\big\}$$

$$\big(\textit{resp. } \big\{(p_1, q_1), \dots, (p_{j-1}, q_{j-1}), (q_{j+1}, 0), \\ (p_j - q_{j+1}, q_j + q_{j-1}), (p_{j+2}, q_{j+2}), \dots, (p_r, q_r)\big\}\big)$$

and $\pi = A(\mathfrak{q}', \nu_1, \dots, \nu_j, \nu_{j-1}, \dots, \nu_r)$ (resp. $\pi = A(\mathfrak{q}', \nu_1, \dots, \nu_{j+1}, \nu_j, \dots, \nu_r)$). If π is in the mediocre range, one has $A_{\mathfrak{q}}(\mu) \cong \pi$.

Proof. The statement follows from Lemmas 3.1 and 3.5 by the explicit computation of tableaux. □

Remark 3.8. One can prove Corollaries 3.6 and 3.7 by the induction in stages of $A_{\mathfrak{q}}(\mu)$ (see [Huang 2025, §5.6; Trapa 2001, Lemma 3.9]). Note that Corollary 3.7 is a special case of [Huang 2025, Proposition 4.9; Trapa 2001, Lemma 9.3].

4. *A*-parameters and main theorem

We recall Mœglin and Renard's description of A-parameters in terms of cohomological inductions, and state the main theorem of this paper.

4.1. ***A-parameters.*** The A-parameters ψ are defined by a formal sum

$$\psi = \bigoplus_{i=1}^{r} \chi_{t_i, s_i} \otimes S_{a_i}, \tag{4-1}$$

where $\chi_{t,s}$ is the character of $\mathbb{C}^\times$ defined by $z \mapsto (z/\bar{z})^{t/2}(z\bar{z})^{s/2}$, S_m is the irreducible representation of $\mathrm{SL}_2(\mathbb{C})$ with dimension m, and the triplets (t, s, a) run over multisets on $(t, s, a) \in \mathbb{Z} \times \sqrt{-1}\,\mathbb{R} \times \mathbb{Z}_{>0}$. When $s = 0$, we write $\chi_t = \chi_{t,s}$. This definition of χ_t differs slightly from that of [Ichino 2022] but is the same as [Mœglin and Renard 2019]. For an A-parameter $\psi = \bigoplus_i \chi_{t_i, s_i} \otimes S_{a_i}$, we say that ψ is good if $\frac{1}{2}(t_i + a_i + N) \in \mathbb{Z}$ and $s_i = 0$ for any i. Associated with an A-parameter ψ, we obtain the component group $\mathfrak{S}_\psi$. It is isomorphic to a free $\mathbb{Z}/2\mathbb{Z}$-module

$$\mathfrak{S}_\psi = (\mathbb{Z}/2\mathbb{Z})e_1 \oplus \cdots \oplus (\mathbb{Z}/2\mathbb{Z})e_r.$$

Let $\Pi(\psi)$ be the A-packet of ψ, that is, the set of semisimple representations of G satisfying the standard and twisted endoscopic character relations (see [Atobe et al. 2024, §1.6; Kaletha et al. 2014, (1.6.1)]).

Remark 4.1. In the usual definition, an A-parameter is a homomorphism $\psi_{\mathbb{R}}$ from $W_{\mathbb{R}} \times \mathrm{SL}_2(\mathbb{C})$ to the L-group of the unitary group. Here, $W_{\mathbb{R}}$ is the real Weil group. By the base change $\psi_{\mathbb{R}}|_{\mathbb{C}^\times \times \mathrm{SL}_2(\mathbb{C})}$, the parameter $\psi_{\mathbb{R}}$ can be identified with the formal sum (4-1). For details, see [Gan et al. 2012, Theorem 8.1].

4.2. ***Mœglin–Renard's construction of A-packets.*** Take a good A-parameter

$$\psi = \bigoplus_{i=1}^{r} \chi_{t_i} \otimes S_{a_i}.$$

We define the infinitesimal character χ_ψ of ψ by the multiset

$$\bigsqcup_{i=1}^{r} \left\{\tfrac{1}{2}(t_i + a_i - 1), \tfrac{1}{2}(t_i + a_i - 3), \dots, \tfrac{1}{2}(t_i - a_i + 1)\right\} = \bigsqcup_{i=1}^{r} \left[\tfrac{1}{2}(t_i - a_i + 1), \tfrac{1}{2}(t_i + a_i - 1)\right]. \tag{4-2}$$

Then, all the representations in $\Pi(\psi)$ have the same infinitesimal character with the Harish-Chandra parameter (4-2). Following Théorème 1.1 in [Mœglin and Renard 2019], we describe the representations in $\Pi(\psi)$. Put

$$\mathcal{D}(\psi) = \left\{ (p_i, q_i)_{i=1,\dots,r} \in (\mathbb{Z}_{\geq 0})^2 \mid p_i + q_i = a_i \text{ for any } i \text{ and } \sum_{i=1}^{r} p_i = p, \sum_{i=1}^{r} q_i = q \right\}.$$

For $\underline{d} \in \mathcal{D}(\psi)$, set

$$\mu_{\underline{d}} = (\underbrace{\mu_1, \dots, \mu_1}_{p_1}, \dots, \underbrace{\mu_r, \dots, \mu_r}_{p_r}, \underbrace{\mu_1, \dots, \mu_1}_{q_1}, \dots, \underbrace{\mu_r, \dots, \mu_r}_{q_r}),$$

where $\mu_i = \frac{1}{2}(t_i + a_i - N) + a_{<i}$ and $a_{<i} = \sum_{j<i} a_j$. Note that there is a typo in [Mœglin and Renard 2019, (4.2)]: $\frac{1}{2}(t_i + a_i - N) - a_{<i}$ should be $\frac{1}{2}(t_i + a_i - N) + a_{<i}$. For $x_{\underline{d}}$ as in (3-1), we define the cohomological induction by

$$\mathscr{A}_{\underline{d}}(\psi) = A_{\mathfrak{q}(x_{\underline{d}})}(\mu_{\underline{d}}) \tag{4-3}$$

and a character $\varepsilon_{\underline{d}}$ on $\mathfrak{S}_\psi$ by

$$\varepsilon_{\underline{d}}(e_i) = (-1)^{p_i a_{<i} + q_i (a_{<i} + 1) + a_i (a_i - 1)/2}$$

for any $e_i \in \mathfrak{S}_\psi$. The following is proved in [Mœglin and Renard 2019, Théorème 1.1]. We choose the same Whittaker datum $\mathfrak{w}$ as [Mœglin and Renard 2019] (see [Atobe 2020, Appendix A]).

Theorem 4.2. *Let ψ be a good A-parameter with $\psi = \bigoplus_{i=1}^{r} \chi_{t_i} \otimes S_{a_i}$. Suppose that $t_1 \geq \cdots \geq t_r$ and $a_i \geq a_{i+1}$ if $t_i = t_{i+1}$. We then have*

$$\Pi(\psi) = \{\mathscr{A}_{\underline{d}}(\psi) \mid \underline{d} \in \mathcal{D}(\psi)\}.$$

The character of $\mathfrak{S}_\psi$ associated with $\mathscr{A}_{\underline{d}}(\psi)$ is equal to $\varepsilon_{\underline{d}}$. Moreover, the multiplicity one holds in $\Pi(\psi)$.

Remark 4.3. The correspondence $\mathscr{A}_{\underline{d}}(\psi)$ to $\varepsilon_{\underline{d}}$ depends on the choice of Whittaker datum $\mathfrak{w}$. We may explicitly calculate the dependence. See [Kaletha et al. 2014, Theorem 1.6.1] for details.

Remark 4.4. The statement does not imply that the cohomological induction $\mathscr{A}_{\underline{d}}(\psi)$ is nonzero. Trapa [2001] gives an algorithm to determine whether cohomological inductions of G are zero. Recently, Huang [2025] and Chengyu [2025] independently considered the nonvanishing of cohomological inductions of real unitary groups with Chengyu treating the "nice" case and Huang the general case.

Remark 4.5. For general ψ, representations in the A-packet $\Pi(\psi)$ consist of the parabolic induction from the representations in the packet $\Pi(\psi_0)$ for certain good A-parameter $\psi_0 \subset \psi$. Here, ψ_0 is a good A-parameter of a unitary group that is a subgroup of G.

4.3. ***Main theorem.*** In this subsection, we state the main theorem following the notation as in the introduction. The lemma below plays a crucial role in stating and proving the main theorem.

Lemma 4.6. *Let $\psi = \bigoplus_{i=1}^{r} \chi_{t_i,s} \otimes S_{a_i}$ be an A-parameter. If $\Pi(\psi)$ contains an irreducible lowest-weight representation π, the parameter ψ is good and $\chi_\psi = \chi_\pi$. Moreover, if $\mathscr{A}_{\underline{d}}(\psi) \in \Pi(\psi)$ is nonzero and lowest weight, there exists j such that $q_i = 0$ for any $i < j$ and $p_\ell = 0$ for any $\ell > j$, i.e., $\underline{d} = \underline{d}_0$ and $\mathfrak{q}_{\underline{d}}$ is holomorphic.*

Proof. The goodness of ψ follows from the construction of general ψ as in [Mœglin and Renard 2019, Proposition 5.2]. The condition $\chi_\psi = \chi_\pi$ is obvious since representations in $\Pi(\psi)$ have the same infinitesimal character χ_ψ. For the second statement, consider the signed tableau $\mathrm{Ann}(\mathscr{A}_{\underline{d}}(\psi))$. If there exist different integers k and ℓ with $p_k q_k p_\ell q_\ell \neq 0$, the signed tableau $\mathrm{Ann}(\mathscr{A}_{\underline{d}}(\psi))$ satisfies either

- there exists a row with three or more boxes, or
- there exists a row with two boxes arranged in the order of $-$ and $+$.

Then $\mathscr{A}_{\underline{d}}(\psi)$ is not of a unitary lowest-weight representation by Corollary 3.3. □

The following is the main theorem of this paper. Note that the nonvanishing condition $\mathscr{A}_{\underline{d}_0}(\psi)$ is determined in Lemma 3.5 (see also Corollary 4.8).

Theorem 4.7. *Let $\lambda = (\lambda_1, \dots, \lambda_N)$ be a Δ_c^+-dominant integral weight and π_λ be the irreducible lowest-weight representation of lowest K-type λ. Let ψ be a good A-parameter with $\chi_\psi = \chi_{\pi_\lambda}$ such that $\mathscr{A}_{\underline{d}_0}(\psi)$ is nonzero.*

(1) *If $N - p' \le \lambda_p - \lambda_{p+1} < N - q'$, the packet $\Pi(\psi)$ contains π_λ if and only if $\left[\lambda_p - \frac{1}{2}(N-1), \lambda_{p+1} + \frac{1}{2}(N-1)\right] \subset \nu_j \subset P'$.*

(2) *If $N - q' \le \lambda_p - \lambda_{p+1} < N - p'$, the packet $\Pi(\psi)$ contains π_λ if and only if either*

- $\nu_{\leq j} = P$, *or*
- $\left[\lambda_p - \frac{1}{2}(N-1), \lambda_{p+1} + \frac{1}{2}(N-1)\right] \subset \nu_j \subset Q'$.

(3) *If* $N-p', N-q' \leq \lambda_p - \lambda_{p+1}$, *the packet* $\Pi(\psi)$ *contains* π_λ *if and only if either*

- $P \subset \nu_{\leq j} \subset P \sqcup I$, *or*
- $I \subset \nu_j \subset Q'$.

(4) *If* $\lambda_p - \lambda_{p+1} < N - p', N - q'$, *the packet* $\Pi(\psi)$ *contains* π_λ *if and only if* $\left[\lambda_p - \frac{1}{2}(N-1), \lambda_{p+1} + \frac{1}{2}(N-1)\right] = \nu_j$.

Our proof, provided in Section 5, is based on the explicit computation of K-types and the associated tableaux of $\mathscr{A}_{\underline{d}}(\psi)$. More precisely, the if part follows from Lemma 3.2 and Corollaries 3.6–3.7. For the only if part, we divide the cases into $q_j = 0$ or $q_j \neq 0$. When $q_j = 0$, the statement follows from Lemma 3.1. When $q_j \neq 0$, we calculate the associated tableau $\mathrm{Ann}(\mathscr{A}_{\underline{d}}(\psi))$ in Lemmas 5.2, 5.3, and 5.4. Then, the theorem follows.

As a consequence of Theorem 4.7, we can determine the lowest K-type of the lowest-weight representation in $\Pi(\psi)$.

Corollary 4.8. *Let* $\psi = \bigoplus_{i=1}^{r} \chi_{t_i} \otimes S_{a_i}$ *be a good A-parameter. The packet* $\Pi(\psi)$ *contains a nonzero unitary lowest-weight representation if and only if both*

- $\nu_{<j}$ *and* $\nu_{>j}$ *are multiplicity free*, *and*
- $\#(\nu_j \cap \nu_{>j}) \leq p_j$ *and* $\#(\nu_j \cap \nu_{<j}) \leq q_j$.

When $\Pi(\psi)$ *contains a nonzero unitary lowest-weight representation* π *in* $\Pi(\psi)$, *the lowest* K*-type* λ *of* π *is given as follows*:

(1) *When* $q_j = 0$, *the lowest* K*-type* λ *of* π *satisfies* $P(\lambda) = \nu_{\leq j}$ *and* $Q(\lambda) = \nu_{>j}$.

(2) *When* $p_j = \#(\nu_j \cap \nu_{>j})$ *and* $q_j \neq 0$, *the lowest* K*-type* λ *of* π *satisfies* $P(\lambda) = \nu_{<j} \sqcup (\nu_j \cap \nu_{>j})$ *and* $Q(\lambda) = \nu_{\geq j} \setminus (\nu_j \cap \nu_{>j})$.

(3) *When* $q_j = \#(\nu_j \cap \nu_{<j}) \neq 0$, *the lowest* K*-type* λ *of* π *satisfies*

$$P(\lambda) = \nu_{\leq j} \setminus (\nu_{<j} \cap \nu_j) = \{\sigma_1, \ldots, \sigma_p\} \quad \text{and} \quad Q(\lambda) = (\nu_j \cap \nu_{<j}) \sqcup \nu_{>j}.$$

(4) *When* $p_j \neq \#(\nu_j \cap \nu_{>j})$ *and* $q_j \neq \#(\nu_j \cap \nu_{<j})$, *set* $\nu_{<j} \sqcup \nu_{>j} = \{\sigma_1, \ldots, \sigma_{N-\#(\nu_j)}\}$. *Let* i_0 *be the minimal integer such that*

$$1 \leq i_0 \leq \#(\nu_j) \quad \text{and} \quad \#(\nu_j) - i_0 + 1 + \#\{x \in \nu_{<j} \sqcup \nu_{>j} \mid x > \nu_{j,i_0}\} = p.$$

Then, the lowest K*-type* $\lambda = (\lambda_1, \ldots, \lambda_N)$ *of* π *is given by*

$$\lambda_i = \begin{cases} \sigma_i - \frac{1}{2}(p-q+1) + i & \text{if } i < p - \#(\nu_j) + i_0, \\ \nu_{j,1} + \frac{1}{2}(N+1) - \#(\nu_j) & \text{if } p - \#(\nu_j) + i_0 \leq i \leq p, \\ \nu_{j,1} - \frac{1}{2}(N-1) & \text{if } p+1 \leq i \leq p + i_0 - 1, \\ \sigma_{i-\#(\nu_j)} - \frac{1}{2}(N+1) - p + i & \text{if } p + i_0 \leq i. \end{cases}$$

Proof. To show the nonvanishing condition, it suffices to consider the case $\underline{d} = \underline{d}_0$ by Lemma 4.6. The nonvanishing condition for $\mathscr{A}_{\underline{d}_0}(\psi)$ is given in Lemma 3.5. Suppose that $\mathscr{A}_{\underline{d}_0}(\psi)$ is nonzero. Then, the statement (1) follows from Lemma 3.1. For (2) and (3), the statements follow from Lemma 3.2 and Corollaries 3.6–3.7. For (4), we have $\lambda_p - \lambda_{p+1} < N - p'(\lambda), N - q'(\lambda)$ by Theorem 4.7. Now, i_0 as in (4) exists. We then have $\nu_j = \left[\lambda_p - \frac{1}{2}(N-1), \lambda_{p+1} + \frac{1}{2}(N-1)\right]$ and $P(\lambda) = \{\nu_{j,i_0}, \nu_{j,i_0+1} \dots, \nu_{j,\#(\nu_j)}\} \sqcup \{x \in \nu \setminus \nu_j \mid x > \nu_{j,i_0}\}$. The statement follows from Lemma 3.2(1) and Corollary 3.7. □

5. Proof of main theorem

We prove the main theorem in Sections 5.2, 5.3, 5.4, and 5.5.

5.1. ***Cohomological induction for holomorphic θ-stable parabolic subalgebras.*** To show the main theorem, we will need Lemma 5.1 below.

By Lemma 4.6, it suffices to consider $\underline{d} = \underline{d}_0$. We already described the nonvanishing conditions of such cohomological inductions in Lemma 3.5. In the following lemma, we will investigate the necessary condition that $\mathscr{A}_{\underline{d}_0}(\psi)$ is isomorphic to π_λ with a given λ. A key point here is that the cohomological induction $\mathscr{A}_{\underline{d}}(\psi)$ is in the weakly fair range.

Lemma 5.1. *Take an irreducible unitary lowest-weight representation π_λ with lowest K-type λ. Let ψ be an A-parameter with $\pi_\lambda \in \Pi(\psi)$. Let ν_i be the segments associated with ψ and $\underline{d}_0$.*

(1) *As multisets, $\nu_{<j}$ and $\nu_{>j}$ are multiplicity free.*

(2) *If $\nu_j \subset \nu_{>j}$, then $q_j = 0$.*

(3) $\nu_j \not\subset \nu_{<j}$.

(4) $I \subset \nu_{<j} \sqcup \nu_{>j}$.

(5) *If $\nu_{<j} \cap \nu_{>j} \neq \varnothing$, then $\nu_j \subset \nu_{>j}$ and $q_j = 0$.*

(6) *If $\nu_{<j} \cap \nu_{>j} \cap I \neq \varnothing$ and $I \cap \nu_j \neq \varnothing$, then $\nu_j \subset I \subset \nu_{>j}$ and $q_j = 0$.*

(7) *If $\nu_{<j} \cap \nu_{>j} \cap I = \varnothing$, then $I \subset \nu_j$.*

(8) *If $I \neq \varnothing$, then $I \cap \nu_j \neq \varnothing$.*

Proof. The statements (1), (2), (3), and (5) follow immediately from Lemma 3.5. Note that $p_j \neq 0$ by definition of j. For (4), consider the multiplicities of each element in ν. The multiplicities of elements in I are two in ν. Thus, $\nu_{<j} \sqcup \nu_{>j}$ contains I, since ν_j is a set.

Set $I = [x, y]$ and $\nu_j = [\alpha, \beta]$. By $\pi_\lambda \in \Pi(\psi)$, one has $P \sqcup Q = \nu$. For (6), assume $\nu_{<j} \cap \nu_{>j} \cap I \neq \varnothing$ and $\nu_j \cap I \neq \varnothing$. Since any element in I has multiplicity two in ν, one has $\nu_{<j} \cap \nu_{>j} \cap I = I \setminus (I \cap \nu_j)$ by (1). Moreover, $I \cap \nu_j$ is a segment since I

and ν_i are segments. Hence, $\nu_{<j} \cap \nu_{>j} \cap I$ contains x or y. When $y \in \nu_{<j} \cap \nu_{>j}$, we denote by z the minimal member in $\nu_{<j} \cap \nu_{>j} \cap I$. The maximal member in $\nu_j \cap I$ is $z-1$. By the weakly fair property, the set $\nu_{<j}$ does not contain $z-1$. We thus have $z-1 \in \nu_{>j}$. Then, the set $\nu_{>j}$ contains I and in particular, $\nu_j \subset I \subset \nu_{>j}$. Here, we use the fact that the real numbers $x-1$ and $y+1$ have multiplicity at most one in ν. When $x \in \nu_{<j} \cap \nu_{>j}$, one has $\nu_j \subset I \subset \nu_{<j}$ by the same discussion. This case does not happen by (3). Hence, we have $\nu_j \subset I \subset \nu_{>j}$ and then $q_j = 0$.

The statement (7) follows immediately from the fact that the multiplicities of elements in I are two.

For (8), suppose $I \neq \varnothing$ and $I \cap \nu_j = \varnothing$. By the proof of Lemma 3.5, we have $\nu_j \subset \nu_{>j}$ and $\alpha \leq \beta < x$. Then, $q_j = 0$, and there exists an element t in $P' \sqcup Q'$ with multiplicity two such that $t < x$. The existence of t implies $\lambda_p - \lambda_{p+1} < N - q'$. By $q_j = 0$ and Lemma 3.2(3), we may assume $N - p' \leq \lambda_p - \lambda_{p+1}$. Then, the set I is equal to Q'. In this case, $x-1 \notin Q$, but $x-1 \in P \cap \nu_{>j}$ by definition. Moreover, the segment $\nu_{>j}$ contains the set $Q \setminus Q'$ and $I = Q'$, and then $Q \subset \nu_{>j}$. Hence,

$$\#\nu_{>j} \geq \#\{x-1\} + \#Q \geq q+1.$$

This contradicts the definition of j. Hence $I \cap \nu_j \neq \varnothing$. □

In the following sections, we complete the proof of Theorem 4.7.

5.2. ***Proof of main theorem: the case $N - p' \leq \lambda_p - \lambda_{p+1} < N - q'$.*** We first show the only if part. By the assumption, one has $I = Q'$. Put $\nu_j = [\alpha, \beta]$.

When $q_j = 0$, by Lemma 3.1, we have $\nu_{\leq j} = P$ and $I = Q' \subset \nu_{>j} = Q$ if $\mathscr{A}_{\underline{d}_0}(\psi) \cong \pi_\lambda$. In this case, one has $\alpha = \lambda_p - \frac{1}{2}(N-1)$. Note that by the weakly fair property, we have $\beta \geq \lambda_{p+1} + \frac{1}{2}(N-1)$. Hence, the segment ν_j contains $\left[\lambda_p - \frac{1}{2}(N-1), \lambda_{p+1} + \frac{1}{2}(N-1)\right]$.

When $q_j \neq 0$, the segment ν_j is not contained in $\nu_{>j}$ by Lemma 5.1 (2). Also, $\nu_{<j} \cap \nu_{>j} = \varnothing$ and ν_j contains I by Lemma 5.1. Set

$$\nu \setminus \nu_j = \{\sigma_1, \dots, \sigma_{N-\#(\nu_j)}\}, \quad \nu_j \cap (\nu_{<j} \sqcup \nu_{>j}) = \{\sigma_{f+1}, \dots, \sigma_g\}$$

with $\sigma_1 > \cdots > \sigma_{N-\#(\nu_j)}$. Define $\underline{d} = \{(p_i, q_i)\}$ and ν'_i by

$$(p_i, q_i) = \begin{cases} (1,0) & \text{if } i < p - p_j, \\ (p_j, q_j) & \text{if } i = p_j, \\ (0,1) & \text{if } i > p_j, \end{cases} \qquad \nu'_i = \begin{cases} \{\sigma_i\} & \text{if } i < p - p_j, \\ \nu_j & \text{if } i = p - p_j, \\ \{\sigma_{i-1}\} & \text{if } i > p - p_j, \end{cases}$$

and

$$\pi(\psi) = A(\mathfrak{q}_{\underline{d}}, \nu'_1, \dots, \nu'_{N-\#(\nu_j)+1}). \tag{5-1}$$

By Corollary 3.6, $\pi(\psi)$ is in the mediocre range and isomorphic to $\mathscr{A}_{\underline{d}_0}(\psi)$.

Lemma 5.2. *If $\pi(\psi)\cong\pi_\lambda$, then* $\left[\lambda_p-\frac{1}{2}(N-1),\lambda_{p+1}+\frac{1}{2}(N-1)\right]\subset\nu_j\subset P'$.

Proof. Since the segment ν_j contains I, it suffices to show $\alpha=\lambda_p-\frac{1}{2}(N-1)$. Suppose $\alpha\geq\lambda_p-\frac{1}{2}(N-1)$. By Lemma 3.5, one has $\#(\nu_{<j}\cap\nu_j)\leq q_j=q-\#\nu_{>j}$. Since the set $(\nu_{<j}\cap\nu_j)\sqcup\nu_{>j}$ contains Q in this case, one has $\#\nu_{>j}+\#(\nu_{<j}\cap\nu_j)\geq q$. Hence, we have $\nu_{>j}\sqcup(\nu_{<j}\cap\nu_j)=Q$ and in particular, $\nu_{>j}\subset Q$. By $I\subset\nu_j$, the multiset ν_j is not multiplicity free or $\nu_{>j}\not\subset Q$ if $\alpha>\lambda_p-\frac{1}{2}(N-1)$. This shows $\alpha\leq\lambda_p-\frac{1}{2}(N-1)$.

It remains to show $\alpha\geq\lambda_p-\frac{1}{2}(N-1)$. Suppose $\alpha<\lambda_p-\frac{1}{2}(N-1)$. To show $\pi(\psi)\not\cong\pi_\lambda$ under this assumption, recall the tableau $\mathrm{Ann}(\pi_\lambda)$. By Lemma 3.2(2), the first column of the tableau $\mathrm{Ann}(\pi_\lambda)$ consists of entries

$$\lambda_1+\tfrac{1}{2}(p-q-1),\lambda_2+\tfrac{1}{2}(p-q-3),\dots,\lambda_p-\tfrac{1}{2}(N-1),\dots$$

and the second column consists of entries

$$\lambda_{p+1}+\tfrac{1}{2}(N-1),\lambda_{p+2}+\tfrac{1}{2}(N-3),\dots,\lambda_{p+\min\{p,q\}}+\tfrac{1}{2}(N+1)-\min\{p,q\}$$

from top to bottom. This is invariant under Trapa's algorithm. We describe the tableau $\mathrm{Ann}(\pi(\psi))$ as follows. By $\#(\nu_{<j})+p_j+q_j-\#(\nu_{<j}\cap\nu_j)>p$ and $\#(\nu_{<j})+p_j=p$, we have $q_j-\#(\nu_{<j}\cap\nu_j)>0$. When there is no $x\in\nu$ with $\beta<x$, the first column of $\pi(\psi)$ consists of

$$\beta=\lambda_1+\tfrac{1}{2}(p-q-1),\beta-1,\dots,\alpha,\dots$$

from top to bottom. In particular, the entry next to $\lambda_p-\frac{1}{2}(N-1)$ is $\lambda_p-\frac{1}{2}(N+1)$. However, the entry next to $\lambda_p-\frac{1}{2}(N-1)$ in the first column of $\mathrm{Ann}(\pi_\lambda)$ does not equal $\lambda_p-\frac{1}{2}(N+1)$. Indeed, if $q\leq p$, there is no such box. If $q>p$, the entry is $\lambda_{2p+1}+\frac{1}{2}(N+1)-(p+1)=\lambda_{2p+1}-\frac{1}{2}(p-q+1)$. We then have

$$\begin{aligned}\lambda_p-\tfrac{1}{2}(N+1)-\big(\lambda_{2p+1}-\tfrac{1}{2}(p-q+1)\big)&=\lambda_p-\lambda_{2p+1}-q\\&=\lambda_p-\lambda_{p+1}-(N-p)+(\lambda_{p+1}-\lambda_{2p+1})\\&\geq\lambda_{p+1}-\lambda_{2p+1}.\end{aligned}$$

For the last line, we use $P=P'$ and $N-p'\leq\lambda_p-\lambda_{p+1}$. By $N-p=N-p'<N-q'$, one has $q'<p$ and then $\lambda_{p+1}-\lambda_{2p+1}>0$. Hence, the tableaux $\mathrm{Ann}(\pi_\lambda)$ and $\mathrm{Ann}(\pi(\psi))$ are different and in particular, the representations are different. We may assume that there exists $x\in\nu$ such that $x>\beta$. Put $f=\#\{x\in\nu\mid x>\beta\}$. Let $m=\min\{f,q_j-\#(\nu_j\cap\nu_{<j})\}$ and i_0 be the maximal integer such that $1\leq i_0\leq g-f$ and $\sigma_{f+i_0}\geq\nu_{j,m+i_0}$. Here, $\nu_j=\{\nu_{j,1},\dots,\nu_{j,a_j}\}$ with $\nu_{j,1}>\cdots>\nu_{j,a_j}$. By assumption, m is positive. By Lemma 3.5, the $(1,2)$-th entry in $\mathrm{Ann}(\pi(\psi))$ is β. Hence, we have $\beta=\lambda_{p+1}+\frac{1}{2}(N-1)$. This shows $i_0\geq\#I$. The second column of $\mathrm{Ann}(\pi(\psi))$ consists of

$$\nu_{j,1},\nu_{j,2},\dots,\nu_{j,m+i_0},\dots$$

from top to bottom. In particular, the entry next to $\nu_{j,\#(I)} = \lambda_{p+1} + \frac{1}{2}(N+1) - q'$ is $\nu_{j,\#(I)+1} = \nu_{j,\#(I)} - 1$. Note that $\lambda_{p+1} + \frac{1}{2}(N-1) - q'$ is in ν by $\lambda_p - \lambda_{p+1} < N - q'$. However, in the second column of $\mathrm{Ann}(\pi_\lambda)$, the entry next to $\lambda_{p+1} + \frac{1}{2}(N+1) - q'$ is $\lambda_{p+q'+1} + \frac{1}{2}(N-1) - q' < \nu_{j,\#(I)+1}$, if it exists. Hence, the representation $\lambda_{p+1} + \frac{1}{2}(N+1) - q'$ is not isomorphic to π_λ since the associated ν-antitableau tableaux are different. □

We show the converse. Suppose that $\mathscr{A}_{\underline{d}_0}(\psi)$ satisfies

$$\left[\lambda_p - \tfrac{1}{2}(N-1), \lambda_{p+1} + \tfrac{1}{2}(N-1)\right] \subset \nu_j \subset P'.$$

The nonvanishing of $\mathscr{A}_{\underline{d}_0}(\psi)$ follows from Lemma 3.5. Since the multiset $\nu \setminus \nu_j = \nu_{<j} \sqcup \nu_{>j}$ is multiplicity free, the representation $\mathscr{A}_{\underline{d}_0}(\psi)$ is isomorphic to π_λ by Lemma 3.2 and Corollaries 3.6–3.7.

5.3. ***Proof of main theorem: the case $N - q' \le \lambda_p - \lambda_{p+1} < N - p'$.*** We first show the only if part. Let $P'' = (P \cap Q') \setminus I$ and $p'' = \#(P'')$. One has $I = P'$ and $I \cap \nu_j \neq \varnothing$ by Lemma 5.1(8). When $q_j = 0$, by Lemma 3.1, the representation $\mathscr{A}_{\underline{d}_0}(\psi)$ is isomorphic to π_λ if and only if $\nu_{\le j} = P$.

We consider the case where $q_j \neq 0$. Then, the multiset $\nu_{<j} \sqcup \nu_{>j}$ is multiplicity free and ν_j is not contained in $\nu_{>j}$. Let $\pi(\psi)$ be the cohomological induction defined in the same way in (5-1). The statement follows from this lemma:

Lemma 5.3. *If $\pi(\psi) \cong \pi_\lambda$, we then have*

$$\left[\lambda_p - \tfrac{1}{2}(N-1), \lambda_{p+1} + \tfrac{1}{2}(N-1)\right] \subset \nu_j \subset Q'.$$

Proof. Put $\nu_j = [\alpha, \beta]$. Since ν_j contains I, one has $\alpha \le \lambda_p - \frac{1}{2}(N-1)$. It remains to show $\beta = \lambda_{p+1} + \frac{1}{2}(N-1)$. Suppose that $\beta < \lambda_{p+1} + \frac{1}{2}(N-1)$. We then have $p_j \ge \#(\nu_j \cap \nu_{>j})$ and the multiset $\nu_{\le j}$ contains a set P properly, since $\nu_{<j}$ contains $\{x \in \nu \mid x > \beta\}$. We show that $\#(\nu_j \cap \nu_{>j}) - p_j$ is positive. By $\#(\nu_j \cap \nu_{>j}) = \#I + \#P'' - \#(\nu_{<j} \cap \nu_j)$, one has

$$\begin{aligned}\#(\nu_j \cap \nu_{>j}) - p_j &= p' + p'' - \#(\nu_j \cap \nu_{<j}) - p_j \\ &= p' + p'' - \#(\nu_j \cap \nu_{<j}) - (p - \#(\nu_{<j})) \\ &= \#(\nu_{<j}) - \#(\nu_j \cap \nu_{<j}) - (p - p' - p'') > 0.\end{aligned}$$

The last inequality follows from $\beta < \lambda_{p+1} + \frac{1}{2}(N-1)$. Hence $\pi(\psi)$ is zero. This is a contradiction. Therefore we have $\beta \ge \lambda_{p+1} + \frac{1}{2}(N-1)$.

It remains to show $\beta \le \lambda_{p+1} + \frac{1}{2}(N-1)$. Put $f = \#\left\{x \in \nu \mid x > \lambda_{p+1} + \frac{1}{2}(N-1)\right\}$. Suppose $\beta > \lambda_{p+1} + \frac{1}{2}(N-1)$. By assumption, we have $q_j > \#(\nu_j \cap \nu_{<j})$ and $f \neq 0$. We recall the tableau $\mathrm{Ann}(\pi_\lambda)$. By Lemmas 3.2 and 3.5, the second column of the

tableau $\mathrm{Ann}(\pi_\lambda)$ consists of

$$\lambda_{p+1}+\tfrac{1}{2}(N-1), \lambda_{p+2}+\tfrac{1}{2}(N-3), \dots, \lambda_{p+\min\{p,q\}+1-i_0}+\tfrac{1}{2}(N-1)-\min\{p,q\}+i_0, \underbrace{\lambda_{p+1-i_0}-\tfrac{1}{2}(N+1)+i_0, \dots, \lambda_p-\tfrac{1}{2}(N-1)}_{i_0}, \dots .$$

Here, i_0 is the maximal positive integer such that

$$\lambda_{p+\min\{p,q\}+1-i_0}+\tfrac{1}{2}(N-1)-\min\{p,q\}+i_0 \geq \lambda_{p+1-i_0}-\tfrac{1}{2}(N+1)+i_0$$

if it exists. If no such i_0 exists, i_0 is defined as 0. In particular, the $(1,2)$-th entry is $\lambda_{p+1}+\frac{1}{2}(N-1)$. In this case, the number of boxes in the second column from the top to the box filled with $\lambda_p-\frac{1}{2}(N-1)$ is greater than $p'+p''$. If there exists $x \in \nu$ with $x > \beta$, the $(1,2)$-th entry of $\mathrm{Ann}(\pi(\psi))$ is β by $q_j > \#(\nu_j \cap \nu_{>j})$. Thus, the tableaux $\mathrm{Ann}(\pi_\lambda)$ and $\mathrm{Ann}(\pi(\psi))$ are different by the assumption $\beta > \lambda_{p+1}+\frac{1}{2}(N-1)$. In other words, the representations π_λ and $\pi(\psi)$ are different. The remaining case is that there is no $x \in \nu$ with $x > \beta$. Then, the second column of $\mathrm{Ann}(\mathscr{A}_{\underline{d}_0}(\psi))$ consists of entries

$$\lambda_{f+1}+\tfrac{1}{2}(p-q-1)-f, \lambda_{f+2}+\tfrac{1}{2}(N-3)-f, \dots, \lambda_p-\tfrac{1}{2}(N-1), \dots$$

from top to bottom. The number of boxes from the top to the box filled with $\lambda_p-\frac{1}{2}(N-1)$ is $p'+p''$, different from that of $\mathrm{Ann}(\pi_\lambda)$. Therefore, the representation $\mathscr{A}_{\underline{d}_0}(\psi)$ is not isomorphic to π_λ. We conclude that $\beta = \lambda_{p+1}+\frac{1}{2}(N-1)$. □

It remains to show the converse. Suppose that $\mathscr{A}_{\underline{d}_0}(\psi)$ satisfies $\nu_{\leq j} = P$. Then, $\mathscr{A}_{\underline{d}_0}(\psi) \cong \pi_\lambda$ by Lemma 3.1. Suppose next that $\mathscr{A}_{\underline{d}_0}(\psi)$ satisfies

$$\left[\lambda_p-\tfrac{1}{2}(N-1), \lambda_{p+1}+\tfrac{1}{2}(N-1)\right] \subset \nu_j \subset Q'.$$

In this case, by the explicit computation of p_j and q_j, the representation $\mathscr{A}_{\underline{d}_0}(\psi)$ is nonzero and isomorphic to π_λ by Lemma 3.2 and Corollaries 3.6–3.7.

5.4. ***Proof of main theorem: the case $N-p', N-q' \leq \lambda_p-\lambda_{p+1}$.*** We first show the only if part. When $q_j = 0$, by Lemma 3.1, we have $\nu_{\leq j} = P$ and $\nu_{>j} = Q$. In particular, $P \subset \nu_{\leq j} \subset P \sqcup I$. When $q_j \neq 0$, the segment ν_j contains I. When $\nu_j \cap (Q' \setminus I) = \varnothing$, we have $\nu_{\leq j} \subset P \sqcup I$ and $P \subset \nu_{\leq j}$ since $\nu_{>j}$ is multiplicity free by Lemma 3.5. We may assume that $\nu_j \cap (Q' \setminus I) \neq \varnothing$ and $\lambda_1+\frac{1}{2}(p-q-1) > \lambda_{p+1}+\frac{1}{2}(N-1)$. In fact, if $\lambda_1+\frac{1}{2}(p-q-1) = \lambda_{p+1}+\frac{1}{2}(N-1)$, the segment ν_j is automatically contained in Q'. Consider the tableaux for $\mathrm{Ann}(\mathscr{A}_{\underline{d}_0}(\psi))$ and $\mathrm{Ann}(\pi_\lambda)$. We show that ν_j is contained in Q' if $\mathscr{A}_{\underline{d}_0}(\psi) \cong \pi_\lambda$. Note that we have $q_j - \#(\nu_j \cap \nu_{<j}) > 0$. Indeed, by assumption, the multiset $\nu_{\leq j}$ contains P properly and then $P \sqcup (\nu_j \cap \nu_{<j}) \subsetneq \nu_{\leq j}$. We then have $p + \#(\nu_j \cap \nu_{<j}) < \#\nu_{<j} + p_j + q_j$

and, in particular, $0 < q_j - \#(\nu_j \cap \nu_{<j})$ by $p_j + \#(\nu_{<j}) = p$. By Lemma 3.2, the first column of $\mathrm{Ann}(\pi_\lambda)$ is

$$\lambda_1 + \tfrac{1}{2}(p-q-1), \lambda_1 + \tfrac{1}{2}(p-q-3), \ldots, \lambda_p - \tfrac{1}{2}(N-1), \ldots$$

and the second column is

$$\lambda_{p+1} + \tfrac{1}{2}(N-1), \lambda_{p+2} + \tfrac{1}{2}(N-3), \ldots, \lambda_{p+\min\{p,q\}} + \tfrac{1}{2}(N+1) - \min\{p,q\}.$$

Put $\nu_j = [\alpha, \beta]$. Suppose $\beta = \lambda_1 + \frac{1}{2}(p-q-1)$. Then

$$P = P' \quad \text{and} \quad \beta > \lambda_{p+1} + \tfrac{1}{2}(N-1).$$

The first column of $\mathrm{Ann}(\mathscr{A}_{\underline{d}_0}(\psi))$ consists of

$$\beta, \beta - 1, \ldots, \alpha, \ldots$$

and the second column consists of

$$\lambda_{p+1} + \tfrac{1}{2}(N-1), \lambda_{p+2} + \tfrac{1}{2}(N-3), \ldots, \lambda_{p+\ell} + \tfrac{1}{2}(N+1) - \ell, \alpha - 1, \ldots$$

from top to bottom. Here, ℓ is the unique positive integer such that

$$\lambda_{p+\ell} + \tfrac{1}{2}(N+1) - \ell = \lambda_p - \tfrac{1}{2}(N-1).$$

Note that in the second column of $\mathrm{Ann}(\mathscr{A}_{\underline{d}_0}(\psi))$, the box next to the box filled with $\lambda_{p+\ell} + \frac{1}{2}(N+1) - \ell$ exists if and only if $\beta > \lambda_{p+1} + \frac{1}{2}(N-1)$ and $\nu_j \sqcup I \neq \nu$. Contrary to this, in the second column of $\mathrm{Ann}(\pi_\lambda)$, the box next to the box filled with $\lambda_{p+\ell} + \frac{1}{2}(N+1) - \ell$ exists if and only if $\lambda_1 + \frac{1}{2}(p-q-1) > \lambda_{p+1} + \frac{1}{2}(N-1)$ and $I \neq Q$. Hence, under our assumption, there exists a box next to the box filled with $\lambda_{p+\ell} + \frac{1}{2}(N+1) - \ell$ in the second column of $\mathrm{Ann}(\pi_\lambda)$ and it is equal to $\lambda_{p+\ell} + \frac{1}{2}(N-1) - \ell$. We may additionally assume $\nu \neq \nu_j \sqcup I$. The entry next to $\lambda_{p+\ell} + \frac{1}{2}(N+1) - \ell$ in the second column of $\mathrm{Ann}(\mathscr{A}_{\underline{d}_0}(\psi))$ is strictly less than $\lambda_{p+\ell} + \frac{1}{2}(N+1) - \ell - 1$. This shows that the tableaux $\mathrm{Ann}(\mathscr{A}_{\underline{d}_0}(\psi))$ and $\mathrm{Ann}(\pi_\lambda)$ are different. Suppose $\beta \neq \lambda_1 + \frac{1}{2}(p-q-1)$. Then, the $(1,2)$-th entry in $\mathrm{Ann}(\mathscr{A}_{\underline{d}_0}(\psi))$ is β by $q_j > \#(\nu_j \cap \nu_{<j})$. Hence, we have $\beta = \lambda_{p+1} + \frac{1}{2}(N-1)$ if $\mathscr{A}_{\underline{d}_0}(\psi) \cong \pi_\lambda$. In other words, one has $I \subset \nu_j \subset Q'$ if $\mathscr{A}_{\underline{d}_0}(\psi) \cong \pi_\lambda$.

