Pacific Journal of Mathematics

Volume 306 No. 1 May 2020

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 2020 is US \$520/year for the electronic version, and \$705/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. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.1

TENSOR STRUCTURE FOR NORI MOTIVES

LUCA BARBIERI-VIALE, ANNETTE HUBER AND MIKE PREST

We construct a tensor product on Freyd's universal abelian category $\mathbf{Ab}(C)$ attached to an additive tensor category or a $\otimes$-quiver and establish a universal property. This is used to give an alternative construction for the tensor product on Nori motives.

Introduction

In the late 1990s, Nori made a spectacular proposal for an unconditional definition of an abelian category of motives and a motivic Galois group over a field of characteristic zero. It has two main inputs:

(1) The existence of a universal abelian category attached to a fixed representation of a quiver.

(2) His basic lemma (known earlier to Beilinson and Vilonen) which shows the existence of an algebraically defined "skeletal filtration" on an affine algebraic variety.

The first part is enough to give the definition of the category. The second is needed

---

Barbieri-Viale acknowledges the support of the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) through the Research Project (PRIN 2010-11) "Arithmetic Algebraic Geometry and Number Theory" and the Freiburg Institute of Advanced Study. Huber thankfully acknowledges support by the Freiburg Institute of Advanced Study during the preparation of this note. Prest gratefully acknowledges the support of the Freiburg Institute of Advanced Study.
MSC2010: primary 14F99; secondary 18E10, 18E30, 03C60.
Keywords: Nori motive, tensor category, free abelian category.

in order to establish the tensor structure. In a third step, we pass from effective motives to all motives and check rigidity.

The motivic Galois group is its Tannaka dual. However, all steps are intrinsically linked together. The proof of the existence of the abelian category is done by constructing a suitable coalgebra. The tensor product is defined by turning this coalgebra into a bialgebra. After localisation, it is shown to be even a Hopf algebra — the Hopf algebra of the motivic Galois group. Indeed, the proof given in full detail in [Huber and Müller-Stach 2017] gives as a byproduct a full proof of Tannaka duality.

Meanwhile there have been a couple of alternative approaches to the first step of the above program; see [Barbieri-Viale et al. 2018; Barbieri-Viale 2017; Barbieri-Viale and Prest 2018; Ivorra 2017]. They are more general and arguably simpler. However, these references did not address tensor products.

In this paper we explain how the approach of [Barbieri-Viale and Prest 2018] can be used to handle tensor categories and tensor functors. We show that if $(C, \otimes)$ is an additive tensor category then Freyd's universal abelian category $\mathrm{Ab}(C)$ carries an induced right-exact tensor structure which is also universal in a certain sense (the exact statement is Proposition 1.10).

Given a module M on C (i.e., an additive functor into an abelian tensor category), this induces, under additional technical assumptions, a tensor structure on the universal abelian category $\mathcal{A}(M)$ for the module M. This is again universal; see Proposition 1.13. The results can also be reformulated in terms of representations of quivers (see Section 2, in particular Theorem 2.10) bringing it even closer to the shape of Nori's original results. Our results are a lot more general in allowing modules with values in quite general abelian categories. We get back Nori's case as the case of the representation of a quiver in the category of modules over a Dedekind domain or even a noetherian ring of homological dimension at most 2.

We also show how to apply our results to Nori motives. This can be done by using his original quiver of good pairs. Alternatively, we start with the more canonical tensor category of geometric motives in the sense of Voevodsky. However, the functor H^0_B used in the definition of Nori motives is *not* a tensor functor, in contrast with the graded functor H^*_B. It remains to check that the Künneth components are motivic. This step of the construction relies on Nori's basic lemma. We give an abstract criterion in Section 3. It is applied to Nori motives in Section 4: we obtain a Tannakian category and define the motivic Galois group as its Tannaka dual. We find this more natural than defining the category as representations of the motivic Galois group.

We feel that the nature of the argument and the role of the basic lemma become a lot clearer in this new description. However, its main advantage is the great generality in choosing the target category $\mathcal{A}$. For example, we can easily define Nori motives over a base by using the Betti-realisation of triangulated motives into constructible sheaves. In the follow-up [Barbieri-Viale and Prest 2020], we take a more axiomatic approach, using many-sorted languages such that (co)homology theories are models of certain regular theories in that language. All this applies to several different geometric situations.

Notation. By a tensor category $(C, \otimes)$ we mean a category C provided with a functor $\otimes : C \times C \to C$ satisfying an associativity constraint and with $\mathbf{1}$ a unit object; in addition, also a commutativity constraint can be required, e.g., see [Deligne and Milne 1982, §1]. This is often called a nonstrict tensor category. By an additive (resp. abelian) tensor category we mean a tensor category $(C, \otimes)$ such that C is additive (resp. abelian) and $\otimes$ is a biadditive functor; see [Deligne and Milne 1982, Definition 1.15]. Tensor functors are not assumed strict. Tensor functors between additive tensor categories are assumed to be additive. We denote by $\mathbb{Q}$-vsp. the tensor category of $\mathbb{Q}$-vector spaces.

If $\mathcal{A}$ is an abelian category, we denote by $\operatorname{gr}\mathcal{A}$ the associated category of $\mathbb{Z}$-graded objects. If, in addition, $(\mathcal{A}, \otimes)$ carries a tensor structure, we equip $(\operatorname{gr}\mathcal{A}, \otimes)$ with the induced tensor structure. If the tensor product is commutative, we choose the commutativity constraint on $\operatorname{gr}\mathcal{A}$ such that the product becomes graded anticommutative.

For an additive category C we shall consider the additive functors from C to the category Ab of abelian groups as (left) C-modules. We shall denote by C-mod the category of finitely presented C-modules; see, e.g., [Prest 2011, Chapters 2 and 3].

1. Universal abelian tensor categories

Let C be an additive category. We denote by $\mathrm{Ab}(C)$ the universal abelian category on C; see [Freyd 1966; Prest 2011, Chapter 4]. We may refer to it as Freyd's abelian category. It comes with a canonical fully faithful functor $C \hookrightarrow \mathrm{Ab}(C)$. Recall that this functor is universal with respect to additive functors into abelian categories, e.g., see [Barbieri-Viale and Prest 2018, Theorem 1.1].

Thus, for $M : C \to \mathcal{A}$ an additive functor into some abelian category $\mathcal{A}$, we obtain an induced exact functor $\widetilde{M} : \mathrm{Ab}(C) \to \mathcal{A}$, unique to natural equivalence.

We denote by $\mathcal{A}(M)$ the quotient of $\mathrm{Ab}(C)$ by the Serre subcategory which is the kernel of $\widetilde{M}$; we also denote by $\widetilde{M} : \mathcal{A}(M) \to \mathcal{A}$ the induced faithful exact functor:

$$\begin{tikzcd} C \arrow[r] \arrow[dd, "M"'] & \mathrm{Ab}(C) \arrow[d] \arrow[ddl, "\widetilde{M}"'] \\ & \mathcal{A}(M) \arrow[dl, "\widetilde{M}"] \\ \mathcal{A} & \end{tikzcd}$$

We shall refer to $\mathcal{A}(M)$ as the *universal abelian category defined by M*, according with [Barbieri-Viale and Prest 2018, §1.1]. In fact, this abelian category $\mathcal{A}(M)$ is universal for (i.e., initial among) all abelian categories together with a faithful exact functor into $\mathcal{A}$ which extends M. Note that, in the case where $\mathcal{A}$ is the category of finitely generated modules over a commutative noetherian ring R, this recovers Nori's abelian category (see [Huber and Müller-Stach 2017, Chapter 7] and compare with [Barbieri-Viale and Prest 2018, §1.2]). For later use, we introduce:

Definition 1.1. Let C be additive and let $C \to \mathrm{Ab}(C)$ be Freyd's abelian category. We denote by $\mathrm{Ab}(C)^\flat$ the smallest full subcategory containing the objects in the image of C and closed under kernels.

Remark 1.2. The universal abelian category $\mathrm{Ab}(C)$ can be constructed explicitly as the category (C-mod)-mod (see, e.g., [Prest 2011, 4.3]). In this construction, $\mathrm{Ab}(C)^\flat$ is, because C-mod has cokernels and every object of C-mod is the cokernel of a morphism between representables, precisely the image of C-mod under the (contravariant) Yoneda embedding into $\mathrm{Ab}(C)$. All objects of $\mathrm{Ab}(C)^\flat$ are representable functors, and hence, by the Yoneda lemma, projective. The above definition is independent of this description.

Let $(C, \otimes)$ be an additive tensor category; see [Deligne and Milne 1982, §1]. Consider an (additive) tensor functor $M : (C, \otimes) \to (\mathcal{A}, \otimes)$ where $(\mathcal{A}, \otimes)$ is an abelian tensor category. We want to equip the above universal abelian category $\mathcal{A}(M)$ with a natural tensor structure $(\mathcal{A}(M), \otimes)$ such that $\widetilde{M} : (\mathcal{A}(M), \otimes) \to (\mathcal{A}, \otimes)$ is turned into a tensor functor. We proceed in several steps.

Multilinear functors. By definition, $\mathrm{Ab}(C)$ has a universal property with respect to additive functors. In fact, this extends to biadditive and even multiadditive functors, even though we lose some properties.

We first recall a well-known property of injective resolutions.

Lemma 1.3. *Let $\mathcal{A}$ be an abelian category, $f : X \to Y$ a morphism in $\mathcal{A}$. Assume*

$$0 \to X \to I_0 \to I_1 \quad \text{and} \quad 0 \to Y \to J_0 \to J_1$$

are exact and that all I_k, J_k are injective. Then there are lifts $f_0 : I_0 \to J_0$, $f_1 : I_1 \to J_1$ making the diagram

$$\begin{tikzcd}
0 \arrow[r] & X \arrow[r] \arrow[d, "f"'] & I_0 \arrow[r, "d"] \arrow[d, "f_0"'] & I_1 \arrow[d, "f_1"'] \\
0 \arrow[r] & Y \arrow[r] & J_0 \arrow[r] & J_1
\end{tikzcd}$$

commute. Moreover, if (g_0, g_1) is a second lift, then there is $h : I_1 \to J_0$ such that

$$f_0 - g_0 = h \circ d.$$

Proof. Our complexes are the starting bits of injective resolutions and h is the beginning of a chain homotopy. The assertion is usually proved as the first step of the proof of existence of a lift of f to an injective resolution and that the lift is unique up to chain homotopy; see for example [Mac Lane 1963, Theorem 6.1]. $\square$

As usual, we also have the dual statement for left resolutions by projectives.

Proposition 1.4. *Let $C_1, \dots, C_n$ be additive categories and let $\mathcal{A}$ be an abelian category.*

(1) *Let $F : C_1 \times \cdots \times C_n \to \mathcal{A}$ be a multilinear functor, i.e., additive in each argument. Then F extends to a multilinear functor*

$$\widetilde{F} : \mathrm{Ab}(C_1) \times \mathrm{Ab}(C_2) \times \cdots \times \mathrm{Ab}(C_n) \to \mathcal{A}$$

which is right-exact in each argument. Fix j and for $i \neq j$ choose $X_i \in \mathrm{Ab}(C_i)^\flat$ (see Definition 1.1). Then $\widetilde{F}(X_1, \dots, -, \dots, X_n)$ is exact as a functor on $\mathrm{Ab}(C_j)$.

(2) *The functor F is uniquely determined up to unique isomorphism of functors by these properties.*

(3) *Let $\alpha : F_1 \to F_2$ be a transformation of multilinear functors $C_1 \times \cdots \times C_n \to \mathcal{A}$ and $\widetilde{F}_1$ and $\widetilde{F}_2$ their extensions to $\mathrm{Ab}(C_1) \times \cdots \times \mathrm{Ab}(C_n)$. Then there is a transformation of functors $\tilde{\alpha} : \widetilde{F}_1 \to \widetilde{F}_2$ extending α. It is unique.*

Proof. Recall that $\mathrm{Ab}(C_i) = (C_i\text{-mod})\text{-mod}$ and that the universal functor factors

$$C_i \to (C_i\text{-mod})^{\mathrm{op}} \to (C_i\text{-mod})\text{-mod}$$

where both steps are given by the Yoneda embedding. As pointed out in Remark 1.2 the subcategory $\mathrm{Ab}(C_i)^\flat$ agrees with the image of $(C_i\text{-mod})^{\mathrm{op}}$.

All statements are shown in two steps. In the first we extend to a functor

$$F' : (C_1\text{-mod})^{\mathrm{op}} \times (C_2\text{-mod})^{\mathrm{op}} \times \cdots \times (C_n\text{-mod})^{\mathrm{op}} \to \mathcal{A}$$

which will be multilinear and left-exact in each argument. In the second step, which is actually dual to the first, we extend F' to $\widetilde{F}$.

We first show uniqueness. This will make clear why the formula that we use in the construction is correct. Let E be any extension of F to $\mathrm{Ab}(C_i)$ with the exactness property of (1). Let $X_i \in (C_i\text{-mod})^{\mathrm{op}}$. We argue by descending induction on the number of X_i which are in the image of C_i, i.e., of the form $(A_i, -)^{\mathrm{op}}$. If they all are, then $E(X_1, \dots, X_n) = F(A_i, \dots, A_n)$ by assumption. Assume that E is uniquely determined if at least m of the X_i are corepresentable. After reordering we have to consider the tuple $(X_1, \dots, X_n)$ with $X_i = (A_i, -)^{\mathrm{op}}$ for $i < m$. By definition, there is an injective corepresentation

$$0 \to X_m \to (A_m, -)^{\mathrm{op}} \to (B_m, -)^{\mathrm{op}}.$$

By (1), the functor $E(X_1,\dots,X_{m-1},-,X_{m+1},\dots,X_n)$ is exact. Hence we have an exact sequence

$$0\to E(X_1,\dots,X_n)\to E(X_1,\dots,X_{m-1},(A_m,-)^{\mathrm{op}},X_{m+1},\dots,X_n)$$
$$\to E(X_1,\dots,X_{m-1},(B_m,-)^{\mathrm{op}},X_{m+1},\dots,X_n).$$

By induction the two terms on the right are uniquely determined up to unique isomorphism. As a kernel, $E(X_1,\dots,X_n)$ is again uniquely determined up to unique isomorphism. By induction, this shows uniqueness if all arguments are in C_i-mod.

The dual argument for a right-exact E and representations gives uniqueness for arguments in $\mathrm{Ab}(C_i)$.

We turn to the construction of F'. Let $X_i\in(C_i\text{-mod})^{\mathrm{op}}$. By definition, these objects have an injective copresentation

$$0\to X_i\to(A_i,-)^{\mathrm{op}}\to(B_i,-)^{\mathrm{op}}.$$

We choose such a presentation for each object $X_i\in(C_i\text{-mod})^{\mathrm{op}}$. The uniqueness proof suggests $F'(X_1,\dots,X_n)\subset F(A_1,\dots,A_n)$ as an iteration of kernels. The same object is given by the formula

$$F'(X_1,\dots,X_n):=\mathrm{Ker}\Big(F(A_1,\dots,A_n)\to\bigoplus_{m=1}^{n}F(A_1,\dots,A_{m-1},B_m,A_{m+1},\dots,A_n)\Big).$$

In other words, applying F to the n-tuple of complexes $(A_i,-)^{\mathrm{op}}\to(B_i,-)^{\mathrm{op}}$ we obtain an n-fold complex. The above is H^0 of its total complex.

Let $1\le i\le n$. For a morphism $X_i\to Y_i$ in $(C_i\text{-mod})^{\mathrm{op}}$ we choose a lift to the copresentations as in Lemma 1.3. This induces a morphism $F'(X_1,\dots,X_n)\to F'(X_1,\dots,Y_i,\dots,X_n)$. It is independent of the lift because any two such morphisms differ by h as in Lemma 1.3. This makes F' a functor in each variable.

The dual argument via projective presentations gives a right-exact extension to $\mathrm{Ab}(C_i)$.

A diagram chase shows that the functor $\widetilde{F}$ has the exactness property claimed in (1) because F' is left-exact, $\widetilde{F}$ right-exact and every object Y of $\mathrm{Ab}(C_i)$ has a projective resolution of the form

$$0\leftarrow Y\leftarrow P^0\leftarrow P^1\leftarrow P^2\leftarrow 0$$

with $P^i\in(C_i\text{-mod})^{\mathrm{op}}$.

Now let $\alpha:F_1\to F_2$ be a transformation of functors. Going through the above construction, we get induced $\alpha':F_1'\to F_2'$ and then $\tilde{\alpha}:\widetilde{F}_1\to\widetilde{F}_2$. The uniqueness argument for the functors also gives the uniqueness of the transformation. □

Remark 1.5. Unexpectedly the extension $\widetilde{F}$ *fails* to be exact in each argument. For a counterexample, see Example 1.12 below.

Remark 1.6. In a general enriched-category setting, lifting monoidal structure to functor categories can be found in [Bunge 1969] and [Day 1970].

This applies in particular to additive tensor categories.

Definition 1.7. Let $(C, \otimes)$ be an additive tensor category. We extend the functor $\otimes : C \times C \to C$ defining

$$\otimes : \mathrm{Ab}(C) \times \mathrm{Ab}(C) \to \mathrm{Ab}(C)$$

as the extension of $C \times C \to C \hookrightarrow \mathrm{Ab}(C)$ of Proposition 1.4.

Proposition 1.8. *Let* $(\mathrm{Ab}(C), \otimes)$ *be Freyd's category together with the functor in Definition 1.7. Then*:

(1) $(\mathrm{Ab}(C), \otimes)$ *is an abelian tensor category.*

(2) *The tensor product is right-exact by construction. The objects in* $\mathrm{Ab}(C)^\flat$ *are flat*, *i.e.*, *acyclic with respect to* $\otimes$.

(3) *If the tensor structure on* C *is commutative then so is the tensor structure on* $\mathrm{Ab}(C)$.

Proof. Right-exactness and acyclicity are special cases of Proposition 1.4. Let $\mathbf{1}$ be the unit object of C. By definition it comes with a transformation of functors $u : \mathbf{1} \otimes - \to \mathrm{id}$ on C. Let $[\mathbf{1}]$ be its image in $\mathrm{Ab}(C)$. In explicit formulas this means $[\mathbf{1}] = ((\mathbf{1}, -), -)$. Then $[\mathbf{1}]$ with the induced transformation is the unit of $\mathrm{Ab}(C)$. The equivalences used to express the associativity constraint on C^3 (see [Deligne and Milne 1982, §1]) induce equivalences on $\mathrm{Ab}(C)^3$. In detail: Let

$$F_1 : C^3 \xrightarrow{\otimes \circ (\mathrm{id}, \otimes)} C \quad \text{and} \quad F_2 : C^3 \xrightarrow{\otimes \circ (\otimes, \mathrm{id})} C.$$

The associativity constraint is a functorial isomorphism $\alpha : F_1 \to F_2$. By abuse of notation we use the same notation for their composition with the inclusion $C \to \mathrm{Ab}(C)$. Note that α is still a functorial isomorphism. The functor $\mathrm{Ab}(C)^3 \to \mathrm{Ab}(C)$ given by $(X, Y, Z) \mapsto X \otimes (Y \otimes Z)$ is right-exact in each argument and exact as a functor in one variable if the other entries are flat. By the uniqueness property of Proposition 1.4 it agrees with $\widetilde{F}_1$. The same argument also applies to F_2. Again by Proposition 1.4, the transformation extends to a transformation $\tilde{\alpha}$. This is our associativity constraint. We need to check that a certain diagram of functors on $\mathrm{Ab}(C)^4$ involving $\otimes$ and $\tilde{\alpha}$ commutes. This holds by the uniqueness part of Proposition 1.4 applied to functors $C^4 \to C$.

We argue similarly for the commutativity constraint if there is one on C. □

Definition 1.9. For an abelian tensor category, with a right-exact tensor product, a $\flat$-*subcategory* is a full additive subcategory of flat objects (i.e., acyclic with respect to the tensor product) which is closed under kernels. If $(\mathcal{A}, \otimes)$ is such an abelian tensor category we shall denote by $\mathcal{A}^\flat \subseteq \mathcal{A}$ some $\flat$-subcategory.

As a consequence of Proposition 1.8 we have that $\mathrm{Ab}(C)^\flat \subset \mathrm{Ab}(C)$ as in Definition 1.1 is a $\flat$-subcategory.

Proposition 1.10 (universal property). *Let C be an additive tensor category. Let $\mathcal{A}$ be an abelian tensor category with a right-exact tensor product. Let $M : (C, \otimes) \to (\mathcal{A}, \otimes)$ be a tensor functor. In addition, assume that M factors via $\mathcal{A}^\flat \subseteq \mathcal{A}$ a $\flat$-subcategory (see Definition 1.9). Then $\widetilde{M} : (\mathrm{Ab}(C), \otimes) \to (\mathcal{A}, \otimes)$ is a tensor functor. The triple $(\mathrm{Ab}(C), \mathrm{Ab}(C)^\flat, \otimes)$ is universal with this property, and in particular unique.*

Proof. Let $M : (C, \otimes) \to (\mathcal{A}, \otimes)$ be a tensor functor. We have to compare

$$\mathrm{Ab}(C) \times \mathrm{Ab}(C) \to \mathrm{Ab}(C) \to \mathcal{A}$$

and

$$\mathrm{Ab}(C) \times \mathrm{Ab}(C) \to \mathcal{A} \times \mathcal{A} \to \mathcal{A}.$$

Both are right-exact in each argument (this is where right-exactness of the tensor product on $\mathcal{A}$ is used) and agree on $C \times C$.

As in the proof of Proposition 1.4, we extend M in two steps: first to $(C\text{-mod})^{\mathrm{op}}$, then to $\mathrm{Ab}(C) = (C\text{-mod})\text{-mod}$. The second step is unproblematic as it only uses the right-exactness. In the first step, we need to check the action on (certain) kernels. Let $X_1, X_2 \in (C\text{-mod})^{\mathrm{op}}$ with resolutions

$$0 \to X_i \to (A_i, -)^{\mathrm{op}} \to (B_i, -)^{\mathrm{op}}.$$

By definition

$$0 \to M'(X_i) \to M(A_i) \to M(B_i)$$

is exact. By assumption $M(A_i), M(B_i)$ and hence also $M'(X_i)$ are in $\mathcal{A}^\flat$. In particular, consider the diagram

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] \\
0 \arrow[r] & M'(X_1) \otimes M'(X_2) \arrow[r] \arrow[d] & M'(X_1) \otimes M(A_2) \arrow[r] \arrow[d] & M'(X_1) \otimes M(B_2) \arrow[d] \\
0 \arrow[r] & M(A_1) \otimes M'(X_2) \arrow[r] \arrow[d] & M(A_1) \otimes M(A_2) \arrow[r] \arrow[d] & M(A_1) \otimes M(B_2) \arrow[d] \\
0 \arrow[r] & M(B_1) \otimes M'(X_2) \arrow[r] & M(B_1) \otimes M(A_2) \arrow[r] & M(B_1) \otimes M(B_2)
\end{tikzcd}$$

All rows and columns are exact because they arise by tensoring an exact sequence with a flat object. This implies

$$\begin{aligned}
M'(X_1)\otimes M'(X_2)\\
&= \operatorname{Ker}\big(M(A_1)\otimes M(A_2) \to (M(A_1)\otimes M(B_2))\oplus(M(B_1)\otimes M(A_2))\big)\\
&= M'(X_1\otimes X_2).
\end{aligned}$$

The triple $(\mathrm{Ab}(C), \mathrm{Ab}(C)^\flat, \otimes)$ itself satisfies the assumptions of the universal property; hence it is universal and as such unique. □

Remark 1.11. There are a number of interesting cases where the assumptions of Proposition 1.10 and Definition 1.9 are satisfied. However, they are not as general as one could hope for.

(1) If $\otimes$ is exact on $\mathcal{A}$, then $\mathcal{A}^\flat = \mathcal{A}$ clearly satisfies the assumptions.

(2) If $C_1 \to C_2$ is a $\otimes$-functor between additive tensor categories, by composition, we may consider $M : C_1 \to \mathcal{A}^\flat = \mathrm{Ab}(C_2)^\flat \subset \mathcal{A} = \mathrm{Ab}(C_2)$ which satisfies the assumptions; then, by the universal property, we get an exact tensor functor $\widetilde{M} : \mathrm{Ab}(C_1) \to \mathrm{Ab}(C_2)$.

(3) The assumptions are satisfied if $\mathcal{A} = R$-mod for a Dedekind domain R where $\mathcal{A}^\flat$ is the $\flat$-subcategory of projective finitely generated R-modules, i.e., torsion free finitely generated modules, and $M : C \to \mathcal{A}^\flat$ is any tensor functor. In particular this is true for $R = \mathbb{Z}$.

(4) They are not satisfied for $\mathcal{A} = R$-mod for a general noetherian commutative ring R and the subcategory of projective finitely generated R-modules, which is not a $\flat$-subcategory if the global dimension of R is > 2. See Example 1.12.

Example 1.12. Let C be the category with objects $(\mathbb{Z}/4)^n$ for $n \geq 0$ and morphisms given by homomorphisms of abelian groups.

Let $\mathcal{A} = \mathbb{Z}/4$-mod. In this case is possible to compute all objects explicitly. The functor $M \mapsto M^\vee = \mathrm{Hom}(M, \mathbb{Z}/4)$ is an antiequivalence of C with itself. We have C-mod $\cong \mathbb{Z}/4$-mod with $C \to \mathbb{Z}/4$-mod given by $M \mapsto M^\vee$. Hence $\mathrm{Ab}(C)$ is the category of finitely presented presheaves on $\mathbb{Z}/4$-mod. Objects are uniquely determined by the values of these presheaves on the groups $\mathbb{Z}/4$ and $\mathbb{Z}/2$. Direct computation will show:

(1) $\otimes$ is not biexact on $\mathrm{Ab}(C)$.

(2) The tensor functor $\mathrm{Ab}(C) \to \mathcal{A}$ induced by the inclusion functor $C \to \mathcal{A}$ is not a tensor functor.

By Auslander–Reiten theory (see, e.g., [Assem et al. 2006, §IV.6, p. 149]) the simple objects of the category $\mathrm{Ab}(C)$ have the form $(X, -)/\mathrm{rad}(X, -)$ for X an indecomposable $\mathbb{Z}/4$-module. So there are two simple objects, S and T say, and these are such that $S(\mathbb{Z}/4) = \mathbb{Z}/2$, $S(\mathbb{Z}/2) = 0$ and $T(\mathbb{Z}/4) = 0$, $T(\mathbb{Z}/2) = \mathbb{Z}/2$. Noting the exact sequence $0 \to \mathbb{Z}/2 \xrightarrow{j} \mathbb{Z}/4 \xrightarrow{p} \mathbb{Z}/2 \to 0$ and considering the maps $(p, -)$ and $(j, -)$ in $\mathrm{Ab}(C)$, it can be easily checked that $\mathrm{rad}(\mathbb{Z}/4, -) = (\mathbb{Z}/2, -)$ and that $(\mathbb{Z}/2, -)$ has length 2, with socle S. The remaining indecomposable objects of $\mathrm{Ab}(C)$ may then be computed (see, for example, [Prest 2012, 4.3]): there are five of them, all of them subquotients of the two representable functors. They are $(\mathbb{Z}/4, -)$, $(\mathbb{Z}/2, -)$, the two simples S, T and $(\mathbb{Z}/4, -)/S$.

Now consider the exact functor $\widetilde{\mathbb{Z}/4} : \mathrm{Ab}(C) \to \mathbb{Z}/4\text{-mod}$. This is evaluation of an object of $\mathrm{Ab}(C)$, considered as a functor on $\mathbb{Z}/4$-mod, at $\mathbb{Z}/4$, hence is 0 only on T among those five indecomposables. Therefore its kernel is the Serre subcategory which consists of direct sums of copies of T. In order to compute $T \otimes T$, we apply the definition of the tensor product on $\mathrm{Ab}(C)$ using the projective presentation

$$(\mathbb{Z}/4, -) \xrightarrow{(j,-)} (\mathbb{Z}/2, -) \xrightarrow{\pi_T} T \to 0$$

of T and, checking

$$(\mathrm{id}_{\mathbb{Z}/2}, -) \otimes (j, -) = 0,$$

we obtain $T \otimes T = (\mathbb{Z}/2, -)$, which is not in the kernel of $\widetilde{\mathbb{Z}/4}$, so this is not a tensor functor. As part of the computation of $T \otimes T$ one sees that

$$T \otimes (\mathbb{Z}/2, -) = (\mathbb{Z}/2, -).$$

So applying $T \otimes -$ to the monomorphism

$$(\mathbb{Z}/2, -) \xrightarrow{(p,-)} (\mathbb{Z}/4, -)$$

gives $(\mathbb{Z}/2, -) \to T$ which is not monic, showing that $\otimes$ is not exact on $\mathrm{Ab}(C)$.

This implies that we cannot expect a different, exact, tensor product on $\mathrm{Ab}(C)$ extending the tensor product on C — by the universal property the identity would have to be a tensor functor.

Tensor structures on $\mathcal{A}(M)$. Consider $(\mathcal{A}, \otimes)$ an abelian tensor category with a right-exact tensor product.

For the sake of exposition we now drop explicit reference to $\otimes$ if unnecessary.

Proposition 1.13. *Let C be an additive tensor category, $\mathcal{A}$ an abelian tensor category with a right-exact tensor product, and $M : C \to \mathcal{A}$ an additive tensor functor. Further assume that M factors through a $\flat$-subcategory $\mathcal{A}^\flat \subset \mathcal{A}$ (see Definition 1.9).*

(1) *Then $\mathcal{A}(M)$ carries a canonical tensor structure such that the faithful exact functor $\widetilde{M} : \mathcal{A}(M) \to \mathcal{A}$ is a tensor functor.*

(2) *If in addition, the tensor structures on C and $\mathcal{A}$ are commutative and the tensor functor is symmetric, then the tensor product on $\mathcal{A}(M)$ is symmetric.*

(3) *If in addition, the tensor structure on C is rigid and the tensor product and the* Hom*-functor on $\mathcal{A}$ are exact in both arguments, the same is true for $\mathcal{A}(M)$.*

Proof. We need to check that the tensor functor on $\mathrm{Ab}(C)$ (see Proposition 1.8) factors via an induced tensor structure on $\mathcal{A}(M)$. We have a commutative diagram

$$\begin{tikzcd}
\mathrm{Ab}(C) \times \mathrm{Ab}(C) \arrow[r, "\otimes"] \arrow[d, "\widetilde{M}\times\widetilde{M}"'] & \mathrm{Ab}(C) \arrow[d, "\widetilde{M}"] \\
\mathcal{A} \times \mathcal{A} \arrow[r, "\otimes"] & \mathcal{A}
\end{tikzcd}$$

by Proposition 1.10. This implies that the kernel of $\mathrm{Ab}(C) \to \mathcal{A}$ is a $\otimes$-ideal. Hence the tensor product induces one on $\mathcal{A}(M)$. Associativity, unit and symmetry are immediate from the properties of the tensor structure on $\mathrm{Ab}(C)$.

We turn to rigidity. By assumption, every object X of C has a strong dual. By the criterion formulated in [Levine 1998, Part I, IV, Proposition 1.1.9] the existence of a dual for X can be characterized by the existence of unit and counit maps satisfying some compatibilities. In particular, this property is functorial, hence the image of X in $\mathcal{A}(M)$ also has a strong dual. Consider the full subcategory of $\mathcal{A}(M)$ consisting of objects with a strong dual. It contains all objects in the image of C. Under our assumptions on $\mathcal{A}$, the tensor product on $\mathcal{A}(M)$ is exact in both arguments and hence the subcategory is closed under kernels and cokernels. Hence it is an abelian subcategory of $\mathcal{A}(M)$ containing the image of C, hence it agrees with $\mathcal{A}(M)$. □

Proposition 1.14 (universal property). *Let C, $\mathcal{A}^\flat \subset \mathcal{A}$, and M be as defined in Proposition 1.13. In addition, let $\mathcal{B}$ be another abelian tensor category with $\flat$-subcategory $\mathcal{B}^\flat$, let $N : C \to \mathcal{B}$ be an additive tensor functor which factors through $\mathcal{B}^\flat$ and let $\phi : \mathcal{B} \to \mathcal{A}$ be a faithful exact functor mapping $\mathcal{B}^\flat$ to $\mathcal{A}^\flat$ such that $\phi \circ N = M$:*

$$\begin{tikzcd}
 & \mathcal{A}(M) \arrow[dr, "\widetilde{M}"] \arrow[dd, dotted] & \\
C \arrow[ur] \arrow[rr, "M"] \arrow[dr, "N"'] & & \mathcal{A} \\
 & \mathcal{B} \arrow[ur, "\phi"'] &
\end{tikzcd}$$

Then there exists a unique faithful exact tensor functor $\Phi : \mathcal{A}(M) \to \mathcal{B}$ making the diagram commute.

In particular, the universal property characterises $(\mathcal{A}(M), \widetilde{M})$ uniquely up to unique equivalence of categories.

Proof. The universal property of $\mathrm{Ab}(C)$ (see Proposition 1.10) gives us a similar statement, but with $\mathrm{Ab}(C)$ instead of $\mathcal{A}(M)$. The kernels of $\mathrm{Ab}(C) \to \mathcal{A}$ and $\mathrm{Ab}(C) \to \mathcal{B}$ agree because ϕ is faithful and exact. Hence $\mathcal{A}(M) = \mathcal{A}(N)$ and $\Phi = \widetilde{N}$. □

By a simple trick that we learned from Arapura [2013], this can be upgraded to a more general one.

Corollary 1.15 (generalised universal property). *Let C, $\mathcal{A}^\flat \subset \mathcal{A}$, and M be as in Proposition 1.13. In addition, let $\mathcal{B}$ be another abelian tensor category with $\flat$-subcategory $\mathcal{B}^\flat$, $N : C \to \mathcal{B}$ be an additive tensor functor which factors through $\mathcal{B}^\flat$. Let $(\mathcal{B}', \mathcal{B}'^\flat)$ be a third abelian tensor category, $\phi : \mathcal{A} \to \mathcal{B}'$ and $\psi : \mathcal{B} \to \mathcal{B}'$ faithful exact tensor functors respecting the $\flat$-subcategories. Finally, let $F : \psi \circ N \to \phi \circ M$ be an isomorphism of functors:*

$$\begin{tikzcd}
 & \mathcal{A}(M) \arrow[dr, "\widetilde{M}"] \arrow[dd, dotted] & \\
C \arrow[ur] \arrow[rr, "M"] \arrow[dr, "N"'] & & \mathcal{A} \arrow[d, "\phi"'] \\
 & \mathcal{B} \arrow[r, "\psi"'] & \mathcal{B}'
\end{tikzcd}$$

Then there exists a faithful exact tensor functor $\Phi : \mathcal{A}(M) \to \mathcal{B}$ making the diagram commute up to isomorphism of functors.

Proof. Let $\mathcal{C}$ be a tensor category with objects of the form (A, B, f) where $A \in \mathcal{A}$, $B \in \mathcal{B}$ and $f : \psi(B) \to \phi'(A)$. This category is abelian with kernels and cokernels taken componentwise. We equip it with a right-exact tensor product $(A, B, f) \otimes (A', B', f') = (A \otimes A', B \otimes B', f \otimes f')$. Let $\mathcal{C}^\flat$ be the subcategory of (A, B, f) with $A \in \mathcal{A}^\flat$, $B \in \mathcal{B}^\flat$.

Let $N' : C \to \mathcal{C}$ be the additive tensor functor $X \mapsto (M(X), N(X), F_X)$. We apply the universal property of Proposition 1.14 to N' and the forgetful functor $\phi : \mathcal{C} \to \mathcal{A}$. We define Φ as $\widetilde{N}'$ composed with the forgetful functor to $\mathcal{B}$. In other words, $\widetilde{N}'(X) = (\widetilde{M}(X), \widetilde{N}(X), F_X)$. The isomorphism of functors is given by F_X. □

Example 1.16. A possible application is with $\mathcal{A}$, $\mathcal{B}$, $\mathcal{B}'$ the categories of k-vector spaces, L-vector spaces and L'-vector spaces, respectively, for field extensions L'/k and L'/L.

Remark 1.17. This is a version of Nori's result on the tensor structure on his abelian category; see [Huber and Müller-Stach 2017, Proposition 8.1.5]. It is much stronger in allowing general abelian categories $\mathcal{A}$ as targets. In loc. cit. it was claimed that the original construction works for functors $C \to R$-proj (where the latter is the category of finitely generated projective modules over a noetherian ring R). However, as Paranjape pointed out, the proof is only correct if kernels of maps between projective modules are projective, i.e., if the global dimension of R is at most 2.

2. Universal ⊗-representation

We want to extend our results to representations of quivers. Given the results of the previous section, this means to extend tensor structures from a quiver to the additive category generated by it.

Recall from [Borceux 1994, Definition 5.1.5] or [Gabriel and Riedtmann 1979] the concept of a quiver "with relations", i.e., a quiver (a collection of vertices and directed edges) with a set of commutativity conditions or linear relations between paths (= compositions of directed edges). In this sense:

Definition 2.1. A $\otimes$-*quiver* is a quiver D with relations, plus the following data $(\mathrm{id}, \otimes, \alpha, \beta, \beta', \mathbf{1}, u)$ with,

(1) for every vertex v, a distinguished self-edge $\mathrm{id} : v \to v$;

(2) for every pair of vertices (v, w), a vertex denoted $v \otimes w$ in D;

(3) for every edge $e : v \to v'$ and vertex w, an edge $e \otimes \mathrm{id} : v \otimes w \to v' \otimes w$ and an edge $\mathrm{id} \otimes e : w \otimes v \to w \otimes v'$;

(4) for every pair of vertices u, v, a distinguished edge $\alpha_{u,v} : u \otimes v \to v \otimes u$;

(5) for every triple of vertices u, v, w, a distinguished edge $\beta_{u,vw} : u \otimes (v \otimes w) \to (u \otimes v) \otimes w$ and also $\beta'_{u,vw} : (u \otimes v) \otimes w \to u \otimes (v \otimes w)$;

(6) a distinguished vertex $\mathbf{1}$;

(7) for every vertex, distinguished edges $u_v : v \to \mathbf{1} \otimes v$ and $u'_v : \mathbf{1} \otimes v \to v$;

and the relations

(1) $\mathrm{id}_v \otimes \mathrm{id}_v = \mathrm{id}_{v \otimes v}$;

(2) $\mathrm{id}_v = e_v$ where e_v is the empty path for every vertex v;

(3) $(e \otimes \mathrm{id}) \circ (\mathrm{id} \otimes e') = (\mathrm{id} \otimes e') \circ (e \otimes \mathrm{id})$ for all pairs of edges e, e';

(4) $\alpha_{v,w} \circ \alpha_{w,v} = \mathrm{id}$ for all vertices v, w;

(5) $(\mathrm{id} \otimes \gamma) \circ \alpha = \alpha \circ (\gamma \otimes \mathrm{id})$ and $(\gamma \otimes \mathrm{id}) \circ \alpha = \alpha \circ (\mathrm{id} \otimes \gamma)$ for all edges γ;

(6) $\beta_{u,vw} \circ \beta'_{uv,w} = \mathrm{id}$ and $\beta'_{uv,w} \circ \beta_{u,vw} = \mathrm{id}$;

(7) $\beta \circ (\gamma \otimes (\mathrm{id} \otimes \mathrm{id})) = ((\gamma \otimes \mathrm{id}) \otimes \mathrm{id}) \circ \beta$ for all edges γ and analogously in the second and third argument;

(8) (pentagon axiom) for all vertices x, y, z, t the relation

$$\begin{tikzcd}
x \otimes (y \otimes (z \otimes t)) \arrow[r, "\beta"] \arrow[d, "\mathrm{id} \otimes \beta"'] & (x \otimes y) \otimes (z \otimes t) \arrow[r, "\beta"] & ((x \otimes y) \otimes z) \otimes t \\
x \otimes ((y \otimes z) \otimes t) \arrow[rr, "\beta"] & & (x \otimes (y \otimes z)) \otimes t \arrow[u, "\beta \otimes \mathrm{id}"']
\end{tikzcd}$$

(9) for all vertices x, y, z the relation

$$\begin{tikzcd} x\otimes(y\otimes z) \arrow[r, "\beta"] \arrow[d, "\mathrm{id}\otimes\alpha"'] & (x\otimes y)\otimes z \arrow[r, "\alpha"] & z\otimes(x\otimes y) \arrow[d, "\beta"] \\ x\otimes(z\otimes y) \arrow[r, "\beta"] & (x\otimes z)\otimes y \arrow[r, "\alpha\otimes\mathrm{id}"] & (z\otimes x)\otimes y \end{tikzcd}$$

(10) $u_v \circ u'_v = \mathrm{id}$ and $u'_v \circ u_v = \mathrm{id}$ for all vertices v;

(11) for all edges $e : v \to v'$ the relation

$$\begin{tikzcd} v' \arrow[r, "u"] & \mathbf{1}\otimes v' \\ v \arrow[u, "e"] \arrow[r, "u"] & \mathbf{1}\otimes v \arrow[u, "\mathrm{id}\otimes e"'] \end{tikzcd}$$

Remark 2.2. This data is modelled after the notion of a commutative product structure on a diagram with identities; see [Huber and Müller-Stach 2017, Definition 8.1.3] and the variant in [loc. cit., Remark 8.1.6]. The axioms for the associativity and commutativity constraint and unitality are the usual ones for a commutative tensor category; see [Deligne and Milne 1982, §1].

It is more general than the notion of a monoidal quiver introduced by Bruguières [2004, Section 5.2].

Recall [Barbieri-Viale and Prest 2018] where a universal representation

$$\Delta : D \to \mathrm{Ab}(D)$$

is constructed for any quiver D. It is given by the composition

$$\Delta : D \to \mathcal{P}(D) \to \mathbb{Z}D \to \mathbb{Z}D^+ \to \mathrm{Ab}(\mathbb{Z}D^+) = \mathrm{Ab}(D),$$

where (in the notation of [Barbieri-Viale and Prest 2018, §1]) $\mathcal{P}(D)$ is the path category, $\mathbb{Z}D$ the preadditive enrichment of $\mathcal{P}(D)$ and $\mathbb{Z}D^+$ its additive completion.

We now repeat the same chain with tensor categories. Let $(D, \otimes)$ be a $\otimes$-quiver. We define the $\otimes$-path category $\mathcal{P}(D)^{\otimes}$ as the quotient of the path category by the relations of $(D, \otimes)$. We define

$$\otimes : \mathcal{P}(D) \times \mathcal{P}(D) \to \mathcal{P}(D)$$

on objects as prescribed by the tensor structure. Let

$$\Gamma = \gamma_1 \circ \cdots \circ \gamma_n$$

and

$$\Delta = \delta_1 \circ \cdots \circ \gamma_m$$

be paths. We define

$$\Gamma \otimes \Delta = (\gamma_1 \otimes \mathrm{id}) \circ \cdots \circ (\gamma_n \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \delta_1) \circ \cdots \circ (\mathrm{id} \otimes \delta_m).$$

For example, for $n = m = 1$ and $\gamma : v \to v',\ \delta : w \to w'$, we have

$$\begin{tikzcd} v \otimes w \arrow[r, "\mathrm{id}\otimes\delta"] \arrow[d, "\gamma\otimes\mathrm{id}"'] \arrow[dr, "\gamma\otimes\delta"] & v \otimes w' \arrow[d, "\gamma\otimes\mathrm{id}"] \\ v' \otimes w \arrow[r, "\delta\otimes\mathrm{id}"'] & v' \otimes w' \end{tikzcd}$$

where we have by definition set the diagonal to be the path via the top right corner. In $\mathcal{P}(D)^{\otimes}$ this agrees with the path via the bottom left corner because of the relation (3).

Lemma 2.3. *Let $(D, \otimes)$ be a $\otimes$-quiver. Then $\mathcal{P}(D)^{\otimes}$ is a tensor category.*

Proof. Property (3) of a tensor structure ensures that $\otimes$ is a functor on $\mathcal{P}(D)^{\otimes}$. The other axioms make sure that the commutativity constraint α and the associativity constraint β are isomorphisms and satisfy the properties of a commutative tensor category. The relations on u_v ensure that $v \to \mathbf{1} \otimes v$ is an isomorphism and the functor $\mathbf{1} \otimes -$ is an equivalence of categories. □

Definition 2.4. Let $(D, \otimes)$ be a $\otimes$-quiver. We put $\mathbb{Z}D^{\otimes}$ and $\mathbb{Z}D^{\otimes,+}$ the preadditive and additive hull of $\mathcal{P}(D)^{\otimes}$. Denote by $\mathrm{Ab}(D)^{\otimes}$ Freyd's abelian category of $\mathbb{Z}D^{\otimes,+}$.

Proposition 2.5. *$\mathbb{Z}D^{\otimes}$, $\mathbb{Z}D^{\otimes,+}$ and $\mathrm{Ab}(D)^{\otimes}$ with the bilinear extension of $\otimes$ are commutative tensor categories. The canonical functor $\mathbb{Z}D^{+} \to \mathbb{Z}D^{\otimes,+}$ induces a Serre quotient $\pi : \mathrm{Ab}(D) \to \mathrm{Ab}(D)^{\otimes}$.*

Proof. The statements on the additive and preadditive category are obvious. The statement on the abelian category is Proposition 1.8. The claim on the Serre quotient is granted by the following general fact. □

Lemma 2.6. *Let D_1 be a quiver with relations and D the underlying quiver. Then $\pi : \mathrm{Ab}(D) \to \mathrm{Ab}(D_1)$ is a Serre quotient.*

This is well known but for the convenience of the reader we give the simple argument directly.

Proof. Consider $\mathrm{Ab}(D)/\operatorname{Ker}(\pi)$. By construction this is an exact subcategory of $\mathrm{Ab}(D_1)$, hence it remains to check that the inclusion is full and essentially surjective.

The quiver D has a canonical representation in $\mathrm{Ab}(D)/\operatorname{Ker}\pi$. All relations in D_1 are satisfied, hence it is even a representation of D_1. By the universal property this yields an exact functor $\mathrm{Ab}(D_1) \to \mathrm{Ab}(D)/\operatorname{Ker}\pi$. By the uniqueness part of the universal property, its composition with the inclusion into $\mathrm{Ab}(D_1)$ is isomorphic to the identity. In particular, the inclusion is full and essentially surjective, and hence an equivalence of categories. □

We now turn to the universal property. The obvious approach is to consider representations $T : D \to \mathcal{A}$ where all relations in D are mapped to identities in $\mathcal{A}$. However, this is too rigid for most applications. We follow the approach of [Huber and Müller-Stach 2017, Definition 8.1.3].

Definition 2.7. Let D be a $\otimes$-quiver, and $\mathcal{A}$ a commutative tensor category. A *tensor representation* or *$\otimes$-representation* for short, is a representation $T : D \to \mathcal{A}$ of the underlying quiver together with the choice of an isomorphism $\kappa_0 : \mathbf{1} \to T(\mathbf{1})$ and of natural isomorphisms

$$\kappa : T(u) \otimes T(v) \xrightarrow{\simeq} T(u \otimes v)$$

for all vertices $u, v \in D$, functorial in each variable and compatible with the associativity and commutativity constraints and the unit in the obvious way.

Proposition 2.8. *Let $(D, \otimes)$ be a $\otimes$-quiver.*

(1) *$D \to \mathcal{P}(D)^{\otimes}$ is the universal $\otimes$-representation into a commutative tensor category.*

(2) *$D \to \mathbb{Z}D^{\otimes,+}$ is the universal $\otimes$-representation into an additive commutative tensor category.*

Proof. The universal properties for $\mathcal{P}(D)^{\otimes}$ and $\mathbb{Z}D^{\otimes,+}$ are obvious. $\square$

Theorem 2.9. *Let $(D, \otimes)$ be a $\otimes$-quiver.*

(1) *The natural assignment $\Delta^{\otimes} : D \to \mathrm{Ab}(D)^{\otimes}$ is a $\otimes$-representation into an abelian tensor category with right-exact tensor product.*

(2) *It takes values in the subcategory $(\mathrm{Ab}(D)^{\otimes})^{\flat}$ of Definition 1.1. Moreover, this is a $\flat$-subcategory (see Definition 1.9).*

(3) *The category $\mathrm{Ab}(D)^{\otimes}$ is universal with this property.*

In detail: Let $T : D \to \mathcal{A}$ be a $\otimes$-representation via κ in an abelian tensor category with a right-exact tensor, which factors through a $\flat$-subcategory $\mathcal{A}^{\flat} \subseteq \mathcal{A}$. Then there is an induced exact tensor functor $\widetilde{M}^{\otimes} : \mathrm{Ab}(D)^{\otimes} \to \mathcal{A}$.

Proof. $\mathcal{A} = \mathrm{Ab}(D)^{\otimes}$ is an abelian tensor category by Proposition 2.5 and $\Delta^{\otimes}$ is a $\otimes$-representation by construction. It factors via the additive category $\mathbb{Z}D^{\otimes,+}$. Hence property (2) follows from Proposition 1.8. To see the second statement, note that if

$$(B, -) \xrightarrow{(f,-)} (A, -)$$

is a morphism in $(\mathrm{Ab}(D)^{\otimes})^{\flat}$ then its kernel is

$$(C, -) \xrightarrow{(g,-)} (B, -)$$

where $B \xrightarrow{g} C$ is the cokernel of $A \xrightarrow{f} B$.

The induced functor

$$M^{\otimes} : \mathbb{Z}D^{\otimes,+} \to \mathcal{A}$$

satisfies the assumptions in Proposition 1.10. Thus it induces the tensor functor $\widetilde{M}^{\otimes} : \mathrm{Ab}(D)^{\otimes} \to \mathcal{A}$ such that $T = \widetilde{M}^{\otimes}\Delta^{\otimes}$. □

Recall the universal representation theorem stated in [Barbieri-Viale and Prest 2018]. For $T : D \to \mathcal{A}$ any representation of a quiver in an abelian category $\mathcal{A}$ there is an induced additive functor

$$M : \mathbb{Z}D^{+} \to \mathcal{A}$$

and a corresponding $\widetilde{M} : \mathrm{Ab}(D) \to \mathcal{A}$ such that

$$\widetilde{T} : D \to \mathcal{A}(T) := \mathcal{A}(M) = \mathrm{Ab}(D)/\operatorname{Ker}\widetilde{M}$$

is the induced universal representation (see [Barbieri-Viale and Prest 2018, §1.3]). For a $\otimes$-quiver D, together with a $\otimes$-representation T in an abelian tensor category $\mathcal{A}$, as in Theorem 2.9, we have now constructed a factorisation via an exact tensor functor $\widetilde{M}^{\otimes}$ on $\mathrm{Ab}(D)^{\otimes}$. Hence we get a tensorial refinement of the universal representation theorem. This also implies the existence of a tensor structure on the universal abelian category $\mathcal{A}(T)$ attached to the representation. Note that this is really $\mathcal{A}(T)$; in contrast to $\mathcal{P}(D)^{\otimes}$ etc., no $\otimes$-adornment is needed.

Theorem 2.10. *Let $T : D \to \mathcal{A}$ be a representation in an abelian tensor category with a right-exact tensor, which factors through a $\flat$-subcategory $\mathcal{A}^{\flat} \subseteq \mathcal{A}$, with the following additional properties:*

(i) *$(D, \otimes)$ is a $\otimes$-quiver.*

(ii) *T is a $\otimes$-representation in $\mathcal{A}^{\flat} \subseteq \mathcal{A}$ via κ.*

Then Nori's universal abelian category $\mathcal{A}(T)$ carries a right-exact tensor product and $\widetilde{M} : \mathcal{A}(T) \to \mathcal{A}$ is a tensor functor (here M is the additive functor induced by T and $\widetilde{M}$ is the faithful exact functor induced by M; see also Proposition 1.13). It is universal among such representations into abelian tensor categories $\mathcal{B}$ compatible with T (cf. the statement of Proposition 1.14) via a faithful exact tensor functor $\mathcal{B} \to \mathcal{A}$.

Proof. By the universal property in Theorem 2.9, there is a canonical exact tensor functor $\widetilde{M}^{\otimes} : \mathrm{Ab}(D)^{\otimes} \to \mathcal{A}$. Hence $\operatorname{Ker}\widetilde{M}^{\otimes}$ is a Serre subcategory and a tensor ideal. Denoting by $\mathcal{A}(T)^{\otimes}$ the Serre quotient $\mathrm{Ab}(D)^{\otimes}/\operatorname{Ker}\widetilde{M}^{\otimes}$ we have obtained a tensor category with the universal property as claimed. Furthermore, by the universal property of $\mathcal{A}(T)$, there is also an exact faithful functor

$$\mathcal{A}(T) \to \mathcal{A}(T)^{\otimes}.$$

We claim that it is an equivalence of abelian categories. The canonical additive functor $\mathbb{Z}D^+ \to \mathbb{Z}D^{\otimes,+}$ induces an exact functor $\pi : \mathrm{Ab}(D) \to \mathrm{Ab}(D)^\otimes$ such that $\widetilde{M}^\otimes \circ \pi = \widetilde{M}$ by the uniqueness in the universal property of Freyd's construction (see [Barbieri-Viale and Prest 2018, Theorem 1.1]). The faithful exact functor $\bar{\pi} : \mathrm{Ab}(D)/\operatorname{Ker}\pi \xrightarrow{\simeq} \mathrm{Ab}(D)^\otimes$ is an equivalence by Proposition 2.5.

Thus, the composition $\mathrm{Ab}(D) \xrightarrow{\pi} \mathrm{Ab}(D)^\otimes \to \mathrm{Ab}(D)^\otimes/\operatorname{Ker}(\widetilde{M}^\otimes)$ is essentially surjective and is equivalent to the composition $\mathrm{Ab}(D) \to \mathcal{A}(T) \to \mathcal{A}(T)^\otimes$ since they have equivalent compositions with the faithful functor $\mathcal{A}(T)^\otimes \to \mathcal{A}$. So $\mathcal{A}(T) \to \mathcal{A}(T)^\otimes$ also is essentially surjective and hence an equivalence. □

Remark 2.11. The universal property can be upgraded analogously to Corollary 1.15.

Remark 2.12. In the special case where $\mathcal{A}$ is the category of finitely generated modules over a Dedekind domain, this gives back Nori's original result as formulated for example in [Huber and Müller-Stach 2017]. The same case (actually in the more restrictive setting of monoidal quivers) is also handled by Bruguières [2004, Theorem 3]. His conditions P1 and P2 are analogous to our factorisation via $\mathcal{A}^\flat$. Both [Huber and Müller-Stach 2017] and [Bruguières 2004] are based on the explicit description of the universal abelian category as comodules or modules.

Signs. In many cases, notably in Nori's original application, we do not start with a tensor representation but with a tensor representation with signs. We explain the necessary modifications, following again the approach of [Huber and Müller-Stach 2017, Definition 8.1.3].

Definition 2.13. A *graded quiver* is a quiver together with a function $|\cdot|$ assigning to each vertex a degree in $\mathbb{Z}/2\mathbb{Z}$. For an edge $e : v \to w$ we put $|e| = |w| - |v|$. A *graded $\otimes$-quiver* is a graded quiver together with the data of a $\otimes$-quiver such that $|v \otimes w| = |v| + |w|$ and $|\mathbf{1}| = 0$. The relations are the same as for a $\otimes$-quiver, except for relation (3), which is replaced by

(3′) $(e \otimes \mathrm{id}) \circ (\mathrm{id} \otimes e') = (-1)^{|e|\,|e'|}(\mathrm{id} \otimes e') \circ (e \otimes \mathrm{id})$ for all pairs of edges e, e'.

The grading on D induces gradings on $\mathcal{P}(D)$, $\mathbb{Z}D$, and $\mathbb{Z}D^+$. In the case of the additive hull this means that every object is equipped with a decomposition into an even and an odd part. Note that morphisms are *not* required to preserve the degree. Recall that part of the data of a $\otimes$-quiver is the choice of edges $\alpha_{v,w} : v \otimes w \to w \otimes v$.

Definition 2.14. Let $(D, \otimes)$ be a graded $\otimes$-quiver.

(1) We define $\mathbb{Z}D^{\otimes,\mathrm{sgn}}$ as the quotient of the category $\mathbb{Z}D$ modulo the relations of a graded $\otimes$-quiver. It is equipped with the tensor product $\otimes^{\mathrm{sgn}}$ which agrees with $\otimes$ on objects and for morphisms $\gamma : v \to v'$, $\delta : w \to w'$,

$$\gamma \otimes^{\mathrm{sgn}} \delta = (-1)^{|\gamma|\,|w|}\gamma \otimes \delta,$$

with associativity constraint $\beta^{\mathrm{sgn}}_{u,vw} = \beta_{u,vw}$ and commutativity constraint given by

$$\alpha^{\mathrm{sgn}}_{v,w} = (-1)^{|v|\,|w|}\alpha_{v,w} : v \otimes w \to w \otimes v$$

for all objects v, w.

(2) Let $\mathbb{Z}D^{\otimes,\mathrm{sgn},+}$ be the category $\mathbb{Z}D^{\otimes,+}$ with tensor structure given by the additive extension from $\mathbb{Z}D^{\otimes,\mathrm{sgn}}$.

(3) Set $\mathrm{Ab}(D)^{\otimes,\mathrm{sgn}} = \mathrm{Ab}(\mathbb{Z}D^{\otimes,\mathrm{sgn},+})$ for the universal abelian category attached to $\mathbb{Z}D^{\otimes,\mathrm{sgn},+}$.

Remark 2.15. Note that $\mathbb{Z}D^{\otimes}$ is different from $\mathbb{Z}D^{\otimes,\mathrm{sgn}}$ even as an additive category.

Lemma 2.16. *$\mathbb{Z}D^{\otimes,\mathrm{sgn}}$ and $\mathbb{Z}D^{\otimes,\mathrm{sgn},+}$ are well-defined tensor categories.*

Proof. It suffices to consider $\mathbb{Z}D^{\otimes,\mathrm{sgn}}$. We have to check that $\otimes^{\mathrm{sgn}}$ satisfies the axioms of a commutative tensor category. Condition (3′) ensures functoriality of $\otimes^{\mathrm{sgn}}$. It is tedious but straightforward to see that β and α are functorial. For example, for $\gamma : x \to x'$, $\delta : y \to y'$ the diagram reads

$$\begin{tikzcd}
x \otimes y \arrow[r, "(-1)^{|x|\,|y|}\alpha"] \arrow[d, "(-1)^{|\gamma||y|}\gamma\otimes\delta"'] & y \otimes x \arrow[d, "(-1)^{|\delta||x|}\delta\otimes\gamma"] \\
x' \otimes y' \arrow[r, "(-1)^{|x'|\,|y'|}\alpha"] & y' \otimes x'
\end{tikzcd}$$

It does not commute on the level of $\mathcal{P}(D)$. In order to check that it commutes in $\mathbb{Z}D^{\otimes,\mathrm{sgn}}$, it is enough to treat the two special cases $\gamma = \mathrm{id}$ or $\delta = \mathrm{id}$ separately because $(\gamma, \delta) = (\gamma, \mathrm{id}) \circ (\mathrm{id}, \delta)$. In each of these cases the diagram commutes in $\mathcal{P}(D)$.

The pentagon axiom (concerning associativity) holds because it is a relation on D and no signs are involved. Unitality is preserved because $\mathbf{1}$ is of degree 0. The hexagon axiom reads

$$\begin{tikzcd}
x \otimes (y \otimes z) \arrow[r, "\beta"] \arrow[d, "\mathrm{id}\otimes(-1)^{|y|\,|z|}\alpha"'] & (x \otimes y) \otimes z \arrow[r, "(-1)^{(|x|+|y|)|z|}\alpha"] & z \otimes (x \otimes y) \arrow[d, "\beta"] \\
x \otimes (z \otimes y) \arrow[r, "\beta"] & (x \otimes z) \otimes y \arrow[r, "(-1)^{|x|\,|z|}\alpha\otimes\mathrm{id}"] & (z \otimes x) \otimes y
\end{tikzcd}$$

It commutes because the hexagon axiom holds for $\otimes$. □

Again, we turn to representations. Following [Huber and Müller-Stach 2017, Definition 8.1.3]:

Definition 2.17. Let $(D, \otimes)$ be a graded $\otimes$-quiver. Let $\mathcal{A}$ be an additive commutative tensor category. A *graded tensor representation* of $(D, \otimes)$ is a representation

$T : D \to \mathcal{A}$ of the underlying quiver together with the choice of an isomorphism $\kappa_0 : \mathbf{1} \to T(\mathbf{1})$ and of natural isomorphisms

$$\kappa : T(u) \otimes T(v) \xrightarrow{\simeq} T(u \otimes v)$$

for all vertices $u, v \in D$, functorial in each variable and compatible with the associativity constraint and the unit in the obvious way and such that:

(1) For all vertices v, w,

$$\begin{tikzcd} T(v \otimes w) \arrow[r, "T(\alpha)"] & T(w \otimes v) \\ T(v) \otimes T(w) \arrow[u, "\kappa"] \arrow[r] & T(w) \otimes T(v) \arrow[u, "\kappa"'] \end{tikzcd}$$

commutes where the bottom arrow is $(-1)^{|v||w|}$ times the commutativity constraint in $\mathcal{A}$.

(2) For all edges $\gamma : v \to v'$ and vertices w,

$$\begin{tikzcd} T(v \otimes w) \arrow[r, "T(\gamma \otimes \mathrm{id})"] & T(v' \otimes w) \\ T(v) \otimes T(w) \arrow[u, "\kappa"] \arrow[r, "T(\gamma) \otimes \mathrm{id}"] & T(v') \otimes T(w) \arrow[u, "\kappa"'] \end{tikzcd}$$

commutes up to the factor $(-1)^{|\gamma||w|}$.

(3) For all edges $\gamma : v \to v'$ and vertices w,

$$\begin{tikzcd} T(w \otimes v) \arrow[r, "T(\mathrm{id} \otimes \gamma)"] & T(w \otimes v') \\ T(w) \otimes T(v) \arrow[u, "\kappa"] \arrow[r, "\mathrm{id} \otimes T(\gamma)"] & T(w) \otimes T(v') \arrow[u, "\kappa"'] \end{tikzcd}$$

commutes (without signs).

The following is a graded analogue of Proposition 2.8(2).

Proposition 2.18. *Let $(D, \otimes)$ be a graded $\otimes$-quiver. The natural map $D \to \mathbb{Z}D^{\otimes,\mathrm{sgn},+}$ is the universal graded $\otimes$-representation of $(D, \otimes)$. In detail: it is a graded $\otimes$-representation and if $T : D \to \mathcal{A}$ is a graded tensor representation in an additive commutative tensor category $\mathcal{A}$ then T factors uniquely through an induced additive tensor functor as shown*

$$\begin{tikzcd} D \arrow[r] \arrow[dr, "T"'] & \mathbb{Z}D^{\otimes,\mathrm{sgn},+} \arrow[d, "M^{\otimes,\mathrm{sgn}}"] \\ & \mathcal{A} \end{tikzcd}$$

Proof. The argument is the same as in the ungraded case. Relation (3′) is forced by the signs in the graded tensor representation. □

Now consider the category $\mathrm{Ab}(D)^{\otimes,\mathrm{sgn}}$ as in Definition 2.14(3).

Theorem 2.19. *The graded analogue of Theorem 2.9 is satisfied by the category* $\mathrm{Ab}(D)^{\otimes,\mathrm{sgn}}$.

Proof. The proof is as in the ungraded case. □

Finally:

Theorem 2.20. *Let $T : D \to \mathcal{A}$ be a representation in an abelian tensor category with a right-exact tensor, which factors through a $\flat$-subcategory $\mathcal{A}^\flat \subset \mathcal{A}$ with the following additional properties:*

(1) *$(D, \otimes)$ is a graded $\otimes$-quiver.*

(2) *T is a graded $\otimes$-representation in $\mathcal{A}^\flat \subset \mathcal{A}$ via κ.*

Then Nori's universal abelian category $\mathcal{A}(T)$ carries a right-exact tensor product and $\widetilde{M} : \mathcal{A}(T) \to \mathcal{A}$ is a tensor functor (here M is the additive functor induced by T and $\widetilde{M}$ is the faithful exact functor induced by M; see also Proposition 1.13). It is universal among such representations into abelian tensor categories $\mathcal{B}$ compatible with T (cf. Proposition 1.14) via a faithful exact tensor functor $\mathcal{B} \to \mathcal{A}$.

Proof. Compare with the proof of Theorem 2.10. If T is such a graded tensor representation, we get $M^{\otimes,\mathrm{sgn}} : \mathbb{Z}D^{\otimes,\mathrm{sgn},+} \to \mathcal{A}$ and also an induced exact tensor functor $\widetilde{M}^{\otimes,\mathrm{sgn}} : \mathrm{Ab}(D)^{\otimes,\mathrm{sgn}} \to \mathcal{A}$. Denote by $\mathcal{A}(T)^{\otimes,\mathrm{sgn}}$ the quotient of $\mathrm{Ab}(D)^{\otimes,\mathrm{sgn}}$ by the kernel of $\widetilde{M}^{\otimes,\mathrm{sgn}}$. We have that $\mathcal{A}(T) \to \mathcal{A}(T)^{\otimes,\mathrm{sgn}}$ is an equivalence. □

Remark 2.21. Again, the universal property can be upgraded analogously to Corollary 1.15.

3. Homological functors

We return to the case of additive categories, but specialise further by considering triangulated categories and homological functors.

Proposition 3.1 [Neeman 2001, Theorem 5.1.18]. *Let $\mathcal{T}$ be a triangulated category. Then there is an abelian category $\mathrm{Ab}^\Delta(\mathcal{T})$ and a homological functor $[-] : \mathcal{T} \to \mathrm{Ab}^\Delta(\mathcal{T})$ such that every homological functor $\mathcal{T} \to \mathcal{A}$ into an abelian category factors uniquely via an exact functor $\mathrm{Ab}^\Delta(\mathcal{T}) \to \mathcal{A}$.*

Neeman uses the notation $\mathcal{A}(\mathcal{T})$, which we have reserved for Nori's abelian category. By construction, $\mathrm{Ab}^\Delta(\mathcal{T})$ is the subcategory of finitely presented objects in the category of presheaves of abelian groups on $\mathcal{T}$. It is obtained from the image of the Yoneda functor by adding all cokernels. Hence every object of $\mathrm{Ab}^\Delta(\mathcal{T})$ is the cokernel of a morphism of objects in the image of the Yoneda functor and this is a projective resolution. Note, however, see [Neeman 2001, §5.2, Appendix C], that the category $\mathrm{Ab}^{\Delta}(\mathcal{T})$ is typically not well-powered. Our constructions and results in Section 1 do not require the initial categories C to be well-powered, so those results apply here.

Proposition 3.2. *Let $\mathcal{T}$ be a tensor triangulated category. Then $\mathrm{Ab}^{\Delta}(\mathcal{T})$ carries a right-exact tensor product. If $\mathcal{A}$ is an abelian tensor category with right-exact tensor and $\mathcal{T} \to \mathcal{A}$ is a homological tensor functor, then the natural functor $\mathrm{Ab}^{\Delta}(\mathcal{T}) \to \mathcal{A}$ is a tensor functor.*

Proof. We extend from representable objects to cokernels by using projective resolutions $[B] \to [A] \to X \to 0$, where $[A]$ denotes the image of A in $\mathrm{Ab}^{\Delta}(\mathcal{T})$. The arguments are dual to the ones used in the first part of the proof of Proposition 1.4. The associativity constraint etc. are constructed in the same way as in the proof of Proposition 1.8. The compatibility of $\mathrm{Ab}^{\Delta}(\mathcal{T}) \to \mathcal{A}$ with the tensor structure holds for representable arguments and extends to cokernels by right-exactness. □

Remark 3.3. (1) This was already proved by Balmer, Krause and Stevenson in [Balmer et al. 2020, Proposition A.14] for compactly generated tensor triangulated categories for the smaller category of all presheaves on $\mathcal{T}^c$ which is universal for all homological functors commuting with colimits, see [Krause 2000, Section 2].

(2) Applying the universal property of $\mathrm{Ab}(\mathcal{T})$ to $\mathcal{T} \to \mathrm{Ab}^{\Delta}(\mathcal{T})$, we obtain an exact functor $\mathrm{Ab}(\mathcal{T}) \to \mathrm{Ab}^{\Delta}(\mathcal{T})$ but it is not clear whether this is a tensor functor. This may well be false since the kernel of a map between representable functors in $\mathrm{Ab}^{\Delta}(\mathcal{T})$ might not be tensor-flat.

Proposition 3.4. *Let $\mathcal{T}$ be a tensor triangulated category, $\mathcal{A}$ an abelian tensor category with a right-exact tensor product, and $M : \mathcal{T} \to \mathcal{A}$ a homological functor and tensor functor.*

(1) *Then $\mathcal{A}(M)$ carries a canonical right-exact tensor structure such that the faithful exact functor $\widetilde{M} : \mathcal{A}(M) \to \mathcal{A}$ is a tensor functor.*

(2) *If in addition, the tensor structures on $\mathcal{T}$ and $\mathcal{A}$ are commutative and the tensor functor is symmetric, then the tensor product on $\mathcal{A}(M)$ is symmetric.*

(3) *If in addition, the tensor structure on $\mathcal{T}$ is rigid and the tensor product and the* Hom*-functor on $\mathcal{A}$ are exact in both arguments, the same is true for $\mathcal{A}(M)$.*

Proof. The proof is the same as for Proposition 1.13, but with $\mathrm{Ab}(\mathcal{T})$ and the tensor product of Proposition 1.8 replaced with $\mathrm{Ab}^{\Delta}(\mathcal{T})$ and the tensor product of Proposition 3.2. □

Remark 3.5. If both Propositions 1.13 and 3.4 apply, then by the universal property of Proposition 3.4 the tensor structures agree because they agree on objects in the image of $\mathcal{T}$.

Künneth components. We now consider the following situation modelled for the application to Nori motives. Let $\mathcal{T}$ be a triangulated category, $\mathcal{A}$ an abelian category and $R:\mathcal{T}\to D^b(\mathcal{A})$ an exact functor. We abbreviate $H^i_R := H^i\circ R$ and $H^*_R := \bigoplus H^i_R$. The latter is understood with values in $\operatorname{gr}\mathcal{A}$. Let $\mathcal{A}(H^*_R)$ be the universal abelian category defined by H^*_R, and $\mathcal{A}(H^0_R)$ that defined by H^0_R. The commutative diagram

$$\begin{tikzcd} \mathcal{T} \arrow[r, "H^*_R"] \arrow[dr, "H^0_R"'] & \operatorname{gr}\mathcal{A} \arrow[d, "(-)^0"] \\ & \mathcal{A} \end{tikzcd}$$

induces a functor $\mathcal{A}(H^*_R)\to\mathcal{A}(H^0_R)$. We also have $\widetilde{H}^*_R:\mathcal{A}(H^*_R)\to\operatorname{gr}\mathcal{A}$.

Definition 3.6. In the above situation let $\mathcal{A}_0(H^*_R)\subset\mathcal{A}(H^*_R)$ be the full subcategory of objects $X\in\mathcal{A}(H^*_R)$ with $\widetilde{H}^*_R(X)\in\operatorname{gr}\mathcal{A}$ concentrated in degree 0.

The subcategory is abelian and closed under subquotients and extensions.

Remark 3.7. We are interested in the case where $\mathcal{T}$ is a triangulated tensor category, $\mathcal{A}$ is an abelian tensor category with an exact tensor product and R is a tensor functor. Then H^*_R is a tensor functor, but H^0_R is not. Hence while $\mathcal{A}(H^*_R)$ is a tensor category by the results of Section 1, this does not follow for $\mathcal{A}(H^0_R)$. It is, however, true for $\mathcal{A}_0(H^*_R)$. In good cases, it will be equivalent to $\mathcal{A}(H^0_R)$, giving the latter the tensor structure that we want.

Proposition 3.8. *Let $\mathcal{T}$, $\mathcal{A}$ and R be as above. Assume in addition that R can be lifted to an exact functor*

$$R:\mathcal{T}\to D^b(\mathcal{A}_0(H^*_R)).$$

Then the natural functor

$$\mathcal{A}_0(H^*_R)\to\mathcal{A}(H^0_R)$$

is an equivalence of categories.

Proof. We abbreviate $\mathcal{A}' := \mathcal{A}_0(H^*_R)$. By assumption, there is a commutative diagram

$$\begin{tikzcd} \mathcal{T} \arrow[r] \arrow[dr] & D^b(\mathcal{A}') \arrow[r] \arrow[d, "H^0"] & D^b(\mathcal{A}) \arrow[d, "H^0"] \\ & \mathcal{A}' \arrow[r] & \mathcal{A} \end{tikzcd}$$

The functor $\widetilde{H^0}_R:\mathcal{A}(H^0_R)\to\mathcal{A}$ is faithful and exact by construction. The same is true for $\widetilde{H_R^*}: \mathcal{A}' \to \operatorname{gr}\mathcal{A}$. By definition, this functor takes values in degree 0; hence $\mathcal{A}' \to \mathcal{A}$ is also faithful and exact. This implies that the universal categories defined by $H^0: \mathcal{T} \to \mathcal{A}'$ and $H_R^0: \mathcal{T} \to \mathcal{A}$ agree. This gives $\mathcal{A}(H_R^0) \to \mathcal{A}'$ inverse to the inclusion. □

Corollary 3.9. *Let $\mathcal{T}$ be a tensor triangulated category. Let $\mathcal{A}$ be an abelian tensor category with an exact tensor product. Let $R: \mathcal{T} \to D^b(\mathcal{A})$ be a tensor triangulated functor. Assume in addition, that R factors via $D^b(\mathcal{A}_0(H_R^*))$. Then $\mathcal{A}(H_R^0)$ carries a natural tensor structure such that $\mathcal{A}(H_R^0) \to \mathcal{A}$ is a tensor functor. If the tensor product on $\mathcal{T}$ is rigid and* $\operatorname{Hom}_{\mathcal{A}}$ *exact in both variables, then the tensor product on $\mathcal{A}(H_R^0)$ is rigid as well.*

Proof. Combining Proposition 3.8 with the strategy of Remark 3.7 gives the tensor structure. If the tensor product on $\mathcal{T}$ is rigid and $\operatorname{Hom}_{\mathcal{A}}$ exact, then by Proposition 1.13 the tensor product on $\mathcal{A}(H_R^*)$ is rigid as well. Hence every object X of $\mathcal{A}_0(H_R^*)$ has a dual $X^\vee$ in $\mathcal{A}(H_R^*)$. The object $X^\vee$ is actually in $\mathcal{A}_0(H_R^*)$, as we can test by applying the forgetful functor to $\operatorname{gr}\mathcal{A}$. □

Remark 3.10. (1) The use of the *bounded* derived category in the above argument is not very important. We can drop the assumption, if arbitrary direct sums exist in $\mathcal{A}$. This is needed in order to write down the Künneth formula or, equivalently, the tensor structure on $D(\mathcal{A})$.

(2) We may also replace $D^b(\mathcal{A})$ by a tensor triangulated category equipped with a t-structure (compatible with the tensor structure) with heart $\mathcal{A}$ without any change in the arguments.

Integral coefficients. What we have done so far does not apply to $\mathcal{A} = \mathbb{Z}$-mod because its tensor product is not exact. However, there is a version of the above criterion for integral coefficients.

Let $\mathcal{T}$ be a triangulated category. Let $\mathcal{A}$ be an abelian tensor category with a right-exact tensor product such that its derivation on $D^b(\mathcal{A})$ exists. Let $\mathcal{A}^\flat \subset \mathcal{A}$ be a $\flat$-subcategory as in Definition 1.9. Let $R: \mathcal{T} \to D^b(\mathcal{A})$ be a tensor functor. Note that $H_R^*: \mathcal{T} \to \operatorname{gr}\mathcal{A}$ is no longer a tensor functor because $H^*: D^b(\mathcal{A}) \to \operatorname{gr}\mathcal{A}$ is not. However, we have the following lemma:

Lemma 3.11. *In this situation, let $\mathcal{T}^\flat \subset \mathcal{T}$ be the full subcategory of objects with H_R^* in* $\operatorname{gr}\mathcal{A}^\flat$. *Then $\mathcal{T}^\flat$ is a tensor category and*

$$H_R^*|_{\mathcal{T}^\flat}: \mathcal{T}^\flat \to \operatorname{gr}\mathcal{A}^\flat$$

is a tensor functor which satisfies the assumptions of the universal property in Proposition 1.10.

Proof. Obviously $\operatorname{gr}\mathcal{A}^\flat \subset \operatorname{gr}\mathcal{A}$ consists of flat objects and is closed under kernels. It remains to check the claim on the tensor functor with $\mathcal{T} = D^b(\mathcal{A})$. This amounts to the naive Künneth formula for these objects. The subcategory $\mathcal{T}^\flat$ is stable under the canonical truncation functor and shift. Hence it suffices to check the formula for objects of $\mathcal{A}^\flat \subset \mathcal{T}^\flat$. They are flat; hence the derived tensor product agrees with the tensor product in $\mathcal{A}^\flat$. As a byproduct of the formula we see that $\mathcal{T}^\flat$ is stable under the derived tensor product. □

We now replace $\mathcal{A}(H_R^*)$ by $\mathcal{A}(H_R^*|_{\mathcal{T}^\flat})$ and set as before $\mathcal{A}_0(H_R^*|_{\mathcal{T}^\flat})$ to be the subcategory of objects concentrated in degree 0.

Corollary 3.12. *Let $\mathcal{T}$ be a tensor triangulated category. Let $\mathcal{A}$ be an abelian tensor category with a right-exact tensor product. Let $\mathcal{A}^\flat \subset \mathcal{A}$ be a $\flat$-subcategory and assume that the derived tensor product exists on $D^b(\mathcal{A})$. Let $R : \mathcal{T} \to D^b(\mathcal{A})$ be a tensor triangulated functor. Let $\mathcal{T}^\flat$ and $\mathcal{A}_0(H_R^*|_{\mathcal{T}^\flat})$ be as above.*

Assume in addition, that R factors via $D^b(\mathcal{A}_0(H_R^*|_{\mathcal{T}^\flat}))$. Then $\mathcal{A}(H_R^0)$ carries a natural tensor structure such that $\mathcal{A}(H_R^0) \to \mathcal{A}$ is a tensor functor.

4. Nori motives

Recall the original definition of Nori. Let k be a field and $\sigma : k \to \mathbb{C}$ be an embedding. Let Sch_k be the category of schemes which are separated and of finite type over the field k. Let D^{Nori} be Nori's quiver on Sch_k having vertices (X, Y, n) where $Y \subseteq X$ is a closed subscheme and $n \in \mathbb{Z}$ and edges $(X', Y', n) \to (X, Y, n)$ for each morphism $f : X \to X'$ in Sch_k such that $f(Y) \subseteq Y'$, and an additional edge $(Y, Z, n) \to (X, Y, n+1)$ for $Z \subseteq Y \subseteq X$ closed subschemes. Let

$$H_B : D^{\mathrm{Nori}} \to \mathbb{Z}\text{-mod}$$

be the representation given by $(X, Y, n) \rightsquigarrow H_B^n(X(\mathbb{C}), Y(\mathbb{C}); \mathbb{Z})$, the relative singular cohomology group after base change to the complex numbers.

Definition 4.1 (Nori; see also [Huber and Müller-Stach 2017, §9]). The abelian category

$$\mathsf{ECM}_k := \mathcal{A}(H_B)$$

is the category of effective cohomological Nori motives. There is a noneffective version that we shall denote NM_k.

Remark 4.2. The diagram D^{Nori} above agrees with the diagram $\mathrm{Pairs}^{\mathrm{eff}}$ of [Huber and Müller-Stach 2017, Definition 9.1.1]. In loc. cit., the abelian categories are denoted by $\mathsf{MM}^{\mathrm{eff}}_{\mathrm{Nori}}(k)$ and $\mathsf{MM}_{\mathrm{Nori}}(k)$, respectively. Noneffective motives are obtained either by localisation of the diagram or of the category with respect to the Lefschetz motive $\mathbf{1}(-1) = (\mathbb{G}_m, \{1\}, 1)$. This is somewhat premature at this point as it involves the tensor structure. We are going to concentrate on the effective case.

Tensor product via graded ⊗-quivers. Let $D^{\mathrm{Nori},\otimes}$ be the same quiver with, in addition, the following structure of a graded ⊗-quiver in the sense of Definition 2.13. The grading is given by

$$(X, Y, n) \mapsto \bar{n} \in \mathbb{Z}/2\mathbb{Z}.$$

For vertices (X, Y, n), (X', Y, n') we put

$$(X, Y, n) \otimes (X', Y', n') := (X \times_k X', X \times_k Y' \cup Y \times_k X', n + n'),$$

making use of the product in Sch_k. We choose the edges id, α, β, β', u, u' and the vertex $\mathbf{1}$ in the canonical way, e.g., the unit $\mathbf{1} = (\mathrm{Spec}(k), \varnothing, 0)$,

$$u : (X, Y, n) \to (\mathrm{Spec}(k), \varnothing, 0) \otimes (X, Y, n)$$

and $u' : (\mathrm{Spec}(k), \varnothing, 0) \otimes (X, Y, n) \to (X, Y, n)$ are the canonical maps. As relations we use the relations required by Definition 2.13. All this is completely parallel to [Huber and Müller-Stach 2017, §9.3]. By construction we obtain a graded ⊗-quiver.

Recall that the singular cohomology H_B^* is provided with a natural cross or external product

$$\kappa^B_{n,n'} : H^n_B(X, Y) \otimes H^{n'}_B(X', Y') \to H^{n+n'}_B(X \times_k X', X \times_k Y' \cup Y \times_k X').$$

Note that the representation H_B is *not* a ⊗-representation since $\kappa^B_{n,n'}$ fails to be an isomorphism, in general.

Following Nori, we set $D^{\mathrm{good},\otimes}$ for the full sub-⊗-quiver of vertices (X, Y, n) such that $H^*_B(X, Y)$ is concentrated in degree n and free as a $\mathbb{Z}$-module.

Lemma 4.3. *The Betti cohomology*

$$H_B^{\mathrm{good}} := H_B|_{D^{\mathrm{good}}} : D^{\mathrm{good},\otimes} \to \mathbb{Z}\text{-}mod$$

is a graded ⊗-representation with values in the subcategory $(\mathbb{Z}\text{-}mod)^\flat$ of free $\mathbb{Z}$-modules of finite type.

Proof. On good pairs, the map $\kappa^B_{n,n'}$ is indeed an isomorphism by the Künneth formula. The relations of the tensor quiver are all mapped to equalities in $\mathbb{Z}$-mod by the standard properties of singular cohomology. Most are checked explicitly in [Huber and Müller-Stach 2017, Proposition 9.3.1]. The remaining ones (e.g., concerning the inverse u' of u) are obvious. □

Indeed, our definition of a graded ⊗-quiver was modelled on this case.

Corollary 4.4. *The abelian category $\mathcal{A}(H_B^{\mathrm{good}})$ carries a natural ⊗-structure compatible with the forgetful functor to $\mathbb{Z}$-mod.*

Proof. See Theorem 2.20. □

Nori's basic lemma comes into play in comparing the universal categories for the two diagrams.

Theorem 4.5 (Nori; see [Huber and Müller-Stach 2017, Theorem 9.2.22]). *The quiver D^{Nori} can be represented in $\mathcal{A}(H_B^{\mathrm{good}})$ in a compatible way with H_B. In particular,*

$$\mathsf{ECM}_k \cong \mathcal{A}(H_B^{\mathrm{good}})$$

carries a natural tensor structure.

In the above, we are copying Nori's approach, but replace his approach to the universal abelian category and its tensor product with the one developed in this paper. In [Barbieri-Viale and Prest 2020] we go further, providing a general axiomatic framework for tensor motivic categories associated to a cohomological functor on a suitable base category; the tensor structure is induced, using our main theorem, by the cartesian tensor structure on the base category via a cohomological Künneth formula.

We now turn to a different approach which does not mention D^{Nori} and D^{good} (at least not obviously so).

Tensor product via triangulated motives. Let $\mathrm{DM}_{\mathrm{gm}}(k, \mathbb{Q})$ be Voevodsky's category of geometric motives over k with rational coefficients. Let

$$R_B : \mathrm{DM}_{\mathrm{gm}}(k, \mathbb{Q}) \to D^b(\mathbb{Q}\text{-vsp.})$$

be the Betti-realisation. It maps the motive of an algebraic variety to its singular cochain complex.

Remark 4.6. The existence of the Betti-realisation is completely straightforward. The first reference with rational coefficients is [Huber 2000] and the correction thereto, where they appear as a byproduct of a functor into mixed realisations. For integral coefficients it is formulated in [Harrer 2016]. In the original literature on motives, realisation functors were usually contravariant. This is also the viewpoint taken in the above references.

More recently, Voevodsky and then Ayoub who, in [Ayoub 2010] constructs Betti-realisations for motives over any base, use the covariant point of view.

For our application, it does not matter which point of view is taken. We fix on the contravariant one because we want to refer to [Harrer 2016] later on.

Definition 4.7. Let $\mathsf{MM}_k := \mathcal{A}(H_B^0)$ be the universal abelian category defined by the Betti-realisation.

Based on a sketch of Nori, Harrer [2016] showed the following:

Theorem 4.8 [Harrer 2016, Theorem 7.3.1]. *The Betti-realisation factors naturally via the bounded derived category of MM_k and even that of $\mathcal{A}_0(H_B^*)$.*

Remark 4.9. The proof is based on Nori's basic lemma: for every affine variety X and subvariety Y, there is a subvariety

$$X \supset Z \supset Y$$

such that the singular cohomology of the pair (X, Z) is concentrated in the degree equal to the dimension of X, i.e., $(X, Z, \dim X)$ is a good pair. As pointed out by Nori, this can be used in order to construct, for every affine variety X, a natural complex of motives. Using Čech-complexes, this extends to all varieties. Harrer's main effort was to establish functoriality of the construction with respect to finite correspondences. When working with rational coefficients (as we do), functoriality with respect to morphisms is enough; see [Ivorra 2016; Huber and Müller-Stach 2017]. Harrer's result is formulated for NM_k, but actually proved for $\mathcal{A}(H_B^{\mathrm{good}})$. The same proof also works without change for $\mathsf{MM}_k = \mathcal{A}(H_B^0)$ and even the refinement $\mathcal{A}_0(H_B^*)$.

Theorem 4.10. *The category* MM_k *carries a natural tensor structure such that* $\mathsf{MM}_k \to \mathbb{Q}$*-vsp. is a tensor functor and* $\mathrm{DM}_{\mathrm{gm}}(k, \mathbb{Q}) \to D^b(\mathsf{MM}_k)$ *is a triangulated tensor functor.*

In particular, MM_k *is Tannakian.*

Proof. We apply Corollary 3.9 to the rigid tensor category $\mathrm{DM}_{\mathrm{gm}}(k, \mathbb{Q})$ and the Betti-realisation. The assumption is satisfied by Theorem 4.8. This makes MM_k a rigid tensor category; the Betti-realisation $\mathsf{MM}_k \to \mathbb{Q}$-vsp. is a fibre functor. □

Definition 4.11. The *motivic Galois group* of k is defined as the Tannakian dual of the MM_k.

Proposition 4.12. MM_k *is naturally equivalent to Nori's original category*, *i.e.*, $\mathsf{NM}_k \cong \mathsf{MM}_k$. *The motivic Galois group is naturally isomorphic to Nori's original motivic Galois group.*

Proof. For the abelian category, this is already shown in [Huber and Müller-Stach 2017]. The tensor structures are based on the Künneth formula. In each case it is uniquely determined by its value for very good pairs, hence they are the same. The statement about the motivic Galois group follows. □

Remark 4.13. The whole argument also works for motives with coefficients in any field (including finite fields) or Dedekind domain (in particular the integers). Corollary 3.12 can be used instead of the more straightforward Corollary 3.9. The integral case is handled in [Harrer 2016].

Acknowledgements

Thanks to Ralph Kaufmann for bringing our attention to the references [Bunge 1969] and [Day 1970]. We thank Paul Balmer for pointing us to [Balmer et al. 2020] and subsequent discussions of the case of homological functors; see Section 3. We thank the referee for her/his suggestions that considerably improved the exposition.

References

[Arapura 2013] D. Arapura, "An abelian category of motivic sheaves", *Adv. Math.* **233** (2013), 135–195. MR Zbl

[Assem et al. 2006] I. Assem, D. Simson, and A. Skowroński, *Elements of the representation theory of associative algebras, I: Techniques of representation theory*, Lond. Math. Soc. Student Texts **65**, Cambridge Univ. Press, 2006. MR Zbl

[Ayoub 2010] J. Ayoub, "Note sur les opérations de Grothendieck et la réalisation de Betti", *J. Inst. Math. Jussieu* **9**:2 (2010), 225–263. MR Zbl

[Balmer et al. 2020] P. Balmer, H. Krause, and G. Stevenson, "The frame of smashing tensor-ideals", *Math. Proc. Cambridge Philos. Soc.* **168**:2 (2020), 323–343. MR Zbl

[Barbieri-Viale 2017] L. Barbieri-Viale, "$\mathbb{T}$-motives", *J. Pure Appl. Algebra* **221**:7 (2017), 1565–1588. MR Zbl

[Barbieri-Viale and Prest 2018] L. Barbieri-Viale and M. Prest, "Definable categories and $\mathbb{T}$-motives", *Rend. Semin. Mat. Univ. Padova* **139** (2018), 205–224. MR Zbl

[Barbieri-Viale and Prest 2020] L. Barbieri-Viale and M. Prest, "Tensor product of motives via Künneth formula", *J. Pure Appl. Algebra* **224**:6 (2020), art. id. 106267. MR Zbl

[Barbieri-Viale et al. 2018] L. Barbieri-Viale, O. Caramello, and L. Lafforgue, "Syntactic categories for Nori motives", *Selecta Math.* (*N.S.*) **24**:4 (2018), 3619–3648. MR Zbl

[Borceux 1994] F. Borceux, *Handbook of categorical algebra, I: Basic category theory*, Encycl. Math. Appl. **50**, Cambridge Univ. Press, 1994. MR Zbl

[Bruguières 2004] A. Bruguières, "On a Tannakian theorem due to Nori", preprint, 2004, Available at https://tinyurl.com/brugnori.

[Bunge 1969] M. C. Bunge, "Relative functor categories and categories of algebras", *J. Algebra* **11** (1969), 64–101. MR Zbl

[Day 1970] B. Day, "On closed categories of functors", pp. 1–38 in *Reports of the Midwest Category Seminar, IV*, edited by S. Mac Lane, Lecture Notes in Math. **137**, Springer, 1970. MR Zbl

[Deligne and Milne 1982] P. Deligne and J. S. Milne, "Tannakian categories", pp. 101–228 in *Hodge cycles, motives, and Shimura varieties*, Lecture Notes in Math. **900**, Springer, 1982. Zbl

[Freyd 1966] P. Freyd, "Representations in abelian categories", pp. 95–120 in *Proc. Conf. Categorical Algebra* (La Jolla, CA, 1965), edited by S. Eilenberg et al., Springer, 1966. MR Zbl

[Gabriel and Riedtmann 1979] P. Gabriel and C. Riedtmann, "Group representations without groups", *Comment. Math. Helv.* **54**:2 (1979), 240–287. MR Zbl

[Harrer 2016] D. Harrer, *Comparison of the categories of motives defined by Voevodsky and Nori*, Ph.D. thesis, University of Freiburg, 2016, Available at https://tinyurl.com/harrerphd.

[Huber 2000] A. Huber, "Realization of Voevodsky's motives", *J. Algebraic Geom.* **9**:4 (2000), 755–799. Correction in **13**:1 (2004), 195–207. MR Zbl

[Huber and Müller-Stach 2017] A. Huber and S. Müller-Stach, *Periods and Nori motives*, Ergebnisse der Mathematik (3) **65**, Springer, 2017. MR Zbl

[Ivorra 2016] F. Ivorra, "Perverse, Hodge and motivic realizations of étale motives", *Compos. Math.* **152**:6 (2016), 1237–1285. MR Zbl

[Ivorra 2017] F. Ivorra, "Perverse Nori motives", *Math. Res. Lett.* **24**:4 (2017), 1097–1131. MR Zbl

[Krause 2000] H. Krause, "Smashing subcategories and the telescope conjecture: an algebraic approach", *Invent. Math.* **139**:1 (2000), 99–133. MR Zbl

[Levine 1998] M. Levine, *Mixed motives*, Math. Surv. Monogr. **57**, Amer. Math. Soc., Providence, RI, 1998. MR Zbl

[Mac Lane 1963] S. Mac Lane, *Homology*, Grundlehren der Math. Wissenschaften **114**, Academic Press, New York, 1963. MR Zbl

[Neeman 2001] A. Neeman, *Triangulated categories*, Ann. Math. Studies **148**, Princeton Univ. Press, 2001. MR Zbl

[Prest 2011] M. Prest, *Definable additive categories: purity and model theory*, Mem. Amer. Math. Soc. **987**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[Prest 2012] M. Prest, "Categories of imaginaries for definable additive categories", preprint, 2012. arXiv 1202.0427

Received February 4, 2019. Revised November 8, 2019.

LUCA BARBIERI-VIALE
DIPARTIMENTO DI MATEMATICA "F. ENRIQUES"
UNIVERSITÀ DEGLI STUDI DI MILANO
MILANO
ITALY

luca.barbieri-viale@unimi.it

ANNETTE HUBER
MATHEMATISCHES INSTITUT
UNIVERSITÄT FREIBURG
FREIBURG
GERMANY

annette.huber@math.uni-freiburg.de

MIKE PREST
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MANCHESTER
MANCHESTER
UNITED KINGDOM

mprest@manchester.ac.uk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.31

ON THE GARDEN OF EDEN THEOREM FOR ENDOMORPHISMS OF SYMBOLIC ALGEBRAIC VARIETIES

TULLIO CECCHERINI-SILBERSTEIN,
MICHEL COORNAERT AND XUAN KIEN PHUNG

Let G be an amenable group and let X be an irreducible complete algebraic variety over an algebraically closed field K. Let A denote the set of K-points of X and let $\tau : A^G \to A^G$ be an algebraic cellular automaton over (G, X, K), that is, a cellular automaton over the group G and the alphabet A whose local defining map is induced by a morphism of K-algebraic varieties. We introduce a weak notion of preinjectivity for algebraic cellular automata, namely $(*)$-preinjectivity, and prove that τ is surjective if and only if it is $(*)$-preinjective. In particular, τ has the Myhill property, i.e., is surjective whenever it is preinjective. Our result gives a positive answer to a question raised by Gromov (*J. Eur. Math. Soc.* 1:2 (1999), 109–197) and yields an analogue of the classical Moore–Myhill Garden of Eden theorem.

1. Introduction

Gromov [1999] brought out fascinating connections between algebraic geometry and symbolic dynamics. He asked the following:

MSC2010: 14A10, 14A15, 37B10, 37B15, 43A07, 68Q80.

Keywords: algebraic cellular automaton, complete algebraic variety, projective algebraic variety, amenable group, Krull dimension, algebraic mean dimension, preinjectivity, Garden of Eden theorem.

Question [Gromov 1999, 8.J.]. Does the Garden of Eden theorem generalize to the proalgebraic category? First, one asks if preinjectivity implies surjectivity, while the reverse implication needs further modification of definitions.

Our goal here is to present some positive answers to Gromov's question. Before stating our main results, we need to recall a few facts related to the classical Garden of Eden theorem, symbolic dynamics, and algebraic geometry (see Section 2 for more details and references).

Fix a set A, called the *alphabet*, and a group G, called the *universe*. The set $A^G := \{c : G \to A\}$, consisting of all maps from G to A, is called the set of *configurations*. Equip A^G with the *G-shift*, i.e., the action of G defined by the map $G \times A^G \to A^G$, $(g, c) \mapsto gc$, where $(gc)(h) := c(g^{-1}h)$ for all $g, h \in G$ and $c \in A^G$.

Given a configuration $c \in A^G$ and a subset $\Omega \subset G$, we write $c|_\Omega$ for the restriction of c to Ω, i.e., the element $c|_\Omega \in A^\Omega$ defined by $c|_\Omega(g) := c(g)$ for all $g \in \Omega$.

A *cellular automaton* over the group G and the alphabet A is a map $\tau : A^G \to A^G$ satisfying the following property: there exist a finite subset $M \subset G$ and a map $\mu : A^M \to A$ such that

$$(\tau(c))(g) = \mu((g^{-1}c)|_M) \quad \text{for all } c \in A^G \text{ and } g \in G. \tag{1-1}$$

Such a set M is then called a *memory set* of τ and μ is called the associated *local defining map* (see [Ceccherini-Silberstein and Coornaert 2010]). Note that it immediately follows from (1-1) that every cellular automaton $\tau : A^G \to A^G$ is *G-equivariant*, i.e., satisfies $\tau(gc) = g\tau(c)$ for all $c \in A^G$ and $g \in G$, and continuous with respect to the prodiscrete topology, that is, the product topology on A^G obtained by taking the discrete topology on each factor A of A^G.

Two configurations $c_1, c_2 \in A^G$ are said to be *almost equal* if the set $\{g \in G : c_1(g) \neq c_2(g)\}$ is finite. A cellular automaton $\tau : A^G \to A^G$ is called *preinjective* if $\tau(c_1) = \tau(c_2)$ implies $c_1 = c_2$ whenever $c_1, c_2 \in A^G$ are almost equal.

Myhill [1963] proved that if A is a finite set and $G = \mathbb{Z}^d$ ($d \in \mathbb{N}$), then every preinjective cellular automaton $\tau : A^G \to A^G$ is surjective. Together with the converse implication, which had been established by Moore [1962], this yields the celebrated Garden of Eden theorem of Moore and Myhill stating that a cellular automaton with finite alphabet over the group $\mathbb{Z}^d$ is preinjective if and only if it is surjective. The Garden of Eden theorem was subsequently extended to cellular automata with finite alphabet over amenable groups in [Ceccherini-Silberstein et al. 1999]. There is also a linear version of the Garden of Eden theorem. More precisely, it is shown in [Ceccherini-Silberstein and Coornaert 2006] (see also [Ceccherini-Silberstein and Coornaert 2010, Theorem 8.9.6]) that if A is a finite-dimensional vector space over a field K and G is an amenable group, then a K-linear cellular automaton $\tau : A^G \to A^G$ is preinjective if and only if it is surjective.

Consider now an algebraic variety X over a field K, i.e., a scheme of finite type over K, and let $A := X(K)$ denote the set of K-points of X, that is, the set consisting of all K-scheme morphisms $\mathrm{Spec}(K) \to X$. We say that a cellular automaton $\tau : A^G \to A^G$ is an *algebraic cellular automaton* over (G, X, K) if τ admits a memory set M with local defining map $\mu : A^M \to A$ such that μ is induced by some K-scheme morphism $f : X^M \to X$, where X^M denotes the K-fibered product of a family of copies of X indexed by M (see Definition 1.1 in [Ceccherini-Silberstein et al. 2019]).

In the present paper, we shall first establish a version of the Myhill part of the Garden of Eden theorem for certain algebraic cellular automata. This yields a positive answer to the first part of Gromov's question. More specifically, we shall prove the following result (see Theorem 7.1 for a more general statement).

Theorem 1.1. *Let G be an amenable group and let X be an irreducible complete algebraic variety over an algebraically closed field K. Let $A := X(K)$ denote the set of K-points of X. Then every preinjective algebraic cellular automaton $\tau : A^G \to A^G$ over (G, X, K) is surjective.*

As injectivity trivially implies preinjectivity, an immediate consequence of Theorem 1.1 is the following.

Corollary 1.2. *Let X be an irreducible complete algebraic variety over an algebraically closed field K and let G be an amenable group. Let $A := X(K)$ denote the set of K-points of X. Then every injective algebraic cellular automaton $\tau : A^G \to A^G$ over (G, X, K) is surjective.*

It is shown in [Ceccherini-Silberstein et al. 2019, Theorem 1.2] that if X is a complete (possibly not irreducible) algebraic variety over an algebraically closed field K and G is a locally residually finite group, then every injective algebraic cellular automaton over (G, X, K) is surjective. Therefore, Corollary 1.2 remains true if the hypothesis that G is *amenable* is replaced by the hypothesis that G is *locally residually finite*. We shall see in Example 8.5 that if G is a free group on two generators, then, given any algebraically closed field K, there exist an irreducible complete K-algebraic variety X and an algebraic cellular automaton over (G, X, K) that is preinjective but not surjective. As a free group on two generators is residually finite, we deduce that Theorem 1.1 becomes false if "amenable" is replaced by "residually finite" in its hypotheses.

Let us note that, as implicitly stated in Gromov's question, the converse implication, i.e., the analogue of the Moore implication, does not hold under the hypotheses of Theorem 1.1. For example, if K is an algebraically closed field whose characteristic is not equal to 2, the projective line $\mathbb{P}^1_K$ is an irreducible complete K-algebraic variety and the morphism $f : \mathbb{P}^1_K \to \mathbb{P}^1_K$ given by $(x : y) \mapsto (x^2 : y^2)$ is surjective but not injective. Taking $A := \mathbb{P}^1_K(K)$, we deduce that, for any group G, the map $\tau : A^G \to A^G$ defined by $(\tau(c))(g) := f(c(g))$ for all $c \in A^G$ and $g \in G$, is an algebraic cellular automaton over (G, X, K) that is surjective but not preinjective.

In order to formulate a version of the Garden of Eden theorem for algebraic cellular automata, we introduce a weak notion of preinjectivity for them, namely $(*)$-*preinjectivity* (see Definition 6.1 below). We shall prove that Theorem 1.1 remains valid if we replace the hypothesis that τ is preinjective by the weaker hypothesis that τ is $(*)$-preinjective. This weak form of preinjectivity also allows us to establish a version of the Moore part of the Garden of Eden theorem for algebraic cellular automata.

Theorem 1.3. *Let G be an amenable group and let X be an irreducible algebraic variety over an algebraically closed field K. Let $A := X(K)$ denote the set of K-points of X. Then every surjective algebraic cellular automaton $\tau : A^G \to A^G$ over (G, X, K) is $(*)$-preinjective.*

Note that X is not assumed to be complete in Theorem 1.3. Combining these results, we obtain the following version of the Garden of Eden theorem (see Theorem 7.1) for algebraic cellular automata.

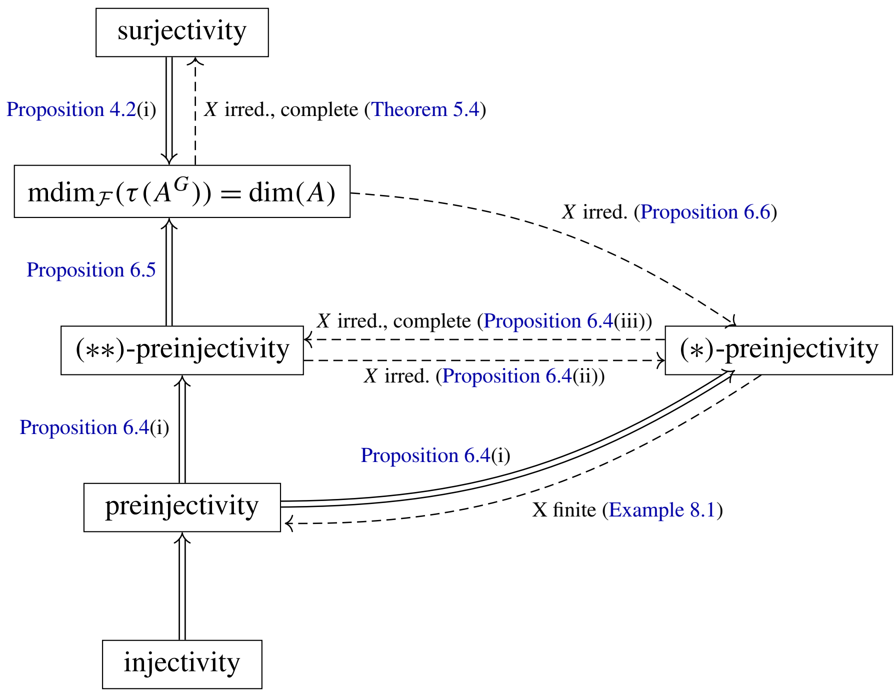
Theorem 1.4. *Let G be an amenable group and let X be an irreducible complete algebraic variety over an algebraically closed field K. Let $A := X(K)$ denote the set of K-points of X and let $\tau : A^G \to A^G$ be an algebraic cellular automaton over (G, X, K). Then the following conditions are equivalent*:

(a) *τ is surjective.*

(b) *τ is $(*)$-preinjective.*

The paper is organized as follows. The next section collects background material on algebraic varieties and amenable groups. Section 3 contains some preliminary results on algebraic cellular automata. In Section 4, we introduce the *algebraic mean dimension* $\mathrm{mdim}_{\mathcal{F}}(\Gamma)$ of a subset $\Gamma \subset A^G$, where G is an amenable group equipped with a Følner net $\mathcal{F}$ and A is the set of K-points of an algebraic variety X over an algebraically closed field K. The definition of algebraic mean dimension is analogous to that of topological entropy. Here $\mathrm{mdim}_{\mathcal{F}}(\Gamma)$ is obtained as a limit of the average Krull dimension of the projection of Γ along the Følner net. It follows in particular that $\mathrm{mdim}_{\mathcal{F}}(\Gamma)$ is always bounded above by the dimension of the variety X and equality holds if $\Gamma = A^G$. In Section 5, we prove that if X is irreducible and complete, then τ is surjective if and only if its image has maximal algebraic mean dimension (Theorem 5.4). In Section 6, we introduce the notions of $(*)$-preinjectivity and $(**)$-preinjectivity, which are both implied by preinjectivity. In the trivial case when A is finite, that is, X is 0-dimensional, every cellular automaton $\tau : A^G \to A^G$ is algebraic over (G, X, K) and $(*)$-preinjectivity is equivalent to preinjectivity (see Example 8.1). We show that if X is irreducible and complete, then τ is $(*)$-preinjective if and only if it is $(**)$-preinjective (see Assertion (iii) in Proposition 6.4). We also establish relations between $(*)$-preinjectivity, $(**)$-preinjectivity, and the fact that the image of τ has maximal algebraic mean dimension. In Section 7, we combine the results of the two previous sections to obtain Theorems 1.3 and 7.1, which extend Theorem 1.1 as well as Theorem 1.4. Another result in this section says that, under suitable conditions, the surjectivity of an algebraic cellular automaton, provided it is defined over an amenable group and an algebraically closed field, is a property that is invariant under base change of the ground field (Theorem 7.2). Several counterexamples are presented in Section 8 showing that the hypotheses in our results are reasonably optimal. Some open questions are formulated in the final section.

Let G be an amenable group with a Følner net $\mathcal{F}$. Let X be an algebraic variety over an algebraically closed field K. Let $A := X(K)$ and suppose that $\tau : A^G \to A^G$ is an algebraic cellular automaton over (G, X, K). We can summarize our results in the following diagram relating the various properties of the algebraic cellular automaton τ:



2. Background material and preliminary results

2A. *Krull dimension and Jacobson spaces.* Let X be a topological space. Given a subset $Y \subset X$, we denote by $\overline{Y}$ the closure of Y in X.

A point $x \in X$ is said to be a *closed* (respectively, *generic*) point of X if $\overline{\{x\}} = \{x\}$ (respectively, $\overline{\{x\}} = X$).

One says that X is *irreducible* if every nonempty open subset of X is dense in X. This amounts to saying that if $X = Y \cup Z$, where Y and Z are closed subsets of Y, then $X = Y$ or $X = Z$.

A subset $Y \subset X$ is called an *irreducible component* of X if Y is irreducible (for the induced topology) and maximal for inclusion among all irreducible subsets of X. As the closure of an irreducible subset of X is irreducible, every irreducible component of X is closed in X. By Zorn's lemma, every irreducible subset of X is contained in some irreducible component of X. Since every singleton of X is irreducible, it follows that X is the union of its irreducible components.

The topological space X is called *Noetherian* if every descending chain of closed subsets of X is stationary. Every subset of a Noetherian topological space is Noetherian for the induced topology. If X is Noetherian, then X is quasicompact and admits only finitely many irreducible components.

The *Krull dimension* of X, denoted by $\dim(X)$, is defined as being the supremum of the lengths of all the strictly ascending chains of closed irreducible subsets of X.

Proposition 2.1. *Let X be a topological space. Then the following hold*:

(i) *If Y is a subset of X, then* $\dim(Y) \leq \dim(X)$.

(ii) *If X is irreducible with* $\dim(X) < \infty$ *and Y is a closed subset of X such that* $\dim(Y) = \dim(X)$, *then one has $Y = X$.*

(iii) *If $(U_\lambda)_{\lambda \in \Lambda}$ is an open cover of X, then one has* $\dim(X) = \sup_{\lambda \in \Lambda} \dim(U_\lambda)$.

(iv) *One has* $\dim(X) = \sup_{Y \in \mathcal{C}(X)} \dim(Y)$, *where $\mathcal{C}(X)$ denotes the set of all irreducible components of X.*

(v) *If X is the union of a finite family $(Z_i)_{i \in I}$ of closed irreducible subsets of X, then every irreducible component of X is equal to one of the Z_i and one has* $\dim(X) = \max_{i \in I} \dim(Z_i)$.

Proof. For (i), (ii), (iii), and (iv), see [Görtz and Wedhorn 2010, Lemma 5.7]. Assertion (v) follows from [Liu 2002, Proposition 2.4.5(c)] and Assertion (iii). □

A subset Y of a topological space X is said to be *very dense* in X if $F \cap Y$ is dense in F for every closed subset F of X.

Proposition 2.2. *Suppose that Y is a very dense subset of a topological space X. Then one has* $\dim(X) = \dim(Y)$.

Proof. By Proposition 2.1, it suffices to prove that $\dim(X) \leq \dim(Y)$. We first observe that if F and F' are closed subsets of X such that $F \cap Y = F' \cap Y$, then $F = F'$ since Y is very dense in X. Note also that $F \cap Y$ is irreducible for every closed irreducible subset F of X. Thus, if $F_0 \subset F_1 \subset \cdots \subset F_n$ is a strictly ascending chain of closed irreducible subsets of X, then $F_0 \cap Y \subset F_1 \cap Y \subset \cdots \subset F_n \cap Y$ is a strictly ascending chain of closed irreducible subsets of Y. It follows that $\dim(X) \leq \dim(Y)$. □

Let X be a topological space. We denote by X_0 the set of closed points of X. One says that the topological space X is *Jacobson* if X_0 is very dense in X. From the result of Proposition 2.2, we immediately deduce the following.

Corollary 2.3. *Let X be a Jacobson space. Then one has* $\dim(X) = \dim(X_0)$.

A subset of a topological space X is said to be *locally closed* if it is the intersection of an open subset and a closed subset of X. A subset of X is called *constructible* if it is a finite union of locally closed subsets of X. The set of constructible subsets of X is closed under finite union, finite intersection, and set difference. Every constructible subset $C \subset X$ contains a dense open subset of $\overline{C}$ (see [An 2012, Lemma 2.1]).

Proposition 2.4. *Let X be a Jacobson topological space and let C be a constructible subset of X. Then the following hold*:

(i) *C is Jacobson.*

(ii) $C_0 = C \cap X_0$.

(iii) $\dim(C) = \dim(C_0) = \dim(C_0 \cap X_0)$.

Proof. See, e.g., [Ceccherini-Silberstein et al. 2019, Lemma A.2] for the proof of (i) and (ii). Assertion (iii) follows from (i), (ii), and Corollary 2.3. □

As immediate consequences of the preceding proposition, we get the following.

Corollary 2.5. *Let X be a Jacobson space. Then the map $C \mapsto C \cap X_0$ yields a bijection from the set of constructible subsets of X onto the set of constructible subsets of X_0. Moreover*, *this map preserves Krull dimension.*

Corollary 2.6. *Let X be a Jacobson space. Then X is irreducible if and only if X_0 is irreducible.*

2B. ***Schemes and algebraic varieties.*** In this subsection, we collect all the material about schemes and algebraic varieties that we shall need in the present paper (see [Görtz and Wedhorn 2010; EGA IV_1 1964; EGA IV_3 1966; Harris 1992; Hartshorne 1977; Liu 2002; Vakil] for more details). All rings are commutative with 1. We recall that a *scheme* is a locally ringed space, that is, a topological space endowed with a sheaf of rings such that the stalk at each point is a local ring. Following a common abuse, if there is no risk of confusion, we shall use the same symbol to denote a scheme and its underlying topological space. The topology on the underlying topological space of a scheme is called the *Zariski topology*.

Every scheme X is *sobre*, i.e., the map $x \mapsto \overline{\{x\}}$ yields a bijection from X onto the set of nonempty closed irreducible subsets of X (see, e.g., Proposition 3.23 in [Görtz and Wedhorn 2010]). In particular, every nonempty closed irreducible subset of a scheme X admits a unique generic point. A scheme is called *irreducible* (respectively, *Jacobson*) if its underlying topological space is irreducible (respectively, Jacobson). The *Krull dimension* $\dim(X)$ of a scheme X is defined as being the dimension of its underlying topological space.

The *spectrum* of a ring R is a scheme whose underlying set consists of all prime ideals of R. The spectrum of a ring R is denoted by $\operatorname{Spec}(R)$ or simply R when there is no risk of confusion. The *Krull dimension* $\dim(R)$ of a ring R is the Krull dimension of its spectrum. It is equal to the supremum of the lengths of all the strictly ascending chains of prime ideals of R.

A scheme X is called *Noetherian* if the space X admits a finite affine open cover $(U_i)_{i \in I}$ such that, for each $i \in I$, one has $U_i = \operatorname{Spec}(R_i)$, where R_i is a Noetherian ring. The underlying topological space of every Noetherian scheme is Noetherian. However, there are schemes that are not Noetherian although their underlying topological spaces are Noetherian.

Let K be a field. An *algebraic variety* over K (or K-*algebraic variety*) is a scheme of finite type over K.

Given an algebraic variety X over a field K, the set of K-*points* of X is the set $X(K)$ consisting of all K-scheme morphisms $\operatorname{Spec}(K) \to X$.

Proposition 2.7. *Let X be an algebraic variety over an algebraically closed field K. Then the map from $X(K)$ into X, that sends each $f \in X(K)$ to the image by f of the unique point of* $\operatorname{Spec}(K)$, *yields a bijection from $X(K)$ onto the set $X_0 \subset X$ consisting of all closed points of X.*

Proof. See [EGA I 1971, Corollaire 6.4.2]. □

Remark 2.8. Proposition 2.7 allows us, in the case when X is an algebraic variety over an algebraically closed field K, to identify $X(K)$ with X_0.

Proposition 2.9. *Let X be an algebraic variety over a field K. Let C and D be constructible subsets of X. Then the following hold*:

(i) *The scheme X is Noetherian.*

(ii) *X is Jacobson.*

(iii) $\dim(X_0) = \dim(X) < \infty$.

(iv) *C is Jacobson.*

(v) $C_0 = C \cap X_0$.

(vi) $\dim(C_0) = \dim(C) = \dim(\overline{C})$.

(vii) *If $C \subset D$ then $C_0 \subset D_0$.*

Proof. Assertions (i) and (ii) follow for instance from Assertions (i) and (iii) in [Ceccherini-Silberstein et al. 2019, Lemma A.17].

Since X is Jacobson, we have that $\dim(X_0) = \dim(X)$ by Corollary 2.3. To prove that $\dim(X) < \infty$, as every scheme is locally affine, we can assume, by virtue of Proposition 2.1(iii), that X is affine. Then $X = \operatorname{Spec}(R)$ for some finitely generated K-algebra R. By the Noether normalization lemma, there exist an integer $d \geq 0$ and an injective K-algebra morphism $K[t_1, \dots, t_d] \to R$ such that R is a finitely generated $K[t_1, \dots, t_d]$-module. This implies $\dim(X) = \dim(R) = d < \infty$ (see [Görtz and Wedhorn 2010, Corollary 5.17]) and completes the proof of (iii).

Assertions (iv) and (v) follow from (i) and Proposition 2.4.

From (i) and Proposition 2.4(iii), we deduce that $\dim(C \cap X_0) = \dim(C)$. Thus, to complete the proof of (vi), it remains only to show that $\dim(C) = \dim(\overline{C})$. To see this, we first observe that C contains an open dense subset U of $\overline{C}$ since C is constructible. Let us equip $\overline{C} \subset X$ with its induced reduced closed subscheme structure. Then U is an open subscheme of $\overline{C}$ and both $\overline{C}$ and U are K-algebraic varieties. Since U is Noetherian, it admits finitely many irreducible components. Let $x_1, \dots, x_n$ denote the generic points of the irreducible components of U and consider their closures $\overline{\{x_1\}}, \dots, \overline{\{x_n\}}$ in X, equipped with their induced reduced closed subscheme structure. As U is dense in $\overline{C}$, we have that $\overline{C} = \bigcup_{1 \leq i \leq n} \overline{\{x_i\}}$. Since each $\overline{\{x_i\}}$ is a closed irreducible subset of $\overline{C}$, we deduce from Proposition 2.1(v) that

$$\dim(\overline{C}) = \max_{1 \leq i \leq n} \dim(\overline{\{x_i\}}). \tag{2-1}$$

Now observe that, for all $1 \leq i \leq n$, the set $U \cap \overline{\{x_i\}}$ is an open subset of $\overline{\{x_i\}}$ that is nonempty since $x_i \in U$. Hence, Theorem 5.22(3) in [Görtz and Wedhorn 2010] applied to the irreducible algebraic varieties $\overline{\{x_i\}}$ implies that

$$\dim(\overline{\{x_i\}}) = \dim(U \cap \overline{\{x_i\}}) \leq \dim(U), \tag{2-2}$$

where the last inequality follows from Proposition 2.1(i). We deduce from (2-1), (2-2), and Proposition 2.1(i) that $\dim(\overline{C}) \leq \dim(U) \leq \dim(C)$. As $\dim(C) \leq \dim(\overline{C})$ by Proposition 2.1(i), we conclude that

$$\dim(U) = \dim(C) = \dim(\overline{C}). \tag{2-3}$$

This completes the proof of (vi).

Assertion (vii) is an immediate consequence of (v). $\square$

Proposition 2.10. *Let X and Y be algebraic varieties over a field K and let $f : X \to Y$ be a K-scheme morphism. Let C be a constructible subset of X. Then the following hold*:

(i) *$f(C)$ is a constructible subset of Y.*

(ii) $f(C_0) = (f(C))_0$.

(iii) $\dim(f(C)) \le \dim(C)$.

(iv) $f(X_0) \subset Y_0$.

(v) $\dim(f(X)) \le \dim(X)$.

(vi) *If E is a constructible subset of X_0, then $f(E)$ is a constructible subset of Y_0 and one has* $\dim(f(E)) \le \dim(E)$.

Proof. Assertion (i) is Chevalley's theorem (see, for example, [EGA IV$_1$ 1964, Théorème 1.8.4], [Hartshorne 1977, p. 93], [Vakil, Theorem 7.4.2]).

For (ii), see, e.g., [Ceccherini-Silberstein et al. 2019, Lemma A.22(v)].

To prove (iii), first observe that $D := f(C)$ is a constructible subset of Y by (i). Therefore D contains a dense open subset V of $\overline{D}$. Let $y_1, \dots, y_m$ denote the generic points of the irreducible components of V (see the proof of Proposition 2.9(vi)). As $V \subset D$, there exist points $x_1, \dots, x_m \in C$ such that $f(x_i) = y_i$ for $1 \le i \le m$. Consider the closure $\overline{\{x_i\}}$ (resp. $\overline{\{y_i\}}$) of the singletons $\{x_i\}$ (resp. $\{y_i\}$) in X (resp. Y), equipped with their induced reduced closed subscheme structures. For $1 \le i \le m$, let $f_i : \overline{\{x_i\}} \to \overline{\{y_i\}}$ be the dominant K-scheme morphism induced by f (see [EGA I 1971, Proposition I.5.2.2]). It follows from Theorem 5.22(3) in [Görtz and Wedhorn 2010] that

$$\dim(\overline{\{y_i\}}) \le \dim(\overline{\{x_i\}}) \quad \text{for all } 1 \le i \le m. \tag{2-4}$$

On the other hand, we have that $\overline{D} = \bigcup_{1 \le i \le m} \overline{\{y_i\}}$ since V is dense in $\overline{D}$, so that

$$\dim(\overline{f(C)}) = \max_{1 \le i \le n} \dim(\overline{\{y_i\}}) \tag{2-5}$$

by applying Proposition 2.1(v). From (2-5) and (2-4), we get

$$\dim(\overline{D}) \le \max_{1 \le i \le n} \dim(\overline{\{x_i\}}) \le \dim(\overline{C}), \tag{2-6}$$

where the last inequality follows from Proposition 2.1(i). Now, as $C \subset X$ and $D \subset Y$ are constructible subsets, we have that $\dim(C) = \dim(\overline{C})$ and $\dim(D) = \dim(\overline{D})$ by Proposition 2.9(vi). Therefore, inequality (2-6) gives us $\dim(D) \le \dim(C)$. This completes the proof of (iii).

Assertions (iv) and (v) are deduced from (ii) and (iii) after taking $C = X$.

Suppose now that E is a constructible subset of X_0. Then $E = C \cap X_0$ for some constructible subset $C \subset X$ by Corollary 2.5. We then have $f(E) = f(C) \cap Y_0$ by virtue of (i), (ii), and Proposition 2.9(v), and hence

$$\begin{aligned} \dim(f(E)) = \dim(f(C) \cap Y_0) &= \dim(f(C)) \\ &\le \dim(C) = \dim(C \cap X_0) = \dim(E) \end{aligned}$$

by using (i), (iii), and Proposition 2.9(vi). This shows (vi). □

Proposition 2.11. *Let X and Y be algebraic varieties over a field K and let $f : X \to Y$ be a K-scheme morphism. For $x \in X$, let $y := f(x)$. Then the following hold*:

(i) *There exists a closed point $x \in X$ such that*

$$\dim(f^{-1}(y)) \geq \dim(X) - \dim(Y). \tag{2-7}$$

(ii) *If X and Y are both irreducible, then inequality (2-7) is satisfied for every closed point $x \in X$.*

Proof. Consider the *geometric fiber* of f at y, that is, the Y-fibered product $X_y := X \times_Y \kappa(y)$, where $\kappa(y)$ is the residue field of Y at y, and recall that the first projection morphism $X_y \to X$ induces a homeomorphism from X_y onto $f^{-1}(y)$ (see [Liu 2002, Proposition 3.1.16]). As X and Y are Noetherian schemes, it follows from Theorem 4.3.12 in [Liu 2002] that

$$\dim(\mathcal{O}_{X_y,x}) \geq \dim(\mathcal{O}_{X,x}) - \dim(\mathcal{O}_{Y,y}) \tag{2-8}$$

for all $x \in X$.

Suppose first that X and Y are irreducible. If x is a closed point of X, then $y = f(x)$ is a closed point of Y (see, e.g., Lemma A.22 in [Ceccherini-Silberstein et al. 2019]). By applying Corollary 2.5.24 in [Liu 2002], we then get

$$\dim(\mathcal{O}_{X,x}) = \dim(X) \quad \text{and} \quad \dim(\mathcal{O}_{Y,y}) = \dim(Y),$$

so that Assertion (ii) follows from (2-8) and the general fact that

$$\dim(\mathcal{O}_{X_y,x}) \leq \dim(X_y) = \dim(f^{-1}(y)).$$

To prove Assertion (i), consider an irreducible component Z of X such that $\dim(Z) = \dim(X)$ and the closure $V = \overline{f(Z)} \subset Y$ of its image. As the closure of every irreducible subset is itself irreducible, V is also irreducible. We equip Z and V with their induced reduced closed subscheme structures and denote by $\iota : Z \to X$ the closed immersion associated with Z. By [EGA I 1971, Proposition I.5.2.2], $f \circ \iota$ induces a K-morphism of irreducible algebraic varieties $h : Z \to V$. Let $x \in Z$ be a closed point and $y = h(x) = f(x)$. Then by what we proved above for the irreducible case (Assertion (ii)), we conclude that

$$\dim(f^{-1}(y)) \geq \dim(h^{-1}(y)) \geq \dim(Z) - \dim(V) \geq \dim(X) - \dim(Y),$$

where the first inequality follows from the inclusion $f^{-1}(y) \supset h^{-1}(y)$ and the last inequality from the inclusion $V \subset Y$. □

Proposition 2.12. *Let X and Y be algebraic varieties over a field K and let $X \times_K Y$ denote their K-fibered product. Then the following hold*:

(i) *$X \times_K Y$ is a K-algebraic variety.*

(ii) $(X \times_K Y)(K) = X(K) \times Y(K)$.

(iii) $\dim(X \times_K Y) = \dim(X) + \dim(Y)$.

Proof. Assertion (i) follows from Lemma 4.22 in [Görtz and Wedhorn 2010].

Assertion (ii) is an immediate consequence of the universal property of K-fibered products.

Assertion (iii) follows from Propositions 5.37 and 5.50 in [Görtz and Wedhorn 2010]. □

Proposition 2.13. *Let X and Y be algebraic varieties over an algebraically closed field K and let $X \times_K Y$ denote their K-fibered product. Let C (resp. D) be a constructible subset of X (resp. Y). Then the following hold*:

(i) $(X \times_K Y)_0 = X_0 \times Y_0$.

(ii) $C_0 \times D_0 \subset (X \times_K Y)_0$.

(iii) *The set $C_0 \times D_0 \subset (X \times_K Y)_0 \subset X \times_K Y$ being equipped with the Zariski topology, one has that* $\dim(C_0 \times D_0) = \dim(C_0) + \dim(D_0)$.

(iv) *If X and Y are irreducible, then $X \times_K Y$ is irreducible.*

Proof. Assertion (i) immediately follows from Remark 2.8 and Assertions (i) and (ii) in Proposition 2.12.

Since $C_0 \subset X_0$ and $D_0 \subset Y_0$ by Proposition 2.9(ii), we have that

$$C_0 \times D_0 \subset X_0 \times Y_0 = (X \times_K Y)_0$$

by using (i). This shows (ii).

Since C (resp. D) is a constructible subset of X (resp. Y), it contains an open dense subset U (resp. V) of $\overline{C}$ (resp. $\overline{D}$). Let us equip $\overline{C}$ and $\overline{D}$ with their induced reduced closed subscheme structure. Thus U, V, $\overline{C}$, and $\overline{D}$ are now viewed as K-algebraic varieties. By properties of base change, $U \times_K V$ is an open subscheme of $\overline{C} \times_K \overline{D}$, which is in turn a closed subscheme of $X \times_K Y$. By applying Proposition 2.9(vii), we have that $U_0 \subset C_0 \subset \overline{C}$ and $V_0 \subset D_0 \subset \overline{D}$, so that

$$(U \times_K V)_0 = U_0 \times V_0 \subset C_0 \times D_0 \subset (\overline{C})_0 \times (\overline{D})_0 \\ = (\overline{C} \times_K \overline{D})_0 \subset \overline{C} \times_K \overline{D}, \tag{2-9}$$

where the two equalities follow from (i). By using Proposition 2.1(i), we deduce from (2-9) that

$$\dim((U \times_K V)_0) \leq \dim(C_0 \times D_0) \\ \leq \dim(\overline{C} \times_K \overline{D}). \tag{2-10}$$

Now since

$$\begin{aligned}\dim((U \times_K V)_0) &= \dim(U \times_K V) && \text{(by Proposition 2.9(iii))}\\ &= \dim(U) + \dim(V) && \text{(by Proposition 2.12(iii))}\\ &= \dim(C) + \dim(D) && \text{(by (2-3))}\\ &= \dim(C_0) + \dim(D_0) && \text{(by Proposition 2.9(vi))}\end{aligned}$$

and

$$\begin{aligned}\dim(\overline{C} \times_K \overline{D}) &= \dim(\overline{C}) + \dim(\overline{D}) && \text{(by Proposition 2.12(iii))}\\ &= \dim(C_0) + \dim(D_0) && \text{(by Proposition 2.9(vi)),}\end{aligned}$$

it follows from (2-10) that $\dim(C_0 \times D_0) = \dim(C_0) + \dim(D_0)$. This shows (iii).

Assertion (iv) follows from Proposition 5.50 in [Görtz and Wedhorn 2010]. □

Remark 2.14. Assertion (iv) in Proposition 2.13 becomes false if we remove the hypothesis that the field K is algebraically closed. For example, $X := \mathrm{Spec}(\mathbb{C})$ is an irreducible algebraic variety over $K := \mathbb{R}$ but $X \times_K X = \mathrm{Spec}(\mathbb{C} \times \mathbb{C})$ is not irreducible.

2C. *Projective varieties.* Let K be a field. A K-algebraic variety X is called a *projective variety* over K if there exists a closed immersion $\iota : X \to \mathbb{P}^N_K$ for some $N \in \mathbb{N}$ (see [Liu 2002, Definition 2.3.47]). In that case, we can identify the underlying topological space of X with its image by ι.

Theorem 2.15 (projective dimension theorem). *Let K be an algebraically closed field. Let Y and Z be closed subschemes of $\mathbb{P}^N_K$ of dimension r and s, respectively. Suppose $r + s \geq N$. Then $Y \cap Z$ is nonempty. Equivalently, $Y(K) \cap Z(K) \subset \mathbb{P}^N(K)$ is nonempty.*

Proof. Up to replacing Y and Z by irreducible components of maximal dimension equipped with their induced reduced closed subscheme structure, we can assume that Y and Z are irreducible. The theorem is then just a reformulation of Theorem I.7.1 in [Hartshorne 1977]. Indeed, $Y \cap Z$ is closed so it is Jacobson. We then equip it with the induced reduced subscheme structure. Hence, $Y \cap Z$ is nonempty if and only if it has a closed point, i.e., $(Y \cap Z)(K) = Y(K) \cap Z(K)$ is nonempty. See also Proposition 5.40 and its corollaries in [Görtz and Wedhorn 2010]. □

Corollary 2.16. *Let $N \in \mathbb{N}$ and let $X \subset \mathbb{P}^N_K$ be a projective variety over an algebraically closed field K. Let $L \subset X$ be a hyperplane section of X, i.e., $L = H \cap X$, where $H \subset \mathbb{P}^N_K$ is a hyperplane not containing X. Let $C \subset X$ be a closed subscheme such that* $\dim(C) \geq 1$. *Then $L \cap C$ is nonempty.*

Proof. By hypothesis, $X \subset \mathbb{P}^N_K$ is a closed subscheme. Hence, C is also a closed subscheme of $\mathbb{P}^N_K$ since C is closed in X. Note $\dim(C) + \dim(H) \geq 1 + N - 1 = n$.

Therefore, we deduce from Theorem 2.15 that $H \cap C \neq \varnothing$. As $C \subset X$, we conclude that $L \cap C = H \cap X \cap C = H \cap C$ is nonempty. □

Remark 2.17. With the notation as in Corollary 2.16, we claim that hyperplane sections of X always exist. Indeed, let $H_0, \dots, H_N$ denote the $N+1$ standard coordinate hyperplanes of $\mathbb{P}^N$. Since $H_0 \cap \dots \cap H_N = \varnothing$ and $X \supset C$ is nonempty, there exists a hyperplane H_i not containing X. This proves the claim.

In fact, let $\iota : X \to \mathbb{P}^N$ be the closed immersion. Let $\mathcal{O}(1)$ denote the Serre line bundle of $\mathbb{P}^N$. Then each global section of the very ample line bundle $\iota^*\mathcal{O}(1)$ of X is a hyperplane section. These global sections, denoted by $H^0(X, \iota^*\mathcal{O}(1))$, form a strictly positive finite-dimensional K-vector space. See Chapters II.5 and III, and Appendix A in [Hartshorne 1977] for more details.

Every projective K-algebraic variety is K-proper, i.e., complete, (see for example [Liu 2002, Theorem 3.3.30]). The converse is not true. However, we have the following consequence of Chow's lemma, which we shall use in Section 6.

Theorem 2.18 (Chow's lemma). *Let X be an irreducible complete algebraic variety over a field K. Then there exist an irreducible projective K-algebraic variety $\tilde{X}$ and a surjective K-morphism $f : \tilde{X} \to X$ with* $\dim(\tilde{X}) = \dim(X)$.

Proof. See [EGA II 1961, Corollaire II.5.6.2]. □

2D. *Amenable groups.* A group G is called *amenable* if there exist a directed set I and a family $(F_i)_{i \in I}$ of nonempty finite subsets of G such that

$$\lim_i \frac{|F_i \setminus F_i g|}{|F_i|} = 0 \quad \text{for all } g \in G, \tag{2-11}$$

(see [Ceccherini-Silberstein and Coornaert 2010, Chapter 4]). Such a family $(F_i)_{i \in I}$ is then called a (*right*) *Følner net* for G.

All finitely generated groups of subexponential growth and all solvable groups are amenable. Moreover, the class of amenable groups is closed under the operations of taking subgroups, quotients, extensions, and direct limits. On the other hand, every group containing a nonabelian free subgroup is nonamenable.

2E. *Tilings.* Let G be a group. Let E and E' be two finite subsets of G. A subset $T \subset G$ is called an (E, E')-*tiling* if it satisfies the following two conditions:

(T-1) The subsets gE, $g \in T$, are pairwise disjoint.

(T-2) $G = \bigcup_{g \in T} gE'$.

The following statement is an immediate consequence of Zorn's lemma (see [Ceccherini-Silberstein and Coornaert 2010, Proposition 5.6.3]).

Proposition 2.19. *Let G be a group. Let E be a nonempty finite subset of G and let $E' := EE^{-1} = \{gh^{-1} : g, h \in E\}$. Then there exists an (E, E')-tiling $T \subset G$.*

We shall need the following estimate on the growth of tilings with respect to Følner nets.

Proposition 2.20. *Let G be an amenable group and $(F_i)_{i \in I}$ be a Følner net for G. Let E and E' be finite subsets of G and suppose that $T \subset G$ is an (E, E')-tiling. For each $i \in I$, define the subset $T_i \subset T$ by $T_i := \{g \in T : gE \subset F_i\}$. Then there exist a real number $\alpha > 0$ and an element $i_0 \in I$ such that $|T_i| \geq \alpha |F_i|$ for all $i \geq i_0$.*

Proof. See [Ceccherini-Silberstein and Coornaert 2010, Proposition 5.6.4]. □

3. Algebraic cellular automata

3A. ***Interiors, neighborhoods, and boundaries.*** Let G be a group and let M be a finite subset of G. The *M-interior* Ω^- and the *M-neighborhood* Ω^+ of a subset $\Omega \subset G$ are the subsets of G defined, respectively, by

$$\Omega^- := \{g \in G : gM \subset \Omega\}$$

and

$$\Omega^+ := \Omega M = \{gh : g \in \Omega \text{ and } h \in M\}.$$

Note that $\Omega^- \subset \Omega \subset \Omega^+$ if $1_G \in M$.

We define the *M-boundary* $\partial\Omega$ of Ω by $\partial M = \Omega^+ \setminus \Omega^-$.

If G is an amenable group and $(F_i)_{i \in I}$ is a Følner net of G, then one has

$$\lim_i \frac{|\partial F_i|}{|F_i|} = 0 \quad \text{for every finite subset } M \subset G \tag{3-1}$$

(see, e.g., [Ceccherini-Silberstein and Coornaert 2010, Proposition 5.4.4]).

Let A be a set and let G be a group. Suppose now that we are given a cellular automaton $\tau : A^G \to A^G$ with memory set M. Let $\Omega \subset G$ and let Ω^- and Ω^+ be defined as above.

The cellular automaton τ induces maps $\tau_\Omega^- : A^\Omega \to A^{\Omega^-}$ and $\tau_\Omega^+ : A^{\Omega^+} \to A^\Omega$ defined, respectively, by

$$\tau_\Omega^-(u) := (\tau(c))|_{\Omega^-} \quad \text{for all } u \in A^\Omega,$$

and

$$\tau_\Omega^+(u) := (\tau(c))|_\Omega \quad \text{for all } u \in A^{\Omega^+},$$

where $c \in A^G$ is any configuration extending u. Observe that the maps τ_Ω^- and τ_Ω^+ are well defined. Indeed, (1-1) implies that $\tau(c)(g)$ only depends on the restriction of c to gM, for all $g \in G$.

3B. ***Cellular automata over algebraic varieties.*** (see [Ceccherini-Silberstein et al. 2019]) Let S be a scheme and let X, Y be S-schemes. We denote by $X(Y)$ the set of Y-points of X, i.e., the set consisting of all S-scheme morphisms $Y \to X$. If E is a finite set, X^E will denote the S-fibered product of a family of copies of X indexed by E. Note that $(X^E)(Y) = (X(Y))^E$ by the universal property of S-fibered products. If $f : Z \to X$ is an S-scheme morphism, then f induces a map $F^{(Y)} : Z(Y) \to X(Y)$ given by $f^{(Y)}(\varphi) = f \circ \varphi$ for all $\varphi \in Z(Y)$.

Let $A := X(Y)$ and let G be a group. Let $\tau : A^G \to A^G$ be a cellular automaton over the alphabet A and the group G. We say that τ is an *algebraic cellular automaton over the group G and schemes* S, X, Y if τ admits a memory set M such that the associated local defining map $\mu_M : A^M \to A$ satisfies the following condition:

($*$) There exists an S-scheme morphism $f : X^M \to X$ such that $\mu_M = f^{(Y)}$.

Remark 3.1. If $X(S) \neq \varnothing$, and condition ($*$) is satisfied for some memory set M of τ, then ($*$) is satisfied for any memory set of τ (see [Ceccherini-Silberstein et al. 2019, Proposition 3.1]). This applies in particular when $S = \operatorname{Spec}(K)$ for some algebraically closed field K since in that case $X(S)$ is (identified with) the set of closed points of X and X is Jacobson.

Lemma 3.2. *Let S be a scheme and let X, Y be S-schemes. Let $A := X(Y)$ and let G be a group. Suppose that $\tau : A^G \to A^G$ is an algebraic cellular automaton over the schemes S, X, Y and let M be a memory set of τ satisfying* ($*$). *Let Ω be a finite subset of G. Then there exist S-scheme morphisms $f_\Omega^- : X^\Omega \to X^{\Omega^-}$ and $f_\Omega^+ : X^{\Omega^+} \to X^\Omega$ such that $f_\Omega^{-(Y)} = \tau_\Omega^-$ and $f_\Omega^{+(Y)} = \tau_\Omega^+$.*

Proof. We prove the assertion for f_Ω^-. The construction of f_Ω^+ is similar. For every $g \in \Omega^-$, we consider the S-scheme projection morphism $p_g : X^\Omega \to X^{gM}$ and the S-scheme isomorphism $i_g : X^{gM} \to X^M$ induced by the bijective map $gM \to M$ given by left-multiplication by g^{-1}. Then the family of S-scheme morphisms $f \circ i_g \circ p_g : X^\Omega \to X$ for $g \in \Omega^-$ yields, by the universal property of S-fibered products, a S-scheme morphism $f_\Omega^- : X^\Omega \to X^{\Omega^-}$. It is clear from this construction that $f_\Omega^{-(Y)} = \tau_\Omega^-$. □

Let G be a group and let K be a field. Let X be a K-algebraic variety and let $A := X(K)$. We say that a cellular automaton $\tau : A^G \to A^G$ is an *algebraic cellular automaton over* (G, X, K) if τ is an algebraic cellular automaton over the group G and the schemes (K, X, K), i.e., for some, or equivalently any (by Remark 3.1), memory set M of τ, there exists a K-scheme morphism $f : X^M \to X$ such that $f^{(K)} : A^M \to A$ is the local defining map of τ associated with M.

Suppose now that the field K is algebraically closed. Recall from Remark 2.8, that A is regarded as the set of closed points of X. Given a finite subset Ω of G, we denote by X^Ω the K-fibered product of a family of copies of X indexed by Ω. It follows from Assertion (ii) in Proposition 2.12 that

$$A^\Omega = (X(K))^\Omega = X^\Omega(K).$$

Thus, A^Ω is the set of closed points of the algebraic variety X^Ω. Note that Proposition 2.9 and Assertion (iii) in Proposition 2.12 imply that

$$\dim(A^\Omega) = \dim(X^\Omega) = |\Omega| \dim(X) < \infty. \tag{3-2}$$

In what follows, every subset of A^Ω, or more generally of X^Ω, is equipped with the topology induced by the Zariski topology on X^Ω.

Remark 3.3. Let $\iota : X_{\mathrm{red}} \to X$ denote the reduced scheme associated to the K-algebraic variety X. Then X_{red} is also a K-algebraic variety and the immersion ι induces the identification $X_{\mathrm{red}}(K) = X(K)$. Moreover, $(X_{\mathrm{red}})^\Omega = (X^\Omega)_{\mathrm{red}}$ for every finite subset $\Omega \subset G$, so that every algebraic cellular automaton over (G, X, K) can be considered as an algebraic cellular automaton over (G, X_{red}, K) (see [Ceccherini-Silberstein et al. 2019, Remark A.11]). Hence, there is no loss of generality to assume that X is reduced.

Proposition 3.4. *Let G be a group and let X be an algebraic variety over an algebraically closed field K. Let $A := X(K)$ and let $\tau : A^G \to A^G$ be an algebraic cellular automaton over (G, X, K). Let $\Gamma \subset A^G$ and let $\Phi := \tau(\Gamma)$ denote the image of Γ under τ. Then the following hold*:

(i) *If Γ_Ω is a constructible subset of A^Ω for every finite subset $\Omega \subset G$, then Φ_Ω is a constructible subset of A^Ω for every finite subset $\Omega \subset G$.*

(ii) *If the variety X is complete and Γ_Ω is closed in A^Ω for every finite subset $\Omega \subset G$, then Φ_Ω is closed in A^Ω for every finite subset $\Omega \subset G$.*

Proof. Let M be a memory set of τ such that the associated local defining map $\mu : A^M \to A$ is induced by some K-scheme morphism $f : X^M \to X$. Let Ω be a finite subset of G and define Ω^+, τ^+_Ω, and f^+_Ω as in Section 3A and Lemma 3.2. Note that $\Phi_\Omega = \tau^+_\Omega(\Gamma_{\Omega^+})$. Thus, if Γ_{Ω^+} is a constructible subset of A^{Ω^+}, then $\Phi_\Omega = f^+_\Omega(\Gamma_{\Omega^+})$ is a constructible subset of A^Ω by Proposition 2.10(vi). This shows (i).

Suppose now that the variety X is complete, i.e., proper over K. Then, X^{Ω^+} and X^Ω are also proper over K since fibered products of proper schemes are proper. As every K-morphism between proper K-schemes is closed, it follows that $f^+_\Omega : X^{\Omega^+} \to X^\Omega$ is closed. Assume now that Γ_{Ω^+} is closed in A^{Ω^+}. This means that there exists a closed subset F of X^{Ω^+} such that $\Gamma_{\Omega^+} = A^{\Omega^+} \cap F$ is the set of closed points of F. We then get, by using Proposition 2.10(ii),

$$\Phi_\Omega = f^+_\Omega(\Gamma_{\Omega^+}) = f^+_\Omega(A^{\Omega^+} \cap F) = A^\Omega \cap f^+_\Omega(F),$$

which implies that Φ_Ω is closed in A^Ω. This shows (ii). □

4. Algebraic mean dimension

The definition of algebraic mean dimension we introduce in this section is analogous to that of topological and measure-theoretic entropy, as well as the various notions of mean dimension introduced by Gromov [1999].

Definition 4.1. Let G be an amenable group and let $\mathcal{F} = (F_i)_{i \in I}$ be a Følner net for G. Let X be an algebraic variety over an algebraically closed field K and let $A := X(K)$ denote the set of K-points of X. The *algebraic mean dimension* of a subset $\Gamma \subset A^G$ with respect to $\mathcal{F}$ is the quantity $\operatorname{mdim}_{\mathcal{F}}(\Gamma)$ defined by

$$\operatorname{mdim}_{\mathcal{F}}(\Gamma) := \limsup_{i \in I} \frac{\dim(\Gamma_{F_i})}{|F_i|}, \tag{4-1}$$

where $\dim(\Gamma_{F_i})$ denotes the Krull dimension of $\Gamma_{F_i} \subset A^{F_i} \subset X^{F_i}$ with respect to the Zariski topology and $|\cdot|$ denotes cardinality.

Here are some immediate properties of algebraic mean dimension.

Proposition 4.2. *Let G be an amenable group and let $\mathcal{F} = (F_i)_{i \in I}$ be a Følner net for G. Let X be an algebraic variety over an algebraically closed field K and let $A := X(K)$. Then the following hold*:

(i) $\operatorname{mdim}_{\mathcal{F}}(A^G) = \dim(X)$.

(ii) *For all subsets* $\Gamma, \Gamma' \subset A^G$ *such that* $\Gamma \subset \Gamma'$, *one has* $\operatorname{mdim}_{\mathcal{F}}(\Gamma) \le \operatorname{mdim}_{\mathcal{F}}(\Gamma')$.

(iii) *For every subset* $\Gamma \subset A^G$, *one has* $\operatorname{mdim}_{\mathcal{F}}(\Gamma) \le \dim(X)$.

Proof. (i) For every $i \in I$, we have that $(A^G)_{F_i} = A^{F_i}$, so that

$$\begin{aligned} \frac{\dim((A^G)_{F_i})}{|F_i|} &= \frac{\dim(A^{F_i})}{|F_i|} \\ &= \frac{|F_i| \dim(X)}{|F_i|} \qquad \text{(by (3-2))} \\ &= \dim(X). \end{aligned}$$

It follows that $\operatorname{mdim}_{\mathcal{F}}(A^G) = \limsup_i \dim(X) = \dim(X)$.

(ii) Suppose that $\Gamma \subset \Gamma' \subset A^G$. Then, for all $i \in I$, we have that $\Gamma_{F_i} \subset \Gamma'_{F_i}$ and hence $\dim(\Gamma_{F_i}) \le \dim(\Gamma'_{F_i})$ by applying Proposition 2.1(i). We deduce that

$$\operatorname{mdim}_{\mathcal{F}}(\Gamma) = \limsup_{i \in I} \frac{\dim(\Gamma_{F_i})}{|F_i|} \le \limsup_{i \in I} \frac{\dim(\Gamma'_{F_i})}{|F_i|} = \operatorname{mdim}_{\mathcal{F}}(\Gamma').$$

Assertion (iii) is an immediate consequence of (i) and (ii). $\square$

5. Algebraic mean dimension and surjectivity

Proposition 5.1. *Let G be an amenable group and let $\mathcal{F} = (F_i)_{i \in I}$ be a Følner net for G. Let X be an algebraic variety over an algebraically closed field K and let $A := X(K)$. Let $\tau : A^G \to A^G$ be an algebraic cellular automaton over (G, X, K). Suppose that a subset $\Gamma \subset A^G$ satisfies the following property: for every finite subset $\Omega \subset G$, the set Γ_Ω is a constructible subset of A^Ω for the Zariski topology. Then*

$$\mathrm{mdim}_{\mathcal{F}}(\tau(\Gamma)) \leq \mathrm{mdim}_{\mathcal{F}}(\Gamma).$$

Proof. Let $\Gamma' := \tau(\Gamma)$. Let $M \subset G$ be a memory set of τ such that the associated local defining map $\mu : A^M \to A$ satisfies $\mu = f^{(K)}$ for some K-scheme morphism $f : X^M \to X$. By Remark 3.1, up to replacing M by $M \cup \{1_G\}$ if necessary, we can assume that $1_G \in M$. Let Ω be a finite subset of G. Observe that, using the notation introduced in Section 3A and Lemma 3.2,

$$\Gamma'_{\Omega^-} = \tau_\Omega^-(\Gamma_\Omega) = f_\Omega^-(\Gamma_\Omega).$$

Therefore, it follows from Proposition 2.10(vi) that

$$\dim(\Gamma'_{\Omega^-}) \leq \dim(\Gamma_\Omega). \tag{5-1}$$

Since $1_G \in M$, we have that $\Omega^- \subset \Omega \subset \Omega^+$ and thus

$$\Gamma'_\Omega \subset \Gamma'_{\Omega^-} \times A^{\Omega \setminus \Omega^-} \subset X^{\Omega^-} \times_K X^{\Omega \setminus \Omega^-} = X^\Omega.$$

Therefore, we find that

$$\begin{aligned}
\dim(\Gamma'_\Omega) &\leq \dim(\Gamma'_{\Omega^-} \times A^{\Omega \setminus \Omega^-}) && \text{(by Proposition 2.1)}\\
&= \dim(\Gamma'_{\Omega^-}) + \dim(A^{\Omega \setminus \Omega^-}) && \text{(by Proposition 2.13(iii))}\\
&= \dim(\Gamma'_{\Omega^-}) + |\Omega \setminus \Omega^-| \dim(A) && \text{(by Proposition 2.13(iii))}\\
&\leq \dim(\Gamma_\Omega) + |\Omega \setminus \Omega^-| \dim(A) && \text{(by (5-1))}.
\end{aligned}$$

Since $\Omega \setminus \Omega^- \subset \partial\Omega$, we deduce that

$$\dim(\Gamma'_\Omega) \leq \dim(\Gamma_\Omega) + |\partial\Omega| \dim(A).$$

Taking $\Omega = F_i$, this gives us

$$\frac{\dim(\Gamma'_{F_i})}{|F_i|} \leq \frac{\dim(\Gamma_{F_i})}{|F_i|} + \frac{|\partial F_i|}{|F_i|} \dim(A).$$

Since

$$\lim_{i \in I} |\partial F_i| / |F_i| = 0$$

by (3-1), we conclude that

$$\mathrm{mdim}_{\mathcal{F}}(\Gamma') = \limsup_{i \in I} \frac{\dim(\Gamma'_{F_i})}{|F_i|} \leq \limsup_{i \in I} \frac{\dim(\Gamma_{F_i})}{|F_i|} = \mathrm{mdim}_{\mathcal{F}}(\Gamma). \qquad \square$$

Lemma 5.2. *Let G be an amenable group and let $\mathcal{F} = (F_i)_{i \in I}$ be a Følner net for G. Let X be an irreducible algebraic variety over an algebraically closed field K and let $A := X(K)$. Suppose that $\Gamma \subset A^G$ satisfies the following condition:*

(C) *There exist finite subsets $E, E' \subset G$ and an (E, E')-tiling $T \subset G$ such that for all $g \in T$, $\Gamma_{gE} \subsetneqq A^{gE}$ is a proper closed subset of A^{gE} for the Zariski topology.*

Then one has $\operatorname{mdim}_{\mathcal{F}}(\Gamma) < \dim(X)$.

Proof. For each $i \in I$, define, as in Proposition 2.20, the subset $T_i \subset T$ by $T_i := \{g \in T : gE \subset F_i\}$ and set

$$F_i^* := F_i \setminus \coprod_{g \in T_i} gE,$$

where $\coprod$ denotes disjoint union. For all $g \in T$, the set Γ_{gE} is a proper closed subset of A^{gE} by our hypothesis (C). As A^{gE} is irreducible since X is irreducible and K is algebraically closed (see Proposition 2.12(iv) and Corollary 2.6), it follows from Proposition 2.1(ii) that

$$\dim(\Gamma_{gE}) \leq \dim(A^{gE}) - 1 = |gE| \dim(A) - 1 = |gE| \dim(X) - 1 \tag{5-2}$$

for all $g \in T$. Now observe that, for each $i \in I$,

$$\Gamma_{F_i} \subset A^{F_i^*} \times \prod_{g \in T_i} \Gamma_{gE} \subset X^{F_i^*} \times_K \prod_{g \in T_i} X^{gE} = X^{F_i},$$

so that

$$\begin{aligned}
\dim(\Gamma_{F_i}) &\leq \dim\Big(A^{F_i^*} \times \prod_{g \in T_i} \Gamma_{gE}\Big) && \text{(by Proposition 2.1(i))} \\
&= |F_i^*| \dim(A) + \sum_{g \in T_i} \dim(\Gamma_{gE}) && \text{(by Proposition 2.13(iii))} \\
&\leq |F_i^*| \dim(A) + \sum_{g \in T_i} (|gE| \dim(A) - 1) && \text{(by (5-2))} \\
&= \Big(|F_i^*| + \sum_{g \in T_i} |gE|\Big) \dim(A) - |T_i| \\
&= |F_i| \dim(A) - |T_i| = |F_i| \dim(X) - |T_i|.
\end{aligned}$$

Now, by virtue of Proposition 2.20, there exist $\alpha > 0$ and $i_0 \in I$ such that $|T_i| \geq \alpha |F_i|$ for all $i \geq i_0$. We deduce that, for all $i \geq i_0$,

$$\frac{\dim(\Gamma_{F_i})}{|F_i|} \leq \dim(X) - \alpha.$$

This implies that

$$\operatorname{mdim}_{\mathcal{F}}(\Gamma) = \limsup_{i \in I} \frac{\dim(\Gamma_{F_i})}{|F_i|} \leq \dim(X) - \alpha < \dim(X). \qquad \square$$

Lemma 5.3. *Let G be an amenable group and let $\mathcal{F} = (F_i)_{i \in I}$ be a Følner net for G. Let X be an irreducible algebraic variety over an algebraically closed field K and let $A := X(K)$. Suppose that a G-invariant subset $\Gamma \subset A^G$ satisfies the following conditions*:

(D1) Γ *is closed in A^G for the prodiscrete topology.*

(D2) *For every finite subset Ω of G, the set Γ_Ω is closed in A^Ω for the Zariski topology.*

(D3) $\operatorname{mdim}_{\mathcal{F}}(\Gamma) = \dim(X)$.

Then one has $\Gamma = A^G$.

Proof. We proceed by contradiction. Suppose that there is a configuration $c \in A^G$ that does not belong to Γ. By (D1), the set $A^G \setminus \Gamma$ is an open subset of A^G for the prodiscrete topology. Thus, we can find a finite subset $E \subset G$ such that $c|_E \notin \Gamma_E$. This implies that $\Gamma_E \subsetneq A^E$. As Γ is G-invariant, we have that

$$\Gamma_{gE} \subsetneq A^{gE} \quad \text{for all } g \in G. \tag{5-3}$$

By Proposition 2.19, there exist a finite subset $E' \subset G$ and an (E, E')-tiling $T \subset G$. Since Γ satisfies the hypotheses of Lemma 5.2 by (D2) and (5-3), we deduce that $\operatorname{mdim}_{\mathcal{F}}(\Gamma) < \dim(X)$, which contradicts (D3). $\square$

We can now prove the main result of this section.

Theorem 5.4. *Let G be an amenable group and let $\mathcal{F}$ be a Følner net for G. Let X be an irreducible complete algebraic variety over an algebraically closed field K and let $A := X(K)$. Let $\tau : A^G \to A^G$ be an algebraic cellular automaton over (G, X, K). Suppose that* $\operatorname{mdim}_{\mathcal{F}}(\tau(A^G)) = \dim(X)$. *Then τ is surjective.*

Proof. Let us check that $\Gamma := \tau(A^G)$ satisfies the hypotheses of Lemma 5.3. Condition (D1), that is, the fact that Γ is closed in A^G for the prodiscrete topology, follows from Theorem 4.1 in [Ceccherini-Silberstein et al. 2019]. On the other hand, Condition (D2), that is, the fact that Γ_Ω is closed in A^Ω with respect to the Zariski topology for every finite subset $\Omega \subset G$, is satisfied by Proposition 3.4(ii). By applying Lemma 5.3, we conclude that $\Gamma = A^G$. This shows that τ is surjective. $\square$

6. Algebraic mean dimension and weak preinjectivity

In this section, we introduce two notions of weak preinjectivity for algebraic cellular automata, i.e., $(*)$-preinjectivity and $(**)$-preinjectivity. We shall see that these notions are equivalent under general hypotheses and are both implied by preinjectivity.

We use the following notation. Given a set A, a group G, a finite subset $\Omega \subset G$, a subset $D \subset A^{\Omega}$, and an element $p \in A^{G \setminus \Omega}$, we write

$$D_p := D \times \{p\} = \{c \in A^G : c|_{\Omega} \in D \text{ and } c|_{G \setminus \Omega} = p\}.$$

We say that a subset $\Gamma \subset A^G$ has *finite support* if $\Gamma = D_p$ for some D, p as above.

Let $\tau : A^G \to A^G$ be a cellular automaton over the group G and the alphabet A with memory set M. Observe that if $\Gamma \subset A^G$ has finite support then $\tau(\Gamma)$ also has finite support. Indeed, $\tau(D_p) = R_s$ for some subset $R \subset A^{\Omega^+}$ and $s = s(\tau, p) \in A^{G \setminus \Omega^+}$. Suppose now that X is an algebraic variety over an algebraically closed field K and $A = X(K)$. Then we write $\dim(D_p) := \dim(D)$, where $\dim(D)$ is the Krull dimension of $D \subset A^{\Omega}$ with respect to the Zariski topology. Note that $\dim(D_p)$ is well defined. Indeed, suppose that $C_q = D_p$ for some $C \subset A^{\Lambda}$ and $q \in A^{G \setminus \Lambda}$, where Λ is a finite subset of G. Then clearly C and D are homeomorphic so that $\dim(C_q) = \dim(C) = \dim(D) = \dim(D_p)$.

Definition 6.1. Let G be a group and let X be an algebraic variety over an algebraically closed field K. Let $A := X(K)$ and let $\tau : A^G \to A^G$ be an algebraic cellular automaton over (G, X, K).

We say that τ is $(*)$-*preinjective* if there do not exist a finite subset $\Omega \subset G$ and a subset $H \subsetneq A^{\Omega}$ that is closed for the Zariski topology such that

$$\tau((A^{\Omega})_p) = \tau(H_p) \quad \text{for all } p \in A^{G \setminus \Omega}.$$

We say that τ is $(**)$-*preinjective* if there does not exist a finite subset $\Omega \subset G$ such that

$$\dim(\tau((A^{\Omega})_p)) < \dim(A^{\Omega}) \quad \text{for all } p \in A^{G \setminus \Omega}.$$

Remark 6.2. Let us note that $(*)$-preinjectivity and $(**)$-preinjectivity as well as preinjectivity itself, are *local* properties. More precisely, using again the notation of Definition 6.1 and Section 3A, let M be a memory set of τ such that $1_G \in M$ and $M = M^{-1}$. Then $(*)$-preinjectivity amounts to saying that, for every finite subset $\Omega \subset G$, there exist no proper closed subsets $H \subsetneq A^{\Omega}$ such that

$$\tau^+_{\Omega}(A^{\Omega} \times \{q\}) = \tau^+_{\Omega}(H \times \{q\}) \quad \text{for all } q \in A^{\Omega^+ \setminus \Omega}.$$

Similarly, $(**)$-preinjectivity means (by Proposition 2.10(iii)) that for every finite subset $\Omega \subset G$, we have

$$\dim(\tau^+_{\Omega}(A^{\Omega} \times \{q\})) = \dim(A^{\Omega}) \quad \text{for some } q \in A^{\Omega^+ \setminus \Omega}.$$

Finally, preinjectivity means that for every finite subset $\Omega \subset G$ and every $q \in A^{\Omega^{++} \setminus \Omega}$ (where $\Omega^{++} = (\Omega^+)^+$), the restriction of

$$\tau^+_{\Omega^+} : A^{\Omega^{++}} \to A^{\Omega^+}$$

to $A^{\Omega} \times \{q\} \subset A^{\Omega^{++}}$ is injective.

In order to establish, in the next proposition, some key relations between preinjectivity, $(*)$-preinjectivity, and $(**)$-preinjectivity, we shall use the following general auxiliary result.

Lemma 6.3. *Let X be an irreducible complete algebraic variety over an algebraically closed field K. Then there exists a proper closed subset $H \subsetneq X$ satisfying the following property*:

(P) *If Y is a K-algebraic variety with* $\dim(Y) < \dim(X)$ *and $h : X \twoheadrightarrow Y$ is a surjective K-scheme morphism, then one has $h(H) = Y$.*

Proof. Since X is irreducible and complete over K, it follows from Chow's lemma (see Theorem 2.18) that there exist an irreducible projective K-algebraic variety $\tilde{X}$ with $\dim(\tilde{X}) = \dim(X)$ and a surjective morphism $f : \tilde{X} \to X$ of K-schemes. Let $\tilde{H}$ be a hyperplane section of the projective variety $\tilde{X}$ (see Corollary 2.16 and Remark 2.17). Let $H := f(\tilde{H}) \subset X$. As f is a morphism between proper schemes, it is proper and hence closed. Thus, H is a closed subset of X. Since $\tilde{H} \subsetneq X$ is a proper closed subset and $\tilde{X}$ is irreducible, we have $\dim(\tilde{H}) < \dim(\tilde{X})$. We deduce from Proposition 2.10(iii) that

$$\dim(X) = \dim(\tilde{X}) > \dim(\tilde{H}) \geq \dim(f(\tilde{H})) = \dim(H).$$

It follows that H is a proper closed subset of X.

Now let Y and $h : X \twoheadrightarrow Y$ be as in the statement of the lemma. As X is irreducible, $Y = h(X)$ is also irreducible. Consider the surjective composite morphism

$$g := h \circ f : \tilde{X} \to Y.$$

By Assertion (ii) of Proposition 2.11 applied to $g : \tilde{X} \to Y$, for every closed point $y \in Y$, the closed subset $g^{-1}(y) \subset \tilde{X}$ satisfies

$$\dim(g^{-1}(y)) \geq \dim(\tilde{X}) - \dim(Y) = \dim(X) - \dim(Y) \geq 1.$$

Hence, we deduce from Corollary 2.16 that the closed subset $\tilde{H} \cap g^{-1}(y)$ is nonempty for every closed point $y \in Y$. Therefore, $f(\tilde{H})$ contains the set of closed points Y_0 of Y. As $g = h \circ f$ and $H = f(\tilde{H})$, it follows that

$$h(H) = h(f(\tilde{H})) = g(\tilde{H}) \supset Y_0. \tag{6-1}$$

Since $h(H)$ is constructible in Y by Chevalley's theorem, $Y \setminus h(H)$ is also constructible and hence Jacobson (see Proposition 2.9(iv)). On the other hand, $Y \setminus h(H)$ does not contain any closed point of Y by (6-1). From Proposition 2.9(iv), we then deduce that the Jacobson space $Y \setminus h(H)$ has no closed points. It follows that $Y \setminus h(H)$ is empty. This shows that $h(H) = Y$. ☐

Proposition 6.4. *Let G be a group and let X be an algebraic variety over an algebraically closed field K. Let $A := X(K)$ and let $\tau : A^G \to A^G$ be an algebraic cellular automaton over (G, X, K). Then the following hold*:

(i) *If τ is preinjective, then τ is both $(*)$-preinjective and $(**)$-preinjective.*

(ii) *If X is irreducible and τ is $(**)$-preinjective, then τ is $(*)$-preinjective.*

(iii) *If X is irreducible and complete over K, then τ is $(*)$-preinjective if and only if it is $(**)$-preinjective.*

Proof. Suppose first that X is irreducible and that τ is not $(*)$-preinjective, i.e., there exists a finite subset $\Omega \subset G$ and a closed subset $H \subsetneq A^\Omega$ such that

$$\tau((A^\Omega)_p) = \tau(H_p) \quad \text{for all } p \in A^{G\setminus\Omega}. \tag{6-2}$$

Since K is algebraically closed, we deduce from Proposition 2.13(iv) that X^Ω and hence A^Ω are irreducible. Thus, it follows from (6-2) that

$$\begin{aligned} \dim(\tau((A^\Omega)_p)) &= \dim(\tau(H_p)) \\ &\leq \dim(H_p) = \dim(H) && \text{(by Proposition 2.10(iii))} \\ &< \dim(A^\Omega) && \text{(by Proposition 2.1(ii))} \end{aligned}$$

for all $p \in A^{G\setminus\Omega}$. Therefore τ is not $(**)$-preinjective. This proves (ii).

Let M be a memory set of τ and $f : X^M \to X$ be a K-scheme morphism such that $f^{(K)} : A^M \to A$ is the local defining map associated with M. After enlarging M if necessary, we can assume $1_G \in M$ and $M = M^{-1}$. We use again the notation introduced in Section 3A and write $\Omega^{++} := (\Omega^+)^+$. Note that $\Omega \subset \Omega^+ \subset \Omega^{++}$, since $1_G \in M$.

For the proof of (i) and (iii), we shall use the following construction. We suppose that τ is not $(**)$-preinjective, i.e., there exists a finite subset $\Omega \subset G$ such that

$$\dim(\tau((A^\Omega)_p)) < \dim(A^\Omega) \quad \text{for all } p \in A^{G\setminus\Omega}. \tag{6-3}$$

Let $p \in A^{G\setminus\Omega}$ with Ω as above and a configuration $c \in (A^\Omega)_p$ extending p. Observe that $\tau(c)|_{G\setminus\Omega^+}$ only depends on p (here we use the fact that $gM \subset G \setminus \Omega$ for all $g \in G \setminus \Omega^+$ since $M^{-1} = M$).

Consider the following closed immersion induced by p:

$$\iota := (\mathrm{Id}_{X^\Omega}, p|_{\Omega^{++}\setminus\Omega}) : X^\Omega = X^\Omega \times_K \prod_{\Omega^{++}\setminus\Omega} \mathrm{Spec}(K) \to X^{\Omega^{++}}.$$

Let $Z := \iota(X^\Omega) \subset X^{\Omega^{++}}$ be the closed image of ι equipped with the reduced closed subscheme structure. Let $j : Z \to X^{\Omega^{++}}$ be the corresponding closed immersion. Since we can assume that X is reduced by Remark 3.3, it follows from Proposition I.5.2.2 of [EGA I 1971] that ι factors through a morphism $\gamma : X^\Omega \to Z$.

Note that Z is homeomorphic to X^Ω. Note also that for any subset $\Gamma \subset A^\Omega$, we have

$$\gamma(\Gamma) = \Gamma \times \{p|_{\Omega^{++}\setminus\Omega}\} = \Gamma_p|_{\Omega^{++}\setminus\Omega}. \tag{6-4}$$

We consider now the K-scheme morphism

$$h := f^+_{\Omega^+} \circ j : Z \to X^{\Omega^+},$$

where $f^+_{\Omega^+} : X^{\Omega^{++}} \to X^{\Omega^+}$ is defined as in Lemma 3.2. Clearly,

$$\sigma := h^{(K)} : Z(K) \to A^{\Omega^+} \tag{6-5}$$

is the restriction of $\tau^+_{\Omega^+}$ to $Z(K) = A^\Omega \times \{p|_{\Omega^{++}\setminus\Omega}\}$.

Let $Y := \overline{\mathrm{Im}(h)} \subset X^{\Omega^+}$ be the closure of $\mathrm{Im}(h)$ in X^{Ω^+}. We equip Y with the induced reduced closed subscheme structure over K. By [EGA I 1971, Proposition I.5.2.2], the morphism h factors through a K-scheme morphism

$$k : Z \to Y.$$

Observe that $\mathrm{Im}(h)$ is constructible in X^{Ω^+} by Chevalley's theorem. We then have the identifications $\sigma(Z(K)) = h(Z_0) = \mathrm{Im}(h)_0$ by Proposition 2.10(ii) and Remark 2.8. From Proposition 2.9(vi),

$$\dim(\sigma(Z(K))) = \dim(\mathrm{Im}(h)_0) = \dim(\mathrm{Im}(h)) = \dim(\overline{\mathrm{Im}(h)}) = \dim(Y).$$

Thus, we deduce from inequality (6-3) and the above equalities that

$$\begin{aligned} \dim(Z) &= \dim(X^\Omega) = \dim(A^\Omega) \\ &> \dim(\tau((A^\Omega)_p)) = \dim(\tau^+_{\Omega^+}(A^\Omega \times \{p|_{\Omega^{++}\setminus\Omega}\})) \\ &= \dim(\sigma(Z(K))) = \dim(Y). \end{aligned} \tag{6-6}$$

In order to show (i), suppose that τ is preinjective but not $(**)$-preinjective. Let $\Omega \subset G$, $p \in A^{G\setminus\Omega}$ and the maps h, k be constructed as above. Since τ is preinjective, the map $\sigma = h^{(K)}$ (see (6-5)) is injective. As K is algebraically closed, we can identify closed points of Z and Y with $Z(K)$ and $Y(K)$, respectively. We deduce (from [Ceccherini-Silberstein et al. 2019, Lemma A.22(iii)] for example) that h and thus k are injective. Proposition 2.11(i) applied to $k : Z \to Y$ shows that there exists a closed point $b \in Y_0 \subset X^{\Omega^+}$ such that

$$\dim(h^{-1}(b)) = \dim(k^{-1}(b)) \geq \dim(Z) - \dim(Y) \geq 1,$$

where the last inequality follows from (6-6). This is a contradiction since $h^{-1}(b) = k^{-1}(b)$ is Jacobson and has at most one closed point by the injectivity of h and k. This proves that when τ is preinjective, it must be $(**)$-preinjective. Since preinjectivity implies trivially $(*)$-preinjectivity by the definition, the point (i) is proved.

We proceed now to the proof of (iii). Suppose that X is irreducible and complete over K and that τ is not $(**)$-preinjective. Let $\Omega \subset G$, $p \in A^{G\setminus\Omega}$, $c \in (A^\Omega)_p$ and the maps h, k, γ be as above. Observe again that X^Ω is irreducible by Proposition 2.13(iv). As X is proper, the varieties X^Ω, X^{Ω^+} and Z are also proper over K. Hence $h : Z \to X^{\Omega^+}$ is a morphism of proper K-schemes. Consequently, h is closed, $Y = \mathrm{Im}(h)$ and thus $k : Z \to Y$ is surjective.

Since X^Ω is irreducible and complete, Lemma 6.3 shows that there exists a proper closed subset $L \subsetneq X^\Omega$ independent of p satisfying:

(P) If V is a K-algebraic variety with $\dim(V) < \dim(X^\Omega)$ and $\Phi : X^\Omega \twoheadrightarrow V$ is a surjective K-scheme morphism, then one has $\Phi(L) = V$.

We consider the set of K-points of L:

$$H := L(K) \subsetneq A^\Omega.$$

Applying Property (P) to the surjective morphism $k \circ \gamma : X^\Omega \twoheadrightarrow Y$, we deduce that $k(\gamma(L)) = Y$. Since a surjective morphism between K-algebraic varieties induces a surjective map between their sets of closed points (see [Ceccherini-Silberstein et al. 2019, Lemma A.22(iv)]), we deduce that

$$h(H \times \{p|_{\Omega^{++}\setminus\Omega}\}) = k((\gamma(H)) = Y(K) = h(A^\Omega \times \{p|_{\Omega^{++}\setminus\Omega}\}). \tag{6-7}$$

Now let $d = h(c|_{\Omega^{++}}) = \tau(c)|_{\Omega^+} \in A^{\Omega^+}$ be identified with a closed point in $Y_0 = Y(K)$. By (6-7), there exists $z \in H = L(K)$ such that $h(z \times \{p|_{\Omega^{++}\setminus\Omega}\}) = d$. Let $c' \in H_p \subset A^G$ be a configuration such that $c'|_\Omega = z$ and $c'|_{G\setminus\Omega} = p$. Then we see that

$$h(c'|_{\Omega^{++}}) = h(c|_{\Omega^{++}}) = d.$$

Hence, $\tau(c')|_{\Omega^+} = \tau(c)|_{\Omega^+} = d$. As $c'|_{G\setminus\Omega} = c|_{G\setminus\Omega} = p$, we have $\tau(c')|_{G\setminus\Omega^+} = \tau(c)|_{G\setminus\Omega^+}$. Thus, we find that $\tau(c') = \tau(c)$. As $p \in A^{G\setminus\Omega}$ and $c \in (A^\Omega)_p$ are arbitrary and since $c' \in H_p$, we deduce that

$$\tau((A^\Omega)_p) = \tau(H_p) \quad \text{for all } p \in A^{G\setminus\Omega}.$$

Thus, τ is not $(*)$-preinjective. Hence, $(*)$-preinjectivity implies $(**)$-preinjectivity when X is irreducible and complete over K. Together with (ii), this completes the proof of (iii). □

Proposition 6.5. *Let G be an amenable group and let $\mathcal{F} = (F_i)_{i\in I}$ be a Følner net for G. Let X be an algebraic variety over an algebraically closed field K and let $A := X(K)$. Let $\tau : A^G \to A^G$ be an algebraic cellular automaton over (G, X, K). Suppose that τ is $(**)$-preinjective. Then one has*

$$\mathrm{mdim}_{\mathcal{F}}(\tau(A^G)) = \dim(X). \tag{6-8}$$

Proof. We proceed by contradiction. Suppose that (6-8) is not satisfied, i.e.,

$$\operatorname{mdim}_{\mathcal{F}}(\tau(A^G)) < \dim(X). \tag{6-9}$$

Let M be a memory set for τ such that $1_G \in M$. Let $\Gamma := \tau(A^G)$. As $\Gamma_{F_i^+}$ is a subset of $\Gamma_{F_i} \times A^{F_i^+ \setminus F_i}$ and Γ_{F_i} is a constructible subset of A^{F_i} (by Assertion (i) of Proposition 3.4), we have that

$$\begin{aligned}
\dim(\Gamma_{F_i^+}) &\leq \dim(\Gamma_{F_i} \times A^{F_i^+ \setminus F_i}) && \text{(by Proposition 2.1(i))}\\
&= \dim(\Gamma_{F_i}) + \dim(A^{F_i^+ \setminus F_i}) && \text{(by Proposition 2.13(iii))}\\
&= \dim(\Gamma_{F_i}) + |F_i^+ \setminus F_i| \dim(X) && \text{(by (3-2))}\\
&\leq \dim(\Gamma_{F_i}) + |\partial(F_i)| \dim(X) && (\text{since } F_i^+ \setminus F_i \subset \partial F_i).
\end{aligned}$$

Hence, we find that

$$\frac{\dim(\Gamma_{F_i^+})}{|F_i|} \leq \frac{\dim(\Gamma_{F_i})}{|F_i|} + \frac{|\partial(F_i)|}{|F_i|} \dim(X).$$

The above inequality together with (6-9) and (3-1) show that there exists $i_0 \in I$ such that

$$\dim(\Gamma_{F_{i_0}^+}) < |F_{i_0}| \dim(X). \tag{6-10}$$

Observe now that for all $p \in A^{G \setminus F_{i_0}}$, we have that $\tau((A^{F_{i_0}})_p) \subset \tau(A^G) \cap (A^{F_{i_0}^+})_q$ where $q = \tau(\tilde{p})|_{G \setminus F_{i_0}^+} \in A^{G \setminus F_{i_0}^+}$ and $\tilde{p} \in A^G$ is an arbitrary configuration that extends p. Therefore, we have for all $p \in A^{G \setminus F_{i_0}}$ that

$$\begin{aligned}
\dim(\tau((A^{F_{i_0}})_p)) &\leq \dim(\tau(A^G)|_{F_{i_0}^+}) = \dim(\Gamma_{F_{i_0}^+})\\
&< |F_{i_0}| \dim(A) = \dim(A^{F_{i_0}}) && \text{(by (6-10))}.
\end{aligned}$$

We can thus conclude that τ is not $(**)$-preinjective. □

As described by the next proposition, the converse of Proposition 6.5 also holds if we replace $(**)$-preinjectivity by $(*)$-preinjectivity.

Proposition 6.6. *Let G be an amenable group and let $\mathcal{F}$ be a Følner net for G. Let X be an algebraic variety over an algebraically closed field K. Let $A := X(K)$ and let $\tau : A^G \to A^G$ be an algebraic cellular automaton over (G, X, K). Suppose that X is irreducible and that*

$$\operatorname{mdim}_{\mathcal{F}}(\tau(A^G)) = \dim(X). \tag{6-11}$$

Then τ is $(*)$-preinjective.

Proof. We proceed by contradiction. Suppose that the cellular automaton τ is not $(*)$-preinjective. Thus, there exist a finite subset $E \subset G$ and a closed proper subset $H \subsetneq A^E$ such that

$$\tau((A^E)_p) = \tau(H_p) \quad \text{for all } p \in A^{G\setminus E}. \tag{6-12}$$

By Proposition 2.19, we can find a finite subset $E' \subset G$ such that G contains an (E, E')-tiling T. For every $t \in T$, we define $H_t \subset A^{tE}$ to be the image of H under the canonical bijective map $A^E \to A^{tE}$ that is induced by the left-multiplication by t^{-1}. Since τ is G-equivariant, we deduce from (6-12) that for each $t \in T$,

$$\tau((A^{tE})_p) = \tau((H_t)_p) \quad \text{for all } p \in A^{G\setminus tE}. \tag{6-13}$$

Consider the subset $\Gamma \subset A^G$ defined by

$$\Gamma := A^{G\setminus TE} \times \prod_{t\in T} H_t.$$

We claim that $\tau(A^G) = \tau(\Gamma)$. Indeed, let $c \in A^G$ be any configuration and let us show that there exists a configuration in Γ whose image under τ is equal to $\tau(c)$.

To see this, consider the set $\Phi \subset A^G$ consisting of all configurations $d \in A^G$ satisfying the following conditions:

(C1) $d|_{G\setminus TE} = c|_{G\setminus TE}$.

(C2) if $t \in T$, then $d|_{tE} = c|_{tE}$ or $d|_{tE} \in H_t$.

(C3) $\tau(d) = \tau(c)$.

Given a configuration $d \in \Phi$, we define the subset $T_d \subset T$ by

$$T_d := \{t \in T : d|_{tE} \in H_t\}.$$

We partially order Φ by the relation $\leq$ defined by

$$d \leq e \iff (T_d \subset T_e \text{ and } d|_{tE} = e|_{tE} \text{ for all } t \in T_d).$$

Let us check that Φ satisfies the hypotheses of Zorn's lemma. The set Φ is not empty since $c \in \Phi$. On the other hand, suppose that Ψ is a nonempty totally ordered subset of Φ. Let us show that Ψ admits an upper bound in Φ. To see this, first observe that, for $t \in T$ fixed, the restrictions $d|_{tE},\ d \in \Psi$, are eventually constant, i.e., there exists $\lambda_t \in A^{tE}$ such that $d|_{tE} = \lambda_t$ for all $d \in \Psi$ large enough (with respect to $\leq$). Consider now the configuration $e \in A^G$ defined by

$$e|_{G\setminus TE} = c|_{G\setminus TE} \quad \text{and} \quad e|_{tE} = \lambda_t \quad \text{for all } t \in T.$$

It is clear that e satisfies (C1) and (C2). If Ω is a finite subset of G, then there are only finitely many $g \in T$ such that $gE \subset \Omega$. It follows that there exists $d \in \Psi$ such that $e|_\Omega = d|_\Omega$. Taking $\Omega = gM$, where $g \in G$ and M is a memory set of τ, we deduce that $\tau(e)(g) = \tau(d)(g) = \tau(c)(g)$ for all $g \in G$. This shows that e also satisfies (C3). Thus, $e \in \Phi$ is an upper bound for Ψ. By Zorn's lemma, Φ admits a maximal element m. We have that $\tau(m) = \tau(c)$ since $m \in \Phi$ satisfies (C3). We also have that $m \in \Gamma$. Indeed, otherwise, there would be some $t \in T$ such that $m|_{tE} \notin H_t$. But then using (6-13), we could modify m on tE and get an element $m' \geq m$ in Φ with $T_{m'} = T_m \cup \{t\}$, contradicting the maximality of m. This completes the proof that $\tau(A^G) = \tau(\Gamma)$.

We then get

$$\begin{aligned} \operatorname{mdim}_{\mathcal{F}}(\tau(A^G)) &= \operatorname{mdim}_{\mathcal{F}}(\tau(\Gamma)) \\ &\leq \operatorname{mdim}_{\mathcal{F}}(\Gamma) && \text{(by Proposition 5.1)} \\ &< \dim(X) && \text{(by Lemma 5.2)}, \end{aligned}$$

which contradicts (6-11). Observe that the hypotheses of Proposition 5.1 are satisfied since Γ_Ω is a closed and hence constructible subset of A^Ω for every finite subset $\Omega \subset G$. Note also that the hypotheses of Lemma 5.2 are satisfied since X is assumed to be irreducible and $\Gamma_{tE} = H_t$ is a proper closed subset of A^{tE} for all $t \in T$. ☐

7. Main results

Proof of Theorem 1.3. The result follows from Propositions 4.2(i) and 6.6. ☐

The following statement contains Theorem 1.1 as well as Theorem 1.4.

Theorem 7.1. *Let G be an amenable group. Let X be an irreducible complete algebraic variety over an algebraically closed field K and let $A := X(K)$. Suppose that $\tau : A^G \to A^G$ is an algebraic cellular automaton over (G, X, K). Then the following conditions are equivalent*:

(a) *τ is surjective.*

(b) *τ is $(*)$-preinjective.*

(c) *τ is $(**)$-preinjective.*

(d) *For some (or, equivalently, any) Følner net $\mathcal{F}$ of the group G, one has* $\operatorname{mdim}_{\mathcal{F}}(\tau(A^G)) = \dim(X)$.

Moreover, if τ is preinjective then it is surjective.

Proof. The fact that (a) implies (d) follows from Proposition 4.2(i) and the converse implication from Theorem 5.4. Thus, (a) and (d) are equivalent. We know (d) implies (b) by Proposition 6.6, and (b) implies (c) by Proposition 6.4(iii). On the other hand, we have that (c) implies (d) by Proposition 6.5. This shows that conditions (b), (c), and (d) are equivalent.

Finally, the last assertion follows from the fact that preinjectivity implies $(*)$-preinjectivity by Proposition 6.4(i) and the implication (b) $\Longrightarrow$ (a). ☐

Let G be a group and $M \subset G$ be a finite subset. Let X be an algebraic variety over an algebraically closed field K and let $f: X^M \to X$ be a K-scheme morphism. For each field extension L/K, let $X_L := X \times_K \mathrm{Spec}(L)$ denote the L-algebraic variety obtained by the base change $\mathrm{Spec}(L) \to \mathrm{Spec}(K)$. Then $X_L(L) = X(L)$. We denote by $\tau^{(L)}: X(L)^G \to X(L)^G$ the algebraic cellular automaton over (G, X_L, L) with memory set M and associated local defining map $f^{(L)}$.

Theorem 7.2. *With the above notation, suppose in addition that G is amenable, X is irreducible and complete, and L is algebraically closed. Then $\tau^{(K)}$ is surjective if and only if $\tau^{(L)}$ is surjective.*

Proof. This follows from Theorem 5.4 and the invariance of dimension of algebraic varieties under base change of the ground field. Indeed, let $\Gamma^{(K)} := \tau^{(K)}(X(K)^G)$ and $\Gamma^{(L)} := \tau^{(L)}(X(L)^G)$. Let $\Omega \subset G$ be a finite subset. Then we have the identifications

$$\Gamma_\Omega^{(K)} = f_\Omega^{+(K)}(X(K)^{\Omega^+}) = f_\Omega^+(X^{\Omega^+})_0,$$
$$\Gamma_\Omega^{(L)} = f_\Omega^{+(L)}(X(L)^{\Omega^+}) = (f_\Omega^+ \times \mathrm{Id}_L)(X_L^{\Omega^+})_0.$$

Thus for all finite subsets $\Omega \subset G$, we find that

$$\dim(\Gamma_\Omega^{(K)}) = \dim(f_\Omega^+(X^{\Omega^+})) = \dim((f_\Omega^+ \times \mathrm{Id}_L)(X_L^{\Omega^+})) = \dim(\Gamma_\Omega^{(L)}).$$

Let $\mathcal{F} = (F_i)_{i \in I}$ be a Følner net of G. Then by the definition of mean dimension,

$$\mathrm{mdim}_{\mathcal{F}}(\Gamma^{(K)}) := \limsup_{i \in I} \frac{\dim(\Gamma_{F_i}^{(K)})}{|F_i|} = \limsup_{i \in I} \frac{\dim(\Gamma_{F_i}^{(L)})}{|F_i|} =: \mathrm{mdim}_{\mathcal{F}}(\Gamma^{(L)}).$$

We can therefore conclude from Theorem 5.4 that $\tau^{(K)}$ is surjective if and only if $\tau^{(L)}$ is surjective. $\square$

8. Counterexamples

In the following example, we shall see that Theorem 1.4, Theorem 5.4, and Proposition 6.4 become false if we remove the hypothesis that X is irreducible, even if X is assumed to be 0-dimensional.

Example 8.1. Let G be a group and K an algebraically closed field.

Suppose that X is a K-algebraic variety with $\dim(X) = 0$. Then $A := X(K)$ is a finite nonempty set. Moreover, every map $A \to A$ is induced by some K-scheme morphism $X \to X$. Conversely, given a finite nonempty set A, there exists a 0-dimensional K-algebraic variety X such that $X(K) = A$. We can take for example the reduced K-algebraic variety X obtained by taking the discrete union of a family indexed by A of copies of $\mathrm{Spec}(K)$.

Let A be a finite nonempty set and X a 0-dimensional K-algebraic variety such that $X(K) = A$. Clearly the cellular automata over the group G and the alphabet A are precisely the algebraic cellular automata over (G, X, K).

Now let $\tau : A^G \to A^G$ be a cellular automaton over the group G and the alphabet A. By the classical Garden of Eden theorem in [Ceccherini-Silberstein et al. 1999], the surjectivity of τ is equivalent to its preinjectivity and is also equivalent to the fact that $\tau(A^G)$ has maximal topological entropy. Note that it immediately follows from the characterization of preinjectivity by the absence of a pair of distinct mutually erasable patterns (see, e.g., [Ceccherini-Silberstein and Coornaert 2010, Proposition 5.5.2]) that τ is preinjective if and only if it is $(*)$-preinjective. Observe also that τ is always $(**)$-preinjective. In the case when G is amenable with a Følner net $\mathcal{F}$ then τ satisfies $\operatorname{mdim}_{\mathcal{F}}(\tau(A^G)) = \dim(X) = 0$ since

$$\dim(A^\Omega) = 0$$

for every finite subset $\Omega \subset G$. The variety X is irreducible if and only if A is a singleton. Otherwise, there exist cellular automata $\tau : A^G \to A^G$ that are not surjective (e.g., the map $\tau : A^G \to A^G$ defined by $\tau(c) = c_0$ for all $c \in A^G$, where $c_0 \in A^G$ is some constant configuration). Such a cellular automaton is $(**)$-preinjective but not $(*)$-preinjective.

The next example shows that we cannot replace the hypothesis that X is irreducible by the weaker hypothesis that it is connected in Theorem 1.4, Theorem 5.4, and Proposition 6.4.

Example 8.2. Let G be an amenable group and let $\mathcal{F}$ be a Følner net for G. Let K be an algebraically closed field. Consider the projective curve X in $\mathbb{P}^2_K$ defined by

$$X := \operatorname{Proj}(K[u, v, w]/(uv)) \subset \mathbb{P}^2_K.$$

Then $X = L_u \cup L_v \subset \mathbb{P}^2_K$ is the union of the two projective coordinate lines

$$L_u := \{u = 0\}, \quad \text{and} \quad L_v := \{v = 0\} \subset \mathbb{P}^2_K.$$

Since X has two irreducible components L_u and L_v, it is not irreducible. However, X is clearly connected. In the principal affine chart $\mathbb{A}^2_K = D_+(w) = \{w \neq 0\} \subset \mathbb{P}^2_K$, we see that X is given by

$$Y = \operatorname{Spec}(K[x, y]/(xy)) = I_x \cup I_y \subset \mathbb{A}^2_K,$$

where $x = u/w$, $y = v/w$, $I_x = \{x = 0\}$ and $I_y = \{y = 0\}$. Let $h : Y \to Y$ be the contraction morphism induced by the morphism of K-algebras:

$$K[x, y]/(xy) \to K[x, y]/(xy), \qquad (x, y) \mapsto (x, 0).$$

It is clear that $h(Y) = h(I_y) = I_y = \operatorname{Spec}(K[x, y]/(y)) \simeq \mathbb{A}^1_K$. By, for example, the valuative criteria of properness (see [Hartshorne 1977, Theorem II.4.7]), there is a K-scheme morphism $f : X \to X$ extending h and that

$$f((0:1:0)) = f((1:0:0)) = (1:0:0) \in L_v.$$

Hence, $f(X) = f(L_v) = L_v$ and thus $\dim(f(X)) = \dim(L_v) = 1$. Clearly f is not surjective and thus $f^{(K)}$ is not surjective either.

Now let $A := X(K)$ and let $\tau : A^G \to A^G$ be the cellular automaton over (G, X, K) with memory set $M = \{1_G\}$ and associated local defining map $f^{(K)} : A \to A$. Observe that τ is not preinjective since f is not injective and $M = \{1_G\}$. Also τ is not $(*)$-preinjective since $f(X) = f(L_v) = L_v$. On the other hand, $\mathrm{mdim}_{\mathcal{F}}(\tau(A^G)) = 1 = \dim(X)$ and τ is $(**)$-preinjective but not surjective.

The following example shows that Theorems 1.4 and 5.4 do not hold in general for irreducible noncomplete algebraic varieties.

Example 8.3. Let G be an amenable group and let $\mathcal{F}$ be a Følner net for G. Let K be an algebraically closed field. Let X be an irreducible algebraic variety over K and let $A := X(K)$. Suppose that $f : X \to X$ is a nonsurjective dominant K-scheme morphism. Observe that f is not injective by the Ax–Grothendieck theorem. Let $\tau : A^G \to A^G$ be the algebraic cellular automaton over (G, X, K) with memory set $M = \{1_G\}$ and associated local defining map $f^{(K)} : A \to A$.

Since f is dominant, Chevalley's theorem and Proposition 2.9(vi) imply that

$$\dim(f(X)) = \dim(X).$$

As K is algebraically closed, we have that

$$\dim(f(A)) = \dim(f(X)) = \dim(X).$$

We deduce that $\mathrm{mdim}_{\mathcal{F}}(\tau(A^G)) = \dim(X)$. It is clear that τ is both $(*)$- and $(**)$-preinjective. However, since f is not surjective, $f^{(K)}$ is not surjective either (see, e.g., [Ceccherini-Silberstein et al. 2019, Lemma A.22(iv)]). Hence, τ is not surjective. Note also that $f^{(K)}$ is not injective since f is not (see, e.g., [Ceccherini-Silberstein et al. 2019, Lemma A.22(iii)]). It follows that τ is not preinjective.

Here is a class of such couples (X, f). Let $X = \mathbb{A}^2 = \mathrm{Spec}(K[x, y])$ be the affine plane over K and consider the morphism $f : \mathbb{A}^2 \to \mathbb{A}^2$ given by the morphism of K-algebras

$$K[x, y] \to K[x, y], \quad (x, y) \mapsto (x^r, x^s P(y)),$$

where $r, s \geq 1$ and $P \in K[y]$ is a nonconstant polynomial in y. It is clear that $f(\mathbb{A}^2) = \mathbb{A}^2 \setminus (\{x = 0\} \setminus \{(0, 0)\})$. Hence f is indeed a nonsurjective dominant K-scheme morphism. This construction can be easily generalized to higher-dimensional affine spaces $\mathbb{A}^n$ for $n \geq 2$ by using, for example, the morphisms of K-algebras given by

$$K[x_1, \dots, x_n] \to K[x_1, \dots, x_n],$$
$$(x_1, \dots, x_n) \mapsto (x_1^{r_1}, x_1^{r_2} P_2(x_2), \dots, x_1^{r_n} P_n(x_n)),$$

where $r_1, \dots, r_n \in \mathbb{N}^*$ and $P_2, \dots, P_n$ are nonconstant polynomials in $x_2, \dots, x_n$, respectively.

We give now an example with nontrivial minimal memory set showing that we cannot omit the hypothesis that X is complete in Theorem 5.4.

Example 8.4. In this example, we take $G := \mathbb{Z}$. Thus G is amenable. Let $X := \mathbb{A}_K^1 = \operatorname{Spec}(K[t])$ be the affine line over an algebraically closed field K and let $A := X(K) = K$. Let $\tau : A^G \to A^G$ be the cellular automaton over (G, X, K) with memory set $M = \{0, 1\} \subset \mathbb{Z}$ and associated local defining map given by $f^{(K)} : X(K)^K \to X(K)$, where $f : X^M \to X$ is the K-scheme morphism induced by the morphism of K-algebras

$$K[t] \to K[x, y], \quad t \mapsto xy.$$

Clearly $\tau : K^{\mathbb{Z}} \to K^{\mathbb{Z}}$ is given by the formula

$$\tau(c)(n) = c(n)c(n+1) \quad \text{for all } c \in K^{\mathbb{Z}} \text{ and } n \in \mathbb{Z}.$$

Consider the configuration $d \in K^{\mathbb{Z}}$ such that $d(-1) = d(1) = 1$ and $d(n) = 0$ if $n \in \mathbb{Z} \setminus \{-1, 1\}$. If there were some configuration $c \in \mathbb{K}^{\mathbb{Z}}$ such that $\tau(c) = d$, then we would have $c(0)c(1) = d(0) = 0$. This would imply $c(0) = 0$ or $c(1) = 0$ and hence $d(-1) = 0$ or $d(1) = 0$, which is a contradiction. We deduce that d has no preimage under τ. Thus τ is not surjective.

We claim that $\operatorname{mdim}(\tau(K^{\mathbb{Z}})) = 1 = \dim(X)$. Indeed, let $\Gamma := \tau(K^{\mathbb{Z}})$. For each $m \in \mathbb{N}$, let $F_m := [-m, m] \cap \mathbb{Z} \subset \mathbb{Z}$. Then $\mathcal{F} := (F_m)_m$ is a Følner sequence for $\mathbb{Z}$. Note that $F_m^+ = [-m, m+1] \cap \mathbb{Z}$. Consider the K-scheme morphism (see Lemma 3.2)

$$f_{F_m}^+ : X^{F_m^+} = \mathbb{A}^{2m+2} \to X^{F_m} = \mathbb{A}^{2m+1}.$$

Then $\tau_{F_m}^+ = f_{F_m}^{+(K)}$. It is immediate that

$$\tau_{F_m}^+(c_{-m}, \dots, c_{m+1}) = (c_{-m}c_{-m+1}, \dots, c_m c_{m+1}).$$

We deduce that the image of $f_{F_m}^+$ contains $\mathbb{A}^{2m+1} \setminus L$ where

$$L = V((x_{-m} \cdots x_m)) \subset \mathbb{A}^{2m+1} = \operatorname{Spec}(K[x_{-m}, \dots, x_m])$$

is the union of the $2m+1$ coordinate hyperplanes given by the equation $x_{-m} \cdots x_m = 0$. Therefore, $\dim(\operatorname{Im}(f_{F_m}^+)) = \dim(\mathbb{A}^{2m+1})$ and thus $\dim(\Gamma_{F_m}) = |F_m| = 2m+1$ for all $m \in \mathbb{N}$. Hence, we conclude that

$$\mathrm{mdim}_{\mathcal{F}}(\Gamma) = \limsup_{m} \frac{\dim(\Gamma_{F_m})}{|F_m|} = 1 = \dim(A)$$

as claimed. Since d is almost equal to the configuration $0 \in K^{\mathbb{Z}}$ and $\tau(d) = \tau(0) = 0$, we see that τ is not preinjective. It follows from Proposition 6.6 that τ is $(*)$-preinjective.

The following example shows that the hypothesis that G is amenable cannot be omitted in Theorem 1.1.

Example 8.5. Let $G = F_2$ be the free group of rank 2 based on the generators a, b. We recall that G is residually finite but not amenable. Let $M := \{a, b, a^{-1}, b^{-1}\} \subset F_2$. Consider an abelian variety $Y = (Y, +)$ over an algebraically closed field K with identity element $e \in Y(K)$. We suppose that Y is nontrivial, so that $\dim(Y) \geq 1$. The K-fibered product $X := Y \times_K Y$ is also a nontrivial abelian variety over K. The set $A := X(K) = Y(K) \times Y(K)$ of K-points of X is a nontrivial abelian group. For $i = 1, 2$ let $q^i : Y^M \times_K Y^M \to Y^M$ be the i-th projection and for $g \in M$, let $q_g : Y^M \to Y$ be the projection on the g-factor. For $i = 1, 2$ and $g \in M$, let $p_g^i : X^M = Y^M \times_K Y^M \to Y$ be the projection defined by $p_g^i := q_g \circ q^i$. Let $h : X^M \to Y$ be the morphism defined by

$$h := p_a^1 + p_{a^{-1}}^1 + p_b^2 + p_{b^{-1}}^2.$$

Let

$$\iota := (\mathrm{Id}_Y, e) : Y = Y \times_K \mathrm{Spec}(K) \to X = Y \times_K Y.$$

Finally, we define $f := \iota \circ h : X^M \to X$.

Let $\tau : A^G \to A^G$ be the algebraic cellular automaton over (G, X, K) with memory set M and associated local defining map $\mu = f^{(K)} : A^M \to A$. Observe that for all $c \in A^G$,

$$\tau(c)(g) = (\pi_1(c(ga)) + \pi_1(c(ga^{-1})) + \pi_2(c(gb)) + \pi_2(c(gb^{-1})), e),$$

where $\pi_i : A \to A$ is given by $\pi_i(u_1, u_2) = (u_i, e)$ for $i = 1, 2$ and $(u_1, u_2) \in Y(K) \times Y(K) = A$. By [Ceccherini-Silberstein and Coornaert 2010, Proposition 5.11], we see that τ is preinjective but not surjective.

9. Questions

Question 1. Can we remove the hypothesis that X is complete (respectively, irreducible) in Theorem 1.1?

Question 2. Does Theorem 1.3 still hold without the assumption that X is irreducible?

Question 3. Does Theorem 7.2 remain valid without the amenability hypothesis on the group G?

Question 4. Does Theorem 1.1 characterize amenable groups?

References

[An 2012] J. An, "Rigid geometric structures, isometric actions, and algebraic quotients", *Geom. Dedicata* **157** (2012), 153–185. MR Zbl

[Ceccherini-Silberstein and Coornaert 2006] T. Ceccherini-Silberstein and M. Coornaert, "The Garden of Eden theorem for linear cellular automata", *Ergodic Theory Dynam. Systems* **26**:1 (2006), 53–68. MR Zbl

[Ceccherini-Silberstein and Coornaert 2010] T. Ceccherini-Silberstein and M. Coornaert, *Cellular automata and groups*, Springer, 2010. MR Zbl

[Ceccherini-Silberstein et al. 1999] T. G. Ceccherini-Silberstein, A. Machì, and F. Scarabotti, "Amenable groups and cellular automata", *Ann. Inst. Fourier* (*Grenoble*) **49**:2 (1999), 673–685. MR Zbl

[Ceccherini-Silberstein et al. 2019] T. Ceccherini-Silberstein, M. Coornaert, and X. K. Phung, "On injective endomorphisms of symbolic schemes", *Comm. Algebra* **47**:11 (2019), 4824–4852. MR Zbl

[EGA I 1971] A. Grothendieck and J. A. Dieudonné, *Eléments de géométrie algébrique, I*, Grundlehren der Mathematischen Wissenschaften **166**, Springer, 1971. MR Zbl

[EGA II 1961] A. Grothendieck, "Éléments de géométrie algébrique, II: Étude globale élémentaire de quelques classes de morphismes", *Inst. Hautes Études Sci. Publ. Math.* **8** (1961), 5–222. MR Zbl

[EGA IV_1 1964] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, I", *Inst. Hautes Études Sci. Publ. Math.* **20** (1964), 5–259. MR Zbl

[EGA IV_3 1966] A. Grothendieck, "Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* **28** (1966), 5–255. MR Zbl

[Görtz and Wedhorn 2010] U. Görtz and T. Wedhorn, *Algebraic geometry*, vol. I: Schemes, Vieweg + Teubner, Wiesbaden, 2010. MR Zbl

[Gromov 1999] M. Gromov, "Endomorphisms of symbolic algebraic varieties", *J. Eur. Math. Soc.* (*JEMS*) **1**:2 (1999), 109–197. MR Zbl

[Harris 1992] J. Harris, *Algebraic geometry: a first course*, Graduate Texts in Mathematics **133**, Springer, 1992. MR Zbl

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, 1977. MR Zbl

[Liu 2002] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Graduate Texts in Mathematics **6**, Oxford University Press, 2002. MR Zbl

[Moore 1962] E. F. Moore, "Machine models of self-reproduction", pp. 17–33 in *Mathematical problems in the biological sciences*, edited by R. Bellman, Proc. Sympos. Appl. Math. **XIV**, American Mathematical Society, Providence, RI, 1962. Zbl

[Myhill 1963] J. Myhill, "The converse of Moore's Garden-of-Eden theorem", *Proc. Amer. Math. Soc.* **14** (1963), 685–686. MR Zbl

[Vakil] R. Vakil, "Math 216: foundations of algebraic geometry", lecture notes, available at https://math216.wordpress.com/.

Received March 29, 2018.

TULLIO CECCHERINI-SILBERSTEIN
DIPARTIMENTO DI INGEGNERIA
UNIVERSITÀ DEL SANNIO
BENEVENTO
ITALY
tullio.cs@sbai.uniroma1.it

MICHEL COORNAERT
CNRS, IRMS UMR 7501
UNIVERSITÉ DE STRASBOURG
STRASBOURG
FRANCE
michel.coornaert@math.unistra.fr

XUAN KIEN PHUNG
CNRS, IRMA UMR 7501
UNIVERSITÉ DE STRASBOURG
STRASBOURG
FRANCE
phung@math.unistra.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.67

BERGMAN KERNELS OF ELEMENTARY REINHARDT DOMAINS

DEBRAJ CHAKRABARTI, AUSTIN KONKEL,
MEERA MAINKAR AND EVAN MILLER

We study the Bergman kernel of certain domains in $\mathbb{C}^n$, called elementary Reinhardt domains, generalizing the classical Hartogs triangle. For some elementary Reinhardt domains, we explicitly compute the kernel, which is a rational function of the coordinates. For some other such domains, we show that the kernel is not a rational function. For a general elementary Reinhardt domain, we obtain a representation of the kernel as an infinite series.

1. Introduction

1A. ***Elementary Reinhardt domains.*** Let $\mathbb{D}^n = \{z \in \mathbb{C}^n \mid |z_j| < 1 \text{ for } 1 \le j \le n\}$ denote the unit polydisc in $\mathbb{C}^n$, $n \ge 2$, and let $k = (k_1, \ldots, k_n) \in \mathbb{Z}^n$ be a multi-index. The goal of this paper is the study of the Bergman kernel of the domain

$$\mathscr{H}(k) = \left\{z \in \mathbb{D}^n \mid z^k \text{ is defined, and } |z^k| < 1\right\}, \tag{1-1}$$

where we use the standard multi-index convention $z^k = z_1^{k_1} z_2^{k_2} \cdots z_n^{k_n}$, and the only way this can fail to be defined is if its evaluation involves division by zero. We will call the domain $\mathscr{H}(k)$ the *elementary Reinhardt domain* associated to the multi-index k (cf. [Jarnicki and Pflug 2008, pp. 33ff.], where this terminology is used, with a slightly different definition). A famous example of such a domain is the *Hartogs triangle*

$$\mathscr{H}(1, -1) = \{|z_1| < |z_2| < 1\} \subset \mathbb{C}^2,$$

a well-known source of counterexamples in several complex variables (see, e.g., [Shaw 2015]).

Debraj Chakrabarti, Austin Konkel and Evan Miller were partially supported by NSF grant DMS-1600371. Austin Konkel was also supported by a Student Research and Creative Endeavors grant from Central Michigan University.

MSC2010: primary 32A25; secondary 32A07.

Keywords: Bergman kernel.

It is easy to see that $\mathscr{H}(k)$ is logarithmically convex, and therefore pseudoconvex (see [Range 1986]). If the multi-index k contains both positive and negative entries, then $\mathscr{H}(k)$ is a Reinhardt domain with the origin as a boundary point, so it follows (see [Chakrabarti 2019]) that each holomorphic function smooth up to the boundary on $\mathscr{H}(k)$ extends to a larger, fixed domain, a property which is classical in the special case of the Hartogs triangle (see [Sibony 1975; Behnke 1933]). Therefore, $\mathscr{H}(k)$ does not have a basis of Stein neighborhoods, and is not a so-called $\mathcal{H}^\infty$-domain of holomorphy. This makes domains such as $\mathscr{H}(k)$ particularly interesting from the point of view of function theory on nonsmooth domains, since each smoothly bounded pseudoconvex domain is in fact an $\mathcal{H}^\infty$-domain of holomorphy [Catlin 1980; Hakim and Sibony 1980].

Recently, the unusual L^p-mapping properties of the Bergman projection on the *generalized Hartogs triangle* $\mathscr{H}(m, -n) \subset \mathbb{C}^2$ (where m, n are coprime positive integers) have received the attention of several authors [Chakrabarti and Zeytuncu 2016; Edholm 2016; Edholm and McNeal 2016; 2017; Chakrabarti et al. 2019]. In many of these investigations, the explicit form of the Bergman kernel of $\mathscr{H}(m, -n) \subset \mathbb{C}^2$ plays a crucial role. The elementary Reinhardt domains are a natural class generalizing the Hartogs triangle. Motivated by this, in this paper, we make a preliminary study of the Bergman kernels of the domains $\mathscr{H}(k)$. In particular, we investigate whether such a Bergman kernel is a rational function of the coordinates, as it indeed is if $n = 2$ (see [Edholm 2016; Edholm and McNeal 2017]). Other recent attempts at higher dimensional generalizations may be found in [Park 2018; Chen 2017; Huo 2018; Chen et al. 2020]. For planar domains, the rationality (or algebraicity) of the Bergman kernel has important function-theoretic repercussions (see [Bell 2005]). It would be interesting to see whether something similar is true for the elementary Reinhardt domains.

From now on we will assume that the multi-index $k = (k_1, \dots, k_n)$ defining the domain (1-1) has the following properties:

(1) At least one of the components of the multi-index is positive, at least one of the components is negative, and no component is zero.

    We will call the number of positive components of k, the *signature* s of the elementary Reinhardt domain $\mathscr{H}(k)$.

(2) If $\mathscr{H}(k)$ has signature s, after renaming the coordinates, we will assume without loss of generality that $k_j > 0$ for $1 \le j \le s$ and $k_j < 0$ if $s+1 \le j \le n$.

(3) We will also assume without loss of generality that the numbers $k_1, \dots, k_n$ are relatively prime.

1B. ***Explicit formula.*** For elementary Reinhardt domains of signature 1, we now give an explicit formula for the Bergman kernel as a rational function of the coordinates.

To state the result, introduce the following notation. For integers λ and μ, let

$$\mathsf{D}_\lambda(\mu) = \begin{cases} 0, & \mu \le -1 \text{ or } \mu \ge 2\lambda - 1, \\ \mu + 1, & 0 \le \mu \le \lambda - 1, \\ 2\lambda - 1 - \mu, & \lambda \le \mu \le 2\lambda - 2. \end{cases} \tag{1-2}$$

It will be seen in Section 4B below, that the seemingly complicated expression $\mathsf{D}_\lambda(\mu)$ arises as the number of solutions in pairs of integers (x, y) of the equation $x + y = \mu$ subject to the constraints $0 \le x, y \le \lambda - 1$ (see (4-23), (4-24), (4-25)).

Theorem 1.1. *Let $n \ge 2$, let $k_1, \dots, k_n$ be relatively prime positive integers, and let*

$$k = (k_1, -k_2, \dots, -k_n) \in \mathbb{Z}^n$$

be a multi-index. The Bergman kernel of the elementary Reinhardt domain $\mathscr{H}(k)$ is given by

$$\mathbb{B}_{\mathscr{H}(k)}(z, w) = \frac{1}{\pi^n L} \cdot \frac{\sum_{\beta \in \mathfrak{G}} C(\beta) t^\beta}{\left(\prod_{b=2}^n t_b^{k_b} - t_1^{k_1}\right)^2 \cdot \prod_{b=2}^n (1 - t_b)^2}, \tag{1-3}$$

where $t = (t_1, \dots, t_n) \in \mathbb{C}^n$ with $t_a = z_a \overline{w}_a$ for $1 \le a \le n$, and

$$C(\beta) = \mathsf{D}_K\big(2K - \ell_1(\beta_1 + 1) - 1\big) \cdot \prod_{b=2}^n \mathsf{D}_{\ell_b}\big(\ell_b(\beta_b + 1) + \ell_1(\beta_1 + 1) - 2K - 1\big), \tag{1-4}$$

where the function $\mathsf{D}_*(\cdot)$ is defined in (1-2) above, with

$$K = \operatorname{lcm}(k_1, \dots, k_n), \qquad \ell_a = \frac{K}{k_a} \quad \textit{for } 1 \le a \le n, \quad \textit{and} \quad L = \prod_{a=1}^n \ell_a,$$

and where the finite collection of multi-indices $\mathfrak{G} \subset \mathbb{Z}^n$ is defined by

$$\mathfrak{G} = \big\{(\beta_1, \dots, \beta_n) \in \mathbb{Z}^n \mid 0 \le \beta_1 \le 2k_1 - 2, \textit{ and } 0 \le \beta_b \le 2k_b \textit{ for each } 2 \le b \le n\big\}. \tag{1-5}$$

While the expression (1-3) is somewhat complicated, it generalizes and extends several known results in the literature. In the special case of the classical Hartogs triangle $\mathscr{H}(1, -1)$, an explicit expression for the Bergman kernel is already found in [Bremermann 1955]. Recently, Edholm [2016] computed using Bell's formula (see (4-2) below) the Bergman kernels of $\mathscr{H}(1, -k)$ and $\mathscr{H}(k, -1)$, where $k \ge 2$ is an integer ("fat and thin generalized Hartogs triangles"), a computational *tour de force* which inspired Theorem 1.1. Edholm and McNeal [2017] studied $\mathscr{H}(m, -n) \subset \mathbb{C}^2$, where m, n are coprime positive integers, and expressed its Bergman kernel as the sum of m "sub-Bergman kernels." The sub-Bergman kernels are obtained by summing subseries of the power series (2-2) representing the Bergman kernel of a Reinhardt domain. These subseries consist of terms with monomials whose exponents are represented by straight lines of different slopes in the lattice point diagram of monomials, resulting in a decomposition of the kernel into convenient pieces, which permits the explicit summation of each of the subkernels in closed form as a rational function, and determination of the L^p-regularity of each piece. However, our formula (1-3) shows that splitting the kernel into the subkernels is unnecessary, and the main L^p estimate of [Edholm and McNeal 2017] could proceed directly from (1-3). Starting from (1-3), we recapture below in Section 4C the special cases considered in [Edholm 2016]. Theorem 1.1 also opens the way to generalize the interesting recent results related to L^p-regularity of the Bergman projection, duality of Bergman spaces etc. (see [Edholm and McNeal 2017; Chakrabarti and Zeytuncu 2016; Chakrabarti et al. 2019]) to higher dimensions.

1C. *Signatures greater than 1.* In signatures $s \geq 2$ (so that the ambient dimension $n \geq 3$) the situation is much less clear, and is worth further study. Here we collect a few observations which seems to indicate that there are some fundamental differences between the cases $s = 1$ and $s \geq 2$. In particular, it seems plausible that the Bergman kernels of elementary Reinhardt domains of signature $s \geq 2$ can not be represented using a simple rational function such as (1-3).

Let $n \geq 2$, and let $1 \leq s \leq n-1$. We denote by $\Omega_{n,s}$ the elementary Reinhardt domain of signature s in $\mathbb{C}^n$, where each component of the defining multi-index is ± 1, i.e.,

$$\Omega_{n,s} = \mathscr{H}(\underbrace{1, \ldots, 1}_{s}, \underbrace{-1, \ldots, -1}_{n-s}), \tag{1-6}$$

so that $\Omega_{n,s} = \{z \in \mathbb{D}^n : |z_1 \cdots z_s| < |z_{s+1} z_{s+2} \cdots z_n|\}$. We will call $\Omega_{n,s}$ the *model elementary domain* of signature s. It is shown in Proposition 2.1 below that the model elementary domains are branched covers of all elementary Reinhardt domains.

In Theorem 3.1 below, we give an account of the coefficients of the power series expansion of $\mathbb{B}_{\Omega_{n,s}}$ by computing the L^2-norms of monomials $e_\alpha(z) = z_1^{\alpha_1} \cdots z_n^{\alpha_n}$. This shows that the coefficient of

$$(z_1 \overline{w}_1)^{\alpha_1} \cdots (z_n \overline{w}_n)^{\alpha_n}$$

in the power series expansion of $\mathbb{B}_{\Omega_{n,s}}$ is a polynomial in $\alpha_1, \ldots, \alpha_n$ only if $s = 1$. For $2 \leq s \leq n-1$, the coefficient is a *rational* function of $\alpha_1, \ldots, \alpha_n$. From this we are able to deduce the following:

Theorem 1.2. *If $n \geq 3$, then $\mathbb{B}_{\Omega_{n,n-1}}$ is not a rational function.*

It seems highly plausible that in fact $\mathbb{B}_{\Omega_{n,s}}$ is a transcendental function of the coordinates unless $s = 1$, though at present we are not in possession of a complete proof. If this conjecture is correct, using the proper map from a model domain to an arbitrary elementary Reinhardt domain, it will follow that the Bergman kernel of an elementary Reinhardt domain of signature $s \geq 2$ is transcendental.

Some further properties of the series representation of the Bergman kernel are explored in Section 3B below.

2. Preliminaries

2A. ***Bergman theory.*** We briefly recall some basic facts about Bergman spaces and kernels and clarify our notation. An extensive modern exposition of this topic from the complex analysis point of view is [Krantz 2013], and from the operator theory point of view is [Duren and Schuster 2004].

Let $\Omega \subset \mathbb{C}^n$ be a domain, i.e., a connected open set. Then $A^2(\Omega)$, the (L^2)-Bergman space of Ω, is the Hilbert space of holomorphic functions which are square integrable with respect to the Lebesgue measure dV. This is a so-called *reproducing kernel Hilbert space*, and its reproducing kernel is the *Bergman kernel*, a function $\mathbb{B}_\Omega : \Omega \times \Omega \to \mathbb{C}$, holomorphic in the first and antiholomorphic in the second input such that for each $f \in A^2(\Omega)$ we have for each $z \in \Omega$ the *reproducing property*:

$$f(z) = \int_\Omega f(w)\mathbb{B}_\Omega(z, w)\, dV(w).$$

A domain $\Omega \subset \mathbb{C}^n$ is *Reinhardt* if whenever $z \in \Omega$, and $\lambda \in \mathbb{T}^n$, where

$$\mathbb{T}^n = \left\{\lambda \in \mathbb{C}^n \mid |\lambda_j| = 1 \text{ for each } 1 \leq j \leq n\right\}$$

is the unit torus, we have $(\lambda_1 z_1, \dots, \lambda_n z_n) \in \Omega$. For a Reinhardt domain, there is a canonical series representation of the Bergman kernel. For each multi-index $\alpha \in \mathbb{Z}^n$, let e_α denote the monomial

$$e_\alpha(z) = z^\alpha = z_1^{\alpha_1} \cdots z_n^{\alpha_n}. \tag{2-1}$$

Then the Bergman kernel of Ω has the series representation converging uniformly on compact subsets of $\Omega \times \Omega$:

$$\mathbb{B}_\Omega(z, w) = \sum_{\alpha \in \mathbb{Z}^n} \frac{1}{\|e_\alpha\|^2} z^\alpha \overline{w^\alpha}, \tag{2-2}$$

where

$$\|e_\alpha\|^2 = \int_\Omega |e_\alpha(z)|^2\, dV(z), \tag{2-3}$$

and if for an $\alpha \in \mathbb{Z}^n$ the integral (2-3) diverges, the coefficient $1/\|e_\alpha\|^2$ in (2-2) is taken to be zero. An immediate consequence of this series representation is the following simple observation: if $\widetilde{\Omega} \subset \mathbb{C}^n$ is the domain

$$\widetilde{\Omega} = \{(z_1\overline{w}_1, \dots, z_n\overline{w}_n) \mid z, w \in \Omega\},$$

then there is a holomorphic function $\widetilde{B}$ on $\widetilde{\Omega}$ such that $\mathbb{B}_\Omega(z,w)=\widetilde{B}(z_1\overline{w}_1,\dots,z_n\overline{w}_n)$, where for $t\in\widetilde{\Omega}$,

$$\widetilde{B}(t)=\sum_{\alpha\in\mathbb{Z}^n}\frac{1}{\|e_\alpha\|^2}t^\alpha. \tag{2-4}$$

Therefore, the Bergman kernel of a Reinhardt domain Ω can be thought of as a holomorphic function on a different domain $\widetilde{\Omega}$, and this simplifies its study.

2B. ***Model domains as branched covers.*** A map $\phi:\mathbb{C}^n\to\mathbb{C}^n$ will be said to be of the *diagonal type* if there are positive integers $\ell_1,\dots,\ell_n$ such that

$$\phi(z_1,\dots,z_n)=\left(z_1^{\ell_1},\dots,z_n^{\ell_n}\right). \tag{2-5}$$

Proposition 2.1. *Let $n\geq 2$ and let H be an elementary Reinhardt domain in $\mathbb{C}^n$ of signature $1\leq s\leq n-1$. Then there is a proper holomorphic map of diagonal type from the model elementary domain $\Omega_{n,s}$ of (1-6) to H.*

Proof. Let $k=(k_1,\dots,k_s,-k_{s+1},\dots,-k_n)$ be the multi-index such that $H=\mathscr{H}(k)$. Let us set $K=\operatorname{lcm}(k_1,\dots,k_n)$, and let $\ell_j=K/k_j$. Define the map ϕ by (2-5). Then ϕ defines a proper holomorphic map from $\mathbb{C}^n$ to itself. To show that ϕ restricts to a proper map from $\Omega_{n,s}$ to H, it suffices to show that $\phi^{-1}(H)=\Omega_{n,s}$. Indeed, if $z\in\mathbb{C}^n$ is such that $\phi(z)\in H$, then we have $|\phi(z)^k|<1$. But since

$$\phi(z)^k=(z_1^{\ell_1})^{k_1}\cdots(z_s^{\ell_s})^{k_s}(z_{s+1}^{\ell_{s+1}})^{-k_{s+1}}\cdots(z_n^{\ell_n})^{-k_n}=(z_1\cdots z_s)^K(z_{s+1}\cdots z_n)^{-K},$$

it follows that $z\in\Omega_{n,s}$ and the result follows. □

Definition 2.2. Let H be an elementary Reinhardt domain in $\mathbb{C}^n$ of signature s. The map $\phi:\Omega_{n,s}\to H$ given by (2-5) will be referred to as the *standard proper map* associated with H.

Note that there may also be proper holomorphic maps from $\Omega_{n,s}$ to H different from the standard map. And for certain elementary Reinhardt domains, biholomorphic maps can even be found. For example, the map from $\Omega_{2,1}=\{|z_1|<|z_2|<1\}\subset\mathbb{C}^2$ to $\mathscr{H}(m,-n)=\{|z_1|^{m/n}<|z_2|<1\}\subset\mathbb{C}^2$ given by $(z_1,z_2)\to(z_1z_2^{n-1},z_2^m)$ is a proper holomorphic map different from the standard map, and a biholomorphism if and only if $m=1$.

3. Norms of monomials

In the following theorem, we describe the coefficients of the series expansion (2-2) of the Bergman kernel of an elementary Reinhardt domain.

Theorem 3.1. *Let $n\geq 2$, let $1\leq s\leq n-1$, and let $\alpha\in\mathbb{Z}^n$. Let $\beta\in\mathbb{Z}^n$ be the multi-index $(\beta_1,\dots,\beta_n)$ such that*

$$\beta_j=\alpha_j+1.$$

Then, on the model domain $\Omega_{n,s}$, we have

(1) $\|e_\alpha\|^2_{\Omega_{n,s}} < \infty$ *if and only if*

$$\beta_j > 0 \quad \text{and} \quad \beta_j + \beta_\ell > 0 \quad \text{for } 1 \le j \le s, \text{ and } s+1 \le \ell \le n. \tag{3-1}$$

(2) *If α is such that $\|e_\alpha\|^2_{\Omega_{n,s}} < \infty$, we have that*

$$\|e_\alpha\|^2_{\Omega_{n,s}} = \pi^n \cdot \frac{R_{n,s}(\beta)}{S_{n,s}(\beta)}, \tag{3-2}$$

where R, S are homogeneous polynomials in n variables with integer coefficients, with

$$S_{n,s}(\beta) = \prod_{j=1}^{s} \beta_j \prod_{\substack{1 \le j \le s \\ s+1 \le \ell \le n}} (\beta_j + \beta_\ell), \tag{3-3}$$

and $R_{n,s}$ is a homogeneous polynomial of total degree $(n-s)(s-1)$ such that $R_{n,s}$ and $S_{n,s}$ have no common factors. We further have $R_{n,1} = 1$.

Recall that the total degree of a monomial is the sum of exponents of each of the variables.

Proof. Fix an $s \ge 1$, and we prove this by induction on n. We will start with the base case of $n = s$, which is not part of the statement of the theorem as stated, but for which the result also holds. Denote by $\mathbb{D}^s$ the unit polydisc $\{|z_j| < 1 : j = 1, \dots, s\}$ in $\mathbb{C}^s$. Notice that

$$\Omega_{s,s} = \{z \in \mathbb{D}^s \mid |z_1 z_2 \cdots z_n| < 1\} = \mathbb{D}^s.$$

In this case we have for $\alpha \in \mathbb{Z}^s$ by direct computation that $\|e_\alpha\|^2 < \infty$ if and only if

$$\beta_j = \alpha_j + 1 > 0, \quad j = 1, \dots, s,$$

and for such α

$$\|e_\alpha\|^2_{\mathbb{D}^s} = \pi^s \frac{1}{(\alpha_1+1)\cdots(\alpha_s+1)} = \pi^s \frac{1}{\beta_1 \cdots \beta_s}. \tag{3-4}$$

Therefore (3-1) is satisfied, and if we take $R_{s,s} = 1$ and $S_{s,s} = \beta_1 \cdots \beta_s$ then (3-3) is satisfied, and $R_{s,s}$ does have degree $(n-s)(s-1) = (s-s)(s-1) = 0$, as needed.

We now proceed by induction. Assume the result is true for some $n \ge s$. For simplicity of notation, let $\mathbf{1} = (1, 1, \dots, 1)$, then we set

$$\mathcal{D}_{n,s}(\beta) = \frac{1}{\pi^n} \|e_{\beta - \mathbf{1}}\|^2_{\Omega_{n,s}}, \tag{3-5}$$

and for $\beta_{n+1} \in \mathbb{Z}$ denote by $(\beta, \beta_{n+1}) \in \mathbb{Z}^{n+1}$ the multi-index

$$(\beta, \beta_{n+1}) = (\beta_1, \dots, \beta_n, \beta_{n+1}).$$

To abbreviate the formulas that follow, let $\beta^* \in \mathbb{Z}^n$ be the multi-index given by

$$\beta_j^* = \begin{cases} \beta_j + \beta_{n+1} & \text{if } 1 \le j \le s, \\ \beta_j - \beta_{n+1} & \text{if } s+1 \le j \le n. \end{cases} \tag{3-6}$$

Notice that β^* actually depends on $\beta \in \mathbb{Z}^n$ and $\beta_{n+1} \in \mathbb{Z}$, though this has been suppressed from the notation. We claim that $\mathcal{D}_{n+1,k}(\beta, \beta_{n+1})$ can be represented as follows

$$\mathcal{D}_{n+1,s}(\beta, \beta_{n+1}) = \frac{1}{\beta_{n+1}}(\mathcal{D}_{n,s}(\beta) - \mathcal{D}_{n,s}(\beta^*)). \tag{3-7}$$

We postpone the proof of the claim to proceed with the induction. Note that $\|e_{\beta-\mathbf{1}}\| < \infty$ is equivalent to $\mathcal{D}_{n,s}(\beta) < \infty$. Let $(\beta, \beta_{n+1}) \in \mathbb{Z}^n$. From (3-7), it follows that $\mathcal{D}_{n+1,s}(\beta, \beta_{n+1}) < \infty$ if and only if $\mathcal{D}_{n,s}(\beta) < \infty$ and $\mathcal{D}_{n,s}(\beta^*) < \infty$, since each of $\mathcal{D}_{n,s}(\beta)$ and $\mathcal{D}_{n,s}(\beta^*)$ is strictly positive. From $\mathcal{D}_{n,s}(\beta) < \infty$, using the induction hypothesis, we see that the conditions (3-1) hold. From $\mathcal{D}_{n,s}(\beta^*) < \infty$ we get the conditions

$$\beta_j^* > 0 \quad \text{and} \quad \beta_j^* + \beta_\ell^* > 0 \quad \text{for } 1 \le j \le s, \text{ and } s+1 \le \ell \le n,$$

which, using the definition of β_j^* in (3-6) becomes

$$\beta_j + \beta_{n+1} > 0, \quad \text{and} \quad \beta_j + \beta_\ell > 0, \quad 1 \le j \le s, \text{ and } s+1 \le \ell \le n+1. \tag{3-8}$$

Now (3-1) and (3-8) together imply that the conclusion (1) of the theorem we are proving holds for $n+1$, provided it holds for n.

Assuming now that $\mathcal{D}_{n+1,s}(\beta, \beta_{n+1}) < \infty$, by (3-7), and the induction hypothesis, we have that

$$\begin{aligned} \mathcal{D}_{n+1,s}(\beta, \beta_{n+1}) &= \frac{1}{\beta_{n+1}}\left(\frac{R_{n,s}(\beta)}{S_{n,s}(\beta)} - \frac{R_{n,s}(\beta^*)}{S_{n,s}(\beta^*)}\right) \\ &= \frac{1}{\beta_{n+1}}\left(\frac{R_{n,s}(\beta)S_{n,s}(\beta^*) - R_{n,s}(\beta^*)S_{n,s}(\beta)}{S_{n,s}(\beta)S_{n,s}(\beta^*)}\right). \end{aligned} \tag{3-9}$$

Using the definition (3-6) of β^*, we have

$$\begin{aligned} S_{n,s}(\beta^*) &= \prod_{j=1}^{s} \beta_j^* \prod_{\substack{1 \le j \le s \\ s+1 \le \ell \le n}} (\beta_j^* + \beta_\ell^*) = \prod_{j=1}^{s} (\beta_j + \beta_{n+1}) \prod_{\substack{1 \le j \le s \\ s+1 \le \ell \le n}} (\beta_j + \beta_\ell) \\ &= \prod_{\substack{1 \le j \le s \\ s+1 \le \ell \le n+1}} (\beta_j + \beta_\ell). \end{aligned} \tag{3-10}$$

Therefore using (3-3) and (3-10):

$$S_{n,s}(\beta)S_{n,s}(\beta^*) = \prod_{j=1}^{s}\beta_j \prod_{\substack{1\le j\le s\\ s+1\le \ell\le n}}(\beta_j+\beta_\ell) \prod_{\substack{1\le j\le s\\ s+1\le \ell\le n+1}}(\beta_j+\beta_\ell) \tag{3-11}$$
$$= S_{n+1,s}(\beta,\beta_{n+1})\cdot \prod_{\substack{1\le j\le s\\ s+1\le \ell\le n}}(\beta_j+\beta_\ell),$$

where $S_{n+1,s}(\beta,\beta_{n+1})$ is as in (3-3). The expression in the numerator of (3-9) is given, using (3-10) and (3-3) by

$$R_{n,s}(\beta)S_{n,s}(\beta^*) - R_{n,s}(\beta^*)S_{n,s}(\beta) \tag{3-12}$$
$$= R_{n,s}(\beta)\cdot \prod_{\substack{1\le j\le s\\ s+1\le \ell\le n+1}}(\beta_j+\beta_\ell) - R_{n,s}(\beta^*)\cdot\prod_{j=1}^{s}\beta_j \prod_{\substack{1\le j\le s\\ s+1\le \ell\le n}}(\beta_j+\beta_\ell)$$
$$= \left(R_{n,s}(\beta)\cdot\prod_{j=1}^{s}(\beta_j+\beta_{n+1}) - R_{n,s}(\beta^*)\cdot\prod_{j=1}^{s}\beta_j\right)\cdot \prod_{\substack{1\le j\le s\\ s+1\le \ell\le n}}(\beta_j+\beta_\ell).$$

Using (3-11) and (3-12) in (3-9), we see that the numerator and denominator of (3-9) share the common factor,

$$\prod_{\substack{1\le j\le s\\ s+1\le \ell\le n}}(\beta_j+\beta_\ell).$$

Removing this common factor we see that

$$\mathcal{D}_{n+1}(\beta,\beta_{n+1}) = \frac{f(\beta_{n+1})/(\beta_{n+1})}{S_{n+1,s}(\beta,\beta_{n+1})},$$

where we now think of $(\beta_1,\dots,\beta_{n+1})$ as indeterminates, and f as a polynomial in the ring $\mathbb{Q}(\beta_1,\dots,\beta_n)[\beta_{n+1}]$ of polynomials in the indeterminate β_{n+1} over the field of rational functions $\mathbb{Q}(\beta_1,\dots,\beta_n)$ in n indeterminates, with f given by

$$f(\beta_{n+1}) = R_{n,s}(\beta)\cdot\prod_{j=1}^{s}(\beta_j+\beta_{n+1}) - R_{n,s}(\beta^*)\cdot\prod_{j=1}^{s}\beta_j.$$

Now the formulas (3-6) defining β^* in terms of $\beta_1,\dots,\beta_{n+1}$ show that if $\beta_{n+1}=0$, then $\beta^*=\beta$. It now follows that $f(0)=0$, so that

$$R_{n+1,s}(\beta,\beta_{n+1}) = f(\beta_{n+1})/\beta_{n+1},$$

is a polynomial in the ring $\mathbb{Q}(\beta_1,\dots,\beta_n)$. But noting further that $f\in\mathbb{Z}[\beta_1,\dots,\beta_{n+1}]$, and the divisor β_{n+1} has leading coefficient 1, we see that $R_{n+1,s}\in\mathbb{Z}[\beta_1,\dots,\beta_{n+1}]$, which we wanted to prove. We therefore have the recursive formula:

$$R_{n+1,s}(\beta, \beta_{n+1}) = \frac{1}{\beta_{n+1}} \cdot \left(R_{n,s}(\beta) \cdot \prod_{j=1}^{s} (\beta_j + \beta_{n+1}) - R_{n,s}(\beta^*) \cdot \prod_{j=1}^{s} \beta_j \right). \tag{3-13}$$

By the induction hypothesis, $R_{n,s}$ is a homogeneous polynomial in the n variables $\beta_1, \dots, \beta_n$ of total degree $(n-s)(s-1)$. By the definition (3-6) of β^*, we see that $R_{n,s}(\beta^*)$ is a also homogeneous polynomial of the $n+1$ variables $\beta_1, \dots, \beta_{n+1}$. The quantity in large parentheses in (3-13) is therefore the difference of two homogeneous polynomials of total degree $(n-s)(s-1)+s$. It is therefore either zero, or itself a homogeneous polynomial of degree $(n-s)(s-1)+s$. But it cannot be zero, since then the norm of a monomial is zero, which is absurd. Finally by (3-13), the polynomial $R_{n+1,s}$ is also homogeneous, being the ratio of two homogeneous polynomials, and has total degree

$$(n-s)(s-1)+s-1 = ((n+1)-s)(s-1).$$

We will now show that $R_{n+1,s}(\beta)$ and $S_{n+1,s}(\beta)$ have no common factors.

By induction hypothesis $S_{n,s}(\beta)$ has no common factors with $R_{n,s}(\beta)$. Since $S_{n,s}(\beta)$ is a product of linear factors β_j and $(\beta_j + \beta_\ell)$ where $1 \le j \le s$, and $s+1 \le \ell \le n$, none of these factors divide $R_{n,s}(\beta)$.

From the symmetry of $\Omega_{n+1,s}$, the definition (3-2), and the symmetry of $S_{n+1,s}$ we know that $R_{n+1,s}(\beta, \beta_n+1)$ is symmetric in variables $\beta_1, \dots, \beta_s$ and variables $\beta_{s+1}, \dots, \beta_{n+1}$. Starting from these facts, we can verify that none of the linear factors of $S_{n+1,s}$ divides the right hand side of (3-13), by noting that even if these linear factors vanish, the right hand side of (3-13) does not. Hence, $R_{n+1,s}(\beta, \beta_{n+1})$ and $S_{n+1,s}(\beta, \beta_{n+1})$ have no common factors.

Therefore, the inductive proof of the theorem is complete, except that we need to establish the claim (3-7) on which the above induction was based. Note that from (3-5), with $\mathbf{1} = (1, \dots, 1)$, we have

$$\mathcal{D}_{n,s}(\beta) = \frac{1}{\pi^n} \|e_{\beta-\mathbf{1}}\|^2 = \frac{1}{\pi^n} \int_{\Omega_{n,s}} |e_{\beta-\mathbf{1}}(z)|^2 \, dV(z).$$

Using polar coordinates $z_j = r_j e^{i\theta_j}$ and using the fact that $dV(z) = \prod_{j=1}^{n} r_j \, dr_j \, d\theta_j = r^{\mathbf{1}} \, dV(r) \, dV(\theta)$, where $r = (r_1, \dots, r_n)$, we have

$$\mathcal{D}_{n,s}(\beta) = \frac{1}{\pi^n} \cdot (2\pi)^n \int_{|\Omega_{n,s}|} r^{2\beta-\mathbf{1}} \, dV(r),$$

where $|\Omega_{n,s}| \subset \mathbb{R}^n$ is the Reinhardt shadow of $\Omega_{n,s}$, i.e., the image of $\Omega_{n,s}$ under the map $z \mapsto (|z_1|, \dots, |z_n|)$. We will make the further change of variables $t_j = r_j^2$, which maps $|\Omega_{n,s}|$ diffeomorphically to itself. The integral now takes the form

$$\mathcal{D}_{n,s}(\beta) = \frac{1}{\pi^n} \cdot (2\pi)^n \int_{|\Omega_{n,s}|} t^{\beta - \mathbf{1}}\, dV(t).$$

We will transform this integral into an n-fold repeated integral. For simplicity of notation we denote repeated integrals with differential in front and integrand after that, so that

$$\int_{x_2=a_2}^{b_2} g(x_2)\left(\int_{x_1=a_1}^{b_1} f(x_1,x_2)dx_1\right)dx_2 = \int_{a_2}^{b_2} dx_2 \cdot g(x_2) \int_{a_1}^{b_1} dx_1 \cdot f(x_1,x_2),$$

and adopt similar notations for multiple repeated integrals, so that the innermost integral in the conventional notation is the rightmost factor. The region of integration over which $t \in \mathbb{R}^n$ ranges is described by the inequalities

$$0 \le t_1 \cdots t_s < t_{s+1} \cdots t_n < 1, \quad 0 \le t_1 < 1, \ldots, 0 \le t_n < 1.$$

Then, $\mathcal{D}_{n,s}(\beta)$ can be expressed explicitly by the following n-fold integral:

$$\int_0^1 dt_1 \cdot t_1^{\beta_1 - 1} \int_0^1 dt_2 \cdot t_2^{\beta_2 - 1} \cdots \int_0^1 dt_s \cdot t_s^{\beta_s - 1} \int_{t_1 \cdots t_s}^1 dt_{s+1} \cdot t_{s+1}^{\beta_{s+1}-1} \cdot \int_{t_1\cdots t_s/t_{s+1}}^1 dt_{s+2} \cdot t_{s+2}^{\beta_{s+2}-1} \cdots \int_{t_1 \cdots t_s/(t_{s+1}\cdots t_{n-1})}^1 dt_n \cdot t_n^{\beta_n - 1}.$$

Similarly,

$$\begin{aligned}
&\mathcal{D}_{n+1,s}(\beta, \beta_{n+1}) \\
&= \int_0^1 dt_1 \cdot t_1^{\beta_1-1} \cdots \int_0^1 dt_s \cdot t_s^{\beta_s-1} \int_{t_1\cdots t_s}^1 dt_{s+1} \cdot t_{s+1}^{\beta_{s+1}-1} \int_{t_1\cdots t_s/t_{s+1}}^1 dt_{s+2} \cdot t_{s+2}^{\beta_{s+2}-1} \\
&\qquad \cdots \int_{t_1\cdots t_s/(t_{s+1}\cdots t_{n-1})}^1 dt_n \cdot t_n^{\beta_n-1} \int_{t_1\cdots t_s/(t_{s+1}\cdots t_n)}^1 dt_{n+1} \cdot t_{n+1}^{\beta_{n+1}-1} \\
&= \int_0^1 dt_1 \cdot t_1^{\beta_1-1} \cdots \int_0^1 dt_s \cdot t_s^{\beta_s-1} \int_{t_1\cdots t_s}^1 dt_{s+1} \cdot t_{s+1}^{\beta_{s+1}-1} \int_{t_1 \cdot t_s/t_{s+1}}^1 dt_{s+2} \cdot t_{s+2}^{\beta_{s+2}-1} \\
&\qquad \cdots \int_{t_1\cdots t_s/(t_{s+1}\cdots t_{n-1})}^1 dt_n \cdot \frac{1}{\beta_{n+1}}\left(1 - \left(\frac{t_1 \cdots t_s}{t_{s+1}\cdots t_n}\right)^{\beta_{n+1}}\right), \\
&\qquad\qquad \text{(where we have evaluated the innermost integral)} \\
&= \frac{1}{\beta_{n+1}}\Bigg(\mathcal{D}_{n,s}(\beta) - \int_0^1 dt_1 \cdot t_1^{\beta_1+\beta_{n+1}-1} \cdots \int_0^1 dt_s \cdot t_s^{\beta_s+\beta_{n+1}-1} \int_{t_1\cdots t_s}^1 dt_{s+1} \cdot t_{s+1}^{\beta_{s+1}-\beta_{n+1}-1} \\
&\qquad \cdot \int_{t_1\cdots t_s/t_{s+1}}^1 dt_{s+2} \cdot t_{s+2}^{\beta_{s+2}-\beta_{n+1}-1} \cdots \int_{t_1\cdots t_s/(t_{s+1}\cdots t_{n-1})}^1 dt_n \cdot t_n^{\beta_n-\beta_{n+1}-1}\Bigg)
\end{aligned}$$

$$= \frac{1}{\beta_{n+1}} \big(\mathcal{D}_{n,s}(\beta) - \mathcal{D}_{n,s}(\beta^*)\big),$$

which completes the proof of (3-7). □

3A. ***Proof of Theorem 1.2.*** From (2-2), we may write

$$\mathbb{B}_{\Omega_{n,n-1}}(z, w) = \sum_{\beta \in \mathcal{T}} \frac{1}{\|e_{\beta-\mathbf{1}}\|^2} t^{\beta-\mathbf{1}},$$

where $\mathbf{1} = (1, \dots, 1)$ and $\mathcal{T}$ is the set of indices corresponding to $s = n-1$ in (3-1), i.e.,

$$\beta_j > 0, \quad \beta_j + \beta_n > 0, \quad \text{for } 1 \le j \le n-1.$$

Furthermore, we have from Theorem 3.1 that

$$\|e_{\beta-\mathbf{1}}\|^2 = \pi^n \frac{R_{n,n-1}(\beta)}{S_{n,n-1}(\beta)},$$

where by (3-3), we have

$$S_{n,n-1}(\beta) = \prod_{j=1}^{n-1} \beta_j \prod_{j=1}^{n-1} (\beta_j + \beta_n) = \prod_{j=1}^{n-1} \beta_j (\beta_j + \beta_n),$$

and using the recursive relation (3-13) and the fact that $R_{n,n} \equiv 1$ (see (3-4)), we see that

$$R_{n,n-1}(\beta) = \frac{1}{\beta_n} \left(\prod_{j=1}^{n-1} (\beta_j + \beta_n) - \prod_{j=1}^{n-1} \beta_j \right).$$

Therefore, with $t_j = z_j \overline{w}_j$, we have

$$\begin{aligned} \mathbb{B}_{\Omega_{n,n-1}}(z, w) &= \widetilde{B}(t_1, \dots, t_n) \\ &= \frac{1}{\pi^n} \sum_{\beta \in \mathcal{T}} \frac{\prod_{j=1}^{n-1} \beta_j (\beta_j + \beta_n)}{\frac{1}{\beta_n} \big(\prod_{j=1}^{n-1} (\beta_j + \beta_n) - \prod_{j=1}^{n-1} \beta_j \big)} t^{\beta-\mathbf{1}}. \end{aligned} \tag{3-14}$$

We now consider the function $\tilde{b}$ of one variable defined by

$$\tilde{b}(t_n) = \widetilde{B}(\underbrace{0, \dots, 0}_{n-1}, t_n).$$

This is defined in the punctured disc $\{0 < |t_n| < 1\}$, and noting that in (3-14) only the terms with $\beta_j = 1,\ 1 \le j \le n-1$ survive if $t_1 = \dots = t_{n-1} = 0$, we conclude that

$$\tilde{b}(t_n) = \sum_{\beta_n=0}^{\infty} \frac{\beta_n(1+\beta_n)^{n-1}}{(1+\beta_n)^{n-1}-1} t_n^{\beta_n-1} \tag{3-15}$$
$$= \frac{t_n^{-1}}{n-1} + \sum_{k=1}^{\infty} k t_n^{k-1} + \sum_{k=1}^{\infty} \frac{k}{(k+1)^{n-1}-1} t_n^{k-1}$$
$$= \frac{t_n^{-1}}{n-1} + \frac{1}{(1-t_n)^2} + \hat{b}(t_n),$$

where

$$\hat{b}(t_n) = \sum_{k=1}^{\infty} \frac{k}{(k+1)^{n-1}-1} \cdot t_n^{k-1}.$$

Since the function $\widetilde{B}$ is holomorphic on the domain

$$\{(z_1\overline{w}_1, \dots, z_n\overline{w}_n) \mid z, w \in \Omega_{n,n-1}\}$$

it follows that $\tilde{b}$ is holomorphic in the punctured disc $\{0 < |t_n| < 1\}$, and therefore $\hat{b}$ is holomorphic in the unit disc $\{|t_n| < 1\}$.

Now for a contradiction, assume that $\mathbb{B}_{\Omega_{n,n-1}}$ is rational. It follows that $\hat{b}$ is a rational function of one variable, holomorphic in the unit disc, and its k-th Taylor coefficient decays as $k^{-(n-2)}$ as $k \to \infty$. Recall that by hypothesis $n \geq 3$, so the coefficients go to zero.

Let $\alpha_1, \dots, \alpha_m$ be the poles of the rational function $\hat{b}$, where $|\alpha_j| \geq 1$ since $\hat{b}$ is holomorphic in the unit disc. It follows by expansion in partial fractions (see [Flajolet and Sedgewick 2009, pp. 256ff.]) that the k-th Taylor coefficient of $\hat{b}$ is of the form $\sum_{j=1}^{m} \alpha_j^{-k} \Pi_j(k)$, where Π_j is a polynomial for each j. Since the coefficients go to zero as $k \to \infty$, we must have $|\alpha_j| > 1$, for each $j = 1, \dots, m$. Therefore, the decay of the coefficients is exponential in k, which contradicts the $k^{-(n-2)}$ decay. Therefore $\hat{b}$ cannot be a rational function, and so $\mathbb{B}_{\Omega_{n,n-1}}$ is not a rational function if $n \geq 3$.

3B. ***Some remarks on the nature of the Bergman Kernel of $\boldsymbol{\Omega_{n,s}}$.*** The form of the coefficients of the series in Theorem 3.1 as well as the argument in the proof of Theorem 1.2 suggest that the Bergman kernel of $\Omega_{n,s}$ is not rational except for $s = 1$, though we do not have a complete proof of this yet. However, Theorem 3.1 is already sufficient to rule out certain hasty conjectures about the form of $\mathbb{B}_{\Omega_{n,s}}$ that one might make based on (1-3) or similar formulas in [Park 2018]. For example, for $s \neq 1$, the kernel $\mathbb{B}_{\Omega_{n,s}}$ *cannot* be written in the form

$$\frac{1}{\pi^n} \frac{P(t)}{\left(\prod_{b=s+1}^{n} t_b^{k_b} - \prod_{a=1}^{s} t_a^{k_a}\right)^2 \cdot \prod_{b=s+1}^{n} (1-t_b)^2},$$

for a polynomial P, since the coefficient of t^α of the Taylor expansion of this function is a polynomial in α. Additionally, we saw above that when $s \neq 1$, the Taylor coefficients of $\mathbb{B}_{\Omega_{n,s}}$ are rational functions of α which are not polynomials. Another interesting algebraic property is given by the following:

Proposition 3.2. *Let $n \geq 2$ and $1 \leq s \leq n-1$. Let $\widetilde{B}$ be the function of $t_j = z_j \overline{w}_j$ associated with the Bergman kernel of $\Omega_{n,s}$, as defined in (2-4). Then there is a nonzero linear differential operator $\mathscr{L}$ with polynomial coefficients, such that $\mathscr{L}\widetilde{B}$ is a polynomial.*

Proof. The case $s = 1$ is trivial, since then by Theorem 1.1, $\widetilde{B}$ is a rational function P/Q, where P, Q are polynomials. Therefore we can simply take $\mathscr{L}$ to be the zeroth order multiplication operator determined by Q.

Notice that we can write, thanks to Theorem 3.1, the series representation

$$\widetilde{B}(t) = \frac{1}{\pi^n} \cdot \frac{1}{t_1 \cdots t_n} \sum_{\beta \in \mathcal{S}} \frac{S_{n,s}(\beta)}{R_{n,s}(\beta)} t^\beta,$$

where $R_{n,s}(\beta_1, \ldots, \beta_n)$ and $S_{n,s}(\beta_1, \ldots, \beta_n)$ are homogeneous polynomials in the variables $\beta_1, \ldots, \beta_n$, and $\mathcal{S}$ is the subset of $\mathbb{Z}^n$ determined by the conditions (3-1). Let $\mathcal{M}$ denote the multiplication operator induced by the polynomial $t_1 \cdots t_n$, and let

$$\mathscr{L}_0 = R_{n,s}\Big(t_1 \frac{\partial}{\partial t_1}, \ldots, t_n \frac{\partial}{\partial t_n}\Big) \circ \mathcal{M}.$$

Then we see that

$$\mathscr{L}_0 \widetilde{B}(t) = \frac{1}{\pi^n} \sum_{\beta \in S} S_{n,s}(\beta) t^\beta. \tag{3-16}$$

Now since the coefficients $S_{n,s}(\beta)$ are polynomials in β and the region of summation $\mathcal{S}$ is the intersection of a finite number of *closed* half-spaces in $\mathbb{Z}^n$ (since the open conditions in (3-1) can be replaced by closed conditions), it follows that the right hand side of (3-16) is a rational function (see the proof of Theorem 1.1 below). If $Q(t)$ is the denominator of this rational function, and $\mathcal{Q}$ is the multiplication operator induced by Q, we can take $\mathscr{L} = \mathcal{Q} \circ \mathscr{L}_0$. □

4. Proof of Theorem 1.1

4A. ***Kernel of model domain.*** We begin by computing the Bergman kernel of the model elementary Reinhardt domain $\Omega_{n,1}$:

Proposition 4.1. *The Bergman kernel of $\Omega_{n,1}$ is given by*

$$\mathbb{B}_{\Omega_{n,1}}(z, w) = \frac{1}{\pi^n} \cdot \frac{\prod_{b=2}^n t_b}{\big(\prod_{b=2}^n t_b - t_1\big)^2 \cdot \prod_{b=2}^n (1 - t_b)^2},$$

where

$$t_b = z_b \overline{w}_b \quad \textit{for } 1 \le b \le n.$$

Proof. From Theorem 3.1, we see that for $\alpha \in \mathbb{Z}^n$, we have $\|e_\alpha\|^2_{\Omega_{n,1}} < \infty$ if and only if

$$\alpha_1 + 1 > 0, \qquad \alpha_1 + \alpha_\ell + 2 > 0, \quad 2 \le \ell \le n,$$

which is equivalent to

$$\alpha_1 \ge 0, \qquad \alpha_1 + \alpha_\ell + 1 \ge 0, \quad 2 \le \ell \le n.$$

Let $\mathcal{S} \subset \mathbb{Z}^n$ be the set of multi-indices satisfying the above condition. Also from Theorem 3.1, it follows that for $\alpha \in \mathcal{S}$ we have

$$\|e_\alpha\|^2_{\Omega_{n,1}} = \pi^n \frac{1}{(\alpha_1+1)\prod_{b=2}^n(\alpha_1+\alpha_b+2)}.$$

Using (2-2) and the abbreviation $t_b = z_b \overline{w}_b$, we have by a direct summation of the series (2-2):

$$\begin{aligned}
\mathbb{B}_{\Omega_{n,1}}(z,w) &= \frac{1}{\pi^n}\sum_{\alpha\in\mathcal{S}}\Bigg((\alpha_1+1)\prod_{b=2}^n(\alpha_1+\alpha_b+2)\Bigg)t^\alpha \\
&= \frac{1}{\pi^n}\cdot\sum_{\alpha_1=0}^\infty(\alpha_1+1)t_1^{\alpha_1}\prod_{b=2}^n\Bigg(\sum_{\alpha_b=-\alpha_1-1}^\infty(\alpha_1+\alpha_b+2)t_b^{\alpha_b}\Bigg) \\
&= \frac{1}{\pi^n}\cdot\prod_{b=2}^n\frac{1}{t_b(1-t_b)^2}\sum_{\alpha_1=0}^\infty(\alpha_1+1)t_1^{\alpha_1}\prod_{b=2}^n t_b^{-\alpha_1} \\
&\quad\Bigg(\text{using the easily proved identity } \sum_{\alpha_b=-\alpha_1-1}^\infty(\alpha_1+\alpha_b+2)t_b^{\alpha_b} = \frac{t_b^{-\alpha_1-1}}{(1-t_b)^2}\Bigg) \\
&= \frac{1}{\pi^n}\cdot\prod_{b=2}^n\frac{1}{t_b(1-t_b)^2}\sum_{\alpha_1=0}^\infty(\alpha_1+1)\rho^{\alpha_1}, \qquad \text{with } \rho = t_1/\textstyle\prod_{b=2}^n t_b \\
&= \frac{1}{\pi^n}\cdot\frac{1}{(1-\rho)^2}\cdot\prod_{b=2}^n\frac{1}{t_b(1-t_b)^2} \\
&= \frac{1}{\pi^n}\cdot\frac{1}{\big(1-t_1/\prod_{b=2}^n t_b\big)^2}\cdot\prod_{b=2}^n\frac{1}{t_b(1-t_b)^2} \\
&= \frac{1}{\pi^n}\cdot\frac{\prod_{b=2}^n t_b}{\big(\prod_{b=2}^n t_b - t_1\big)^2\cdot\prod_{b=2}^n(1-t_b)^2},
\end{aligned}$$

where we have used the identity $\sum_{\alpha_1=0}^\infty(\alpha_1+1)\rho^{\alpha_1} = 1/(1-\rho)^2$ which holds since $|\rho| < 1$. □

4B. ***Explicit kernel.*** The following simple arithmetical fact will be used:

Lemma 4.2. *Let $k_1, \dots, k_n$ be positive integers such that* $\gcd(k_1, \dots, k_n) = 1$, *i.e.*, *$k_1, \dots, k_n$ are relatively prime. Let $K = \operatorname{lcm}(k_1, \dots, k_n)$ and $\ell_j = K/k_j$ with $1 \le j \le n$. Then*

$$\operatorname{lcm}(\ell_1, \dots, \ell_n) = K.$$

Proof. Let $k_j = \prod_{p \in \text{Primes}} p^{v_j(p)}$ be the prime factoring of k_j. Then

$$K = \prod_{p \in \text{Primes}} p^{N(p)}, \quad \text{where } N(p) = \max_{1 \le j \le n} (v_j(p)).$$

Now

$$\ell_j = \frac{K}{k_j} = \prod_{p \in \text{Primes}} p^{N(p) - v_j(p)}.$$

So,

$$\begin{aligned} \operatorname{lcm}(\ell_1, \dots, \ell_n) &= \prod_{p \in \text{Primes}} p^{\max_j (N(p) - v_j(p))} = \prod_{p \in \text{Primes}} p^{N(p) - \min_j (v_j(p))} \\ &= \prod_{p \in \text{Primes}} p^{N(p)} = K, \end{aligned}$$

where we have used the fact that since $\gcd(k_1, \dots, k_n) = 1$, it follows that

$$\min_{1 \le j \le n} (v_j(p)) = 0. \qquad \square$$

Proof of Theorem 1.1. Let $\phi : \Omega_{n,1} \to \mathscr{H}(k)$ be the standard proper holomorphic map which was constructed in Proposition 2.1. Notice that this map is given by the formula

$$\phi(z_1, \dots, z_n) = (z_1^{\ell_1}, \dots, z_n^{\ell_n}), \tag{4-1}$$

where ℓ_j has exactly the same meaning as in the statement of our result. Now by the famous Bell transformation formula [Bell 1982]:

$$u(z) \cdot \mathbb{B}_{\mathscr{H}(k)}(\phi(z), w) = \sum_j \mathbb{B}_{\Omega_{n,1}}(z, \Phi_j(w)) \cdot \overline{U_j(w)}, \tag{4-2}$$

where $u = \det(\phi')$, the Φ_j are local branches of ϕ^{-1}, and $U_j = \det(\Phi_j')$. The Jacobian determinant of ϕ is given by

$$u(z) = \det \phi'(z) = \det \operatorname{diag}(\ell_1 z_1^{\ell_1 - 1}, \dots, \ell_n z_n^{\ell_n - 1}) = \prod_{a=1}^{n} \ell_a z_a^{\ell_a - 1}.$$

The map ϕ has $L = \prod_{a=1}^{n} \ell_a$ local inverses. To enumerate them, introduce the set of multi-indices

$$\mathfrak{B} = \{(j_1, \dots, j_n) \in \mathbb{Z}^n \mid 0 \le j_a \le \ell_a - 1, \text{ for } a = 1, \dots, n\}, \tag{4-3}$$

then for each multi-index $j \in \mathfrak{B}$, there is a branch Φ_j of the local inverse of ϕ given by

$$\Phi_j(z_1, \dots, z_n) = \big(\zeta_1^{j_1} z_1^{1/\ell_1}, \zeta_2^{j_2} z_2^{1/\ell_2}, \dots, \zeta_n^{j_n} z_n^{1/\ell_b}\big),$$

where

$$\zeta_a = e^{2\pi i/\ell_a}, \quad \text{for each } 1 \le a \le n$$

is an ℓ_a-th root of unity, and the root functions $z_1^1/\ell_1, \dots, z_n^1/\ell_n$ exist locally off the critical locus. We then have for each $j \in \mathfrak{B}$

$$\begin{aligned} U_j(w) = \det \Phi_j'(w) &= \det \operatorname{diag}\left(\frac{\zeta_1^{j_1}}{\ell_1} w_1{}^{1/\ell_1 - 1}, \dots, \frac{\zeta_n^{j_n}}{\ell_n} w_n{}^{1/\ell_n - 1}\right) \\ &= \prod_{a=1}^{n} \frac{\zeta_a{}^{j_a}}{\ell_a} w_a{}^{1/\ell_a - 1}, \end{aligned}$$

where $\operatorname{diag}(\cdot)$ denotes a diagonal matrix with the specified diagonal entries. Therefore by Bell's formula (4-2) we have

$$\begin{aligned} &\prod_{a=1}^{n} \ell_a z_a^{\ell_a - 1} \cdot \mathbb{B}_{\mathscr{H}(k)}(\phi(z), w) \\ &= \sum_{j \in \mathfrak{B}} \mathbb{B}_{\Omega_{n,1}}(z, \Phi_j(w)) \cdot \overline{\prod_{a=1}^{n} \frac{\zeta_a^{j_a}}{\ell_a} w_a{}^{1/\ell_a - 1}} \\ &= \frac{1}{\pi^n} \cdot \sum_{j \in \mathfrak{B}} \frac{\prod_{b=2}^{n} \bar{\zeta}_b^{j_b} z_b \overline{w}_b^{1/\ell_b}}{\big(\prod_{b=2}^{n} \bar{\zeta}_b^{j_b} z_b \overline{w}_b^{1/\ell_b} - \zeta_1^{j_1} z_1 \overline{w}_1^{1/\ell_1}\big)^2 \cdot \prod_{b=2}^{n} \big(1 - \bar{\zeta}_b^{j_b} z_b w_b^{1/\ell_b}\big)^2} \\ &\qquad\qquad \cdot \prod_{a=1}^{n} \frac{\bar{\zeta}_a^{j_a}}{\ell_a} \overline{w}_a^{1/\ell_a - 1}, \end{aligned}$$

where we have used the formula in Proposition 4.1 for the Bergman kernel of $\Omega_{n,1}$. Introduce the abbreviations

$$r_a = z_a \overline{w}_a^{1/\ell_a}, \quad a = 1, \dots, n, \tag{4-4}$$

so that we have from the above (recall that $L = \prod_{j=1}^{n} \ell_j$)

$$\begin{aligned}\mathbb{B}_{\mathscr{H}(k)}(\phi(z), w) &= \frac{1}{\pi^n L^2} \sum_{j \in \mathfrak{B}} \frac{\prod_{a=1}^{n} \bar{\zeta}_a^{j_a} r_a^{1-\ell_a} \cdot \prod_{b=2}^{n} \bar{\zeta}_b^{j_b} r_b}{\left(\prod_{b=2}^{n} \bar{\zeta}_b^{j_b} r_b - \bar{\zeta}_1^{j_1} r_1\right)^2 \prod_{b=2}^{n} (1 - \bar{\zeta}_b^{j_b} r_b)^2} \\ &= \frac{1}{\pi^n L^2} \sum_{j \in \mathfrak{B}} \frac{\bar{\zeta}_1^{j_1} r_1^{1-\ell_1} \cdot \prod_{b=2}^{n} \bar{\zeta}_b^{2j_b} r_b^{2-\ell_b}}{\left(\prod_{b=2}^{n} \bar{\zeta}_b^{j_b} r_b - \bar{\zeta}_1^{j_1} r_1\right)^2 \prod_{b=2}^{n} (1 - \bar{\zeta}_b^{j_b} r_b)^2}.\end{aligned} \tag{4-5}$$

Let $\widehat{B}(r_1, \ldots, r_n)$ denote the quantity in (4-5). We claim that the function $\widehat{B}$ of n variables has the following invariance property, which will be needed later: for each c with $1 \le c \le n$, we have

$$\widehat{B}(r_1, \ldots, \bar{\zeta}_c r_c, \ldots, r_n) = \widehat{B}(r_1, \ldots, r_n). \tag{4-6}$$

To see this, notice that we have, for each c with $2 \le c \le n$, that

$$\begin{aligned}&\widehat{B}(r_1, r_2, \ldots, \bar{\zeta}_c r_c, \ldots, r_n) \\ &= \frac{1}{\pi^n L^2} \sum_{j \in \mathfrak{B}} \frac{\bar{\zeta}_1^{j_1} r_1^{1-\ell_1} \cdot \bar{\zeta}_c^{2(j_c+1)} r_c^{2-\ell_c} \cdot \prod_{\substack{2 \le b \le n \\ b \ne c}} \bar{\zeta}_b^{2j_b} r_b^{2-\ell_b}}{\left(\bar{\zeta}_c^{j_c+1} r_c \prod_{\substack{2 \le b \le n \\ b \ne c}} \bar{\zeta}_b^{j_b} r_b - \bar{\zeta}_1^{j_1} r_1\right)^2 (1 - \bar{\zeta}_c^{j_c+1} r_c)^2 \prod_{\substack{2 \le b \le n \\ b \ne c}} (1 - \bar{\zeta}_b^{j_b} r_b)^2}.\end{aligned}$$

Notice that the above sum is precisely the same as $\widehat{B}(r_1, \ldots, r_n)$, since changing j_c to $j_c + 1$ simply amounts to a reindexing of the sum, thanks to the fact that the ℓ_c-th roots of unity form a cyclic group generated by ζ_c.

In a similar way, $\widehat{B}(\bar{\zeta}_1 r_1, r_2, \ldots, r_n)$ is precisely the same as $\widehat{B}(r_1, \ldots, r_n)$, since changing j_1 to $j_1 + 1$ simply amounts to a reindexing of the sum, thanks to the fact that the ℓ_1-th roots of unity form a cyclic group generated by ζ_1. These two observations combined establish (4-6).

Now let

$$\Delta = \left(\left(\prod_{b=2}^{n} r_b \right)^K - r_1^K \right)^2 \cdot \prod_{b=2}^{n} (1 - r_b^{\ell_b})^2, \tag{4-7}$$

where $K = \operatorname{lcm}(k_1, \ldots, k_n)$ as in the statement of the theorem. Then we can write

$$\begin{aligned}&\widehat{B}(r_1, \ldots, r_n) \\ &= \frac{1}{\pi^n L^2 \Delta} \sum_{j \in \mathfrak{B}} \left(\bar{\zeta}_1^{j_1} r_1^{1-\ell_1} \prod_{b=2}^{n} \bar{\zeta}_b^{2j_b} r_b^{2-\ell_b} \cdot \frac{\left(\left(\prod_{b=2}^{n} r_b\right)^K - r_1^K\right)^2}{\left(\prod_{b=2}^{n} \bar{\zeta}_b^{j_b} r_b - \bar{\zeta}_1^{j_1} r_1\right)^2} \cdot \prod_{b=2}^{n} \frac{(1 - r_b^{\ell_b})^2}{(1 - \bar{\zeta}_b^{j_b} r_b)^2} \right)\end{aligned}$$

(4-8)

$$= \frac{1}{\pi^n L^2 \Delta} \sum_{j \in \mathfrak{B}} \left(\bar{\zeta}_1^{j_1} r_1^{1-\ell_1} \prod_{b=2}^{n} \bar{\zeta}_b^{2j_b} r_b^{2-\ell_b} \cdot \left(\sum_{\nu=0}^{K-1} \left(\prod_{b=2}^{n} \bar{\zeta}_b^{j_b} r_b \right)^{\nu} (\bar{\zeta}_1^{j_1} r_1)^{K-\nu-1} \right)^2 \times \prod_{b=2}^{n} \left(\sum_{m_b=0}^{\ell_b-1} (\bar{\zeta}_b^{j_b} r_b)^{m_b} \right)^2 \right)$$

(4-9)

$$= \frac{1}{\pi^n L^2 \Delta} \sum_{\alpha_1=0}^{2K-2} \sum_{\alpha_2=0}^{2K+2\ell_2-4} \cdots \sum_{\alpha_n=0}^{2K+2\ell_n-4} A(\alpha) r_1^{\alpha_1+1-\ell_1} \prod_{b=2}^{n} r_b^{\alpha_b+2-\ell_b}$$

(4-10)

$$= \frac{1}{\pi^n L^2 \Delta} \sum_{\alpha_1=1-\ell_1}^{2K-\ell_1-1} \sum_{\alpha_2=2-\ell_2}^{2K+\ell_2-2} \cdots \sum_{\alpha_n=2-\ell_n}^{2K+\ell_n-2} \tilde{A}(\alpha) r^{\alpha},$$

where in (4-9), for simplicity of notation, we have expressed the quantity under the summation sign in (4-8) as a (Laurent) polynomial in the n variables $(r_1, \dots, r_n)$ with coefficients $A(\alpha) \in \mathbb{C}$. In (4-10), we have reindexed the sum, and we denote $r^\alpha = r_1^{\alpha_1} \cdots r_n^{\alpha_n}$. Also, $\tilde{A}(\alpha) = A(\alpha_1 + \ell_1 - 1, \alpha_2 + \ell_2 - 2, \dots, \alpha_n + \ell_n - 2)$.

Notice that (4-10) is a multivariable polynomial in $(r_1, \dots, r_n)$. Then, by the invariance of $\widehat{B}$ shown in (4-6), we can replace the variable r_a, with $1 \le a \le n$ by $\bar{\zeta}_a r_a$, and the value of the polynomial remains unchanged

$$\frac{1}{\pi^n L^2 \Delta} \sum_{\alpha_1=1-\ell_1}^{2K-\ell_1-1} \sum_{\alpha_2=2-\ell_2}^{2K+\ell_2-2} \cdots \sum_{\alpha_n=2-\ell_n}^{2K+\ell_n-2} \tilde{A}(\alpha) r^{\alpha} = \frac{1}{\pi^n L^2 \Delta} \sum_{\alpha_1=1-\ell_1}^{2K-\ell_1-1} \sum_{\alpha_2=2-\ell_2}^{2K+\ell_2-2} \cdots \sum_{\alpha_n=2-\ell_n}^{2K+\ell_n-2} \bar{\zeta}^{\alpha_a} \tilde{A}(\alpha) r^{\alpha}.$$

Looking at the difference of the two sides of the above equation, we see that for each $r = (r_1, \dots, r_n)$ and each $1 \le a \le n$, we have

$$\sum_{\alpha_1=1-\ell_1}^{2K-\ell_1-1} \sum_{\alpha_2=2-\ell_2}^{2K+\ell_2-2} \cdots \sum_{\alpha_n=2-\ell_n}^{2K+\ell_n-2} (\bar{\zeta}^{\alpha_a} - 1) \tilde{A}(\alpha) r^{\alpha} = 0.$$

This is a polynomial in r which vanishes identically, so each of its coefficients is zero. This implies that for a fixed α, the quantity $\tilde{A}(\alpha)$ can be nonzero only if $(\bar{\zeta}^{\alpha_a} - 1) = 0$. Since this holds for each $1 \le a \le n$, the only terms in (4-10) that survive are the ones in which the monomial $r^\alpha = r_1^{\alpha_1} r_2^{\alpha_2} \cdots r_n^{\alpha_n}$ is of the form

$$\alpha = \ell \cdot \beta =: (\ell_1 \beta_1, \ell_2 \beta_2, \dots, \ell_n \beta_n),$$

for some $\beta \in \mathbb{Z}^n$. From the bounds on the indices α_c in (4-10), this implies that the indices corresponding to possibly nonzero terms are the following multiples of ℓ_c:

$$\alpha_c = 0, \ell_c, \dots, 2K \quad \text{for each } 2 \le c \le n \quad \text{if } \ell_c \ne 1, \tag{4-11}$$

$$\alpha_c = 1, \dots, 2K-1 \quad \text{for each } 2 \le c \le n \quad \text{if } \ell_c = 1, \tag{4-12}$$

since for these (and only these) α_c, we have $2-\ell_c \le \alpha_c \le 2K+\ell_c-2$, and α_c is divisible by ℓ_c. Recall here that by Lemma 4.2, the integer $K = \operatorname{lcm}(k_1, \dots, k_n)$ is divisible by ℓ_c, since we also have $K = \operatorname{lcm}(\ell_1, \dots, \ell_n)$. Similar arguments also show that the indices α_1 for which we can have possibly nonzero terms in (4-10) are

$$\alpha_1 = 0, \ell_1, \dots, 2K-2\ell_1. \tag{4-13}$$

Using the representation $\alpha = \ell \cdot \beta = (\ell_1\beta_1, \dots, \ell_n\beta_n)$, we see that these same indices are also described by the collection $\mathfrak{G}^*(k)$ of $\beta \in \mathbb{Z}^n$ such that

$$0 \le \beta_1 \le \frac{2K}{\ell_1} - 2 = 2k_1 - 2, \tag{4-14}$$

and for each $2 \le b \le n$

$$\begin{cases} 0 \le \beta_b \le 2K/\ell_b = 2k_b & \text{if } \ell_b \ne 1, \\ 1 \le \beta_b \le 2K-1 = 2k_b - 1 & \text{if } \ell_b = 1. \end{cases} \tag{4-15}$$

Notice that the set $\mathfrak{G}$ of (1-5) is contained in $\mathfrak{G}^*(k)$. We can now write

$$\widehat{B}(r_1, \dots, r_n) = \text{(4-10)} = \frac{1}{\pi^n \cdot L^2 \cdot \Delta} \sum_{\beta \in \mathfrak{G}^*(k)} \tilde{A}(\ell \cdot \beta) r^{\ell \cdot \beta}, \tag{4-16}$$

which follows from combining Equations (4-11) through (4-13).

We now proceed to compute the coefficients $\tilde{A}(\ell \cdot \beta)$. Introduce a set of indices $\mathfrak{C} \subset \mathbb{Z}^{n-1}$ by setting

$$\mathfrak{C} = \left\{ (m_2, \dots, m_n) \in \mathbb{Z}^{n-1} \mid 0 \le m_b \le \ell_b - 1 \text{ for } 2 \le b \le n \right\}. \tag{4-17}$$

Now, in (4-8), we rewrite the first square factor as a product of two sums over indices ν and N:

$$\begin{aligned} &\left(\sum_{\nu=0}^{K-1} \left(\prod_{b=2}^{n} \bar{\zeta}_b^{\,j_b} r_b \right)^{\nu} (\bar{\zeta}_1^{\,j_1} r_1)^{K-\nu-1} \right)^2 \\ &\qquad = \left(\sum_{\nu=0}^{K-1} \left(\prod_{b=2}^{n} \bar{\zeta}_b^{\,j_b} r_b \right)^{\nu} (\bar{\zeta}_1^{\,j_1} r_1)^{K-\nu-1} \right) \left(\sum_{N=0}^{K-1} \left(\prod_{b=2}^{n} \bar{\zeta}_b^{\,j_b} r_b \right)^{N} (\bar{\zeta}_1^{\,j_1} r_1)^{K-N-1} \right). \end{aligned}$$

Similarly, writing each of the other $(n-2)$ square factors $\left(\sum_{m_b=0}^{\ell_b-1} (\bar{\zeta}_b^{\,j_b} r_b)^{m_b}\right)^2$ for $2 \le b \le n$ in (4-8) as a product of sums over different indices m_b and M_b and then expanding the products we can rewrite (4-8) as

$$\begin{aligned}(4\text{-}18)\quad \widehat{B}(r_1,\dots,r_n) &= \frac{1}{\pi^n L^2\Delta}\sum_{j\in\mathfrak{B}}\sum_{m,M\in\mathfrak{C}}\sum_{\nu,N=0}^{K-1}\bar{\zeta}_1^{\,j_1(2K-\nu-N-1)}r_1^{2K-\nu-N-\ell_1-1}\\ &\qquad\cdot\prod_{b=2}^{n}\bar{\zeta}_b^{\,j_b(\nu+N+m_b+M_b+2)}r_b^{m_b+M_b+\nu+N-\ell_b+2},\end{aligned}$$

where, in the sum above, $j=(j_1,\dots,j_n)$ ranges over the set $\mathfrak{B}$ of (4-3), and $m=(m_2,\dots,m_n)$ and $M=(M_2,\dots,M_n)$ are multi-indices that range over the set $\mathfrak{C}$ of (4-17), and the indices ν and N each go independently from 0 to $K-1$. To find $\tilde{A}(\ell\cdot\beta)$, note that in the sum (4-18), we are considering those terms in which the power of r_1 is $\ell_1\beta_1$ and the power of r_b is $\ell_b\beta_b$ for $2\le b\le n$. Notice that for these powers of r_j, the powers of the ζ_j are each 1. Therefore, comparing the two expressions (4-18) and (4-16) for $\widehat{B}(r_1,\dots,r_n)$, we conclude that for each $\beta\in\mathfrak{G}^*(k)$ we have

$$\tilde{A}(\ell\cdot\beta)=\sum{}'\,\bar{\zeta}_1^{\,j_1(2K-\nu-N-1)}\prod_{b=2}^{n}\bar{\zeta}_b^{\,j_b(\nu+N+m_b+M_b+2)} \tag{4-19}$$

$$=\sum{}'\,1, \tag{4-20}$$

where $\sum'$ denotes a sum extending over the set of indices

$$j=(j_1,j_2,\dots,j_n),\quad m=(m_2,\dots,m_n),\quad M=(M_2,\dots,M_n)\quad\text{and}\quad \nu,N$$

ranging over

$$\begin{cases} j\in\mathfrak{B},\quad m,M\in\mathfrak{C},\\ 0\le\nu,N\le K-1,\\ m_b+M_b+\nu+N+2-\ell_b=\beta_b\ell_b,\quad\text{for each } 2\le b\le n,\\ 2K-\nu-N-\ell_1-1=\beta_1\ell_1.\end{cases}$$

The expression in (4-20) follows from (4-19) since for each such index, the summand is clearly 1. Observe now that in the range of summation described above, the indices $j=(j_1,j_2,\dots,j_n)\in\mathfrak{B}$ (with $\mathfrak{B}$ as in (4-3)) vary freely without any interaction with the other indices m,M,ν,N. Therefore,

$$\tilde{A}(\ell\cdot\beta)=\text{(4-20)}=\sum_{j\in\mathfrak{B}}C(\beta)=|\mathfrak{B}|\cdot C(\beta)=L\cdot C(\beta), \tag{4-21}$$

where as in the statement of the theorem, $L=\prod_{a=1}^{n}\ell_a$, and $C(\beta)$ is the number of solutions in integers $m=(m_2,\dots,m_n)$, $M=(M_2,\dots,M_n)$, ν,N of the system of equations and inequalities given by

$$\begin{cases} 0 \le m_b, M_b \le \ell_b - 1 & \text{for each } 2 \le b \le n, \\ 0 \le \nu, N \le K - 1, & \\ m_b + M_b + \nu + N = \ell_b(\beta_b + 1) - 2 & \text{for each } 2 \le b \le n, \\ \nu + N = 2K - \ell_1(\beta_1 + 1) - 1. & \end{cases}$$

To find $C(\beta)$, we first note that the third equation may be replaced (with the help of the last equation) by the equivalent equation

$$m_b + M_b = \ell_b(\beta_b + 1) + \ell_1(\beta_1 + 1) - 2K - 1 \quad \text{for each } 2 \le b \le n. \tag{4-22}$$

Consequently, the number of solutions $C(\beta)$ of the system can be obtained by multiplying together the number of solutions of

$$\nu + N = 2K - \ell_1(\beta_1 + 1) - 1, \quad 0 \le \nu, N \le K - 1$$

with the number of solutions for each b, with $2 \le b \le n$ to

$$m_b + M_b = \ell_b(\beta_b + 1) + \ell_1(\beta_1 + 1) - 2K - 1, \quad 0 \le m_b, M_b \le \ell_b - 1.$$

To represent these numbers, for integers λ, μ, define $\mathsf{D}_\lambda(\mu)$ to be the number of integer solutions $(x, y) \in \mathbb{Z}^2$ of the system of equations and inequalities:

$$x + y = \mu, \tag{4-23}$$

$$0 \le x \le \lambda - 1, \tag{4-24}$$

$$0 \le y \le \lambda - 1. \tag{4-25}$$

Then clearly we have

$$C(\beta) = \mathsf{D}_K\big(2K - \ell_1(\beta_1 + 1) - 1\big) \cdot \prod_{b=2}^{n} \mathsf{D}_{\ell_b}\big(\ell_b(\beta_b + 1) + \ell_1(\beta_1 + 1) - 2K - 1\big). \tag{4-26}$$

Claim. *The numbers* $\mathsf{D}_\lambda(\mu)$ *are given by the formula* (1-2) *that precedes the statement of Theorem 1.1.*

Indeed, if $\mu \le -1$, then by (4-23), we have $x + y \le -1$. However, from (4-24) and (4-25) in the definition of $\mathsf{D}_\lambda(\mu)$, this is impossible. Hence, $\mathsf{D}_\lambda(\mu) = 0$. Similarly, if $\mu \ge 2\lambda - 1$, then by (4-23), $x + y \ge 2\lambda - 1$. However, from (4-24) and (4-25) in the definition of $\mathsf{D}_\lambda(\mu)$, this is impossible. Hence, $\mathsf{D}_\lambda(\mu) = 0$.

In the other cases, it is easy to enumerate the solutions. If $0 \le \mu \le \lambda - 1$, then

$$\mathsf{D}_\lambda(\mu) = \big|\{(x, \mu - x) : 0 \le x \le \mu\}\big| = \mu + 1,$$

and if $\lambda \le \mu \le 2\lambda - 2$, then

$$\mathsf{D}_\lambda(\mu) = \Big|\big\{(x, \mu - x) : \mu - \lambda + 1 \le x \le \lambda - 1\big\}\Big| = 2\lambda - 1 - \mu,$$

completing the proof of the claim.

From (4-16) and (4-21) we see that

$$\mathbb{B}_{\mathscr{H}(k)}(\phi(z),w)=\widehat{B}(r_1,\dots,r_n)=\frac{1}{\pi^n L^2\Delta}\sum_{\beta\in\mathfrak{G}^*(k)} L\cdot C(\beta)r^{\ell\cdot\beta} = \frac{1}{\pi^n L\Delta}\sum_{\beta\in\mathfrak{G}^*(k)} C(\beta)r^{\ell\cdot\beta}. \tag{4-27}$$

Now

$$\phi(z)=(\phi_1(z),\dots,\phi_n(z))=(z_1^{\ell_1},\dots,z_n^{\ell_n}).$$

Therefore, recalling the definition (4-4), we see that

$$r^{\ell\cdot\beta}=(r_1^{\ell_1})^{\beta_1}\cdots(r_n^{\ell_n})^{\beta_n}=(z_1^{\ell_1}\overline{w}_1)^{\beta_1}\cdots(z_n^{\ell_n}\overline{w}_n)^{\beta_n}=(\phi_1(z)\overline{w}_1)^{\beta_1}\cdots(\phi_n(z)\overline{w}_n)^{\beta_n}.$$

Also, remembering that $\ell_b=K/k_b$ for each b, we have

$$r_b^K=(z_b^{\ell_b})^{k_b}\overline{w}_b^{k_b}=\phi_b(z)^{k_b}\overline{w}_b^{k_b},$$

and

$$r_b^{\ell_b}=z_b^{\ell_b}\overline{w}_b=\phi_b(z)\overline{w}_b.$$

Therefore, recalling the definition (4-7), we have

$$\begin{aligned}\Delta&=\left(\left(\prod_{b=2}^n r_b\right)^K-r_1^K\right)^2\cdot\prod_{b=2}^n(1-r_b^{\ell_b})^2\\&=\left(\left(\prod_{b=2}^n \phi_b(z)^{k_b}\overline{w}_b^{k_b}\right)-\phi_1(z)^{k_1}\overline{w}_1^{k_1}\right)^2\cdot\prod_{b=2}^n(1-\phi_b(z)\overline{w}_b)^2.\end{aligned}$$

Therefore, if we replace $\phi(z)$ by z in the first member of (4-27), we see that the last member is transformed to a function of $(t_1,\dots,t_n)$, where $t_a=z_a\overline{w}_a$. In fact, we get (1-3), thus completing the proof of the result, except that in the numerator of (1-3) we have obtained the polynomial $\sum_{\beta\in\mathfrak{G}^*(k)}C(\beta)t^\beta$ instead of $\sum_{\beta\in\mathfrak{G}}C(\beta)t^\beta$. Therefore, to complete the proof, we need to show that if $\beta\in\mathfrak{G}^*(k)\setminus\mathfrak{G}$ then $C(\beta)=0$. Now for such a β, there exists a $2\le b\le n$ such that $\ell_b=1$ and β_b is either 0 or $2k_b$. First assume that $\beta_b=0$. Then the factor $\mathsf{D}_{\ell_b}(\ell_b(\beta_b+1)+\ell_1(\beta_1+1)-2K-1)$ in the formula (1-4) reduces to $\mathsf{D}_1(\ell_1(\beta_1+1)-2K)$. By the definition (1-2) of D, this is not zero if and only if $\ell_1(\beta_1+1)-2K=0$. However, in the latter case, we have the first factor of (1-4) equal to zero, since it equals $\mathsf{D}_K(-1)$.

In the other case $\beta_b=2k_b=2K$ we see that the factor

$$\mathsf{D}_{\ell_b}\big(\ell_b(\beta_b+1)+\ell_1(\beta_1+1)-2K-1\big)$$

reduces to $\mathsf{D}_1(\ell_1(\beta_1+1))=0$. $\square$

4C. ***Recapturing the special cases $\mathscr{H}(1,-k)$ and $\mathscr{H}(k,-1)$.*** We now show that the results of [Edholm 2016] on explicit Bergman kernels of fat and thin Hartogs triangles are special cases of Theorem 1.1.

4C1. $\mathscr{H}(1,-k)$, $k \geq 1$. We follow the notation used in Theorem 1.1. For $\mathscr{H}(1,-k)$ we have $k_1 = 1$ and $k_2 = k$. Hence $K = \operatorname{lcm}(1,k) = k$ and $L = k$. We then have

$$\mathfrak{G} = \left\{(\beta_1, \beta_2) \in \mathbb{Z}^2 \mid \beta_1 = 0,\ 0 \leq \beta_2 \leq 2k\right\}.$$

For $(0, \beta_2) \in \mathfrak{G}$, we compute $C(0, \beta_2)$, where $0 \leq \beta_2 \leq 2k$. By (1-4), we have

$$C(0,\beta_2) = \mathsf{D}_k(k-1)\mathsf{D}_1(\beta_2 - k).$$

Now from (1-2), we have $\mathsf{D}_k(k-1) = k$ and

$$\mathsf{D}_1(\beta_2 - k) = \begin{cases} 0, & 0 \leq \beta_2 \leq k-1, \\ 1, & \beta_2 = k, \\ 0, & k+1 \leq \beta_2 \leq 2k. \end{cases}$$

Hence for $\beta = (0, \beta_2) \in \mathfrak{G}$, $C(\beta) \neq 0$ if and only if $\beta_2 = k$ and in this case, $C(\beta) = k$. Hence the formula (1-3) gives

$$\mathbb{B}_{\mathscr{H}(1,-k)}(z,w) = \frac{1}{\pi^2 k} \cdot \frac{k t_2^k}{(t_2^k - t_1)^2(1-t_2)^2} = \frac{1}{\pi^2} \cdot \frac{t_2^k}{(t_2^k - t_1)^2(1-t_2)^2},$$

which precisely is the content of [Edholm 2016, Theorem 1.4].