The converse follows from Lemmas 3.2–3.5 and Corollaries 3.6–3.7. This completes the proof.

5.5. ***Proof of main theorem: the case $\lambda_p - \lambda_{p+1} < N - p', N - q'$.*** We first show the only if part. Put $\nu_j = [\alpha, \beta]$. In this case, one has $q_j > 0$ by Lemma 3.2(3). Then, ν_j contains I and $\nu_{<j} \sqcup \nu_{>j}$ is multiplicity free. Let $\pi(\psi)$ be the cohomological induction defined in the same way in (5-1). The statement follows from this lemma:

Lemma 5.4. *If $\pi(\psi) \cong \pi_\lambda$, one has $\nu_j = \left[\lambda_p - \frac{1}{2}(N-1), \lambda_{p+1} + \frac{1}{2}(N-1)\right]$.*

Proof. The statement follows from the explicit calculation of the associated tableaux. Recall the tableau $\mathrm{Ann}(\pi_\lambda)$. Put $f = \#\{x \in \nu \mid x > \lambda_{p+1} + \frac{1}{2}(N-1)\}$ and $m = \min\{f, N - (\lambda_p - \lambda_{p+1}) - p' - p''\}$, where $p'' = \#((P \cap Q') \setminus I)$. When $f \neq 0$, the $(1,2)$-th entry in $\mathrm{Ann}(\pi_\lambda)$ is $\lambda_{p+1} + \frac{1}{2}(N-1)$ by Lemmas 3.2(1) and 3.5. Let i_0 be the maximal positive integer such that $\lambda_{f+i_0} + \frac{1}{2}(p-q+1) - (f+i_0) \geq \lambda_{p+1} + \frac{1}{2}(N+1) - m - i_0$, if it exists. If there is no such i_0, set $i_0 = 0$. Then, the first column of $\mathrm{Ann}(\pi_\lambda)$ consists of

$$\underbrace{\lambda_1 + \tfrac{1}{2}(p-q-1), \lambda_2 + \tfrac{1}{2}(p-q-3), \dots, \lambda_f + \tfrac{1}{2}(p-q+1) - f,}_{f}$$

$$\left.\begin{aligned}&\lambda_{f+1} + \tfrac{1}{2}(p-q+1) - (f+1), \lambda_{f+2} + \tfrac{1}{2}(p-q+1) - (f+2),\\ &\qquad\qquad\qquad\qquad \dots, \lambda_{f+i_0} + \tfrac{1}{2}(p-q+1) - (f+i_0)\end{aligned}\right\} i_0$$

$$\begin{aligned}&\lambda_{p+m+i_0+1} + \tfrac{1}{2}(N-1) - (m+i_0), \lambda_{p+m+i_0+2} + \tfrac{1}{2}(N-1) - (m+i_0+1),\\ &\qquad\qquad\qquad\qquad\qquad\qquad \dots, \lambda_p - \tfrac{1}{2}(N-1), \dots\end{aligned}$$

from top to bottom. The entry adjacent to $\lambda_p - \frac{1}{2}(N-1)$ is strictly less than $\lambda_p - \frac{1}{2}(N-1) - 1$ since $p - f$ is strictly greater than the number of elements in ν with multiplicity two. When $f = 0$, the $(1,2)$-th entry in $\mathrm{Ann}(\pi_\lambda)$ is the maximal member in ν with multiplicity two. This is greater than or equal to $\lambda_p - \frac{1}{2}(N-1) + p'$. Then the first column consists of

$$\lambda_{p+1} + \tfrac{1}{2}(N-1), \lambda_{p+1} + \tfrac{1}{2}(N-3), \dots, \lambda_p - \tfrac{1}{2}(N-3), \lambda_p - \tfrac{1}{2}(N-1), \dots$$

from top to bottom.

We claim $\beta = \lambda_{p+1} + \frac{1}{2}(N-1)$. Note that $q_j > \#(\nu_j \cap \nu_{<j})$ and $p_j > \#(\nu_j \cap \nu_{>j})$ if $\pi(\psi) \cong \pi_\lambda$ by Lemma 3.2 and Corollaries 3.6–3.7. Suppose first that $\beta < \lambda_{p+1} + \frac{1}{2}(N-1)$. Then, there exists $x \in \nu$ such that $x > \beta$. In this case, the $(1,2)$-th entry in $\pi(\psi)$ is β by $q_j > \#(\nu_j \cap \nu_{<j})$. If $\pi(\psi) \cong \pi_\lambda$, there exists no $x \in \nu$ with $x > \lambda_{p+1} + \frac{1}{2}(N-1)$, i.e., $f = 0$, and β is the maximal member in ν with multiplicity two. Consider the number of boxes from the top to the box filled with $\lambda_{p+1} + \frac{1}{2}(N-1) - q'$ in the first column. For π_λ, by $f = 0$, this number is q', but for $\pi(\psi)$, it is strictly less than q' by $\beta < \lambda_{p+1} + \frac{1}{2}(N-1)$ unless $\beta = \lambda_p - \frac{1}{2}(N+1) + p'$. If $\beta = \lambda_p - \frac{1}{2}(N+1) + p'$, consider the location of the box filled with the unique entry $\lambda_{p+1} + \frac{1}{2}(N-1) - q'$. This is in the first column for $\mathrm{Ann}(\pi_\lambda)$ and in the second column for $\mathrm{Ann}(\pi(\psi))$. Hence, the representations π_λ and $\mathscr{A}_{\underline{d}_0}(\psi)$ are different. Suppose next that $\beta > \lambda_{p+1} + \frac{1}{2}(N-1)$. Then, $f > 0$. If there exists $x \in \nu$ with $x > \beta$, the $(1,2)$-th entry in $\pi(\psi)$ is β. Then, one has $\beta = \lambda_{p+1} + \frac{1}{2}(N-1)$. This is a contradiction. When there exists no $x \in \nu$ with $x > \beta$, we compare the ν-antitableaux of π_λ and $\pi(\psi)$. In this case, for π_λ, the number of boxes from the top to the box filled with $\lambda_{p+1} + \frac{1}{2}(N+1) - q'$ in the first column is at most $\max\{-q + q' + (\lambda_p - \lambda_{p+1}), q'\}$, that is strictly less than $f+q'$ by $f \neq 0$. Note that $-q+q'+(\lambda_p-\lambda_{p+1})$ is equal to the integer a such that $\lambda_a+\frac{1}{2}(p-q+1)-a=\lambda_{p+1}+\frac{1}{2}(N+1)-q'$. For $\pi(\psi)$, the number of boxes from the top to the box filled with $\lambda_{p+1}+\frac{1}{2}(N+1)-q'$ in the first column is $f+q'$, since ν_j contains I. Hence, the tableaux are different. We conclude that β is equal to $\lambda_{p+1}+\frac{1}{2}(N-1)$ if $\pi_\lambda \cong \mathscr{A}_{\underline{d}_0}(\psi)$. In the following, we may assume $\beta=\lambda_{p+1}+\frac{1}{2}(N-1)$.

It remains to show $\alpha=\lambda_p-\frac{1}{2}(N-1)$. If there exists no $x \in \nu$ with $x > \lambda_{p+1}+\frac{1}{2}(N-1)$, the first column of $\mathrm{Ann}(\pi(\psi))$ consists of

$$\lambda_{p+1}+\tfrac{1}{2}(N-1)=\beta,\ldots,\alpha,\ldots$$

from top to bottom. Then, the entry next to α is strictly less than $\alpha-1$ since $q_j > \#(\nu_j \cap \nu_{<j})$. For π_λ, in the first column, the entry next to $\lambda_p-\frac{1}{2}(N-1)$ is strictly less than $\lambda_p-\frac{1}{2}(N-1)-1$ and the entry next to x with $\lambda_p-\frac{1}{2}(N-1)<x\leq \lambda_{p+1}+\frac{1}{2}(N-1)$ is $x-1$. Hence, if $\pi_\lambda \cong \mathscr{A}_{\underline{d}_0}(\psi)$, we have $\alpha=\lambda_p-\frac{1}{2}(N-1)$. We may assume that there exists $x \in \nu$ with $x > \lambda_{p+1}+\frac{1}{2}(N-1)$, i.e., $f \neq 0$. Recall that by $\pi_\lambda \cong \mathscr{A}_{\underline{d}_0}(\psi)$, we have $p_j > \#(\nu_4')$. This shows that the entry next to α in the first column is strictly less than $\alpha-1$ if the box exists. By the description of $\mathrm{Ann}(\pi_\lambda)$ and the same discussion above, we have $\alpha=\lambda_p-\frac{1}{2}(N-1)$. □

For the converse, apply Lemma 3.2(1) and Corollaries 3.6–3.7. We then have $\pi_\lambda \cong \mathscr{A}_{\underline{d}_0}(\psi)$. This completes the proof.

Acknowledgements

We wish to thank Takuya Yamauchi for invaluable comments and encouragement. We would also like to thank Du Chengyu for sharing ideas on the proof of the main theorem and clarifying our questions about the properties of $A_{\mathfrak{q}}(\mu)$. The author is partially supported by JSPS KAKENHI Grant Number 23K12965.

References

[Adams 1987] J. Adams, "Unitary highest weight modules", *Adv. in Math.* **63**:2 (1987), 113–137. MR

[Atobe 2020] H. Atobe, "On the non-vanishing of theta liftings of tempered representations of $\mathrm{U}(p,q)$", *Adv. Math.* **363** (2020), art. id. 106984. MR

[Atobe et al. 2024] H. Atobe, W. T. Gan, A. Ichino, T. Kaletha, A. Mínguez, and S. W. Shin, "Local intertwining relations and co-tempered A-packets of classical groups", preprint, 2024. arXiv 2410.13504

[Barbasch and Vogan 1983] D. Barbasch and D. Vogan, "Weyl group representations and nilpotent orbits", pp. 21–33 in *Representation theory of reductive groups* (Park City, UT, 1982), edited by P. C. Trombi, Progr. Math. **40**, Birkhäuser, Boston, MA, 1983. MR

[Chengyu 2025] D. Chengyu, "On the nonvanishing condition for $A_{\mathfrak{q}}(\lambda)$ of $U(p,q)$ in the mediocre range", preprint, 2025. arXiv 2405.03216

[Collingwood and McGovern 1993] D. H. Collingwood and W. M. McGovern, *Nilpotent orbits in semisimple Lie algebras*, Van Nostrand Reinhold, New York, 1993. MR

[Enright et al. 1983] T. Enright, R. Howe, and N. Wallach, "A classification of unitary highest weight modules", pp. 97–143 in *Representation theory of reductive groups* (Park City, UT, 1982), edited by P. C. Trombi, Progr. Math. **40**, Birkhäuser, Boston, MA, 1983. MR

[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**, Soc. Math. France, Paris, 2012. MR

[Horinaga et al. 2025] S. Horinaga, Y. Maeda, and T. Yamauchi, "The Kodaira dimension of even dimensional ball quotients", preprint, 2025. arXiv 2507.22203

[Huang 2025] C. Huang, "Non-zero condition on Mœglin–Renard's parametrization for Arthur packets of $\mathrm{U}(p,q)$", preprint, 2025. arXiv 2409.09358

[Huang and Pandžić 2006] J.-S. Huang and P. Pandžić, *Dirac operators in representation theory*, Birkhäuser, Boston, MA, 2006. MR

[Humphreys 2008] J. E. Humphreys, *Representations of semisimple Lie algebras in the BGG category $\mathcal{O}$*, Graduate Studies in Mathematics **94**, Amer. Math. Soc., Providence, RI, 2008. MR

[Ichino 2022] A. Ichino, "Theta lifting for tempered representations of real unitary groups", *Adv. Math.* **398** (2022), art. id. 108188. MR

[Kaletha et al. 2014] T. Kaletha, A. Minguez, S. W. Shin, and P.-J. White, "Endoscopic classification of representations: inner forms of unitary groups", preprint, 2014. arXiv 1409.3731

[Knapp and Vogan 1995] A. W. Knapp and D. A. Vogan, Jr., *Cohomological induction and unitary representations*, Princeton Mathematical Series **45**, Princeton Univ. Press, 1995. MR

[Mœglin and Renard 2019] C. Mœglin and D. Renard, "Sur les paquets d'Arthur des groupes unitaires et quelques conséquences pour les groupes classiques", *Pacific J. Math.* **299**:1 (2019), 53–88. MR

[Trapa 2001] P. E. Trapa, "Annihilators and associated varieties of $A_{\mathfrak{q}}(\lambda)$ modules for $\mathrm{U}(p,q)$", *Compositio Math.* **129**:1 (2001), 1–45. MR

[Vogan 1997] D. A. Vogan, Jr., "Cohomology and group representations", pp. 219–243 in *Representation theory and automorphic forms* (Edinburgh, 1996), edited by T. N. Bailey and A. W. Knapp, Proc. Sympos. Pure Math. **61**, Amer. Math. Soc., Providence, RI, 1997. MR

[Wong and Zhang 2024] K. D. Wong and H. Zhang, "The unitary dual of $U(n,2)$", *Int. Math. Res. Not.* **2024**:14 (2024), 10678–10707. MR

Received March 10, 2025. Revised June 19, 2025.

SHUJI HORINAGA
NTT INSTITUTE FOR FUNDAMENTAL MATHEMATICS
NTT COMMUNICATION SCIENCE LABORATORIES
NTT INC.
JAPAN
shorinaga@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 338, No. 2, 2025
https://doi.org/10.2140/pjm.2025.338.325

AN EVOLUTION OF MATRIX-VALUED ORTHOGONAL POLYNOMIALS

ERIK KOELINK, PABLO ROMÁN AND WADIM ZUDILIN

We establish new explicit connections between classical (scalar) and matrix Gegenbauer polynomials, which result in new symmetries of the latter and further give access to several properties that have been out of reach before: generating functions, distribution of zeros for individual entries of the matrices and new type of differential-difference structure. We further speculate about other potentials of the connection formulas found. Part of our proofs makes use of creative telescoping in a matrix setting — the strategy which is not yet developed algorithmically.

1. Introduction

When it comes to a topic as classical as orthogonal polynomials, one can hardly expect spectacular novelties. But they do happen: for example, in situations when a *natural* generalization is found. Decades ago, in the middle of the 20th century, M. G. Krein introduced matrix-valued orthogonal polynomials, in his study of higher-order differential operators and of the corresponding moment problem. This new topic progressed at a slow pace and mainly focused on analogs of classical results for scalar orthogonal polynomials; the overview in [7] gives an introduction and extensive literature up to 2008. Despite of the theoretical development, not so many concrete examples of matrix-valued orthogonal polynomials have been encountered. Those that have been found in the last decades using insights from representation theory demonstrate a rich structure, not always observable for their scalar originals. At the same time the polynomials are not easily accessible from a computational perspective; this makes it hard to draw further connections to other mathematics areas, for example, to number theory and analysis of special functions. In this paper we aim at changing this perception and demonstrating that naturally defined matrix-valued orthogonal polynomials are much closer to their scalar prototypes than expected. It is this closeness and its numerous consequences

MSC2020: primary 33C45; secondary 33C47, 33E30, 33F10.

Keywords: experimental mathematics, orthogonal polynomials, matrix-valued polynomials, differential-difference operators, zeros of polynomials, creative telescoping.

that we refer to as an *evolution* in the title. We are mainly concerned with matrix-valued analogs of the Gegenbauer polynomials

$$C_n^{(\nu)}(x) = \frac{(2\nu)_n}{n!} \sum_{k=0}^{n} \frac{(-n)_k (2\nu+n)_k}{k!\,\left(\nu+\frac{1}{2}\right)_k} \left(\frac{1-x}{2}\right)^k = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k \frac{\Gamma(\nu+n-k)}{k!\,(n-2k)!\,\Gamma(\nu)} (2x)^{n-2k},$$

also known as ultraspherical polynomials, where $(a)_k = \Gamma(a+k)/\Gamma(a)$ denotes the Pochhammer symbol. These matrix-valued polynomials $P_n^{(\nu)}(x)$ were introduced in [18]. For a special value of $\nu = 1$, they are matrix-valued analogs of the Chebyshev polynomials of the second kind and they had been previously defined using matrix-valued spherical functions, see [16; 17], building on earlier work of Koornwinder [21]. For the 2×2 case, a more general set of matrix-valued Gegenbauer polynomials is introduced by Pacharoni and Zurrián [23], and the entries are directly given as sum of two scalar Gegenbauer polynomials. The connection between these sets of 2×2-matrix-valued Gegenbauer polynomials is discussed in [18, Remark 3.8]. In this paper we give an extension of the expansion in scalar Gegenbauer polynomials for arbitrary size, see Theorem 3.6.

A connection between the scalar and matrix-valued Gegenbauer polynomials is somewhat intimate: the former appear in description of the matrix-valued weight for the latter, see [18] and Section 2 below. At the same time, the scalar Gegenbauer polynomials can be naturally promoted to matrix-valued orthogonal polynomials and their span connected with the span of the matrix-valued Gegenbauer polynomials. Such connections between different sets of matrix-valued orthogonal polynomials have been never recorded before; in Section 3 we give two explicit connection formulas for the scalar and matrix-valued Gegenbauer polynomials. With their help we can further explore the structure of related difference and differential equations. In particular, we construct in Section 4 new type of differential equations satisfied by the matrix-valued orthogonal polynomials. These equations utilize the non-commutative structure of the matrix-valued orthogonal polynomials and, therefore, degenerate in the scalar setting; this is perhaps a reason for their nonappearance in the literature.

The expressions in Section 3 lead to a fairly simple computational access to the matrix-valued polynomials. For example, they allow us to discuss generating functions of the matrix-valued Gegenbauer polynomials using known generating functions of the scalar ones; this is done in Section 5. The explicit formulas and numerical check suggest further that the zeros of entries of the matrix-valued polynomials follow distinguished patterns — those serve as a generalization of the property of scalar orthogonal polynomials to have all zeros located in the convex hull of the orthogonality measure on the real line. We speculate about these observations in Section 6, also towards other known matrix-valued orthogonal polynomials. Finally, in Section 7 we highlight some further potential applications of the connection formulas from Section 3.

The lineup of our exposition is somewhat misleading for how our results were actually found. We first looked for distribution of the zeros of entries of Gegenbauer polynomials of “reasonable” matrix size utilizing their expressions from [18] as triple hypergeometric sums. We observed that, for certain entries, the zeros were real and interlacing with those of the corresponding polynomial entry from the previous-indexed Gegenbauer. Such considerations helped us to suspect an explicit connection with scalar orthogonal polynomials and to realize that simpler expressions exist; making use of basic principles of experimental mathematics [3] we managed to recognize new symmetries of the matrix-valued polynomials and figure out what now comes as Theorem 3.6. In order to prove the corresponding formulas we needed to invent a matrix generalization of the famous Wilf–Zeilberger algorithm of creative telescoping [25; 26]; as no implementation of such exists at the moment of writing, the related linear algebra calculations were manually performed. Theorem 3.6 suggests that the formulas can be inverted, to express scalar Gegenbauer polynomials in terms of matrix-valued ones; after another round of experimentation we ended up with what is now Theorem 3.4. Its proof is more in “classical spirits” — in contrast with the other proof, we could not find a creative-telescoping argument. Our analysis of the zeros of entries of Gegenbauer polynomials played an important role in the execution and further development of this project. Though we only possess a limited explanation of the structure of these zeros, we feel a need for sharing our observations with the reader, so we display them in a condensed form in Section 6.

It seems to be worthy of pointing out in this introduction that a potential use of creative telescoping in the matrix- or vector-valued (noncommutative!) setup may give access to interesting hypergeometric identities, not necessarily linked to matrix-valued orthogonal polynomials. We hope that related algorithms will be designed and made efficient in the near future, with several nice applications that we cannot foresee at the moment.

2. Gegenbauer polynomials

We start this section with an overview of classical Gegenbauer polynomials and then review known facts from [18] about their matrix-valued mates.

The scalar Gegenbauer polynomials form a subfamily of the Jacobi polynomials. Their definition above can be put in the hypergeometric form

$$C_n^{(\nu)}(x) = \frac{(2\nu)_n}{n!}\,{}_2F_1\left(\begin{matrix}-n,\, 2\nu+n\\ \nu+\frac{1}{2}\end{matrix};\frac{1-x}{2}\right), \tag{2-1}$$

where we conventionally follow the standard notation, see [1; 2; 13; 15]. The Gegenbauer polynomials satisfy the connection formula

$$C_m^{(\nu)}(x) = \sum_{s=0}^{\lfloor m/2\rfloor} \frac{(\lambda+m-2s)(\nu)_{m-s}}{(\lambda)_{m-s+1}} \frac{(\nu-\lambda)_s}{s!} C_{m-2s}^{(\lambda)}(x). \tag{2-2}$$

In the particular case $\lambda = \nu + N$, $N \in \mathbb{N}$, the sum has a natural upper bound:

$$C_m^{(\nu)}(x) = \sum_{s=0}^{\lfloor m/2\rfloor \wedge N} \frac{(\nu+N+m-2s)(\nu)_{m-s}}{(\nu+N)_{m-s+1}} \frac{(-N)_s}{s!} C_{m-2s}^{(\nu+N)}(x), \tag{2-3}$$

where $\lfloor m/2\rfloor \wedge N$ denotes the minimum of $\lfloor m/2\rfloor$ and N. The linearization formula for the Gegenbauer polynomials reads

$$\begin{aligned} C_k^{(\nu)}(x)\,C_l^{(\nu)}(x) &\\ = \sum_{p=0}^{k\wedge l} \frac{k+l+\nu-2p}{k+l+\nu-p} & \frac{(\nu)_p(\nu)_{k-p}(\nu)_{l-p}(2\nu)_{k+l-p}}{p!\,(k-p)!\,(l-p)!\,(\nu)_{k+l-p}} \frac{(k+l-2p)!}{(2\nu)_{k+l-2p}} C_{k+l-2p}^{(\nu)}(x). \end{aligned} \tag{2-4}$$

The orthogonality relations

$$\int_{-1}^{1} C_k^{(\nu)}(x)\,C_n^{(\nu)}(x)(1-x^2)^{\nu-\frac{1}{2}}\,dx = \delta_{k,n} \frac{\pi\, 2^{1-2\nu}\,\Gamma(n+2\nu)}{\Gamma(\nu)^2(n+\nu)\,n!} \tag{2-5}$$

hold for $\nu > -\frac{1}{2}$, with a slightly different normalization required when $\nu = 0$. Since the Gegenbauer polynomials are orthogonal, they satisfy a three-term recurrence relation; it is

$$2(n+\nu)\,x C_n^{(\nu)}(x) = (n+1)\,C_{n+1}^{(\nu)}(x) + (n+2\nu-1)\,C_{n-1}^{(\nu)}(x), \tag{2-6}$$

which determines the polynomials from the initial conditions $C_{-1}^{(\nu)}(x) = 0$, $C_0^{(\nu)}(x) = 1$. We also use the Gegenbauer polynomials for negative ν, in which case we follow the convention in [5]. We refer to [1; 2; 13; 15] for these results and for more information on the Gegenbauer polynomials; see, in particular, Askey [2] for the history and importance of the connection and linearization formulas (2-2), (2-4).

We now review the matrix-valued Gegenbauer polynomials. For $\ell \in \frac{1}{2}\mathbb{N}$ and $\nu > 0$, following [18, Definition 2.1] their matrix-valued weight is the $(2\ell+1)\times(2\ell+1)$-matrix-valued function $W(x) = W^{(\nu)}(x)$ given by

$$(W(x))_{i,j} = (1-x^2)^{\nu-\frac{1}{2}} \sum_{k=0\vee i+j-2\ell}^{i\wedge j} \alpha_k(i,j)\,C_{i+j-2k}^{(\nu)}(x), \tag{2-7}$$

$$\begin{aligned} \alpha_k(i,j) = (-1)^k \frac{i!\,j!\,(i+j-2k)!}{k!\,(2\nu)_{i+j-2k}(\nu)_{i+j-k}} & \frac{(\nu)_{i-k}(\nu)_{j-k}}{(i-k)!\,(j-k)!} \frac{i+j-2k+\nu}{i+j-k+\nu} \\ & \times \frac{(2\ell-i)!\,(2\ell-j)!}{(2\ell+k-i-j)!}(-2\ell-\nu)_k \frac{(2\ell+\nu)}{(2\ell)!}, \end{aligned}$$

where $i, j \in \{0, 1, \dots, 2\ell\}$ and the notation $a \vee b$, $a \wedge b$ stands for $\min\{a, b\}$ and $\max\{a, b\}$, respectively. We slightly alter the expression for α_k from [18] to make $W^{(\nu)}$ transparently symmetric. Finally, put $W^{(\nu)}(x) = (1-x^2)^{\nu-1/2} W^{(\nu)}_{\mathrm{pol}}(x)$ (correcting a typo in [18, p. 463], where a superfluous $(1-x^2)^{\nu-1/2}$ appears in the first line). It turns out that $W^{(\nu)}(x)$ is positive definite; this follows from the LDU-decomposition [18, Theorem 2.1], which states

$$W^{(\nu)}(x) = L^{(\nu)}(x)\, T^{(\nu)}(x)\, L^{(\nu)}(x)^t, \quad x \in (-1, 1), \tag{2-8}$$

where $L^{(\nu)} : [-1, 1] \to M_{2\ell+1}(\mathbb{C})$ is the unipotent lower triangular matrix-valued polynomial,

$$(L^{(\nu)}(x))_{m,k} = \begin{cases} 0 & \text{if } m < k, \\ \dfrac{m!}{k!\,(2\nu+2k)_{m-k}} C^{(\nu+k)}_{m-k}(x) & \text{if } m \ge k, \end{cases}$$

and $T^{(\nu)} : (-1, 1) \to M_{2\ell+1}(\mathbb{C})$ is the diagonal matrix-valued function,

$$(T^{(\nu)}(x))_{k,k} = t^{(\nu)}_k (1-x^2)^{k+\nu-\frac{1}{2}}, \quad t^{(\nu)}_k = \frac{k!\,(\nu)_k}{\left(\nu+\frac{1}{2}\right)_k} \frac{(2\nu+2\ell)_k\,(2\ell+\nu)}{(2\ell-k+1)_k\,(2\nu+k-1)_k}.$$

Let J be the matrix with 1s along the antidiagonal, $J_{i,j} = \delta_{i+j,2\ell}$. Then $J W^{(\nu)}(x) = W^{(\nu)}(x)\, J$ for all $x \in (-1, 1)$ by [18, Proposition 2.6].

The monic matrix-valued Gegenbauer polynomials $P^{(\nu)}_n(x)$, $n \in \mathbb{N}$, are orthogonal with respect to the matrix-valued measure (2-7), i.e.,

$$\int_{-1}^{1} P^{(\nu)}_n(x)\, W^{(\nu)}(x) (P^{(\nu)}_m(x))^* \, dx = \delta_{n,m} H^{(\nu)}_n, \tag{2-9}$$

$$P^{(\nu)}_n(x) = x^n P^{(\nu)}_{n,n} + x^{n-1} P^{(\nu)}_{n,n-1} + \cdots + x P^{(\nu)}_{n,1} + P^{(\nu)}_{n,0}, \quad P^{(\nu)}_{n,n} = \mathbf{1},\ P^{(\nu)}_{n,i} \in M_{2\ell+1}(\mathbb{C});$$

a formula for the squared norm $H^{(\nu)}_n$, a positive definite diagonal matrix, can be found in [18, Theorem 3.1(i)]. The matrix conjugation $*$ in our real setting simply means the transpose.

An important property of the matrix-valued Gegenbauer polynomials is

$$\frac{dP^{(\nu)}_n}{dx}(x) = n P^{(\nu+1)}_{n-1}(x); \tag{2-10}$$

see [18, Theorem 3.1(ii)].

Similar to the scalar case, matrix-valued orthogonal polynomials satisfy a three-term recurrence relation. In the case of matrix-valued Gegenbauer polynomials the relation assumes the explicit form [18, Theorem 3.3]

$$x P^{(\nu)}_n(x) = P^{(\nu)}_{n+1}(x) + B^{(\nu)}_n P^{(\nu)}_n(x) + C^{(\nu)}_n P^{(\nu)}_{n-1}(x), \tag{2-11}$$

where the matrices $B_n^{(\nu)}$ and $C_n^{(\nu)}$ are given by

$$B_n^{(\nu)} = \sum_{j=1}^{2\ell} \frac{j(j+\nu-1)}{2(j+n+\nu-1)(j+n+\nu)} E_{j,j-1} + \sum_{j=0}^{2\ell-1} \frac{(2\ell-j)(2\ell-j+\nu-1)}{2(2\ell-j+n+\nu-1)(2\ell+n-j+\nu)} E_{j,j+1},$$

$$C_n^{(\nu)} = \sum_{j=0}^{2\ell} \frac{n(n+\nu-1)(2\ell+n+\nu)(2\ell+n+2\nu-1)}{4(2\ell+n+\nu-j-1)(2\ell+n+\nu-j)(j+n+\nu-1)(j+n+\nu)} E_{j,j}.$$

Here $E_{i,j}$ denotes the matrix with 1 on the (i, j)-entry and 0 elsewhere. With the initial conditions $P_{-1}^{(\nu)}(x) = 0$, $P_0^{(\nu)}(x) = \mathbf{1}$, the recurrence (2-11) determines the family $(P_n^{(\nu)}(x))_{n\in\mathbb{N}}$ completely.

The matrix-valued Gegenbauer polynomials can actually be made symmetric, a feature which seems to be rather uncommon. We define

$$\hat{P}_n^{(\nu)}(x) = D_n^{(\nu)} P_n^{(\nu)}(x), \quad (D_n^{(\nu)})_{i,j} = \delta_{i,j} \binom{2\ell}{i} \frac{(\nu+n)_i}{(\nu+n+2\ell-i)_i}, \tag{2-12}$$

and we show in Corollary 3.7 that $(\hat{P}_n^{(\nu)}(x))^t = \hat{P}_n^{(\nu)}(x)$, so that $\hat{P}_n^{(\nu)}(x)$ is symmetric.

Remark 2.1. (i) Notice that the diagonal matrix $D_n^{(\nu)}$ depends only on $\nu+n$; in particular, $D_n^{(\nu)} = D_{n-1}^{(\nu+1)}$ and by (2-10) we have

$$\frac{d}{dx} \hat{P}_n^{(\nu)}(x) = n D_n^{(\nu)} (D_{n-1}^{(\nu+1)})^{-1} \hat{P}_{n-1}^{(\nu+1)}(x) = n \hat{P}_{n-1}^{(\nu+1)}(x). \tag{2-13}$$

(ii) $J D_n^{(\nu)} = D_n^{(\nu)} J$, where as above J is the matrix with 1s along the antidiagonal.

3. The expansion of Gegenbauer polynomials in matrix-valued Gegenbauer polynomials

The goal of this section is to establish two special cases of a connection formula between the matrix-valued Gegenbauer polynomials and the scalar Gegenbauer polynomials. To start with, we define the matrices $F_{k,n}^{(\nu)}$, $k \in \{0, 1, \ldots, n\}$, by

$$\hat{P}_n^{(\nu)}(x) = \sum_{k=0}^{n} F_{k,n}^{(\nu)} C_{n-k}^{(\nu+2\ell)}(x) \tag{3-1}$$

and in dual setting we define $G_{k,n}^{(\nu)}$, $k \in \{0, 1, \ldots, n\}$, by

$$C_m^{(\nu)}(x)\mathbf{1} = \sum_{r=0}^{m} G_{r,m}^{(\nu)} \hat{P}_{m-r}^{(\nu)}(x). \tag{3-2}$$

We show that the summation ranges in (3-1), (3-2) are bounded by 2ℓ, so that the number of nonzero terms in (3-1), (3-2) is at most the size of the matrix-valued polynomials. We first describe the relation arising from (2-13).

Lemma 3.1. *For $n, m \in \mathbb{N}$, $n \geq 1$, $m \geq 1$, we have*

$$F^{(\nu)}_{k,n} = \frac{n}{2(\nu+2\ell)} F^{(\nu+1)}_{k,n-1}, \quad 0 \leq k \leq n-1,$$
$$G^{(\nu)}_{r,m} = \frac{2\nu}{m-r} G^{(\nu+1)}_{r,m-1}, \qquad 0 \leq r \leq m-1.$$

Proof. Differentiate (3-1) using $\frac{d}{dx} C^{(\lambda)}_n(x) = 2\lambda C^{(\lambda+1)}_{n-1}(x)$ and (2-13) to find out that

$$\begin{aligned} n \hat{P}^{(\nu+1)}_{n-1}(x) &= \sum_{k=0}^{2\ell} F^{(\nu)}_{k,n}\, 2(\nu+2\ell)\, C^{(\nu+1+2\ell)}_{n-1-k}(x) \\ &= 2(\nu+2\ell) \sum_{k=0}^{2\ell} (D^{(\nu)}_n)^{-1} D^{(\nu+1)}_{n-1} F^{(\nu)}_{k,n}\, C^{(\nu+1+2\ell)}_{n-1-k}(x). \end{aligned}$$

On the other hand, applying (3-1) with (n, ν) replaced by $(n-1, \nu+1)$ gives

$$F^{(\nu+1)}_{k,n-1} = \frac{2(\nu+2\ell)}{n} F^{(\nu)}_{k,n}$$

by uniqueness of the expansion. This proves the first statement, and the second one follows analogously. □

In particular, it follows from Lemma 3.1 that

$$G^{(\nu)}_{m,m+k} = \frac{2^k (\nu)_k}{k!} G^{(\nu+k)}_{m,m} \quad \text{and} \quad F^{(\nu)}_{m,m+k} = \frac{(m+1)_k}{2^k(\nu+2\ell)_k} F^{(\nu+k)}_{m,m} \tag{3-3}$$

for $m, k \in \mathbb{N}$. Therefore, viewing $(G^{(\nu)}_{r,m})_{r,m\in\mathbb{N}}$ and $(F^{(\nu)}_{r,m})_{r,m\in\mathbb{N}}$ as infinite matrices we see that they are upper triangular, with (3-3) showing that each element in the r-th row is determined by the element on the diagonal in the r-th row.

Next we look at $G^{(\nu)}_{m,m}$ and $F^{(\nu)}_{m,m}$. From (3-3) we see that the upper bound of the sum in (3-2), respectively (3-1), can be replaced by $2\ell \wedge m = \min(2\ell, m)$, respectively $2\ell \wedge n = \min(2\ell, n)$, if we show that $G^{(\nu)}_{m,m} = F^{(\nu)}_{m,m} = 0$ for $m > 2\ell$.

Lemma 3.2. $F^{(\nu)}_{m,m} = 0$ *for* $m > 2\ell$.

Proof. Note that the (i, j)-th entry of the matrix $F^{(\nu)}_{m,m}$ is a multiple of the integral

$$\int_{-1}^{1} (P^{(\nu)}_m(x))_{i,j}\, (1-x^2)^{\nu+2\ell-\frac{1}{2}}\, dx \tag{3-4}$$

and we need to show that this integral vanishes for $m > 2\ell$. By [18, Theorem 3.4],

$$(P^{(\nu)}_m(x))_{i,j} = \sum_{p=j}^{2\ell\wedge m+i} z_p\, C^{(\nu+p)}_{m+i-p}(x)\, C^{(1-\nu-p)}_{p-j}(x)$$

for some explicit constants z_p, so that the integral (3-4) equals

$$\sum_{p=j}^{2\ell\wedge m+i} z_p \int_{-1}^{1} C^{(\nu+p)}_{m+i-p}(x) \{C^{(1-\nu-p)}_{p-j}(x)(1-x^2)^{2\ell-p}\} (1-x^2)^{\nu+p-\frac{1}{2}}\, dx.$$

Using the orthogonality (2-5) and the fact that the term in the parentheses is a polynomial of degree $4\ell - p - j$, we see that the p-th term vanishes for $m+i-p > 4\ell - p - j$, that is, for $m > 4\ell - i - j$. Therefore, the integral (3-4) vanishes for $m > 2\ell$ when $i + j \leq 2\ell$.

Because J commutes with $W^{(\nu)}(x)$ for all x, see [18, Proposition 2.6], we also have that the monic polynomials $P_n^{(\nu)}(x)$ commute with J by [19, Lemma 3.1(2)]. In particular, $(P_m^{(\nu)}(x))_{2\ell-i,2\ell-j} = (P_m^{(\nu)}(x))_{i,j}$ and the integral in (3-4) is zero for $m > 2\ell$ for all (i, j). □

In principle the approach of Lemma 3.2 can also be used to calculate $F_{m,m}^{(\nu)}$ for $m \leq 2\ell$; this gives an explicit but rather complicated expression for $F_{k,m}^{(\nu)}$ with the help of (3-3). In Theorem 3.6 we give a simple formula for $F_{k,m}^{(\nu)}$. We leave it open whether the summation formula induced from these two different calculations is of interest on its own.

For the matrices $G_{k,m}^{(\nu)}$ we first recall that we can rewrite (2-9) as

$$\int_{-1}^{1} \hat{P}_n^{(\nu)}(x)\, W(x) (\hat{P}_m^{(\nu)}(x))^*\, dx = \delta_{m,n}\, \hat{H}_n^{(\nu)}, \tag{3-5}$$

with $\hat{H}_n^{(\nu)} = D_n^{(\nu)} H_n^{(\nu)} (D_n^{(\nu)})^*$ being a diagonal matrix as well. Then

$$\begin{aligned}\int_{-1}^{1} C_m^{(\nu)}(x)\, W(x) (D_0^{(\nu)})^*\, dx &= \int_{-1}^{1} \sum_{r=0}^{m} G_{r,m}^{(\nu)}\, \hat{P}_{m-r}^{(\nu)}(x)\, W(x) (\hat{P}_0^{(\nu)}(x))^*\, dx \\ &= G_{m,m}^{(\nu)}\, \hat{H}_0^{(\nu)}\end{aligned}$$

and

$$G_{m,m}^{(\nu)}\, D_0^{(\nu)} H_0^{(\nu)} = \int_{-1}^{1} C_m^{(\nu)}(x)\, W(x)\, dx. \tag{3-6}$$

Note that $D_0^{(\nu)} H_0^{(\nu)}$ is an explicit invertible diagonal matrix, so that all we need to do is to calculate the matrix entries of

$$\int_{-1}^{1} C_m^{(\nu)}(x)\, W(x)_{i,j}\, dx$$

to determine $G_{m,m}^{(\nu)}$. Observe that $W(x)_{i,j} = W(x)_{j,i}$, since W is Hermitian and real-valued for $x \in (-1, 1)$. Furthermore, $W(x)_{i,j} = W(x)_{2\ell-i,2\ell-j}$ by Proposition 2.6 in [18]. This reduces our calculation to evaluating the integral

$$\int_{-1}^{1} C_m^{(\nu)}(x)\, W(x)_{i,j}\, dx$$

for the case $i \geq j,\ i + j \leq 2\ell$.

Lemma 3.3. *Assume $j \ge i$, $i+j \le 2\ell$. Then*

$$\int_{-1}^{1} C_m^{(\nu)}(x)\, W(x)_{i,j}\, dx = 0,$$

when $m < j-i$ or $i+j \not\equiv m \bmod 2$ or $m > 2\ell$.

In particular, Lemma 3.3 shows that the r-th term of the sum (3-2) vanishes when $r > 2\ell \wedge m$.

Proof. Recall formula (2-7). Using the orthogonality relations (2-5) and the fact that the matrix-valued Gegenbauer polynomials are symmetric, the statement follows. □

Theorem 3.4. *Define*

$$\begin{aligned}
&\phi(\nu; i, j, r)\\
&= \phi_\ell(\nu; i, j, r)\\
&= \binom{2\ell}{r}\binom{r}{\frac{1}{2}(r+i-j)}\binom{2\ell-r}{\frac{1}{2}(i+j-r)}\binom{2\ell}{i}^{-1}\binom{2\ell}{j}^{-1}(\nu-r+j)(\nu+2\ell-r-j)\\
&\qquad\times \frac{\Gamma(\nu+2\ell-r)\,\Gamma\bigl(\nu-\frac{1}{2}(r-i+j)\bigr)\,\Gamma\bigl(\nu-\frac{1}{2}(r+i-j)\bigr)}{\Gamma\bigl(\nu+1-\frac{1}{2}(r-i-j)\bigr)\,\Gamma\bigl(\nu+2\ell+1-\frac{1}{2}(r+i+j)\bigr)}
\end{aligned}$$

if $i+j \equiv r \bmod 2$ and $\phi(\nu; i, j, r) = 0$ otherwise. Then for $0 \le r \le 2\ell$, $m \in \mathbb{N}$, the matrices

$$(G_{r,m}^{(\nu)})_{i,j} = \frac{2^{m-r}}{(m-r)!\,\Gamma(\nu)} \cdot \phi(\nu+m; i, j, r), \quad i, j \in \{0, 1, \ldots, 2\ell\},$$

satisfy

$$C_m^{(\nu)}(x)\mathbf{1} = \sum_{r=0}^{m\wedge 2\ell} G_{r,m}^{(\nu)}\, \hat{P}_{m-r}^{(\nu)}(x).$$

Proof. The zero entries of the matrix in (3-6) given in Lemma 3.3 can be ignored. Employing the formula

$$(H_0^{(\nu)})_{j,j} = \sqrt{\pi}\frac{\Gamma\bigl(\nu+\frac{1}{2}\bigr)}{\Gamma(\nu+1)}(2\ell+\nu)\frac{j!\,(2\ell-j)!\,(\nu+1)_{2\ell}}{(2\ell)!\,(\nu+1)_j(\nu+1)_{2l-j}}$$

for the squared norm at $n=0$ and (2-7) for $\alpha_k(i, j)$, using the expression for $D_n^{(\nu)}$ from (2-12) and the orthogonality relations (2-5) in (3-6), we can evaluate $G_{m,m}^{(\nu)}$. Then the formula for $G_{m,r}^{(\nu)}$ follows from the shift property (3-3). A straightforward calculation gives the result. □

As a corollary to the proof of Theorem 3.4, we find symmetry properties for $G_{r,m}^{(\nu)}$.