4C2. $\mathscr{H}(k,-1)$, $k \geq 2$. In this case, $k_1 = k$ and $k_2 = 1$. Hence $K = k$ and $L = k$. We then have

$$\mathfrak{G} = \left\{(\beta_1, \beta_2) \in \mathbb{Z}^2 \mid 0 \leq \beta_1 \leq 2k-2,\ 0 \leq \beta_2 \leq 2\right\},$$

and

$$C(\beta) = \begin{cases} \mathsf{D}_k(2k-\beta_1-2)\mathsf{D}_k(\beta_1 - k), & \beta = (\beta_1, 0), \\ \mathsf{D}_k(2k-\beta_1-2)\mathsf{D}_k(\beta_1), & \beta = (\beta_1, 1), \\ \mathsf{D}_k(2k-\beta_1-2)\mathsf{D}_k(\beta_1 + k), & \beta = (\beta_1, 2). \end{cases}$$

We compute the D_k.

$$\mathsf{D}_k(2k-\beta_1-2) = \begin{cases} \beta_1 + 1, & 0 \leq \beta_1 \leq k-1, \\ 2k - \beta_1 - 1, & k \leq \beta_1 \leq 2k-2. \end{cases}$$

$$\mathsf{D}_k(\beta_1 - k) = \begin{cases} 0, & 0 \leq \beta_1 \leq k-1, \\ \beta_1 - k + 1, & k \leq \beta_1 \leq 2k-2. \end{cases}$$

$$\mathsf{D}_k(\beta_1) = \begin{cases} \beta_1 + 1, & 0 \leq \beta_1 \leq k-1, \\ 2k - 1 - \beta_1, & k \leq \beta_1 \leq 2k-2. \end{cases}$$

$$\mathsf{D}_k(\beta_1 + k) = \begin{cases} k - \beta_1 - 1, & 0 \leq \beta_1 \leq k-1, \\ 0, & k \leq \beta_1 \leq 2k-2. \end{cases}$$

Hence,

$$\sum_{\beta\in\mathfrak{G}} C(\beta)t^\beta = \underbrace{\sum_{\beta_1=k}^{2k-2}(2k-\beta_1-1)(\beta_1-k+1)t_1^{\beta_1}}_{\beta_2=0}$$
$$+\underbrace{\left(\sum_{\beta_1=0}^{k-1}(\beta_1+1)^2 t_1^{\beta_1}t_2+\sum_{\beta_1=k}^{2k-2}(2k-\beta_1-1)^2 t_1^{\beta_1}t_2\right)}_{\beta_2=1}$$
$$+\underbrace{\sum_{\beta_1=0}^{k-1}(\beta_1+1)(k-\beta_1-1)t_1^{\beta_1}t_2^2}_{\beta_2=2}.$$

We rewrite the terms corresponding to $\beta_2 = 0, 1, 2$ as follows. In the term for $\beta_2 = 0$, by making the substitution $\ell = \beta_1 - k + 1$, we obtain

$$\left(\sum_{\ell=1}^{k-\ell}(k-\ell)\ell\cdot t_1^{\ell-1}\right)t_1^k.$$

In the first sum of the second term (which corresponds to $\beta_2 = 1$) we make the substitution $\ell = \beta_1 + 1$, which transforms it into $\sum_{\ell=1}^{k}\ell^2\cdot t_1^{\ell-1}t_2$. In the second sum, we make the substitution $\ell = \beta_1 - k + 1$, which transforms it into $\sum_{\ell=1}^{k}(k-\ell)^2\cdot t_1^{k+\ell-1}t_2$. Combining the two we can represent the second term as

$$\sum_{\ell=1}^{k}\ell^2 t_1^{\ell-1}t_2+\sum_{\ell=1}^{k}(k-\ell)^2 t_1^{k+\ell-1}t_2=\left(\sum_{\ell=1}^{k}(\ell^2+(k-\ell)^2)t_1^k)t_1^{\ell-1}\right)t_2.$$

Similarly using the substitution $\ell = \beta_1 + 1$, the last term becomes

$$\sum_{\beta_\ell=0}^{k-1}(\beta_1+1)(k-\beta_1-1)t_1^{\beta_1}t_2^2=\sum_{\ell=1}^{k}\ell(k-\ell)t_1^{\ell-1}t_2^2.$$

Therefore we get the expression for the Bergman kernel for $\mathscr{H}(k,-1)$ as

$$\frac{1}{\pi^2 k}\cdot\frac{1}{(t_2-t_1^k)^2(1-t_2)^2}\cdot\left(\left(\sum_{\ell=1}^{k-1}(k-\ell)\cdot\ell\cdot t_1^{\ell-1}\right)t_1^k+\left(\sum_{\ell=1}^{k}(\ell^2+(k-\ell)^2)t_1^k t_1^{\ell-1}\right)t_2\right.$$
$$\left.+\left(\sum_{\ell=1}^{k}\ell\;(k-\ell)t_1^{\ell-1}\right)t_2^2\right).$$

The above expression is precisely the statement of [Edholm 2016, Theorem 1.2].

Acknowledgements

We gratefully acknowledge the helpful comments of Doron Zeilberger, Jeff McNeal and Luke Edholm. We also thank the anonymous referee for many excellent suggestions which led to significant improvements.

References

[Behnke 1933] H. Behnke, "Zur Theorie der Singularitäten der Funktionen mehrerer komplexen Veränderlichen", *Math. Ann.* **108**:1 (1933), 91–104. MR Zbl

[Bell 1982] S. R. Bell, "The Bergman kernel function and proper holomorphic mappings", *Trans. Amer. Math. Soc.* **270**:2 (1982), 685–691. MR Zbl

[Bell 2005] S. R. Bell, "The Bergman kernel and quadrature domains in the plane", pp. 61–78 in *Quadrature domains and their applications*, edited by P. Ebenfelt et al., Oper. Theory Adv. Appl. **156**, Birkhäuser, Basel, 2005. MR Zbl

[Bremermann 1955] H. J. Bremermann, "Holomorphic continuation of the kernel function and the Bergman metric in several complex variables", pp. 349–383 in *Lectures on functions of a complex variable* (Ann Arbor, MI, 1953), edited by W. Kaplan, Univ. Michigan Press, 1955. MR Zbl

[Catlin 1980] D. Catlin, "Boundary behavior of holomorphic functions on pseudoconvex domains", *J. Differential Geom.* **15**:4 (1980), 605–625. MR Zbl

[Chakrabarti 2019] D. Chakrabarti, "On an observation of Sibony", *Proc. Amer. Math. Soc.* **147**:8 (2019), 3451–3454. MR Zbl

[Chakrabarti and Zeytuncu 2016] D. Chakrabarti and Y. E. Zeytuncu, "L^p mapping properties of the Bergman projection on the Hartogs triangle", *Proc. Amer. Math. Soc.* **144**:4 (2016), 1643–1653. MR Zbl

[Chakrabarti et al. 2019] D. Chakrabarti, L. D. Edholm, and J. D. McNeal, "Duality and approximation of Bergman spaces", *Adv. Math.* **341** (2019), 616–656. MR Zbl

[Chen 2017] L. Chen, "The L^p boundedness of the Bergman projection for a class of bounded Hartogs domains", *J. Math. Anal. Appl.* **448**:1 (2017), 598–610. MR Zbl

[Chen et al. 2020] L. Chen, S. G. Krantz, and Y. Yuan, "L^p regularity of the Bergman projection on domains covered by the polydisc", *J. Funct. Anal.* **279**:2 (2020), art. id. 108522. MR Zbl

[Duren and Schuster 2004] P. Duren and A. Schuster, *Bergman spaces*, Math. Surv. Monogr. **100**, Amer. Math. Soc., Providence, RI, 2004. MR Zbl

[Edholm 2016] L. D. Edholm, "Bergman theory of certain generalized Hartogs triangles", *Pacific J. Math.* **284**:2 (2016), 327–342. MR Zbl

[Edholm and McNeal 2016] L. D. Edholm and J. D. McNeal, "The Bergman projection on fat Hartogs triangles: L^p boundedness", *Proc. Amer. Math. Soc.* **144**:5 (2016), 2185–2196. MR Zbl

[Edholm and McNeal 2017] L. D. Edholm and J. D. McNeal, "Bergman subspaces and subkernels: degenerate L^p mapping and zeroes", *J. Geom. Anal.* **27**:4 (2017), 2658–2683. MR Zbl

[Flajolet and Sedgewick 2009] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge Univ. Press, 2009. MR Zbl

[Hakim and Sibony 1980] M. Hakim and N. Sibony, "Spectre de $A(\bar{\Omega})$ pour des domaines bornés faiblement pseudoconvexes réguliers", *J. Funct. Anal.* **37**:2 (1980), 127–135. MR Zbl

[Huo 2018] Z. Huo, "L^p estimates for the Bergman projection on some Reinhardt domains", *Proc. Amer. Math. Soc.* **146**:6 (2018), 2541–2553. MR Zbl

[Jarnicki and Pflug 2008] M. Jarnicki and P. Pflug, *First steps in several complex variables: Reinhardt domains*, Eur. Math. Soc., Zürich, 2008. MR Zbl

[Krantz 2013] S. G. Krantz, *Geometric analysis of the Bergman kernel and metric*, Grad. Texts in Math. **268**, Springer, 2013. MR Zbl

[Park 2018] J.-D. Park, "The explicit forms and zeros of the Bergman kernel for 3-dimensional Hartogs triangles", *J. Math. Anal. Appl.* **460**:2 (2018), 954–975. MR Zbl

[Range 1986] R. M. Range, *Holomorphic functions and integral representations in several complex variables*, Grad. Texts in Math. **108**, Springer, 1986. MR Zbl

[Shaw 2015] M.-C. Shaw, "The Hartogs triangle in complex analysis", pp. 105–115 in *Geometry and topology of submanifolds and currents* (Norman, OK, 2012/Stillwater, OK, 2013), edited by W. Li and S. W. Wei, Contemp. Math. **646**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Sibony 1975] N. Sibony, "Prolongement des fonctions holomorphes bornées et métrique de Carathéodory", *Invent. Math.* **29**:3 (1975), 205–230. MR Zbl

Received September 6, 2019. Revised January 3, 2020.

DEBRAJ CHAKRABARTI
DEPARTMENT OF MATHEMATICS
CENTRAL MICHIGAN UNIVERSITY
MT PLEASANT, MI
UNITED STATES
chakr2d@cmich.edu

AUSTIN KONKEL
DEPARTMENT OF MATHEMATICS
CENTRAL MICHIGAN UNIVERSITY
MT PLEASANT, MI
UNITED STATES
konke1am@cmich.edu

MEERA MAINKAR
DEPARTMENT OF MATHEMATICS
CENTRAL MICHIGAN UNIVERSITY
MT PLEASANT, MI
UNITED STATES
maink1m@cmich.edu

EVAN MILLER
DEPARTMENT OF MATHEMATICS
CENTRAL MICHIGAN UNIVERSITY
MT PLEASANT, MI
UNITED STATES
mille7em@cmich.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.95

CENTRAL SPLITTING OF MANIFOLDS WITH NO CONJUGATE POINTS

JAMES DIBBLE

Each compact Riemannian manifold with no conjugate points admits a family of functions whose integrals vanish exactly when central Busemann functions split linearly. These functions vanish when all central Busemann functions are sub- or superharmonic. When central Busemann functions are convex or concave, they must be totally geodesic. These yield generalizations of the splitting theorems of O'Sullivan and Eberlein for manifolds with no focal points and, respectively, nonpositive curvature.

1. Introduction

A key insight in Burago and Ivanov's [1994] proof that each Riemannian torus with no conjugate points is flat is that the asymptotic norm of its fundamental group $\pi_1(\mathbb{T}^k) \cong \mathbb{Z}^k$ is Riemannian, in the sense that it is generated by an inner product. In the case of an arbitrary compact Riemannian manifold N whose fundamental group has center $Z(\pi_1(N))$ of rank k, O'Sullivan [1974] showed that, when N has no focal points, it must be foliated by totally geodesic and flat k-dimensional toruses and, moreover, be covered by an isometric product with a flat $\mathbb{T}^k$. This generalized a theorem of Wolf [1964] about manifolds with nonpositive sectional curvature. It was later shown by Eberlein [1982] that each compact manifold with nonpositive sectional curvature and $Z(\pi_1(N))$ of rank k is finitely covered by a diffeomorphic product with $\mathbb{T}^k$. The aim of this paper is to explore conditions, between having no conjugate points and no focal points, that allow for variations on these results.

Associated to each $z_0, z_1 \in Z(\pi_1(N))$ is a function $F_{z_0 z_1} : N \to \mathbb{R}$ defined by (3-1); the linear splitting of central Busemann functions on $\hat{N}$, where $\pi : \hat{N} \to N$ is the universal covering map, is one of a number of conditions equivalent to the vanishing of all $\int_N F_{z_0 z_1}\, d\,\mathrm{vol}_N$.

Theorem 1.1. *Let N be a compact Riemannian manifold with no conjugate points and Z a subgroup of $Z(\pi_1(N))$. Then, the following are equivalent*:

(i) $\int_N F_{z_0 z_1}\, d\,\mathrm{vol}_N = 0$ *for all* $z_0, z_1 \in Z$.

MSC2010: primary 53C20, 53C24; secondary 53C22.
Keywords: no conjugate points, no focal points, Busemann function, center theorem.

(ii) $B(z_0, z_1) = 1/\mathrm{vol}(N) \int_N h(\omega_{z_0}, \omega_{z_1})\, d\,\mathrm{vol}_N$ *for all* $z_0, z_1 \in Z$.

(iii) $h(\omega_{z_0}(x), \omega_{z_1}(x)) = B(z_0, z_1)$ *for all* $x \in N$ *and all* $z_0, z_1 \in Z$.

(iv) $\omega_{\sum_i m_i z_i} = \sum_i m_i \omega_{z_i}$ *for all* $z_i \in Z$ *and all* $m_i \in \mathbb{Z}$.

The notation that appears in Theorem 1.1 is defined in the next two sections. The proof is partly an application of Green's identity.

It follows from the main result of [Dibble 2019a] that any subgroup Z of $Z(\pi_1(N))$ is isomorphic to $\mathbb{Z}^k$ for some $0 \le k \le \dim(N)$. When statements (i)–(iv) in Theorem 1.1 hold, one may define two integral formulations of an inner product that generates the asymptotic norm of Z. Moreover, a number of topological properties, which are known in the case of no focal points, hold (cf. [O'Sullivan 1974]).

Theorem 1.2. *Let* N *be a compact Riemannian manifold with no conjugate points and* Z *a subgroup of* $Z(\pi_1(N))$ *for which statements* (i)–(iv) *in Theorem 1.1 hold. Then, so do the following*:

(a) *The asymptotic norm* $\|\cdot\|_\infty$ *of* Z *with respect to any isomorphism* $Z \to \mathbb{Z}^k$ *is Riemannian.*

(b) *If* $w_1, \dots, w_k$ *generate* Z *and* $H_1, \dots, H_k$ *are corresponding horospheres, then* $\hat{H} = \bigcap_{i=1}^k H_i$ *is a simply connected submanifold of* $\hat{N}$.

(c) $\hat{N}$ *is diffeomorphic to* $\hat{H} \times \mathbb{R}^k$.

(d) *There exists a sequence of normal covering maps*

$$\hat{H} \times \mathbb{R}^k \xrightarrow{\psi_0} N_0 \times \mathbb{T}^k \xrightarrow{\phi_0} N,$$

with respective deck transformation groups $\pi_1(N_0) \times Z$ *and* Γ, *such that* ψ_0 *is a product map*, N_0 *is orientable*, $\pi_1(N_0)$ *is a normal subgroup of* $\pi_1(N)$ *containing the commutator subgroup* $[\pi_1(N), \pi_1(N)]$, *and the sequences*

$$0 \to \pi_1(N_0) \times Z \to \pi_1(N) \to \Gamma \to 0$$

and

$$0 \to (\pi_1(N_0)/[\pi_1(N), \pi_1(N)]) \times Z \to H_1(N, \mathbb{Z}) \to \Gamma \to 0$$

are exact.

Because $[\pi_1(N), \pi_1(N)] \subseteq \pi_1(N_0)$, the covering map ϕ_0 is abelian.

It is clear from (3-1) that the integrals $\int_N F_{z_0 z_1}\, d\,\mathrm{vol}_N$ vanish when central Busemann functions are harmonic. Lemma 5.1 states that sub- or superharmonic central Busemann functions must be harmonic, which implies the following.

Corollary 1.3. *Let* N *be a compact Riemannian manifold with no conjugate points and* Z *a subgroup of* $Z(\pi_1(N))$. *If each Busemann function associated with* Z *is sub- or superharmonic*, *then statements* (i)–(iv) *in Theorem 1.1 and the conclusions of Theorem 1.2 hold.*

Since they must be harmonic, concave or convex central Busemann functions must have vanishing Hessian and therefore be totally geodesic, in the sense that they map geodesics to geodesics. It is well known that Busemann functions are convex when N has no focal points or, more narrowly, nonpositive sectional curvature [Eschenburg 1977]. Thus, the following generalizes the splitting theorems of O'Sullivan and Eberlein.

Theorem 1.4. *Let N be a compact Riemannian manifold with no conjugate points and Z a subgroup of $Z(\pi_1(N))$. If each Busemann function associated with Z is convex or concave, then, in addition to the conclusions of Theorem 1.2, the following hold*:

(a) *$\hat{N}$ is isometric to $\hat{H} \times \mathbb{R}^k$, where Z acts on each $\mathbb{R}^k$-fiber by translations.*

(b) *N is foliated by totally geodesic and flat k-dimensional toruses.*

(c) *There exists a sequence of Riemannian covering maps*

$$\hat{H} \times \mathbb{R}^k \xrightarrow{\psi_1} N_1 \times \mathbb{T}^k \xrightarrow{\phi_1} N,$$

where N_1 is orientable, ψ_1 is a product, ϕ_1 has finitely many sheets, and each map restricts on each $\mathbb{R}^k$- or $\mathbb{T}^k$-fiber to a totally geodesic and locally isometric immersion onto a leaf of the $\mathbb{T}^k$-foliation below.

It is not claimed in Theorem 1.4(c) that $N_1 \times \mathbb{T}^k$ has a product metric, and the statement cannot be strengthened to the isometric splitting of a finite cover, as there are examples in [Lawson and Yau 1972; Eberlein 1980; 1981] of compact $N_1 \times \mathbb{T}^k$ with nonpositive curvature that are not finitely covered by isometric products with flat k-dimensional toruses.

Eberlein additionally proved that every compact manifold having nonpositive sectional curvature and fundamental group with nontrivial center is of a canonical form. This result also generalizes, provided one modifies the definition of a canonical manifold in [Eberlein 1982] appropriately. Let $\hat{H}$ be a complete and simply connected manifold with no conjugate points, Γ_0 a properly discontinuous group of isometries of $\hat{H}$, where $\hat{H}/\Gamma_0$ is compact, and $\rho : \Gamma_0 \to \mathbb{T}^k$ a homomorphism whose kernel contains no nontrivial elements with fixed points. Define an action of Γ_0 on $\hat{H} \times \mathbb{T}^k$ by $\Phi(\xi, \hat{x}) = (\Phi(\hat{x}), \rho(\Phi) \cdot \xi)$ for all $\Phi \in \Gamma_0$, $\xi \in \mathbb{T}^k$, and $\hat{x} \in \hat{H}$. Then, the quotient $(\hat{H} \times \mathbb{T}^k)/\Gamma_0$ is called a *k-canonical* manifold.

Theorem 1.5. *Let N be a compact Riemannian manifold with no conjugate points and Z a subgroup of $Z(\pi_1(N))$. If each Busemann function associated with Z is convex or concave, then N is a k-canonical manifold for $k = \operatorname{rank} Z$. When $Z = Z(\pi_1(N))$, Γ_0 is centerless.*

The proof of Theorem 1.5 will be omitted, as it exactly follows Eberlein's argument, using Theorem 1.4(a) in the place of Lemma 1 of [Eberlein 1982].

2. Preliminaries

Throughout this paper, (N, h) will denote a compact, connected, n-dimensional, and C^r Riemannian manifold for $r \geq 2$. The covering metric on $\hat{N}$ will be denoted by $\hat{h}$. The metric h will have no conjugate points, which by definition means that the exponential map on each tangent space is nonsingular. It follows that, for each $p \in \hat{N}$, $\exp_p : T_p\hat{N} \to \mathbb{R}^n$ is a diffeomorphism. Thus, N is aspherical and, consequently, $\pi_1(N)$ is torsion free [Hurewicz 1995]. It is well known that a complete manifold with nonpositive sectional curvature has no focal points, that a complete manifold with no focal points has no conjugate points, and that neither of these implications is reversible within the space of compact manifolds of any fixed dimension at least two [Gulliver 1975].

For any tangent vector v, denote by γ_v the geodesic $t \mapsto \exp(tv)$. Corresponding to each unit vector $v \in T\hat{N}$ is the Busemann function $b_v : \hat{N} \to \mathbb{R}$ defined by $b_v(x) = \lim_{t\to\infty}[t - d(\gamma_v(t), x)]$. It will be convenient to generalize the idea of a Busemann function to arbitrary tangent vectors by setting

$$b_v = \begin{cases} \|v\| b_{v/\|v\|} & \text{if } v \neq 0, \\ 0 & \text{if } v = 0, \end{cases}$$

for any $v \in T\hat{N}$. It was essentially shown by Busemann [1955] that associated to each $z \in Z(\pi_1(N))$ is a unique constant-length vector field ω_z on N with the property that each $\gamma_{\omega_z(x)}|_{[0,1]}$ is a closed geodesic representing z in $\pi_1(N, x)$. Denote by $\hat{\omega}_z$ the lift of ω_z to $\hat{N}$. A Busemann function b_v is *central* if $v = \hat{\omega}_z(p)$ for some $z \in Z(\pi_1(N))$ and $p \in \hat{N}$. Following the arguments in [Burago and Ivanov 1994] and [Heber 1994], one finds that, for each $z \in Z(\pi_1(N))$, the corresponding central Busemann functions have gradient field $\hat{\omega}_z$. Arbitrary Busemann functions are known only to be C^1, except for manifolds with bounded asymptote, a class which includes those with no focal points, where they are C^2 [Eschenburg 1977]. However, for $z \in Z(\pi_1(N))$, the inverse function theorem implies that ω_z is C^{r-1} and, consequently, that each central Busemann function is C^r [Dibble 2019b].

For each $v \in T\hat{N}$, a horosphere of v is, by definition, a level set of the Busemann function b_v. When $z \in Z(\pi_1(N))$, the horospheres of $\hat{\omega}_z(\hat{x})$ are the leaves of the normal distribution to $\hat{\omega}_z$ for any $\hat{x} \in \hat{N}$ and, consequently, form a C^r foliation of $\hat{N}$ by hypersurfaces. In this way, one may speak of a horosphere H of z itself. For each such H, the normal bundle NH is a trivial line bundle; since no point of $\hat{N}$ is focal to H, the exponential map on NH is a diffeomorphism onto $\hat{N}$. In this way, $\hat{N} \cong H \times \mathbb{R}$.

The action of $\pi_1(N)$ on $\hat{N}$ by deck transformations will be denoted $(\alpha, x) \mapsto \alpha(x)$. The following is a special case of an important lemma of Ivanov and Kapovitch [2014]. However, as they point out, the methods of Croke and Schroeder [1986] suffice to prove the lemma for central elements. In particular, the narrow statement here requires only that the metric be C^2. The more general statement in [Ivanov and Kapovitch 2014] requires C^k regularity for some k depending on n.

Lemma 2.1. *Let $\alpha \in \pi_1(N)$ and $z \in Z(\pi_1(N))$. If γ_v is an axis of z, then $b_v(\alpha(\hat{x})) - b_v(\hat{x})$ is independent of $\hat{x} \in \hat{N}$.*

For any choices of $p, q \in \hat{N}$, $b_{\omega_{z_1}(p)}$ and $b_{\omega_{z_1}(q)}$ differ by a constant. Thus, the function $B(z_0, z_1) = b_{\omega_{z_1}(p)}(z_0(x)) - b_{\omega_{z_1}(p)}(\hat{x})$, defined on $Z(\pi_1(N)) \times Z(\pi_1(N))$, is also independent of the choice of p. Loosely speaking, $B(z_0, z_1)$ is the change in the Busemann functions of z_1 in the direction of z_0.

Lemma 2.1 implies, by the argument in [Ivanov and Kapovitch 2014], the virtual splitting of cyclic subgroups of $Z(\pi_1(N))$.

Theorem 2.2 (Ivanov and Kapovitch). *For each nontrivial $z \in Z(\pi_1(N))$, there exists a finite-index subgroup G of $\pi_1(N)$ isomorphic to a direct product $G' \times \mathbb{Z}$, under which identification z corresponds to a generator of the $\mathbb{Z}$-factor.*

A simple consequence is that, when $Z(\pi_1(N))$ has rank at least k, a finite-index subgroup of $\pi_1(N)$ splits as a product with $\mathbb{Z}^k$.

Corollary 2.3. *If $z_1, \dots, z_k$ are independent elements of $Z(\pi_1(N))$, then there exists a finite-index subgroup G of $\pi_1(N)$ isomorphic to a direct product $G' \times \mathbb{Z}^k$, under which identification the $\mathbb{Z}^k$-factor is generated by elements of the form $z_1^{m_1}, \dots, z_k^{m_k}$ for $m_i \geq 1$.*

Proof. Let $\eta_i = (\pi_i, \sigma_i) : G_i \to G_i' \times \mathbb{Z}$ be the isomorphisms guaranteed by Theorem 2.2, where each G_i' is taken to be a subgroup of G_i on which π_i is the identity. Let $G = \bigcap_{i=1}^k G_i$, $G' = \bigcap_{i=1}^k G_i'$, and $\pi' = \pi_k \circ \dots \circ \pi_1$. Then, $\eta = (\pi', \sigma_1, \dots, \sigma_k) : G \to G' \times \mathbb{Z}^k$ is a homomorphism. Note that

$$\operatorname{Ker} \eta \subseteq \bigcap_{i=1}^k \operatorname{Ker} \sigma_i \subseteq \bigcap_{i=1}^k G_i' = G'.$$

Since each π_i is the identity on G', η must be one-to-one. Since G has finite index, there exist smallest $m_i \geq 1$ such that $z_i^{m_i} \in G$. It follows that $\eta(G)$ is isomorphic to $G' \times \mathbb{Z}^k$ in such a way that the $z_i^{m_i}$ generate the $\mathbb{Z}^k$-factor. □

For any $z \in Z(\pi_1(N))$, write $\|z\|_\infty = B(z, z)^{1/2}$. If Z is any subgroup of $Z(\pi_1(N))$, then $Z \cong \mathbb{Z}^k$ for some $0 \leq k \leq n$ [Dibble 2019a]. With respect to a fixed isomorphism $Z \to \mathbb{Z}^k$, $\|\cdot\|_\infty$ agrees with the asymptotic norm of the orbit metric on Z obtained from its action on $\hat{N}$. That is, if d denotes the induced $\mathbb{Z}^k$-equivariant distance function on $\mathbb{Z}^k$, then, up to the identification of Z with $\mathbb{Z}^k$,

$$\|z\|_\infty = \lim_{m \to \infty} \frac{d(0, mz)}{m},$$

and this extends uniquely to a norm on $\mathbb{R}^k$ (see Section 8.5 of [Burago et al. 2001] for details). Following [Burago and Ivanov 1994], one would like to define an inner product on $\mathbb{R}^k$ that extends B with respect to this identification. In fact, elementary arguments show that B satisfies many of the properties of an inner product: $B(z_0, mz_1) = mB(z_0, z_1) = B(z_0, mz_1)$ for all $m \in \mathbb{Z}$; by Corollary 4.2 of [Ivanov and Kapovitch 2014], $B(z_0 + z_1, z_2) = B(z_0, z_2) + B(z_1, z_2)$; and, significantly, B satisfies the Cauchy–Schwarz inequality,

$$|B(z_0, z_1)| \leq \|z_0\|_\infty \|z_1\|_\infty, \tag{2-1}$$

with equality if and only if z_0 and z_1 are rationally related, in the sense that $mz_0 = \ell z_1$ for some $m, \ell \in \mathbb{Z}$ not both zero. However, it is not clear that, in general, B is symmetric or additive in its second slot. For the purpose of showing that $\|\cdot\|_\infty$ is generated by an inner product, those two conditions are equivalent: additivity would follow from symmetry, and, if additivity held, then the symmetrization of B would extend to an inner product generating $\|\cdot\|_\infty$.

This section ends with a technical lemma about the horospheres of central Busemann functions. Although it is in many ways unclear how well behaved they are, they may in some cases be used to construct fundamental domains of π. A preliminary lemma is first presented.

Lemma 2.4. *If $z \in Z(\pi_1(N))$ is primitive, in the sense that $z \neq mz'$ for all $|m| > 1$ and $z' \in \pi_1(N)$, then, for each $\hat{x} \in \hat{N}$, $\pi \circ \gamma_{\hat{\omega}(\hat{x})}$ is injective on $[0, \|z\|_\infty)$.*

Proof. Assume that $\pi \circ \gamma_{\hat{\omega}(x)}(t_0) = \pi \circ \gamma_{\hat{\omega}(x)}(t_1)$ for $0 \leq t_0 < t_1 < \|z\|_\infty$. Write $T = t_1 - t_0$. Without loss of generality, replace x with $\gamma_{\hat{\omega}(x)}(t_0)$, so that

$$\pi \circ \gamma_{\hat{\omega}(x)}(0) = \pi \circ \gamma_{\hat{\omega}(x)}(T) = \pi \circ \gamma_{\hat{\omega}(x)}(\|z\|_\infty).$$

Since $\mathrm{inj}(\pi(x)) > 0$, $T = (a/b)\|z\|_\infty$ for some relatively prime $a, b \in \mathbb{Z}$ with $|b| > 1$. Because a and b are relatively prime, there exist $m, n \in \mathbb{Z}$ such that $ma = nb + 1$. Since $\pi \circ \gamma_{\hat{\omega}(x)}((ma/b)\|z\|_\infty) = \pi \circ \gamma_{\hat{\omega}(x)}(0)$, $\gamma_{\omega(\pi(x))}|_{[0,(1/b)\|z\|_\infty]}$ is a closed geodesic representing an element w of $\pi_1(N)$ with $z = bw$, which is a contradiction. $\square$

Lemma 2.5. *Let $z \in Z(\pi_1(N))$ be primitive and H be a horosphere of z. Then, there exist disjoint open subsets $U_1, \dots, U_K$ of H such that, for $V_i = U_i \times [0, \|z\|_\infty)$ and $V = \bigcup_{i=1}^K V_i$, π is injective on V and* $\mathrm{vol}_h(N \setminus \pi(V)) = 0$.

Proof. It follows from $\|z\|_\infty$-periodicity that there exists $\varepsilon > 0$ such that, for each $x \in H$, the exponential map is injective on $B(x, \varepsilon) \times [0, \|z_0\|_\infty) \subseteq NH$. By compactness, one may choose a finite subset $\{W_1, \dots, W_K\}$ of $\{B(x, \varepsilon) \mid x \in H\}$ with the property that $\{\pi(W_i \times [0, \|z_0\|_\infty])\}$ is an open cover of N. Let $U_1 = W_1$. The remaining U_i are defined inductively: Suppose $U_1, \dots, U_{i-1}$ have been defined and satisfy $U_j \subseteq W_j$. For each $1 \leq j \leq i-1$, there exist at most finitely many $\alpha_{j,1}, \dots, \alpha_{j,M_j} \in \pi_1(N)$

such that $\alpha_{j,m}(\hat{U}_j \times [0, \|z\|_\infty]) \cap W_i \neq \varnothing$. Let

$$U_i = W_i \setminus \bigcup_{j=1}^{i-1} \bigcup_{m=1}^{M_j} \alpha_{j,m}(\hat{U}_j \times [0, \|z\|_\infty]).$$

The sets $U_1, \dots, U_K$ constructed in this way have the desired properties. $\square$

3. Integral formulations of B

Let $z_0, z_1 \in Z(\pi_1(N))$. The function $F_{z_0 z_1} : N \to \mathbb{R}$ in Theorem 1.2 will be defined by an equivariant construction on $\hat{N}$. Denote by Δ the Laplace–Beltrami operator. Fix $\hat{p} \in \hat{N}$ and $v_i = \hat{\omega}_{z_i}(\hat{p})$. For $x \in N$, let $\hat{x} \in \pi^{-1}(x)$ and $v = \hat{\omega}_{z_0}(\hat{x})$, and define $F_{z_0 z_1}(x)$ to be the average value of $b_{v_1} \Delta b_{v_0}$ along $\gamma_v|_{[0,1]}$ with respect to arclength:

$$F_{z_0 z_1}(x) = \fint_{\gamma_v|_{[0,1]}} b_{v_1} \Delta b_{v_0} \, ds. \tag{3-1}$$

More specifically, $F_{z_0 z_1} = 0$ if z_0 is the identity, and

$$F_{z_0 z_1}(x) = \frac{1}{\|z_0\|_\infty} \int_{\gamma_v|_{[0,1]}} b_{v_1} \Delta b_{v_0} \, ds$$

otherwise. To see that $F_{z_0 z_1}$ is well defined, first recall the following classical result.

Lemma 3.1 (Green's identity). *Let (M, g) be a C^2 Riemannian manifold with boundary ∂M and ν an outward-pointing unit normal vector field along ∂M. If $\phi, \psi : M \to \mathbb{R}$ are C^2 functions, then*

$$\int_M \phi \Delta \psi \, d\mathrm{vol}_M + \int_M g(\nabla \phi, \nabla \psi) \, d\mathrm{vol}_M = \int_{\partial M} \phi g(\nabla \psi, \nu) \, d\mathrm{vol}_{\partial M}.$$

This generalizes in a straightforward way to manifolds with corners. A simple consequence is that, for a central Busemann function b_v, Δb_v measures the change in volume of its horospheres under its gradient flow.

Lemma 3.2. *Let $z \in Z(\pi_1(N))$ be nontrivial. Let b_v be a corresponding central Busemann function and $T \in \mathbb{R}$. For each $t \in \mathbb{R}$, denote by H_t the horosphere of z along which $b_v = t$. Let $U \subset H_T$ be any open set with C^r boundary and compact closure, and let $U_t = U \times [T, T+t]$ with respect to the splitting $\hat{N} \cong H_T \times \mathbb{R}$. Then,*

$$\frac{d}{dt} \mathrm{vol}_{H_{T+t}}(U \times \{T+t\}) = \frac{1}{\|z\|_\infty} \int_{U \times \{T+t\}} \Delta b_v \, d\mathrm{vol}_{H_{T+t}}.$$

Proof. Denote by ν the outward-pointing unit normal vector field along the C^r portion of ∂U_t, so that, except at corners, $\nu = -\hat{\omega}_z / \|z\|_\infty$ along $U \times \{T\}$, $\nu = \hat{\omega}_z / \|z\|_\infty$

along $U \times \{T+t\}$, and $\hat{h}(\nu, \hat{\omega}_z) = 0$ along $\partial U \times [0, \|z\|_\infty]$. By Green's identity,

$$\begin{aligned}\int_{U_t} \Delta b_v \, d\mathrm{vol}_{\hat{N}} &= \int_{\partial U_t} \hat{h}(\hat{\omega}_z, \nu) \, d\mathrm{vol}_{\partial U_t} \\ &= \|z\|_\infty [\mathrm{vol}_{H_{T+t}}(U \times \{T+t\}) - \mathrm{vol}_{H_T}(U \times \{T\})].\end{aligned}$$

It follows from the coarea formula that

$$\int_{U_t} \Delta b_v \, d\mathrm{vol}_{\hat{N}} = \int_T^{T+t} \int_{U \times \{T+s\}} \Delta b_v \, d\mathrm{vol}_{H_{T+s}} \, ds.$$

The result follows immediately. □

If $\hat{x}' \in \pi^{-1}(x)$ and $v' = \hat{\omega}_{z_0}(\hat{x}')$, then $\hat{x}' = \alpha(x)$ for some $\alpha \in \pi_1(N)$. Note that, for all t, $\Delta b_{v_0}(\gamma_v(t)) = \Delta b_{v_0}(\gamma_{v'}(t))$ and, by Lemma 2.1,

$$b_{v_1}(\gamma_v(t)) - b_{v_1}(\gamma_{v'}(t)) = B(\alpha, z_1) \in \mathbb{R}.$$

Moreover, for $T = b_{v_0}(\hat{x})/\|z_0\|_\infty^2$, $U^\varepsilon = B(\gamma_v(-T), \varepsilon) \subseteq H_0$, and $U^\varepsilon_{\|z_0\|_\infty} = U^\varepsilon \times [T, T + \|z_0\|_\infty]$, an application of Lemma 3.2 shows that

$$\begin{aligned}\int_{\gamma_v|_{[0,1]}} \Delta b_{v_0} dt &= \lim_{\varepsilon \to 0} \int_{U^\varepsilon_{\|z_0\|_\infty}} \Delta b_{v_0} \, d\mathrm{vol}_{\hat{N}} \\ &= \lim_{\varepsilon \to 0} \|z_0\|_\infty [\mathrm{vol}_{H_{T+\|z_0\|_\infty}}(U^\varepsilon \times \{T + \|z_0\|_\infty\}) - \mathrm{vol}_{H_T}(U^\varepsilon \times \{T\})] = 0.\end{aligned}$$

Thus,

$$\begin{aligned}\int_{\gamma_{v'}|_{[0,1]}} b_{v_1} \Delta b_{v_0} dt &= \int_{\gamma_v|_{[0,1]}} [b_{v_1} + B(\alpha, z_1)] \Delta b_{v_0} ds \\ &= \int_{\gamma_v|_{[0,1]}} b_{v_1} \Delta b_{v_0} ds + B(\alpha, z_1) \int_{\gamma_v|_{[0,1]}} \Delta b_{v_0} ds = \int_{\gamma_v|_{[0,1]}} b_{v_1} \Delta b_{v_0} ds.\end{aligned}$$

So, $F_{z_0 z_1}$ is well defined. Similarly, $F_{z_0 z_1}$ is independent of the choice of $\hat{p}$, as replacing $\hat{p}$ with $\hat{q}$ induces a constant change in b_{v_1} of $b_{v_1}(\hat{q}) - b_{v_1}(\hat{p})$ while leaving Δb_{v_0} unchanged.

For any $\hat{p}$ as above, let S be a fundamental domain of π constructed using a horosphere H_T of v_0 as in Lemma 2.5, and write $S_t = S \times [t, t + \|z_0\|_\infty]$ for each $t \in [T, T + \|z_0\|_\infty]$. Then,

$$\begin{aligned}\int_N F_{z_0 z_1} \, d\mathrm{vol}_N &= \int_{S_T} \int_{\gamma_{\hat{\omega}_{z_0}(\hat{x})}|_{[0,1]}} b_{v_1} \Delta b_{v_0} ds \; d\mathrm{vol}_{\hat{N}} \\ &= \int_T^{T+\|z_0\|_\infty} \int_{S \times \{t\}} \int_{\gamma_{\hat{\omega}_{z_0}(\hat{x})}|_{[0,1]}} b_{v_1} \Delta b_{v_0} ds \; d\mathrm{vol}_{H_t} \; dt \\ &= \int_T^{T+\|z_0\|_\infty} \int_{S_t} b_{v_1} \Delta b_{v_0} \, d\mathrm{vol}_{\hat{N}} \; dt.\end{aligned} \tag{3-2}$$

Theorem 3.3. *Let $z_0, z_1 \in Z(\pi_1(N))$. Then,*

$$B(z_0, z_1) = \frac{1}{\mathrm{vol}(N)} \int_N [h(\omega_{z_0}, \omega_{z_1}) + F_{z_0 z_1}]\, d\mathrm{vol}_N\,.$$

Proof. By Theorem 2.2, there exist a finite covering map $\psi : \tilde{N} \to N$, with covering metric $\tilde{h}$, and a primitive $\tilde{z}_0 \in Z(\pi_1(\tilde{N}))$ such that $\psi_*(\tilde{z}_0) = z_0$. For each $i = 1, 2$, denote by $\tilde{\omega}_i$ the lift of ω_i to $\tilde{N}$. Then,

$$\frac{1}{\mathrm{vol}(N)} \int_N h(\omega_{z_0}, \omega_{z_1})\, d\mathrm{vol}_N = \frac{1}{\mathrm{vol}_{\tilde{h}}(\tilde{N})} \int_{\tilde{N}} \tilde{h}(\tilde{\omega}_0, \tilde{\omega}_1)\, d\mathrm{vol}_{\tilde{h}}\,.$$

Fix $\hat{x} \in \hat{N}$ and $v_i = \hat{\omega}_{z_i}(\hat{x})$. Let $T \in \mathbb{R}$. For the horospheres H_t of z_0, let $U_1, \ldots, U_K$ be the open subsets of H_T obtained by applying Lemma 2.5 to the covering $\hat{N} \to \tilde{N}$, and write $S = \bigcup_{i=1}^K U_i$ and $S_t = S \times [t, t + \|z\|_\infty]$. If each U_i has C^r boundary, then Green's identity shows that

$$\int_{S_t} \hat{h}(\hat{\omega}_{z_0}, \hat{\omega}_{z_1})\, d\mathrm{vol}_{\hat{N}} = \int_{\partial S_t} b_{v_1} \hat{h}(\hat{\omega}_{z_0}, \nu)\, d\mathrm{vol}_{\partial S_t} - \int_{S_t} b_{v_1} \Delta b_{v_0}\, d\mathrm{vol}_{\hat{N}}\,. \tag{3-3}$$

Applying Lemma 2.1, one has that

$$\begin{aligned}\int_{\partial S_t} b_{v_1} \hat{h}(\hat{\omega}_{z_0}, \nu)\, d\mathrm{vol}_{\partial S_t} &= \|z_0\|_\infty \left[\int_{S \times \{t + \|z_0\|_\infty\}} b_{v_1}\, d\mathrm{vol}_{H_{\|z_0\|_\infty}} - \int_{S \times \{t\}} b_{v_1}\, d\mathrm{vol}_{H_0} \right] \\ &= \|z_0\|_\infty B(z_0, z_1)\, \mathrm{vol}_{H_t}(S \times \{t\}).\end{aligned}$$

Suppose that z_0 is not the identity, as the result holds otherwise. Substituting the above into (3-3), integrating both sides from T to $T + \|z_0\|_\infty$, and using the fact that $\int_{S_t} \hat{h}(\hat{\omega}_{z_0}, \hat{\omega}_{z_1})\, d\mathrm{vol}_{\hat{N}} = \int_{\tilde{N}} \tilde{h}(\tilde{\omega}_0, \tilde{\omega}_1)\, d\mathrm{vol}_{\tilde{h}}$ yields

$$\begin{aligned} B(z_0, z_1) &= \frac{1}{\mathrm{vol}(N)} \int_N h(\omega_{z_0}, \omega_{z_1})\, d\mathrm{vol}_N \\ &\qquad + \frac{1}{\|z_0\|_\infty \mathrm{vol}(N)} \int_T^{T + \|z_0\|_\infty} \int_{S_t} b_{v_1} \Delta b_{v_0}\, d\mathrm{vol}_{\hat{N}}\, dt \\ &= \frac{1}{\mathrm{vol}(N)} \int_N h(\omega_{z_0}, \omega_{z_1})\, d\mathrm{vol}_N + \frac{1}{\mathrm{vol}(N)} \int_N F_{z_0 z_1}\, d\mathrm{vol}_N\,, \end{aligned} \tag{3-4}$$

the latter equality following from (3-2). In the general case, one may apply a similar argument to a union of open sets $U_{i,m} \subseteq U_i$ that have C^r boundary and whose measures converge to that of U_i as $m \to \infty$. $\square$

Corollary 3.4. *Let $z_0, z_1 \in Z(\pi_1(N))$. Then, the following hold*:

(a) $B(z_0, z_1) = B(z_1, z_0)$ *if and only if* $\int_N F_{z_0 z_1}\, d\mathrm{vol}_N = \int_N F_{z_1 z_0}\, d\mathrm{vol}_N$.

(b) $B(z_0, z_1) = 1/\mathrm{vol}(N) \int_N h(\omega_{z_0}, \omega_{z_1})\, d\mathrm{vol}_N$ *if and only if* $\int_N F_{z_0 z_1}\, d\mathrm{vol}_N = 0$.

Let Z be any subgroup of $Z(\pi_1(N))$ of rank k, and fix an isomorphism $D: Z \to \mathbb{Z}^k$. Denote by $e_1, \dots, e_k$ the standard basis for $\mathbb{R}^k$, and let $w_i = D^{-1}(e_i)$. Let $f: \mathbb{T}^k \to N$ be any fixed C^1 map satisfying $f_*(\alpha_i) = w_i$, where $\alpha_1, \dots, \alpha_k$ are generators for $\pi_1(\mathbb{T}^k)$. Since N is aspherical, and therefore an Eilenberg–Mac Lane space, such maps are guaranteed to exist, although in the case of no conjugate points they may be constructed more concretely by iteratively exponentiating around loops (see Section 3.3 of [Dibble 2019b]).

Theorem 3.5. *Let $z_0, z_1 \in Z$ and $\beta_1 = f^*(\omega_{z_1}^\flat)$ (i.e., $\beta_1(v) = h(\omega_{z_1}, f_*(v))$ for all $v \in T\mathbb{T}^k$). Then, $B(z_0, z_1) = \int_{\mathbb{T}^k} \beta_1 \circ D(z_0)\, d\mathrm{vol}_g$ for the standard flat metric g on $\mathbb{T}^k$.*

Proof. Suppose $k \geq 2$, as otherwise the result is clear. Denote by $\phi: \mathbb{R}^k \to \mathbb{T}^k$ the covering map that quotients by $\mathbb{Z}^k$. Let $V = D(z_0) = \sum_{i=1}^m v_i e_i$ be a constant vector field on $\mathbb{T}^k$. Suppose $V \neq 0$, so that some $v_i \neq 0$. Let $u_i = V$ and, for each $j \neq i$, let u_j be the vector in $\mathbb{R}^k$ whose only nonzero entries are $-v_j$ in the i-th component and v_i in the j-th component. Let P be the parallelepiped in $\mathbb{R}^k$ determined by $u_1, \dots, u_k$. Then, P is the union of $v_i^{k-2} \sum_{j=1}^k v_j^2 = v_i^{k-2} \|V\|_{\mathbb{R}^k}^2$ fundamental domains of ϕ. Denote by Q the face of P that contains the origin but not u_i, and, for each $p \in Q$, denote by α_p the geodesic $t \mapsto p + tV/\|V\|_{\mathbb{R}^k}$.

The map f lifts to a map $F: \mathbb{R}^k \to \hat{N}$ such that $f \circ \phi = \pi \circ F$. Let $\hat{V} = \phi^{-1}(V)$, so that

$$h(\omega_{z_1}, f_*(V)) = \hat{h}(\hat{\omega}_{z_1}, F_*(\hat{V})) = \hat{h}(\hat{\omega}_{z_1}, F_*(\alpha_x'(t))).$$

The coarea formula implies that

$$\begin{aligned}
\int_{\mathbb{T}^k} h(\omega_{z_1}, f_*(V))d\,\mathrm{vol}_g &= \frac{1}{v_i^{k-2}\|V\|_{\mathbb{R}^k}} \int_Q \left[\int_0^{\|V\|_{\mathbb{R}^k}} \hat{h}(\hat{\omega}_{z_1}, F_*(\alpha_x'(t)))dt\right] d\mathrm{vol}_Q \\
&= \frac{1}{v_i^{k-2}\|V\|_{\mathbb{R}^k}} \int_Q [b_{z_1} \circ F(x+V) - b_{z_1} \circ F(x)]\, d\mathrm{vol}_Q \\
&= \frac{1}{v_i^{k-2}\|V\|_{\mathbb{R}^k}} \int_Q [b_{z_1}(z_0(F(x))) - b_{z_1}(F(x))]\, d\mathrm{vol}_Q \\
&= \frac{1}{v_i^{k-2}\|V\|_{\mathbb{R}^k}} \int_Q B(z_0, z_1)\, d\mathrm{vol}_Q \\
&= B(z_0, z_1).
\end{aligned}$$

If $V = 0$, then z_0 is the identity, and one obtains the same equality. □

Remark 3.6. For $N = \mathbb{T}^k$ and $z \in \mathbb{Z}^k$, $z = D(\omega_z) = \int_{\mathbb{T}^k} \omega_z\, d\mathrm{vol}_g$, where g is the standard flat metric on $\mathbb{T}^k$. This continuously extends the direction at infinity in [Burago and Ivanov 1994] to the leaves of the Heber foliation [1994] of the unit sphere bundle of $\mathbb{T}^k$.

4. Proof of Theorems 1.1 and 1.2

As before, let Z be a subgroup of $Z(\pi_1(N))$, fix an isomorphism $D : Z \to \mathbb{Z}^k$, and let $w_i = D^{-1}(e_i)$, so that $w_1, \dots, w_k$ generate Z. Write $\mathscr{B}_0 = \{\omega_z \mid z \in Z\}$, and denote by $\mathscr{B}$ the vector space of finite formal linear combinations of elements of $\mathscr{B}_0$ with real coefficients. (For each $\hat{x} \in \hat{N}$, $\mathscr{B}$ may be identified with the space of finite formal combinations of Busemann functions in Z that vanish at $\hat{x}$.) Roughly speaking, one may extend D to a linear direction at infinity $\mathscr{D} : \mathscr{B} \to \mathbb{R}^k$ by setting $\mathscr{D}\left(\sum_i a_i \omega_{z_i}\right) = \sum_i a_i D(z_i)$. In the other direction, there is the inclusion $\iota : \mathbb{Z}^k \to \mathscr{B}$ defined by $\iota\left(\sum_i m_i e_i\right) = \omega_{\sum_i m_i w_i}$.

For each $\zeta_1, \zeta_2 \in \mathscr{B}$ (i.e., $\zeta_i = \sum_j a_{ij} \omega_{z_{ij}}$ for $z_{ij} \in Z$ and $a_{ij} \in \mathbb{R}$), let $\beta_i = f^*(\zeta_i^\flat)$. Define

$$G(\zeta_1, \zeta_2) = \frac{1}{2} \int_{\mathbb{T}^k} [\beta_1 \circ \mathscr{D}(\zeta_2) + \beta_2 \circ \mathscr{D}(\zeta_1)] \, d\mu_g.$$

It is clear that G is symmetric and bilinear. By Corollary 3.4(a) and Theorem 3.5, when (i) holds,

$$G(\omega_{z_0}, \omega_{z_1}) = B(z_0, z_1) = B(z_1, z_0).$$

At the same time, one may define a semi-inner product H on $\mathscr{B}$ by

$$H(\zeta_1, \zeta_2) = \frac{1}{\operatorname{vol}(N)} \int_N h(\zeta_1, \zeta_2) \, d\operatorname{vol}_N .$$

Again assuming (i), Corollary 3.4(b) implies that

$$H(\omega_{z_0}, \omega_{z_1}) = B(z_0, z_1) = B(z_1, z_0).$$

In this case, since they agree on $\mathscr{B}_0 \times \mathscr{B}_0$, G and H define the same semi-inner product on $\mathscr{B}$ and, consequently, the same seminorm. Overloading notation, write $\|\cdot\|_\infty = \|\cdot\|_G = \|\cdot\|_H$.

Lemma 4.1. *Let Z be a subgroup of $Z(\pi_1(N))$ such that $\omega_{\sum_i m_i z_i} = \sum_i m_i \omega_{z_i}$ for all $z_i \in Z$ and all $m_i \in \mathbb{Z}$. Then, for all $z_0, z_1 \in Z$, the following hold*:

(a) $[\omega_{z_0}, \omega_{z_1}] = 2\nabla_{\omega_{z_0}} \omega_{z_1}$ *(i.e.,* $\nabla_{\omega_{z_0}} \omega_{z_1} = -\nabla_{\omega_{z_1}} \omega_{z_0}$*).*

(b) $h([\omega_{z_0}, \omega_{z_1}], \omega_{z_0}) = h([\omega_{z_0}, \omega_{z_1}], \omega_{z_1}) = 0.$

(c) $h(\omega_{z_0}(x), \omega_{z_1}(x)) = B(z_0, z_1)$ *for all* $x \in N$.

Proof. One has that

$$\begin{aligned} 0 = \nabla_{\omega_{z_0+z_1}} \omega_{z_0+z_1} &= \nabla_{\omega_{z_0}} \omega_{z_0} + \nabla_{\omega_{z_0}} \omega_{z_1} + \nabla_{\omega_{z_1}} \omega_{z_0} + \nabla_{\omega_{z_1}} \omega_{z_1} \\ &= \nabla_{\omega_{z_0}} \omega_{z_1} + \nabla_{\omega_{z_1}} \omega_{z_0}, \end{aligned}$$

so $\nabla_{\omega_{z_0}} \omega_{z_1} = -\nabla_{\omega_{z_1}} \omega_{z_0}$. This proves (a). Moreover,

$$0 = \omega_{z_1}[h(\omega_{z_0}, \omega_{z_0})] = 2h(\nabla_{\omega_{z_1}} \omega_{z_0}, \omega_{z_0}) = -2h(\nabla_{\omega_{z_0}} \omega_{z_1}, \omega_{z_0}) = -2\omega_{z_0}[h(\omega_{z_1}, \omega_{z_0})].$$

One may deduce (b) from the first three equalities above. Let $\hat{x} \in \hat{N}$, and write $\gamma = \gamma_{\hat{\omega}_{z_0}(\hat{x})}$. Since $h(\hat{\omega}_{z_1}, \hat{\omega}_{z_0})$ is constant along γ, one finds that

$$\begin{aligned} B(z_0, z_1) &= b_{z_1}(z_0(x)) - b_{z_1}(x) \\ &= \int_0^1 h((\hat{\omega}_{z_0} \circ \gamma)(t), (\hat{\omega}_{z_1} \circ \gamma)(t))\, dt = h(\hat{\omega}_{z_0}(x), \hat{\omega}_{z_1}(x)), \end{aligned}$$

which proves (c). □

It is now possible to prove Theorem 1.1. Statements (i) and (ii) are equivalent by Corollary 3.4(b). By Lemma 4.1(c), (iv) implies (iii), and it is clear that (iii) implies (ii). Therefore, the proof can be completed by showing that (i) implies (iv).

Suppose (i) holds. Write $\alpha = \omega_{\sum_i m_i z_i}$, $\beta = \sum_i m_i \omega_{z_i}$, and $\zeta = \alpha - \beta$. Then,

$$\begin{aligned} \|\zeta\|_\infty^2 &= G(\zeta, \zeta) \\ &= \int_{\mathbb{T}^k} h(\zeta, D(\zeta))\, d\mathrm{vol}_g = \int_{\mathbb{T}^k} h(\zeta, 0)\, d\mathrm{vol}_g = 0, \end{aligned}$$

which means that

$$0 = \mathrm{vol}(N) H(\zeta, \zeta) = \int_N h(\zeta, \zeta)\, d\mathrm{vol}_h = \int_N [\|\alpha\|_h^2 + \|\beta\|_h^2 - 2h(\alpha, \beta)]\, d\mathrm{vol}_h\,.$$

Thus,

$$\int_N h(\alpha, \beta)\, d\mathrm{vol}_h = \int_N \tfrac{1}{2}(\|\alpha\|_h^2 + \|\beta\|_h^2)\, d\mathrm{vol}_h\,. \tag{4-1}$$

By the Cauchy–Schwarz inequality,

$$h(\alpha, \beta) \leq \|\alpha\|_h \|\beta\|_h,$$

so

$$\int_N (\|\alpha\|_h^2 + \|\beta\|_h^2)\, d\mathrm{vol}_h \leq \int_N 2\|\alpha\|_h^2 \|\beta\|_h^2\, d\mathrm{vol}_h\,.$$

At the same time, $2\|\alpha\|_h \|\beta\|_h \leq \|\alpha\|_h^2 + \|\beta\|_h^2$. Thus,

$$\int_N 2\|\alpha\|_h^2 \|\beta\|_h^2\, d\mathrm{vol}_h = \int_N (\|\alpha\|_h^2 + \|\beta\|_h^2)\, d\mathrm{vol}_h\,,$$

which implies that $\|\alpha\|_h = \|\beta\|_h$ on N. Substituting into (4-1), one finds that

$$\int_N h(\alpha, \beta)\, d\mathrm{vol}_h = \int_N \|\alpha\|_h^2 \|\beta\|_h^2\, d\mathrm{vol}_h$$

and, consequently, that $h(\alpha, \beta) = \|\alpha\|_h \|\beta\|_h$ on N. It follows that $\alpha = \beta$, which is (iv). This proves Theorem 1.1.

Theorem 1.2(a) is a consequence of the following lemma and the fact that, when (i)–(iv) hold, G agrees with H.

Lemma 4.2. *Let Z be a subgroup of $Z(\pi_1(N))$ of rank k. If G is positive semidefinite, then*

$$G(\iota(z_0),\iota(z_1)) = G(\omega_{z_0},\omega_{z_1}) = \tfrac{1}{2}[B(z_0,z_1)+B(z_1,z_0)]$$

extends to an inner product on $\mathbb{R}^k$ that induces the asymptotic norm $\|\cdot\|_\infty$ of Z with respect to the isomorphism $D : Z \to \mathbb{Z}^k$.

Proof. Since G is positive semidefinite, it is a semi-inner product and induces a seminorm $\|\cdot\|_G$ on $\mathscr{B}$. If $\zeta \in \operatorname{Ker}(\mathscr{D})$, then it follows from the definition of G that $\|\zeta\|_G = 0$. Conversely, suppose $\|\zeta\|_G = 0$. Since $\|\iota(\,\cdot\,)\|_G = \|\cdot\|_\infty$, there exists $c > 0$ such that $\|\iota(x)\|_G \ge c\|x\|_{\mathbb{R}^k}$ for all $x \in \mathbb{Z}^k$. Write $W = \operatorname{span}\{\omega_{w_1},\dots,\omega_{w_k}\}$. Since $\mathscr{D}|_W : W \to \mathbb{R}^k$ is an invertible linear map, there exists $C > 0$ such that $\|w\|_G \le C\|\mathscr{D}(w)\|_{\mathbb{R}^k}$ for all $w \in W$. There exist $K_i \to \infty$ and $v_i \in W$ with $\|\mathscr{D}(v_i)\|_{\mathbb{R}^k} \le 1$ such that $\mathscr{D}(K_i\zeta + v_i) \in \mathbb{Z}^k$. Since $\mathscr{D}\circ\iota$ is the identity map on $\mathbb{Z}^k$, one has that

$$\mathscr{D}(K_i\zeta + v_i - \iota\circ\mathscr{D}(K_i\zeta+v_i)) = 0,$$

and, consequently,

$$|\,\|K_i\zeta+v_i\|_G - \|\iota\circ\mathscr{D}(K_i\zeta+v_i)\|_G\,| \le \|K_i\zeta+v_i-\iota\circ\mathscr{D}(K_i\zeta+v_i)\|_G = 0.$$

It follows that

$$c\|\mathscr{D}(K_i\zeta+v_i)\|_{\mathbb{R}^k} \le \|\iota\circ\mathscr{D}(K_i\zeta+v_i)\|_G = \|K_i\zeta+v_i\|_G \le \|K_i\zeta\|_G + \|v_i\|_G \le C.$$

Therefore, $\|\mathscr{D}(\zeta)\|_{\mathbb{R}^k} \le (1+C/c)/K_i \to 0$ as $i \to \infty$ and, consequently, $\mathscr{D}(\zeta) = 0$. Thus, $\operatorname{Ker}(\mathscr{D}) = \{\zeta \in \mathscr{B} \mid \|\zeta\|_G = 0\}$.

Let $\tilde{\mathscr{B}}$ be the vector space obtained as the quotient of $\mathscr{B}$ by $\operatorname{Ker}(\mathscr{D})$, $\tilde{G}$ the corresponding inner product on $\mathscr{B}$, and $\|\cdot\|_{\tilde{G}}$ the induced norm. The map $\mathscr{D}$ descends to a map $\tilde{\mathscr{D}}$ on $\tilde{\mathscr{B}}$ with trivial kernel. By construction, $\tilde{\mathscr{D}}([\omega_{w_i}]) = e_i$ for each i, so $\tilde{\mathscr{D}}$ is surjective. From this, we have that $\tilde{\mathscr{D}}$ is a linear isomorphism, and $\tilde{\mathscr{D}}^{-1}(\mathbb{Z}^k) = \left\{\sum_i m_i[\omega_{w_i}] \mid m_i \in \mathbb{Z}\right\} = \{[\omega_z] \mid z \in Z\}$ projects to a dense subset of the unit sphere in $\tilde{\mathscr{B}}$. Since

$$\|[\omega_z]\|_{\tilde{G}}^2 = \tilde{G}([\omega_z],[\omega_z]) = G(\omega_z,\omega_z) = B(z,z) = \|z\|_\infty^2 = \|\tilde{\mathscr{D}}([\omega_z])\|_\infty^2$$

for all $z \in Z$, it follows by continuity that $\tilde{\mathscr{D}} : (\tilde{\mathscr{B}}, \|\cdot\|_{\tilde{G}}) \to (\mathbb{R}^k, \|\cdot\|_\infty)$ is an isomorphism of normed spaces. Consequently, $\tilde{\mathscr{D}}_*(\tilde{G})$ is an inner product on $\mathbb{R}^k$ that induces $\|\cdot\|_\infty$. On $\mathbb{Z}^k$, $\tilde{\mathscr{D}}_*(\tilde{G}) = G\circ\iota$. $\square$

By the following lemma, when (i)–(iv) hold, the involutive C^{r-1} distribution $\bigcap_{i=1}^k \hat{\omega}_{w_i}^{\perp}$ has constant dimension $n-k$, and, by Frobenius's theorem, it foliates $\hat{N}$ by C^r $(n-k)$-dimensional submanifolds, each of which is contained in an intersection of the form $\bigcap_{i=1}^k H_i$ for horospheres $H_1,\dots,H_k$ of $w_1,\dots,w_k$, respectively.

Lemma 4.3. *Let Z be a subgroup of $Z(\pi_1(N))$ of rank k for which statements* (i)–(iv) *of Theorem 1.1 hold. Then, for each $\hat{x} \in \hat{N}$, the set $\{\hat{\omega}_{w_1}(\hat{x}), \ldots, \hat{\omega}_{w_k}(\hat{x})\}$ is linearly independent.*

Proof. Suppose that, for some $c_1, \ldots, c_k \in \mathbb{R}$, $\sum_{i=1}^{k} c_i \omega_{w_i}(\hat{x}) = 0$. Then,

$$0 = \left\| \sum_{i=1}^{k} c_i \omega_{w_i}(\hat{x}) \right\|_h^2 = \sum_{i,j=1}^{k} c_i c_j h(\omega_{w_i}(\hat{x}), \omega_{w_j}(\hat{x}))$$
$$= \sum_{i,j=1}^{k} c_i c_j h(\omega_{w_i}(\hat{y}), \omega_{w_j}(\hat{y})) = \left\| \sum_{i=1}^{k} c_i \omega_{w_i}(\hat{y}) \right\|_h^2$$

for all $\hat{y} \in N$. Thus, $\sum_{i=1}^{k} c_i \omega_{w_i} = 0$ on N. Therefore,

$$0 = \left\| \sum_{i=1}^{k} c_i \omega_{w_i} \right\|_H = \left\| \sum_{i=1}^{k} c_i \omega_{w_i} \right\|_G$$

and, consequently,

$$0 = \mathscr{D}\left(\sum_{i=1}^{k} c_i \omega_{w_i} \right) = \sum_{i=1}^{k} c_i e_i .$$

This forces $c_i = 0$ for all i. □

The next lemma implies that each intersection $\bigcap_{i=1}^{k} H_i$ is contained in one leaf of the foliation and, as a consequence, its leaves are exactly those intersections.

Lemma 4.4. *Let Z be a subgroup of $Z(\pi_1(N))$ for which statements* (i)–(iv) *in Theorem 1.1 hold. Fix $\hat{x} \in \hat{N}$, and, for each i, denote by H_i the horosphere of w_i through $\hat{x}$. Then, $\hat{H} = \bigcap_{i=1}^{k} H_i$ is connected.*

Proof. The proof is by induction. Write $v_j = \hat{\omega}_{w_j(\hat{x})}$. Since b_{v_1} has nonzero gradient, the projection $\hat{N} \to \hat{H}_{z_1}$ along the integral curves of b_{v_1} is a continuous surjection, which implies that H_1 is connected. If the result holds for $\bigcap_{i=1}^{j} H_i$, then, by Lemma 4.1(c), the restriction of $b_{v_{j+1}}$ to $\bigcap_{i=1}^{j} H_i$ has nonzero gradient, so $\bigcap_{i=1}^{j+1} H_i$ is similarly connected. □

Define a map $\Psi : \hat{H} \times \mathbb{R}^k \to \hat{N}$ in the following way: For each $(\hat{y}, s_1, \ldots, s_k)$ in $\hat{H} \times \mathbb{R}^k$, let $\hat{x}_0 = \hat{y}$, and inductively define $\hat{x}_{i+1} = \gamma_{\hat{\omega}_{i+1}(\hat{x}_i)}(s_{i+1}) = \exp_{\hat{x}_i}(s_{i+1}\hat{\omega}_{i+1}(\hat{x}_i))$. Let $\Psi(\hat{y}, s_1, \ldots, s_k) = \hat{x}_k$. The splittings $\hat{N} \cong H_i \times \mathbb{R}$ ensure that $D\Psi$ is nonsingular, and, consequently, the inverse function theorem implies that Ψ is a local diffeomorphism.

Lemma 4.5. Ψ *is proper.*

Proof. It follows from Corollary 3.4 that $[B(w_i, w_j)]$ is a positive-definite symmetric matrix, so $\sum_{i=1}^{k} |b_i \circ \Psi(\hat{y}, m_1, \dots, m_k)| \to \infty$ uniformly as $\sum_{i=1}^{k} |m_i| \to \infty$ for $m_i \in \mathbb{Z}$. By periodicity, there exists a Lipschitz constant, uniform in $\hat{y} \in \hat{N}$, for all maps of the form $\Psi(\hat{y}, \cdot)$. Thus, $\sum_{i=1}^{k} |b_i \circ \Psi(\hat{y}, s_1, \dots, s_k)| \to \infty$ uniformly as $\sum_{i=1}^{k} |s_i| \to \infty$. Let $X \subset \hat{N}$ be compact. Then $\sum_{i=1}^{k} |b_i|$ is bounded on X, which implies that the projection of $\Psi^{-1}(X)$ onto the $\mathbb{R}^k$-factor is compact. At the same time, the projection of $\Psi^{-1}(X)$ onto the $\hat{H}$-factor is contained within a closed ball around X, so it too is compact. Thus, $\Psi^{-1}(X)$ is compact. □

Hadamard's global inverse function theorem implies that Ψ is a diffeomorphism, which proves parts (c) and, in turn, (b) of Theorem 1.2.

For a fixed $\hat{x} \in \hat{N}$, let $\hat{H}$ be the intersection of horospheres $\bigcap_{i=1}^{k} H_i$ containing $\hat{x}$. Following the argument in [O'Sullivan 1974], set $G_0 = \{g \in \pi_1(N) \mid g(\hat{x}) \in \hat{H}\}$. By Lemma 2.1, G_0 is normal, contains the commutator subgroup $[\pi_1(N), \pi_1(N)]$, and acts freely and properly discontinuously on $\hat{H}$ by isometries. Note that the subgroup G_0' of G_0 consisting of orientation-preserving elements has all of those same properties, and that the quotient space $N_0 = \hat{H}/G_0'$ is orientable. The subgroup G generated by G_0' and Z is isomorphic to $G_0' \times Z$, and the quotient space $\hat{H}/G$ is diffeomorphic to $N_0 \times \mathbb{T}^k$. One obtains normal covering maps

$$\hat{H} \times \mathbb{R}^k \xrightarrow{\psi_0} N_0 \times \mathbb{T}^k \xrightarrow{\phi_0} N.$$

Note that ψ_0 may be assumed to be a product and that $G_0' = \pi_1(N_0)$. Let Γ denote the quotient group $\pi_1(N)/G$. By construction,

$$0 \to \pi_1(N_0) \times Z \to \pi_1(N) \to \Gamma \to 0$$

is a short exact sequence, which implies that

$$0 \to (\pi_1(N_0)/[\pi_1(N), \pi_1(N)]) \times Z \to H_1(N, \mathbb{Z}) \to \Gamma \to 0$$

is as well. This proves Theorem 1.2(d).

5. Proof of Theorem 1.4

A straightforward volume argument shows that, whenever a central Busemann function is everywhere subharmonic or everywhere superharmonic, it must be harmonic.

Lemma 5.1. *Let N be a compact Riemannian manifold with no conjugate points and b_v a central Busemann function on $\hat{N}$. If b_v is either sub- or superharmonic, then it is harmonic.*

Proof. If b_v vanishes identically, the result is clear. Suppose b_v corresponds to a nontrivial element of $Z(\pi_1(N))$. Let T, U, and, for each $t \in \mathbb{R}$, H_t be as in

Lemma 3.2. Then,

$$\frac{d}{dt}\,\mathrm{vol}_{H_{T+t}}(U\times\{T+t\})=\frac{1}{\|z\|_\infty}\int_{U\times\{T+t\}}\Delta b_v\,d\,\mathrm{vol}_{H_{T+t}}.$$

Since the right-hand side is either nonnegative or nonpositive, $\mathrm{vol}_{H_{T+t}}(U\times\{T+t\})$ is $\|z\|_\infty$-periodic, and U and T are arbitrary, Δb_v must vanish identically. □

If b_v is a convex or concave central Busemann function, then it is subharmonic or, respectively, superharmonic. By Lemma 5.1, it is therefore harmonic. Since every convex or concave harmonic function has vanishing Hessian, one obtains the following.

Lemma 5.2. *Let b_v be a central Busemann function on $\hat{N}$. If b_v is convex or concave, then it is totally geodesic.*

For the remainder of this section, the hypotheses of Theorem 1.4 will be assumed. In this case, the conclusions of Theorem 1.1 hold. As before, let $w_1,\dots,w_k$ generate Z.

Fix $\hat{x}\in N$, and, for each i, let $v_i=\hat{\omega}_{w_i}(\hat{x})$. Since b_{v_i} is totally geodesic, each of its horospheres is a totally geodesic submanifold of $\hat{N}$. This and the fact that $\nabla_{\hat{\omega}_{w_i}}\hat{\omega}_{w_i}=0$ imply that $\hat{\omega}_{w_i}$ is parallel. Thus, the k-dimensional distribution $\mathscr{S}=\mathrm{span}\{\hat{\omega}_{w_1},\dots,\hat{\omega}_{w_1}\}$ is involutive and, by Frobenius's theorem, foliates $\hat{N}$ by k-dimensional submanifolds. Because $\mathscr{S}$ restricts on each leaf of this foliation to a globally parallel orthonormal frame, its leaves are flat and totally geodesic Euclidean spaces. It follows from de Rham's splitting theorem [1952] that $\hat{N}$ is isometric to $\hat{H}\times\mathbb{R}^k$ for any intersection of horospheres $\hat{H}$ as in Theorem 1.1. Note that Z acts on each $\mathbb{R}^k$-fiber by translations, which completes the proof of Theorem 1.4(a). The proof of part (b) follows exactly as in [O'Sullivan 1974]: the distribution $\mathscr{D}$ projects to a parallel distribution on N, the leaves of which are compact, flat, totally geodesic, and without holonomy; consequently, they must be toruses.

Unlike in the case of nonpositive curvature, it is not known that a nontrivial subgroup of the fundamental group of a compact manifold with no conjugate points has center consisting of its Clifford translations (cf. Proposition 2.3 of [Chen and Eberlein 1980] and Lemma 3 of [Eberlein 1982]). However, Theorem 1.4(c) may still be proved using Theorem 2.2 and the argument in Remark 1 of [Eberlein 1982]. Corollary 2.3 implies that N is finitely covered by a manifold $\tilde{N}$ with $\pi_1(\tilde{N})\cong G'\times\mathbb{Z}^k$ for a subgroup G' of $\pi_1(N)$, where $w_1^{m_1},\dots,w_k^{m_k}$ generate the $\mathbb{Z}^k$-factor. Without loss of generality, one may replace G' with its orientation-preserving elements. By Theorem 1.4(a) and Lemma 2.1, elements of $\pi_1(N)$ preserve the horizontal and vertical foliations of $\hat{H}\times\mathbb{R}^k$. Thus, G' acts on $\hat{H}$. If this action were not free and properly discontinuous, one could construct, by an argument similar to the proof of

Lemma 2.4, elements of $\pi_1(\tilde{N})$ of arbitrarily small displacement, contradicting the compactness of N. Thus, the quotient space $N_1 = \hat{H}/G'$ is an orientable manifold.

An elementary argument shows that $\tilde{N}$ has the structure of a $\mathbb{T}^k$-bundle, with totally geodesic and flat $\mathbb{T}^k$-fibers, over N_1. The proof of Lemma 4 of [Eberlein 1982], except with C^r regularity, passes through exactly as written. Thus, one may construct a C^r section of the $\mathbb{T}^k$-bundle as in Remark 1 of [Eberlein 1982], which then implies that $\tilde{N}$ is diffeomorphic to $N_1 \times \mathbb{T}^k$ and that the covering map $N_1 \times \mathbb{T}^k \to N$ restricts on each $\mathbb{T}^k$-fiber to a totally geodesic and locally isometric immersion onto a leaf of the $\mathbb{T}^k$-foliation of N. This proves part (c).

6. Additional results and questions

This work was largely motivated by the question of whether the center theorem of Wolf [1964] and O'Sullivan [1974] generalizes to the case of no conjugate points.

Question 6.1. Let N be a compact Riemannian manifold with no conjugate points with $Z(\pi_1(N))$ of rank k.

(a) Is N foliated by totally geodesic and flat k-toruses?

(b) Does the universal covering space $\hat{N}$ split isometrically as $\hat{H} \times \mathbb{R}^k$?

As a starting point, one might try to show that the asymptotic norm on $Z(\pi_1(N))$ is Riemannian.

It is natural to look for geometric conditions, weaker than having no focal points, that ensure the linear splitting of central Busemann functions in Theorem 1.1. The following argument shows that it suffices to control the asymptotic geometry of distance spheres on $\hat{N}$: Fix a nontrivial $z \in Z(\pi_1(N))$. The C^{r-1} regularity of ω_z guarantees that the stable Jacobi tensor along each $\gamma_{\hat{\omega}_z(\hat{x})}$ is bounded. The argument in Proposition 5 of [Eschenburg 1977] shows that, at each point $\hat{x} \in \hat{N}$, the second fundamental forms of the distance spheres $\partial B(\gamma_v z(-t), t)$, for $v = \hat{\omega}_z(\hat{x})/\|\hat{\omega}_z(\hat{x})\|$, converge to that of the horosphere of z through $\hat{x}$.

For any unit vector $v \in T_{\hat{x}}\hat{N}$, define the distance function $\rho_{\hat{x}}(\,\cdot\,) = d_{\hat{N}}(\,\cdot\,, \hat{x})$ to be *asymptotically subharmonic in the direction of* v (respectively, *superharmonic*) if $\liminf_{t\to\infty} \Delta\rho_{\hat{x}}(\gamma_v(t)) \geq 0$ (respectively, $\limsup_{t\to\infty} \Delta\rho_{\hat{x}}(\gamma_v(t)) \leq 0$). On $\hat{N} \setminus \{\hat{x}\}$, let $\kappa_{\hat{x}}^-$ and $\kappa_{\hat{x}}^+$ be the functions equal to the smallest and, respectively, largest eigenvalues of $\operatorname{Hess} \rho_{\hat{x}}$, and similarly define $\rho_{\hat{x}}$ to be *asymptotically convex in the direction of* v (respectively, *concave*) if $\liminf_{t\to\infty} \kappa_{\hat{x}}^-(\gamma_v(t)) \geq 0$ (respectively, $\limsup_{t\to\infty} \kappa_{\hat{x}}^+(\gamma_v(t)) \leq 0$). Lemmas 5.1 and 5.2 imply the following.

Proposition 6.2. *Let $z \in Z(\pi_1(N))$. Then, the following hold*:

(a) *If, for all $\hat{x} \in \hat{N}$, $\rho_{\hat{x}}$ is asymptotically subharmonic in the direction of $v = \hat{\omega}_z(\hat{x})$, then, for each such v, b_v is harmonic.*

(b) *If, for all* $\hat{x} \in \hat{N}$, $\rho_{\hat{x}}$ *is asymptotically convex in the direction of* $v = \hat{\omega}_z(\hat{x})$, *then, for each such* v, b_v *is totally geodesic.*

Similar results hold for asymptotically superharmonic or asymptotically concave central Busemann functions.

Question 6.3. Are there natural geometric conditions, other than having no focal points, that ensure central Busemann functions are asymptotically subharmonic or asymptotically convex?

Natural conditions that have been considered are Ricci curvature bounds, the effects and limitations of which are discussed thoroughly in [Eschenburg and O'Sullivan 1980].

It's also unclear whether the Heber foliation of the unit sphere bundle of a torus with no conjugate points [Heber 1994] generalizes to the case where Busemann functions in Z split linearly.

Question 6.4. Let N be a compact Riemannian manifold with no conjugate points and Z a subgroup of $Z(\pi_1(N))$ such that statements (i)–(iv) in Theorem 1.1 hold. Write

$$\mathscr{W} = \left\{ \sum_{i=1}^{k} a_i \hat{\omega}_{w_i} \mid a_i \in \mathbb{R} \right\},$$

and, for any $\hat{x} \in \hat{N}$, write

$$\mathscr{V}_{\hat{x}} = \left\{ \nabla b_v \mid v = \sum_{i=1}^{k} a_i \hat{\omega}_{w_i}(\hat{x}) \text{ for } a_i \in \mathbb{R} \right\}.$$

Is $\mathscr{V}_{\hat{x}} = \mathscr{W}$?

An elementary argument shows that, for each nonzero $a = (a_1, \ldots, a_k) \in \mathbb{R}^k$, there is a unique C^r horofunction h_a such that any sequence of rational directions that converge to a induces a sequence of Busemann functions that vanish at $\hat{x}$ and converge uniformly on compact sets to h_a. Moreover, the gradient flow of h_a is through geodesics, and its level sets are the integral submanifolds of the codimension-one involutive distribution

$$\left(\sum_{i=1}^{k} a_i \hat{\omega}_{w_i} \right)^{\perp}.$$

However, without the linear divergence of geodesics in [Heber 1994], it's not clear that, when a is irrational, $h_a = b_v$ for $v = \sum_{i=1}^{k} a_i \hat{\omega}_{w_i}(\hat{x})$.

References

[Burago and Ivanov 1994] D. Burago and S. Ivanov, "Riemannian tori without conjugate points are flat", *Geom. Funct. Anal.* **4**:3 (1994), 259–269. MR Zbl

[Burago et al. 2001] D. Burago, Y. Burago, and S. Ivanov, *A course in metric geometry*, Graduate Studies in Mathematics **33**, American Mathematical Society, Providence, RI, 2001. MR Zbl

[Busemann 1955] H. Busemann, *The geometry of geodesics*, Academic Press, New York, 1955. MR Zbl

[Chen and Eberlein 1980] S. S. Chen and P. Eberlein, "Isometry groups of simply connected manifolds of nonpositive curvature", *Illinois J. Math.* **24**:1 (1980), 73–103. MR Zbl

[Croke and Schroeder 1986] C. B. Croke and V. Schroeder, "The fundamental group of compact manifolds without conjugate points", *Comment. Math. Helv.* **61**:1 (1986), 161–175. MR Zbl

[Dibble 2019a] J. Dibble, "Abelian subgroups of the fundamental group of a space with no conjugate points", preprint, 2019. To appear in *Groups Geom. Dyn.* arXiv

[Dibble 2019b] J. Dibble, "Totally geodesic maps into manifolds with no focal points", *Bull. Lond. Math. Soc.* **51**:3 (2019), 443–458. MR Zbl

[Eberlein 1980] P. Eberlein, "Lattices in spaces of nonpositive curvature", *Ann. of Math.* (2) **111**:3 (1980), 435–476. MR Zbl

[Eberlein 1981] P. Eberlein, "Geodesic rigidity in compact nonpositively curved manifolds", *Trans. Amer. Math. Soc.* **268**:2 (1981), 411–443. MR Zbl

[Eberlein 1982] P. Eberlein, "A canonical form for compact nonpositively curved manifolds whose fundamental groups have nontrivial center", *Math. Ann.* **260**:1 (1982), 23–29. MR Zbl

[Eschenburg 1977] J.-H. Eschenburg, "Horospheres and the stable part of the geodesic flow", *Math. Z.* **153**:3 (1977), 237–251. MR Zbl

[Eschenburg and O'Sullivan 1980] J.-H. Eschenburg and J. J. O'Sullivan, "Jacobi tensors and Ricci curvature", *Math. Ann.* **252**:1 (1980), 1–26. MR Zbl

[Gulliver 1975] R. Gulliver, "On the variety of manifolds without conjugate points", *Trans. Amer. Math. Soc.* **210** (1975), 185–201. MR Zbl

[Heber 1994] J. Heber, "On the geodesic flow of tori without conjugate points", *Math. Z.* **216**:2 (1994), 209–216. MR Zbl

[Hurewicz 1995] W. Hurewicz, *Collected works of Witold Hurewicz*, edited by K. Kuperberg, American Mathematical Society, Providence, RI, 1995. MR Zbl

[Ivanov and Kapovitch 2014] S. Ivanov and V. Kapovitch, "Manifolds without conjugate points and their fundamental groups", *J. Differential Geom.* **96**:2 (2014), 223–240. MR Zbl

[Lawson and Yau 1972] H. B. Lawson, Jr. and S. T. Yau, "Compact manifolds of nonpositive curvature", *J. Differential Geometry* **7** (1972), 211–228. MR Zbl

[O'Sullivan 1974] J. J. O'Sullivan, "Manifolds without conjugate points", *Math. Ann.* **210** (1974), 295–311. MR Zbl

[de Rham 1952] G. de Rham, "Sur la reductibilité d'un espace de Riemann", *Comment. Math. Helv.* **26** (1952), 328–344. MR Zbl

[Wolf 1964] J. A. Wolf, "Homogeneity and bounded isometries in manifolds of negative curvature", *Illinois J. Math.* **8** (1964), 14–18. MR Zbl

Received July 30, 2018. Revised September 18, 2019.

JAMES DIBBLE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF IOWA
IOWA CITY, IA
UNITED STATES

james-dibble@uiowa.edu
Current address:
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF SOUTHERN MAINE
PORTLAND, ME
UNITED STATES

james.dibble@maine.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.115

ON THE GLOBAL GAN–GROSS–PRASAD CONJECTURE FOR GENERAL SPIN GROUPS

MELISSA EMORY

We formulate a global Gan–Gross–Prasad conjecture for general spin groups. That is, we formulate a conjecture on a relation between periods of certain automorphic forms on $\mathrm{GSpin}_{n+1} \times \mathrm{GSpin}_n$ along the diagonal subgroup GSpin_n and some L-values. To support the conjecture, we show that the conjecture holds for $n = 2$ and 3 and for certain cases for $n = 4$.

1. Introduction

Gross and Prasad [1992] conjectured that the nonvanishing of periods of automorphic forms on $\mathrm{SO}_{n+1} \times \mathrm{SO}_n$ along the diagonal subgroup SO_n is equivalent to the nonvanishing of certain automorphic L-functions at the central critical value. Gan, Gross and Prasad [2012] extended this conjecture, now known as the *Gan–Gross–Prasad* (*GGP*) *conjecture*, to the remaining classical groups. The original Gross–Prasad conjecture was refined in [Ichino and Ikeda 2010], where an explicit relationship was conjectured between the period integral and the central critical L-values. An analogous conjecture was developed for unitary groups by N. Harris [2014]. The purpose of this paper is to formulate a similar conjecture for a nonclassical group known as the general spin group (GSpin), and to verify the conjecture for the first three cases essentially by interpreting the following known results: the Waldspurger formula [1985] for $n = 2$, Ichino's triple product formula [2008] for $n = 3$, and a result of Gan and Ichino [2011] for $n = 4$.

Let us first recall the original global Gross–Prasad conjecture. Let F be a number field and $\mathbb{A}$ the ring of adeles over F. Let $(V_n, q_n) \subset (V_{n+1}, q_{n+1})$ be an inclusion of quadratic spaces of respective dimensions n and $n+1$ over F, so that $q_{n+1}|_{V_n} = q_n$, where we assume n is at least 2 and V_n is not isomorphic to the hyperbolic plane. Then we have the natural inclusion $\mathrm{SO}_n \subset \mathrm{SO}_{n+1}$ of the corresponding special orthogonal groups $\mathrm{SO}_n := \mathrm{SO}(V_n)$ and $\mathrm{SO}_{n+1} := \mathrm{SO}(V_{n+1})$ over F, which gives rise to the inclusion $\mathrm{SO}_n(\mathbb{A}) \subset \mathrm{SO}_{n+1}(\mathbb{A})$. Let π_n and π_{n+1} be irreducible tempered cuspidal automorphic representations of $\mathrm{SO}_n(\mathbb{A})$ and $\mathrm{SO}_{n+1}(\mathbb{A})$, respectively. The original global Gross–Prasad conjecture is as follows.

MSC2010: primary 11F70; secondary 22E55.
Keywords: periods of automorphic forms, L-values, Gan–Gross–Prasad conjecture.

Conjecture 1.1 (original global Gross–Prasad conjecture [1992]). *Assume that for every place v of F, $\mathrm{Hom}_{\mathrm{SO}_n(F_v)}(\pi_{n+1,v} \otimes \pi_{n,v}, \mathbb{C}) \neq \{0\}$. Then there exist vectors $\phi \in V_{\pi_{n+1}}$ and $f \in V_{\pi_n}$ such that*

$$\int_{\mathrm{SO}_n(F)\backslash \mathrm{SO}_n(\mathbb{A})} \phi(g) f(g)\, dg \neq 0$$

if and only if the tensor product L-function $L(1/2, \pi_{n+1} \times \pi_n)$ does not vanish.

Ichino and Ikeda [2010] refined this conjecture by writing down an explicit (conjectural) relationship between the period integral and $L(1/2, \pi_{n+1} \times \pi_n)$ as follows. First define

$$\mathcal{P} : V_{\pi_{n+1}} \otimes V_{\pi_n} \to \mathbb{C}$$

by

$$\mathcal{P}(\phi, f) = \int_{\mathrm{SO}_n(F)\backslash \mathrm{SO}_n(\mathbb{A})} \phi(g) f(g)\, dg,$$

for $\phi \in V_{\pi_{n+1}}$ and $f \in V_{\pi_n}$, where dg is the Tamagawa measure of $\mathrm{SO}_n(\mathbb{A})$. This is, of course, nothing but the period integral of the above original Gross–Prasad conjecture. The basic idea of Ichino and Ikeda is to define a "local period" α_v by using the matrix coefficients of the local representations $\pi_{n+1,v}$ and $\pi_{n,v}$ so that the infinite product $\prod_v \alpha_v$ is defined. They then conjecture that the global $|\mathcal{P}(\phi, f)|^2$ is proportional to the product $\prod_v \alpha_v(\phi_v, f_v)$ for factorizable $\phi = \otimes_v \phi_v$ and $f = \otimes_v f_v$ and the L-value $L\big(\frac{1}{2}, \pi_{n+1} \times \pi_n\big)$ appears in the constant of proportionality.

To state their conjecture more precisely, first let

$$\mathcal{B}_{\pi_{n+1}} : V_{\pi_{n+1}} \otimes \overline{V}_{\pi_{n+1}} \to \mathbb{C} \quad \text{and} \quad \mathcal{B}_{\pi_n} : V_{\pi_n} \otimes \overline{V}_{\pi_n} \to \mathbb{C}$$

be the Petersson pairings defined as usual via the Tamagawa measures. Then fix isomorphisms

$$\pi_n \cong \otimes_v \pi_{n,v} \quad \text{and} \quad \pi_{n+1} \cong \otimes_v \pi_{n+1,v},$$

and decompositions

$$\mathcal{B}_{\pi_{n+1}} = \prod_v \mathcal{B}_{\pi_{n+1,v}} \quad \text{and} \quad \mathcal{B}_{\pi_n} = \prod_v \mathcal{B}_{\pi_{n,v}},$$

where

$$\mathcal{B}_{\pi_{n+1,v}} : \pi_{n+1,v} \otimes \bar{\pi}_{n+1,v} \to \mathbb{C} \quad \text{and} \quad \mathcal{B}_{\pi_{n,v}} : \pi_{n,v} \otimes \bar{\pi}_{n,v} \to \mathbb{C}$$

are local pairings. Also fix a decomposition $dg = \prod_v dg_v$ of the Tamagawa measure dg on $\mathrm{SO}_n(\mathbb{A})$.

Then define an $\mathrm{SO}_n(F_v) \times \mathrm{SO}_n(F_v)$-invariant functional

$$\alpha_v^{\natural} : (\pi_{n+1,v} \boxtimes \bar{\pi}_{n+1,v}) \otimes (\pi_{n,v} \boxtimes \overline{\pi}_{n,v}) \to \mathbb{C} \tag{1-1}$$

by

$$\alpha_v^\natural(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v}) := \int_{\mathrm{SO}_n(F_v)} \mathcal{B}_{\pi_{n+1,v}}(\pi_{n+1,v}(g_v)\phi_{1,v}, \phi_{2,v})\mathcal{B}_{\pi_{n,v}}(\pi_{n,v}(g_v) f_{1,v}, f_{2,v})\, dg_v$$

for $\phi_{1,v}, \phi_{2,v}$ in $\pi_{n+1,v}$ and $f_{1,v}, f_{2,v}$ in $\pi_{n,v}$. Ichino and Ikeda have proven that if $\pi_{i,v}$ is tempered then the integral for $\alpha_v^\natural$ converges absolutely, and

$$\alpha_v^\natural(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v}) = \Delta_{\mathrm{SO}_{n+1,v}} \frac{L_v(1/2, \pi_{n,v} \times \pi_{n+1,v})}{L_v(1, \pi_{n,v}, \mathrm{Ad})L_v(1, \pi_{n+1,v}, \mathrm{Ad})}$$

for almost all v, where

$$\Delta_{\mathrm{SO}_{n+1,v}} := \begin{cases} \zeta_v(2)\zeta_v(4)\cdots\zeta_v(2m) & \text{if } \dim V_{n+1,v} = 2m+1, \\ \zeta_v(2)\zeta_v(4)\cdots\zeta_v(2m-2)\cdot L_v(m, \chi_{V_{n+1,v}}) & \text{if } \dim V_{n+1,v} = 2m, \end{cases}$$

where $\chi_{V_{n+1,v}}$ is the character associated with the discriminant of the quadratic form associated to $V_{n+1,v}$. Accordingly, for all v define the normalized $\mathrm{SO}_{n+1}(F_v) \times \mathrm{SO}_n(F_v)$-invariant functional

$$\alpha_v : (\pi_{n+1,v} \boxtimes \bar{\pi}_{n+1,v}) \otimes (\pi_{n,v} \boxtimes \bar{\pi}_{n,v}) \to \mathbb{C}$$

by setting

$$\alpha_v = \Delta_{\mathrm{SO}_{n+1,v}}^{-1} \frac{L_v(1, \pi_{n,v}, \mathrm{Ad})L_v(1, \pi_{n+1,v}, \mathrm{Ad})}{L_v(1/2, \pi_{n,v} \times \pi_{n+1,v})} \alpha_v^\natural,$$

so that the infinite product

$$\prod_v \alpha_v(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v})$$

is well defined. Also we write

$$\alpha_v(\phi_v, f_v) := \alpha_v(\phi_v, \phi_v; f_v, f_v)$$

for $\phi_v \in \pi_{n+1,v}$ and $f_v \in \pi_{n,v}$.

Using this notation, we can state the Ichino–Ikeda refinement of the global Gross–Prasad conjecture for the special orthogonal groups as follows.

Conjecture 1.2 (Ichino–Ikeda refinement). *Assume π_{n+1} and π_n are tempered cuspidal automorphic representations of $\mathrm{SO}_{n+1}(\mathbb{A})$ and $\mathrm{SO}_n(\mathbb{A})$, respectively, and π_{n+1} and π_n appear with multiplicity one in the discrete spectrum. Then for each factorizable $\phi = \otimes_v \phi_v \in V_{\pi_{n+1}}$ and $f = \otimes_v f_v \in V_{\pi_n}$, we have*

$$|\mathcal{P}(\phi, f)|^2 = \frac{\Delta_{\mathrm{SO}_{n+1}}}{2^\beta} \frac{L(1/2, \pi_n \times \pi_{n+1})}{L(1, \pi_n, \mathrm{Ad})L(1, \pi_{n+1}, \mathrm{Ad})} \prod_v \alpha_v(\phi_v, f_v),$$

where 2^β is the product of cardinalities of the component groups attached to the

L-packets for π_{n+1} and π_n and

$$\Delta_{\mathrm{SO}_{n+1}} := \begin{cases} \zeta(2)\zeta(4)\cdots\zeta(2m) & \text{if } \dim V_{n+1} = 2m+1, \\ \zeta(2)\zeta(4)\cdots\zeta(2m-2)\cdot L(m, \chi_{V_{n+1}}) & \text{if } \dim V_{n+1} = 2m, \end{cases}$$

where $\chi_{V_{n+1}}$ is the quadratic character associated with V_{n+1}.

Remark 1.3. It should be noted that what is denoted by $\mathcal{P}(\phi, f)$ in [Ichino and Ikeda 2010] is our $|\mathcal{P}(\phi, f)|^2$. See [Xue 2017, Conjecture 6.2.1] for a similar conjecture if n is odd and π_{n+1} appears with multiplicity two in the discrete spectrum.

This refined conjecture for $n = 2$ follows from the well-known Waldspurger formula [1985] and the one for $n = 3$ follows from Ichino's triple product formula [2008]. Also for $n = 4$, Gan and Ichino [2011] have proven the conjecture under certain assumptions for π_{n+1} and π_n.

Our goal in this paper is to generalize this conjecture for the general spin groups and verify it for the cases $n = 2$, 3 and 4 by using [Waldspurger 1985; Ichino 2008; Gan and Ichino 2011], respectively, as above.

Let us first briefly recall some generalities of the general spin group. Let (V_n, q_n) be a quadratic space over F of dimension n. The general spin group associated with (V_n, q_n), which we denote by $\mathrm{GSpin}(V_n)$ or simply by GSpin_n, is a reductive group over F such that we have the short exact sequence

$$1 \to \mathrm{GL}_1 \to \mathrm{GSpin}(V_n) \to \mathrm{SO}(V_n) \to 1.$$

It should be noted that GL_1 is in the center of $\mathrm{GSpin}(V_n)$ and is the connected component Z_n° of the center if $n > 2$. If $n = 2$ then GSpin_2 is commutative and hence the connected component of the center is larger than this GL_1. However, as a convention in this paper, we set

$$Z_n^\circ = \mathrm{GL}_1$$

even when $n = 2$. Also the group $\mathrm{GSpin}(V_n)$ is equipped with a homomorphism

$$N : \mathrm{GSpin}(V_n) \to \mathrm{GL}_1,$$

which is called the spinor norm. Note that for $z \in \mathrm{GL}_1 \subseteq Z_n^\circ$ we have $N(z) = z^2$.

Next assume we have an inclusion $(V_n, q_n) \subseteq (V_{n+1}, q_{n+1})$ of quadratic spaces. Then we have the natural inclusion $\mathrm{GSpin}(V_n) \subseteq \mathrm{GSpin}(V_{n+1})$, which makes the diagram

$$\begin{array}{ccccccccc} 1 & \longrightarrow & Z_{n+1}^\circ & \longrightarrow & \mathrm{GSpin}(V_{n+1}) & \longrightarrow & \mathrm{SO}(V_{n+1}) & \longrightarrow & 1 \\ & & \| & & \cup\!| & & \cup\!| & & \\ 1 & \longrightarrow & Z_n^\circ & \longrightarrow & \mathrm{GSpin}(V_n) & \longrightarrow & \mathrm{SO}(V_n) & \longrightarrow & 1 \end{array}$$

commute.

Now, let π_{n+1} and π_n be tempered cuspidal automorphic representations of $\mathrm{GSpin}_{n+1}(\mathbb{A})$ and $\mathrm{GSpin}_n(\mathbb{A})$, respectively, and let $\omega_{\pi_{n+1}}$ and ω_{π_n} be the restrictions to $Z_n^\circ(\mathbb{A}) = \mathbb{A}^\times$ of the central characters of π_{n+1} and π_n, respectively, so that $\omega_{\pi_{n+1}}$ and ω_{π_n} are Hecke characters. Furthermore, assume that the product $\omega_{\pi_{n+1}}\omega_{\pi_n}$ has a square root; namely there exists a Hecke character $\omega : \mathbb{A}^\times \to \mathbb{C}^1$ such that

$$\omega^2 = \omega_{\pi_{n+1}}\omega_{\pi_n}.$$

Note that such an ω is not unique and two such ω's differ by a quadratic character. For each such ω, we define the global period

$$\mathcal{P}_\omega : V_{\pi_{n+1}} \otimes V_{\pi_n} \to \mathbb{C}$$

by

$$\mathcal{P}_\omega(\phi, f) := \int_{Z_n^\circ(\mathbb{A})\,\mathrm{GSpin}_n(F)\backslash \mathrm{GSpin}_n(\mathbb{A})} \phi(g) f(g) \omega(N(g))^{-1}\, dg,$$

where $\phi \in V_{\pi_{n+1}}$ and $f \in V_{\pi_n}$, and dg is the Tamagawa measure of $\mathrm{GSpin}_n(\mathbb{A})$. Because of the assumption on the central characters that $\omega^2 = \omega_{\pi_{n+1}}\omega_{\pi_n}$, this integral is well defined. Then in the same way as in the SO case, we define the local period $\alpha_{\omega_v}(\phi_v, f_v)$ in such a way that $\alpha_{\omega_v}(\phi_v, f_v) = 1$ for almost all v so that the product $\prod_v \alpha_{\omega_v}(\phi_v, f_v)$ makes sense, and we make an analogous conjecture.

To be precise, by fixing isomorphisms

$$\pi_n \cong \otimes_v \pi_{n,v} \quad \text{and} \quad \pi_{n+1} \cong \otimes_v \pi_{n+1,v}$$

and decompositions

$$\mathcal{B}_{\pi_{n+1}} = \prod_v \mathcal{B}_{\pi_{n+1,v}} \quad \text{and} \quad \mathcal{B}_{\pi_n} = \prod_v \mathcal{B}_{\pi_{n,v}},$$

where the global $\mathcal{B}_{\pi_{n+1}}$ and $\mathcal{B}_{\pi_n}$ are the Petersson pairings defined by the Tamagawa measures, we define the $\mathrm{GSpin}_{n+1}(F_v) \times \mathrm{GSpin}_n(F_v)$-invariant functional

$$\alpha^\natural_{\omega_v} : (\pi_{n+1,v} \boxtimes \bar{\pi}_{n+1,v}) \otimes (\pi_{n,v} \boxtimes \bar{\pi}_{n,v}) \to \mathbb{C}$$

by

$$\begin{aligned}\alpha^\natural_{\omega_v}&(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v}) \\ &:= \int_{Z_n^\circ(F_v)\backslash \mathrm{GSpin}_n(F_v)} \mathcal{B}_{\pi_{n+1,v}}(\pi_{n+1,v}(g_v)\phi_{1,v}, \phi_{2,v}) \\ &\qquad\qquad \times \mathcal{B}_{\pi_{n,v}}(\pi_{n,v}(g_v) f_{1,v}, f_{2,v}) \omega_v(N(g_v))^{-1}\, dg_v\end{aligned}$$

for $\phi_{1,v}, \phi_{2,v}$ in $\pi_{n+1,v}$ and $f_{1,v}, f_{2,v}$ in $\pi_{n,v}$, where we also fix the factorization $dg = \prod_v dg_v$ of the Tamagawa measure on $\mathrm{GSpin}_n(\mathbb{A})$.

Then we prove that if $\pi_{n+1,v}$ and $\pi_{n,v}$ are tempered, the integral for $\alpha^\natural_{\omega_v}(\phi_v, f_v)$ converges absolutely, and

$$\alpha^\natural_{\omega_v}(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v}) = \Delta_{\mathrm{SO}_{n+1,v}} \frac{L_v(1/2, \pi_{n,v} \times \pi_{n+1,v} \otimes \omega_v^{-1})}{L_v(1, \pi_{n,v}, \mathrm{Ad}) L_v(1, \pi_{n+1,v}, \mathrm{Ad})}$$

for almost all v, where $\Delta_{\mathrm{SO}_{n+1,v}}$ as before. Hence if we normalize $\alpha^{\natural}_{\omega_v}$ by setting

$$\alpha_{\omega_v} := \Delta^{-1}_{\mathrm{SO}_{n+1,v}} \frac{L_v(1, \pi_{n,v}, \mathrm{Ad}) L_v(1, \pi_{n+1,v}, \mathrm{Ad})}{L_v(1/2, \pi_{n,v} \times \pi_{n+1,v} \otimes \omega_v^{-1})} \alpha^{\natural}_{\omega_v},$$

then the infinite product

$$\prod_v \alpha_{\omega_v}(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v})$$

is well defined. Also we write

$$\alpha_{\omega_v}(\phi_v, f_v) := \alpha_{\omega_v}(\phi_v, \phi_v; f_v, f_v)$$

for $\phi_v \in \pi_{n+1,v}$ and $f_v \in \pi_{n,v}$.

Then we make the following conjecture, which we call the global GGP conjecture for GSpin.

Conjecture 1.4 (the global GGP conjecture for GSpin). *Let $\pi_n \cong \otimes_v \pi_{n,v}$ and $\pi_{n+1} \cong \otimes_v \pi_{n+1,v}$ be irreducible tempered cuspidal automorphic representations of $G_n(\mathbb{A})$ and $G_{n+1}(\mathbb{A})$, respectively, and we assume π_{n+1} and π_n appear with multiplicity one in the discrete spectrum. Assume there exists ω such that $\omega^2 = \omega_{\pi_{n+1}} \omega_{\pi_n}$. Then for each factorizable $\phi = \otimes_v \phi_v \in V_{\pi_{n+1}}$ and $f = \otimes_v f_v \in V_{\pi_n}$,*

$$|\mathcal{P}_\omega(\phi, f)|^2 = \frac{\Delta_{\mathrm{SO}_{n+1}}}{2^\beta} \frac{L(1/2, \pi_{n+1} \times \pi_n \otimes \omega^{-1})}{L(1, \pi_{n+1}, \mathrm{Ad}) L(1, \pi_n, \mathrm{Ad})} \prod_v \alpha_{\omega_v}(\phi_v, f_v),$$

where $\Delta_{\mathrm{SO}_{n+1}}$ is as before.

Let us mention that the L-function $L(\pi_{n+1} \times \pi_n \otimes \omega^{-1})$ is conjecturally holomorphic at $s = 1/2$ by, say, [Piatetski-Shapiro and Rallis 1987, Theorem 5.1]. Also the adjoint L-functions $L(s, \pi_{n+1}, \mathrm{Ad})$ and $L(1, \pi_n, \mathrm{Ad})$ are conjecturally nonzero holomorphic at $s = 1$ as in [Lapid and Mao 2015, (3.2) and page 483].

An interesting quantity in the above conjecture is 2^β, which is conjecturally related to the cardinalities of the component groups attached to the L-parameters of π_{n+1} and π_n. To discuss this issue, we first set up some general notations. Let G be a reductive group over our number field F and let π be a cuspidal automorphic representation of $G(\mathbb{A})$. Furthermore, let $\mathcal{L}_F$ be the hypothetical global Langlands group of F and let $\phi = \phi_\pi : \mathcal{L}_F \to {}^L G = \widehat{G} \rtimes \Gamma_F$ be the hypothetical global Langlands parameter of π, where $\Gamma_F = \mathrm{Gal}(\overline{F}/F)$. Set $S_\phi := \mathrm{Cent}(\mathrm{Im}(\phi), \widehat{G})$ and define

$$\mathcal{S}_\phi := S_\phi / S_\phi^0 Z(\widehat{G})^{\Gamma_F},$$

where S_ϕ^0 is the identity component of the complex reductive group S_ϕ, $Z(\widehat{G})$ is the center of $\widehat{G}$, and $Z(\widehat{G})^{\Gamma_F}$ is the subgroup of invariants in $Z(\widehat{G})$ under the natural action of Γ_F. Denote by $\widehat{G}_{\mathrm{sc}}$ the simply connected cover of the derived group $\widehat{G}_{\mathrm{der}}$

of $\widehat{G}$, and by $S_{\phi,\mathrm{sc}}$ the full preimage of S_ϕ in $\widehat{G}_{\mathrm{sc}}$. We then define

$$\mathcal{S}_{\phi,\mathrm{sc}} := S_{\phi,\mathrm{sc}}/S^{\circ}_{\phi,\mathrm{sc}}.$$

Using this notation, we make the following conjecture.

Conjecture 1.5. *Let ϕ_{n+1} and ϕ_n be the (conjectural) global Langlands parameters of π_{n+1} and π_n, respectively. If π_{n+1} and π_n appear with multiplicity one in the discrete spectrum, then*

$$2^\beta = 4|\mathcal{S}_{\phi_n}|\,|\mathcal{S}_{\phi_{n+1}}| = \tfrac{1}{2}|\mathcal{S}_{\phi_n,\mathrm{sc}}|\,|\mathcal{S}_{\phi_{n+1},\mathrm{sc}}|.$$

Remark 1.6. If π_{n+1} or π_n appear with multiplicity greater than one in the discrete spectrum, then Conjecture 1.5 should be modified in a similar way as for special orthogonal groups (see Conjecture 6.2.1 in [Xue 2017]).

As the last thing in this introduction, let us mention that if the central characters of π_{n+1} and π_n are both trivial, so that $\omega_{\pi_{n+1}} = \omega_{\pi_n} = 1$, then π_{n+1} and π_n can be seen as automorphic representations of $\mathrm{SO}_{n+1}(\mathbb{A})$ and $\mathrm{SO}_n(\mathbb{A})$, respectively. In this case, if one chooses $\omega = 1$, one can readily see that our conjecture reduces to that of Ichino and Ikeda. Hence our conjecture should be considered as a "generalization" of the Ichino–Ikeda conjecture rather than an analogue of it.

This paper is organized as follows. In Section 2, we review the general theory of GSpin and discuss Conjecture 1.5. In Section 3 we establish the convergence of the integral and then compute the integral for unramified data. In Section 4, we wrap-up our formulation of the conjecture, and then establish the conjecture for the $n = 2$, 3 and 4 cases.

Notation. If π is a representation of a group G, we denote the space of π by V_π. If π admits a central character we write ω_π for the central character of π restricted to the connected component of the center of G. Assume that the space V_π is a space of functions or maps on the group G and π is a representation of G on V_π defined by right translation (for example, when π is an automorphic subrepresentation). Let H be a subgroup of G. We define $\pi\|_H$ to be the representation of H realized in the space

$$V_{\pi\|_H} := \{f|_H : f \in V_\pi\}$$

of restrictions of $f \in V_\pi$ to H on which H acts by right translation. Namely $\pi\|_H$ is the representation obtained by restricting the functions in V_π.

For a reductive group G over F, we usually identify the group G with its F-rational points $G(F)$. We denote by Γ_F the absolute Galois group of F. If (V, q) is a quadratic space over F, we denote its discriminant by $\mathrm{disc}(V)$, which is always viewed in $F^\times/F^{\times 2}$. We denote by $\mathbb{H}$ the hyperbolic plane over F, namely the unique 2-dimensional split quadratic space.

2. The general spin group

In this section, we begin by defining the general spin group over any field F and then explicitly compute the first five cases. We then let F be a number field, and define the local and global L-functions for GSpin. The section is concluded with the component groups for the Langlands parameters for GSpin and we discuss Conjecture 1.5.

2A. *The group* GSpin. In the past literature such as [Asgari 2000; Asgari and Shahidi 2006; Asgari and Choiy 2017], the quasisplit general spin group is often defined in terms of roots and coroots, which is useful, for example, when one computes the dual group. In this paper, however, we give an alternate definition in terms of the Clifford algebra, which can be done even for the nonquasisplit case, and which more naturally gives the inclusion $\mathrm{GSpin}_n \subseteq \mathrm{GSpin}_{n+1}$.

Let (V, q) be a quadratic space with quadratic form q over an arbitrary field F of characteristic different from 2 with $\dim V = n$. (Of course we are interested in the case when F is local or global, but in this subsection F can be arbitrary.) Let $T(V)$ denote the tensor algebra of V, that is,

$$T(V) := \bigoplus_{k=0}^{\infty} V^{\otimes k} = F \oplus V \oplus (V \otimes V) \oplus (V \otimes V \otimes V) \oplus \cdots .$$

Let $I(q)$ be the two-sided ideal of $T(V)$ generated by the elements of the form

$$v \otimes v - q(v) \cdot 1,$$

where $v \in V$, and define

$$C(V) := C(V, q) = T(V)/I(q),$$

called the Clifford algebra associated with (V, q). We write $v_1 \otimes v_2 \otimes \cdots \otimes v_k = v_1 v_2 \cdots v_k$ and in particular $v^2 = q(v)$.

Lemma 2.1. *Let x and y be orthogonal in V; namely $B_q(x, y) = 0$, where B_q is the bilinear form associated with q. Then in $C(V)$,*

$$x \cdot y = -y \cdot x.$$

Proof. By definition of B_q, $B_q(x, y) = 0$ implies $q(x + y) - q(x) - q(y) = 0$, namely $q(x + y) = q(x) + q(y)$. Thus, we have

$$(x + y)(x + y) = x \cdot x + y \cdot y.$$

The lemma follows. □

Let $\{x_1, x_2, \ldots, x_n\}$ be an orthogonal basis of V. Then each element in $C(V)$ is a linear combination of the elements of the form $x_{i_1} x_{i_2} \cdots x_{i_k}$, where $x_{i_j} \in \{x_1, x_2, \ldots, x_n\}$. Furthermore, by the above lemma along with $v^2 = q(v) \in F$, we can readily see that each element in $C(V)$ is a linear combination of the elements of the form $x_{i_1}\cdots x_{i_k}$ with $i_1<i_2<\cdots<i_k$. We then define

$$C^k(V):=\Big\{\sum_i c_i x_{i_1}\cdots x_{i_k}\ :\ i_1<i_2<\cdots<i_k\Big\},$$

so that we have

$$C(V)=C^0(V)\oplus C^1(V)\oplus\cdots\oplus C^n(V).$$

Further, we define

$$C^+(V)=C^0(V)\oplus C^2(V)\oplus C^4(V)\oplus\cdots$$

and

$$C^-(V)=C^1(V)\oplus C^3(V)\oplus C^5(V)\oplus\cdots,$$

and call them the even Clifford algebra and the odd Clifford algebra, respectively, so that we have

$$C(V)=C^+(V)\oplus C^-(V).$$

Let us note that

$$\dim C^+(V)=\dim C^-(V)=2^{n-1}$$

and so $\dim C(V)=2^n$ (see, e.g., [Lam 2005, Theorem 1.8 and Corollary 1.9].)

With this said, the general spin group $\mathrm{GSpin}(V)$ associated with (V,q) is defined as

$$\mathrm{GSpin}(V):=\{g\in C^+(V): g^{-1}\text{ exists and } gVg^{-1}=V\}$$

(see, for example, [Deligne 1972, Section 3.2]). We sometimes write $\mathrm{GSpin}_n=\mathrm{GSpin}(V)$ when V is clear from the context.

The Clifford algebra $C(V)$ is equipped with a natural involution (which we call the canonical involution on $C(V)$) defined by

$$(v_1\ldots v_k)^*=v_k\ldots v_1,$$

where $v_i\in V$, giving rise to the map

$$N: C(V)\to C(V),\qquad N(x)=xx^*,$$

for $x\in C(V)$. It is immediate that $C^+(V)$ is closed under the canonical involution thanks to Lemma 2.1. Now if $g\in\mathrm{GSpin}(V)$ then we have $N(g)\in C^0(V)^\times=F^\times$, say, by [Scharlau 1985, Lemma 3.2, page 335], and we obtain the group homomorphism

$$N:\mathrm{GSpin}(V)\to F^\times,$$

which we call the spinor norm on $\mathrm{GSpin}(V)$.

Theorem 2.2. *Let $(V_n,q_n)\subset(V_{n+1},q_{n+1})$ be an inclusion of quadratic spaces, so that $q_{n+1}|_{V_n}=q_n$ Then we have the natural inclusion* $\mathrm{GSpin}(V_n)\subset\mathrm{GSpin}(V_{n+1})$.

Proof. If $V_n \subset V_{n+1}$ is an inclusion of quadratic spaces then $C^+(V_n) \subset C^+(V_{n+1})$. Now suppose $g \in C^+(V_n)$ is such that $gV_ng^{-1} \subseteq V_n$. Then it suffices to show that $gV_{n+1}g^{-1} \subseteq V_{n+1}$. We can choose an orthogonal basis such that

$$V_{n+1} = \mathrm{Span}\{v_1, v_2, \ldots, v_n, v_{n+1}\}$$

and such that $\{v_1, v_2, \ldots, v_n\}$ is an orthogonal basis for V_n. Then we can write

$$g = \sum_i \alpha_i v_{i_1} \ldots v_{i_k},$$

where $i_k \leq n$ and $\alpha_i \in F^\times$. Since g is in the even Clifford algebra, there are an even number of vectors appearing and by Lemma 2.1, $gv_{n+1} = v_{n+1}g$, and the claim follows. □

There is a natural homomorphism $p : \mathrm{GSpin}(V) \to \mathrm{SO}(V)$ sending $g \in \mathrm{GSpin}(V)$ to the map $v \mapsto gvg^{-1}$, giving the short exact sequence of algebraic groups

$$1 \longrightarrow \mathrm{GL}_1 \longrightarrow \mathrm{GSpin}(V) \xrightarrow{\;p\;} \mathrm{SO}(V) \longrightarrow 1 \tag{2-1}$$

where $\ker p = \mathrm{GL}_1 = C^0(V)^\times$ (see, for example, [Deligne 1972, Section 3]). Hence if $(V, q) \subset (V', q')$ is the inclusion of quadratic spaces as in the above corollary, we have the commutative diagram

$$\begin{array}{ccccccccc}
1 & \longrightarrow & \mathrm{GL}_1 & \longrightarrow & \mathrm{GSpin}(V') & \longrightarrow & \mathrm{SO}(V') & \longrightarrow & 1 \\
 & & \| & & \cup\!\mid & & \cup\!\mid & & \\
1 & \longrightarrow & \mathrm{GL}_1 & \longrightarrow & \mathrm{GSpin}(V) & \longrightarrow & \mathrm{SO}(V) & \longrightarrow & 1
\end{array}$$

because we have the obvious equality $C^0(V) = C^0(V')$.

The following should be mentioned.

Lemma 2.3. *Assume* $\dim V = n > 2$ *and let* Z° *be the identity component of the center of* $\mathrm{GSpin}(V)$. *Then* $\ker p = Z^\circ$, *where* $p : \mathrm{GSpin}(V) \to \mathrm{SO}(V)$ *is as above*, *and hence* $Z^\circ = \mathrm{GL}_1$.

Proof. First it is clear that $\ker p \subseteq Z^\circ$ because $\ker p = \mathrm{GL}_1$ is in the center of $\mathrm{GSpin}(V)$ and GL_1 is connected. So it suffices to show $Z^\circ \subseteq \ker p$. So let $a \in Z^\circ$. Then $p(a)$ is in the center Z_{SO} of $\mathrm{SO}(V)$. Now if n is odd, then $Z_{\mathrm{SO}} = 1$ and hence $p(a) = 1$. If n is even, then $Z_{\mathrm{SO}} = \{\pm 1\}$. But since $\{\pm 1\}$ is disconnected and Z° is connected (as an algebraic group), we also have $p(a) = 1$. So in either case, we have $a \in p^{-1}(1) = \ker p$. □

Let us note that if $\dim V \leq 2$ then it is well known that the entire group $\mathrm{SO}(V)$ is commutative. Similarly, as we will see in the next subsection, the general spin group $\mathrm{GSpin}(V)$ is also commutative.

2B. *Low rank* GSpin. In this subsection, we will explicitly compute $\mathrm{GSpin}(V_n)$ when $n = \dim V_n$ is small. This is done by using that $\mathrm{GSpin}(V_n)$ is a subgroup of the group of similitudes

$$\mathrm{Sim}_n := \{g \in C^+(V_n) : N(g) \in C^0(V)^\times = F^\times\}$$

(see, for example, [Knus et al. 1998, Proposition 13.10]) along with the following result.

Theorem 2.4. *Let $d = \mathrm{disc}(V_n)$ and $E = F(\sqrt{d})$. Then we have*

$$C^+(V_n) = \begin{cases} A & \text{if } n \text{ is odd;} \\ A \times A & \text{if } n \text{ is even and } d = 1; \\ A_E & \text{if } n \text{ is even and } d \neq 1, \end{cases}$$

where A is a central simple algebra over F and A_E is a central simple algebra over E. Note that since $\dim_F C^+(V_n) = 2^{n-1}$ we must have $\dim_F A = 2^{n-1}$ if n is odd, and $\dim_F A = \dim_E A_E = 2^{n-2}$ if n is even.

Also,

$$\text{the involution } * \text{ is} \begin{cases} \text{unitary} & \text{if } n \equiv 2, 6 \bmod 8; \\ \text{symplectic} & \text{if } n \equiv 3, 4, 5 \bmod 8; \\ \text{orthogonal} & \text{if } n \equiv 0, 1, 7 \bmod 8, \end{cases}$$

and furthermore if $n \equiv 0, 4 \bmod 8$ and $C^+(V_n) = A \times A$ then $*$ is of orthogonal or symplectic type on each factor of $C^+(V_n)$.

Proof. See [Lam 2005, Theorems 2.4, 2.5; Knus et al. 1998, (8.4) Proposition]. □

Though known to the experts, using that GSpin_n is a connected subgroup of Sim_n we can easily compute $\mathrm{GSpin}_n = \mathrm{GSpin}(V_n)$ for $n = 1, 2, 3, 4$ and the split case of GSpin_5 by showing for these low rank cases that $\dim \mathrm{GSpin}_n = \dim \mathrm{Sim}_n$ and hence $\mathrm{GSpin}_n \cong \mathrm{Sim}_n$ as follows:

***n* = 1 :** If $n = 1$ then by Theorem 2.4, $C^+(V_1)$ is a central simple algebra over F of dimension 1. Hence, $C^+(V_1) = F$, and

$$\begin{aligned} \mathrm{Sim}_1 &= \{g \in C^+(V_1) : N(g) \in F^\times\} \\ &= \{g \in F : N(g) \in F^\times\} \\ &= \{g \in F : gg^* \in F^\times\} \\ &= \{g \in F : g^2 \in F^\times\} \\ &\cong F^\times \cong \mathrm{GSpin}_1 . \end{aligned}$$

***n* = 2 :** If $n = 2$ then by Theorem 2.4 there are two cases to consider.

Case 1: Assume $d = 1$, so that $V_2 = \mathbb{H}$ (hyperbolic plane). Then $C^+(V_2) = A \times A$, where A is a central simple algebra over F with $\dim_F A = 1$, $C^+(V_2) = F \times F$.

Then by Theorem 2.4 we know that the involution $*$ is given by $(a, b)^* = (b, a)$ for $(a, b) \in F \times F$. The fixed field of this involution is $\Delta F = \{(a, a)\}$, which is equal to $C^0(V_2)$. Hence

$$\begin{aligned}\mathrm{Sim}_2 &= \{(a, b) \in F \times F : (a, b)(a, b)^* \in \Delta F^\times\}\\ &= \{(a, b) \in F \times F : (a, b)(b, a) \in \Delta F^\times\}\\ &= \{(a, b) \in F \times F : (ab, ab) \in \Delta F^\times\}\\ &\cong F^\times \times F^\times \cong \mathrm{GL}_1 \times \mathrm{GL}_1 \cong \mathrm{GSpin}_2 .\end{aligned}$$

Case 2: Assume $d \neq 1$, so that $V_2 = E$ and $q_2 = N_{E/F}$. Then $C^+(V_2)$ is a central simple algebra over $E =$ with $\dim_F C^+(V_2) = 2$, which implies $C^+(V_2) = E$. Again by Theorem 2.4 the involution $*$ is the Galois conjugate of g. Then

$$\mathrm{Sim}_2 = \{g \in E : gg^* \in F^\times\} \cong E^\times \cong \mathrm{Res}_{E/F}\,\mathrm{GL}_1 \cong \mathrm{GSpin}_2 .$$

***n* = 3 :** If $n = 3$ then $C^+(V_3)$ is a central simple algebra over F and $\dim_F C^+(V_3) = 4$, which means $C^+(V_3)$ is a quaternion algebra D over F. Moreover, by Theorem 2.4, the involution $*$ is symplectic, which implies that $*$ is the quaternion conjugation. Then

$$\mathrm{Sim}_3 = \{g \in C^+(V_3) : N(g) \in F^\times\} = \{g \in D : g\bar{g} \in F^\times\} = D^\times \cong \mathrm{GSpin}_3,$$

where the bar indicates the quaternion conjugation. In particular, if D is split then $\mathrm{GSpin}_3 = \mathrm{GL}_2$.

***n* = 4 :** If $n = 4$ then by Theorem 2.4 there are two cases to consider.

Case 1: Assume $d = 1$, so that $C^+(V_4) = A \times A$ where A is a central simple algebra over F with $\dim_F A = 4$. Then $C^+(V_4) = D \times D$ where D is a quaternion algebra over F, and the involution $*$ is symplectic on each factor D. Hence the involution $*$ is given by

$$(x, y)^* = (\bar{x}, \bar{y})$$

for $(x, y) \in D$, where the bar is the quaternion conjugation as above. Hence

$$\begin{aligned}\mathrm{Sim}_4 &= \{(x, y) \in D \times D : (x, y)(\bar{x}, \bar{y}) \in \Delta F^\times\}\\ &= \{(x, y) \in D \times D : (x\bar{x}, y\bar{y}) \in \Delta F^\times\}\\ &= \{(x, y) \in D \times D : x\bar{x} = y\bar{y} \in F^\times\}\\ &= \{(x, y) \in D \times D : N_D(x) = N_D(y) \in F^\times\}\\ &\cong \mathrm{GSpin}_4,\end{aligned}$$

where N_D is the reduced norm on D. In particular, if $D = M_2(F)$ then

$$\mathrm{GSpin}_4 = \{(x, y) \in \mathrm{GL}_2(F) \times \mathrm{GL}_2(F) : \det x = \det y\}.$$

Case 2: Assume $d \neq 1$, so that $C^+(V_4)$ is a central simple algebra over $E = F(\sqrt{d})$ and $\dim_F C^+(V_4) = 8$ so $\dim_E C^+(V_4) = 4$. Thus, $C^+(V_4)$ is a quaternion algebra D_E over E and the involution $*$ is the quaternion conjugation on D_E, which fixes E pointwise. Then

$$\mathrm{Sim}_4 = \{g \in C^+(V_5) : gg^* \in F^\times\} = \{g \in D_E : N_{D_E}(g) \in F^\times\} \cong \mathrm{GSpin}_4,$$

where N_{D_E} is the reduced norm on the quaternion algebra D_E and the isomorphism follows from the equal dimensions of the respective Lie algebras of $\{g \in D_E : N_{D_E}(g) \in F^\times\}$ and GSpin_4. In particular, if $D_E = M_2(E)$, then $N_{D_E}(g) = \det g$, so $\mathrm{GSpin}_4 = \{g \in \mathrm{GL}_2(E) : \det g \in F^\times\}$.

***n* = 5 :** If $n = 5$ then $C^+(V_5)$ is a central simple algebra of dimension 16 over F. Now, for our purposes we need only the case when V_5 is of the form

$$V_5 = \mathbb{H} \perp \mathbb{H} \perp \langle a\rangle,$$

where $a \in F^\times$. Then we will show $C^+(V_5) = M_4(F)$ and $\mathrm{GSpin}_5 = \mathrm{GSp}_4$. First assume $a = -1$, so that $V_5 = \mathbb{H} \perp \mathbb{H} \perp \langle -1\rangle$. Then by [Lam 2005, Corollary 2.10, page 112],

$$C^+(V_5) = C^+(\langle -1\rangle \perp (\mathbb{H} \perp \mathbb{H})) = C(\mathbb{H} \perp \mathbb{H}) = M_4(F).$$

For general $a \in F^\times$, we have $aV_5 = \mathbb{H} \perp \mathbb{H} \perp \langle -a\rangle$, and $C^+(aV_5) = C^+(V_5)$ by [Lam 2005, Corollary 2.11, page 112]. Hence, $C^+(V_5) = M_4(F)$.

The involution $*$ on $C^+(V_5)$ is symplectic involution, which by definition means that there exists a 4-dimensional symplectic space $(V, \langle -, -\rangle)$ over F with $\dim V = 4$ and an isomorphism $M_4(F) \cong \mathrm{End}_F V$ such that the involution $*$ is the pullback of the adjoint involution on V induced by the symplectic form $\langle -, -\rangle$. Hence we have

$$\mathrm{Sim}_5 = \{g \in M_4(F) : gg^* \in F^\times\} = \{g \in \mathrm{End}_F V = gg^* \in F^\times\},$$

where by g^* for $g \in \mathrm{End}_F V$ we mean the adjoint with respect to the symplectic form on V. Then viewed inside $\mathrm{End}_F V$ we have

$$\langle gv, gv'\rangle = \langle v, g^* gv'\rangle = N(g)\langle v, v'\rangle$$

for all $v, v' \in V$, which implies $\mathrm{Sim}_5 \cong \mathrm{GSp}_4$, and $N(g)$ is the similitude factor of $g \in \mathrm{GSp}_4$. The equality of dimensions of $\mathrm{Sim}_4 \cong \mathrm{GSp}_4$ and GSpin_5 gives

$$\mathrm{GSpin}_5 \cong \mathrm{GSp}_4\,.$$

Remark 2.5. Sometimes in the literature, GSpin_1 is "defined" as GL_1. (See, for example, [Asgari 2002, page 678].) However, this actually "follows from" our definition of GSpin. Also the case $n = 4$ is also proven in [Asgari and Choiy 2017, Proposition 2.1] by using roots and coroots.)

2C. ***On certain L-functions.*** In this subsection, we review the basics of the L-group of GSpin and certain L-functions attached to a cuspidal automorphic representation of GSpin. Accordingly, we let F be a number field and we simply write $\mathrm{GSpin}_n = \mathrm{GSpin}(V_n)$.

First recall that the Langlands dual group $\widehat{\mathrm{GSpin}_n}$ is defined as

$$\widehat{\mathrm{GSpin}_n} = \begin{cases} \mathrm{GSp}_{2m}(\mathbb{C}) & \text{if } n = 2m+1; \\ \mathrm{GSO}_{2m}(\mathbb{C}) & \text{if } n = 2m. \end{cases}$$

Next assume that GSpin_n is quasisplit. Then (the Galois form of) the global L-group ${}^L\mathrm{GSpin}_n$ is defined as

$${}^L\,\mathrm{GSpin}_n = \begin{cases} \mathrm{GSp}_{2m}(\mathbb{C}) \times \Gamma_F & \text{if } n = 2m+1; \\ \mathrm{GSO}_{2m}(\mathbb{C}) \rtimes \Gamma_F & \text{if } n = 2m, \end{cases}$$

where Γ_F is the absolute Galois group of F. Note that for $n = 2m$ the action of Γ_F is trivial on Γ_E where $F = E(\sqrt{d})$ and $d = \mathrm{disc}(V_n)$, so that we have the natural surjection

$$\mathrm{GSO}_{2m}(\mathbb{C}) \rtimes \Gamma_F \to \mathrm{GSO}_{2m} \rtimes \mathrm{Gal}(E/F) \subseteq \mathrm{GO}_{2m}(\mathbb{C}),$$

where in the case $d \neq 1$ the nontrivial element in $\mathrm{Gal}(E/F)$ acts as in, say [Hundley and Sayag 2016, Section 4.3], so that we have the inclusion in $\mathrm{GO}_{2m}(\mathbb{C})$. If GSpin_n is not quasisplit, there exists a unique quasisplit inner form GSpin_n^* of GSpin_n, and we define

$${}^L\mathrm{GSpin}_n = {}^L\mathrm{GSpin}_n^*.$$

Now for each place v of F, we define the local L-group ${}^L\mathrm{GSpin}_n(F_v)$ analogously as above by replacing F by F_v.

Let $\pi = \otimes_v \pi_v$ be a cuspidal automorphic representation of $\mathrm{GSpin}_n(\mathbb{A})$. Assuming the (conjectural) local Langlands correspondence for GSpin_n, we have the local Langlands parameter

$$\phi_{\pi_v} : WD_{F_v} \to {}^L\mathrm{GSpin}_n(F_v),$$

where WD_{F_v} is the Weil–Deligne group of F_v. Now for each homomorphism

$$\rho : {}^L\mathrm{GSpin}_n(F_v) \to \mathrm{GL}_N(\mathbb{C}),$$

the local L-factor is defined as

$$L_v(s, \pi_v, \rho) := L_v(s, \rho \circ \pi_{\phi_v}),$$

where the right-hand side is the local L-factor of Artin type associated with the N-dimensional Galois representation $\rho \circ \pi_{\phi_v}$. We then define the global automorphic L-function by

$$L(s, \pi, \rho) = \prod_v L_v(s, \rho \circ \phi_{\pi_v}),$$

and of course it is expected that this product converges for $\Re(s)$ sufficiently large and admits meromorphic continuation and a functional equation.

There are three cases of ρ we are interested in: the standard representation, adjoint representation and tensor product representation, where the last one actually involves another GSpin_{n+1}.

Firstly, we have the standard representation $\rho = \mathrm{std}_n$, which is given by

$$\mathrm{std}_n : {}^L\mathrm{GSpin}_n(F_v) \to \mathrm{GL}_{2m}(\mathbb{C}),$$

where m is such that $n = 2m+1$ or $n = 2m$, depending on the parity of n. To be more precise, we have the natural maps

$$\mathrm{std}_n : {}^L\,\mathrm{GSpin}_{2m+1}(F_v) = \mathrm{GSp}_{2m}(\mathbb{C}) \times \Gamma_{F_v} \to \mathrm{GSp}_{2m}(\mathbb{C}) \subseteq \mathrm{GL}_{2m}(\mathbb{C}),$$

and

$$\mathrm{std}_n : {}^L\,\mathrm{GSpin}_{2m}(F_v) \to \mathrm{GSO}_{2m} \rtimes \mathrm{Gal}(E_v/F_v) \subseteq \mathrm{GO}_{2m}(\mathbb{C}) \subseteq \mathrm{GL}_{2m}(\mathbb{C}),$$

where E_v is interpreted as $F_v \times F_v$ for split v. We have the L-functions $L(s,\pi,\mathrm{std}_n) = \prod_v L(s,\pi_v,\mathrm{std}_n)$ called the standard L-function, which we simply write $L(s,\pi)$.

Secondly, we need to consider the adjoint representation. Note that we have the adjoint representations

$$\mathrm{Ad}_{\mathfrak{gsp}} : {}^L\mathrm{GSpin}_{2m+1}(F_v) \to \mathrm{GSp}_{2m}(\mathbb{C}) \to \mathrm{GL}(\mathfrak{gsp}_{2m}(\mathbb{C}))$$

and

$$\mathrm{Ad}_{\mathfrak{gso}} : {}^L\mathrm{GSpin}_{2m}(F_v) \to \mathrm{GO}_{2m}(\mathbb{C}) \to \mathrm{GL}(\mathfrak{gso}_{2m}(\mathbb{C})),$$

where $\mathfrak{gsp}_{2m}(\mathbb{C})$ and $\mathfrak{gso}_{2m}(\mathbb{C})$ are the Lie algebras of the corresponding groups as usual. Since

$$\mathfrak{gsp}_{2m}(\mathbb{C}) = \mathfrak{sp}_{2n}(\mathbb{C}) \oplus \mathbb{C} \quad \text{and} \quad \mathfrak{gso}_{2m}(\mathbb{C}) = \mathfrak{so}_{2n}(\mathbb{C}) \oplus \mathbb{C},$$

we have

$$\mathrm{Ad}_{\mathfrak{gsp}} = \mathrm{Ad}_{\mathfrak{sp}} \oplus \mathbb{1} \quad \text{and} \quad \mathrm{Ad}_{\mathfrak{gso}} = \mathrm{Ad}_{\mathfrak{so}} \oplus \mathbb{1}.$$

Accordingly, we have

$$L_v(s,\pi_v,\mathrm{Ad}_{\mathfrak{gsp}}) = L_v(s,\pi_v,\mathrm{Ad}_{\mathfrak{sp}})\zeta_v(s)$$

and

$$L_v(s,\pi_v,\mathrm{Ad}_{\mathfrak{gso}}) = L_v(s,\pi_v,\mathrm{Ad}_{\mathfrak{so}})\zeta_v(s),$$

which, of course, give

$$L(s,\pi,\mathrm{Ad}_{\mathfrak{gsp}}) = L(s,\pi,\mathrm{Ad}_{\mathfrak{sp}})\zeta(s) \quad \text{and} \quad L(s,\pi,\mathrm{Ad}_{\mathfrak{gso}}) = L(s,\pi,\mathrm{Ad}_{\mathfrak{so}})\zeta(s),$$

by taking the product over all v. Then we define

$$L(s,\pi,\mathrm{Ad}) := \begin{cases} L(s,\pi,\mathrm{Ad}_{\mathfrak{sp}}) & \text{if } n = 2m+1; \\ L(s,\pi,\mathrm{Ad}_{\mathfrak{so}}) & \text{if } n = 2m, \end{cases}$$

and call it the adjoint L-function of π. The adjoint L-function $L(s,\pi,\mathrm{Ad})$ is conjecturally nonzero and holomorphic as in [Lapid and Mao 2015, (3.2) and page 483].

Thirdly, we consider the tensor product representation. For this we also consider GSpin_{n+1} and the standard representation std_{n+1}. Then we have the tensor product $\mathrm{std}_n \otimes \mathrm{std}_{n+1}$ of the two standard representations, which is of the form

$$\mathrm{std}_n \otimes \mathrm{std}_{n+1} : {}^L\,\mathrm{GSpin}_n(F_v) \times {}^L\,\mathrm{GSpin}_{n+1}(F_v) \to \begin{cases} \mathrm{GL}_{n^2}(\mathbb{C}) & \text{if } n = 2m; \\ \mathrm{GL}_{(n-1)n}(\mathbb{C}) & \text{if } n = 2m+1. \end{cases}$$

Then for cuspidal automorphic representations π_n and π_{n+1} of $\mathrm{GSpin}_n(\mathbb{A})$ and GSpin_{n+1}, respectively, we write

$$L(s, \pi_n \times \pi_{n+1}) = L(s, \pi_n \boxtimes \pi_{n+1}, \mathrm{std}_n \otimes \mathrm{std}_{n+1})$$

and call it the tensor product L-function. Let us mention that the L-function $L(s, \pi_n \times \pi_{n+1})$ is conjecturally holomorphic; see, for example, [Piatetski-Shapiro and Rallis 1987, Theorem 5.1].

2D. ***Component groups for Langlands parameters.*** The goal of this section is to discuss the component groups for the (conjectural) global Langlands parameters in relation to Conjecture 1.5. As we did in the introduction, let $\mathcal{L}_F$ be the hypothetical global Langlands group of our number field F and let $\phi : \mathcal{L}_F \to {}^L G = \widehat{G} \rtimes \Gamma_F$ be a global Langlands parameter. Set $S_\phi = \mathrm{Cent}(\mathrm{Im}(\phi), \widehat{G})$, and define

$$\mathcal{S}_\phi = S_\phi / S_\phi^0 Z(\widehat{G})^{\Gamma_F},$$

where S_ϕ^0 is the identity component of the complex reductive group S_ϕ, $Z(\widehat{G})$ is the center of $\widehat{G}$ and $Z(\widehat{G})^{\Gamma_F}$ is the subgroup of invariants in $Z(\widehat{G})$ under the natural action of Γ_F.

Now, let π_n and π_{n+1} be cuspidal automorphic representations of $\mathrm{GSpin}_n(\mathbb{A})$ and $\mathrm{GSpin}_{n+1}(\mathbb{A})$, respectively. We then have made the following conjecture (Conjecture 1.5):

$$2^\beta = 4 |\mathcal{S}_{\phi_n}|\, |\mathcal{S}_{\phi_{n+1}}|,$$

where ϕ_n and ϕ_{n+1} are the (conjectural) global Langlands parameters of π_n and π_{n+1}, respectively.

Lemma 2.6. *The group $\mathcal{S}_{\phi_n}$ is an elementary abelian 2-group, so in particular its order is a power of 2.*

Proof. First consider the case $n = 2m+1$, so that the (conjectural) global Langlands parameter ϕ_n is of the form

$$\phi_n : \mathcal{L}_F \to {}^L\mathrm{GSpin}_{2m+1} = \mathrm{GSp}_{2m}(\mathbb{C}) \times \Gamma_F$$

such that $\mathrm{Im}(\phi_n)$ is not in a proper parabolic subgroup of $\mathrm{GSp}_{2m}(\mathbb{C})$. Since the action of Γ_F on $\mathrm{GSp}_{2m}(\mathbb{C})$ is trivial, we may consider ϕ_n as the composite

$$\phi_n : \mathcal{L}_F \to \mathrm{GSp}_{2m}(\mathbb{C}) \times \Gamma_F \to \mathrm{GSp}_{2m}(\mathbb{C}),$$

where the second map is the obvious projection. Moreover, by writing $\mathrm{GSp}_{2m}(\mathbb{C}) = \mathrm{GSp}(W)$, where $(W, \langle -, - \rangle)$ is an m-dimensional symplectic space over F, we view ϕ_n as a representation of $\mathcal{L}_F$ acting on W.

Now, let $W_1 \subseteq W$ be an irreducible subspace of ϕ_n, and $W_1^{\perp}$ the orthogonal complement of W_1 with respect to the symplectic form $\langle -, - \rangle$. One can readily see that $W_1^{\perp}$ is a subrepresentation of ϕ_n.

Assume $W_1 \cap W_1^{\perp} \neq 0$. Then $W_1 \cap W_1^{\perp}$ is a nonzero subrepresentation of W_1, which, by irreducibility of W_1, implies that $W_1 \cap W_1^{\perp} = W_1 \subseteq W_1^{\perp}$. Hence W_1 is totally isotropic, and so the group $\mathrm{Im}(\phi_m)$ stabilizes the flag $0 \subseteq W_1 \subseteq W$ which implies $\mathrm{Im}(\phi_n)$ is in the proper parabolic subgroup $P(W_1) \subseteq \mathrm{GSp}(W)$, which contradicts our assumption that $\mathrm{Im}(\phi_n)$ is not in a proper parabolic subgroup of $\mathrm{GSp}(W)$.