Corollary 3.5. *We have $G_{r,m}^{(\nu)} J = J G_{r,m}^{(\nu)}$, and the matrix $G_{r,m}^{(\nu)} D_0^{(\nu)} H_0^{(\nu)}$ is symmetric.*

Proof. The explicit expression of $D_n^{(\nu)}$ in (2-12) shows that $D_n^{(\nu)}$ commutes with J. Furthermore, $W(x)$ commutes with J by [18, Proposition 2.6]. This implies that J commutes with $P_n^{(\nu)}(x)$ and $H_n^{(\nu)}$ by [19, Lemma 3.1] and the fact that $J^*=J$. In turn, this means that J also commutes with $\hat{P}_n^{(\nu)}(x)$ and $\hat{H}_n^{(\nu)}$. With the help of (3-6) we deduce that

$$\begin{aligned} G_{m,m}^{(\nu)} &= \int_{-1}^{1} C_m^{(\nu)}(x)\, W^{(\nu)}(x)\, dx\, (D_0^{(\nu)})^*(\hat{H}_0^{(\nu)})^{-1} \\ &= \int_{-1}^{1} C_m^{(\nu)}(x)\, W^{(\nu)}(x)\, dx\, (H_0^{(\nu)})^{-1}(D_0^{(\nu)})^{-1}, \end{aligned}$$

so that $G_{m,m}^{(\nu)}$ commutes with J. By (3-3) we see that $G_{r,m}^{(\nu)}$ commutes with J. This also implies that $G_{m,m}^{(\nu)} D_0^{(\nu)} H_0^{(\nu)}$ is a symmetric matrix, and hence the statement for $G_{r,m}^{(\nu)}$ follows from (3-3). □

Our next objective is to determine $F_{k,n}^{(\nu)}$ using Lemma 3.2. We follow an indirect approach and prove (3-1) by showing that the right-hand side of (3-1) satisfies the three-term recurrence relation for the polynomials $\hat{P}_n^{(\nu)}(x)$. In other words, we want to show that these polynomials satisfy the three-term recurrence relation

$$x\hat{P}_n^{(\nu)}(x) = \hat{A}_n^{(\nu)}\hat{P}_{n+1}^{(\nu)}(x) + \hat{B}_n^{(\nu)}\hat{P}_n^{(\nu)}(x) + \hat{C}_n^{(\nu)}\hat{P}_{n-1}^{(\nu)}(x), \tag{3-7}$$

with $\hat{A}_n^{(\nu)} = D_n^{(\nu)}(D_{n+1}^{(\nu)})^{-1}$, $\hat{B}_n^{(\nu)} = D_n^{(\nu)} B_n^{(\nu)}(D_n^{(\nu)})^{-1}$ and $\hat{C}_n^{(\nu)} = D_n^{(\nu)} C_n^{(\nu)}(D_{n-1}^{(\nu)})^{-1}$ with the matrices $B_n^{(\nu)}$, $C_n^{(\nu)}$ as in (2-11). Note that $\hat{A}_n^{(\nu)}$ and $\hat{C}_n^{(\nu)}$ are symmetric, however $\hat{B}_n^{(\nu)}$ is not in general.

Theorem 3.6. *For $i, j, k \in \{0, 1, \dots, 2\ell\}$, define*

$$\begin{aligned} \gamma(\nu; i, j, k) &= \gamma_\ell(\nu; i, j, k) \\ &= (-1)^k(\nu+2\ell)(\nu+2\ell-k)\binom{2\ell}{k}\binom{k}{\frac{1}{2}(k+i-j)}\binom{2\ell-k}{\frac{1}{2}(i+j-k)} \\ &\quad\times \frac{\Gamma\big(\nu-\frac{1}{2}(k-i-j)\big)\,\Gamma\big(\nu+2\ell-\frac{1}{2}(k+i+j)\big)}{\Gamma(\nu)\,\Gamma\big(\nu+2\ell+1-\frac{1}{2}(k-i+j)\big)\,\Gamma\big(\nu+2\ell+1-\frac{1}{2}(k+i-j)\big)} \end{aligned}$$

if $i+j \equiv k \bmod 2$ and $\gamma(\nu; i, j, k) = 0$ otherwise. Let the matrix $F_{k,n}^{(\nu)}$ be given by

$$(F_{k,n}^{(\nu)})_{i,j} = \frac{n!\,\Gamma(\nu+2\ell)}{2^n}\gamma(\nu+n; i, j, k);$$

then

$$\hat{P}_n^{(\nu)}(x) = \sum_{k=0}^{n\wedge 2\ell} F_{k,n}^{(\nu)} C_{n-k}^{(\nu+2\ell)}(x).$$

Since $\binom{n}{k}$ is nonzero only for $n, k \in \mathbb{N}$ with $0 \le k \le n$, we see that $F_{0,n}^{(\nu)}$ is a diagonal matrix. Similarly, $F_{1,n}^{(\nu)}$ is zero on the diagonal and has nonzero sub- and superdiagonals.

Since the matrices $F^{(\nu)}_{k,n}$ are symmetric, the following corollary is immediate.

Corollary 3.7. *The polynomials $\hat{P}^{(\nu)}_n(x)$ are symmetric, i.e., $(\hat{P}^{(\nu)}_n(x))^t = \hat{P}^{(\nu)}_n(x)$. In particular, $(P^{(\nu)}_n(x))^t = D^{(\nu)}_n P^{(\nu)}_n(x)(D^{(\nu)}_n)^{-1}$.*

Proof of Theorem 3.6. Lemma 3.2 gives us the bound $n \wedge 2\ell$ for the number of terms in the expansion. We prove the latter by showing that the right-hand side satisfies the same three-term recurrence (3-7) and initial conditions for $n = -1$ and $n = 0$. For $n = -1$, we have that the right-hand side is the zero matrix. For $n = 0$, the right-hand side reduces to the diagonal matrix $F^{(\nu)}_{0,0}$ and by inspection

$$\begin{aligned}(F^{(\nu)}_{0,0})_{i,i} &= \Gamma(\nu+2\ell)(\nu+2\ell)^2\binom{2\ell}{i}\frac{\Gamma(\nu+i)\,\Gamma(\nu+2\ell-i)}{\Gamma(\nu)\,\Gamma(\nu+2\ell+1)^2}\\ &= \binom{2\ell}{i}\frac{(\nu)_i}{(\nu+2\ell-i)_i} = (D^{(\nu)}_0)_{i,i}.\end{aligned}$$

Thus, the initial values match, and it remains to show that the three-term recurrence relation (3-7) is satisfied by $\sum_{k=0}^{n\wedge 2\ell} F^{(\nu)}_{k,n} C^{(\nu+2\ell)}_{n-k}(x)$.

Using the three-term recurrence relation (2-6) for the Gegenbauer polynomials and the explicit expressions we see that (i, j)-entry of $x\sum_{k=0}^{n\wedge 2\ell} F^{(\nu)}_{k,n} C^{(\nu+2\ell)}_{n-k}(x)$ consists of two sums in terms of the Gegenbauer polynomials $C^{(\nu+2\ell)}_{n-k}(x)$ for $k \geq 0$. Similarly, since the matrices $\hat{A}^{(\nu)}_n$, $\hat{B}^{(\nu)}_n$ and $\hat{C}^{(\nu)}_n$ that appears on the right-hand side of (3-7) are either diagonal or have two nonzero diagonals, we see that the right-hand side involves four sums in terms of Gegenbauer polynomials $C^{(\nu+2\ell)}_{n-k}(x)$ with $k \geq 0$. Comparing the coefficients of $C^{(\nu+2\ell)}_{n-k}(x)$ on both sides, we see that we need the equality

$$\begin{aligned}&\frac{n-k}{\nu+2\ell+n-k-1}\gamma(\nu+n; i, j, k+1) + \frac{2(\nu+2\ell)+n-k}{\nu+2\ell+n-k+1}\gamma(\nu+n; i, j, k-1)\\ &\quad= \frac{(n+1)(\nu+n)(\nu+2\ell+n)}{(\nu+n+i)(\nu+2\ell+n-i)}\gamma(\nu+n+1; i, j, k+1)\\ &\qquad+ \frac{(\nu+i-1)(2\ell-i+1)}{(\nu+n+i)(\nu+2\ell+n-i)}\gamma(\nu+n; i-1, j, k)\\ &\qquad+ \frac{(i+1)(\nu+2\ell-i-1)}{(\nu+n+i)(\nu+2\ell+n-i)}\gamma(\nu+n; i+1, j, k)\\ &\qquad+ \frac{(\nu+2\ell+n)(2\nu+2\ell+n-1)}{(\nu+n+i)(\nu+2\ell+n-i)(\nu+2\ell+n-1)}\gamma(\nu+n-1; i, j, k-1)\end{aligned}$$

to be valid, where we have divided by the normalizing constant $n!\,\Gamma(\nu+2\ell)/2^{n+1}$. This identity is trivially true in case $i + j \not\equiv k \mod 2$, since all six terms equal zero. In the general case, we divide by $\gamma(\nu+n; i, j, k+1)$ and the required identity becomes an identity involving rational functions in the parameters. This rational identity is checked (by computer algebra) to be valid. □

The matrices $G^{(\nu)}_{r,m}$ and $F^{(\nu)}_{k,n}$ determined in Theorems 3.4 and 3.6 are related to expansion and summation formulas, which we state below.

Corollary 3.8. *The matrix-valued polynomials* $\hat{P}^{(\nu)}_n(x)$ *satisfy the expansion*

$$\hat{P}^{(\nu)}_n(x) = \sum_{t=0}^{n\wedge 4\ell} M^{(\nu)}_t \hat{P}^{(\nu+2\ell)}_{n-t}(x), \quad \textit{where } M^{(\nu)}_t = \sum_{k=0\vee t-2\ell}^{n\wedge 2\ell} F^{(\nu)}_{k,n} G^{(\nu+2\ell)}_{t-k,n-k}.$$

Corollary 3.8 follows immediately by first applying Theorem 3.6 and next Theorem 3.4. Note that Corollary 3.8 is a matrix analog of a very specific case of (2-3).

Corollary 3.9. *The following double summation result holds*: *for* $s \in \mathbb{N}$ *with* $0 \le s \le \lfloor m/2 \rfloor$ *and indices* i, j, *we have*

$$\sum_{p=0}^{2\ell} \sum_{r=0\vee 2s-2\ell}^{m\wedge 2\ell} (G^{(\nu)}_{r,m})_{i,p} (F^{(\nu)}_{2s-r,m-r})_{p,j} = \delta_{i,j} \frac{\nu+2\ell+m-2s}{\nu+2\ell} \frac{(\nu)_{m-s}}{(\nu+2\ell+1)_{m-s}} \frac{(-2\ell)_s}{s!}.$$

Proof. First apply Theorem 3.4 and then Theorem 3.6 to deduce

$$C^{(\nu)}_m(x)\mathbf{1} = \sum_{r=0}^{m\wedge 2\ell} \sum_{k=0}^{(m-r)\wedge 2\ell} G^{(\nu)}_{r,m} F^{(\nu)}_{k,m-r} C^{(\nu+2\ell)}_{m-p};$$

this connection formula of the scalar Gegenbauer polynomials is known: see (2-3) with $N = 2\ell$. Comparing the two we conclude that

$$\sum_{r=0\vee t-2\ell}^{m\wedge 2\ell} G^{(\nu)}_{r,m} F^{(\nu)}_{t-r,m-r} = \mathbf{0}$$

for t odd (which is already clear from the definitions in Theorems 3.4 and 3.6), and

$$\sum_{r=0\vee 2s-2\ell}^{m\wedge 2\ell} G^{(\nu)}_{r,m} F^{(\nu)}_{2s-r,m-r} = \frac{\nu+2\ell+m-2s}{\nu+2\ell} \frac{(\nu)_{m-s}}{(\nu+2\ell+1)_{m-s}} \frac{(-2\ell)_s}{s!}\mathbf{1}$$

for $t = 2s$ even. Taking the (i, j)-th entry we obtain the desired result. □

The identity in Corollary 3.9 gives an example of a matrix hypergeometric summation formula; we are not aware of its proof using classical hypergeometric identities.

4. Differential and difference equations related to matrix-valued Gegenbauer polynomials

Since the matrix-valued Gegenbauer polynomials $\hat{P}^{(\nu)}_n$ are symmetric by Corollary 3.7, we can take the transposed version of identities for the polynomials $\hat{P}^{(\nu)}_n$ and compare the resulting identity to the original one. This gives various identities for the matrix-valued polynomials $\hat{P}_n^{(\nu)}$; with the help of Theorem 3.6 we can rewrite those in terms of recurrence relations for the matrices $F_{k,n}^{(\nu)}$. Therefore, the fact that $\hat{P}_n^{(\nu)}$ are symmetric polynomials gives a series of identities which can be viewed as mixed differential-difference equations, where the differential and difference operators can act from both sides. This procedure can be performed for any set of difference, respectively differential, equations that occur in the left, respectively right, Fourier algebras associated to the matrix-valued Gegenbauer polynomials, see [6] for the definition. The results obtained do *not* have any classical analogs, since in the scalar case all commutators vanish. The resulting identities have a true matrix nature.

In this section we outline this procedure for three explicit situations. First we consider the three-term recursion for the matrix-valued Gegenbauer polynomials $\hat{P}_n^{(\nu)}$, and next we deal with the two matrix differential operators that have the matrix-valued Gegenbauer polynomials $\hat{P}_n^{(\nu)}$ as eigenfunctions, see [18, Theorems 2.3, 3.2].

We start with the three-term recursion (3-7), which is the basic example of an element in the left Fourier algebra. The symmetry of the matrix-valued orthogonal polynomials $\hat{P}_n^{(\nu)}(x)$ shows that the polynomials also satisfy a three-term recurrence relation with matrix multiplication by matrices depending on n from the right, i.e.,

$$x\,\hat{P}_n^{(\nu)}(x) = \hat{P}_{n+1}^{(\nu)}(x)\,\hat{A}_n^{(\nu)} + \hat{P}_n^{(\nu)}(x)(\hat{B}_n^{(\nu)})^t + \hat{P}_{n-1}^{(\nu)}(x)\,\hat{C}_n^{(\nu)}, \tag{4-1}$$

considering that the diagonal matrices $\hat{A}_n^{(\nu)}$ and $\hat{C}_n^{(\nu)}$ are automatically symmetric.

Proposition 4.1. *The symmetric matrix-valued Gegenbauer polynomials satisfy*

$$[\hat{P}_{n+1}^{(\nu)}(x),\, \hat{A}_n^{(\nu)}] + \hat{P}_n^{(\nu)}(x)(\hat{B}_n^{(\nu)})^t - \hat{B}_n^{(\nu)}\,\hat{P}_n^{(\nu)}(x) + [\hat{P}_{n-1}^{(\nu)}(x),\, \hat{C}_n^{(\nu)}] = 0.$$

In turn, the matrices $F_{k,n}^{(\nu)}$ defined in Theorem 3.6 satisfy

$$[F_{k+1,n+1}^{(\nu)},\, \hat{A}_n^{(\nu)}] + F_{k,n}^{(\nu)}(\hat{B}_n^{(\nu)})^t - \hat{B}_n^{(\nu)}\,F_{k,n}^{(\nu)} + [F_{k-1,n-1}^{(\nu)},\, \hat{C}_n^{(\nu)}] = 0,$$

with the convention that $F_{k,n}^{(\nu)} = 0$ for $k > n \wedge 2\ell$ or for $k < 0$, and $\hat{C}_0^{(\nu)} = 0$.

Entrywise the recursion for $F_{k,n}^{(\nu)}$ boils down to

$$\begin{aligned}((\hat{A}_n^{(\nu)})_{j,j} - (\hat{A}_n^{(\nu)})_{i,i})(F_{k+1,n+1}^{(\nu)})_{i,j} + (\hat{B}_n^{(\nu)})_{j,j-1}(F_{k,n}^{(\nu)})_{i,j-1} \\ + (\hat{B}_n^{(\nu)})_{j,j+1}(F_{k,n}^{(\nu)})_{i,j+1} - (\hat{B}_n^{(\nu)})_{i,i-1}(F_{k,n}^{(\nu)})_{i-1,j} - (\hat{B}_n^{(\nu)})_{i,i+1}(F_{k,n}^{(\nu)})_{i+1,j} \\ + ((\hat{C}_n^{(\nu)})_{j,j} - (\hat{C}_n^{(\nu)})_{i,i})(F_{k-1,n-1}^{(\nu)})_{i,j} = 0,\end{aligned}$$

with the explicit matrix entries for $\hat{A}_n^{(\nu)}$, $\hat{B}_n^{(\nu)}$ and $\hat{C}_n^{(\nu)}$ recorded in (2-11), (3-7).

Proof. The first part follows by subtracting (3-7) from (4-1). Note that this is a polynomial identity of degree n, since the leading coefficient $D_{n+1}^{(\nu)}$ of $\hat{P}_{n+1}^{(\nu)}$ commutes with $\hat{A}_n^{(\nu)}$ as both are diagonal.

The statement for the matrices $F_{k,n}^{(\nu)}$ then follows by plugging Theorem 3.6 in the identity from the first part. This procedure leads to an expansion in terms of Gegenbauer polynomials $C_m^{(\nu+2\ell)}$; collecting the coefficients of a Gegenbauer polynomial of fixed degree gives the result. □

Our next example arises from the second-order matrix differential operator of hypergeometric type for which the matrix-valued Gegenbauer polynomials are eigenfunctions. At the same time, the matrix-valued Gegenbauer polynomials are eigenfunctions for a first-order matrix differential equation. These two matrix differential operators are in the right Fourier algebra for the matrix-valued Gegenbauer polynomials. We recall the operators explicitly from [18, Theorem 3.2].

The second-order matrix hypergeometric differential operator for which the polynomials are eigenfunctions is given as follows, see [18, Theorems 2.3, 3.2], where we switched to the notation $\mathcal{D}^{(\nu)}$ in order to avoid confusion with the diagonal matrix $D_n^{(\nu)}$ used in this paper:

$$P_n^{(\nu)} \cdot \mathcal{D}^{(\nu)} = \Lambda_n(\mathcal{D}^{(\nu)}) P_n^{(\nu)}, \quad \Lambda_n(\mathcal{D}^{(\nu)}) = -n(2\ell+2\nu+n)\mathbf{1} - V, \tag{4-2}$$

with

$$\mathcal{D}^{(\nu)} = \frac{d^2}{dx^2}(1-x^2)\mathbf{1} + \frac{d}{dx}\big(C - x(2\ell+2\nu+1)\mathbf{1}\big) - V,$$
$$C = \sum_{i=0}^{2\ell-1}(2\ell-i)E_{i,i+1} + \sum_{i=1}^{2\ell} i E_{i,i-1}, \quad V = -\sum_{i=0}^{2\ell} i(2\ell-i)E_{i,i}.$$

It follows that

$$\hat{P}_n^{(\nu)} \cdot \mathcal{D}^{(\nu)} = D_n^{(\nu)} \Lambda_n(\mathcal{D}^{(\nu)})(D_n^{(\nu)})^{-1} \hat{P}_n^{(\nu)} = \Lambda_n(\mathcal{D}^{(\nu)}) \hat{P}_n^{(\nu)}, \tag{4-3}$$

since $D_n^{(\nu)}$ and $\Lambda_n(\mathcal{D}^{(\nu)})$ commute, being diagonal matrices.

Proposition 4.2. *In the notation above, the symmetric matrix-valued Gegenbauer polynomials satisfy*

$$\frac{d\hat{P}_n^{(\nu)}}{dx}(x)C - C^t \frac{d\hat{P}_n^{(\nu)}}{dx}(x) = -2[V, \hat{P}_n^{(\nu)}(x)].$$

In turn, the matrices $F_{k,n}^{(\nu)}$ defined in Theorem 3.6 satisfy

$$F_{k,n}^{(\nu)}C - C^t F_{k,n}^{(\nu)} = \frac{1}{\nu+2\ell+n-k-1}[F_{k+1,n}^{(\nu)}, V] - \frac{1}{\nu+2\ell+n-k+1}[F_{k-1,n}^{(\nu)}, V].$$

Observe that the first identity of Proposition 4.2 is a matrix-valued polynomial identity of degree $n-1$, since the leading coefficient of $\hat{P}_n^{(\nu)}$ is diagonal and commutes with V. Furthermore, notice that the participating matrices C and V depend on neither degree n nor parameter ν.

Entrywise the recursion for $F_{k,n}^{(\nu)}$ in Proposition 4.2 reads

$$\frac{V_{j,j}-V_{i,i}}{\nu+2\ell+n-k-1}(F_{k+1,n}^{(\nu)})_{i,j}-\frac{V_{j,j}-V_{i,i}}{\nu+2\ell+n-k+1}(F_{k-1,n}^{(\nu)})_{i,j}$$
$$=C_{j-1,j}(F_{k,n}^{(\nu)})_{i,j-1}+C_{j+1,j}(F_{k,n}^{(\nu)})_{i,j+1}-C_{i-1,i}(F_{k,n}^{(\nu)})_{i-1,j}-C_{i+1,i}(F_{k,n}^{(\nu)})_{i+1,j},$$

with the explicit matrix entries for C and V given above. Again, C and V are independent of n and ν.

Proof. Equation (4-3) translates into

$$\frac{d^2\hat{P}_n^{(\nu)}}{dx^2}(x)(1-x^2)\mathbf{1}+\frac{d\hat{P}_n^{(\nu)}}{dx}(x)\big(C-x(2\ell+2\nu+1)\mathbf{1}\big)-\hat{P}_n^{(\nu)}(x)V$$
$$=\Lambda_n(\mathcal{D}^{(\nu)})\hat{P}_n^{(\nu)}(x).$$

We take the transpose of this identity using the fact that V and $\Lambda(\mathcal{D}^{(\nu)})$ are diagonal, and therefore symmetric, and then subtract one from the other, keeping in mind that multiples of the identity $\mathbf{1}$ commute with any matrix:

$$\frac{d\hat{P}_n^{(\nu)}}{dx}(x)C-C^t\frac{d\hat{P}_n^{(\nu)}}{dx}(x)-[\hat{P}_n^{(\nu)}(x),V]=[\Lambda_n(\mathcal{D}^{(\nu)}),\hat{P}_n^{(\nu)}(x)].$$

The commutator on the right side does not depend on n in the eigenvalue $\Lambda_n(\mathcal{D}^{(\nu)})$, as all the dependence on n in $\Lambda(\mathcal{D}^{(\nu)})$ is in the multiple $-n(2\ell+2\nu+n)\mathbf{1}$ of the identity; it reduces to $-[V,\hat{P}_n^{(\nu)}(x)]$ and gives the first identity of the proposition.

For the second identity we implement Theorem 3.6 in the identity just proven and use $\frac{d}{dx}C_{n-k}^{(\nu+2\ell)}(x)=2(\nu+2\ell)C_{n-k-1}^{(\nu+2\ell+1)}(x)$ (see, e.g., [1; 2; 13; 15]) to obtain

$$\sum_{k=0}^{n\wedge 2\ell}(F_{k,n}^{(\nu)}C-C^tF_{k,n}^{(\nu)})\,2(\nu+2\ell)\,C_{n-k-1}^{(\nu+2\ell+1)}(x)=-2\sum_{k=0}^{n\wedge 2\ell}[V,F_{k,n}^{(\nu)}]\,C_{n-k}^{(\nu+2\ell)}(x).$$

The case $N=1$ of (2-3) leads to

$$\text{(4-4)}\qquad \frac{\nu+2\ell+n-k}{\nu+2\ell}C_{n-k}^{(\nu+2\ell)}(x)=C_{n-k}^{(\nu+2\ell+1)}(x)-C_{n-k-2}^{(\nu+2\ell+1)}(x)$$

and to an expansion in Gegenbauer polynomials $C_{n-k}^{(\nu+2\ell+1)}$ on the right-hand side. It remains to compare the coefficients in Gegenbauer polynomials $C_{n-k}^{(\nu+2\ell+1)}$ on both sides to deduce the second identity in the proposition. □

We perform the same procedure for the first-order matrix differential operator $\mathcal{E}^{(\nu)}$ contained in the right Fourier algebra. The operator is given by

$$\text{(4-5)}\qquad P_n^{(\nu)}\cdot\mathcal{E}^{(\nu)}=\Lambda_n(\mathcal{E}^{(\nu)})P_n^{(\nu)},\quad \Lambda_n(\mathcal{E}^{(\nu)})=A_0^{(\nu)}+nB_1,$$

where

$$\mathcal{E}^{(\nu)}=\frac{d}{dx}(xB_1+B_0)+A_0^{(\nu)},\quad 2\ell B_0=\sum_{i=0}^{2\ell-1}(2\ell-i)E_{i,i+1}-\sum_{i=1}^{2\ell}iE_{i,i-1},$$
$$\ell B_1=-\sum_{i=0}^{2\ell}(\ell-i)E_{i,i},\quad \ell A_0^{(\nu)}=\sum_{i=0}^{2\ell}\big((\ell+1)(i-2\ell)-(\nu-1)(\ell-i)\big)E_{i,i}.$$

Then the symmetric Gegenbauer polynomials satisfy

$$\hat{P}^{(\nu)}_n \cdot \mathcal{E}^{(\nu)} = D^{(\nu)}_n \Lambda_n(\mathcal{E}^{(\nu)})(D^{(\nu)}_n)^{-1}\hat{P}^{(\nu)}_n = \Lambda_n(\mathcal{E}^{(\nu)})\hat{P}^{(\nu)}_n, \tag{4-6}$$

since $D^{(\nu)}_n$ and $\Lambda_n(\mathcal{E}^{(\nu)})$ commute, being diagonal matrices.

Proposition 4.3. *The symmetric matrix-valued Gegenbauer polynomials satisfy*

$$x\left[\frac{d\hat{P}^{(\nu)}_n}{dx}(x),\, B_1\right] + \frac{d\hat{P}^{(\nu)}_n}{dx}(x)B_0 - B_0^t\frac{d\hat{P}^{(\nu)}_n}{dx}(x) = [2A^{(\nu)}_0 + nB_1,\, \hat{P}^{(\nu)}_n(x)].$$

Furthermore, the matrices $F^{(\nu)}_{k,n}$ defined in Theorem 3.6 satisfy

$$\left[\frac{F^{(\nu)}_{k-2,n}}{\nu+2\ell+n-k+2},\, (2-k+2\nu+4\ell)B_1 - 2A^{(\nu)}_0\right] \\ + \left[\frac{F^{(\nu)}_{k,n}}{\nu+2\ell+n-k},\, (2n-k)B_1 + 2A^{(\nu)}_0\right] = 2(B_0^t F^{(\nu)}_{k-1,n} - F^{(\nu)}_{k-1,n}B_0).$$

Note that the first identity of Proposition 4.3 is a matrix-valued polynomial identity of degree $n-1$, since the leading coefficient of $\hat{P}^{(\nu)}_n$ is diagonal and commutes with B_1 and $A^{(\nu)}_0$. We refrain from writing the second identity in terms of matrix coefficients.

Proof. Note that (4-6) means

$$\frac{d\hat{P}^{(\nu)}_n}{dx}(x)(xB_1+B_0) + \hat{P}^{(\nu)}_n(x)A^{(\nu)}_0 = \Lambda_n(\mathcal{E}^{(\nu)})\hat{P}^{(\nu)}_n(x);$$

we take the transpose of this identity taking into account that B_1, $A^{(\nu)}_0$ and $\Lambda(\mathcal{E}^{(\nu)})$ are diagonal, and therefore symmetric. Subtracting one from the other gives

$$x\left[\frac{d\hat{P}^{(\nu)}_n}{dx}(x),\, B_1\right] + \frac{d\hat{P}^{(\nu)}_n}{dx}(x)B_0 - B_0^t\frac{d\hat{P}^{(\nu)}_n}{dx}(x) + [\hat{P}^{(\nu)}_n(x),\, A^{(\nu)}_0] \\ = [\Lambda_n(\mathcal{E}^{(\nu)}),\, \hat{P}^{(\nu)}_n(x)].$$

It remains to collect the two commutators to obtain the first identity. Using it together with Theorem 3.6, plugging in the derivative of the scalar Gegenbauer polynomial and applying the connection formula (4-4) leads to

$$\sum_{k=0}^{n\wedge 2\ell}[F^{(\nu)}_{k,n},\, B_1]\,2xC^{(\nu+2\ell+1)}_{n-k-1}(x) + \sum_{k=0}^{n\wedge 2\ell}(F^{(\nu)}_{k,n}B_0 - B_0^tF^{(\nu)}_{k,n})\,2C^{(\nu+2\ell+1)}_{n-k-1}(x) \\ = \sum_{k=0}^{n\wedge 2\ell}\frac{[2A^{(\nu)}_0 + nB_1,\, F^{(\nu)}_{k,n}]}{\nu+2\ell+n-k}\big(C^{(\nu+2\ell+1)}_{n-k}(x) - C^{(\nu+2\ell+1)}_{n-k-2}(x)\big).$$

Finally, we use the three-term recursion (2-6) for the scalar Gegenbauer polynomials to write both sides as expansions in terms of $C_p^{(\nu+2\ell+1)}(x)$; comparing the coefficients of these expansions gives us the second identity of the proposition. □

5. Generating function for matrix-valued Gegenbauer polynomials

Several generating functions are known for the scalar Gegenbauer polynomials; they depend on a related normalization factor. The "pure" generating function is simply

$$\sum_{n=0}^{\infty} C_n^{(\lambda)}(x)\,t^n = \frac{1}{(1-2xt+t^2)^{\lambda}}. \tag{5-1}$$

In contrast, generating functions are not known for matrix-valued orthogonal polynomials, in particular, for the matrix-valued Gegenbauer polynomials that we discuss. In this short section we explain how the explicit formulas in Theorem 3.6 allow one to write down the generating function for suitably normalized matrix-valued Gegenbauer polynomials when ℓ is fixed. For this purpose consider the polynomials

$$\widetilde{P}_n^{(\nu)}(x) = \sum_{k=0}^{2\ell} \widetilde{F}_{k,n}^{(\nu)} \cdot C_{n-k}^{(\nu+2\ell)}(x),$$

where

$$(\widetilde{F}_{k,n}^{(\nu)})_{i,j} = \Gamma(\nu+n+2)(\nu+n+1)\,\gamma(\nu+n;\,i,\,j,\,k).$$

This normalization of the Gegenbauer polynomials does not affect their symmetry and is chosen in such a way that the entries of new matrices $\widetilde{F}_{k,n}^{(\nu)}$ are *polynomials* in $\nu+n$, in fact of degree $\lfloor \ell \rfloor$; indeed, the latter integer counts the maximal number of scalar Gegenbauer polynomials that show up in a linear combination of every entry of the matrix $\widetilde{P}_n^{(\nu)}(x)$. Notice that from (5-1) we have

$$\begin{aligned}\sum_{n=0}^{\infty}(\lambda+n)^j C_n^{(\lambda)}(x)\,t^n &= t^{-\lambda}\sum_{n=0}^{\infty}(\lambda+n)^j C_n^{(\lambda)}(x)\,t^{\lambda+n}\\ &= t^{-\lambda}\left(t\frac{\mathrm{d}}{\mathrm{d}t}\right)^j \sum_{n=0}^{\infty} C_n^{(\lambda)}(x)\,t^{\lambda+n}\\ &= t^{-\lambda}\left(t\frac{\mathrm{d}}{\mathrm{d}t}\right)^j \left(\frac{t}{1-2xt+t^2}\right)^{\lambda};\end{aligned}$$

for example,

$$\sum_{n=0}^{\infty}(\lambda+n)\,C_n^{(\lambda)}(x)\,t^n = \frac{\lambda(1-t^2)}{(1-2xt+t^2)^{\lambda+1}}.$$

Using these formulas for $j = 0, 1, \dots, \lfloor \ell \rfloor$ we can write explicitly the generating series

$$M(x;t) = \sum_{n=0}^{\infty} \widetilde{P}_n^{(\nu)}(x)\,t^n$$

for every particular choice of ℓ. We will limit ourselves to the illustration for $\ell = 1$, when the Gegenbauer polynomials are 3×3 matrices. In this case we obtain

$$\widetilde{F}^{(\nu)}_{0,n} = \begin{pmatrix} \lambda-1 & 0 & 0 \\ 0 & 2\lambda-4 & 0 \\ 0 & 0 & \lambda-1 \end{pmatrix}\Bigg|_{\lambda=(\nu+2)+n},$$

$$\widetilde{F}^{(\nu)}_{1,n} = \begin{pmatrix} 0 & -2\lambda & 0 \\ -2\lambda & 0 & -2\lambda \\ 0 & -2\lambda & 0 \end{pmatrix}\Bigg|_{\lambda=(\nu+2)+(n-1)},$$

$$\widetilde{F}^{(\nu)}_{2,n} = \begin{pmatrix} 0 & 0 & \lambda+1 \\ 0 & 2\lambda+4 & 0 \\ \lambda+1 & 0 & 0 \end{pmatrix}\Bigg|_{\lambda=(\nu+2)+(n-2)}.$$

Therefore, the generating function of the 3×3 Gegenbauer polynomials reads

$$\begin{aligned}
M(x;t) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \sum_{n=0}^{\infty} ((\nu+2)+n)\, C_n^{(\nu+2)}(x)\, t^n - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \sum_{n=0}^{\infty} C_n^{(\nu+2)}(x)\, t^n \\
&\quad - 2t \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \sum_{n=1}^{\infty} ((\nu+2)+(n-1))\, C_{n-1}^{(\nu+2)}(x)\, t^{n-1} \\
&\quad + t^2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \sum_{n=2}^{\infty} ((\nu+2)+(n-2))\, C_{n-2}^{(\nu+2)}(x)\, t^{n-2} \\
&\quad + t^2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 0 \end{pmatrix} \sum_{n=2}^{\infty} C_{n-2}^{(\nu+2)}(x)\, t^{n-2} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \frac{(\nu+2)(1-t^2)}{(1-2xt+t^2)^{\nu+3}} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \frac{1}{(1-2xt+t^2)^{\nu+2}} \\
&\quad - \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \frac{2(\nu+2)t(1-t^2)}{(1-2xt+t^2)^{\nu+3}} \\
&\quad + \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \frac{(\nu+2)t^2(1-t^2)}{(1-2xt+t^2)^{\nu+3}} + \begin{pmatrix} 0 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 0 \end{pmatrix} \frac{t^2}{(1-2xt+t^2)^{\nu+2}},
\end{aligned}$$

so that $M(x;t) \cdot (1-2xt+t^2)^{\nu+3}$ is a 3×3 matrix with entries from $\mathbb{Z}[\nu, x, t]$.

The computation shows that, for general $\ell > 0$, the generating function $M(x; t)$ is a $(2\ell+1) \times (2\ell+1)$ matrix multiple of the generating function (5-1) with $\lambda = \nu+2\ell+\lfloor \ell \rfloor$, whose all entries are polynomials in ν, x and t with integer coefficients.

6. Zeros of matrix-valued Gegenbauer polynomials

It is common to understand zeros of matrix-valued polynomials as those of their *determinants*, see [7] for further discussion. Such zeros serve as a natural analog of those for scalar polynomials: in the case of matrix-valued orthogonal polynomials for the measure supported on a real interval, their determinants have all zeros on the (internal part of the) interval (of multiplicities that do not exceed the size of the matrix); see [10, Theorem 1.1; 11, Corollary 4.4]. For our matrix-valued Gegenbauer polynomials, the support of the measure is the interval $[-1, 1]$, and hence all the zeros of their determinants lie in the interior of this interval (and are symmetric with respect to the origin). Theorem 3.6 provides us with access to *individual entries* of these matrix-valued polynomials, and we report on our findings in this section. At the moment, these considerations do not seem to provide any particular insight on a connection between entrywise zeros and the zeros of the determinant of the matrix-valued polynomial. They may be part of a different analytic phenomenon that shows up in the matrix-valued setting. On the other hand, there is a clear interest in zeros of suitable linear combinations of scalar orthogonal polynomials, see, e.g., Beardon and Driver [4] and Durán [9]. The linear combinations of Gegenbauer polynomials arising as entries of the matrix-valued Gegenbauer polynomials for which we study the structure of the zeros give intriguing examples, which are typically outside the classes studied in [4; 9].

For the clarity of exposition in this section, we introduce the notion of *echelon* for entries of an $(2\ell+1) \times (2\ell+1)$-matrix $(a_{ij})_{0 \le i,j \le 2\ell}$. This corresponds to a "distance of the entry a_{ij} to the boundary of the matrix", given explicitly by $\mathrm{ech}_\ell(i, j) = 1 + \min\{i, 2\ell - i, j, 2\ell - j\}$. For example, entries located in the first or last rows, or in the first or last columns are referred to as being from the "first echelon": if $i \in \{0, 2\ell\}$ or $j \in \{0, 2\ell\}$ then $\mathrm{ech}_\ell(i, j) = 1$.

The binomial factor

$$\binom{2\ell}{k}\binom{k}{\frac{1}{2}(k+i-j)}\binom{2\ell-k}{\frac{1}{2}(i+j-k)}$$

in the definition of $\gamma_\ell(\nu; i, j, k)$ dictates the presence of at most $\mathrm{ech}_\ell(i, j)$ polynomials $C_m^{(\nu+2\ell)}(x)$ in the linear combination expressing the entry $(\hat{P}_n^{(\nu)}(x))_{i,j}$. In particular, the entries from the "first echelon" (when $\mathrm{ech}_\ell(i, j) = 1$) are multiples of corresponding scalar Gegenbauer polynomials, so that their zeros are real and lie on the interval $(-1, 1)$. They even inherit the zero interlacing property when we consider the same first-echelon entry of two consecutive matrix-valued Gegenbauer polynomials.

The zeros of the entries from the "second echelon" (when $\mathrm{ech}_\ell(i, j) = 2$) are also real and belong to the measure support interval $[-1, 1]$. To see this, notice that these entries are always linear combinations of $C_m^{(\lambda)}(x)$ and $C_{m-2}^{(\lambda)}(x)$, where $\lambda = \nu + 2\ell$, with coefficients *of the same sign*. On the other hand, the two scalar Gegenbauer polynomials in such combinations can be translated into *consecutive* Jacobi polynomials[1] $J_n^{(\alpha,\beta)}(x)$ with the help of expressions

$$C_{2n}^{(\lambda)}(x) = \frac{(\lambda)_n}{\left(\frac{1}{2}\right)_n} J_n^{(\lambda-\frac{1}{2},-\frac{1}{2})}(2x^2-1) \quad \text{and} \quad C_{2n+1}^{(\lambda)}(x) = \frac{(\lambda)_{n+1}}{\left(\frac{1}{2}\right)_{n+1}} x J_n^{(\lambda-\frac{1}{2},\frac{1}{2})}(2x^2-1).$$

Finally, the zeros of two consecutive Jacobi polynomials (labeled by the same pair of parameters (α, β)) lie on the interval $(-1, 1)$ and interlace, so that the zeros of their linear combination with nonnegative coefficients lie on the same interval by the Hermite–Kakeya–Obreschkoff theorem (see, for example, [24, Theorem 6.3.8] or [22, Proposition 2.10] for a more general statement when multiple zeros and nonstrict interlacing are allowed). This justifies the location of zeros of entries $(\hat{P}_n^{(\nu)}(x))_{i,j}$ with $\mathrm{ech}_\ell(i, j) = 2$.

Remark 6.1. For the middle $(1, 1)$-entry in the 3×3 case ($\ell = 1$), Theorem 3.6 gives the following explicit expression:

$$(\hat{P}_n^{(\nu)}(x))_{1,1} = \frac{(\nu+n+2)\, n!\, \Gamma(\nu+2)}{2^{n-1}(\nu+n+1)^2\Gamma(\nu+n)} \cdot \left(\frac{C_n^{(\nu+2)}(x)}{\nu+n+2} + \frac{C_{n-2}^{(\nu+2)}(x)}{\nu+n} \right),$$

so that the polynomial is proportional to the sum $\widetilde{C}_n^{(\nu+2)}(x) + \widetilde{C}_{n-2}^{(\nu+2)}(x)$, where $\widetilde{C}_n^{(\lambda)}(x) = C_n^{(\lambda)}(x)/(\lambda+n)$. Though the argument from the last paragraph explains why the zeros of $\widetilde{C}_n^{(\lambda)}(x) + \widetilde{C}_{n-k}^{(\lambda)}(x)$ lie on the interval $-1 < x < 1$ for $k = 1$ and 2, experimentally we have observed that this is also the case for other choices of shift $k \in \mathbb{Z}$.

The first case when the "third echelon" shows up is the middle $(2, 2)$-entry of the 5×5 Gegenbauer polynomials ($\ell = 2$); we get

$$\begin{aligned}
&(\hat{P}_n^{(\nu)}(x))_{2,2} \\
&= \frac{3(\nu+n+4)\, n!\, \Gamma(\nu+4)}{2^{n-1}\Gamma(\nu+n)} \\
&\quad \times \left(\frac{C_n^{(\nu+4)}(x)}{(\nu+n+2)^2(\nu+n+3)^2(\nu+n+4)} \right. \\
&\qquad \left. + \frac{4C_{n-2}^{(\nu+4)}(x)}{(\nu+n+1)^2(\nu+n+2)(\nu+n+3)^2} + \frac{C_{n-4}^{(\nu+4)}(x)}{(\nu+n)(\nu+n+1)^2(\nu+n+2)^2} \right).
\end{aligned}$$

[1] They are usually called $P_n^{(\alpha,\beta)}(x)$ but we try to avoid a conflicting notation with our matrix-valued Gegenbauer polynomials.

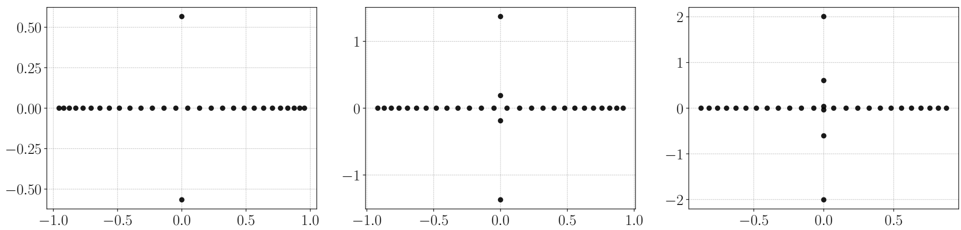

Figure 1. Zeros of the middle entry $(\hat{P}_{30}^{(3)}(x))_{\ell,\ell}$ for $\ell = 2$ (left), $\ell = 4$ (center) and $\ell = 6$ (right).

The structure of zeros of these polynomials is somewhat peculiar: There is one zero $x = 0$ for $n = 1$; then a pair of purely imaginary (conjugate) zeros for $n = 2$; then another pair and $x = 0$ for $n = 3$. Beyond this n we start witnessing real zeros (always on the interval $-1 < x < 1$) and a pair of imaginary conjugates. The real zeros interlace when passing from $n-1$ to n, while the upper imaginary zero increase (up to a certain constant strictly less than 1) with n.

Similar structures are present for "higher echelons" when more pairs of purely imaginary zeros come up, while all other zeros remain real — see Figure 1 for selected instances. So far we have not figured out reasons of this behavior. Through a complicated route, our discussions of this phenomenon led us to consider general linear combinations of scalar orthogonal polynomials and finally resulted in a separate project — the details of this side investigation can be found in [20].

We have not explored carefully the theme of zero loci of individual entries for other known families of *symmetrizable* orthogonal polynomials. But the peculiar structure of zeros seems to persist, for instance, in "randomly chosen" examples of matrix-valued Hermite polynomials from [14]; the fact that they can be normalized as symmetric is numerically supported in all these examples. At the same time, we do not expect the zero location to be a "universal" phenomenon. As has been pointed out to us by Arno Kuijlaars, the off-diagonal entries of matrix-valued orthogonal polynomials can have a very exotic distribution already in 2×2 situations; examples arising from a beautiful combinatorics of periodic hexagon tilings with period 2 are discussed in detail in [12] — the off-diagonal zeros follow a quite unexpected pattern.

7. Matrix matters and discussion

The explicit connection formulas for the matrix-valued and scalar Gegenbauer polynomials in Theorems 3.4 and 3.6 raise a question about how general this phenomenon is for matrix-valued orthogonal polynomials. They also suggest an approach to look for new families of matrix-valued polynomials by imposing the length in such expansions to depend only on the matrix size but not on the degree. Decompositions of this type, in particular Theorem 3.6, hint at the possibility to investigate asymptotics of matrix-valued orthogonal polynomials using known ones for their scalar bases. These formulas also offer numerous further directions for study and applications of matrix-valued polynomials.

The expressions in Theorem 3.6 allow one to pursue analysis of a related Padé problem for the matrix-valued generating function of the moments of the weight (2-7). The corresponding theoretical setup goes in complete parallel with the scalar version [8]. This gives room to potential applications of these matrix-valued orthogonal polynomials in number theory, to arithmetic properties of the values of the generating function at rationals — the values that can be viewed as both matrices and entrywise. This arithmetic direction seems to be under-explored at the moment.

It would be also of great interest to understand the new differential-difference structure for the matrix-valued Gegenbauer polynomials given in Section 4 more conceptually. We stress on the fact that it completely degenerates in the scalar case (for 1×1 matrices), so that it does not represent any classically familiar setting.

We are confident that the mathematics story of this note will continue in diverse — and quite remarkable! — directions.

Acknowledgements

It is our pleasure to thank Tom Koornwinder, Arno Kuijlaars, Andrei Martínez-Finkelshtein and Maarten van Pruijssen for helpful discussions of several aspects of the work. Furthermore, we thank Andrei Martínez-Finkelshtein for supplying us with the argument we use in the paragraph preceding Remark 6.1. We also thank Antonio Durán for a conversation about zero locations of entries of matrix-valued orthogonal polynomials and the referee for pointing out numerous stylistic inconsistencies in the earlier version. This work would be hardly possible without live interactions between the authors, which took occasion at several events on orthogonal polynomials (in Nijmegen, Leuven, Córdoba) and during extended visits of Koelink to Argentina and Román to the Netherlands: Erik Koelink thanks Universidad Nacional Córdoba for its hospitality and Pablo Román thanks the Radboud University for its hospitality. Part of the work of Zudilin was done during his visit in the Max-Planck Institute of Mathematics (in Bonn); Wadim Zudilin thanks the institute for excellent working conditions provided. This work was supported in part from the NWO grant OCENW.KLEIN.006.

References

[1] G. E. Andrews, R. Askey, and R. Roy, *Special functions*, Encyclopedia of Mathematics and its Applications **71**, Cambridge Univ. Press, 1999. MR

[2] R. Askey, *Orthogonal polynomials and special functions*, CBMS-NSF Regional Conference Series in Applied Mathematics **21**, SIAM, Philadelphia, PA, 1975. MR

[3] D. H. Bailey, J. M. Borwein, N. J. Calkin, R. Girgensohn, D. R. Luke, and V. H. Moll, *Experimental mathematics in action*, A K Peters, Wellesley, MA, 2007. MR

[4] A. F. Beardon and K. A. Driver, "The zeros of linear combinations of orthogonal polynomials", *J. Approx. Theory* **137**:2 (2005), 179–186. MR

[5] L. Cagliero and T. H. Koornwinder, "Explicit matrix inverses for lower triangular matrices with entries involving Jacobi polynomials", *J. Approx. Theory* **193** (2015), 20–38. MR

[6] W. R. Casper and M. Yakimov, "The matrix Bochner problem", *Amer. J. Math.* **144**:4 (2022), 1009–1065. MR

[7] D. Damanik, A. Pushnitski, and B. Simon, "The analytic theory of matrix orthogonal polynomials", *Surv. Approx. Theory* **4** (2008), 1–85. MR

[8] A. J. Durán, "Markov's theorem for orthogonal matrix polynomials", *Canad. J. Math.* **48**:6 (1996), 1180–1195. MR

[9] A. J. Durán, "Zeros of linear combinations of orthogonal polynomials", preprint, 2025. arXiv 2505.11956

[10] A. J. Durán and P. López-Rodríguez, "Orthogonal matrix polynomials: zeros and Blumenthal's theorem", *J. Approx. Theory* **84**:1 (1996), 96–118. MR

[11] A. J. Durán, P. López-Rodríguez, and E. B. Saff, "Zero asymptotic behaviour for orthogonal matrix polynomials", *J. Anal. Math.* **78** (1999), 37–60. MR

[12] A. Groot and A. B. J. Kuijlaars, "Matrix-valued orthogonal polynomials related to hexagon tilings", *J. Approx. Theory* **270** (2021), art. id. 105619, 36 pp. MR

[13] M. E. H. Ismail, *Classical and quantum orthogonal polynomials in one variable*, Encyclopedia of Mathematics and its Applications **98**, Cambridge Univ. Press, 2009. MR

[14] M. E. H. Ismail, E. Koelink, and P. Román, "Matrix valued Hermite polynomials, Burchnall formulas and non-abelian Toda lattice", *Adv. in Appl. Math.* **110** (2019), 235–269. MR

[15] R. Koekoek, P. A. Lesky, and R. F. Swarttouw, *Hypergeometric orthogonal polynomials and their q-analogues*, Springer, 2010. MR

[16] E. Koelink, M. van Pruijssen, and P. Román, "Matrix-valued orthogonal polynomials related to $(\mathrm{SU}(2)\times\mathrm{SU}(2), \mathrm{diag})$", *Int. Math. Res. Not.* **2012**:24 (2012), 5673–5730. MR

[17] E. Koelink, M. van Pruijssen, and P. Román, "Matrix-valued orthogonal polynomials related to $(\mathrm{SU}(2)\times\mathrm{SU}(2), \mathrm{diag})$, II", *Publ. Res. Inst. Math. Sci.* **49**:2 (2013), 271–312. MR

[18] E. Koelink, A. M. de los Ríos, and P. Román, "Matrix-valued Gegenbauer-type polynomials", *Constr. Approx.* **46**:3 (2017), 459–487. MR

[19] E. Koelink and P. Román, "Orthogonal vs. non-orthogonal reducibility of matrix-valued measures", *Symmetry Integrability Geom. Methods Appl.* **12** (2016), art. id. 008, 9 pp. MR

[20] E. Koelink, P. Román, and W. Zudilin, "A partial-sum deformation for a family of orthogonal polynomials", *Indag. Math.* (online publication May 2025).

[21] T. H. Koornwinder, "Matrix elements of irreducible representations of $\mathrm{SU}(2)\times\mathrm{SU}(2)$ and vector-valued orthogonal polynomials", *SIAM J. Math. Anal.* **16**:3 (1985), 602–613. MR

[22] A. Martínez-Finkelshtein, R. Morales, and D. Perales, "Real roots of hypergeometric polynomials via finite free convolution", *Int. Math. Res. Not.* **2024**:16 (2024), 11642–11687. MR

[23] I. Pacharoni and I. Zurrián, "Matrix Gegenbauer polynomials: the 2×2 fundamental cases", *Constr. Approx.* **43**:2 (2016), 253–271. MR

[24] Q. I. Rahman and G. Schmeisser, *Analytic theory of polynomials*, London Mathematical Society Monographs (N. S.) **26**, Oxford Univ. Press, 2002. MR

[25] H. S. Wilf and D. Zeilberger, "An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities", *Invent. Math.* **108**:3 (1992), 575–633. MR

[26] D. Zeilberger, "The method of creative telescoping", *J. Symbolic Comput.* **11**:3 (1991), 195–204. MR

Received December 6, 2024. Revised July 2, 2025.

ERIK KOELINK
IMAPP
RADBOUD UNIVERSITY
NIJMEGEN
NETHERLANDS
e.koelink@math.ru.nl

PABLO ROMÁN
FAMAF-CIEM
UNIVERSIDAD NACIONAL DE CÓRDOBA
CÓRDOBA
ARGENTINA
pablo.roman@unc.edu.ar

WADIM ZUDILIN
IMAPP
RADBOUD UNIVERSITY
NIJMEGEN
NETHERLANDS
w.zudilin@math.ru.nl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 338, No. 2, 2025
https://doi.org/10.2140/pjm.2025.338.349

DEFECT RELATION OF $n+1$ COMPONENTS THROUGH THE GCD METHOD

MIN RU AND JULIE TZU-YUEH WANG

We study the defect relation through the GCD method. In particular, among other results, we extend the defect relation result of Chen, Huynh, Sun and Xie (2025) to moving targets. The truncated defect relation is also studied. Furthermore, we obtain the degeneracy locus, which can be determined effectively and is independent of the maps under the consideration.

1. Motivation

Let $f:\mathbb{C}\to\mathbb{P}^n(\mathbb{C})$ be a holomorphic map. It is known that (see [10; 14; 15]) if $f(\mathbb{C})$ omits $n+2$ smooth hypersurfaces D_j, $1\le j\le n+2$, of $\mathbb{P}^n(\mathbb{C})$, where $D:=D_1+\cdots+D_{n+2}$ is located in general position, then f must be algebraically degenerate (that is, $f(\mathbb{C})$ is contained in a proper subvariety of $\mathbb{P}^n(\mathbb{C})$). In [11], J. Noguchi, J. Winkelmann, and K. Yamanoi showed that the number $n+2$ of the omitting hypersurfaces could be reduced to $n+1$ when $\deg D\ge n+2$. Their proof relies on their earlier result for holomorphic maps from $\mathbb{C}$ in the semiabelian variety $A:=(\mathbb{C}^*)^n$, which is stated as follows.

Theorem A [12]. *Let D be an effective divisor on $A:=(\mathbb{C}^*)^n$. Let $f:\mathbb{C}\to A$ be an algebraically nondegenerate holomorphic map. Then there exists a smooth compactification of A independent of f, such that, for any $\epsilon>0$,*

$$N_f(D,r)-N_f^{(1)}(D,r)\le_{\text{exc}}\epsilon T_{f,\overline{D}}(r). \tag{1-1}$$

Using the above theorem, Noguchi, Winkelmann, and Yamanoi showed that one can reduce the number of omitting divisors by one (i.e., from $n+2$ to $n+1$). Their argument is similar to ours described in Section 5. We briefly outline the argument here: Assume that $D_j=\{Q_j=0\}$ and $f(\mathbb{C})$ omits D_j for $1\le j\le n+1$. Consider a morphism $\pi:\mathbb{P}^n(\mathbb{C})\to\mathbb{P}^n(\mathbb{C})$ given by $\boldsymbol{x}\mapsto[Q_1^{a_1}(\boldsymbol{x}):\cdots:Q_{n+1}^{a_{n+1}}(\boldsymbol{x})]$, where

The research of Min Ru is supported in part by Simon Foundations grant award #531604 and #521604. Wang is supported in part by Taiwan's NSTC grant 113-2115-M-001-011-MY3.
MSC2020: primary 30D35; secondary 32H30, 32Q45.
Keywords: defect, hyper surfaces, moving targets, GCD methods.

$a_i := \operatorname{lcm}(\deg Q_1, \ldots, \deg Q_{n+1})/\deg Q_i$. Let

$$G := \det\left(\frac{\partial Q_i}{\partial x_j}\right)_{\substack{1\le i\le n+1\\ 0\le j\le n}} \in \mathbb{C}[x_0, \ldots, x_n].$$

By taking out a nonconstant irreducible factor $\tilde{G}$ of G in $\mathbb{C}[x_0, \ldots, x_n]$, one produces an additional hypersurface $\tilde{D}_{n+2} = \{\tilde{G} = 0\}$ in $\mathbb{P}^n(\mathbb{C})$. Furthermore, one can show that $D_1, \ldots, D_{n+1}, \tilde{D}_{n+2}$ are located in general position, and, by using (1-1), one can show that $N_f(\tilde{D}_{n+2}, r) \le_{\text{exc}} \epsilon T_f(r)$. Thus we can apply the second main theorem obtained by the first author [15] to get the conclusion.

In a recent manuscript by Z. Chen, D. T. Huynh, R. Sun and S. Y. Xie (see [1]), the result of Noguchi, Winkelmann, and Yamanoi mentioned above was further extended to the following defect relation.

Theorem B (Chen, Huynh, Sun and Xie [1]). *Let $\{D_i\}_{i=1}^{n+1}$ be $n+1$ hypersurfaces in $\mathbb{P}^n(\mathbb{C})$ with total degree $\sum_{i=1}^{n+1} \deg D_i \ge n+2$ satisfying one precise generic condition* (*see* (4.3) *in* [1]). *Then, for every algebraically nondegenerate entire holomorphic curve $f : \mathbb{C} \to \mathbb{P}^n(\mathbb{C})$, the following defect relation holds*:

$$\sum_{i=1}^{n+1} \delta_f(D_j) < n+1.$$

In the omitting case, we have that $Q_j(f)$ is nowhere zero for all $1 \le j \le n+1$ where $D_j = \{Q_j = 0\}$, so that one can reduce it to the semiabelian variety case $(\mathbb{C}^*)^n$ by considering

$$F := \left(\frac{Q_1(f)}{Q_{n+1}(f)}, \ldots, \frac{Q_n(f)}{Q_{n+1}(f)}\right) \in (\mathbb{C}^*)^n,$$

assuming that $\deg Q_1 = \cdots = \deg Q_{n+1}$. Hence Theorem A could be applied directly. However, in their "proof-by-contradiction" argument of the proof of Theorem B, the condition that $\sum_{i=1}^{n+1} \delta_f(D_j) = n+1$ only implies that $N_f(r, D_i) = o(T_f(r))$ (rather than $f(\mathbb{C})$ omitting D_j). To overcome this difficulty, they used the "parabolic Nevanlinna theory" developed by M. Păun and N. Sibony (see [13]), by considering the holomorphic mapping $f : Y \to \mathbb{P}^n(\mathbb{C})$ with $Y := \mathbb{C} \setminus f^{-1}(D)$, which leads to the omitting case after restricting f to Y. The key ingredient in their paper is to show that Y is an open parabolic Riemann surface with exhaustion function σ satisfying

$$\limsup_{r\to\infty} \frac{\mathfrak{X}_\sigma(r)}{T_f(r)} = 0.$$

While the method of Chen, Huynh, Sun, and Xie is very interesting and creative, it still relies on the result of Noguchi, Winkelmann, and Yamanoi (Theorem A), which greatly depends on the geometry of semiabelian varieties. For example, it is very hard to generalize the result to the moving target case.

This paper studies the defect relation through the GCD method. We don't use Theorem A. Indeed, we give and prove a variant and more general version of Theorem A by using the GCD theorem established by Aaron Levin and the second author [9]. This allows us to get a much more general defect relation (for example, the moving target case). The method was initiated by P. Corvaja and U. Zannier (see [2]), where they studied the $n = 2$ case. After Aaron Levin and the second author [9] established the general GCD theorem, it has been successfully used in a series papers by the second author and her coauthors; see [4; 5; 6; 7]. The purpose of this paper is to further use the ideas developed in [4; 5; 6; 7] to extend the defect relation, for example, to the moving target case, by using the GCD method. Furthermore, the truncated defect relation is also studied. We also pay attention on the degenerate locus. In particular, we can relax the condition that f is algebraically nondegenerate to the condition that the image of f is not contained in a subvariety Z which can be effectively predetermined and is independent of f, in the spirit of the strong Green–Griffiths–Lang conjecture.

2. Statement of the results

We use the standard notation in Nevanlinna theory (see [16] or [4; 5; 6; 7]). Let $\boldsymbol{g} = (g_0, \dots, g_n) : \mathbb{C} \to \mathbb{P}^n(\mathbb{C})$ be a holomorphic curve, where $g_0, \dots, g_n$ are entire functions without common zero. We recall that the *small function field with respect to* $\boldsymbol{g}$ is given by

$$K_{\boldsymbol{g}} := \{a : a \text{ is a meromorphic function on } \mathbb{C} \text{ with } T_a(r) = o(T_{\boldsymbol{g}}(r))\}. \tag{2-1}$$

Let K be a subfield of the field $\mathcal{M}$ of meromorphic functions. We say that *Z is a Zariski closed subset in $\mathbb{P}^n$ defined over K* if there exists a nonconstant homogeneous polynomial $F \in K[x_0, \dots, x_n]$ such that

$$Z = \{[f_0 : \cdots : f_n] \in \mathbb{P}^n(\mathcal{M}) : F(f_0, \dots, f_n) \equiv 0\}.$$

We say a holomorphic map $\boldsymbol{g} : \mathbb{C} \to \mathbb{P}^n$ is not contained in Z if $F(\boldsymbol{g})$ is not identically zero. In particular, when $K = \mathbb{C}$, the Zariski closed set is defined over $\mathbb{C}$, that is, $F \in \mathbb{C}[x_0, \dots, x_n]$, and $\boldsymbol{g}$ is not contained in Z is equivalent to $F(\boldsymbol{g}) \not\equiv 0$. For each homogeneous polynomial $G = \sum_I a_I \boldsymbol{x}^I \in K[x_0, \dots, x_n]$, where $I = (i_0, \dots, i_n) \in \mathbb{Z}_{\geq 0}^{n+1}$ and $\boldsymbol{x}^I = x_0^{i_0} \cdots x_n^{i_n}$, we define $G(z_0) := \sum_I a_I(z_0)\boldsymbol{x}^I$ if all the coefficients of G are holomorphic at z_0 and do not vanish simultaneously at z_0. Let $G_1, \dots, G_q$ be nonconstant homogeneous polynomials in $K[x_0, \dots, x_n]$. We say that they are *in weakly general position* if there exists a point $z_0 \in \mathbb{C}$ such that each $G_i(z_0)$, $1 \leq i \leq q$, can be defined as above and the union of the zero loci of $G_i(z_0)$ (as a divisor in $\mathbb{P}^n(\mathbb{C})$), $1 \leq i \leq q$, is in general position.

Theorem 1. *Let K be a subfield of the field $\mathcal{M}$ of meromorphic functions. Let F be a nonconstant homogeneous polynomial in $K[x_0, \dots, x_n]$ with no monomial factors and no repeated factors. Denote by H_i, $0 \le i \le n$, the coordinate hyperplanes of $\mathbb{P}^n(\mathbb{C})$. Then, for any $\epsilon > 0$, there exists a proper Zariski closed subset Z of $\mathbb{P}^n$ defined over K such that for any nonconstant holomorphic curve $\boldsymbol{g} : \mathbb{C} \to \mathbb{P}^n(\mathbb{C})$ with $K \subset K_{\boldsymbol{g}}$, $N_{\boldsymbol{g}}(H_i, r) = o(T_{\boldsymbol{g}}(r))$ for $0 \le i \le n$ and $\boldsymbol{g}$ not contained in Z, we have*

$$N_{\boldsymbol{g}}([F=0], r) - N_{\boldsymbol{g}}^{(1)}([F=0], r) \le_{\text{exc}} \epsilon T_{\boldsymbol{g}}(r). \tag{2-2}$$

If we assume furthermore that the hypersurface defined by F in $\mathbb{P}^n$ and the coordinate hyperplanes are in weakly general position, then

$$N_{\boldsymbol{g}}^{(1)}([F=0], r) \ge_{\text{exc}} (\deg F - \epsilon) \cdot T_{\boldsymbol{g}}(r). \tag{2-3}$$

Moreover, the exceptional set Z can be expressed as the zero locus of a finite set $\Sigma \subset K[x_0, \dots, x_n]$ with the following properties:

(Z1) *Σ depends on ϵ and F only and can be determined explicitly;*

(Z2) *the degree of each polynomial in Σ can be effectively bounded from above in terms of ϵ, n, and the degree of F.*

We apply Theorem 1 to derive the following version of the strong Green–Griffiths–Lang conjecture for moving targets.

Theorem 2. *Let K be a subfield of the field of meromorphic functions. Let F_i, $1 \le i \le n+1$, be homogeneous irreducible polynomials of positive degree in $K[x_0, \dots, x_n]$ such that $\sum_{i=1}^{n+1} \deg F_i \ge n+2$. Assume that there exists $z_0 \in \mathbb{C}$ such that all the coefficients of all F_i, $1 \le i \le n+1$, are holomorphic at z_0 and the zero locus of F_i evaluated at z_0, $1 \le i \le n+1$, intersect transversally. Then there exists a nontrivial homogeneous polynomial $B \in K[x_0, \dots, x_n]$ such that for any nonconstant holomorphic map $\boldsymbol{f} : \mathbb{C} \to \mathbb{P}^n(\mathbb{C})$ with $K \subset K_{\boldsymbol{f}}$ and $N_{F_i(\boldsymbol{f})}(0, r) = o(T_{\boldsymbol{f}}(r))$ for $1 \le i \le n+1$, we have $B(\boldsymbol{f}) \equiv 0$. Furthermore, B can be determined effectively and its degree can be effectively bounded from above in terms of n, and the degrees of F_i, $1 \le i \le n+1$.*

As a consequence, we obtain the following defect relation for moving targets.

Corollary 3 (defect relation for moving targets). *With the same notation and assumptions as in Theorem 2, let $D_i = [F_i = 0]$ for $1 \le i \le n+1$. Then for any nonconstant holomorphic map $\boldsymbol{f} : \mathbb{C} \to \mathbb{P}^n(\mathbb{C})$ with $K \subset K_{\boldsymbol{f}}$ and $B(\boldsymbol{f}) \not\equiv 0$, the following defect inequality holds:*

$$\sum_{i=1}^{n+1} \delta_{\boldsymbol{f}}(D_i) < n+1,$$

where $D_i = [F_i = 0]$. *Additionally, if* $n = 2$, *then*

$$\sum_{i=1}^{3} \delta_f^{(1)}(D_i) < 3,$$

where, for a divisor D *with* $d = \deg D$,

$$\delta_f^{(1)}(D) = 1 - \limsup_{r\to\infty} \frac{N_f^{(1)}(D, r)}{d T_f(r)}.$$

When $K = \mathbb{C}$, the following strong defect relation improves Theorem B by giving an explicit exceptional set and a truncated defect bound when $n = 2$.

Corollary 4. *Let* D_i, $1 \le i \le n+1$, *be* $n+1$ *hypersurfaces in* $\mathbb{P}^n(\mathbb{C})$, *not all being hyperplanes. Assume* D_i, $1 \le i \le n+1$, *intersect transversally. Then there exists a Zariski closed subset* Z *in* $\mathbb{P}^n(\mathbb{C})$, *which can be determined effectively and its degree can be effectively bounded from above in terms of* n, *and the degree of* D_i, *such that for any nonconstant holomorphic map* $f : \mathbb{C} \to \mathbb{P}^n(\mathbb{C})$ *whose image is not contained in* Z, *the following defect inequality holds*:

$$\sum_{i=1}^{n+1} \delta_f(D_i) < n+1.$$

Additionally, if $n = 2$, *then*

$$\sum_{i=1}^{3} \delta_f^{(1)}(D_i) < 3.$$

3. Some preliminaries and the GCD theorem

3.1. ***Preliminaries.*** We now introduce some basic notation and definitions from Nevanlinna theory, and recall fundamental results. For further details, we refer the reader to [16]. Let f be a meromorphic function, $z \in \mathbb{C}$ be a complex number, and m be a positive integer. Define the valuation functions $v_z(f) := \mathrm{ord}_z(f)$,

$$v_z^+(f) := \max\{0, v_z(f)\}, \quad \text{and} \quad v_z^-(f) := -\min\{0, v_z(f)\}.$$

Let $n_f(\infty, r)$ (respectively, $n_f^{(m)}(\infty, r)$) denote the number of poles of f in the set $\{z : |z| \le r\}$, counting multiplicity (respectively, ignoring multiplicity larger than $m \in \mathbb{N}$). The associated *counting function* and *truncated counting function* of f of order m at ∞ are

$$N_f(\infty, r) := \int_0^r \frac{n_f(\infty, t) - n_f(\infty, 0)}{t}\, dt + n_f(\infty, 0)\log r,$$

$$N_f^{(m)}(\infty, r) := \int_0^r \frac{n_f^{(m)}(\infty, t) - n_f^{(m)}(\infty, 0)}{t}\, dt + n_f^{(m)}(\infty, 0)\log r.$$

For $a \in \mathbb{C}$, the *counting function* and *truncated counting function* of f with respect to a are defined as

$$N_f(a, r) := N_{1/(f-a)}(r, \infty) \quad \text{and} \quad N_f^{(m)}(a, r) := N_{1/(f-a)}^{(m)}(\infty, r).$$

The *proximity function* $m_f(\infty, r)$ is given by

$$m_f(\infty, r) := \int_0^{2\pi} \log^+ |f(re^{i\theta})| \frac{d\theta}{2\pi},$$

where $\log^+ x = \max\{0, \log x\}$ for $x \geq 0$. The *characteristic function* is defined by

$$T_f(r) := m_f(\infty, r) + N_f(\infty, r).$$

Let $f_1, \ldots, f_n$ be meromorphic functions with $n \geq 2$. Define the local gcd multiplicity function by

$$n(f_1, \ldots, f_n, r) := \sum_{|z| \leq r} \min_{1 \leq i \leq n} \{v_z^+(f_i)\}$$

and the associated gcd counting function by

$$N_{\text{gcd}}(f_1, \ldots, f_n, r) := \int_0^r \frac{n(f_1, \ldots, f_n, t) - n(f_1, \ldots, f_n, 0)}{t} \, dt + n(f_1, \ldots, f_n, 0) \log r.$$

Let $\boldsymbol{f} : \mathbb{C} \to \mathbb{P}^n(\mathbb{C})$ be a holomorphic map and $(f_0, \ldots, f_n)$ be a reduced representation of $\boldsymbol{f}$, i.e., $f_0, \ldots, f_n$ are entire functions on $\mathbb{C}$ without common zeros. The *Nevanlinna–Cartan characteristic function* $T_{\boldsymbol{f}}(r)$ is defined by

$$T_{\boldsymbol{f}}(r) = \int_0^{2\pi} \log \max\{|f_0(re^{i\theta})|, \ldots, |f_n(re^{i\theta})|\} \frac{d\theta}{2\pi}.$$

Let $D = [F = 0]$ be a divisor in $\mathbb{P}^n(\mathbb{C})$ defined by a homogeneous polynomial $F \in \mathbb{C}[x_0, \ldots, x_n]$. The counting function with respect to D is defined by $N_{\boldsymbol{f}}(D, r) = N_{F(\boldsymbol{f})}(0, r)$.

We will make use of the following elementary inequality (see [16]).

Proposition 5. *Let $\boldsymbol{f} = (f_0, \ldots, f_n) : \mathbb{C} \to \mathbb{P}^n(\mathbb{C})$ be holomorphic curve, where $f_0, \ldots, f_n$ are entire functions without common zeros. Assume that f_0 is not identically zero. Then*

$$T_{f_j/f_0}(r) + O(1) \leq T_{\boldsymbol{f}}(r) \leq \sum_{j=1}^n T_{f_j/f_0}(r) + O(1).$$

Combining Proposition 5 with [17, Theorem 2.1], we obtain the following result.

Theorem 6 [17, Theorem 2.1]. *Let $f_0, \dots, f_n$ be entire functions with no common zeros. Assume that f_{n+1} is the holomorphic function such that*

$$f_0 + \cdots + f_n + f_{n+1} = 0.$$

If $\sum_{i \in I} f_i \neq 0$ for any proper subset $I \subset \{0, \dots, n+1\}$, then

$$T_{f_j/f_i}(r) \leq T_{\boldsymbol{f}}(r) + O(1) \leq_{\text{exc}} \sum_{i=0}^{n+1} N_{f_i}^{(n)}(0, r) + O(\log T_{\boldsymbol{f}}(r))$$

for any pair $0 \leq i, j \leq n$, where $\boldsymbol{f} := (f_0, \dots, f_n)$.

We will use the following version of the Hilbert Nullstellensatz, reformulated from [8, Chapter IX, Theorem 3.4]. See also [3, Proposition 2.1; 18, Chapter XI].

Proposition 7. *Let K be a subfield of the field of meromorphic functions. Let $\{Q_i\}_{i=1}^{n+1}$ be a set of homogeneous polynomials in $K[x_0, \dots, x_n]$ in weakly general position and with* $\deg Q_j = d_j \geq 1$. *Then there exist a positive integer s, an element $R \in K$ which is not identically zero and $P_{ji} \in K[x_0, \dots, x_n]$, $1 \leq i, j \leq n+1$, such that, for each $0 \leq j \leq n$,*

$$x_j^s \cdot R = \sum_{i=1}^{n+1} P_{ji} Q_i.$$

The following is a version of the Borel lemma for small functions. The proof can easily be obtained with some slightly modifications from [4, Lemma 3.3].

Lemma 8. *Let $f_0, \dots, f_n$ be nontrivial entire functions with no common zero and let $\boldsymbol{f} := (f_0, \dots, f_n)$. Assume that*

$$N_{f_i}^{(1)}(0, r) = o(T_{\boldsymbol{f}}(r)) \quad \text{for } 0 \leq i \leq n.$$

If $f_0, \dots, f_n$ are linearly dependent over $K_{\boldsymbol{f}}$, then for each $i \in \{0, \dots, n\}$ there exists $j \in \{0, \dots, n\}$ with $j \neq i$ such that $f_i/f_j \in K_{\boldsymbol{f}}$.

3.2. *The GCD theorem.*

Theorem 9 (the GCD theorem). *Let $g_0, g_1, \dots, g_n$ be entire functions without common zeros and let $\boldsymbol{g} = [g_0 : g_1 : \cdots : g_n]$. Let $F, G \in K_{\boldsymbol{g}}[x_0, \dots, x_n]$ be nonconstant coprime homogeneous polynomials. Assume that one of the following holds*:

(a) $N_{g_i}(0, r) = o(T_{\boldsymbol{g}}(r))$ *for* $0 \leq i \leq n$;

(b) $N_{g_i}^{(1)}(0, r) = o(T_{\boldsymbol{g}}(r))$ *for $0 \leq i \leq n$ and one of the hypersurfaces defined by $G = 0$ or $F = 0$ in $\mathbb{P}^n(K)$ is in weakly general position with the $n+1$ coordinate hyperplanes.*

Then, for any $\epsilon > 0$, there exists a positive integer m independent of $\mathbf{g}$ such that we have either

$$N_{\text{gcd}}(F(g_0,\dots,g_n), G(g_0,\dots,g_n), r) \le_{\text{exc}} \epsilon T_{\mathbf{g}}(r), \tag{3-1}$$

or

$$T_{(g_1/g_0)^{m_1}\cdots(g_n/g_0)^{m_n}}(r) = o(T_{\mathbf{g}}(r)) \tag{3-2}$$

for some nontrivial tuple of integers $(m_1,\dots,m_n)$ with $|m_1|+\cdots+|m_n| \le 2m$.

For the convenience of later application, we state the following result for $n = 1$.

Proposition 10. *Let g_0, g_1 be entire functions without common zeros and let $\mathbf{g} = (g_0, g_1)$. Assume that $\mathbf{g}$ is not constant. Let $F, G \in K_{\mathbf{g}}[x_0, x_1]$ be nonconstant coprime homogeneous polynomials. Then*

$$N_{\text{gcd}}(F(g_0,g_1), G(g_0,g_1), r) \le o(T_{\mathbf{g}}(r)). \tag{3-3}$$

Proof. Since F and G are coprime homogeneous polynomials in $K_{\mathbf{g}}[x_0, x_1]$, we may apply Proposition 7 to find an integer s, $R \in K_{\mathbf{g}} \setminus \{0\}$ and $H_i \in K_{\mathbf{g}}[x_0, x_1]$, $1 \le i \le 4$, such that

$$x_0^s \cdot R = H_1F + H_2G \quad \text{and} \quad x_1^s \cdot R = H_3F + H_4G. \tag{3-4}$$

Here, we may assume that H_i, $1 \le i \le 4$, are homogeneous polynomials with degree equal to $s - \deg F$. By evaluating (3-4) at (g_0, g_1), we have

$$\begin{aligned} g_0^s \cdot R &= H_1(g_0,g_1)F(g_0,g_1) + H_2(g_0,g_1)G(g_0,g_1),\\ g_1^s \cdot R &= H_3(g_0,g_1)F(g_0,g_1) + H_4(g_0,g_1)G(g_0,g_1). \end{aligned} \tag{3-5}$$

Since g_0 and g_1 have no common zeros, we observe that

$$\min\{v_z^+(F(g_0,g_1)), v_z^+(G(g_0,g_1))\} \le v_z^+(R) + \sum_{\alpha\in I} v_z^-(\alpha) \tag{3-6}$$

for each $z \in \mathbb{C}$. Here I is the set of nontrivial coefficients of H_i, $1 \le i \le 4$. Hence,

$$N_{\text{gcd}}(F(g_0,g_1), G(g_0,g_1), r) \le N_R(0,r) + \sum_{\alpha\in I} N_\alpha(\infty, r) \le o(T_{\mathbf{g}}(r)), \tag{3-7}$$

as R and the coefficients of F_i are in $K_{\mathbf{g}}$. □

To prove Theorem 9, we use the following fundamental result by Levin and the second author for $n \ge 2$.

Theorem 11 [9, Theorem 5.7]. *Let $g_0, g_1, \dots, g_n$ be entire functions without common zeros with $n \ge 2$ and let $\mathbf{g} = [g_0 : g_1 : \cdots : g_n]$. Let $F, G \in K_{\mathbf{g}}[x_0, x_1, \dots, x_n]$ be coprime homogeneous polynomials of the same degree $d > 0$. Let I be the set of exponents $\mathbf{i}$ such that $\mathbf{x}^{\mathbf{i}}$ appears with a nonzero coefficient in either F or G. Let*

$m \ge d$ be a positive integer. Suppose that $\{g_0^{i_0}\cdots g_n^{i_n} : i_0+\cdots+i_n = m\}$ is linearly independent over $K_{\boldsymbol{g}}$. Then, for any $\epsilon > 0$, there is a positive integer L such that

$$
\begin{aligned}
\text{(3-8)}\quad & MN_{\mathrm{gcd}}(F(\boldsymbol{g}), G(\boldsymbol{g}), r) \\
&\le_{\mathrm{exc}} c_{m,n,d}\sum_{i=1}^{n} N_{g_i}^{(L)}(0,r) + \left(\frac{m}{n+1}\binom{m+n}{n} - c_{m,n,d} - M'm\right)\sum_{i=1}^{n} N_{g_i}(0,r) \\
&\quad + \binom{m+n-2d}{n} N_{\mathrm{gcd}}(\{\boldsymbol{g}^{\boldsymbol{i}}\}_{\boldsymbol{i}\in I}, r) + \left(M'mn + \epsilon m + \frac{M\epsilon}{2}\right) T_{\boldsymbol{g}}(r) + o(T_{\boldsymbol{g}}(r)),
\end{aligned}
$$

where $c_{m,n,d} = 2\binom{m+n-d}{n+1} - \binom{m+n-2d}{n+1}$, $M = 2\binom{m+n-d}{n} - \binom{m+n-2d}{n}$, and M' is an integer of order $O(m^{n-2})$, where $\le_{\mathrm{exc}}$ means the inequality holds for all $r \in (0, \infty)$ except for a set E of finite measure.

We note that $M' := \dim K_{\boldsymbol{g}}[x_0, \ldots, x_n]_m/(F, G)_m \le d^2\binom{m+n-2}{n-2}$.

Proof of Theorem 9. Without loss of generality, we assume that $\deg F = \deg G$. We first prove when $n \ge 2$. Let $\epsilon > 0$. To establish (3-1) or (3-2), we can assume that ϵ is sufficiently small. We can choose a real $C_1 \ge 1$ independent of ϵ and $\boldsymbol{g}$ such that $m = C_1\epsilon^{-1} \ge 2d$,

$$
\text{(3-9)}\quad \frac{M'mn}{M} \le \frac{\epsilon}{4}, \quad \text{and} \quad \frac{1}{M}\left(\frac{m}{n+1}\binom{m+n}{n} - c_{m,n,d} - M'm\right) \le \frac{\epsilon}{4(n+1)}.
$$

We may assume that each g_i is not identically zero; otherwise, (3-2) holds trivially. Suppose that the set $\{g_0^{i_0}\cdots g_n^{i_n} : i_0+\cdots+i_n = m\}$ is linearly independent over $K_{\boldsymbol{g}}$. We aim at concluding (3-1) under assumption (a) or (b). Suppose (a) holds, i.e., $N_{g_i}(0,r) = o(T_{\boldsymbol{g}}(r))$ for $0 \le i \le n$. Then (3-8) implies that

$$
\text{(3-10)}\quad N_{\mathrm{gcd}}(F(\boldsymbol{g}), G(\boldsymbol{g}), r) \le_{\mathrm{exc}} \left(\frac{M'mn}{M} + \epsilon\frac{m}{M} + \frac{\epsilon}{2}\right) T_{\boldsymbol{g}}(r) + o(T_{\boldsymbol{g}}(r)) < \epsilon T_{\boldsymbol{g}}(r).
$$

If (b) holds, then

$$
N_{g_i}^{(L)}(0,r) \le L N_{g_i}^{(1)}(0,r) = o(T_{\boldsymbol{g}}(r))
$$

for $0 \le i \le n$. The assumption that one of $[G = 0]$ or $[F = 0]$ is in weakly general position with the $n+1$ coordinate hyperplanes in $\mathbb{P}^n$ implies that the set $\{(d, 0, \ldots, 0), \ldots, (0, \ldots, 0, d)\}$ is a subset of I. Since $g_0, \ldots, g_n$ are entire function with no common zero, we have

$$
N_{\mathrm{gcd}}(\{\boldsymbol{g}^{\boldsymbol{i}}\}_{\boldsymbol{i}\in I}, r) = 0
$$

when (b) holds. Then by (3-8), (3-9) and that $N_{g_i}(0,r) \le T_{\boldsymbol{g}}(r)$, we obtain (3-1).

Finally, if the set $\{g_0^{i_0}\cdots g_n^{i_n} : i_0+\cdots+i_n = m\}$ is dependent over $K_{\boldsymbol{g}}$, then we may apply Lemma 8 to derive that there exists a nontrivial n-tuple of integers $(j_1, \ldots, j_n)$ with $|j_1|+\cdots+|j_n| \le 2m$ such that

$$
T_{(g_1/g_0)^{j_1}\cdots(g_n/g_0)^{j_n}}(r) = o(T_{\boldsymbol{g}}(r)). \qquad \square
$$

4. Proof of Theorem 1

4.1. *Some lemmas.* We recall some lemmas from [7].

Lemma 12. *Let $n \geq 2$ and let $(m_1, \dots, m_n)$ be a nonzero vector in $\mathbb{Z}^n$ with* $\gcd(m_1, \dots, m_n) = 1$. *Then there exist $\boldsymbol{v}_i = (v_{i,1}, \dots, v_{i,n}) \in \mathbb{Z}^n$ for $1 \leq i \leq n-1$ such that*

$$|v_{i,j}| \leq \max\{|m_j|, 1\} \quad \textit{for } 1 \leq j \leq n$$

and $(m_1, \dots, m_n)$ together with the $\boldsymbol{v}_i$'s form a basis of $\mathbb{Z}^n$.

Let k be a field and let q and r be positive integers. We write $\boldsymbol{t} := (t_1, \dots, t_q)$ and $\boldsymbol{x} := (x_1, \dots, x_r)$. For $\boldsymbol{i} = (i_1, \dots, i_r) \in \mathbb{Z}^r$, define $\boldsymbol{x}^{\boldsymbol{i}} = x_1^{i_1} \cdots x_r^{i_r}$ and $\boldsymbol{t}^{\boldsymbol{i}} = t_1^{i_1} \cdots t_r^{i_r}$. For $\Sigma \subseteq k[\boldsymbol{t}]$, let $\mathcal{Z}(\Sigma) = \{\lambda \in k^q : f(\lambda) = 0 \text{ for every } f \in \Sigma.\}$.