Thus we necessarily have $W_1 \cap W_1^{\perp} = 0$, so that $W = W_1 \oplus W_1^{\perp}$. By induction,

$$W = W_1 \oplus W_2 \oplus \cdots \oplus W_k,$$

where each $(W_i, \langle \cdot, \cdot \rangle)$ is a smaller symplectic space. Then

$$\mathrm{Im}(\phi_n) \subseteq (\mathrm{GSp}(W_1) \times \cdots \times \mathrm{GSp}(W_k)) \cap \mathrm{GSp}(W),$$

which forces

$$\mathrm{Im}(\phi_n) \subseteq \{(g_1, \cdots, g_k) \in \mathrm{GSp}(W_1) \times \cdots \times \mathrm{GSp}(W_k) : N_1(g_1) = \cdots = N_k(g_k)\},$$

where each N_i is the similitude character on $\mathrm{GSp}(W_i)$. Then we see that

$$\begin{aligned} S_{\phi_n} &= \mathrm{Cent}(\mathrm{Im}(\phi_\pi), \mathrm{GSp}_{2m}(\mathbb{C})) \\ &= \{(a_1 I_{W_1}, \ldots, a_k I_{W_k}) : a_1^2 = \cdots = a_k^2\} \\ &= \{(\pm a I_{W_1}, \cdots, \pm a I_{W_k}) : a \in \mathbb{C}^\times\}, \end{aligned} \tag{2-2}$$

where I_{W_i} is the identity on W_i. The identity component is then

$$S_{\phi_n}^\circ = \{(a I_{W_1}, \cdots, a I_{W_k}) : a \in \mathbb{C}^\times\} = Z(\mathrm{GSp}_{2m}(\mathbb{C})).$$

Hence we have

$$S_{\phi_n} / S_{\phi_n}^\circ Z(\hat{G})^\Gamma = \{(\pm 1, \ldots, \pm 1)\}.$$

Finally, we have that

$$\mathcal{S}_{\phi_n} = \{(\pm 1, \ldots, \pm 1)\} = (\mathbb{Z}/2\mathbb{Z})^k,$$

which finishes the proof for $n = 2m + 1$.

Next assume $n = 2m$. Then the Langlands parameter is of the form

$$\phi_n : \mathcal{L}_F \to {}^L\mathrm{GSpin}_{2m+1} = \mathrm{GSO}_{2m}(\mathbb{C}) \rtimes \Gamma_F.$$

Note that Γ_F acts trivially on the center $Z(\mathrm{GSO}_{2m})$. Furthermore, as we have seen in the previous subsection, the action of Γ_F is trivial on Γ_E, where $E = F(\sqrt{d})$, and hence we may consider ϕ_n as a map

$$\phi_n : \mathcal{L}_F \to \mathrm{GSO}_{2m}(\mathbb{C}) \rtimes \mathrm{Gal}(E/F) = \mathrm{GO}_{2m}(\mathbb{C}).$$

Again by writing $\mathrm{GO}_{2m}(\mathbb{C}) = \mathrm{GO}(W)$ for a complex symmetric bilinear space W, we can consider ϕ_n as a representation of $\mathcal{L}_F$ acting on W. Then we can argue as before. $\square$

Next, let us set up some general notation. Let G be a reductive group over F. Set $\widehat{G}_{\mathrm{der}}$ to be the derived group of $\widehat{G}$ and $\widehat{G}_{\mathrm{sc}}$ to be the simply connected cover of $\widehat{G}_{\mathrm{der}}$, so that we have the maps

$$\widehat{G}_{\mathrm{sc}} \twoheadrightarrow \widehat{G}_{\mathrm{der}} \subseteq \widehat{G} \to \widehat{G}/Z(\widehat{G})^{\Gamma}. \tag{2-3}$$

For each global Langlands parameter $\phi : \mathcal{L}_F \to \widehat{G} \rtimes \Gamma_F$, we set $S_{\phi,\mathrm{sc}} \subseteq \widehat{G}_{\mathrm{sc}}$ to be the full preimage of $S_\phi \subseteq \widehat{G}$ under the map $\widehat{G}_{\mathrm{sc}} \to \widehat{G}$ as above. We then define the larger component group by

$$\mathcal{S}_{\phi,\mathrm{sc}} := S_{\phi,\mathrm{sc}}/S^0_{\phi,\mathrm{sc}},$$

which is a central extension of $\mathcal{S}_\phi$ by

$$\widehat{Z}_{\phi,\mathrm{sc}} = \widehat{Z}_{\mathrm{sc}}/\widehat{Z}_{\mathrm{sc}} \cap S^\circ_{\phi,\mathrm{sc}}, \tag{2-4}$$

where $\widehat{Z}_{\mathrm{sc}}$ is the center of $\widehat{G}_{\mathrm{sc}}$ and $S_{\phi,\mathrm{sc}}$ is the full inverse image of S_ϕ in $\widehat{G}_{\mathrm{sc}}$; namely we have the short exact sequence

$$1 \longrightarrow \widehat{Z}_{\phi,\mathrm{sc}} \longrightarrow \mathcal{S}_{\phi,\mathrm{sc}} \longrightarrow \mathcal{S}_\phi \longrightarrow 1, \tag{2-5}$$

(see, e.g., [Arthur 2013, (9,2.2)]). It should be noted that this immediately implies

$$|\mathcal{S}_{\phi,\mathrm{sc}}| = |\widehat{Z}_{\phi,\mathrm{sc}}|\,|\mathcal{S}_\phi|. \tag{2-6}$$

Let us note that if $G = \mathrm{GSpin}_{2m+1}$ then

$$\widehat{G} = \mathrm{GSp}_{2m}(\mathbb{C}), \qquad \widehat{G}_{\mathrm{der}} = \mathrm{Sp}_{2m}(\mathbb{C}), \qquad \widehat{G}_{\mathrm{sc}} = \mathrm{Sp}_{2m}(\mathbb{C}),$$

and if $G = \mathrm{GSpin}_{2m}$ then

$$\widehat{G} = \mathrm{GSO}_{2m}(\mathbb{C}), \qquad \widehat{G}_{\mathrm{der}} = \mathrm{SO}_{2m}(\mathbb{C}), \qquad \widehat{G}_{\mathrm{sc}} = \mathrm{Spin}_{2m}(\mathbb{C}).$$

With this said, we first have the following.

Lemma 2.7. *Assume $G = \mathrm{GSpin}_n$. In the above notation, we have*

$$|\mathcal{S}_{\phi,\mathrm{sc}}| = \begin{cases} 2|\mathcal{S}_\phi| & \text{if } n = 2m+1, \\ 4|\mathcal{S}_\phi| & \text{if } n = 2m, \end{cases}$$

where ϕ is, of course, a hypothetical global Langlands parameter for GSpin_n.

Proof. Assume $n = 2m+1$. Then (2-3) is written as

$$\mathrm{Sp}_{2m}(\mathbb{C}) \twoheadrightarrow \mathrm{Sp}_{2m}(\mathbb{C}) \subseteq \mathrm{GSp}_{2m}(\mathbb{C}).$$

From (2-2), one can see that $S_{\phi,\mathrm{sc}} = S_\phi \cap \mathrm{Sp}_{2m}(\mathbb{C}) = \{\pm 1, \pm 1, \dots, \pm 1\}$. So $S^\circ_{\phi,\mathrm{sc}} = 1$. Hence, we can see

$$\widehat{Z}_{\mathrm{sc}} = Z(\mathrm{Sp}_{2m}(\mathbb{C})) = \mu_2 \quad \text{and} \quad \widehat{Z}_{\mathrm{sc}} \cap S^\circ_{\phi,\mathrm{sc}} = 1.$$

Thus (2-4) and (2-6) prove the lemma.

Assume $n = 2m$. Then (2-3) is written as

$$\mathrm{Spin}_{2m}(\mathbb{C}) \twoheadrightarrow \mathrm{SO}_{2m}(\mathbb{C}) \subseteq \mathrm{GSO}_{2m}(\mathbb{C}).$$

Hence from (2-2) one can see

$$S_\phi \cap \mathrm{SO}_{2m}(\mathbb{C}) = \{\pm 1, \pm 1, \dots, \pm 1\},$$

which gives

$$S^\circ_{\phi,\mathrm{sc}} = 1.$$

Since $Z(\mathrm{Spin}_{2m}(\mathbb{C})) = \mu_2 \times \mu_2$ or μ_4, we have the lemma from (2-4) and (2-6). □

This lemma immediately implies the following, which is a part of Conjecture 1.5 in the introduction.

Corollary 2.8. *For the* (*conjectural*) *global Langlands parameters ϕ_n and ϕ_{n+1} of π_n and π_{n+1}, respectively, we have*

$$4|\mathcal{S}_{\phi_n}|\,|\mathcal{S}_{\phi_{n+1}}| = \tfrac{1}{2}|\mathcal{S}_{\phi_n,\mathrm{sc}}|\,|\mathcal{S}_{\phi_{n+1},\mathrm{sc}}|.$$

3. Local integrals of matrix coefficients

In this section, we take up the local intertwining map $\alpha^\natural_{\omega_v}$ in (1-1) and firstly prove that the integral that defines $\alpha^\natural_{\omega_v}$ converges at every v, assuming the representations are tempered, and then secondly compute the integral for unramified data, which essentially follows from [Ichino and Ikeda 2010].

In this section, everything is purely local, and hence we suppress the subscript v from our notation, and in particular F will denote a local field.

3A. ***Some general lemmas.*** In this first subsection, we will prove a couple of lemmas in harmonic analysis, which apply to any connected reductive group over local F. (Though those two lemma might be known to experts, we will give our proofs here because we are not able to locate them in the literature.) Accordingly, in this subsection we let G be any connected reductive group over F.

First, as usual, we define

$$G^1 := \bigcap_{\chi \in \mathrm{Rat}(G)} \ker|\chi|_F,$$

where $\mathrm{Rat}(G)$ is the set of all rational characters on G. Then we have the following.

Lemma 3.1. *Let G be a reductive group and let Z° be the identity component of the center of G. Let $f : G \to \mathbb{C}$ be a measurable function such that $f(zg) = f(g)$ for all $z \in Z^\circ$ and $g \in G$. Then $\int_{Z^\circ\backslash G} f(g)\,dg$ converges absolutely if and only if $\int_{G^1} f(g)\,dg$ converges absolutely.*

Proof. Set $Z^1 = Z^\circ \cap G^1$ and let $\mathrm{pr} : Z^\circ\backslash G \to G^1 Z^\circ\backslash G$ be the natural projection. Then one can readily see that $\ker(\mathrm{pr}) = Z^\circ\backslash G^1 Z^\circ \cong Z^1\backslash G^1$, which implies

$$G^1 Z^\circ\backslash G \cong (Z^1\backslash G^1)\backslash(Z^\circ\backslash G).$$

Thus, we can compute

$$\begin{aligned}\int_{Z^\circ\backslash G} f(g)\,dg &= \int_{(Z^1\backslash G^1)\backslash(Z^\circ\backslash G)} \int_{Z^1\backslash G^1} f(g_1\bar{g})\,dg_1 d\bar{g}\\ &= \int_{G^1 Z^\circ\backslash G} \int_{Z^1\backslash G^1} f(g_1\bar{g})\,dg_1 d\bar{g}\\ &= \sum_{\bar{g}\in G^1 Z^\circ\backslash G} \int_{Z^1\backslash G^1} f(g_1\bar{g})\,dg_1,\end{aligned}$$

where the last equality follows because $G^1 Z^\circ\backslash G$ is finite. Hence we have

$$\int_{Z^\circ\backslash G} |f(g)|\,dg < \infty \quad \text{if and only if} \quad \int_{Z^1\backslash G^1} |f(g)|\,dg < \infty.$$

On the other hand, using the invariance of the measure we also have

$$\begin{aligned}\int_{G^1} f(g)\,dg &= \int_{Z^1\backslash G^1} \int_{Z^1} f(ag)\,da\,dg\\ &= \int_{Z^1\backslash G^1} \int_{Z^1} f(g)\,da\,dg\\ &= \mathrm{Vol}(Z^1) \int_{Z^1\backslash G^1} f(g)\,dg,\end{aligned}$$

since $Z^1 = Z^\circ \cap G^1$ is compact. Hence

$$\int_{Z^1\backslash G^1} |f(g)|\,dg < \infty \quad \text{if and only if} \quad \int_{G^1} |f(g)|\,dg < \infty. \qquad \square$$

To state the second lemma, let us fix a special maximal compact subgroup K of G, and a Levi part $M_{\min}$ of a minimal parabolic of G. Then we have a Cartan decomposition $G = K M_{\min}^+ K$ where

$$M_{\min}^+ := \{m \in M_{\min} : |\alpha_i(m)| \le 1 \text{ for } i = 1, \dots, r\},$$

where the α_i correspond to the simple roots. From equation (4) of [Waldspurger 2003], we have

$$G = \coprod_{m \in M_{\min}^{+}/M_{\min}^{1}} KmK, \tag{3-1}$$

where

$$M_{\min}^{1} := M_{\min} \cap G^{1}.$$

Then we have the following lemma.

Lemma 3.2. *Let F be nonarchimedean. For $f \in L^1(G)$,*

$$\int_G f(g)\,dg = \int_{M_{\min}^{+}} \mu(m) \int_{K \times K} f(k_1 m k_2)\,dk_1 dk_2 dm$$

where

$$\mu(m) = \frac{\mathrm{Vol}(KmK)}{\mathrm{Vol}(M_{\min}^{1})} = C \cdot \mathrm{Vol}(KmK),$$

for some positive constant C.

Proof. Using [Silberger 1979, page 149] we have

$$\begin{aligned}
\int_G f(x)\,dx &= \sum_{m \in M_{\min}^{+}/M_{\min}^{1}} \int_{KmK} f(x)\,dx \\
&= \sum_{m \in M_{\min}^{+}/M_{\min}^{1}} \int_{K \times K} \mathrm{Vol}(KmK) f(k_1 m k_2)\,dk_1 dk_2 \\
&= \sum_{m \in M_{\min}^{+}/M_{\min}^{1}} \mathrm{Vol}(KmK) \int_{K \times K} f(k_1 m k_2)\,dk_1 dk_2 \\
&= \mathrm{Vol}(M_{\min}^{1}) \sum_{m \in M_{\min}^{+}/M_{\min}^{1}} \mu(m) \int_{K \times K} f(k_1 m k_2)\,dk_1 dk_2 \\
&= \int_{M_{\min}^{+}/M_{\min}^{1}} \int_{M_{\min}^{1}} \mu(m m_1) \int_{K \times K} f(k_1 m m_1 k_2)\,dk_1 dk_2 dm_1 dm \\
&= \int_{M_{\min}^{+}} \mu(m) \int_{K \times K} f(k_1 m k_2)\,dk_1 dk_2 dm,
\end{aligned}$$

where we used that $\mathrm{Vol}(M_{\min}^{1})$ is finite because $M_{\min}^{1}$ is compact. □

Remark 3.3. In the archimedean case, a similar integral formula as in Lemma 3.2 holds (see, for example, [Helgason 2000, Theorem 5.8]).

3B. ***Convergence of the integral.*** By using the two lemmas in the previous subsection we are now in a position to prove the convergence of the integral in (1-1). Hence in this subsection, we specialize to $G = \mathrm{GSpin}(V_n)$, where (V_n, q_n) is an n-dimensional quadratic space over F. In this case, we have

$$G^1 = \{g \in \mathrm{GSpin}(V_n) : |N(g)| = 1\}.$$

Also note that we have a Witt decomposition $V = X \oplus V_n^{\mathrm{an}} \oplus Y$, where X and Y are totally isotropic spaces and V_n^{an} is the anisotropic part. By fixing a basis for X, we obtain a minimal parabolic subgroup $P_{\min} = M_{\min} N_{\min}$ of G with

$$M_{\min} = \underbrace{\mathrm{GL}_1 \times \mathrm{GL}_1 \cdots \times \mathrm{GL}_1}_{r-\text{times}} \times \mathrm{GSpin}(V^{\mathrm{an}}),$$

where r is the Witt rank of V_n, which is by definition the dimension of X. Then one can see that

$$M_{\min}^{+} = \{(x_1, x_2, \dots, x_r, g_{\mathrm{an}}) : |x_i| \le |x_{i+1}|\}.$$

We define

$$M_{\min}^{+,1} = \{m \in M_{\min}^{+} : |\nu(m)| = 1\} = \{m \in M_{\min}^{+} : |\nu(g_{an})| = 1\}.$$

The maximal torus $A_{\min}$ of $M_{\min}$ is of the form

$$A_{\min} = \underbrace{\mathrm{GL}_1 \times \mathrm{GL}_1 \cdots \times \mathrm{GL}_1}_{r-\text{times}} \times \mathrm{GL}_1 = \{(x_1, x_2, \dots, x_r, x_0)\}.$$

We then define

$$\begin{aligned} A_{\min}^{+,1} &:= \{a \in A_{\min} : |\alpha_i(a)| \le 1 \text{ for } i = 1, 2, \dots r \text{ and } a \in G^1\} \\ &= \{(x_1, x_2, \dots, x_r, x_0) : |x_i| \le |x_{i+1}|,\ 1 \le i \le n-1 \text{ and } |x_0| = 1\}. \end{aligned}$$

Let $\delta_{P_{\min}}$ be the modulus character of $P_{\min}$. Then

$$\delta_{P_{\min}}(x) = \prod_{i=1}^{r} |x_i|^{d+2r-2i}.$$

Now, let us get to the integral we would like to show to be convergent. Assume we have $(V_n, q_n) \subseteq (V_{n+1}, q_{n+1})$, so that we have $\mathrm{GSpin}(V_n) \subseteq \mathrm{GSpin}(V_{n+1})$. For simplicity, we write $G_n = \mathrm{GSpin}(V_n)$ and $G_{n+1} = \mathrm{GSpin}(V_{n+1})$. Let π_i be a tempered representation of G_i such that $\omega_{\pi_n} \omega_{\pi_{n+1}} = \omega^2$ for some ω. Then the integrand for $\alpha_\omega^\natural$ in (1-1) is a product of matrix coefficients of π_n and π_{n+1} together with $\omega^{-1}(N(g))$. Hence the convergence of the integral boils down to the following.

Proposition 3.4. *Keep the above notation and assumption, so in particular assume π_n and π_{n+1} are tempered. Then for all the matrix coefficients Φ_{n+1} and Φ_n of*

π_{n+1} *and* π_n, *respectively*, *the integral*

$$\int_{Z_n^\circ \backslash G_n} \Phi_{n+1}(g)\Phi_n(g)\omega^{-1}(N(g))\,dg$$

is absolutely convergent, *recalling that* $Z_n^\circ = \mathrm{GL}_1$ *even when* $n = 2$.

Proof. Let us first mention that, in this proof, for each G_i we denote the various subgroups introduced above by $P_{i,\min}$, $A_{i,\min}$, $M^1_{i,\min}$, $M^{+,1}_{i,\min}$, $A^{+,1}_{i,\min}$, etc., and also we denote the Witt index of V_i by r_i.

Assume $n > 2$. Then Z_n° is indeed the identity component of the center and so by Lemma 3.1 it suffices to show the absolute convergence of

$$\int_{G_n^1} \Phi_{n+1}(g)\Phi_n(g)\omega^{-1}(N(g))\,dg.$$

Using (3-1) we have

$$G_n^1 = \Bigl(\coprod_{m\in M^+_{n,\min}/M^1_{n,\min}} K_n m K_n\Bigr)\bigcap G_n^1 = \coprod_{m\in (M^+_{n,\min}/M^1_{n,\min})\cap G_n^1} K_n m K_n.$$

Then by Lemma 3.2, the convergence of the integral is reduced to the convergence of

$$\int_{M^{+,1}_{n,\min}} \mu(m)\int_{K_n\times K_n} |\Phi_{n+1}(k_1mk_2)|\,|\Phi_n(k_1mk_2)|\,dk_1dk_2dm$$

where $\mu(m) = C\cdot \mathrm{Vol}(K_n m K_n)$ for some positive constant C.

Furthermore, since Φ_n and Φ_{n+1} are matrix coefficients of tempered representations, they satisfy, for any $g \in G_i$,

$$|\Phi_i(g)| \le A\,\Xi_i(g)(1+\sigma_i(g))^B$$

for some positive constants A and B, where Ξ_i and σ_i are, respectively, Harish-Chandra's spherical function and a height function on G_i. (See [Waldspurger 2003, page 274].) Note that here we may and do assume $\sigma_{n+1}|_{G_n} = \sigma_n$ and simply write σ for both. Since both Ξ_i and σ are K_i-bi-invariant, and $K_n \times K_n$ is compact, the convergence of the integral reduces to the convergence of

$$\int_{M^{+,1}_{n,\min}} \mu(m)\,\Xi_{n+1}(m)\,\Xi_n(m)(1+\sigma(m))^{2B}\,dm.$$

By [Silberger 1979, Theorem 4.2.1, page 154; Waldspurger 2003, Lemma II.1.1], there exist positive constants A and B such that

$$A^{-1}\delta^{1/2}_{P_{i,\min}}(m) \le \Xi_i(m) \le A\delta^{1/2}_{P_{i,\min}}(m)(1+\sigma(m))^B$$

for any $m \in M_{i,\min}^{+,1}$. So the convergence of the integral is reduced to the convergence of

$$\int_{M_{n,\min}^{+,1}} \mu(m)(\delta_{P_{n+1,\min}})^{1/2}(m)(\delta_{P_{n,\min}})^{1/2}(m)(1+\sigma(m))^{2B}\,dm.$$

Moreover, there exists a positive constant A such that $\mu(m) \le A\delta_{P_{n,\min}}^{-1}$ for any $m \in M_{n,\min}^{+,1}$ [Waldspurger 2003, page 241]. So it is enough to show that

$$\int_{M_{n,\min}^{+,1}} (\delta_{P_{n,\min}})^{-1/2}(m)(\delta_{P_{n+1,\min}})^{1/2}(m)(1+\sigma(m))^{2B}\,dm$$

converges absolutely.

When n is even, $A_{n,\min}^{+,1}$ sits inside of $A_{n+1,\min}^{+,1}$. Thus, the convergence of the integral is reduced to the convergence of

$$\int_{A_{n,\min}^{+,1}} (\delta_{P_{n,\min}})^{-1/2}(m)(\delta_{P_{n+1,\min}})^{1/2}(m)(1+\sigma(m))^{2B}\,dm.$$

Hence, the convergence of the integral is reduced to the convergence of

$$\begin{aligned}\int_{|x_0|=1}\int_{|x_1|\le\cdots\le|x_{r_n}|\le 1} |x_1\cdots x_{r_n}|^{1/2}\biggl(1-\sum_{j=1}^{r_n}\log|x_j|\biggr)^{2B} d^\times x_1\cdots d^\times x_{r_n} d^\times x_0 \\ = \int_{|x_0|=1} d^\times x_0 \int_{|x_1|\le\cdots\le|x_{r_n}|\le 1} |x_1\cdots x_{r_n}|^{1/2}\biggl(1-\sum_{j=1}^{r_n}\log|x_j|\biggr)^{2B} d^\times x_1\cdots d^\times x_{r_n}.\end{aligned}$$

Since $\int_{|x_0|=1} d^\times x_0$ is an integral over a compact set, the convergence of the integral is reduced to the convergence of

$$\int_{|x_1|\le\cdots\le|x_{r_n}|\le 1} |x_1\cdots x_{r_n}|^{1/2}\biggl(1-\sum_{j=1}^{r_n}\log|x_j|\biggr)^{2B} d^\times x_1\cdots d^\times x_{r_n}$$

which is precisely the integral considered in [Ichino and Ikeda 2010, page 1388].

When n is odd, $A_{n,\min}^{+,1}$ is not a subset of $A_{n+1,\min}^{+,1}$. Hence, the convergence of the integral in this case is reduced to the convergence of

$$\begin{aligned}\int_{|x_0|=1}\int_{|x_1|\le\cdots\le|x_{r_n}|\le 1} |x_1\cdots x_{r_n}|^{1/2}\biggl(1-\sum_{j=1}^{r_n}\log|x_j|\biggr)^{2B} d^\times x_1\cdots d^\times x_{r_n} \\ + \int_{|x_1|\le\cdots\le|x_{r_n-1}|\le|x_{r_n}|^{-1}\le 1} |x_1\cdots x_{r_n-1}x_{r_n}^{-1}|^{1/2} \\ \biggl(1-\sum_{j=1}^{r_n-1}\log|x_j|+\log|x_{r_n}|\biggr)^{2B} d^\times x_1\cdots d^\times x_{r_n} d^\times x_0\end{aligned}$$

$$= \int_{|x_0|=1} d^\times x_0 \int_{|x_1|\le\cdots\le|x_{r_n}|\le 1} |x_1\cdots x_{r_n}|^{1/2}\Bigl(1-\sum_{j=1}^{r_n}\log|x_j|\Bigr)^{2B} d^\times x_1\cdots d^\times x_{r_n}$$
$$+\int_{|x_1|\le\cdots\le|x_{r_n-1}|\le|x_{r_n}|^{-1}\le 1} |x_1\cdots x_{r_n-1}x_{r_n}^{-1}|^{1/2}$$
$$\Bigl(1-\sum_{j=1}^{r_n-1}\log|x_j|+\log|x_{r_n}|\Bigr)^{2B} d^\times x_1\cdots d^\times x_{r_n} d^\times x_0.$$

Since $\int_{|x_0|=1} d^\times x_0$ is an integral over a compact set, the convergence of the integral is reduced to the convergence of

$$\int_{|x_1|\le\cdots\le|x_{r_n}|\le 1} |x_1\cdots x_{r_n}|^{1/2}\Bigl(1-\sum_{j=1}^{r_n}\log|x_j|\Bigr)^{2B} d^\times x_1\cdots d^\times x_{r_n}$$
$$+\int_{|x_1|\le\cdots\le|x_{r_n-1}|\le|x_{r_n}|^{-1}\le 1} |x_1\cdots x_{r_n-1}x_{r_n}^{-1}|^{1/2}$$
$$\Bigl(1-\sum_{j=1}^{r_n-1}\log|x_j|+\log|x_{r_n}|\Bigr)^{2B} d^\times x_1\cdots d^\times x_{r_n}.$$

This integral is precisely the one considered in [Ichino and Ikeda 2010, page 1388].

Lastly, assume $n=2$. In this case, we have seen $G_2 = F^\times\times F^\times$ or $E^\times$. If $G_2 = E^\times$ then $\mathrm{GL}_1\backslash G_2 = E^1$, which is compact, and hence the convergence of the integral is immediate. If $G_2 = F^\times\times F^\times$ then $\mathrm{GL}_1\backslash G_2 = F^\times$, in which case we can apply the above argument by using the estimate of the matrix coefficient, and indeed the computation is easier and left to the reader. □

3C. ***Calculation of integrals in the unramified case.*** In this subsection, we consider the unramified integral. Accordingly, we assume that all the data are unramified. To be precise, we assume

(1) G_i is unramified over F;

(2) K_i is a hyperspecial maximal compact subgroup of G_i;

(3) $K_n\subset K_{n+1}$;

(4) π_i is an unramified representation of G_i;

(5) $\int_{K_i} dg_i = 1$.

Furthermore, let ω be the unramified character such that $\omega^2 = \omega_{\pi_{n+1}}\omega_{\pi_n}$. Note that there is a unique such ω.

Then we have the following.

Proposition 3.5. *Under the above assumptions, let $\phi^\circ \in \pi_{n+1}$ and $f^\circ \in \pi_n$ be the spherical vectors such that*

$$\mathcal{B}_{\pi_{n+1}}(\phi^\circ, \phi^\circ) = 1 \quad \textit{and} \quad \mathcal{B}_{\pi_n}(f^\circ, f^\circ) = 1.$$

Then we have

$$\alpha^\sharp_\omega(\phi^\circ, f^\circ) = \Delta_{\mathrm{SO}_{n+1}} \cdot \frac{L(1/2, \pi_n \times \pi_{n+1} \otimes \omega^{-1})}{L(1, \pi_{n+1}, \mathrm{Ad}) L(1, \pi_n, \mathrm{Ad})}.$$

Proof. Let

$$\Phi^\circ_{\pi_{n+1}}(g) := \mathcal{B}_{\pi_{n+1}}(\pi_{n+1}(g)\phi^\circ, \phi^\circ) \quad \text{and} \quad \Phi^\circ_{\pi_n}(g) := \mathcal{B}_{\pi_n}(\pi_n(g) f^\circ, f^\circ);$$

namely they are the normalized spherical matrix coefficients so that $\Phi^\circ_{n+1}(1) = \Phi^\circ_n(1) = 1$. Since $\omega_{\pi_{n+1}}$ and ω_{π_n} are unramified, there exist unique unramified square roots $\omega^{1/2}_{\pi_{n+1}}$ and $\omega^{1/2}_{\pi_n}$. Let us denote

$$\overline{\pi_{n+1}} := \pi_{n+1} \otimes \omega^{-1/2}_{\pi_{n+1}} \circ N \quad \text{and} \quad \overline{\pi_n} := \pi_n \otimes \omega^{-1/2}_{\pi_n} \circ N,$$

which have trivial central characters and hence viewed as representations of SO_{n+1} and SO_n, respectively. Then one can readily see that

$$\begin{aligned}\Phi^\circ_{\pi_{n+1}}(g)\omega^{-1/2}_{\pi_{n+1}}(N(g)) &= \mathcal{B}_{\pi_{n+1}}(\pi_{n+1}(g)\omega^{-1/2}_{\pi_{n+1}}(N(g))\phi^\circ, \phi^\circ) \\ &= \Phi^\circ_{\overline{\pi_{n+1}}}(g),\end{aligned}$$

where $\Phi^\circ_{\overline{\pi_{n+1}}}$ is the normalized spherical matrix coefficient of $\overline{\pi_{n+1}}$ which gives $\Phi^\circ_{\overline{\pi_{n+1}}}(1) = 1$. Similarly, we have

$$\Phi^\circ_{\pi_{n+1}}(g)\omega^{-1/2}_{\pi_{n+1}}(N(g)) = \Phi^\circ_{\overline{\pi_{n+1}}}(g).$$

Hence by using Theorem 1.2 in [Ichino and Ikeda 2010] we have the following:

$$\begin{aligned}\alpha^\sharp_\omega(\phi^\circ, f^\circ) &= \int_{Z^\circ_n \backslash G_n} \Phi^\circ_{\pi_{n+1}}(g)\Phi^\circ_{\pi_{n+1}}(g)\omega^{-1/2}_{\pi_{n+1}}(N(g))\omega^{-1/2}_{\pi_n}(N(g))\, dg \\ &= \int_{\mathrm{SO}_n} \Phi^\circ_{\overline{\pi_{n+1}}}(g)\Phi^\circ_{\overline{\pi_n}}(g)\, dg \\ &= \Delta_{\mathrm{SO}_{n+1}} \cdot \frac{L(1/2, \overline{\pi_n} \times \overline{\pi_{n+1}})}{L(1, \overline{\pi_{n+1}}, \mathrm{Ad}) L(1, \overline{\pi_n}, \mathrm{Ad})}.\end{aligned}$$

Now one can readily see that

$$L(s, \overline{\pi_n} \times \overline{\pi_{n+1}}) = L(s, \pi_n \otimes \omega^{-1/2}_{\pi_n} \times \pi_{n+1} \otimes \omega^{-1/2}_{\pi_{n+1}}) = L(s, \pi_n \times \pi_{n+1} \otimes \omega^{-1})$$

because $\omega^{-1/2}_{\pi_n}\omega^{-1/2}_{\pi_{n+1}} = \omega^{-1}$. Also one can see that

$$L(s, \overline{\pi_{n+1}}, \mathrm{Ad}) = L(s, \pi_{n+1}, \mathrm{Ad}) \quad \text{and} \quad L(s, \overline{\pi_{n+1}}, \mathrm{Ad}) = L(s, \pi_{n+1}, \mathrm{Ad})$$

by definition of the adjoint L-function. The proposition follows. $\square$

Remark 3.6. Although the local calculations in the unramified section follow from [Ichino and Ikeda 2010], this is only possible because the square root always exists for the unramified case. In the ramified or global cases, we may not assume this.

4. Wrap-up of the conjecture and low rank cases

In this section, let us first wrap up our conjecture and then prove low rank cases. So in this section we let F be a number field and $\mathbb{A}$ the ring of adeles.

4A. ***Wrap-up.*** Assume π_n and π_{n+1} are tempered cuspidal automorphic representations of $G_n(\mathbb{A}) = \mathrm{GSpin}_n(\mathbb{A})$ and $G_{n+1}(\mathbb{A}) = \mathrm{GSpin}_{n+1}(\mathbb{A})$ such that there exists a Hecke character ω with $\omega^2 = \omega_{\pi_n}\omega_{\pi_{n+1}}$. Fix the tensor product decompositions $\pi_n = \otimes_v \pi_{n,v}$ and $\pi_{n+1} = \otimes_v \pi_{n+1,v}$ and fix factorizable $\phi = \otimes_v \phi_v$ in π_{n+1} and $f = \otimes_v f_v$ in π_n.

First of all, for almost all v, the assumptions of Proposition 3.5 are satisfied and hence we have

$$\alpha_{\omega_v}(\phi_v, f_v) = \frac{L(1, \pi_{n+1}, \mathrm{Ad})L(1, \pi_n, \mathrm{Ad})}{\Delta_{\mathrm{SO}_{n+1}} \cdot L(1/2, \pi_n \times \pi_{n+1} \otimes \omega^{-1})} \alpha^{\natural}_{\omega_v}(\phi_v, f_v) = 1,$$

where ϕ_v and f_v are the spherical vectors as in the previous section. Thus the infinite product

$$\prod_v \alpha_{\omega_v}(\phi_v, f_v)$$

is well defined.

Accordingly, we can and do form the following conjecture:

Conjecture 4.1 (the global GGP conjecture for GSpin). *With the assumptions stated above, if π_{n+1} and π_n appear with multiplicity one in the discrete spectrum, then*

$$|\mathcal{P}_\omega(\phi, f)|^2 = \frac{\Delta_{\mathrm{SO}_{n+1}}}{2^\beta} \frac{L(1/2, \pi_n \times \pi_{n+1} \otimes \omega^{-1})}{L(1, \pi_{n+1}, \mathrm{Ad})L(1, \pi_n, \mathrm{Ad})} \prod_v \alpha_{\omega_v}(\phi_v, f_v),$$

where

$$\Delta_{\mathrm{SO}_{n+1}} := \begin{cases} \zeta(2)\zeta(4)\cdots\zeta(2m) & \textit{if } \dim V_{n+1} = 2m+1, \\ \zeta(2)\zeta(4)\cdots\zeta(2m-2)\cdot L(m, \chi_{V_{n+1}}) & \textit{if } \dim V_{n+1} = 2m, \end{cases}$$

where $\chi_{V_{n+1}}$ is the quadratic character associated with V_{n+1}.

Here we are assuming the L-function $L(s, \pi_n \times \pi_{n+1} \otimes \omega^{-1})$ is holomorphic at $s = \frac{1}{2}$ and the adjoint L-functions $L(s, \pi_{n+1}, \mathrm{Ad})$ and $L(s, \pi_{n+1}, \mathrm{Ad})$ are nonzero and holomorphic at $s = 1$. As discussed in the introduction, we also make the following conjecture regarding the constant 2^β:

Conjecture 4.2. *Let ϕ_{n+1} and ϕ_n be the* (*conjectural*) *global Langlands parameters of π_{n+1} and π_n, respectively. If π_{n+1} and π_n appear with multiplicity one in the discrete spectrum*, *then*

$$2^\beta = 4|\mathcal{S}_{\phi_n}|\,|\mathcal{S}_{\phi_{n+1}}| = \tfrac{1}{2}|\mathcal{S}_{\phi_n,\mathrm{sc}}|\,|\mathcal{S}_{\phi_{n+1},\mathrm{sc}}|.$$

Now let us note the relation between our conjecture and that of Ichino and Ikeda. Assume π_n and π_{n+1} both have the trivial central character, so that we can choose $\omega = \mathbf{1}$. Then π_n and π_{n+1} can be viewed as automorphic representations of $\mathrm{SO}_n(\mathbb{A})$ and $\mathrm{SO}_{n+1}(\mathbb{A})$. (Conjecturally, this means that if ϕ_n is the global L-parameter of π_n then the image of ϕ_n is already in $\mathrm{Sp}_{2m}(\mathbb{C})$ or $\mathrm{O}_{2m}(\mathbb{C})$ depending on the parity of n, and similarly for π_{n+1}.) Then one can see that the tensor product L-function $L(s, \pi_n \times \pi_{n+1} \otimes \mathbf{1})$ as the L-function for GSpin is equal to the tensor product L-function $L(s, \pi_n \times \pi_{n+1})$ as the L-function for SO, and similarly for the adjoint L-functions. Hence in this case, our conjecture is precisely that of Ichino and Ikeda. In this sense, our conjecture should be interpreted as a generalization of that of Ichino and Ikeda instead of an analogue of it.

4B. ***Conjecture for* $\mathbf{GSpin_2 \times GSpin_3}$ (*Waldspurger formula case*).** Let us consider the lowest rank case, so we let V_2 and V_3 be quadratic spaces of dimensions 2 and 3. Note then that $V_2 = E$, where E is a quadratic extension of F equipped with the norm form or $V_2 = \mathbb{H}$ (hyperbolic plane), and there exists $V_3 = D_0$, where D_0 is the set of trace zero elements of a (not necessarily division) quaternion algebra D equipped with the norm form. Recall in Section 2B we have computed

$$\begin{aligned} \mathrm{GSpin}_2 &= \mathrm{GSpin}(V_2) = \begin{cases} \mathrm{Res}_{E/F}\,\mathrm{GL}_1 & \text{if } V_2 = E; \\ \mathrm{GL}_1 \times \mathrm{GL}_1 & \text{if } V_2 = \mathbb{H}, \end{cases} \\ \mathrm{GSpin}_3 &= \mathrm{GSpin}(V_3) = D^\times. \end{aligned}$$

In this subsection, we consider the case $V_2 = E$, and assume D is such that we have an inclusion $E \subseteq D_0$, which gives the inclusion $\mathrm{Res}_{E/F}\,\mathrm{GL}_1 \subseteq D^\times$. This is essentially the case treated in [Waldspurger 1985] and the resulting formula is normally known as the Waldspurger formula.

So we let π_2 be a cuspidal automorphic representation on $\mathrm{GSpin}_2(\mathbb{A}) = \mathbb{A}_E^\times$, namely a Hecke character χ on $\mathbb{A}_E^\times$ and let π_3 be a tempered cuspidal automorphic representation of $D^\times(\mathbb{A})$ such that there exists a Hecke character ω on $\mathbb{A}$ with

$$\omega^2 = \omega_{\pi_3}\chi|_{\mathbb{A}_F^\times}.$$

Consider $\pi_3 \otimes \omega^{-1}$, which is an automorphic representation of $D^\times(\mathbb{A})$ with the central character $\omega_{\pi_3\otimes\omega^{-1}} = \omega^{-2}\omega_{\pi_3}$, so that

$$\omega_{\pi_3\otimes\omega^{-1}} \cdot \chi|_{\mathbb{A}^\times} = 1.$$

Then for each $\phi \in V_{\pi_3}$ and $\chi \in V_\chi$, our period integral is

$$\mathcal{P}_\omega(\phi, \chi) = \int_{E^\times \backslash \mathbb{A}_E^\times / \mathbb{A}^\times} \phi(g) f(g) \omega(N(g))^{-1} \, dg = \int_{E^\times \backslash \mathbb{A}_E^\times / \mathbb{A}^\times} (\phi \cdot \omega^{-1})(g) \chi(g) \, dg,$$

which is nothing but the period integral considered by Waldspurger for $\pi_3 \otimes \omega^{-1}$ and χ. Hence by using the Waldspurger formula, we obtain

$$\begin{aligned} |\mathcal{P}_\omega(\phi, \chi)|^2 &= \frac{\zeta(2) L(1/2, BC(\pi_3 \otimes \omega^{-1}) \otimes \chi)}{4L(1, \mu)^2 L(1, \pi_3 \otimes \omega^{-1}, \mathrm{Ad})} \prod_v \alpha_{\omega_v}(\phi_v, \chi_v) \\ &= \frac{\zeta(2) L(1/2, BC(\pi_3 \otimes \omega^{-1}) \otimes \chi)}{4L(1, \chi, \mathrm{Ad}) L(1, \pi_3, \mathrm{Ad})} \prod_v \alpha_{\omega_v}(\phi_v, \chi_v) \\ &= \frac{\zeta(2) L(s, \pi_3 \otimes \omega^{-1} \times \chi)}{4L(1, \chi, \mathrm{Ad}) L(1, \pi_3, \mathrm{Ad})} \prod_v \alpha_{\omega_v}(\phi_v, \chi_v), \end{aligned}$$

where for the first equality we used the Waldspurger formula with μ the quadratic character for the extension E/F, and for the last equality we used [Bump 1997, page 102]. This confirms Conjecture 1.4. Moreover, $|\mathcal{S}_{\phi_2}| = |\mathcal{S}_{\phi_3}| = 1$, so $2^\beta = 4|\mathcal{S}_{\phi_2}|\,|\mathcal{S}_{\phi_3}|$, confirming Conjecture 1.5.

Remark 4.3. Waldspurger assumed the central character of π is trivial, but in [Yuan et al. 2013], this condition was removed. Also the Waldspurger formula we used in the above looks slightly different from the original in [Waldspurger 1985] or from the Waldspurger formula listed in [Yuan et al. 2013, Theorem 1.4 in Section 1.4.2]. This is due to the following: Waldspurger chose the global Haar measure used in the period integral such that the $\mathrm{Vol}(E^\times \backslash \mathbb{A}_E^\times / \mathbb{A}^\times) = 2L(1, \mu)$. Yuan et al. [2013] chose the global Haar measure such that the volume is 1. Then each chose local measures to be compatible with their choice of global measure. With our choice of measures, our formulation is equivalent.

4C. ***Conjecture for*** $\mathbf{GSpin_2 \times GSpin_3}$ **(*Jacquet–Langlands case*).** Next consider the case $V_2 = \mathbb{H}$, so that $\mathrm{GSpin}_2 = \mathrm{GL}_1 \times \mathrm{GL}_1$. In this case we have an embedding $\mathrm{GSpin}_2 \subseteq \mathrm{GSpin}_3$ only when $\mathrm{GSpin}_3 = \mathrm{GL}_2$; namely D in the previous subsection is split.

Before moving on, let us mention that this case is actually excluded from our conjecture in the first place. Indeed, as we will see, even though we can obtain a similar formula by using the well-known Jacquet–Langlands theory, the resulting formula is not exactly as in our conjecture. We consider this case merely as a low rank exception.

Let π_2 be a tempered cuspidal automorphic representation of $\mathrm{GL}_1(\mathbb{A}) \times \mathrm{GL}_1(\mathbb{A})$, so that we have $\pi_2 = \chi_1 \boxtimes \chi_2$ where χ_1, χ_2 are both unitary Hecke characters of $\mathbb{A}^\times$, and let π_3 be a tempered cuspidal automorphic representation of $\mathrm{GL}_2(\mathbb{A})$

with central character ω_{π_3}. By our assumption, there exists a Hecke character ω such that

$$\omega^2 = \chi_1\chi_2\omega_{\pi_3}.$$

Proposition 4.4. *Keep the above notation and assumption. Then for $\phi \in V_{\pi_3}$ and $\chi_1 \boxtimes \chi_2 \in V_{\pi_2}$, we have*

$$|\mathcal{P}_\omega(\phi, \chi_1 \boxtimes \chi_2)|^2 = \frac{\zeta(2)L(1/2, \pi_3 \times \pi_2 \otimes \omega^{-1})}{2|\mathcal{S}_{\phi_2}|\,|\mathcal{S}_{\phi_3}|L(1, \pi_3, \mathrm{Ad})} \prod_v \alpha_{\omega_v}(\phi_v, \chi_{1,v} \boxtimes \chi_{2,v}).$$

Proof. This is a standard exercise using the well-known Jacquet–Langlands theory as well as [Waldspurger 1985, Proposition 6, page 208] and that $|\mathcal{S}_{\phi_2}| = |\mathcal{S}_{\phi_3}| = 1$. The details are left to the reader. $\square$

Note that Proposition 4.4 is similar to Conjectures 1.4 and 1.5.

4D. ***Conjecture for* GSpin$_4$ × GSpin$_3$ *(triple product formula).*** In this section we prove the conjecture for GSpin_4 and GSpin_3, which essentially boils down to Ichino's triple product formula. So we let V_3 and V_4 be quadratic spaces of dimension three and four, respectively and write $G_3 = \mathrm{GSpin}(V_3)$ and $G_4 = \mathrm{GSpin}(V_4)$. Then there exists a (not necessarily division) quaternion algebra D such that

$$G_3(\mathbb{A}) = D^\times(\mathbb{A}),$$
$$G_4(\mathbb{A}) = \begin{cases} \{(g_1, g_2) \in D^\times(\mathbb{A}) \times D^\times(\mathbb{A}) : N_D(g_1) = N_D(g_2) \in \mathbb{A}^\times\}; \\ \{g \in D^\times(\mathbb{A}_E) : N(g) \in \mathbb{A}^\times\}, \end{cases}$$

where for G_4 the first is the case if $\mathrm{disc}(V_4) = 1$ and the second is the case if $\mathrm{disc}(V_4) \neq 1$. Here, to be more precise, we are assuming that V_3 and V_4 are such that the corresponding G_3 and G_4 are as above with the same D, so that we have the inclusion $G_3(\mathbb{A}) \subseteq G_4(\mathbb{A})$.

To utilize Ichino's triple product formula, we need to introduce the group

$$\tilde{G}_4(\mathbb{A}) = \begin{cases} D^\times(\mathbb{A}) \times D^\times(\mathbb{A}) & \text{if disc } V_4 = 1; \\ D^\times(\mathbb{A}_E) & \text{if disc } V_4 \neq 1, \end{cases}$$

so that we have

$$G_3(\mathbb{A}) \subseteq G_4(\mathbb{A}) \subseteq \tilde{G}_4(\mathbb{A}).$$

Now let π_i be a tempered cuspidal automorphic representation of G_i for $i = 3, 4$ such that there exists a Hecke character ω with $\omega^2 = \omega_{\pi_4}\omega_{\pi_3}$.

First, we assume $\mathrm{disc}\, V_4 = 1$. To use Ichino's triple product formula, we need to relate π_4 with an automorphic representation of $\tilde{G}_4(\mathbb{A}) = D^\times(\mathbb{A}) \times D^\times(\mathbb{A})$ as follows. By [Hiraga and Saito 2012, Theorem 4.13], there exists an irreducible cuspidal automorphic representation $\sigma_1 \boxtimes \sigma_2$ of $\tilde{G}(\mathbb{A}) = D^\times(\mathbb{A}) \times D^\times(\mathbb{A})$ on the space $V_{\sigma_1\boxtimes\sigma_2}$ such that $V_{\pi_4} \subseteq V^1_{\sigma_1\boxtimes\sigma_2}|_{G_4(\mathbb{A})}$ and $\sigma_1 \boxtimes \sigma_2\|_{\mathrm{G}_4(\mathbb{A})} = \pi_4$, where $V^1_{\sigma_1\otimes\sigma_2}$ is the subspace of $V_{\sigma_1\boxtimes\sigma_2}$ on which the group

$$\mathfrak{X}_{\sigma_1\boxtimes\sigma_2} = \{\gamma \in (Z^\circ_{\tilde{G}}(\mathbb{A})G_4(\mathbb{A})\tilde{G}(F)\backslash\tilde{G}(\mathbb{A}))^D : (\sigma_1\boxtimes\sigma_2)\otimes\gamma \cong \sigma_1\boxtimes\sigma_2\}$$

acts trivially; here the superscript D in the above set indicates the Pontryagin dual; namely $\sigma_1\boxtimes\sigma_2$ is an automorphic representation which "lies above π_4". (Recall that the $\|$ notation is defined in the notation section.) Moreover, Let χ_1, χ_2 be the central characters of σ_1, σ_2, respectively. Then $(\chi_1\boxtimes\chi_2)|_{Z^\circ_4} = \chi_1\chi_2 = \omega_{\pi_4}$. Since $\omega_{\pi_4}\omega_{\pi_3} = \omega^2$, we have $\chi_1\chi_2\omega_{\pi_3} = \omega^2$.

Now, let $\phi \in V_{\pi_4}$ and $f \in V_{\pi_3}$. Since $V_{\pi_4} \subseteq V_{\sigma_1\boxtimes\sigma_2}|_{G_4(\mathbb{A})}$, we may assume $\phi = (\phi_1\otimes\phi_2)|_{G_4(\mathbb{A})}$ for some $\phi_i \in V_{\sigma_i}$. Then our period integral is of the form

$$\begin{aligned}\mathcal{P}_\omega(\phi, f) &= \int_{Z^\circ_3(\mathbb{A})G_3(F)\backslash G_3(\mathbb{A})} \phi(g)f(g)\omega(N(g))^{-1}\,dg\\ &= \int_{Z^\circ_3(\mathbb{A})G_3(F)\backslash G_3(\mathbb{A})} \phi_1(g)\phi_2(g)(f\cdot\omega^{-1})(g)\,dg\end{aligned}$$

because $\phi(g) = \phi_1(g)\phi_2(g)$ for $g \in G_3(\mathbb{A})$, which is precisely the triple product integral considered by Ichino for the automorphic representation $\sigma_1\boxtimes\sigma_2\boxtimes(\pi_3\otimes\omega^{-1})$ of $D^\times(\mathbb{A})\times D^\times(\mathbb{A})\times D^\times(\mathbb{A})$.

However, the local integral that Ichino considers is different from our local integral. For $\phi_{1,v}, \phi_{2,v}$ in $\pi_{4,v}$ and $f_{1,v}, f_{2,v}$ in $\pi_{3,v}$ our local integral is of the form

$$\begin{aligned}&\alpha^\natural_{\omega_v}(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v})\\ &= \int_{Z^\circ_3(F_v)\backslash \mathrm{G}_3(F_v)} \mathcal{B}_{\pi_{4,v}}(\pi_{4,v}(g_v)\phi_{1,v}, \phi_{2,v})\mathcal{B}_{\pi_{3,v}}(\pi_{3,v}(g_v)f_{1,v}, f_{2,v})\omega_v(N(g_v))^{-1}\,dg_v\\ &= \int_{Z^\circ_3(F_v)\backslash \mathrm{G}_3(F_v)} \mathcal{B}_{\pi_{4,v}}((\sigma_1\otimes\sigma_2)_v(g_v)\phi_{1,v}, \phi_{2,v})\mathcal{B}_{\pi_{3,v}}((\pi_{3,v}\cdot\omega_v^{-1})(g_v)f_{1,v}, f_{2,v})\,dg_v\end{aligned}$$

because for $\sigma_1\boxtimes\sigma_2\|_{\mathrm{G}_4(\mathbb{A})} = \pi_4$ we have that $\pi_{4,v} \subseteq (\sigma_1\otimes\sigma_2)_v\|_{\mathrm{G}_4(F_v)}$.

On the other hand, the local integral considered by Ichino is

$$\begin{aligned}&I_v(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v})\\ &\quad= \int_{Z^\circ_3(F_v)\backslash \mathrm{G}_3(F_v)} \mathcal{B}_{(\sigma_1\otimes\sigma_2)_v}((\sigma_1\otimes\sigma_2)_v(g_v)\phi_{1,v}, \phi_{2,v})\mathcal{B}_{\pi_{3,v}}(\pi_{3,v}(g_v)f_{1,v}, f_{2,v})\,dg_v\end{aligned}$$

so we need to relate $\mathcal{B}_{(\sigma_1\otimes\sigma_2)_v}$ with $\mathcal{B}_{\pi_{4,v}}$. Since $\pi_{4,v} \subseteq (\sigma_1\otimes\sigma_2)_v$ for all v we choose $\mathcal{B}_{\pi_{4,v}} = \mathcal{B}_{(\sigma_1\otimes\sigma_2)_v}|_{\pi_{4,v}\times\pi_{4,v}}$ for all v except one. We pick one place v_o and set $\mathcal{B}_{\pi_{4,v_0}} = \mathcal{B}_{(\sigma_1\otimes\sigma_2)_v}|_{\pi_{4,v}\times\pi_{4,v}}\cdot C$ for some constant C which gives $\prod_v \mathcal{B}_{\pi_{4,v}} = \prod_v \mathcal{B}_{(\sigma_1\otimes\sigma_2)_v}\cdot C$ so that $\mathcal{B}_{\pi_4} = C\cdot\mathcal{B}_{(\sigma_1\otimes\sigma_2)}$. This constant C which relates $\mathcal{B}_{(\sigma_1\otimes\sigma_2)}$ with $\mathcal{B}_{\pi_4}$is given in [Hiraga and Saito 2012, Remark 4.20]; namely,

$$C = |\mathfrak{X}_{\sigma_1\otimes\sigma_2}| \frac{\mathrm{Vol}(Z_4^\circ(\mathbb{A})\,\mathrm{G}_4(F)\backslash\,\mathrm{G}_4(\mathbb{A}))}{\mathrm{Vol}(Z_{\tilde{G}}^\circ(\mathbb{A})\tilde{G}(F)\backslash\tilde{G}(\mathbb{A}))}.$$

Noting that $Z_4^\circ(\mathbb{A})\mathrm{G}_4(F)\backslash\mathrm{G}_4(\mathbb{A}) \cong \mathrm{SO}(F)\backslash\mathrm{SO}_4(\mathbb{A})$ and $Z_{\tilde{G}}^\circ(\mathbb{A})\tilde{G}(F)\backslash\tilde{G}(\mathbb{A}) \cong \mathrm{SO}_3(F)\backslash\mathrm{SO}_3(\mathbb{A}) \times \mathrm{SO}_3(F)\backslash\,\mathrm{SO}_3(\mathbb{A})$, we have

$$\mathrm{Vol}(Z_{\mathrm{G}_4}^\circ(\mathbb{A})\mathrm{G}_4(F)\backslash\mathrm{G}_4(\mathbb{A})) = 2 \quad\text{and}\quad \mathrm{Vol}(Z_{\tilde{G}}^\circ(\mathbb{A})\tilde{G}(F)\backslash\tilde{G}(\mathbb{A})) = 4,$$

so that $C = \frac{1}{2}|\mathfrak{X}_{\sigma_1\otimes\sigma_2}|$ and normalizing $\alpha_{\omega_v}^\sharp$ as in the introduction gives

$$\prod_v \alpha_{\omega_v} = \frac{2}{|\mathfrak{X}_{\sigma_1\otimes\sigma_2}|}\cdot\prod_v I_v.$$

Thus, Ichino's triple product formula ([Ichino 2008, Theorem 1.1]) applied to $\sigma_1 \boxtimes \sigma_2 \boxtimes (\pi_3 \otimes \omega^{-1})$ with $c = 3$ gives

$$\begin{aligned}
|\mathcal{P}_\omega(\phi, f)|^2 &= \frac{2}{|\mathfrak{X}_{\sigma_1\otimes\sigma_2}|}\cdot\frac{\zeta(2)L(1/2, (\sigma_1\otimes\sigma_2)\boxtimes(\pi_3\otimes\omega^{-1}))}{8L(1, \sigma_1\otimes\sigma_2, \mathrm{Ad})L(1, \pi_3, Ad)}\prod_v \alpha_{\omega_v}(\phi_v, f_v) \\
&= \frac{\Delta_{G_4}L(1/2, (\sigma_1\otimes\sigma_2)\boxtimes(\pi_3\otimes\omega^{-1}))}{4|\mathfrak{X}_{\sigma_1\otimes\sigma_2}|L(1, \sigma_1\otimes\sigma_2, \mathrm{Ad})L(1, \pi_3, Ad)}\prod_v \alpha_{\omega_v}(\phi_v, f_v) \\
&= \frac{\Delta_{G_4}L(1/2, \pi_4\times\pi_3\otimes\omega^{-1})}{4|\mathfrak{X}_{\sigma_1\otimes\sigma_2}|L(1, \pi_4, \mathrm{Ad})L(1, \pi_3, \mathrm{Ad})}\prod_v \alpha_{\omega_v}(\phi_v, f_v).
\end{aligned}$$

Also we have that $|\mathcal{S}_{\phi_3}| = 1$ and $|\mathcal{S}_{\phi_4}| = |\mathfrak{X}_{\sigma_1\otimes\sigma_2}|$ by [Lapid and Mao 2015, Section 6.3]. Hence, Conjectures 1.4 and 1.5 hold.

Next assume $\operatorname{disc} V_4 \neq 1$. Using [Hiraga and Saito 2012, Theorem 4.13] again, there exists an irreducible cuspidal automorphic representation τ of $\tilde{G}(\mathbb{A}) = \mathrm{GL}_2(\mathbb{A}_E)$ on the space V_τ such that $V_{\pi_4} \subseteq V_\tau^1|_{G_4}$ and $\tau\|_{\mathrm{G}_4} = \pi_4$; namely τ "lies above π_4". Moreover, V_τ^1 is the subspace of V_τ on which the group

$$\mathfrak{X}_\tau = \{\gamma \in (Z_{\tilde{G}}^\circ(\mathbb{A})G_4(\mathbb{A})\tilde{G}(F)\backslash\tilde{G}(\mathbb{A}))^D : \tau\otimes\gamma \cong \tau\}$$

acts trivially. Note $\omega_\tau|_{Z_4^\circ(\mathbb{A})} = \omega_{\pi_4}$, and since $\omega_{\pi_4}\omega_{\pi_3} = \omega^2$ we have $\omega_\tau|_{Z_4^\circ(\mathbb{A})}\omega_{\pi_3} = \omega^2$. Now let $\phi \in V_{\pi_4}$ and $f \in V_{\pi_3}$. Since $V_{\pi_4} \subseteq V_\tau|_{G_4(\mathbb{A})}$, we can write $\phi = \tilde{\phi}|_{G_4(\mathbb{A})}$ for some $\tilde{\phi} \in V_\tau$. Then our period integral is of the form

$$\begin{aligned}
\mathcal{P}_\omega(\phi, f) &= \int_{Z_4^\circ(\mathbb{A})G_3(F)\backslash G_3(\mathbb{A})} \phi(g)f(g)\omega(N(g))^{-1}\,dg \\
&= \int_{Z_4^\circ(\mathbb{A})G_3(F)\backslash G_3(\mathbb{A})} \tilde{\phi}(g)(f\cdot\omega^{-1})(g)\,dg
\end{aligned}$$

because $\phi(g) = \tilde{\phi}(g)$ for $g \in G_3(\mathbb{A})$, which is precisely the period integral considered by Ichino for the automorphic representation $\tau \boxtimes (\pi_3 \otimes \omega^{-1})$ of $D^\times(\mathbb{A}_E) \times D^\times(\mathbb{A})$.

However, again the local integral that Ichino considers is different than our local integral. For $\phi_{1,v}, \phi_{2,v}$ in $\pi_{4,v}$ and $f_{1,v}, f_{2,v}$ in $\pi_{3,v}$ our local integral is of the form

$$\begin{aligned}
&\alpha^\natural_{\omega_v}(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v}) \\
&= \int_{Z_3^\circ(F_v)\backslash \mathrm{G}_3(F_v)} \mathcal{B}_{\pi_{4,v}}(\pi_{4,v}(g_v)\phi_{1,v}, \phi_{2,v})\mathcal{B}_{\pi_{3,v}}(\pi_{3,v}(g_v) f_{1,v}, f_{2,v})\omega_v(N(g_v))^{-1} dg_v \\
&= \int_{Z_3^\circ(F_v)\backslash \mathrm{G}_3(F_v)} \mathcal{B}_{\pi_{4,v}}(\tau_v(g_v)\phi_{1,v}, \phi_{2,v})\mathcal{B}_{\pi_{3,v}}((\pi_{3,v}\cdot\omega_v^{-1})(g_v) f_{1,v}, f_{2,v}) dg_v
\end{aligned}$$

because for $\tau\|_{\mathrm{G}_4(\mathbb{A})} = \pi_4$ we have that $\pi_{4,v} \subseteq \tau_v\|_{\mathrm{G}_4(F_v)}$.

On the other hand, the local integral considered by Ichino is

$$\begin{aligned}
&I_v(\phi_{1,v}, \phi_{2,v}; f_{1,v}, f_{2,v}) \\
&\qquad = \int_{Z_3^\circ(F_v)\backslash \mathrm{G}_3(F_v)} \mathcal{B}_{\tau_v}(\tau_v(g_v)\phi_{1,v}, \phi_{2,v})\mathcal{B}_{\pi_{3,v}}(\pi_{3,v}(g_v) f_{1,v}, f_{2,v}) dg_v,
\end{aligned}$$

so we need to relate $\mathcal{B}_{\tau_v}$ with $\mathcal{B}_{\pi_{4,v}}$. Since $\pi_{4,v} \subseteq \tau_v$ for all v we choose $\mathcal{B}_{\pi_{4,v}} = \mathcal{B}_{\tau_v}|_{\pi_{4,v}\times\pi_{4,v}}$ for all v except one and pick one place v_o and set $\mathcal{B}_{\pi_{4,v_0}} = \mathcal{B}_{\tau_v}|_{\pi_{4,v}\times\pi_{4,v}} \cdot C'$ for some constant C' which gives $\prod_v \mathcal{B}_{\pi_{4,v}} = \prod_v \mathcal{B}_{\tau_v} \cdot C'$ and so $\mathcal{B}_{\pi_4} = C' \cdot \mathcal{B}_\tau$ where C' is given in [Hiraga and Saito 2012, Remark 4.20]; namely,

$$C' = |\mathfrak{X}_\tau| \frac{\mathrm{Vol}(Z_4^\circ(\mathbb{A})\,\mathrm{G}_4(F)\backslash\,\mathrm{G}_4(\mathbb{A}))}{\mathrm{Vol}(Z_{\tilde{G}}^\circ(\mathbb{A})\tilde{G}(F)\backslash\tilde{G}(\mathbb{A}))},$$

where the volumes are

$$\mathrm{Vol}(Z_{G_4}^\circ(\mathbb{A})G_4(F)\backslash G_4(\mathbb{A}_F)) = \mathrm{Vol}(Z_{\tilde{G}}^\circ(\mathbb{A})\tilde{G}(F)\backslash\tilde{G}(\mathbb{A}_\mathbb{F})) = 2,$$

and so $C' = |\mathfrak{X}_\tau|$ in this case.

Then Ichino's triple product formula applied to $\tau \boxtimes (\pi_3 \otimes \omega^{-1})$ with $c = 2$ gives

$$\begin{aligned}
|\mathcal{P}_\omega(\phi, f)|^2 &= \frac{1}{\mathfrak{X}_\tau} \cdot \frac{\zeta(2) L(1/2, \tau \boxtimes (\pi_3 \otimes \omega^{-1}))}{4L(1, \tau, \mathrm{Ad})L(1, \pi_3, Ad)} \prod_v \alpha_{\omega_v}(\phi_v, f_v) \\
&= \frac{\Delta_{G_4} L(1/2, \tau \boxtimes (\pi_3 \otimes \omega^{-1}))}{4|\mathfrak{X}_\tau| L(1, \tau, \mathrm{Ad})L(1, \pi_3, Ad)} \prod_v \alpha_{\omega_v}(\phi_v, f_v) \\
&= \frac{\Delta_{G_4} L(1/2, \pi_4 \times \pi_3 \otimes \omega^{-1})}{4|\mathfrak{X}_\tau| L(1, \pi_4, \mathrm{Ad})L(1, \pi_3, \mathrm{Ad})} \prod_v \alpha_{\omega_v}(\phi_v, f_v).
\end{aligned}$$

Moreover, $|\mathcal{S}_{\phi_3}| = 1$ and $|\mathcal{S}_{\phi_4}| = |\mathfrak{X}_\tau|$ by [Lapid and Mao 2015, Section 6.4]. Hence, Conjectures 1.4 and 1.5 hold for this case, too.

Remark 4.5. In Ichino's triple product formula, the two cases $d = 1$ and $d \neq 1$ are different. However, once we consider Ichino's formula as an instance of the GGP conjecture for GSpin, these two cases can be considered as one case as above.

4E. ***Conjecture for*** **$\mathrm{GSpin}_5 \times \mathrm{GSpin}_4$ (*Gan–Ichino formula*).** Finally, we consider the conjecture for $n = 5$. For this case, we will not be able to prove the conjecture in full generality but only for some special cases as we will explain in what follows.

Firstly we consider only the case where GSpin_5 is split and GSpin_4 is quasisplit, namely

$$\mathrm{GSpin}_4 = \begin{cases} \{(g_1, g_2) \in \mathrm{GL}_1 \times \mathrm{GL}_1 : \det g_1 = \det g_2\}; \\ \{g \in \mathrm{Res}_{E/F} \mathrm{GL}_2 : N(g) \in \mathrm{GL}_1\}, \end{cases}$$
$$\mathrm{GSpin}_5 = \mathrm{GSp}_4 .$$

Or equivalently, $\mathrm{GSpin}_4 = \mathrm{GSpin}(V_4)$ with $V_4 = \mathbb{H} \oplus \mathbb{H}$ or $V_4 = \mathbb{H} \oplus E$ for a quadratic extension E of F equipped with the norm form, and $\mathrm{GSpin}_5 = \mathrm{GSpin}(V_5)$ with $\mathbb{H} \oplus \mathbb{H} \oplus \langle 1 \rangle$. Let us note that we have the natural inclusion $V_4 \subseteq V_5$ of the quadratic forms, which gives rise to the natural inclusion $\mathrm{GSpin}_4 \subseteq \mathrm{GSpin}_5$.

Secondly, for our (tempered) cuspidal automorphic representations π_5 and π_4 of $\mathrm{GSpin}_5(\mathbb{A})$ and $\mathrm{GSpin}_4(\mathbb{A})$, respectively, we only consider the following special cases. For π_5, we assume that π_5 is the theta lift of a cuspidal automorphic representation σ of $\mathrm{GO}(V)(\mathbb{A})$, where V is a 4-dimensional quadratic space. (Let us note that σ has to satisfy certain technical conditions such that the theta lift to $\mathrm{GSp}_4(\mathbb{A})$ is nonzero and cuspidal. See [Gan and Ichino 2011, page 236] for details.) Here, let us denote the discriminant algebra of V by K, which is the étale quadratic algebra over F defined by

$$K = \begin{cases} F \times F & \text{if } \mathrm{disc}(V) = 1; \\ F(\sqrt{\mathrm{disc}(V)}) & \text{if } \mathrm{disc}(V) \neq 1. \end{cases}$$

As for π_4, let τ be a cuspidal automorphic representation which “lies above π_4” as in the previous subsection. To be precise, let

$$\tilde{G}_4 := \begin{cases} \mathrm{GL}_2 \times \mathrm{GL}_2 & \text{if } V_4 = \mathbb{H} \oplus \mathbb{H}; \\ \mathrm{Res}_{E/F} \mathrm{GL}_2 & \text{if } V_4 = \mathbb{H} \oplus E. \end{cases}$$

By Theorem 4.13 of [Hiraga and Saito 2012], there exists an irreducible unitary cuspidal automorphic representation τ of $\tilde{G}_4(\mathbb{A})$ such that $\tau \|_{\mathrm{G}_4} = \pi_4$ and $V_{\pi_4} \subset V_\tau^1|_{G_4}$, where V_τ^1 is the subspace of V_τ such that

$$\mathfrak{X}_\tau = \{\gamma \in (Z^\circ_{\tilde{G}}(\mathbb{A}) G_4(\mathbb{A}) \tilde{G}_4(F) \backslash \tilde{G}_4(\mathbb{A})^D : \tau \otimes \gamma \cong \tau\}$$

acts trivially. Then we assume that the base change τ_K of τ to $\tilde{G}(\mathbb{A}_K)$ is cuspidal, where K is the discriminant algebra of V as above, and the Jacquet–Langlands transfer of τ_K to $D^\times(\mathbb{A}_{E \otimes K})$ exists. (See the top of page 237 of [Gan and Ichino 2011] for details.)

Then Gan and Ichino essentially proved the following.

Theorem 4.6 (Gan and Ichino). *Let π_5 and π_4 be as above. Assume there exists a Hecke character ω such that $\omega^2 = \omega_{\pi_5}\omega_{\pi_4}$. Then for factorizable $\phi \in V_{\pi_5}$ and $f \in V_{\pi_4}$, we have*

$$|\mathcal{P}_\omega(\phi, f)|^2 = \frac{\zeta(2)\zeta(4)L(1/2, \pi_5 \times \pi_4 \otimes \omega^{-1})}{2^\beta |\mathfrak{X}_\tau| L(1, \pi_5, \mathrm{Ad})(L(1, \pi_4, \mathrm{Ad})} \prod_v \alpha_{\omega_v}(\phi_v, f_v),$$

where

$$\beta = \begin{cases} 3 & \textit{if } \operatorname{disc}(V) = 1; \\ 2 & \textit{if } \operatorname{disc}(V) \neq 1. \end{cases}$$

Proof. This is essentially Theorem 1.1 of [Gan and Ichino 2011] with the notation adjusted to ours by setting $\pi = \pi_5$ and $\pi' = \pi_4 \otimes \omega^{-1}$, where π and π' are as in [Gan and Ichino 2011]. But it should be mentioned that if π_4 satisfies the above mentioned conditions, then so does $\pi_4 \otimes \omega^{-1}$, and hence we can use the Gan–Ichino formula for $\pi_5 \times \pi_4 \otimes \omega^{-1}$. □

Now, let us take care of the constant $2^\beta|\mathfrak{X}_\tau|$. By [Lapid and Mao 2015, Section 6], we know $|\mathfrak{X}_\tau| = |\mathcal{S}_{\phi_4}|$. For the representation π_5 considered here, Roberts [2001] essentially verified that

$$|\mathcal{S}_{\phi_5}| = \begin{cases} 2 \text{ if } \operatorname{disc}(V) = 1; \\ 1 \text{ if } \operatorname{disc}(V) \neq 1. \end{cases}$$

Hence if $\operatorname{disc}(V) = 1$ we have

$$2^\beta|\mathfrak{X}_\tau| = 2^3|\mathfrak{X}_\tau| = 2^2|\mathcal{S}_{\phi_5}|\,|\mathcal{S}_{\phi_4}|,$$

and if disc $V \neq 1$ we have

$$2^\beta|\mathfrak{X}_\tau| = 2^2|\mathfrak{X}_\tau| = 2^2|\mathcal{S}_{\phi_5}|\,|\mathcal{S}_{\phi_4}|.$$

Thus in either case the above theorem confirms Conjectures 1.4 and 1.5.

Remark 4.7. In the above theorem, Gan and Ichino assumed that F and E are totally real number fields. This assumption was to utilize the Siegel–Weil formula. (See [Gan and Ichino 2011, Remark 1.3].) However, the condition is no longer necessary thanks to the work of Gan, Qui and Takeda in [Gan et al. 2014]. Further, Roberts [2001] assumed that F and E are totally real essentially for the same reason, and hence this assumption is not necessary, either.

Also Gan and Ichino do not assume that π_5 and π_4 are tempered. This is because for the case at hand the convergence of the local integral as in Proposition 3.4 can be shown by using the Kim–Shahidi estimate as in [Gan and Ichino 2011, Lemma 9.1]. Hence we do not even need to assume π_5 and π_4 are tempered.

Acknowledgements

This work stems from the author's PhD thesis. The author thanks her advisor Shuichiro Takeda for suggesting this problem, and his helpful advice and support. The author also thanks Atsushi Ichino for helpful discussions and for directing her to [Xue 2017] and [Lapid and Mao 2015]; as well as Hang Xue for discussions regarding Conjecture 1.5 and Remark 1.6.

References

[Arthur 2013] J. Arthur, *The endoscopic classification of representations: orthogonal and symplectic groups*, American Mathematical Society Colloquium Publications **61**, American Mathematical Society, Providence, RI, 2013. MR Zbl

[Asgari 2000] M. Asgari, *On holomorphy of local Langlands L-functions*, Ph.D. thesis, Purdue University, 2000, available at https://search.proquest.com/docview/304618163. MR

[Asgari 2002] M. Asgari, "Local L-functions for split spinor groups", *Canad. J. Math.* **54**:4 (2002), 673–693. MR Zbl

[Asgari and Choiy 2017] M. Asgari and K. Choiy, "The local Langlands conjecture for p-adic GSpin_4, GSpin_6, and their inner forms", *Forum Math.* **29**:6 (2017), 1261–1290. MR Zbl

[Asgari and Shahidi 2006] M. Asgari and F. Shahidi, "Generic transfer for general spin groups", *Duke Math. J.* **132**:1 (2006), 137–190. MR Zbl

[Bump 1997] D. Bump, *Automorphic forms and representations*, Cambridge Studies in Advanced Mathematics **55**, Cambridge University Press, 1997. MR Zbl

[Deligne 1972] P. Deligne, "La conjecture de Weil pour les surfaces $K3$", *Invent. Math.* **15** (1972), 206–226. MR Zbl

[Gan and Ichino 2011] W. T. Gan and A. Ichino, "On endoscopy and the refined Gross–Prasad conjecture for $(\mathrm{SO}_5, \mathrm{SO}_4)$", *J. Inst. Math. Jussieu* **10**:2 (2011), 235–324. MR Zbl

[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. 171–312 in *Sur les conjectures de Gross et Prasad*, vol. I, Astérisque **346**, Société Mathématique de France, Paris, 2012. MR Zbl

[Gan et al. 2014] W. T. Gan, Y. Qiu, and S. Takeda, "The regularized Siegel–Weil formula (the second term identity) and the Rallis inner product formula", *Invent. Math.* **198**:3 (2014), 739–831. MR Zbl

[Gross and Prasad 1992] B. H. Gross and D. Prasad, "On the decomposition of a representation of SO_n when restricted to SO_{n-1}", *Canad. J. Math.* **44**:5 (1992), 974–1002. MR Zbl

[Harris 2014] R. N. Harris, "The refined Gross–Prasad conjecture for unitary groups", *Int. Math. Res. Not.* **2014**:2 (2014), 303–389. MR Zbl

[Helgason 2000] S. Helgason, *Groups and geometric analysis: integral geometry, invariant differential operators, and spherical functions*, Mathematical Surveys and Monographs **83**, American Mathematical Society, Providence, RI, 2000. MR Zbl

[Hiraga and Saito 2012] K. Hiraga and H. Saito, *On L-packets for inner forms of SL_n*, Mem. Amer. Math. Soc. **1013**, American Mathematical Society, Providence, RI, 2012. MR Zbl

[Hundley and Sayag 2016] J. Hundley and E. Sayag, *Descent construction for GSpin groups*, Mem. Amer. Math. Soc. **1148**, American Mathematical Society, Providence, RI, 2016. MR Zbl

[Ichino 2008] A. Ichino, "Trilinear forms and the central values of triple product L-functions", *Duke Math. J.* **145**:2 (2008), 281–307. MR Zbl

[Ichino and Ikeda 2010] A. Ichino and T. Ikeda, "On the periods of automorphic forms on special orthogonal groups and the Gross–Prasad conjecture", *Geom. Funct. Anal.* **19**:5 (2010), 1378–1425. MR Zbl

[Knus et al. 1998] M.-A. Knus, A. Merkurjev, M. Rost, and J.-P. Tignol, *The book of involutions*, American Mathematical Society Colloquium Publications **44**, American Mathematical Society, Providence, RI, 1998. MR Zbl

[Lam 2005] T. Y. Lam, *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics **67**, American Mathematical Society, Providence, RI, 2005. MR Zbl

[Lapid and Mao 2015] E. Lapid and Z. Mao, "A conjecture on Whittaker–Fourier coefficients of cusp forms", *J. Number Theory* **146** (2015), 448–505. MR Zbl

[Piatetski-Shapiro and Rallis 1987] I. Piatetski-Shapiro and S. Rallis, "Rankin triple L functions", *Compositio Math.* **64**:1 (1987), 31–115. MR Zbl

[Roberts 2001] B. Roberts, "Global L-packets for GSp(2) and theta lifts", *Doc. Math.* **6** (2001), 247–314. MR Zbl

[Scharlau 1985] W. Scharlau, *Quadratic and Hermitian forms*, Grundlehren der Mathematischen Wissenschaften **270**, Springer, 1985. MR Zbl

[Silberger 1979] A. J. Silberger, *Introduction to harmonic analysis on reductive p-adic groups*, Mathematical Notes **23**, Princeton University Press, 1979. MR Zbl

[Waldspurger 1985] J.-L. Waldspurger, "Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie", *Compositio Math.* **54**:2 (1985), 173–242. MR Zbl

[Waldspurger 2003] J.-L. Waldspurger, "La formule de Plancherel pour les groupes p-adiques (d'après Harish-Chandra)", *J. Inst. Math. Jussieu* **2**:2 (2003), 235–333. MR Zbl

[Xue 2017] H. Xue, "Refined global Gan–Gross–Prasad conjecture for Fourier–Jacobi periods on symplectic groups", *Compos. Math.* **153**:1 (2017), 68–131. MR Zbl

[Yuan et al. 2013] X. Yuan, S.-W. Zhang, and W. Zhang, *The Gross–Zagier formula on Shimura curves*, Annals of Mathematics Studies **184**, Princeton University Press, 2013. MR Zbl

Received February 19, 2019. Revised October 22, 2019.

MELISSA EMORY
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TORONTO
TORONTO ON
CANADA
memory@math.toronto.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.153

SCHUR ALGEBRAS FOR THE ALTERNATING GROUP AND KOSZUL DUALITY

THANGAVELU GEETHA, AMRITANSHU PRASAD AND SHRADDHA SRIVASTAVA

We introduce the alternating Schur algebra $\mathrm{AS}_F(n, d)$ as the commutant of the action of the alternating group A_d on the d-fold tensor power of an n-dimensional F-vector space. When F has characteristic different from 2, we give a basis of $\mathrm{AS}_F(n, d)$ in terms of bipartite graphs, and a graphical interpretation of the structure constants. We introduce the abstract Koszul duality functor on modules for the even part of any $Z/2Z$-graded algebra. The algebra $\mathrm{AS}_F(n, d)$ is $Z/2Z$-graded, having the classical Schur algebra $\mathrm{S}_F(n, d)$ as its even part. This leads to an approach to Koszul duality for $\mathrm{S}_F(n, d)$-modules that is amenable to combinatorial methods. We characterize the category of $\mathrm{AS}_F(n, d)$-modules in terms of $\mathrm{S}_F(n, d)$-modules and their Koszul duals. We use the graphical basis of $\mathrm{AS}_F(n, d)$ to study the dependence of the behavior of derived Koszul duality on n and d.

1. Introduction

1A. ***Schur–Weyl duality and its variants.*** Frobenius [1900] determined the irreducible characters of the symmetric group S_d over C, the field of complex numbers. Building on this, Schur classified the irreducible polynomial representations of $\mathrm{GL}_n(C)$ and computed their characters in his PhD thesis [Schur 1901]. The group $\mathrm{GL}_n(C)$ acts on the factors of $(C^n)^{\otimes d}$, while S_d permutes the tensor factors. Schur [1927] used these commuting actions to reprove the results of his dissertation. Following Weyl's expositions of this method [Weyl 1931; Weyl 1939], it is known as Schur–Weyl duality.

Over the years, several variants of Schur–Weyl duality have emerged. Shrinking $\mathrm{GL}_n(C)$ to the orthogonal group $\mathrm{O}_n(C)$, Brauer [1937] obtained the duality between Brauer algebras $\mathrm{Br}_d(n)$ and $\mathrm{O}_n(C)$. Motivated by the Potts model in statistical mechanics, Jones [1994] and Martin [1991] further shrunk $\mathrm{O}_n(C)$ down to S_n, obtaining the partition algebras $\mathrm{P}_d(n)$. Bloss [2005] reduced S_n to A_n to obtain an algebra $\mathrm{AP}_d(n)$ which coincides with the partition algebra when $n \geq 2d+2$. We take the smallest possible step in the opposite direction: we reveal what takes the place

MSC2010: 05E10, 20G05, 20G43.

Keywords: Schur algebra, Koszul duality, Schur–Weyl duality, alternating group.

$F^{n\otimes d}$		
??	this article	$\boldsymbol{A}_d$
$\cup$		$\cap$
$\mathrm{GL}_n(F)$	Schur–Weyl	$\boldsymbol{S}_d$
$\cup$		$\cap$
$\mathrm{O}_n(F)$	Brauer	$\mathrm{Br}_d(n)$
$\cup$		$\cap$
$\boldsymbol{S}_n$	Martin and Jones	$\mathrm{P}_d(n)$
$\cup$		$\cap$
$\boldsymbol{A}_n$	Bloss	$\mathrm{AP}_d(n)$

Table 1. Dualities arising from tensor space.

of the polynomial representations of $\mathrm{GL}_n(\boldsymbol{C})$ when the action of the symmetric group $\boldsymbol{S}_d$ is restricted to the alternating group $\boldsymbol{A}_d$. The situation is summarized in Table 1. The significance of this investigation lies in its connection with the Koszul duality functor on the category of homogeneous polynomial representations of $\mathrm{GL}_n(\boldsymbol{C})$ of degree d.

1B. ***Schur algebras for the alternating group.*** Motivated by Green [2007, Theorem 2.6c], define the Schur algebra as

$$\mathrm{S}_F(n,d) = \mathrm{End}_{\boldsymbol{S}_d}((F^n)^{\otimes d})$$

for any field F, and positive integers n and d. When F is infinite, then $\mathrm{S}_F(n,d)$-modules are the same as homogeneous polynomial representations of $\mathrm{GL}_n(F)$ of degree d (see [Green 2007, Section 2.4] and [Prasad 2015, Section 6.2]). Define the alternating Schur algebra $\mathrm{AS}_F(n,d)$ by replacing $\boldsymbol{S}_d$ by $\boldsymbol{A}_d$ in the definition above:

$$\mathrm{AS}_F(n,d) = \mathrm{End}_{\boldsymbol{A}_d}((F^n)^{\otimes d}).$$

When F has characteristic different from 2, this algebra has a decomposition (see Lemma 2.1)

$$\mathrm{AS}_F(n,d) = \mathrm{S}_F(n,d) \oplus \mathrm{S}_F^-(n,d) \tag{1}$$

as a $\mathbb{Z}/2\mathbb{Z}$-graded algebra. Here $\mathrm{S}_F^-(n,d) = \mathrm{Hom}_{\boldsymbol{S}_d}((F^n)^{\otimes d}, (F^n)^{\otimes d} \otimes \mathrm{sgn})$, where sgn is used to denote the sign character of $\boldsymbol{S}_d$. The subspace $\mathrm{S}_F^-(n,d)$ is an $(\mathrm{S}_F(n,d), \mathrm{S}_F(n,d))$-bimodule.

When $n^2 < d$, then $\mathrm{S}_F^-(n,d) = 0$, and $\mathrm{AS}_F(n,d) = \mathrm{S}_F(n,d)$, as observed by Regev [2002, Theorem 1]. But when $n^2 \geq d$, $\mathrm{S}_F^-(n,d) \neq 0$, and in Lemma 2.2, we note that $\mathrm{S}_F^-(n,d)$ is a full tilting left $\mathrm{S}_F(n,d)$-module as studied by Donkin [1993, Section 3].

1C. ***Bases and structure constants.*** Schur [1927] gave a combinatorial description of a basis and the corresponding structure constants of the Schur algebra (see also [Green 2007, Section 2.3]). By indexing Schur's basis of $\mathrm{S}_F(n, d)$ by bipartite multigraphs with $n+n$ vertices and d edges, Méndez [2001] (see also [Geetha and Prasad 2014]) gave a graphic interpretation of the structure constants. We describe a basis of $\mathrm{S}_F^-(n, d)$ in terms of bipartite simple graphs in Theorem 2.24. So from the decomposition (1), a basis of $\mathrm{AS}_F(n, d)$ is obtained. A graphic interpretation of the structure constants of $\mathrm{AS}_F(n, d)$ is given in Theorems 2.14 and 2.25. This will be used to derive properties of $\mathrm{AS}_F(n, d)$, its bimodule $\mathrm{S}_F^-(n, d)$, and Koszul duality.

1D. ***Koszul duality and modules.*** The term *Koszul duality* is used for several constructions which interchange the roles of exterior and symmetric powers.

The earliest notion of Koszul duality was introduced by Priddy [1970]. It applies to *pre-Koszul algebras*, which are also called *quadratic algebras*. A pre-Koszul algebra is a quotient of a tensor algebra

$$T(V) = \bigoplus_{n \geq 0} \otimes^n V$$

by a two-sided ideal I that is generated in degree two. Its Koszul dual is the algebra

$$T(V^*)/(I \cap (V \otimes V))^{\perp};$$

the quotient of the dual tensor algebra by the annihilator in degree two of I. In this setting the Koszul dual of the symmetric algebra of V is the exterior algebra of V^*.

Bernstein, Gelfand, and Gelfand [Bernstein et al. 1978, Theorem 3] introduced an equivalence between the bounded derived categories of graded modules over symmetric and exterior algebras, which was called the Koszul duality functor by Beilinson, Ginsburg, and Schectman [Beilinson et al. 1988].

Friedlander and Suslin [1997] introduced the category of *strict polynomial functors* of degree d as the representations of the Schur category of degree d, for each nonnegative integer d (see Section 4A). The category of strict polynomial functors of degree d unifies the categories of homogeneous polynomial representations of $\mathrm{GL}_n(F)$ of degree d across all n. Standard examples of strict polynomial functors of degree d are the d-th tensor power functor $\otimes^d$, the d-th symmetric power functor Sym^d, and the d-th exterior power functor $\wedge^d$. Evaluating a strict polynomial functor of degree d at F^n gives an $\mathrm{S}_F(n, d)$-module for each n. Friedlander and Suslin showed that this evaluation functor is an equivalence of categories when $n \geq d$. Chałupnik [2008] and Touzé [2014] used the term Koszul duality to refer to a functor on the category of strict polynomial functors of degree d which takes the Schur functor associated to the partition λ of d to the Weyl functor associated with the partition λ' *conjugate* to λ. Krause [2013] discovered an internal tensor product on the category of strict polynomial functors of fixed degree d. Given such a tensor product it was then natural for him to define Koszul duality in this category as the tensor product with $\wedge^d$. This definition is different from the Koszul duality functors defined earlier by Chałupnik and Touzé. Those coincide with a duality defined by Ringel [1991] using tilting modules for quasihereditary algebras. This tilting module was described by Donkin [1993] in the case of Schur algebras.

We introduce the term *abstract Koszul duality* to refer to a very simple functor which makes sense for any $\mathbf{Z}/2\mathbf{Z}$-graded algebra $\mathrm{AS} = \mathrm{S} \oplus \mathrm{S}^-$. The abstract Koszul dual of an S-module V is defined as

$$D(V) = \mathrm{S}^- \otimes_{\mathrm{S}} V.$$

The multiplication operation on AS gives rise to an (S, S)-bimodule homomorphism $\phi : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \to S$ and hence a natural transformation from $D \circ D$ to the identity functor on the category of S-modules. We prove (Theorem 3.4) that the category of AS-modules is the same as the category of pairs (M, θ_M) where M is an S-module and $\theta_M : D(M) \to M$ is compatible with ϕ in the sense of (16).

In Section 4, we specialize to the case $\mathrm{AS}_F(n, d) = \mathrm{S}_F(n, d) \oplus \mathrm{S}_F^-(n, d)$ to obtain a Koszul duality functor D on the category of $\mathrm{S}_F(n, d)$-modules. In Theorem 4.5, we show that the evaluation at F^n of the Koszul duality functor of Krause is naturally isomorphic to our Koszul duality functor when $n \geq d$. In this sense, our abstract Koszul duality functor on Schur algebras coincides with Krause's Koszul duality.

Our description of the structure constants of $\mathrm{AS}_F(n, d)$ allows us to give a direct combinatorial proof of the well-known fact that, when $n \geq d$ and when the characteristic of F is 0 or greater than d, then abstract Koszul duality is an equivalence (Theorem 4.2).

Krause [2013] showed that the derived Koszul duality functor is an autoequivalence of the unbounded derived category of strict polynomial functors. Since the evaluation functor is an equivalence, this implies that derived Koszul duality is an autoequivalence at the level of the unbounded derived category of $\mathrm{S}_F(n, d)$-modules when $n \geq d$. However, this does not address the case where $n < d$. Using our combinatorial methods, we show that derived Koszul duality is not an equivalence when $n < d$ (Theorem 4.9). This proof uses a criterion of Happel [1987] for a tensor functor to be a derived equivalence. In the context of derived Koszul duality, this criterion requires that the canonical algebra homomorphism $\mathrm{S}_F(n, d) \to \mathrm{End}_{\mathrm{S}_F(n,d)}(\mathrm{S}_F^-(n, d))$ is an isomorphism. Donkin [1993, Proposition 3.7] proved this for $n \geq d$. For the case when the characteristic of F is not 2 we give a combinatorial proof of Donkin's result, and also show that it fails when $n < d$ (Theorem 4.7). Figure 1 on page 181 describes the behavior of Koszul duality for all values of the parameters n and d.

We conclude this paper by discussing a possible application of our techniques to Bloss's alternating partition algebra, and a diagrammatic interpretation of the Schur category (Section 5).

2. The alternating Schur algebra

Let F be a field of characteristic different from 2, and n and d be positive integers. The symmetric group $\boldsymbol{S}_d$ acts on the tensor space $(F^n)^{\otimes d}$ by permuting the tensor factors. The Schur algebra can be defined as

$$\mathrm{S}_F(n,d) := \mathrm{End}_{\boldsymbol{S}_d}((F^n)^{\otimes d}).$$

By restricting the action of $\boldsymbol{S}_d$ to the alternating group $\boldsymbol{A}_d$, define the *alternating Schur algebra* as

$$\mathrm{AS}_F(n,d) := \mathrm{End}_{\boldsymbol{A}_d}((F^n)^{\otimes d}).$$

Clearly, $\mathrm{S}_F(n,d)$ is a subalgebra of $\mathrm{AS}_F(n,d)$.

Lemma 2.1. *For any representations V and W of $\boldsymbol{S}_d$,*

$$\mathrm{Hom}_{\boldsymbol{A}_d}(V,W) = \mathrm{Hom}_{\boldsymbol{S}_d}(V,W) \oplus \mathrm{Hom}_{\boldsymbol{S}_d}(V, W\otimes \mathrm{sgn}). \tag{2}$$

Here $W\otimes \mathrm{sgn}$ denotes the twist of W by the sign character $\mathrm{sgn} : \boldsymbol{S}_d \to \{\pm 1\}$.

Define

$$\mathrm{S}_F^-(n,d) := \mathrm{Hom}_{\boldsymbol{S}_d}((F^n)^{\otimes d}, (F^n)^{\otimes d}\otimes \mathrm{sgn}). \tag{3}$$

Lemma 2.1 gives a $\boldsymbol{Z}/2\boldsymbol{Z}$-grading of $\mathrm{AS}_F(n,d)$ in the sense of Bourbaki [1974, Chapter III, Section 3.1]:

$$\mathrm{AS}_F(n,d) = \mathrm{S}_F(n,d) \oplus \mathrm{S}_F^-(n,d). \tag{4}$$

The summand $\mathrm{S}_F^-(n,d)$ is an $(\mathrm{S}_F(n,d), \mathrm{S}_F(n,d))$-bimodule. Recall that a *weak composition* of d with n parts is a vector $\lambda = (\lambda_1, \dots, \lambda_n)$ of nonnegative integers summing to d. Let $\Lambda(n,d)$ denote the set of weak compositions of d with n parts. For each $\lambda = (\lambda_1,\dots,\lambda_n) \in \Lambda(n,d)$, define

$$\wedge^\lambda F^n = \wedge^{\lambda_1} F^n \otimes \cdots \otimes \wedge^{\lambda_n} F^n,$$

where, for a nonnegative integer s, $\wedge^s F^n$ is the s-th exterior power of F^n. As a left $\mathrm{S}_F(n,d)$-module,

$$\mathrm{S}_F^-(n,d) = \bigoplus_{\lambda\in\Lambda(n,d)} \wedge^\lambda F^n. \tag{5}$$

For each partition λ of d with at most n parts, let $\Delta(\lambda)$ denote the $\mathrm{S}_F(n,d)$-module known as the *Weyl module* with highest weight λ as in [Donkin 1993, Section 1]. A *tilting module* is an $\mathrm{S}_F(n,d)$-module V such that both V and its dual V^* have filtrations by the Weyl modules $\Delta(\lambda)$. Ringel [1991] showed that, for every such λ, there exists an indecomposable tilting module $M(\lambda)$ with unique highest weight λ. A *full tilting module* is a tilting module that contains $M(\lambda)$ as a direct summand for every partition λ of d with at most n parts [Donkin 1993, Section 3]. By Donkin [1993, Lemma 3.4], (5) implies:

Lemma 2.2. *The left module $\mathrm{S}_F^-(n,d)$ is a full tilting module of $\mathrm{S}_F(n,d)$.*

2A. ***Twisted permutation representations.*** Let X be a finite set on which a group G acts on the right (henceforth called a G-set). The space $F[X]$ of F-valued functions on X may be regarded as a representation of G:

$$\rho_X(g)f(x) = f(x\cdot g) \quad \text{for } x\in X,\ g\in G, \text{ and } f\in F[X]. \tag{6}$$

Let χ be a multiplicative character $G\to F^*$. We may twist the representation (6) by χ:

$$\rho_X^\chi(g)f(x) = \chi(g)f(x\cdot g). \tag{7}$$

Denote the representation space of this twisted action as $F[X]\otimes\chi$.

Suppose that X and Y are finite G-sets. Given a function $\kappa : X\times Y\to F$, the *integral operator* $\xi_\kappa : F[Y]\to F[X]$ associated to κ is defined as

$$\xi_\kappa f(x) = \sum_{y\in Y}\kappa(x,y)f(y) \quad \text{for } f\in F[Y]. \tag{8}$$

The function κ is known as the *integral kernel* of ξ_κ.

If Z is another finite G-set, $\kappa' : X\times Y\to F$ and $\kappa'' : Y\times Z\to F$ are functions. Then,

$$\xi_{\kappa'}\circ\xi_{\kappa''} = \xi_{\kappa'*\kappa''},$$

where $\kappa'*\kappa'' : X\times Z\to F$ is the *convolution product*

$$\kappa'*\kappa''(x,z) = \sum_{y\in Y}\kappa'(x,y)\kappa''(y,z). \tag{9}$$

We have (see [Prasad 2015, Section 4.2]):

Theorem 2.3. *For any finite G-spaces X and Y, and any multiplicative character $\chi : G\to F^*$,*

$$\begin{aligned}\operatorname{Hom}_G(F[Y], F[X]\otimes\chi)\\ = \{\xi_\kappa \mid \kappa : X\times Y\to F \text{ such that } \kappa(x\cdot g, y\cdot g) = \chi(g)\kappa(x,y)\}.\end{aligned} \tag{10}$$

The identity (10) implies that

$$\dim\operatorname{Hom}_G(F[Y], F[X]\otimes\chi) \le |(X\times Y)/G|,$$

with equality holding if χ is the trivial character. However, if $g\in G$, and $(x,y)\in X\times Y$ are such that $(x\cdot g, y\cdot g) = (x,y)$, then if $\xi_\kappa\in\operatorname{Hom}_G(F[Y], F[X]\otimes\chi)$,

$$\kappa(x,y) = \kappa(x\cdot g, y\cdot g) = \chi(g)\kappa(x,y),$$

so that either $\chi(g) = 1$ or κ vanishes on the G-orbit of (x,y).

For each element $x \in X$, let $G_x = \{g \in G \mid g \cdot x = x\}$ be the stabilizer of x in G.

Definition 2.4 (transverse pair). A pair $(x, y) \in X \times Y$ is said to be *transverse* with respect to χ if $G_x \cap G_y \subset \ker \chi$. If (x, y) is a transverse pair with respect to χ, we write $x \pitchfork y$.

If (x, y) is a transverse pair, then

$$\kappa(x \cdot g, y \cdot g) := \chi(g)\kappa(x, y)$$

is a well-defined nonzero function on the G-orbit of (x, y). Let

$$X \pitchfork Y = \{(x, y) \in X \times Y \mid x \pitchfork y\}.$$

Then $X \pitchfork Y$ is stable under the diagonal action of G on $X \times Y$. We have (see [Prasad 2015, Theorem 4.2.3]):

Theorem 2.5. *Let X and Y be finite G-sets, and $\chi : G \to F^*$ be a multiplicative character. For each orbit $O \in (X \pitchfork Y)/G$, choose a base point $(x_O, y_O) \in O$. Define*

$$\kappa_O(x, y) = \begin{cases} \chi(g) & \text{if } x = x_O \cdot g \text{ and } y = y_O \cdot g \text{ for some } g \in G, \\ 0 & \text{otherwise.} \end{cases}$$

For simplicity, write ξ_O for ξ_{κ_O}. Then the set

$$\{\xi_O \mid O \in (X \pitchfork Y)/G\}$$

is a basis for $\mathrm{Hom}_G(F[Y], F[X] \otimes \chi)$. *Consequently,*

$$\dim \mathrm{Hom}_G(F[Y], F[X] \otimes \chi) = |(X \pitchfork Y)/G)|.$$

In the special case where χ is the trivial character, we get:

Corollary 2.6. *Let X and Y be finite G-sets. For each orbit O in $(X \times Y)/G$ define*

$$\kappa_O(x, y) = \begin{cases} 1 & \text{if } (x, y) \in O, \\ 0 & \text{otherwise.} \end{cases}$$

Write $\xi_O = \xi_{\kappa_O}$. Then the set

$$\{\xi_O \mid O \in (X \times Y)/G\}$$

is a basis for $\mathrm{Hom}_G(F[Y], F[X])$. *Consequently,*

$$\dim \mathrm{Hom}_G(F[Y], F[X]) = |(X \times Y)/G|.$$

Given a function $\kappa : X \times Y \to F$, define

$$\kappa^*(y, x) = \kappa(x, y) \quad \text{for } x \in X,\ y \in Y.$$

The following is easy to see:

Lemma 2.7. *For any G-set X, the map $\xi_\kappa \mapsto \xi_{\kappa^*}$ is an anti-involution on the algebra* $\mathrm{End}_G(F[X])$.

2B. ***Structure constants of the Schur algebra.*** We recall the combinatorial interpretation of structure constants of the Schur algebra from [Geetha and Prasad 2014]. Let $[n] = \{1, \dots, n\}$ and

$$I(n, d) = \{\underline{i} := (i_1, \dots, i_d) \mid i_s \in [n]\}.$$

An element $w \in \boldsymbol{S}_d$ acts on $I(n, d)$ by permuting the coordinates:

$$(i_1, \dots, i_d) \cdot w = (i_{w(1)}, \dots, i_{w(d)}).$$

For $\underline{i} = (i_1, \dots, i_d) \in I(n, d)$, define

$$e_{\underline{i}} = e_{i_1} \otimes \cdots \otimes e_{i_d},$$

where e_i is the i-th coordinate vector in F^n. The vector space $(F^n)^{\otimes d}$ has a basis

$$\{e_{\underline{i}} \mid \underline{i} \in I(n, d)\}$$

and $w \in \boldsymbol{S}_d$ acts on a basis vector $e_{\underline{i}}$ as follows:

$$w \cdot e_{\underline{i}} = w \cdot (e_{i_1} \otimes \cdots \otimes e_{i_d}) = e_{i_{w^{-1}(1)}} \otimes \cdots \otimes e_{i_{w^{-1}(d)}}.$$

Let $F[I(n, d)]$ denote the space of all F-valued functions on $I(n, d)$. Mapping $e_{\underline{i}}$ to the indicator function of $\underline{i} \in I(n, d)$ defines an isomorphism of $(F^n)^{\otimes d}$ onto $F[I(n, d)]$. Thus $(F^n)^{\otimes d}$ can be regarded as a permutation representation of $\boldsymbol{S}_d$.

Let $B(n, d)$ denote the set of all configurations of d distinguishable balls, numbered $1, \dots, d$ in n boxes, numbered $1, \dots, n$. The symmetric group $\boldsymbol{S}_d$ acts on such configurations by permuting the d balls. An element of $B(n, d)$ is a set partition

$$\{1, \dots, d\} = S_1 \coprod \cdots \coprod S_n,$$

where S_i is the set of balls in the i-th box.

Lemma 2.8. *Given $\underline{i} \in I(n, d)$, let $b(\underline{i})$ denote the balls-in-boxes configuration in $B(n, d)$ where the i-th box contains the balls $\{s \mid i_s = i\}$. Then*

$$b : I(n, d) \to B(n, d)$$

is an $\boldsymbol{S}_d$-equivariant bijection of $I(n, d)$ onto $B(n, d)$.

By Corollary 2.6, a basis for $\mathrm{S}_F(n, d)$ is indexed by orbits for the diagonal action of $\boldsymbol{S}_d$ on $B(n, d) \times B(n, d)$.

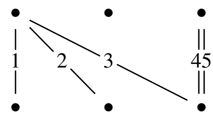
Definition 2.9 (labelled bipartite multigraph). Let $[n] = \{1, \dots, n\}$ (as before) and $[n'] = \{1', \dots, n'\}$. A labelling of a bipartite multigraph with vertex set $[n'] \coprod [n]$ and d edges is a function $l : [d] \to [n'] \times [n]$ such that, for each $(i', j) \in [n'] \times [n]$, the cardinality of $l^{-1}(i', j)$ is the number of edges joining i' and j. In other words, labels are assigned to edges without distinguishing between edges joining the same pair of vertices.

Given a pair $S = (S_1, \dots, S_n)$ and $T = (T_1, \dots, T_n)$ in $B(n, d)$, define a labelled bipartite graph $\gamma_{S,T}$ with multiple edges on the vertex set $[n'] \coprod [n]$ as follows:

> There are $|S_j \cap T_i|$ edges between i' and j, labelled by the numbers of the balls in $S_j \cap T_i$.

The bipartite multigraph is always drawn in two rows, with the vertices from $[n']$ in the upper row and vertices from $[n]$ in the lower row, numbered from left to right. Since the vertices are always labelled in this manner, the vertex labels can be omitted in the drawing.

Example 2.10. When $S = (\{1\}, \{2\}, \{3, 4, 5\})$ and $T = (\{1, 2, 3\}, \varnothing, \{4, 5\})$, the associated labelled multigraph is



Clearly, $(S, T) \mapsto \gamma_{S,T}$ is a bijection from $B(n, d) \times B(n, d)$ onto the set of labelled bipartite multigraphs with vertex set $[n'] \coprod [n]$ and d edges. The symmetric group $\boldsymbol{S}_d$ acts on $B(n, d) \times B(n, d)$ by permuting labels. Therefore the $\boldsymbol{S}_d$ orbits in $B(n, d) \times B(n, d)$ are obtained by simply forgetting the labels, leaving only the underlying bipartite multigraph. We write $\Gamma_{S,T}$ for the bipartite multigraph underlying $\gamma_{S,T}$. Such a graph can also be represented by its *adjacency matrix* (whose (i, j)-th entry is the number of edges joining i' and j), which is a matrix of nonnegative integers that sum to d.

In view of Corollary 2.6, we recover a result of [Geetha and Prasad 2014; Méndez 2001]:

Theorem 2.11. *Let $M(n, d)$ denote the set of all bipartite multigraphs with vertex set $[n'] \coprod [n]$ and d edges. For each $\Gamma \in M(n, d)$, define $\xi_\Gamma \in \mathrm{S}_F(n, d) = \mathrm{End}_{\boldsymbol{S}_d}(F[B(n, d)])$ by*

$$\xi_\Gamma f(S) = \sum_{\{T \mid \Gamma_{S,T} = \Gamma\}} f(T).$$

Then

$$\{\xi_\Gamma \mid \Gamma \in M(n, d)\}$$

is a basis for $\mathrm{S}_F(n, d)$.

Remark 2.12. If $(\underline{i}, \underline{j})$ has image Γ under the composition $I(n, d)^2 \to B(n, d)^2 \to M(n, d)$, then the basis element $\xi_{\underline{i},\underline{j}}$ of [Green 2007, Section 2.6] coincides with the basis element ξ_Γ of Theorem 2.11.

The structure constants $c^{\Gamma}_{\Gamma_1 \Gamma_2}$ are defined by

$$\xi_{\Gamma_1} \xi_{\Gamma_2} = \sum_{\Gamma \in M(n,d)} c^{\Gamma}_{\Gamma_1 \Gamma_2} \xi_\Gamma.$$

Definition 2.13. Let l, l_1 and l_2 be labellings of graphs Γ, Γ_1 and Γ_2 in $M(n, d)$, respectively. We say that (l_1, l_2) is compatible with l if, for all $s = 1, \ldots, d$, if we write $l_1(s) = (i'_1, j_1)$ and $l_2(s) = (i'_2, j_2)$, then

(2.13.a) $i'_2 = j_1$, and

(2.13.b) $l(s) = (i'_1, j_2)$.

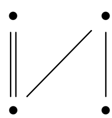
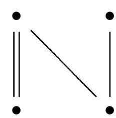
We obtain yet another enumerative description of the structure constants of the Schur algebra (see also [Green 2007, 2.3(b)] and [Geetha and Prasad 2014; Méndez 2001]).

Theorem 2.14. *Let l be any labelling of Γ. The structure constant $c^{\Gamma}_{\Gamma'\Gamma''}$ is the number of pairs (l_1, l_2) of labellings of Γ_1 and Γ_2 that are compatible with l.*

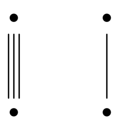
Before giving a proof, we illustrate the theorem with a few examples.

Example 2.15. Let $w \in \boldsymbol{S}_d$ be a permutation, and assume that $n \geq d$. Let $\Gamma(w)$ denote the bipartite graph where (i', j) is an edge if and only if $1 \leq i \leq d$ and $w(i) = j$. Then, for all $w_1, w_2 \in \boldsymbol{S}_d$, $\xi_{\Gamma(w_1)}\xi_{\Gamma(w_2)} = \xi_{\Gamma(w_1 w_2)}$.

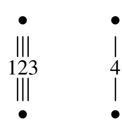
Example 2.16. Consider

$\Gamma_1 =$  $\Gamma_2 =$ 

To find $c^{\Gamma}_{\Gamma_1\Gamma_2}$, with

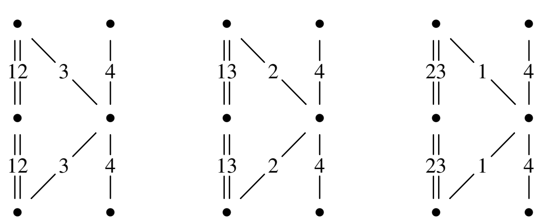
$\Gamma = \Gamma_3 =$ 

choose any labelling of Γ, such as

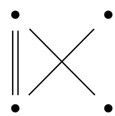


For this there are clearly three pairs of compatible labellings of Γ_1 and Γ_2, namely, we can choose which of the first three balls ends up in the second box of the middle row:

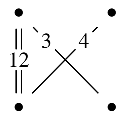
(11)



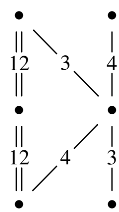
On the other hand, if

$\Gamma = \Gamma_4 =$ 

we may take the labelling

(12) 

For this labelling, the only compatible labellings of Γ_1 and Γ_2 are



It turns out that for no other $\Gamma \in \Gamma(n, d)$ is it possible to find even one compatible way of labelling Γ_1 and Γ_2, so we have

$$\xi_{\Gamma_1}\xi_{\Gamma_2} = 3\xi_{\Gamma_3} + \xi_{\Gamma_4}.$$

Example 2.17. Let $F_{n,n}$ denote the complete bipartite graph which has vertex set $[n'] \coprod [n]$, where every vertex in $[n']$ is connected to every vertex in $[n]$. Then the coefficient of $\xi_{F_{n,n}}$ in $\xi_{F_{n,n}}\xi_{F_{n,n}}$ is the number of Latin squares of order n [OEIS, Sequence A002860].

To see this, let l be any labelling of the edges of $F_{n,n}$. Given labellings l_1 and l_2 of $F_{n,n}$ that are compatible with l, define the (i, j)-th entry of the Latin square associated to (l_1, l_2) to be k if $l^{-1}(i', j) = l_2^{-1}(k', j) = l_1^{-1}(i', k)$. Remarkably, the number of Latin squares of order n is known only for $n = 1, \dots, 11$.

Proof of Theorem 2.14. Given a labelling l of Γ, define:

$$S_j = \cup_{i=1}^{n} l^{-1}(i', j) \quad \text{and} \quad U_i = \cup_{j=1}^{n} l^{-1}(i', j).$$

Then $S = (S_1, \dots, S_n)$, and $U = (U_1, \dots, U_n)$ are elements of $B(n, d)$, and by construction $\Gamma_{S,U} = \Gamma$. Now (9) implies that

(13) $$c^{\Gamma}_{\Gamma_1\Gamma_2} = \#\{T \in B(n, d) \mid \Gamma_{S,T} = \Gamma_1 \text{ and } \Gamma_{T,U} = \Gamma_2\}.$$

Given $T \in B(n, d)$ contributing to the above count, define labellings l_1 and l_2 of Γ_1 and Γ_2 by:

$$l_1^{-1}(i', j) = S_j \cap T_i \quad \text{and} \quad l_2^{-1}(i', j) = T_j \cap U_i.$$

Then (l_1, l_2) is compatible with l. Conversely, for every pair (l_1, l_2) compatible with l, take $T = (T_1, \ldots, T_n)$ where

$$T_k = \bigcup_{i'=1}^{n} l_1^{-1}(i', k) = \bigcup_{j=1}^{n} l_2^{-1}(k', j).$$

Then T contributes to the count in (13). $\square$

Example 2.18. In Example 2.16, the three compatible pairs of labels in (11) correspond to taking T as $(\{1, 2\}, \{3, 4\})$, $(\{1, 3\}, \{2, 4\})$, and $(\{2, 3\}, \{1, 4\})$, respectively, and the compatible pair of labels in (12) corresponds to $T = (\{1, 2\}, \{3, 4\})$.

2C. ***A basis for*** $\mathbf{S}_F^-(\boldsymbol{n}, \boldsymbol{d})$**.** By Theorem 2.5, a basis of $\mathrm{S}_F^-(n, d)$ is indexed by orbits in $B(n, d) \pitchfork B(n, d)/\boldsymbol{S}_d$. Here $\pitchfork$ denotes transversality with respect to the sign character $\mathrm{sgn} : S_n \to \{\pm 1\}$ (see Definition 2.4).

Lemma 2.19. *A pair $(S, T) \in B(n, d)^2$ lies in $B(n, d) \pitchfork B(n, d)$ if and only if $\gamma_{S,T}$ is a simple bipartite graph.*

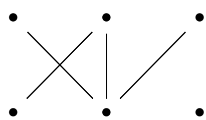
Proof. Let $S = (S_1, \ldots, S_n)$ and $T = (T_1, \ldots, T_n)$. If $\gamma_{S,T}$ is not simple, then there exist indices i and j such that $S_j \cap T_i$ contains at least two elements, say k and l. The transposition $(kl) \in \boldsymbol{S}_d$ stabilizes (S, T) but has $\mathrm{sgn}((kl)) = -1$, so $(S, T) \notin B(n, d) \pitchfork B(n, d)$.

However, if $\gamma_{S,T}$ is simple, then the simultaneous stabilizer of S and T in $\boldsymbol{S}_d$ is trivial, so $(S, T) \in B(n, d) \pitchfork B(n, d)$. $\square$

In order to specify a basis for $\mathrm{S}_F^-(n, d)$ using Theorem 2.5, we need to choose a base point for each $\boldsymbol{S}_d$-orbit in $B(n, d) \pitchfork B(n, d)$. We do this using Definition 2.20.

Definition 2.20 (standard labelling of a bipartite simple graph). Given a bipartite simple graph Γ with vertex set $[n'] \coprod [n]$, label each edge by its index when the edges (i', j) are arranged in increasing lexicographic order, with priority given to the upper index, i.e., $(i', j) < (r', s)$ if either $i' < r'$ or $i' = r'$ and $j < s$.

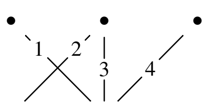
Example 2.21. Take

$\Gamma =$ 

The edges, written in lexicographic order, are

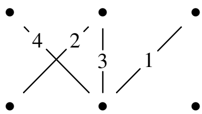
$$(1', 2), (2', 1), (2', 2), (3', 2).$$

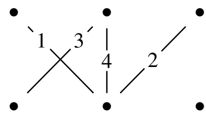
Therefore the standard labelling is



Definition 2.22 (sign of a labelling of a bipartite simple graph). Let l_0 denote the standard labelling of a simple bipartite graph Γ on $[n'] \coprod [n]$. Let $l : [d] \to [n'] \times [n]$ be a labelling of Γ (see Definition 2.9). The sign $\epsilon(\Gamma, l)$ of l is the sign of the permutation on $[d]$ which takes $l_0(i)$ to $l(i)$ for each i.

Example 2.23. For the graph from Example 2.21, the labellings

$l_1 =$ 

and $l_2 =$ 

give rise to permutations 4231 and 1342 respectively, so that $\epsilon(\Gamma, l_1) = -1$, and $\epsilon(\Gamma, l_2) = +1$.

Recall, from Section 2B, that $\gamma_{S,T}$ is a labelled bipartite graph associated to $(S, T) \in B(n, d) \times B(n, d)$, whose underlying unlabelled graph is denoted by $\Gamma_{S,T}$. Let $l_{S,T} : [d] \to [n] \times [n']$ denote the labelling of $\gamma_{S,T}$, and write $\epsilon(\gamma_{S,T})$ for $\epsilon(\Gamma_{S,T}, l_{S,T})$.

Theorem 2.24. *Let $N(n, d)$ denote the set of all bipartite simple graphs with vertex set $[n'] \coprod [n]$ and d edges. For each $\Gamma \in N(n, d)$, define $\zeta_\Gamma \in \mathrm{S}^-_F(n, d) =$ $\mathrm{Hom}_{S_d}(F[B(n, d)], F[B(n, d)] \otimes \mathrm{sgn})$ by*

$$\zeta_\Gamma f(S) = \sum_{\{T \mid \Gamma_{S,T} = \Gamma\}} \epsilon(\gamma_{S,T}) f(T).$$

The set

$$\{\zeta_\Gamma \mid \Gamma \in N(n, d)\}$$

forms a basis of $\mathrm{S}^-_F(n, d)$.

Proof. Recall that we choose the pair (S_0, T_0) corresponding to the standard labelling l_0 of Γ as the base point of the orbit associated to Γ. A pair (S, T) is in the orbit of (S_0, T_0) if and only if $\Gamma_{S,T} = \Gamma$. And the sign of the permutation $w \in S_d$ such that $S = S_0.w$ and $T = T_0.w$ is the sign of the labelled bipartite graph $\gamma_{S,T}$. So the integral kernel κ_Γ of the operator ζ_Γ is

$$\kappa_\Gamma(S, T) = \begin{cases} \epsilon(\gamma_{S,T}) & \text{if } \Gamma_{S,T} = \Gamma, \\ 0 & \text{otherwise.} \end{cases}$$

So the theorem follows from Theorem 2.5. □

Theorem 2.14 tells us how to multiply two elements of the subalgebra $\mathrm{S}_F(n, d)$ of $\mathrm{AS}_F(n, d)$. The remaining structure constants are given by the following theorem.

Theorem 2.25. *The remaining structure constants are given as follows*:

(2.25.a) *Given* $\Gamma_1 \in M(n,d)$ *and* $\Gamma_2 \in N(n,d)$,

$$\xi_{\Gamma_1}\zeta_{\Gamma_2} = \sum_{\Gamma \in N(n,d)} c^{\Gamma}_{\Gamma_1\Gamma_2}\zeta_{\Gamma},$$

where

$$c^{\Gamma}_{\Gamma_1\Gamma_2} = \sum_{l_1,l_2} \epsilon(\Gamma_2, l_2),$$

and the sum runs over all labellings l_1 *and* l_2 *of* Γ_1 *and* Γ_2*, respectively, that are compatible with the standard labelling* l *of* Γ.

(2.25.b) *Given* $\Gamma_1 \in N(n,d)$ *and* $\Gamma_2 \in M(n,d)$,

$$\zeta_{\Gamma_1}\xi_{\Gamma_2} = \sum_{\Gamma \in N(n,d)} c^{\Gamma}_{\Gamma_1\Gamma_2}\zeta_{\Gamma},$$

where

$$c^{\Gamma}_{\Gamma_1\Gamma_2} = \sum_{l_1,l_2} \epsilon(\Gamma_1, l_1),$$

and the sum runs over all labellings l_1 *and* l_2 *of* Γ_1 *and* Γ_2*, respectively, that are compatible with the standard labelling* l *of* Γ.

(2.25.c) *Given* $\Gamma_1 \in N(n,d)$ *and* $\Gamma_2 \in N(n,d)$,

$$\zeta_{\Gamma_1}\zeta_{\Gamma_2} = \sum_{\Gamma \in M(n,d)} c^{\Gamma}_{\Gamma_1\Gamma_2}\xi_{\Gamma},$$

where

$$c^{\Gamma}_{\Gamma_1\Gamma_2} = \sum_{l_1,l_2} \epsilon(\Gamma_1, l_1)\epsilon(\Gamma_2, l_2),$$

and the sum runs over all labellings l_1 *and* l_2 *of* Γ_1 *and* Γ_2*, respectively, that are compatible with a fixed labelling* l *of* Γ.

Proof. Given a labelling l of Γ, construct S and U in $B(n,d)$ as in the proof of Theorem 2.14. Define $\kappa_{\Gamma_2} : B(n,d) \times B(n,d) \to F$ by

$$\kappa_{\Gamma_2}(T,U) = \epsilon(\gamma_{T,U}).$$

Then ζ_{Γ_2} is the integral operator $\xi_{\kappa_{\Gamma_2}}$, as in (8). Then, by (9), the structure constant in (2.25.a) of the theorem is given by

$$c^{\Gamma}_{\Gamma_1\Gamma_2} = \sum_{T} \zeta_{\Gamma_2}(T,U),$$

where the sum runs over all $T \in B(n,d)$ such that $\Gamma_{S,T} = \Gamma_1$ and $\Gamma_{T,U} = \Gamma_2$. Defining labellings l_1 and l_2 of Γ_1 and Γ_2 as in the proof of Theorem 2.14, we find that $\zeta_{\Gamma_2}(T,U) = \epsilon(\Gamma_2, l_2)$, proving (2.25.a). The proofs of the remaining assertions are similar. $\square$

Definition 2.26. Given $\Gamma \in M(n,d) \sqcup N(n,d)$, we define Γ^* to be the horizontal reflection of Γ, i.e., i' is connected to j in Γ^* if and only if j' is connected to i in Γ. The operation $*$ on the set $M(n,d) \sqcup N(n,d)$ is an involution.

Lemma 2.27. *For every $\Gamma \in N(n,d)$, let l_0 denote its standard labelling. Let l_0^* denote the labelling of Γ^* given by $l_0^*(i',j) = l_0(j',i)$. Then the linear map $\mathrm{AS}_F(n,d) \to \mathrm{AS}_F(n,d)$ defined by*

$$\xi_\Gamma \mapsto \xi_{\Gamma^*} \text{ for } \Gamma \in M(n,d),$$
$$\zeta_\Gamma \mapsto \epsilon(\Gamma^*, l_0^*)\zeta_{\Gamma^*} \text{ for } \Gamma \in N(n,d)$$

is an anti-involution of $\mathrm{AS}_F(n,d)$.

Remark 2.28. The above involution, when restricted to the Schur algebra, is the same as the one described by Green [2007, Section 2.7].

Proof. We show that the linear map in Lemma 2.27 is the same as the anti-involution in Lemma 2.7 with $X = B(n,d)$ and $G = \boldsymbol{A}_d$.

For $\Gamma \in M(n,d)$, ξ_Γ is the integral operator with kernel

$$\kappa_\Gamma(S,T) = \begin{cases} 1 & \text{if } \Gamma_{S,T} = \Gamma, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\Gamma_{T,S} = \Gamma^*_{S,T}$, $\kappa_\Gamma^* = \kappa_{\Gamma^*}$.

For $\Gamma \in N(n,d)$, ζ_Γ is the integral operator with kernel

$$\kappa_\Gamma(S,T) = \begin{cases} \epsilon(\gamma_{S,T}) & \text{if } \Gamma_{S,T} = \Gamma, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, if $\gamma_{S,T} = (\Gamma, l_0)$, then $\gamma_{T,S} = (\Gamma^*, l_0^*)$. Therefore,

$$\begin{aligned} \kappa_{\Gamma^*}(T,S) &= \epsilon(\gamma_{T,S}) \\ &= \epsilon(\Gamma^*, l_0^*)\kappa_\Gamma(S,T). \end{aligned}$$

So the kernels κ_{Γ^*} and $\epsilon(\Gamma^*, l_0^*)\kappa_\Gamma^*$ coincide at (T,S), and hence on its entire $\boldsymbol{S}_d$-orbit in $B(n,d)$. □

We illustrate the above results with an example that will be used in the proof of Lemma 4.1.

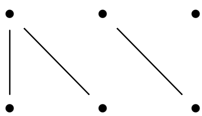
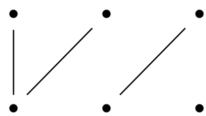
Example 2.29. Recall that $\Lambda(n,d)$ denotes the set of all weak compositions of d with at most n parts. For $\lambda \in \Lambda(n,d)$ with $n \geq d$, let $\Gamma_\lambda \in N(n,d)$ denote the bipartite graph where i' is connected to j if

$$\lambda_1 + \cdots + \lambda_{i'-1} < j \leq \lambda_1 + \cdots + \lambda_{i'}.$$

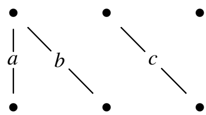
Then we have

$$\zeta_{\Gamma_\lambda}\zeta_{\Gamma_\lambda^*} = \lambda_1! \cdots \lambda_n! \xi_{\Gamma_\lambda^0},$$

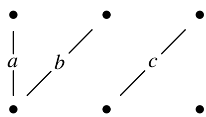
where $\Gamma_\lambda^0 \in M(n,d)$ is the bipartite multigraph where i' is connected to i by λ_i edges. For example,

$\Gamma^*_{(2,1)} =$  and $\Gamma_{(2,1)} =$ 

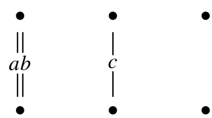
For a labelling

$l_1 =$ 

of $\Gamma^*_{(2,1)}$ only the labelling

$l_2 =$ 

of $\Gamma_{(2,1)}$, and the labelling

$l =$ 

of $\Gamma^0_{(2,1)}$ are such that (l_1, l_2) are compatible with l. Moreover, $\epsilon(\Gamma^*_{(2,1)}, l_1) = \epsilon(\Gamma_{(2,1)}, l_2)$. Interchanging the labels a and b in l_1 and l_2, respectively, gives another pair of labels compatible with l, so that $\zeta_{\Gamma_\lambda}\zeta_{\Gamma^*_\lambda} = 2\xi_{\Gamma^0_\lambda}$.

The remaining results in this section help us understand the structure of $S_F^-(n,d)$ as an $S_F(n,d)$-module. Some of them will play an important role in understanding Koszul duality (Section 4).

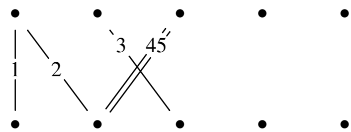
The notion of standard labelling (Definition 2.20) of graphs in $N(n,d)$ can be extended to graphs in $M(n,d)$ as follows: when an edge (i', j) occurs with multiplicity m, it is simply listed m times when the edges are arranged in lexicographic order with priority given to the upper index. Example 2.10 is the standard labelling of its underlying graph.

Definition 2.30. For $n \geq d$, define the following simple bipartite graphs associated to $\Gamma \in M(n,d)$:

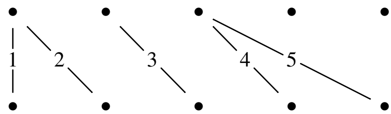
(2.30.a) Let $D(\Gamma) \in N(n,d)$ be the graph with edges (i', s) for every edge (i', j) with label s under the standard labelling of Γ.

(2.30.b) Let $U(\Gamma) \in N(n,d)$ be the graph with edges (s', j) for every edge (i', j) with label s under the standard labelling of Γ.

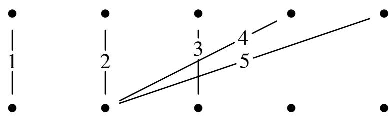
Example 2.31. Let $n = 5$, $d = 5$, and Γ (with its standard labelling) be given by



Then

$D(\Gamma) =$ 

and

$U(\Gamma) =$ 

The significance of the elements $U(\Gamma)$ and $D(\Gamma)$, for $\Gamma \in M(n, d)$, is elaborated in the following lemmas.

Lemma 2.32. *Let $n \geq d$ and $\Gamma \in N(n, d)$. Then $\zeta_\Gamma = \xi_{U(\Gamma)}\zeta_{\Gamma_{\lambda_0}}\xi_{D(\Gamma)}$, where $\lambda_0 = (1^d, 0^{n-d}) \in \Lambda(n, d)$. Consequently, $\mathrm{S}_F^-(n, d)$ is a cyclic $(\mathrm{S}_F(n, d), \mathrm{S}_F(n, d))$-bimodule.*

Proof. This can be done in two steps. Firstly, $\zeta_{\Gamma_{\lambda_0}}\xi_{D(\Gamma)} = \zeta_{D(\Gamma)}$, and secondly $\xi_{U(\Gamma)}\zeta_{D(\Gamma)} = \zeta_\Gamma$. We indicate the proof of the second identity (the first is similar): Let l_0, l_1, and l_2 be the standard labellings of Γ, $U(\Gamma)$ and $D(\Gamma)$, respectively. The labellings (l_1, l_2) of $D(\Gamma)$ and $U(\Gamma)$ are the only ones that are compatible with l_0. This is because, for the edge (i', j) of Γ labelled s, s is the unique vertex such that i' is connected to s in $D(\Gamma)$ and s' is connected to j in $U(\Gamma)$. The identity now follows from (2.25.a). $\square$

Similarly, we get Lemma 2.33:

Lemma 2.33. *For $n \geq d$ and $\Gamma \in N(n, d)$, we have $\zeta_\Gamma = \zeta_{U(\Gamma)}\xi_{D(\Gamma)}$ and $\zeta_\Gamma = \xi_{U(\Gamma)}\zeta_{D(\Gamma)}$.*

Corollary 2.34. *As a left $\mathrm{S}_F(n, d)$-module, $\mathrm{S}_F^-(n, d)$ is generated by*

$$\{\zeta_{\Gamma_\lambda^*} \mid \lambda \in \Lambda(n, d)\},$$

and as a right $\mathrm{S}_F(n, d)$-module, it is generated by $\{\zeta_{\Gamma_\lambda} \mid \lambda \in \Lambda(n, d)\}$. Here Γ_λ is the graph associated to λ in Example 2.29.

Proof. For any $\Gamma \in M(n, d)$, $D(\Gamma)$ is of the form Γ_λ^* for some $\lambda \in \Lambda(n, d)$, so the statement for left modules follows from the second identity in Lemma 2.33. The statement for right modules follows by applying Lemma 2.27 to the first identity in Lemma 2.33. $\square$

Lemma 2.35. *Let $n \geq d$ and $\Gamma \in M(n, d)$. Then, for $\Gamma' \in M(n, d)$, the structure constant of $\zeta_{U(\Gamma)}$ in the product $\xi_{\Gamma'}\zeta_{D(\Gamma)^*}$ is $\delta_{\Gamma',\Gamma}$.*

Proof. The edge (i', j) with label s in the standard labelling of Γ gives rise to an edge (s', j) with standard label s in $U(\Gamma)$. The graph $D(\Gamma)^*$ has only one edge originating at s', namely (s', i). Therefore, for any compatible pair (l_1, l_2) of labellings of $D(\Gamma)^*$ and Γ', this edge must have label s. Thus Γ' must have an edge (i', j) labelled s. In other words, $\Gamma' = \Gamma$ and l_2 is its standard labelling. □

3. Abstract Koszul duality

3A. ***The algebra.*** Recall [Bourbaki 1974, Chapter III, Section 3.1] that a $\boldsymbol{Z}/2\boldsymbol{Z}$ grading on a ring AS is a decomposition $\mathrm{AS} = \mathrm{S} \oplus \mathrm{S}^-$ into additive subgroups such that S is a subring, S^- is closed under left and right multiplication by elements of S, and for any $\alpha, \beta \in \mathrm{S}^-$, $\alpha\beta \in \mathrm{S}$. This $\boldsymbol{Z}/2\boldsymbol{Z}$-grading gives rise to

(a) an (S, S)-bimodule structure on S^-,

(b) and an (S, S)-bimodule homomorphism $\phi : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \to \mathrm{S}$ (induced by the S-balanced bilinear map $(\alpha, \beta) \mapsto \alpha\beta$ for $\alpha, \beta \in \mathrm{S}^-$).

Example 3.1. We may take $\mathrm{AS} = \mathrm{AS}_F(n, d) = \mathrm{End}_{A_d}((F^n)^{\otimes d})$ and $\mathrm{S} = \mathrm{S}_F(n, d) = \mathrm{End}_{\mathrm{S}_d}((F^n)^{\otimes d})$, for any field F with characteristic different from 2.

3B. ***Modules.*** Let M be an AS-module. The AS-module structure can be viewed as a linear map:

$$\mathrm{AS} \otimes_{\boldsymbol{Z}} M = (\mathrm{S} \oplus \mathrm{S}^-) \otimes_{\boldsymbol{Z}} M = (\mathrm{S} \otimes_{\boldsymbol{Z}} M) \oplus (\mathrm{S}^- \otimes_{\boldsymbol{Z}} M) \to M.$$

So M is an S-module, and restriction of the module action $\mathrm{AS} \otimes_{\boldsymbol{Z}} M \to M$ to $\mathrm{S}^- \otimes_{\boldsymbol{Z}} M$ induces an S-module homomorphism

$$\theta_M : \mathrm{S}^- \otimes_{\mathrm{S}} M \to M. \tag{14}$$

Furthermore, this homomorphism θ_M has the property that the diagram

$$\begin{tikzcd} \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \otimes_{\mathrm{S}} M \arrow[r, "\phi\otimes\mathrm{id}_M"] \arrow[d, "\mathrm{id}_{\mathrm{S}^-}\otimes\theta_M"'] & S \otimes_{\mathrm{S}} M \arrow[d, equal] \\ \mathrm{S}^- \otimes_{\mathrm{S}} M \arrow[r, "\theta_M"'] & M \end{tikzcd} \tag{15}$$

commutes.

Definition 3.2. Given an (S, S)-bimodule S^- and an (S, S)-bimodule homomorphism $\phi : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \to \mathrm{S}$, for an S-module N, an S-module homomorphism

$\theta : \mathrm{S}^- \otimes_{\mathrm{S}} N \to N$ is said to be compatible with ϕ if the diagram

$$\begin{tikzcd} \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \otimes_{\mathrm{S}} N \arrow[r, "\phi\otimes\mathrm{id}_N"] \arrow[d, "\mathrm{id}_{\mathrm{S}^-}\otimes\theta"'] & S \otimes_{\mathrm{S}} N \arrow[d, equal] \\ \mathrm{S}^- \otimes_{\mathrm{S}} N \arrow[r, "\theta"'] & N \end{tikzcd} \tag{16}$$

commutes.

3C. ***Duality.*** Let $\mathrm{AS}=\mathrm{S}\oplus\mathrm{S}^-$ be as before. Define a functor $D:\mathrm{S}\text{-Mod}\to\mathrm{S}\text{-Mod}$ by

$$D(M) = \mathrm{S}^- \otimes_{\mathrm{S}} M,$$

for every S-module M. Given S-modules M and N, and an S-module homomorphism $f : M \to N$, let $D(f) = \mathrm{id}_{\mathrm{S}^-} \otimes f : D(M) \to D(N)$. We call the resulting functor $D : \mathrm{S}\text{-Mod} \to \mathrm{S}\text{-Mod}$ an *abstract Koszul duality functor*. In Section 4 it will be shown that, in the setting of Example 3.1 (the alternating Schur algebra), abstract Koszul duality is essentially the Koszul duality functor of Krause [2013].

The commutative diagram (16), defining the compatibility of θ with ϕ, can be rewritten in terms of abstract Koszul duality as

$$\begin{tikzcd} D^2(N) \arrow[r, "\phi\otimes\mathrm{id}_N"] \arrow[d, "D(\theta)"'] & N \\ D(N) \arrow[ur, "\theta"'] & \end{tikzcd} \tag{17}$$

Definition 3.3. Given an (S, S)-bimodule S^- and an (S, S)-bimodule homomorphism $\phi : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \to \mathrm{S}$, let (S, ϕ)-Mod denote the category whose objects are pairs (N, θ), where N is an S-module, and $\theta : D(N) \to N$ is compatible with ϕ. A morphism $(N, \theta) \to (N', \theta')$ is an S-module homomorphism $f : N \to N'$ such that the diagram

$$\begin{tikzcd} D(N) \arrow[r, "\theta"] \arrow[d, "D(f)"'] & N \arrow[d, "f"] \\ D(N') \arrow[r, "\theta'"] & N' \end{tikzcd}$$

commutes.

Theorem 3.4. *Given an* AS*-module* M, *let* θ_M *be as in* (14). *Then* $M \mapsto (M, \theta_M)$ *is an isomorphism of categories* $\mathrm{AS}\text{-Mod} \to (\mathrm{S}, \phi)\text{-Mod}$.

Proof. Given an object (N, θ) in (S, ϕ)-Mod, the compatibility of θ with ϕ allows the S-module structure on N to be extended to an AS-module structure. This constructs the inverse of the functor in the theorem. $\square$

Given an S-module N, the morphism $\phi : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \to \mathrm{S}$ gives rise to a natural transformation $\eta : D^2 \to \mathrm{id}_{\text{S-Mod}}$, defined as the composition

$$\begin{tikzcd} D^2(N) \arrow[r, "\phi\otimes\mathrm{id}_N"] \arrow[dr, "\eta_N"'] & \mathrm{S}\otimes_{\mathrm{S}} N \arrow[d, equal] \\ & N \end{tikzcd} \tag{18}$$

Theorem 3.5. *Let* AS, S, S^- *and* ϕ *be as in Section 3A. The following are equivalent*:

(3.5.a) *The map* $\phi : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \to \mathrm{S}$ *is an isomorphism.*

(3.5.b) *The natural transformation* $\eta : D^2 \to \mathrm{id}_{\text{S-Mod}}$ *is a natural isomorphism.*

(3.5.c) *For every object* (N, θ) *in* (S, ϕ)-Mod*,* $\theta : \mathrm{S}^- \otimes_{\mathrm{S}} N \to N$ *is an isomorphism of* S*-modules.*

Proof. To see that (3.5.a) implies (3.5.b), observe from the diagram (18) that if ϕ is an isomorphism, then η_N is an isomorphism for every N. It follows that η is a natural isomorphism. For the converse, taking $N = \mathrm{S}$, the commutativity of (18) shows that ϕ is an isomorphism.

To see that (3.5.a) implies (3.5.c), note that the commutativity of (16) implies that, if ϕ is an isomorphism, then θ is an epimorphism, and $\mathrm{id}_{\mathrm{S}^-} \otimes \theta$ is a monomorphism. Since tensoring is a right-exact functor, it follows that $\mathrm{id}_{\mathrm{S}^-} \otimes \theta$ is also an epimorphism, and hence an isomorphism. Since $\phi \otimes \mathrm{id}_N$ is also an isomorphism, the inverse of θ can be constructed by reversing the arrows in (16). For the converse, just take $N = \mathrm{S}$ in (3.5.c). □

3D. ***Abstract Ringel duality.*** Let S^- be an (S, S)-bimodule. Denote the left S-module S^- by ${}_{\mathrm{S}}\mathrm{S}^-$. For a left S-module M, the homomorphism space $\mathrm{Hom}_{\mathrm{S}}({}_{\mathrm{S}}\mathrm{S}^-, M)$ inherits the structure of a left S-module from the right S-module structure on S^-. Motivated by [Ringel 1991, Section 6], we call the functor

$$\mathrm{Hom}_{\mathrm{S}}({}_{\mathrm{S}}\mathrm{S}^-, -) : \text{S-Mod} \to \text{S-Mod}$$

the abstract Ringel duality functor on S-Mod. It is clear that the abstract Koszul duality functor is the left adjoint of the abstract Ringel duality functor.

3E. ***Abstract simple modules.*** In general, it is not clear how simple AS-modules can be classified using simple S-modules and Koszul duality. In this section, we give some results in this direction. These are enough to give a complete solution in the semisimple case.

Let M be a simple S-module. We consider the following cases:

3E1. *DM is isomorphic to M.* If $\eta_M : D^2M \to M$ is zero, then $(M, 0)$ (where 0 is the zero map from $DM \to M$) is the unique ϕ-compatible morphism. Otherwise, any nonzero morphism $\theta : DM \to M$ is an isomorphism. Schur's lemma implies that $\theta \circ D\theta = a\eta_M$ for some $a \in (\mathrm{End}_\mathrm{S}\, M)^*$ (the multiplicative group of nonzero elements in the division algebra $\mathrm{End}_\mathrm{S}\, M$). If a has a square root in $(\mathrm{End}_\mathrm{S}\, M)^*$, then θ can be normalized to make it a ϕ-compatible morphism. Moreover, after normalization, $\pm\theta$ are two ϕ-compatible morphisms, leading to two nonisomorphic AS-modules. Also, in this case, $(M, \pm\theta)$ are simple, because their restrictions to S are simple. On the other hand, if a does not have a square root in $\mathrm{End}_\mathrm{S}(M)$, then there is no simple AS-module whose restriction to S is isomorphic to M.

3E2. $DM = 0$. In this case $(M, 0)$ is the unique AS-module whose restriction to S is isomorphic to M.