Lemma 13. *Assume that k is infinite. Let $f(\boldsymbol{t}, \boldsymbol{x}) \in k[\boldsymbol{t}, \boldsymbol{x}]$ be a polynomial with no monomial factor and no repeated irreducible factor in $k[\boldsymbol{t}, \boldsymbol{x}]$. Then there exists an effectively computable nonempty finite set $\Sigma \subset k[\boldsymbol{t}] \setminus \{0\}$ such that for every $\lambda \in k^q \setminus \mathcal{Z}(\Sigma)$, the polynomial $f(\lambda, \boldsymbol{x})$ has no monomial or repeated irreducible factor. Moreover, the cardinality of Σ and the degree of each polynomial in Σ can be bounded effectively in terms of q, r, and the degree of f. Furthermore, if $f(\boldsymbol{t}, \boldsymbol{x}) \in k_0[\boldsymbol{t}, \boldsymbol{x}]$ for k_0 being a subfield of k, then Σ is defined over k_0.*

4.2. *Preliminary theorem.* Let $\boldsymbol{g} = (g_0, \dots, g_n)$, where $g_i \not\equiv 0$, $0 \leq i \leq n$, are entire functions without common zeros. Let $u_i = g_i/g_0$, for $1 \leq i \leq n$. We observe that

$$\max_{1 \leq j \leq n} \{T_{u_j}(r)\} \leq T_{\boldsymbol{g}}(r) \leq n \max_{1 \leq j \leq n} \{T_{u_j}(r)\}, \tag{4-1}$$

and

$$N_{u_i}(0, r) + N_{u_i}(\infty, r) \leq N_{g_i}(0, r) + N_{g_0}(0, r) \tag{4-2}$$

for each $1 \leq i \leq n$.

Recall that

$$K_{\boldsymbol{g}} := \{a : a \text{ is a meromorphic function on } \mathbb{C} \text{ with } T_a(r) \leq o(T_{\boldsymbol{g}}(r))\},$$

which is the field of meromorphic functions of slow growth with respect to $\boldsymbol{g}$. We note that $a' \in K_{\boldsymbol{g}}$ if $a \in K_{\boldsymbol{g}}$. Furthermore, $u_i'/u_i \in K_{\boldsymbol{g}}$ if

$$N_{u_i}^{(1)}(0, r) + N_{u_i}^{(1)}(\infty, r) \leq o\Big(\max_{1 \leq j \leq n} \{T_{u_j}(r)\}\Big).$$

Let $\boldsymbol{x} := (x_1, \dots, x_n)$ and $\boldsymbol{u} = (u_1, \dots, u_n)$. For $\boldsymbol{i} = (i_1, \dots, i_n) \in \mathbb{Z}^n$, we let $\boldsymbol{x}^{\boldsymbol{i}} := x_1^{i_1} \cdots x_n^{i_n}$ and $\boldsymbol{u}^{\boldsymbol{i}} := u_1^{i_1} \cdots u_n^{i_n}$. For a nonconstant polynomial

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{i}} a_{\boldsymbol{i}} \boldsymbol{x}^{\boldsymbol{i}} \in K_{\boldsymbol{g}}[\boldsymbol{x}] := K_{\boldsymbol{g}}[x_1, \dots, x_n],$$

we define

$$D_{\boldsymbol{u}}(F)(\boldsymbol{x}) := \sum_{\boldsymbol{i}} \frac{(a_{\boldsymbol{i}}\boldsymbol{u}^{\boldsymbol{i}})'}{\boldsymbol{u}^{\boldsymbol{i}}}\boldsymbol{x}^{\boldsymbol{i}} = \sum_{\boldsymbol{i}} \left(a_{\boldsymbol{i}}' + a_{\boldsymbol{i}} \cdot \sum_{j=1}^{n} i_j \frac{u_j'}{u_j} \right) \boldsymbol{x}^{\boldsymbol{i}} \in K_{\boldsymbol{g}}[\boldsymbol{x}]. \tag{4-3}$$

A direct computation shows that

$$F(\boldsymbol{u})' = D_{\boldsymbol{u}}(F)(\boldsymbol{u}), \tag{4-4}$$

and that the product rule

$$D_{\boldsymbol{u}}(FG) = D_{\boldsymbol{u}}(F)G + FD_{\boldsymbol{u}}(G) \tag{4-5}$$

holds for $F, G \in K_{\boldsymbol{g}}[\boldsymbol{x}]$.

Lemma 14 [5, Lemma 3.1]. *Let F be a nonconstant polynomial in $K_{\boldsymbol{g}}[\boldsymbol{x}]$ with no monomial factors and no repeated factors. Assume that*

$$N_{u_i}^{(1)}(0, r) + N_{u_i}^{(1)}(\infty, r) = o\big(\max_{1 \le j \le n} \{T_{u_j}(r)\}\big)$$

for each $1 \le i \le n$. Then F and $D_{\boldsymbol{u}}(F)$ are coprime in $K_{\boldsymbol{g}}[\boldsymbol{x}]$ unless there exists a nontrivial tuple of integers $(m_1, \dots, m_n)$ with $\sum_{i=1}^{n} |m_i| \le 2 \deg F$ such that $T_{u_1^{m_1} \cdots u_n^{m_n}}(r) = o\big(\max_{1 \le j \le n}\{T_{u_j}(r)\}\big)$.

We now state a preliminary theorem in affine form.

Theorem 15. *Let K be a subfield of the field of meromorphic functions. Let G be a nonconstant polynomial in $K[x_1, \dots, x_n]$ with no monomial factors and no repeated factors. Assume one of the following holds*:

(a) $N_{u_i}(0, r) + N_{u_i}(\infty, r) = o\big(\max_{1 \le j \le n}\{T_{u_j}(r)\}\big)$ *for each $1 \le i \le n$, or*

(b) $N_{u_i}^{(1)}(0, r) + N_{u_i}^{(1)}(\infty, r) = o\big(\max_{1 \le j \le n}\{T_{u_j}(r)\}\big)$ *for each $1 \le i \le n$, and that $[G = 0]$ and the $n+1$ coordinate hyperplanes are in weakly general position in $\mathbb{P}^n$.*

For any $\epsilon > 0$, there exists a positive integer m such that for any n-tuple of meromorphic functions $\boldsymbol{u} = (u_1, \dots, u_n)$ satisfying $K \subset K_{\boldsymbol{g}}$, where $\boldsymbol{g} = [1 : u_1 : \cdots : u_n]$, we have either

$$T_{{u_1}^{m_1} \cdots {u_n}^{m_n}}(r) = o\big(\max_{1 \le j \le n} \{T_{u_j}(r)\}\big) \tag{4-6}$$

for a nontrivial n-tuple $(m_1, \dots, m_n)$ of integers with $\sum_{i=0}^{n} |m_i| \le m$, or

$$N_{G(\boldsymbol{u})}(0, r) - N_{G(\boldsymbol{u})}^{(1)}(0, r) \le_{\text{exc}} \epsilon \max_{1 \le j \le n} \{T_{u_j}(r)\}. \tag{4-7}$$

Proof. Let $z_0 \in \mathbb{C}$. If $v_{z_0}(G(\boldsymbol{u})) \ge 2$, then it follows from (4-4) that $v_{z_0}(D_{\boldsymbol{u}}(G)(\boldsymbol{u})) = v_{z_0}(G(\boldsymbol{u})) - 1$. Hence,

$$\min\{v_{z_0}^+(G(\boldsymbol{u})), v_{z_0}^+(D_{\boldsymbol{u}}(G)(\boldsymbol{u}))\} \ge v_{z_0}^+(G(\boldsymbol{u})) - \min\{1, v_{z_0}^+(G(\boldsymbol{u}))\}.$$

Consequently,

$$N_{\mathrm{gcd}}(G(\boldsymbol{u}), D_{\boldsymbol{u}}(G)(\boldsymbol{u}), r) \geq N_{G(\boldsymbol{u})}(0, r) - N^{(1)}_{G(\boldsymbol{u})}(0, r). \tag{4-8}$$

By Lemma 14, G and $D_{\boldsymbol{u}}(G)$ are either coprime or (4-6) holds for $m = 2\deg G$. Therefore, we assume that G and $D_{\boldsymbol{u}}(G)$ are coprime. By Theorem 9, we find a positive integer m depending only on ϵ, n and $\deg G$ such that either (4-6) holds or

$$N_{\mathrm{gcd}}(G(\boldsymbol{u}), D_{\boldsymbol{u}}(G)(\boldsymbol{u}), r) \leq_{\mathrm{exc}} \epsilon \max_{1\leq j\leq n}\{T_{u_j}(r)\}. \tag{4-9}$$

Together with (4-8), we obtain (4-7). □

4.3. *Further refinement.* We will prove the following theorem by finding an exceptional set in Theorem 15.

Theorem 16. *Let K be a subfield of the field of meromorphic functions. Let G be a nonconstant polynomial in $K[x_1, \dots, x_n]$ with no monomial factors and no repeated factors. For any $\epsilon > 0$, there exists a nonconstant polynomial H in $K[x_1, \dots, x_n]$ such that for any n-tuple of meromorphic functions $\boldsymbol{u} = (u_1, \dots, u_n)$ satisfying*

$$N_{u_i}(0, r) + N_{u_i}(\infty, r) = o\Big(\max_{1\leq j\leq n}\{T_{u_j}(r)\}\Big) \quad \text{for each } 1 \leq i \leq n, \tag{4-10}$$

and $K \subset K_{\boldsymbol{g}}$, where $\boldsymbol{g} = [1 : u_1 : \cdots : u_n]$, we have either $H(\boldsymbol{u}) \equiv 0$ or

$$N_{G(\boldsymbol{u})}(0, r) - N^{(1)}_{G(\boldsymbol{u})}(0, r) \leq_{\mathrm{exc}} \epsilon \max_{1\leq j\leq n}\{T_{u_j}(r)\}. \tag{4-11}$$

Moreover, H can be determined effectively and the degree of H can be bounded effectively in terms of ϵ, n and the degree of G.

Remark. The effectiveness of determining H follows from the application of Lemma 13 in the induction process. Moreover, the estimate for the degree of H depends on the fact that the constant m in Theorem 9 can be determined effectively, as well as on the proof of Theorem 9 and the content of Lemma 12. While a rough bound for the degree of H can, in principle, be obtained by tracing these dependencies, carrying this out would involve substantial additional technical detail beyond the scope of the present work.

Proof. The proof of [7, Theorem 4] can be adapted to suit the current situation. We will closely adhere to their arguments and notation. We first fix some notation:

(i) For a matrix $A = (a_{ij})$ with complex-valued entries, let

$$\|A\|_\infty = \max_i \sum_j |a_{ij}|$$

be the maximum of the absolute row sums.

(ii) We say that a nontrivial meromorphic function β *has small zeros and poles with respect to $\boldsymbol{g}$* if $N_\beta(0, r) + N_\beta(\infty, r) = o(T_{\boldsymbol{g}}(r))$.

Let $G \in K[x_1, \dots, x_n] \setminus K$ with no monomial factors and no repeated factors. Let $\epsilon > 0$. In the following we consider a n-tuple of meromorphic functions $\boldsymbol{u} = (u_1, \dots, u_n)$ satisfying

$$N_{u_i}(0, r) + N_{u_i}(\infty, r) = o\big(\max_{1\le j\le n} \{T_{u_j}(r)\}\big) = o(T_{\boldsymbol{g}}(r))$$

for each $1 \le i \le n$, and $K \subset K_{\boldsymbol{g}}$, where $\boldsymbol{g} = [1 : u_1 : \cdots : u_n]$. We note that $\lambda \in K_{\boldsymbol{g}}$ if and only if $T_\lambda(r) = o\big(\max_{1\le j\le n}\{T_{u_j}(r)\}\big)$ by (4-1).

When $n = 1$, the theorem is a direct consequence of Theorem 15 since u_1 is constant if (4-6) holds.

From this point, we let $n \ge 2$. We will effectively construct a nonconstant polynomial H in $K[x_1, \dots, x_n]$ such that (4-11) holds if $H(u_1, \dots, u_n) \not\equiv 0$.

The arguments are carried out inductively in several steps. In the following, the $c_{i,j}$'s and M_i's denote positive real numbers depending only on ϵ, n, $\deg G$, and the previously defined $c_{i',j'}$ and $M_{i'}$.

Step 1: We apply Theorem 15. The condition (a) in Theorem 15 holds under our assumption, so if (4-7) holds then we are done. Otherwise, there exists an n-tuple of integers $(m_1, \dots, m_n) \ne (0, \dots, 0)$ with $\sum |m_i| \le M_1$ such that

$$\lambda_1 := u_1^{m_1} \cdots u_n^{m_n} \in K_{\boldsymbol{g}}. \tag{4-12}$$

We may assume $\gcd(m_1, \dots, m_n) = 1$. By Lemma 12, $(m_1, \dots, m_n)$ extends to a basis $(m_1, \dots, m_n)$, $(a_{21}, \dots, a_{2n}), \dots, (a_{n1}, \dots, a_{nn})$ of $\mathbb{Z}^n$ such that

$$|a_{i1}| + \cdots + |a_{in}| \le M_1 + n \quad \text{for } 2 \le i \le n. \tag{4-13}$$

Consider the change of variables

$$\Lambda_1 := x_1^{m_1} \cdots x_n^{m_n} \qquad \text{and} \qquad X_{1,i} := x_1^{a_{i1}} \cdots x_n^{a_{in}} \quad \text{for } 2 \le i \le n \tag{4-14}$$

and put

$$\beta_{1,i} = u_1^{a_{i1}} \cdots u_n^{a_{in}} \quad \text{for } 2 \le i \le n. \tag{4-15}$$

Let A_1 denote the $n \times n$ matrix whose rows are the above basis of $\mathbb{Z}^n$. Then we formally express the above identities as

$$(\Lambda_1, X_{1,2}, \dots, X_{1,n}) = (x_1, \dots, x_n)^{A_1}, \quad (\lambda_1, \beta_{1,2}, \dots, \beta_{1,n}) = (u_1, \dots, u_n)^{A_1}. \tag{4-16}$$

Let $B_1 = A_1^{-1}$. The entries of B_1 can be bounded from above in terms of M_1 and n. We have

$$(x_1, \dots, x_n) = (\Lambda_1, X_{1,2}, \dots, X_{1,n})^{B_1}, \quad (u_1, \dots, u_n) = (\lambda_1, \beta_{1,2}, \dots, \beta_{1,n})^{B_1}. \tag{4-17}$$

Let $G_1(\Lambda_1, X_{1,2}, \dots, X_{1,n}) \in K[\Lambda_1, X_{1,2}, \dots, X_{1,n}]$ with no monomial factors and

$$G((\Lambda_1, X_{1,2}, \dots, X_{1,n})^{B_1}) = \Lambda_1^{d_1} X_{1,2}^{d_2} \cdots X_{1,n}^{d_n} G_1(\Lambda_1, X_{1,2}, \dots, X_{1,n}) \tag{4-18}$$

for some integers d_i, $1 \le i \le n$. Since the transformations in (4-16) and (4-17) are invertible of each other and G has no repeated irreducible factors, we have that G_1 has no repeated irreducible factors either. The coefficients of G_1 are the same as the coefficients of G and $\deg G_1$ can be bounded from above explicitly in terms of M_1, n, and $\deg G$. Consider $G_1(\lambda_1, X_{1,2}, \dots, X_{1,n}) \in K(\lambda_1)[X_{1,2}, \dots, X_{1,n}]$; by using (4-12) we have

$$K(\lambda_1) \subset K_{\boldsymbol{g}}. \tag{4-19}$$

For the particular change of variables in (4-16), (4-17), and (4-18) (that depends on the matrix A_1), we apply the Lemma 13 with k being the field of meromorphic functions $\mathcal{M}$ and $k_0 = K$ and (4-14) to find a nonconstant polynomial $H_1' \in K[x_1, \dots, x_n]$ such that $G_1(\lambda_1, X_{1,2}, \dots, X_{1,n})$ has neither monomial nor repeated irreducible factors if $H_1'(u_1, \dots, u_n) \not\equiv 0$. We now take H_1 to be the product of all such H_1' where A_1 ranges over the finitely many elements of $\mathrm{GL}_n(\mathbb{Z})$ with $\|A_1\|_\infty \le M_1 + n$. From Lemma 13, $\deg H_1$ depends only on ϵ, n and $\deg G$.

Since the u_i's, λ_1, and $\beta_{1,j}$'s have small zero and pole with respect to $\boldsymbol{g}$, we have

$$\begin{aligned} N_{G(\boldsymbol{u})}(0, r) - N^{(1)}_{G(\boldsymbol{u})}(0, r) \\ = N_{G_1(\lambda_1, \beta_{1,2}, \dots, \beta_{1,n})}(0, r) - N^{(1)}_{G_1(\lambda_1, \beta_{1,2}, \dots, \beta_{1,n})}(0, r) + o(T_{\boldsymbol{g}}(r)) \end{aligned} \tag{4-20}$$

by (4-16) and (4-18). From (4-16), (4-17) and (4-12), we have

$$\max_{1 \le i \le n} \{T_{u_i}(r)\} = O\big(\max\{T_{\lambda_1}(r), T_{\beta_{1,2}}(r), \dots, T_{\beta_{1,n}}(r)\}\big) = O\big(\max_{2 \le i \le n} \{T_{\beta_{1,i}}(r)\}\big). \tag{4-21}$$

In conclusion, at the end of this step we have

$$\max_{2 \le i \le n} \{T_{\beta_{1,i}}(r)\} = O\big(\max_{1 \le i \le n} \{T_{u_i}(r)\}\big). \tag{4-22}$$

Furthermore, it remains to consider the case when

$$N_{G_1(\lambda_1, \beta_{1,2}, \dots, \beta_{1,n})}(0, r) - N^{(1)}_{G_1(\lambda_1, \beta_{1,2}, \dots, \beta_{1,n})}(0, r) <_{\text{exc}} \epsilon \max_{1 \le i \le n} \{T_{u_i}(r)\} \tag{4-23}$$

fails to hold under the assumption that $H_1(u_1, \dots, u_n) \not\equiv 0$.

There are $n-1$ many steps in total. Hence if $n \ge 3$, we proceed with the following $n-2$ many more steps.

Step 2: We include this step in order to illustrate the transition from Step $s-1$ to Step s below. Since the various estimates and constructions are similar to those in Step 1, we skip some of the details. Suppose $H_1(u_1, \dots, u_n) \not\equiv 0$ so that $G_1(\lambda_1, X_{1,2}, \dots, X_{1,n})$ has neither monomial nor repeated factors.

We apply Theorem 15, assuming (4-23) fails to hold for $G_1(\lambda_1, X_{1,2}, \dots, X_{1,n})$ and $(\beta_{1,2}, \dots, \beta_{1,n})$, and use (4-19), (4-22), to get an $(n-1)$-tuple $(m'_2, \dots, m'_n) \neq (0, \dots, 0)$ with $\sum |m'_i| \leq M_2$ such that

$$\lambda_2 := \beta_{1,2}^{m'_2} \cdots \beta_{1,n}^{m'_n} \in K_{\boldsymbol{g}}. \tag{4-24}$$

We may assume $\gcd(m'_2, \dots, m'_n) = 1$. By Lemma 12, $(m'_2, \dots, m'_n)$ extends to a basis of $\mathbb{Z}^{n-1}$ in which each vector has ℓ_1-norm at most $M_2 + n$.

Let A'_2 be the $(n-1)\times(n-1)$ matrix whose rows are the above basis of $\mathbb{Z}^{n-1}$. We make the transformation

$$(\Lambda_2, X_{2,3}, \dots, X_{2,n}) = (X_{1,2}, \dots, X_{1,n})^{A'_2}, \quad (\lambda_2, \beta_{2,3}, \dots, \beta_{2,n}) = (\beta_{1,2}, \dots, \beta_{1,n})^{A'_2}.$$

Let $A_2 = (1) \oplus A'_2$ be the $n\times n$ block diagonal matrix with the $(1, 1)$-entry 1 and the matrix A'_2 in the remaining $(n-1)\times(n-1)$ block. We have

$$(\Lambda_1, \Lambda_2, \dots, X_{2,n}) = (\Lambda_1, X_{1,2}, \dots, X_{1,n})^{A_2},$$
$$(\lambda_1, \lambda_2, \dots, \beta_{2,n}) = (\lambda_1, \beta_{1,2}, \dots, \beta_{1,n})^{A_2}.$$

Combining this with (4-16), we have

$$\begin{aligned}(\Lambda_1, \Lambda_2, X_{2,3}, \dots, X_{2,n}) &= (x_1, \dots, x_n)^{A_2 A_1},\\ (\lambda_1, \lambda_2, \beta_{2,3}, \dots, \beta_{2,n}) &= (u_1, \dots, u_n)^{A_2 A_1}.\end{aligned} \tag{4-25}$$

Let $B_2 = (A_2 A_1)^{-1}$. Let $G_2(\Lambda_1, \Lambda_2, X_{2,3}, \dots, X_{2,n})$ be the polynomial with no monomial factors such that

$$G_0((\Lambda_1, \Lambda_2, X_{2,3}, \dots, X_{2,n})^{B_2}) = \Lambda_1^{d'_1} \Lambda_2^{d'_2} X_{2,3}^{d'_3} \cdots X_{2,n}^{d'_n} G_2(\Lambda_1, \Lambda_2, X_{2,3}, \dots, X_{2,n})$$

for some $d'_1, \dots, d'_n \in \mathbb{Z}$. We have that $\deg G_2$ can be bounded from above explicitly in terms of M_2, M_1, n, and $\deg G$. As before, we regard $G_2(\lambda_1, \lambda_2, X_{2,3}, \dots, X_{2,n})$ as a polynomial in $X_{2,3}, \dots, X_{2,n}$ with coefficients in $K_{\boldsymbol{g}}$ using (4-12) and (4-24).

For a particular A_1 and A_2, we apply Lemma 13 with $k = \mathcal{M}$ and $k_0 = K$ and use (4-12) and (4-24) to get a nonconstant polynomial H'_2 in $K[x_1, \dots, x_n]$ such that $G_2(\lambda_1, \lambda_2, X_{2,3}, \dots, X_{2,n})$ has neither monomial nor repeated factors. We now take H_2 to be the product of all such H'_2 where A_1 and A_2 range over the finitely many unimodular matrices with $\|A_1\|_\infty \leq M_1 + n$ and $\|A_2\|_\infty \leq M_2 + n$. By using similar estimates, at the end of this step, we have

$$\max_{3\leq i\leq n} \{T_{\beta_{1,i}}(r)\} = O\big(\max_{1\leq i\leq n} \{T_{u_i}(r)\}\big) \tag{4-26}$$

and

$$N_{G_2(\lambda_1,\lambda_2,\beta_{2,3},\dots,\beta_{2,n})}(0, r) - N^{(1)}_{G_2(\lambda_1,\lambda_2,\beta_{2,3},\dots,\beta_{2,n})}(0, r) <_{\text{exc}} \epsilon \max_{1\leq i\leq n} \{T_{u_i}(r)\} \tag{4-27}$$

fails to hold.

Let $2 \le s \le n-1$ and suppose that we have completed Step $s-1$. This includes the construction of $H_{s-1} \in K[x_1, \dots, x_n]$ with degree depends on ϵ, n and $\deg G$ only. We then complete Step s in the same manner Step 2 is carried out after Step 1. The last one is Step $n-1$ resulting in $H_{n-1} \in K[x_1, \dots, x_n]$. We now define $H = H_1 \cdots H_{n-1}$. Then $\deg H$ depends only on ϵ, n and $\deg G$ since each H_i does so. Suppose $H(u_1, \dots, u_n) \not\equiv 0$. Assume we go through all the above $n-1$ steps to get the polynomial

$$P(X_{n-1,n}) := G_{n-1}(\lambda_1, \dots, \lambda_{n-1}, X_{n-1,n}) \in K_{\boldsymbol{g}}[X_{n-1,n}]$$

such that its degree can be bounded explicitly in terms of $M_{n-1}, \dots, M_1$, n, and $\deg G$. At the end of Step $n-1$, we have that $\beta_{n-1,n}$ has small zero and pole with respect to $\boldsymbol{g}$, so it satisfies

$$T_{\beta_{n-1,n}}(r) = O\Big(\max_{1\le i\le n}\{T_{u_i}(r)\}\Big). \tag{4-28}$$

If

$$N_{P(\beta_{n-1,n})}(0,r) - N^{(1)}_{P(\beta_{n-1,n})}(0,r) <_{\text{exc}} \epsilon \max_{1\le i\le n}\{T_{u_i}(r)\}, \tag{4-29}$$

then we are done. Otherwise, since $H_{n-1}(u_1, \dots, u_n) \not\equiv 0$, the polynomial $P(X_{n-1,n})$ has neither monomial nor repeated irreducible factors, according to Theorem 15, there exists a nonzero integer m such that, by using (4-28),

$$T_{\beta^m_{n-1,n}}(r) = o(T_{\beta_{n-1,n}}(r)), \tag{4-30}$$

which is not possible since $\beta_{n-1,n}$ is not constant. □

4.4. *Proof of Theorem 1.*

Proof of Theorem 1. Let $F \in K[x_0, \dots, x_n]$. Consider a holomorphic curve $\boldsymbol{g} = (g_0, \dots, g_n)$, where $g_0, \dots, g_n$ are entire functions with no common zeros, such that $K \subset K_{\boldsymbol{g}}$ and $N_{\boldsymbol{g}}(H_i, r) = o(T_{\boldsymbol{g}}(r))$ for $0 \le i \le n$. Let $u_i = g_i/g_0$ for $0 \le i \le n$, $\boldsymbol{u} = (u_1, \dots, u_n)$, and $G := F(1, x_1, \dots, x_n) \in K[x_1, \dots, x_n]$. Then

$$\begin{aligned} N_{u_i}(0,r) + N_{u_i}(\infty,r) &\le N_{g_i}(0,r) + N_{g_0}(0,r) \\ &= N_{\boldsymbol{g}}(H_i,r) + N_{\boldsymbol{g}}(H_0,r) = o(T_{\boldsymbol{g}}(r)) \end{aligned} \tag{4-31}$$

for each $1 \le i \le n$, and, by (4-1),

$$\max_{1\le i\le n}\{T_{u_i}(r)\} = O(T_{\boldsymbol{g}}(r)). \tag{4-32}$$

Since $F(\boldsymbol{g}) = F(g_0, \dots, g_n) = g_0^d G(\boldsymbol{u})$, we have

$$N_{F(\boldsymbol{g})}(0,r) = N_{G(\boldsymbol{u})}(0,r) + o(T_{\boldsymbol{g}}(r)), \quad N^{(1)}_{F(\boldsymbol{g})}(0,r) = N^{(1)}_{G(\boldsymbol{u})}(0,r) + o(T_{\boldsymbol{g}}(r)). \tag{4-33}$$

Consequently, we may apply Theorem 16 for any given positive real ϵ to find a nontrivial polynomial $Q \in K[x_1, \dots, x_n]$ such that (2-2) holds, that is,

$$N_{F(\boldsymbol{g})}(0,r) - N^{(1)}_{F(\boldsymbol{g})}(0,r) \le \epsilon T_{\boldsymbol{g}}(r), \tag{4-34}$$

when $Q(\boldsymbol{u}) \not\equiv 0$. In addition, the polynomial Q can be determined effectively and the degree of Q can be bounded effectively in terms of ϵ, n and the degree of F. At this step, we take Z to be the zero locus of the homogeneous polynomial

$$x_0^{\deg Q} \cdot Q\left(\frac{x_1}{x_0}, \dots, \frac{x_n}{x_0}\right) \in K[x_0, \dots, x_n].$$

Let $F = \sum_{\boldsymbol{i} \in I_F} \alpha_{\boldsymbol{i}} \boldsymbol{x}^{\boldsymbol{i}} \in K[x_0, \dots, x_n]$, and let W be the Zariski closed subset that is the union of hypersurfaces of $\mathbb{P}^n$ of the form $\sum_{\boldsymbol{i} \in J} \alpha_{\boldsymbol{i}} \boldsymbol{x}^{\boldsymbol{i}} = 0$, where J is a nonempty subset of I_F. The Zariski closed set $Z \cup W$ satisfies (Z1) and (Z2) since both Z and W do so. We now prove (2-3) holds (after possibly enlarging Z) by further assuming that the hypersurface $[F = 0]$ and the coordinate hyperplanes in $\mathbb{P}^n$ are in weakly general position. Therefore, we may write

$$F(\boldsymbol{g}) = \sum_{0 \le i \le n} \alpha_{\boldsymbol{i}_i} g_i^d + \sum_{\boldsymbol{i} \in I_G \setminus I} \alpha_{\boldsymbol{i}} \boldsymbol{g}^{\boldsymbol{i}}, \tag{4-35}$$

where $\alpha_{\boldsymbol{i}_i} \ne 0$ for $0 \le i \le n$ and $I = \{\boldsymbol{i}_0 := (d, 0, \dots, 0), \dots, \boldsymbol{i}_n := (0, \dots, 0, d)\}$.

For $\boldsymbol{g}$ with $\boldsymbol{g}(\mathbb{C})$ not contained in $Z \cup W$, we may use Theorem 6 to show that

$$dT_{\boldsymbol{g}}(r) \le N_{F(\boldsymbol{g})}(0,r) + o(T_{\boldsymbol{g}}(r))$$

since $\alpha_{\boldsymbol{i}_i} \in K_{\boldsymbol{g}}$ and $N_{g_i}(0,r) = o(T_{\boldsymbol{g}}(r))$ for $0 \le i \le n$. Together with (4-34), we arrive at $N^{(1)}_{F(\boldsymbol{g})}(0,r) \ge (d - \epsilon) T_{\boldsymbol{g}}(r)$. By letting $Z \cup W$ be the desired exceptional set Z, we finish the proof. ☐

5. Proof of Theorem 2

We will adapt the proof strategy employed in [6, Theorem 1.2] to suit the current situation and subsequently apply Theorem 1.

Proof. Let $z_0 \in \mathbb{C}$ such that all the coefficients of all F_i, $1 \le i \le n+1$, are holomorphic at z_0 and the zero locus of F_i, $1 \le i \le n+1$, evaluating at z_0, denoted by $D_i(z_0)$, intersect transversally. These conditions imply that z_0 is not a common zero of the coefficients of F_i, for each $1 \le i \le n+1$.

Since the zero locus of $F_i(z_0)$, $1 \le i \le n+1$, intersect transversally, they are in general position; thus the set of polynomials F_i, $1 \le i \le n+1$, is in weakly general position. Then Proposition 7 implies that the only $(x_0, \dots, x_n) \in \mathcal{M}^{n+1}$ with $F_i(x_0, \dots, x_n) \equiv 0$ for each $1 \le i \le n+1$ is $(0, \dots, 0)$. Thus the association

$\boldsymbol{x} \mapsto [F_1^{a_1}(\boldsymbol{x}) : \cdots : F_{n+1}^{a_{n+1}}(\boldsymbol{x})]$, where $a_i := \operatorname{lcm}(\deg F_1, \ldots, \deg F_{n+1})/\deg F_i$, defines a morphism $\pi : \mathbb{P}^n(\mathcal{M}) \to \mathbb{P}^n(\mathcal{M})$ over K. Let

$$G := \det\left(\frac{\partial F_i}{\partial x_j}\right)_{\substack{1\le i\le n+1\\ 0\le j\le n}} \in K[x_0, \ldots, x_n].$$

Define $\pi|_{z_0} = [F_1^{a_1}(z_0) : \cdots : F_{n+1}^{a_{n+1}}(z_0)] : \mathbb{P}^n(\mathbb{C}) \to \mathbb{P}^n(\mathbb{C})$, which is a morphism since $F_1(z_0), \ldots, F_{n+1}(z_0)$ are in general position. As proved in [6, Theorem 1.2], we have that $[G(z_0) = 0]$ (the zero locus of $G(z_0)$), $D_1(z_0), \ldots, D_{n+1}(z_0)$ are in general position (in $\mathbb{P}^n(\mathbb{C})$). Hence, there is a nonconstant irreducible factor $\tilde{G}$ of G in $K[x_0, \ldots, x_n]$ such that $\tilde{G}$, $F_1, \ldots, F_{n+1}$ is in weakly general position. Denote by Y the zero locus of $\tilde{G}$ in $\mathbb{P}^n(\overline{K})$. We note that Y is contained in the ramification divisor of π since $\tilde{G}$ is a factor of the determinant of the Jacobian matrix associated with the map π. Then there exists an irreducible homogeneous polynomial $A \in K[y_0, \ldots, y_n]$ such that the vanishing order of $\pi^* A$ along Y is at least 2. Then this construction gives $\pi^* \circ A = \tilde{G}^2 H$ for some $H \in K[x_0, \ldots, x_n]$. Since the divisors defined by $\tilde{G}(z_0), F_1(z_0), \ldots, F_{n+1}(z_0)$ are in general position, their images are also in general position. Therefore, A and y_i, $0 \le i \le n$, are in weakly general position.

Now let $\boldsymbol{f} = (f_0, \ldots, f_n) : \mathbb{C} \to \mathbb{P}^n$ be a holomorphic map, where $f_0, \ldots, f_n$ are entire functions without common zeros. Assume that $K \subset K_{\boldsymbol{f}}$. Let $\boldsymbol{g} := \pi(\boldsymbol{f}) = (F_1(\boldsymbol{f})^{a_1}, \ldots, F_{n+1}(\boldsymbol{f})^{a_{n+1}})$, where each $F_i(\boldsymbol{f})^{a_i}$, $1 \le i \le n+1$, is an entire function with no zeros. Then

$$T_{\boldsymbol{g}}(r) = d_1 T_{\boldsymbol{f}}(r) + o(T_{\boldsymbol{f}}(r)), \tag{5-1}$$

where $d_1 = \deg F_1 \cdot a_1$. From $A(\boldsymbol{g}) = (\pi^* \circ A)(\boldsymbol{f}) = \tilde{G}^2(\boldsymbol{f}) H(\boldsymbol{f})$, it follows that for each $z \in \mathbb{C}$ with $v_z(\tilde{G}(\boldsymbol{f})) > 0$, we have

$$\begin{aligned} \max\{0, v_z(A(\boldsymbol{g}))\} &\ge 2 v_z(\tilde{G}(\boldsymbol{f})) + \min\{0, v_z(H(\boldsymbol{f}))\} \\ &\ge v_z(\tilde{G}(\boldsymbol{f})) + 1 + \min\{0, v_z(H(\boldsymbol{f}))\}. \end{aligned} \tag{5-2}$$

Since $f_0, \ldots, f_n$ are entire functions, the nonnegative number $-\min\{0, v_z(H(\boldsymbol{f}))\}$ is bounded by the number of poles of the coefficients of H at z. Since the coefficients of H are in K and $N_\beta(\infty, r) \le T_\beta(r) + O(1) = o(T_{\boldsymbol{f}}(r))$ for any $\beta \in K$, it follows from (5-2) that

$$N_{\tilde{G}(\boldsymbol{f})}(0, r) \le N_{A(\boldsymbol{g})}(0, r) - N^{(1)}_{A(\boldsymbol{g})}(0, r) + o(T_{\boldsymbol{f}}(r)). \tag{5-3}$$

Assume furthermore that $N_{F_i(\boldsymbol{f})}(0, r) = o(T_{\boldsymbol{f}}(r))$ for $1 \le i \le n+1$. Then $N_{\boldsymbol{g}}(H_i, r) = o(T_{\boldsymbol{f}}(r))$ ($= o(T_{\boldsymbol{g}}(r))$ by (5-1)) for coordinate hyperplanes H_i, $0 \le i \le n$, of $\mathbb{P}^n$. We now apply Theorem 1 for $\epsilon = 1/(4d_1)$. Then we can find a homogeneous polynomial $B_0 \in K[y_0, \ldots, y_n]$ such that for any nonconstant holomorphic map

$\boldsymbol{f} = (f_0, \dots, f_n) : \mathbb{C} \to \mathbb{P}^n$ such that $K \subset K_{\boldsymbol{f}}$ and $N_{F_i(\boldsymbol{f})}(0,r) = o(T_{\boldsymbol{f}}(r))$ for $1 \le i \le n+1$, with $B_0(\boldsymbol{g}) = B_0(\pi(\boldsymbol{f}))$ not identically zero, we have

$$N_{A(\boldsymbol{g})}(0,r) - N^{(1)}_{A(\boldsymbol{g})}(0,r) \le_{\mathrm{exc}} \epsilon T_{\boldsymbol{g}}(r) \tag{5-4}$$

and

$$N^{(1)}_{A(\boldsymbol{g})}(0,r) \ge_{\mathrm{exc}} (\deg A - \epsilon) \cdot T_{\boldsymbol{g}}(r). \tag{5-5}$$

Combining (5-3) and (5-4), we have

$$N_{\tilde{G}(\boldsymbol{f})}(0,r) \le_{\mathrm{exc}} \epsilon T_{\boldsymbol{g}}(r). \tag{5-6}$$

Since $[\tilde{G}=0] \le \pi^*([A=0])$ as divisors, we can derive, from the functorial property of Weil functions,

$$m_{\boldsymbol{f}}([\tilde{G}=0],r) \le m_{\boldsymbol{g}}([A=0],r) = \deg A \cdot T_{\boldsymbol{g}}(r) - N_{A(\boldsymbol{g})}(0,r) + o(T_{\boldsymbol{g}}(r)). \tag{5-7}$$

Then by (5-5),we have

$$m_{\boldsymbol{f}}([\tilde{G}=0],r) \le_{\mathrm{exc}} \epsilon T_{\boldsymbol{g}}(r). \tag{5-8}$$

Combining (5-6), (5-8) and (5-1), we have

$$T_{[\tilde{G}=0],\boldsymbol{f}}(r) \le_{\mathrm{exc}} 2\epsilon T_{\boldsymbol{g}}(r) = 2\epsilon \cdot d_1 T_{\boldsymbol{f}}(r) + o(T_{\boldsymbol{f}}(r)), \tag{5-9}$$

On the other hand, the first main theorem implies that

$$\deg \tilde{G} \cdot T_{\boldsymbol{f}}(r) = T_{[\tilde{G}=0],\boldsymbol{f}}(r) + o(T_{\boldsymbol{f}}(r)). \tag{5-10}$$

Therefore, we have

$$T_{\boldsymbol{f}}(r) \le_{\mathrm{exc}} 2\epsilon \cdot d_1 T_{\boldsymbol{f}}(r) + o(T_{\boldsymbol{f}}(r)), \tag{5-11}$$

which is not possible since $\epsilon = 1/(4d_1)$. This shows that $B_0(\boldsymbol{g})$ is identically zero. Let $B := \pi^*(B_0) = B_0(F_1^{a_1}, \dots, F_{n+1}^{a_{n+1}}) \in K[x_0, \dots, x_n]$, which is not identically zero since π is a finite morphism. Then $B(\boldsymbol{f})$ is identically zero as asserted. □

The defect relation stated in Corollary 3 directly follows from Theorem 2 by noticing that $\sum_{i=1}^{n+1} \delta_{\boldsymbol{f}}(D_i) = n+1$ if and only if $N_{\boldsymbol{f}}(D_i, r) = o(T_{\boldsymbol{f}}(r))$ for each i. To establish the truncated defect relation for $n = 2$, we relax the assumption to $N^{(1)}_{\boldsymbol{g}}(H_i, r) = o(T_{\boldsymbol{g}}(r))$ for $0 \le i \le 2$. In order to apply this relaxed condition (b) in Theorem 15, one must assume that the hypersurface $[G = 0]$ and the $n+1$ coordinate hyperplanes are in weakly general position in $\mathbb{P}^n$. Unfortunately, this geometric condition does not persist under the induction process. We state a modified version of Theorem 1 below to demonstrate that Theorem 2 remains valid under these relaxed assumptions.

Theorem 17. *Let K be a subfield of the field $\mathcal{M}$ of meromorphic functions. Let G be a nonconstant homogeneous polynomial in $K[x_0, x_1, x_2]$ with no monomial factors and no repeated factors. Let $H_i = [x_{i-1} = 0]$, $1 \le i \le 3$, be the coordinate hyperplane divisors of $\mathbb{P}^2$. Assume that the plane curve $[G = 0]$ and H_i, $1 \le i \le 3$, are in weakly general position. Then for any $\epsilon > 0$, there exists a proper Zariski closed subset Z of $\mathbb{P}^2$ defined over K such that for any nonconstant holomorphic curve $\boldsymbol{g} = (g_0, g_1, g_2) : \mathbb{C} \to \mathbb{P}^2(\mathbb{C})$ such that $N^{(1)}_{\boldsymbol{g}}(H_i, r) = o(T_{\boldsymbol{g}}(r))$ for $0 \le i \le 2$ with $\boldsymbol{g}$ not contained in Z, we have*

$$N_{G(\boldsymbol{g})}(0, r) - N^{(1)}_{G(\boldsymbol{g})}(0, r) \le_{\text{exc}} \epsilon T_{\boldsymbol{g}}(r) \tag{5-12}$$

and

$$N^{(1)}_{G(\boldsymbol{g})}(0, r) \ge_{\text{exc}} (\deg G - \epsilon) \cdot T_{\boldsymbol{g}}(r). \tag{5-13}$$

Furthermore, the exceptional set Z is a finite union of closed subsets given by homogenization equations of the form $x_1^{n_1} x_2^{n_2} = \lambda$, where $\lambda \in K^*$ and (n_1, n_2) is a pair of integers with $\max\{|n_1|, |n_2|\}$ bounded from above by an effectively computable integer m.

Proof of Corollary 3. Since $0 \le \delta_{\boldsymbol{f}}(D_i) \le 1$ for $1 \le i \le n+1$, it is clear that $\sum_{i=1}^{n+1} \delta_{\boldsymbol{f}}(D_i) = n+1$ if and only if $\delta_{\boldsymbol{f}}(D_i) = 1$ for each i. On the other hand, $\delta_{\boldsymbol{f}}(D_i) = 1$ if and only if $N_{\boldsymbol{f}}(D_i, r) = o(T_{\boldsymbol{f}}(r))$. Therefore, we have either $\sum_{i=1}^{n+1} \delta_{\boldsymbol{f}}(D_i) < n+1$ or there exists a homogeneous polynomial $B \in K[x_0, \dots, x_n]$ as described Theorem 2 such that $B(\boldsymbol{f})$ is identically zero.