3E3. *DM is simple, but not isomorphic to M, $\eta_M \neq 0$.* Let $\tilde{M} = DM \oplus M$. We have $D\tilde{M} = D^2M \oplus DM$. Any morphism $\theta : D\tilde{M} \to \tilde{M}$ can be written in matrix form as

$$\theta = \begin{pmatrix} X & Y \\ Z & W \end{pmatrix},$$

where $X : D^2M \to DM$, $Y : DM \to DM$, $Z : D^2M \to M$, and $W : DM \to M$. By Schur's lemma, $W = 0$. The compatibility of θ with ϕ becomes

$$\begin{pmatrix} X & Y \\ Z & 0 \end{pmatrix}\begin{pmatrix} DX & DY \\ DZ & 0 \end{pmatrix} = \begin{pmatrix} \eta_{DM} & 0 \\ 0 & \eta_M \end{pmatrix}.$$

Multiplying out the left-hand side gives

$$\begin{pmatrix} XDX + YDZ & XDY \\ ZDX & ZDY \end{pmatrix} = \begin{pmatrix} \eta_{DM} & 0 \\ 0 & \eta_M \end{pmatrix}.$$

Since $\eta_M \neq 0$, $DY \neq 0$, and so $Y \neq 0$. Since DM is simple, by Schur's lemma, Y is invertible. Since D is a functor, DY is also invertible. Hence, equality of top right entries implies that $X = 0$. Moreover, $Z = \eta_M DY^{-1}$. In other words, θ is of the form

$$\theta_Y = \begin{pmatrix} 0 & Y \\ \eta_M DY^{-1} & 0 \end{pmatrix}.$$

Lemma 3.6. *For all $Y, Y' \in (\mathrm{End}_\mathrm{S}\, M)^*$, $\mathrm{Hom}_{\mathrm{AS}}((\tilde{M}, \theta_Y), (\tilde{M}, \theta_{Y'}))$ is nonzero, and $\mathrm{End}_{\mathrm{AS}}(\tilde{M}, \theta_Y)$ is a division ring.*

Proof. Any AS-module morphism $(\tilde{M}, \theta_y) \to (\tilde{M}, \theta_{y'})$ can be written in matrix form as

$$\begin{pmatrix} X & 0 \\ 0 & W \end{pmatrix}, \text{ where } X \in \mathrm{End}_\mathrm{S}\, DM \text{ and } W \in \mathrm{End}_\mathrm{S}\, M,$$

and must satisfy

$$\begin{pmatrix} X & 0 \\ 0 & W \end{pmatrix}\begin{pmatrix} 0 & Y \\ \eta_M DY^{-1} & 0 \end{pmatrix} = \begin{pmatrix} 0 & Y' \\ \eta_M DY'^{-1} & 0 \end{pmatrix}\begin{pmatrix} DX & 0 \\ 0 & DW \end{pmatrix}.$$

We get $XY = Y'DW$ and $W\eta_M DY^{-1} = \eta_M DY'^{-1}DX$. Taking $W = \mathrm{id}_M$ and $X = Y'Y^{-1}$ gives a nonzero element of $\mathrm{Hom}_{\mathrm{AS}}((\tilde{M}, \theta_Y), (\tilde{M}, \theta_{Y'}))$. When $Y = Y'$, then we have $X = YDWY^{-1}$, so that X is nonzero (and hence invertible) if and only if W is. It follows that every nonzero element of $\mathrm{End}_{\mathrm{AS}}(\tilde{M}, \theta_Y)$ is invertible. □

Lemma 3.7. *Let M be an* S*-module. Then the* AS*-module* $\mathrm{AS} \otimes_{\mathrm{S}} M$ *is isomorphic to* $(DM \oplus M, \theta)$, *where θ is given by the matrix*

$$\theta = \begin{pmatrix} 0 & \mathrm{id}_{DM} \\ \eta_M & 0 \end{pmatrix}.$$

Proof. Note that

$$\mathrm{AS} \otimes_{\mathrm{S}} M = (\mathrm{S}^- \otimes_{\mathrm{S}} M) \oplus (\mathrm{S} \otimes_{\mathrm{S}} M) = DM \oplus M.$$

The map θ comes from the action of S^- on this AS-module, which gives $\eta_M : D^2M \to M$ on the first summand, and $\mathrm{id}_{DM} : DM \to DM$ on the second. □

Theorem 3.8. *The* AS*-module* $(DM \oplus M, \theta_Y)$ *defined above is isomorphic to* $\mathrm{AS} \otimes_{\mathrm{S}} M$ *for every* $Y \in (\mathrm{End}_{\mathrm{S}}\, DM)^*$. *Consequently, whenever M and DM are simple, nonisomorphic* S*-modules, and* $\eta_M \neq 0$, *then* $\mathrm{AS}\otimes_{\mathrm{S}} M$ *is, up to isomorphism, the unique simple* AS*-module whose restriction to* S *contains M.*

Proof. To see that $\mathrm{AS} \otimes_{\mathrm{S}} M = DM \oplus M$ is simple, note that its only proper nontrivial S-submodules are M and DM. But M is not AS-invariant because S^- maps M onto DM. Also, DM is not AS-invariant, because S^- maps DM onto D^2M. Since $\eta_M \neq 0$, D^2M cannot be contained in DM. The theorem now follows from Lemma 3.6. □

3E4. *The case where ϕ is an isomorphism.* When $\phi : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \to \mathrm{S}$ is an isomorphism, the preceding results, using Theorem 3.5, can be summarized in the following form:

Theorem 3.9. *Suppose that* AS *is endowed with a* $\mathbf{Z}/2\mathbf{Z}$*-grading* $\mathrm{AS} = \mathrm{S} \oplus \mathrm{S}^-$, *and* $\phi : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \to \mathrm{S}$ (*as defined in Section 3A*) *is an isomorphism. Let M be a simple* S*-module. Then*:

(3.9.a) *Suppose there exists an isomorphism $\theta : DM \to M$. Then θ can be scaled to become compatible with ϕ. There exist at most two isomorphism classes of simple* AS*-modules* $(M, \pm\theta)$ *whose restrictions to* S *are isomorphic to M. If* $(\mathrm{End}_{\mathrm{S}}\, M)^*$ *is a* 2*-divisible group, then these two classes always exist.*

(3.9.b) *Otherwise, up to isomorphism,* $\mathrm{AS} \otimes_{\mathrm{S}} M$ *is the unique simple* AS*-module whose restriction to* S *contains* M *as a submodule. Also,* $\mathrm{AS} \otimes_{\mathrm{S}} M$ *and* $\mathrm{AS} \otimes_{\mathrm{S}} DM$ *are isomorphic as* AS*-modules.*

Corollary 3.10. *Suppose* F *is an algebraically closed field of characteristic different from* 2*. Let* $\mathrm{AS} = \mathrm{S} \oplus \mathrm{S}^-$ *be a* $\mathbf{Z}/2\mathbf{Z}$*-graded* F*-algebra. A complete set of isomorphism classes of simple* AS*-modules is given by*

(3.10.c) $(M, \pm\theta)$ *(defined in Section 3E1), as* M *runs over isomorphism classes of simple* S*-modules such that* DM *is isomorphic to* M*,*

(3.10.d) $\mathrm{AS} \otimes_{\mathrm{S}} M$*, as* M *runs over isomorphism classes of all unordered pairs* $\{M, M'\}$ *of nonisomorphic mutually dual simple* S*-modules.*

4. Koszul duality for modules over Schur algebra

In this section, let S denote the Schur algebra $\mathrm{S}_F(n, d)$, and let S^- denote the (S, S)-bimodule $\mathrm{S}^-_F(n, d)$. We now use our combinatorial methods from Section 2 to determine when abstract Koszul duality is an equivalence.

Lemma 4.1. *When the characteristic of* F *is* 0 *or greater than* d*, and* $n \geq d$*, the map* $\phi : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \to \mathrm{S}$ *is an isomorphism.*

Proof. For each $\lambda \in \Lambda(n, d)$, let $\Gamma^0_\lambda \in M(n, d)$ be the bipartite multigraph with λ_i edges from i' to i (and no other edges), as in Example 2.29. Then

$$\mathrm{id}_{\mathrm{S}} = \sum_{\lambda \in \Lambda(n,d)} \xi_{\Gamma^0_\lambda}. \tag{19}$$

Therefore by Example 2.29,

$$\mathrm{id}_S = \sum_{\lambda \in \Lambda(n,d)} \frac{1}{\lambda_1! \cdots \lambda_n!} \zeta_{\Gamma_\lambda} \zeta^*_{\Gamma_\lambda}. \tag{20}$$

Therefore the image of ϕ, which is a two-sided ideal of S, contains the identity element, and therefore is all of S.

The injectivity of ϕ can be proved using a dimension count. Let N^d (resp., N_d) denote the graphs in $N(n, d)$ with upper (resp., lower) degree sequence $(1^d, 0^{n-d})$. Let $\Gamma(w) \in N(n, d)$ be as in Example 2.15. For $\Gamma \in N^d$, define $\Gamma \cdot w \in N^d$ by $\xi_\Gamma \xi_{\Gamma(w)} = \xi_{\Gamma \cdot w}$. Similarly, for $\Gamma \in N_d$, define $w \cdot \Gamma \in N_d$ by $\xi_{\Gamma(w)} \xi_\Gamma = \xi_{w \cdot \Gamma}$. Consider the equivalence relation on $N^d \times N_d$ where $(\Gamma, \Gamma') \sim (\Gamma \cdot w^{-1}, w \cdot \Gamma')$ for $w \in \boldsymbol{S}_d$. Let $N^d \times_{\boldsymbol{S}_d} N_d$ denote the set of equivalence classes.

Now, given $(\Gamma', \Gamma'') \in N^d \times N_d$, define $\Gamma = \Phi(\Gamma', \Gamma'') \in M(n, d)$ to be the graph for which the number of edges joining (i', j) is the number of indices $1 \leq k \leq n$ such that (i', k) is an edge of Γ'' and (k', j) is an edge of Γ'. This map induces an injective function $\overline{\Phi}: N^d \times_{S_d} N_d \to M(n,d)$. Moreover, $\Gamma = \Phi(U(\Gamma), D(\Gamma))$, so $\overline{\Phi}: N^d \times_{S_d} N_d \to M(n,d)$ is a bijection.

The elements $\zeta_{\Gamma'} \otimes \zeta_{\Gamma''}$ as Γ' and Γ'' run over $N(n,d)$, span $\mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^-$. We have

$$\begin{aligned}\zeta_\Gamma \otimes \zeta_{\Gamma'} &= \zeta_{U(\Gamma)}\xi_{D(\Gamma)} \otimes \xi_{U(\Gamma')}\zeta_{D(\Gamma')} \qquad \text{(from Lemma 2.33)}\\ &= \zeta_{U(\Gamma)} \otimes \xi_{D(\Gamma)}\xi_{U(\Gamma')}\zeta_{D(\Gamma')}.\end{aligned}$$

Now $U(\Gamma) \in N^d$ and $\xi_{D(\Gamma)}\xi_{U(\Gamma')}\zeta_{D(\Gamma')}$ lies in the span of $\zeta_{\Gamma'''}$, for $\Gamma''' \in N_d$. Therefore $\zeta_{\Gamma'} \otimes \zeta_{\Gamma''}$ span $\mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^-$ as $(\Gamma', \Gamma'') \in N^d \times N_d$. Moreover, $\zeta_{\Gamma' \cdot w} \otimes \zeta_{w^{-1}\cdot \Gamma''} = \zeta_{\Gamma'} \otimes \zeta_{\Gamma''}$, so $\dim \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{S}^- \leq |N^d \times_{S_d} N_d| = |M(n,d)|$. □

Now, using Theorem 3.5 we have established a direct combinatorial proof of the following theorem:

Theorem 4.2. *For a field F of characteristic* 0 *or greater than d, and $n \geq d$, the Koszul duality functor D* : S-Mod → S-Mod *is an equivalence of categories.*

4A. ***Strict polynomial functor.*** Friedlander and Suslin [1997] introduced strict polynomial functors in order to establish the finite generation of the full cohomology ring of a finite group scheme. They also showed that the strict polynomial functors of degree d unify modules over the Schur algebras $\mathrm{S}_F(n,d)$ across all n. In this section, we briefly recall the definition of strict polynomial functors and some useful functors on the category of strict polynomial functors.

Following [Krause 2013; van der Kallen 2015], define the *Schur category* (also known as the *divided power category*) $\mathbf{\Gamma}^d_F$ as the category whose objects are finite-dimensional vector spaces over F. For objects V and W, the morphism space is

$$\mathrm{Hom}_{\mathbf{\Gamma}^d_F}(V, W) := \mathrm{Hom}_{S_d}(V^{\otimes d}, W^{\otimes d}).$$

The category $\mathrm{Rep}\,\mathbf{\Gamma}^d_F$ of strict polynomial functors is the functor category

$$\mathrm{Func}(\mathbf{\Gamma}^d_F, F\text{-Mod}).$$

Thus it is an abelian, complete, and cocomplete category.

Example 4.3. Let V and W be objects of $\mathbf{\Gamma}^d_F$. Some examples of strict polynomial functors are:

(4.3.a) The d-th tensor power functor $\otimes^d : \mathbf{\Gamma}^d_F \to F$-Mod. On objects, $\otimes^d(V) = V^{\otimes d}$. On the morphism space, the map

$$\mathrm{Hom}_{S_d}(V^{\otimes d}, W^{\otimes d}) \to \mathrm{Hom}_{S_d}(V^{\otimes d}, W^{\otimes d})$$

is the identity map.

(4.3.b) The d-th divided power functor $\Gamma^d : \mathbf{\Gamma}^d_F \to F$-Mod. On objects $\Gamma^d(V) = (V^{\otimes d})^{S_d}$ and on the morphism space, the map

$$\mathrm{Hom}_{\mathbf{\Gamma}^d_F}(V, W) \to \mathrm{Hom}_{\boldsymbol{S}_d}((V^{\otimes d})^{\boldsymbol{S}_d}, (W^{\otimes d})^{\boldsymbol{S}_d})$$

is given by the restriction.

(4.3.c) Similarly, the d-th exterior power functor $\wedge^d : \mathbf{\Gamma}^d_F \to F$-Mod and the d-th symmetric power functor $\mathrm{Sym}^d : \mathbf{\Gamma}^d_F \to F$-Mod are strict polynomial functors of degree d.

(4.3.d) Let U be an object in $\mathbf{\Gamma}^d_F$. Then define $\boldsymbol{h}_U : \mathbf{\Gamma}^d_F \to F$-Mod as follows:

$$\boldsymbol{h}_U(W) = \mathrm{Hom}_{\mathbf{\Gamma}^d_F}(U, W) = \mathrm{Hom}_{\boldsymbol{S}_d}(U^{\otimes d}, W^{\otimes d}).$$

The functor $\boldsymbol{h}_U \in \mathrm{Rep}\,\mathbf{\Gamma}^d_F$ is called a *representable functor*. The functor $\boldsymbol{h} : U \mapsto \boldsymbol{h}_U$ is the contravariant Yoneda embedding.

(4.3.e) For any object U of $\mathbf{\Gamma}^d_F$ and any $X \in \mathrm{Rep}\,\mathbf{\Gamma}^d_F$, we define a functor $X^U : \mathbf{\Gamma}^d_F \to F$-Mod by

$$X^U(W) = X(\mathrm{Hom}_F(U, W)). \tag{21}$$

When $X = \Gamma^d,\ X^U = h_U$.

Given a strict polynomial functor $X,\ X(F^n)$ inherits the structure of an $\mathrm{S}_F(n, d)$-module. For every nonnegative integer n, we have the evaluation functor $\mathrm{ev}_n : \mathrm{Rep}\,\mathbf{\Gamma}^d_F \to \mathrm{S}_F(n, d)$-Mod as

$$\mathrm{ev}_n(X) = X(F^n) \quad \text{for } X \in \mathrm{Rep}\,\mathbf{\Gamma}^d_F.$$

Theorem 4.4 [Friedlander and Suslin 1997, Theorem 3.2]. *The functor* $\mathrm{ev}_n : \mathrm{Rep}\,\mathbf{\Gamma}^d_F \to \mathrm{S}_F(n, d)$-Mod *is an equivalence of categories whenever* $n \geq d$.

4B. ***Koszul duality of strict polynomial functors:*** Krause [2013] defined an internal tensor product ($\underline{\otimes}$) on the category of strict polynomial functors of a fixed degree d. Kulkarni, Srivastava, and Subrahmanyam [Kulkarni et al. 2018], and independently, Aquilino and Reischuk [2017] showed that this internal tensor product, via the Schur functor, is related to the Kronecker tensor product of representations of the symmetric group $\boldsymbol{S}_d$. Krause used this internal tensor product to introduce Koszul duality as the functor $(\wedge^d \underline{\otimes} -) : \mathrm{Rep}\,\mathbf{\Gamma}^d_F \to \mathrm{Rep}\,\mathbf{\Gamma}^d_F$. We can think about this functor as follows: for the representable functor $\boldsymbol{h}_V \in \mathrm{Rep}\,\mathbf{\Gamma}^d_F$, we have, using the notation of (21),

$$\wedge^d \underline{\otimes}\, \boldsymbol{h}_V = \wedge^{d,V}. \tag{22}$$

For arbitrary $X \in \operatorname{Rep} \mathbf{\Gamma}_F^d$, following [Krause 2013], we exploit a theorem of Mac Lane [1998, III.7, Theorem 1], namely:

$$X = \operatorname*{colim}_{\boldsymbol{h}_V \to X} \boldsymbol{h}_V. \tag{23}$$

Using this we have

$$\wedge^d \underline{\otimes} X = \operatorname*{colim}_{\boldsymbol{h}_V \to X} \wedge^d \underline{\otimes} \boldsymbol{h}_V = \operatorname*{colim}_{\boldsymbol{h}_V \to X} \wedge^{d,V}. \tag{24}$$

In the following theorem, we relate the abstract Koszul duality of Schur algebra with the Koszul duality of strict polynomial functors.

Theorem 4.5. *Consider the functors*

$$\begin{aligned}(\mathrm{S}^- \otimes_S \mathrm{ev}_n(-)):\operatorname{Rep} \mathbf{\Gamma}_F^d &\to \text{S-Mod},\\ \mathrm{ev}_n(\wedge^d \underline{\otimes} -):\operatorname{Rep} \mathbf{\Gamma}_F^d &\to \text{S-Mod}.\end{aligned}$$

Then there exists a natural transformation

$$\eta : (\mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{ev}_n(-)) \to \mathrm{ev}_n(\wedge^d \underline{\otimes} -),$$

which is an isomorphism when $n \geq d$.

Proof. Let $X = \boldsymbol{h}_V$. Then,

$$\begin{aligned}\mathrm{ev}_n(\wedge^d \underline{\otimes} \boldsymbol{h}_V) &= \mathrm{ev}_n(\wedge^{d,V}) && \text{(by (22))}\\ &= \wedge^{d,V}(F^n)\\ &= \wedge^d \operatorname{Hom}_F(V, F^n) && \text{(by (21))}\\ &\simeq \operatorname{Hom}_{S_d}(V^{\otimes d}, (F^n)^{\otimes d} \otimes \mathrm{sgn}).\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{ev}_n(\boldsymbol{h}_V) &= \mathrm{S}^- \otimes_{\mathrm{S}} \boldsymbol{h}_V(F^n)\\ &= \operatorname{Hom}_{S_d}((F^n)^{\otimes d}, (F^n)^{\otimes d} \otimes \mathrm{sgn}) \otimes_{\mathrm{S}} \operatorname{Hom}_{S_d}(V^{\otimes d}, (F^n)^{\otimes d}).\end{aligned}$$

Using these identifications, $\eta_{\boldsymbol{h}_V}(g_1 \otimes g_2) = g_1 \circ g_2$, for $g_1 \in \mathrm{S}^-$ and $g_2 \in \mathrm{ev}_n(\boldsymbol{h}_V)$, defines an S-linear map

$$\eta_{\boldsymbol{h}_V} : \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{ev}_n(\boldsymbol{h}_V) \to \mathrm{ev}_n(\wedge^d \underline{\otimes} \boldsymbol{h}_V). \tag{25}$$

For arbitrary $X \in \operatorname{Rep} \mathbf{\Gamma}_F^d$, we construct η_X using (23):

$$\eta_X = \operatorname*{colim}_{\boldsymbol{h}_V \to X} \eta_{\boldsymbol{h}_V}.$$

From the Yoneda lemma [Mac Lane 1998, page 59], every morphism $\boldsymbol{h}_V \to \boldsymbol{h}_W$ between the representable functors is of the form $\boldsymbol{h}_f$ for a unique morphism

$f \in \mathrm{Hom}_{\boldsymbol{F}_F^d}(W, V)$. The following diagram commutes:

$$\begin{tikzcd} \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{ev}_n(\boldsymbol{h}_V) \arrow[r, "\eta_{\boldsymbol{h}_V}"] \arrow[d, "\mathrm{id}_{\mathrm{S}^-}\otimes \mathrm{ev}_n(\boldsymbol{h}_f)"'] & \mathrm{ev}_n(\wedge^d \underline{\otimes}\, \boldsymbol{h}_V) \arrow[d, "\mathrm{ev}_n(\mathrm{id}_{\wedge^d} \underline{\otimes}\, \boldsymbol{h}_f)"] \\ \mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{ev}_n(\boldsymbol{h}_W) \arrow[r, "\eta_W"] & \mathrm{ev}_n(\wedge^d \underline{\otimes}\, \boldsymbol{h}_V) \end{tikzcd}$$

Taking colimits then gives the naturality of η.

If $n \geq d$, $\boldsymbol{h}_{F^n}$ is a small projective generator of $\mathrm{Rep}\,\boldsymbol{\Gamma}_F^d$, i.e., every object has a presentation by $\boldsymbol{h}_{F^n}$ (see [Krause 2013]). Note that the map $\eta_{\boldsymbol{h}_{F^n}}$ (25) is surjective because $\eta_{\boldsymbol{h}_{F^n}}(f \otimes \mathrm{id}_{\mathrm{S}}) = f$ for $f \in \mathrm{S}^-$ and hence an isomorphism because $\mathrm{ev}_n(\wedge^d \underline{\otimes}\, \boldsymbol{h}_{F^n})$ is isomorphic to S^-. By the construction of η_X, this implies that each η_X is an isomorphism for $X \in \mathrm{Rep}\,\boldsymbol{\Gamma}_F^d$. □

4C. *Derived abstract Koszul duality.* For a finite-dimensional associative algebra A, let $\mathcal{D}(A\text{-Mod})$ be the unbounded derived category of A-Mod. For an (A, A)-bimodule M, the functor $(M \otimes_A -)$ is a right exact functor so the total left derived functor $(M \otimes_A^{\boldsymbol{L}} -) : \mathcal{D}(A\text{-Mod}) \to \mathcal{D}(A\text{-Mod})$ exists. From Happel [1987], we recall necessary and sufficient conditions for the functor $(M \otimes_A^{\boldsymbol{L}} -)$ to be an equivalence of categories.

For each $x \in A$, let $\psi_x \in \mathrm{End}_A(M_A)$ be defined by

$$\psi_x(y) = xy.$$

Taking x to ψ_x gives rise to a homomorphism of algebras:

$$\psi : A \to \mathrm{End}_A(M_A). \tag{26}$$

Theorem 4.6 [Happel 1987]. *For a finite-dimensional algebra A and an (A, A)-bimodule M, the functor $(M \otimes_A^{\boldsymbol{L}} -) : \mathcal{D}(A\text{-Mod}) \to \mathcal{D}(A\text{-Mod})$ is an equivalence of categories if and only if*:

(4.6.a) *The module M_A admits a finite resolution by finitely generated projective right modules over A.*

(4.6.b) *The canonical map $\psi : A \to \mathrm{End}_A(M_A)$ is an isomorphism, and for $i \geq 1$, $\mathrm{Ext}_A^i(M, M) = 0$.*

(4.6.c) *There exists an exact sequence consisting of right A-modules:*

$$0 \to A \to M_1 \to \cdots \to M_l \to 0,$$

where for $1 \leq i \leq l$, M_i is a direct summand of finite direct sum of copies of M.

Theorem 4.7. *Let $A = \mathrm{S}$ and M be the (S, S)-bimodule S^-. Then the map ψ in (26) is an isomorphism if and only if $n \geq d$.*

Remark 4.8. When $n \geq d$, it is known that ψ is an isomorphism, even for q-Schur algebras (see Donkin [1998, page 82]).

Proof. Suppose $n < d$. Consider the labelled bipartite multigraph

$$\Gamma_1 = \begin{array}{cccc} \bullet & \bullet & \cdots & \bullet \\ \overset{|||}{\underset{|||}{{\scriptstyle 12\cdots d}}} & & & \\ \bullet & \bullet & \cdots & \bullet \end{array}$$

Let $\Gamma_2 \in N(n, d)$. Since Γ_2 is a simple bipartite graph, any labelling of Γ_2 which satisfies Definition 2.13(a) requires d balls to place into d distinct boxes out of n. This is not possible as $n < d$. Thus we get $\xi_{\Gamma_1}\zeta_{\Gamma_2} = 0$. Since ζ_Γ for $\Gamma \in N(n, d)$ forms a basis of S^- we get that ξ_{Γ_1} is in the kernel of ψ, and so ψ is not injective.

For the converse, suppose $n \geq d$. That the map ψ is an isomorphism is known from [Donkin 1993, Proposition 3.7], but we give a combinatorial proof here. For $\theta \in \mathrm{End}_{\mathrm{S}}(\mathrm{S}_{\mathrm{S}}^-)$, we denote the coefficient of ζ_{Γ_1} in $\theta(\zeta_{\Gamma_2})$ by $\langle \theta(\zeta_{\Gamma_2}), \zeta_{\Gamma_1}\rangle$.

To see that ψ is injective, note that any element of S is of the form $s = \sum_{\Gamma \in M(n,d)} \alpha_\Gamma \xi_\Gamma$. Now $\psi(s) = 0$ if and only if

$$\sum_{\Gamma \in M(n,d)} \alpha_\Gamma \xi_\Gamma \zeta_{\Gamma'} = 0,$$

for every $\Gamma' \in N(n, d)$. Fix $\Gamma_1 \in M(n, d)$ and let $\Gamma' = D(\Gamma_1)^*$. Then by Lemma 2.35,

$$\alpha_{\Gamma_1} = \Big\langle \sum_{\Gamma \in M(n,d)} \alpha_\Gamma \xi_\Gamma \zeta_{D(\Gamma_1)^*}, \zeta_{U(\Gamma_1)} \Big\rangle.$$

Thus $\alpha_{\Gamma_1} = 0$.

To see that ψ is surjective, we will show that $\dim_F \mathrm{End}_{\mathrm{S}}(\mathrm{S}_{\mathrm{S}}^-) \leq |M(n, d)|$. Firstly, by Lemma 2.33, $\mathrm{S}_{\mathrm{S}}^-$ is generated by $G = \{\zeta_\Gamma \mid \Gamma \in N^d\}$. Recall that N^d denotes the set of graphs in $N(n, d)$ with upper degree sequence $(1^d, 0^{n-d})$. Therefore any $\theta \in \mathrm{End}_{\mathrm{S}}(\mathrm{S}_{\mathrm{S}}^-)$ is determined by its values on this set. Since θ is an S-module homomorphism, $\theta(\zeta_\Gamma)$ again lies in the span of G. Therefore θ is completely determined by the values

$$\{\langle \theta(\zeta_\Gamma), \zeta_{\Gamma'}\rangle \mid \Gamma, \Gamma' \in N^d\}.$$

Moreover, for any $w \in \boldsymbol{S}_d$,

$$\langle \theta(\zeta_\Gamma), \zeta_{\Gamma'}\rangle = \langle \theta(\zeta_{\Gamma \cdot w}), \zeta_{\Gamma' \cdot w}\rangle.$$

Therefore $\dim_F \mathrm{End}_{\mathrm{S}}(\mathrm{S}_{\mathrm{S}}^-) \leq |(N^d \times N^d)/\boldsymbol{S}_d| = |M(n, d)|$. □

Theorem 4.9. *Let F be any field of characteristic different from* 2. *The functor*

$$(\mathrm{S}^- \otimes_{\mathrm{S}}^{\boldsymbol{L}} -) : \mathcal{D}(\text{S-Mod}) \to \mathcal{D}(\text{S-Mod}) \tag{27}$$

is an equivalence of categories if and only if $n \geq d$.

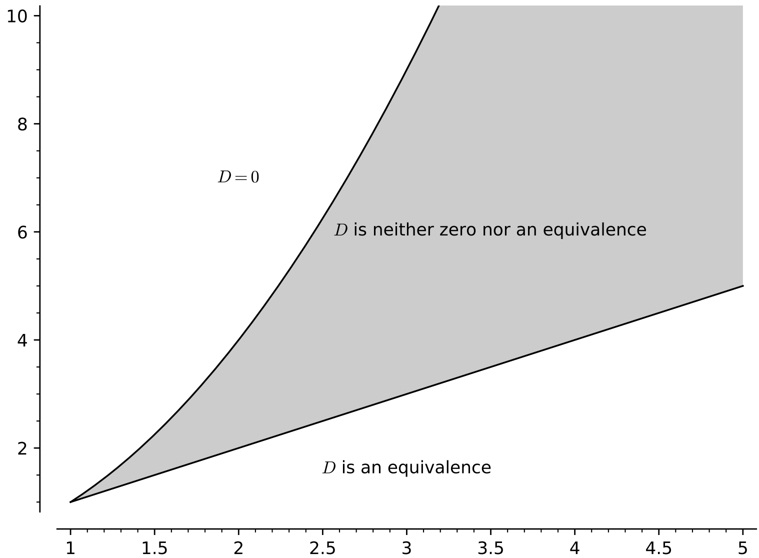


Figure 1. Dependence of derived Koszul duality on n and d.

Proof. If $n \geq d$ then from Theorem 4.5, $\mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{ev}_n(-)$ is isomorphic to $\mathrm{ev}_n(\wedge^d \underline{\otimes} -)$. Since the functor ev_n is exact, the total left derived functors of $\mathrm{S}^- \otimes_{\mathrm{S}} \mathrm{ev}_n(-)$ and $\mathrm{ev}_n(\wedge^d \underline{\otimes} -)$ are isomorphic to $\mathrm{S}^- \otimes^{\boldsymbol{L}}_{\mathrm{S}} \mathrm{ev}_n(-)$ and $\mathrm{ev}_n(\wedge^d \underline{\otimes}^{\boldsymbol{L}} -)$, respectively. From [Krause 2013, Theorem 4.9], $(\wedge^d \underline{\otimes}^{\boldsymbol{L}} -)$ is an equivalence. From Theorem 4.4 $\mathrm{ev}_n(-)$ is an equivalence. Therefore $\mathrm{S}^- \otimes^{\boldsymbol{L}}_{\mathrm{S}} \mathrm{ev}_n(-)$ is an equivalence. So $\mathrm{ev}_n(-)$ being an equivalence forces $(\mathrm{S}^- \otimes^{\boldsymbol{L}}_{\mathrm{S}} -)$ to be an equivalence.

For the converse, suppose $(\mathrm{S}^- \otimes^{\boldsymbol{L}}_{\mathrm{S}} -)$ is an equivalence. Then from Theorem 4.6 the map $\psi : \mathrm{S} \to \mathrm{End}_{\mathrm{S}}(\mathrm{S}^-)$ is an isomorphism. Theorem 4.7 now implies $n \geq d$. □

The behavior of derived Koszul duality for different values of n and d is summarized in Figure 1.

5. Concluding remarks

5A. ***Toward alternating partition algebras.*** The centralizer algebra $\mathrm{End}_{\boldsymbol{S}_n}((F^n)^{\otimes d})$ is a quotient of the partition algebra $\mathrm{P}_d(n)$ of [Jones 1994] and [Martin 1991]. Further restricting the action of $\boldsymbol{S}_n$ to the alternating group $\boldsymbol{A}_n$, we get the *alternating partition algebra* $\mathrm{AP}_d(n) = \mathrm{End}_{\boldsymbol{A}_n}((F^n)^{\otimes d})$, which from the isomorphism (2) decomposes as follows:

$$\mathrm{End}_{\boldsymbol{A}_n}((F^n)^{\otimes d}) = \mathrm{End}_{\boldsymbol{S}_n}((F^n)^{\otimes d}) \oplus \mathrm{Hom}_{\boldsymbol{S}_n}((F^n)^{\otimes d}, (F^n)^{\otimes d} \otimes \mathrm{sgn}) \tag{28}$$

Letting $\mathrm{P}^-_d(n) = \mathrm{Hom}_{\boldsymbol{S}_n}((F^n)^{\otimes d}, (F^n)^{\otimes d} \otimes \mathrm{sgn})$, $\mathrm{P}^-_d(n)$ becomes a $(\mathrm{P}_d(n), \mathrm{P}_d(n))$-bimdoule by inflation. Bloss [2005] showed $\mathrm{P}^-_d(n)$ is nonzero if and only if

$n < 2d + 2$. So we get an abstract Koszul duality

$$(\mathrm{P}_d^-(n) \otimes_{\mathrm{P}_d(n)} -)$$

on the category of modules over the partition algebra $\mathrm{P}_d(n)$ when $n < 2d+2$. Many of the ideas and techniques in this article can be used to study bases, structure constants, and abstract Koszul duality for partition algebras. For $F = \boldsymbol{C}$, the dimensions of simple modules of $\mathrm{AP}_d(n)$ are given combinatorially by Benkart, Halverson, and Harman; see [Benkart et al. 2017].

5B. ***A diagrammatic interpretation of the Schur category.*** The following is one possible way to define the notion of a diagram category in the spirit of [Martin 2008].

Definition 5.1 (diagram category). A category $\mathcal{C}$ is called a *diagram category* if there exists a sequence $\{V_n\}_{n\geq 0}$ of objects which constitute a skeleton of $\mathcal{C}$, and for each pair (m, n) of nonnegative integers, a class of "diagrams" $M(m, n)$, a basis

$$\mathcal{B}_{m,n} = \{\xi_\Gamma \mid \Gamma \in M(m, n)\}$$

of $\mathrm{Hom}_{\mathcal{C}}(V_n, V_m)$, and a combinatorial rule for computing the structure constants $c^{\Gamma}_{\Gamma'\Gamma''}$ that are defined by

$$\xi_{\Gamma'} \circ \xi_{\Gamma''} = \sum_{\Gamma} c^{\Gamma}_{\Gamma'\Gamma''} \xi_\Gamma$$

for $\Gamma' \in M(l, m)$, $\Gamma'' \in M(m, n)$ and $\Gamma \in M(l, n)$.

Remark 5.2. In the examples discussed by Martin [2008], given diagrams Γ' and Γ'', there can exist more than one diagram Γ such that $c^{\Gamma}_{\Gamma'\Gamma''} > 0$. This is not a requirement in the above definition.

Consider the Schur category $\boldsymbol{\Gamma}^d_F$ defined in Section 4A. Take $V_n = F^n$. Define $M_d(m, n)$ to be the set of all bipartite multigraphs with vertex set $[n'] \coprod [m]$ with d edges. Mimicking the discussion in Section 2B, one may endow the Schur category with the structure of a diagram category in the sense of Definition 5.1.

Acknowledgements

Geetha was supported by the Humboldt Foundation, Institute of Algebra and Number Theory, University of Stuttgart, and by SERB MATRICS grant (MTR/2017/000424) from the Department of Science & Technology, India. Prasad was supported by a Swarnajayanti fellowship (DST/SJF/MSA-02/2014-15) from the Department of Science & Technology, India. Srivastava was supported by a national postdoctoral fellowship (PDF/2017/000861) from the Department of Science & Technology, India. The authors thank Steffen König and Upendra Kulkarni for many helpful suggestions.

References

[Aquilino and Reischuk 2017] C. Aquilino and R. Reischuk, "The monoidal structure on strict polynomial functors", *J. Algebra* **485** (2017), 213–229. MR Zbl

[Beilinson et al. 1988] A. A. Beilinson, V. A. Ginsburg, and V. V. Schechtman, "Koszul duality", *J. Geom. Phys.* **5**:3 (1988), 317–350. MR Zbl

[Benkart et al. 2017] G. Benkart, T. Halverson, and N. Harman, "Dimensions of irreducible modules for partition algebras and tensor power multiplicities for symmetric and alternating groups", *J. Algebraic Combin.* **46**:1 (2017), 77–108. MR Zbl

[Bernstein et al. 1978] I. N. Bernstein, I. M. Gelfand, and S. I. Gelfand, "Algebraic vector bundles on $\mathbb{P}^n$ and problems of linear algebra", *Funktsional. Anal. i Prilozhen.* **12**:3 (1978), 66–67. In Russian; translated in *Funct. Anal. Appl.* **12**:3 (1978), 212–214. MR Zbl

[Bloss 2005] M. Bloss, "The partition algebra as a centralizer algebra of the alternating group", *Comm. Algebra* **33**:7 (2005), 2219–2229. MR Zbl

[Bourbaki 1974] N. Bourbaki, *Algebra, I: Chapters 1–3*, Hermann, Paris, 1974. MR Zbl

[Brauer 1937] R. Brauer, "On algebras which are connected with the semisimple continuous groups", *Ann. of Math.* (2) **38**:4 (1937), 857–872. MR Zbl

[Chałupnik 2008] M. Chałupnik, "Koszul duality and extensions of exponential functors", *Adv. Math.* **218**:3 (2008), 969–982. MR Zbl

[Donkin 1993] S. Donkin, "On tilting modules for algebraic groups", *Math. Z.* **212**:1 (1993), 39–60. MR Zbl

[Donkin 1998] S. Donkin, *The q-Schur algebra*, Lond. Math. Soc. Lect. Note Ser. **253**, Cambridge Univ. Press, 1998. MR Zbl

[Friedlander and Suslin 1997] E. M. Friedlander and A. Suslin, "Cohomology of finite group schemes over a field", *Invent. Math.* **127**:2 (1997), 209–270. MR Zbl

[Frobenius 1900] G. Frobenius, "Über die Charaktere der symmetrischen Gruppe", *Sitzungsberichte Akad. Berlin* **1900** (1900), 516–534. Zbl

[Geetha and Prasad 2014] T. Geetha and A. Prasad, "Graphic interpretation of the structure constants of the Schur algebra", lecture notes from Int. Cong. Women Math., 2014. arXiv

[Green 2007] J. A. Green, *Polynomial representations of* GL_n, 2nd augmented ed., Lecture Notes in Math. **830**, Springer, 2007. MR Zbl

[Happel 1987] D. Happel, "On the derived category of a finite-dimensional algebra", *Comment. Math. Helv.* **62**:3 (1987), 339–389. MR Zbl

[Jones 1994] V. F. R. Jones, "The Potts model and the symmetric group", pp. 259–267 in *Subfactors* (Kyuzeso, Japan, 1993), edited by H. Araki et al., World Sci., River Edge, NJ, 1994. MR Zbl

[van der Kallen 2015] W. van der Kallen, "Lectures on bifunctors and finite generation of rational cohomology algebras", pp. 41–65 in *Lectures on functor homology* (Nantes, France, 2012), edited by V. Franjou and A. Touzé, Progr. Math. **311**, Birkhäuser, Cham, 2015. MR Zbl

[Krause 2013] H. Krause, "Koszul, Ringel and Serre duality for strict polynomial functors", *Compos. Math.* **149**:6 (2013), 996–1018. MR Zbl

[Kulkarni et al. 2018] U. Kulkarni, S. Srivastava, and K. V. Subrahmanyam, "Relating tensor structures on representations of general linear and symmetric groups", *Transform. Groups* **23**:2 (2018), 437–461. Zbl

[Mac Lane 1998] S. Mac Lane, *Categories for the working mathematician*, 2nd ed., Grad. Texts in Math. **5**, Springer, 1998. MR Zbl

[Martin 1991] P. Martin, *Potts models and related problems in statistical mechanics*, Ser. Adv. Stat. Mech. **5**, World Sci., Teaneck, NJ, 1991. MR Zbl

[Martin 2008] P. Martin, "On diagram categories, representation theory and statistical mechanics", pp. 99–136 in *Noncommutative rings, group rings, diagram algebras and their applications* (Chennai, India, 2006), edited by S. K. Jain and S. Parvathi, Contemp. Math. **456**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[Méndez 2001] M. A. Méndez, "Directed graphs and the combinatorics of the polynomial representations of $\mathrm{GL}_n(\mathbb{C})$", *Ann. Comb.* **5**:3-4 (2001), 459–478. MR Zbl

[OEIS] N. J. A. Sloane, "Number of Latin squares of order n", in *The on-line encyclopedia of integer sequences*.

[Prasad 2015] A. Prasad, *Representation theory: a combinatorial viewpoint*, Cambridge Stud. Adv. Math. **147**, Cambridge Univ. Press, 2015. MR Zbl

[Priddy 1970] S. B. Priddy, "Koszul resolutions", *Trans. Amer. Math. Soc.* **152** (1970), 39–60. MR Zbl

[Regev 2002] A. Regev, "Double centralizing theorems for the alternating groups", *J. Algebra* **250**:1 (2002), 335–352. MR Zbl

[Ringel 1991] C. M. Ringel, "The category of modules with good filtrations over a quasi-hereditary algebra has almost split sequences", *Math. Z.* **208**:2 (1991), 209–223. MR Zbl

[Schur 1901] I. Schur, *Über eine Klasse von Matrizen, die sich einer gegebenen Matrix zuordnen lassen*, Ph.D. thesis, Friedrich-Wilhelms-Universität Berlin, 1901, Available at https://tinyurl.com/ischurphd. Zbl

[Schur 1927] I. Schur, "Über die rationalen Darstellungen der allgemeinen linearen Gruppe", *Sitzungsberichte Akad. Berlin* **1927** (1927), 58–75. Zbl

[Touzé 2014] A. Touzé, "Bar complexes and extensions of classical exponential functors", *Ann. Inst. Fourier (Grenoble)* **64**:6 (2014), 2563–2637. MR Zbl

[Weyl 1931] H. Weyl, *The theory of groups and quantum mechanics*, Dutton, New York, 1931. Zbl

[Weyl 1939] H. Weyl, *The classical groups: their invariants and representations*, Princeton Univ. Press, 1939. MR Zbl

Received May 15, 2019. Revised December 20, 2019.

THANGAVELU GEETHA
SCHOOL OF MATHEMATICS
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH
THIRUVANANTHAPURAM
INDIA
geetha_curie@yahoo.co.in

AMRITANSHU PRASAD
INSTITUTE OF MATHEMATICAL SCIENCES (HBNI)
CHENNAI
INDIA
amri@imsc.res.in

SHRADDHA SRIVASTAVA
INSTITUTE OF MATHEMATICAL SCIENCES (HBNI)
CHENNAI
INDIA
maths.shraddha@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.185

A POSITIVE MASS THEOREM FOR MANIFOLDS WITH BOUNDARY

SVEN HIRSCH AND PENGZI MIAO

We derive a positive mass theorem for asymptotically flat manifolds with boundary whose mean curvature satisfies a sharp estimate involving the conformal Green's function. The theorem also holds if the conformal Green's function is replaced by the standard Green's function for the Laplacian operator. As an application, we obtain an inequality relating the mass and harmonic functions that generalizes H. Bray's mass-capacity inequality in his proof of the Riemannian Penrose conjecture.

1. Introduction

One of the fundamental results in mathematical relativity is the Riemannian positive mass theorem:

Theorem 1.1 [Schoen and Yau 2017, Theorem 5.3]. *Suppose* (M^n, g) *is an* n*-dimensional*, $n \geq 3$, *complete*, *asymptotically flat manifold with nonnegative scalar curvature. Then the ADM mass of* (M, g) *is nonnegative and is zero if and only if* (M^n, g) *is isometric to the Euclidean space* $(\mathbb{R}^n, \delta)$.

This result was first proven in low dimensions by R. Schoen and S.-T. Yau [1979a; 1979b] using minimal surface methods and later by E. Witten [1981] for spin manifolds exploiting the Bochner–Weitzenböck formula for the Dirac operator (see also [Parker and Taubes 1982; Bartnik 1986]). Recently, Schoen and Yau [2017] extended their arguments to arbitrary dimensions (see [Lohkamp 2015; 2016] for a different approach). The minimal surface method applies to the case in which the manifold has nonempty boundary with nonpositive mean curvature, i.e., the mean curvature vector vanishes or points toward the infinity (see [Schoen and Yau 1979a]). If the boundary has positive mean curvature, the positivity of the mass is a more subtle question as the boundary no longer acts as a barrier for minimal surfaces. For instance, consider a manifold M obtained by cutting out a rotationally symmetric ball in a negative-mass Schwarzschild manifold. In this case, M has

MSC2010: primary 53Z05; secondary 83C99.
Keywords: positive mass theorem with boundary.

negative mass and a zero area singularity (see [Bray and Jauregui 2013]) is shielded by the boundary due to the relatively large positive mean curvature of ∂M.

The spinor method also adapts to the setting of manifolds with boundary. If the boundary mean curvature is nonpositive, positivity of the mass was shown in [Gibbons et al. 1983] (see also [Herzlich 1998]). Refining the spinor analysis, M. Herzlich [1997] relaxed the requirement of nonpositive mean curvature and showed that a condition $H \le 4\sqrt{\pi/|\partial M|}$ in dimension 3 implies the positivity of the mass. For higher-dimensional spin manifolds, Herzlich [2002] proved a similar result.

In this work, we consider asymptotically flat manifolds of dimensions $n \ge 3$, with boundary allowed to have positive mean curvature. We show that if the mean curvature satisfies an upper estimate involving the conformal Green's function, then the manifold has nonnegative mass. More precisely, we have:

Theorem 1.2. *Let (M^n, g) be an n-dimensional, asymptotically flat manifold with nonnegative scalar curvature R, with boundary Σ. Let H be the mean curvature of Σ with respect to the ∞-pointing unit normal ν. Let u be the conformal Green's function given by*

$$\begin{cases} \Delta u - \frac{n-2}{4(n-1)} R u = 0 & \text{in } M, \\ u \to 0 & \text{at } \infty, \\ u = 1 & \text{at } \Sigma. \end{cases}$$

Then the estimate

$$H(x) \le -\frac{n-1}{n-2} \nabla_\nu u(x) \quad \text{for all } x \in \Sigma \tag{1-1}$$

implies positivity of the ADM mass of (M^n, g); moreover, (M, g) has zero mass if and only if (M^n, g) is isometric to $(\mathbb{R}^n, \delta)$ minus a round ball.

By the assumption $R \ge 0$ and the maximum principle, we have an immediate corollary:

Corollary 1.3. *The same result as above holds true if one replaces u with the standard Green's function*

$$\begin{cases} \Delta v = 0 & \text{in } M, \\ v \to 0 & \text{at } \infty, \\ v = 1 & \text{at } \Sigma, \end{cases}$$

and requires

$$H(x) \le -\frac{n-1}{n-2} \nabla_\nu v(x) \quad \text{for all } x \in \Sigma. \tag{1-2}$$

H. Bray [2001] proved the following theorem which plays a key role in his proof of the Riemannian Penrose inequality.

Theorem 1.4. *Let (M^3, g) be a complete, asymptotically flat 3-manifold with nonnegative scalar curvature, with boundary Σ which has zero mean curvature. Let $\varphi(x)$ be a function on (M^3, g) which satisfies*

$$\begin{cases} \Delta\varphi = 0 & \text{in } M, \\ \varphi \to 1 & \text{at } \infty, \\ \varphi = 0 & \text{at } \Sigma. \end{cases}$$

Then

$$\mathfrak{m} \geq \mathcal{C}, \tag{1-3}$$

where $\mathfrak{m}$ is the ADM mass of (M^3, g) and $\mathcal{C} > 0$ is the constant in the asymptotic expansion

$$\varphi(x) = 1 - \frac{\mathcal{C}}{|x|} + o(|x|^{-1}) \quad \text{as } x \to \infty.$$

Moreover, equality in (1-3) holds if and only if (M^3, g) is isometric to a spatial Schwarzschild manifold outside its horizon.

Using Corollary 1.3, we obtain the following generalization of Bray's theorem.

Theorem 1.5. *Let (M^n, g) be an n-dimensional, asymptotically flat manifold, with nonnegative scalar curvature, with boundary Σ. Let H be the mean curvature of Σ with respect to the ∞-pointing unit normal ν. Let $\phi(x)$ be a function on (M, g) which satisfies*

$$\begin{cases} \Delta\phi = 0 & \text{in } M, \\ \phi \to 1 & \text{at } \infty, \\ \phi = c & \text{at } \Sigma, \end{cases}$$

where $c > -1$ is a constant and $c \neq 1$. If

$$\frac{2c}{1-c^2}\frac{\partial\phi}{\partial\nu} \geq \frac{n-2}{n-1}H, \tag{1-4}$$

then

$$\mathfrak{m} \geq C, \tag{1-5}$$

where $\mathfrak{m}$ is the ADM mass of (M^n, g) and C is the constant in

$$\phi = 1 - \frac{C}{|x|^{n-2}} + o(|x|^{2-n}) \quad \text{as } x \to \infty.$$

Moreover, equality in (1-5) holds if and only if (M^n, g) is isometric to the exterior region outside a rotationally symmetric sphere in an n-dimensional spatial Schwarzschild manifold, which is

$$\left(\mathbb{R}^n/\{|x| < r_0\}, \left(1 + \frac{\mathfrak{m}}{2|x|^{n-2}}\right)^{4/(n-2)} \delta_{ij}\right)$$

for some constants $r_0 > 0$.

Remark 1.6. The mass $\mathfrak{m}$ of the Schwarzschild manifold in the rigidity statement of Theorem 1.5 can be arbitrary, and in particular $\mathfrak{m}$ can be negative. Moreover, if $\mathfrak{m} > 0$, r_0 can be arbitrary, meaning that Σ can be either outside the Schwarzschild horizon or inside the Schwarzschild horizon.

Remark 1.7. If $H \leq 0$, one can take $c = 0$. In this case, Theorem 1.5 reduces to Theorem 1.4.

Remark 1.8. An equivalent formulation of Theorem 1.5 shows Corollary 1.3 corresponds to a version of Theorem 1.5 if $c = 1$. See Theorem 5.1 in Section 5 for details.

Remark 1.9. In application, given a harmonic function ϕ that goes to 1 at infinity on an asymptotically flat manifold with nonnegative scalar curvature, one can consider the level sets $\Sigma_c = \phi^{-1}(c)$. As long as condition (1-4) holds at some Σ_c, one will have an estimate of the mass in terms of ϕ.

We now outline the proof of Theorem 1.2. The idea is simply to show the manifold (M, g) in Theorem 1.2 admits a legitimate compact fill-in $(\Omega, \tilde{g})$ so that, if $(\Omega, \tilde{g})$ is glued to (M, g), the resulting manifold satisfies the assumptions of the positive mass theorem. To produce such a fill-in, one conformally deforms (M, g) using the given function u (or v). The idea of conformally deforming an asymptotically flat manifold with boundary to produce a compact piece appeared in the pioneer work of G. Bunting and A. Masood-ul-Alam [1987] and was used by Bray [2001] in his proof of the Riemannian Penrose inequality. A recent application of this idea to obtain capacity estimates was given by C. Mantoulidis, L.-F. Tam and Miao [2019].

Once Theorem 1.2 is proved, Theorem 1.5 follows by considering a conformally deformed metric $\tilde{g} = \left(\frac{1}{2}(1+\phi)\right)^{4/(n-2)} g$ on M and applying Corollary 1.3 to $(M^n, \tilde{g})$.

The rest of this paper is organized as follows. We provide some background material in Section 2. In Section 3, we prove Theorem 1.2 for harmonically flat metrics. The general case of Theorem 1.2 is proved in Section 4. In Section 5, we use Corollary 1.3 to prove Theorem 1.5. In Section 6, we give some examples and discussions.

2. Prerequisites

We start by recalling the definition of a manifold being asymptotically flat (see [Schoen and Yau 2017; Bray and Lee 2009] for instance). Here we only consider manifolds with one end, and the general case can be treated in the same fashion.

Definition 2.1. Let $n \geq 3$. A Riemannian manifold (M^n, g) is asymptotically flat if there is a compact set $K \subset M$ such that M/K is diffeomorphic to $\mathbb{R}^n$ minus a ball and in this coordinate chart, g satisfies

$$g = \delta + \mathcal{O}(|x|^{-\tau}), \quad \partial g = \mathcal{O}(|x|^{-\tau-1}), \quad \partial^2 g = \mathcal{O}(|x|^{-\tau-2}),$$

and $R = \mathcal{O}(|x|^{-q})$, where $\tau > \frac{1}{2}(n-2)$ and $q > n$. Here R is the scalar curvature of g.

On an asymptotically flat (M^n, g), the ADM mass [Arnowitt et al. 1959] m is given by

$$\mathfrak{m} = \frac{1}{2(n-1)\omega_{n-1}} \lim_{r\to\infty} \int_{|x|=r} (g_{ij,j} - g_{jj,i})\nu^i,$$

where ω_{n-1} is the volume of the unit sphere in $\mathbb{R}^n$, and the unit normal ν and volume integral are with respect to the Euclidean metric. The fact that m is a geometric invariant of (M, g) was shown by Bartnik [1986] and by Chruściel [1986] independently.

We next address some regularity questions regarding the positive mass theorem. Originally the result was proved for smooth manifolds, but since then there have been also several singular cases, proven in [Miao 2002; Shi and Tam 2002; Lee 2013; 2018; McFeron and Székelyhidi 2012; Lee and LeFloch 2015; Li and Mantoulidis 2019]. Combined with Theorem 1.1, results and proofs in [Miao 2002; McFeron and Székelyhidi 2012; Shi and Tam 2018] in particular give the following theorems:

Theorem 2.2 [Miao 2002, Theorem 1; McFeron and Székelyhidi 2012, Theorem 2]. *Let (M^n, g) be an asymptotically flat manifold. Suppose $\Sigma^{n-1} \subset M^n$ is a closed hypersurface which divides M into an exterior part M_1 and an interior part M_2. Suppose that the metric g is smooth up to Σ on both sides of Σ and has nonnegative scalar curvature away from Σ. Let H_1 and H_2 be the mean curvature of Σ with respect to the unit normal pointing to infinity in (M_1, g) and (M_2, g), respectively. If*

$$H_1 \leq H_2,$$

then the ADM mass of (M^n, g) is nonnegative and is zero if and only if (M_1, g) is isometric to $(\mathbb{R}^n/\Omega, \delta)$ for a bounded domain $\Omega \subset \mathbb{R}^n$ with smooth boundary and (M_2, g) is isometric to (Ω, δ).

Theorem 2.3 [Shi and Tam 2018, Theorem 7.2]. *Let (M^n, g) be an asymptotically flat manifold such that $g \in W^{1,p}_{\mathrm{loc}}$ for some $p > n$ and is smooth with nonnegative scalar curvature away from a point q. Then $m \geq 0$ and $m = 0$ only if M^n is diffeomorphic to $\mathbb{R}^n$ and g is flat away from q.*

3. The harmonically flat case

As in [Bray 2001], we first prove Theorem 1.2 for the case $g = \xi^{4/(n-2)}\delta$ near ∞, where ξ is a Euclidean harmonic function. This property is often known as (M^n, g) being harmonically flat (see [Bray 2001]).

Proof. Let $(\tilde{M}^n, \tilde{g})$ be the one point compactification of $(M^n, u^{4/(n-2)}g)$ where u is the conformal Green's function from Theorem 1.2. Note that in particular

$\tilde{g}|_\Sigma = g|_\Sigma$ and by the maximum principle $u > 0$. Also, we are able to extend $\tilde{g}$ smoothly to the point at ∞ as in Bray's proof of Theorem 1.4, also see the detailed exposition in Section 2 of [Mantoulidis et al. 2019]. This argument is based on the removable singularity theorem for harmonic functions and thus crucially requires the manifold to be harmonically flat. Note that in this case u is harmonic near ∞. In the general setting this conformal blow-down may produce a singular metric which must be smoothed which is performed in Section 4 following [Shi and Tam 2018] and [Mantoulidis et al. 2019].

The conformally transformed manifold $\tilde{M}$ has zero scalar curvature $\tilde{R}$ due to the well-known formula

$$\tilde{R} = u^{-4/(n-2)} R - \frac{4(n-1)}{n-2} u^{-4/(n-2)-1} \Delta u. \tag{3-6}$$

Moreover, the mean curvature transforms under conformal transformation and change of unit normal via

$$\tilde{H} = -u^{-2/(n-2)} H - \frac{2(n-1)}{n-2} u^{-n/(n-2)} \nabla_\nu u. \tag{3-7}$$

Note that we have to change the sign of the unit normal of $\tilde{\Sigma}$ so that it corresponds with the normal of Σ after gluing $\tilde{M}$ into M as carried out below. In combination with $u|_\Sigma = 1$, (3-7) leads to

$$\tilde{H} = -H - \frac{2(n-1)}{n-2} \nabla_\nu u. \tag{3-8}$$

Next we glue M and $\tilde{M}$ together along Σ to obtain a manifold $\hat{M}$ with corner Σ. More precisely, we define $(\hat{M}, \hat{g})$ by $\hat{M} = (M \sqcup \tilde{M})/_{\Sigma\sim\tilde{\Sigma}}$ and $\hat{g}|_M = g,\ \hat{g}|_{\tilde{M}} = \tilde{g}$. Then $\hat{M}$ is asymptotically flat, has nonnegative scalar curvature and no boundary. Furthermore, due to (1-1) and (3-8), we have

$$\tilde{H} = -H - \frac{2(n-1)}{n-2} \nabla_\nu u \geq H.$$

Thus all the assumptions of the positive mass theorem with corners, Theorem 2.2, are satisfied and we obtain $\hat{\mathfrak{m}} \geq 0$. Since we did not change the asymptotic behavior of the metric during the construction of $(\hat{M}, \hat{g})$ we have $\mathfrak{m} = \hat{\mathfrak{m}} \geq 0$ as desired.

Next, we apply the rigidity part of the positive mass theorem with corners [Miao 2002; McFeron and Székelyhidi 2012] to deduce that in the case of $\mathfrak{m} = 0$, $(\hat{M}, \hat{g})$ is isometric to $(\mathbb{R}^n, \delta)$ where δ is the standard metric. Thus (M^n, g) is isometric to $(R^n/\Omega, \delta)$, and $(\tilde{M}^n, \tilde{g})$ is isometric to (Ω, δ) where $\Omega \subset \mathbb{R}^n$ is some smooth, open set. Our goal is to show that $\Omega = B_\rho(x_0)$, $\rho > 0$, where $x_0 \in \tilde{M}$ is the compactified point from ∞. We begin by observing that the function

$$\tilde{u} := \frac{1}{u}$$

is harmonic on $\tilde{M}$ with $\tilde{u}|_\Sigma = 1$ and

$$\nabla_{\tilde{\nu}}\tilde{u} = -\nabla_\nu \tilde{u} = \nabla_\nu u.$$

Again, note how the unit normal of Σ changes sign after the conformal transformation. Therefore, we can define $\hat{u} \in C^1(\hat{M})$ by

$$\begin{cases} \hat{u}(x) = u(x) & \text{for } x \in M, \\ \hat{u}(x) = \tilde{u}(x) & \text{for } x \in \tilde{M}. \end{cases}$$

Since $\hat{u}$ is harmonic outside Σ and C^1 across Σ, it follows from standard PDE arguments that $\hat{u}$ is also harmonic across Σ. Thus, the uniqueness of solutions to the Laplace equation implies

$$\hat{u}(x) = \frac{\rho^{n-2}}{|x - x_0|^{n-2}}$$

for some $\rho > 0$ which results in $\Sigma = S_\rho(x_0)$ as desired. Alternatively, we could also directly apply Theorem 5.1 of [Mantoulidis et al. 2019] to $(\tilde{M}, \tilde{g}, \tilde{u})$. □

4. The general case

In this section, we prove Theorem 1.2 in the general case. Let u be the conformal Green's function from Theorem 1.2, i.e., $\mathcal{L}u = 0$ where

$$\mathcal{L} = \Delta - \frac{n-2}{4(n-1)} R$$

is the conformal Laplacian with respect to g. As in Appendix A of [Mantoulidis et al. 2019] we have the following asymptotic behavior of u at ∞:

Lemma 4.1. *There exists a constant $\mathcal{D}$ such that*

$$u(x) = \frac{\mathcal{D}}{|x|^{n-2}} + \mathcal{O}(|x|^{-\gamma}),$$

where $\gamma := \min(q-2, n+\tau-2, n-1) > n-2$.

Proof. Let $\mathbb{R}^n / B_{R_1}(0)$ be the asymptotically flat end. We start by constructing a barrier to the conformal Green's function. For this purpose let $\psi = ar^{-n+2} - r^{-n+2-\epsilon}$ with $a, \epsilon > 0$. Then as in the proof of Lemma A2 in [Mantoulidis et al. 2019],

$$\mathcal{L}\psi = (-n+2-\epsilon)(-n+1-\epsilon)r^{-n-\epsilon} + \mathcal{O}(|x|^{-\kappa}),$$

where $\kappa := \min(n+\tau, q) > n$. Choosing $\epsilon < \min(\tau, n-q)$ there is an $R_2 \geq R_1$ independent of a such that $\Delta\psi \leq 0$ on $\mathbb{R}^3 / B_{R_2}(0)$. Moreover, we set $a \gg 1$ such that $\psi > u$ on the complement of $\mathbb{R}^n / B_{R_2}(0)$ in M. Hence, by the maximum principle, $u \leq \psi \leq ar^{-n+2}$ and similarly $u \geq -ar^{-n+2}$.

By the Schauder estimates and Theorem 6.2 in [Gilbarg and Trudinger 1998],

$$|u| \leq C|x|^{-\kappa}, \quad |\nabla u| \leq C|x|^{-\kappa-1}, \quad |\nabla^2 u| \leq C|x|^{-\kappa-2},$$

where C denotes some constant which may change in the following computations from line to line. We extend u smoothly onto $\mathbb{R}^n$ such that $u \leq C$ in $B_{R_0}(0)$ with $R_0 \leq R_1$. Let $f = \hat{\Delta} u = \mathcal{O}(|x|^{-\kappa})$, where $\hat{\Delta}$ is the Euclidean Laplacian, and define

$$w(x) := -\frac{1}{n(n-2)\alpha_n} \int_{\mathbb{R}^n} \frac{1}{|x-y|^{n-2}} f(y)\, dy,$$

where α_n is the volume of the unit ball in $\mathbb{R}^n$. This is well defined due to the decay properties of $f = \hat{\Delta} u$. Also, observe that $\hat{\Delta} w = f$. Next, as in the proof of Lemma A2 in Appendix A of [Mantoulidis et al. 2019], denoting $|x| = r$, we compute

$$\int_{B_r(x)/B_r(0)} \frac{1}{|x-y|^{n-2}} f(y)\, dy \leq Cr^{-\kappa} \int_{B_r(x)/B_r(0)} \frac{1}{|x-y|^{n-2}}\, dy \leq Cr^{-\kappa+2} \tag{4-9}$$

and

$$\begin{aligned} \int_{\mathbb{R}^n/(B_r(0)\cup B_r(x))} \frac{1}{|x-y|^{n-2}} f(y)\, dyi &\leq Cr^{-n+2} \int_{\mathbb{R}^3/(B_r(0)\cup B_r(x))} |y|^{-n-\tau} \\ &\leq Cr^{-\kappa+2}. \end{aligned} \tag{4-10}$$

Further, we additionally split the integral over $B_r(0)$ into $B_r(0)/B_{R_0}(0) \cup B_{R_0}(0)$:

$$\begin{aligned} &\int_{B_r(0)/B_{R_0}(0)} \left(\frac{1}{|x-y|^{n-2}} - \frac{1}{|x|^{n-2}} \right) f(y)\, dy \\ &\qquad = \int_{B_{R_0}(0)/B_\rho(0)} \frac{|x|^{2n-4} - |x-y|^{2n-4}}{|x|^{n-2}|x-y|^{n-2}(|x|^{n-2} + |x-y|^{n-2})} f(y)\, dy \\ &\qquad \leq \sum_{j=1}^{2n-4} \int_{B_r(0)/B_{R_0}(0)} C \frac{|y|^{-\kappa+j}}{|x|^{n-2+j}} \leq Cr^{-n+1}. \end{aligned} \tag{4-11}$$

Similarly, we have

$$\begin{aligned} &\int_{B_{R_0}(0)} \left(\frac{1}{|x-y|^{n-2}} - \frac{1}{|x|^{n-2}} \right) f(y)\, dy \\ &\qquad = \int_{B_{R_0}(0)} \frac{|x|^{2n-4} - |x-y|^{2n-4}}{|x|^{n-2}|x-y|^{n-2}(|x|^{n-2} + |x-y|^{n-2})} f(y)\, dy \\ &\qquad \leq \sum_{j=1}^{2n-4} \int_{B_{R_0}(0)} C \frac{|y|^k}{|x|^{n-2+k}} \leq Cr^{-n+1}. \end{aligned} \tag{4-12}$$

Lastly, we note

$$\frac{1}{|x|^{n-2}}\int_{\mathbb{R}^n/B_r(0)} f(y)dy \le Cr^{-\kappa+2}. \tag{4-13}$$

Combining equations (4-9)–(4-13), we obtain

$$-n(n-2)\alpha_n w = \frac{\int_{\mathbb{R}^n}\hat{\Delta}u}{r^{n-2}} + \mathcal{O}(|x|^{-\min(q-2,n+\tau-2,n-1)}).$$

In particular, by the maximum principle, we have $w = u$. As in [Mantoulidis et al. 2019], Schauder theory [Gilbarg and Trudinger 1998, Theorem 6.3] gives the higher order estimates. □

Without loss of generality we may assume $R_1 = 1$, i.e., the asymptotically flat end is diffeomorphic to $\mathbb{R}^n/B_1(0)$. Next, we introduce the Kelvin transform $\mathcal{K}$ and prove the following elementary lemma:

Lemma 4.2. *Let (M^n, g) be asymptotically flat, i.e., $g_{ij} = \delta_{ij} + \sigma_{ij}$ where $\sigma_{ij} = \mathcal{O}_2(|x|^{-\tau}) = \mathcal{O}_2(|y|^{\tau})$ and $\mathcal{K}(x) := x/|x|^2$. Then for $y = \mathcal{K}(x)$ and $h_{ij} = g(\partial_i y, \partial_j y)$, we have*

$$h_{ij} = |y|^{-4}\delta_{ij} + \mathcal{O}(|y|^{\tau-4}).$$

Moreover, we have

$$\partial_k h_{ij} = -4|y|^{-5}\partial_k|y|\delta_{ij} + \mathcal{O}(|y|^{\tau-5}).$$

Proof. We compute

$$\begin{aligned} h_{ij} &= g\left(\frac{\partial_i x}{|x|^2} - 2\frac{\langle \partial_i x, x\rangle x}{|x|^4}, \frac{\partial_j x}{|x|^2} - 2\frac{\langle \partial_j x, x\rangle x}{|x|^4}\right) \\ &= |y|^{-4}[\delta_{ij} + \sigma_{ij} + |y|^{-4}(4y_i y_j y_k y_l \sigma^{kl} - 2|y|^2(y_i y^k \sigma_{jk} + y_j y^k \sigma_{ik}))] \\ &= |y|^{-4}\delta_{ij} + \mathcal{O}(|y|^{\tau-4}). \end{aligned}$$

The second statement follows analogously. □

Next, we study the metric $\tilde{g} = u^{4/(n-2)}g$. A priori $\tilde{g}$ is defined on $\mathbb{R}^n/B_1(0)$ but by inverting the coordinates via the Kelvin transform $\mathcal{K}$ we may view $\tilde{g}$ as metric on $B_1(0)$. Thereby the point ∞ of the one point compactification corresponds to the origin under the Kelvin transform and we wish to extend $\tilde{g}$ there. This is done in the following lemma which is based on Lemma 6.1 in [Mantoulidis et al. 2019]. Moreover, we also obtain some regularity for $\tilde{g}$ which is necessary in order to perform the smoothing procedure below.

Lemma 4.3. *The metric $\tilde{g}$ extends continuously across the origin to a $W^{1,p}$ metric for some $p > n$.*

Proof. We compute the coordinates $y = \mathcal{K}(x)$ for h_{ij},

$$u^{4/(n-2)} h_{ij} = \mathcal{D}\delta_{ij} + \mathcal{O}(|y|^{\gamma+2-n}).$$

Hence h is continuous in the origin. Next, we compute in a similar fashion

$$\begin{aligned}\partial_k(u^{4/(n-2)} h_{ij}) &= (4\mathcal{D}|y|^{n-7} + \mathcal{O}(|y|^{\gamma-5}))(\mathcal{D}|y|^{n-2} + \mathcal{O}(|y|^{\gamma}))^{(6-n)/(n-2)}\partial_k|y| \\ &\qquad + (-4\mathcal{D}|y|^{n-7} + \mathcal{O}(|y|^{\gamma-5})(\mathcal{D}|y|^{n-2} + \mathcal{O}(|y|^{\gamma}))^{(6-n)/(n-2)}\partial_k|y| \\ &= \mathcal{O}(|y|^{\gamma-n+1}).\end{aligned}$$

Thus $u^4 h_{ij} \in W^{1,p}(B_1(0))$ for $p = n/(\min(q-n-1, \tau-1)) > n$. □

Now we are in a position to verify the following proposition, which is known to people who are familiar with the work in [Miao 2002; Shi and Tam 2002; 2018; Lee 2013; McFeron and Székelyhidi 2012; Li and Mantoulidis 2019; Mantoulidis et al. 2019].

Proposition 4.4. *Suppose (M^n, g) has corner singularity along a hypersurface Σ with $H_1 \le H_2$ as in Theorem 2.2. Also, suppose there is a point singularity $q \in M_2$ where g is in $W^{1,p}$ for some $p > n$ near q. If g has nonnegative scalar curvature away from Σ and $\{q\}$, then $\mathfrak{m} \ge 0$ with equality if and only if (M_1, g) is isometric to $(\mathbb{R}^n/\Omega, \delta)$ for a bounded domain Ω with smooth boundary.*

Proof. Due to the assumption $H_1 \le H_2$ at Σ, we may exactly follow Proposition 3.1 in [Miao 2002] to approximate g by a family of smooth metrics g_δ such that $g_\delta(x) = g(x)$ for $\mathrm{dist}(\Sigma, x) \ge \delta$ and the scalar curvature R_δ satisfies

$$R_\delta(x) \ge -C \tag{4-14}$$

for $\mathrm{dist}(\Sigma, x) < \delta$. Equation (4-14) in particular shows that the integral of the negative part of the scalar curvature can be made arbitrarily small during the approximation process.

Near the point singularity q, we approximate g as in [Shi and Tam 2018, Lemma 4.1] and [Mantoulidis et al. 2019, Lemma 3.6] to obtain smooth metrics $\{g_\epsilon\}$ so that $g_\epsilon = g$ outside $B_\epsilon(\{q\})$, and $\|g_\epsilon\|_{W^{1,p}(B_\epsilon(\{q\}))} \le C$ where C is independent of ϵ. By Lemma 3.7 in [Mantoulidis et al. 2019], the uniform $W^{1,p}$ bound on g_ϵ implies that the integral of the negative part of R_{g_ϵ} over $B_\epsilon(\{q\})$ becomes arbitrarily small. (We note that Lemma 3.7 in [Mantoulidis et al. 2019] is stated in a slightly more general version and for our purpose the result already follows from the estimate on the term $\boldsymbol{I}_B$ in Lemma 3.7 in [Mantoulidis et al. 2019].)

Since both approximations g_ϵ and g_δ are local, we can perform them simultaneously to approximate g by a smooth metric with $\int_M R^-$ becoming arbitrarily small. Here R^- denotes the negative part of the scalar curvature.

Now we proceed as usual and conformally transform M such that $R \geq 0$ everywhere; see [Schoen and Yau 1979b] and Section 4 in [Miao 2002]. Thereby the mass converges as in [Miao 2002] and we may deduce that the mass is nonnegative by the standard positive mass theorem.

If the mass is zero, we need to apply the argument from [McFeron and Székelyhidi 2012]. This is due to [McFeron and Székelyhidi 2012] relying on Ricci flow which has the advantage that the mass stays constant during the smoothing process so we can apply the rigidity statement of the positive mass theorem to the flow solution with initial data g (see also Section 7 of [Shi and Tam 2018]). This shows the rigidity of the proposition. □

Proof of Theorem 1.2. Due to Lemma 4.3, the conformally filled-in manifold $(\hat{M}, \hat{g})$ constructed in Section 3 satisfies the conditions of Proposition 4.4. Hence Theorem 1.2 follows from Proposition 4.4 in the same way that the harmonically flat case is proven in Section 3. □

5. Application

In this section, we prove Theorem 1.5, which generalizes Bray's result in Theorem 1.4.

Proof of Theorem 1.5. By the maximum principle, $\phi > -1$ on M. Define

$$w = \frac{2}{1+\phi} \quad \text{and} \quad \tilde{g} = w^{-4/(n-2)} g.$$

$(M^n, \tilde{g})$ is asymptotically flat with nonnegative scalar curvature. The fact $\Delta w^{-1} = 0$ implies

$$\tilde{\Delta} w = 0.$$

Moreover, $w \to 1$ at ∞ and $w = 2/(1+c)$ at Σ. Next, define

$$v = \frac{1+c}{1-c}(w-1).$$

Then

$$\tilde{\Delta} v = 0, \quad v \to 0 \text{ at } \infty, \quad \text{and} \quad v = 1 \text{ at } \Sigma.$$

Let $\tilde{H}$ be the mean curvature of Σ in $(M, \tilde{g})$ with respect to the ∞-pointing unit normal $\tilde{\nu}$. By (3-7), it follows that

$$\tilde{H} = w^{2/(n-2)} \left[H + \frac{2(n-1)}{(n-2)} w \nabla_\nu w^{-1} \right],$$
$$\tilde{\nu} = w^{2/(n-2)} \nu,$$

and

$$\nabla_{\tilde{\nu}} v = w^{2/(n-2)} \nabla_\nu v.$$

Hence, at Σ, by (1-4),

$$\begin{aligned}\tilde{H}+\frac{n-1}{n-2}\nabla_\nu v &= w^{2/(n-2)}\left[H+\frac{(n-1)}{(n-2)}(2w\nabla_\nu w^{-1}+\nabla_\nu v)\right]\\ &= w^{2/(n-2)}\left[H-\frac{(n-1)}{(n-2)}\frac{2c}{(1-c^2)}\nabla_\nu v\right]\\ &\le 0.\end{aligned}$$

Thus, by Corollary 1.3,

$$\tilde{\mathfrak{m}} \ge 0, \tag{5-15}$$

where $\tilde{\mathfrak{m}}$ is the ADM mass of $(M^n, \tilde{g})$. Since by the definition of mass $\tilde{\mathfrak{m}}$ and $\mathfrak{m}$ are related by $\tilde{\mathfrak{m}} = \mathfrak{m} - C$, we conclude from (5-15) that

$$\mathfrak{m} \ge C.$$

If $\mathfrak{m} = C$, then $\tilde{\mathfrak{m}} = 0$. By Corollary 1.3, $(M^n, \tilde{g})$ is isometric to $(\mathbb{R}^n/\{|x| < r_0\}, \delta)$ for some constant $r_0 > 0$. In this case, w is a Euclidean harmonic function that goes to 1 at ∞ and equals a positive constant $2/(1+c)$ at $\Sigma = \{|x| = r_0\}$. Hence,

$$w = 1 + \frac{\mathfrak{m}}{2|x|^{n-2}},$$

where $\mathfrak{m}$ is a constant satisfying $\mathfrak{m} = (1-c)/(1+c)2r_0^{n-2}$. □

It is worth pointing out the following equivalent form of Theorem 1.5.

Theorem 5.1. *Let (M^n, g) be an n-dimensional, asymptotically flat manifold, with nonnegative scalar curvature, with boundary Σ. Let $\varphi(x)$ be a function on (M, g) which satisfies*

$$\begin{cases}\Delta\varphi = 0 & \text{in } M,\\ \varphi \to 1 & \text{at } \infty,\\ \varphi = 0 & \text{at } \Sigma.\end{cases}$$

Let H be the mean curvature of Σ in (M^n, g) with respect to the ∞-pointing normal ν. If there is a constant $c > -1$ such that

$$\frac{2c}{1+c}\nabla_\nu\varphi \ge \frac{n-2}{n-1}H, \tag{5-16}$$

then

$$\mathfrak{m} \ge (1-c)\mathcal{C}, \tag{5-17}$$

where $\mathfrak{m}$ is the ADM mass of (M^n, g) and $\mathcal{C}$ is the constant in

$$\varphi = 1 - \frac{\mathcal{C}}{|x|^{n-2}} + o(|x|^{2-n}) \quad \text{as } x \to \infty.$$

Moreover, equality in (5-17) holds if and only if (M^n, g) is isometric to an n-dimensional spatial Schwarzschild manifold outside a rotationally symmetric sphere, that is,

$$\left(\mathbb{R}^n/\{|x|<r_0\}, \left(1+\frac{\mathfrak{m}}{2|x|^{n-2}}\right)^{4/(n-2)}\delta_{ij}\right) \quad \textit{for some constants } r_0>0.$$

Proof. If $c \neq 1$, the theorem follows directly from Theorem 1.5 by letting $\varphi = 1/(1-c)(\phi - c)$. If $c = 1$, the theorem reduces to Corollary 1.3. ☐

Remark 5.2. For a constant $c > -1$ satisfying (5-16) to exist, one needs to have

$$2\nabla_\nu \varphi > \frac{n-2}{n-1} H.$$

If $n = 3$, this coincides with the condition in [Mantoulidis et al. 2019, Theorem 1.5].

6. Examples and discussions

Example 6.1 (boundary with $H \leq 0$). Every manifold (M^n, g) whose boundary ∂M has nonpositive mean curvature satisfies condition (1-1) and thus has positive mass m. More precisely, $m \geq \mathcal{C} > 0$ by Theorem 1.5.

Example 6.2 (regions in Schwarzschild manifold). Let

$$(M^n, g) = \left(\mathbb{R}^n/B_{r_0}(0), \left(1+\frac{m}{2r^{n-2}}\right)^{4/(n-2)}\delta\right)$$

be a spacelike slice in Schwarzschild spacetime of mass m. Note, m is allowed to be negative. We begin by computing the mean curvature of the boundary $S_{r_0}(0)$ using (3-7):

$$H = (2^{n/(n-2)} r_0^{n-1} - 2^{2/(n-2)} m r_0) \frac{n-1}{(2r_0^{n-2}+m)^{n/(n-2)}}.$$

Next, we observe that the Green's function is given by

$$u(r) = \frac{2r_0^{n-2}+m}{2r^{n-2}+m}. \tag{6-18}$$

Combining this with the observation $|\partial_r|_g = (1+m/(2r^{n-2}))^{2/(n-2)}$, at S_{r_0} we have

$$-\frac{n-1}{n-2}\nabla_\nu u = 2^{n/(n-2)} r_0^{n-1} \frac{n-1}{(2r_0^{n-2}+m)^{n/(n-2)}}.$$

Thus, as predicted by the rigidity statement of Theorem 1.5, we have

$$H = -\frac{2c(n-1)}{(1+c)(n-2)}\nabla_\nu u$$

for

$$c = \frac{2r_0^{n-2}}{2r_0^{n-2}+m}.$$

Moreover, $2c/(c+1)$ can approach 1 while maintaining negative mass which shows that Theorem 1.2 is sharp in this sense.

Example 6.3 (conformal minimal boundary). Suppose that $g = \phi^{4/(n-2)}\tilde{g}$ such that Σ is a minimal surface with respect to $\tilde{g}$ where ϕ solves

$$\begin{cases} \Delta\phi = 0 & \text{in } M, \\ \phi \to 2 & \text{at } \infty, \\ \phi = 1 & \text{at } \Sigma. \end{cases}$$

Observe that by (3-6), $\tilde{R}$ also has nonnegative scalar curvature. Note that $v\phi$ is a harmonic function on $(M^n, \tilde{g})$ satisfying $(\phi v)|_\Sigma = 1$ and $(\phi v)(x) \to 0$ for $|x| \to \infty$. Precisely the same properties hold true for the function $(2-\phi)$ and thus, by the uniqueness property for harmonic functions, we deduce $\phi v = 2 - \phi$. Therefore, at Σ we have

$$\nabla_\nu v = -\frac{2}{v^2}\nabla_\nu \phi = -2\nabla_\nu \phi.$$

Next, using (3-7) we compute

$$H = \tilde{H}\phi^{-2/(n-2)} + \frac{2(n-1)}{n-2}\phi^{-n/(n-2)}\nabla_\nu\phi = \frac{2(n-1)}{n-2}\nabla_\nu\phi = -\frac{n-1}{n-2}\nabla_\nu v.$$

Hence condition (1-2) is satisfied with equality and we have $m \geq 0$. Conversely, suppose we have $H \equiv -(n-1)/(n-2)\nabla_\nu v$. Then, we define $\phi := 2/(1+v)$ and $\tilde{g} = \phi^{-4/(n-2)}g$. Applying (3-7) again, we obtain

$$\tilde{H} = H\phi^{2/(n-2)} + \frac{2(n-1)}{n-2}\phi^{n/(n+2)}\nabla_\nu(\phi^{-1}) = H + \frac{n-1}{n-2}\nabla_\nu v = 0.$$

Thus Σ is a minimal surface with respect to $\tilde{g}$. Indeed, this calculation shows that condition (1-2) holds if and only if $\tilde{H} \leq 0$ under the above conformal transformation. In particular, this suggests Bray's result and proof of Theorem 1.4 in [Bray 2001] can also be applied to derive Corollary 1.3 and Theorem 1.5 in the harmonically flat case.

Next, we want to relate Theorem 1.2, Theorem 1.5 and Example 6.3 to the Riemannian Penrose conjecture. This conjecture was shown by G. Huisken and T. Ilmanen [2001] (one black hole) and by Bray [2001] (multiple black holes). The latter proof has been extended up to dimension 7 by Bray and D. Lee [2009].

One of the key steps in Bray's proof is to prove the mass-capacity estimate, Theorem 1.4. There an asymptotically flat manifold with nonnegative scalar curvature is first reflected across its horizon, then conformally transformed via the Green's function so the positive mass theorem may be applied. In view of the positive mass theorem for manifolds with boundary, the reflection argument can now be dropped. More precisely, the asymptotically flat manifold can be directly conformally transformed via the Green's function as presented in Example 6.3.

As it turns out, Theorem 1.4 is the reason the proof of the Penrose inequality breaks down in higher dimensions as discovered by Bray. This is due to minimal hypersurfaces being no longer regular which then in particular complicates the reflection argument. Therefore, showing a positive mass theorem for manifolds with singular boundary may be considered a strategy for proving the higher-dimensional Penrose inequality.

We also recall that Herzlich [1997; 2002] showed a positive mass theorem for asymptotically flat manifolds with boundary ∂M whose mean curvature satisfies an upper bound which depends on the area of ∂M in dimension 3 and depends on the Yamabe invariant $\mathcal{Y}$ of ∂M in higher dimensions. More precisely, the latter case assumes

$$H \leq |\Sigma|^{-1/(n-1)}\sqrt{\frac{n-1}{n-2}\mathcal{Y}(\Sigma)}.$$

The proofs in [Herzlich 1997; 2002] make use of estimates of the first eigenvalue of the Dirac operator by C. Bär [1992] and by O. Hijazi [1986; 1991]. Results in [Herzlich 1997; 2002] have been extended to the asymptotically hyperbolic setting by Hijazi, S. Montiel and S. Raulot [Hijazi et al. 2015]. It is plausible that Theorem 1.2 or Corollary 1.3 may have an analogue in the asymptotically hyperbolic setting.

Acknowledgements

This work was initiated when Hirsch was visiting the University of Miami (UM) in October 2018. He is grateful for the hospitality of the department of mathematics at UM. He also wants to thank H. Bray for many encouraging discussions and interest in this work. Miao's research is partially supported by Simons Foundation Collaboration Grant for Mathematicians #585168.

References

[Arnowitt et al. 1959] R. Arnowitt, S. Deser, and C. W. Misner, "Dynamical structure and definition of energy in general relativity", *Phys. Rev.* (2) **116** (1959), 1322–1330. MR Zbl

[Bär 1992] C. Bär, "Lower eigenvalue estimates for Dirac operators", *Math. Ann.* **293**:1 (1992), 39–46. MR Zbl

[Bartnik 1986] R. Bartnik, "The mass of an asymptotically flat manifold", *Comm. Pure Appl. Math.* **39**:5 (1986), 661–693. MR Zbl

[Bray 2001] H. L. Bray, "Proof of the Riemannian Penrose inequality using the positive mass theorem", *J. Differential Geom.* **59**:2 (2001), 177–267. MR Zbl

[Bray and Jauregui 2013] H. L. Bray and J. L. Jauregui, "A geometric theory of zero area singularities in general relativity", *Asian J. Math.* **17**:3 (2013), 525–559. MR Zbl

[Bray and Lee 2009] H. L. Bray and D. A. Lee, "On the Riemannian Penrose inequality in dimensions less than eight", *Duke Math. J.* **148**:1 (2009), 81–106. MR Zbl

[Bunting and Masood-ul Alam 1987] G. L. Bunting and A. K. M. Masood-ul Alam, "Nonexistence of multiple black holes in asymptotically Euclidean static vacuum space-time", *Gen. Relativity Gravitation* **19**:2 (1987), 147–154. MR Zbl

[Chruściel 1986] P. Chruściel, "Boundary conditions at spatial infinity from a Hamiltonian point of view", pp. 49–59 in *Topological properties and global structure of space-time* (Erice, 1985), edited by P. G. Bergmann and V. De Sabbata, NATO Adv. Sci. Inst. Ser. B Phys. **138**, Plenum, New York, 1986. MR Zbl

[Gibbons et al. 1983] G. W. Gibbons, S. W. Hawking, G. T. Horowitz, and M. J. Perry, "Positive mass theorems for black holes", *Comm. Math. Phys.* **88**:3 (1983), 295–308. MR

[Gilbarg and Trudinger 1998] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, revised 2nd ed., Springer, 1998. MR Zbl

[Herzlich 1997] M. Herzlich, "A Penrose-like inequality for the mass of Riemannian asymptotically flat manifolds", *Comm. Math. Phys.* **188**:1 (1997), 121–133. MR Zbl

[Herzlich 1998] M. Herzlich, "The positive mass theorem for black holes revisited", *J. Geom. Phys.* **26**:1-2 (1998), 97–111. MR Zbl

[Herzlich 2002] M. Herzlich, "Minimal surfaces, the Dirac operator and the Penrose inequality", pp. 9–16 in *Séminaire de Théorie Spectrale et Géométrie*, Sémin. Théor. Spectr. Géom. **20**, Univ. Grenoble I, Saint-Martin-d'Hères, 2002. MR Zbl

[Hijazi 1986] O. Hijazi, "A conformal lower bound for the smallest eigenvalue of the Dirac operator and Killing spinors", *Comm. Math. Phys.* **104**:1 (1986), 151–162. MR Zbl

[Hijazi 1991] O. Hijazi, "Première valeur propre de l'opérateur de Dirac et nombre de Yamabe", *C. R. Acad. Sci. Paris Sér. I Math.* **313**:12 (1991), 865–868. MR Zbl

[Hijazi et al. 2015] O. Hijazi, S. Montiel, and S. Raulot, "A positive mass theorem for asymptotically hyperbolic manifolds with inner boundary", *Internat. J. Math.* **26**:12 (2015), art. id. 1550101, 17 pp. MR Zbl

[Huisken and Ilmanen 2001] G. Huisken and T. Ilmanen, "The inverse mean curvature flow and the Riemannian Penrose inequality", *J. Differential Geom.* **59**:3 (2001), 353–437. MR Zbl

[Lee 2013] D. A. Lee, "A positive mass theorem for Lipschitz metrics with small singular sets", *Proc. Amer. Math. Soc.* **141**:11 (2013), 3997–4004. MR Zbl

[Lee and LeFloch 2015] D. A. Lee and P. G. LeFloch, "The positive mass theorem for manifolds with distributional curvature", *Comm. Math. Phys.* **339**:1 (2015), 99–120. MR Zbl

[Li and Mantoulidis 2019] C. Li and C. Mantoulidis, "Positive scalar curvature with skeleton singularities", *Math. Ann.* **374**:1-2 (2019), 99–131. MR Zbl

[Lohkamp 2015] J. Lohkamp, "Skin structures in scalar curvature geometry", preprint, 2015. arXiv

[Lohkamp 2016] J. Lohkamp, "The higher dimensional positive mass theorem I", preprint, 2016. arXiv

[Mantoulidis et al. 2019] C. Mantoulidis, P. Miao, and L.-F. Tam, "Capacity, quasi-local mass, and singular fill-ins", *J. Reine Angew. Math.* (online publication December 2019).

[McFeron and Székelyhidi 2012] D. McFeron and G. Székelyhidi, "On the positive mass theorem for manifolds with corners", *Comm. Math. Phys.* **313**:2 (2012), 425–443. MR Zbl

[Miao 2002] P. Miao, "Positive mass theorem on manifolds admitting corners along a hypersurface", *Adv. Theor. Math. Phys.* **6**:6 (2002), 1163–1182. MR

[Parker and Taubes 1982] T. Parker and C. H. Taubes, "On Witten's proof of the positive energy theorem", *Comm. Math. Phys.* **84**:2 (1982), 223–238. MR Zbl

[Schoen and Yau 1979a] R. Schoen and S. T. Yau, "On the proof of the positive mass conjecture in general relativity", *Comm. Math. Phys.* **65**:1 (1979), 45–76. MR Zbl

[Schoen and Yau 1979b] R. M. Schoen and S. T. Yau, "Complete manifolds with nonnegative scalar curvature and the positive action conjecture in general relativity", *Proc. Nat. Acad. Sci. U.S.A.* **76**:3 (1979), 1024–1025. MR Zbl

[Schoen and Yau 2017] R. Schoen and S.-T. Yau, "Positive scalar curvature and minimal hypersurface singularities", preprint, 2017. arXiv

[Shi and Tam 2002] Y. Shi and L.-F. Tam, "Positive mass theorem and the boundary behaviors of compact manifolds with nonnegative scalar curvature", *J. Differential Geom.* **62**:1 (2002), 79–125. MR Zbl

[Shi and Tam 2018] Y. Shi and L.-F. Tam, "Scalar curvature and singular metrics", *Pacific J. Math.* **293**:2 (2018), 427–470. MR Zbl

[Witten 1981] E. Witten, "A new proof of the positive energy theorem", *Comm. Math. Phys.* **80**:3 (1981), 381–402. MR Zbl

Received March 19, 2019. Revised July 6, 2019.

SVEN HIRSCH
DUKE UNIVERSITY
DURHAM, NC
UNITED STATES
sven.hirsch@duke.edu

PENGZI MIAO
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MIAMI
CORAL GABLES, FL
UNITED STATES
pengzim@math.miami.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.203

CIRCLE PATTERNS ON SURFACES OF FINITE TOPOLOGICAL TYPE REVISITED

YUE-PING JIANG, QIANGHUA LUO AND ZE ZHOU

Using topological degree theory, we obtain the existence of circle patterns with prescribed combinatorial type and obtuse exterior intersection angles on surfaces of finite topological type. As consequences, several generalizations of circle pattern theorem are obtained.

1. Introduction

To study the geometry and topology of 3-manifolds, Thurston [1979, Chapter 13] states the circle pattern theorem regarding the existence and uniqueness of circle patterns on higher genus surfaces with prescribed combinatorial type and nonobtuse exterior intersection angles. In recent years, the circle patterns have played significant roles in various problems in combinatorics [Schramm 1992; 1993; Liu and Zhou 2016], discrete and computational geometry [Stephenson 2005; Dai et al. 2008], deformation theory [He and Liu 2013; Huang and Liu 2017], minimal surfaces [Bobenko et al. 2006], and many others.

Assume that $\mathcal{T}$ is a triangulation of a compact oriented surface S (possibly with boundary) with a constant curvature metric μ. A circle pattern $\mathcal{P}$ on (S, μ) is a collection of oriented circles. We say $\mathcal{P}$ is $\mathcal{T}$-type if there exists a geodesic triangulation $\mathcal{T}(\mu)$ of (S, μ) such that $\mathcal{T}(\mu)$ is isotopic to $\mathcal{T}$ and the vertices of $\mathcal{T}(\mu)$ coincide with the centers of the circles in $\mathcal{P}$. Let V, E, F be the sets of vertices, edges and triangles of $\mathcal{T}$. In this paper, we will focus on these $\mathcal{T}$-type circle patterns $\mathcal{P} = \{C_v : v \in V\}$ such that C_u and C_w intersect with each other whenever there exists an edge between u and w. In this situation we have the exterior intersection angle $\Theta(e) \in [0, \pi)$ for every $e \in E$. We refer to Stephenson's monograph [2005] for more details on circle patterns.

Given a function $\Theta : E \to [0, \pi)$ defined on the edge set of $\mathcal{T}$, let us consider the following question: Does there exist a $\mathcal{T}$-type circle pattern whose exterior intersection angle function is given by Θ? In addition, if it does exist, to what extent

Jiang is supported by NSFC (No.11631010). Luo and Zhou are supported by NSFC (No.11601141, No.11631010).
MSC2010: 52C25, 52C26.
Keywords: circle patterns, topological degree.

is the circle pattern unique? Under the condition that S has empty boundary and $0 \le \Theta \le \pi/2$, a celebrated answer to this question is the following circle pattern theorem.

Theorem 1.1 [Thurston 1979, Chapter 13]. *Let $\mathcal{T}$ be a triangulation of an oriented closed surface S of genus $g > 0$. Suppose that $\Theta : E \to [0, \pi/2]$ is a function satisfying the following conditions*:

(i) *If e_1, e_2, e_3 form a null-homotopic closed curve in S, and if $\sum_{l=1}^{3} \Theta(e_l) \ge \pi$, then these three edges form the boundary of a triangle of $\mathcal{T}$.*

(ii) *If e_1, e_2, e_3, e_4 form a null-homotopic closed curve in S, then $\sum_{l=1}^{4} \Theta(e_l) < 2\pi$.*

Then there exists a constant curvature (*equal to 0 for $g = 1$ and equal to -1 for $g > 1$*) *metric μ on S such that (S, μ) supports a $\mathcal{T}$-type circle pattern $\mathcal{P}$ with the exterior intersection angles given by Θ. Moreover, the pair $(\mu, \mathcal{P})$ is unique up to isometries if $g > 1$, and up to similarities if $g = 1$.*

Recently, Ge, Hua and Zhou [Ge et al. $\ge$ 2020] obtained a more general result, which considered the case that S had possibly nonempty boundary and Θ was possibly larger than $\pi/2$. Before stating the result, let us introduce some terminologies. Assume that S is of the topological type (g, h), i.e., S is of genus g and has boundary consisting of h disjoint simple closed curves. A closed (not necessarily simple) curve γ in S is said to be *pseudo-Jordan* if $S \setminus \gamma$ contains at least one simply connected component $\mathbb{K}_\gamma$ such that $\partial \mathbb{K}_\gamma = \gamma$, and an *enclosing set* $A_\gamma \subset V$ of γ consists of all vertices covered by $\mathbb{K}_\gamma$. Similarly, an arc λ in S is said to be *semi-pseudo-Jordan* if there exists an open arc $\tilde{\lambda} \subset \partial S$ such that $\lambda \cup \tilde{\lambda}$ is a pseudo-Jordan curve in S, and a *semi-enclosing set* W_λ of λ consists of all vertices covered by $\mathbb{K}_{\lambda \cup \tilde{\lambda}} \cup \tilde{\lambda}$. What is more, we say a pseudo-Jordan (or semi-pseudo-Jordan) curve or arc is *nonvacant* if one of its enclosing sets (or semi-enclosing sets) is nonempty. Denote by V_∂ and E_∂ the sets of boundary vertices and edges of $\mathcal{T}$. Ge, Hua and Zhou's result [$\ge$ 2020] was stated as follows.

Theorem 1.2. *Let $\mathcal{T}$ be a triangulation of a surface S of topological type (g, h). Suppose that $\Theta : E \to [0, \pi)$ and $\varphi : V_\partial \to [0, \pi)$ are two functions satisfying the following properties*:

(C1) *If the edges e_1, e_2, e_3 form the boundary of a triangle of $\mathcal{T}$, then*

$$I(e_1) + I(e_2)I(e_3) \ge 0, \quad I(e_2) + I(e_1)I(e_3) \ge 0, \quad I(e_3) + I(e_1)I(e_2) \ge 0,$$

where $I(e_\eta) = \cos\Theta(e_\eta)$ for $\eta = 1, 2, 3$.

(C2) *The Gauss–Bonnet inequality* (*resp. equality*) *holds*:

$$\sum_{v \in V_\partial} \varphi(v) \; > \; 2\pi\chi(S) \quad \Bigg(\textit{resp.} \; \sum_{v \in V_\partial} \varphi(v) = 2\pi\chi(S)\Bigg).$$

(C3) *When the edges $e_1, e_2, \ldots, e_s$ form a nonvacant pseudo-Jordan curve in S, then $\sum_{l=1}^{s} \Theta(e_l) < (s-2)\pi$.*

(C4) *When the edges $e_1, \ldots, e_s$ form a nonvacant semi-pseudo-Jordan arc λ in S, for any nonempty semi-enclosing set W_λ of λ, then*

$$\sum_{v \in W_\lambda \cap V_\partial} \varphi(v) + \sum_{l=1}^{s} (\pi - \Theta(e_l)) > \pi.$$

Then there exists a hyperbolic (resp. Euclidean) metric μ on S so that (S, μ) has boundary consisting of h disjoint simple piecewise-geodesic closed curves with turning angles assigned by φ and supports a $\mathcal{T}$-type circle pattern $\mathcal{P}$ whose exterior intersection angle function is Θ. Moreover, the pair $(\mathcal{P}, \mu)$ is unique up to isometries (resp. similarities).

In this paper, we shall prove the following result which generalizes the existence part of the above theorem.

Theorem 1.3. *Let $\mathcal{T}$ be a triangulation of a surface S of topological type (g, h). Suppose that $\Theta : E \to [0, \pi)$ and $\varphi : V_\partial \to [0, \pi)$ are two functions such that* (C2), (C3), (C4) *and the following condition are satisfied.*

(R1) *If the edges e_1, e_2, and e_3 form the boundary of a triangle of $\mathcal{T}$, and if $\sum_{i=1}^{3} \Theta(e_i) > \pi$, then $\Theta(e_1)+\Theta(e_2) < \pi+\Theta(e_3)$, $\Theta(e_2)+\Theta(e_3) < \pi+\Theta(e_1)$, $\Theta(e_3)+\Theta(e_1) < \pi+\Theta(e_2)$.*

Then there exists a hyperbolic (resp. Euclidean) metric on S so that (S, μ) has boundary consisting of h disjoint simple piecewise-geodesic closed curves with turning angles assigned by φ and supports a $\mathcal{T}$-type circle pattern $\mathcal{P}$ whose exterior intersection angle function is Θ.

Remark 1.4. The condition (C1) implies (R1). On the other hand, we can easily find examples showing that the converse does not hold. That means our result is strictly stronger than the existence part of Theorem 1.2.

Let M and N be two manifolds with the same dimension and let f be a map between M and N. Let $\Lambda \subset M$ be a relatively compact open subset. Suppose that $y \in N \setminus f(\partial\Lambda)$ is a regular value of f and $\Lambda \cap f^{-1}(y) = \{x_1, x_2, \ldots, x_m\}$. The topological degree of y and f in Λ is

$$\deg(f, \Lambda, y) := \sum_{l=1}^{m} \operatorname{sgn}(f, x_l),$$

where $\operatorname{sgn}(f, x_i) = 1$ in the case that the tangent map d_{x_i} preserves orientation and $\operatorname{sgn}(f, x_i) = -1$ in the other case. In general, if $\deg(f, \Lambda, y) \neq 0$, then $\Lambda \cap f^{-1}(y) \neq \varnothing$. More details can be found in [Guillemin and Pollack 1974; Hirsch 1976; Milnor 1965] and the Appendix.

For each $r = (\rho_1, \dots, \rho_{|V|}) \in \mathbb{R}_+^{|V|}$, Thurston's famous construction produces a corresponding curvature $\mathrm{Th}(\Theta, r)$. What is more, there is a curvature $K[\varphi]$ produced by the function φ above (see Section 3 for the details). The strategy to prove our result above is to show that there is a relatively compact open subset Ω of $\mathbb{R}_+^{|V|}$ such that

$$\deg(\mathrm{Th}(\Theta, \cdot), \Omega, K[\varphi]) \neq 0.$$

The paper is organized as follows: In Section 2, we introduce basic properties of three-circle configurations. In Section 3, using topological degree theory, we prove Theorem 1.3. In Section 4, we establish some results on planar circle patterns. The last section is an appendix regarding some results from manifold theory.

Throughout this paper, we denote by $|\cdot|$ the cardinality of a set and denote by $\chi(\cdot)$ the Euler characteristic of a topological space. We denote by $d_E(\cdot,\cdot)$ and $d_H(\cdot,\cdot)$ the distances in Euclidean and hyperbolic geometries, respectively.

2. Preliminaries

In this section we collect several lemmas on three-circle configurations. It should be pointed out that the nonobtuse versions of these results have been established by Thurston [1979].

Lemma 2.1. *Suppose that $\Theta_1, \Theta_2, \Theta_3 \in [0, \pi)$ are three angles satisfying*

$$\Theta_1 + \Theta_2 + \Theta_3 \leq \pi,$$

or

$$\Theta_1+\Theta_2+\Theta_3 > \pi, \quad \Theta_1+\Theta_2 < \pi+\Theta_3, \quad \Theta_2+\Theta_3 < \pi+\Theta_1, \quad \Theta_3+\Theta_1 < \pi+\Theta_2.$$

Then for any three positive numbers ρ_1, ρ_2, ρ_3, there exists a configuration of three intersecting circles in both Euclidean and hyperbolic geometries, unique up to isometry, having radii ρ_1, ρ_2, ρ_3 and meeting in exterior intersection angles $\Theta_1, \Theta_2, \Theta_3$.

Proof. In Euclidean geometry, set

$$l_1 = \sqrt{\rho_2^2 + \rho_3^2 + 2\rho_2\rho_3 \cos\Theta_1},$$
$$l_2 = \sqrt{\rho_1^2 + \rho_3^2 + 2\rho_1\rho_3 \cos\Theta_2},$$
$$l_3 = \sqrt{\rho_1^2 + \rho_2^2 + 2\rho_1\rho_2 \cos\Theta_3}.$$

To prove the lemma, it suffices to verify that l_1, l_2, l_3 satisfy the triangle inequalities. We divide the proof into the following cases:

(E1) $\Theta_1+\Theta_2+\Theta_3 \leq \pi$. To show the triangle equalities is equivalent to checking

$$\begin{aligned}(2\text{-}1)\quad & \sin^2\Theta_1\rho_2^2\rho_3^2+\sin^2\Theta_2\rho_1^2\rho_3^2+\sin^2\Theta_3\rho_1^2\rho_2^2 \\ &\qquad +2\xi_1\rho_1^2\rho_2\rho_3+2\xi_2\rho_1\rho_2^2\rho_3+2\xi_3\rho_1\rho_2\rho_3^2 > 0,\end{aligned}$$

where $\xi_i=\cos\Theta_i+\cos\Theta_j\cos\Theta_k$ for distinct subscripts $i,\ j,\ k\in\{1,2,3\}$. Note that

$$\xi_1=\cos\Theta_1+\cos(\Theta_2+\Theta_3)+\sin\Theta_2\sin\Theta_3 \geq 2\cos\frac{\Theta_1+\Theta_2+\Theta_3}{2}\cos\frac{\Theta_1-\Theta_2-\Theta_3}{2}\geq 0.$$

Similarly, $\xi_2\geq 0$ and $\xi_3\geq 0$. Thus (2-1) holds.

(E2) $\Theta_1+\Theta_2+\Theta_3>\pi,\ \Theta_1+\Theta_2<\pi+\Theta_3,\ \Theta_1+\Theta_3<\pi+\Theta_2,\ \Theta_3+\Theta_2<\pi+\Theta_1$. In the complex plan $\mathbb{C}$ we find the three points

$$z_1=\rho_1\exp(i\pi-i\Theta_2),\quad z_2=\rho_2\exp(i\Theta_1-i\pi),\quad z_3=\rho_3.$$

It is easy to see

$$\begin{aligned} d_E(z_2,z_3)&=\sqrt{\rho_2^2+\rho_3^2+2\rho_2\rho_3\cos\Theta_1}=l_1,\\ d_E(z_1,z_3)&=\sqrt{\rho_1^2+\rho_3^2+2\rho_1\rho_3\cos\Theta_2}=l_2,\\ d_E(z_1,z_2)&=\sqrt{\rho_1^2+\rho_2^2-2\rho_1\rho_2\cos(\Theta_1+\Theta_2)}>\sqrt{\rho_1^2+\rho_2^2+2\rho_1\rho_2\cos\Theta_3}=l_3.\end{aligned}$$

As a result, we obtain

$$l_1+l_2=d_E(z_2,z_3)+d_E(z_1,z_3)\geq d_E(z_1,z_2)>l_3.$$

Similarly, $l_1+l_3>l_2$ and $l_2+l_3>l_1$. We thus prove the triangle equalities.

In hyperbolic geometry, the lemma has been established in [Zhou 2017] by computation. For completeness, here we give an independent proof. Set

$$\begin{aligned} l_1&=\cosh^{-1}(\cosh\rho_2\cosh\rho_3+\sinh\rho_2\sinh\rho_3\cos\Theta_1),\\ l_2&=\cosh^{-1}(\cosh\rho_1\cosh\rho_3+\sinh\rho_1\sinh\rho_3\cos\Theta_2),\\ l_3&=\cosh^{-1}(\cosh\rho_1\cosh\rho_2+\sinh\rho_1\sinh\rho_2\cos\Theta_3).\end{aligned}$$

It suffices to verify that l_1,l_2,l_3 satisfy the triangle inequalities. As before, we divide the proof into the following cases.

(H1) $\Theta_1+\Theta_2+\Theta_3\leq\pi$. To obtain the proof, we need to show

$$\begin{aligned}(2\text{-}2)\quad & \sin^2\Theta_1x_2^2x_3^2+\sin^2\Theta_2x_1^2x_3^2+\sin^2\Theta_3x_1^2x_2^2\\ &\qquad +(2+2\cos\Theta_1\cos\Theta_2\cos\Theta_3)x_1^2x_2^2x_3^2+2\xi_1a_2a_3x_1^2x_2x_3\\ &\qquad\qquad +2\xi_2a_1a_3x_1x_2^2x_3+2\xi_3a_1a_2x_1x_2x_3^2 > 0,\end{aligned}$$

where $a_\eta=\cosh\rho_\eta$ and $x_\eta=\sinh\rho_\eta$ for $\eta=1,2,3$. Under the assumption that $\Theta_1+\Theta_2+\Theta_3\leq\pi$, we have shown that $\xi_1\geq 0,\ \xi_2\geq 0$ and $\xi_3\geq 0$. Hence (2-2) holds.

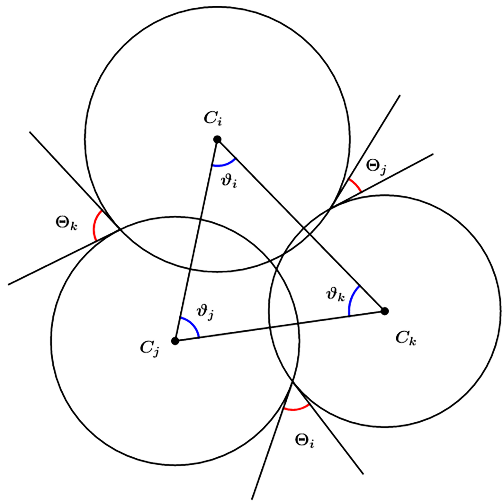


Figure 1. A configuration of three circles.

(H2) $\Theta_1+\Theta_2+\Theta_3>\pi,\ \Theta_1+\Theta_2<\pi+\Theta_3,\ \Theta_1+\Theta_3<\pi+\Theta_2,\ \Theta_2+\Theta_3<\pi+\Theta_1.$
In the hyperbolic disk $\mathbb{D}$ we find the three points

$$w_1=\tanh\frac{\rho_1}{2}\exp(i\pi-i\Theta_2),\quad w_2=\tanh\frac{\rho_2}{2}\exp(i\Theta_1-i\pi),\quad w_3=\tanh\frac{\rho_3}{2}.$$

Similarly, it is easy to see

$$\begin{aligned}
d_H(w_2,w_3)&=\cosh^{-1}(\cosh\rho_2\cosh\rho_3+\sinh\rho_2\sinh\rho_3\cos\Theta_1)=l_1,\\
d_H(w_1,w_3)&=\cosh^{-1}(\cosh\rho_1\cosh\rho_3+\sinh\rho_1\sinh\rho_3\cos\Theta_2)=l_2,\\
d_H(w_1,w_2)&=\cosh^{-1}(\cosh\rho_1\cosh\rho_2-\sinh\rho_1\sinh\rho_2\cos(\Theta_1+\Theta_2))\\
&\qquad>\cosh^{-1}(\cosh\rho_1\cosh\rho_2+\sinh\rho_1\sinh\rho_2\cos\Theta_3)=l_3.
\end{aligned}$$

As a result, we obtain

$$l_1+l_2=d_H(w_2,w_3)+d_H(w_1,w_3)\geq d_H(w_1,w_2)>l_3.$$

Similarly, $l_1+l_3>l_2$ and $l_2+l_3>l_1$. We thus finish the proof. □

As in Figure 1, let ϑ_1, ϑ_2, ϑ_3 denote the corresponding inner angles of the triangle formed by the centers of the three circles. The following two lemmas were obtained in [Zhou 2017]:

Lemma 2.2. *Assume that* $a,b,c\in(0,+\infty]$. *In both Euclidean and hyperbolic geometries*, *we have*

$$\lim_{(\rho_1,\rho_2,\rho_3)\to(0,a,b)}\vartheta_1=\pi-\Theta_1,$$

$$\lim_{(\rho_1,\rho_2,\rho_3)\to(0,0,c)}\vartheta_1+\vartheta_2=\pi,$$

$$\lim_{(\rho_1,\rho_2,\rho_3)\to(0,0,0)}\vartheta_1+\vartheta_2+\vartheta_3=\pi.$$

Lemma 2.3. *Fixing* $\Theta_1, \Theta_2, \Theta_3 \in [0, \pi)$, *in hyperbolic geometry, for any* $\epsilon > 0$, *there exists a positive number* L *such that for any positive* ρ_1, ρ_2, ρ_3 *satisfying* $\rho_1 > L$,

$$\vartheta_1 < \epsilon.$$

As a consequence, we have

$$\lim_{\rho_1 \to +\infty} \vartheta_1 = 0.$$

Remark 2.4. Let $\Omega \subset [0, \pi)^3$ be the set of all vectors $(\Theta_1, \Theta_2, \Theta_3)$ satisfying $\Theta_1 + \Theta_2 + \Theta_3 \leq \pi$ or $\Theta_1 + \Theta_2 + \Theta_3 > \pi$, $\Theta_1 + \Theta_2 < \pi + \Theta_3$, $\Theta_2 + \Theta_3 < \pi + \Theta_1$, $\Theta_1 + \Theta_3 < \pi + \Theta_2$. Lemma 2.3 still holds if $(\Theta_1, \Theta_2, \Theta_3)$ varies in a compact subset of Ω.

3. Proof of Theorem 1.3

Thurston's construction. Recall that $\mathcal{T}$ is a triangulation of S with the vertex set V, the edge set E and the face set F and Θ is a weight associated to the edge set. Assume that $V = \{v_1, \ldots, v_{|V|}\}$. For each radius vector $r = (\rho_1, \ldots, \rho_{|V|}) \in \mathbb{R}_+^{|V|}$, it assigns each vertex v_i a positive number ρ_i. Using the two systems of data Θ and r, we obtain a hyperbolic (or Euclidean) cone metric on the surface S as follows.

Each triangle $\triangle(v_i v_j v_k)$ of $\mathcal{T}$ is associated with a hyperbolic (or Euclidean) triangle formed by centers of three hyperbolic circles of radii ρ_i, ρ_j, ρ_k with exterior intersection angles $\Theta([v_i, v_j]), \Theta([v_j, v_k]), \Theta([v_k, v_i])$. Precisely, let l_{ij}, l_{jk}, l_{ki} be the three lengths of this triangle. Then

$$l_{ij} = \cosh^{-1}(\cosh \rho_i \cosh \rho_j + \sinh \rho_i \sinh \rho_j \cos \Theta([v_i, v_j]))$$

in hyperbolic background geometry, or

$$l_{ij} = \sqrt{\rho_i^2 + \rho_j^2 + 2\rho_i \rho_j \cos \Theta([v_i, v_j])}$$

in Euclidean background geometry. Similarly, we can get the formulas for l_{jk} and l_{ki}. Under the condition (R1), Lemma 2.1 implies that l_{ij}, l_{jk}, l_{ki} satisfy the triangle inequalities. Thus the above procedure works well.

Gluing all these hyperbolic triangles along the common edges produces a hyperbolic cone metric on S with possible cone singularities at vertices of $\mathcal{T}$. For $i = 1, 2, \ldots, |V|$, set K_i as the discrete curvature at v_i. More precisely,

$$K_i := \begin{cases} 2\pi - \sigma(v_i) & \text{if } v_i \in V \setminus V_\partial, \\ \pi - \sigma(v_i) & \text{if } v_i \in V_\partial, \end{cases}$$

where $\sigma(v_i)$ is the sum of inner angles at v_i for all hyperbolic (or Euclidean) triangles incident to v_i. Clearly, $K_1, \ldots, K_{|V|}$ are smooth functions of Θ and r.

This gives rise to the map $\mathrm{Th}(\Theta, \cdot)$:

$$\mathbb{R}_+^{|V|} \to \mathbb{R}^{|V|},$$
$$(\rho_1, \rho_2, \dots, \rho_{|V|}) \mapsto (K_1, K_2, \dots, K_{|V|}).$$

Using the prescribed function $\varphi : V_\partial \to [0, \pi)$ in Theorem 1.3, we define a vector

$$K[\varphi] = (k_1, \dots, k_{|V|}),$$

where

$$k_i = \begin{cases} 0 & \text{if } v_i \in V \setminus V_\partial, \\ \varphi(v_i) & \text{if } v_i \in V_\partial. \end{cases}$$

The main goal now is to show that $K[\varphi]$ belongs to the image of the map $\mathrm{Th}(\Theta, \cdot)$. Actually, if there exists a radius vector

$$r^* = (\rho_1^*, \dots, \rho_{|V|}^*)$$

such that $\mathrm{Th}(\Theta, r^*) = K[\varphi]$, then it produces a hyperbolic (or Euclidean) metric on S. Drawing the circle centered at v_i with the radius ρ_i^* for each v_i, we obtain the desired circle pattern realizing $(\mathcal{T}, \Theta)$.

Topological degree. To this end, we will use the topological degree theory. The readers can refer to the Appendix for some basic knowledge on this subject.

First let us deal with the hyperbolic geometry case. We need to find a relatively compact open set $\Omega \subset \mathbb{R}_+^{|V|}$ and determine the topological degree

$$\deg(\mathrm{Th}(\Theta, \cdot), \Omega, K[\varphi]).$$

Once we show

$$\deg(\mathrm{Th}(\Theta, \cdot), \Omega, K[\varphi]) \neq 0,$$

Theorem A.9 then implies that $K[\varphi]$ is in the image of $\mathrm{Th}(\Theta, \cdot)$. Now let us compute $\deg(\mathrm{Th}(\Theta, \cdot), \Omega, K[\varphi])$ via homotopy method. For any $t \in [0, 1]$, let $\Theta(t) = t\Theta$. Because $\Theta(t)$ satisfies the condition (R1), Lemma 2.1 implies that $\mathrm{Th}(\Theta(t), \cdot)$ is well defined. Setting $\mathrm{Th}_t = \mathrm{Th}(\Theta(t), \cdot)$, gives a continuous homotopy from $\mathrm{Th}(\Theta, \cdot)$ to Th_0, where $\mathrm{Th}_0 = \mathrm{Th}(0, \cdot)$.

Lemma 3.1. *In hyperbolic geometry, there exists a relatively compact open set $\Omega \subset \mathbb{R}_+^{|V|}$ such that*

$$\mathrm{Th}_t(\mathbb{R}_+^{|V|} \setminus \Omega) \subset \mathbb{R}^{|V|}\{K[\varphi]\} \quad \textit{for all } t \in [0, 1].$$

Proof. Let us exhaust $\mathbb{R}_+^{|V|}$ by an increasing sequence of relatively compact open sets $\{\Omega_n\}$. Assuming that the lemma is not true, then for each n, we obtain $t_n \in [0, 1]$ and

$$r_n = (\rho_{1,n}, \dots, \rho_{|V|,n}) \in \mathbb{R}_+^{|V|} \setminus \Omega_n$$

such that

$$\mathrm{Th}(\Theta(t_n), r_n) = K[\varphi].$$

Namely,

$$K_i(\Theta(t_n), r_n) = \begin{cases} 0 & \text{if } v_i \in V \setminus V_\partial, \\ \varphi(v_i) & \text{if } v_i \in V_\partial. \end{cases} \tag{3-1}$$

Because $\{\Omega_n\}$ exhausts $\mathbb{R}_+^{|V|}$, there exists $i_0 \in \{1, 2, \ldots, |V|\}$ such that

$$\rho_{i_0,n} \to +\infty \quad \text{or} \quad \rho_{i_0,n} \to 0.$$

In the first case, it follows from Lemma 2.3 and Remark 2.4, that

$$K_{i_0}(\Theta(t_n), r_n) \to \begin{cases} 2\pi & \text{if } v_{i_0} \in V \setminus V_\partial, \\ \pi & \text{if } v_{i_0} \in V_\partial. \end{cases}$$

which contradicts (3-1).

It remains to consider the second case. Without loss of generality, assume that $\{t_n\}$ converges to a number $t_* \in [0, 1]$. Otherwise, we pick up a convergent subsequence of $\{t_n\}$. Let $A \subset V$ be the set of vertices v_i for which $\rho_{i,n} \to 0$. Then A is a nonempty subset of V. We denote by $G(A)$ the union of η-cells ($\eta = 0, 1, 2$) of $\mathcal{T}$ that have at least one vertex in A, and denote by $\mathrm{Lk}(A)$ the set of pairs (e, v) of an edge e and a vertex v such that

$$v \in A; \qquad \partial e \cap A = \varnothing; \qquad e, v \text{ form a triangle of } \mathcal{T}.$$

Due to (3-1) and Proposition 3.2, we obtain

$$\sum_{v_i \in A\cap V_\partial} \varphi(v_i) = -\sum_{(e,v)\in Lk(A)} (\pi - \Theta(t_*)(e)) + 2\pi\chi(G(A)\setminus\partial S) + \pi(\chi(G(A)\cap\partial S)).$$

From Thurston's construction, the radius vector r_n produces a $\mathcal{T}$*-type* circle pattern pair $(\mu_n, \mathcal{P}_n)$ realizing $\Theta(t_n)$. Without loss of generality, suppose that $G(A) \setminus \partial S$ is connected and is of topological type (g_0, h_0). Note that

$$\chi(G(A) \setminus \partial S) = 2 - 2g_0 - h_0.$$

Meanwhile, $\chi(G(A) \cap \partial S)$ is equal to $-m$, where m is the number of open arc components of $G(A) \cap \partial S$. This yields

$$\sum_{v_i \in A\cap V_\partial} \varphi(v_i) = -\sum_{(e,v)\in Lk(A)} (\pi - \Theta(t_*)(e)) + 2\pi(2 - 2g_0 - h_0) - m\pi. \tag{3-2}$$

If $g_0 \geq 1$ or $h_0 \geq 2$, the right side of (3-2) is negative, which contradicts the condition that $\varphi(v_i) \geq 0$ for any $v_i \in A$.

Let $g_0 = 0$, $h_0 = 1$. Then $G(A) \setminus \partial S$ is simply connected. Suppose that $\mathrm{Lk}(A) = \{(e_l, v_l)\}_{l=1}^{s}$. Using (3-2), we have

$$\sum_{v_i \in A \cap V_\partial} \varphi(v) = -\sum_{l=1}^{s} (\pi - \Theta(t_*)(e_l)) + 2\pi - m\pi. \tag{3-3}$$

We divide it into the following cases:

(i) $m \geq 2$. Similar arguments to the above part lead to a contradiction.

(ii) $m = 0$. Here, either $G(A) \cap \partial S = \varnothing$, or $G(A) \setminus \partial S$ is bounded by $G(A) \cap \partial S$. Because $G(A) \setminus \partial S$ is simply connected, the latter case occurs if and only if $(g, h) = (0, 1)$ and $A = V$. Applying the Gauss–Bonnet formula, it is easy to see

$$\sum_{i=1}^{|V|} K_i(\Theta(t_n), r_n) = 2\pi\chi(S) + \mathrm{Area}(S, \mu_n).$$

As $n \to \infty$, the radius of every circle tends to zero. Therefore

$$\mathrm{Area}(S, \mu_n) \to 0.$$

Combining this with (3-1), we obtain

$$\sum_{v_i \in V_\partial} \varphi(v) = 2\pi\chi(S),$$

which contradicts the condition (C2).Thus $G(A) \cap \partial S = \varnothing$, and we have $A \cap V_\partial = \varnothing$. By (3-3), we have

$$0 = -\sum_{l=1}^{s} (\pi - \Theta(t_*)(e_l)) + 2\pi. \tag{3-4}$$

However, the edges $e_1, e_2, \ldots, e_s$ form a *nonvacant pseudo-Jordan* curve. According to the condition (C3), we obtain

$$-\sum_{l=1}^{s} (\pi - \Theta(t_*)(e_l)) + 2\pi \;\leq \sum_{l=1}^{s} \Theta(e_l) - (s-2)\pi \;<\; 0,$$

which contradicts (3-4).

(iii) $m = 1$. Similarly, it follows from (3-3) that

$$\sum_{v_i \in A \cap V_\partial} \varphi(v_i) = -\sum_{l=1}^{s} (\pi - \Theta(t_*)(e_l)) + \pi. \tag{3-5}$$

But the edges $e_1, e_2, \ldots, e_s$ form a *semi-pseudo-Jordan* arc so that A is a nonempty *semi-enclosing* set. Under the condition (C4), we have

$$\sum_{v_i\in A\cap V_\partial}\varphi(v_i)+\sum_{l=1}^{s}(\pi-\Theta(t_*)(e_l)) \geq \sum_{v_i\in A\cap V_\partial}\varphi(v_i)+\sum_{l=1}^{s}(\pi-\Theta(e_l)) > \pi,$$

which contradicts (3-5). □

The following result was obtained by Ge, Hua and Zhou [≥ 2020]. It plays a significant role in the above part.

Proposition 3.2. *Let* A, $\mathrm{Lk}(A)$ *and* $G(A)$ *be as above. Then*

$$\sum_{v_i\in A}K_i(\Theta(t_n),r_n)\to -\sum_{(e,v)\in \mathrm{Lk}(A)}(\pi-\Theta(t_*)(e))+2\pi\chi(G(A)\setminus\partial S)+\pi\chi(G(A)\cap\partial S).$$

Remark 3.3. The Euler characteristic $\chi(G(A)\cap\partial S)$ is equal to the opposite of the number of open arc components of $G(A)\cap\partial S$.

It remains to consider the Euclidean geometry case. We easily derive the following combinatorial Gauss–Bonnet formula:

$$\sum_{i=1}^{|V|}K_i=2\pi\chi(S). \tag{3-6}$$

We use $Y\subset\mathbb{R}^{|V|}$ to denote the hyperplane determined by (3-6) and use $X\subset\mathbb{R}_+^{|V|}$ to denote the set of radius vectors $r=(\rho_1,\dots,\rho_{|V|})$ satisfying

$$\sum_{i=1}^{|V|}\rho_i=1.$$

It is easy to see that

$$\dim(X)=\dim(Y)=|V|-1.$$

We have the restriction curvature map $\mathrm{Rh}(\Theta,\cdot)$:

$$X\to Y,$$
$$(\rho_1,\rho_2,\dots,\rho_{|V|})\mapsto (K_1,K_2,\dots,K_{|V|}).$$

Analogously, we can construct the homotopy $\mathrm{Rh}_t=\mathrm{Rh}(\Theta(t),\cdot)$ from $\mathrm{Rh}(\Theta,\cdot)$ to $\mathrm{Rh}(0,\cdot)$. Furthermore, similar arguments to the proof of Lemma 3.1 give the following lemma.

Lemma 3.4. *In Euclidean geometry, there exists a relatively compact open set* $\Lambda\subset X$ *such that*

$$\mathrm{Rh}_t(X\setminus\Lambda)\subset Y\setminus\{K[\varphi]\}\quad \textit{for all } t\in[0,1].$$

We use the strategy of Zhou [2017] for the proof of the following result.

Theorem 3.5. *Suppose that Θ and φ satisfy the conditions of Theorem 1.3. In hyperbolic geometry, we have*

$$\deg(\mathrm{Th}(\Theta,\cdot),\Omega,K[\varphi])=1 \quad or \quad \deg(\mathrm{Th}(\Theta,\cdot),\Omega,K[\varphi])=-1.$$

In Euclidean geometry, we have

$$\deg(\mathrm{Rh}(\Theta,\cdot),\Lambda,K[\varphi])=1 \quad or \quad \deg(\mathrm{Rh}(\Theta,\cdot),\Lambda,K[\varphi])=-1.$$

Proof. In hyperbolic background geometry, due to Theorem A.8,

$$\deg(\mathrm{Th}(\Theta,\cdot),\Omega,K[\varphi])=\deg(\mathrm{Th}_0,\Omega,K[\varphi]).$$

According to Theorem 1.2, $\Omega\cap\mathrm{Th}_0^{-1}(K[\varphi])$ consists of a unique point. Meanwhile, from [Ge et al. $\geq$ 2020], the Jacobian matrix of the map Th_0 is diagonally dominant. Hence $K[\varphi]$ is a regular value of Th_0. As a result,

$$\deg(\mathrm{Th}_0,\Omega,K[\varphi])=1 \quad \text{or} \quad \deg(\mathrm{Th}_0,\Omega,K[\varphi])=-1.$$

In Euclidean geometry, the conclusion follows verbatim from the hyperbolic case. □

Proof of Theorem 1.3. It follows from Theorem 3.5 and Theorem A.9 that $K[\varphi]$ lies in the image of the map $\mathrm{Th}(\Theta,\cdot)$. By Thurston's construction, there exists a constant curvature metric μ on S such that (S,μ) supports a $\mathcal{T}$-type circle pattern $\mathcal{P}$ with the exterior intersection angles given by Θ. □

4. Further discussion

Cone metrics with prescribed curvatures. For a nonempty subset A of V, recall that $G(A)$ is the union of η-cells ($\eta=0,1,2$) of $\mathcal{T}$ that have at least one vertex in A, and $\mathrm{Lk}(A)$ is the set of pairs (e,v) of an edge e and a vertex v such that

$$v\in A,\qquad \partial e\cap A=\varnothing,\qquad \text{the set of } e,v \text{ form a triangle of } \mathcal{T}.$$

The following result is a generalization of Theorem 1.3.

Theorem 4.1. *Let $\mathcal{T}$ be a triangulation of a surface S of topological type (g,h). Assume that $\Theta: E\to[0,\pi)$ is a function satisfying* (R1) *and $K=(k_1,k_2,\dots,k_{|V|})$ is a vector satisfying*

$$k_i < \begin{cases} 2\pi & \text{if } v_i\in V\setminus V_\partial,\\ \pi & \text{if } v_i\in V_\partial,\end{cases} \tag{4-1}$$

and

$$\sum_{v_i\in A} k_i > -\sum_{(e,v)\in \mathrm{Lk}(A)} (\pi-\Theta(e)) + 2\pi\chi(G(A)\setminus\partial S)+\pi\chi(G(A)\cap\partial S) \tag{4-2}$$

for any nonempty subset A of V. Then there exists a hyperbolic cone metric μ in S

such that (S, μ) *has discrete curvatures assigned by* K *and supports a* $\mathcal{T}$*-type circle pattern* $\mathcal{P}$ *with exterior intersection angles given by* Θ*.*

Proof. We still use the homotopy $\mathrm{Th}_t(\cdot) = \mathrm{Th}(\Theta(t), \cdot)$. Similar arguments to the proof of Lemma 3.1 imply that there exists a relatively compact open set $\Omega_K \subset \mathbb{R}_+^{|V|}$ such that

$$\mathrm{Th}_t(\mathbb{R}_+^{|V|} \setminus \Omega_K) \subset \mathbb{R}_+^{|V|} \setminus \{K\} \quad \text{for all } t \in [0, 1].$$

Set $\mathrm{Th}_0(\cdot) = \mathrm{Th}(0, \cdot)$. Because K satisfies the inequalities (4-1) and (4-2), it follows from [Ge et al. $\geq$ 2020, Theorem 0.4] that K is a regular value of the map Th_0 and $\Omega_K \cap \mathrm{Th}_0^{-1}(K)$ consists of a unique point. Therefore,

$$\deg(\mathrm{Th}_0, \Omega_K, K) = 1 \text{ or } \deg(\mathrm{Th}_0, \Omega_K, K) = -1.$$

Due to Theorem A.8, we obtain

$$\deg(\mathrm{Th}(\Theta, \cdot), \Omega_K, K) = 1 \quad \text{or} \quad \deg(\mathrm{Th}(\Theta, \cdot), \Omega_K, K) = -1.$$

Theorem A.9 then implies that K is in the image of $\mathrm{Th}(\Theta, \cdot)$. Applying Thurston's construction, the statement follows. □

Similarly, we obtain the following result.

Theorem 4.2. *Let* $\mathcal{T}$ *be a triangulation of a surface* S *of topological type* (g, h)*. Assume that* $\Theta : E \to [0, \pi)$ *is a function satisfying* (R1) *and* $K = (k_1, k_2, \dots, k_{|V|})$ *is a vector satisfying* (4-1) *and*

$$\sum_{v_i \in A} k_i \geq - \sum_{(e,v) \in \mathrm{Lk}(A)} (\pi - \Theta(e)) + 2\pi\chi(G(A) \setminus \partial S) + \pi\chi(G(A) \cap \partial S) \tag{4-3}$$

for any nonempty subset A *of* V*, where the equality holds if and only if* $A = V$*. Then there exists a Euclidean cone metric* μ *in* S *such that* (S, μ) *has discrete curvatures assigned by* K *and supports a* $\mathcal{T}$*-type circle pattern* $\mathcal{P}$ *with exterior intersection angles given by* Θ*.*

Planar circle patterns. Setting $(g, h) = (0, 1)$, Theorem 1.3 gives the following result.

Theorem 4.3. *Let* $\mathcal{T}$ *be a triangulation of a closed topological disk. Suppose that* $\Theta : E \to [0, \pi)$ *and* $\varphi : V_\partial \to [0, \pi)$ *are two functions such that* (R1), (C3) *and the following conditions are satisfied*:

(T1) *The Gauss–Bonnet inequality* (*resp. equality*) *holds*:

$$\sum_{v \in V_\partial} \varphi(v) > 2\pi \quad \left(\textit{resp. } \sum_{v \in V_\partial} \varphi(v) = 2\pi \right).$$

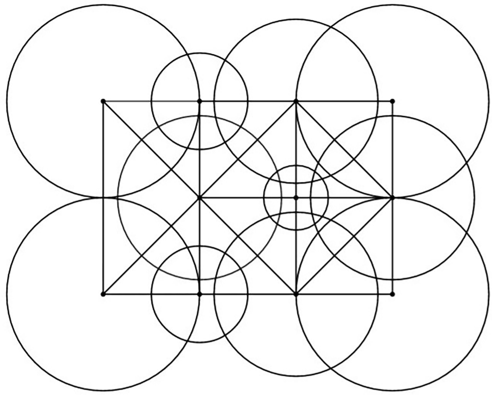


Figure 2. A circle pattern on a rectangle.

(T2) *When the edges $e_1, \dots, e_s$ form a simple arc, joining two distinct boundary vertices, with a nonempty semi-enclosing set W, then*

$$\sum_{v \in W \cap V_\partial} \varphi(v) + \sum_{i=1}^{s} (\pi - \Theta(e_i)) > \pi.$$

Then there exists a convex hyperbolic (*resp. Euclidean*) *polygon Q, so that Q has external angles given by φ and supports a $\mathcal{T}$-type circle pattern $\mathcal{P}$ whose exterior intersection angle function is Θ.*

Turning to the limiting case that $\varphi \equiv \pi$, the following result is straightforward.

Corollary 4.4. *Let $\mathcal{T}$ be a triangulation of a closed topological disk with a function $\Theta : E \to [0, \pi)$ satisfying* (R1), (C3). *Then there exists a $\mathcal{T}$-type circle pattern $\mathcal{P}$ in the hyperbolic disk, whose exterior intersection angle function is Θ and whose boundary circles are horocycles.*

Another corollary deals with circle patterns on rectangles (see Figure 2).

Corollary 4.5. *Let $\mathcal{T}$ be a triangulation of a quadrangle. Suppose $\Theta : E \to [0, \pi)$ is a function satisfying* (R1), (C3) *and the following conditions*:

(Q1) *When the edges $e_1, \dots, e_s$ form a simple arc joining a pair of opposite corner vertices, then $\sum_{i=1}^{s} \Theta(e_i) < (s - 1/2)\pi$.*

(Q2) *When the edges $e_1, \dots, e_s$ form a simple arc which does not join a pair of opposite corner vertices, then $\sum_{i=1}^{s} \Theta(e_i) < (s - 1)\pi$.*

Then there exists a rectangle R which supports a circle pattern $\mathcal{P}$ with exterior intersection angles given by Θ.

Proof. Let $V_* \subset V_\partial$ be the set of corner vertices. We define a function $\varphi : V_\partial \to [0, \pi)$ by setting

$$\varphi(v) = \begin{cases} \pi/2 & \text{if } v \in V_*, \\ 0 & \text{if } v \in V_\partial \setminus V_*. \end{cases}$$

Under the conditions (Q1) and (Q2), it is easy to see that (T1) and (T2) are satisfied. By Theorem 4.3, the conclusion holds. □

Finally, the following result generalizes Marden and Rodin's theorem [1990].

Theorem 4.6. *Let $\mathcal{T}$ be a triangulation of the sphere. Suppose that $\Theta : E \to [0, \pi)$ is a function satisfying* (R1) *and the following conditions*:

(a) *There is a triangle of $\mathcal{T}$ with boundary edges e_a, e_b, and e_c such that*

$$\sum_{l=a,b,c} \Theta(e_l) < \pi.$$

(b) *When the edges $e_1, e_2, \ldots, e_s$ form a simple closed curve which separates the vertices of $\mathcal{T}$, then $\sum_{l=1}^{s} \Theta(e_l) < (s-2)\pi$.*

Then there exists a $\mathcal{T}$-type *circle pattern $\mathcal{P}$ on the Riemann sphere $\hat{\mathbb{C}}$ with the exterior intersection angles given by Θ.*

A circle pattern $\mathcal{P}$ on the sphere can be stereographically projected to the plane. Since stereographic projection is conformal, the exterior intersection angles are the same. More precisely, we can construct the circle pattern on the sphere by constructing the corresponding planar pattern and then projecting it back to the sphere. Thus, to prove Theorem 4.6, we only need to construct a proper planar circle pattern.

Proof of Theorem 4.6. Let $\triangle_\infty$ be the triangle of $\mathcal{T}$ with edges e_a, e_b, e_c and let v_a, v_b, v_c be the vertices opposite to e_a, e_b, e_c. By stereographic projection, then $\mathcal{T}' = \mathcal{T} \setminus \triangle_\infty$ gives a triangulation of $\triangle_\infty$. Moreover, we can regard $\Theta : E \to [0, \pi)$ as a function defined on the edge of T'. Because of the stereographic projection construction, it suffices to show that there exists a $\mathcal{T}'$-type circle pattern $\mathcal{P}$ on the unit disk with the exterior intersection angles given by Θ, which is a result of Corollary 4.4. □

Appendix

In this part we shall give a simple introduction to some results on manifolds, especially the topological degree theory. The reader can refer to [Guillemin and Pollack 1974; Hirsch 1976; Milnor 1965] for more background.

Let M, N be two oriented smooth manifolds. A point $x \in M$ is called critical for a C^1 map $f : M \mapsto N$ if the tangent map $df : T_x \mapsto N_{f(x)}$ is not surjective. Let C_f denote the set of critical points of f, and define $N \setminus f(C_f)$ to be the set of regular values of f.

Theorem A.1 (regular value theorem). *Let $f : M \mapsto N$ be a C^r $(r \geq 1)$ map, and let $y \in N$ be a regular value of f. Then $f^{-1}(y)$ is a closed C^r submanifold of M. If $y \in \mathrm{im}(f)$, then the codimension of $f^{-1}(y)$ is equal to the dimension of N.*

Theorem A.2 (Sard's theorem). *Let M, N be manifolds of dimensions m, n and let $f : M \mapsto N$ be a C^r map. If*

$$r \geq \max\{1, m-n+1\},$$

then $f(C_f)$ has zero measure in N.

Assume that M and N have the same dimension. Let $\Lambda \subset M$ be a relatively compact open subset. Precisely, Λ has compact closure $\overline{\Lambda}$ in M. For a continuous map $f : M \mapsto N$ and a point $y \in N \setminus f(\partial\Lambda)$, we shall define a topological invariant $\deg(f, \Lambda, y)$, called the topological degree of y and f in Λ.

First, suppose that $f \in C^0(\overline{\Lambda}, N) \cap C^\infty(\Lambda, N)$ and y is a regular value. When the set $\Lambda \cap f^{-1}(y)$ is empty, we set $\deg(f, \Lambda, y) = 0$. If $\Lambda \cap f^{-1}(y)$ is nonempty, the regular value theorem implies that it consists of finite points. For a point $x \in \Lambda \cap f^{-1}(y)$, the sign $\mathrm{sgn}(f, x)$ is equal to $+1$ if the tangent map $d_x f : M_x \mapsto N_y$ preserves orientation. Otherwise, $\mathrm{sgn}(f, x) = -1$.

Definition A.3. Suppose that $\Lambda \cap f^{-1}(y) = \{x_1, x_2, \ldots, x_m\}$. We define

$$\deg(f, \Lambda, y) := \sum_{l=1}^{m} \mathrm{sgn}(f, x_l).$$

For simplicity, in what follows we use the notation $f \pitchfork_\Lambda y$ to denote that y is a regular value of the restriction map $f : \Lambda \mapsto N$.

Proposition A.4. *Suppose that $f_i \in C^0(\overline{\Lambda}, N) \cap C^\infty(\Lambda, N)$, $f_i \pitchfork_\Lambda y$ and $f_i(\partial\Lambda) \subset N \setminus \{y\}$, $i = 0, 1$. If there exists a homotopy*

$$H \in C^0(I \times \overline{\Lambda}, N)$$

such that $H(0, \cdot) = f_0(\cdot)$, $H(1, \cdot) = f_1(\cdot)$, and $H(I \times \partial\Lambda) \subset N \setminus \{y\}$, then

$$\deg(f_0, \Lambda, y) = \deg(f_1, \Lambda, y).$$

The following lemma is a consequence of Sard's theorem.

Lemma A.5. *For any $f \in C^0(\overline{\Lambda}, N)$ and $y \in N$, if $f(\partial\Lambda) \subset N \setminus \{y\}$, then there exist $g \in C^0(\overline{\Lambda}, N) \cap C^\infty(\Lambda, N)$ and $H \in C^0(I \times \overline{\Lambda}, N)$ such that*

(i) $g \pitchfork_\Lambda y$;

(ii) $H(0, \cdot) = f(\cdot)$, $H(1, \cdot) = g(\cdot)$;

(iii) $H(I \times \partial\Lambda) \subset N \setminus \{y\}$.

Now we are ready to define the topological degrees for general continuous maps. Recall that $f \in C^0(\overline{\Lambda}, N)$ and $f(\partial\Lambda) \subset N \setminus \{y\}$.

Definition A.6. The topological degree of y and f in Λ is defined as

$$\deg(f, \Lambda, y) = \deg(g, \Lambda, y),$$

where g is given in Lemma A.5.

Due to Proposition A.4, $\deg(f, \Lambda, y)$ is well defined. Below we list several properties of the topological degree.

Theorem A.7. *Let $\gamma \subset N$ be a continuous curve such that $f(\partial\Lambda) \subset N \setminus \{\gamma\}$. Then*

$$\deg(f, \Lambda, \gamma(t)) = \deg(f, \Lambda, \gamma(0)) \quad \textit{for all } t \in [0, 1].$$

For $i = 0, 1$, suppose that $f_i \in C^0(\overline{\Lambda}, N)$ satisfies $f_i(\partial\Lambda) \subset N \setminus \{y\}$.

Theorem A.8. *If there exists $H \in C^0(I \times \overline{\Lambda}, N)$ such that*

(i) $H(0, \cdot\,) = f_0(\cdot\,)$, $H(1, \cdot\,) = f_1(\cdot\,)$,

(ii) $H(I \times \partial\Lambda) \subset N \setminus \{y\}$,

then

$$\deg(f_0, \Lambda, y) = \deg(f_1, \Lambda, y).$$

Finally, it follows from the definition that:

Theorem A.9. *If $\deg(f, \Lambda, y) \neq 0$, then $\Lambda \cap f^{-1}(y) \neq \varnothing$.*

References

[Bobenko et al. 2006] A. I. Bobenko, T. Hoffmann, and B. A. Springborn, "Minimal surfaces from circle patterns: geometry from combinatorics", *Ann. of Math.* (2) **164**:1 (2006), 231–264. MR Zbl

[Dai et al. 2008] J. Dai, X. D. Gu, and F. Luo, *Variational principles for discrete surfaces*, Advanced Lectures in Mathematics **4**, International Press, Somerville, MA, 2008. MR Zbl

[Ge et al. ≥ 2020] H. Ge, B. Hua, and Z. Zhou, "Circle patterns on surfaces of finite topological type", to appear in *Am. J. Math.*

[Guillemin and Pollack 1974] V. Guillemin and A. Pollack, *Differential topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974. MR Zbl

[He and Liu 2013] Z. He and J. Liu, "On the Teichmüller theory of circle patterns", *Trans. Amer. Math. Soc.* **365**:12 (2013), 6517–6541. MR Zbl

[Hirsch 1976] M. W. Hirsch, *Differential topology*, Graduate Texts in Mathematics **33**, Springer, 1976. MR Zbl

[Huang and Liu 2017] X. Huang and J. Liu, "Characterizations of circle patterns and finite convex polyhedra in hyperbolic 3-space", *Math. Ann.* **368**:1-2 (2017), 213–231. MR Zbl

[Liu and Zhou 2016] J. Liu and Z. Zhou, "How many cages midscribe an egg", *Invent. Math.* **203**:2 (2016), 655–673. MR Zbl

[Marden and Rodin 1990] A. Marden and B. Rodin, "On Thurston's formulation and proof of Andreev's theorem", pp. 103–115 in *Computational methods and function theory* (Valparaíso, 1989), edited by S. Ruscheweyh et al., Lecture Notes in Math. **1435**, Springer, 1990. MR Zbl

[Milnor 1965] J. W. Milnor, *Topology from the differentiable viewpoint*, University Press of Virginia, Charlottesville, VA, 1965. MR Zbl

[Schramm 1992] O. Schramm, "How to cage an egg", *Invent. Math.* **107**:3 (1992), 543–560. MR Zbl

[Schramm 1993] O. Schramm, "Square tilings with prescribed combinatorics", *Israel J. Math.* **84**:1-2 (1993), 97–118. MR Zbl

[Stephenson 2005] K. Stephenson, *Introduction to circle packing: the theory of discrete analytic functions*, Cambridge University Press, 2005. MR Zbl

[Thurston 1979] W. P. Thurston, "The geometry and topology of three-manifolds", lecture notes, Princeton University, 1979, Available at http://msri.org/publications/books/gt3m.

[Zhou 2017] Z. Zhou, "Circle patterns with obtuse exterior intersection angles", preprint, 2017. arXiv

Received May 13, 2019. Revised December 20, 2019.

YUE-PING JIANG
KEY LAB OF INTELLIGENT INFORMATION PROCESSING AND APPLIED MATHEMATICS
SCHOOL OF MATHEMATICS
HUNAN UNIVERSITY
CHANGSHA
CHINA
ypjiang@hnu.edu.cn

QIANGHUA LUO
KEY LAB OF INTELLIGENT INFORMATION PROCESSING AND APPLIED MATHEMATICS
SCHOOL OF MATHEMATICS
HUNAN UNIVERSITY
CHANGSHA
CHINA
luo.qh@hnu.edu.cn

ZE ZHOU
SCHOOL OF MATHEMATICS
HUNAN UNIVERSITY
CHANGSHA
CHINA
zhouze@hnu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.221

ON SOME CONJECTURES OF HEYWOOD

DONG LI

We prove several sharp interpolation inequalities for Dirichlet Laplacians on general domains. We show optimality of the interpolation constants (which are independent of the underlying domain) and present several proofs. This settles several questions recent questions of Heywood *Ann. Univ. Ferrara Sez. VII Sci. Mat.* 60:1 (2014), 149–167.

1. Introduction

In a recent series of papers, Heywood and Xie [Heywood 2014; 2000; Heywood and Xie 1997; Xie 1991; 1995; 1997] embarked on a program to build a wellposedness theory of Navier–Stokes equations requiring no regularity assumptions on the boundary. Xie [1992] conjectured the inequality

$$\|\boldsymbol{u}\|^2_{L^\infty(\Omega)} \le \frac{1}{3\pi}\|\nabla \boldsymbol{u}\|_{L^2(\Omega)}\|\tilde{\Delta}\boldsymbol{u}\|_{L^2(\Omega)}, \tag{1-1}$$

where Ω is an arbitrary open set in $\mathbb{R}^3$, $\boldsymbol{u}$ is any vector field in the completion of divergence free C_c^∞ vector fields in the Dirichlet norm $\|\nabla\cdot\|$, and $\tilde{\Delta}$ is the Stokes operator. Xie [1991] considered the model Poisson problem and proved

$$\|u\|^2_{L^\infty(\Omega)} \le \frac{1}{2\pi}\|\nabla u\|_{L^2(\Omega)}\|\Delta u\|_{L^2(\Omega)} \tag{1-2}$$

MSC2010: 35Q35.
Keywords: Heywood conjecture, Dirichlet Laplacian, interpolation inequality, sharp constant.

for the Laplacian case. The key argument in [Xie 1991] is the use of a maximum principle

$$\|G_\mu(\cdot\,,y)\|_{L^2(\Omega)} \le \|g_\mu(\cdot\,,y)\|_{L^2(\mathbb{R}^3)},$$

where G_μ is the Green's function for $-\Delta_D+\mu$ (Δ_D is the Dirichlet Laplacian) and g_μ is the Green's function on $\mathbb{R}^3$. Heywood [2000; 2014] reformulated several deep conjectures seeking to prove the Stokes case (1-1) for which the maximum principle is no longer available. In order to reach further understanding of these problems Heywood [2014] formulated several analogous conjectures for the Fourier sine series and cosine series. The purpose of this paper is to settle some of these questions. Our first result concerns one-dimensional Dirichlet Laplacian operators.

Theorem 1.1. *For $\Omega=(0,\pi)$, if f is real-valued and $f\in H_0^1(\Omega)\cap H^2(\Omega)$, then*

$$\|f\|^2_{L^\infty(\Omega)} \le 3\|(-\Delta_D)^{1/6}f\|_{L^2(\Omega)}\cdot\|(-\Delta_D)^{1/3}f\|_{L^2(\Omega)},$$

where Δ_D is the Dirichlet Laplacian on Ω. Furthermore the constant 3 is sharp.

An equivalent reformulation of Theorem 1.1 is the following corollary.

Corollary 1.2 (conjecture by Heywood [2014]). *Denote $\Omega=(-\pi,\pi)$. For a Fourier series $f=\sum_{n=1}^\infty f_n\sin nx$, we have*

$$\|f\|^2_{L^\infty(\Omega)} \le \tfrac{3}{2}\|f^{(1/3)}\|_{L^2(\Omega)}\|f^{(2/3)}\|_{L^2(\Omega)},$$

where $f^{(1/3)}$ (resp. $f^{(2/3)}$) denotes the fractional derivative of order $\frac13$ $\left(\text{resp. } \frac23\right)$ for f.[1] *Furthermore the constant $\frac32$ is sharp.*

Remark. Note that for $f=\sum_{n=1}^\infty f_n\sin nx$, $\|f^{(1/3)}\|_{L^2(-\pi,\pi)}=\sqrt{2}\|f^{(1/3)}\|_{L^2(0,\pi)}$ and hence the factor $\frac32$ in Corollary 1.2.

Theorem 1.3. *Let $\Omega=(-\pi,\pi)$. For a Fourier cosine series $f=\sum_{n=1}^\infty f_n\cos nx$, we have*

$$\|f\|^2_{L^\infty(\Omega)} \le 3\|f^{(1/3)}\|_{L^2(\Omega)}\|f^{(2/3)}\|_{L^2(\Omega)}. \tag{1-3}$$

Furthermore the constant 3 is sharp. Similarly, for a Fourier series

$$f=\sum_{n=1}^\infty f_n^{(1)}\cos nx+\sum_{n=1}^\infty f_n^{(2)}\sin nx,$$

we have

$$\|f\|^2_{L^\infty(\Omega)} \le 3\|f^{(1/3)}\|_{L^2(\Omega)}\|f^{(2/3)}\|_{L^2(\Omega)}. \tag{1-4}$$

The constant 3 is also sharp.

[1] Here in the context of Fourier series, one can regard the fractional derivative of order α as the Fourier multiplier n^α.

Remark. Corollary 1.2 was conjectured by Heywood [2014, Section 8]. The inequality (1-3) was proved by Heywood [2014] and he conjectured that the constant 3 is sharp [2014, Conjecture 1]. Here we settle these questions in the affirmative.

Remark. The main issue in Theorems 1.1 and 1.3 is the sharpness of the constants. Of course if one does not insist on having sharp constants, the above inequalities can be obtained by a usual splitting argument.

Theorem 1.1 can certainly be generalized further. Below we give a generalization to all dimensions with explicit interpolation constants.

Theorem 1.4. *Let Ω be a bounded domain in $\mathbb{R}^d$ with smooth boundary and denote by Δ_D the Dirichlet Laplacian on Ω. Denote $A=(-\Delta_D)^{1/2}$ and assume $(d-1)/2 \le \alpha < d/2$. Let $N=(d+3)/2$ if d is odd, and $N=(d+2)/2$ if d is even. Then for any $f:\ \Omega \to \mathbb{R}$ with $f \in H^N(\Omega) \cap H_0^1(\Omega)$, we have*

$$\|f\|_{L^\infty(\Omega)}^2 \le C_{\alpha,d} \|A^\alpha f\|_{L^2(\Omega)} \|A^{d-\alpha} f\|_{L^2(\Omega)},$$

where

$$C_{\alpha,d} = \frac{d}{d-2\alpha} \cdot \frac{1}{2^d \pi^{d/2-1} \Gamma\left(\frac{d}{2}+1\right)}.$$

The constant $C_{\alpha,d}$ is sharp.

Remark. The same result holds when $\Omega = \mathbb{R}^d$. A remarkable feature is that the constant $C_{\alpha,d}$ is independent of the domain.

Remark. It is easy to check that if $d=1$ and $\alpha=\frac{1}{3}$, then $C_{\alpha,d}=3$. If $\alpha=1$ and $d=3$, then $C_{\alpha,d}=1/(2\pi)$ and we recover the inequality (1-2).

The interpolation exponent $\frac{1}{2}$ is not special and we have the following generalization for all $0<\theta<1$.

Theorem 1.5. *Let Ω be a smooth bounded domain in $\mathbb{R}^d$ and denote by Δ_D the Dirichlet Laplacian on Ω. Denote $A=(-\Delta_D)^{1/2}$. Assume $\alpha \ge 0$ and $\beta > 0$, $0<\theta<1$, satisfy*:

- $0 \le \alpha < d/2$.
- $\alpha\theta + \beta(1-\theta) = d/2$.
- $0 < \beta - \alpha \le 1$.

Let $N=(d+3)/2$ if d is odd and $N=(d+2)/2$ if d is even. Then for any $f:\Omega \to \mathbb{R}$ with $f \in H^N(\Omega) \cap H_0^1(\Omega)$, we have

$$\|f\|_{L^\infty(\Omega)}^2 \le C_{\alpha,\beta,\theta,d} \left(\|A^\alpha f\|_{L^2(\Omega)}^\theta \|A^\beta f\|_{L^2(\Omega)}^{1-\theta}\right)^2,$$

where

$$C_{\alpha,\beta,\theta,d} = (2\pi)^{-d} \cdot \frac{1}{\beta-\alpha} \cdot \frac{2\pi^{d/2+1}}{\Gamma\left(\frac{d}{2}\right)} \cdot \frac{1}{\sin \pi\theta}.$$

The constant $C_{\alpha,\beta,\theta,d}$ is sharp.

Remark. Again the same inequality holds when $\Omega = \mathbb{R}^d$. In this case one can characterize all functions attaining the sharp interpolation constant. Modulo dilation and translation symmetries, such functions take the form

$$f(x) = \text{const} \cdot ((-\Delta)^{-\alpha/2}(1 + (-\Delta)^{\beta-\alpha})^{-1}\delta_0)(x),$$

where δ_0 is the usual Dirac distribution on $\mathbb{R}^d$ with mass centered at the origin. See Remark 8.4 for a proof and more details.

Remark 1.6. The constant $C_{\alpha,\beta,\theta,d}$ is sharp in the sense that there exists $f_n \in C_c^\infty(\Omega)$, such that

$$\frac{\|f_n\|^2_{L^\infty(\Omega)}}{(\|A^\alpha f_n\|^\theta_{L^2(\Omega)}\|A^\beta f_n\|^{1-\theta}_{L^2(\Omega)})^2} \to C_{\alpha,\beta,\theta,d}, \quad \text{as } n \to \infty.$$

See Remark 7.3 for an explicit construction. In particular, such functions can be constructed to be essentially independent of the underlying domain. On the other hand, a natural question is whether for any smooth bounded domain Ω there can exist a function f achieving the sharp interpolation constant. We shall disprove this at the end of Section 8. In other words, as long as Ω is not the whole space $\mathbb{R}^d$, then for any $f \in H^N(\Omega) \cap H^1_0(\Omega)$ which is not identically zero, we actually have the strict inequality

$$\|f\|^2_{L^\infty(\Omega)} < C_{\alpha,\beta,\theta,d} \cdot \left(\|A^\alpha f\|^\theta_{L^2(\Omega)}\|A^\beta f\|^{1-\theta}_{L^2(\Omega)}\right)^2.$$

We shall give three different proofs of Theorem 1.1. In the first proof, we shall follow the variational approach initiated in the aforementioned papers by Heywood and Xie. This approach is remarkably effective for the L^∞ setting. In the simplest scenario, one can see Lemma 2.1 which can be regarded as a device to reduce the original infinite-dimensional problem to a finite-dimensional one. One can then prove the sharp interpolation inequality (as well as showing the sharpness of interpolation constants) by proving certain limiting asymptotics of Lagrange multipliers μ_N. Specifically for Theorem 1.1 the main step is the proof of Proposition 3.1 which will occupy most of this paper. Obtaining this proposition is the key to showing the sharpness of the interpolation constant $\frac{3}{2}$ for the sine series case. In Sections 3–6 we give the proof of Proposition 3.1 with the help of a series of lemmas for different regimes of parameters. Here we exploit certain monotonicity and convexity properties of the Fourier coefficients, as well as some sharp bounds on the polylog function. In Section 7 we will give a second (and slightly shorter) proof of Theorem 1.1 (and Theorems 1.4 and 1.5) by using a comparison argument. Here the main point is to prove an upper bound directly without using any asymptotics of μ_N. The sharpness of the interpolation constant is proved separately by using a "blowup" argument. In Section 8, we give a third proof (which is also the shortest), which does not rely on any variational argument. The key technical input is a novel splitting of the operator symbol saturating the comparison principle for fractional operators. Thanks to this new proof we are able to characterize the set of functions attaining the sharp constant. It is possible that these ideas can be suitably adapted to much more general settings.

We now gather some notation used in this paper.

Notation. For any $1 \le p \le \infty$, we denote by $\|f\|_{L^p(\Omega)}$ or $\|f\|_{L^p_x(\Omega)}$ (to stress that f is a function of x) the usual Lebesgue norm of a function f over a region $\Omega \subset \mathbb{R}^d$.

We denote by $\mathcal{S}(\mathbb{R}^d)$ the usual Schwartz functions on $\mathbb{R}^d$.

We shall adopt the following convention for Fourier transforms. For $f : \mathbb{R}^d \to \mathbb{R}$, denote the Fourier transform

$$\hat{f}(\xi) = (\mathcal{F}f)(\xi) = \int_{\mathbb{R}^d} f(x)e^{-ix\cdot\xi}\,dx.$$

The inverse Fourier transform is then given by the formula

$$f(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \hat{f}(\xi)e^{ix\cdot\xi}\,d\xi.$$

Note that with this convention of Fourier transform, the usual Parseval formula reads as

$$\|f\|_{L^2_x(\mathbb{R}^d)} = (2\pi)^{-d/2}\|\hat{f}\|_{L^2_\xi(\mathbb{R}^d)}.$$

For $s \in \mathbb{R}$, the fractional Laplacian operator $|\nabla|^s = (-\Delta)^{s/2}$ is denoted by

$$\widehat{|\nabla|^s f}\,(\xi) = |\xi|^s\hat{f}(\xi).$$

For $s \in \mathbb{R}$, the Sobolev norm $\|f\|_{\dot{H}^s(\mathbb{R}^d)}$ is computed as

$$\|f\|_{\dot{H}^s(\mathbb{R}^d)} = \||\nabla|^s f\|_{L^2(\mathbb{R}^d)} = (2\pi)^{-d/2}\||\xi|^s\hat{f}(\xi)\|_{L^2_\xi(\mathbb{R}^d)},$$

whenever it is well defined.

Let Ω be any bounded domain of $\mathbb{R}^d$ with smooth boundary. We denote by $0 < \lambda_1 < \lambda_2 \le \cdot \le \lambda_k \to \infty$ the ordered eigenvalues (counted with multiplicity) with corresponding L^2-normalized eigen-functions ϕ_j for the Dirichlet Laplacian Δ_D on Ω. Sometimes to stress the dependence on Ω we write Δ_Ω in place of Δ_D. For any $s \in \mathbb{R}$ and any f with the expansion (over the L^2 orthonormal basis $(\phi_j)_{j=1}^\infty$)

$$f = \sum_{j=1}^\infty f_j\phi_j,$$

one can define the Sobolev norm $\|f\|_{\dot H^s(\Omega)}$ as (noting that $-\Delta_D\phi_j=\lambda_j\phi_j$)

$$\|f\|_{\dot H^s(\Omega)}=\|(-\Delta_D)^{s/2}f\|_{L^2(\Omega)}=\left(\sum_{j=1}^{\infty}|f_j|^2\lambda_j^s\right)^{1/2},$$

provided the series converges. The fractional Laplacian operator $(-\Delta_D)^s$ is introduced via the usual functional calculus, or sometimes using the quadratic form. Also one can view it as the multiplier λ^s with the help of eigen-function expansions. The inhomogeneous Sobolev norm is simply defined as $\|f\|_{H^s(\Omega)}=\|f\|_{L^2(\Omega)}+\|f\|_{\dot H^s(\Omega)}$.

For any $x\in\mathbb{R}$, we denote by $[x]$ the closest integer to x. For $x\in\mathbb{C}$ with $\mathrm{Re}(x)>0$, the usual Gamma function is defined by

$$\Gamma(x)=\int_0^\infty e^{-t}t^{x-1}\,dt.$$

We shall not need the analytic continuation of Γ into the left half complex plane.

2. Preliminary reductions and proof of Theorems 1.1 and 1.3

Lemma 2.1. *Let $(a_n)_{n=1}^N$, $(v_n)_{n=1}^N$, $(w_n)_{n=1}^N$ be given real numbers. Assume $v_n>0$, $w_n>0$ for all $1\le n\le N$. For any real numbers $b_1,\dots,b_N$ with $(b_1,\dots,b_N)\neq(0,\dots,0)$, consider*

$$R(b_1,\dots,b_N)=\frac{\left(\sum_{n=1}^N a_nb_n\right)^2}{\left(\sum_{n=1}^N b_n^2v_n\right)^{1/2}\left(\sum_{n=1}^N b_n^2w_n\right)^{1/2}}.$$

Then

$$\max_{0\neq(b_1,\dots,b_N)\in\mathbb{R}^N}R(b_1,\dots,b_N)=\mu^{-\frac12}\sum_{n=1}^N\frac{4a_n^2v_n}{(v_n+w_n/\mu)^2},$$

where $\mu>0$ is a root solving the equation

$$\sum_{n=1}^N\frac{4a_n^2}{(v_n+w_n/\mu)^2}(v_n\mu-w_n)=0.$$

Proof. Without loss of generality we may assume $a_n\ge0$ for all $1\le n\le N$ and $(a_1,\dots,a_N)\neq(0,\dots,0)$. In view of homogeneity the supremum of R is attained at some finite $b_i=\bar b_i\ge0$, $1\le i\le N$. Denote $u=\sum_{n=1}^N a_n\bar b_n$, $A=\sum_{n=1}^N\bar b_n^2v_n$, $B=\sum_{n=1}^N\bar b_n^2w_n$, and $\mu=B/A$. Differentiating $\log R$ at $b_n=\bar b_n$ then gives

$$\frac{2a_n}{u}=\frac{\bar b_nv_n}{A}+\frac{\bar b_nw_n}{B}=\frac{\bar b_n}{A}\left(v_n+\frac{w_n}{\mu}\right),$$

or

$$\frac{2a_n}{v_n + w_n/\mu} = \frac{u}{A}\bar{b}_n.$$

This then yields two identities:

$$\sum_{n=1}^{N} \frac{4a_n^2 v_n}{(v_n + w_n/\mu)^2} = \frac{u^2}{A^2}\sum_{n=1}^{N} \bar{b}_n^2 v_n = \frac{u^2}{A},$$
$$\sum_{n=1}^{N} \frac{4a_n^2 w_n}{(v_n + w_n/\mu)^2} = \frac{u^2}{A^2} B = \mu\frac{u^2}{A}.$$

Clearly then the maximum of R is given by

$$R = \frac{u^2}{A^{1/2}B^{1/2}} = \mu^{-\frac{1}{2}}\frac{u^2}{A} = \sum_{n=1}^{N} \frac{4a_n^2 v_n}{(\mu^{1/4}v_n + \mu^{-3/4}w_n)^2},$$

where $\mu > 0$ solves

$$\sum_{n=1}^{N} \frac{4a_n^2}{(v_n + w_n/\mu)^2}(v_n\mu - w_n) = 0. \qquad \square$$

The l^2-summation or the interpolation exponent $\frac{1}{2}$ are not special. Below we record a similar lemma for all l^p-summations and interpolation exponents $\theta \in (0, 1)$. The proof is essentially the same and therefore omitted.

Lemma 2.2. *Let $(a_n)_{n=1}^N$, $(v_n)_{n=1}^N$, $(w_n)_{n=1}^N$ be given real numbers. Assume $v_n > 0$, $w_n > 0$ for all $1 \le n \le N$. Let $1 < p < \infty$ and $0 < \theta < 1$. For any real numbers $b_1, \ldots, b_N$ with $(b_1, \ldots, b_N) \ne (0, \ldots, 0)$, consider*

$$R(b_1, \ldots, b_N) = \frac{\left(\sum_{n=1}^N a_n b_n\right)^2}{\left(\sum_{n=1}^N |b_n|^p v_n\right)^{2\theta/p}\left(\sum_{n=1}^N |b_n|^p w_n\right)^{2(1-\theta)/p}}.$$

Then

$$\max_{0\ne(b_1,\ldots,b_N)\in\mathbb{R}^N} R(b_1, \ldots, b_N) = \mu^{-2(1-\theta)/p}\left(\sum_{n=1}^{N} \frac{|a_n|^{p/(p-1)} v_n}{(\theta v_n + (1-\theta)w_n/\mu)^{p/(p-1)}}\right)^{2(p-1)/p},$$

where $\mu > 0$ is a root solving the equation

$$\sum_{n=1}^{N} \frac{|a_n|^{p/(p-1)}}{(\theta v_n + (1-\theta)w_n/\mu)^{p/(p-1)}}(v_n\mu - w_n) = 0.$$

Remark. For the special case $v_n \equiv w_n$, one can easily check that $\mu = 1$ and we recover the usual Hölder's inequality:

$$\left|\sum_{n=1}^{N} a_n b_n\right| \le \left(\sum_{n=1}^{N} |b_n|^p w_n\right)^{1/p} \left(\sum_{n=1}^{N} |a_n|^{p/(p-1)} w_n^{-1/(p-1)}\right)^{(p-1)/p}.$$

Of course this is equivalent to the case $w_n \equiv 1$ by a change of variable

$$a_n w_n^{-1/p} \to a_n \quad \text{and} \quad b_n w_n^{1/p} \to b_n.$$

A "continuous" consequence of Lemma 2.2 is the following proposition.

Proposition 2.3. *Let* $1 < p < \infty$ *and* $0 < \theta < 1$. *Assume* $-d/p < a < (p-1)/pd < b < \infty$ *satisfy*

$$d\left(1 - \frac{1}{p}\right) = a\theta + b(1-\theta).$$

For any $g : \mathbb{R}^d \to \mathbb{C}$ *with* $g \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, *if* $|x|^a g \in L^p$ *and* $|x|^b g \in L^p$, *then* $g \in L^1$, *and*

$$\|g\|_{L^1(\mathbb{R}^d)} \le C_1 \cdot \| |x|^a g\|^{\theta}_{L^p(\mathbb{R}^d)} \cdot \| |x|^b g\|^{1-\theta}_{L^p(\mathbb{R}^d)},$$

where

$$C_1 = \left(\int_{\mathbb{R}^d} |x|^{-pa/(p-1)} (\theta + (1-\theta)|x|^{p(b-a)})^{-p/(p-1)}\, dx\right)^{(p-1)/p}.$$

Remark. For $d = 1,\ p = 2,\ \theta = \frac{1}{2},\ a = 0$ and $b = 1$, we get

$$\|g\|_{L^1(\mathbb{R})} \le \sqrt{2\pi} \cdot \|g\|^{1/2}_{L^2(\mathbb{R})} \|xg\|^{1/2}_{L^2(\mathbb{R})}. \tag{2-1}$$

The sharp constant is achieved by choosing $g(x) = 1/\pi \cdot 1/(1+x^2)$. On the other hand, the usual splitting argument (i.e., splitting the L^1 norm of g into $|x| \le R$ and $\ge R$, Cauchy–Schwartz and then optimizing in R) only gives the nonsharp inequality

$$\|g\|_{L^1(\mathbb{R})} \le 2\sqrt{2} \cdot \|g\|^{1/2}_{L^2(\mathbb{R})} \|xg\|^{1/2}_{L^2(\mathbb{R})}.$$

Note also that if g is an even function on $\mathbb{R}$, then (2-1) becomes

$$\|g\|_{L^1(\mathbb{R}^+)} \le \sqrt{\pi} \|g\|^{1/2}_{L^2(\mathbb{R}^+)} \|xg\|^{1/2}_{L^2(\mathbb{R}^+)}.$$

This inequality will appear again in (2-3) below.

Remark. A different (and considerably shorter) proof is available. See Remark 8.3.

Remark. Alternatively one can also carry out a variational argument minimizing the energy functional $E(g) = \|g\|_2^2 + \|xg\|_2^2$ subject to the constraint $\|g\|_1 = 1$. The minimizer is then given by the Cauchy distribution. We leave it to interested readers to verify the details.

Proof of Proposition 2.3. First we consider the case $0 \le a < (p-1)/(p)d < b < \infty$. Without loss of generality, we can assume $g \ge 0$, $g \in L^1$ and g is compactly supported. By mollification one can further assume that $g \in C_c^\infty(\mathbb{R}^d)$. By scaling we can assume $\operatorname{supp}(g) \subset [-1, 1]^d$. Now for each integer $J \ge 10$ define

$$\begin{aligned}\Omega_J &= \left\{ \left(\frac{j_1}{J}, \frac{j_2}{J}, \dots \frac{j_d}{J} \right) : 1 \le |j_i| \le J-1,\ 1 \le i \le d \right\} \\ &= \{x_n : 1 \ne n \le N_J := (2(J-1))^d\}.\end{aligned}$$

The set Ω_J provides a partition of the cube $[-1, 1]^d$. Denote $v(x) = |x|^{ap}$ and $w(x) = |x|^{bp}$. We then consider

$$R_{N_J}(b_1, \dots, b_{N_J}) = \frac{\left(\sum_{n=1}^{N_J} b_n\right)^2}{\left(\sum_{n=1}^{N_J} |b_n|^p v(x_n)\right)^{2\theta/p} \left(\sum_{n=1}^{N_J} |b_n|^p w(x_n)\right)^{2(1-\theta)/p}}.$$

By Lemma 2.2, we get for some $\mu_J > 0$:

$$R_{N_J} \le \left(\sum_{n=1}^{N_J} h\left(\frac{x_n}{t_J}\right) \cdot \frac{1}{(t_J J)^d} \right)^{2(p-1)/p},$$

where $t_J = \mu_J^{1/((b-a)p)}$ and

$$h(x) = |x|^{-ap/(p-1)} (\theta + (1-\theta)|x|^{(b-a)p})^{-p/(p-1)}.$$

Note that $\partial_{x_j} h(x) < 0$ on $(0, \infty)^d$. Recall that for any one-variable function $f : (0, \infty) \to [0, \infty)$ with $f'(x) \le 0$, we have

$$\sum_{j=1}^{M} f\left(\frac{j}{M}\right) \frac{1}{M} \le \int_0^1 f(x)\, dx.$$

By iteratively using the above inequality along each coordinate axis, we have

$$\begin{aligned}\sum_{n=1}^{N_J} h\left(\frac{x_n}{t_J}\right) \cdot \frac{1}{(t_J J)^d} &= 2^d \sum_{j_1, \dots, j_d = 1}^{J-1} h\left(\frac{j_1}{t_J J}, \dots, \frac{j_d}{t_J J}\right) \cdot \frac{1}{(t_J J)^d} \\ &\le \int_{\mathbb{R}^d} h(x/t_J) \cdot \frac{1}{t_J^d}\, dx = \int_{\mathbb{R}^d} h(x)\, dx.\end{aligned}$$

The desired inequality then follows by taking $b_n = g(x_n)$, normalizing the corresponding Riemann sum and sending J to infinity. This finishes the case $a \ge 0$.

Next we discuss how to modify the argument for the case $-d/p < a < 0$. Without loss of generality, one can assume $g \ge 0$, $g \in L^1(\mathbb{R}^d)$, and is compactly supported. Let $\eta \in C_c^\infty(\mathbb{R}^d)$ be such that $\eta(x) \equiv 1$ for $|x| \le 1$ and $\eta(x) = 0$ for $|x| \ge 2$. Define $\eta_\epsilon(x) = \epsilon^{-d} \eta(x/\epsilon)$ and $g_\epsilon = g * \eta_\epsilon$. Note that $|x|^a g_\epsilon(x) \le \text{const} \cdot |x|^a \cdot (\mathcal{M} g)(x)$

(here $\mathcal{M}g$ is the usual Hardy–Littlewood maximal function) and $|x|^a \cdot (\mathcal{M}g) \in L^p$ (note that $|x|^{ap}$ for $-d/p < a < (p-1)/(p)d$ is an A_p weight), thus by Lebesgue dominated convergence we have $\||x|^a g_\epsilon\|_p \to \||x|^a g\|_p$ as $\epsilon \to 0$. Similarly noting that $|x|^b$ is bounded on the support of g, we have $\||x|^b g_\epsilon\|_p \to \||x|^b g\|_p$ as $\epsilon \to 0$. Thus we have shown that it suffices to prove the inequality when $g \in C_c^\infty(\mathbb{R}^d)$ and $g \geq 0$. By scaling one can assume $\operatorname{supp}(g) \subset [-1, 1]^d$.

Now fix $\epsilon_0 > 0$ which later will tend to zero. For each integer $J \geq 10$, define

$$\begin{aligned}\Omega'_J &= \left\{\left(\frac{j_1}{J}, \frac{j_2}{J}, \dots, \frac{j_d}{J}\right) : \epsilon_0 J \leq |j_i| \leq J,\, 1 \leq i \leq d\right\}\\ &= \{x_n : 1 \leq n \leq N_J\},\end{aligned}$$

where N_J is the cardinality of Ω'_J. Recall $v(x) = |x|^{ap},\ w(x) = |x|^{bp}$ and consider

$$R_{N_J}(b_1, \dots, b_{N_J}) = \frac{\left(\sum_{n=1}^{N_J} b_n\right)^2}{\left(\sum_{n=1}^{N_J} |b_n|^p v(x_n)\right)^{2\theta/p} \left(\sum_{n=1}^{N_J} |b_n|^p w(x_n)\right)^{2(1-\theta)/p}}.$$

By Lemma 2.2, for some $\mu = \mu_J > 0$ satisfying

$$\sum_{n=1}^{N_J} \frac{1}{(\theta v(x_n) + (1-\theta) w(x_n)/\mu)^{p/(p-1)}} (v(x_n)\mu - w(x_n)) = 0, \tag{2-2}$$

we have

$$R_{N_J} \leq \left(\sum_{n=1}^{N_J} h\left(\frac{x_n}{t_J}\right) \cdot \frac{1}{(t_J J)^d}\right)^{2(p-1)/p},$$

where $t_J = \mu_J^{1/((b-a)p)}$ and

$$h(x) = |x|^{-ap/(p-1)} (\theta + (1-\theta)|x|^{(b-a)p})^{-p/(p-1)}.$$

Since $|x_n| \geq \epsilon_0$ for any $x_n \in \Omega'_J$, we have $w(x_n)/v(x_n) \geq \epsilon_0^{(b-a)p} > 0$. Thus by (2-2), we have

$$\mu_J \geq \epsilon_0^{(b-a)p} > 0.$$

Now regarding h as radial function, $h'(\rho) = 0$ for some $\rho = \rho_0$ satisfying

$$a + \frac{(1-\theta)(b-a)p}{\theta + (1-\theta)\rho^{(b-a)p}} = 0.$$

In the Riemann sum, one can then bound the transition region (where h' switches its sign) easily as:

$$\sum_{\substack{x_n \in \Omega'_J \\ |x_n/t_J - \rho_0| < 10d/t_J J}} h\left(\frac{x_n}{t_J}\right) \cdot \frac{1}{(t_J J)^d} \leq \text{const} \cdot (J t_J)^{-1} \to 0 \quad \text{as } J \to \infty.$$

In the region where h' has a fixed sign, one can then easily make a comparison argument and get the inequality

$$\limsup_{J\to\infty} R_{N_J} \le \left(\int_{\mathbb{R}^d} h(x)\,dx\right)^{2(p-1)/p} + B_{\epsilon_0},$$

where $B_{\epsilon_0} \to 0$ as $\epsilon_0 \to 0$. The desired inequality then follows. □

An immediate corollary of Proposition 2.3 is the following sharp interpolation inequality.

Corollary 2.4. *Let the dimension d be greater than or equal to* 1. *Let* $0 < \theta < 1$ *and* $-d/2 < a < d/2 < b < \infty$ *satisfy*

$$\frac{d}{2} = a\theta + b(1-\theta).$$

Then for any real-valued $f \in \mathcal{S}(\mathbb{R}^d)$, *we have*

$$\begin{aligned}\|f\|_{L_x^\infty(\mathbb{R}^d)} &\le \frac{1}{(2\pi)^d}\|\hat{f}(\xi)\|_{L_\xi^1(\mathbb{R}^d)} \\ &\le C_2\, \|\,|\nabla|^a f\|_{L_x^2(\mathbb{R}^d)}^{\theta} \cdot \|\,|\nabla|^b f\|_{L_x^2(\mathbb{R}^d)}^{1-\theta},\end{aligned}$$

where $|\nabla| = \sqrt{-\Delta}$ *and*

$$C_2 = (2\pi)^{-d/2}\cdot\left(\int_{\mathbb{R}^d} |x|^{-2a}(\theta + (1-\theta)|x|^{2(b-a)})^{-2}\,dx\right)^{1/2}.$$

Remark. For the special case $d = 3,\ a = 1,\ b = 2$ and $\theta = 1/2$, we get $C_2 = 1/\sqrt{2\pi}$ and we recover (1-2) on $\mathbb{R}^3$. In this case, it is also possible to find an explicit function achieving the sharp constant C_2. Indeed by using the Fourier transform, one only needs to find a radial function f, such that its Fourier transform $\hat{f}(\xi) = \int_{\mathbb{R}^3} f(x)e^{-ix\xi}\,dx \ge 0$ satisfies

$$\begin{aligned}(\|f\|_\infty = 1):& \int_0^\infty \hat{f}(\rho)\rho^2\,d\rho = 2\pi^2,\\ (\|\nabla f\|_2\cdot\|\Delta f\|_2 = 2\pi):& \left(\int_0^\infty (\hat{f}(\rho)\rho^2)^2\,d\rho\right)\cdot\left(\int_0^\infty (\hat{f}(\rho)\rho^2)^2\rho^2\,d\rho\right) = 16\pi^6.\end{aligned}$$

Denote $\hat{f}(\rho)\rho^2 = 2\pi^2 g(\rho)$. Then $\int_0^\infty g(\rho) = 1$ and

$$\int_0^\infty g(\rho)^2\,d\rho \cdot \int_0^\infty g(\rho)^2\rho^2\,d\rho = \pi^{-2}. \tag{2-3}$$

Choosing $g(\rho) = 2/\pi\cdot(1+\rho^2)^{-1}$ then achieves both equalities.

Proof of Corollary 2.4. Use Fourier transforms and Proposition 2.3. □

Lemma 2.5 (for Theorem 1.1). *Fix $0 < y < \pi$. For any $N \geq 2$, suppose $\mu_N > 0$ solves*

$$\sum_{n=1}^{N} \frac{\sin^2 ny}{(n^{2/3} + \mu_N^{-1} n^{4/3})^2} (n^{2/3}\mu_N - n^{4/3}) = 0.$$

Then

(1) $$\lim_{N\to\infty} \mu_N = \infty.$$

(2) $$\lim_{N\to\infty} \frac{\mu_N}{N^{2/3}} = 0.$$

(3) $$\lim_{N\to\infty} \sum_{n=1}^{N} \frac{8n^{2/3}\sin^2 ny}{(\mu_N^{1/4} n^{2/3} + \mu_N^{-3/4} n^{4/3})^2} = 3\pi.$$

The proof of Lemma 2.5 will be deferred to Section 3. We now complete the proof of Theorem 1.1.

Proof of Theorem 1.1. Fix $y \in (0, \pi)$. For any $N \geq 2$, consider

$$u(y) = \sum_{n=1}^{N} f_n \phi_n(y), \quad \phi_n(y) = \sqrt{\frac{2}{\pi}} \sin ny.$$

Clearly,

$$\|(-\Delta_D)^{1/6} u\|_{L^2(\Omega)} = \left(\sum_{n=1}^{N} f_n^2 n^{2/3}\right)^{1/2},$$

$$\|(-\Delta_D)^{1/3} u\|_{L^2(\Omega)} = \left(\sum_{n=1}^{N} f_n^2 n^{4/3}\right)^{1/2}.$$

Then

$$\frac{|u(y)|^2}{\|(-\Delta_D)^{1/6} u\|_{L^2(\Omega)} \|(-\Delta_D)^{1/3} u\|_{L^2(\Omega)}} = \frac{\left|\sum_{n=1}^{N} f_n\phi_n(y)\right|^2}{\left(\sum_{n=1}^{N} f_n^2 n^{2/3}\right)^{1/2} \left(\sum_{n=1}^{N} f_n^2 n^{4/3}\right)^{1/2}} =: R_N.$$

By Lemma 2.1 with $f_n = b_n$, $\phi_n(y) = a_n$, $v_n = n^{2/3}$, $w_n = n^{4/3}$, we have

$$\max_{0\neq(f_1,\ldots,f_N)\in\mathbb{R}^N} R_N = \sum_{n=1}^{N} \frac{4\phi_n^2(y) n^{2/3}}{(\mu_N^{1/4} n^{2/3} + \mu_N^{-3/4} n^{4/3})^2},$$

where $\mu_N > 0$ solves

$$\sum_{n=1}^{N} \frac{4\phi_n^2(y)}{(n^{2/3} + \mu_N^{-1} n^{4/3})^2} (n^{2/3}\mu_N - n^{4/3}) = 0.$$

The result then follows from Lemma 2.5. Note that the proof here shows that the constant 3 is sharp since the minimizer for each R_N exists. □

Proof of (1-3) *in Theorem 1.3.* Note that for any $f = \sum_{n=1}^N f_n \cos nx$,

$$\|f\|_{L^\infty(\Omega)} \le \sum_{n=1}^{\infty} |f_n|,$$

$$\|f^{(1/3)}\|_{L^2(\Omega)} = \sqrt{\pi}\left(\sum_{n=1}^{N} f_n^2 n^{2/3}\right)^{1/2},$$

$$\|f^{(2/3)}\|_{L^2(\Omega)} = \sqrt{\pi}\left(\sum_{n=1}^{N} f_n^2 n^{4/3}\right)^{1/2}.$$

We then only need to show the discrete inequality

$$\left(\sum_{n=1}^{N} |f_n|\right)^2 \le 3\pi\left(\sum_{n=1}^{N} f_n^2 n^{2/3}\right)\left(\sum_{n=1}^{N} f_n^2 n^{4/3}\right).$$

By Lemma 2.1, it suffices to prove

$$\lim_{N\to\infty} \sum_{n=1}^{N} \frac{4n^{2/3}}{(\mu_N^{1/4} n^{2/3} + \mu_N^{-3/4} n^{4/3})^2} = 3\pi,$$

where $\mu = \mu_N > 0$ solves

$$\sum_{n=1}^{N} 4/(n^{2/3} + n^{4/3}/\mu)^2 (n^{2/3}\mu - n^{4/3}) = 0.$$

By Lemma 5.4 and an argument similar to that in Lemma 3.2, $\lim_{N\to\infty} \mu_N = \infty$. By Lemma 3.5, we also have $\lim_{N\to\infty} \mu_N / N^{2/3} = 0$. It follows that

$$\begin{aligned}
\lim_{N\to\infty} \sum_{n=1}^{N} \frac{4n^{2/3}}{(\mu_N^{1/4} n^{2/3} + \mu_N^{-3/4} n^{4/3})^2} &= \lim_{\mu\to\infty} \sum_{n=1}^{\infty} \frac{4}{(\mu^{1/4} n^{1/3} + \mu^{-3/4} n)^2} \\
&= \lim_{\mu\to\infty} 4 \sum_{n=1}^{\infty} \left(\left(\frac{n}{\mu^{3/2}}\right)^{1/3} + \frac{n}{\mu^{3/2}}\right)^{-2} \cdot \mu^{-3/2} \\
&= 4 \int_0^\infty (x^{1/3} + x)^{-2}\, dx = 3\pi
\end{aligned}$$

□

Proof of (1-4) *in Theorem 1.3.* This is similar to the proof of (1-3). We omit details. □

3. Proof of Lemma 2.5

The proof of statement (1) in Lemma 2.5 can be reduced to the following proposition.

Proposition 3.1. *For any* $0 < y < \pi,\ \ 0 < \mu < \infty$, *we have*

$$\sum_{n=1}^{\infty} \frac{\sin^2 ny}{(\mu + n^{2/3})^2} \left(\frac{\mu}{n^{2/3}} - 1 \right) < 0.$$

Or equivalently, *for any* $0 < t < 1,\ \ 0 < y \le \pi$,

$$\sum_{n=1}^{\infty} \frac{(nt)^{-2/3} - 1}{(1 + (nt)^{2/3})^2} (1 - \cos ny) < 0.$$

The proof of Proposition 3.1 will be deferred to later sections. In this section we shall assume it holds true and then complete the proof of Lemma 2.5. We shall assume the same notation as in Lemma 2.5.

Lemma 3.2 (lower estimate of μ_N). *We have*

$$\lim_{N \to \infty} \mu_N = \infty.$$

Proof of Lemma 3.2. Suppose for some $N_j \to \infty$ we have $\lim_{j \to \infty} \mu_{N_j} = \mu_* < \infty$. Since

$$\sum_{n=1}^{N_j} \frac{\sin^2 ny}{(\mu_{N_j} + n^{2/3})^2} \left(\frac{\mu_{N_j}}{n^{2/3}} - 1 \right) = 0,$$

taking the limit $j \to \infty$ then yields

$$\sum_{n=1}^{\infty} \frac{\sin^2 ny}{(\mu_* + n^{2/3})^2} \left(\frac{\mu_*}{n^{2/3}} - 1 \right) = 0.$$

Now clearly $\mu_* \neq 0$. This is then a contradiction to Proposition 3.1. Thus we must have $\lim_{N \to \infty} \mu_N = \infty$. □

Next we shall assume statement (2) in Lemma 2.5 holds and show how to prove statement (3). For this we need a simple formula.

Lemma 3.3. *Let* $0 < \theta < 2\pi$. *Suppose* $\{a_n\}_{n=1}^{\infty}$ *is a sequence of complex numbers such that*

$$\sum_{n=1}^{\infty} |a_n - a_{n+1}| < \infty \quad \textit{and} \quad \lim_{n \to \infty} a_n = 0.$$

Then

$$\sum_{n=1}^{\infty} a_n e^{in\theta} = \sum_{n=1}^{\infty} e^{i(n+1)/(2)\theta} \cdot \frac{\sin \frac{n\theta}{2}}{\sin \frac{\theta}{2}} (a_n - a_{n+1}), \tag{3-1}$$

and

$$\sum_{n=1}^{\infty} a_n \cos n\theta = \sum_{n=1}^{\infty}\left(-\frac{1}{2}+\frac{\sin\left(n+\frac{1}{2}\right)\theta}{2\sin\frac{\theta}{2}}\right)(a_n-a_{n+1}).$$

In particular, if in addition a_n is positive and monotonically decreasing (or negative and monotonically increasing), then

$$\left|\sum_{n=1}^{\infty} a_n e^{in\theta}\right| \le \frac{|a_1|}{\sin\frac{\theta}{2}}.$$

Proof. Recall the usual angle formula

$$S_N=\sum_{n=1}^{N} e^{in\theta}=\frac{e^{i\theta}(1-e^{in\theta})}{1-e^{i\theta}}=e^{i(n+1)/(2)\theta}\cdot\frac{\sin\frac{n\theta}{2}}{\sin\frac{\theta}{2}}.$$

Note that

$$|S_N|\le\frac{1}{\sin\frac{\theta}{2}}$$

so that the series on the RHS of (3-1) converges absolutely. Denote $S_0=0$. Then

$$\sum_{n=1}^{N} a_n(S_n-S_{n-1})=\sum_{n=1}^{N} a_nS_n-\sum_{n=1}^{N-1} a_{n+1}S_n \tag{3-2}$$

$$=\sum_{n=1}^{N-1} S_n(a_n-a_{n+1})+a_NS_N.$$

Sending $N\to\infty$ then yields (3-1). The last estimate is obvious. □

We now complete the proof of statement (3) in Lemma 2.5.

Proof of Lemma 2.5(3). Since we assume $\lim_{N\to\infty}\mu_N/N^{2/3}=0$, we get

$$\sum_{n>N}\frac{8n^{2/3}\sin^2 ny}{(\mu_N^{1/4}n^{2/3}+\mu_N^{-3/4}n^{4/3})^2}\le\mu_N^{3/2}\sum_{n>N}\frac{8}{n^2}\le\text{const}\cdot\mu_N^{3/2}\cdot N^{-1}\to 0,\quad\text{as } N\to\infty.$$

Then (since $\lim_{N\to\infty}\mu_N=\infty$) it suffices to prove

$$\lim_{\mu\to\infty}\sum_{n=1}^{\infty}\frac{8n^{2/3}\sin^2 ny}{(\mu^{1/4}n^{2/3}+\mu^{-3/4}n^{4/3})^2}=3\pi.$$

Write $2\sin^2 ny=1-\cos 2ny$. By Lemma 3.3, we have

$$\left|\sum_{n=1}^{\infty}\frac{4}{(\mu^{1/4}n^{1/3}+\mu^{-3/4}n)^2}\cos 2ny\right|\le\frac{4\mu^{-1/2}}{\sin y}\to 0\quad\text{as }\mu\to\infty.$$

On the other hand,

$$\sum_{n=1}^{\infty}\frac{4}{(\mu^{1/4}n^{1/3}+\mu^{-3/4}n)^2}=4\sum_{n=1}^{\infty}\left(\left(\frac{n}{\mu^{3/2}}\right)^{1/3}+\frac{n}{\mu^{3/2}}\right)^{-2}\cdot\mu^{-3/2}$$
$$\to 4\int_0^{\infty}(x^{1/3}+x)^{-2}\,dx=3\pi \quad \text{as } \mu\to\infty. \qquad \square$$

The rest of this section is devoted to the proof of statement (2) of Lemma 2.5. Denote $t_N=\mu_N^{-3/2}$. Clearly $t=t_N$ is a solution to the equation

$$\sum_{n=1}^{N}\frac{\sin^2 ny}{(1+(nt)^{2/3})^2}((nt)^{-2/3}-1)=0.$$

To complete the proof of statement (2), it suffices to prove the following lemma.

Lemma 3.4 (upper estimate of $\mu_N=t_N^{-2/3}$). *Let $f(z)=(z^{-2/3}-1)/(1+z^{2/3})^2$ and fix $y\in(0,\pi)$. Suppose $t=t_N>0$ is a solution to the equation*

$$\sum_{n=1}^{N}f(nt)\sin^2 ny=0.$$

Then

$$\lim_{N\to\infty}\frac{1}{Nt_N}=0.$$

Proof of Lemma 3.4. It suffices for us to show that for any integer $N_0\geq 100$, there exists N_1 sufficiently large, such that if $m\geq N_1$ and $mt\leq N_0$, then

$$\sum_{n=1}^{m}f(nt)\sin^2 ny>0.$$

Now note that $\sin^2 ny=(1-\cos n\tilde{y})/2$, $\tilde{y}=2y\in(0,2\pi)$. In view of the symmetry of $\cos n\tilde{y}$ under the change of variable $\tilde{y}\to 2\pi-\tilde{y}$, we may assume $\tilde{y}\in(0,\pi]$. Furthermore we will drop the tilde and consider only the expression

$$S_m=\sum_{n=1}^{m}f(nt)(1-\cos ny), \qquad y\in(0,\pi].$$

Since $f(z)>0$ for $0<z\leq 1$, we may assume $mt>1$. Note $t\leq N_0/m\leq N_0/N_1\to 0$ if $N_1\to\infty$. We shall think of t as the main parameter. Using the fact that $f(nt)<0$ if $n>1/t$, we have

$$S_m\geq\sum_{n\leq N_0/t}f(nt)(1-\cos ny)=\sum_{n\leq N_0/t}f(nt)-\sum_{n\leq N_0/t}f(nt)\cos ny.$$

Now by (3-2) and denoting $m_0=[N_0/t]$ ($[x]$ denotes the integer part of x),

$$\sum_{n\le N_0/t} f(nt)\cos ny$$
$$=\sum_{n\le m_0-1}\left(-\frac{1}{2}+\frac{\sin\left(n+\frac{1}{2}\right)y}{2\sin\frac{1}{2}y}\right)(f(nt)-f((n+1)t))+f(m_0t)\left(-\frac{1}{2}+\frac{\sin\left(m_0+\frac{1}{2}\right)y}{2\sin\frac{1}{2}y}\right).$$

Thus

$$\left|\sum_{n\le N_0/t} f(nt)\cos ny\right|$$
$$\le\left(\frac{1}{2}+\frac{1}{2\sin\frac{1}{2}y}\right)\sum_{n\le m_0-1}|f(nt)-f((n+1)t)|+\left(\frac{1}{2}+\frac{1}{2\sin\frac{1}{2}y}\right)|f(m_0t)|.$$

Note that $F(x)=f(x)-x^{-2/3}=-(1+x^{2/3})^{-2}(3+x^{2/3})$ is monotonically increasing on $(0,\infty)$. Then

$$\sum_{n\le m_0-1}|f(nt)-f((n+1)t)|$$
$$\le\sum_{n\le m_0-1}(F((n+1)t)-F(nt))+\sum_{n\le m_0-1}((nt)^{-2/3}-((n+1)t)^{-2/3})$$
$$\le F(m_0t)-F(t)+t^{-2/3}-(m_0t)^{-2/3}\le 3\cdot(1+t^{-2/3}).$$

Thus

$$\left|\sum_{n\le N_0/t} f(nt)\cos ny\right|\le C_2\cdot(1+t^{-2/3}),$$

where $C_2>0$ depends only on y.

Now let $\epsilon_0>0$ be sufficiently small (depending on N_0) such that

$$\int_0^{\epsilon_0} f(z)\,dz<-\frac{1}{2}\int_{N_0}^{\infty} f(z)\,dz.$$

It follows easily that

$$S_m\ge\sum_{\epsilon_0/t\le n\le N_0/t} f(nt)-C_2\cdot(1+t^{-2/3})$$
$$\ge\frac{1}{t}\left(\int_{\epsilon_0}^{N_0} f(z)\,dz-C_3t\right)-C_2\cdot(1+t^{-2/3}),$$

where $C_3>0$ depends only on (ϵ_0,N_0). Here we used a standard estimate on the Riemann sum of the function on the interval $[\epsilon_0,N_0]$ with mesh size t.

Since $\int_0^\infty f(z)\,dz = 0$, we then have

$$\begin{aligned}\int_{\epsilon_0}^{N_0} f(z)\,dz &= -\int_{N_0}^{\infty} f(z)\,dz - \int_0^{\epsilon_0} f(z)\,dz \\ &> -\frac{1}{2}\int_{N_0}^{\infty} f(z)\,dz =: C_{N_0} > 0.\end{aligned}$$

Then

$$S_m \geq \frac{C_{N_0}}{t} - C_3 - C_2 \cdot (1 + t^{-2/3}) > 0,$$

if t is sufficiently small. The lemma is proved. □

There is a version of Lemma 3.4 needed for the cosine series case. We record it here and sketch the proof.

Lemma 3.5. *For each integer $m \geq 10$ suppose $\mu = \mu_m > 0$ is a solution to the equation*

$$\sum_{n=1}^{m} \frac{1}{(\mu + n^{2/3})^2}\left(\frac{\mu}{n^{2/3}} - 1\right) = 0.$$

Then

$$\lim_{m\to\infty} \frac{\mu_m}{m^{2/3}} = 0.$$

Proof of Lemma 3.5. Denote $t_m = \mu_m^{-3/2}$. Then $t = t_m$ solves

$$\sum_{n=1}^{m} f(nt) = 0, \qquad f(z) = \frac{z^{-2/3} - 1}{(1 + z^{2/3})^2}$$

and we need to show

$$\lim_{m\to\infty} \frac{1}{mt_m} = 0.$$

It suffices to show that for any integer $N_0 \geq 100$, there exists N_1 sufficiently large, such that if $m \geq N_1$ and $mt \leq N_0$, then

$$S_m = \sum_{n=1}^{m} f(nt) > 0.$$

Since $f(z) > 0$ for $0 < z < 1$, we may assume $mt > 1$. Then similar to the proof in Lemma 3.4, we have

$$S_m \geq \sum_{n \leq N_0/t} f(nt) > \frac{C_{N_0}}{t} - D_{N_0},$$

where $C_{N_0} > 0$, $D_{N_0} > 0$ are constants depending only on N_0. Taking t sufficiently small then yields the result. □

4. Continuous analogue of Proposition 3.1

In this section we prove a continuous analogue of Proposition 3.1. It is formulated as the following theorem.

Theorem 4.1. *For any $y > 0$ and $0 < \mu < \infty$, we have*

$$\tilde{I} = \int_0^\infty \frac{\sin^2(xy)}{(\mu + x^{2/3})^2}\left(1 - \frac{\mu}{x^{2/3}}\right) dx \geq 0.01\mu^{-1/2} \min\{\mu^{-1/2} y^{-1/3}, \mu^3 y^2\} > 0.$$

Proof of Theorem 4.1. By a change of variable $x = \mu^{3/2}\tilde{x}$, we have

$$\tilde{I} = -\mu^{-1/2} \int_0^\infty \sin^2(\mu^{3/2} y\tilde{x}) f(\tilde{x})\, d\tilde{x},$$

where

$$f(x) = (x^{-2/3} - 1)/(1 + x^{2/3})^2.$$

It is easy to check (by making a change of variable and using contour integrals) that

$$\int_0^\infty f(\tilde{x})\, d\tilde{x} = 0.$$

Then

$$\tilde{I} = \mu^{-1/2} \cdot \frac{1}{2} \int_0^\infty \cos(2\mu^{3/2} y\tilde{x}) f(\tilde{x})\, d\tilde{x}.$$

The result then follows from Theorem 4.2. □

Theorem 4.2. *For any $\mu > 0$, we have*

$$I = \int_0^\infty \cos(\mu x) f(x)\, dx \geq 0.03 \min\{\mu^{-1/3}, \mu^2\} > 0,$$

where $f(x) = (x^{-2/3} - 1)/(1 + x^{2/3})^2$.

Proof of Theorem 4.2. We proceed in a few steps.

Step 1: Let $C_1 = \pi/\Gamma(2/3) \approx 2.32003$ ($\Gamma(\cdot)$ is the usual Gamma function). Denote

$$\mu_0 = \left(\frac{3 - C_1}{C_1}\right)^{3/2} \approx 0.15867.$$

We first show $I \geq 0.03\mu^{-1/3}$ for $\mu > 1.1\mu_0$. With $f(x) - x^{-2/3} := F(x)$, write

$$I = \int_0^\infty \cos(\mu x) x^{-2/3}\, dx + \int_0^\infty \cos(\mu x)(f(x) - x^{-2/3})\, dx,$$

It is easy to check that

$$F(x) = -(1+x^{2/3})^{-2}(3+x^{2/3}), \tag{4-1}$$

$$F'(x) = \tfrac{2}{3}x^{-1/3}(1+x^{2/3})^{-3}(5+x^{2/3}), \tag{4-2}$$

$$\int_0^\infty \cos(\mu x) x^{-2/3}\,dx = C_1 \cdot \mu^{-1/3}.$$

Then

$$I = C_1 \mu^{-1/3} - \frac{1}{\mu}\int_0^\infty \sin(\mu x) F'(x)\,dx.$$

Using the inequality $|\sin z| \le \min\{|z|, 1\}$, we have (since $F' > 0$)

$$\begin{aligned}\left|\int_0^\infty \sin(\mu x) F'(x)\,dx\right| &\le \int_0^{1/\mu} \mu x F'(x)\,dx + \int_{1/\mu}^\infty F'(x)\,dx \\ &= -\mu \int_0^{1/\mu} F(x)\,dx = 3\mu^{2/3}(1+\mu^{2/3})^{-1}.\end{aligned}$$

It follows that

$$\begin{aligned} I &\ge C_1 \mu^{-1/3} - 3\mu^{-1/3}(1+\mu^{2/3})^{-1} \\ &\ge \mu^{-1/3}(C_1 - 3(1+(1.1\mu_0)^{2/3})^{-1}) \ge 0.03\mu^{-1/3}.\end{aligned}$$

Step 2: Let $0 < \mu \le 1.1\mu_0$. Recall that $f(x) < 0$ when $x > 1$. Then

$$\begin{aligned} I &= \int_0^2 \cos(\mu x) f(x)\,dx + \int_2^\infty \cos(\mu x) f(x)\,dx \\ &\ge \int_0^2 (\cos(\mu x) - 1) f(x)\,dx + \int_0^2 f(x)\,dx - \int_2^\infty |f(x)|\,dx \\ &= \int_0^2 (\cos(\mu x) - 1) f(x)\,dx + \int_0^\infty f(x)\,dx \\ &= \int_0^2 (\cos(\mu x) - 1) f(x)\,dx, \end{aligned}$$

where in the last step above we used the fact $\int_0^\infty f(x)\,dx = 0$. Now we use the inequality

$$\left|\cos z - 1 + \tfrac{1}{2}z^2\right| \le \tfrac{1}{24}z^4.$$

Then

$$I \ge -\mu^2 \frac{1}{2}\int_0^2 x^2 f(x)\,dx - \frac{\mu^4}{24}\int_0^2 x^4 |f(x)|\,dx.$$

One can explicitly compute

$$\frac{1}{24}\int_0^2 x^4|f(x)|\,dx = \frac{8697-3212\cdot 2^{1/3}-952\cdot 2^{2/3}-2310\pi+4620\arctan(2^{1/3})}{3080} \approx 0.0128004;$$

$$-\frac{1}{2}\int_0^2 x^2 f(x)\,dx = \tfrac{3}{5}(-6+13\cdot 2^{1/3}+2\cdot 2^{2/3}-15\arctan(2^{1/3})) \approx 0.0330903.$$

Then

$$\geq \mu^2(0.03309-0.012801\mu^2)I \geq \mu^2(0.03309-0.012801(1.1\mu_0)^2)I \geq 0.03\mu^2. \qquad \square$$

5. Proof of Proposition 3.1

5A. ***The case where $3^{-\frac{3}{2}} \approx 0.19245 \leq t < 1$.*** Write

$$S = f(t)(1-\cos y) + \sum_{n=2}^{\infty} f(nt)(1-\cos ny) = f(t)\Bigl(1-\cos y + \sum_{n=2}^{\infty}\frac{f(nt)}{f(t)}(1-\cos ny)\Bigr),$$

where $f(x) = (x^{-2/3}-1)/(1+x^{2/3})^2$. We first need a lemma.

Lemma 5.1 (monotonicity of coefficients). *For any $n \geq 2$, $f(nt)/f(t)$ is monotonically decreasing in t for $3^{-3/2} \leq t < 1$.*

Proof of Lemma 5.1. Write $f(x) = x^{-2/3}g(x^{2/3})$, where $g(x) = (1-x)/(1+x)^2$. Then

$$\frac{f(nt)}{f(t)} = n^{-2/3}\cdot\frac{g(n^{2/3}t^{2/3})}{g(t^{2/3})}.$$

Denote $\lambda = n^{2/3} \geq 2^{2/3}$. It suffices for us to show

$$\tilde{g}(x) = g(\lambda x)/g(x)$$

is decreasing for $\frac{1}{3} \leq x < 1$. It is easy to check that

$$\tilde{g}'(x) = -\frac{(\lambda-1)(1+x)(3\lambda x^2-(1+\lambda)x+3)}{(1-x)^2(1+\lambda x)^2}.$$

Clearly for $\frac{1}{3} \leq x < 1$,

$$3\lambda x^2-(1+\lambda)x+3 = \lambda x(3x-1)+3-x > 0. \qquad \square$$

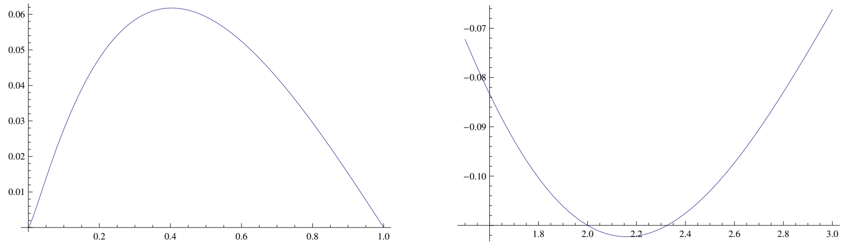


Figure 1. $g(x)$ for $0 \le x \le 1$ and $1.5 \le x \le 3$.

Denote

$$\tilde{S}(t, y) = 1 - \cos y + \sum_{n=2}^{\infty} \frac{f(nt)}{f(t)}(1 - \cos ny).$$

By Lemma 5.1, we get for $3^{-3/2} \le t < 1$,

$$\tilde{S}(t, y) \le \tilde{S}(3^{-3/2}, y).$$

Note that $\tilde{S}(t, y)$ has the same sign as S for all $0 < t < 1$. Thus it suffices to prove $S < 0$ for $0 < t \le 3^{-3/2} \approx 0.19245$. This will be achieved in the next steps.

5B. ***The case where $0 < t \le 3^{-\frac{3}{2}}$ and $y/t \le 1.5$.*** Denote $\epsilon = y/t \le \frac{3}{2}$. It suffices for us to show the inequality

$$\sum_{n \le 3/t} f(nt)(1 - \cos(\epsilon nt)) < 0, \qquad \text{for all } 0 < t \le 3^{-3/2},$$

where we recall $f(x) = (x^{-2/3} - 1)/(1 + x^{2/3})^2$.

Lemma 5.2. *Denote* $g(x) = f(x)\big(1 - \cos \frac{3}{2}x\big)$. *Then*

$$\max_{0 \le x \le 1} g(x) < 0.062, \qquad \max_{1.5 \le x \le 3} g(x) = g(3) < -0.066.$$

It follows that for any $0 < x_1 \le 1$ *and* $1.5 \le x_2 \le 3$, *we have*

$$f(x_1)\big(1 - \cos \tfrac{3}{2}x_1\big) + f(x_2)\big(1 - \cos \tfrac{3}{2}x_2\big) < 0.$$

Proof of Lemma 5.2. By the plot of $g(x)$ in Figure 1, it is clear that the maximum of $g(x)$ on $[0, 1]$ occurs near $0.3 \le x \le 0.6$, and the maximum of $g(x)$ on $[1.5, 3]$ occurs at $x = 3$. One can then easily check the maximum of $g(x)$ is bounded by 0.062 and -0.066 respectively. The whole argument is an easy exercise of numerical analysis together with routine numerical computation. We omit details. □

Lemma 5.3. *For any* $0 < x_1 \le 1$, $1.5 \le x_2 \le 3$, *we have*

$$I(\epsilon) = f(x_1)(1 - \cos \epsilon x_1) + f(x_2)(1 - \cos \epsilon x_2) < 0, \quad \textit{for all } 0 < \epsilon \le \tfrac{3}{2}.$$

Proof of Lemma 5.3. The idea is to show that the sign of $I(\epsilon)$ is dictated by $I(\epsilon = 3/2)$. Note that

$$\frac{1-\cos\epsilon x_2}{1-\cos\epsilon x_1} = \left(\frac{\sin\frac{1}{2}\epsilon x_2}{\sin\frac{1}{2}\epsilon x_1}\right)^2.$$

Consider the ratio $\sin\frac{1}{2}\epsilon x_2/\sin\frac{1}{2}\epsilon x_1$. Note that $\frac{1}{2}\cdot\frac{3}{2}\cdot 3 = \frac{9}{4} < \pi$ so that $\sin\frac{1}{2}\epsilon x_2$ is positive. Denote $b = x_2/x_1 > 1$, $\eta = \frac{1}{2}\epsilon x_1$ and note that $b\eta \le \frac{9}{4}$. Consider

$$h(\eta) = \frac{\sin b\eta}{\sin\eta}.$$

Since $f(x_2) < 0$ for $1.5 \le x_2 \le 3$, it suffices for us to show that $h(\eta)$ is monotonically decreasing in η. Clearly

$$h'(\eta) = \frac{1}{(\sin\eta)^2}(b\cos b\eta\sin\eta - \sin b\eta\cos\eta).$$

Note that if $\frac{9}{4} \ge b\eta \ge \pi/2$ we have $h' < 0$. Then it remains for us to check the inequality for $b\eta < \pi/2$ and $0 < \eta \le \frac{3}{4}$,

$$g(\eta) = \tan b\eta - b\tan\eta > 0.$$

It is easy to compute that

$$g'(\eta) = b(\sec^2 b\eta - \sec^2\eta) > 0.$$

Here we used the fact that $0 < \cos(b\eta) < \cos(\eta)$. Since $g(0) = 0$, clearly then $g(\eta) > 0$ for $0 < \eta < \min\{\pi/(2b), 3/4\}$. □

By Lemma 5.3, we have

$$\sum_{n\le 3/t} f(nt)(1-\cos\epsilon nt) < \sum_{n<1/t} f(nt)(1-\cos\epsilon nt) + \sum_{1.5/t\le n\le 3/t} f(nt)(1-\cos\epsilon nt) < 0,$$

since the first sum contains at most $[1/t]$ terms and the second sum contains obviously $\ge [1.5/t]$ terms.

5C. ***The case where $0 < t \le 3^{-\frac{3}{2}}$ and $y/t \ge 1.5$, reduction to an inequality.*** In this section, we shall show that to finish the proof of Proposition 3.1 for the case $0 < t \le 3^{-3/2}$ and $y/t \ge 1.5$, it suffices to prove the following inequality: for any $1.5t \le y \le \pi$, $0 < x = t^{2/3} \le 1/3$,

$$\frac{3}{1+x} - 1 + G_0(y) - \left(\frac{1}{2\sin\frac{1}{2}y} - \frac{1}{2}\right)\cdot\frac{3x+x^2}{(1+x)^2} > 0, \tag{5-1}$$

where

$$G_0(y) = \sum_{n=1}^{\infty} n^{-2/3}\cos ny. \tag{5-2}$$

We begin with an easy lemma.

Lemma 5.4. *Recall $f(x) = (x^{-2/3} - 1)/(1 + x^{2/3})^2$. For any $t > 0$, we have*

$$S = \sum_{n=1}^{\infty} f(nt) < t^{-2/3}\left(1 - \frac{3}{1 + t^{2/3}}\right).$$

In particular, $S < 0$ for all $0 < t < \infty$.

Proof of Lemma 5.4. Using the fact that $\int_0^\infty f(x)\,dx = 0$ and $(f(x) - x^{-2/3})' > 0$, we have

$$\begin{aligned}
tS &= \sum_{n=1}^{\infty} \int_{(n-1)t}^{nt} (x-(n-1)t) f'(x)\,dx \\
&= \int_0^t x f'(x)\,dx + \sum_{n=2}^{\infty} \int_{(n-1)t}^{nt} (x-nt) f'(x)\,dx + t\int_t^\infty f'(x)\,dx \\
&= -\int_0^t f(x)\,dx + \sum_{n=2}^{\infty} \int_{(n-1)t}^{nt} (x-nt)(f(x)-x^{-2/3})'\,dx \\
&\qquad\qquad + \sum_{n=2}^{\infty} \int_{(n-1)t}^{nt} (x-nt)(x^{-2/3})'\,dx \\
&\le -\int_0^t f(x)\,dx + t^{1/3} = -\frac{3t^{1/3}}{1+t^{2/3}} + t^{1/3}.
\end{aligned}$$

Now observe that if $t > 1$, then $f(nt) < 0$ for any $n \ge 1$ which implies $S < 0$. For $0 < t \le 1$, we clearly have $1-3/(1+t^{2/3}) < 0$. Thus $S < 0$ in all cases. □

Lemma 5.5. $\sum_{n=1}^{\infty} f(nt)\cos ny \ge t^{-2/3} G_0(y) + \left(-\frac{1}{2} + \frac{1}{2\sin\frac{1}{2}y}\right) F(t),$

where $F(t) = f(t) - t^{-2/3} = -(1+t^{2/3})^{-2}(3+t^{2/3})$ and $G_0(y)$ was defined in (5-2).

Proof of Lemma 5.5. Note that for $\lim_{x\to\infty} F(x) = 0$ and for any $t > 0$,

$$\sum_{n=1}^{\infty} |F(nt) - F((n+1)t)| < \infty.$$

By Lemma 3.3, we have

$$\begin{aligned}
\sum_{n=1}^{\infty} f(nt)\cos ny &= \sum_{n=1}^{\infty} (nt)^{-2/3} \cos ny + \sum_{n=1}^{\infty} F(nt)\cos ny \\
&= t^{-2/3} G_0(y) + \sum_{n=1}^{\infty} \left(-\frac{1}{2} + \frac{\sin\left(n+\frac{1}{2}\right)y}{2\sin\frac{1}{2}y}\right)(F(nt) - F((n+1)t)).
\end{aligned}$$

Clearly

$$\sum_{n=1}^{\infty}\left(-\tfrac{1}{2}\right)(F(nt)-F((n+1)t)) = -\tfrac{1}{2}F(t).$$

On the other hand, since $F(x) = f(x) - x^{-2/3}$ is negative and monotonically increasing on $(0, \infty)$, we have

$$\sum_{n=1}^{\infty}\left|\frac{\sin(n+\frac{1}{2})y}{2\sin\frac{1}{2}y}\right| |F(nt)-F((n+1)t)| \\ \leq \frac{1}{2\sin\frac{1}{2}y}\sum_{n=1}^{\infty}(F((n+1)t)-F(nt)) = -\frac{F(t)}{2\sin\frac{1}{2}y}.$$

Clearly then

$$\sum_{n=1}^{\infty} f(nt)\cos ny \geq t^{-2/3}G_0(y) + \left(-\frac{1}{2}+\frac{1}{2\sin\frac{1}{2}y}\right)F(t). \qquad \square$$

By Lemma 5.4 and Lemma 5.5, we get

$$\sum_{n=1}^{\infty} f(nt)(1-\cos ny) \\ < t^{-2/3}\left(1-\frac{3}{1+t^{2/3}}\right) - t^{-2/3}G_0(y) - \left(\frac{1}{2\sin\frac{1}{2}y}-\frac{1}{2}\right)\cdot(-1)\cdot\frac{3+t^{2/3}}{(1+t^{2/3})^2}.$$

It is then clear that we only need to prove (5-1).

6. Proof of (5-1)

6A. ***The case where*** $\boldsymbol{0.208 \leq y \leq \pi}$**.** We first establish some properties of the polygamma function $G_0(y)$.

Lemma 6.1. *Recall* $G_0(y) = \sum_{n=1}^{\infty} n^{-2/3}\cos ny$. *We have for* $0 < y \leq \pi$,

$$G_0(y) = g_0(\cos y),$$

where for $-1 \leq x < 1$,

$$g_0(x) = \frac{1}{\Gamma\left(\frac{2}{3}\right)}\int_0^{\infty} t^{-1/3}\cdot\frac{e^t x - 1}{e^{2t}-2e^t x + 1}\,dt \\ = \frac{1}{\Gamma\left(\frac{2}{3}\right)}\int_0^{\infty} t^{-1/3}\cdot\frac{1}{2}\left(-1+\frac{\sinh t}{\cosh t - x}\right)dt.$$

In particular $g_0'(x) > 0$ *for* $-1 \leq x < 1$ *and* $G_0(y)$ *is monotonically decreasing for* $0 < y \leq \pi$.

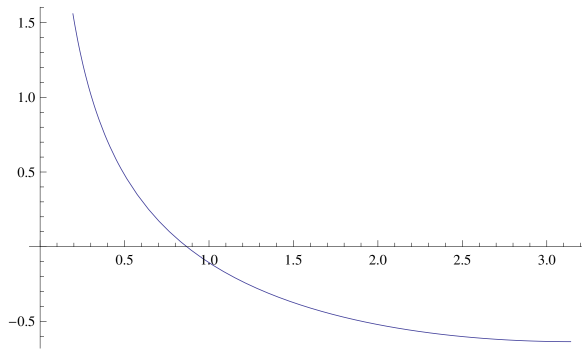


Figure 2. Plot of $G_0(y)$.

Proof of Lemma 6.1. By using the identity

$$n^{-s} = \frac{1}{\Gamma(s)} \int_0^\infty e^{-tn} t^{s-1}\,dt, \quad s > 0,$$

we get, for $z = e^{i\alpha}$ with $0 < \alpha \le \pi$,

$$\sum_{n=1}^{N} n^{-s} z^n = \frac{1}{\Gamma(s)} \int_0^\infty \frac{ze^{-t} - z^{N+1} e^{-(N+1)t}}{1 - ze^{-t}} t^{s-1}\,dt.$$

Since $|z| = 1$ and $z \neq 1$, it is easy to check that

$$\lim_{N\to\infty} z^{N+1} \int_0^\infty \frac{e^{-(N+1)t}}{1 - ze^{-t}} t^{s-1}\,dt = 0.$$

We then obtain

$$\sum_{n=1}^{\infty} n^{-s} z^n = \frac{1}{\Gamma(s)} \int_0^\infty \frac{1}{e^t/z - 1} t^{s-1}\,dt.$$

The desired identity then follows from

$$\mathrm{Re}\left(\frac{1}{e^t e^{-iy} - 1}\right) = \frac{e^t \cos y - 1}{e^{2t} - 2e^t \cos y + 1}.$$

To show $g_0'(x) > 0$ it suffices to consider the integrand

$$h(t,x) = \frac{\sinh t}{\cosh t - x}.$$

Note that $h(t,x)$ has no singularities since $-1 \le x < 1$ and $t > 0$. Clearly $\partial/\partial x\, h(t,x) > 0$ which implies $g_0'(x) > 0$. $\square$

Lemma 6.2. *For any* $0.208 \le y \le \pi$, *we have*

$$\tilde{S}(y) = \tfrac{25}{16} + G_0(y) - \tfrac{5}{16} \cdot \frac{1}{\sin \frac{1}{2} y} > 0.$$

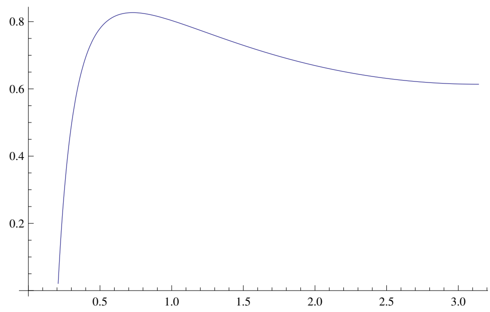


Figure 3. Plot of $\tilde{S}(y)$.

Proof of Lemma 6.2. Although from Figure 3, it is transparent that $\tilde{S}(y) > 0$, we shall still give a more "analytic" proof. (Since $G_0(y)$ is not an elementary function we try to reduce the numerical evaluation of $G_0(y)$ to as few sample points as possible).

By Lemma 6.1, $G_0(y)$ is monotonically decreasing on $(0, \pi]$. Observe that $1/(\sin(1/2)y)$ is also monotonically decreasing on $(0, \pi]$. Then clearly for any $0 < h \le \pi$,

$$\min_{h/2 \le y \le h} \tilde{S}(y) \ge \tfrac{25}{16} + G_0(h) - \tfrac{5}{16} \cdot \frac{1}{\sin \frac{1}{4}h} =: \tilde{G}(h).$$

We first show that $\tilde{S}(y) > 0$ for $\pi/8 \le y \le \pi$. By explicit numerical evaluations, we have

$$\tilde{G}(\pi) \approx 0.484381 > 0, \quad \tilde{G}(\pi/2) \approx 0.345131 > 0, \quad \tilde{G}(\pi/4) \approx 0.0401163 > 0.$$

Thus $\tilde{S}(y) > 0$ on $[\pi/8, \pi]$.

Next we consider the regime $0.208 \le y \le \pi/8 \approx 0.392699$. Again by monotonicity, we have for $0.208 \le b_1 \le y \le b_2 \le \pi/8$,

$$\tilde{S}(y) \ge \tfrac{25}{16} + G_0(b_2) - \tfrac{5}{16} \cdot \frac{1}{\sin \frac{1}{2}b_1}.$$

By iteratively using the above inequality, we get:

If $0.3 \le y \le \pi/8$, then $\tilde{S}(y) \ge 0.195 > 0$.

If $0.25 \le y \le 0.3$, then $\tilde{S}(y) \ge 0.0758 > 0$.

If $0.23 \le y \le 0.25$, then $\tilde{S}(y) \ge 0.07 > 0$.

If $0.22 \le y \le 0.23$, then $\tilde{S}(y) \ge 0.05 > 0$.

If $0.215 \le y \le 0.22$, then $\tilde{S}(y) \ge 0.04 > 0$.

If $0.21 \le y \le 0.215$, then $\tilde{S}(y) \ge 0.006 > 0$.

If $0.208 \le y \le 0.21$, then $\tilde{S}(y) \ge 0.008 > 0$. $\square$

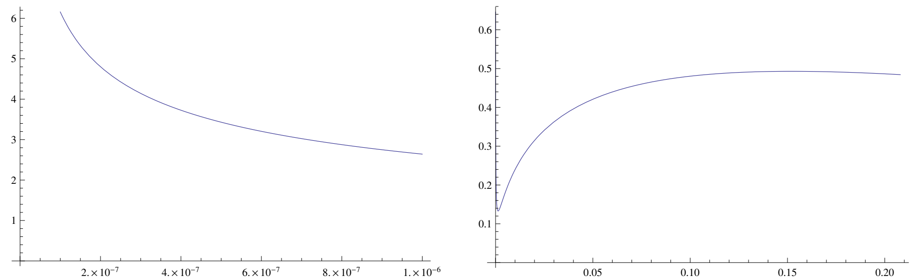


Figure 4. $\tilde{G}_1(y)$ for $10^{-7} \le y \le 10^{-6}$ and $10^{-6} \le y \le 0.208$.

We now complete the proof of (5-1) in the case $0.208 \le y \le \pi$. For any constant $C > 0$, denote

$$\tilde{F}(x) = \frac{3}{1+x} - C \cdot \frac{3x+x^2}{(x+1)^2} = \frac{3}{1+x} - C \cdot \left(1 + \frac{1}{x+1} - \frac{2}{(x+1)^2}\right).$$

Clearly

$$\tilde{F}'(x) = -\frac{3}{(1+x)^2} - C \cdot \frac{3-x}{(x+1)^3} < 0 \quad \text{for } 0 < x < 1.$$

Then for $x \le \frac{1}{3}$, we have

$$\tilde{F}(x) \ge \tilde{F}\left(\tfrac{1}{3}\right) = \tfrac{9}{4} - \tfrac{5}{8}C.$$

It follows easily that

$$\begin{aligned}\text{LHS of (5-1)} &\ge \frac{5}{4} + G_0(y) - \left(-\frac{1}{2} + \frac{1}{2\sin\frac{1}{2}y}\right) \cdot \frac{5}{8}\\ &= \frac{25}{16} + G_0(y) - \frac{5}{16} \cdot \frac{1}{\sin\frac{1}{2}y}.\end{aligned}$$

The result then follows from Lemma 6.2.

6B. ***The case where* $\mathbf{10^{-6} \le y \le 0.208}$ *and* $\mathbf{0 < t \le \frac{2}{3}y}$.** By the fact that for $C > 0$,

$$\tilde{F}(x) = \frac{3}{1+x} - C\left(1 + \frac{1}{1+x} - \frac{2}{(1+x)^2}\right)$$

is monotonically decreasing in x, we only need to prove the inequality for $10^{-6} \le y \le 0.208$:

$$\tilde{G}_1(y) := A\left(\left(\tfrac{2}{3}y\right)^{2/3}\right) + G_0(y) - \left(-\frac{1}{2} + \frac{1}{2\sin\frac{1}{2}y}\right)B\left(\left(\tfrac{2}{3}y\right)^{2/3}\right) > 0, \tag{6-1}$$

where

$$A(x) = \frac{3}{1+x} - 1, \qquad B(x) = \frac{3x+x^2}{(1+x)^2} = 1 + \frac{1}{x+1} - \frac{2}{(x+1)^2}.$$

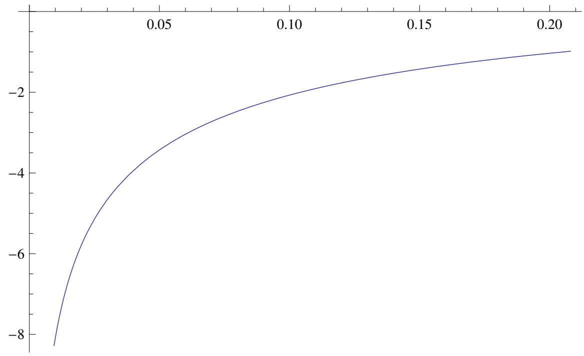


Figure 5. $B_3(y)$ for $10^{-6} \le y \le 0.208$.

From the graph of $\tilde{G}(y)$ (see Figure 4), it is transparent that $\tilde{G}_1(y) > 0$. However we shall give a more "analytic" proof at the expense of a finite number of precise numerical evaluations.

Lemma 6.3. *Set*

$$B_3(y) = A\Big(\Big(\frac{2}{3}y\Big)^{2/3}\Big) - \Big(-\frac{1}{2} + \frac{1}{2\sin\frac{1}{2}y}\Big)B\Big(\Big(\frac{2}{3}y\Big)^{2/3}\Big).$$

Then $B_3(y)$ is monotonically increasing for $10^{-6} \le y \le 0.208$.

Proof of Lemma 6.3. See Figure 5. Note that $B_3(y)$ consists only of smooth elementary functions and one can easily prove the monotonicity by constructing a suitable numerical approximation and rigorous error estimates. We omit such standard arguments. □

Now to complete the proof of (6-1), we note that by the monotonic decreasing property of $G_0(y)$ (Lemma 6.1) and the monotonic increasing property of $B_3(y)$ (Lemma 6.3), we have for any $10^{-6} \le b_1 \le y \le b_2 \le 0.208$,

$$\tilde{G}_1(y) \ge G_0(b_2) + B_3(b_1) =: \tilde{G}_2(b_1, b_2).$$

The idea is to choose finitely many $[b_1^j, b_2^j]$ such that $G_0(b_2^j) + B_3(b_1^j) > 0$ and $\bigcup_j [b_1^j, b_2^j]$ covers the whole interval $[10^{-6}, 0.208]$.

We now record the (rigorous) numerical computation below.

$$b_1 = 0.15, \quad b_2 = 0.208, \quad \tilde{G}_2 \approx 0.425414 > 0.$$

$$b_1 = 0.11, \quad b_2 = 0.15, \quad \tilde{G}_2 \approx 0.00988472 > 0.$$

$$b_1 = 0.084, \quad b_2 = 0.11, \quad \tilde{G}_2 \approx 0.0137149 > 0.$$

$$b_1 = 0.066, \quad b_2 = 0.084, \quad \tilde{G}_2 \approx 0.00547362 > 0.$$

$$b_1 = 0.06, \quad b_2 = 0.066, \quad \tilde{G}_2 \approx 0.254453 > 0.$$

Next we consider $\tilde{G}(0.06\theta^{j+1}, 0.06\theta^{j})$ for $\theta = 0.95$ and $0 \le j \le 40$. The values are listed below:

0.332451, 0.325437, 0.318268, 0.310954, 0.303506, 0.295935, 0.28825,
0.280461, 0.272577, 0.264606, 0.256557, 0.248438, 0.240257, 0.232021,
0.223737, 0.215412, 0.207053, 0.198666, 0.190256, 0.181829, 0.17339,
0.164945, 0.156499, 0.148055, 0.139618, 0.131192, 0.122782, 0.114389,
0.106019, 0.0976739, 0.0893566, 0.0810702, 0.0728171, 0.0645998, 0.0564205,
0.0482813, 0.0401842, 0.0321307, 0.0241227, 0.0161614, 0.00824834.

Note that $0.06 \cdot 0.95^{41} \approx 0.00732519$ and we then only need to consider the interval $[10^{-6}, 0.0074]$. For this we compute $\tilde{G}(0.0074\theta^{j+1}, 0.0074\theta^{j})$ for $\theta = 0.98$ and $0 \le j \le 75$. The obtained numerical values are:

0.129451, 0.127181, 0.124925, 0.122681, 0.120452, 0.118236, 0.116034,
0.113845, 0.111671, 0.109511, 0.107365, 0.105232, 0.103115, 0.101011,
0.0989223, 0.0968478, 0.0947878, 0.0927426, 0.090712, 0.0886962, 0.0866952,
0.0847091, 0.082738, 0.080782, 0.078841, 0.0769151, 0.0750044, 0.073109,
0.0712288, 0.069364, 0.0675146, 0.0656806, 0.0638621, 0.0620591, 0.0602716,
0.0584998, 0.0567436, 0.0550031, 0.0532783, 0.0515693, 0.049876, 0.0481986,
0.0465371, 0.0448914, 0.0432616, 0.0416478, 0.04005, 0.0384682, 0.0369024,
0.0353527, 0.033819, 0.0323015, 0.0308, 0.0293148, 0.0278457, 0.0263928,
0.0249561, 0.0235357, 0.0221315, 0.0207436, 0.019372, 0.0180166, 0.0166777,
0.015355, 0.0140487, 0.0127588, 0.0114853, 0.0102282, 0.00898749, 0.00776321,
0.00655539, 0.00536401, 0.00418911, 0.00303069, 0.00188876, 0.00076333

Note that $0.0074 \cdot 0.98^{76} \approx 0.00159373$, so we reduce to the interval $[10^{-6}, 0.0016]$. Finally we compute $\tilde{G}(0.0016\theta^{j+1}, 0.0016\theta^{j})$ for $\theta = 0.985$ and $0 \le j \le 500$.

0.0337677, 0.0331133, 0.0324689, 0.0318346, 0.0312105, 0.0305965, 0.0299926,
0.0293989, 0.0288153, 0.028242, 0.0276788, 0.0271257, 0.0265829, 0.0260503,
0.0255279, 0.0250157, 0.0245137, 0.024022, 0.0235405, 0.0230693, 0.0226083,
0.0221576, 0.0217171, 0.021287, 0.0208672, 0.0204576, 0.0200584, 0.0196695,
0.0192909, 0.0189226, 0.0185647, 0.0182172, 0.01788, 0.0175532, 0.0172368,
0.0169308, 0.0166351, 0.0163499, 0.0160751, 0.0158107, 0.0155567, 0.0153132,
0.0150802, 0.0148576, 0.0146455, 0.0144439, 0.0142527, 0.0140721, 0.013902,
0.0137424, 0.0135933, 0.0134548, 0.0133268, 0.0132094, 0.0131025, 0.0130063,
0.0129206, 0.0128456, 0.0127811, 0.0127273, 0.0126841, 0.0126516, 0.0126297,
0.0126185, 0.012618, 0.0126282, 0.0126491, 0.0126807, 0.0127231, 0.0127762,
0.01284, 0.0129146, 0.013, 0.0130962, 0.0132032, 0.013321, 0.0134496,
0.0135891, 0.0137395, 0.0139007, 0.0140728, 0.0142558, 0.0144498, 0.0146546,
0.0148704, 0.0150972, 0.015335, 0.0155837, 0.0158434, 0.0161142, 0.016396,
0.0166888, 0.0169927, 0.0173077, 0.0176338, 0.017971, 0.0183193, 0.0186788,
0.0190495, 0.0194313, 0.0198244, 0.0202286, 0.0206441, 0.0210708, 0.0215089,
0.0219582, 0.0224188, 0.0228907, 0.023374, 0.0238687, 0.0243747, 0.0248922,
0.025421, 0.0259613, 0.0265131, 0.0270764, 0.0276511, 0.0282374, 0.0288353,
0.0294447, 0.0300657, 0.0306983, 0.0313426, 0.0319985, 0.032666, 0.0333453,
0.0340363, 0.0347391, 0.0354536, 0.0361799, 0.036918, 0.037668, 0.0384299,
0.0392036, 0.0399892, 0.0407868, 0.0415964, 0.0424179, 0.0432515, 0.0440971,
0.0449547, 0.0458245, 0.0467064, 0.0476004, 0.0485066, 0.0494251, 0.0503557,
0.0512986, 0.0522538, 0.0532213, 0.0542012, 0.0551934, 0.056198, 0.0572151,
0.0582446, 0.0592866, 0.0603412, 0.0614082, 0.0624879, 0.0635802, 0.0646852,
0.0658028, 0.0669331, 0.0680762, 0.069232, 0.0704007, 0.0715822, 0.0727766,
0.0739838, 0.0752041, 0.0764373, 0.0776835, 0.0789427, 0.0802151, 0.0815005,
0.0827992, 0.084111, 0.085436, 0.0867743, 0.0881259, 0.0894908, 0.0908691,
0.0922608, 0.093666, 0.0950846, 0.0965168, 0.0979625, 0.0994219, 0.100895,
0.102382, 0.103882, 0.105396, 0.106924, 0.108466, 0.110022, 0.111591,
0.113175, 0.114773, 0.116384, 0.11801, 0.11965, 0.121304, 0.122973,
0.124655, 0.126352, 0.128064, 0.129789, 0.13153, 0.133284, 0.135054,
0.136838, 0.138636, 0.140449, 0.142277, 0.14412, 0.145978, 0.14785,
0.149738, 0.15164, 0.153558, 0.15549, 0.157438, 0.159401, 0.161379,
0.163373, 0.165381, 0.167406, 0.169445, 0.1715, 0.173571, 0.175658,
0.17776, 0.179877, 0.182011, 0.18416, 0.186325, 0.188507, 0.190704,
0.192917, 0.195147, 0.197392, 0.199654, 0.201932, 0.204227, 0.206538,
0.208865, 0.211209, 0.213569, 0.215946, 0.21834, 0.220751, 0.223179,
0.225623, 0.228084, 0.230563, 0.233058, 0.235571, 0.238101, 0.240648,
0.243213, 0.245795, 0.248394, 0.251011, 0.253646, 0.256298, 0.258968,
0.261656, 0.264362, 0.267086, 0.269828, 0.272588, 0.275366, 0.278163,
0.280977, 0.28381, 0.286662, 0.289532, 0.292421, 0.295329, 0.298255,

```
0.3012,    0.304164, 0.307147, 0.310149, 0.31317,  0.316211, 0.319271,
0.32235,   0.325448, 0.328566, 0.331704, 0.334862, 0.338039, 0.341236,
0.344453,  0.34769,  0.350948, 0.354225, 0.357523, 0.360841, 0.364179,
0.367538,  0.370918, 0.374318, 0.37774,  0.381182, 0.384645, 0.388129,
0.391634,  0.39516,  0.398708, 0.402277, 0.405868, 0.40948,  0.413114,
0.41677,   0.420448, 0.424148, 0.427869, 0.431613, 0.435379, 0.439168,
0.442979,  0.446812, 0.450668, 0.454547, 0.458449, 0.462374, 0.466321,
0.470292,  0.474286, 0.478304, 0.482345, 0.486409, 0.490497, 0.494609,
0.498745,  0.502905, 0.507089, 0.511297, 0.515529, 0.519785, 0.524067,
0.528372,  0.532703, 0.537058, 0.541439, 0.545844, 0.550275, 0.55473,
0.559212,  0.563718, 0.568251, 0.572809, 0.577393, 0.582003, 0.586639,
0.591301,  0.59599,  0.600705, 0.605446, 0.610215, 0.61501,  0.619832,
0.624681,  0.629557, 0.634461, 0.639392, 0.64435,  0.649337, 0.654351,
0.659393,  0.664463, 0.669561, 0.674688, 0.679843, 0.685026, 0.690238,
0.69548,   0.70075,  0.706049, 0.711378, 0.716735, 0.722123, 0.72754,
0.732987,  0.738464, 0.743971, 0.749508, 0.755075, 0.760673, 0.766302,
0.771961,  0.777651, 0.783373, 0.789125, 0.794909, 0.800725, 0.806572,
0.812451,  0.818362, 0.824305, 0.83028,  0.836288, 0.842328, 0.848401,
0.854507,  0.860646, 0.866818, 0.873023, 0.879262, 0.885534, 0.891841,
0.898181,  0.904555, 0.910964, 0.917407, 0.923885, 0.930398, 0.936945,
0.943528,  0.950146, 0.956799, 0.963489, 0.970213, 0.976974, 0.983771,
0.990604,  0.997474, 1.00438,  1.01132,  1.0183,   1.02532,  1.03238,
1.03947,   1.0466,   1.05377,  1.06097,  1.06822,  1.0755,   1.08282,
1.09018,   1.09758,  1.10502,  1.1125,   1.12001,  1.12757,  1.13517,
1.14281,   1.15048,  1.1582,   1.16596,  1.17376,  1.1816,   1.18948,
1.19741,   1.20537,  1.21338,  1.22143,  1.22952,  1.23766,  1.24584,
1.25406,   1.26232,  1.27063,  1.27898,  1.28737,  1.29581,  1.30429,
1.31282,   1.32139,  1.33001,  1.33867,  1.34738,  1.35613,  1.36493,
1.37377,   1.38266,  1.3916,   1.40058,  1.40961,  1.41869,  1.42782,
1.43699,   1.44621,  1.45548,  1.46479,  1.47416,  1.48357,  1.49304,
1.50255,   1.51211,  1.52172,  1.53138,  1.5411,   1.55086,  1.56067,
1.57054,   1.58045,  1.59042,  1.60044
```

Note that $0.0016 \cdot 0.985^{501} \approx 8.23515 \times 10^{-7}$. Thus we have proved the inequality in the range $10^{-6} \le y \le 0.208$.

6C. ***The case where*** $\mathbf{0 < y \le 10^{-6}}$ ***and*** $\mathbf{0 < t \le \frac{2}{3} y}$**.** By the same reasoning as in the (beginning of the) previous section, we only need to prove the inequality (6-1) for $0 < y \le 10^{-6}$. We begin with a lemma on the estimate of the polylog function $G_0(y)$.

Lemma 6.4. *For any* $0 < y \le 10^{-6}$, *we have*

$$G_0(y) > 2.32003 y^{-1/3} - 2.45392.$$

Proof of Lemma 6.4. By Lemma 6.1, we have $G_0(y) = g_0(\cos y)$, where

$$g_0(s) = \frac{1}{\Gamma\left(\frac{2}{3}\right)} \int_0^\infty t^{-1/3} \cdot \frac{1}{2}\left(-1 + \frac{\sinh t}{\cosh t - s}\right) dt.$$

where $\sinh t / \cosh t - s =: h(s)$. Observe that $h'(s) > 0$ for $-1 \le s < 1$. Denote $s_0 = \cos 10^{-6}$ and $\epsilon = 1 - s$ (note $\cos 10^{-6} \le s < 1$). Then

$$\begin{aligned}\Gamma\left(\tfrac{2}{3}\right) g_0(s) &\ge \int_0^1 t^{-1/3}\left(-\frac{1}{2}+\frac{1}{2}\cdot\frac{\sinh t}{\cosh t - s}\right) dt + \int_1^\infty t^{-1/3}\left(-\frac{1}{2}+\frac{1}{2}\frac{\sinh t}{\cosh t - s_0}\right) dt \\ &\ge \frac{1}{2}\int_0^1 t^{-1/3}\frac{\sinh t}{\cosh t - s}\, dt - 0.36394,\end{aligned}$$

where we have evaluated the explicit value of the integral containing s_0.

Now denote $z = \cosh t$ and note that $z = \frac{1}{2}e^t + e^{-t} > \frac{1}{2}2 + t^2 = 1 + \frac{1}{2}t^2$. This implies $t < (2(z-1))^{1/2}$ and

$$t^{-1/3} > 2^{-1/6}(z-1)^{-1/6}.$$

Then

$$\begin{aligned}
&\int_0^1 t^{-1/3}\frac{\sinh t}{\cosh t - s}\,dt \\
&\quad = \int_1^{\cosh 1} t^{-1/3}(z-s)^{-1}\,dz > \int_1^{\cosh 1} 2^{-1/6}(z-1)^{-1/6}(z-s)^{-1}\,dz \\
&\quad = \int_1^{\infty} 2^{-1/6}(z-1)^{-1/6}(z-s)^{-1}\,dz - \int_{\cosh 1}^{\infty} 2^{-1/6}(z-1)^{-1/6}(z-1)^{-1}\,dz \\
&\quad = (2\epsilon)^{-1/6}\int_0^{\infty} \tilde{z}^{-1/6}(\tilde{z}+1)^{-1}\,d\tilde{z} - \int_{\cosh 1}^{\infty} 2^{-1/6}(z-1)^{-1/6}(z-1)^{-1}\,dz,
\end{aligned}$$

where in the last line we used the change of variable $z - 1 = (1-s)\tilde{z} = \epsilon\tilde{z}$. It is easy to check that

$$\int_0^{\infty} \tilde{z}^{-1/6}(\tilde{z}+1)^{-1}\,d\tilde{z} = 2\pi,$$

$$\int_{\cosh 1}^{\infty} 2^{-1/6}(z-1)^{-1/6}(z-1)^{-1}\,dz \approx 5.91792.$$

Then

$$\int_0^1 t^{-1/3}\frac{\sinh t}{\cosh t - s}\,dt > 2(\pi(2\epsilon)^{-1/6} - 2.95896).$$

and

$$\Gamma\left(\tfrac{2}{3}\right)g_0(s) \geq \pi(2\epsilon)^{-1/6} - 3.3229.$$

Now note that $0 < y \leq 10^{-6}$ and

$$s = \cos y = 1 - \frac{1}{2}y^2 + \frac{\cos(\xi)}{24}y^4 > 1 - \frac{y^2}{2}, \qquad \xi = \xi(y) \in (0, 10^{-6}).$$

So $2\epsilon = 2(1-s) < y^2$. Then

$$G_0(y) > \frac{1}{\Gamma\left(\frac{2}{3}\right)}(\pi \cdot y^{-1/3} - 3.3229) \geq 2.32003y^{-1/3} - 2.45392. \qquad \square$$

Lemma 6.5. *Recall* $A(x) = 3/(1+x) - 1$ *and* $B(x) = (3+x)/(1+x)^2 x$. *Then for all* $0 < y \leq 10^{-6}$,

$$A\left(\left(\frac{2}{3}y\right)^{2/3}\right) + 2.32003y^{-\frac{1}{3}} - 2.45392 - \left(-\frac{1}{2} + \frac{1}{2\sin\frac{1}{2}y}\right)B\left(\left(\frac{2}{3}y\right)^{2/3}\right) > 0.$$

Proof of Lemma 6.5. Clearly

$$A\big((\tfrac{2}{3}y)^{2/3}\big) \geq A\big((\tfrac{2}{3}\cdot 10^{-6})^{2/3}\big) \geq 1.99977,$$
$$B\big((\tfrac{2}{3}y)^{2/3}\big) \leq 3\cdot(\tfrac{2}{3}y)^{2/3}.$$

For $0 < y \leq 10^{-6}$, it is easy to check that

$$y/\big(2\sin\tfrac{1}{2}y\big) < 1 + 10^{-12}.$$

Then

$$\begin{aligned} A\left(\left(\frac{2}{3}y\right)^{2/3}\right) &+ 2.32003y^{-1/3} - 2.45392 - \left(-\frac{1}{2} + \frac{1}{2\sin\frac{1}{2}y}\right)B\left(\left(\frac{2}{3}y\right)^{2/3}\right) \\ &> -0.454149 + 2.32003y^{-1/3} - \left(-\frac{1}{2} + \frac{1+10^{-12}}{y}\right)\cdot 3\cdot\left(\frac{2}{3}y\right)^{2/3} \\ &> -0.454149 + 2.32003y^{-1/3} - y^{-1/3}\cdot 2.28943 \\ &> 0.03y^{-1/3} - 0.454149 \geq 0.03\cdot 10^{2} - 0.454149 > 0. \end{aligned}$$ □

Now clearly the proof of (6-1) for $0 < y \leq 10^{-6}$ follows from Lemma 6.4 and Lemma 6.5.

7. Second proof of Theorem 1.1 and proofs of Theorems 1.4–1.5

In this section we give an alternative proof of Theorem 1.1 and also Theorems 1.4 and 1.5. We shall first prove the inequality with sharp constants, and defer the proof of the sharpness of the interpolation constants to the end of this section. To isolate the main point of the argument we need to establish an auxiliary lemma which gives a comparison principle between the Dirichlet Laplacian Δ_D and the whole space Laplacian Δ. Note that the domain of Δ_D is the usual space $H_0^1(\Omega)\cap H^2(\Omega)$. For any $f \in H_0^1(\Omega)$ we naturally identify it as a function on $\mathbb{R}^d$ by defining its values outside Ω to be zero.

Lemma 7.1 (comparison between Δ_D and Δ). *Let Ω be a smooth bounded domain in $\mathbb{R}^d$. The following hold*:

(1) *For any $t > 0$ and any $f \in H_0^1(\Omega)$ with $f \geq 0$, we have*

$$0 \leq e^{t\Delta_D} f \leq e^{t\Delta} f.$$

(2) *For any $t > 0$, $0 < \alpha < 1$ and any $f \in H_0^1(\Omega)$ with $f \geq 0$, we have*

$$0 \leq e^{-t(-\Delta_D)^\alpha} f \leq e^{-t(-\Delta)^\alpha} f.$$

(3) *For any constant* $c_1 > 0$, $0 < \alpha \le 1$, $m > 0$ *and any* $f \in H_0^1(\Omega)$ *with* $f \ge 0$,

$$0 \le \frac{1}{(c_1 + (-\Delta_D)^\alpha)^m} f \le \frac{1}{(c_1 + (-\Delta)^\alpha)^m} f.$$

(4) *For any* $0 < \beta < d/2$, *and any* $f \in C_c^\infty(\Omega)$ *with* $f \ge 0$,

$$0 \le (-\Delta_D)^{-\beta} f \le (-\Delta)^{-\beta} f.$$

(5) *Let* $x_0 \in \Omega$. *For any* $c_1 > 0$, $0 < \beta < d/4$, $0 < \alpha \le 1$, $m > 0$, *with* $m\alpha + \beta > d/4$, *we have*

$$\begin{aligned} &\|(-\Delta_D)^{-\beta}(c_1 + (-\Delta_D)^\alpha)^{-m}(\delta(x - x_0))\|_{L^2(\Omega)} \\ &\qquad\qquad \le \|(-\Delta)^{-\beta}(c_1 + (-\Delta)^\alpha)^{-m}(\delta(x - x_0))\|_{L^2(\mathbb{R}^d)}. \end{aligned}$$

Remark. In (5), the condition $\beta < d/4$ is to ensure the low frequency part of the (delta) function lies in L^2 and the condition $m\alpha + \beta > d/4$ is to ensure the high frequency part of the function lies in L^2.

Proof. The first inequality follows from a maximum principle argument on heat equations. For the second inequality, we need to use the usual subordination principle, namely for any $0 < \gamma < 1$, $\lambda \ge 0$, we have

$$e^{-\lambda^\gamma} = \int_0^\infty e^{-\lambda\tau} p_\gamma(\tau)\, d\tau,$$

where $p_\gamma(\tau)$ is a probability density function on $[0, \infty)$. Then

$$e^{-t(-\Delta_D)^\alpha} = \int_0^\infty e^{\tau t^{1/\alpha} \Delta_D} p_\alpha(\tau)\, d\tau$$

and the second inequality easily follows from this expression and the first inequality.

The third and fourth inequalities follow from the identities

$$\begin{aligned} \frac{1}{(c_1 + (-\Delta_D)^\alpha)^m} &= \frac{1}{\Gamma(m)} \int_0^\infty \tau^{m-1} e^{-\tau c_1} e^{-\tau(-\Delta_D)^\alpha}\, d\tau, \\ (-\Delta_D)^{-\beta} &= \frac{1}{\Gamma(\beta)} \int_0^\infty e^{\tau\Delta_D} \tau^{\beta-1}\, d\tau, \end{aligned}$$

where $\Gamma(\cdot)$ is the usual Gamma function. Note that the condition $0 < \beta < d/2$ guarantees that $(-\Delta)^{-\beta} f \in L^\infty(\mathbb{R}^d)$ since $f \in C_c^\infty(\Omega)$.

Finally for the fifth inequality, we first note that for any $f \in H_0^1(\Omega)$ with $f \ge 0$,

$$\begin{aligned} (-\Delta_D)^{-\beta}(c_1 + (-\Delta_D)^\alpha)^{-m} f &\le (-\Delta)^{-\beta}((c_1 + (-\Delta_D)^\alpha)^{-m} f) \\ &\le (-\Delta)^{-\beta}((c_1 + (-\Delta)^\alpha)^{-m} f). \end{aligned}$$

The corresponding result for $\delta(x - x_0)$ then follows from this and a standard approximation argument. $\square$

We now continue the proof of Theorem 1.1.

For any $N \geq 1$, consider

$$u(x) = \sum_{n=1}^{N} u_n \phi_n(x),$$

where $\phi_n(x) = \sqrt{2/\pi}\,\sin(nx)$. Clearly

$$\|(-\Delta_D)^{1/6} u\|_{L^2(\Omega)} = \left(\sum_{n=1}^{N} u_n^2 n^{2/3}\right)^{1/2},$$

$$\|(-\Delta_D)^{1/3} u\|_{L^2(\Omega)} = \left(\sum_{n=1}^{N} u_n^2 n^{4/3}\right)^{1/2}.$$

By Lemma 2.1, we have for any $y \in \Omega$,

$$\begin{aligned}\left|\sum_{n=1}^{N} u_n \phi_n(y)\right|^2 &\leq R_N \cdot \left(\sum_{n=1}^{N} u_n^2 n^{2/3}\right)^{1/2} \cdot \left(\sum_{n=1} u_n^2 n^{4/3}\right)^{1/2} \\ &= R_N \cdot \|(-\Delta_D)^{1/6} u\|_{L^2(\Omega)} \cdot \|(-\Delta_D)^{1/3} u\|_{L^2(\Omega)},\end{aligned}$$

where

$$R_N = \sum_{n=1}^{N} \frac{4\phi_n^2(y) n^{2/3}}{(\mu^{1/4} n^{2/3} + \mu^{-3/4} n^{4/3})^2},$$

and $\mu > 0$ is a constant. Now by Parseval and Lemma 7.1, we have

$$\begin{aligned} R_N &\leq \left\| \frac{2(-\Delta_D)^{1/6}}{\mu^{1/4}(-\Delta_D)^{1/3} + \mu^{-3/4}(-\Delta_D)^{2/3}} (\delta(x-y)) \right\|_{L_x^2(\Omega)}^2 \\ &= \left\| \frac{2}{(-\Delta_D)^{1/6}(\mu^{1/4} + \mu^{-3/4}(-\Delta_D)^{1/3})} (\delta(x-y)) \right\|_{L_x^2(\Omega)}^2 \\ &\leq \left\| \frac{2}{(-\Delta)^{1/6}(\mu^{1/4} + \mu^{-3/4}(-\Delta)^{1/3})} (\delta(x-y)) \right\|_{L_x^2(\mathbb{R})}^2 \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{4}{|\xi|^{2/3}(\mu^{1/4} + \mu^{-3/4}|\xi|^{2/3})^2}\, d\xi. \end{aligned}$$

By an easy change of variable, the last integral is equal to

$$\frac{12}{\pi} \int_0^{\infty} \frac{1}{(1+\rho^2)^2}\, d\rho = 3,$$

where the explicit value is computed by a standard contour integral computation.

The proof of Theorem 1.4 is similar to the proof above. Hence we shall only sketch it below.

Proof of Theorem 1.4. Denote $-\Delta_D \phi_n = \lambda_n^2 \phi_n$, where λ_n^2 is the n-th eigenvalue and ϕ_n is the eigen-function. Clearly for $u = \sum_{n=1}^N u_n \phi_n$,

$$\|A^\alpha u\|_{L^2(\Omega)} = \left(\sum_{n=1}^N u_n^2 \lambda_n^{2\alpha} \right)^{1/2}, \quad \|A^\beta u\|_{L^2(\Omega)} = \left(\sum_{n=1}^N u_n^2 \lambda_n^{2\beta} \right)^{1/2},$$

where $\beta = d - \alpha$. Denote

$$R_N = \frac{\left| \sum_{n=1}^N u_n \phi_n(y) \right|^2}{\|A^\alpha u\|_2 \|A^\beta u\|_2}.$$

Then by Lemmas 2.1 and 7.1, we have

$$\begin{aligned}
R_N &\le \sum_{n=1}^N \frac{4\phi_n^2(y)\lambda_n^{2\alpha}}{(\mu^{1/4}\lambda_n^{2\alpha} + \mu^{-3/4}\lambda_n^{2\beta})^2} \\
&\le 4 \cdot \mu^{-1/2} \|A^{-\alpha} \cdot (1 + A^{2(\beta-\alpha)}/\mu)^{-1}(\delta(x-y))\|_{L^2(\Omega)}^2 \\
&\le 4 \cdot \mu^{-1/2} \||\nabla|^{-\alpha} \cdot (1 + |\nabla|^{2(\beta-\alpha)}/\mu)^{-1}(\delta(x-y))\|_{L^2(\Omega)}^2 \\
&\le \frac{4}{\sqrt{\mu}} \cdot \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\xi|^{-2\alpha} \cdot (1 + |\xi|^{2(\beta-\alpha)}/\mu)^{-2} \, d\xi \\
&= \frac{4}{(2\pi)^d} \int_{\mathbb{R}^d} |\xi|^{-2\alpha} (1 + |\xi|^{2(d-2\alpha)})^{-2} \, d\xi.
\end{aligned}$$

Evaluating the last integral in polar coordinates then yields the result. □

Proof of Theorem 1.5. The proof of Theorem 1.5 is essentially the same as the proof of Theorem 1.4. We shall only point out the change of numerology. Namely

$$R_N = \frac{\left| \sum_{n=1}^N u_n \phi_n(y) \right|^2}{\|A^\alpha u\|_2^{2\theta} \|A^\beta u\|_2^{2(1-\theta)}}$$

and

$$\begin{aligned}
R_N &\le \mu^{-(1-\theta)} \cdot \|A^{-\alpha} \cdot (\theta + (1-\theta)A^{2(\beta-\alpha)}/\mu)^{-1}(\delta(x-y))\|_{L^2(\Omega)}^2 \\
&\le \mu^{-(1-\theta)} \cdot \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\xi|^{-2\alpha} (\theta + (1-\theta)|\xi|^{2(\beta-\alpha)}/\mu)^{-2} \, d\xi \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\xi|^{-2\alpha} (\theta + (1-\theta)|\xi|^{2(\beta-\alpha)})^{-2} \, d\xi.
\end{aligned}$$

One can then use polar coordinates, a change of variable and rescaling, and a simple identity

$$\int_0^\infty \rho^{-\gamma}(1+\rho)^{-2} \, d\rho = \frac{\pi\gamma}{\sin(\pi\gamma)}, \quad 0 < \gamma < 1,$$

to get the result. Note that the above identity is a simple exercise using contour integrals. □

We now prove the sharpness of the interpolation constants in Theorems 1.1–1.5. For this we will need a convergence result. For any bounded domain $\Omega \subset \mathbb{R}^d$ with smooth boundary, fix some $x_0 \in \Omega$ and define for $\lambda > 0$,

$$\Omega_\lambda = \{x \in \mathbb{R}^d : (x - x_0)/\lambda \in \Omega\}. \tag{7-1}$$

In other words, Ω_λ is the λ-dilation of Ω. We shall denote by Δ_Ω and Δ_{Ω_λ} the Dirichlet Laplacian on Ω and Ω_λ respectively.

Lemma 7.2. *For any $f \in C_c^\infty(\Omega)$ and any $a > 0$,*

$$\|(-\Delta_{\Omega_\lambda})^a f\|_{L^2(\Omega_\lambda)} \to \|(-\Delta)^a f\|_{L^2(\mathbb{R}^d)} \quad \text{as } \lambda \to \infty.$$

Proof of Lemma 7.2. Without loss of generality we can assume $0 < a < 1$. Note that for $f \in C_c^\infty(\Omega)$, we have $\Delta f \in C_c^\infty(\Omega)$ and thus

$$\begin{aligned}(-\Delta_{\Omega_\lambda})^a f &= (-\Delta_{\Omega_\lambda})^{-(1-a)}(-\Delta f) \\ &= \frac{1}{\Gamma(1-a)} \int_0^\infty t^{-a} e^{t\Delta_{\Omega_\lambda}}(-\Delta f)\,dt.\end{aligned}$$

Note that for t large,

$$\|e^{t\Delta_{\Omega_\lambda}}(-\Delta f)\|_{L^2(\Omega_\lambda)} = \|\Delta_{\Omega_\lambda} e^{t\Delta_{\Omega_\lambda}} f\|_{L^2(\Omega_\lambda)} \le t^{-1}\|f\|_{L^2(\mathbb{R}^d)},$$

and for t small,

$$\|e^{t\Delta_{\Omega_\lambda}}(-\Delta f)\|_{L^2(\Omega_\lambda)} \le \|\Delta f\|_{L^2(\mathbb{R}^d)}.$$

These two estimates allow us to localize in t. It then suffices to prove for single $t \sim 1$ and any $g \in C_c^\infty(\Omega)$ that

$$\|e^{t\Delta_{\Omega_\lambda}} g - e^{t\Delta} g\|_{L^2(\Omega_\lambda)} + \|e^{t\Delta} g\|_{L^2(\mathbb{R}^d \setminus \Omega_\lambda)} \to 0 \quad \text{as } \lambda \to \infty. \tag{7-2}$$

Now without loss of generality assume $0 \in \Omega$ and Ω_λ is the dilation of Ω with respect to the origin. Set $\eta = e^{t\Omega_\lambda} g - e^{t\Delta} g$. Note that $\|e^{t\Delta} g\|_{L^\infty(|x| \gtrsim \lambda)} \lesssim \lambda^{-M}$ for any $M > 0$. A simple maximum principle estimate on η then yields (7-2). □

We now show the sharpness of the interpolation constant in Theorem 1.4 (the argument for Theorems 1.1–1.5 is similar). The idea is as follows. Fix Ω and without loss of generality assume $0 \in \Omega$. Denote Ω_λ as in Lemma 7.2. Suppose

$$C_\Omega = \sup_{f \in H^N(\Omega) \cap H_0^1(\Omega)} \frac{\|f\|^2_{L^\infty(\Omega)}}{\|A^\alpha f\|_{L^2(\Omega)} \|A^{d-\alpha} f\|_{L^2(\Omega)}}, \quad N = [(d+1)/2].$$

By the proof of Lemma 7.2, $C_\Omega \le C_{\alpha,d}$ ($C_{\alpha,d}$ is the constant in Theorem 1.4). Alternatively we can write

$$\|f\|^2_{L^\infty(\Omega)} \le C_\Omega \|A^\alpha f\|_{L^2(\Omega)} \|A^{d-\alpha} f\|_{L^2(\Omega)}.$$

Clearly the same inequality holds with Ω_λ replacing Ω and the same interpolation constant C_Ω. Now fix $f \in C_c^\infty(\mathbb{R}^d)$. By taking $\lambda \to \infty$ and using Lemma 7.2, we then obtain

$$\|f\|^2_{L^\infty(\mathbb{R}^d)} \le C_\Omega \|(-\Delta)^{\alpha/2} f\|_{L^2(\mathbb{R}^d)} \|(-\Delta)^{(d-\alpha)/2} f\|_{L^2(\mathbb{R}^d)}.$$

Now for the whole space case, by Proposition 8.2 and Remark 8.4, there exists $f_* \in \dot{H}^\alpha(\mathbb{R}^d) \cap \dot{H}^{d-\alpha}(\mathbb{R}^d)$ attaining the sharp constant $C_{\alpha,d}$. By smoothly truncating f_*, one can find $f_n \in C_c^\infty(\mathbb{R}^d)$ which approaches $C_{\alpha,d}$ in the limit. Clearly then $C_\Omega \ge C_{\alpha,d}$ and thus $C_\Omega = C_{\alpha,d}$.

Remark 7.3. The preceding argument can be adapted to give an explicit sequence of functions f_n attaining the sharp interpolation constant $C_{\alpha,d}$ in the limit $n \to \infty$. We again illustrate for the case of Theorem 1.4. Without loss of generality assume $0 \in \Omega$ and define Ω_λ as in (7-1). Let $\chi \in C_c^\infty(\mathbb{R}^d)$ be such that $\chi(x) = 1$ for $|x| \le 1$ and $\chi(x) = 0$ for $|x| \ge 2$. Fix any integer $n \ge 1$. Define $g_R(x) = f_*(x)\chi(x/R)$. By taking R sufficiently large, we have

$$C_{\alpha,d} \cdot \left(1 - \frac{1}{n}\right) \le \frac{\|g_R\|^2_{L^\infty(\mathbb{R}^d)}}{\|(-\Delta)^{\alpha/2} g_R\|_{L^2(\mathbb{R}^d)} \|(-\Delta)^{(d-\alpha)/2} g_R\|_{L^2(\mathbb{R}^d)}} \le C_{\alpha,d}.$$

Now fix R and define $h_\lambda(x) = g_R(\lambda x)$. Now observe that for any $\lambda > 0$ sufficiently large, we have $h_\lambda \in C_c^\infty(\Omega)$. Also for any $s > 0$,

$$\|(-\Delta_\Omega)^{s/2}(h_\lambda)\|_{L^2(\Omega)} = \lambda^{s-d/2} \|(-\Delta_{\Omega_\lambda})^{s/2}(g_R)\|_{L^2(\Omega_\lambda)}.$$

By Lemma 7.2, it follows easily that

$$\begin{aligned}
&\frac{\|h_\lambda\|^2_{L^\infty(\Omega)}}{\|(-\Delta_\Omega)^{\alpha/2} h_\lambda\|_{L^2(\Omega)} \|(-\Delta_\Omega)^{(d-\alpha)/2} h_\lambda\|_{L^2(\Omega)}} \\
&\qquad = \frac{\|g_R\|^2_{L^\infty(\Omega_\lambda)}}{\|(-\Delta_{\Omega_\lambda})^{\alpha/2} g_R\|_{L^2(\Omega_\lambda)} \|(-\Delta_{\Omega_\lambda})^{(d-\alpha)/2} g_R\|_{L^2(\Omega_\lambda)}} \\
&\qquad \to \frac{\|g_R\|^2_{L^\infty(\mathbb{R}^d)}}{\|(-\Delta)^{\alpha/2} g_R\|_{L^2(\mathbb{R}^d)} \|(-\Delta)^{(d-\alpha)/2} g_R\|_{L^2(\mathbb{R}^d)}} \quad \text{as } \lambda \to \infty.
\end{aligned}$$

One can then choose λ_n sufficiently large and define $f_n = h_{\lambda_n}$.

8. Third proof of Theorem 1.1

Lemma 8.1 (equivalence of constants). *Let the dimension $d \ge 1$. Let $0 < \theta < 1$, $1 < p < \infty$, $-\infty < a < (p-1)/(p)d < b < \infty$ satisfy*

$$d\left(1 - \frac{1}{p}\right) = \theta a + (1-\theta) b.$$

Then

$$\int_{\mathbb{R}^d} |x|^{-pa/(p-1)}(\theta+(1-\theta)|x|^{p(b-a)})^{-p/(p-1)}\,dx \\ = \int_{\mathbb{R}^d} |x|^{-pa/(p-1)}(\theta+(1-\theta)|x|^{p(b-a)})^{-1/(p-1)}\,dx.$$

Proof of Lemma 8.1. In polar coordinates, the identity takes the form

$$\int_0^\infty \rho^{-pa/(p-1)+d-1}(\theta+(1-\theta)\rho^{p(b-a)})^{-p/(p-1)}\,d\rho \\ = \int_0^\infty \rho^{-pa/(p-1)+d-1}(\theta+(1-\theta)\rho^{p(b-a)})^{-1/(p-1)}\,d\rho.$$

Since $d(1-1/p)=a+(b-a)(1-\theta)$, we can write

$$-\frac{pa}{p-1}+d-1=\frac{p}{p-1}(b-a)(1-\theta)-1.$$

By a change of variable $(1-\theta)/(\theta)\rho^{p(b-a)} \to \rho$, we reduce to proving

$$\int_0^\infty \rho^{\alpha(1-\theta)-1}(1+\rho)^{-1-\alpha}\,d\rho = \theta \int_0^\infty \rho^{\alpha(1-\theta)-1}(1+\rho)^{-\alpha}\,d\rho,$$

where $\alpha = 1/(p-1)$. The above is equivalent to

$$(1-\theta)\int_0^\infty \rho^{\alpha(1-\theta)-1}(1+\rho)^{-1-\alpha}\,d\rho = \theta\int_0^\infty \rho^{\alpha(1-\theta)}(1+\rho)^{-1-\alpha}\,d\rho.$$

Using integration by parts, the left-hand side of this equation gives

$$\frac{1+\alpha}{\alpha}\int_0^\infty \rho^{\alpha(1-\theta)}(1+\rho)^{-2-\alpha}\,d\rho,$$

while the right-hand side gives

$$\frac{1+\alpha}{\alpha}\int_0^\infty \rho^{\alpha\theta}\cdot\left(\frac{1+\rho}{\rho}\right)^{-2-\alpha}\cdot\rho^{-2}\,d\rho = \frac{1+\alpha}{\alpha}\int_0^\infty \tilde{\rho}^{\alpha(1-\theta)}(1+\tilde{\rho})^{-2-\alpha}\,d\tilde{\rho},$$

with $\tilde{\rho}=1/\rho$, concluding the proof. □

Proposition 8.2. *Let* $1<p<\infty$ *and* $0<\theta<1$. *Assume* $-\infty<a<(p-1)/(p)d<b<\infty$ *satisfy*

$$d\left(1-\frac{1}{p}\right)=a\theta+b(1-\theta).$$

For any $g:\mathbb{R}^d\to\mathbb{C}$ *in* $L^1_{\mathrm{loc}}(\mathbb{R}^d)$, *if* $|x|^a g\in L^p$ *and* $|x|^b g\in L^p$, *then* $g\in L^1$, *and*

$$\|g\|_{L^1(\mathbb{R}^d)} \le C_2\cdot \||x|^a g\|^{\theta}_{L^p(\mathbb{R}^d)}\cdot \||x|^b g\|^{1-\theta}_{L^p(\mathbb{R}^d)},$$

where

$$C_2 = \left(\int_{\mathbb{R}^d} |x|^{-pa/(p-1)} (\theta + (1-\theta)|x|^{p(b-a)})^{-1/(p-1)}\, dx \right)^{(p-1)/p}.$$

The constant C_2 is sharp with the equality achieved by

$$g(x) = g_*(x) = |x|^{-pa} (\theta + (1-\theta)|x|^{p(b-a)})^{-1/(p-1)}.$$

Furthermore, for any $g(x)$ achieving the equality, there exist $a_1 \in \mathbb{R}$, $a_2 > 0$ and measurable $\omega : \mathbb{R}^d \to \mathbb{R}$ such that

$$g(x) = a_1 g_*(a_2 x) \cdot e^{i\omega(x)} \quad \textit{a.e. in } x \in \mathbb{R}^d.$$

Remark 8.3. By Lemma 8.1, we have $C_2 = C_1$ where C_1 is the constant in Proposition 2.3. Hence these two propositions are essentially equivalent (except for the wider range for the parameter a in the latter).

Remark 8.4. For the special case $p = 2$, it follows that the sharp constant to the inequality

$$\|f\|_{L^\infty(\mathbb{R}^d)} \le C \cdot \| |\nabla|^a f\|^{\theta}_{L^2(\mathbb{R}^d)} \| |\nabla|^b f\|^{1-\theta}_{L^2(\mathbb{R}^d)} \quad \text{for all } f \in \dot{H}^a \cap \dot{H}^b,$$

with $-\infty < a < d/2 < b < \infty$, $0 < \theta < 1$ and $d/2 = a\theta + b(1-\theta)$, is achieved by the function $f = f_*$ with Fourier transform

$$\hat{f}_*(\xi) = \text{const} \cdot |\xi|^{-2a} \cdot (\theta + (1-\theta)|\xi|^{2(b-a)})^{-1}.$$

Furthermore, by examining the condition for equality (we also need to use the condition for $\|f\|_{L^\infty_x} = (2\pi)^{-d} \|f\|_{L^1_\xi}$ to simplify the phase factor $e^{i\omega(\xi)}$), we have that any function f attaining the sharp interpolation constant must be equal to f_* up to multiplication by a constant factor, dilation or translation (in x).

Proof of Proposition 8.2. By scaling we may assume $\| |x|^a g\|_p = \| |x|^b g\|_p = 1$. Clearly then $\|wg\|_p = 1$ with

$$w(x) = |x|^a (\theta + (1-\theta)|x|^{p(b-a)})^{1/p}.$$

By the Hölder inequality, we have

$$\|g\|_1 = \|w^{-1} \cdot wg\|_1 \le \|w^{-1}\|_{p/(p-1)} \|wg\|_p = C_2.$$

The equality is clearly achieved by $|g| = \text{const} \cdot w^{-p/(p-1)}$ (a.e. in $x \in \mathbb{R}^d$). Note that such a function achieving the inequality must be unique and equal to $g_*(x)$ (up to dilations and multiplication by a constant factor). $\square$

Inspired by the proof of Proposition 8.2, we now give a short proof of the inequality (1-2). Note that we do not make use of any variational arguments (although the argument is “deeply inspired” by the aforementioned variational proof). The proof of the (more general) Theorem 1.5 will only differ by some numerology.

Proof of (1-2). Without loss of generality, assume $\|\nabla f\|_2 = 1$ and $\|\Delta f\|_2 = \gamma > 0$. Denote $A = (-\Delta_D)^{1/2}$ and

$$B = \left(\frac{\gamma^2 A^2 + A^4}{2\gamma}\right)^{1/2} = A\left(\frac{\gamma^2 - \Delta_D}{2\gamma}\right)^{1/2}.$$

Then by using eigen-function expansion, it is easy to check that $\|Bf\|_2 = \sqrt{\gamma}$. Clearly

$$\|f\|^2_{L^\infty(\Omega)} = \|B^{-1}Bf\|^2_{L^\infty(\Omega)}.$$

So it suffices for us to prove the following inequality for $g \in H^1_0(\Omega)$:

$$\|B^{-1}g\|_{L^\infty(\Omega)} \le (2\pi)^{-1/2}\|g\|_{L^2(\Omega)}.$$

Denote $\tilde{g}(x) = |g(x)|$ if $x \in \Omega$, and $\tilde{g}(x) = 0$ for $x \in \mathbb{R}^3 \setminus \Omega$. Denote

$$\widetilde{B} = \left(\frac{-\gamma^2\Delta + \Delta^2}{2\gamma}\right)^{1/2} = |\nabla|\left(\frac{\gamma^2 - \Delta}{2\gamma}\right)^{1/2}.$$

By using a comparison argument (as in Lemma 7.1), it is easy to check that

$$\begin{aligned} \text{(8-1)} \qquad \|B^{-1}g\|_{L^\infty(\Omega)} &\le \|\widetilde{B}^{-1}\tilde{g}\|_{L^\infty(\mathbb{R}^3)} \\ &\le (2\pi)^{-3}\|\widehat{\widetilde{B}^{-1}\tilde{g}}\,(\xi)\|_{L^1_\xi(\mathbb{R}^3)} \le (2\pi)^{-1/2}\|\tilde{g}\|_2, \end{aligned}$$ □

Finally we give a short proof of Theorem 1.5.

Proof of Theorem 1.5. Without loss of generality, $\|A^\alpha f\|_2 = 1$ and $\|A^\beta f\|_2 = b > 0$. We then consider

$$B = \left(\frac{b^2 A^{2\alpha} + A^{2\beta}}{2b^{2\theta}}\right)^{1/2}.$$

Note that $B^{-1} = \sqrt{2}b^\theta A^{-\alpha}(b^2 + A^{2(\beta-\alpha)})^{-1/2}$. In view of the assumption $\beta - \alpha \le 1$ and by Lemma 7.1, one can still use the comparison argument and bound B^{-1} by the corresponding full space operator $\widetilde{B}^{-1}$. The sharp constant is then given by the expression

$$\begin{aligned} C &= \|\widetilde{B}^{-1}\delta_0\|^2_{L^2(\mathbb{R}^d)} \\ &= (2\pi)^{-d}|S^{d-1}| \cdot \int_0^\infty \frac{2}{1+\rho^{2(\beta-\alpha)}} \cdot \rho^{d-1-2\alpha}\,d\rho, \end{aligned}$$

where $|S^{d-1}|$ denotes the surface area of a $(d-1)$-dimensional hypersphere. Note that by Lemma 8.1 and the proof of Theorem 1.5 in Section 7, the constant C

is the same as the constant $C_{\alpha,\beta,\theta,d}$ in Theorem 1.5. By using the fact $d-2\alpha = 2(1-\theta)(\beta-\alpha)$ and the change of variable $t=\rho^{2(\beta-\alpha)}$, we can easily calculate C as

$$C = (2\pi)^{-d}\cdot\frac{1}{\beta-\alpha}\cdot|S^{d-1}|\cdot\int_0^\infty (1+t)^{-1}t^{-\theta}\,dt$$
$$= (2\pi)^{-d}\cdot\frac{1}{\beta-\alpha}\cdot\frac{2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)}\cdot\frac{\pi}{\sin\pi\theta}. \qquad \square$$

Finally we show that for any smooth bounded domain $\Omega\subset\mathbb{R}^d$, there does not exist any function achieving the equality with sharp interpolation constant. First we need a simple lemma.

Lemma 8.5. *Let δ_0 be the usual Dirac distribution on $\mathbb{R}^d$. Assume $0<\alpha_0<d$ and $0<\beta_0\le 2$. Then there exists a constant $C_{d,\alpha_0,\beta_0}>0$, such that*

$$\big(|\nabla|^{-\alpha_0}(1+|\nabla|^{\beta_0})^{-1}\delta_0\big)(x) \ge C_{d,\alpha_0,\beta_0}(1+|x|)^{-(d-\alpha_0)} \quad \textit{for all } x\in\mathbb{R}^d.$$

Proof of Lemma 8.5. Denote $g=(1+|\nabla|^{\beta_0})^{-1}\delta_0$. Clearly $g\in L^1(\mathbb{R}^d)$, $g\ge 0$, and there exists $R_0=R_0(\beta_0,d)>0$ such that

$$\int_{|y|<R_0} g(y)\,dy > \tfrac{1}{2}.$$

Clearly then

$$\int_{|y|<R_0} |x-y|^{-(d-\alpha_0)}g(y)\,dy \ge \text{const}\cdot(1+|x|)^{-(d-\alpha_0)}.$$

Since $|\nabla|^{-\alpha_0}g = \text{const}\cdot|x|^{-(d-\alpha_0)} * g$, the result easily follows. $\square$

Proof of Remark 1.6. To avoid complicated numerology, consider first the case of inequality (1-2). Assume there exists a function attaining the sharp interpolation constant. A close examination of the last inequality in (8-1) shows that to have inequality we must have

$$\hat{\tilde{g}}(\xi) = c_1\cdot e^{i\omega(\xi)}\cdot|\xi|^{-1}(\gamma^2+|\xi|^2)^{-1/2},$$

where $c_1>0$ is a constant and $\omega:\mathbb{R}^d\to\mathbb{R}$. By examining the second inequality in (8-1), it is not difficult to check that we must have $e^{i\omega(\xi)}\equiv 1$ (a.e. in ξ) in order to have equality. By Lemma 8.5, we then must have

$$\tilde{g}(x)\ge \text{const}\cdot(1+|x|)^{-2} \quad \text{for all } x\in\mathbb{R}^3.$$

But this contradicts the fact that $\tilde{g}(x)=0$ for $x\in\mathbb{R}^3\setminus\Omega$. Thus there can be no such function for the case of inequality (1-2).

The proof for the general case (Theorem 1.5) follows along similar lines. We omit details. $\square$

Acknowledgments

Li was supported in part by Hong Kong RGC grant GRF 16307317 and 16309518. We thank the anonymous referees for very helpful remarks and suggestions.

References

[Heywood 2000] J. G. Heywood, "An alternative to Xie's conjecture concerning the Stokes problem in nonsmooth domains", *Ann. Univ. Ferrara Sez. VII* (*N.S.*) **46** (2000), 267–284. MR Zbl

[Heywood 2014] J. G. Heywood, "Seeking a proof of Xie's inequality: analogues for series and Fourier series", *Ann. Univ. Ferrara Sez. VII Sci. Mat.* **60**:1 (2014), 149–167. MR Zbl

[Heywood and Xie 1997] J. G. Heywood and W. Xie, "Smooth solutions of the vector Burgers equation in nonsmooth domains", *Differential Integral Equations* **10**:5 (1997), 961–974. MR Zbl

[Xie 1991] W. Xie, "A sharp pointwise bound for functions with L^2-Laplacians and zero boundary values of arbitrary three-dimensional domains", *Indiana Univ. Math. J.* **40**:4 (1991), 1185–1192. MR Zbl

[Xie 1992] W. Xie, "On a three-norm inequality for the Stokes operator in nonsmooth domains", pp. 310–315 in *The Navier–Stokes equations II—theory and numerical methods* (Oberwolfach, 1991), edited by J. G. Heywood et al., Lecture Notes in Math. **1530**, Springer, 1992. MR Zbl

[Xie 1995] W. Xie, "Integral representations and L^∞ bounds for solutions of the Helmholtz equation on arbitrary open sets in $\mathbb{R}^2$ and $\mathbb{R}^3$", *Differential Integral Equations* **8**:3 (1995), 689–698. MR Zbl

[Xie 1997] W. Xie, "Sharp Sobolev interpolation inequalities for the Stokes operator", *Differential Integral Equations* **10**:2 (1997), 393–399. MR Zbl

Received February 16, 2018. Revised January 8, 2019.

DONG LI
HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
CLEAR WATER BAY
HONG KONG
madli@ust.hk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.265

KNAPP–STEIN DIMENSION THEOREM FOR FINITE CENTRAL COVERING GROUPS

CAIHUA LUO

It is folklore that the Knapp–Stein dimension theorem should be extended word by word to general covering groups. But we note that such a proof does not exist in the literature. For completeness, we provide a proof of the classical Knapp–Stein dimension theorem for finite central covering groups. As an example, we obtain the R-group structure for Mp_{2n} based on Gan and Savin's work on the local theta correspondence for $(Mp_{2n}, \mathrm{SO}_{2n+1})$.

1. Introduction

Let G be a connected reductive group defined over a nonarchimedean local field F. By abuse of notation, we also write G for $G(F)$, and denote by $\tilde{G}$ the associated finite central covering group of $G(F)$ by μ_n, i.e.,

$$1 \to \mu_n \to \tilde{G} \xrightarrow{p} G(F) \to 1.$$

Representation-theoretically, one of the fundamental problems is to understand the classification of irreducible admissible representations of G. Notably, we have the diagram

$$\Pi_{s.c}(G) \overset{\text{(iii)}}{\subset} \Pi_{d.s}(G) \overset{\text{(ii)}}{\subset} \Pi_{\text{temp}}(G) \overset{\text{(i)}}{\subset} \Pi(G).$$

Here $\Pi(G)$, $\Pi_{\text{temp}}(G)$, $\Pi_{d.s}(G)$ and $\Pi_{s.c}(G)$ stand for the set of isomorphism classes of irreducible admissible representations, tempered representations, discrete series and supercuspidal representations of G, respectively. Recall that

- (i) is the Langlands classification [1989]. The covering case is established by Ban and Jantzen [2013],
- (ii) is the Knapp–Zuckerman classification [1982]. The covering case follows from [Waldspurger 2003, Proposition III.4.1] (see [Li 2012]).
- (iii) is the Moeglin–Tadić classification for classical groups [2002]. The covering case is unknown so far.

MSC2010: 22E35.
Keywords: finite central covering group, R-group, intertwining operator, Knapp–Stein dimension theorem.

Note that in (ii), as the normalized induced representation is unitary, we may investigate its finer structure, i.e., the so-called R-group theory (see [Knapp and Stein 1971; 1980; Silberger 1978]). Indeed, the R-group theory not only determines the decomposition of induced representations, but also plays an essential role in the endoscopy theory. In view of this, it is necessary to extend the R-group theory to covering groups. This is exactly what we will do in the paper.

In what follows, we give a rough outline of the main results. Exact definitions and notation are given in the body of the paper. For a standard parabolic $P = MN \subset G$ and $\sigma \in \Pi_{d.s}(M)$, we denote by $I_P^G(\sigma)$ the associated normalized induced representation of G, and by $W(M)$ the relative Weyl group of M in G. Let $W(\sigma) := \{\omega \in W(M) : \omega.\sigma = \sigma\}$. Fix a section of $W(M)$ in K_{good} which is a special compact open subgroup of G. For simplicity, we use the same letter w for the fixed lifting of $w \in W(M)$ if no confusion arises. For $\omega \in W(\sigma)$, we may define an unnormalized intertwining operator $\gamma(\omega) = A(\omega) \circ M(\omega, \sigma)$ on $I_P^G(\sigma)$ as in Lemma 2.2, and then define $W^0(\sigma) := \{\omega \in W(\sigma) : A(\omega) \circ M(\omega, \sigma) \text{ is a scalar}\}$. It is well known that $W^0(\sigma)$ is a normal subgroup of $W(\sigma)$, and the following exact sequence splits:

$$1 \to W^0(\sigma) \to W(\sigma) \to R(\sigma) \to 1.$$

Main Theorem (Theorem 2.4). Modifying the above notation for $\tilde{G}$, we have

$$\dim \operatorname{End}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})) = |R(\tilde{\sigma})|.$$

Classically, Knapp and Stein [1971; 1980] established the dimension theorem for tempered induced representations of semisimple Lie groups and later Silberger [1978] extended the Knapp–Stein dimension theorem to p-adic reductive groups. Notice that the strategies to prove the Knapp–Stein dimension theorem in [Silberger 1978] are as follows.

(i) The Harish-Chandra commuting algebra theorem; that is, $\operatorname{End}_G(I_P^G(\sigma)) = \operatorname{Span}\{\gamma(\omega) : \omega \in W(\sigma)\}$.

(ii) $\operatorname{End}_G(I_P^G(\sigma)) = \operatorname{Span}\{\gamma(\omega) : \omega \in R(\sigma)\}$.

(iii) The multiplicity of a given exponent in the tempered Jacquet module $J_{\bar{P}}^{\omega}(I_P^G(\sigma))$ is no greater than the cardinality of $W^0(\sigma)$. This in turn implies that

$$\dim \operatorname{End}_G(I_P^G(\sigma)) = [W(\sigma) : W^0(\sigma)] = |R(\sigma)|.$$

So basically, we adapt the same argument as in [Silberger 1978] to extend the Knapp–Stein dimension theorem for finite central covering groups. But instead of showing the multiplicity of a given exponent is bounded by $W^0(\sigma)$ in the weak Jacquet module $J_{\bar{P}}^{\omega}(I_P^G(\sigma))$, we follow D. Ban's argument [2004] to directly show the linear independence property of these $\gamma(\omega)$ with $\omega \in R(\sigma)$. Herein, we should mention that the Knapp–Stein dimension theorem for finite central covering groups has been announced by Wen-Wei Li [2012] and D. Szpruch [2013].

2. Knapp–Stein dimension theorem

2A. ***Notation and conventions.*** In this section, we first recall some necessary definitions and properties in [Waldspurger 2003; Li 2012] for our purpose.

Recall that F is a nonarchimedean field, G is a connected reductive group defined over F, and $\tilde{G}$ is the associated finite central covering group of $G(F)$ by μ_n, i.e.,

$$1 \to \mu_n \to \tilde{G} \xrightarrow{p} G(F) \to 1.$$

Fix a maximal split torus T of G and the associated minimal parabolic subgroup $P_0 = M_0 N_0$ with the Levi subgroup M_0 containing T. Denote by $W_G(T) := N_G(T)/C_G(T)$ the Weyl group of G with respect to T, and by $\Phi = \Phi(G, T)$ the set of relative roots of T in G. The choice of P_0 determines the set of relative simple roots Δ and the set of relative positive roots $\Phi^+ \subset \Phi$. If $\alpha \in \Phi^+$, we write $\alpha > 0$.