When $n = 2$, the conclusion of Theorem 2 holds under a weaker assumption that $N^{(1)}_{\boldsymbol{f}}(D_i, r) = o(T_{\boldsymbol{f}}(r))$ for $i = 0, 1, 2$ by replacing the use of Theorem 1 with Theorem 17. Therefore, the above arguments show that $\sum_{i=1}^{3} \delta^{(1)}_{\boldsymbol{f}}(D_i) < 3$ or there exists a homogeneous polynomial $B \in K[x_0, x_1, x_2]$ as described Theorem 2 such that $B(\boldsymbol{f})$ is identically zero. □

Proof of Theorem 17. Let $\boldsymbol{g} = (g_0, g_1, g_2)$ with $N^{(1)}_{g_i}(0, r) = o(T_{\boldsymbol{g}}(r))$, $0 \le i \le 2$, where g_0, g_1, g_2 have no common zeros. We prove (5-12) first. Under our assumption, the condition (b) in Theorem 15 holds. Hence, by Theorem 15, we only need to consider the case that

$$T_{(g_1/g_0)^{n_1}(g_2/g_0)^{n_2}}(r) = o(T_{\boldsymbol{g}}(r)). \tag{5-14}$$

We may assume that n_1 and n_2 are coprime. Consequently, there exist integers a and b such that $n_1 a + n_2 b = 1$. Consider the variables

$$\Lambda = X^{n_1} Y^{n_2} \quad \text{and} \quad T = X^b Y^{-a}. \tag{5-15}$$

Then, we may express

$$X = \Lambda^a T^{n_2} \quad \text{and} \quad Y = \Lambda^b T^{-n_1}. \tag{5-16}$$

Let $G_1(X,Y)=G(1,X,Y)$. Define $B(\Lambda,T)\in K[\Lambda,T]$ as the polynomial with no monomial factors and such that

$$G_1(X,Y)=G_1(\Lambda^a T^{n_2},\Lambda^b T^{-n_1})=T^{M_1}\Lambda^{M_2}B(\Lambda,T) \tag{5-17}$$

for some integers M_1 and M_2.

Let $u_1=g_1/g_0$, $u_2=g_2/g_0$, and $\lambda:=u_1^{n_1}u_2^{n_2}$. Then we have

$$T_\lambda(r)=o(T_{\boldsymbol{g}}(r)). \tag{5-18}$$

To prove (5-12), we will reduce the problem to one-variable polynomials $B(\lambda,T)$ for all possible $\lambda\in K$ that satisfy (5-18) but not (5-12). Our objective is to eliminate those λ values with $B(\lambda,T)$ containing a factor of T or having repeated factors, so that we can apply the GCD theorem after eliminating those λ. Since T is not a factor of $B(\Lambda,T)$, it follows that $B(\Lambda,0)\in K[\Lambda]$ is not identically zero. Consequently, there exist at most finite $\gamma_1,\dots,\gamma_s\in K$ such that $B(\gamma_i,0)=0$ for $1\le i\le s$. Therefore, T is not a factor of $B(\lambda,T)$ if $\lambda\ne\gamma_i$, $1\le i\le s$. Regarding repeated factors, let's express $B(\Lambda,T)=B_\Lambda(T)\in K[\Lambda][T]$. Since the transformation in (5-15) establishes to a bijection between the sets $\{X^{t_1}Y^{t_2}:t_1,t_2\in\mathbb{Z}\}$ and $\{\Lambda^{a_1}T^{a_2}:a_1,a_2\in\mathbb{Z}\}$, it is evident that $B(\Lambda,T)\in K[\Lambda,T]$ is square free, given that G is square free. Consequently, the resultant $R(B_\Lambda,B'_\Lambda)$ of B_Λ and $B'_\Lambda(T)$ is a polynomial in $K[\Lambda]$, which is not identically zero. Let

$$\alpha_i\in K,\ 1\le i\le t,\ \text{be the zeros of the resultant } R(B_\Lambda,B'_\Lambda). \tag{5-19}$$

It is clear that $B(\lambda,T)$ has no multiple factors in $K[\lambda][T]$ if $\lambda\ne\alpha_i$ for any $1\le i\le t$. Therefore, it is clear that we need to consider those λ with $\lambda\ne\alpha_i$ for any $1\le i\le t$ and $\lambda\ne\gamma_j$ for any $1\le j\le s$. Assuming such, let $B(T):=\lambda^{M_2}B(\lambda,T)$ as in (5-17). Let $\beta:=u_1^bu_2^{-a}$ and define $D_\beta(B)\in K_{\boldsymbol{g}}[T]$ as in (4-3). By Lemma 14, the polynomials B and $D_\beta(B)$ are coprime in $K_{\boldsymbol{g}}[T]$. Let $\tilde{B}\in K(\lambda)[Z,U]$ and $\tilde{D}_\beta(B)$ be the homogenization of B and $D_\beta(B)$, respectively. Write $\beta=\beta_1/\beta_0$, where β_0 and β_1 are entire functions without common zeros. Then by Proposition 10

$$N_{\mathrm{gcd}}(\tilde{B}(\beta_0,\beta_1),\tilde{D}_\beta(B)(\beta_0,\beta_1),r)\le o(T_{\boldsymbol{g}}(r)) \tag{5-20}$$

since β is not constant. On the other hand, from the proof of [4, Proposition 5.3], there exists a proper Zariski closed set W of $\mathbb{P}^2(\mathbb{C})$, independent of $\boldsymbol{g}$, such that, if image of $\boldsymbol{g}$ is contained in W,

$$N_{G(\boldsymbol{g})}(0,r)-N^{(1)}_{G(\boldsymbol{g})}(0,r)\le_{\mathrm{exc}} N_{\mathrm{gcd}}(\tilde{B}(\beta_0,\beta_1),\tilde{D}_\beta(B)(\beta_0,\beta_1),r). \tag{5-21}$$

Furthermore, W can be described in Theorem 17. We conclude the proof of (5-12) by combining (5-20) and (5-21).

We now proceed to prove (5-13). Let $G=\sum_{\boldsymbol{i}\in I_G}\alpha_{\boldsymbol{i}}\boldsymbol{x}^{\boldsymbol{i}}\in K[x_0,x_1,x_2]$. Since the hypersurface $[G=0]$ and the coordinate hyperplanes in $\mathbb{P}^2$ are in weakly general position, may write

$$G(\boldsymbol{g}) = \sum_{0\le i\le 2} \alpha_{\boldsymbol{i}_i} g_i^d + \sum_{\boldsymbol{i}\in I_G\setminus I} \alpha_{\boldsymbol{i}} \boldsymbol{g}^{\boldsymbol{i}}, \tag{5-22}$$

where $\alpha_{\boldsymbol{i}_i} \neq 0$ for $0 \le i \le 2$ and $I = \{(d, 0, 0), (0, d, 0), (0, 0, d)\}$.

Let's express $B(\Lambda, T)$ in the form

$$B(\Lambda, T) = \sum_{i\in I_B} b_i(\Lambda)T^i \in K[\Lambda][T], \tag{5-23}$$

where $b_i \neq 0$ if $i \in I_B$. We define $J \subset K[\Lambda]$ as the finite set containing all $b_i(\Lambda)$ for $i \in I_B$ and all of their proper subsums. Set $\mathcal{R} := \{r \in K \mid h(r) = 0$ for some $h \in J\}$. It is crucial that the proof of Theorem 1 has already demonstrated that (5-13) holds if neither $G(\boldsymbol{g})$ nor any proper subsum of (5-22) is zero. Therefore, when evaluating $B(\Lambda, T)$ at $\Lambda = \lambda \notin \mathcal{R}$ and $T = \beta$, we need to consider equations of the type

$$\sum_{i\in I_B} a_i(\lambda)\beta^i = 0, \tag{5-24}$$

where $a_i(\Lambda)$ is a subsum of $b_i(\Lambda)$, and there are at least two nontrivial a_i in the left-hand side of (5-24) since $\lambda \notin \mathcal{R}$. Hence,

$$T_\beta(r) \le c_3 T_\lambda(r) = o(T_{\boldsymbol{g}}(r)).$$

This, however, leads to a contradiction. □

References

[1] Z. Chen, D. T. Huynh, R. Sun, and S. Y. Xie, "Entire holomorphic curves into $\mathbb{P}^n(\mathbb{C})$ intersecting $n+1$ general hypersurfaces", *Science China Math.* (online publication March 2025).

[2] P. Corvaja and U. Zannier, "Some cases of Vojta's conjecture on integral points over function fields", *J. Algebraic Geom.* **17**:2 (2008), 295–333. MR

[3] G. Dethloff and T. V. Tan, "A second main theorem for moving hypersurface targets", *Houston J. Math.* **37**:1 (2011), 79–111. MR

[4] J. Guo and J. T.-Y. Wang, "A complex case of Vojta's general ABC conjecture and cases of Campana's orbifold conjecture", *Trans. Amer. Math. Soc.* **377**:7 (2024), 4961–4991. MR

[5] J. Guo, C.-L. Sun, and J. T.-Y. Wang, "On the dth roots of exponential polynomials and related problems arising from the Green–Griffiths–Lang conjecture", *J. Geom. Anal.* **31**:5 (2021), 5201–5218. MR

[6] J. Guo, C.-L. Sun, and J. T.-Y. Wang, "A truncated second main theorem for algebraic tori with moving targets and applications", *J. Lond. Math. Soc.* (2) **106**:4 (2022), 3670–3686. MR

[7] J. Guo, K. D. Nguyen, C.-L. Sun, and J. T.-Y. Wang, "Vojta's abc conjecture for algebraic tori and applications over function fields", *Adv. Math.* **476** (2025), art. id. 110358. MR

[8] S. Lang, *Algebra*, 3rd ed., Graduate Texts in Mathematics **211**, Springer, 2002. MR

[9] A. Levin and J. T.-Y. Wang, "Greatest common divisors of analytic functions and Nevanlinna theory on algebraic tori", *J. Reine Angew. Math.* **767** (2020), 77–107. MR

[10] J. Noguchi and J. Winkelmann, "Holomorphic curves and integral points off divisors", *Math. Z.* **239**:3 (2002), 593–610. MR

[11] J. Noguchi, J. Winkelmann, and K. Yamanoi, "Degeneracy of holomorphic curves into algebraic varieties", *J. Math. Pures Appl.* (9) **88**:3 (2007), 293–306. MR

[12] J. Noguchi, J. Winkelmann, and K. Yamanoi, "The second main theorem for holomorphic curves into semi-abelian varieties, II", *Forum Math.* **20**:3 (2008), 469–503. MR

[13] M. Păun and N. Sibony, "Value distribution theory for parabolic Riemann surfaces", pp. 13–72 in *Hyperbolicity properties of algebraic varieties*, edited by S. Diverio, Panor. Synthèses **56**, Soc. Math. France, Paris, 2021. MR

[14] M. Ru, "Integral points and the hyperbolicity of the complement of hypersurfaces", *J. Reine Angew. Math.* **442** (1993), 163–176. MR

[15] M. Ru, "A defect relation for holomorphic curves intersecting hypersurfaces", *Amer. J. Math.* **126**:1 (2004), 215–226. MR

[16] M. Ru, *Nevanlinna theory and its relation to Diophantine approximation*, 2nd ed., World Scientific, Hackensack, NJ, 2021. MR

[17] M. Ru and J. T.-Y. Wang, "Truncated second main theorem with moving targets", *Trans. Amer. Math. Soc.* **356**:2 (2004), 557–571. MR

[18] B. L. van der Waerden, *Algebra, II*, Springer, New York, 1970. MR

Received April 16, 2024. Revised June 10, 2025.

MIN RU
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON
HOUSTON, TX
UNITED STATES
minru@math.uh.edu

JULIE TZU-YUEH WANG
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI
TAIWAN
jwang@math.sinica.edu.tw

PACIFIC JOURNAL OF MATHEMATICS
Vol. 338, No. 2, 2025
https://doi.org/10.2140/pjm.2025.338.373

THE TANGENT SPACES OF TEICHMÜLLER SPACE FROM AN ENERGY-CONSCIOUS PERSPECTIVE

DIVYA SHARMA AND MICHAEL S. WEISS

The Teichmüller space of a closed oriented (real) surface of genus at least 2 is a moduli space of complex structures on the surface, but can also be defined as a space of certain representations of the fundamental group of the surface in the group of orientation-preserving isometries of the hyperbolic plane. As a consequence the tangent spaces of Teichmüller space admit two rather different descriptions. We use harmonic vector fields (defined as infinitesimal analogs of harmonic maps) on the hyperbolic plane to make a bridge between these descriptions.

1. Introduction

1.1. ***Teichmüller space.*** Let Σ be an oriented, connected, closed and smooth surface of genus ≥ 2. Write $\mathrm{diff}_0(\Sigma)$ for the group of diffeomorphisms $\Sigma \to \Sigma$ which are isotopic to the identity. By the Korn–Lichtenstein theorem, the choice of a complex structure on Σ is equivalent to the choice of a complex structure $J : T\Sigma \to T\Sigma$ on the tangent bundle. (J is a smooth vector bundle automorphism covering $\mathrm{id} : \Sigma \to \Sigma$, and it satisfies $J^2 = -\mathrm{id}$. Such a J can be called an *almost complex structure on* Σ.)

Definition 1.1.1. Teichmüller space $\mathscr{T}(\Sigma)$ is the space of complex structures J on $T\Sigma$, modulo the right action of $\mathrm{diff}_0(\Sigma)$ given by $(J \cdot f)_x := (df_x)^{-1} \circ J_{f(x)} \circ df_x$ for $f \in \mathrm{diff}_0(\Sigma)$ and $x \in \Sigma$.

We will rely more on the "metric" definition of $\mathscr{T}(\Sigma)$. Let $\Sigma' \to \Sigma$ be a universal covering with deck transformation group Γ. We also refer to Γ as the *fundamental group* of Σ. A complex structure on Σ determines a complex structure on Σ', and an embedding of Γ into the group of complex automorphisms of Σ'. By uniformization

Sharma held an Ada Lovelace PhD studentship awarded by the University of Münster, 2017–2021. In the early stages, this work was supported by the Alexander von Humboldt Foundation through a Humboldt Professorship award to Weiss; in the later stages, 2019–2022, by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044-390685587, Mathematics Münster: Dynamics–Geometry–Structure.
MSC2020: 30F60.
Keywords: Teichmüller space, harmonic, quadratic differential, Poisson kernel.

theory, Σ' is isomorphic as a complex manifold to the open unit disk $\mathbb{D} \subset \mathbb{C}$. In this way, a complex structure on Σ gives us an embedding ρ of Γ into the group of complex automorphisms of $\mathbb{D}$, which we can also view as the group $\mathrm{isom}_+(\mathbb{D})$ of orientation-preserving isometries of $\mathbb{D}$ equipped with the Poincaré metric. The homomorphism $\rho : \Gamma \to \mathrm{isom}_+(\mathbb{D})$ is not well defined as such, but it is well defined up to conjugation by an element of $\mathrm{isom}_+(\mathbb{D})$. This leads us to the metric definition of $\mathscr{T}(\Sigma)$.

Definition 1.1.2. Teichmüller space $\mathscr{T}(\Sigma)$ is the space of injective homomorphisms $\rho : \Gamma \to \mathrm{isom}_+(\mathbb{D})$ with discrete and cocompact image $\rho(\Gamma)$, modulo the left action of $\mathrm{isom}_+(\mathbb{D})$ by conjugation.

The equivalence of the two definitions of $\mathscr{T}(\Sigma)$ is well known. The metric definition does not *use* uniformization theory, although it is explained by uniformization theory. One has to do some work to show that $\mathscr{T}(\Sigma)$, according to that definition, is a smooth manifold of real dimension $-3\chi(\Sigma)$.

1.2. ***The tangent spaces of Teichmüller space.*** We continue in the notation of the previous section. In particular, $\Sigma' \to \Sigma$ is a universal covering, J is a complex structure on $T\Sigma$ and $\rho : \Gamma \to \mathrm{isom}_+(\mathbb{D})$ is an injective homomorphism with discrete and cocompact image.

Definition 1.2.1. A quadratic differential on (Σ, J) is a (continuous) section of the complex line bundle $\hom_{\mathbb{C}}(T\Sigma \otimes_{\mathbb{C}} T\Sigma, \mathbb{C})$ on Σ.

In one description, the tangent space of $\mathscr{T}(\Sigma)$ at a point J (complex structure on $T\Sigma$) "is" the vector space of holomorphic quadratic differentials on Σ. This can be justified as follows. There is an $\mathbb{R}$-linear injective map $\phi \mapsto \mathrm{Re}(\phi)$ from the complex 1-dimensional vector bundle $\hom_{\mathbb{C}}(T\Sigma \otimes_{\mathbb{C}} T\Sigma, \mathbb{C})$ on Σ to the real 3-dimensional vector bundle of symmetric $\mathbb{R}$-bilinear forms on $T\Sigma$. We have already seen that J determines a complex structure on $\Sigma' \cong \mathbb{D}$ and so a hyperbolic metric on Σ' invariant under the action of Γ, and so a hyperbolic (Riemannian) metric g on Σ itself. A holomorphic section ϕ of $\hom_{\mathbb{C}}(T\Sigma \otimes_{\mathbb{C}} T\Sigma, \mathbb{C})$, indeed any smooth section ϕ of that vector bundle, determines a 1-parameter family of Riemannian metrics on Σ by $t \mapsto g + t\mathrm{Re}(\phi)$ for $t \in \mathbb{R}$ close enough to 0. Each of the Riemannian metrics $g + t\mathrm{Re}(\phi)$ determines a conformal structure on Σ, hence a complex structure J_t on $T\Sigma$. This gives a (germ of a) smooth curve $t \mapsto J_t$ in the Teichmüller space, with $J_0 = J$. The velocity vector of that curve at $t = 0$ is the tangent vector which we associate to ϕ. This procedure gives an isomorphism from the vector space of holomorphic quadratic differentials on Σ to that tangent space. It uses Definition 1.1.1. See [Imayoshi and Taniguchi 1992] for more details.

The other popular description of the tangent spaces of $\mathscr{T}(\Sigma)$ relies on the metric definition of $\mathscr{T}(\Sigma)$. Instead of selecting a datum J, we begin with $\rho : \Gamma \to \mathrm{isom}_+(\mathbb{D})$.

Let $\mathfrak{g}$ be the tangent space of $\mathrm{isom}_+(\mathbb{D})$ at the identity element, a.k.a. the Lie algebra of $\mathrm{isom}_+(\mathbb{D})$. The homomorphism ρ determines a left action of Γ on $\mathfrak{g}$ by conjugation, $\gamma \cdot v := \mathrm{ad}(\rho(\gamma))(v)$ for $\gamma \in \Gamma$ and $v \in \mathfrak{g}$. We may write $\mathfrak{g}_\rho$ to specify this action.

We select a tangent vector to $\mathscr{T}(\Sigma)$ at ρ by choosing a smooth 1-parameter family $(\rho_t)_{t\in[-\varepsilon,+\varepsilon]}$ of homomorphisms $\Gamma \to \mathrm{isom}_+(\mathbb{D})$ such that $\rho_0 = \rho$. (We could insist that the homomorphisms ρ_t are all injective with discrete and cocompact image, like ρ_0, but by [Weil 1960] this is automatically satisfied for t close enough to 0.) Then we can form

$$\frac{d(\rho_t \cdot \rho_0^{-1})}{dt}\Big|_{t=0},$$

which gives us a map from Γ to $\mathfrak{g}_\rho$. It turns out to be a 1-cocycle. Its class in $H^1(\Gamma; \mathfrak{g}_\rho)$ is well defined. We arrive at the following description of the tangent space of $\mathscr{T}(\Sigma)$ at the point determined by the homomorphism ρ: it is $H^1(\Gamma; \mathfrak{g}_\rho)$. It is a small disadvantage of this description that $H^1(\Gamma; \mathfrak{g}_\rho)$ seems to depend on ρ itself, not just on the representation (conjugacy class of homomorphisms) determined by ρ. We leave it to the reader to come to terms with this.

1.3. ***Harmonic vector fields on the hyperbolic plane.*** Let $f : M \to N$ be a smooth map between Riemannian manifolds. The map f has a "Laplacian" $\tau(f)$ which is a section, defined on M, of the vector bundle $f^*(TN)$. It has a coordinate-free definition as the "trace" of the total second derivative of f. (The total second derivative is a fiberwise bilinear map over M from $TM \times_M TM$ to $f^*(TN)$.) This definition comes from [Eells and Sampson 1964]. The letter τ stands for *tension* more than for *trace*. Following [loc. cit.], the map f is considered *harmonic* if $\tau(f)$ is everywhere zero. There is also a characterization of harmonic maps as critical points of the energy functional

$$f \mapsto \int_M \tfrac{1}{2}\|df\|^2\, d\mu, \tag{1.3.1}$$

where μ is the measure on M determined by the Riemannian metric. That characterization needs exegesis if M is not compact. In the case $N = \mathbb{R}$, the Eells–Sampson definition of a harmonic map agrees with the standard definition of a harmonic function on M, and in the case $M = \mathbb{R}$, the harmonic maps are the geodesics in N.

Let ξ be a smooth vector field on M, where M is still a Riemannian manifold. It is always possible to find a smooth flow $(\phi_t : M \to M)_{t\in\mathbb{R}}$ such that

$$\frac{d\phi_t}{dt}\Big|_{t=0} = \xi.$$

The vector field

$$\tau(\xi) := \frac{d\tau(\phi_t)}{dt}\Big|_{t=0}$$

depends only on ξ, and we may view it as an infinitesimal variant of the Eells–Sampson Laplacian for maps. Consequently we say that ξ is *harmonic* if $\tau(\xi)$ is everywhere zero. (Warning: *Harmonic vector field* can mean very different things to different people, but here we use it in the spirit of [Eells and Sampson 1964]. See [Dodson et al. 2002] for some foundational results on harmonic vector fields.)

For us the case where M is an oriented Riemannian 2-manifold with Riemannian metric g is important. In that case M is also a complex 1-manifold. The 3-dimensional real vector bundle E of symmetric $\mathbb{R}$-bilinear forms on TM has a canonical splitting

$$E = E_1 + E_2,$$

where E_1 is a real line bundle and E_2 has a preferred structure of complex (holomorphic) line bundle. Namely, E_1 is the real line subbundle spanned by the everywhere nonzero section g of E, and E_2 is the image of the vector bundle monomorphism which was mentioned before: $\hom_{\mathbb{C}}(TM \otimes_{\mathbb{C}} TM, \mathbb{C}) \to E$ given by $\phi \mapsto \mathrm{Re}(\phi)$. We like to call E_1 the *scalar* summand of E, and E_2 the *trace-free* summand. The following recognition principle for harmonic vector fields appears to be well known, and so we state it without proof. It has an analog for smooth maps between 2-dimensional oriented Riemannian manifolds [Jost 1984, Lemma 1.1; Gerstenhaber and Rauch 1954a; 1954b].

Proposition 1.3.1. *A smooth vector field ξ on the 2-dimensional oriented Riemannian manifold (M, g) is harmonic if and only if the trace-free component of the Lie derivative $\mathscr{L}_\xi(g)$ is holomorphic*; *i.e., if it is* $\mathrm{Re}(\phi)$ *for a* ***holomorphic*** *section ϕ of* $\hom_{\mathbb{C}}(TM \otimes_{\mathbb{C}} TM, \mathbb{C})$.

Remark 1.3.2. The complex line bundle $\hom_{\mathbb{C}}(TM \otimes_{\mathbb{C}} TM, \mathbb{C})$ in Proposition 1.3.1 comes with a preferred hermitian metric. We use this to equip each fiber with a norm. (It is well known that a hermitian inner product on a finite-dimensional complex vector space is determined by its real part, and so by the associated norm. For $z \in M$, the preferred norm on $\hom_{\mathbb{C}}(T_zM \otimes_{\mathbb{C}} T_zM, \mathbb{C})$ is given by $\|f\| := |f(v \otimes v)|$ for $f \in \hom_{\mathbb{C}}(T_zM \otimes_{\mathbb{C}} T_zM, \mathbb{C})$ and $v \in T_zM$ such that $\|v\| = 1$. This does not depend on the choice of v.)

Now we specialize by letting $M = \mathbb{D}$ (with the Poincaré metric, which will still be called g). In order to state our first main result, Theorem I, we introduce some more vocabulary. In the definition that follows, λ is the ordinary Lebesgue measure on $\mathbb{R}^2$ and λ_0 is the unnormalized Haar measure on S^1, so that $\lambda_0(S^1) = 2\pi$.

Definition 1.3.3. Let ξ be a continuous vector field on $\mathbb{D}$ and let ζ be an L^2-vector field (with values in $\mathbb{R}^2$) along S^1. (For the purposes of this definition, ξ and ζ could be regarded as functions from $\mathbb{D}$ and S^1, respectively, to $\mathbb{R}^2$.) We say that ζ is a *distributional boundary* for ξ if for every continuous vector field α on $\mathbb{D} \cup S^1$

the function

$$s \mapsto \int_{z\in\mathbb{D},\ |z|\le s} \xi(z)\cdot\alpha(z)\,d\lambda \tag{1.3.2}$$

defined on [0, 1) has an extension to all of [0, 1] which is differentiable at $s = 1$, with derivative there equal to

$$\int_{S^1} \zeta(z)\cdot\alpha(z)\,d\lambda_0.$$

In this situation, ζ is determined by ξ. If in addition ζ is tangential, which means that $\zeta(z)\cdot z = 0$ for (almost) all $z \in S^1$, then we say that ξ is *boundary controlled*.

The matching condition relating ξ and ζ in Definition 1.3.3 is invariant under the preferred right action(s) of $\mathrm{diff}(\mathbb{D}\cup S^1)$. See Proposition A.2.2. The preferred right action of $\mathrm{diff}(\mathbb{D}\cup S^1)$ on the space of continuous vector fields on $\mathbb{D}$ is given by $(\xi\cdot h)(x) := (dh(x))^{-1}\big(\xi(h(x))\big)$ for $x \in \mathbb{D}$ and $h \in \mathrm{diff}(\mathbb{D}\cup S^1)$. The preferred right action of $\mathrm{diff}(\mathbb{D}\cup S^1)$ on the space of L^2-vector fields on S^1 is similar. It can be enlightening to write $h^*\xi$ and $h^*\zeta$ instead of $\xi\cdot h$ and $\zeta\cdot h$.

Definition 1.3.4. A smooth vector field ξ on $\mathbb{D}$ is *conformal*, respectively *quasi-conformal*, if the trace-free component of $\mathscr{L}_\xi(g)$ is zero everywhere, respectively uniformly bounded in the norm of Remark 1.3.2.

Examples. The conformal vector fields on $\mathbb{D}$ are the holomorphic vector fields. See Lemma A.2.1. Every conformal vector field is harmonic. Every Killing vector field ξ on $\mathbb{D}$ (element of $\mathfrak{g}$) is conformal and boundary controlled. Indeed, ξ has a smooth extension to a vector field on $\mathbb{D}\cup S^1$ whose restriction to S^1 is tangential to S^1.

Theorem I. *For every holomorphic quadratic differential ϕ on $\mathbb{D}$ which is uniformly bounded in the norm of Remark 1.3.2, there exists a smooth and boundary controlled vector field ξ on $\mathbb{D}$ such that the trace-free component of $\mathscr{L}_\xi(g)$ is* $\mathrm{Re}(\phi)$. *In the case where $\phi \equiv 0$, the vector field ξ must be a Killing vector field.*

The theorem can be reformulated as follows. There is a short exact sequence of real vector spaces and $\mathbb{R}$-linear maps

$$0 \to \mathfrak{g} \xrightarrow{\text{(a)}} U_\infty \xrightarrow{\text{(b)}} V_\infty \to 0, \tag{1.3.3}$$

where V_∞ is the space of all holomorphic quadratic differentials on $\mathbb{D}$ which are bounded in the norm of Remark 1.3.2, and U_∞ is the space of all harmonic, boundary controlled and quasiconformal smooth vector fields on $\mathbb{D}$. The map (b) takes $\xi \in U_\infty$ to ϕ, where $\mathrm{Re}(\phi)$ is the trace-free component of $\mathscr{L}_\xi(g)$. The map (a) is an inclusion.

The proof of Theorem I takes up most of Section 2. Using the theorem, we can explain (*without* relying on uniformization theory) how the two descriptions of the tangent space to $\mathscr{T}(\Sigma)$ at the point determined by some $\rho : \Gamma \to \mathrm{isom}_+(\mathbb{D})$ are related. We identify Σ with the orbit space $\mathbb{D}/\rho = \mathbb{D}/\rho(\Gamma)$, so that a holomorphic quadratic differential on Σ is tantamount to a holomorphic quadratic differential ϕ on $\mathbb{D}$ which is invariant under the subgroup $\rho(\Gamma)$ of $\mathrm{isom}_+(\mathbb{D})$. Such a ϕ is automatically bounded! Choose ξ as in Theorem I such that the trace-free component of $\mathscr{L}_\xi(g)$ is $\mathrm{Re}(\phi)$. Then the formula $h \mapsto h^*\xi - \xi$ (for $h \in \rho(\Gamma) \cong \Gamma$) defines a 1-cocycle on Γ with values in $\mathfrak{g}_\rho$. Its class in $H^1(\Gamma; \mathfrak{g}_\rho)$ does not depend on the choice of ξ. In the case where that class is zero, it is easy to see that we can reconsider the choice of ξ so as to make it invariant under $\rho(\Gamma)$. Then ξ descends to a harmonic vector field on Σ. By [Dodson et al. 2002, Theorem 3.1] this implies $\xi \equiv 0$ and so $\phi \equiv 0$. In other words, we have an *injective* linear map from the vector space of holomorphic quadratic differentials on $\Sigma = \mathbb{D}/\rho$ to $H^1(\Gamma; \mathfrak{g}_\rho)$. By dimension counting, it must be a linear isomorphism.

The interesting aspect of Theorem I is that it makes only a boundedness assumption on ϕ, not an assumption of invariance under a discrete subgroup of $\mathrm{isom}_+(\mathbb{D})$. It is reminiscent of universal Teichmüller theory. See, for example, [Markovic and Šarić 2009; Markovic 2017].

1.4. ***Vector fields from boundary data.*** We have a partial converse to Theorem I which is inspired by the Poisson formula, as in [Ransford 1995, Theorem 1.2.4]. It is our second main result.

Theorem II. *There is a unique continuous linear map F from the vector space of continuous tangential vector fields on S^1 to the vector space of continuous vector fields on $\mathbb{D}$ satisfying the following conditions.*

(i) *$F(\xi)$ is harmonic, for every continuous tangential vector field ξ on S^1.*

(ii) *ξ and $F(\xi)$ together make up a continuous vector field on $\mathbb{D} \cup S^1$.*

(iii) *F is equivariant for the actions of* $\mathrm{isom}(\mathbb{D})$ *on domain and codomain.*

Moreover F extends uniquely to a continuous linear map from the vector space of tangential L^2-vector fields on S^1 to the vector space of continuous vector fields on $\mathbb{D}$. In this setting property (ii) *turns into the following*: *$F(\xi)$ is boundary controlled with distributional boundary ξ. Properties* (i) *and* (iii) *remain intact.*

Remark. In the first part of the statement (before “Moreover…”), the vector space W of continuous tangential vector fields on S^1 is equipped with the compact-open C^0 topology. The vector space Y of continuous vector fields on $\mathbb{D}$ is also equipped with the compact-open C^0 topology (throughout). In the second part of the statement, the vector space $\mathcal{X}$ of tangential L^2-vector fields on S^1 is viewed as a Hilbert space. The statement “Moreover F extends…” is slightly imprecise because W is not a subspace of $\mathcal{X}$. There is an inclusion $W \hookrightarrow \mathcal{X}$ of sets which is linear and continuous as a map of topological vector spaces. The image of W in $\mathcal{X}$ is dense; therefore the “extension” of F from W to $\mathcal{X}$ is certainly unique, if it exists. Properties (i) and (iii) are enough to characterize $F : W \to Y$ up to multiplication by a real scalar. We do not know whether properties (i) and (ii) are enough to characterize $F : W \to Y$.

Lemma 1.4.1. *Let $\xi \in \mathfrak{g}$ be a Killing vector field on $\mathbb{D}$. Let ζ be the matching tangential vector field on S^1, so that ξ and ζ together define a smooth vector field on $\mathbb{D} \cup S^1$. Then $F(\zeta) = \xi$.*

See the Appendix for the proof.

As before, let W be the vector space of continuous vector fields on $\mathbb{D}$. If we have $\rho : \Gamma \to \mathrm{isom}_+(\mathbb{D})$, then we have a preferred right action of Γ on W determined by ρ, and we may write W_ρ to specify the action. The following is again well known, and we omit the proof.

Lemma 1.4.2. *The map $H^1(\Gamma; \mathfrak{g}_\rho) \to H^1(\Gamma; W_\rho)$ induced by the inclusion of the Killing vector fields, $\mathfrak{g} \hookrightarrow W$, is zero.*

Proposition 1.4.3. *Let ψ be a continuous vector field on $\mathbb{D}$ such that $h^*\psi - \psi$ is in $\mathfrak{g}$, for all $h \in \rho(\Gamma)$. Then ψ is boundary controlled. The distributional boundary depends only on the 1-cocycle $h \mapsto h^*\psi - \psi$.*

Proof. (This uses Theorem I.) Let C be the cochain complex normally used to define the cohomology groups $H^j(\Gamma; W_\rho)$ for $j \geq 0$, and write $\delta : C^j \to C^{j+1}$ for the differential in C. Let $D \subset C$ be the cochain subcomplex corresponding to $\mathfrak{g} \subset W$. We have $\psi \in C^0$ and we are assuming $\delta\psi \in D^1$. Theorem I and the dimension-counting argument at the end of Section 1.3 imply that for the 1-cocycle $\delta\psi$ in D there exists $\xi \in C^0$, harmonic and boundary controlled, such that $\delta\xi = \delta\psi$. Therefore $\xi - \psi \in C^0$ is a 0-cocycle. This means that it is invariant under Γ. It follows by inspection that $\xi - \psi$ is boundary controlled with distributional boundary zero. Therefore ψ is boundary controlled and has the same distributional boundary as ξ. □

Now we explain very briefly how Theorem II can help us to make the passage from $H^1(\Gamma; \mathfrak{g}_\rho)$ to the vector space of holomorphic quadratic differentials on $\Sigma = \mathbb{D}/\rho$. We begin with some $v \in H^1(\Gamma; \mathfrak{g}_\rho)$. By Lemma 1.4.2, and in the notation used in the proof of Proposition 1.4.3, the class v can be represented by a cocycle (with values in $\mathfrak{g}_\rho$) of the form $\delta\psi$, where ψ is a continuous vector field on $\mathbb{D}$. By Proposition 1.4.3, the vector field ψ is boundary controlled. Let ζ be its distributional boundary. Then $\xi := F(\zeta)$ is a *harmonic* vector field on $\mathbb{D}$. On the basis of Lemma 1.4.1 and Theorem II it is easy to verify that $\delta\xi$ is a 1-cocycle with values in $\mathfrak{g}$. Moreover it agrees with $\delta\psi$ since both $\delta\xi$ and $\delta\psi$ are "matches" for $\delta\zeta$. Therefore ξ can be viewed as an improvement on ψ. The trace-free component of $\mathscr{L}_\xi(g)$ is $\mathrm{Re}(\phi)$ for a quadratic differential ϕ on $\mathbb{D}$ which is holomorphic by Proposition 1.3.1. The quadratic differential ϕ is also invariant under the group $\rho(\Gamma) \subset \mathrm{isom}_+(\mathbb{D})$ because of Lemma 1.4.1 and condition (iii) in Theorem II.

(Remember that $\delta\xi = \delta\psi$.) Therefore we have a holomorphic quadratic differential on $\Sigma = \mathbb{D}/\rho$.

The above procedure based on Theorem II which takes us from $H^1(\Gamma; \mathfrak{g}_\rho)$ to the vector space of holomorphic quadratic differentials on $\mathbb{D}/\rho$ is the inverse of the other one, based on Theorem I. The verification should be mechanical.

2. Constructing harmonic vector fields from quadratic differentials

2.1. ***Harmonic vector fields in isothermal coordinates.*** Suppose that ξ is defined on $U \subset \mathbb{R}^2$ and U is equipped with a Riemannian metric of the form $ds^2 = \lambda(x, y)(dx^2 + dy^2)$.

If the flow $(\phi_t)_{t\in[0,\varepsilon)}$ and the vector field ξ are related as above, then we can describe ϕ_t to first order in terms of ξ:

$$\phi_t(z) \approx z + t\xi(z) \quad \text{for } z \in U.$$

We define a family of Riemannian metrics on U as follows:

$$t \mapsto \rho_t = \phi_t^* g. \tag{2.1.1}$$

More precisely the map in (2.1.1) has the form

$$t \mapsto (D\phi_t : T_pU \to T_{\phi_t(p)}\mathbb{H})^* g. \tag{2.1.2}$$

To the first order, (2.1.2) can be expressed as follows:

$$t \mapsto \big(\mathrm{id} + t \cdot D\xi : T_pU \to T_{\phi_t(p)}\mathbb{H}\big)^* g,$$

where $D\xi$ is the total derivative of ξ (the latter being viewed as a smooth map from an open set in $\mathbb{R}$ or $\mathbb{C}$ to $\mathbb{R}^2$ or $\mathbb{C}$). Continuing in this manner, we get

$$\begin{aligned}\rho_t &\approx (\mathrm{id} + t \cdot D\xi)^T (g + t \cdot Dg(\xi))\,(\mathrm{id} + t \cdot D\xi)\\ &\approx g + t \cdot D\xi^T g + tDg(\xi) + t \cdot D\xi \cdot g\\ &= g + (t \cdot D\xi^T + t \cdot D\xi) \cdot g + tDg(\xi).\end{aligned}$$

Calculating

$$\left.\frac{d\rho_t}{dt}\right|_{t=0}$$

gives us a section of the vector bundle of (real) symmetric bilinear forms on TU and this is denoted by $\mathscr{L}_\xi(g)$, the Lie derivative of g along ξ. Therefore,

$$\mathscr{L}_\xi g = (D\xi^T + D\xi)g + Dg(\xi) \tag{2.1.3}$$

in our preferred coordinates. We can write $D\xi$ in matrix form,

$$D\xi = \begin{bmatrix} \xi_x^1 & \xi_y^1 \\ \xi_x^2 & \xi_y^2 \end{bmatrix}.$$

Then (2.1.3) turns into

$$\mathscr{L}_\xi(g) = \lambda \begin{bmatrix} 2\xi_x^1 & \xi_y^1+\xi_x^2 \\ \xi_x^2+\xi_y^1 & 2\xi_y^2 \end{bmatrix} + \begin{bmatrix} \langle D\lambda, \xi\rangle & 0 \\ 0 & \langle D\lambda, \xi\rangle \end{bmatrix}$$

$$= \underbrace{\lambda \begin{bmatrix} \xi_x^1-\xi_y^2 & \xi_y^1+\xi_x^2 \\ \xi_y^1+\xi_x^2 & \xi_y^2-\xi_x^1 \end{bmatrix}}_{\text{TF}} + \lambda \begin{bmatrix} \xi_x^1+\xi_y^2 & 0 \\ 0 & \xi_x^1+\xi_y^2 \end{bmatrix} + \begin{bmatrix} \langle D\lambda, \xi\rangle & 0 \\ 0 & \langle D\lambda, \xi\rangle \end{bmatrix}.$$

Recall from Section 1.3 that the vector bundle E of (real) symmetric $\mathbb{R}$-bilinear forms on TU splits into a 1-dimensional real vector subbundle (the scalar summand) spanned by the everywhere nonzero section g and a 1-dimensional complex line bundle (the trace-free summand) which is the image of the embedding

$$\hom_{\mathbb{C}}(TU \otimes_{\mathbb{C}} TU, \mathbb{C}) \to E$$

taking ϕ to $\operatorname{Re}(\phi)$. Since the local coordinates that we are using here are in agreement with the conformal structure determined by the Riemannian metric g, our calculation implies that the trace-free component of $\mathscr{L}_\xi(g)$ is the summand with the label TF,

$$\lambda \begin{bmatrix} \xi_x^1-\xi_y^2 & \xi_y^1+\xi_x^2 \\ \xi_y^1+\xi_x^2 & \xi_y^2-\xi_x^1 \end{bmatrix}. \tag{2.1.4}$$

Indeed, TF is $\operatorname{Re}(f\cdot(dz\otimes dz))$, for $f := \text{TF}_{11} - \iota\text{TF}_{12}$, where $\iota = \sqrt{-1}$.

Proposition 2.1.1. *Let ξ be a smooth vector field on $U \subset \mathbb{R}^2$, and let g be a Riemannian metric on U of the form $ds^2 = \lambda(x,y)(dx^2+dy^2)$. Then the trace-free component of $\mathscr{L}_\xi(g)$ is* $\operatorname{Re}(f\cdot(dz\otimes_{\mathbb{C}} dz))$, *where $f : U \to \mathbb{C}$ is determined by*

$$\overline{f(z)} = 2\lambda \frac{\partial\xi}{\partial\bar z}(z). \tag{2.1.5}$$

Proof. Most of this has already been established. The formula for $\overline{f(z)}$ needs to be unraveled. The Wirtinger derivative (for differentiable functions from U to $\mathbb{C}$) is

$$\partial/\partial\bar z = \tfrac{1}{2}(\partial/\partial x + \iota\,\partial/\partial y), \tag{2.1.6}$$

where we allow ourselves to write $z = x+\iota y$. In particular

$$2\lambda\frac{\partial\xi}{\partial\bar z} = \lambda\cdot(\xi_x + \iota\xi_y).$$

It is understood that $\xi_x = \xi_x^1 + \iota\xi_x^2$ and $\xi_y = \xi_y^1 + \iota\xi_y^2$, so that we obtain

$$2\lambda\frac{\partial\xi}{\partial\bar z} = \lambda(\xi_x^1 + \iota\xi_x^2 - \xi_y^2 + \iota\xi_1^y) = \lambda(\xi_x^1 - \xi_y^2 + \iota(\xi_x^2+\xi_y^1)) = \text{TF}_{11} + \iota\text{TF}_{12},$$

which is conjugate to $\text{TF}_{11} - \iota\text{TF}_{12}$, our definition of f, given just after (2.1.4). □

Remark 2.1.2. Think of Proposition 2.1.1 as a complement to Proposition 1.3.1.