As is well known, the standard parabolic subgroups of G are uniquely determined by a subset Θ of Δ. For such a subset $\Theta \subset \Delta$, let $P_\Theta = M_\Theta N_\Theta$ be the associated standard parabolic subgroup of G with Levi subgroup $M_\Theta \supset M_0$, and $W(M_\Theta) := N_G(M)/M$ be the relative Weyl group with respect to M_Θ in G. Let T_{M_Θ} be the split component of the center of M_Θ, and $X(M_\Theta)_F := \mathrm{Hom}_{F-\mathrm{grp}}(M_\Theta, \mathbb{G}_m)$ be the group of all F-rational characters of M_Θ. Denote $\mathfrak{a}_{M_\Theta} := \mathrm{Hom}(X(M_\Theta)_F, \mathbb{R}) = \mathrm{Hom}(X(T_{M_\Theta})_F, \mathbb{R})$. Recall that in [Luo 2017] we have defined the Harish-Chandra homomorphism $H_{\tilde{M}_\Theta} : \tilde{M}_\Theta \to \mathfrak{a}_{M_\Theta}$ by $H_{M_\Theta} \circ p$, and $\tilde{M}_\Theta^1 := \mathrm{Ker}(H_{\tilde{M}_\Theta})$. Denote $X(\tilde{M}_\Theta) := \mathrm{Hom}(\tilde{M}_\Theta/\tilde{M}_\Theta^1, \mathbb{C}^\times)$. As $X(\tilde{M}_\Theta) = X(M_\Theta)$, we may attach a complex algebraic variety structure on $X(\tilde{M}_\Theta)$ via the surjective homomorphism

$$\mathfrak{a}^*_{\tilde{M}_\Theta,\mathbb{C}} := X(M_\Theta)_F \bigotimes_{\mathbb{Z}} \mathbb{C} \xrightarrow{\iota} X(M_\Theta)$$

given by $\chi \otimes s \mapsto |\chi(\cdot)|^s$. Notice that $\mathrm{Ker}(\iota)$ is of the form $2\pi i/(\log q)L$ with L a lattice in $X(M_\Theta)_F \otimes_{\mathbb{Z}} \mathbb{Q}$; it does make sense to define the notion of real part $\mathrm{Re}(\chi)$ of $\chi \in X(M_\Theta)$. Denote $\mathrm{Im}\, X(\tilde{M}_\Theta) := \mathrm{Im}\, X(M_\Theta) = \{\chi \in X(M_\Theta) : \mathrm{Re}(\chi) = 0\}$.

Fix the canonical section of unipotent elements in G to $\tilde{G}$ as in [Mœglin and Waldspurger 1995, Appendix I] or [Li 2014, Proposition 2.2.1]. By abuse of notation, for a unipotent subgroup $N \subset G$, we also write N for its canonical section in $\tilde{G}$. For a standard parabolic subgroup $P = MN$ of G, let $\tilde{P} = \tilde{M}N$ be the preimage of P in $\tilde{G}$. Denote by $\Pi(\tilde{M})$ (resp. Π_2, $\Pi_{\mathrm{temp}}(\tilde{M})$) the set of isomorphism classes of genuine irreducible admissible (resp. discrete series, tempered) representations of $\tilde{M}$. Notice that the complex torus $X(\tilde{M})$ acts naturally on $\Pi(\tilde{M})$ by $(\chi, \tilde{\sigma}) \mapsto \tilde{\sigma} \otimes \chi$, where $\chi \in X(\tilde{M})$ and $\tilde{\sigma} \in \Pi(\tilde{M})$. The induced action of $\mathrm{Im}\, X(\tilde{M})$ preserves $\Pi_2(\tilde{M})$. For an orbit $\mathcal{O}$ in $\Pi_2(\tilde{M})$ under the action of $\mathrm{Im}(X(\tilde{M}))$, we may chose a base point $\tilde{\sigma} \in \mathcal{O}$, and then furnish the orbit with $\mathcal{O}$ a C^∞ variety structure via the isomorphism

$$\mathrm{Im}\, X(\tilde{M})/\,\mathrm{Stab}_{\mathrm{Im}\, X(\tilde{M})}(\tilde{\sigma}) \xrightarrow{\sim} \mathcal{O} : \chi \mapsto \tilde{\sigma} \otimes \chi.$$

Here $\mathrm{Stab}_{\mathrm{Im}\,X(\tilde{M})}(\tilde{\sigma})$ is a finite group. Analogously, we may attach a complex algebraic variety structure on an orbit $\mathcal{O}_{\mathbb{C}}$ for the action of $X(\tilde{M})$ on $\Pi(\tilde{M})$. In this case, we may talk about C^∞, regular and rational functions on $\mathcal{O}_{\mathbb{C}}$. For a genuine discrete series representation $\tilde{\sigma} \in \Pi_2(\tilde{M})$, we define the action of $W(M)$ on $\tilde{\sigma} \in \Pi_2(\tilde{M})$ and a parabolic subgroup $\tilde{P} = \tilde{M}N$ by

$$w.\tilde{\sigma}(\tilde{m}) := \tilde{\sigma}(\tilde{w}^{-1}\tilde{m}\tilde{w})$$

and $w.\tilde{P} = \tilde{w}\tilde{P}\tilde{w}^{-1}$ for $\tilde{w} \in \tilde{G}$ a representative of $w \in W(M)$ and any $\tilde{m} \in \tilde{M}$. We set $W(\tilde{\sigma}) := \{\omega \in W(M) : \tilde{w}.\tilde{\sigma} = \tilde{\sigma}\}$. In order to introduce the Knapp–Stein dimension theorem for finite central covering groups, we would like to summarize some necessary definitions and the associated properties in [Waldspurger 2003; Li 2012; Ban and Jantzen 2013] which are rather standard and will be needed later on.

Bruhat decomposition. (see [Ban and Jantzen 2013, Lemma 2.6]): For $\Theta \subset \Delta$, we attach a parabolic subgroup $P_\Theta = M_\Theta N_\Theta$ of G and the associated $\tilde{P}_\Theta = \tilde{M}_\Theta N$ of $\tilde{G}$. Denote

$${}^{P_\Theta}W^{P_\Theta} := W_{M_\Theta}(T)\backslash W_G(T)/W_{M_\Theta}(T) = \{\omega \in W_G(T) : \omega^{-1}.\Theta \subset \Phi^+, \omega.\Theta \subset \Phi^+\}.$$

Then we have the following decomposition of $\tilde{G}$:

$$\tilde{G} = \bigsqcup_{\omega \in {}^{P_\Theta}W^{P_\Theta}} \tilde{P}_\Theta \tilde{\omega} \tilde{P}_\Theta,$$

where $\tilde{\omega}$ is an arbitrary lifting of ω in $\tilde{G}$.

Bernstein–Zelevinsky geometric lemma (see [Ban and Jantzen 2013, Proposition 3.3] or [Waldspurger 2003, Section I.3]). Let $P = MU$ and $Q = LV$ be standard parabolic subgroups of G. For $\tilde{\sigma} \in \Pi(\tilde{M})$, we have, in the Grothendieck group,

$$J_{\tilde{Q}} \circ I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma}) = \sum_{\omega \in {}^Q W^P} I_{\tilde{L}\cap\tilde{\omega}.\tilde{P}}^{\tilde{L}} \circ \tilde{\omega} \circ J_{\tilde{M}\cap\tilde{\omega}^{-1}.\tilde{Q}}(\tilde{\sigma}).$$

Here $J_{\tilde{Q}}$ stands for the normalized Jacquet functor with respect to $\tilde{Q}$, and $I_{\tilde{P}}^{\tilde{G}}$ for the normalized induced representation with respect to $\tilde{P}$. In particular, for $\tilde{\sigma} \in \Pi_2(\tilde{M})$, the geometric lemma gives rise to

$$\dim \mathrm{End}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})) \le |W(M)|.$$

Intertwining operator. Fix two semistandard parabolic subgroups $P_1 = MU_1$ and $P_2 = MU_2$ of G, and $(\tilde{\sigma}, V) \in \Pi(\tilde{M})$. Under some continuation on $\mathcal{O}_{\mathbb{C}} \ni \tilde{\sigma}$ (see [Waldspurger 2003; Li 2012]), we may define an intertwining operator $J_{\tilde{P}_2|\tilde{P}_1}(\tilde{\sigma})$ as

$$I_{\tilde{P}_1}^{\tilde{G}}(\tilde{\sigma}) \xrightarrow{J_{\tilde{P}_2|\tilde{P}_1}(\tilde{\sigma})} I_{\tilde{P}_2}^{\tilde{G}}(\tilde{\sigma}) : f(\tilde{g}) \mapsto J_{\tilde{P}_2|\tilde{P}_1}(\tilde{\sigma})(f)(\tilde{g}) = \int_{U_1\cap U_2\backslash U_2} f(u\tilde{g})\,du.$$

For later use, we mention the following properties of $J_{\tilde{P}_2|\tilde{P}_1}$ in [Waldspurger 2003, Section IV.3]:

• (Plancherel measure) For $\tilde{\sigma} \in \Pi(\tilde{M})$, and a semistandard parabolic $P = MN$ of G, denote by $\overline{P} = M\overline{N}$ the unique opposite parabolic subgroup of P in G. By the generic irreducibility property of $\mathrm{Ind}(\tilde{\sigma})$ (see [Renard 2010, Section VI.8.5]), we may define

$$j_{\tilde{P}}(\tilde{\sigma}) := J_{\tilde{P}|\bar{\tilde{P}}}(\tilde{\sigma}) J_{\bar{\tilde{P}}|\tilde{P}}(\tilde{\sigma})$$

as a scalar which does not depend on $P \supset M$, and denoted by $j(\tilde{\sigma})$. For simplicity, for $\tilde{\sigma} \in \Pi_2(\tilde{M})$, we define the Plancherel measure attached to $\tilde{\sigma}$ as $\mu^{\tilde{G}}(\tilde{\sigma}) := j(\tilde{\sigma})^{-1}$. Then we have

– $\mu^{\tilde{G}}(\tilde{\sigma}) \geq 0$;
– $\mu^{\tilde{G}}(\tilde{\sigma}) = \prod_\alpha \mu^{\tilde{M}_\alpha}(\tilde{\sigma})$, where α runs over all the reduced roots of T_M, up to sign.
– $\mu^{\tilde{G}}(\tilde{\omega}.\tilde{\sigma}) = \mu^{\tilde{G}}(\tilde{\sigma})$, where $\omega \in W_G(T)$.
– $\mu^{\tilde{G}}(\tilde{\sigma}) = \mu^{\tilde{G}}(\check{\tilde{\sigma}})$. Here $\check{\tilde{\sigma}}$ is the contragredient of $\tilde{\sigma}$.

• (associativity) For $P_i = MU_i,\ i = 1, 2, 3$,

$$J_{\tilde{P}_3|\tilde{P}_2}(\tilde{\sigma}) \cdot J_{\tilde{P}_2|\tilde{P}_1}(\tilde{\sigma}) = \Bigl(\prod j_\alpha(\tilde{\sigma})\Bigr) J_{\tilde{P}_3|\tilde{P}_1}(\tilde{\sigma}),$$

where the product is taken over $\Phi_{\mathrm{red}}(P_1) \cap \Phi_{\mathrm{red}}(P_3) \cap \Phi_{\mathrm{red}}(\overline{P}_2)$. In particular, for $\omega \in W(M)$,

$$J_{\tilde{P}|\tilde{\omega}.\tilde{P}}(\tilde{\sigma}) J_{\tilde{\omega}.\tilde{P}|\tilde{P}}(\tilde{\sigma}) = \prod_{\alpha \in \Phi_{\mathrm{red}}(P) \cap \Phi_{\mathrm{red}}(\overline{\omega.P})} j_\alpha(\tilde{\sigma}),$$

where $j_\alpha(\tilde{\sigma})$ is the j-constant with respect to $(M, \tilde{\sigma})$ with M_α in place of G.

Harish-Chandra c-function. Recall that for $(\tilde{\sigma}, V) \in \mathcal{O}_{\mathbb{C}} \subset \Pi(\tilde{M})$, we denote by $(\check{\tilde{\sigma}}, \check{V})$ the contragredient representation of $(\tilde{\sigma}, V)$. For a semistandard parabolic subgroup $P = MU$ of G, we define

$$L(\tilde{\sigma}, \tilde{P}) := I_{\tilde{P}}^{\tilde{G}}(V) \otimes I_{\tilde{P}}^{\tilde{G}}(\check{V}) \hookrightarrow \mathrm{End}(I_{\tilde{P}}^{\tilde{G}}(V)),$$

and the matrix coefficient map

$$E_{\tilde{P}}^{\tilde{G}} : L(\tilde{\sigma}, \tilde{P}) \to C^\infty(\tilde{G}) : \nu \otimes \check{\nu} \mapsto \langle \tilde{\pi}(\tilde{g})\nu, \check{\nu} \rangle,$$

where $\tilde{\pi} = I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$. For $\omega \in W(M)$, we define the Harish-Chandra c-function as

$$c_{\tilde{P}|\tilde{P}}(\omega, \tilde{\sigma}) : L(\tilde{\sigma}, \tilde{P}) \xrightarrow{\lambda(\omega)} L(\tilde{\omega}.\tilde{\sigma}, \tilde{\omega}.\tilde{P}) \xrightarrow{J_{\tilde{P}|\tilde{\omega}.\tilde{P}}(\tilde{\omega}.\tilde{\sigma}) \otimes J_{\bar{\tilde{P}}|\tilde{\omega}.\tilde{P}}(\tilde{\omega}.\tilde{\sigma})} L(\tilde{\omega}.\tilde{\sigma}, \tilde{P});$$

here $\lambda(\omega) : \phi(\,\cdot\,) \mapsto \phi(\tilde{\omega}^{-1}\,\cdot\,)$, and $\tilde{\omega} \in \tilde{K}_{\mathrm{good}}$ where K_{good} is a special compact open subgroup of G which is in good position with respect to M_0. Furthermore, we define

$${}^0c_{\tilde{P}|\tilde{P}}(\omega, \tilde{\sigma}) := c_{\tilde{P}|\tilde{P}}(1, \tilde{\omega}.\tilde{\sigma})^{-1} c_{\tilde{P}|\tilde{P}}(\omega, \tilde{\sigma}) \in \mathrm{Hom}_{\tilde{G}\times\tilde{G}}(L(\tilde{\sigma}, \tilde{P}), L(\tilde{\omega}.\tilde{\sigma}, \tilde{P})).$$

Note that these operators ${}^0c_{\tilde{P}|\tilde{P}}(\omega, \tilde{\sigma})$ for $\omega \in W(\tilde{\sigma})$ have the following properties (see [Waldspurger 2003, Section V.3] or [Li 2012, Section 2.5]).

- (regularity) They are regular on $\mathcal{O}$ and unitary operators.
- (associativity) For $\omega_1, \omega_2 \in W(M)$, $\tilde{\sigma} \in \mathcal{O} \subset \Pi_2(\tilde{M})$ and $P_i = MU_i$ with $i = 1, 2, 3$, based on the associativity of $J_{\tilde{P}'|\tilde{P}}$, we have the equality
$$ {}^0c_{\tilde{P}_3|\tilde{P}_2}(\omega_1, \tilde{\omega}_2.\tilde{\sigma})\, {}^0c_{\tilde{P}_2|\tilde{P}_1}(\omega_2, \tilde{\sigma}) = {}^0c_{\tilde{P}_3|\tilde{P}_1}(\omega_1\omega_2, \tilde{\sigma}). $$

Plancherel formula (see [Li 2012, Section 2.6]). Let $\mathcal{C}(\tilde{G})$ be the Schwartz–Harish-Chandra function space of $\tilde{G}$, and $C^\infty(\mathcal{O}, \tilde{P})$ the space of C^∞-functions on $\mathcal{O}$, i.e., $\psi[\mathcal{O}, P] : \tilde{\sigma} \to \psi[\mathcal{O}, P]_{\tilde{\sigma}} \in L(\tilde{\sigma}, \tilde{P})$, such that it is compatible with the isomorphism class of $\tilde{\sigma}$. Denote by Θ the set of pairs $(\mathcal{O}, P = MU)$, where P is a semistandard parabolic subgroup of G and $\mathcal{O} \subset \Pi_2(\tilde{M})$ an orbit under the action of $\operatorname{Im} X(\tilde{M})$. Let $C^\infty(\Theta) := \bigoplus_{(\mathcal{O},P)\in\Theta} C^\infty(\mathcal{O}, \tilde{P})$, and write an element $\psi \in C^\infty(\Theta)$ in the form of $\psi = (\psi[\mathcal{O}, P])_{\mathcal{O},P}$. We set $C^\infty(\Theta)^{\mathrm{inv}}$ to be the subspace of $C^\infty(\Theta)$ consisting of the elements ψ such that
$$ \psi[\tilde{\omega}.\mathcal{O}, P']_{\tilde{\omega}.\tilde{\sigma}} = {}^0c_{\tilde{P}'|\tilde{P}}(\omega, \tilde{\sigma})\psi[\mathcal{O}, P]_{\tilde{\sigma}}, $$
for all $(\mathcal{O}, P) \in \Theta$ and all P'. For $f \in \mathcal{C}(\tilde{G})$, we define a map $\iota : \mathcal{C}(\tilde{G}) \to C^\infty(\Theta)^{\mathrm{inv}}$ as
$$ f \mapsto (\psi_f[\mathcal{O}, P] : \tilde{\sigma} \mapsto d(\tilde{\sigma}) I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})(\check{f}))_{\mathcal{O},P}, $$
where $d(\tilde{\sigma})$ is the formal degree of $\tilde{\sigma}$, and $\check{f}(\tilde{g}) = f(\tilde{g}^{-1})$. In rough terms, given both $\mathcal{C}(\tilde{G})$ and $C^\infty(\Theta)^{\mathrm{inv}}$ the natural direct limit topology with respect to open compact subgroups, the Plancherel formula says that ι is an isomorphism of topological spaces, and the associated inverse map κ is given by
$$ \psi \mapsto \sum_{(\mathcal{O},P)\in\Theta} \gamma(G|M)|W_M(T)|\,|W_G(T)|^{-1}|\mathcal{P}(M)|^{-1} f_{\psi[\mathcal{O},P]}, $$
where $\mathcal{P}(M)$ is the set of all parabolic subgroups $P = MU$ with Levi group M,
$$ \gamma(G|M) = \int_{\bar{U}} \delta_P(m_p(\bar{u}))\, d\bar{u} \quad \text{and} \quad f_{\psi[\mathcal{O},P]}(\tilde{g}) = \int_{\mathcal{O}} \mu^{\tilde{G}}(\tilde{\sigma})(E_{\tilde{P}}^{\tilde{G}} \psi[\mathcal{O}, P]_{\tilde{\sigma}})(\tilde{g})\, d\tilde{\sigma}. $$

2B. ***Knapp–Stein dimension theorem.*** Now we are ready to prove the Knapp–Stein dimension theorem for finite central covering groups.

For $(\tilde{\sigma}, V) \in \mathcal{O} \subset \Pi_2(\tilde{M})$, and $\omega \in W(\tilde{\sigma})$, denote by $A(\omega)$ the unique isomorphism, up to scalars of $|\cdot| = 1$, between $\tilde{\omega}.\tilde{\sigma}$ and $\tilde{\sigma}$, and extend $A(\omega)$ to be an isomorphism between $I_{\tilde{P}}^{\tilde{G}}(\tilde{\omega}.\tilde{\sigma})$ and $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$. Note that
$$ A(\omega) \circ {}^0c_{\tilde{P}|\tilde{P}}(\omega, \tilde{\sigma}) \in \operatorname{End}_{\tilde{G}\times\tilde{G}}(L(\tilde{\sigma}, \tilde{P})) $$
is regular on $\mathcal{O}$ and unitary. Note that the $\tilde{G}\times\tilde{G}$-equivalent space $\operatorname{End}_{\tilde{G}\times\tilde{G}}(L(\tilde{\sigma}, \tilde{P}))$

is finite-dimensional. Applying the Skolem–Noether theorem, based on the associativity property of ${}^0c_{\tilde P|\tilde P}(\omega,\tilde\sigma)$, we define a projective unitary representation $\omega\mapsto\gamma(\omega)$ of $W(\tilde\sigma)$ on the underling vector space of the induced representation $I_{\tilde P}^{\tilde G}(\tilde\sigma)$, such that $A(\omega)\circ{}^0c_{\tilde P|\tilde P}(\omega,\tilde\sigma)=\mathrm{Ad}(\gamma(\omega))$ on $L(\tilde\sigma,\tilde P)\hookrightarrow\mathrm{End}(I_{\tilde P}^{\tilde G}(V))$. Notice that the adjoint $\gamma(\omega)^*$ of $\gamma(\omega)$ exists and equals $\gamma(\omega^{-1})$ up to a scalar. So the vector space $\Gamma:=\mathrm{Span}\{\gamma(\omega):\omega\in W(\tilde\sigma)\}$ is a selfadjoint algebra, and hence semisimple. Before turning to the Harish-Chandra commuting algebra theorem, we first discuss the explicit form of $\gamma(\omega)$. Denote $\Phi(\tilde\sigma):=\{\alpha\in\Phi_{\mathrm{red}}(P):\mu^{M_\alpha}(\tilde\sigma)=0\}$; then $W^0(\tilde\sigma):=\langle S_\alpha:\alpha\in\Phi(\tilde\sigma)\rangle\subset W(\tilde\sigma)$, where S_α is the simple reflection associated to α (see [Waldspurger 2003, Proposition IV.2.2]).

Lemma 2.1. *Indeed, $\Phi(\tilde\sigma)$ is a subroot system, which in turn says that $W^0(\tilde\sigma)$ is a normal subgroup of $W(\tilde\sigma)$.*

Proof. For the convenience of the reader, we include the argument in [Harish-Chandra 1976, Section 40] as follows. Notice that this is equivalent to showing $\Phi(\tilde\sigma)$ is stable under the action of $W(\tilde\sigma)$, i.e., $\omega.\alpha\in\Phi(\tilde\sigma)$ for $\alpha\in\Phi(\tilde\sigma)$ and $\omega\in W(\tilde\sigma)$. The upshot is for $\lambda\in\mathrm{Im}\,\mathfrak{a}^*_{M,\mathbb{C}}$, and $\alpha\in\Phi_{\mathrm{red}}(P)$,

$$\mu^{\tilde M_\alpha}(\tilde\sigma\otimes\lambda)>0\text{ unless }\langle\alpha,\lambda\rangle=0.$$

Take $P=MU$, and let $\alpha_1,\dots,\alpha_r$ be all the distinct elements in $\Phi(\tilde\sigma)$. Assume $\alpha=\alpha_1$. Let j be the unique index such that $\omega.\alpha=\pm\alpha_j$. Fix $\lambda\in\mathrm{Im}\,\mathfrak{a}^*_{M,\mathbb{C}}$, such that

$$\langle\alpha,\lambda\rangle=0\quad\text{and}\quad\langle\alpha_i,\lambda\rangle\neq0,\quad 2\le i\le r.$$

This implies $\mu^{\tilde M_\alpha}(\tilde\sigma\otimes\lambda)=\mu^{\tilde M_\alpha}(\tilde\sigma)=0$. Thus $\mu^{\tilde G}(\tilde\sigma\otimes\lambda)=\prod_{1\le i\le r}\mu^{\tilde M_{\alpha_i}}(\tilde\sigma\otimes\lambda)=0$. Note $\tilde\omega.\tilde\sigma=\tilde\sigma$, so $0=\mu^{\tilde G}(\tilde\sigma\otimes\lambda)=\mu^{\tilde G}(\tilde\omega.\tilde\sigma\otimes\omega.\lambda)=\mu^{\tilde G}(\tilde\sigma\otimes\omega.\lambda)$. This implies $\mu^{\tilde M_{\alpha_i}}(\tilde\sigma\otimes\omega.\lambda)=0$ for some i. But it is clear that $\langle\alpha_i,\omega.\lambda\rangle\neq0$ unless $i=j$, which is to say $0=\mu^{\tilde M_{\alpha_j}}(\tilde\sigma\otimes\omega.\lambda)=\mu^{\tilde M_{\alpha_j}}(\tilde\sigma)$, whence $\alpha_j\in\Phi(\tilde\sigma)$. □

In view of this, we may define the R-group

$$R(\tilde\sigma):=W(\tilde\sigma)/W^0(\tilde\sigma)=\{\omega\in W(\tilde\sigma):\omega.\Delta(\tilde\sigma)=\Delta(\tilde\sigma)\},$$

where $\Delta(\tilde\sigma)$ is the set of simple roots in $\Phi(\tilde\sigma)$.

Lemma 2.2. *Keeping the same notation as above, for $\omega\in W^0(\tilde\sigma)$, we have*

$${}^0c_{\tilde P|\tilde P}(\omega,\tilde\sigma)=\mathrm{id},\ \textit{i.e.},\ \gamma(\omega)=\mathrm{id}.$$

On the other hand, for $\omega\in R(\tilde\sigma)$, we have

$$\gamma(\omega)=A(\omega)\circ J_{\tilde P|\tilde\omega.\tilde P}(\tilde\omega.\tilde\sigma)\circ\lambda(\omega),$$

where $\lambda(\omega):I_{\tilde P}^{\tilde G}(\tilde\sigma)\xrightarrow{\sim}I_{\tilde\omega.\tilde P}^{\tilde G}(\tilde\omega.\tilde\sigma)$ is the canonical isomorphism given by $\phi(\cdot)\mapsto\phi(\tilde\omega^{-1}\cdot)$.

Proof. For the first part, by the associativity property of the Harish-Chandra c-function ${}^0c_{\tilde{P}|\tilde{P}}$, it suffices to prove that ${}^0c_{\tilde{P}|\tilde{P}}(S_\alpha,\tilde{\sigma}) = \mathrm{id}$ for $\alpha \in \Phi(\tilde{\sigma})$. Setting $P_\alpha = P \cap M_\alpha$, the functoriality property of ${}^0c_{\tilde{P}|\tilde{P}}$ says that

$${}^0c_{\tilde{P}|\tilde{P}}(S_\alpha,\tilde{\sigma}) = {}^0c_{\tilde{P}_\alpha|\tilde{P}_\alpha}(S_\alpha,\tilde{\sigma})|_{L(\tilde{\sigma},\tilde{P})}.$$

On the other hand, applying Savin's results on the maximal parabolic subgroup case in [Savin 2017, Proposition 2], we then have

$${}^0c_{\tilde{P}|\tilde{P}}(S_\alpha,\tilde{\sigma}) = {}^0c_{\tilde{P}_\alpha|\tilde{P}_\alpha}(S_\alpha,\tilde{\sigma})|_{L(\tilde{\sigma},\tilde{P})} = \mathrm{id}.$$

As for the second part, this results from the associativity of $J_{\tilde{P}|\tilde{\omega}.\tilde{P}}(\,\cdot\,)$, i.e.,

$$J_{\tilde{P}|\tilde{\omega}.\tilde{P}}(\,\cdot\,)J_{\tilde{\omega}.\tilde{P}|\tilde{P}}(\,\cdot\,) = \prod_{\alpha\in\Phi_{\mathrm{red}}(P)\cap\Phi_{\mathrm{red}}(\overline{\omega.P})} j_\alpha(\,\cdot\,).$$

Notice that for $\alpha \in \Phi_{\mathrm{red}}(P)\cap\Phi_{\mathrm{red}}(\overline{\omega.P})$, $\mu^{\tilde{M}_\alpha}(\tilde{\omega}.\tilde{\sigma}) = \mu^{\tilde{M}_\alpha}(\tilde{\sigma}) \neq 0$, so $J_{\tilde{P}|\tilde{\omega}.\tilde{P}}(\tilde{\omega}.\tilde{\sigma})$ and $J_{\tilde{\omega}.\tilde{P}|\tilde{P}}(\tilde{\omega}.\check{\tilde{\sigma}})$ are holomorphic at $\tilde{\sigma}$, and hence invertible. In this case, based on the general associativity property of $J_{\tilde{P}_2|\tilde{P}_1}$, we have

$$A(\omega)\circ {}^0c_{\tilde{P}|\tilde{P}}(\omega,\tilde{\sigma}) = A(\omega)\circ J_{\tilde{P}|\tilde{\omega}.\tilde{P}}(\tilde{\omega}.\tilde{\sigma})\circ\lambda(\omega) \bigotimes A(\omega)\circ J_{\tilde{\omega}.\tilde{P}|\tilde{P}}(\tilde{\omega}.\check{\tilde{\sigma}})\circ\lambda(\omega),$$

which in turn implies that

$$\gamma(\omega) = A(\omega)\circ J_{\tilde{P}|\tilde{\omega}.\tilde{P}}(\tilde{\omega}.\tilde{\sigma})\circ\lambda(\omega). \qquad \square$$

Remark. We note that [Savin 2017, Proposition 2] concerns unitary supercuspidal representations, but the argument applies to discrete series as well based on the following facts (see [Waldspurger 2003, Lemme III.3.1, Corollaire III.7.3]):

- $\mathrm{Hom}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}\tilde{\sigma}, I_{\tilde{P}}^{\tilde{G}}\tilde{\sigma}) = \mathrm{Hom}_{\tilde{M}}(J_{\tilde{M}}^{\omega}(I_{\tilde{P}}^{\tilde{G}}\tilde{\sigma}),\tilde{\sigma})$, where $J_{\tilde{M}}^{\omega}$ stands for the tempered part of the normalized Jacquet module $J_{\tilde{M}}$.
- $J_{\tilde{M}}^{\omega}(I_{\tilde{P}}^{\tilde{G}}\tilde{\sigma}) = \sum_{\omega\in W(M)}\omega.\tilde{\sigma}$ as virtual representations.

Note that $A(\omega)\circ J_{\tilde{P}|\tilde{\omega}.\tilde{P}}(\tilde{\omega}.\tilde{\sigma})\circ\lambda(\omega)$ is nothing but the well-known intertwining operator $A(\omega)\circ M(\omega,\tilde{\sigma})$ which is defined by

$$I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma}) \xrightarrow{M(\omega,\tilde{\sigma})} I_{\tilde{P}}^{\tilde{G}}(\tilde{\omega}.\tilde{\sigma}) : f \mapsto \int_{U\cap\tilde{\omega}U\tilde{\omega}^{-1}\backslash U} f(\tilde{\omega}^{-1}u\tilde{g})\,du.$$

In the following lemma, we show the linear independence of $\{\gamma(\omega) : \omega \in R(\tilde{\sigma})\}$ adapting the argument in [Ban 2004].

Lemma 2.3. *For $\omega \in R(\tilde{\sigma})$, $\gamma(\omega)$ are linearly independent on $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$.*

Proof. For the convenience of the reader, we recall the argument in [Ban 2004, Theorem 4.3] as follows. The upshot is to construct a function $f \in I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$ with the following "separation" property for the nontrivial $\omega \in W(M)$:

- $f(1) = 0$.

- $M(\omega,\tilde{\sigma})f(1)$ is absolutely convergent and nonzero.
- For any $\omega_1 \in W(M)$ with $\omega_1 \not\geq \omega$, $M(\omega_1,\tilde{\sigma})f(1)=0$.

Such a function is constructed as follows. Fix a nonzero element ν in $(\tilde{\sigma}, V)$. Let K be a compact open subgroup of G splitting in $\tilde{G}$, such that $\delta^{1/2}(k)\tilde{k}\nu=\nu$ for all $k \in K \cap \tilde{M}$. In addition, we may assume K is invariant under conjugation by $\omega \in W(M)$. Notice that $\tilde{P}(\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega}\cap K)$ is open in $\tilde{P}(\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega})$. Hence we may choose a compact subgroup $K_0 \subset K$ which is invariant under conjugation by $\omega \in W(M)$, such that

$$K_0\cap\tilde{P}(\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega})\subset\tilde{P}(\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega}\cap K)\quad\text{and}\quad\tilde{\omega}^{-1}K_0\subset\tilde{G}_{\omega^{-1}},$$

where $\tilde{G}_{\omega^{-1}} := \bigcup_{\omega'\geq\omega^{-1}}\tilde{P}\tilde{\omega}'\tilde{P}$. Then we may define the "separation" function f as

$$f(\tilde{g})=\begin{cases}\delta^{1/2}(m)\tilde{\sigma}(\tilde{m})\nu & \text{if } \tilde{g}=\tilde{m}u\tilde{\omega}^{-1}k\in\tilde{P}\tilde{\omega}^{-1}K_0,\\ 0 & \text{otherwise.}\end{cases}$$

It is easy to see such an f is well defined and belongs to $I_{\tilde{P}}^{\tilde{G}}(V)$: If $\tilde{p}_1\tilde{\omega}^{-1}k_1=\tilde{p}_2\tilde{\omega}^{-1}k_2$, for $\tilde{p}_1,\tilde{p}_2\in\tilde{P}$ and $k_1,k_2\in K_0$, we then have

$$\tilde{p}_2^{-1}\tilde{p}_1=\tilde{\omega}^{-1}k_2k_1^{-1}\tilde{\omega}\in K_0,$$

which in turn implies that, as $\delta^{1/2}(k)\tilde{\sigma}(k)\nu=\nu$ for $k\in K$,

$$\delta^{1/2}(\tilde{p}_1)\tilde{\sigma}(\tilde{p}_1)\nu=\delta^{1/2}(\tilde{p}_2)\tilde{\sigma}(\tilde{p}_2)\nu.$$

On the other hand, we have

$$\operatorname{supp}(f)\subset\tilde{P}\tilde{\omega}^{-1}K_0\subset\tilde{G}_{\omega^{-1}}:=\bigcup_{\omega'\geq\omega^{-1}}\tilde{P}\tilde{\omega}'\tilde{P},$$

and $\operatorname{supp}(f)\cap\tilde{P}=\varnothing$. Observe that

$$\begin{aligned}M(\omega,\tilde{\sigma})f(1)&=\int_{U\cap\tilde{\omega}U\tilde{\omega}^{-1}\backslash U}f(\tilde{\omega}^{-1}u)\,du\\&=\int_{\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega}}f(u\tilde{\omega}^{-1})\,du=\int_{\tilde{P}K_0\cap\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega}}f(u\tilde{\omega}^{-1})\,du.\end{aligned}$$

Notice that for $u\in\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega}\cap\tilde{P}K_0$, we write $u=\tilde{p}k_0$; then

$$\tilde{p}^{-1}u=k_0\in K_0\cap\tilde{P}(\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega})\subset\tilde{P}(\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega}\cap K),$$

thus $u\in\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega}\cap K\subset K$ and $\tilde{p}\in K$, therefore

$$f(u\tilde{\omega}^{-1})=f(\tilde{p}K_0\tilde{\omega}^{-1})=\delta^{1/2}(\tilde{p})\tilde{\sigma}(\tilde{p})\nu=\nu,$$

which in turn says that

$$M(\omega,\tilde{\sigma})f(1)=\operatorname{mes}(\overline{U}\cap\tilde{\omega}^{-1}U\tilde{\omega}\cap PK_0)\nu\neq 0.$$

The remaining vanishing statement is easy. ☐

To finish the proof of Knapp–Stein dimension theorem for finite central covering groups, it remains to prove Harish-Chandra commuting algebra theorem. For simplicity, we state Harish-Chandra commuting algebra theorem and Knapp–Stein dimension theorem together as follows.

Theorem 2.4. *Keeping the same notation as before, we have*

$$\mathrm{End}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})) = \mathrm{Span}\{\gamma(\omega) : \omega \in W(\tilde{\sigma})\} = \mathrm{Span}\{\gamma(\omega) : \omega \in R(\tilde{\sigma})\}.$$

Therefore, $\dim \mathrm{End}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})) = |R(\tilde{\sigma})|$.

Proof. This follows from the Plancherel formula stated in Section 2A using the same argument as in [Silberger 1979, Theorem 5.5.3.2]. For the convenience of the reader, we sketch the main ideas as follows. Denote

$$L^0(\tilde{\sigma}, \tilde{P}) := \{\phi \in L(\tilde{\sigma}, \tilde{P}) : {}^0c_{\tilde{P}|\tilde{P}}(\omega, \tilde{\sigma})\phi = \phi \text{ for all } \omega \in W(\tilde{\sigma})\}.$$

Recall $\Gamma = \mathrm{Span}\{\gamma(\omega) : \omega \in W(\tilde{\sigma})\} \subset \mathrm{End}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma}))$.

Step 1. The centralizer $C_{L(\tilde{\sigma},P)}(\Gamma)$ of Γ in $L(\tilde{\sigma}, P)$ satisfies $C_{L(\tilde{\sigma},P)}(\Gamma) = L^0(\tilde{\sigma}, \tilde{P})$, and $L^0(\tilde{\sigma}, \tilde{P}) = I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})(\mathcal{C}(\tilde{G}))$. The latter plays the key role which follows from the Plancherel formula, i.e., the isomorphism

$$\mathcal{C}(\tilde{G}) \xrightarrow{\sim} C^\infty(\Theta)^{\mathrm{inv}}.$$

Step 2. $C_{L(\tilde{\sigma},P)}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})(\mathcal{C}(\tilde{G}))) = \mathrm{End}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma}))$. This follows from the definition.

Step 3. The Wedderburn double centralizer theorem says that

$$\Gamma = C_{L(\tilde{\sigma},P)}(C_{L(\tilde{\sigma},P)}(\Gamma)) = \mathrm{End}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})).$$

In order to apply the Wedderburn double centralizer theorem, one has to consider some finite-dimensional subspaces $L(\sigma, P)^{K\times K}$ of $L(\sigma, P)$ consisting of $K \times K$-invariant vectors, where K is an open-compact subgroup of G which splits in $\tilde{G}$ and is small enough, as in [Silberger 1979].

To be precise, let K be a sufficiently small open-compact subgroup of G which splits in $\tilde{G}$, then we have

$$\mathrm{End}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})) = \mathrm{End}_{\mathcal{C}(\tilde{G})}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})) = \mathrm{End}_{\mathcal{C}_K(\tilde{G})}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})^K),$$

where $\mathcal{C}_K(\tilde{G})$ is the subspace of double K-invariant functions in $\mathcal{C}(\tilde{G})$, and $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})^K$ is the subspace of K-invariant vectors in $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$.

Denoting by Γ_K the restriction of the action of Γ to $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})^K$, we have $\dim \Gamma_K = \dim \Gamma$. As $\Gamma \subset L(\tilde{\sigma}, P)$ is semisimple, thus

$$\Gamma_K = C_{L(\tilde{\sigma},P)^{K\times K}}(C_{L(\tilde{\sigma},P)^{K\times K}}(\Gamma_K)) = C_{L(\tilde{\sigma},P)^{K\times K}}(L^0(\tilde{\sigma}, P)^{K\times K}).$$

On the other hand, Harish-Chandra's Plancherel formula implies that

$$L^0(\tilde{\sigma}, P)^{K\times K} = I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})(\mathcal{C}_K(\tilde{G})),$$

which in turn implies that

$$\begin{aligned}\Gamma_K &= C_{L(\tilde{\sigma},P)^{K\times K}}(L^0(\tilde{\sigma}, P)^{K\times K})\\ &= C_{L(\tilde{\sigma},P)^{K\times K}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})(\mathcal{C}_K(\tilde{G}))) = \mathrm{End}_{\mathcal{C}_K(\tilde{G})}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})^K),\end{aligned}$$

whence

$$\Gamma = \Gamma_K = \mathrm{End}_{\mathcal{C}_K(\tilde{G})}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})^K) = \mathrm{End}_{\tilde{G}}(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})).$$ □

2C. ***An example: R-group for genuine unramified principal series.*** The decomposition of tempered induced representation $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$ is determined by our $R(\tilde{\sigma})$-group, especially intertwining operators. Recall that the associativity property of intertwining operators says that

$$J_{\tilde{P}_3|\tilde{P}_2}(\tilde{\sigma})J_{\tilde{P}_2|\tilde{P}_1}(\tilde{\sigma}) = \Big(\prod j_\alpha(\tilde{\sigma})\Big)J_{\tilde{P}_3|\tilde{P}_1}(\tilde{\sigma}).$$

We normalize those intertwining operators $J_{P'|P}(\tilde{\sigma})$ for covering groups by a factor $r_{P'|P}(\tilde{\sigma})$ as done by Arthur for linear groups (please refer to [Li 2012] for the details) so that our normalized intertwining operators

$$R_{P'|P}(\tilde{\sigma}) := r_{P'|P}(\tilde{\sigma})^{-1}J_{P'|P}(\tilde{\sigma})$$

satisfy some natural properties, for example

$$R_{\tilde{P}_3|\tilde{P}_2}(\tilde{\sigma})R_{\tilde{P}_2|\tilde{P}_1}(\tilde{\sigma}) = R_{\tilde{P}_3|\tilde{P}_1}(\tilde{\sigma}).$$

As in [Arthur 1993], we then define the normalized intertwining operators

$$R(w, \tilde{\sigma}) := A(w)\circ\lambda(w)\circ R_{w^{-1}.P|P}(\tilde{\sigma}) : I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma}) \to I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$$

which satisfy

$$\Gamma := \mathrm{Span}\{\gamma(w) : w \in R(\tilde{\sigma})\} = \mathrm{Span}\{R(w, \tilde{\sigma})\}.$$

On the other hand, the definition of $A(w)\circ\lambda(w)$ depends on the lift of w in $\tilde{K}_{\text{good}}$. For simplicity, we use the same letter w to be the fixed lifting of $w \in W_G(T)$ if no confusion arises. In general, $w \to R(w, \tilde{\sigma})$ is not a homomorphism, but we have the formula

$$R(w_1w_2, \tilde{\sigma}) = \eta_{\tilde{\sigma}}(w_1, w_2)R(w_1, \tilde{\sigma})R(w_2, \tilde{\sigma}), \quad w_1, w_2 \in R(\tilde{\sigma}),$$

where

$$\eta_{\tilde{\sigma}}(w_1, w_2) = A(w_1w_2)\circ\lambda(w_1w_2)\circ\lambda(w_2)^{-1}\circ A(w_2)^{-1}\circ\lambda(w_1)^{-1}\circ A(w_1)^{-1}$$

is a 2-cocycle for $R(\tilde{\sigma})$ with values in $\mathbb{C}^\times$. Thus the image $\bar{\eta}_{\tilde{\sigma}}$ of $\eta_{\tilde{\sigma}}$ in $H^2(R(\tilde{\sigma}), \mathbb{C}^\times)$

gives the obstruction of extending the representation $\tilde{\sigma}$ to the groups generated by $\tilde{M}$ and $\{\tilde{w} \in \tilde{K}_{\text{good}} : w \in R(\tilde{\sigma})\}$.

As in [Arthur 1993, page 87], we deal with the problem by fixing a finite central extension

$$1 \to Z_{\tilde{\sigma}} \to \bar{R}_{\tilde{\sigma}} \to R(\tilde{\sigma}) \to 1$$

over which $\eta_{\tilde{\sigma}}$ splits. We then choose a function $\kappa_{\tilde{\sigma}} : \bar{R}_{\tilde{\sigma}} \to \mathbb{C}^\times$ such that $\eta_{\tilde{\sigma}}$ splits, i.e.,

$$\eta_{\tilde{\sigma}}(r_1, r_2) = \kappa_{\tilde{\sigma}}(r_1 r_2)\kappa_{\tilde{\sigma}}(r_2)^{-1}\kappa_{\tilde{\sigma}}(r_1)^{-1}, \qquad r_1, r_2 \in \bar{R}_{\tilde{\sigma}},$$

where $\eta_{\tilde{\sigma}}$ is identified with its pullback to $\bar{R}_{\tilde{\sigma}} \times \bar{R}_{\tilde{\sigma}}$. It follows that

$$\kappa_{\tilde{\sigma}}(zr) = \chi_{\tilde{\sigma}}(z)\kappa_{\tilde{\sigma}}(r), \quad z \in Z_{\tilde{\sigma}},\ r \in \bar{R}_{\tilde{\sigma}},$$

where $\chi_{\tilde{\sigma}}$ is a character on the central subgroup $Z_{\tilde{\sigma}}$. We twist our intertwining operators by $\kappa_{\tilde{\sigma}}$, i.e.,

$$\bar{R}(r, \tilde{\sigma}) := \kappa_{\tilde{\sigma}}(r)^{-1} R(r, \tilde{\sigma}), \quad r \in \bar{R}_{\tilde{\sigma}}$$

which gives rise to a homomorphism of $\bar{R}_{\tilde{\sigma}}$ to the group of unitary intertwining operators for $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$ satisfying

$$\bar{R}(zr, \tilde{\sigma}) = \chi_{\tilde{\sigma}}(z)^{-1}\bar{R}(r, \tilde{\sigma}), \quad z \in Z_{\tilde{\sigma}},\ r \in \bar{R}_{\tilde{\sigma}}.$$

Therefore we obtain a representation R of $\bar{R}_{\tilde{\sigma}} \times \tilde{G}$ on the underling vector space $\mathcal{H}_{\tilde{\sigma}}$ of $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$, i.e.,

$$R(r, g) := \bar{R}(r, \tilde{\sigma}) I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma}, g), \quad r \in \bar{R}_{\tilde{\sigma}},\ g \in \tilde{G}.$$

Thus our Knapp-Stein dimension theorem, i.e., Theorem 2.4, implies that

$$R = \bigoplus_{\rho} \check{\rho} \otimes \pi_\rho,$$

where ρ runs over the set $\Pi(\bar{R}_{\tilde{\sigma}})_{\chi_{\tilde{\sigma}}}$ of irreducible representations of $\bar{R}_{\tilde{\sigma}}$ with $Z_{\tilde{\sigma}}$-central character $\chi_{\tilde{\sigma}}$, and $\check{\rho}$ is the contragredient representation of ρ, while $\pi_\rho \in JH(I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma}))$.

It is well known that such a 2-cocycle $\bar{\eta}_{\tilde{\sigma}}$ is trivial if $\Pi(\bar{R}_{\tilde{\sigma}})_{\chi_{\tilde{\sigma}}}$ contains a one-dimensional representation, thus giving:

Lemma 2.5 (D. Keys). *Keep the notions as above. If the tempered induction $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$ contains a constituent which is of multiplicity one, then the* 2*-cocycle $\bar{\eta}_{\tilde{\sigma}}$ is trivial.*

Proof. Since $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$ contains a constituent which is of multiplicity one, the decomposition

$$R = \bigoplus_{\rho \in \Pi(\bar{R}_{\tilde{\sigma}})_{\chi_{\tilde{\sigma}}}} \check{\rho} \otimes \pi_\rho$$

implies $\Pi(\bar{R}_{\tilde{\sigma}})_{\chi_{\tilde{\sigma}}}$ contains a one-dimensional representation, so our claim holds. □

A typical example of such a situation is when $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$ is an unramified genuine unitary principal series, that is:

Corollary 2.6. *Keep the notions as above. For genuine unramified unitary principal series $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$, the representation R of $R_{\tilde{\sigma}} \times \tilde{G}$ on the underling vector space $\mathcal{H}_{\tilde{\sigma}}$ of $I_{\tilde{P}}^{\tilde{G}}(\tilde{\sigma})$ decomposes as*

$$R = \bigoplus_{\rho \in \Pi(R_{\tilde{\sigma}})} \check{\rho} \otimes \pi_\rho.$$

In what follows, we would like to investigate R-groups for genuine unramified unitary principal series of the tame Brylinski–Deligne n-fold covering group $\tilde{G}$ of a split simply connected group G defined over the nonarchimedean local field F, where tame means that n and p are coprime. Under this setting, those R-groups are isomorphic to the associated R-groups of the incarnation split linear group G_n which gives rise to the same Langlands dual group $\tilde{G}^\vee$; moreover G_n is an isogeny to G' which has the same or dual root system of G depending on the cover (see [Savin 2004; Gan and Gao 2018; Weissman 2018]). Note that R-groups for G are well known (see [Keys 1982]). So it reduces to investigating the relation of R-groups under isogeny. Let $p : G \to G_n$ be the isogeny map. Restricting to their maximal torus gives $p : T \to T_n$ and $p* : \Pi(T_n) \to \Pi(T)$, i.e., $\chi_n \mapsto \chi := \chi_n \circ p$. Therefore $W_{\chi_n} := \{w \in W_G(T) : w.\chi_n = \chi_n\} < W_\chi := \{w \in W_G(T) : w.\chi = \chi\}$ as the map p is $W_G(T)$-equivalent. On the other hand, for $\chi \in \Pi_2(T)$ and a root $\alpha \in \Phi := \Phi(G, T)$, it is well known that the corank one Plancherel measure $\mu_\alpha(\chi) := \mu^{M_\alpha}(\chi)$ is equal to 0 if and only if $\chi_\alpha := \chi \circ H_{\alpha^\vee} = 1$ (see [Winarsky 1978]), where $H_{\alpha^\vee}$ is the one-parameter subgroup given by α under Harish-Chandra homomorphism (see [Waldspurger 2003]). A similar criterion holds for covering groups (please refer to [Goldberg and Szpruch 2016] for the details). Note that $\chi_\alpha = 1$ says that $(\chi_n)_\alpha = 1$, which implies that $S_\alpha.\chi_n = \chi_n$. In either PGL_2 or SL_2, such a corank one unitary induction is always irreducible, thus $\mu_\alpha(\chi_n) = 0$, i.e., $W_{\chi_n}^0 := \langle S_\alpha : \mu_\alpha(\chi_n) = 0, \alpha \in \Phi(G, T)\rangle = W_\chi^0 := \langle S_\alpha : \mu_\alpha(\chi) = 0, \alpha \in \Phi(G, T)\rangle$.

Lemma 2.7. *Retain the notions as above. We have*

$$R_{\chi_n} := W_{\chi_n}/W_{\chi_n}^0 \lhd R_\chi := W_\chi/W_\chi^0.$$

Proof. This is equivalent to showing that $W_{\chi_n} \lhd W_\chi$, i.e., for any $w \in W_\chi$ and $w_n \in W_{\chi_n}$, $w_n.(w.\chi_n) = (w.\chi_n)$. But $w \in W_\chi$ implies that $w.\chi_n = \chi_n \chi_c$ for some $\chi_c \in \Pi_2(T_n/p(T))$. Note that any $\chi_c \in \Pi_2(T_n/p(T))$ is $W_G(T)$-invariant which follows from the fact that $S_\alpha.y - y$ lies in the coroot lattice of T for any coroot y of T_n and $\alpha \in \Phi(G, T)$. Thus $w_n.(w.\chi_n) = w_n(\chi_n \chi_c) = w_n.\chi_n \chi_c = (w.\chi_n)$. □

It is also well known that R-groups for unitary unramified principal series of adjoint groups are trivial (see [Li 1992, Corollary 2.6]). Thus the nontrivial R-

groups for split semisimple groups which have not been discussed in [Keys 1982] and [Goldberg 1994] are as follows:

Corollary 2.8 (Keys). *$G_n^\vee = SL_\iota(\mathbb{C})/\mu_m : R \simeq \mathbb{Z}/d\mathbb{Z}$ with $d|m|\iota$.*

2D. *An example: R-group of Mp_{2n}.* In what follows, we discuss some properties of R-groups for Mp_{2n}. Let us first introduce a simple fact. Recall that we have the following decomposition of $\tilde{R} \times G$ acting on $I_P^G(\sigma)$:

$$I_P^G(\sigma) = \bigoplus_{\rho \in \Pi(\tilde{R})_{\chi_\sigma}} \check{\rho} \bigotimes \pi_\rho.$$

As an easy corollary, we have the following criterion on the abelian property of R.

Corollary 2.9. *If $\Pi_{\chi_\sigma}(\tilde{R})$ consists of one-dimensional representations, then $R \simeq \tilde{R}/Z$ is abelian.*

Proof. This results from the following fact: For a finite group G and a subgroup $H < Z(G)$, fix a character χ of H, if as G-modules

$$(\star) \qquad \mathrm{Ind}_H^G(\chi) = \bigoplus_{i=1}^{|G/H|} \chi_i;$$

then G/H is abelian.

Note that if χ is trivial, then this is quite obvious. As for χ nontrivial, we may consider a new set of characters $S := \{\chi_1^{-1} \cdot \chi_i : i = 1, \dots, |G/H|\}$. It is easy to see $\chi_1^{-1}\chi_i \neq \chi_1^{-1}\chi_j$ for $i \neq j$, and these are the characters of G which are trivial on H, which in turn says

$$\mathrm{Ind}_H^G(1) = \bigoplus_{i=1}^{|G/H|} \chi_1^{-1}\chi_i,$$

whence $(\star)$ holds. □

In view of the above corollary, based on Gan and Savin's work on local theta correspondence [2012], we have:

Corollary 2.10. *Keeping the notation as before, $R(\tilde{\sigma})$ is abelian, and*

$$R(\tilde{\sigma}) = R(\Theta(\tilde{\sigma})).$$

Here $(\tilde{\sigma}, \Theta(\tilde{\sigma}))$ is a Howe duality pair under the local theta correspondence for (Mp_{2n}, SO_{2n+1}).

Proof. The first part follows from the preservation of multiplicities in tempered inductions under the local theta correspondence. The second part follows from the preservation of Plancherel measures under the local theta correspondence. □

Remark. We recently learned that M. Hanzer [2019] had described the R-group for Mp_{2n} using the local theta correspondence.

Acknowledgements

This is the last part of the author's thesis. He is much indebted to Professor Wee Teck Gan for his guidance and numerous discussions on various topics. The paper was initially meant to be an appendix to [Gao et al. 2019], but it turned out to be insufficiently related. The author also wants to thank the referee for detailed comments.

References

[Arthur 1993] J. Arthur, "On elliptic tempered characters", *Acta Math.* **171**:1 (1993), 73–138. MR Zbl

[Ban 2004] D. Ban, "Linear independence of intertwining operators", *J. Algebra* **271**:2 (2004), 749–767. MR Zbl

[Ban and Jantzen 2013] D. Ban and C. Jantzen, "The Langlands quotient theorem for finite central extensions of p-adic groups", *Glas. Mat. Ser. III* **48(68)**:2 (2013), 313–334. MR Zbl

[Gan and Gao 2018] W. T. Gan and F. Gao, "The Langlands–Weissman program for Brylinski–Deligne extensions", pp. 187–275 in *L-groups and the Langlands program for covering groups*, Astérisque **398**, Société Mathématique de France, Paris, 2018. MR

[Gan and Savin 2012] W. T. Gan and G. Savin, "Representations of metaplectic groups I: Epsilon dichotomy and local Langlands correspondence", *Compos. Math.* **148**:6 (2012), 1655–1694. MR Zbl

[Gao et al. 2019] F. Gao, F. Shahidi, and D. Szpruch, "Local coefficients and gamma factors for principal series of covering groups", preprint, 2019. arXiv

[Goldberg 1994] D. Goldberg, "Reducibility of induced representations for $\mathrm{Sp}(2n)$ and $\mathrm{SO}(n)$", *Amer. J. Math.* **116**:5 (1994), 1101–1151. MR Zbl

[Goldberg and Szpruch 2016] D. Goldberg and D. Szpruch, "Plancherel measures for coverings of p-adic $\mathrm{SL}_2(F)$", *Int. J. Number Theory* **12**:7 (2016), 1907–1936. MR Zbl

[Hanzer 2019] M. Hanzer, "R-groups for metaplectic groups", *Israel J. Math.* **231**:1 (2019), 467–488. MR Zbl

[Harish-Chandra 1976] Harish-Chandra, "Harmonic analysis on real reductive groups, III: The Maass–Selberg relations and the Plancherel formula", *Ann. of Math.* (2) **104**:1 (1976), 117–201. MR Zbl

[Keys 1982] D. Keys, "Reducibility of unramified unitary principal series representations of p-adic groups and class-1 representations", *Math. Ann.* **260**:4 (1982), 397–402. MR Zbl

[Knapp and Stein 1971] A. W. Knapp and E. M. Stein, "Intertwining operators for semisimple groups", *Ann. of Math.* (2) **93** (1971), 489–578. MR Zbl

[Knapp and Stein 1980] A. W. Knapp and E. M. Stein, "Intertwining operators for semisimple groups, II", *Invent. Math.* **60**:1 (1980), 9–84. MR Zbl

[Knapp and Zuckerman 1982] A. W. Knapp and G. J. Zuckerman, "Classification of irreducible tempered representations of semisimple groups", *Ann. of Math.* (2) **116**:2 (1982), 389–455. MR Zbl

[Langlands 1989] R. P. Langlands, "On the classification of irreducible representations of real algebraic groups", pp. 101–170 in *Representation theory and harmonic analysis on semisimple Lie groups*, edited by P. J. Sally, Jr., Math. Surveys Monogr. **31**, Amer. Math. Soc., Providence, RI, 1989. MR Zbl

[Li 1992] J.-S. Li, "Some results on the unramified principal series of p-adic groups", *Math. Ann.* **292**:4 (1992), 747–761. MR Zbl

[Li 2012] W.-W. Li, "La formule des traces pour les revêtements de groupes réductifs connexes, II: Analyse harmonique locale", *Ann. Sci. Éc. Norm. Supér.* (4) **45**:5 (2012), 787–859. MR Zbl

[Li 2014] W.-W. Li, "La formule des traces pour les revêtements de groupes réductifs connexes, I: Le développement géométrique fin", *J. Reine Angew. Math.* **686** (2014), 37–109. MR Zbl

[Luo 2017] C. Luo, "Howe finiteness conjecture for covering groups", *Math. Res. Lett.* **24**:6 (2017), 1759–1774. MR Zbl

[Mœglin and Tadić 2002] C. Mœglin and M. Tadić, "Construction of discrete series for classical p-adic groups", *J. Amer. Math. Soc.* **15**:3 (2002), 715–786. MR Zbl

[Mœglin and Waldspurger 1995] C. Mœglin and J.-L. Waldspurger, *Spectral decomposition and Eisenstein series: a paraphrase of scripture*, Cambridge Tracts in Mathematics **113**, Cambridge University Press, 1995. MR

[Renard 2010] D. Renard, *Représentations des groupes réductifs p-adiques*, Cours Spécialisés **17**, Société Mathématique de France, Paris, 2010. MR Zbl

[Savin 2004] G. Savin, "On unramified representations of covering groups", *J. Reine Angew. Math.* **566** (2004), 111–134. MR Zbl

[Savin 2017] G. Savin, "An addendum to Casselman's notes", (2017). Appendix to "Principal series representations of metaplectic groups" by Shiang Tang. arXiv

[Silberger 1978] A. J. Silberger, "The Knapp–Stein dimension theorem for p-adic groups", *Proc. Amer. Math. Soc.* **68**:2 (1978), 243–246. Correction in **76**:1 (1979), 169–170. MR Zbl

[Silberger 1979] A. J. Silberger, *Introduction to harmonic analysis on reductive p-adic groups*, Mathematical Notes **23**, Princeton University Press, 1979. MR Zbl

[Szpruch 2013] D. Szpruch, "Some irreducibility theorems of parabolic induction on the metaplectic group via the Langlands–Shahidi method", *Israel J. Math.* **195**:2 (2013), 897–971. MR Zbl

[Waldspurger 2003] J.-L. Waldspurger, "La formule de Plancherel pour les groupes p-adiques (d'après Harish-Chandra)", *J. Inst. Math. Jussieu* **2**:2 (2003), 235–333. MR Zbl

[Weissman 2018] M. H. Weissman, "L-groups and parameters for covering groups", pp. 33–186 in *L-groups and the Langlands program for covering groups*, Astérisque **398**, Société Mathématique de France, Paris, 2018. MR

[Winarsky 1978] N. Winarsky, "Reducibility of principal series representations of p-adic Chevalley groups", *Amer. J. Math.* **100**:5 (1978), 941–956. MR Zbl

Received November 10, 2019. Revised December 21, 2019.

CAIHUA LUO
DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
GOTEBORG
SWEDEN

chluo@amss.ac.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.281

SOME CLASSIFICATIONS OF BIHARMONIC HYPERSURFACES WITH CONSTANT SCALAR CURVATURE

SHUN MAETA AND YE-LIN OU

We give some classifications of biharmonic hypersurfaces with constant scalar curvature. These include biharmonic Einstein hypersurfaces in space forms, compact biharmonic hypersurfaces with constant scalar curvature in a sphere, and some complete biharmonic hypersurfaces of constant scalar curvature in space forms and in a nonpositively curved Einstein space. Our results provide additional cases (Theorem 2.3 and Proposition 2.8) that support the conjecture that a biharmonic submanifold in S^{m+1} has constant mean curvature, and two more cases that support Chen's conjecture on biharmonic hypersurfaces (Corollaries 2.2 and 2.7).

1. Introduction

Biharmonic maps, as a generalization of harmonic maps, are maps between Riemannian manifolds which are critical points of the bienergy functional. Biharmonic submanifolds are the images of biharmonic isometric immersions and they include minimal submanifolds as special cases. As in the study of minimal submanifolds, a fundamental problem in the study of biharmonic submanifolds is to classify nonminimal biharmonic submanifolds (called proper biharmonic submanifolds) in a model space. For example, when the ambient space is a space form, we have the following conjectures which are still far beyond our reach.

Conjecture (Chen's conjecture on biharmonic submanifolds of Euclidean space [1991; 2015]). *Any biharmonic submanifold in a Euclidean space is minimal.*

Maeta is supported partially by the Grant-in-Aid for Young Scientists(B), No.15K17542, Japan Society for the Promotion of Science, and partially by JSPS Overseas Research Fellowships 2017-2019 No. 70. The work was done while he was visiting the Department of Mathematics of Texas A & M University–Commerce as a Visiting Scholar and he is grateful to the department and the university for the hospitality he received during the visit. Ou is supported by a grant from the Simons Foundation (#427231, Ye-Lin Ou).

MSC2010: primary 58E20; secondary 53C12.

Keywords: biharmonic hypersurfaces, Einstein manifolds, constant scalar curvature, constant mean curvature.

Conjecture (Balmuş, Montaldo and Oniciuc's conjectures on biharmonic submanifolds of spheres [2008; 2012]). (1) *A proper biharmonic submanifold of a sphere has constant mean curvature.*

(2) *The only proper biharmonic hypersurface of* S^{m+1} *is a part of* $S^m(1/\sqrt{2})$ *or* $S^p(1/\sqrt{2}) \times S^q(1/\sqrt{2})$ *with* $p+q=m$ *and* $p \neq q$.

A lot of work related to these conjectures has been done since 2000; some recent developments, partial results and open problems in the topics can be found in the recent survey [Ou 2016].

For a hypersurface $\varphi : (M^m, g) \to (N^{m+1}, h)$ in a Riemannian manifold, i.e., an isometric immersion of codimension 1, we choose a local unit normal vector field ξ with respect to which, we have the second fundamental form $B(X, Y) = b(X, Y)\xi$, where $b : TM \times TM \to C^\infty(M)$ is the function-valued second fundamental form. The Gauss and the Weingarten formulae read, respectively,

$$\tilde{\nabla}_X Y = (\tilde{\nabla}_X Y)^\top + (\tilde{\nabla}_X Y)^\perp = \nabla_X Y + b(X, Y)\xi, \tag{1}$$

and

$$\tilde{\nabla}_X \xi = (\tilde{\nabla}_X \xi)^\top + (\tilde{\nabla}_X \xi)^\perp = -AX + \nabla^\perp_X \xi = -AX, \tag{2}$$

where A is the shape operator of the hypersurface with respect to the unit normal vector ξ.

We use R^M, Ric^M, and Scal^M (respectively, R^N, Ric^N, and Scal^N) to denote the Riemannian, Ricci and scalar curvatures of (M^m, g) (respectively, of (N, h)) with the following conventions:

$$\begin{aligned}
\mathrm{R}^M(X, Y, Z, W) &= \langle \mathrm{R}^M(Z, W)Y, X\rangle,\\
\mathrm{R}^M(Z, W)Y &= \nabla_Z\nabla_W Y - \nabla_W\nabla_Z Y - \nabla_{[Z,W]}Y,\\
\mathrm{Ric}^M(X, Y) &= \sum_{i=1}^m \mathrm{R}^M(X, e_i, Y, e_i) \quad \text{for an orthonormal base } \{e_i\}.
\end{aligned}$$

Using these, together with (1) and (2), we have the Gauss equation

$$\mathrm{R}^N(X, Y, Z, W) = \mathrm{R}^M(X, Y, Z, W) + b(X, W)b(Y, Z) - b(X, Z)b(Y, W), \tag{3}$$

where

$$b(X, Y) = \langle AX, Y\rangle = \langle B(X, Y), \xi\rangle, \tag{4}$$

for any $X, Y, Z, W \in TM$.

From (3) we have the relationship between the Ricci curvatures of the hypersurface and the ambient space

$$\mathrm{Ric}^N(X, Y) = \mathrm{Ric}^M(X, Y) + \langle AX, AY\rangle - mH\langle AX, Y\rangle + \mathrm{R}^N(X, \xi, Y, \xi), \tag{5}$$

and the relationship between the scalar curvatures is

$$\mathrm{Scal}^N = \mathrm{Scal}^M + |A|^2 - m^2 H^2 + 2\,\mathrm{Ric}^N(\xi,\xi). \tag{6}$$

It was proved in [Ou 2010] that a hypersurface $\varphi : M^m \to N^{m+1}$ with mean curvature vector $\eta = H\xi$ is biharmonic if and only if

$$\begin{cases} \Delta H - H|A|^2 + H\mathrm{Ric}^N(\xi,\xi) = 0, \\ 2A\,(\mathrm{grad}\, H) + \frac{m}{2}\,\mathrm{grad}\ H^2 - 2\,H\,(\mathrm{Ric}^N\,(\xi))^\top = 0, \end{cases} \tag{7}$$

where $\mathrm{Ric}^N : T_q N \to T_q N$ denotes the Ricci operator of the ambient space defined by $\langle \mathrm{Ric}^N\,(Z),\, W\rangle = \mathrm{Ric}^N(Z,\, W)$.

In particular, for biharmonic hypersurfaces in an Einstein space, we have:

Corollary 1.1 [Ou 2010]. *A hypersurface $\varphi : M^m \to (N^{m+1}, h)$ of an Einstein manifold with $\mathrm{Ric}^N = \lambda h$ is biharmonic if and only if its mean curvature function H solves the equations*

$$\begin{cases} \Delta H - H(|A|^2 - \lambda) = 0, \\ A\,(\mathrm{grad}\ H) + \frac{m}{2} H\,\mathrm{grad}\ H = 0. \end{cases} \tag{8}$$

In this note, we give some classifications of biharmonic hypersurfaces with constant scalar curvature. These include biharmonic Einstein hypersurfaces in space forms, compact biharmonic hypersurfaces with constant scalar curvature in a sphere, and some complete biharmonic hypersurfaces with constant scalar curvature in space forms and in a nonpositively curved Einstein space. Our results provide a further case (Theorem 2.3) that supports Balmuş, Montaldo and Oniciuc's conjecture that a biharmonic submanifold in S^{m+1} has constant mean curvature, and two more cases that support Chen's conjecture on biharmonic hypersurfaces (Corollaries 2.2 and 2.7).

2. Biharmonic hypersurfaces with constant scalar curvature in Einstein spaces

Recall that a Riemannian manifold (M^m, g) is called an Einstein space if its Ricci curvature is proportional to the metric, i.e., $\mathrm{Ric}^M = \lambda g$. It is well known that every 2-dimensional manifold is an Einstein space; any space form is of Einstein, and any 3-dimensional Einstein space has to be a space form. One can also check that for $m \geq 3$, if (M^m, g) is an Einstein space with $\mathrm{Ric}^M = \lambda g$, then λ has to be a constant and hence an Einstein space (M^m, g) of dimension $m \geq 3$ has constant scalar curvature since $\mathrm{Scal}^M = m\lambda$.

Our first result is the following theorem which gives a classification of biharmonic Einstein hypersurfaces in a space form.

Theorem 2.1. *An Einstein hypersurface $M^m \hookrightarrow (N^{m+1}(C), h)$ $(m \geq 3)$ in a space form is biharmonic if and only if it is minimal or $|A|^2 = mC$. In the latter case, the hypersurface has positive scalar curvature, i.e.,* $\mathrm{Scal}^M = m(m-2)C + m^2H^2 > 0$.

Proof. Suppose that the mean curvature function H is not constant; then there exists an open neighborhood U of M on which $\nabla H \neq 0$. Substituting $X = Y = \nabla H$ into (5) gives

$$\mathrm{Ric}^M(\nabla H, \nabla H) = (m-1)C|\nabla H|^2 + mHg(A(\nabla H), \nabla H) - g(A(\nabla H), A(\nabla H)). \tag{9}$$

Substituting the second equation $A(\nabla H) = -\frac{m}{2}H\nabla H$ of the biharmonic equation for a hypersurface into the above equation we have

$$\mathrm{Ric}^M(\nabla H, \nabla H) = \big((m-1)C - \tfrac{3}{4}m^2H^2\big)|\nabla H|^2. \tag{10}$$

If (M^m, g) is an Einstein hypersurface with $\mathrm{Ric}^M = \mu g$ for some constant μ, then (10) becomes

$$\mu|\nabla H|^2 = \big((m-1)C - \tfrac{3}{4}m^2H^2\big)|\nabla H|^2. \tag{11}$$

Since $\nabla H \neq 0$ on $U \subset M$, we have $\mu = (m-1)C - \frac{3}{4}m^2H^2$, which implies H is a constant on $U \subset M$ since μ is a constant. This contradicts the assumption that $\nabla H \neq 0$ on $U \subset M$. The contradiction shows that the mean curvature function H has to be a constant. It follows from the first equation $\Delta H - H(|A|^2 - mC) = 0$ of (8) that either $H = 0$ and the hypersurface is minimal, or $|A|^2 = mC$, and in this case, $C > 0$. Using the scalar curvature $\mathrm{Scal}^N = (m+1)mC > 0$ of $(N^{m+1}(C), h)$, $|A|^2 = mC$ and (6) we have $\mathrm{Scal}^M = m(m-2)C + m^2H^2 > 0$.

Thus, we obtain the theorem. □

As an immediate consequence of Theorem 2.1, we have:

Corollary 2.2. *A biharmonic Einstein hypersurface of Euclidean space $\mathbb{R}^{m+1}$ or hyperbolic space H^{m+1} is minimal. A proper biharmonic Einstein hypersurface in S^{m+1} has constant mean curvature and $|A|^2 = m$.*

It was proved in [Chen 1993] (see also [Balmuş et al. 2012]) that for $m \geq 2$, if a compact biharmonic hypersurface in sphere $S^{m+1}(1)$ with the squared norm of the second fundamental form satisfies $|A|^2 \leq m$, then $|A|^2 = 0$, or $|A|^2 = m$ and it has constant mean curvature. Also, Fu [2015] proved that a biharmonic hypersurface with constant scalar curvature in 5-dimensional space forms $M^5(C)$ has constant mean curvature, and in [Fu and Hong 2018], it was proved that a biharmonic hypersurface with constant scalar curvature and at most six distinct principal curvatures in space forms $M^{m+1}(C)$ has constant mean curvature. Our next theorem shows that the same result as in [Fu and Hong 2018] holds when we replace the principal curvature assumption by the compactness of the hypersurface.

Theorem 2.3. *A compact hypersurface with constant scalar curvature* $(M^m, g) \hookrightarrow S^{m+1}$ *in a sphere is biharmonic if and only if it is minimal, or it has nonzero constant mean curvature, and* $|A|^2 = m$.

Proof. If the scalar curvature Scal^M of the hypersurface is constant, then, by (6), we have $|A|^2 = m^2H^2 + \text{constant}$, from which we have

$$\nabla|A|^2 = 2m^2 H\nabla H. \tag{12}$$

Using this, together with the first equation of (8) with $\lambda = m$, we have

$$\nabla\Delta H = \nabla[(|A|^2 - m)H] = (|A|^2 - m + 2m^2H^2)\nabla H. \tag{13}$$

On the other hand, we have the following estimate of the squared norm of the Hessian of H:

$$\begin{aligned}|\nabla \mathrm{d}H|^2 &= \sum_{i,j=1}^m [\nabla \mathrm{d}H(e_i, e_j)]^2 \geq \sum_{i=1}^m [\nabla \mathrm{d}H(e_i, e_i)]^2 \geq \frac{1}{m}\left(\sum_{i=1}^m \nabla \mathrm{d}H(e_i, e_i)\right)^2\\ &= \frac{1}{m}(\Delta H)^2.\end{aligned} \tag{14}$$

Substituting (10), (13) and (14) into the Bochner formula for $|\nabla H|^2$, we have

$$\begin{aligned}\tfrac{1}{2}\Delta|\nabla H|^2 &= |\nabla \mathrm{d}H|^2 + \mathrm{Ric}^M(\nabla H, \nabla H) + \langle \nabla H, \nabla\Delta H\rangle\\ &\geq \frac{1}{m}(\Delta H)^2 + \big(|A|^2 - 1 + \tfrac{5}{4}m^2H^2\big)|\nabla H|^2.\end{aligned} \tag{15}$$

Since M is compact, we integrate both sides of (15) to get

$$\int_M \Big[\tfrac{1}{m}(\Delta H)^2 + \big(|A|^2 - 1 + \tfrac{5}{4}m^2H^2\big)|\nabla H|^2\Big] dv_g \leq \int_M \tfrac{1}{2}\Delta|\nabla H|^2 dv_g = 0. \tag{16}$$

Using compactness of M, (13), and the divergence theorem we have

$$\begin{aligned}\frac{1}{m}\int_M \Delta H\Delta H dv_g &= \frac{1}{m}\int_M \Delta H \operatorname{div}(\nabla H) dv_g = -\frac{1}{m}\int_M \langle\nabla\Delta H, \nabla H\rangle dv_g\\ &= -\frac{1}{m}\int_M (|A|^2 - m + 2m^2H^2)|\nabla H|^2 dv_g,\end{aligned} \tag{17}$$

Substituting this into (16) we have

$$0 \leq \int_M \Big(\Big(1 - \frac{1}{m}\Big)|A|^2 + \frac{m}{4}(5m-8)H^2\Big)|\nabla H|^2 dv_g \leq 0. \tag{18}$$

It follows that $\big[\big(1 - \frac{1}{m}\big)|A|^2 + \frac{m}{4}(5m-8)H^2\big]|\nabla H|^2 = 0$ from which, together with Newton's inequality $|A|^2 \geq mH^2$, we have

$$0 = \Big[\Big(1 - \frac{1}{m}\Big)|A|^2 + \frac{m}{4}(5m-8)H^2\Big]|\nabla H|^2 \geq \tfrac{1}{16}(5m^2 - 4m - 4)|\nabla H^2|^2. \tag{19}$$

From this, we conclude that $H = \text{constant}$. In the case where $H = \text{constant} \neq 0$, we use the first equation of (8) to have $|A|^2 = m$. Thus, we obtain the theorem. $\square$

Remark 1. (i) Notice that Theorem 2.3, together with the results in [Chen 1993] and [Balmuş et al. 2012], implies that for a compact biharmonic hypersurface in a sphere, if one of the data Scal^M, H and $|A|^2$ is constant, then so are the other two.

(ii) We would like to point out that according to a classical result of Chern, do Carmo and Kobayashi [Chern et al. 1970], any compact minimal hypersurface in S^{m+1} with $|A|^2 = m$ is locally the Clifford tori $S^p \times S^{m-p}$. On the other hand, a hypersurface of S^{m+1} with constant mean curvature and $|A|^2 = m$ is biharmonic. So it would be very interesting to classify hypersurfaces of S^{m+1} with constant mean curvature and $|A|^2 = m$. This would be an important case to solve the second of Balmuş, Montaldo and Oniciuc's conjectures.

(iii) Finally, we remark that there are infinitely many compact hypersurfaces of constant scalar curvatures in a sphere. In fact, it was proved in [Ejiri 1982] that there exist countably many isometric immersions of $S^1 \times S^{m-1}$ into a sphere S^{m+1} so that $S^1 \times S^{m-1}$ is a warped product of constant scalar curvature $m(m-1)$ with respect to the induced metric.

Theorem 2.4. *A compact Einstein hypersurface $M^m \hookrightarrow (N^{m+1}, h)$ in an Einstein manifold with $\mathrm{Ric}^N = \lambda h$ is biharmonic if and only if it is minimal or $|A|^2 = \lambda$. Furthermore, in the latter case, the hypersurface has positive scalar curvature, i.e.,*

$$\mathrm{Scal}^M = (m-2)\lambda + m^2 H^2 > 0.$$

Proof. Suppose the hypersurface M is Einstein with $\mathrm{Ric}^M = \mu g$, then, by (6),

$$(m+1)\lambda = m\mu + |A|^2 - m^2 H^2 + 2\lambda.$$

Hence,

$$\mu = \left(1 - \frac{1}{m}\right)\lambda - \frac{1}{m}|A|^2 + mH^2. \tag{20}$$

It follows that

$$\begin{aligned}\langle \nabla H, \nabla \Delta H \rangle &= \langle \nabla H, \nabla (H|A|^2 - \lambda H) \rangle \\ &= |A|^2 |\nabla H|^2 + H \langle \nabla H, \nabla |A|^2 \rangle - \lambda |\nabla H|^2 \\ &= (|A|^2 - \lambda + 2m^2 H^2)|\nabla H|^2,\end{aligned} \tag{21}$$

where the first equality was obtained by using the first equation of (8), and the third equality follows from (6) and the fact that the scalar curvature of an Einstein hypersurface is constant.

Using these and a similar computation used in obtaining (15) we have,

$$\begin{aligned}(22)\qquad \tfrac{1}{2}\Delta|\nabla H|^2 &= |\nabla \mathrm{d}H|^2+\mathrm{Ric}^M(\nabla H,\nabla H)+\langle\nabla H,\nabla\Delta H\rangle\\ &\geq \frac{1}{m}(\Delta H)^2+\mu|\nabla H|^2+(|A|^2-\lambda+2m^2H^2)|\nabla H|^2\\ &= \frac{1}{m}(\Delta H)^2+\Big(-\frac{1}{m}\lambda+\Big(1-\frac{1}{m}\Big)|A|^2+m(2m+1)H^2\Big)|\nabla H|^2.\end{aligned}$$

Similar to (17), we have

$$\begin{aligned}(23)\qquad \frac{1}{m}\int_M \Delta H\Delta H dv_g &= \frac{1}{m}\int_M \Delta H \mathrm{div}(\nabla H)dv_g = -\frac{1}{m}\int_M\langle\nabla\Delta H,\nabla H\rangle dv_g\\ &= -\frac{1}{m}\int_M(|A|^2-\lambda+2m^2H^2)|\nabla H|^2 dv_g.\end{aligned}$$

Integrating both sides of (22) and using the compactness of M and (23) we obtain

$$\begin{aligned}(24)\qquad 0 &\geq \int_M\Big(\frac{1}{m}(m-2)|A|^2+m(2m-1)H^2\Big)|\nabla H|^2 dv_g\\ &\geq \tfrac{1}{4}m(2m-1)\int_M|\nabla H^2|^2 dv_g.\end{aligned}$$

It follows that H is constant. If $H=0$, then M is minimal. If $H\neq 0$, by the first equation of (8), we have $|A|^2=\lambda$. Thus, $\mathrm{Scal}^M=m\mu=(m-2)\lambda+m^2H^2>0$. $\square$

For the classification of complete biharmonic hypersurfaces, we notice that it was proved in [Nakauchi and Urakawa 2011], [Maeta 2014a] and [Luo 2015] that a complete hypersurface $M^m\hookrightarrow(N^{m+1},h)$ in a manifold of nonpositive Ricci curvature with mean curvature function $H\in L^p(M)$ for some $0<p<\infty$ is minimal. We will give a classification of biharmonic hypersurfaces of constant scalar curvatures in an Einstein manifold using a condition on the maximum rate of change of the mean curvature function. For that purpose, we will need the following maximum principles:

Theorem 2.5 [Yau 1976; Karp and Li 1982]. *Let M be a complete Riemannian manifold and f be a smooth function on M. If one of the following conditions is satisfied, then f is constant.*

(i) *M has Ricci curvature bounded from below and $\Delta f\geq\varepsilon f$ (for some $\varepsilon>0$) for an upper bounded function $f\geq 0$.*

(ii) *$\Delta f\geq 0$ for $f\geq 0$ and $f\in L^p(M)$ for some $1<p<\infty$.*

(iii) *$f\in L^1(M)$, $\Delta f\geq 0$, $f\geq 0$ and the Ricci curvature is bounded from below by $-c\{1+r^2(x)\}$, where $r(x)$ is the distance function on M.*

Now we are ready to give some classifications of biharmonic hypersurfaces with constant scalar curvature in a nonpositively curved Einstein manifold.

Proposition 2.6. *A complete biharmonic hypersurface $M^m \hookrightarrow (N^{m+1}, h)$ of constant scalar curvature in a nonpositively curved Einstein manifold (N^{m+1}, h) is minimal if one of the following occurs*:

(a) $|\nabla H| \in L^p$ *for some* $2 < p < \infty$.

(b) $|\nabla H| \in L^2$ *and the Ricci curvature is bounded from below by* $-c\{1 + r^2(x)\}$, *where* $r(x)$ *is the distance function on M.*

Proof. If the ambient space (N^{m+1}, h) is a nonpositively curved Einstein space with $\widetilde{\mathrm{Ric}} = \lambda h$, then from (5) we have

$$\begin{aligned}\text{(25)}\qquad \mathrm{Ric}(\nabla H, \nabla H) &\geq \lambda |\nabla H|^2 + mH\langle A(\nabla H), \nabla H\rangle - \langle A(\nabla H), A(\nabla H)\rangle \\ &= \bigl(\lambda - \tfrac{3}{4}m^2H^2\bigr)|\nabla H|^2,\end{aligned}$$

where the equality was obtained by using the second equation of the biharmonic hypersurface equation (8).

Substituting (25) and (21) into the Bochner formula we have

$$\begin{aligned}\text{(26)}\qquad \Delta|\nabla H|^2 &= 2\{|\nabla \mathrm{d}H|^2 + \mathrm{Ric}(\nabla H, \nabla H) + \langle \nabla H, \nabla \Delta H\rangle\} \\ &\geq 2\{\mathrm{Ric}(\nabla H, \nabla H) + \langle \nabla H, \nabla \Delta H\rangle\} \\ &= 2\bigl[\bigl(\lambda - \tfrac{3}{4}m^2H^2\bigr) + (|A|^2 - \lambda + 2m^2H^2)\bigr]|\nabla H|^2 \\ &= 2\bigl(\tfrac{5}{4}m^2H^2 + |A|^2\bigr)|\nabla H|^2 \geq 2\bigl(\tfrac{5}{4}m^2H^2 + mH^2\bigr)|\nabla H|^2 \\ &= \tfrac{1}{8}m(5m+4)|\nabla H^2|^2 \geq 0.\end{aligned}$$

Using maximum principles (ii) and (iii) in Theorem 2.5 we have that $|\nabla H|$ is constant. Using this and (26) again we conclude that $|\nabla H^2| = 0$. It follows that H is constant. If $H = \text{constant} \neq 0$, then, by the first equation of (8), we have $|A|^2 - \lambda = 0$. It follows that $|A|^2 = \lambda \leq 0$ since N is nonpositively curved. From this, we have $|A|^2 = 0$, which means that M is totally geodesic and hence minimal, which is a contradiction. So we are left with the only conclusion that $H = 0$, that is, the hypersurface is minimal. Thus, we complete the proof of the proposition. □

Remark 2. (i) For a classification of complete biharmonic submanifolds with Ricci curvature bounded from below in a nonpositively curved manifold see [Maeta 2014b].

(ii) Our Theorem 2.4 and Proposition 2.6 give some classifications of biharmonic hypersurfaces in an Einstein space. We refer to [Inoguchi and Sasahara 2016; 2017] for some examples and classifications of constant mean curvature proper biharmonic hypersurfaces in a special class of Einstein spaces — the compact Riemannian symmetric space with a G-invariant metric. Also, for some classifications of f-biharmonic hypersurfaces in an Einstein space, see [Ou 2017].

As a corollary of Proposition 2.6, we have an affirmative partial answer to Chen's conjecture:

Corollary 2.7. *Any complete biharmonic hypersurface with constant scalar curvature and $|\nabla H| \in L^p(M)$ for some $2 < p < \infty$ in a Euclidean space is minimal.*

Proposition 2.8. *Let $M^m \hookrightarrow S^{m+1}$ be a complete biharmonic hypersurface with constant scalar curvature in a sphere with $H^2 \geq (2\varepsilon + 4)/(m(5m+4))$ (for some $\varepsilon > 0$). If one of the following is satisfied, then the mean curvature is constant.*

(A) *The Ricci curvature of M is bounded from below, and $|\nabla H|$ from above.*

(B) *$|\nabla H| \in L^p$ for $2 < p < \infty$.*

(C) *$|\nabla H| \in L^2$ and the Ricci curvature is bounded from below by $-c\{1 + r^2(x)\}$, where $r(x)$ is the distance function on M.*

Proof. It follows from (15) that

$$\Delta|\nabla H|^2 \geq 2\left(-1 + \tfrac{5}{4}m^2H^2 + |A|^2\right)|\nabla H|^2.$$

Using Newton's formula $|A|^2 \geq mH^2$ for a hypersurface we have

$$\Delta|\nabla H|^2 \geq 2\left(-1 + \frac{m}{4}(5m+4)H^2\right)|\nabla H|^2.$$

From this, together with the assumption that $H^2 \geq (2\varepsilon+4)/(m(5m+4))$, we have

$$\Delta|\nabla H|^2 \geq \varepsilon|\nabla H|^2. \tag{27}$$

Using the maximum principles (i), (ii), and (iii) in Theorem 2.5 with $f = |\nabla H|^2$ we have obtain the proposition. □

Acknowledgments

We would like to thank C. Oniciuc and the referee for some very useful comments that helped to improve the manuscript.

References

[Balmuş et al. 2008] A. Balmuş, S. Montaldo, and C. Oniciuc, "Classification results for biharmonic submanifolds in spheres", *Israel J. Math.* **168** (2008), 201–220. MR Zbl

[Balmuş et al. 2012] A. Balmuş, S. Montaldo, and C. Oniciuc, "New results toward the classification of biharmonic submanifolds in $\mathbb{S}^n$", *An. Ştiinţ. Univ. "Ovidius" Constanţa Ser. Mat.* **20**:2 (2012), 89–114. MR Zbl

[Chen 1991] B.-Y. Chen, "Some open problems and conjectures on submanifolds of finite type", *Soochow J. Math.* **17**:2 (1991), 169–188. MR Zbl

[Chen 1993] J. H. Chen, "Compact 2-harmonic hypersurfaces in $S^{n+1}(1)$", *Acta Math. Sinica* **36**:3 (1993), 341–347. In Chinese. MR Zbl

[Chen 2015] B.-Y. Chen, *Total mean curvature and submanifolds of finite type*, 2nd ed., Series in Pure Mathematics **27**, World Scientific, Hackensack, NJ, 2015. MR Zbl

[Chern et al. 1970] 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, Springer, 1970. MR Zbl

[Ejiri 1982] N. Ejiri, "Some compact hypersurfaces of constant scalar curvature in a sphere", *J. Geom.* **19**:2 (1982), 197–199. MR Zbl

[Fu 2015] Y. Fu, "On biharmonic hypersurfaces with constant scalar curvatures in $\mathbb{S}^5$", *Proc. Amer. Math. Soc.* **143**:12 (2015), 5399–5409. MR Zbl

[Fu and Hong 2018] Y. Fu and M.-C. Hong, "Biharmonic hypersurfaces with constant scalar curvature in space forms", *Pacific J. Math.* **294**:2 (2018), 329–350. MR Zbl

[Inoguchi and Sasahara 2016] J.-i. Inoguchi and T. Sasahara, "Biharmonic hypersurfaces in Riemannian symmetric spaces I", *Hiroshima Math. J.* **46**:1 (2016), 97–121. MR Zbl

[Inoguchi and Sasahara 2017] J.-i. Inoguchi and T. Sasahara, "Biharmonic hypersurfaces in Riemannian symmetric spaces II", *Hiroshima Math. J.* **47**:3 (2017), 349–378. MR Zbl

[Karp and Li 1982] L. Karp and P. Li, "The heat equation on complete Riemannian manifolds", preprint, 1982.

[Luo 2015] Y. Luo, "On biharmonic submanifolds in non-positively curved manifolds", *J. Geom. Phys.* **88** (2015), 76–87. MR Zbl

[Maeta 2014a] S. Maeta, "Biharmonic maps from a complete Riemannian manifold into a non-positively curved manifold", *Ann. Global Anal. Geom.* **46**:1 (2014), 75–85. MR Zbl

[Maeta 2014b] S. Maeta, "Properly immersed submanifolds in complete Riemannian manifolds", *Adv. Math.* **253** (2014), 139–151. MR Zbl

[Nakauchi and Urakawa 2011] N. Nakauchi and H. Urakawa, "Biharmonic hypersurfaces in a Riemannian manifold with non-positive Ricci curvature", *Ann. Global Anal. Geom.* **40**:2 (2011), 125–131. MR Zbl

[Ou 2010] Y.-L. Ou, "Biharmonic hypersurfaces in Riemannian manifolds", *Pacific J. Math.* **248**:1 (2010), 217–232. MR Zbl

[Ou 2016] Y.-L. Ou, "Some recent progress of biharmonic submanifolds", pp. 127–139 in *Recent advances in the geometry of submanifolds*, edited by B. D. Suceavă et al., Contemp. Math. **674**, Amer. Math. Soc., Providence, RI, 2016. MR Zbl

[Ou 2017] Y.-L. Ou, "f-biharmonic maps and f-biharmonic submanifolds II", *J. Math. Anal. Appl.* **455**:2 (2017), 1285–1296. MR Zbl

[Yau 1976] S. T. Yau, "Some function-theoretic properties of complete Riemannian manifold and their applications to geometry", *Indiana Univ. Math. J.* **25**:7 (1976), 659–670. MR Zbl

Received November 4, 2018. Revised September 8, 2019.

SHUN MAETA
DEPARTMENT OF MATHEMATICS
SHIMANE UNIVERSITY
MATSUE
JAPAN
maeta@riko.shimane-u.ac.jp

YE-LIN OU
DEPARTMENT OF MATHEMATICS
TEXAS A & M UNIVERSITY
COMMERCE, TX
UNITED STATES
yelin_ou@tamu-commerce.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.291

SURFACE DIFFUSION FLOW OF ARBITRARY CODIMENSION IN SPACE FORMS

DONG PU AND HONGWEI XU

Dedicated to Professor Duanzhuang Qian on his 110th anniversary.

We first prove that a properly immersed static n-dimensional submanifold ($\Delta H = 0$) with restricted growth of the curvature at infinity in $\mathbb{F}^{n+p}(c)$ ($c \geq 0$) is totally umbilical if the Willmore functional is pinched by a positive constant depending only on n. Secondly, we obtain a global rigidity theorem for Willmore surfaces in the sphere. Thirdly, we give a lower bound on the lifespan of the surface diffusion flow in $\mathbb{F}^{2+p}(c)$. Finally, we get the longtime existence and convergence of the surface diffusion flow in $\mathbb{F}^{2+p}(c)$ ($c \geq 0$) under the small initial Willmore energy condition.

1. Introduction

An important problem in global differential geometry is the study of curvature and topology of Riemannian manifolds and submanifolds. As we know, curvature flows are powerful tools in the study of sphere theorems as in [Andrews and Baker 2010; Brendle and Schoen 2009; Hamilton 1982; Huisken 1984; Gu et al. 2017; Liu et al. 2018; Wang 2008; Gu and Xu 2012; Xu and Gu 2013], etc. For instance, Brendle and Schoen [2009] proved the remarkable differentiable $\frac{1}{4}$-pinching sphere theorem via the Ricci flow, which had been open for half a century.

Let $\mathbb{F}^{n+p}(c)$ be the $(n+p)$-dimensional complete and simply connected space form with constant curvature c. In this paper, we study the motion of an immersed submanifold with the normal velocity $-\Delta H$ in the space form $\mathbb{F}^{n+p}(c)$. More precisely, let $f : M^n \to \mathbb{F}^{n+p}(c)$ be a compact immersed submanifold. The diffusion flow for submanifolds is a fourth order flow

$$\frac{\partial f}{\partial t} = -\Delta H, \tag{1-1}$$

where $\nabla_X \phi = (D_X \phi)^\perp$ for a tangent vector field X and a normal vector field ϕ

Research supported by the National Natural Science Foundation of China, Grant Nos. 11531012, 11371315.

MSC2010: 53C24, 53C44.

Keywords: surface diffusion, Willmore surface, gap theorem, convergence theorem.

on M, $\Delta = \nabla_i \nabla_i$ and H is the mean curvature vector of M^n. Denote by A the second fundamental form and by $A^\circ = A - \frac{1}{n} g \otimes H$ the trace free part of the second fundamental form. When $\nabla H = 0$, we call M a submanifold with parallel mean curvature. In particular, M is minimal if $H = 0$. When $\Delta H = 0$, we call M a static submanifold for the diffusion flow which is a generalization of the submanifold with parallel mean curvature.

For $n = 2$, surface diffusion flow (1-1) in $\mathbb{R}^3$ was proposed by Mullins [1957] to describe thermal grooving in material sciences. From the view of geometric analysis it appears naturally as the gradient flow of the area functional with respect to the inner product of H^{-1}; see [Mayer 2001; Taylor and Cahn 1994]. Escher, Mayer and Simonett [Escher et al. 1998] showed that solutions that start out close to spheres with respect to the $C^{2+\beta}$-topology for hypersurfaces exist globally and converge exponentially fast to a sphere. Wheeler [2012, Chapter 3] obtained the following convergence theorem.

Theorem A. *Let $f : \Sigma \to \mathbb{R}^3$ be a compact surface. There exists an absolute constant $C_1 > 0$ such that if*

$$\int_\Sigma |A^\circ|^2 \, d\mu < C_1,$$

then the surface diffusion flow with initial data f exists smoothly for all time and converges exponentially to a round sphere as $t \to \infty$.

Willmore flow is another important fourth order flow that is closely related to the surface diffusion flow. Letting M be an n-dimensional compact submanifold in the space form, the Willmore functional is

$$\mathcal{W}(f) = \int_M |A^\circ|^n \, d\mu. \tag{1-2}$$

It is invariant under the conformal (or Moebius) transformations of the ambient space. When $n = 2$, the Willmore functional is also called Willmore energy. For a compact immersed surface $f : \Sigma \to \mathbb{R}^{2+p}$, the associated Euler–Lagrange operator is

$$W(f) = \Delta H + Q(A^\circ)H, \qquad Q(A^\circ)H = A^\circ(e_i, e_j)\langle A^\circ(e_i, e_j), H\rangle. \tag{1-3}$$

When $W(f) = 0$, we call it the Willmore surface. The L_2-gradient flow of $\mathcal{W}(f)$ for surfaces, briefly called Willmore flow, is a quasilinear geometric evolution equation

$$\frac{\partial f}{\partial t} = -W(f). \tag{1-4}$$

Kuwert and Schätzle [2002] gave a lower bound on the lifespan of a smooth solution, which depends only on how much the curvature of the initial surface is concentrated, and in [Kuwert and Schätzle 2001] they proved the following theorem:

Theorem B. *Let $f : \Sigma \to \mathbb{R}^{2+p}$ be a compact surface. There exists a constant $C_2(p) > 0$ such that if*

$$\int_\Sigma |A^\circ|^2 \, d\mu < C_2(p),$$

then the Willmore flow with initial data f exists smoothly for all time and converges exponentially to a round sphere as $t \to \infty$.

Kuwert and Schätzle [2004] investigated point singularities of Willmore surfaces and obtained that the Willmore flow of spheres in $\mathbb{R}^3$ with energy less than 8π exists at all times and converges to a round sphere.

For other fourth order flows, Bernard, Wheeler and Wheeler [Bernard et al. 2019] introduced the so-called Chen's flow and investigated the lifespan theorem and finite-time singularities for such a flow. Moreover, one class of fourth order flows derived from the critical point of some constructed functionals in different background manifolds. Metzger, Wheeler and Wheeler [2013] investigated the gradient flow of $\int_\Sigma |H|^2 \, d\mu$ and proved the lifespan theorem in Riemannian 3-manifold. Link [2013] also studied this flow in bounded Riemannian manifolds. Magni [2015] studied the gradient flow of $\int_\Sigma |A|^2 \, d\mu$ and obtained a smooth convergence theorem in three-dimensional Riemannian manifolds.

The local existence of these fourth order curvature flows is standard as we can see in [Escher et al. 1998], for example. In this paper, we first obtain a gap theorem for an n-dimensional properly immersed static submanifold M^n with restricted growth of the curvature at infinity in $\mathbb{F}^{n+p}(c)(c \geq 0)$ under the pinching condition for the Willmore functional, which generalizes the gap lemma in [Wheeler 2012] to any dimension and codimension.

Theorem 1.1. *Let $f : M^n \to \mathbb{F}^{n+p}(c)$ $(c \geq 0)$ be a properly immersed submanifold with $\Delta H = 0$. There exists a positive constant $\sigma(n)$ depending only on n such that if*

$$\int_M |A^\circ|^n \, d\mu < \sigma(n),$$

$$\liminf_{\rho \to \infty} \frac{1}{\rho^4} \int_{f^{-1}(B_\rho(x))} |A|^2 \, d\mu = 0 \quad \textit{for } x \in \mathbb{F}^{n+p}(c),$$

then M is totally umbilical.

Notice that the properness of the immersion implies the completeness of M. For a complete immersion, if we choose the cutoff function on intrinsic balls, the properness of the immersion isn't essential. For higher dimensions, there are few results about the convergence of fourth order flow, and the above gap theorem provides a feasible method to handle this problem. Xu and Gu [2007] proved a gap theorem for complete surfaces with parallel mean curvature in a space form under the pinching condition $\int_M |A^\circ|^4 \, d\mu < D(|H|, c)$, where $c + |H|^2/4 > 0$. Moreover, they proposed an open problem asking if there is global rigidity for surfaces with parallel mean curvature under the pinching condition for the Willmore functional. As a consequence of Theorem 1.1, we have solved this open problem for compact surfaces in $\mathbb{F}^{2+p}(c)$ $(c \geq 0)$.

For compact Willmore surfaces in a sphere, Xu and Yang [2016] proved a global rigidity theorem under the pinching condition for $\int_M |A^\circ|^4 \, d\mu$. Motivated by the convergence result in Theorem B, we obtain the following global rigidity theorem for compact Willmore surfaces in a sphere under the pinching condition for the Willmore functional.

Theorem 1.2. *Let M be a compact Willmore surface in the unit sphere $\mathbb{S}^{2+p}$. There exists an absolute positive constant σ_1 $(= 6.23 \times 10^{-8})$ such that if*

$$\int_M |A^\circ|^2 \, d\mu < \sigma_1,$$

then M is totally umbilical.

Next we obtain a prior estimate on the lifespan of the surface diffusion flow in space forms in terms of the concentration of curvature at the initial time. Our proof is standard for the analysis of Willmore flow in Euclidean spaces and later in bounded Riemannian manifolds [Link 2013; Metzger et al. 2013].

Theorem 1.3. *Let $f : \Sigma \to \mathbb{F}^{2+p}(c)$ be a compact surface. There exist constants $\epsilon_1 > 0$, $C < \infty$ depending only on p, such that if $\rho > 0$ is chosen with*

$$\int_{f^{-1}(B_\rho(x))} |A|^2 \, d\mu \leq \epsilon \leq \epsilon_1(p) \quad \textit{for any } x \in \mathbb{F}^{2+p}(c),$$

then the maximal time T for the surface diffusion flow with initial data f satisfies

$$T \geq \frac{1}{C}\rho^4,$$

and one has the estimate

$$\int_{f^{-1}(B_\rho(x))} |A|^2 \, d\mu \leq C\epsilon \quad \textit{for } 0 \leq t \leq \frac{1}{C}\rho^4.$$

Then we perform a blowup at an assumed singularity and construct a static surface as a limit. Putting the above results together we get the following convergence theorem.

Theorem 1.4 (main theorem). *Let $f : \Sigma \to \mathbb{F}^{2+p}(c)$ $(c \geq 0)$ be a compact surface. There exists a positive constant $\epsilon_0(p)$ such that if*

$$\int_\Sigma |A^\circ|^2 \, d\mu < \epsilon_0(p), \tag{1-5}$$

then the surface diffusion flow with initial data f exists smoothly for all time and converges exponentially to a round sphere as $t \to \infty$.

In particular, the convergence theorem for the surface diffusion flow implies the differentiable sphere theorem and the method can be also applied to the Willmore flow in a sphere. The key ingredient of the proof is to establish regularity and stability results in a space form. As we know, the equation of the diffusion flow is a fourth order equation so that tools related to the maximum principle are not available. Here we use the Sobolev inequality due to Michael and Simon [1973] and the Hölder inequality to classify the curvature terms especially for $\int_M |A^\circ|^n \, d\mu$, then we get the gap theorem for static submanifolds under the pinching condition for the Willmore functional. However, it contradicts the blowup limit at a finite time curvature singularity and thus we obtain the convergence theorem for the surface diffusion flow in the space form $\mathbb{F}^{2+p}(c)(c \geq 0)$.

2. Preparation

Let $(\mathbb{F}^{n+p}(c), \bar{g})$ be the space form with constant curvature c. For an immersion $f : M^n \to \mathbb{F}^{n+p}(c)$, the basic geometric data associated to f is the induced metric $g = f^*\bar{g}$ and $g(X, Y) = \langle Df \cdot X,\ Df \cdot Y\rangle$ with corresponding Levi-Civita connection ∇. Let $\{e_i\}$ be a set of locally defined orthogonal bases and the summation over repeated indices is used. The second fundamental form is given by $A(X, Y) = D^2_{X,Y} f = D_X(D_Y f) - Df \cdot \nabla_X Y$ with mean curvature vector given by the trace $H = A(e_i, e_i)$ and the trace free part $A^\circ(X, Y) = A(X, Y) - \frac{1}{n} g(X, Y)H$. We have the normal connection $\nabla_X \phi = (D_X \phi)^\perp$ which acts on the normal vector field ϕ along f. In (1-1) the Laplace operator $\Delta\phi = -\nabla^*\nabla\phi$ is understood with respect to the normal connection, where ∇^* denotes the formal adjoint of ∇.

When computing tensor identities we freely use vector fields with first derivative vanishing at a given point. We define the curvature by $R^\perp(X, Y)\phi = \nabla^2_{X,Y}\phi - \nabla^2_{Y,X}\phi$ and the equations of Codazzi, Gauss and Ricci are

$$
\begin{aligned}
&(\nabla_X A)(Y, Z) = (\nabla_Y A)(X, Z),\\
&\nabla H = -\nabla^* A = -\frac{n}{n-1}\nabla^* A^\circ,\\
&R(X, Y, Z, W) = c\langle X, Z\rangle\langle Y, W\rangle - c\langle X, W\rangle\langle Y, Z\rangle\\
&\qquad\qquad\qquad + A(X, Z)A(Y, W) - A(X, W)A(Y, Z),\\
&R^\perp(X, Y)\phi = A^\circ(e_i, X)\langle A^\circ(e_i, Y), \phi\rangle - A^\circ(e_i, Y)\langle A^\circ(e_i, X), \phi\rangle.
\end{aligned}
$$

The Codazzi equation implies that ∇A and $\nabla^2 A$ can be expressed by ∇A° and $\nabla^2 A^\circ$, respectively. In particular one has inequalities

$$
|\nabla A| \leq C|\nabla A^\circ|, \quad |\nabla^2 A| \leq C|\nabla^2 A^\circ|. \tag{2-1}
$$

If ϕ, ψ are normal forms, we denote by $\phi * \psi$ any normal-valued, multilinear form depending on ϕ, ψ in a universal, bilinear way. In particular, we have the properties $|\phi * \psi| \le C|\phi|\,|\psi|$ and $\nabla(\phi * \psi) = \nabla\phi * \psi + \phi * \nabla\psi$. We use the notation P_r^m for any term of the type

$$P_r^m = \sum_{i_1+\cdots+i_r=m} \nabla^{i_1} A * \nabla^{i_2} A * \cdots * \nabla^{i_r} A. \tag{2-2}$$

Now we derive a Sobolev type inequality in $\mathbb{S}^{n+p}$.

Lemma 2.1 [Michael and Simon 1973]. *Let M^n $(n \ge 2)$ be a compact submanifold with or without boundary in the Euclidean space $\mathbb{R}^{n+p}$ with $p \ge 1$. For a nonnegative function $g \in C^1(M)$ such that $g|_{\partial M} = 0$ if $\partial M \ne 0$, we have*

$$\left[\int_M g^{\frac{n}{n-1}}\, d\mu\right]^{\frac{n-1}{n}} \le D(n) \int_M (|\nabla g| + |H| g)\, d\mu,$$

where $D(n) = 4^{n+1}\sigma_n^{-1/n}$, and σ_n is the volume of the unit ball in $\mathbb{R}^n$.

Let M^n be a compact submanifold in $\mathbb{S}^{n+p}$. We consider the composition of isometric immersions $M^n \hookrightarrow \mathbb{S}^{n+p} \subset \mathbb{R}^{n+p+1}$ and denote by $\overline{H}$ the mean curvature vector of M^n as a submanifold in $\mathbb{R}^{n+p+1}$. Then $|\overline{H}|^2 = |H|^2 + n^2$. Hence we have for any nonnegative function $f \in C^1(M)$,

$$\begin{aligned}\left[\int_M g^{\frac{n}{n-1}}\, d\mu\right]^{\frac{n-1}{n}} &\le D(n) \int_M (|\nabla g| + |\overline{H}| g)\, d\mu \\ &\le D(n) \int_M (|\nabla g| + (n + |H|) g)\, d\mu.\end{aligned}$$

3. Gap theorem

In this section, using the Sobolev inequality we get some integral estimates, and obtain the gap theorem. First, we need the following lemma.

Lemma 3.1 [Kuwert and Schätzle 2002]. *Let $f : M^n \to \mathbb{F}^{n+p}(c)$ be a complete submanifold, then for any l-linear normal form ϕ we have*

$$\begin{aligned}((\nabla\nabla^* - \nabla^*\nabla)\phi)&(X_1, X_2 \ldots X_l) \\ &= -(\nabla^* T)(X_1, X_2 \ldots X_l) + (R^{\perp}(e_i, X_l)\phi)(e_i, X_2 \ldots X_l) \\ &\qquad - \sum_{k=1}^{l} \phi(X_i, \ldots, R(e_i, X_l)X_k, \ldots, X_l),\end{aligned}$$

where $T(X_0, X_1 \ldots X_l) = (\nabla_{X_0}\phi)(X_1, X_2 \ldots X_l) - (\nabla_{X_1}\phi)(X_0, X_2 \ldots X_l)$.

We have the following Simons' identity in space form: $\mathbb{F}^{n+p}(c)$.

$$\Delta h_{ij}=\nabla^2 H+H\cdot h_{ip}h_{pj}-h_{ij}\cdot h_{pq}h_{pq}+h_{jq}\cdot h_{ip}h_{pq}-h_{jq}\cdot h_{pq}h_{pi}+cnA^\circ+R^\perp * A.$$

Therefore, we can write

$$\Delta A^\circ=\nabla^2 H+\frac{H^2}{n}A^\circ+A^\circ * A^\circ * A+cnA^\circ-\frac{\Delta H}{n}g. \tag{3-1}$$

Here we need to estimate the terms of $A^\circ * A^\circ * A$ and cnA°, which are different from the equation of surfaces in the Euclidean space. Taking $\phi=\nabla H$ in Lemma 3.1, and using the Gauss equation, we have

$$((\nabla\nabla^*-\nabla^*\nabla)\nabla H)(e_1)=A^\circ * A^\circ * \nabla A^\circ-\nabla H(R(e_i,e_1)e_i).$$

Since

$$\begin{aligned}-\langle R(e_i,e_1)e_i,e_k\rangle\nabla_k H&=c(n-1)\nabla_1 H+H\cdot h_{1k}\nabla_k H-h_{1i}\cdot h_{ik}\nabla_k H\\&=c(n-1)\nabla H(e_1)+\frac{n-1}{n^2}H^2\nabla H(e_1)+A*A^\circ*\nabla A^\circ,\end{aligned}$$

we have

$$\nabla^*(\nabla^2 H)+c(n-1)\nabla H=\nabla(\nabla^*\nabla H)-\frac{n-1}{n^2}H^2\nabla H+A*A^\circ*\nabla A^\circ. \tag{3-2}$$

Taking $\phi=\nabla A^\circ$ in Lemma 3.1, we have

$$\nabla^*(\nabla^2 A^\circ)=\nabla(\nabla^*\nabla A^\circ)+A*A*\nabla A^\circ+cC\nabla A^\circ. \tag{3-3}$$

Replacing ϕ by $\nabla\phi$, we have

$$\Delta(\nabla\phi)-\nabla(\Delta\phi)=A*A*\nabla\phi+A*\nabla A*\phi+cC\nabla\phi. \tag{3-4}$$

Since the last terms of (3-2)–(3-4) are new compared with the equations in the Euclidean space, the next lemma is important in our proof.

Lemma 3.2. *If $f: M^n\to\mathbb{F}^{n+p}(c)$ is a complete submanifold with $\Delta H=F$ and $r\in C_c^1(M^n)$ satisfies $|\nabla r|\le\Lambda$, then*

$$\begin{aligned}&\int|\nabla A^\circ|^2r^2\,d\mu+c\int|A^\circ|^2r^2\,d\mu\\&\qquad\le\frac{C}{\Lambda^2}\int|F|^2r^4\,d\mu+\frac{C}{\Lambda^2}\int|A^\circ|^6r^4\,d\mu+C\Lambda^2\int_{[r>0]}|A|^2\,d\mu,\end{aligned} \tag{3-5}$$

$$\begin{aligned}&\int|\nabla A^\circ|^2r^4\,d\mu+c\int|A^\circ|^2r^4\,d\mu\\&\qquad\le C\int|\nabla H|^2r^4\,d\mu+C\int|A^\circ|^4|A|^2r^4\,d\mu+C\Lambda^4\int_{[r>0]}|A|^2\,d\mu.\end{aligned} \tag{3-6}$$

Proof. Multiplying (3-1) *by* $r^2 A^\circ$, *we have*

$$\begin{aligned}
&\int \Big(|\nabla A^\circ|^2 + cn|A^\circ|^2 + \frac{1}{n} H^2 |A^\circ|^2 \Big) r^2\, d\mu \\
&= \frac{n-1}{n} \int |\nabla H|^2 r^2\, d\mu + C \int |A^\circ|^3 |A| r^2\, d\mu + \int A^\circ * \nabla A^\circ * r * \nabla r\, d\mu \\
&\le -\frac{n-1}{n} \int \langle H, \Delta H \rangle r^2\, d\mu + \frac{C}{\Lambda^2} \int |A^\circ|^6 r^4\, d\mu + \Lambda^2 \int_{[r>0]} |A|^2\, d\mu + \tfrac{1}{2} \int |\nabla A^\circ|^2 r^2\, d\mu \\
&\le \frac{C}{\Lambda^2} \int |F|^2 r^4\, d\mu + \frac{C}{\Lambda^2} \int |A^\circ|^6 r^4\, d\mu + \Lambda^2 \int_{[r>0]} |A|^2\, d\mu + \tfrac{1}{2} \int |\nabla A^\circ|^2 r^2\, d\mu.
\end{aligned}$$

Multiplying (3-1) *by* $r^4 A^\circ$, *we get another inequality.* □

Lemma 3.3. *Under the assumption of Lemma 3.2 we have for* $\eta = r^4$,

$$\begin{aligned}
\int (|\nabla^2 H|^2 + |A|^2 |\nabla A|^2 + |A|^4 |A^\circ|^2) \eta\, d\mu + c \int |\nabla H|^2 \eta\, d\mu \\
\le C \int (|A^\circ|^2 |\nabla A^\circ|^2 + |A^\circ|^6) \eta\, d\mu \\
+ C \int |F|^2 \eta\, d\mu + \Lambda^4 \int_{[r>0]} |A|^2\, d\mu.
\end{aligned}$$

Proof. Multiplying (3-2) by $\nabla H \eta$, we have

$$\begin{aligned}
\text{(3-7)} \quad & \int (|\nabla^2 H|^2 + c(n-1)|\nabla H|^2) \eta\, d\mu \\
& \le \int \Big(|\Delta H|^2 - \frac{n-1}{n^2} H^2 |\nabla H|^2\Big) \eta\, d\mu \\
& \quad + C \int |A|\, |A^\circ|\, |\nabla A^\circ|^2 \eta\, d\mu + \int r^3 * \nabla r * \nabla H \nabla^2 H\, d\mu,
\end{aligned}$$

and the proof is similar to Lemma 2.3 in [Kuwert and Schätzle 2002]. □

Proposition 3.4. *Let* $f : M^n \to \mathbb{F}^{n+p}(c)$ $(c \ge 0)$ *be a complete submanifold with* $F = \Delta H$ *and* $r \in C_c^1(M)$ *satisfying* $s \ge 4$, *and let* $\Lambda = \|\nabla r\|_\infty$, *then*

$$\begin{aligned}
\int (|\nabla^2 A|^2 + |A|^2 |\nabla A|^2 + |A|^4 |A^\circ|^2)\, r^s\, d\mu + c \int (|A^\circ|^2 + |\nabla H|^2) r^s\, d\mu \\
\le C \int |F|^2 r^s\, d\mu + C\Lambda^4 \int_{[r>0]} |A|^2\, d\mu \\
+ C \int (|A^\circ|^2 |\nabla A^\circ|^2 + |A^\circ|^6) r^s\, d\mu.
\end{aligned}$$

Proof. Multiplying (3-3) by $\nabla A^\circ \eta$, $\eta = r^4$, using (3-1) we have

$$\begin{aligned}\int |\nabla^2 A^\circ|^2 \eta \, d\mu &= \int (|\Delta A^\circ|^2 + |A|^2 |\nabla A^\circ|^2 + cC|\nabla A^\circ|^2 + r^{-1} * \nabla r * \nabla^2 A^\circ * \nabla A^\circ)\eta \, d\mu \\ &\le \int (|\nabla^2 H|^2 + |A|^4 |A^\circ|^2 + |A|^2 |\nabla A^\circ|^2 + |F|^2)\eta \, d\mu \\ &\qquad + \tfrac{1}{2} \int |\nabla^2 A^\circ|^2 \eta \, d\mu + \Lambda^2 \int |\nabla A^\circ|^2 r^2 \, d\mu + cC \int (|\nabla A^\circ|^2 + |A^\circ|^2)\eta \, d\mu.\end{aligned}$$

Using Lemma 3.2, we have

$$\begin{aligned}\int |\nabla^2 A^\circ|^2 \eta \, d\mu + c \int |A^\circ|^2 \eta \, d\mu &\le \int (|\nabla^2 H|^2 + |A|^2 |\nabla A^\circ|^2 + |A|^4 |A^\circ|^2 + c|\nabla H|^2)\eta \, d\mu \\ &\qquad + \int |F|^2 \eta \, d\mu + \int |A^\circ|^6 \eta \, d\mu + \Lambda^4 \int_{[r>0]} |A|^2 \, d\mu.\end{aligned}$$

The assertion follows from Lemma 3.3. □

Now we can prove Theorem 1.1.

Proof of Theorem 1.1. It follows from Proposition 3.4 that

$$\begin{aligned}\int (|\nabla^2 A|^2 + |A|^2 |\nabla A|^2 + |A|^4 |A^\circ|^2)\, r^s \, d\mu + c \int (|A^\circ|^2 + |\nabla H|^2) r^s \, d\mu \\ \le C \int |F|^2 r^s \, d\mu + C\Lambda^4 \int_{[r>0]} |A|^2 \, d\mu + C \int (|A^\circ|^2 |\nabla A^\circ|^2 + |A^\circ|^6) r^s \, d\mu.\end{aligned}$$

Then we have

$$\int |A^\circ|^2 |\nabla A^\circ|^2 r^s \, d\mu \le C \int |A^\circ|^6 r^s \, d\mu + C \int |\nabla A^\circ|^3 r^s \, d\mu,$$

and

$$\begin{aligned}\int |\nabla A^\circ|^3 r^s \, d\mu &\le \int |A^\circ| \, |\nabla^2 A^\circ| \, |\nabla A^\circ| r^s \, d\mu + \Lambda \int |A^\circ| \, |\nabla A^\circ|^2 r^{s-1} \, d\mu \\ &\le \int (\delta |\nabla^2 A^\circ|^2 + C_\delta |A^\circ|^2 |\nabla A^\circ|^2) r^s \, d\mu \\ &\qquad + \Lambda \int |A^\circ|^{\frac{1}{2}} |\nabla A^\circ|^{\frac{1}{2}} |A^\circ|^{\frac{1}{2}} |\nabla A^\circ|^{\frac{3}{2}} r^{s-1} \, d\mu \\ &\le \int (\delta |\nabla^2 A^\circ|^2 + C_\delta |A^\circ|^6 + \delta |\nabla A^\circ|^3) r^s \, d\mu \\ &\qquad + \Lambda^{\frac{4}{3}} \int |A^\circ|^{\frac{2}{3}} |\nabla A^\circ|^2 r^{s-\frac{4}{3}} \, d\mu \\ &\le \delta \int |\nabla^2 A^\circ|^2 r^s \, d\mu + C_\delta \int |A^\circ|^6 r^s \, d\mu + \Lambda^4 \int_{[r>0]} |A|^2 \, d\mu.\end{aligned} \tag{3-8}$$

Therefore,

$$\int |A^\circ|^2|\nabla A^\circ|^2 r^s\,d\mu \le \delta \int |\nabla^2 A^\circ|^2 r^s\,d\mu + C_\delta \int |A^\circ|^6 r^s\,d\mu + \Lambda^4 \int_{[r>0]} |A|^2\,d\mu.$$

When $c=0$, the Sobolev inequality is

$$\left[\int g^{\frac{n}{n-1}}\,d\mu\right]^{\frac{n-1}{n}} \le C(n)\int (|\nabla g| + |H|g)\,d\mu.$$

If we take $g = |A^\circ|^{6(n-1)/n} r^{(n-1)/(n)s}$, then

$$\begin{aligned}
(3\text{-}9)\quad \left(\int |A^\circ|^6 r^s\,d\mu\right)^{\frac{n-1}{n}}
&= \left[\int (|A^\circ|^{\frac{6(n-1)}{n}} r^{\frac{n-1}{n}s})^{\frac{n}{n-1}}\,d\mu\right]^{\frac{n-1}{n}}\\
&\le C(n)\int \left[\nabla(|A^\circ|^{\frac{6(n-1)}{n}} r^{\frac{n-1}{n}s}) + |H|\,|A^\circ|^{\frac{6(n-1)}{n}} r^{\frac{n-1}{n}s}\right] d\mu\\
&\le \int |\nabla A^\circ|\,|A^\circ|^{\frac{5n-6}{n}} r^{\frac{n-1}{n}s}\,d\mu + \Lambda \int |A^\circ|^{\frac{6(n-1)}{n}} r^{\frac{n-1}{n}s-1}\,d\mu\\
&\qquad + \int |A|\,|A^\circ|^{\frac{6(n-1)}{n}} r^{\frac{n-1}{n}s}\,d\mu = a+b+c.
\end{aligned}$$

Here

$$\begin{aligned}
a &= \int |\nabla A^\circ|\,|A^\circ|^{\frac{n}{3(n-1)}} r^{\frac{1}{3}s} |A^\circ|^{\frac{(2n-3)(7n-6)}{3n(n-1)}} r^{\frac{2n-3}{3n}s}\,d\mu\\
&\le \int (|\nabla A^\circ|^{\frac{3(n-1)}{n}} |A^\circ| + |A^\circ|^{\frac{7n-6}{n}}) r^{\frac{n-1}{n}s}\,d\mu\\
&\le \left(\int_{[r>0]} |A^\circ|^n\,d\mu\right)^{\frac{1}{n}} \left[\left(\int |\nabla A^\circ|^3 r^s\,d\mu\right)^{\frac{n-1}{n}} + \left(\int |A^\circ|^6 r^s\,d\mu\right)^{\frac{n-1}{n}}\right],
\end{aligned}$$

$$\begin{aligned}
b &= \Lambda \int |A|^\circ |A^\circ|^{\frac{5n-6}{n}} r^{\frac{n-1}{n}s-1}\,d\mu\\
&\le \left(\int_{[r>0]} |A^\circ|^n\,d\mu\right)^{\frac{1}{n}} \left(\Lambda^{\frac{n}{n-1}} \int |A^\circ|^{\frac{5n-6}{n-1}} r^{s-\frac{n}{n-1}}\,d\mu\right)^{\frac{n-1}{n}}\\
&= \left(\int_{[r>0]} |A^\circ|^n\,d\mu\right)^{\frac{1}{n}} \left(\Lambda^{\frac{n}{n-1}} \int |A^\circ|^{\frac{n}{2(n-1)}} |A^\circ|^{\frac{3(3n-4)}{2(n-1)}} r^{s-\frac{n}{n-1}}\,d\mu\right)^{\frac{n-1}{n}}\\
&\le \left(\int_{[r>0]} |A^\circ|^n\,d\mu\right)^{\frac{1}{n}} \left[\Lambda^4\left(\int_{[r>0]} |A^\circ|^2\,d\mu\right) + \left(\int |A^\circ|^6 r^s\,d\mu\right)\right]^{\frac{n-1}{n}},
\end{aligned}$$

$$
\begin{aligned}
c &= \int |A|\,|A^\circ|^{\frac{6(n-1)}{n}} r^{\frac{n-1}{n}s}\,d\mu \\
&\le \Big(\int_{[r>0]} |A^\circ|^n\,d\mu\Big)^{\frac{1}{n}} \Big[\int (|A|\,|A^\circ|^{\frac{5n-6}{n}})^{\frac{n}{n-1}} r^s\,d\mu\Big]^{\frac{n-1}{n}} \\
&\le \Big(\int_{[r>0]} |A^\circ|^n\,d\mu\Big)^{\frac{1}{n}} \Big(\int |A|^2|A^\circ|^4 r^s\,d\mu\Big)^{\frac{n-1}{n}}.
\end{aligned}
$$

When $c=1$, the Sobolev inequality is

$$
\Big[\int g^{\frac{n}{n-1}}\,d\mu\Big]^{\frac{n-1}{n}} \le C(n)\int (|\nabla g|+|H|g+g)\,d\mu.
$$

If we take $g=|A^\circ|^{6(n-1)/n} r^{(n-1)/(n)s}$, then the third term in the Sobolev inequality is

$$
\begin{aligned}
\int |A^\circ|^{\frac{6(n-1)}{n}} r^{\frac{n-1}{n}s}\,d\mu &= \int |A|^\circ |A^\circ|^{\frac{5n-6}{n}} r^{\frac{n-1}{n}s}\,d\mu \\
&\le \Big(\int_{[r>0]} |A^\circ|^n\,d\mu\Big)^{\frac{1}{n}} \Big(\int |A^\circ|^{\frac{5n-6}{n-1}} r^s\,d\mu\Big)^{\frac{n-1}{n}} \\
&\le \Big(\int_{[r>0]} |A^\circ|^n\,d\mu\Big)^{\frac{1}{n}} \Big[\Big(\int |A^\circ|^2 r^s\,d\mu\Big)+\Big(\int |A^\circ|^6 r^s\,d\mu\Big)\Big]^{\frac{n-1}{n}}.
\end{aligned}
$$

Combining the above estimates, we have

$$
\begin{aligned}
\text{(3-10)}\quad & \int (|A^\circ|^2|\nabla A^\circ|^2+|A^\circ|^6) r^s\,d\mu \\
&\le \Lambda^4(1+\|A^\circ\|_{n,[r>0]}^{\frac{n}{n-1}})\int_{[r>0]} |A|^2\,d\mu+\delta\int |\nabla^2 A^\circ|^2 r^s\,d\mu \\
&\quad +\|A^\circ\|_{n,[r>0]}^{\frac{n}{n-1}}\int_M (|\nabla^2 A^\circ|^2+|A^\circ|^6+|A|^2|A^\circ|^4+c|A^\circ|^2) r^s\,d\mu.
\end{aligned}
$$

Now take the cutoff function $r(q)=\varphi(\gamma_x(f(q))/\rho)$, where γ_x is the distance function with respect to the fixed point $x\in\mathbb{F}^{2+p}$, $\varphi\in C^1(\mathbb{R})$ and $\chi_{B_{1/2}(x)}\le\varphi\le\chi_{B_1}(x)$; then $\Lambda=\frac{C}{\rho}$. From Proposition 3.4, $\|A^\circ\|_{n,[r>0]}^n<\sigma(n)$, and $s\ge 4$, we have

$$
\begin{aligned}
\text{(3-11)}\quad & \int (|\nabla^2 A|^2+|A|^2|\nabla A|^2+|A|^4|A^\circ|^2)\, r^s\,d\mu + c\int |A^\circ|^2 r^s\,d\mu \\
&\le \int |F|^2 r^s\,d\mu + \Big(\frac{C}{\rho}\Big)^4\int_{[r>0]} |A|^2\,d\mu.
\end{aligned}
$$

Since f is a proper immersion, taking $\rho\to\infty$ and using a similar argument as in Theorem 8 of [Wheeler 2012], we finish the proof. □

A surface in the sphere is a Willmore surface if and only if it satisfies

$$W(f) = \Delta H + Q(A^\circ)H = 0,$$

or else we get (4-1). The estimate of $W(f)$ is easy to handle as in [Kuwert and Schätzle 2002], so we also have the gap theorem for Willmore surfaces in a sphere.

Proof of Theorem 1.2. For $n = 2$ and $c = 1$, the gradient of the second fundamental form can be controlled (see [Andrews and Baker 2010]) by

$$\tfrac{1}{4}|\nabla H|^2 \le \tfrac{1}{3}|\nabla A|^2 \le |\nabla A^\circ|^2. \tag{3-12}$$

The precise forms of (3-1), (3-2) and (3-3) are

$$\nabla^*\nabla A^\circ + 2KA^\circ = -\nabla^2 H - R^\perp(e_k,\cdot)A(e_k,\cdot) + \frac{\Delta H}{2} g(\cdot,\cdot), \tag{3-13}$$

$$\nabla^*(\nabla^2 H) + K\nabla H = \nabla(\nabla^*\nabla H) - R^\perp(e_k,\cdot)\nabla H(e_k), \tag{3-14}$$

$$\begin{aligned}\langle \nabla^*(\nabla^2 A^\circ), \nabla A^\circ\rangle \le \langle \nabla(\nabla^*\nabla A^\circ), \nabla A^\circ\rangle - 5|\nabla A^\circ|^2 \\ + 2|\nabla H|^2 + 12|A|^2|\nabla A|^2,\end{aligned} \tag{3-15}$$

where the section curvature $K = 1 + |H|^2/4 - |A^\circ|^2/2$.

Multiplying (3-13) by $r^2 A^\circ$ with $r \in C_c^1(M^2)$ satisfies $\|\nabla r\|_\infty = \Lambda$. We have

$$\begin{aligned}
&\int (|\nabla A^\circ|^2 + 2|A^\circ|^2 + \tfrac{1}{2}H^2|A^\circ|^2) r^2\, d\mu \\
&\quad \le \tfrac{1}{2}\int |\nabla H|^2 r^2\, d\mu + 2\int |A^\circ|^4 r^2\, d\mu + 6\Lambda \int |A^\circ|\,|\nabla A^\circ| r\, d\mu \\
&\quad \le -\tfrac{1}{2}\int \langle H, W\rangle r^2\, d\mu + \tfrac{1}{2}\int \langle H, Q(A^\circ)H\rangle r^2\, d\mu \\
&\qquad\quad + \frac{1}{128\Lambda^2}\int |A^\circ|^6 r^4\, d\mu + C\Lambda^2 \int_{[r>0]} |A|^2\, d\mu + \tfrac{1}{2}\int |\nabla A^\circ|^2 r^2\, d\mu \\
&\quad \le \frac{C}{\Lambda^2}\int |W|^2 r^4\, d\mu + \tfrac{1}{2}\int H^2 |A^\circ|^2 r^2\, d\mu \\
&\qquad\quad + \frac{1}{128\Lambda^2}\int |A^\circ|^6 r^4\, d\mu + C\Lambda^2 \int_{[r>0]} |A|^2\, d\mu + \tfrac{1}{2}\int |\nabla A^\circ|^2 r^2\, d\mu,
\end{aligned}$$

and

$$\begin{aligned}&\int (|\nabla A^\circ|^2 + 4|A^\circ|^2) r^2\, d\mu \\ &\qquad \le \frac{C}{\Lambda^2}\int |W|^2 r^4\, d\mu + \frac{1}{64\Lambda^2}\int |A^\circ|^6 r^4\, d\mu + C\Lambda^2 \int_{[r>0]} |A|^2\, d\mu.\end{aligned} \tag{3-16}$$

Multiplying (3-13) by $r^4 A^\circ$, we have

$$\tag{3-17} \int (|\nabla A^\circ|^2 + |A^\circ|^2) r^4 \, d\mu \\ \leq 2 \int |\nabla H|^2 r^4 \, d\mu + 2 \int |A^\circ|^6 r^4 \, d\mu + C\Lambda^4 \int_{[r>0]} |A|^2 \, d\mu.$$

Similarly, from the proof of Proposition 3.4 we have

$$\tag{3-18} \int (\tfrac{3}{4} |\nabla^2 A|^2 + |A^\circ|^2 + 20 |\nabla A^\circ|^2 + 40 |A|^2 |\nabla A|^2 + |A|^4 |A^\circ|^2) \, r^4 \, d\mu \\ \leq \int (3546 |A^\circ|^2 |\nabla A^\circ|^2 + 5230 |A^\circ|^6) r^4 \, d\mu \\ + C \int |W|^2 r^4 \, d\mu + C\Lambda^4 \int_{[r>0]} |A|^2 \, d\mu.$$

Taking $g = |A^\circ|^3 r^2$ in the Sobolev inequality, we get

$$\tag{3-19} \frac{1}{D^2(2)} \int |A^\circ|^6 r^4 \, d\mu \\ \leq \left[\int (3|A^\circ|^2 |\nabla A^\circ| r^2 + 2\Lambda |A^\circ|^3 r + |H| \, |A^\circ|^3 r^2 + 2|A^\circ|^3 r^2) \, d\mu \right]^2 \\ \leq 5 \int_{[r>0]} |A^\circ|^2 \, d\mu \int (9|A^\circ|^2 |\nabla A^\circ|^2 + 2|A^\circ|^6 + 2|A|^2 |A^\circ|^4 + |A^\circ|^2) r^4 \, d\mu \\ + C\Lambda^4 \left(\int_{[r>0]} |A^\circ|^2 \, d\mu \right)^2.$$

Thus

$$\tag{3-20} \left(\frac{1}{5D^2(2)} - 2\sigma_1 \right) \int |A^\circ|^6 r^4 \, d\mu \\ \leq \int_{[r>0]} |A^\circ|^2 \, d\mu \int (9|A^\circ|^2 |\nabla A^\circ|^2 + 2|A|^2 |A^\circ|^4 + |A^\circ|^2) r^4 \, d\mu \\ + C\Lambda^4 \left(\int_{[r>0]} |A^\circ|^2 \, d\mu \right)^2.$$

Taking $g = |A^\circ| |\nabla A^\circ| r^2$ in the Sobolev inequality, we also obtain

$$\tag{3-21} \frac{1}{5D^2(2)} \int |A^\circ|^2 |\nabla A^\circ|^2 r^4 \, d\mu \\ \leq \int_{[r>0]} |A^\circ|^2 \, d\mu \int \big(\tfrac{5}{2} |\nabla^2 A^\circ|^2 + 2|A|^2 |\nabla A^\circ|^2 + 4|\nabla A^\circ|^2 \big) r^4 \, d\mu \\ + C\Lambda^4 \left(\int_{[r>0]} |A^\circ|^2 \, d\mu \right)^2.$$

It follows from (3-18), (3-20) and (3-21) and the assumption $\|A^\circ\|_2^2 < 6.23\times 10^{-8}$ that

$$\int (|\nabla^2 A|^2 + |A|^2|\nabla A|^2 + |A|^4|A^\circ|^2)\, r^s\, d\mu + \int |A^\circ|^2 r^s\, d\mu \\ \le C\int |W|^2 r^s\, d\mu + C\Lambda^4 \int_{[r>0]} |A|^2\, d\mu.$$

We take the cutoff function $r(q)=\varphi(\gamma_x(f(q))/(\rho))$, where γ_x is a distance function with respect to the fixed point $x\in\mathbb{S}^{2+p}$, $\varphi\in C^1(\mathbb{R})$ and $\chi_{B_{1/2}(x)}\le\varphi\le\chi_{B_1}(x)$; then $\Lambda = C/\rho$. Taking $\rho\to\infty$, we finish the proof. □

4. Lifespan theorem

In this section, we give a lower bound on the lifespan of a smooth solution of the surface diffusion flow, which depends only on how much the curvature of the initial surface is concentrated. First, we recall these equations from [Kuwert and Schätzle 2002], and we can see the proof in [Link 2013].

Lemma 4.1. *Let $f : M^n\times[0,T)\to\mathbb{F}^{n+p}(c)$ be a smooth variation with normal velocity $\partial_t f = -V$, then*

$$\begin{aligned}
\partial_t g(X\ Y) &= 2\langle A(X\ Y), V\rangle,\\
\partial_t(d\mu) &= \langle H, V\rangle\, d\mu,\\
\partial_t^\perp(\nabla_X\phi) &= \nabla_X\partial_t^\perp\phi - A(X\ e_i)\langle\nabla_{e_i}V,\ \phi\rangle - \nabla_{e_i}V\langle A(X\ e_i),\phi\rangle,\\
\partial_t^\perp(\nabla_X Y) &= \langle(\nabla_{e_i}A)(X,\ Y),\ V\rangle e_i - \langle A(X\ Y),\nabla_{e_i}V\rangle e_i\\
&\qquad + \langle A(X\ e_i),\nabla_Y V\rangle e_i + \langle A(Y\ e_i),\nabla_X V\rangle e_i,\\
\partial_t^\perp A(X\ Y) &= -\nabla^2_{X,Y}V + A(e_i,X)\langle A(e_i,Y),V\rangle - cg(X\ Y)V,\\
\partial_t^\perp H &= -\Delta V - A(e_i,e_j)\langle A(e_i,e_j),V\rangle - ncV.
\end{aligned}$$

Using Gauss equation $2R_{1212}=2c+H^2-|A|^2$ and Lemma 4.1, the first variation formula for the Willmore functional in space forms in the normal direction -ϕ is

$$\begin{aligned}
\text{(4-1)}\quad &\frac{d}{d\epsilon}\mathcal{W}(f-\epsilon\phi)|_{\epsilon=0}\\
&= \frac{d}{d\epsilon}\int_\Sigma |A^\circ|^2\, d\mu|_{\epsilon=0} = \frac{d}{d\epsilon}\int_\Sigma \Big(\frac{|H|^2}{2}+2c\Big)\, d\mu|_{\epsilon=0}\\
&= \int_\Sigma \langle H, -\Delta\phi - A(e_i,e_j)\langle A(e_i,e_j),\phi\rangle - 2c\phi + 2c\phi\rangle\, d\mu + \tfrac12\int_\Sigma \langle H^3,\phi\rangle\, d\mu\\
&= -\int_\Sigma \langle\phi, \Delta H + Q(A^\circ)H\rangle\, d\mu,
\end{aligned}$$

which proves $\mathrm{grad}_{L^2}\mathcal{W}(f) = \Delta H + Q(A^\circ)H$.

The following lemmas will be needed for computing the evolution of derivatives of the curvature.

Lemma 4.2. *Let ϕ be an $(l-1)$-form with normal values along a variation $f : M^n \times I \to \mathbb{F}^{n+p}(c)$ with normal velocity $\partial_t f = -V$. If $\partial_t^\perp \phi + \Delta^2 \phi = Y$, then $\psi = \nabla \phi$ satisfies an equation*

$$\text{(4-2)}\quad \partial_t^\perp \psi + \Delta^2 \psi = \nabla Y + \sum_{i+j+k=3} \nabla^i A * \nabla^j A * \nabla^k \phi + A * \nabla V * \phi + \nabla A * V * \phi + C(\Delta(\nabla\phi) + \nabla(\Delta\phi)).$$

Proof. Let $X_1, \dots, X_l$ be independent of t and such that $\nabla X_k = 0$ at a given point and a given time. Using Lemma 4.1 we have

$$\begin{aligned}(\partial_t^\perp \psi)(X_1, \dots, X_l) &= \partial_t^\perp((\nabla_{X_1}\phi)(X_2, \dots, X_l)) \\ &= \partial_t^\perp(\nabla_{X_1}\phi(X_2, \dots, X_l)) - \partial_t^\perp \sum_{k=2}^{l} \phi(X_2, \dots, \nabla_{X_1} X_k, \dots, X_l) \\ &= (\nabla(\partial_t^\perp \phi))(X_1, \dots, X_l) + A * \nabla V * \phi + \nabla A * V * \phi.\end{aligned}$$

So,

$$\text{(4-3)}\quad \partial_t^\perp \psi + \Delta^2 \psi - \nabla Y = \Delta^2(\nabla\phi) - \nabla(\Delta^2\phi) + A * \nabla V * \phi + \nabla A * V * \phi.$$

Using (3-4), we have

$$\begin{aligned}\Delta^2(\nabla\phi) - \nabla(\Delta^2\phi) &= \Delta(\Delta(\nabla\phi) - \nabla(\Delta\phi)) + \Delta(\nabla(\Delta\phi)) - \nabla(\Delta(\Delta\phi)) \\ &= \Delta(A * A * \nabla\phi + A * \nabla A * \phi + C\nabla\phi) + A * A * \nabla(\Delta\phi) \\ &\qquad + A * \nabla A * (\Delta\phi) + C\nabla(\Delta\phi) \\ &= \sum_{i+j+k=3} \nabla^i A * \nabla^j A * \nabla^k \phi + C(\Delta(\nabla\phi) + \nabla(\Delta\phi)). \qquad \square\end{aligned}$$

Lemma 4.3. *Let $f : \Sigma \times [0, T) \to \mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow, then*

$$\partial_t^\perp(\nabla^m A) + \Delta^2(\nabla^m A) = P_3^{m+2}(A) + P_1^{m+2}(A),$$

for any $m \in \mathbb{N}_0$.

Proof. We proceed by induction on m, starting with $m = 0$,

$$\begin{aligned}&\Delta(\nabla^2 H) - \nabla^2(\Delta H) \\ &\quad = (\nabla\nabla^* - \nabla^*\nabla)\nabla(\nabla H) + \nabla(\nabla\nabla^* - \nabla^*\nabla)\nabla H \\ &\quad = A * A * \nabla^2 H + A * \nabla A * \nabla H + \nabla(A * A * \nabla H + A * \nabla A * \nabla H) + C\nabla^2 H,\end{aligned}$$

and

$$\partial_t^{\perp} A = -\nabla^2(\Delta H) + A*A*\Delta H - c\Delta H g = -\Delta(\nabla^2 H) + P_3^2(A) + P_1^2(A) \\ = -\Delta^2(A) + P_3^2(A) + P_1^2(A). \tag{4-4}$$

Now let $m \geq 1$ and conclude from (4-2), that we have

$$\begin{aligned}\partial_t^{\perp}(\nabla^m A) &+ \Delta^2(\nabla^m A) \\ &= \nabla(P_3^{m+1}(A) + P_1^{m+1}(A)) + \sum_{i+j+k=3} \nabla^i A * \nabla^j A * \nabla^k(\nabla^{m-1}A) \\ &\qquad + A * \nabla(\Delta H) * \nabla^{m-1}A + \nabla A * \Delta H * \nabla^{m-1}A + \nabla^3(\nabla^{m-1}A),\end{aligned}$$

which yields the result. □

In the following we assume $r = \tilde{r} \circ f$, where $\tilde{r} \in C_c^1(B_\rho)$ and $\|\tilde{r}\|_{C^2} \leq C < \infty$. This implies

$$|\nabla r| \leq C, \quad |\nabla^2 r| \leq C(1 + |A|), \tag{4-5}$$

and for a ball $B_\rho = B_\rho(x_0) \subset \mathbb{F}^{2+p}(c)$ we use the notion

$$\Sigma_\rho(x_0) = f^{-1}(B_\rho(x_0)). \tag{4-6}$$

We denote by $\epsilon_1(p)$ the upper bound of the concentration of curvature for the initial surface, which is a positive constant depending only on the codimension p.

The next lemma can be proved much as in [Kuwert and Schätzle 2002].

Lemma 4.4. *Let $f : \Sigma \times [0, T) \to \mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow, then for $\phi = \nabla^m A$ with $m \in \mathbb{N}$ and $s \geq 2m + 4$, we have*

$$\begin{aligned}\frac{d}{dt}\int |\phi|^2 r^s \, d\mu &+ \tfrac{3}{4}\int |\nabla^2\phi|^2 r^s \, d\mu \\ &\leq \int (P_3^{m+2}(A) + P_5^m(A) + P_1^{m+2}(A)) * \phi \, r^s \, d\mu + C\int |A|^2 r^{s-4-2m} \, d\mu.\end{aligned}$$

Remark 4.5. Suppose $\chi_{B_{\rho(x)}} \leq \tilde{r} \leq \chi_{B_{2\rho(x)}}$ and $\|D^j \tilde{r}\|_{L^\infty} \leq C'\rho^{-j}$ for $j = 1, 2$. Then we have $C = C'/(\rho^{4+2m})$ in Lemma 4.4.