2.2. ***Solving the potential equation.*** Taking $U = \mathbb{H}$ and $\lambda(x, y) = y^{-2}$ in (2.1.5), we obtain

$$\overline{f(z)} = \frac{-8}{(z-\bar{z})^2}\frac{\partial\xi}{\partial\bar{z}}(z). \tag{2.2.1}$$

This is well known in Teichmüller theory. Scott Wolpert [1987, §2.1] calls it *the potential equation*. He has it in the form

$$(z-\bar{z})^2\,\overline{f(z)} = -8\frac{\partial\xi}{\partial\bar{z}}(z) \tag{2.2.2}$$

except for the factor -8 in the right-hand side, which he does not have. (One may ask why he does not have it. We believe this is explained by different conventions, e.g., regarding the preferred metric on $\mathbb{H}$.) He calls the left-hand side a *harmonic Beltrami differential*, and probably he views both sides as Beltrami differentials, consciously or unconsciously multiplying both sides with a standard Beltrami differential denoted $d\bar{z}/dz$ for better or worse.

Wolpert [1987, §2.3] has a very elegant solution for the potential equation. (Here we assume that f is "known", defined on all of $\mathbb{H}$ and holomorphic. The vector field ξ on $\mathbb{H}$ is the unknown. By Proposition 2.1.1 it must be a harmonic vector field.) His solution, in our notation, is

$$\xi(z) = -\frac{1}{8}\,\overline{\int_{z_0}^{z}(\bar{z}-s)^2 f(s)\,ds}, \tag{2.2.3}$$

where $z_0 \in \mathbb{H}$ is fixed. The integral is a complex path integral, along some path in $\mathbb{H}$ from z_0 to z. Since the integrand is holomorphic, the value of the integral is independent of the path selected. He writes: *the potential equation* (2.2.2) *is an immediate consequence of differentiation under the integral*. He is right.

For us, (2.2.3) is not the perfect solution. We need a solution which extends to the ideal boundary of $\mathbb{H}$ under conditions on f which we find reasonable. Therefore we will adopt a more laborious method to solve (2.2.2).

Let η be the constant vector field on $\mathbb{H}$ defined by $\eta(z) = 1 \in \mathbb{C}$ for all z.

Lemma 2.2.1. *Let $f : \mathbb{H} \to \mathbb{C}$ be a holomorphic function and let c be a positive real number. The vector field ξ_c defined by*

$$\xi_c(z) = \left(\int_{\mathrm{Im}(z)}^{c} \iota t^2\,\overline{f(\bar{z}+2\iota t)}\,dt\right)\eta(z)$$

for z subject to $\mathrm{Im}(z) < 2c$ *solves* (2.2.2).

We have to assume $\mathrm{Im}(z) < 2c$ in the definition of $\xi_c(z)$ to ensure that $f(\bar{z}+2\iota t)$ is defined for all $t \in [c, \mathrm{Im}(z)]$.

Proof. Write ξ instead of ξ_c in this proof (and drop the constant factor η, so that ξ becomes a function with values in $\mathbb{C}$). Write D for the total derivative acting on such functions, so that the values of D are real 2×2-matrices. Write $\Phi(z, t)$ for the integrand in the definition of ξ.

The function ξ is a composition $u \circ v$, where $v(z) = (\mathrm{Im}(z), z)$ for $z \in \mathbb{H}$ and u is a function of a real and a complex variable:

$$u(r, z) = \int_r^c \Phi(z, t)\, dt.$$

Applying the chain rule in this situation gives

$$D\xi(z) = -\Phi(z, \mathrm{Im}(z)) \cdot \begin{bmatrix} 0 & 1 \end{bmatrix} + \int_{\mathrm{Im}(z)}^c D(\Phi(z, t))\, dt \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{2.2.4}$$

where $\Phi(z, \mathrm{Im}(z))$ must be viewed as a real $2{\times}1$-matrix. The Wirtinger derivative $\partial/\partial\bar{z}$ is $Q(D)$ for a certain linear map W_2 acting on real $2{\times}2$-matrices,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \frac{1}{2} \begin{bmatrix} a - d \\ b + c \end{bmatrix}.$$

Therefore

$$\frac{\partial\xi}{\partial\bar{z}}(z) = W_2\big(-\Phi(z, \mathrm{Im}(z)) \cdot \begin{bmatrix} 0 & 1 \end{bmatrix}\big) + \int_{\mathrm{Im}(z)}^c W_2(D(\Phi(z, t))\, dt.$$

But $\Phi(z, t)$ is a holomorphic function of z, so that $W_2(D(\Phi(z, t))$ is everywhere zero. So we obtain

$$W_2(D\xi(z)) = W_2\big(-\Phi(z, \mathrm{Im}(z)) \cdot \begin{bmatrix} 0 & 1 \end{bmatrix}\big).$$

Now $\Phi(z, \mathrm{Im}(z)) = \iota y^2 \overline{f(z)}$, where we have written y for $\mathrm{Im}(z)$. Therefore

$$-\Phi(z, \mathrm{Im}(z)) \cdot \begin{bmatrix} 0 & 1 \end{bmatrix} = y^2 \begin{bmatrix} 0 & -\mathrm{Im}(f(z)) \\ 0 & -\mathrm{Re}(f(z)) \end{bmatrix},$$

and so

$$\frac{\partial\xi}{\partial\bar{z}}(z) = W_2(D\xi(z)) = \frac{1}{2} y^2 \begin{bmatrix} \mathrm{Re}(f(z)) \\ -\mathrm{Im}(f(z)) \end{bmatrix} = \frac{1}{2} y^2\, \overline{f(z)} = -\frac{1}{8}(z - \bar{z})^2\, \overline{f(z)}. \quad \square$$

Theorem 2.2.2. *Let f be a holomorphic function on $\mathbb{H}$ which satisfies the condition $|f(z)| \cdot \mathrm{Im}(z)^2 \leq b_0$, where b_0 is a positive constant, $z \in \mathbb{H}$ arbitrary. Then the formula*

$$\xi^{\mathrm{reg}}(z) := \lim_{c \to \infty} \left(\xi_c(z) - \left(\xi_c(\iota) + \frac{\partial \xi_c}{\partial z}(\iota) \cdot (z - \iota) \right) \right), \tag{2.2.5}$$

where ξ_c is as in Lemma 2.2.1, defines a smooth vector field ξ^{reg} on $\mathbb{H}$ which solves (2.2.2).

Remarks. The superscript reg is for *regularization*, the art of making divergent things convergent. The formula for ξ^{reg} uses the Wirtinger derivative $\partial/\partial z = \frac{1}{2}(\partial/\partial x - \iota\,\partial/\partial y)$. The "regularizing" term $\xi_c(\iota) + (\partial\xi_c/\partial z)(\iota)\cdot(z-\iota)$ which we subtract from $\xi_c(z)$ can be regarded as the holomorphic part of the first Taylor approximation to ξ_c at $z = \iota$.

The function $z \mapsto |f(z)|\cdot \text{Im}(z)^2$ on $\mathbb{H}$ is bounded if and only if the quadratic differential $f\cdot(dz \otimes_{\mathbb{C}} dz)$ on $\mathbb{H}$ is bounded in the metric of $\mathbb{H}$ and Remark 1.3.2. Indeed the pointwise norm of the 1-form dz at $z_0 \in \mathbb{H}$ is $\text{Im}(z_0)$, and the pointwise norm of $dz \otimes_{\mathbb{C}} dz$ at z_0 is therefore $\text{Im}(z_0)^2$.

The vector field ξ^{reg} is automatically harmonic if it solves (2.2.2), since f is holomorphic by assumption.

The proof of Theorem 2.2.2 is quite long. We begin by isolating some technicalities and generalities.

Lemma 2.2.3. *If $f : \mathbb{H} \to \mathbb{C}$ is holomorphic and satisfies $|f(z)|\cdot\text{Im}(z)^2 \le b_0$ for all $z \in \mathbb{H}$, where b_0 is a positive constant, then there exist positive constants $b_1, b_2, b_3, \ldots$ such that $|f^{(n)}(z)|\cdot \text{Im}(z)^{n+2} \le b_n$ for all $z \in \mathbb{H}$ and $n = 1, 2, 3, \ldots$.*

Proof. Fix $z_0 \in \mathbb{H}$ and let γ be a smooth curve describing (counterclockwise) a circle of radius $r = \text{Im}(z_0)/2$ about z_0. Then by the Cauchy integral formulas,

$$f^{(n)}(z_0) = \frac{n!}{2\pi\iota}\oint_\gamma \frac{f(z)}{(z-z_0)^{n+1}}\,dz.$$

The values of $|f|$ on the circle are bounded by $b_0\cdot(\text{Im}(z_0)/2)^{-2}$. The circumference of the circle is $\pi\,\text{Im}(z_0)$. Therefore

$$\begin{aligned}|f^{(n)}(z_0)| &\le \frac{n!}{2\pi}(\pi\,\text{Im}(z_0))\cdot b_0(\text{Im}(z_0)/2)^{-2}\cdot(\text{Im}(z_0)/2)^{-(n+1)}\\ &= 2^{n+2}n!\,\pi\, b_0(\text{Im}(z_0))^{-(n+2)}.\end{aligned}$$

We can take $b_n = 2^{n+2}n!\,\pi\, b_0$. □

Lemma 2.2.4 (notation of Lemma 2.2.1). *The Wirtinger derivative $\partial/\partial\bar{z}$ of ξ_c is independent of c, where defined.*

Proof. Make two choices for c, say c_1 and c_2, where $c_2 > c_1 > \frac{1}{2}$. Then

$$\xi_{c_2}(z) - \xi_{c_1}(z) = \left(\int_{c_1}^{c_2} \iota t^2\,\overline{f(\bar{z}+2\iota t)}\,dt\right)\eta(z)$$

is a holomorphic vector field. It is defined for z with $\text{Im}(z) < 2c_1$. □

Remark 2.2.5. Let $\lambda : U \to \mathbb{C}$ be a differentiable function, where U is open in $\mathbb{C}$. The Wirtinger derivatives $\partial/\partial z$ and $\partial/\partial\bar{z}$ of λ together determine the total derivative $D\lambda$. The relationship is

$$D\lambda = \frac{\partial\lambda}{\partial z} + \frac{\partial\lambda}{\partial\bar{z}} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

This is meant as an equation between functions taking values in the algebra of real 2×2-matrices. (A complex number determines a real 2×2-matrix since multiplication by that number is an $\mathbb{R}$-linear operator on $\mathbb{R}^2$.)

Proposition 2.2.6. *Let $(\zeta_n)_{n\geq 0}$ be a sequence of harmonic vector fields defined on an open set $U \subset \mathbb{H}$, and converging in the compact-open C^0 topology to a vector field ζ_∞. Suppose that the trace-free component of $\mathscr{L}_{\zeta_n}(g)$ is the same for all n. Then the sequence (ζ_n) converges to ζ_∞ in the compact-open C^∞-topology. Hence ζ_∞ is harmonic, and the trace-free component of $\mathscr{L}_{\zeta_\infty}(g)$ agrees with the trace-free component of $\mathscr{L}_{\zeta_n}(g)$ for all n.*

Proof. The sequence $(\zeta_n - \zeta_0)_{n\geq 0}$ is a sequence of holomorphic vector fields which can also be viewed as a sequence of holomorphic functions. Therefore the Weierstrass convergence theorem applies. □

Proof of Theorem 2.2.2. Write D for total derivatives (of $\mathbb{C}$-valued functions, with respect to a variable $z \in \mathbb{H}$ or $z \in \mathbb{C}$). Write $\kappa(z) = \max\{1, \mathrm{Im}(z)\}$. As in the proof of Lemma 2.2.1 write

$$\Phi(z,t) := \iota t^2\, \overline{f(\bar{z} + 2\iota t)}.$$

We do not insist on the distinction between vector fields (on open subsets of $\mathbb{H}$) and $\mathbb{C}$-valued functions, because it is irrelevant here. In particular ξ_c will be viewed as a function. The first step is to show

$$\xi_c(z) - \left(\xi_c(\iota) + \frac{\partial\xi_c}{\partial z}(\iota)\cdot(z-\iota)\right) = \int_{\kappa(z)}^{c} \Phi(z,t) - \Phi(\iota,t) - D\Phi(\iota,t)\cdot(z-\iota)\,dt + F(z), \tag{2.2.6}$$

where F is a continuous function of the variable z, independent of c. As a matter of language, we might say (in this proof) that two continuous functions of z and c are *equivalent* if their difference is a function of z only. Then $(z,c) \mapsto \xi_c(z)$ is equivalent to

$$(z,c) \mapsto \int_{\kappa(z)}^{c} \Phi(z,t)\,dt$$

and $(z,c) \mapsto \xi_c(\iota)$ is equivalent to

$$(z,c) \mapsto \int_{\kappa(z)}^{c} \Phi(\iota,t)\,dt.$$

By Lemma 2.2.4 and Remark 2.2.5, the function $(z, c) \mapsto (\partial\xi_c/\partial z)(\iota) \cdot (z - \iota)$ is equivalent to $(z, c) \mapsto D\xi_c(\iota) \cdot (z - \iota)$. As in the proof of Lemma 2.2.1 we have

$$D\xi_c(\iota) = -\Phi(\iota, 1) \cdot \begin{bmatrix} 0 & 1 \end{bmatrix} + \int_1^c D(\Phi(\iota, t))\, dt,$$

so that $(z, c) \mapsto D\xi_c(\iota) \cdot (z - \iota)$ is equivalent to

$$(z, c) \mapsto \int_{\kappa(z)}^c D(\Phi(\iota, t)) \cdot (z - \iota)\, dt.$$

Using all these equivalences, we obtain (2.2.6). And now the second step is clearly to show that the improper integral

$$\int_{\kappa(z)}^{\infty} \Phi(z, t) - \big(\Phi(\iota, t) + D\Phi(\iota, t) \cdot (z - \iota)\big)\, dt \tag{2.2.7}$$

exists. It is a good idea to think of the expression $\Phi(\iota, t) + D\Phi(\iota, t) \cdot (z - \iota)$ as the first Taylor polynomial at $z = \iota$ of the function $z \mapsto \Phi(z, t)$, for fixed t. The "integral form of the remainder" in Taylor's formula gives us

$$\Phi(z, t) - \big(\Phi(\iota, t) + D\Phi(\iota, t) \cdot (z - \iota)\big) = \frac{1}{2} \int_0^1 u''(s) \cdot (1 - s)\, ds,$$

where $u(s) := \Phi(\iota + s(z - \iota), t) = \iota t^2\, \overline{f(-\iota + s(\bar{z} + \iota) + 2\iota t)}$ for fixed t and z. Now Lemma 2.2.3 gives the estimate

$$|u''(s)| \le t^2 \frac{b_2 |z - \iota|^2}{|\iota + s(z - \iota) - 2\iota t|^4}.$$

Here $s \in [0, 1]$, and z is fixed, and b_2 is a positive constant. If $t \ge 1 + |z|$, then we may conclude

$$|u''(s)| \le \frac{b_2 |z - \iota|^2}{t^2}.$$

Therefore if $1 + |z| \le A_1 \le A_2$, then

$$\begin{aligned} &\left| \int_{A_1}^{A_2} \Phi(z, t) - \big(\Phi(\iota, t) + D\Phi(\iota, t) \cdot (z - \iota)\big)\, dt \right| \\ &\qquad \le \int_{A_1}^{A_2} \frac{1}{2} \int_0^1 \frac{b_2 |z - \iota|^2}{t^2} \cdot (1 - s)\, ds\, dt \\ &\qquad = \frac{1}{4} b_2 |z - \iota|^2 \int_{A_1}^{A_2} t^{-2}\, dt = \frac{1}{4} b_2 |z - \iota|^2 (A_1^{-1} - A_2^{-1}) \le \frac{b_2 |z - \iota|^2}{4A_1}. \end{aligned}$$

This means that the improper integral (2.2.7) converges. In other words, $\xi^{\mathrm{reg}}(z)$ is well defined by means of formula (2.2.5) for every $z \in \mathbb{H}$. But we have shown more: the convergence is uniform on compact sets. More precisely, writing

$$V(c,z) := \xi_c(z) - \left(\xi_c(\iota) + \frac{\partial \xi_c}{\partial z}(\iota)\cdot(z-\iota)\right),$$

we have shown that

$$|V(c_1,z) - V(c,z)| \le \frac{b_2|z-\iota|^2}{4c} \tag{2.2.8}$$

under the condition $c_1 \ge c > 1+|z|$. Now Proposition 2.2.6 can be applied. It follows that ξ^{reg} is a harmonic vector field on $\mathbb{H}$ and a solution to (2.2.2). □

2.3. ***Boundary values.*** The standard isometry from the upper half-plane model $\mathbb{H}$ of the hyperbolic plane to the disk model $\mathbb{D}$ is the Cayley transform C, defined by $C(z) := (z-\iota)/(z+\iota)$. The (complex) derivative of C is $z \mapsto (z+\iota)^{-2}$.

Theorem 2.3.1. *The vector field $C_*(\xi^{\text{reg}})$ on $\mathbb{D}$ obtained by applying the Cayley transform to ξ^{reg} of Theorem 2.2.2 admits a continuous extension to $\mathbb{D}\cup S^1$. The value of the extended vector field at $1 \in S^1$ is* 0*.*

Remark 2.3.2. The restriction to S^1 of the extended vector field on $\mathbb{D}\cup S^1$ is not claimed to be tangential to S^1. From our point of view this is an issue. It will be addressed in Section 2.4, which has no other purpose.

As a preparation for the proof, we translate the statement so that we can continue to work in the upper half-plane setting. It is convenient to show first that ξ^{reg} extends to a continuous vector field defined on all of $\mathbb{H}\cup\mathbb{R}$. Under the Cayley transform this corresponds to a continuous vector field on $\mathbb{D}\cup S^1 \smallsetminus \{1\}$. We extend this to all of $\mathbb{D}\cup S^1$ by *defining* the missing value (at $1\in S^1$) to be 0. Then we still have to establish continuity at $1 \in \mathbb{D}\cup S^1$. By the formula for the derivative of C, this is equivalent to the following statement:

$$\lim_{z\in\mathbb{H},\, |z|\to\infty} \frac{\xi^{\text{reg}}(z)}{|z|^2} = 0. \tag{2.3.1}$$

Proof of Theorem 2.3.1. First we show that ξ^{reg} extends to a continuous vector field on $\mathbb{H}\cup\mathbb{R}$. Select some $z\in\mathbb{R}$. The integral

$$\int_{\text{Im}(z)}^{c} \iota t^2\, \overline{f(\bar z + 2\iota t)}\, dt \tag{2.3.2}$$

is an improper integral because the integrand is not defined for $t = 0 = \text{Im}(z)$. But for $t > 0$ the integrand is defined and moreover

$$|\iota t^2 \overline{f(\bar z+2\iota t)}| \le |\bar z + 2\iota t|^2 |f(\bar z+2\iota t)| \le b_0$$

by our condition on f. It follows that, for $z\in\mathbb{R}$, the improper integral (2.3.2) converges. So $\xi_c(z)$ is defined or definable for *all* z such that $0 \le \text{Im}(z) < 2c$, and is continuous as a function of z. Inequality (2.2.8) remains meaningful and valid if we allow $z \in \mathbb{H} \cup \mathbb{R}$. Therefore $\xi^{\text{reg}}(z)$ has a continuous extension to all of $\mathbb{H} \cup \mathbb{R}$.

It remains to prove the claim (2.3.1). Fix some $z \in \mathbb{H}$. We will use (2.2.8), but we have some freedom in choosing c, and we decide $c = 2 + |z|$. This gives

$$\left|\xi^{\text{reg}}(z) - \xi_c(z) - \left(\xi_c(\iota) + \frac{\partial \xi_c}{\partial z}(\iota) \cdot (z - \iota)\right)\right| \le \frac{b_2|z-\iota|^2}{4c} = \frac{b_2|z-\iota|^2}{8+4|z|}.$$

Since the right-hand side of this, divided by $|z|^2$, tends to zero for $|z| \to \infty$, the same is true for the left-hand side. Therefore it is enough to show that the fraction

$$\frac{\xi_c(z) - \big(\xi_c(\iota) + (\partial \xi_c/\partial z)(\iota) \cdot (z - \iota)\big)}{|z^2|}, \tag{2.3.3}$$

evaluated *only* on pairs (c, z) where $c = 2 + |z|$, tends to zero for $|z| \to \infty$. By Lemma 2.2.3 and elementary integration (see Remark 2.3.3 for more details), the following estimates are available:

$$|\xi_c(z)| < A \cdot |z|, \quad |\xi_c(\iota)| < A \cdot |z|, \quad \left|\frac{\partial \xi_c}{\partial z}(\iota)\right| < A \ln|z| + B$$

for some positive constants A and B. Therefore

$$\left|\frac{\partial \xi_c}{\partial z}(\iota) \cdot (z - \iota)\right| < (A \ln|z| + B) \cdot (|z| + 1).$$

Using these estimates in (2.3.3), we see that it does tend to zero for $|z| \to \infty$. □

Remark 2.3.3. Estimate for $\xi_c(z)$: We have

$$\xi_c(z) = \int_{\text{Im}(z)}^{c} \iota t^2\, \overline{f(\bar{z} + 2\iota t)}\, dt,$$

where $|f(\bar{z} + 2\iota t)| \le b_0|\bar{z} + 2\iota t|^{-2}$. Since $t \in [\text{Im}(z), c]$, where $c = 2 + |z|$, we have $t \ge \text{Im}(z)$ and so $|\bar{z} + 2\iota t| \ge \text{Im}(\bar{z} + 2\iota t) = 2t - \text{Im}(z) \ge t$. Therefore

$$|\xi_c(z)| \le \int_{\text{Im}(z)}^{c} t^2 b_0 |\bar{z} + 2\iota t|^{-2}\, dt \le b_0 \int_{\text{Im}(z)}^{c} 1\, dt = b_0(c - \text{Im}(z)) \le b_0(2 + |z|).$$

Estimate for $\xi_c(\iota)$: It must be remembered that $c = 2 + |z|$, where z has little to do with ι. We have

$$\xi_c(\iota) = \int_1^c \iota t^2\, \overline{f(\bar{\iota} + 2\iota t)}\, dt,$$

where $|f(\bar{\iota} + 2\iota t)| \le b_0|2t - 1|^{-2}$. Since $t \ge 1$ we have $|2t - 1| \ge t$ and so

$$|\xi_c(\iota)| \le \int_1^c t^2 b_0 |\bar{\iota} + 2\iota t|^{-2}\, dt \le b_0 \int_1^c 1\, dt = b_0(c - 1) = b_0(1 + |z|).$$

Estimate for $(\partial\xi_c/\partial z)(\iota)$: Again it must be remembered that $c = 2 + |z|$, where z has little to do with ι. We try the total derivative $D\xi_c$ first. In the notation and conventions of (2.2.4) we have

$$D\xi_c(\iota) = -\Phi(\iota, 1) \cdot \begin{bmatrix} 0 & 1 \end{bmatrix} + \int_1^c D(\Phi(\iota, t))\, dt,$$

which means

$$D\xi_c(\iota) = -\iota\overline{f(\iota)} \cdot \begin{bmatrix} 0 & 1 \end{bmatrix} + \int_1^c \iota t^2 (Df_t)(\iota)\, dt, \tag{2.3.4}$$

where f_t is the (holomorphic) function $z \mapsto \overline{f(\bar{z} + 2\iota t)}$. We note therefore that the second summand in the right-hand side of (2.3.4) will contribute in full to $(\partial\xi_c/\partial z)(\iota)$; the other one may not contribute in full but it will make a constant contribution (independent of c). Therefore we have to estimate the second summand only. By the chain rule, $(Df_t)(\iota) = (Df_0)(\iota - 2\iota t)$. This is a complex number (which can also be viewed as a real 2×2-matrix). Now $|(Df_0)(\iota - 2\iota t)| = |Df(2\iota t - \iota)|$ and by Lemma 2.2.3 we have $|Df(2\iota t - \iota)| =\leq b_1|2t - 1|^{-3} \leq b_1 t^{-3}$. Therefore

$$\left| \int_1^c \iota t^2 (Df_t)(\iota)\, dt \right| \leq \int_1^c b_1 t^{-1}\, dt = b_1 \ln c = b_1 \ln(2 + |z|).$$

Remark 2.3.4. It seems to us that Wolpert's solution (2.2.3) of (2.2.2) admits a continuous extension to $\mathbb{H} \cup \mathbb{R}$ if f satisfies the conditions of Theorem 2.2.2. But we were unable to show that it has good growth behavior for $z \to \infty$, e.g., similar to (2.3.1), especially where z approaches ∞ on a horizontal through z_0.

2.4. ***Hardy space to the rescue.*** Let $\mathcal{H} = L^2(S^1, \mathbb{C})$ be the complex Hilbert space of square integrable complex valued functions on S^1. The hermitian inner product is

$$\langle v, w \rangle = \frac{1}{2\pi} \int_0^{2\pi} \langle v(e^{\iota t}), w(e^{\iota t}) \rangle\, dt$$

for $v, w \in \mathcal{H}$. Consequently $\mathcal{H}$ has a standard Hilbert basis consisting of the vectors $v_k := (z \mapsto z^k)$ for $k \in \mathbb{Z}$. It has an orthogonal splitting $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$, where $\mathcal{H}_1$ is the closure of the span of the v_k for $k \geq 0$, and $\mathcal{H}_2$ is the closure of the span of the v_k for $k < 0$. The closed subspace $\mathcal{H}_1$ is known as the *Hardy* space. Each $w \in \mathcal{H}_1$ has a canonical "extension" to a function $w^e : \mathbb{D} \cup S^1 \to \mathbb{C}$ which is holomorphic in $\mathbb{D}$. Namely, if $w = \sum_{k=0}^{\infty} a_k v_k$ in $\mathcal{H}_1$, then the power series $\sum_{k=0}^{\infty} a_k z_k$ converges locally uniformly in $\mathbb{D}$ and represents a holomorphic function there. Note that w can be recovered from the restriction of w^e to $\mathbb{D}$. It is the limit (in the metric of $\mathcal{H}_1$) of the functions $z \mapsto w^e(sz)$, for $z \in S^1$, as $s \in [0, 1)$ tends to 1.

Let $\mathcal{X} \subset \mathcal{H}$ be the closed $\mathbb{R}$-linear subspace consisting of the functions on S^1 which have the form $z \mapsto \iota z \cdot u(z)$, where u is a *real*-valued square-integrable function on S^1. It is an exercise in linear algebra to show that $\mathcal{X} + \mathcal{H}_1 = \mathcal{H}$ and $\dim_{\mathbb{R}}(\mathcal{X} \cap \mathcal{H}_1) = 3$. In fact $\mathcal{X} \cap \mathcal{H}_1$ has an orthonormal (real) basis consisting of the functions

$$z \mapsto \iota z, \quad z \mapsto \tfrac{1}{2}\iota z \cdot \iota(z^{-1} - z), \quad z \mapsto \tfrac{1}{2}\iota z \cdot (z^{-1} + z).$$

This fact has a more illuminating formulation if we think of $\mathcal{H}$ as the vector space of L^2-vector fields on S^1 (where the vectors in the vector fields are elements of $\mathbb{C}$). Then $\mathcal{X}$ is precisely the space of *tangential* L^2-vector fields on S^1. The more illuminating formulation is as follows.

Lemma 2.4.1. $\mathcal{X} + \mathcal{H}_1 = \mathcal{H}$ *and* $\mathcal{X} \cap \mathcal{H}_1 = \mathfrak{g}$, *where* $\mathfrak{g}$ *is the space of Killing vector fields on* S^1 (see explanation below).

It makes sense to speak of Killing vector fields on $\mathbb{D}$; these are the "infinitesimal isometries" of $\mathbb{D}$ with the Poincaré metric. They form a 3-dimensional Lie algebra $\mathfrak{g}$ under the Lie product, which we identify with the Lie algebra of $\mathrm{isom}_+(\mathbb{D})$. It is well known that the Killing vector fields extend to smooth vector fields on $\mathbb{D} \cup S^1$ which are tangential to S^1. Restricting these to S^1 is a faithful operation, and so we may allow ourselves occasionally to think of $\mathfrak{g}$ as a 3-dimensional Lie algebra of tangential smooth vector fields on S^1. (The Lie product will not be of any importance here.)

Proof of Theorem I in Section 1. Let h be the Poincaré metric on $\mathbb{D}$. Let ϕ be a quadratic differential on $\mathbb{D}$ which is uniformly bounded in the norm of Remark 1.3.2. By Theorems 2.2.2 and 2.3.1, there exists a continuous vector field ζ on $\mathbb{D} \cup S^1$ which is smooth on $\mathbb{D}$, and such that the trace-free component of $\mathscr{L}_\zeta(h)$ is ϕ. (In more detail: Let $C^*(\phi)$ be the pullback of ϕ under the Cayley transform C, a holomorphic quadratic differential on $\mathbb{H}$. Write $C^*(\phi) = f \cdot (dz \otimes_{\mathbb{C}} dz)$, so that f is holomorphic and satisfies the conditions of Theorem 2.2.2. Find ξ^{reg} as in the said theorem. Let $C_*(\xi^{\mathrm{reg}})$ be the corresponding vector field on $\mathbb{D}$. Let ζ be the extension of $C_*(\xi^{\mathrm{reg}})$ to $\mathbb{D} \cup S^1$ which exists by Theorem 2.3.1.)

The vector field ζ need not be tangential along S^1. But by Lemma 2.4.1 there exists $\psi \in \mathcal{H}_1$ such that

$$\zeta|_{S^1} - \psi \in \mathcal{X}.$$

Because ψ belongs to the Hardy space $\mathcal{H}_1$, it has a canonical extension ψ^e to $\mathbb{D} \cup S^1$ which is holomorphic on $\mathbb{D}$. It turns out that $\xi := \zeta - \psi^e$, restricted to $\mathbb{D}$, is a smooth and boundary-controlled vector field on $\mathbb{D}$ such that the trace-free component of $\mathcal{L}_\xi(g)$ is $\mathrm{Re}(\phi)$. We will verify the conditions one by one.

(i) The trace-free component of the Lie derivative of the Riemannian metric h along $\zeta - \psi^e$ is ϕ because the trace-free component of $\mathscr{L}_\zeta(h)$ is ϕ and ψ^e is holomorphic on $\mathbb{D}$.

(ii) The vector field $\zeta - \psi^e$ (restricted to $\mathbb{D}$) is boundary controlled because we have the extension to $\mathbb{D} \cup S^1$ by construction, and the extension is tangential along S^1 by construction. The matching condition which must be satisfied by the restrictions of $\zeta - \psi$ to $\mathbb{D}$ and S^1, respectively, is indeed satisfied because it is satisfied separately for ζ and ψ^e.

It remains to understand what happens if $\phi \equiv 0$. In this case, ξ must be a holomorphic vector field on $\mathbb{D}$. It can be written uniquely as a power series $\sum_{k=0}^{\infty} a_k z^k$ which converges locally uniformly in $\mathbb{D}$. Since ξ is boundary controlled, it has a distributional boundary $\beta \in \mathcal{X}$. The matching condition relating ξ and β means that the Fourier coefficients of β are precisely the numbers $a_0, a_1, a_2, \ldots$, the Taylor coefficients of ξ. It follows that $\beta \in \mathcal{H}_1$, and so $\beta \in \mathcal{H}_1 \cap \mathcal{X}$. Therefore β is a Killing vector field on S^1 by Lemma 2.4.1. This has consequences for the Fourier coefficients $a_0, a_1, \ldots$ of β (for example, only a_0, a_1, a_2 can be nonzero). Then we can conclude $\xi \in \mathfrak{g}$. □

3. Constructing harmonic vector fields from boundary data

3.1. ***Symmetry properties of the classical Poisson formula.*** The symmetry group that we have in mind is $\mathrm{isom}(\mathbb{D})$, where $\mathbb{D}$ has the Poincaré metric as usual. Of course this acts on $\mathbb{D}$, but it also acts on S^1 and (continuously) on the compact manifold with boundary $\mathbb{D} \cup S^1$. It acts (on the right, by precomposition) on the vector spaces $C^0(S^1;\mathbb{R})$ and $C^0(\mathbb{D};\mathbb{R})$. (Both $C^0(S^1;\mathbb{R})$ and $C^0(\mathbb{D};\mathbb{R})$ are to be equipped with the compact-open C^0-topology. One of them is a Banach space, the other is just a topological vector space.)

Let $\mathscr{F}$ be the real vector space of continuous, $\mathbb{R}$-linear and $\mathrm{isom}(\mathbb{D})$-equivariant maps from $C^0(S^1;\mathbb{R})$ to $C^0(\mathbb{D};\mathbb{R})$

Proposition 3.1.1. $$\dim_{\mathbb{R}}(\mathscr{F}) = 1.$$

Proof. Let F be a nonzero element of $\mathscr{F}$. For each $z \in \mathbb{D}$, let ev_z from $C^0(\mathbb{D};\mathbb{R})$ to $\mathbb{R}$ be the map *evaluation at* z, a linear functional. Because $\mathrm{isom}(\mathbb{D})$ acts transitively on $\mathbb{D}$, the $\mathrm{isom}(\mathbb{D})$-equivariant map F is determined by the composition $\mathrm{ev}_0 \circ F$. The map $\mathrm{ev}_0 \circ F$ can no longer be claimed to be equivariant for the action(s) of $\mathrm{isom}(\mathbb{D})$, but it is equivariant w.r.t. the subgroup $\mathrm{O}(2) \subset \mathrm{isom}(\mathbb{D})$ consisting of the elements which fix $0 \in \mathbb{D}$. Therefore $\mathrm{ev}_0 \circ F$ is an $\mathrm{O}(2)$-invariant linear functional on $C^0(S^1;\mathbb{R})$. It is well known that the real vector space of these is 1-dimensional, generated by the Haar integral. Therefore $\dim_{\mathbb{R}}(\mathscr{F}) \le 1$.

On the other hand, we can use the equivariance condition to construct $P \in \mathscr{F}$ such that $\mathrm{ev}_0 \circ P$ is the nonzero linear functional taking $v \in C^0(S^1;\mathbb{R})$ to

$$\frac{1}{2\pi}\int_0^{2\pi} v(e^{\iota t})\,dt.$$

Choose $h \in \mathrm{isom}(\mathbb{D})$. By equivariance we must have $(P(v))(h(0)) = (P(v \circ h))(0)$, which comes down to

$$P(v)(h(0)) = \frac{1}{2\pi}\int_0^{2\pi} v(h(e^{\iota t}))\,dt. \tag{3.1.1}$$

The value of the integral depends only on $h(0)$, since h is determined by $h(0)$ up to precomposition with an element of $\mathrm{O}(2) \subset \mathrm{isom}(\mathbb{D})$. Therefore we have a definition of P in (3.1.1). It follows that $\dim(\mathscr{F}) \geq 1$. □

Remark 3.1.2. It is not obvious from (3.1.1) that $P(v)$ is harmonic, and that v and $P(v)$ together define a *continuous* function on $\mathbb{D} \cup S^1$. We are not going to justify these statements fully here (because they are well known). But in Section 3.2 we will encounter similar statements in a slightly different setup, and we will have to justify those. Therefore we sketch an argument. Let us try to make sense of $K := P(\delta_1)$, where δ_1 is a Dirac distribution at $1 \in S^1$. Think of δ_1 as the limit of a sequence of step functions w_n for $n \geq 2$, where w_n is zero outside the short arc I_n in S^1 with endpoints $e^{-\iota/2n}$ and $e^{\iota/2n}$, and has the constant value n for points in the arc. Then $P(\delta_1)$ should be the limit of the $P(w_n)$, and for $P(w_n)$ we expect

$$P(w_n)(h(0)) = \frac{1}{2\pi}\int_0^{2\pi} w_n(h(e^{\iota t}))\,dt$$

although w_n is not continuous. Here we may assume $h \in \mathrm{isom}_+(\mathbb{D})$, which implies that h is holomorphic so that we can use complex calculus notation. Since $w_n \circ h$ is constant on $h^{-1}(I_n)$ with constant value n, and zero elsewhere, the integral is n times the length of $h^{-1}(I_n)$. The length of $h^{-1}(I_n)$ is well approximated by the modulus of the complex number $(h^{-1})'(1)$, times the length of I_n, which is $2\pi/n$. We arrive at

$$K(h(0)) = P(\delta_1)(h(0)) = |(h^{-1})'(1)| = \frac{1}{|h'(h^{-1}(1))|}.$$

This is well defined, which means: dependent only on $h(0)$. Next, it is an interesting exercise to show that the map taking $v \in C^0(S^1;\mathbb{R})$ to the convolution of v and K is an element of $\mathscr{F}$. Therefore we have found another definition of P. The two claims about P can now be reformulated as claims about K. In other words it remains to show that K is harmonic, and that for each $v \in C^0(S^1;\mathbb{R})$, the Hilbert space inner product of v and the function $z \mapsto K(sz)$ (for $z \in S^1$ variable and $s \in [0,1)$ fixed) tends to $v(1)$ as s tends to 1.

3.2. ***Vector fields on the circle as boundary data.*** Let $C^0_v(S^1; TS^1)$ be the topological real vector space (with the compact-open topology) of continuous tangential vector fields on S^1, a.k.a. the space of continuous sections of $TS^1 \to S^1$. An element ψ of $C^0_v(S^1; TS^1)$ can also be viewed as a continuous map from S^1 to $\mathbb{C}$, subject to some conditions, because S^1 is a smooth submanifold of $\mathbb{C}$. This point of view is used in the next lemma. Note that U(1) acts on the right of $C^0_v(S^1; TS^1)$ by $(\psi, A) \mapsto A^*\psi = A^{-1}\cdot(\psi \circ A)$ for $A \in \mathrm{U}(1)$. (Of course A is nothing but a complex number of modulus 1, but we are tend to think of it as a $\mathbb{C}$-linear endomorphism of $\mathbb{C}$. And although it is sometimes convenient to view ψ as a function with values in $\mathbb{C}$, the formula for $A^*\psi$ is what it is because ψ is a vector field after all.)

Lemma 3.2.1. *Every continuous and* $\mathbb{R}$*-linear map* $\Lambda : C^0_v(S^1; TS^1) \to \mathbb{C}$ *which has* $\Lambda(A^*\psi) = A^{-1}\Lambda(\psi)$ *for all* $\psi \in C^0_v(S^1; TS^1)$ *and* $A \in \mathrm{U}(1)$ *is of the form*

$$\Lambda(\psi) = a \cdot \int_0^{2\pi} \psi(e^{\iota t})\, dt \tag{3.2.1}$$

for some $a \in \mathbb{C}$.

Proof. Given such Λ, we extend it to a map $\Lambda^{\mathbb{C}} : C^0_v(S^1; TS^1 \otimes_{\mathbb{R}} \mathbb{C}) \to \mathbb{C}$ as follows. Elements of $C^0_v(S^1; TS^1 \otimes_{\mathbb{R}} \mathbb{C})$ can be written uniquely in the form $\psi = \psi_1 + \iota\psi_2$, where both ψ_1 and ψ_2 are tangential to S^1. Then we let

$$\Lambda^{\mathbb{C}}(\psi_1 + \iota\psi_2) := \Lambda(\psi_1) + \iota\Lambda(\psi_2).$$

There are now two commuting actions of U(1) on $C^0_v(S^1; TS^1 \otimes_{\mathbb{R}} \mathbb{C})$. One is given by $(\psi, A) \mapsto A^*\psi$. The other is pointwise, $(\psi, A) \mapsto A \cdot \psi$. By our assumption on Λ, the map $\Lambda^{\mathbb{C}}$ intertwines the first action with the conjugate of the standard action of U(1) on $\mathbb{C}$. By construction, it intertwines the second action with the standard action of U(1) on $\mathbb{C}$. It follows that $\Lambda^{\mathbb{C}}$ is *invariant* under the operation

$$\psi \mapsto A \cdot A^*\psi,$$

where $\psi \in C^0_v(S^1; TS^1 \otimes_{\mathbb{R}} \mathbb{C})$ and $A \in \mathrm{U}(1)$. Here we can also write $A \cdot A^*\psi = \psi \circ A$ if we think of ψ as a function with values in $\mathbb{C}$. (The inclusion of the tangent space T_zS^1 in $\mathbb{C}$ extends uniquely to a $\mathbb{C}$-linear isomorphism $T_zS^1 \otimes_{\mathbb{R}} \mathbb{C} \to \mathbb{C}$.) Briefly, $\Lambda^{\mathbb{C}}$ satisfies $\Lambda^{\mathbb{C}}(\psi \circ A) = \Lambda^{\mathbb{C}}(\psi)$ for all ψ and all $A \in \mathrm{U}(1)$. Because it is also $\mathbb{C}$-linear by construction, it must have the form

$$\psi \mapsto a \cdot \int_0^{2\pi} \psi(e^{\iota t})\, dt$$

for some $a \in \mathbb{C}$. In particular this is valid for $\psi \in C^0_v(S^1; TS^1)$. □

Remark 3.2.2. Clearly $\Lambda : C^0_v(S^1; TS^1) \to \mathbb{C}$ satisfies the condition $\Lambda(A^*\psi) = A^{-1}\Lambda(\psi)$ for all $\psi \in C^0_v(S^1; TS^1)$ and $A \in \mathrm{U}(1)$ if it is of the form (3.2.1). But if the complex number a in (3.2.1) is a real number, and only then, it will satisfy the same condition for all $A \in \mathrm{O}(2)$. The proof is by direct verification.

Let $C^0_v(\mathbb{D}; T\mathbb{D})$ be the real vector space of continuous vector fields on $\mathbb{D}$, equipped with the compact-open C^0 topology. The group isom($\mathbb{D}$) acts on the right of $C^0_v(\mathbb{D}; T\mathbb{D})$ by $(\xi, h) \mapsto h^*\xi$, which means $(h_*\xi)(z) = Dh(z)^{-1}\big(\xi(h(z))\big)$ for $z \in \mathbb{D}$. The action is $\mathbb{R}$-linear.

Let $\mathscr{F}_{vf}$ be the real vector space of continuous, $\mathbb{R}$-linear and isom($\mathbb{D}$)-equivariant maps from $C^0_v(S^1; TS^1)$ to $C^0_v(\mathbb{D}; T\mathbb{D})$.

Corollary 3.2.3. $$\dim_{\mathbb{R}}(\mathscr{F}_{vf}) = 1.$$

Proof. Let F be a nonzero element of $\mathscr{F}_{vf}$. For each $z \in \mathbb{D}$, let ev_z from $C^0_v(\mathbb{D}; T\mathbb{D})$ to $\mathbb{C} = T_0\mathbb{D}$ be the map *evaluation at z*, an $\mathbb{R}$-linear map. Because isom($\mathbb{D}$) acts transitively on $\mathbb{D}$, the isom($\mathbb{D}$)-equivariant map F is determined by the composition $\mathrm{ev}_0 \circ F$. The map $\Lambda := \mathrm{ev}_0 \circ F$ can no longer claim to be equivariant for the action(s) of isom($\mathbb{D}$), but it satisfies the condition of Lemma 3.2.1 for $A \in \mathrm{U}(1)$, and the stronger condition of Remark 3.2.2 which allows $A \in \mathrm{O}(2)$. Therefore $\Lambda = \mathrm{ev}_0 \circ F$ has the form (3.2.1) for some *real* number a. It follows that $\dim_{\mathbb{R}}(\mathscr{F}_{vf}) \leq 1$.

We can use the equivariance condition to construct an $F \in \mathscr{F}_{vf}$ such that $\mathrm{ev}_0 \circ F$ is the nonzero linear functional taking $\psi \in C^0_v(S^1; TS^1)$ to

$$\frac{1}{2\pi}\int_0^{2\pi} \psi(e^{\iota t})\,dt.$$

Choose $h \in \mathrm{isom}(\mathbb{D})$. By equivariance we must have $(Dh(0))^{-1}\big((F(\psi))(h(0))\big) = (F(h^*\psi))(0)$, which comes down to

$$F(\psi)(h(0)) = \frac{1}{2\pi}(Dh(0))\left(\int_0^{2\pi} (h^*\psi)(e^{\iota t})\,dt\right). \tag{3.2.2}$$

It is easy to see that the right-hand side depends only on $h(0)$, because h is determined by $h(0)$ up to precomposition with an element of $\mathrm{O}(2) \subset \mathrm{isom}(\mathbb{D})$). Therefore F in (3.2.2) is well defined. It is easily seen to be nonzero. It follows that $\dim(\mathscr{F}_{vf}) \geq 1$. □

We want to show that $F(\psi)$, with the definition of F in (3.2.2), is always a harmonic vector field and that ψ and $c \cdot F(\psi)$ together define a continuous vector field on $\mathbb{D} \cup S^1$, where c is a positive real constant factor independent of ψ. Imitating the strategy outlined in Remark 3.1.2, we begin by making sense of

$$K_{vf} := F(\delta_1 \cdot \omega),$$

where δ_1 is the Dirac distribution and $\omega \in C^0_v(S^1, TS^1)$ is the tangential vector field defined by $\omega(z) = \iota z$ for $z \in S^1$. To simplify this, we remark that for every $z \in \mathbb{D}$ there exists a *unique* $h \in \mathrm{isom}_+(\mathbb{D})$ such that $h(0) = z$ and $h(1) = 1$. For such an $h \in \mathrm{isom}_+(\mathbb{D})$ we first need to make sense of $h^*(\delta_1 \cdot \omega) = h^*\delta_1 \cdot h^*\omega$. It is not hard to see (and the calculation of $h^*\delta_1$ can be seen in Remark 3.1.2) that this is $(h'(1))^{-2}\delta_1 \cdot \omega$. Therefore we make the *definition*

$$K_{vf}(h(0)) = \iota h'(0) \cdot (h'(1))^{-2} \tag{3.2.3}$$

on the understanding that $h \in \mathrm{isom}_+(\mathbb{D})$ and $h(1) = 1$. If we specify $z = h(0)$, then we can write

$$h(w) = \frac{yw + z}{yw\bar{z} + 1},$$

where $y = (1-z)/(1-\bar{z}) = (1-z^2)/|1-z|^2$. Then $h'(w) = y(1-|z|^2)/(yw\bar{z}+1)^2$. Therefore $h'(0) = y(1-|z|^2)$ and

$$h'(1) = \frac{y(1-|z|^2)}{(y\bar{z}+1)^2} = \frac{y(1-|z|^2)}{(z+y)^2}$$

(where we have used $h(w) = 1$, so that $y\bar{z} + 1 = z + y$) and

$$h'(0)/(h'(1))^2 = y(1-|z|^2) \cdot \frac{(z+y)^4}{y^2(1-|z|^2)^2} = \frac{(z+y)^4}{y(1-|z|^2)} = \frac{|1-z|^2(z+y)^4}{(1-z)^2(1-|z|^2)}.$$

Now we use $z + y = (z - |z|^2 + 1 - z)/(1-\bar{z}) = (1-|z|^2)/(1-\bar{z})$ and obtain

$$h'(0)/(h'(1))^2 = \frac{(1-|z|^2)^3}{(1-\bar{z})^2|1-z|^2},$$

so that

$$K_{vf}(z) = \iota \cdot \frac{(1-|z|^2)^3}{(1-\bar{z})^2|1-z|^2}. \tag{3.2.4}$$

Lemma 3.2.4. *Suppose that $h \in \mathrm{isom}_+(\mathbb{D})$ satisfies $h(1) = 1$. Then*

$$h^* K_{vf} = c \cdot K_{vf},$$

where $c = (h'(1))^{-2}$.

Proof. Loosely and provisionally we defined ξ as $F(\delta_1 \cdot \omega)$. Therefore $h^*\xi = h^*F(\delta_1 \cdot \omega) = F(h^*(\delta_1 \cdot \omega))$ by the equivariance property of F. Here we can see that $h^*(\delta_1 \cdot \omega) = c \cdot \delta_1 \cdot \omega$ for some constant c, which turns out to be $(h'(1))^{-2}$. A more orderly proof can be given using the definition (3.2.3) and the chain rule. Then we have to allow two $h_1, h_2 \in \mathrm{isom}_+(\mathbb{D})$ such that $h_1(1) = h_2(1) = 1$. We obtain

$$\begin{aligned}(h_2^*\xi)(h_1(0)) &= \big(Dh_2(h_1(0))\big)^{-1} F\big(h_2(h_1(0))\big) \\ &= \big(Dh_2(h_1(0))\big)^{-1} \iota (h_2h_1)'(0)/((h_2h_1)'(1))^2 \\ &= \iota h_1'(0)/((h_2h_1)'(1))^2 \\ &= c \cdot \iota h_1'(0)/(h_1'(1))^2 \\ &= c \cdot \xi(h_1(0)),\end{aligned}$$

where $c = (h_2'(1))^{-2}$. □

Theorem 3.2.5. *The vector field K_{vf} is harmonic on $\mathbb{D}$.*

Proof. The first step will be to show that $\xi := K_{vf}$ is harmonic at the origin in $\mathbb{D}$. For that we can use the second order Taylor approximation of ξ at the origin:

$$\begin{aligned}
\xi(z) &= \frac{\iota(1-|z|^2)^3}{|1-\bar{z}|^2\cdot(1-\bar{z})^2}\\
&= \iota(1-|z|^2)^3(1-\bar{z})^{-3}(1-z)^{-1}\\
&\approx \iota(1-3|z|^2)(1+\bar{z}+\bar{z}^2)^3(1+z+z^2)\\
&\approx \iota(1-3|z|^2)(1+3\bar{z}+3\bar{z}^2+3\bar{z}^2)(1+z+z^2)\\
&\approx \iota(1+3\bar{z}+6\bar{z}^2-3|z|^2)(1+z+z^2)\\
&\approx \iota(1+3\bar{z}+6\bar{z}^2-3|z|^2+z+3|z|^2+z^2)\\
&= \iota(1+z+3\bar{z}+z^2+6\bar{z}^2)\\
&= \iota\big(1+(x+\iota y)+3(x-\iota y)+x^2-y^2+2\iota xy+6(x^2-y^2)-12\iota xy\big)\\
&= \iota(1+4x-2\iota y+7x^2-7y^2-10\iota xy)\\
&= (2y+10xy,\ 1+4x+7x^2-7y^2).
\end{aligned}$$

Let g be the Poincaré metric on $\mathbb{D}$, which we regard as a function from $\mathbb{D}$ to the vector space of symmetric 2×2-matrices. It has the form $g(w)=u(w)\cdot g^E$, where $u:\mathbb{D}\to\mathbb{R}$ is a smooth function which has $u(0)=2$ and g^E is the Euclidean (Riemannian) metric, i.e., a constant, the constant value being the identity 2×2-matrix. The product rule gives us

$$\mathscr{L}_\xi(g)=\mathscr{L}_\xi(u)\cdot g^E+u\cdot\mathscr{L}_\xi(g^E).$$

The first summand is a contribution to the scalar summand. We may neglect it. As to the second summand, we are only interested in the first Taylor approximation at 0, and since the first Taylor approximation of u at 0 is a constant 2, we can replace the second summand by $2\mathscr{L}_\xi(g^E)$. Now we can use (2.1.3):

$$\mathscr{L}_\xi(g^E)=(D\xi^T+D\xi)g^E+Dg(\xi)=D\xi^T+D\xi,$$

which in terms of the above Taylor approximation turns into

$$\begin{bmatrix}20y & 6+24x\\ 6+24x & -28y\end{bmatrix}.$$

The trace-free part is

$$\begin{bmatrix}24y & 6+24x\\ 6+24x & -24y\end{bmatrix},$$

which, as a symmetric bilinear form, agrees with $\mathrm{Re}((24y-\iota(6+24)x)\,dz\otimes_{\mathbb{C}}dz)$. This completes the verification that $\xi=K_{vf}$ is harmonic at the origin, because the first order polynomial map $z\mapsto 24y-\iota(6+24)x$ (for $z=x+\iota y$) is holomorphic.

To finish the proof, we want to argue that “harmonic at the origin” is enough. For other $z \in \mathbb{D}$ we can find $h \in \mathrm{isom}_+(\mathbb{D})$ such that $h(0) = z$ and $h(1) = 1$. Showing that ξ is harmonic at z is equivalent to showing that $h^*\xi$ is harmonic at 0. By Lemma 3.2.4, we can write $h^*\xi = c \cdot \xi$. □

Our next goal is to show that the linear map F in (3.2.2) has another description as something very close to *convolution with* K_{vf}. Let $M_t : \mathbb{C} \to \mathbb{C}$ be the multiplication by $e^{\iota t} \in S^1$; we will also view this as an element of $\mathrm{isom}_+(\mathbb{D})$. Let ψ be a tangential vector field on S^1. Write $\psi = u \cdot \omega$, where ω is the standard tangential vector field $z \mapsto \iota z$ and $u : S^1 \to \mathbb{R}$ is a continuous function. (It is allowed to write $u := \psi/\omega$.) Write $\eta(t) := e^{\iota t}$ for $t \in \mathbb{R}$. The new formula for F that we have in mind is

$$(F(\psi))(z) \stackrel{?}{=} \frac{1}{2\pi} \int_0^{2\pi} (M_{-t}^* K_{vf})(z) \cdot u(e^{\iota t})\, dt. \tag{3.2.5}$$

Unraveling the right-hand side, we obtain

$$\frac{1}{2\pi} \int_0^{2\pi} e^{\iota t} \cdot K_{vf}(e^{-\iota t} z) \cdot u(e^{\iota t})\, dt = \frac{1}{2\pi} \int_0^{2\pi} K_{vf}(e^{-\iota t} z) \cdot e^{\iota t} u(e^{\iota t})\, dt.$$

But $e^{\iota t} u(e^{\iota t})$ is the same as $-\iota \cdot \psi(e^{\iota t})$. So we may write

$$-\iota \frac{1}{2\pi} \int_0^{2\pi} K_{vf}(e^{-\iota t} z) \cdot \psi(e^{\iota t})\, dt$$

and then also $-\iota \cdot (K_{vf} * \psi)(z)$, where the star is for *convolution*.

Proposition 3.2.6. *The linear map F in* (3.2.2) *has the alternative description*

$$\psi \mapsto -\iota \cdot (K_{vf} * \psi).$$

Proof. Take $z \in \mathbb{D}$. Then

$$-\iota \cdot (K_{vf} * \psi)(z) = -\iota \frac{1}{2\pi} \int_0^{2\pi} K_{vf}(e^{-\iota t} z) \cdot \psi(e^{\iota t} z)\, dt.$$

Choose h such that $h(0) = z$ and $h(1) = 1$. If we choose $s \in \mathbb{R}$ appropriately, depending on t, then $M_{-t} h M_s(0) = M_{-t}(z) = e^{-\iota t} z$ and $M_{-t} h M_s(1) = 1$. (We can determine s later.) Therefore by (3.2.3),

$$\begin{aligned} K_{vf}(e^{-\iota t} z) &= \iota (M_{-t} h M_s)'(0) \cdot ((M_{-t} h M_s)'(1))^{-2} \\ &= \iota e^{\iota(s-t)} h'(0) \cdot (h'(e^{\iota s}))^{-2} \cdot e^{-2\iota(s-t)} \\ &= \iota e^{\iota(t-s)} h'(0) \cdot (h'(e^{\iota s}))^{-2}, \end{aligned}$$

so that the formula for $-\iota \cdot (K_{vf} * \psi)(z)$ simplifies to

$$\frac{1}{2\pi} \int_0^{2\pi} e^{\iota(t-s)} h'(0) \cdot (h'(e^{\iota s}))^{-2} \cdot \psi(e^{\iota t})\, dt.$$

Now is the moment to observe that s and t are related by $h(e^{\iota s}) = e^{\iota t}$. Therefore it is "locally" justified to write $t = -\iota \cdot \ln(h(e^{\iota s}))$ and in any case

$$\frac{dt}{ds} = -\iota \frac{h'(e^{\iota s}) \cdot e^{\iota s} \cdot \iota}{h(e^{\iota s})} = \frac{h'(e^{\iota s}) \cdot e^{\iota s}}{h(e^{\iota s})} = h'(e^{\iota s}) \cdot e^{\iota(s-t)}.$$

With the substitution of $h(e^{\iota s})$ for $e^{\iota t}$ and $h'(e^{\iota s}) \cdot e^{\iota(s-t)}\, ds$ for dt the above integral turns into

$$\frac{1}{2\pi} \int_0^{2\pi} h'(0) \cdot (h'(e^{\iota s}))^{-1} \cdot \psi(h(e^{\iota s}))\, ds.$$

This agrees with $F(\psi)(z) = (F(\psi))(h(0))$ according to (3.2.2). □

Corollary 3.2.7. *$F(\psi)$ is harmonic, for every $\psi \in C^0_v(S^1; TS^1)$.*

Proof. By Proposition 3.2.6 and the discussion preceding it, (3.2.5) is correct. By Theorem 3.2.5, the "kernel" K_{vf} is harmonic and so $M^*_{-t} K_{vf}$ is also harmonic, for arbitrary $t \in \mathbb{R}$. Hence the right-hand side of (3.2.5) is harmonic. □

Lemma 3.2.8. *Let ψ be a continuous tangential vector field on S^1. Then $F(\psi)$ and ψ together make up a continuous vector field on the closed unit disk $\mathbb{D} \cup S^1$.*

Proof. For $s \in [0, 1)$, we define a continuous (but not tangential) vector field κ_s on S^1 by $\kappa_s(z) := K_{vf}(sz)$. Then for $\varepsilon \in (0, 1]$ and $z \in S^1$ we have

$$\begin{aligned} \kappa_{1-\varepsilon}(z) &= \frac{\iota(1-(1-\varepsilon)^2)^3}{(1-(1-\varepsilon)\bar{z})^3 \cdot (1-(1-\varepsilon)z)} \\ &= \frac{\iota z^3(1-(1-\varepsilon)^2)^3}{(z-(1-\varepsilon))^3 \cdot (1-(1-\varepsilon)z)} \\ &= \frac{8\iota z^3 \varepsilon^3 V(\varepsilon)}{(1-(1-\varepsilon)z) \cdot (z-(1-\varepsilon))^3}, \end{aligned}$$

where V is a real polynomial of degree 3 with constant coefficient 1.

Let $\lambda_z = |z - (1-\varepsilon)|$. Then $\lambda_z \geq \varepsilon$ and

$$|\kappa_{1-\varepsilon}(z)| = \frac{8\varepsilon^3 V(\varepsilon)}{\lambda_z^4} \leq \frac{8V(\varepsilon)}{\varepsilon}. \tag{3.2.6}$$

This gives us an upper bound for $|\kappa_{1-\varepsilon}(z)|$ which is independent of z, but more importantly it tells us that $\kappa_{1-\varepsilon}$ is very small outside the arc of length $(2\varepsilon)^{1/2}$ centered at 1. Therefore it is enough to show that

$$\lim_{\varepsilon \to 0} \left(\frac{1}{2\pi} \int_0^{2\pi} \kappa_{1-\varepsilon}(e^{\iota t})\, dt \right) = \iota.$$

Since $\kappa_{1-\varepsilon}$ is so very small outside the arc of length $\sqrt{2\varepsilon}$ centered at 1, we may replace the ordinary integral by the complex path integral

$$\oint_\gamma \kappa_{1-\varepsilon}(z)\,dz$$

(where γ is a smooth curve describing the unit circle) at the price of dividing by ι. We may also write s for $1-\varepsilon$. Then

$$\begin{aligned}\lim_{\varepsilon\to 0}\left(\frac{1}{2\pi}\int_0^{2\pi}\kappa_{1-\varepsilon}(e^{\iota t})\,dt\right) &= \lim_{\varepsilon\to 0}\left(\frac{1}{2\pi\iota}\oint_\gamma \kappa_{1-\varepsilon}(z)\,dz\right)\\ &= \lim_{\varepsilon\to 0}\left(\frac{1}{2\pi\iota}\oint_\gamma \frac{8\iota z^3\varepsilon^3 V(\varepsilon)}{(1-sz)\cdot(z-s)^3}\,dz\right)\\ &= \lim_{\varepsilon\to 0}\left(\frac{8\iota\varepsilon^3 V(\varepsilon)}{2\pi\iota}\big(2\pi\iota\cdot \operatorname{Res}(f,s)\big)\right)\\ &= \lim_{\varepsilon\to 0}\big(8\iota\varepsilon^3(\operatorname{Res}(f,s))\big),\end{aligned}$$

where $f(z)=(z^3)(1-sz)^{-1}(z-s)^{-3}$. Now

$$\operatorname{Res}(f,s)=\frac{6s-12s^3+8s^5-2s^7}{2(1-s^2)^4}=\frac{4\varepsilon+2\varepsilon^2+2\varepsilon^3-30\varepsilon^4+34\varepsilon^5-14\varepsilon^6+2\varepsilon^7}{2(16\varepsilon^4-32\varepsilon^5+20\varepsilon^6-8\varepsilon^7+\varepsilon^8)}.$$

Therefore

$$\begin{aligned}&\lim_{\varepsilon\to 0}\big(\iota 8\varepsilon^3(\operatorname{Res}(f,s))\big)\\ &\quad=\lim_{\varepsilon\to 0}\left(8\iota\cdot\frac{4\varepsilon+2\varepsilon^2+2\varepsilon^3-30\varepsilon^4+34\varepsilon^5-14\varepsilon^6+2\varepsilon^7}{2(16\varepsilon-32\varepsilon^2+20\varepsilon^3-8\varepsilon^4+\varepsilon^5)}\right)=\iota.\end{aligned}$$

□

Lemma 3.2.9. *The map F in (3.2.2) has a Lipschitz property. More precisely, there is $c>0$ such that the following holds: if $\psi\in C^0(S^1;TS^1)$ satisfies $\|\psi(z)\|\le 1$ for all $z\in S^1$, in the euclidean norm, then $\|F(\psi)(z)\|\le c$ for all $z\in\mathbb{D}$, again in the euclidean norm.*

Proof. In the proof of Lemma 3.2.8 we learned

$$\lim_{s\to 1}\frac{1}{2\pi}\int_0^{2\pi}|\kappa_s(e^{\iota t})|\,dt=1.$$

This implies that the function

$$s\mapsto\frac{1}{2\pi}\int_0^{2\pi}|\kappa_s(e^{\iota t})|\,dt$$

is defined and continuous for $s\in[0,1]$. Then it has a maximum $c>0$ on the interval. Now let ψ be a continuous tangential vector field on S^1 and suppose that

$\|\psi(z)\| \leq 1$ for all $z \in S^1$. By Proposition 3.2.6 we have, for $z \in S^1$ and $s \in [0,1)$:

$$(F(\psi))(sz) = -\iota \frac{1}{2\pi} \int_0^{2\pi} \kappa_s(e^{-\iota t} z) \cdot \psi(e^{\iota t})\, dt.$$

Therefore

$$\|(F(\psi))(sz)\| \leq \frac{1}{2\pi} \int_0^{2\pi} |\kappa_s(e^{-\iota t} z)|\, dt \leq c. \qquad \square$$

We come to the last part of Theorem II (in the introduction). This is about extending F in (3.2.2) and Proposition 3.2.6 to a continuous map defined on the Hilbert space $\mathcal{X}$ of tangential L^2-vector fields on S^1. The extension formula as such is obvious: both (3.2.2) and the formula in Proposition 3.2.6 tolerate a tangential L^2-vector field ψ instead of a continuous one. Both of these extensions are clearly continuous, and because they agree on a dense subspace of $\mathcal{X}$ they agree on all of $\mathcal{X}$.

Lemma 3.2.10. *For $\zeta \in \mathcal{X}$, let $\xi := F(\zeta) \in C^0_v(\mathbb{D}; T\mathbb{D})$. Then ξ is boundary controlled and ζ is the distributional boundary of ξ; see Definition 1.3.3.*

Proof. We use a method from [Shubin 2020, Section 5.3]. Fix α, continuous vector field on $\mathbb{D} \cup S^1$. Write

$$w(s) := \frac{1}{2\pi} \int_0^{2\pi} s \cdot \xi(se^{\iota t}) \cdot \alpha(se^{\iota t})\, dt$$

for $s \in [0,1)$. It is more than enough to show that

$$\lim_{s \to 1^-} w(s) = \frac{1}{2\pi} \int_0^{2\pi} \zeta(e^{\iota t}) \cdot \alpha(e^{\iota t})\, dt,$$

because w is, except for the constant factor $(2\pi)^{-1}$, the derivative of the function in (1.3.2). Write $\alpha_s : S^1 \to \mathbb{R}^2$ for the map $z \mapsto \alpha(sz)$, where $s \in [0,1]$, and write $\kappa_s : S^1 \to \mathbb{R}^2$ for the map $z \mapsto K_{vf}(sz)$ as in the proof of Lemma 3.2.8, assuming $s \in [0,1)$. By (3.2.2) in the general form which allows a tangential L^2-vector field as input for F, we can write

$$w(s) = \langle -\iota s(\kappa_s * \zeta), \alpha_s \rangle.$$

This notation uses the standard real L^2 inner product (based on the standard inner product in $\mathbb{R}^2$, a.k.a. real part of the hermitian inner product in $\mathbb{C}$). Therefore

$$w(s) = \langle \zeta, -\iota s(\kappa_s * \alpha_s) \rangle.$$

Here we can say that

$$\lim_{s \to 1^-} -\iota s(\kappa_s * \alpha_s) = \lim_{s \to 1^-} -\iota(\kappa_s * \alpha_1) = \alpha_1$$

by Lemmas 3.2.9 and 3.2.8. (These limits take place in $C^0_v(\mathbb{D}, T\mathbb{D})$ or in $C^0(\mathbb{D}, \mathbb{R}^2)$ depending on point of view.) It follows that

$$\lim_{s\to 1^-} w(s) = \langle \zeta, \alpha_1 \rangle. \qquad \square$$

Appendix: Odds and ends

A.1. ***Connections on the Teichmüller bundle.*** A *connection* on a smooth fiber bundle $f : E \to M$ with vertical tangent bundle $T^v E$ is a smooth vector subbundle $T^h E$ (the *horizontal* tangent bundle) of the tangent bundle TE such that $T^h E \oplus T^v E = TE$. Equivalently, a connection on $f : E \to M$ is a smooth vector bundle homomorphism $f^*TM \to TE$ such that the composition

$$f^*TM \to TE \to TE/T^v E$$

is the identity. We are interested in connections on the Teichmüller surface bundle, a.k.a. *universal Teichmüller curve* (the fibers can either be viewed as real surfaces or as complex curves). To describe the bundle we fix $\Gamma = \Gamma_g$ (fundamental group of a surface Σ of genus $g \geq 2$) and $G = \mathrm{isom}_+(\mathbb{H})$ as usual. Let $\mathrm{hom}_0(\Gamma, G)$ be the space of injective homomorphisms with discrete image and compact quotient space $G/\rho(\Gamma)$. Let $\mathrm{rep}_0(\Gamma, G)$ be the quotient of $\mathrm{hom}_0(\Gamma, G)$ obtained by passing to orbits for the conjugation action of G. (This was called $\mathscr{T}(\Sigma)$ in Definition 1.1.2.) There are two commuting left actions of Γ and G, respectively, on the product $\mathrm{hom}_0(\Gamma, G) \times \mathbb{H}$.

The action of Γ is given by $\gamma \cdot (\rho, z)) := (\rho, \rho(\gamma)(z))$ for $\gamma \in \Gamma$. The action of G is given by $A \cdot (\rho, z) := (A\rho A^{-1}, A(z))$ for $A \in G$. Therefore we obtain a commutative diagram

$$\begin{array}{ccc} E := \dfrac{\mathrm{hom}_0(\Gamma, G) \times \mathbb{H}}{\Gamma} & \xrightarrow{\text{proj.}} & \mathrm{hom}_0(\Gamma, G) \\ \big\downarrow & & \big\downarrow \\ E_G := \dfrac{\mathrm{hom}_0(\Gamma, G) \times \mathbb{H}}{G \times \Gamma} & \xrightarrow{\text{proj.}} & \mathrm{rep}_0(\Gamma, G) \end{array} \tag{A.1.1}$$

where the vertical arrows are quotient maps and the horizontal ones are appropriate projections. The two horizontal arrows are surface bundle projections (or complex curve bundle projections). The vertical arrows are principal G-bundle projections. The diagram is a pullback diagram. In the top row, the fiber over $\rho \in \mathrm{hom}_0(\Gamma, G)$ is the surface $\mathbb{H}/\rho(\Gamma)$.

The lower horizontal arrow $E_G \to \mathrm{rep}_0(\Gamma, G)$ in (A.1.1) is the Teichmüller bundle. We take the view that we can construct connections on it by constructing connections on the bundle defined by the upper horizontal arrow $E \to \mathrm{hom}_0(\Gamma, G)$

and imposing conditions which ensure that these connections descend to connections on the bundle defined by the lower horizontal arrow.

By the general remarks on connections in smooth fiber bundles, choosing a connection on the trivial bundle

$$\hom_0(\Gamma, G) \times \mathbb{H} \to \hom_0(\Gamma, G) \tag{A.1.2}$$

amounts to choosing a smooth vector field $\xi(\rho, c, -)$ on $\mathbb{H}$ for every $\rho \in \hom_0(\Gamma, G)$ and 1-cocycle $c \in T_\rho(\hom_0(\Gamma, G)) \cong Z^1(\Gamma; \mathfrak{g}_\rho)$. (This should ideally depend smoothly on ρ and c.) But we want to choose a connection that respects the Γ-action on each fiber of the bundle in (A.1.2), since we want a connection for $E \to \hom_0(\Gamma, G)$ in diagram (A.1.1). This translates into the following condition on $\xi(\rho, c, -)$:

(i) $\delta\xi(\rho, c, -) = c$, where δ is the coboundary operator associated with the $\mathbb{R}\Gamma$-module of smooth vector fields on $\mathbb{H}$. (The module structure depends on ρ.)

Proposition A.1.1. *A connection* $\xi = (\xi(\rho, c, -))$ *for* $E \to \hom_0(\Gamma, G)$ *in* (A.1.1) *descends to a connection for* $E_G \to \mathrm{rep}_0(\Gamma, G)$ *if and only if it satisfies the following additional conditions*:

(ii) *it is invariant under the left action of* G;

(iii) $\xi(\rho, \delta\kappa, -) = \kappa$ *for every* $\kappa \in \mathfrak{g}_\rho$.

Here $\kappa \in \mathfrak{g}_\rho$ should be viewed as a Killing vector field on $\mathbb{H}$. Condition (iii) does not follow from (ii) and (i). It is easy to produce counterexamples.

Proof. Suppose the connection ξ for $E \to \hom_0(\Gamma, G)$ descends. Then condition (ii) is satisfied. Let K be any G-orbit in $\hom_0(\Gamma, G)$. Then the connection ξ restricted to $E|_K \to K$ is the connection determined by the trivialization of $E|_K \to K$ produced by the action of G on $E|_K$ (which is free). This translates into condition (iii). Conversely, suppose that (ii) and (iii) are satisfied by a connection $\xi = (\xi(\rho, c, -))$ on $E \to \hom_0(\Gamma, G)$. Then, by (iii), the restricted connection on $E|_K \to K$ reflects the trivialization of $E|_K \to K$ produced by the action of G on E_K, or equivalently, by the composition $E|_K \hookrightarrow E \to E_G$. Choose a smooth section

$$s : \mathrm{rep}_0(\Gamma, G) \to \hom_0(\Gamma, G)$$

of the projection $\hom_0(\Gamma, G) \to \mathrm{rep}_0(\Gamma, G)$. The section s is covered uniquely by a smooth map $\bar{s} : E_G \to E$ which is a section of the projection $E \to E_G$. The pullback along s and $\bar{s}$ of the connection ξ on $E \to \hom_0(\Gamma, G)$ is a connection θ on $E_G \to \mathrm{rep}_0(\Gamma, G)$. Conditions (ii) and (iii) ensure that ξ is also the pullback of θ along the projections $E \to E_G$ and $\hom_0(\Gamma, G) \to \mathrm{rep}_0(\Gamma, G)$. □

We can meet all of these conditions as follows. Given ρ and c, find a smooth vector field ψ as promised in Lemma 1.4.2 such that $\delta\psi = c$. Let ζ^c be the distributional boundary of ψ as promised in Proposition 1.4.3 and let $\xi(\rho, c, -) := F(\zeta^c)$, with F as in Theorem II. This is well defined and the conditions are easily verified. We authors believe that $\xi(\rho, c, -)$ depends continuously, indeed smoothly, on ρ and c, but the proof could be laborious and perhaps it deserves a separate treatment. (We also believe that this candidate for a connection ξ is identical with the connection which is standard in Teichmüller theory. This can be seen, for example, in [Wolpert 1986, §5]. Evidence for the suspected agreement was given in Section 2.2.)

A.2. *Some postponed proofs.*

Lemma A.2.1. *The conformal vector fields on $\mathbb{D}$ or open subsets of $\mathbb{D}$ are precisely the holomorphic vector fields.*

Proof. By definition, a vector field ξ on $\mathbb{D}$ (or on an open subset of $\mathbb{D}$) is conformal if and only if the trace-free component of $\mathscr{L}_\xi(g)$ is zero. By (2.1.3) and the calculations immediately following it, this happens if and only if $\xi^1_x = \xi^2_y$ and $\xi^1_y = -\xi^2_x$ (where the subscripts indicate partial derivatives). These are exactly the Cauchy–Riemann equations for ξ. □

Proof of Lemma 1.4.1. Let F_1 be the restriction of F in (3.2.2) to the 3-dimensional real vector space of Killing vector fields on S^1. Let F_2 be the linear map which is defined on the same vector space and which takes a Killing vector field on S^1 to the unique matching Killing vector field on $\mathbb{D}$. We need to show $F_1 = F_2$. Since both F_1 and F_2 are equivariant for the right actions of $\mathrm{isom}_+(\mathbb{D})$, it is enough to show that $(F_1(\psi))(0) = (F_2(\psi))(0)$ for all Killing vector fields on S^1. (Follow the reasoning in the proof of Corollary 3.2.3.) By definition, $(F_1(\psi))(0)$ is

$$\frac{1}{2\pi}\int_0^{2\pi} \psi(e^{it})\,dt.$$

In other words it is the mean value of ψ on the circle. But $(F_2(\psi))(0)$ is also the mean value of ψ on the circle. Indeed, ψ and $F_2(\psi)$ are the restrictions to S^1 and $\mathbb{D}$, respectively, of one and the same holomorphic function (alias holomorphic vector field) on $\mathbb{C}$, and this must satisfy the Cauchy integral formula. □

Proposition A.2.2. *The matching condition relating the vector fields ξ and ζ in Definition 1.3.3 is invariant under the canonical right action(s) of the diffeomorphism group of $\mathbb{D} \cup S^1$.*

Proof. We will write $S(\xi, \zeta)$ for the statement *the matching condition holds for ξ and ζ*. Therefore we assume $S(\xi, \zeta)$, and we have to show $S(h^*\xi, h^*\zeta)$, where h is an arbitrary diffeomorphism from $\mathbb{D} \cup S^1$ to itself.

The diffeomorphism h can be written as a composition $h_a \circ h_n$, where h_n is norm-preserving in a neighborhood of the boundary S^1, and h_a is argument-preserving. (By *argument-preserving*, we mean that there is a smooth function

$N : S^1 \times [0, 1] \to [0, 1]$ such that $h_a(sz) = N(z, s) \cdot z$ for all $(z, s) \in S^1 \times [0, 1]$. For convenience we also require $N(z, s) = s$ for s close to 0.) Using integration by substitution, it is easy to show that $S(h^*\xi, h^*\zeta)$ is equivalent to $S(h_a^*\xi, h_a^*\zeta)$. It is also obvious that $h_a^*\zeta = \zeta$. Therefore we may assume from now on that h is argument-preserving, and we have to show $S(h^*\xi, \zeta)$, knowing that $S(\xi, \zeta)$ holds.

Let $G_\xi : S^1 \times [0, 1) \to \mathbb{R}^2$ be defined by $(z, s) \mapsto \int_0^s t \cdot \xi(tz)\, dt$. Then we have

$$\int_{|z|=s} G_\xi \, d\lambda_0 = \int_{z \in \mathbb{D}_s} \xi(z)\, d\lambda$$

for $s < 1$. We use this to reformulate $S(\xi, \zeta)$. Namely, it is equivalent to the following, which we denote by $T(\xi, \zeta)$.

> *The map G_ξ has an extension to $S^1 \times [0, 1]$ which admits a distributional partial derivative $\partial/\partial s$ along $\{(z, s) \mid s = 1\}$, and the latter is equal to ζ.*

This is rather concise. To clarify, the map $z \mapsto G_\xi(z, 1)$ from S^1 to $\mathbb{R}^2$ has two coordinate functions. They are meant to be Lebesgue measurable and integrable functions, and a such defined for almost all $z \in S^1$.

To show that $S(\xi, \zeta)$ implies $T(\xi, \zeta)$, we may use the coordinate functions ξ^1 and ξ^2. We can write $\xi^1 = \xi^1_+ - \xi^1_-$, where ξ^1_+ and ξ^1_- are nonnegative everywhere, and similarly $\xi^2 = \xi^2_+ - \xi^2_-$. For fixed $s \in [0, 1)$, define a function $g^1_{s,+}$ on S^1 by

$$z \mapsto \int_0^s t \cdot \xi^1_+(tz)\, dt.$$

Its integral over S^1 is equal to $\int_{z \in \mathbb{D}_s} \xi^1_+(z)\, d\lambda$. The limit of this for $s \to 1$ exists and is finite. In fact by $S(\xi, \zeta)$ it is equal to

$$\int_{z \in S^1} \zeta^1_+(z)\, d\lambda_0.$$

Therefore we can apply the monotone convergence theorem (Beppo Levi) and conclude that $g^1_{s,+}$ is also defined for $s = 1$, as a measurable and Lebesgue integrable nonnegative function on S^1. We can proceed similarly for ξ^1_-, ξ^2_+ and ξ^2_-. Then we define

$$G_\xi(z, 1) := \begin{bmatrix} g^1_{s,+}(z, 1) - g^1_{s,-}(z, 1) \\ g^2_{s,+}(z, 1) - g^2_{s,-}(z, 1) \end{bmatrix}$$

for $z \in S^1$. The statement concerning the distributional partial derivative $\partial/\partial s$ along $\{(z, s) \mid s = 1\}$ is then clear. The implication $T(\xi, \zeta) \Rightarrow S(\xi, \zeta)$ is also clear.

If h is an argument-preserving diffeomorphism $\mathbb{D} \cup S^1 \to \mathbb{D} \cup S^1$, we can write $h(sz) = N(z, s) \cdot z$ as in the definition of *argument-preserving*. Let H be the diffeomorphism from $S^1 \times [0, 1]$ to $S^1 \times [0, 1]$ defined by $H(z, s) = (z, N(z, s))$. (This satisfies $q \circ H = h \circ q$, where $q : S^1 \times [0, 1] \to \mathbb{D} \cup S^1$ is defined by $(z, s) \mapsto sz$.)

Integration by substitution implies that

$$G_{h^*\xi} = G_{f\cdot\xi} \circ H$$

for a "suitable" smooth map $f : \mathbb{D} \cup S^1 \to \mathrm{End}_{\mathbb{R}}(\mathbb{R}^2)$. The formula for f is complicated. It is easier to describe $f \circ h \circ q$: namely,

$$f\big(h(q(z,s))\big) = \frac{s}{N(z,s)} \cdot \left(\frac{dN}{ds}(z,s)\right)^{-1} \cdot (dh(sz))^{-1}$$

for $z \in S^1$ and $s \in [0, 1]$. (The first two factors in this product of three are real numbers, but the third one is a linear map.) The important features are: $f(z)$ is the unit of $\mathrm{End}_{\mathbb{R}}(\mathbb{R}^2)$ for z close to the origin, $f(z)$ always respects the linear subspace of $\mathbb{R}^2$ spanned by z, and if $|z| = 1$ it also respects the linear subspace perpendicular to z and restricts to the identity there.

This allows us to argue as follows: $T(\xi, \zeta)$ implies $T(f\cdot\xi, f\cdot\zeta)$ easily, and from there we can deduce $T(h^*\xi, \zeta)$ by an application of the chain rule. □

A.3. *Questions and suggestions.*

A.3.1. Let ξ be a harmonic vector field on $\mathbb{D}$ (with the Poincaré metric) which is boundary controlled (notation as in Definition 1.3.3). If the distributional boundary is identically zero, does it follow that ξ is identically zero?

A.3.2. Find a more direct proof of Proposition 1.4.3, i.e., one which does not rely on Theorem I. (Do use Lemma 1.4.2 and look up a proof of this.)

A.3.3. Find a practical characterization of the tangential L^2-vector fields ζ on S^1 such that the vector field $F(\zeta)$ on $\mathbb{D}$ (as in Theorem II) is quasiconformal (Definition 1.3.4).

Authorship

This work is in all essentials the PhD thesis of Divya Sharma (PhD 2021, University of Münster), lightly revised by Weiss, who was the research supervisor at the time. Since Sharma is no longer engaged in mathematics research, it fell to Weiss to make arrangements for getting the work published.

References

[Dodson et al. 2002] C. T. J. Dodson, M. Trinidad Pérez, and M. E. Vázquez-Abal, "Harmonic-Killing vector fields", *Bull. Belg. Math. Soc. Simon Stevin* **9**:4 (2002), 481–490. MR

[Eells and Sampson 1964] J. Eells, Jr. and J. H. Sampson, "Harmonic mappings of Riemannian manifolds", *Amer. J. Math.* **86** (1964), 109–160. MR

[Gerstenhaber and Rauch 1954a] M. Gerstenhaber and H. E. Rauch, "On extremal quasi-conformal mappings, I", *Proc. Nat. Acad. Sci. U.S.A.* **40** (1954), 808–812. MR

[Gerstenhaber and Rauch 1954b] M. Gerstenhaber and H. E. Rauch, "On extremal quasi-conformal mappings, II", *Proc. Nat. Acad. Sci. U.S.A.* **40** (1954), 991–994. MR

[Imayoshi and Taniguchi 1992] Y. Imayoshi and M. Taniguchi, *An introduction to Teichmüller spaces*, Springer, 1992. MR

[Jost 1984] J. Jost, *Harmonic maps between surfaces*, Lect. Notes Math. **1062**, Springer, 1984. MR

[Markovic 2017] V. Markovic, "Harmonic maps and the Schoen conjecture", *J. Amer. Math. Soc.* **30**:3 (2017), 799–817. MR

[Markovic and Šarić 2009] V. Markovic and D. Šarić, "The universal properties of Teichmüller spaces", pp. 261–294 in *Surveys in differential geometry*, *XIV*: *Geometry of Riemann surfaces and their moduli spaces*, International Press, Somerville, MA, 2009. MR

[Ransford 1995] T. Ransford, *Potential theory in the complex plane*, London Math. Soc. Student Texts **28**, Cambridge Univ. Press, 1995. MR

[Shubin 2020] M. Shubin, *Invitation to partial differential equations*, Grad. Stud. Math. **205**, Amer. Math. Soc., Providence, RI, 2020. MR

[Weil 1960] A. Weil, "On discrete subgroups of Lie groups", *Ann. of Math.* (2) **72** (1960), 369–384. MR

[Wolpert 1986] S. A. Wolpert, "Chern forms and the Riemann tensor for the moduli space of curves", *Invent. Math.* **85**:1 (1986), 119–145. MR

[Wolpert 1987] S. A. Wolpert, "Geodesic length functions and the Nielsen problem", *J. Differential Geom.* **25**:2 (1987), 275–296. MR

Received November 7, 2024. Revised June 30, 2025.

DIVYA SHARMA
UNESCO
PARIS
FRANCE
sanu1793@gmail.com

MICHAEL S. WEISS
MATHEMATISCHES INSTITUT
UNIVERSITÄT MÜNSTER
MÜNSTER
GERMANY
m.weiss@uni-muenster.de

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 journals+pjm@msp.org 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.

At submission time a PDF file is required, and it must be derived from a LaTeX source file in which the command `\documentclass[12pt]{amsart}` is active. No changes to horizontal or vertical margins, or line spacing, should be made. Please include in the preamble the command `\usepackage[pagewise]{lineno}\linenumbers` — this will spare referees tedious effort in line counting.

Carefully preserve all related 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 journals+pjm@msp.org.

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 338 No. 2 October 2025