Proposition 4.6. *Let $f : \Sigma \times [0, T) \to \mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow, then for $\phi = \nabla^m A$ with $m \in \mathbb{N}$ and $s \geq 2m + 4$, we have*

$$\begin{aligned}\frac{d}{dt}\int |\phi|^2 r^s \, d\mu &+ \tfrac{1}{2}\int |\nabla^2\phi|^2 r^s \, d\mu \\ &\leq C(\|A\|^4_{\infty\,[r>0]} + 1)\|A\|^2_{2\,[r>0]} + C\|A\|^4_{\infty\,[r>0]}\int |\phi|^2 r^s \, d\mu.\end{aligned} \tag{4-7}$$

Proof. According to Lemma 4.4, we only need to show

$$\int (P_3^{m+2}(A)+P_5^m(A)+P_1^{m+2}(A))*\phi\, r^s\,d\mu+C\int_{[r>0]}|A|^2\,d\mu$$
$$\le \tfrac14\int|\nabla^2\phi|^2 r^s\,d\mu+C(\|A\|^4_{\infty\,[r>0]}+1)\|A\|^2_{2\,[r>0]}+C\|A\|^4_{\infty\,[r>0]}\int|\phi|^2\,r^s\,d\mu,$$

and the proof is similar to Proposition 4.5 in [Kuwert and Schätzle 2002] with

$$\int P_1^{m+2}(A)*\phi r^s\,d\mu\le\tau\int|\nabla^2\phi|^2r^s\,d\mu+C(\tau)\|A\|^2_{2\,[r>0]}. \tag{4-8}$$

It follows from Corollaries 5.3 and 5.5 in [Kuwert and Schätzle 2002]. □

Since the Sobolev inequality is similar when in $\mathbb{F}^{2+p}(c)$ for $c\le 0$, in the next lemma we just check the situation $c>0$ (see [Kuwert and Schätzle 2002]).

Lemma 4.7. *For any normal l-form ϕ on Σ and r as in (4-5),*

$$\|\phi\|^4_{\infty\,[r=1]}\le C\|\phi\|^2_{2\,[r>0]}(\|\nabla^2\phi\|^2_{2\,[r>0]}+\|\phi^2|A|^4\|_{1\,[r>0]}+\|\phi\|^2_{2\,[r>0]}). \tag{4-9}$$

Moreover, if $\phi=A$ and $\|A\|^2_{2\,[r>0]}\le\epsilon_1(p)$ for some ϵ_1 small enough depending on the constants in (4-5), then

$$\|A\|^4_{\infty\,[r=1]}\le C\|A\|^2_{2\,[r>0]}(\|\nabla^2A\|^2_{2\,[r>0]}+\|A\|^2_{2\,[r>0]}). \tag{4-10}$$

Proposition 4.8. *Let $f:\Sigma\times[0,T]\to\mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow, then there exists an $\epsilon_1(p)>0$ such that if*

$$\epsilon=\sup_{[0,T]}\int_{[r>0]}|A|^2\,d\mu\le\epsilon_1,$$

then for any $t\in[0,T]$ we have

$$\int_{[r=1]}|A|^2\,d\mu+\frac14\int_0^t\int_{[r=1]}(|\nabla^2A|^2+|A|^2|\nabla A|^2+|A|^6)\,d\mu\,dt \le\int_{[r_0>0]}|A_0|^2\,d\mu_0+C\epsilon t. \tag{4-11}$$

Proof. We know from (3-10) that

$$\int(|A|^2|\nabla A|^2+|A|^6)r^4\,d\mu$$
$$\le C\int_{[r>0]}|A|^2\,d\mu+\tfrac14\int_M|\nabla^2A|^2r^4\,d\mu$$
$$+\|A\|^2_{2,[r>0]}\int_M(|\nabla^2A|^2+|A|^2|\nabla A|^2+|A|^6+|A|^2)r^4\,d\mu.$$

Choosing $m = 0$ in Lemma 4.4, we get

$$\frac{d}{dt}\int |A|^2 r^4\,d\mu + \tfrac{3}{4}\int |\nabla^2 A|^2 r^4\,d\mu \\ \leq \int (P_3^2(A) + P_5^0(A) + P_1^2(A)) * A\, r^4\,d\mu + C\int_{[r>0]} |A|^2\,d\mu.$$

Combining these two inequalities, we have

$$\begin{aligned}\frac{d}{dt}\int |A|^2 r^4\,d\mu &+ \tfrac{3}{4}\int (|\nabla^2 A|^2 + |A|^2|\nabla A|^2 + |A|^6) r^4\,d\mu \\ &\leq C\int (|A|^3|\nabla^2 A| + |A|^2|\nabla A|^2 + |A|^6 + |\nabla^2 A|\,|A|) r^4 + C\int_{[r>0]} |A|^2\,d\mu \\ &\leq \|A\|^2_{2,[r>0]}\int (|\nabla^2 A|^2 + |A|^2|\nabla A|^2 + |A|^6) r^4\,d\mu + \tfrac{1}{2}\int |\nabla^2 A|^2\,d\mu \\ &\qquad + C\int_{[r>0]} |A|^2\,d\mu + C\Big(\int_{[r>0]} |A|^2\,d\mu\Big)^2.\end{aligned}$$

Since $\int_{[r>0]} |A|^2\,d\mu \leq \epsilon$, it yields

$$\frac{d}{dt}\int |A|^2 r^4\,d\mu + \tfrac{1}{4}\int (|\nabla^2 A|^2 + |A|^2|\nabla A|^2 + |A|^6) r^4\,d\mu \leq C\epsilon. \tag{4-12}$$

The proposition follows by integrating over $[0, t]$. □

This proposition is the local estimate for the flow, from which we can see the variation of curvature is locally small. Furthermore, we have the following estimate as Proposition 3.4 in [Kuwert and Schätzle 2001] for the surface diffusion flow in $\mathbb{F}^{2+p}(c)$ $(c \geq 0)$ with small initial Willmore energy.

Proposition 4.9. *Let $f : \Sigma \times [0, T] \to \mathbb{F}^{2+p}(c)$ $(c \geq 0)$ be a compact surface diffusion flow with $\int_\Sigma |A|^2\,d\mu \leq \aleph$, then there exist constants $\epsilon_0(p) \geq 0$ and $c_1 = C(p)/\aleph > 0$, such that if $\rho > 0$ is chosen with*

$$\int_{\Sigma_\rho} |A^\circ|^2\,d\mu \leq \epsilon < \epsilon_0(p) \text{ at time } t = 0 \text{ for all } \Sigma_\rho \subset \mathbb{F}^{2+p},$$

then for any time $0 \leq t < t_1 = \min\{c_1\rho^4, T\}$, we have

$$\int_{\Sigma_\rho} |A^\circ|^2\,d\mu + \int_0^t \int_{\Sigma_\rho} (|\nabla^2 A|^2 + |A|^2|\nabla A|^2 + |A|^6 + c|\nabla H|^2)\,d\mu dt \leq C(\epsilon + \aleph\rho^{-4} t).$$

Applying Proposition 4.6, Lemma 4.7 and Proposition 4.8, we get the higher order derivative estimate and the proof is similar to Proposition 4.6 in [Kuwert and Schätzle 2002].

Proposition 4.10. *Let $f : \Sigma \times [0, T] \to \mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow. If*

$$\sup_{[0,T]} \int_{[r>0]} |A|^2 \, d\mu \le \epsilon_1(p),$$

where $\epsilon_1(p)$ is small enough, then

$$\|\nabla^m A\|_{\infty\ [r=1]} \le C(m, T, \alpha_0(m+2)),$$

where $\alpha_0(m) = \sum_{j=0}^{m} \|\nabla^j A_0\|_{2\ [r_0>0]}$.

Furthermore, we get the following proposition:

Proposition 4.11 (interior estimates). *Let $f : \Sigma \times (0, T] \to \mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow satisfying the condition*

$$\sup_{(0,T]} \int_{\Sigma_\rho(x)} |A|^2 \, d\mu \le \epsilon < \epsilon_1(p) \quad \textit{for } x \in \mathbb{F}^{2+p}(c),$$

where $T \le C\rho^4$, then for any $k \in \mathbb{N}_0$, at time $t \in (0, T]$, we have the estimates

$$\|\nabla^k A\|_{L^\infty(\Sigma_{\rho/2}(x))} \le c(k)\sqrt{\epsilon}\, t^{-\frac{k+1}{4}},$$
$$\|\nabla^k A\|_{L^2(\Sigma_{\rho/2}(x))} \le c(k)\sqrt{\epsilon}\, t^{-\frac{k}{4}}.$$

These are the higher order derivative estimate, which are localized in time. The proof follows similar to Theorem 3.5 in [Kuwert and Schätzle 2001].

Now we prove the lifespan theorem for the surface diffusion flow.

Proof of Theorem 1.3. We may assume that $\rho = 1$ by rescaling. Put

$$\epsilon(t) = \sup_{x \in \mathbb{F}^{2+p}} \int_{\Sigma_1(x)} |A|^2 \, d\mu.$$

By a trivial covering argument, we get

$$\epsilon(t) \le \Gamma \cdot \sup_{x \in \mathbb{F}^{2+p}} \int_{\Sigma_{\frac{1}{2}}(x)} |A|^2 \, d\mu,$$

for some $\Gamma = \Gamma(p)$. The function $\epsilon : [0, T) \to \mathbb{R}$ is continuous by the compactness of $f(\Sigma \times [0, t])$ for $t < T$. Now let $\lambda > 0$ be a parameter, and define

$$t_0 : \sup\{0 \le t \le \min(T, \lambda) : \epsilon(\tau) \le 3\Gamma\epsilon \text{ for } 0 \le \tau < t\}.$$

The continuity of $\epsilon(t)$ implies $t_0 \ge 0$ and

$$\epsilon(t_0) = 3\Gamma\epsilon \text{ if } t_0 < \min(T, \lambda). \tag{4-13}$$

Fix a cutoff function $\tilde{r} \in C^2(\mathbb{F}^{2+p})$ with $\|\tilde{r}\|_{C^2(\mathbb{F}^{2+p})} \le C$ and $\chi_{B_{1/2}(x)} \le \tilde{r} \le \chi_{B_1(x)}$, then $r = \tilde{r} \circ f$ satisfies condition (4-5). Thus, it follows from Proposition 4.8 that

$$\int_{\Sigma_{\frac{1}{2}}(x)} |A|^2(t)\, d\mu \le \int_{\Sigma_1(x)} |A|^2(0)\, d\mu_0 + C\Gamma\epsilon t \le 2\epsilon \quad \text{for } 0 \le t \le t_0.$$

Taking $\lambda = (C\Gamma)^{-1}$, we conclude that

$$\epsilon(t) \le 2\Gamma\epsilon, \quad 0 \le t \le t_0,$$

and (4-13) implies $t_0 = \min(T, (C\Gamma)^{-1})$. Now if $t_0 = (C\Gamma)^{-1}$, we prove the proposition with a contraction for

$$t_0 = T. \tag{4-14}$$

We can apply Proposition 4.10 to obtain

$$\|\nabla^m A\|_\infty \le C(p, m, f_0).$$

With the same argument as in [Magni 2015], we can get

$$\|\partial^m f\|_\infty, \|\partial^m \partial_t f\|_\infty \le C(p, m, f_0).$$

Then $f(t)$ converges in $C^m(\Sigma)$ as $t \to T$ to a smooth function $f(T)$. By short time existence, we can extend the flow f to an interval $[0, T + \delta)$, which is contrary to the maximality of T. Hence it contradicts with (4-14). This proves Theorem 1.3. □

5. Blowup analysis

In this section, we rescale the surface diffusion flow in $\mathbb{F}^{2+p}(c)$ $(c \ge 0)$ at an assumed singularity, thereby constructing a static surface as a limit. First we need the following local area bound due to L. Simon [1993] in the Euclidean space, which has been generalized by F. Link [2013] in Riemannian manifolds.

Lemma 5.1. *Let $f : \Sigma \to \mathbb{F}^{2+p}(c)$ be a compact surface. Then for $0 < \rho < \infty$ and $\Sigma_\rho = \Sigma_\rho(x_0)$, one has*

$$\frac{\mu(\Sigma_\rho)}{\rho^2} \le C\left(\int_\Sigma |A^\circ|^2\, d\mu + 4\pi\chi(\Sigma)\right). \tag{5-1}$$

Moreover, the global area estimate is given as follows:

Proposition 5.2 (area estimate). *Let $f : \Sigma \times [0, T] \to \mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow with $\int_\Sigma |A^\circ|^2\, d\mu \le \epsilon \le \epsilon_0(p)$, where $\epsilon_0(p)$ is as in Proposition 4.9, then*

$$(1 - C\epsilon)\mu_0(\Sigma) \le \mu(\Sigma) \le \mu_0(\Sigma). \tag{5-2}$$

Proof. Using Lemma 4.1, we have

$$\frac{d}{dt}\int d\mu = -\int |\nabla H|^2\,d\mu < 0,$$

so, $\mu(\Sigma) \le \mu_0(\Sigma)$, and the area is decreasing. On the other hand, by the Sobolev inequalities we have

$$\int |\nabla H|^2\,d\mu \le C(n)\mu(\Sigma)\int |\nabla^2 H|^2 + |H|^2|\nabla H|^2 + c|\nabla H|^2\,d\mu.$$

We obtain

$$\frac{d}{dt}\mu(\Sigma) \ge -C\mu(\Sigma)\int |\nabla^2 H|^2 + |H|^2|\nabla H|^2 + c|\nabla H|^2\,d\mu.$$

Using Proposition 4.9 with $\rho = \infty$ implies

$$\mu(\Sigma) \ge \mu_0(\Sigma)e^{-C\int_0^t\int_\Sigma |\nabla^2 H|^2+|H|^2|\nabla H|^2+c|\nabla H|^2\,d\mu dt} \ge \mu_0(\Sigma)(1-C\epsilon). \qquad \square$$

Now we state the required compactness theorem, which was originally proved by Langer [1985] for surfaces. Recently it has been generalized by Breuning [2015] to any dimension and codimension in the Euclidean space and even to Riemannian setting, where the ambient space can be isometrically embedded into some Euclidean space with bounded second fundamental form. Here we use a simplified version from [Cooper 2011; Magni 2015].

Theorem 5.3 [Cooper 2011, Theorem 1.2]. *Given a compact surface Σ, a sequence of complete Riemannian manifolds $\{(N^{2+p},\bar g_j)\}_{j\in\mathbb{N}}$ with uniformly bounded geometry and two sequences of points $\{q_j\}_{j\in\mathbb{N}} \subset \Sigma$ and $\{x_j\}_{j\in\mathbb{N}} \subset N^{2+p}$, let $f_j : (\Sigma, g_j) \to (N^{2+p},\bar g_j)$ be a sequence of isometric proper immersions such that $f_j(q_j) = x_j$. Suppose that*

$$\mu_j(\Sigma_R(x_j)) \le C(R) \quad \text{for any } R > 0, \tag{5-3}$$

$$\|\nabla^k A_j\|_{L^\infty(\Sigma_R)} \le C_k(R) \quad \text{for any } R > 0 \text{ and any } k \in \mathbb{N}_0. \tag{5-4}$$

Then there exist a surface $\hat\Sigma$, a complete Riemannian manifold $(M^{2+p},\bar g)$ and two points $q \in \hat\Sigma$ and $x \in M^{2+p}$ such that:

- *There exists an increasing exhaustion $\{U_j\}_{j\in\mathbb{N}}$ of $\hat\Sigma$ made of open relatively compact sets, and there are diffeomorphisms $\phi_j : U_j \to \Sigma$ with $\phi_j(q) = q_j$, such that for any $R > 0$ we have $\Sigma_R^{g_j}(q_j) \subset \phi_j(U_j)$, for all $j \ge j_0(R)$.*
- *There is an increasing exhaustion $\{V_j\}_{j\in\mathbb{N}}$ of M^{2+p} made of open relatively compact sets, and diffeomorphisms $\psi_j : V_j \to N^{2+p}$ with $\psi_j(x) = x_j$, such that for any $R > 0$ we have $B_R^{\bar g_j}(x_j) \subset \psi_j(V_j)$, for all $j \ge j_0(R)$.*
- *$\phi_j(U_j) \subset \psi_j(V_j)$ and $\psi_j^*\bar g_j \to \bar g$ smoothly.*

- *There exists a proper immersion $\hat{f} : \hat{\Sigma} \to M^n$ such that $\psi_j^{-1} \circ f_j \circ \phi_j \to \hat{f}$ smoothly with respect to a global isometric embedding of $(M^{2+p}, \bar{g})$ into a suitable Euclidean space $\mathbb{R}^K$. The immersion $\hat{f}$ also satisfies (5-3) and (5-4) with respect to x.*

Here, the sequence of proper immersions f_j converging as in Theorem 5.3 to a proper immersion $\hat{f} : \hat{\Sigma} \to M^{2+p}$, will be denoted by $f_j \to \hat{f}$. Let $f : \Sigma \times [0, T) \to \mathbb{F}^{2+p}(c)$ be a surface diffusion flow defined on a compact surface Σ, where $0 < T \le \infty$. Define

$$\aleph(r, t) = \sup_{x \in \mathbb{F}^{2+p}} \int_{\Sigma_r(x)} |A(t)|^2 \, d\mu_t, \quad \Sigma_r(x) = f^{-1}(B_r(x)).$$

Choose an arbitrary sequence $r_j \searrow 0$ and assume concentration in the sense that for all j,

$$t_j = \inf\{t \ge 0 : \aleph(r_j, t) > \epsilon_2\} < T, \tag{5-5}$$

where $\epsilon_2 = \epsilon_1 / C$, and ϵ_1 and C are the constants from the lifespan theorem. Clearly,

$$\int_{\Sigma_{r_j}(x)} |A(t_j)|^2 \, d\mu_{t_j} \le \epsilon_2 \quad \text{for any } x \in \mathbb{F}^{2+p}(c).$$

On the other hand, choosing an appropriate sequence of balls at times $\tau \searrow t_j$, we find a point $x_j \in \mathbb{F}^{2+p}$ satisfying

$$\int_{f^{-1}(\overline{B_{r_j}(x_j)})} |A(t_j)|^2 \, d\mu_{t_j} \ge \epsilon_2.$$

Now we rescale by considering

$$\begin{gathered} f_j : (\Sigma, g_j) \times [\, -r_j^{-4} t_j, \ r_j^{-4}(T - t_j)) \to (\mathbb{F}^{2+p}, \bar{g}_j), \\ f_j(p, t) = f(p, \ t_j + r_j^4 t), \end{gathered}$$

where $\bar{g}_j = r_j^{-2} \bar{g}$ and $g_j = f_j(\,\cdot\,, t)^* \bar{g}_j$. From Theorem 5.2 in [Magni 2015], we know that the rescaled flows converge locally smoothly on $\hat{\Sigma} \times \mathbb{R}$ to a static surface diffusion flow represented by a static properly immersed surface $\hat{f} : \hat{\Sigma} \to \mathbb{R}^{2+p}$, open sets with the property $\aleph_j(1, t) \le \epsilon_2$, for all $t \le 0$ and

$$\int_{\hat{f}^{-1}(\overline{B_1(0)})} |\hat{A}|^2 d\hat{\mu} \ge \epsilon_2. \tag{5-6}$$

More precisely, the lifespan theorem yields $r_j^{-4}(T - t_j) \ge c_0$ and in fact

$$\aleph_j(1, t) \le \epsilon_1 \quad \text{for } 0 < t \le c_0.$$

We may now apply Proposition 4.11 on parabolic cylinders $B_1(x) \times (t-1, t]$ to obtain

$$\|\nabla^k A\|_{L^\infty} \le c(k) \quad \text{for } r_j^{-4} + 1 \le t \le c_0.$$

Then we can apply the compactness theorem to the sequence

$$f_j = f_j(\,\cdot\,, 0) : \Sigma \to \mathbb{F}^{2+p},$$

thus obtaining a limit immersion $\hat{f}_0 : \hat{\Sigma} \to \mathbb{R}^{2+p}$. Then by reparametrization $f_j(\phi_j, \cdot)$ is a surface diffusion flow and has initial data converging locally in C^k to the immersion $\hat{f}_0$. By standard estimates for geometric evolution equations in [Kuwert and Schätzle 2002], we deduce the locally smooth convergence $f_j \to \hat{f}$, where

$$\hat{f} : \hat{\Sigma} \times [0, c_0] \to \mathbb{R}^{2+p}$$

is a surface diffusion flow with initial data $\hat{f}_0$.

In order to get the static blowup limit, we need the monotonicity of the Willmore functional.

Lemma 5.4. *Let $f : \Sigma \times (0, T] \to \mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow with small initial Willmore energy satisfying Proposition 5.2; we have*

$$\begin{aligned}\frac{d}{dt}\int |A^\circ|^2\, d\mu &\le \frac{1}{2}\frac{d}{dt}\int |A|^2\, d\mu \\ &\le \frac{1}{2}\frac{d}{dt}\int |H|^2\, d\mu \le -\frac{1}{2}\int |\Delta H|^2\, d\mu.\end{aligned}$$

Proof. The first two inequalities are obvious. From (3-7) and (3-8) with $r = 1$, we know

$$\begin{aligned}c\int |\nabla H|^2\, d\mu \le C\int |A\|A^\circ\|\nabla A^\circ|^2\, d\mu &\le C\int |A|^3|A^\circ|^3 + |\nabla A^\circ|^3\, d\mu \\ &\le C\int |A|^4|A^\circ|^2\, d\mu + \delta\int |\nabla^2 A^\circ|^2\, d\mu + C_\delta\int |A^\circ|^6\, d\mu.\end{aligned}$$

From (3-11) we know

$$\int (|\nabla^2 A|^2 + |A|^2|\nabla A|^2 + |A|^4|A^\circ|^2)\, d\mu + c\int |A^\circ|^2\, d\mu \le \int |\Delta H|^2\, d\mu,$$

and (3-10) implies

$$\begin{aligned}\int |A|^2|A^\circ|^4\, d\mu + c\int |\nabla H|^2\, d\mu &\le \int (\delta|A|^4|A^\circ|^2 + C_\delta|A^\circ|^6 + c|\nabla H|^2)\, d\mu \\ &\le (\delta + c\epsilon + \epsilon C_\delta)\int |\Delta H|^2\, d\mu.\end{aligned}$$

So,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\int |H|^2\,d\mu &= -\int (\Delta H + H|A^\circ|^2 + 2cH)(\Delta H)\,d\mu \\ &\leq -\frac{3}{4}\int |\Delta H|^2\,d\mu + 2\int |A|^2|A^\circ|^4\,d\mu + 2c\int |\nabla H|^2\,d\mu \\ &\leq -\frac{1}{2}\int |\Delta H|^2\,d\mu.\end{aligned}$$

□

Proposition 5.5. *Let $f : \Sigma \times (0, T] \to \mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow with small initial Willmore energy, then the blow up $\hat{f}$ which is constructed above is static.*

Proof. Let $U \Subset \hat{\Sigma}$ be an open set and ϕ_j be the diffeomorphisms in Theorem 5.3, then from Lemma 5.4 and the scale invariance of $\|A\|_2^2$,

$$\begin{aligned}&\int_0^{c_0}\int_U |\Delta H(f_j(\phi_j, t))|^2\,d\mu_{f_j}(\phi_j,\cdot)\,dt = \int_0^{c_0}\int_{\phi_j(U)} |\Delta H_j|^2\,d\mu_j\,dt \\ &\leq \int_\Sigma |A_j(0)|^2\,d\mu_j - \int_\Sigma |A_j(c_0)|^2\,d\mu_j = \int_\Sigma |A(t_j)|^2\,d\mu - \int_\Sigma |A(t_j + r_j^4 c_0)|^2\,d\mu,\end{aligned}$$

and it converges to zero as $j \to \infty$, Therefore $\Delta H(\hat{f}) \equiv 0$ and the blow up $\hat{f}$ is static, which means that $\hat{f}(\,\cdot\,, t) = \hat{f}_0$. Furthermore (5-6) implies

$$\int_{\hat{f}^{-1}(\overline{B_1(0)})} |\hat{A}|^2 d\hat{u} \geq \epsilon_2 > 0.$$

Thus $\hat{f}$ is not totally geodesic. □

Lemma 5.6. *Let $\hat{f} : \hat{\Sigma} \to \mathbb{R}^{2+p}$ be the blowup constructed above. If $\hat{\Sigma}$ contains a compact component C, then in fact $\hat{\Sigma} = C$ and Σ is diffeomorphic to C.*

Proposition 5.7 (nontriviality of the blowup). *Let $\hat{f} : \hat{\Sigma} \to \mathbb{R}^{2+p}$ be the blowup of a compact surface diffusion flow as constructed above. Then none of the components of $\hat{f}$ is compact. In particular, the blowup has a component which is a noncompact nonumbilical surface with $\Delta H \equiv 0$.*

Proof. Assume that there is a compact umbilical component of $\hat{f}$, then Lemma 5.6 implies $\hat{f}$ has no further components. It follows that, up to diffeomorphism $\phi_i : \hat{\Sigma} \to \Sigma$, we have

$$\int_\Sigma |A^\circ(t_j)|^2\,d\mu = \int_\Sigma |A_j^\circ(0)|^2\,d\mu_j \to 0,$$
$$\mu(t_j)(\Sigma) = r_j \mu_j(0)(\Sigma) \to 0.$$

This contradicts the area estimate. □

Now we prove the longtime existence. Combining the previous theorems, we can finally rule out concentration of curvature in finite time.

Proposition 5.8. *Let $f : \Sigma \times [0, T) \to \mathbb{F}^{2+p}(c)$ be a compact surface diffusion flow, then there exists a constant $\epsilon_0(p)$ such that if*

$$\int_\Sigma |A^\circ|^2 \, d\mu|_{t=0} < \epsilon_0 \le \sigma(2), \tag{5-7}$$

then $T = \infty$ and there exists a radius $r_0 > 0$ such that

$$\int_{\Sigma_{r_0}(x)} |A|^2 \, d\mu < \epsilon_2, \tag{5-8}$$

for all $x \in \mathbb{F}^{2+p}$, $t \in [0, +\infty)$, where $\epsilon_2 > 0$ as in (5-5).

Proof. First, if curvature concentrates at time T, then we perform a blowup as above at T and get a static surface $\hat{f}$ with small tracefree curvature. Now the gap theorem implies $\hat{f}$ must be a plane or sphere, which contradicts the nontriviality of the blowup. There does not exist a finite time when curvature concentrates, so $T = \infty$.

Then we claim there exists a radius $r_0 > 0$, such that

$$\int_{\Sigma_{r_0}(x)} |A|^2 \, d\mu < \epsilon_2 \quad \text{for all } x \in \mathbb{F}^{2+p}, \quad t \in [0, \infty),$$

where $\epsilon_2 > 0$ is as in(5-5).

If not, in the same way we can perform a blowup as above, and get a static surface $\hat{f} : \hat{\Sigma} \to \mathbb{R}^{2+p}$ with

$$\int_{\hat{f}^{-1}(\overline{B_1(0)})} |\hat{A}|^2 d\hat{\mu} \ge \epsilon_2,$$

whereas

$$\int_{\hat{\Sigma}} |\hat{A}^\circ|^2 d\hat{\mu} \le \epsilon_0,$$

due to the scale invariance of $\|A^\circ\|_2^2$. Now the gap theorem implies $\hat{f}$ must be totally umbilical, which contradicts the nontriviality of the blowup. □

Using a similar argument as in Proposition 5.5, we get the smooth convergence from Propositions 4.11 and 5.8.

Proof of Theorem 1.4. For any sequence $t_j \to \infty$, there exist $\phi_j \in \mathrm{Diff}(\Sigma)$ such that, after passing to a subsequence, $f(\phi_j, t_j)$ converges smoothly to a static surface. The gap theorem implies it is a union of planes and spheres. After excluding several components as in [Kuwert and Schätzle 2001] and investigating the asymptotic behavior as in [Wheeler 2012], we get the global existence and exponential convergence for the surface diffusion flow with small initial Willmore energy. □

In particular, the convergence theorem implies a differentiable sphere theorem.

Corollary 5.9. *Let $f : \Sigma \to \mathbb{F}^{2+p}(c)$ $(c \geq 0)$ be a compact surface. If there exists a positive constant $\epsilon_0(p)$ such that*

$$\int_\Sigma |A^\circ|^2 \, d\mu \ < \epsilon_0(p),$$

then Σ is diffeomorphic to the unit sphere.

The proof of Theorem 1.4 also applies to the Willmore flow in a sphere.

6. Open problems

In this section, we propose several problems for the convergence of the curvature flow. Applying the Morse theory of submanifolds, Shiohama and Xu [2000] obtained a topological sphere theorem for n-dimensional compact submanifolds in $\mathbb{F}^{n+p}(c)$ $(c \geq 0)$ under the pinching condition for the Willmore functional. Thus, a natural problem is whether or not M^n is diffeomorphic to $\mathbb{S}^n$ under the pinching condition for the Willmore functional. More precisely, we propose the following problem.

Conjecture 6.1. *Let $f : M^n \to \mathbb{F}^{n+p}(c)$ be a compact submanifold. There exists a positive constant $C_3(n)$ such that if*

$$\int_M |A^\circ|^n \, d\mu < C_3(n), \tag{6-1}$$

then the diffusion flow for submanifolds with initial data f exists smoothly for all time and converges to a round sphere as $t \to \infty$. In particular, M is diffeomorphic to $\mathbb{S}^n$.

Liu, Xu, Ye and Zhao [2018] investigated the convergence of the mean curvature flow of compact n-dimensional submanifolds in $\mathbb{R}^{n+p}$. They proved if the initial submanifold satisfies some suitable integral curvature conditions, then along the mean curvature flow it will shrink to a round point in finite time.

T. J. Willmore [1968] proved the Willmore inequality about the total mean curvature $\int |H|^2 \, d\mu$ for compact surfaces in $\mathbb{R}^3$, and then B. Y. Chen [1971] obtained a general version of the Willmore inequality for compact submanifolds in $\mathbb{R}^{n+p}$, as follows:

Theorem 6.2. *If M is an n-dimensional compact submanifold in $\mathbb{R}^{n+p}$, then*

$$n^n \operatorname{Vol}(\mathbb{S}^n) \leq \int_M |H|^n \, d\mu,$$

where the equality holds if and only if $M^n = \mathbb{S}^n(r)$.

Therefore, the total mean curvature for compact submanifolds in $\mathbb{R}^{n+p}$ has a natural lower bound. Applying the Morse theory of submanifolds, Xu [2007] obtained a topological sphere theorem under the total mean curvature pinching condition. For other topological sphere theorems we can see [Shiohama and Xu 1994; 2000]. Similarly, we have the following problem for the diffusion flow under the total mean curvature pinching condition.

Conjecture 6.3. *Let $f : M^n \to \mathbb{R}^{n+p}$ be a compact submanifold. There exists a positive constant $C_4(n)$ such that if*

$$\int_M |H|^n \, d\mu < n^n \operatorname{Vol}(\mathbb{S}^n) + C_4(n),$$

then the diffusion flow for submanifolds with initial data f exists smoothly for all time and converges to a round sphere as $t \to \infty$. In particular, M is diffeomorphic to $\mathbb{S}^n$.

For surfaces in $\mathbb{R}^3$, the small Willmore energy assumption means a pinching condition for $\int |H|^2 \, d\mu$ and the convergence of the surface diffusion flow has answered these two problems. The Willmore conjecture verified by Marques and Neves [2014] says the integral of the square of the mean curvature of a torus in $\mathbb{R}^3$ is at least $8\pi^2$. In fact, they proved the integral inequality of the square of the mean curvature for any compact surfaces with genus greater than or equal to one. Thus, a strong version of the problem related to the Marques and Neves's theorem is: what is the best pinching constant in Conjecture 6.3 for compact surfaces in $\mathbb{R}^3$?

Conjecture 6.4. *Let $f : \Sigma \to \mathbb{R}^3$ be a compact surface satisfying*

$$\int_\Sigma |H|^2 \, d\mu < 8\pi^2,$$

then the surface diffusion flow with initial data f exists smoothly for all time and converges to a round sphere as $t \to \infty$. In particular, Σ is diffeomorphic to $\mathbb{S}^2$.

Conjecture 6.5. *Let $f : \Sigma \to \mathbb{R}^3$ be a compact surface satisfying*

$$\int_\Sigma |H|^2 \, d\mu < 8\pi^2,$$

then the Willmore flow with initial data f exists smoothly for all time and converges to a round sphere as $t \to \infty$. In particular, Σ is diffeomorphic to $\mathbb{S}^2$.

For higher dimensions, there are seldom results about the diffusion flow. However it provides a feasible method to handle the above problems so we can study the gradient flow of some constructed functionals. Li [2002] investigated the rigidity of the Willmore submanifold, which is the critical point of the Willmore functional for n-dimensional compact submanifolds in a sphere. As the Willmore flow for surfaces, we can also study the higher-dimensional Willmore flow, which is the gradient flow of the Willmore functional for n-dimensional compact submanifolds in space forms. More precisely, we will study the following higher-dimensional Willmore flow

$$\frac{\partial f}{\partial t} = -\operatorname{grad} \mathcal{W}(f), \tag{6-2}$$

where $\operatorname{grad} \mathcal{W}(f)$ is the gradient vector field of the Willmore functional (1-2). We propose the following problems.

Conjecture 6.6. *Let $f : M^n \to \mathbb{F}^{n+p}(c)$ be a compact submanifold. There exists a positive constant $C_5(n)$ such that if*

$$\int_M |A^\circ|^n \, d\mu < C_5(n), \tag{6-3}$$

then the Willmore flow for submanifolds with initial data f exists smoothly for all time and converges to a round sphere as $t \to \infty$. In particular, M is diffeomorphic to $\mathbb{S}^n$.

Conjecture 6.7. *Let $f : M^n \to \mathbb{R}^{n+p}$ be a compact submanifold. There exists a positive constant $C_6(n)$ such that if*

$$\int_M |H|^n \, d\mu < n^n \operatorname{Vol}(\mathbb{S}^n) + C_6(n),$$

then the Willmore flow for submanifolds with initial data f exists smoothly for all time and converges to a round sphere as $t \to \infty$. In particular, M is diffeomorphic to $\mathbb{S}^n$.

Acknowledgment. The authors would like to thank the referee for valuable suggestions.

References

[Andrews and Baker 2010] B. Andrews and C. Baker, "Mean curvature flow of pinched submanifolds to spheres", *J. Differential Geom.* **85**:3 (2010), 357–395. MR Zbl

[Bernard et al. 2019] Y. Bernard, G. Wheeler, and V.-M. Wheeler, "Concentration-compactness and finite-time singularities for Chen's flow", *J. Math. Sci. Univ. Tokyo* **26**:1 (2019), 55–139. MR Zbl

[Brendle and Schoen 2009] S. Brendle and R. Schoen, "Manifolds with 1/4-pinched curvature are space forms", *J. Amer. Math. Soc.* **22**:1 (2009), 287–307. MR Zbl

[Breuning 2015] P. Breuning, "Immersions with bounded second fundamental form", *J. Geom. Anal.* **25**:2 (2015), 1344–1386. MR Zbl

[Chen 1971] B.-y. Chen, "On a theorem of Fenchel–Borsuk–Willmore–Chern–Lashof", *Math. Ann.* **194** (1971), 19–26. MR Zbl

[Cooper 2011] A. A. Cooper, "A compactness theorem for the second fundamental form", preprint, 2011. arXiv

[Escher et al. 1998] J. Escher, U. F. Mayer, and G. Simonett, "The surface diffusion flow for immersed hypersurfaces", *SIAM J. Math. Anal.* **29**:6 (1998), 1419–1433. MR Zbl

[Gu and Xu 2012] J.-R. Gu and H.-W. Xu, "The sphere theorems for manifolds with positive scalar curvature", *J. Differential Geom.* **92**:3 (2012), 507–545. MR Zbl

[Gu et al. 2017] J.-R. Gu, H.-W. Xu, Z.-Y. Xu, and E.-T. Zhao, "A survey on rigidity problems in geometry and topology of submanifolds", pp. 79–99 in *Proceedings of the Sixth International Congress of Chinese Mathematicians*, vol. II, edited by C.-S. Lin et al., Adv. Lect. Math. **37**, International Press, Somerville, MA, 2017. MR Zbl

[Hamilton 1982] R. S. Hamilton, "Three-manifolds with positive Ricci curvature", *J. Differential Geometry* **17**:2 (1982), 255–306. MR Zbl

[Huisken 1984] G. Huisken, "Flow by mean curvature of convex surfaces into spheres", *J. Differential Geom.* **20**:1 (1984), 237–266. MR Zbl

[Kuwert and Schätzle 2001] E. Kuwert and R. Schätzle, "The Willmore flow with small initial energy", *J. Differential Geom.* **57**:3 (2001), 409–441. MR Zbl

[Kuwert and Schätzle 2002] E. Kuwert and R. Schätzle, "Gradient flow for the Willmore functional", *Comm. Anal. Geom.* **10**:2 (2002), 307–339. MR Zbl

[Kuwert and Schätzle 2004] E. Kuwert and R. Schätzle, "Removability of point singularities of Willmore surfaces", *Ann. of Math.* (2) **160**:1 (2004), 315–357. MR Zbl

[Langer 1985] J. Langer, "A compactness theorem for surfaces with L_p-bounded second fundamental form", *Math. Ann.* **270**:2 (1985), 223–234. MR Zbl

[Li 2002] H. Li, "Willmore submanifolds in a sphere", *Math. Res. Lett.* **9**:5-6 (2002), 771–790. MR Zbl

[Link 2013] F. Link, *Gradient flow for the Willmore functional in Riemannian manifolds of bounded geometry*, Ph.D. thesis, Albert-Ludwigs-Universität Freiburg, 2013. Zbl arXiv

[Liu et al. 2018] K. Liu, H. Xu, F. Ye, and E. Zhao, "The extension and convergence of mean curvature flow in higher codimension", *Trans. Amer. Math. Soc.* **370**:3 (2018), 2231–2262. MR Zbl

[Magni 2015] A. Magni, "A convergence result for the gradient flow of $\int |A|^2$ in Riemannian manifolds", *Geom. Flows* **1**:1 (2015), 1–10. MR Zbl

[Marques and Neves 2014] F. C. Marques and A. Neves, "Min-max theory and the Willmore conjecture", *Ann. of Math.* (2) **179**:2 (2014), 683–782. MR Zbl

[Mayer 2001] U. F. Mayer, "Numerical solutions for the surface diffusion flow in three space dimensions", *Comput. Appl. Math.* **20**:3 (2001), 361–379. MR Zbl

[Metzger et al. 2013] J. Metzger, G. Wheeler, and V.-M. Wheeler, "Willmore flow of surfaces in Riemannian spaces I: Concentration-compactness", preprint, 2013. arXiv

[Michael and Simon 1973] J. H. Michael and L. M. Simon, "Sobolev and mean-value inequalities on generalized submanifolds of R^n", *Comm. Pure Appl. Math.* **26** (1973), 361–379. MR Zbl

[Mullins 1957] W. W. Mullins, "Theory of thermal grooving", *J. Appl. Phys.* **28**:3 (1957), 333–339.

[Shiohama and Xu 1994] K. Shiohama and H. W. Xu, "Rigidity and sphere theorems for submanifolds", *Kyushu J. Math.* **48**:2 (1994), 291–306. MR Zbl

[Shiohama and Xu 2000] K. Shiohama and H. W. Xu, "Rigidity and sphere theorems for submanifolds, II", *Kyushu J. Math.* **54**:1 (2000), 103–109. MR Zbl

[Simon 1993] L. Simon, "Existence of surfaces minimizing the Willmore functional", *Comm. Anal. Geom.* **1**:2 (1993), 281–326. MR Zbl

[Taylor and Cahn 1994] J. E. Taylor and J. W. Cahn, "Linking anisotropic sharp and diffuse surface motion laws via gradient flows", *J. Statist. Phys.* **77**:1-2 (1994), 183–197. MR Zbl

[Wang 2008] M.-T. Wang, "Lectures on mean curvature flows in higher codimensions", pp. 525–543 in *Handbook of geometric analysis*, vol. 1, edited by L. Ji et al., Adv. Lect. Math. **7**, International Press, Somerville, MA, 2008. MR Zbl

[Wheeler 2012] G. Wheeler, "Surface diffusion flow near spheres", *Calc. Var. Partial Differential Equations* **44**:1-2 (2012), 131–151. MR Zbl

[Willmore 1968] T. J. Willmore, "Mean curvature of immersed surfaces", *An. Şti. Univ. "All. I. Cuza" Iaşi Secţ. I a Mat. (N.S.)* **14** (1968), 99–103. MR Zbl

[Xu 2007] H. W. Xu, "Mean value theorem for critical points and sphere theorems", pp. 203–217 in *Proceedings of the 4th International Congress of Chinese Mathematicians*, vol. 2, 2007.

[Xu and Gu 2007] H.-W. Xu and J.-R. Gu, "L^2-isolation phenomenon for complete surfaces arising from Yang–Mills theory", *Lett. Math. Phys.* **80**:2 (2007), 115–126. MR Zbl

[Xu and Gu 2013] H.-W. Xu and J.-R. Gu, "Geometric, topological and differentiable rigidity of submanifolds in space forms", *Geom. Funct. Anal.* **23**:5 (2013), 1684–1703. MR Zbl

[Xu and Yang 2016] H. Xu and D. Yang, "Rigidity theorem for Willmore surfaces in a sphere", *Proc. Indian Acad. Sci. Math. Sci.* **126**:2 (2016), 253–260. MR Zbl

Received April 10, 2019. Revised December 11, 2019.

DONG PU
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
CHINA
pudong@zju.edu.cn

HONGWEI XU
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
CHINA
xuhw@zju.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.321

NONVANISHING SQUARE-INTEGRABLE AUTOMORPHIC COHOMOLOGY CLASSES: THE CASE GL(2) OVER A CENTRAL DIVISION ALGEBRA

JOACHIM SCHWERMER

Let k be a totally real algebraic number field, and let D be a central division algebra of degree d over k. The connected reductive algebraic k-group $\mathrm{GL}(2, D)/k$ has k-rank one; it is an inner form of the split k-group $\mathrm{GL}(2d)/k$. We construct automorphic representations π of $\mathrm{GL}(2d)/k$ which occur nontrivially in the discrete spectrum of $\mathrm{GL}(2d, k)$ and which have specific local components at archimedean as well as nonarchimedean places of k so that there exist automorphic representations π' of $\mathrm{GL}(2, D)(\mathbb{A}_k)$ with $\Xi(\pi') = \pi$ under the Jacquet–Langlands correspondence. These requirements depend on the finite set V_D of places of k at which D does not split, and on the quest to construct representations π' of $\mathrm{GL}(2, D)(\mathbb{A}_k)$ which either represent cuspidal cohomology classes or give rise to square-integrable classes which are not cuspidal, that is, are eventually represented by a residue of an Eisenstein series. The demand for cohomological relevance gives strong constraints at the archimedean components.

1. Introduction

1A. ***The square-integrable cohomology groups $H^*_{(\mathrm{sq})}(G, \mathbb{C})$.*** Let G be a reductive algebraic group over a totally real algebraic number field k, and suppose that G modulo its radical has k-rank greater than zero. We write G_∞ for the group $R_{k/\mathbb{Q}}(G)(\mathbb{R})$ of real points of the algebraic $\mathbb{Q}$-group $R_{k/\mathbb{Q}}(G)$ obtained from G by restriction of scalars, and K_∞ for a maximal compact subgroup of G_∞. Within the framework of the automorphic cohomology $H^*(G, \mathbb{C})$ of a reductive algebraic group G over k one has the notion of square-integrable cohomology (see Section 8 for details). This subspace of $H^*(G, \mathbb{C})$, to be denoted $H^*_{(\mathrm{sq})}(G, \mathbb{C})$, reflects the

This work was partly carried through during my stay in the fall term 2017 at the School of Mathematics, Institute for Advanced Study, Princeton, and then pursued at the Max–Planck Institute for Mathematics, Bonn; I gratefully acknowledge the funding at the IAS, provided by the Charles Simonyi Endowment, as well as the support at the MPIM.
MSC2010: primary 11F70, 11F75; secondary 22E40, 22E55.
Keywords: cohomology of arithmetic groups, square-integrable automorphic representations, general linear group over a central division algebra.

contribution of the discrete spectrum $L^2_{\mathrm{disc},J}(G)$ of G to the cohomology. It contains the cuspidal spectrum $L^2_{\mathrm{cusp},J}(G)$. In fact, there is a decomposition

$$L^2_{\mathrm{disc},J}(G) = L^2_{\mathrm{cusp},J}(G) \oplus L^2_{\mathrm{res},J}(G) \tag{1-1}$$

where the complement $L^2_{\mathrm{res},J}(G)$ denotes the residual spectrum of G. Each constituent of $L^2_{\mathrm{res},J}(G)$ can be structurally described in terms of residues of Eisenstein series attached to irreducible representations occurring in the discrete spectra of the Levi components of proper parabolic k-subgroups of G.

On the cohomological level, this presents itself as a chain of inclusions

$$H^*_{\mathrm{cusp}}(G, \mathbb{C}) \subset H^*_!(G, \mathbb{C}) \subset H^*_{(\mathrm{sq})}(G, \mathbb{C}) \subset H^*(G, \mathbb{C}). \tag{1-2}$$

where $H^*_{\mathrm{cusp}}(G, \mathbb{C})$, the cuspidal cohomology of G, corresponds to the cuspidal spectrum. The so-called interior cohomology $H^*_!(G, \mathbb{C})$, a topologically defined object, is sandwiched between two analytically defined cohomology groups.

1B. ***Nonvanishing results for the square-integrable cohomology of* GL(2, *D*).** The question arises as to how one can detect nonvanishing square-integrable cohomology classes in $H^*_{(\mathrm{sq})}(G, \mathbb{C})$ and related automorphic representations. In this paper we study this problem in the case of the general linear group $\mathrm{GL}(2, D)$ over a finite-dimensional central division algebra D of degree $d > 1$, defined over a totally real algebraic number field k. The group $\mathrm{GL}(2, D)/k$ is an inner form of the general linear group $\mathrm{GL}(2d)/k$.

Two ingredients are essential in this investigation: Firstly, the global Jacquet–Langlands correspondence by Badulescu [2008] and Badulescu and Renard [2010] which relates via an injective map, to be denoted Ξ, the set of the irreducible constituents of the discrete spectrum of $\mathrm{GL}(2, D)(\mathbb{A}_k)$ with the set of the irreducible constituents of the discrete spectrum of $\mathrm{GL}(2d, k)$ (see Section 3 for details).

Secondly, we have to construct automorphic representations $\pi = \otimes'_{v\in V_k}\pi_v$ of $\mathrm{GL}(2d)/k$ which occur nontrivially in the discrete spectrum of $\mathrm{GL}(2d, k)$ and which have specific local components at archimedean as well as nonarchimedean places of k so that there exists a corresponding automorphic representation $\pi' = \otimes'_{v\in V_k}\pi'_v$ of $\mathrm{GL}(2, D)(\mathbb{A}_k)$ with $\Xi(\pi') = \pi$. These requirements depend on D, more precisely, on the finite set V_D of places of k at which D does not split, and on the quest to construct representations π' of $\mathrm{GL}(2, D)(\mathbb{A}_k)$ which either represent cuspidal cohomology classes or give rise to square-integrable classes which are not cuspidal, that is, are eventually represented by a residue of an Eisenstein series. The demand for cohomological relevance gives strong constraints at the archimedean components of $\pi_v,\ v \in V_{k,\infty}$.

We finally construct three different kinds of nonvanishing square-integrable cohomology classes in $H^*_{(\mathrm{sq})}(G, \mathbb{C})$ (see Theorems 8.1, 7.3, and 8.5):

(a) Classes in the cuspidal cohomology $H^*_{(\mathrm{cusp})}(\mathrm{GL}(2, D), \mathbb{C})$ which correspond to a cuspidal representation of $\mathrm{GL}(2d, k)$.

(b) Classes in the cuspidal cohomology $H^*_{(\mathrm{cusp})}(\mathrm{GL}(2, D), \mathbb{C})$ which correspond to a residual representation of $\mathrm{GL}(2d, k)$.

(c) Noncuspidal classes in $H^*_{(\mathrm{sq})}(\mathrm{GL}(2, D), \mathbb{C})$ which correspond to a residual representation of $\mathrm{GL}(2d, k)$ of a type different from the one occurring in (b).

1C. ***Example.*** We illustrate these results by the following example: Let k be a totally real field of degree $[k : \mathbb{Q}] = 4$, and let D be a central division algebra of degree 2 over k. Suppose that $|V_D| = 6$ and that $V_{\infty,k} \subset V_D$. Then the cuspidal representation π' of $\mathrm{GL}(2, D)(\mathbb{A}_k)$ constructed in Theorem 8.1 contributes nontrivially to the cuspidal cohomology $H^*_{\mathrm{cusp}}(H', \mathbb{C})$ of $H' = \mathrm{GL}(2, D)$ in degrees 8, 9, 10, 11, 12, that is, in a range of degrees centered around the middle dimension 10. By contrast, the cuspidal representation constructed in Theorem 7.3 contributes nontrivially in degrees 4, 7, 10, 13, 16. The cohomology class obtained via the residual spectrum contributes in degree 4 to the square-integrable cohomology. Note that these residual classes are carried at the archimedean components by the same irreducible nontempered unitary representation as the nontempered cuspidal classes.

1D. We describe two of the results obtained in a more precise way.

Theorem 1.1. *Given a totally real number field k of degree ℓ, let D be a finite-dimensional central division algebra over k of degree $d > 1$. Let $V_D \subset V_k$ be the finite set of places of k at which D does not split. Let t denote the number of archimedean places in V_D, and suppose that $t > 0$. Then there exist automorphic representations $\pi' = \otimes'_{v \in V_k} \pi_v$ of* $\mathrm{GL}(2, D)(\mathbb{A}_k)$ *which occur as irreducible constituents in the residual spectrum of* $\mathrm{GL}(2, D)(\mathbb{A}_k)$, *whose archimedean components π'_v are irreducible nontempered unitary representations of* $\mathrm{GL}(d, \mathbb{H})$ *for $v \in V_D \cap V_{k,\infty}$ (resp. of* $\mathrm{GL}(2d, \mathbb{R})$ *for $v \in V_{k,\infty}$, $v \notin V_D$), and which give rise to a nontrivial cohomology class in $H^*_{(\mathrm{sq})}(\mathrm{GL}(2, D), \mathbb{C})$ that is not cuspidal.*

The proof relies on the description of the residual spectrum of $\mathrm{GL}(2d, \mathbb{A}_k)$ in [Mœglin and Waldspurger 1989] and an explicit construction of an irreducible tempered representation of $\mathrm{GL}(d, \mathbb{A}_k)$ with prescribed local and global properties, aligned with the demands (see Theorem 3.4) defined by the Jacquet–Langlands correspondence.

We will give the archimedean components π'_v, $v \in V_{k,\infty}$, of the representation π' in the theorem in a precise form in Section 4, denoted by $J_{\mathbb{R}}(2, \theta)$ in the case of the group $\mathrm{GL}(2d, \mathbb{R})$, and denoted by $J'_{\mathbb{R}}(2, \theta')$ in the case of $\mathrm{GL}(d, \mathbb{H})$. These two representations correspond to one another by the local Jacquet–Langlands correspondence. In the latter case, the Poincaré polynomial is given in Proposition 4.2. This permits us, for example, to conclude the following result.

Corollary. *Suppose D does not split at all archimedean places, i.e., $V_{k,\infty} \subset V_D$. Then there exists a nonvanishing noncuspidal cohomology class of degree*

$$q = \ell \cdot \tfrac{1}{2} d(2d-3)$$

in the square-integrable cohomology $H^*_{(\mathrm{sq})}(\mathrm{GL}(2, D))$.

With regard to the construction of cuspidal cohomology classes we discuss the case of nontempered classes. For the other case we refer to Theorem 8.1

Theorem 1.2. *Let k be a totally real number field, and let D be a finite-dimensional central division algebra over k of degree $d > 1$. Suppose that the set V_D of places of D at which D does not split contains at least one archimedean place. Then there exist cuspidal automorphic representations $\pi' = \otimes \pi'_v$ of $H'(\mathbb{A}_k) = \mathrm{GL}(2, D)(\mathbb{A}_k)$ with $\Xi(\pi') =: \pi$ a residual representation of the group $H(\mathbb{A}_k) = \mathrm{GL}(2d, \mathbb{A}_k)$ under the Jacquet–Langlands correspondence Ξ so that the archimedean components π'_v, $v \in V_{\infty,k}$, have the following form: If $v \in V_D \cap V_{k,\infty}$, that is, $H'_v \cong \mathrm{GL}(d, \mathbb{H})$, then $\pi'_v \cong J'_{\mathbb{R}}(2, \theta')$, and if $v \in V_{k,\infty}$, $v \notin V_D$, that is, $H'_v \cong \mathrm{GL}(2d, \mathbb{R})$, then $\pi'_v \cong J_{\mathbb{R}}(2, \theta)$. In both cases the archimedean component is a nontempered representation of H'_v. The representation π' represents a nontrivial class in $H^*_{\mathrm{cusp}}(\mathrm{GL}(2, D), \mathbb{C})$.*

For the proof of this result, consider the uniquely determined standard maximal parabolic k-subgroup Q_d of $H/k = \mathrm{GL}(2d, k)$ which is conjugate to its opposite. We construct a specific residual automorphic representation of $H(\mathbb{A}_k)$, essentially via the residue of an Eisenstein series attached to a cuspidal automorphic representation of the Levi component $L_{Q_d} \cong \mathrm{GL}(d)/k \times \mathrm{GL}(d)/k$ of Q_d. Required by the description of the image of the map Ξ, this cuspidal representation has to satisfy some local conditions at places in V_D as well as at the archimedean places. Thus, secondly, we use the process of global automorphic induction (see [Henniart 2012]) to construct such a cuspidal representation. It is decisive that the global automorphic induction is compatible with the local automorphic induction. We refer to Section 5B for the construction.

Remark 1.3. The archimedean components of the cuspidal representations as constructed above are of the form $J'_{\mathbb{R}}(2, \theta')$. This nontempered unitary representation of $\mathrm{GL}(d, \mathbb{H})$ also appears as an archimedean component (for a place $v \in V_D \cap V_{k,\infty}$) of the global automorphic representation of the adèle group $\mathrm{GL}(2, D)(\mathbb{A}_k)$ which contributes to the noncuspidal cohomology. However, in the former case, the contribution to cohomology is over the full range of cohomological degrees associated with the representation $J'_{\mathbb{R}}(2, \theta')$, whereas in the latter case the degree $q = \ell \cdot \frac{1}{2} d(2d-3)$ is just the minimal degree in which the representation has cohomology. Whether there are higher degrees in which there is a contribution to the noncuspidal square-integrable cohomology is an important question. If $v \in V_{k,\infty}$, $v \notin V_D$, and thus, $\pi'_v \cong J_{\mathbb{R}}(2, \theta)$, so the same question arises (see Remarks 8.6 and 8.7 for details).

Notation and conventions. Let k be an algebraic number field, i.e., an arbitrary finite extension $k/\mathbb{Q}$ of the field $\mathbb{Q}$ of rational numbers, and let $\mathcal{O}_k$ denote its ring of integers. The set of places of k will be denoted by V_k, and $V_{k,\infty}$ (resp. $V_{k,f}$) refers to the subsets of archimedean (resp. nonarchimedean) places of k. Given a place $v \in V_k$, the completion of k with respect to v is denoted by k_v. For a finite place $v \in V_{k,f}$ we write $\mathcal{O}_{k,v}$ for the valuation ring in k_v. If the field k is fixed, we write $V = V_k$ etc.

Let $\mathbb{A}_k$ (resp. $\mathbb{I}_k$) be the ring of adèles (resp. the group of idèles) of k. We denote by $\mathbb{A}_{k,\infty} = \prod_{v \in V_{k,\infty}} k_v$ the archimedean component of the ring $\mathbb{A}_k$, and by $\mathbb{A}_{k,f}$ the finite adèles of k. There is the usual decomposition of $\mathbb{A}_k$ into the archimedean and the nonarchimedean part $\mathbb{A} = \mathbb{A}_{k,\infty} \times \mathbb{A}_{k,f}$.

2. Generalities

In this section, mainly to fix notation, we recollect some background material regarding the general linear group over a finite-dimensional central division algebra defined over some algebraic number field.

2A. ***The algebraic k-group* GL(q, D).** Let A be a central simple algebra of degree d over an algebraic number field k. Given a positive integer q, let $\mathrm{GL}(q, A)$ be the connected reductive algebraic k-group whose group $\mathrm{GL}(q, A)(l)$ of rational points over a commutative k-algebra l containing k equals the group

$$\mathrm{GL}_q(A_l) = \{x \in M_q(A_l) \mid \mathrm{nrd}_{M_q(A_l)}(x) \neq 0\}, \tag{2-1}$$

where $A_l = A \otimes_k l$, and $\mathrm{nrd}_{M_q(A_l)}$ is the reduced norm on the algebra $M_q(A_l)$ of $(q \times q)$-matrices with entries in A_l. If $q = 1$ then $\mathrm{GL}_1(A_l)$ is the group $A_l^\times$ of invertible elements in the l-algebra A_l. The reduced norm defines a surjective k-morphism $\mathrm{GL}(q, A) \to \mathbb{G}_m$ of k-groups, whose kernel is a connected semisimple algebraic k-group, to be denoted $\mathrm{SL}(q, A)$.

If $A = D$ is a central division k-algebra of degree d, that is, $\dim_k D = d^2$, then the connected reductive k-group $\mathrm{GL}(q, D)$ is of semisimple k-rank $q - 1$. Let l be a splitting field of D, thus, there is an isomorphism $\psi : D \otimes_k l \to M_d(l)$ of l-algebras. We fix this isomorphism ψ once and for all. We denote by the same letter the isomorphism

$$\psi : \mathrm{GL}(q, D) \times_k l \to \mathrm{GL}(qd, l), \tag{2-2}$$

of algebraic l-groups induced by ψ. The group $\mathrm{GL}(q, D)/k$ is a k-form of the general linear k-group $H := \mathrm{GL}(qd, k)/k$.

In the specific case of the connected reductive k-group $\mathrm{GL}(2, D)$, the group $Z'(k)$ of k-rational points of its center Z' is given by

$$Z'(k) = \{g = \mathrm{diag}(\lambda, \lambda) \mid \lambda \in k^\times 1_D\}$$

of scalar diagonal matrices. We fix a maximal k-split torus $S' \subset \mathrm{GL}(2, D)$ subject to

$$S'(k) = \left\{ g = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \;\middle|\; \lambda, \mu \in k^\times 1_D \right\}.$$

The centralizer $L' := Z_{\mathrm{GL}(2,D)}(S')$ of S' is given by

$$L'(k) = \left\{ g = \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \;\middle|\; x, y \in D^\times \right\}.$$

Note that L' is isomorphic to the k-group $\mathrm{GL}(1, D) \times \mathrm{GL}(1, D)$.

Let $\Phi'_k = \Phi(\mathrm{GL}(2, D), S') \subset X^*(S')$ be the set of roots of $\mathrm{GL}(2, D)$ with respect to S'. A basis of Φ'_k is given by the nontrivial character $\alpha : S'/k \to \mathbb{G}_m/k$, defined by the assignment $\left(\begin{smallmatrix} \lambda & 0 \\ 0 & \mu \end{smallmatrix}\right) \mapsto \lambda\mu^{-1}$. The corresponding minimal parabolic k-subgroup determined by $\{\alpha\}$ is denoted by Q'. Its Levi factor is $L_{Q'} = L'$, and we have a Levi decomposition of Q' into the semidirect product $L_{Q'}N_{Q'}$ of its unipotent radical $N_{Q'}$ by $L_{Q'}$.

2B. ***Splitting.*** Given a place $v \in V_k$, there exist a positive number r_v and a central division algebra Δ_v over k_v of degree $d_v \geq 1$ (uniquely determined up to isomorphism) so that $D \otimes_k k_v \cong M_{r_v}(\Delta_v)$ with $r_v d_v = d$. We say that a given central division algebra D over k splits at the place $v \in V_k$ if $D \otimes_k k_v \cong M_d(k_v)$. Let V_D be the finite set of places of k at which D does not split, that is, $d_v > 1$. Note that, if $v \in V_{k,\infty}$ is an archimedean place which is complex, then necessarily $d_v = 1$, that is, $\Delta_v = \mathbb{C}$. If there exists a real place $v \in V_D \cap V_{k,\infty}$, then Δ_v is isomorphic to the Hamilton quaternion algebra $\mathbb{H}$, hence $d_v = 2$, and, by $r_v d_v = d$, we get that d is even in this case.

2C. ***Parabolic k-subgroups and Levi subgroups in* $\mathbf{GL}(\boldsymbol{n}, \boldsymbol{k})$.** Let Q_0 denote the minimal parabolic k-subgroup of $\mathrm{GL}(n, k)$, $n \geq 1$, consisting of upper triangular nonsingular matrices, and let $Q_0 = L_0 N_0$ be its Levi decomposition where L_0 denotes the maximal torus of diagonal matrices and N_0 denotes the unipotent radical of Q_0. Let Φ, Φ^+, Δ denote the corresponding sets of roots, positive roots, simple roots, respectively. The set Δ is given as $\Delta = \{\alpha_1, \alpha_2, \ldots, \alpha_{n-1}\}$ where α_i denotes the usual projection $L_0 \to k^\times$ given by the assignment $\mathrm{diag}(t_1, \ldots, t_n) \mapsto t_i/t_{i+1}$. The conjugacy classes with respect to $\mathrm{GL}(n, k)$ in the set $\mathcal{P}(\mathrm{GL}(n))$ of parabolic k-subgroups are in one-to-one correspondence with the subsets of Δ. The class corresponding to $J \subset \Delta$ is the class represented by the standard parabolic subgroup Q_J. We define $S_J = \left(\bigcap_{\alpha \in J} \ker \alpha\right)^\circ$, and we write $L_{Q_J} := Z_{\mathrm{GL}(n)}(S_J)$ for its centralizer. The group L_{Q_J} is reductive, a so-called Levi subgroup of Q_J, and Q_J is the semidirect product of its unipotent radical N_{Q_J} by L_{Q_J}.

We use the following description: Let $\rho = (r_1, \ldots, r_s)$ be an ordered partition of n into positive integers, i.e., an ordered sequence of positive integers so that $r_1 + \cdots + r_s = n$. The corresponding standard parabolic subgroup Q_ρ consists of all matrices in $\mathrm{GL}(n, k)$ admitting a block decomposition in the form $(p_{i,j})$ with $p_{i,j}$

an $(r_i \times r_j)$-matrix, and $p_{i,j} = 0$ for $i > j$. Every parabolic subgroup of $\mathrm{GL}(n,k)$ is conjugate to a subgroup of this type. More precisely, Q_ρ is of type

$$J_\rho = \Delta \setminus \{\alpha_{r_1+\ldots+r_i} : i = 1, \ldots, n-1\},$$

and the assignment $\rho \mapsto J_\rho$ defines a bijection between ordered partitions of n and subsets of Δ. The standard Levi subgroup L_{Q_ρ} of Q_ρ is the subgroup of matrices in Q_ρ where each block above the block diagonal is zero, i.e., $p_{i,j} = 0$ for $i < j$. Thus, there is an isomorphism $L_{Q_\rho} \cong \mathrm{GL}(r_1) \times \cdots \times \mathrm{GL}(r_s)$. By definition, a cuspidal parabolic subgroup corresponds up to conjugacy to the case where $r_i = 1$ or 2 for $i = 1, \ldots, s$.

3. The global Jacquet–Langlands correspondence

3A. ***The global correspondence.*** Let k be a totally real number field of degree ℓ, and let D be a central division algebra over k of degree $d > 1$. Let $V_D \subset V_k$ be the finite set of places of k at which D does not split. Let t denote the number of archimedean places in V_D. Denote by H' the connected reductive algebraic k-group $\mathrm{GL}(2, D)$. This group is of semisimple k-rank one; it is an inner form of the algebraic k-group $H := \mathrm{GL}(2d, k)$. Let Z denote the center of one of the two groups H/k or H'/k. In both cases the locally compact group $Z(\mathbb{A}_k)$ is isomorphic to the group of ideles $\mathbb{I}_k$. The isomorphism is provided by assigning to an element $a \in \mathbb{I}_k$ the scalar matrix of the appropriate size with a on the diagonal. Thus, we may view a unitary character of $Z(k)\backslash Z(\mathbb{A}_k)$ as a unitary character of $k^\times \backslash \mathbb{I}_k$. We fix such a character ω.

The global Jacquet–Langlands correspondence due to Badulescu [2008] and Badulescu and Renard [2010] relates the discrete spectrum of $H(\mathbb{A}_k) = \mathrm{GL}(2d, \mathbb{A}_k)$ and the discrete spectrum of $H'(\mathbb{A}_k) = \mathrm{GL}(2, D)(\mathbb{A}_k)$. The definition uses the local Jacquet–Langlands correspondence (see [Badulescu 2008, Section 3]). It is defined for Harish-Chandra modules at infinite places $v \in V_{k,\infty}$, and smooth representations at finite places $v \in V_{k,f}$. Hence, one should have in mind, when dealing with irreducible constituents of the discrete spectrum, that we actually pass to the underlying $(\mathfrak{g}, K_\mathbb{R}; G(\mathbb{A}_f))$-module without mentioning that explicitly. Note that, by definition, the adèle group $H'(\mathbb{A}_k)$ of the group H'/k is the restricted product $H'(\mathbb{A}_k) = \prod'_{v \in V_k} H'(k_v)$ with respect to the maximal compact subgroups $H'(\mathcal{O}_{k,v}) \subset H'(k_v)$, for almost all $v \in V_{k,f}$. If $v \in V_D \cap V_{k,\infty}$, then $H'(k_v) \cong \mathrm{GL}(d, \mathbb{H})$. If $v \in V_{k,\infty}$, $v \notin V_D$, then $H'(k_v) \cong \mathrm{GL}(2d, \mathbb{R})$.

To be more precise, as in [Badulescu 2008, 5.1], we have the following

Definition 3.1. We say that an irreducible constituent of $L^2_{\mathrm{disc}}(H, \omega)$ is (globally) compatible with respect to D if every local component π_v of π at a place $v \in V_D$ is locally compatible as a unitary representation of $H(k_v) \cong \mathrm{GL}_{2d}(k_v)$, i.e., there is a unitary representation π'_v of $H'(k_v) \cong \mathrm{GL}_2(D_v)$ corresponding to π_v by the local

Jacquet–Langlands correspondence (see [Badulescu and Renard 2010, Section 13], [Badulescu 2008, Section 3]).

In our case at hand, the main result regarding the Jacquet–Langlands correspondence is as follows (see [Badulescu 2008, Theorem 5.1]):

Theorem 3.2. *There is a unique map, to be denoted Ξ, from the set of irreducible constituents of $L^2_{\mathrm{disc}}(H', \omega)$ to the set of irreducible constituents of $L^2_{\mathrm{disc}}(H, \omega)$, such that if $\pi = \Xi(\pi')$, with $\pi = \otimes_{v \in V_k} \pi_v$ and $\pi' = \otimes_{v \in V_k} \pi'_v$, then*

- *π is compatible (with respect to D),*
- *$\pi_v \cong \pi'_v$ for $v \notin V_D$,*
- *π_v corresponds to π'_v by the local Jacquet–Langlands correspondence at $v \in V_D$.*

The map Ξ is injective, and the image of Ξ consists of all compatible constituents of $L^2_{\mathrm{disc}}(H, \omega)$ with respect to D.

3B. ***The classical correspondence.*** Suppose that B is a central division algebra of degree $d = 2$ over k, that is, B is a quaternion division k-algebra. Note that all irreducible automorphic representations of $B^\times_{\mathbb{A}_k}$ are cuspidal. The original correspondence due to Jacquet and Langlands [1970] is a bijection between (cuspidal) automorphic representations of $B^\times_{\mathbb{A}_k}$ which are not one-dimensional and cuspidal automorphic representations of $\mathrm{GL}(2, \mathbb{A}_k)$ with square-integrable local component at each place where B does not split, such that if $\pi' \cong \otimes_v \pi'_v$ corresponds to $\pi \cong \otimes_v \pi_v$, then $\pi'_v \cong \pi_v$ at $v \notin V_B$, and π'_v corresponds to π_v by the local Jacquet–Langlands correspondence at $v \in V_B$. This is extended in [Badulescu 2008; Badulescu and Renard 2010] to an injective map Ξ, analogous to the one described in Theorem 3.2. In particular, Ξ maps a one-dimensional representation, given by $\chi \circ \mathrm{nrd}$ with χ a unitary character of $k^\times \backslash \mathbb{I}_k$, to $\chi \circ \det$.

Remark 3.3. We refer to Section 4A where one finds a description of the local Jacquet–Langlands correspondence between $\mathrm{GL}(2, \mathbb{R})$ and $\mathbb{H}^\times$ in the specific case of unitary square-integrable representations δ of $\mathrm{GL}(2, \mathbb{R})$.

3C. ***The residual spectrum of $H(\mathbb{A}_k) = \mathrm{GL}(2d, \mathbb{A}_k)$.*** The discrete spectrum

$$L^2_{\mathrm{disc}}(H, \omega)$$

of $H(\mathbb{A}_k)$ with respect to ω decomposes into a direct Hilbert space sum

$$L^2_{\mathrm{disc}}(H, \omega) \cong L^2_{\mathrm{cusp}}(H, \omega) \oplus L^2_{\mathrm{res}}(H, \omega) \tag{3-1}$$

of the cuspidal spectrum and the residual spectrum of $H(\mathbb{A}_k)$. The cuspidal spectrum $L^2_{\mathrm{cusp}}(H, \omega)$ is the direct Hilbert space sum of irreducible cuspidal automorphic representations of $H(\mathbb{A}_k)$ with central character ω, each appearing with multiplicity one (see [Shalika 1974]). By the work of Moeglin and Waldspurger [1989], the residual spectrum of $H(\mathbb{A}_k)$ decomposes along the cuspidal support into

$$L^2_{\mathrm{res}}(H,\omega) \cong \oplus_\rho L^2_{\mathrm{res},\{Q_\rho\}}(H,\omega), \tag{3-2}$$

where the sum ranges over the associate classes of all proper k-parabolic subgroups Q_ρ corresponding to a partition $\rho = (r_1, \dots, r_s)$ subject to the condition $r_1 = \dots = r_s$. Let us denote this value by r. Thus, the Levi subgroup L_{Q_ρ} is a direct product of l copies of $\mathrm{GL}(r)$ with $r \cdot s = 2d$. The summand corresponding to the associate class $\{Q_\rho\}$ of k-parabolic subgroups has the following structure: it is given by the sum $L^2_{\mathrm{res},\{Q_\rho\}}(H,\omega) \cong \oplus_\sigma J(s,\sigma)$, where $J(s,\sigma)$ denotes the unique irreducible quotient of an induced representation[1]

$$\sigma|\det|^{(s-1)/2} \times \sigma|\det|^{(s-3)/2} \times \cdots \times \sigma|\det|^{-(s-1)/2} \tag{3-3}$$

with σ an irreducible cuspidal representation of $\mathrm{GL}(r, \mathbb{A}_k)$ whose central character ω_σ equals ω. Note that there is exactly one associate class of maximal parabolic k-subgroups of $\mathrm{GL}(2d)/k$ which can contribute to the decomposition (3-2) of the residual spectrum $L^2_{\mathrm{res}}(H,\omega)$. It is the class of the maximal parabolic k-subgroup Q_d whose Levi subgroup is isomorphic to $\mathrm{GL}(d) \times \mathrm{GL}(d)$. In this case, we have

$$L^2_{\mathrm{res},\{Q_d\}}(H,\omega) \cong \oplus_\sigma J(2,\sigma), \tag{3-4}$$

where σ ranges over the irreducible cuspidal representation of $\mathrm{GL}(d, \mathbb{A}_k)$ with central character $\omega_\sigma = \omega$.

We denote by Q' the minimal parabolic k-subgroup of upper triangular matrices in $\mathrm{GL}(2, D)$. Its Levi subgroup is isomorphic to the k-group $\mathrm{GL}(1, D) \times \mathrm{GL}(1, D)$, and we have a Levi decomposition of Q' into the semidirect product $L'N'$ of its unipotent radical N' by L'. The image of the l-group $Q' \times_k l$ under the map ψ as given in (2-2) is the maximal parabolic l-subgroup $Q_d = Q_{\Delta\setminus\{\alpha_d\}}$. Its Levi subgroup L_{Q_d}/l is isomorphic to $\mathrm{GL}(d)/l \times \mathrm{GL}(d)/l$.

The following result, which concerns one summand in the decomposition of $L^2_{\mathrm{res},\{Q_d\}}(H,\omega)$, is a consequence of the general work of Badulescu and Renard regarding the global Jacquet–Langlands correspondence. By [Badulescu 2008; Badulescu and Renard 2010, Proposition 18.2] we have:

Theorem 3.4. *Suppose that the central division algebra D over k is of even degree, say $d = 2h$. If $\pi \cong J(2,\sigma)$, where $\sigma = \otimes_{v \in V_k} \sigma_v$ is a cuspidal automorphic representation of $\mathrm{GL}(d, \mathbb{A}_k)$, is a summand of $L^2_{\mathrm{res},\{Q_d\}}(H,\omega)$, then π is always*

[1] Here we use the standard notation: Given a partition $(m_1, \dots, m_s)$ of the natural number $m \geq 1$ and given for each m_i, $i = 1, \dots, s$, an automorphic representation π_i of $\mathrm{GL}(m_i, \mathbb{A}_k)$, we denote by $\pi_1 \times \cdots \times \pi_s$ the automorphic representation obtained by parabolic induction from $\pi_1 \otimes \cdots \otimes \pi_s$ on the Levi subgroup L_{P_ρ} of the parabolic subgroup P_ρ attached to the ordered partition $(m_1, \dots, m_s)$ in $\mathrm{GL}(m)$.

compatible with respect to D, that is, $\pi \cong J(2,\sigma)$ occurs in the image of Ξ. One has to distinguish the following two cases in the correspondence via Ξ:

(1) *If there is a place $v_0 \in V_D$ such that σ_{v_0} is not square-integrable, then π corresponds to a cuspidal automorphic representation π' of $H'(\mathbb{A}_k)$.*

(2) *If σ_v is square-integrable at all nonsplit places $v \in V_D$, let σ' be the cuspidal automorphic representation of $D^\times_{\mathbb{A}_k}$ corresponding to σ by the Jacquet–Langlands correspondence. Note that σ' is not one-dimensional. Then π corresponds to the residual representation $J'(2,\sigma')$ of $H'(\mathbb{A}_k)$ which is constructed in analogy to $J(2,\sigma)$ and occurs in the residual spectrum $L^2_{\mathrm{res}}(H',\omega)$.*

4. Some cohomological representations of $\mathbf{GL}(d,\mathbb{H})$ and $\mathbf{GL}(2d,\mathbb{R})$

The investigation of the global injective map Ξ from the set of irreducible constituents of $L^2_{\mathrm{disc}}(H',\omega)$ to the set of irreducible constituents of $L^2_{\mathrm{disc}}(H,\omega)$ involves, if $\pi = \Xi(\pi')$, with $\pi = \otimes_{v\in V_k}\pi_v$ and $\pi' = \otimes_{v\in V_k}\pi'_v$, a precise knowledge of the local Jacquet–Langlands correspondence between π_v and π'_v at places $v \in V_D$. Since we aim to construct global automorphic representations of $H'(\mathbb{A}_k)$ which are of cohomological relevance, this question, in particular, concerns the archimedean places $v \in V_D \cap V_{k,\infty}$. We are interested in those representations π so that the local group in question, say $H(k_v)$ (resp. $H'(k_v)$), $v \in V_{k,\infty}$, has nontrivial continuous cohomology with coefficients in $\pi_v \otimes \mathbb{C}$ (resp. $\pi'_v \otimes \mathbb{C}$). The case of the groups $\mathrm{GL}(1,\mathbb{H}) \cong \mathbb{H}^\times$ and $\mathrm{GL}(2,\mathbb{R})$ provides the basic ingredients in dealing with the general case. Results of Vogan and Zuckerman concerning the general classification of irreducible unitary representations of a real reductive Lie group with nonzero continuous cohomology as established in [Vogan and Zuckerman 1984] (resp. [Vogan 1997]) are fundamental in this study.

4A. ***Discrete series representation of* GL(2, $\mathbb{R}$) *and the local correspondence.*** Let $V(r), r \geq 2$, denote the irreducible two-dimensional representation of the orthogonal group $O(2)$ which is fully induced by the character $k_\theta \mapsto e^{ir\theta}$ of the subgroup $\mathrm{SO}(2)$ of rotations k_θ, $\theta \in [0,2\pi]$, in $O(2)$ of index two. Given an integer $m \geq 2$, we denote by D_m the discrete series representation of $\mathrm{GL}(2,\mathbb{R})$ of lowest $O(2)$-type m. The representation D_m is square-integrable and characterized by the fact that its restriction to the maximal compact subgroup $O(2)$ of $\mathrm{GL}(2,\mathbb{R})$ decomposes as an algebraic sum of the form

$$D_m|_{O(2)} \cong \oplus_{r\in\Sigma(m)} V(r), \qquad \Sigma(m) = \{l \in \mathbb{Z} \mid l \equiv m \bmod 2,\, l \geq m\}.$$

In this labeling of the discrete series representations of $\mathrm{GL}(2,\mathbb{R})$ the Harish-Chandra parameter of D_m, $m \geq 2$, is $m-1$.

The local Jacquet–Langlands correspondence between $\mathrm{GL}(1,\mathbb{H}) = \mathbb{H}^\times$ and $\mathrm{GL}(2,\mathbb{R})$ is as follows: Let δ be a unitary square-integrable representation of

$\mathrm{GL}(2, \mathbb{R})$, and let χ be a unitary character of $\mathbb{R}^\times$. If $\delta = D_2(\chi \circ \det_2)$ is of lowest $O(2)$-type 2, then it corresponds to the character $\chi \circ \mathrm{nrd}_1$ of $\mathbb{H}^\times$. Observe that D_2 corresponds to the trivial character of $\mathbb{H}^\times$, denoted by $\mathbf{1}_{\mathbb{H}^\times}$. Next, if $\delta = D_m(\chi \circ \det_2)$ is of lowest $O(2)$-type $m > 2$, then it corresponds to $\delta' = D'_m(\chi \circ \mathrm{nrd}_1)$, where D'_m is the representation of $\mathbb{H}^\times$ which corresponds to D_m, and nrd_1 denotes the reduced norm on $\mathbb{H}^\times$. The representation δ' is not one-dimensional.

4B. ***Nonvanishing continuous cohomology for D_m.*** Let (σ_k, F_k), $k \geq 0$, be the irreducible finite-dimensional representation of $\mathrm{GL}(2, \mathbb{R})$ of highest weight $\mu_k = k \cdot \omega$ (where ω denotes the fundamental dominant weight of $\mathrm{GL}(2, \mathbb{R})$), thus, $\dim F_k = k + 1$. The continuous cohomology $H^*_{\mathrm{ct}}(\mathrm{GL}(2, \mathbb{R}), D_m \otimes F_k)$ vanishes if $k \neq m - 2$ since the infinitesimal character χ_{D_m} differs from the one of the contragredient representation of (σ_k, F_k). In the case $k = m - 2$ one has

$$H^q_{\mathrm{ct}}(\mathrm{GL}(2, \mathbb{R}), D_m \otimes F_{m-2}) = \mathbb{C}$$

for $q = 1$; it vanishes otherwise.

4C. ***The classification of Vogan and Zuckerman*: *the general case.*** It is necessary, mainly to fix notation, to recall some results of Vogan and Zuckerman concerning the general classification of irreducible unitary representations of a connected real reductive Lie group with nonzero continuous cohomology as established in [Vogan and Zuckerman 1984; Vogan 1997].[2] This constructive approach is algebraic in nature. We fix a maximal compact subgroup $K \subset G$, denote by $X = X_G$ the associated symmetric space, and write θ_K for the corresponding Cartan involution. Write $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$ for the corresponding Cartan decomposition of the Lie algebra $\mathfrak{g}$ of G. By definition, a θ_K-stable parabolic subalgebra of $\mathfrak{g}$ is a parabolic subalgebra $\mathfrak{q} \subset \mathfrak{g}_\mathbb{C}$ such that $\theta_K \mathfrak{q} = \mathfrak{q}$, and $\overline{\mathfrak{q}} \cap \mathfrak{q} = \mathfrak{l}_\mathbb{C}$ is a Levi subalgebra of $\mathfrak{q}$ where the bar refers to complex conjugation with regard to the real form $\mathfrak{g}$ of $\mathfrak{g}_\mathbb{C}$. The Levi subalgebra $\mathfrak{l}_\mathbb{C}$ is necessarily defined over $\mathbb{R}$, and the real subalgebra $\mathfrak{l}$ is stable under the Cartan involution. We define the Levi subgroup L attached to $\mathfrak{q}$ by

$$L = \{g \in G \mid \mathrm{Ad}(g)(\mathfrak{q}) \subset \mathfrak{q}\}. \tag{4-1}$$

It is a connected real reductive group of the same rank as G. The Cartan involution θ_K preserves L, and the restriction $\theta_{K|L}$ to L is a Cartan involution of L. The fact that L contains a maximal torus $T \subset K$ is essential in the classification of θ-stable parabolic subalgebras of $\mathfrak{g}$ up to conjugation by K. Given a θ_K-stable parabolic subalgebra $\mathfrak{q}$ of $\mathfrak{g}$ with Levi subgroup L, write $\mathfrak{u}$ for the nil radical of $\mathfrak{q}$, and $R(\mathfrak{q}) := \dim(\mathfrak{u} \cap \mathfrak{p}_\mathbb{C})$. Attached to $\mathfrak{q}$ there is an irreducible unitary representation $\pi_\mathfrak{q}$ of G. Up to infinitesimal equivalence, $\pi_\mathfrak{q}$ depends only on the K-conjugacy class

[2]If G is nonconnected but still in Harish-Chandra's class the notation is slightly more complicated but there arise no essential new difficulties.

of $\mathfrak{q}$. Notice that there are only finitely many K-conjugacy classes of θ_K-stable parabolic subalgebras of $\mathfrak{g}$. If we write $A_{\mathfrak{q}}$ for the Harish-Chandra module of $\pi_{\mathfrak{q}}$, then the continuous cohomology of G with coefficients in $\pi_{\mathfrak{q}}$ coincides with the relative Lie algebra cohomology with respect to $A_{\mathfrak{q}}$, and we have, using [Vogan and Zuckerman 1984, Theorem 3.3],

$$H^p_{\mathrm{ct}}(G, \pi_{\mathfrak{q}} \otimes \mathbb{C}) \cong H^p(\mathfrak{g}, K; A_{\mathfrak{q}}) \cong H^{p-R(\mathfrak{q})}(\mathfrak{l}, L \cap K; \mathbb{C}). \tag{4-2}$$

The right-hand side is isomorphic to $\mathrm{Hom}_{\mathfrak{l}\cap\mathfrak{k}}(\Lambda^{p-R(\mathfrak{q})}(\mathfrak{l}\cap\mathfrak{p}), \mathbb{C})$ so that we get

$$H^p_{\mathrm{ct}}(G, \pi_{\mathfrak{q}} \otimes \mathbb{C}) \cong \mathrm{Hom}_{\mathfrak{l}\cap\mathfrak{k}}(\Lambda^{p-R(\mathfrak{q})}(\mathfrak{l}\cap\mathfrak{p}), \mathbb{C}). \tag{4-3}$$

Thus, the cohomology group $H^*_{\mathrm{ct}}(G, \pi_{\mathfrak{q}} \otimes \mathbb{C})$ vanishes in degrees below $R(\mathfrak{q})$ and above $R(\mathfrak{q}) + \dim(\mathfrak{l}\cap\mathfrak{p})$. Now interpret the right-hand side of (4-3) in the following way: Let L_u be the compact form of the real Levi subgroup L, and let $X_{L,u}$ be the compact dual of the space $L/(K \cap L)$. Then we have (see, for example, [Schwermer 2010, Section 7.1])

$$\mathrm{Hom}_{\mathfrak{l}\cap\mathfrak{k}}(\Lambda^{p-R(\mathfrak{q})}(\mathfrak{l}\cap\mathfrak{p}), \mathbb{C}) \cong H^{p-R(\mathfrak{q})}(L^0_u/(L_u \cap K)^0, \mathbb{C}). \tag{4-4}$$

By Poincaré duality, we obtain $R(\mathfrak{q}) = \left(\frac{1}{2}\right)(\dim X - \dim X_{L,u})$.

We denote by $P(\pi_{\mathfrak{q}}, t)$ the Poincaré polynomial of the cohomology space $H^*_{\mathrm{ct}}(G, \pi_{\mathfrak{q}} \otimes \mathbb{C})$. Then, by the preceding argument, we obtain the formula

$$P(\pi_{\mathfrak{q}}, t) = t^{R(\mathfrak{q})} P(X_{L,u}, t), \tag{4-5}$$

where $P(X_{L,u}, t)$ denotes the Poincaré polynomial of the compact dual $X_{L,u}$ of the space $L/(K \cap L)$.

Suppose (π, H_π) is an irreducible unitary representation of G so that the continuous cohomology of G with coefficients in (π, H_π) does not vanish. Then, by [Vogan and Zuckerman 1984, Theorem 4.1], there is a θ-stable parabolic subalgebra $\mathfrak{q}$ of $\mathfrak{g}$ so that $\pi_{\mathfrak{q}} \cong \pi$, and thus also for the corresponding Harish-Chandra module $H_{\pi,K} \cong A_{\mathfrak{q}}$. Given a θ-stable parabolic subalgebra $\mathfrak{q}$, the corresponding irreducible unitary representation $\pi_{\mathfrak{q}}$ is a discrete series representation if and only if $\mathfrak{l} \subset \mathfrak{k}$. It is a fundamental series representation if and only if $[\mathfrak{l}, \mathfrak{l}] \subset \mathfrak{k}$. If $[\mathfrak{l}, \mathfrak{l}]$ is not contained in $\mathfrak{k}$ then $\pi_{\mathfrak{q}}$ is not tempered (see [Vogan and Zuckerman 1984, p. 58]). If the θ-stable parabolic subalgebra $\mathfrak{q}$ coincides with $\mathfrak{g}$, then $L = G$, hence the corresponding representation $\pi_{\mathfrak{q}}$ is the trivial representation.

Let $\mathfrak{q}$ be a θ-stable parabolic subalgebra of $\mathfrak{g}$ and let $\pi_{\mathfrak{q}}$ be the corresponding unique irreducible representation of G so that the continuous cohomology of G with coefficients in $\pi_{\mathfrak{q}}$ is nonzero. In the cases of interest for us, it is necessary to determine how these representations fit into the Langlands classification (see [Langlands 1989]) of irreducible admissible representations of G. Fundamentally, the idea behind the classification is to inductively parametrize the irreducible admissible representations of G in terms of irreducible tempered representations of Levi subgroups L of G.

Thus, given a θ-stable parabolic subalgebra of $\mathfrak{g}$, we have to describe the corresponding so-called Langlands quotient, characterized by its uniquely determined data (P, σ, ν), namely, a (standard) parabolic subgroup P of G with decomposition $P = MA_PN$, σ an irreducible tempered representation of M, and $\nu \in \mathfrak{a}_P^*$ such that $\langle \operatorname{Re} \nu, \alpha\rangle > 0$ for all roots α in $\mathfrak{n}$. The final general result, with regard to the choice of P obtained in a process of two steps, is described in [Vogan and Zuckerman 1984, Theorem 6.16]. In the case $\mathrm{GL}(2r, \mathbb{R})$, one can partially read it off from [Speh 1983, Section 4].

4D. ***The classification of Vogan–Zuckerman*: *the cases* $\mathbf{GL(2r, \mathbb{R})}$ *and* $\mathbf{GL(r, \mathbb{H})}$.** Firstly, we consider the case of the nonconnected real reductive group $G = \mathrm{GL}(2r, \mathbb{R})$, $r \geq 1$, $K = O(2r)$, and $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ the Cartan decomposition which corresponds to θ_K. Let $m_0 \geq 0$ be an integer, and let $m_1, \ldots, m_s$ be positive integers with $r = m_0 + m_1 + \ldots + m_s$. Note, $m_0 = r$ in the case $s = 0$. There corresponds a θ_K-stable parabolic subalgebra $\mathfrak{q} = \mathfrak{l} \oplus \mathfrak{u}$ whose corresponding real Levi subalgebra is

$$\mathfrak{l} = \mathfrak{gl}(2m_0, \mathbb{R}) \oplus \mathfrak{gl}(m_1, \mathbb{C}) \oplus \ldots \oplus \mathfrak{gl}(m_s, \mathbb{C}). \tag{4-6}$$

Thus, the possible corresponding Levi subgroups L are

$$L = \mathrm{GL}(2m_0, \mathbb{R}) \times \mathrm{GL}(m_1, \mathbb{C}) \times \ldots \times \mathrm{GL}(m_s, \mathbb{C}). \tag{4-7}$$

Secondly, let G be the connected real reductive group $\mathrm{GL}(r, \mathbb{H})$, $K = \mathrm{Sp}(r)$. Let $n_0 \geq 0$ be an integer, and let $n_1, \ldots, n_s$ be positive integers with $r = n_0 + n_1 + \ldots + n_s$. Note, $n_0 = r$ in the case $s = 0$. There corresponds a θ_K-stable parabolic subalgebra $\mathfrak{q} = \mathfrak{l} \oplus \mathfrak{u}$ of $\mathfrak{g}$ whose corresponding real Levi subalgebra is

$$\mathfrak{l} = \mathfrak{gl}(n_0, \mathbb{H}) \oplus \mathfrak{gl}(n_1, \mathbb{C}) \oplus \ldots \oplus \mathfrak{gl}(n_s, \mathbb{C}). \tag{4-8}$$

Thus, the possible corresponding Levi subgroups L are

$$L = \mathrm{GL}(n_0, \mathbb{H}) \times \mathrm{GL}(n_1, \mathbb{C}) \times \ldots \times \mathrm{GL}(n_s, \mathbb{C}). \tag{4-9}$$

This result is based on an explicit constructive procedure similar to the one carried through in the analogous case of the Lie group $\mathrm{SL}(r, \mathbb{H})$, also denoted by $\mathrm{SU}^*(2r)$, in [Schwermer and Waldner 2011].

4E. ***Tempered cohomological representations.*** Suppose that n is an even positive integer, say, $n = 2r$. Within the family of irreducible unitary tempered representations of the real Lie group $\mathrm{GL}(2r, \mathbb{R})$ there is exactly one representation (θ, H_θ) (up to infinitesimal equivalence) so that the continuous cohomology $H^*_{\mathrm{ct}}(\mathrm{GL}(2r, \mathbb{R}), H_\theta \otimes \mathbb{C})$ of $\mathrm{GL}(2r, \mathbb{R})$ with coefficients in $\theta \otimes \mathbb{C}$ is nonzero. This representation can be described in the following way (see [Schwermer 1986, Section 3]).

Let P_{δ_n} be the cuspidal parabolic subgroup of $\mathrm{GL}(n, \mathbb{R})$ given by the partition $\delta_n = (2, \dots, 2)$ of n. Its Levi subgroup $L_{P_{\delta_n}}$ is isomorphic to $\mathrm{GL}(2, \mathbb{R}) \times \cdots \times \mathrm{GL}(2, \mathbb{R})$. Consider the representation $\tau = \otimes \tau_i,\ i = 1, \dots, r$, of $L_{P_{\delta_n}}$ whose i-th component τ_i is a discrete series representation of $\mathrm{GL}(2, \mathbb{R})$ of lowest $O(2)$-type $2i,\ i = 1, \dots, r$. Then

$$\mathrm{Ind}(P_{\delta_n}, \tau) \cong D_2 \times D_4 \times \cdots D_n \tag{4-10}$$

is an irreducible unitary representation of $\mathrm{GL}(n, \mathbb{R})$, the unique one (up to infinitesimal equivalence) that is tempered and so that

$$H^*_{\mathrm{ct}}(\mathrm{GL}(n, \mathbb{R}), \mathrm{Ind}(P_{\delta_n}, \tau) \otimes \mathbb{C}) \neq \{0\}.$$

The continuous cohomology does not vanish in a range of length

$$\mathrm{rk}\, \mathrm{GL}(n, \mathbb{R}) - \mathrm{rk}\, O(n)$$

around the middle dimension of the underlying symmetric space (see [Borel and Wallach 1980, III, Proposition 5.3]).

For the sake of completeness we ascertain that the θ-stable parabolic subalgebra $\mathfrak{q}$ of $\mathfrak{gl}(2r, \mathbb{R})$ which corresponds to the representation $\mathrm{Ind}(P_{\delta_n}, \tau)$ is of the form such that the real Levi subalgebra is

$$\mathfrak{l} \cong \mathfrak{gl}(1, \mathbb{C}) \oplus \ldots \oplus \mathfrak{gl}(1, \mathbb{C}) \cong \mathfrak{gl}(1, \mathbb{C})^r. \tag{4-11}$$

Thus, its parameter is $(m_0; m_1, \dots, m_r) = (0; 1, \dots, 1)$.

We now determine a representation $(\theta', H_{\theta'})$ of $\mathrm{GL}(r, \mathbb{H})$ which corresponds under the local Jacquet–Langlands correspondence to the representation $(\theta, H_\theta) := \mathrm{Ind}(P_{\delta_n}, \tau)$. We denote by $P'_{\delta_r} = L'_{\delta_r} N'_{\delta_r}$ the standard minimal parabolic subgroup of $\mathrm{GL}(r, \mathbb{H})$ whose Levi subgroup consists of r copies of $\mathrm{GL}(1, \mathbb{H}) \cong \mathbb{H}^\times$. Then the representation

$$\mathrm{Ind}(P'_{\delta_r}, \tau') \cong \mathbf{1}_{\mathbb{H}^\times} \times D'_4 \times \cdots D'_{2r} \tag{4-12}$$

is an irreducible unitary representation of $\mathrm{GL}(r, \mathbb{H})$. In fact, $\theta' := \mathrm{Ind}(P'_{\delta_r}, \tau')$ is the only irreducible unitary representation of $\mathrm{GL}(r, \mathbb{H})$ which is tempered and so the continuous cohomology of $\mathrm{GL}(r, \mathbb{H})$ with coefficients in $\theta' \otimes \mathbb{C}$ is nonzero in a certain range.

Proposition 4.1. *Let $n = 2r$ be even. The irreducible tempered representation $\theta' := \mathrm{Ind}(P'_{\delta_r}, \tau')$ of $\mathrm{GL}(r, \mathbb{H})$ corresponds under the local Jacquet–Langlands correspondence to the irreducible tempered representation $(\theta, H_\theta) := \mathrm{Ind}(P_{\delta_n}, \tau)$ of $\mathrm{GL}(n, \mathbb{R})$. The continuous cohomology of $\mathrm{GL}(r, \mathbb{H})$ with coefficients in $\theta' \otimes \mathbb{C}$ is nonzero. More precisely, the continuous cohomology does not vanish in a range of length $\mathrm{rk}\, \mathrm{GL}(r, \mathbb{H}) - \mathrm{rk}\, \mathrm{Sp}(r)$ around the middle dimension of the underlying symmetric space.*

Proof. Since the local Jacquet–Langlands correspondence at real archimedean places (see [Badulescu and Renard 2010, Section 13]) commutes with parabolic induction and the process of forming tensor products of representations, the assertion is an immediate consequence of the construction of both representations where the building blocks match under the correspondence.

We relate the representation $\theta' = \mathrm{Ind}(P'_{\delta_r}, \tau')$ of $\mathrm{GL}(r, \mathbb{H})$ to the corresponding data within the classification of irreducible unitary representations of the Lie group $\mathrm{GL}(r, \mathbb{H})$ with nonzero continuous cohomology as described in Section 4C. The representation θ' is equivalent to the representation $\pi_{\mathfrak{q}'}$ which corresponds to the θ_K-stable parabolic algebra $\mathfrak{q}'$ of the Lie algebra $\mathfrak{g}'$ of $\mathrm{GL}(r, \mathbb{H})$ so that the real Levi subalgebra $\mathfrak{l}'$ of $\mathfrak{q}'$ is given by

$$\mathfrak{l}' \cong \mathfrak{gl}(1, \mathbb{C})^r. \tag{4-13}$$

Since $[\mathfrak{l}', \mathfrak{l}'] = \{0\} \subset \mathfrak{k}'$, it follows that θ' is tempered. The basis for the coincidence of the form of $\mathfrak{l}'$ in this case with the form of the real Levi subalgebra as given in formula (4-11), and hence finally for the Jacquet–Langlands correspondence, is the fact that the Lie algebras $\mathfrak{g}_{\mathbb{C}}$ and $\mathfrak{g}'_{\mathbb{C}}$ attached to the two groups $\mathrm{GL}(2r, \mathbb{R})$ and $\mathrm{GL}(r, \mathbb{H})$ share Levi subalgebras of θ-stable parabolic subalgebras which are products of $\mathfrak{gl}(m_i, \mathbb{C})$. □

4F. ***A specific nontempered representation of* GL(*d*, ℍ).** Given our global context, that is, a central division algebra D over k of even degree, say $d = 2h$, the representation (θ, H_θ) of $\mathrm{GL}(d, \mathbb{R})$ as well as the representation $(\theta', H_{\theta'})$ of $\mathrm{GL}(h, \mathbb{H})$ give rise to two other representations which are decisive in our construction of nonvanishing square integrable cohomology classes for the group $\mathrm{GL}(2, D)$.

Let $J_{\mathbb{R}}(2, \theta)$ with $\theta = \mathrm{Ind}(P_{\delta_d}, \tau)$ as defined in Section 4E denote the unique irreducible quotient of the induced representation of $\mathrm{GL}(2d, \mathbb{R})$ of the form

$$\theta|\det|^{1/2} \times \theta|\det|^{-1/2}. \tag{4-14}$$

Under the local Jacquet–Langlands correspondence, this representation $J_{\mathbb{R}}(2, \theta)$ of $\mathrm{GL}(2d, \mathbb{R})$ corresponds to the analogous representation $J'_{\mathbb{R}}(2, \theta')$ with $\theta' = \mathrm{Ind}(P'_{\delta_h}, \tau')$ as defined in Section 4E for $\mathrm{GL}(d, \mathbb{H})$, given as the unique irreducible quotient $J'_{\mathbb{R}}(2, \theta')$ of the induced representation of the form

$$\theta' \,\mathrm{nrd}^{1/2} \times \theta' \,\mathrm{nrd}^{-1/2}. \tag{4-15}$$

Proposition 4.2. *The irreducible unitary representation $J'_{\mathbb{R}}(2, \theta')$ of* $\mathrm{GL}(d, \mathbb{H})$, *$d$ even, say $d = 2h$, is a nontempered representation. It corresponds under the local Jacquet–Langlands correspondence to the irreducible nontempered representation $J_{\mathbb{R}}(2, \theta)$ of* $\mathrm{GL}(2d, \mathbb{R})$, *and the continuous cohomology of* $\mathrm{GL}(d, \mathbb{H})$ *with coefficients in $J'_{\mathbb{R}}(2, \theta') \otimes \mathbb{C}$ is nonzero. The Poincaré polynomial of the* *representation* $J'_{\mathbb{R}}(2,\theta')$ *of* $\mathrm{GL}(d,\mathbb{H})$ *has the form*

$$P(J'_{\mathbb{R}}(2,\theta'),t)=t^{\frac{1}{2}d(2d-3)}\cdot\prod_{s=1}^{d/2}\prod_{i=1}^{m_s}(1+t^{2i-1}). \tag{4-16}$$

with $m_i=2,\ i=1,\dots,d/2$. *The lowest degree* p *in which the continuous cohomology* $H^*_{\mathrm{ct}}(\mathrm{GL}(d,\mathbb{H}),J'_{\mathbb{R}}(2,\theta'))$ *of* $\mathrm{GL}(d,\mathbb{H})$ *with coefficients in* $J'_{\mathbb{R}}(2,\theta')$ *does not vanish is* $\frac{1}{2}d(2d-3)$.

Proof. We describe the representation $J'_{\mathbb{R}}(2,\theta')$ of $\mathrm{GL}(d,\mathbb{H})$ in terms of the classification of irreducible unitary representations of $\mathrm{GL}(d,\mathbb{H})$ with nonzero continuous cohomology as established in [Vogan and Zuckerman 1984] and [Vogan 1997] (see Section 4D above). Thus, given the irreducible unitary representation $J'_{\mathbb{R}}(2,\theta')$ of $\mathrm{GL}(d,\mathbb{H})$, we proceed as follows to obtain the corresponding algebraic data in this framework. Since, in the given case, we already know the Langlands data by construction, going backwards in the line of arguments in [Vogan and Zuckerman 1984, Section 6]) we can identify the corresponding θ_K-stable parabolic subalgebra $\mathfrak{q}$ in $\mathfrak{g}$: the real Levi subalgebra $\mathfrak{l}$ of $\mathfrak{q}$ turns out to be

$$\mathfrak{l}\cong\mathfrak{gl}(2,\mathbb{C})\oplus\mathfrak{gl}(2,\mathbb{C})\oplus\ldots\oplus\mathfrak{gl}(2,\mathbb{C})\cong\mathfrak{gl}(2,\mathbb{C})^h. \tag{4-17}$$

By (4-5) the Poincaré polynomial of $J'_{\mathbb{R}}(2,\theta')$ has, in terms of the corresponding θ-stable parabolic subalgebra $\mathfrak{q}$, the form

$$P(J'_{\mathbb{R}}(2,\theta'),t)=t^{R(\mathfrak{q})}P(X_{L,u},t). \tag{4-18}$$

The compact dual $X_{L,u}$ is a product of h copies of the symmetric space

$$(U(2)\times U(2))/U(2).$$

Thus, using [Greub et al. 1976, Theorem IX], we obtain

$$P(X_{L,u},t)=\prod_{s=1}^{h}\prod_{i=1}^{m_s}(1+t^{2i-1}), \tag{4-19}$$

where $(m_0,m_1,\dots,m_h):=(0,2,\dots,2)$ is the partition of d attached to the θ_K-stable parabolic subalgebra $\mathfrak{q}$ in question. The shift $R(\mathfrak{q})=\dim(\mathfrak{u}\cap\mathfrak{p}_{\mathbb{C}})=\frac{1}{2}(\dim X_{\mathrm{GL}(d,\mathbb{H})}-\dim X_{L,u})$ is given by

$$R(\mathfrak{q})=\tfrac{1}{2}[(2d^2-d)-h\cdot 2^2]=\tfrac{1}{2}d(2d-3). \tag{4-20}$$

Thus, the Poincaré polynomial of the representation $J'_{\mathbb{R}}(2,\theta')$ of $\mathrm{GL}(d,\mathbb{H})$ has the form

$$P(J'_{\mathbb{R}}(2,\theta'),t)=t^{\frac{1}{2}d(2d-3)}\cdot\prod_{s=1}^{d/2}\prod_{i=1}^{m_s}(1+t^{2i-1}) \tag{4-21}$$

with $m_i=2,\ i=1,\dots,d/2$. ∎

5. Construction of cohomological cuspidal representations for $\mathrm{GL}(n)/k$ with prescribed local behavior

This section falls into two parts. First, using the transfer of irreducible cuspidal representations between $\mathrm{GL}(n)/k$ and $\mathrm{SL}(n)/k$ as proved in [Labesse and Schwermer 1986; 2019], and their actual construction in [Borel et al. 1996] in the case $\mathrm{SL}(n)/k$, we construct such representations $\pi = \otimes'_{v \in V_k} \pi_v$ for the former group such that its archimedean component π_∞ is cohomological with regard to the trivial coefficient system, and, given a finite set $S \subset V_{f,k}$ of nonarchimedean places, the corresponding local components π_v, $v \in S$, are Steinberg representations. Second, using the concept of automorphic induction, we construct irreducible cuspidal representations of $\mathrm{GL}(n)/k$, n even, with π_∞ cohomological with regard to the trivial coefficient system, and, given a fixed nonarchimedean place $v_0 \in V_{f,k}$, the local component π_{v_0} is not a square-integrable representation of $\mathrm{GL}(n, k_{v_0})$.

5A. *Via transfer.*

Theorem 5.1. *Let k be a totally real number field, and let* $\mathrm{GL}(n)$ *be the general linear group defined over k. Given a finite set $S \subset V_{f,k}$ of finite places of k, there exists a cuspidal automorphic representation $\pi = \otimes'_{v \in V_k} \pi_v$ occurring nontrivially in $L^2_{\mathrm{cusp}}(\mathrm{GL}(n, k) \backslash \mathrm{GL}(n, \mathbb{A}_k))$ so that the local component π_v, $v \in S$, is the Steinberg representation of* $\mathrm{GL}(n, k_v)$ *and so that the local components π_v, $v \in V_{\infty,k}$, of the representation π_∞ are (up to equivalence) the only irreducible tempered representation* $\mathrm{Ind}(P_{\delta_n}, \tau)$ *of* $\mathrm{GL}(n, \mathbb{R})$ *with nontrivial continuous cohomology* $H^*_{\mathrm{ct}}(\mathrm{GL}(n, \mathbb{R}), V_{\pi_\infty} \otimes \mathbb{C})$.

Proof. By [Borel et al. 1996, 11.3] the assertion is valid in the case of the special linear group $\mathrm{SL}(n)/k$. Thus, there exists a cuspidal automorphic representation $\pi = \otimes'_{v \in V} \pi_v$, occurring nontrivially in $L^2_{\mathrm{cusp}}(\mathrm{SL}(n, k) \backslash \mathrm{SL}(n, \mathbb{A}_k))$, so that the local component π_v, $v \in S$, is the Steinberg representation of $\mathrm{SL}(n, k_v)$ and so that the representation π_∞ has nontrivial continuous cohomology $H^*_{\mathrm{ct}}(\mathrm{SL}(n)_\infty, V_{\pi_\infty} \otimes \mathbb{C})$. Using the global transfer between $\mathrm{GL}(n)$ and $\mathrm{SL}(n)$ in terms of L-packets in the automorphic context as proved in [Labesse and Schwermer 1986, Proposition 3.5] (see the correction [Labesse and Schwermer 2019]), there exists a cuspidal automorphic representation $\tilde{\pi} = \otimes'_{v \in V} \tilde{\pi}_v$, occurring nontrivially in $L^2_{\mathrm{cusp}}(\mathrm{GL}(n, k) \backslash \mathrm{GL}(n, \mathbb{A}_k))$.

Given a place $v \in S$, the local component $\tilde{\pi}_v$, $v \in S$, is the Steinberg representation of $\mathrm{GL}(n, k_v)$, since the restriction of the Steinberg representation of the local group $\mathrm{GL}(n, k_v)$ is the Steinberg representation of the group $\mathrm{SL}(n, k_v)$, that is, the corresponding L-packet consists of one element.

At an archimedean place $v \in V_\infty$, by the results recalled in Section 4E, the local component $\tilde{\pi}_v$ is equivalent to the unique irreducible cohomological representation $\mathrm{Ind}(P_{\delta_n}, \tau_n)$ of $\mathrm{GL}(n, \mathbb{R})$.

Finally, note that the restriction $\tilde{\pi}_{v|\,\mathrm{SL}(n,k_v)}$ of an unramified representation $\tilde{\pi}_v$, $v \in V_{f,k}$, contains a uniquely determined constituent that is unramified. □

5B. ***Via automorphic induction*: *the case* GL(2).** We turn to the second construction. More specifically, we call for cuspidal representations $\pi = \otimes'_{v\in V_k}\pi_v$ of $\mathrm{GL}(2, \mathbb{A}_k)$ whose archimedean component π_∞ is cohomological and, given a fixed nonarchimedean place $v_0 \in V_{f,k}$, the component π_{v_0} is not a square-integrable representation of $\mathrm{GL}(2, k_{v_0})$.

In view of this task it is necessary to recall, through the cohomological lens, some facts regarding the compatibility of discrete series representations of $\mathrm{GL}(2, \mathbb{R})$ and the irreducible finite-dimensional algebraic representation (η, E) of $\mathrm{GL}(2)_\infty$ in a complex vector space E. As before we assume that this representation originates from an algebraic representation of the algebraic k-group $\mathrm{GL}(2)$. Its highest weight can be written as $\mu = (\mu)_{\iota_v}$, $v \in V_{\infty,k}$, where ι_v ranges over the embeddings $\iota_v : k \to \mathbb{R}$ corresponding to $v \in V_{\infty,k}$. Each of the weights $(\mu)_{\iota_v}$ is of the form $\mu_v\omega_v$, $\mu_v \in \mathbb{Z}, \mu_v \geq 0$, where ω_v denotes the fundamental dominant weight of the group $G_v \cong \mathrm{GL}(2, \mathbb{R})$, $v \in V_{\infty,k}$. Given a highest weight $\mu = (\mu)_{\iota_v}$, $v \in V_{\infty,k}$, we say that a family $\{D_{m_v}\}$, $m_v \in \mathbb{Z}, m_v \geq 2$, of discrete series representations of $\mathrm{GL}(2, \mathbb{R})$, parametrized by $v \in V_{\infty,k}$, is compatible with μ if $\mu_v\omega_v = (m_v - 2)\omega_v$ for all $v \in V_{\infty,k}$.

Theorem 5.2. *Let k be a totally real algebraic number field, and let (η, E) be an irreducible finite-dimensional algebraic representation of the archimedean component $G_\infty = \prod_{v\in V_{\infty,k}} G_v$, with $G_v \cong \mathrm{GL}(2, \mathbb{R})$, of $\mathrm{GL}(2, \mathbb{A}_k)$. The highest weight of (η, E) is denoted by $\mu = (\mu_v\omega_v)_{v\in V_{\infty,k}}$, where $\mu_v \in \mathbb{Z}$, $\mu_v \geq 0$. Given a fixed nonarchimedean place $v_0 \in V_{f,k}$, there exists an irreducible cuspidal automorphic representation π of $\mathrm{GL}(2, \mathbb{A}_k)$ whose archimedean component $\pi_\infty = \otimes_{v\in V_{\infty,k}}\pi_v$ is of the form*

$$\pi_\infty = \otimes_{v\in V_{\infty,k}} D_{m_v},$$

where the family $\{D_{m_v}\}_{v\in V_{\infty,k}}$ of discrete series representations of $\mathrm{GL}(2, \mathbb{R})$ is compatible with the highest weight μ, that is, $m_v = \mu_v + 2$ for all $v \in V_{\infty,k}$, and where the component π_{v_0} is not a square-integrable representation of $\mathrm{GL}(2, k_{v_0})$. The representation π of $\mathrm{GL}(2, \mathbb{A}_k)$ contributes nontrivially to the cuspidal cohomology $H^*_{\mathrm{cusp}}(\mathrm{GL}(2), E)$ in degree $d = [k : \mathbb{Q}]$.

Proof. The irreducible cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_k)$ we ask for will be constructed via automorphic induction from a Hecke character of an imaginary quadratic extension of k. Given a fixed nonarchimedean place $v_0 \in V_{f,k}$, we may choose such an extension field F of k such that v_0 splits in F. If $\ell = [k : \mathbb{Q}]$, then $[F : \mathbb{Q}] = 2\ell$. The extension $F/\mathbb{Q}$ is usually called a CM-field extension. Fix a CM-type Φ of F, that is, a set $\Phi \subset \mathrm{Hom}(F, \overline{\mathbb{Q}})$, say $\Phi = \{\sigma_w\}_{w\in V_{\infty,F}}$, such that, if $\sigma \in \Phi$, then its conjugate σ^c under the unique nontrivial element of the Galois group $\mathrm{Gal}(F/k)$ does not belong to Φ.

Given a unitary Hecke character $\theta : \mathbb{I}_F \to \mathbb{C}^\times$ of the group of ideles $\mathbb{I}_F$ of F, we denote by $\pi(\theta)$ the automorphic induction of θ to $\mathrm{GL}_2(\mathbb{A}_k)$. It is defined by $\pi(\theta) = \otimes'_v \pi(\theta)_v$, where

(1) if $v \in V_k$ splits in F, then $\pi(\theta)_v$ is the principal series representation of $\mathrm{GL}_2(k_v)$ induced from the character $\theta_{w_1} \otimes \theta_{w_2}$ of the torus, where w_1 and w_2 are the two places of F above v;

(2) if v does not split in F, then $\pi(\theta)_v$ is the local automorphic induction of θ_w to a representation of $\mathrm{GL}_2(k_v)$, where w is the unique place of F lying above v.

Since F is an imaginary quadratic extension field of k, all archimedean places of k do not split in F, thus, the second case is valid at places $v \in V_{\infty,k}$.

The discrete series representation $D_{\kappa+2}$ of $\mathrm{GL}(2, \mathbb{R})$ corresponds, via the local Langlands correspondence, to the two-dimensional irreducible representation of the Weil group $W_\mathbb{R}$ obtained by induction from the character of $W_\mathbb{C} = \mathbb{C}^*$ given by the assignment

$$z \mapsto (z/|z|)^{\kappa+1}, \quad z \in \mathbb{C}^*,$$

where $|z| = \sqrt{z \cdot \bar{z}}$. Hence, $D_{\kappa+2}$ is the local automorphic induction of that character.

Let $\{\mu_v\}_{v \in V_{\infty,k}}$ be the set of integers $\mu_v \in \mathbb{Z}$, $\mu_v \geq 0$, that originates with the highest weight $\mu = (\mu_v \omega_v)_{v \in V_{\infty,k}}$ of (η, E). It is a basic observation of Weil [1956], using a result of Chevalley [1951], that, since F is a CM-field, there is a unitary Hecke character θ of F with archimedean components given by

$$\theta_w(z_w) = (\sigma_w(z_w)/|\sigma_w(z_w)|)^{\mu_w+1} \quad \text{for all } w \in \Phi.$$

In turn, the discrete series representation D_{μ_w+2} is the automorphic induction $\pi(\theta)_v$ of θ_w, with $w \in \Phi$ the only place above $v \in V_{\infty,k}$. Note that $\theta \neq \theta^c$, with $c \in \mathrm{Gal}(F/k)$, $c \neq 1$, since this is correct already for the archimedean components. Thus, the unitary Hecke character θ does not factor through the norm map $N_{F/k}$. As a consequence, by [Arthur and Clozel 1989, Chapter 3, Section 6], the automorphic induction $\pi(\theta)$ of θ is a cuspidal automorphic representation of $\mathrm{GL}(2, \mathbb{A}_k)$.

Since, by the very choice of the CM-field F, the place v_0 of k splits in F, the local component $\pi(\theta)_v$ is the principal series representation of $\mathrm{GL}_2(k_v)$ induced from the character $\theta_{w_1} \otimes \theta_{w_2}$ on the torus in $\mathrm{GL}(2, k_{v_0})$, where w_1 and w_2 are the two places of F above v. Here we have identified k_{v_0} with F_{w_1} or F_{w_2}. Hence, $\pi(\theta)_{v_0}$ is not a square-integrable representation of $\mathrm{GL}(2, k_{v_0})$. □

5C. ***The case* GL(2*n*).** A slight extension by an additional step within the automorphic induction used in the proof above allows us to construct cuspidal representations of $\mathrm{GL}(2n)/k$ with specific local properties. This approach uses a totally real extension with cyclic Galois group.[3] In this case, the global automorphic induction relies on the work of Henniart [2012] and the proof of its compatibility with the local automorphic induction. This compatibility (over a local nonarchimedean field of characteristic zero) is dealt with in [Henniart and Herb 1995]. However, the decisive argument is the case $\mathrm{GL}(2)$. For the sake of simplicity we only deal with the trivial representation as a coefficient system, that is, $E = \mathbb{C}$.

Theorem 5.3. *Let k be a totally real algebraic number field, and let $v_0 \in V_{f,k}$ be a fixed nonarchimedean place of k. Then there exists an irreducible cuspidal automorphic representation π of $\mathrm{GL}(2n, \mathbb{A}_k)$ whose archimedean component $\pi_\infty = \otimes_{v \in V_{\infty,k}} \pi_v$ consists of the local components*

$$\pi_v = \mathrm{Ind}(P_{\delta_{2n}}, \sigma_v), \qquad v \in V_{\infty,k},$$

with P the parabolic subgroup of type $\delta_{2n} = (2, \dots, 2)$ and the representation $\sigma_v = \otimes \sigma_{v,i}$, $i = 1, \dots, n$, of $L_{P_{\delta_{2n}}}$ where $\sigma_{v,i}$ is the discrete series representation of $\mathrm{GL}(2, \mathbb{R})$ of lowest $O(2)$-type $2n - 2i + 2$, $i = 1, \dots, n$. Moreover, the component π_{v_0} corresponding to the fixed place $v_0 \in V_{k,f}$ is not a square-integrable representation of $\mathrm{GL}(2n, k_{v_0})$.

Proof. Given $n \in \mathbb{N}$, $n \geq 2$, there exists a totally real Galois extension $L'/\mathbb{Q}$ with cyclic Galois group of order n. Indeed, Dirichlet's theorem on arithmetical progressions asserts that any progression $a, a+q, a+2q, \dots$ where $(a, q) = 1$ contains infinitely many primes. So, we can a find a prime p with $p = 1 \mod 2n$. The p-th cyclotomic field $\mathbb{Q}(\zeta_p)$ contains the maximal totally real subfield $\mathbb{Q}(\zeta_p + \zeta_p^{-1})$. Since $2n$ divides $p-1$, n divides the degree $(p-1)/2$ of the cyclic extension $\mathbb{Q}(\zeta_p + \zeta_p^{-1})/\mathbb{Q}$. It follows that there is a Galois extension $L'/\mathbb{Q}$ with $L' \subset \mathbb{Q}(\zeta_p + \zeta_p^{-1})$ and cyclic Galois group $\mathrm{Gal}(L'/\mathbb{Q})$ of order n. Since we have infinitely many options to choose the prime p we may (and will) assume that the totally real fields k and L' are linearly disjoint within the algebraic closure $\overline{\mathbb{Q}}$. We denote their compositum by L. Then L/k is a Galois extension of degree n. Let γ denote a generator of the Galois group $\mathrm{Gal}(L/k)$. Choose an imaginary extension F' of L', and form the compositum F of L and F'. Then $F/\mathbb{Q}$ is a CM-field extension, with k its maximal totally real subfield.

We fix a CM-type Φ' of F'. Each element in Φ' extends an archimedean place $V_{L',\infty}$, that is, $|\Phi'| = n$. Following the argument in the proof of Theorem 5.2 there is a unitary Hecke character $\theta' : \mathbb{I}_{F'} \to \mathbb{C}^\times$ whose archimedean component is of the form

$$((z/|z|)^1, (z/|z|)^3, \dots, (z/|z|)^{2n-1}).$$

[3]This idea is taken from Clozel [1987] who deals with the case $\mathrm{SL}(2n)$. Aside from that, we have to deal with the additional local property which is required at a given finite place.

We denote by $\theta : \mathbb{I}_F \to \mathbb{C}^\times$ the character obtained as the composite $\theta' \circ N_{F/F'}$ where $N_{F/F'}$ denotes the norm map. Then, as a first step, there exists a cuspidal representation $\pi(\theta)$ of $\mathrm{GL}(2, \mathbb{A}_L)$, the automorphic induction of θ. In the second step, given $\pi(\theta)$, there exists an automorphic representation $\Pi(\theta)$ of $\mathrm{GL}(2n, \mathbb{A}_k)$ (unique up to isomorphism) whose base change (under the extension L/k) is given as

$$\Psi := \pi(\theta) \times \pi(\theta)^{\gamma} \times \cdots \times \pi(\theta)^{\gamma^{n-1}}.$$

Again we see that the representation $\pi(\theta)$ is not fixed by the elements in $\mathrm{Gal}(L/k)$, thus, $\Pi(\theta)$ is a cuspidal representation of $\mathrm{GL}(2n, A_k)$.

The process of global automorphic induction is compatible with the local process [Henniart 2012, Theorem 5]. More precisely, given a place $v \in V_k$, let $L_v = L \otimes_k k_v$ be the k_v-algebra cyclic under $\mathrm{Gal}(L/k)$. Then $\Pi(\theta)_v$ is the local automorphic induction of the representation $\pi(\theta)_v$ of $\mathrm{GL}(L_v)$. At an archimedean place $v \in V_{\infty,k}$, the representation $\Pi(\theta)_v$ of $\mathrm{GL}(2n, k_v)$ is therefore (up to isomorphism) of the form

$$\Pi(\theta)_v \cong D_2 \times D_4 \times \ldots \times D_{2n}$$

since the unitary Hecke character $\theta' : \mathbb{I}_{F'} \to \mathbb{C}^\times$ we started with had the archimedean component $((z/|z|)^1, (z/|z|)^3, \ldots, (z/|z|)^{2n-1})$. We observe that this is exactly the unique irreducible tempered representation of $\mathrm{GL}(2n, \mathbb{R})$, denoted $\mathrm{Ind}(P_{\delta_{2n}}, \tau)$ in Section 4E, that has nontrivial continuous cohomology

$$H^*_{\mathrm{ct}}(\mathrm{GL}(2n, \mathbb{R}), \mathrm{Ind}(P_{\delta_{2n}}, \tau) \otimes \mathbb{C})$$

with regard to the trivial representation as a coefficient system.

Next, let $v_0 \in V_{f,k}$ be a fixed nonarchimedean place of k, and let $\tilde{v}_0 \in V_{f,L}$ be a place above v_0. Then, as proved in Theorem 5.2, we can get that the local representation $\pi(\theta)_{\tilde{v}_0}$ is not square-integrable. By the compatibility of the global and local automorphic induction, the property descends to the local representation $\Pi(\theta)_{v_0}$. □

6. Construction of residual automorphic representations for GL(2, D)/k

Let k be a totally real algebraic number field of degree ℓ, and let D be a finite-dimensional central division algebra over k of degree $d > 1$. We suppose that there is at least one archimedean place at which D does not split. Then, by Section 2B, it follows that d is even. Let H' denote the algebraic k-group $\mathrm{GL}(2, D)$, and let H denote the k-group $\mathrm{GL}(2d)/k$ as before. Now, based on the existence of cuspidal automorphic representations with certain prescribed local behavior as proved in Theorem 5.1, we construct automorphic representations π' of $H'(\mathbb{A}_k)$ which occur as irreducible constituents in the residual spectrum of $H'(\mathbb{A}_k)$ and which eventually contribute nontrivially to the square-integrable cohomology of H'/k (see Section 8).

Theorem 6.1. *Let k be a totally real number field of degree ℓ, and let D be a finite-dimensional central division algebra over k of degree $d > 1$. Let $V_D \subset V_k$ be the finite set of places of k at which D does not split. Let t denote the number of archimedean places in V_D, and suppose that $t > 0$. Then there exist automorphic representations $\pi' = \otimes'_{v \in V_k} \pi_v$ of $H'(\mathbb{A}_k)$ which occur as irreducible constituents in the residual spectrum of $H'(\mathbb{A}_k)$ and whose archimedean component π_v at a place $v \in V_D \cap V_{k,\infty}$ is equivalent to the irreducible unitary representation $J'_{\mathbb{R}}(2, \theta')$ of* $\mathrm{GL}(d, \mathbb{H})$.

Proof. Consider the maximal parabolic k-subgroup Q_d of $H = \mathrm{GL}(2d)$ whose Levi subgroup is isomorphic to $\mathrm{GL}(d) \times \mathrm{GL}(d)$. Since d is even, say $d = 2h$, there exists an irreducible cuspidal automorphic representation $\sigma = \otimes'_{v \in V_k} \sigma_v$ of $\mathrm{GL}(d, \mathbb{A}_k)$ which satisfies the following two conditions (see Theorem 5.1):

Firstly, the archimedean components of σ are of the form

$$\sigma_v \cong D_2 \times D_4 \times \cdots D_{2h}.$$

This is exactly the irreducible unitary representation of $\mathrm{GL}(d, \mathbb{R})$, which we denoted $\mathrm{Ind}(P_{\delta_d}, \tau)$ in Section 4E, the unique one (up to infinitesimal equivalence) that is tempered and so that the continuous cohomology $H^*_{\mathrm{ct}}(\mathrm{GL}(d, \mathbb{R}), \mathrm{Ind}(P_{\delta_d}, \tau) \otimes \mathbb{C})$ does not vanish.

Secondly, for all finite places $v \in V_D \cap V_{k,f}$, the local component σ_v is the Steinberg representation of $\mathrm{GL}(d, k_v)$. This representation is an irreducible admissible representation of $\mathrm{GL}(d, k_v)$ and it is square-integrable modulo the center [Casselman 1995, Section 7].

Then there is the automorphic representation of $\mathrm{GL}(2d, \mathbb{A}_k)$, to be denoted

$$\sigma |\det|^{1/2} \times \sigma |\det|^{-1/2}, \tag{6-1}$$

obtained by parabolic induction from $\sigma \otimes \sigma$ on the Levi subgroup

$$L_{Q_d} \cong \mathrm{GL}(d) \times \mathrm{GL}(d)$$

of the parabolic subgroup Q_d. By the work of Moeglin and Waldspurger [1989], this representation has a unique irreducible unitary quotient $J(2, \sigma)$ which contributes nontrivially to the summand

$$L^2_{\mathrm{res},\{Q_d\}}(H, \omega_\sigma) \tag{6-2}$$

of the decomposition of the residual spectrum $L^2_{\mathrm{res}}(H, \omega_\sigma)$ of $H(\mathbb{A}_k) = \mathrm{GL}(2d, \mathbb{A}_k)$.

With regard to this constituent $J(2, \sigma)$ the works of Badulescu and Renard, in particular, [Badulescu 2008] and [Badulescu and Renard 2010, Proposition 18.2], imply that, if $\pi \cong J(s, \sigma)$, where $\sigma = \otimes_{v \in V} \sigma_v$ is a cuspidal automorphic representation of $\mathrm{GL}(d, \mathbb{A}_k)$, is a summand of $L^2_{\mathrm{res},\{Q_d\}}(H, \omega_\sigma)$, then π is always compatible with respect to D, that is, $\pi \cong J(s, \sigma)$ occurs in the image of Ξ.

Moreover, by construction, σ_v is square-integrable at all nonsplit places $v \in V_D$. Let σ' be the cuspidal automorphic representation of $\mathrm{GL}(1, D)(\mathbb{A}_k) = D^\times_{\mathbb{A}_k}$ corresponding to σ by the Jacquet–Langlands correspondence. Note that σ' is not one-dimensional. Then $J(s, \sigma)$ corresponds under Ξ to the representation $J'(s, \sigma')$ of $H'(\mathbb{A}_k)$. The latter representation $J'(s, \sigma')$ is obtained as the unique irreducible unitary quotient of the automorphic representation

$$\sigma' \,\mathrm{nrd}^{1/2} \times \sigma' \,\mathrm{nrd}^{-1/2}, \tag{6-3}$$

and it occurs nontrivially in the summand

$$L^2_{\mathrm{res},\{Q'\}}(H', \omega_\sigma) \tag{6-4}$$

of the residual spectrum of $\mathrm{GL}(2, D)(\mathbb{A}_k)$. In this construction, by Proposition 4.2, the archimedean component $J'(s, \sigma')_v$ of $J'(s, \sigma')$ at a place $V_D \cap V_{k,\infty}$ is equivalent to the irreducible unitary representation $J'_{\mathbb{R}}(2, \theta')$ of $\mathrm{GL}(d, \mathbb{H})$. □

Remark 6.2. Note that, if $v \in V_{k,\infty}$ is a place at which the central division algebra splits, that is $D_v \otimes k_v \cong M_d(\mathbb{R})$, and hence $H'_v \cong \mathrm{GL}(2d, \mathbb{R})$, then the corresponding component of π as constructed is of the form $\pi_v \cong J_{\mathbb{R}}(2, \theta)$.

7. Construction of cuspidal automorphic representations for GL(2, D)/k

In this section we use the results of Section 5 to prove the existence of cuspidal automorphic representations of the group $\mathrm{GL}(2, D)(\mathbb{A}_k)$ which are of cohomological relevance. These representations occur in two different forms. One consists of cuspidal representations whose archimedean components are tempered and whose construction relies, via the general Jacquet–Langlands correspondence, on Theorem 5.1. The other form consists of cuspidal representations whose archimedean components are nontempered. Since these latter cuspidal representations are nearly equivalent to the residual automorphic representations constructed in Theorem 6.1, they may be viewed as shadows of Eisenstein series.

7A. ***The tempered case.*** We recall that, given a central division algebra D over k of degree d, the center Z of both of the two groups $H/k = \mathrm{GL}(2d)/k$ and $H'/k = \mathrm{GL}(2, D)/k$ is isomorphic to the group of ideles $\mathbb{I}_k$. Thus, we may view a unitary character of $Z(k)\backslash Z(\mathbb{A}_k)$ as a unitary character of $k^\times\backslash \mathbb{I}_k$. We fix such a character ω. It is preserved by the global Jacquet–Langlands correspondence.

Theorem 7.1. *Let k be a totally real number field of degree ℓ, and let D be a central division algebra over k of degree d. Let V_D be the finite set of places of k at which D does not split. Let t denote the number of archimedean places of k at which D does not split. Denote by H' the algebraic k-group* $\mathrm{GL}(2, D)$, *an inner form of the algebraic k-group $H = \mathrm{GL}(2d)$. Then there exist cuspidal automor-*

phic representations π' of $H'(\mathbb{A}_k)$ with $\Xi(\pi') =: \pi$ cuspidal under the Jacquet–Langlands correspondence Ξ, and whose archimedean components π'_v, $v \in V_{k,\infty}$, are irreducible tempered representations of H'_v with $H^*_{\mathrm{ct}}(H'_v, V_{\pi_v} \otimes \mathbb{C}) \neq \{0\}$.

Proof. We denote by S the finite set $V_D \cap V_{k,f}$, that is, the finite set of nonarchimedean places at which D does not split. By Theorem 5.1 there exists a cuspidal automorphic representation $\pi = \otimes'_{v\in V_k} \pi_v$ occurring nontrivially in

$$L^2_{\mathrm{cusp}}(\mathrm{GL}(2d, k)\backslash \mathrm{GL}(2d, \mathbb{A}_k))$$

so that the local component π_v, $v \in S$, is the Steinberg representation of $\mathrm{GL}(2d, k_v)$ and so that the local components π_v, $v \in V_{k,\infty}$, of the representation π_∞ are (up to equivalence) the only irreducible representation $\mathrm{Ind}(P_{\delta_{2d}}, \sigma)$ of $\mathrm{GL}(2d, \mathbb{R})$ with nontrivial continuous cohomology $H^*_{\mathrm{ct}}(\mathrm{GL}(2d, \mathbb{R}), V_{\pi_v} \otimes \mathbb{C})$. The representation π is compatible with D, since at the places $v \in S$ the local component is square-integrable. Thus, the representation π is in the image of the injective map Ξ, that is, there exists an irreducible constituent π' of $L^2_{\mathrm{cusp}}(\mathrm{GL}(2, D)(\mathbb{A}_k))$ with $\Xi(\pi') = \pi$. Under the local correspondence, for $v \notin V_D$, clearly $\pi_v \cong \pi'_v$, and π_v corresponds to π'_v by the local Jacquet–Langlands correspondence at $v \in V_D$. In particular, let $v \in V_D \cap V_{k,\infty}$, then necessarily d is even and $D_v \cong M_{d/2}(\mathbb{H})$, thus $H'_v \cong \mathrm{GL}(d, \mathbb{H})$. Within the classification (up to infinitesimal equivalence) of the irreducible unitary representations of $\mathrm{GL}(d, \mathbb{H})$ with nonvanishing continuous cohomology with regard to the trivial coefficient system, one finds the so-called fundamental series representation. It is constructed as follows: We denote by $P'_{\delta_d} = L'_{\delta_d} N'_{\delta_d}$ the standard minimal parabolic subgroup of $\mathrm{GL}(d, \mathbb{H})$ whose Levi subgroup L'_{δ_d} consists of d copies of $\mathrm{GL}(1, \mathbb{H})$. As pointed out in Remark 3.3 the local Jacquet–Langlands correspondence between $\mathrm{GL}(2, \mathbb{R})$ and $\mathbb{H}^\times$ asserts that if $\delta = D_2(\chi \circ \det_2)$ is of lowest $O(2)$-type 2, then it corresponds to the character $\chi \circ \mathrm{nrd}_1$ of $\mathbb{H}^\times$, where χ is a unitary character of $\mathbb{R}^\times$. Thus, D_2 corresponds to the trivial character of $\mathbb{H}^\times$.

If $\delta = D_m(\chi \circ \det_2)$ is of lowest $O(2)$-type $m > 2$, then it corresponds to $\delta' = D'_m(\chi \circ \mathrm{nrd}_1)$, which is not a one-dimensional representation of $\mathbb{H}^\times$. Given the discrete series representation D_m, $m > 2$, of $\mathrm{GL}(2, \mathbb{R})$, we denote by D'_m the corresponding representation of $\mathbb{H}^\times$. The local representation

$$\mathrm{Ind}(P'_{\delta_d}, \sigma') \cong \mathbf{1}_{\mathbb{H}^\times} \times D'_4 \times \ldots \times D'_{2d}$$

of $\mathrm{GL}(d, \mathbb{H})$ is an irreducible unitary representation, the unique one that is tempered and has nonvanishing continuous cohomology with regard to the coefficient system $\mathbb{C}$. By [Badulescu and Renard 2010, Section 13], under the local Jacquet–Langlands correspondence this representation corresponds to the irreducible tempered representation $\mathrm{Ind}(P_{\delta_{2d}}, \sigma) \cong D_2 \times D_4 \times \ldots \times D_{2d}$ of $\mathrm{GL}(2d, \mathbb{R})$. $\square$

7B. *The nontempered case.* We retain the notation of the previous subsection. As before we suppose that the set V_D of places of D at which D does not split contains at least one archimedean place. Then it follows that d is even. We write $d = 2h$. Let H' denote the algebraic k-group $\mathrm{GL}(2, D)$. In the following we construct cuspidal automorphic representations π' of $H(\mathbb{A}_k')$ which eventually contribute nontrivially to the cuspidal cohomology $H^*_{\mathrm{cusp}}(H', \mathbb{C})$ and which are CAP-representations. For the sake of clarity we recall this notion.

Definition 7.2. We call an irreducible cuspidal representation τ of a quasi-split connected reductive k-group G a CAP-representation with respect to a parabolic k-subgroup P of G if τ is nearly equivalent to an irreducible constituent of an induced representation $\mathrm{Ind}_P^G\sigma$ where σ is a cuspidal representation of the Levi subgroup of P.

If G' is an inner form of a quasi-split group G as above, a modification of this notion of being CAP is necessary (see, for example, [Gan 2008, 3.9, 3.10] or [Jiang 2010, Section 6]). Since the local groups G_v' and G_v are isomorphic for almost all $v \in V_k$, it makes sense to say that a representation τ' of $G'(\mathbb{A}_k)$ is nearly equivalent to a representation of $G(\mathbb{A}_k)$. Thus, we call an irreducible cuspidal representation τ' of $G'(\mathbb{A}_k)$ a CAP representation with respect to a parabolic k-subgroup of G if τ' is nearly equivalent to an irreducible constituent of an induced representation $\mathrm{Ind}_P^G\sigma$ where σ is a cuspidal representation of the Levi subgroup of P.

Theorem 7.3. *Let k be a totally real number field, and let D be a finite-dimensional central division algebra over k of degree d. Suppose that the set V_D of places of D at which D does not split contains at least one archimedean place. Let H'/k denote the algebraic k-group* $\mathrm{GL}(2, D)/k$. *Then there exist cuspidal automorphic representations π' of $H'(\mathbb{A}_k)$ with $\Xi(\pi') =: \pi$ a residual representation of the group $H(\mathbb{A}_k)$ attached to the split group $H/k = \mathrm{GL}(2d)/k$ under the Jacquet–Langlands correspondence Ξ so that the archimedean components π_v', $v \in V_{\infty,k}$, have the following form: If $v \in V_D \cap V_{k,\infty}$, that is, $H_v' \cong \mathrm{GL}(d, \mathbb{H})$, then $\pi_v' \cong J_{\mathbb{R}}'(2, \theta')$, and if $v \in V_{k,\infty}$, $v \notin V_D$, that is, $H_v' \cong \mathrm{GL}(2d, \mathbb{R})$, then $\pi_v' \cong J_{\mathbb{R}}(2, \theta)$. In both cases the archimedean component is a nontempered representation of H_v'.*

The representation π' is a CAP-representation with respect to the (*maximal*) *parabolic k-subgroup $Q_d = Q_{\Delta\setminus\{\alpha_d\}}$ of* $\mathrm{GL}(2d)/k$.

Proof. The group $H' = \mathrm{GL}(2, D)/k$ is a k-form of the general linear k-group $H = \mathrm{GL}(2d)/k$. Let l be a splitting field of D, thus, there is an isomorphism

$$\psi : \mathrm{GL}(2, D) \times_k l \to \mathrm{GL}(2d)/l$$

of algebraic l-groups. Let Q' be the minimal parabolic k-subgroup of $\mathrm{GL}(2, D)$ fixed in Section 2A. The image of the l-group $Q' \times_k l$ under ψ is the standard parabolic l-subgroup $Q_d = Q_{\Delta\setminus\{\alpha_d\}}$ of $\mathrm{GL}(2d)/\ell$. Its Levi subgroup L_{Q_d}/l is isomorphic to $\mathrm{GL}(d)/l \times \mathrm{GL}(d)/l$. Since d is even, say $d = 2h$, we can use the construction of cuspidal representations carried through in Theorem 5.3 for each of these factors. Thus, there exists an irreducible cuspidal automorphic representation τ of $\mathrm{GL}(2h, \mathbb{A}_k)$ whose archimedean component $\tau_\infty = \otimes_{v \in V_{k,\infty}} \tau_v$ consists of the local components

$$\tau_v = \mathrm{Ind}(P_{\delta_{2h}}, \sigma_v), \quad v \in V_{k,\infty},$$

with P the parabolic subgroup of type $\delta_{2h} = (2, \dots, 2)$ and the representation $\sigma_v = \otimes \sigma_{v,i},\ i = 1, \dots, h$, of $L_{P_{\delta_{2h}}}$ where $\sigma_{v,i}$ is the discrete series representation of $\mathrm{GL}(2, \mathbb{R})$ of lowest $O(2)$-type $2h - 2i + 2,\ i = 1, \dots, h$. Moreover, by Theorem 5.3, we may assume that at a fixed place $v_0 \in V_{f,k}$ the component τ_{v_0} is not a square-integrable representation of $\mathrm{GL}(2h, k_{v_0})$.

We denote by $\mathrm{Ind}(2, \tau)$ the representation of $\mathrm{GL}(2d, \mathbb{A}_k)$ induced from the representation

$$\tau |\det|^{\frac{1}{2}} \otimes \tau |\det|^{-\frac{1}{2}}$$

of the Levi factor $L_{Q_d}(\mathbb{A}_k)$. As proved in [Mœglin and Waldspurger 1989, I. 11], this representation has a unique irreducible quotient to be denoted by $J(2, \tau)$. It is a representation of $\mathrm{GL}(2d, \mathbb{A}_k)$ which occurs in the residual spectrum (see [Jacquet and Shalika 1981]). The representation $J(2, \tau)$ is compatible with respect to D. Thus, there exists a unique irreducible automorphic representation π' of $H'(\mathbb{A}_k)$ with $\Xi(\pi') = J(2, \tau)$. Since the local representation τ_{v_0} is not a square-integrable representation of $\mathrm{GL}(2h, k_{v_0})$, it follows, by [Badulescu and Renard 2010, Proposition 18. 2] (see Theorem 3.4 in this paper), that the representation π' is cuspidal.

Let $v \in V_{k,\infty}$ be an archimedean place of k. By construction, the local component of the cuspidal representation τ is of the form

$$\tau_v = \mathrm{Ind}(P_{\delta_{2h}}, \sigma_v), \quad v \in V_{\infty,k},$$

with P the parabolic subgroup of type $\delta_{2h} = (2, \dots, 2)$ and the representation $\sigma_v = \otimes \sigma_{v,i},\ i = 1, \dots, h$, of $L_{P_{\delta_{2h}}}$ where $\sigma_{v,i}$ is the discrete series representation of $\mathrm{GL}(2, \mathbb{R})$ of lowest $O(2)$-type $2h - 2i + 2,\ i = 1, \dots, h$. Thus, the archimedean component $J(2, \tau_v)$ of $J(2, \tau)$, $v \in V_{k,\infty}$, is equivalent to the irreducible nontempered representation $J_{\mathbb{R}}(2, \theta)$ of $\mathrm{GL}(2d, \mathbb{R})$ (in the notation of Section 4F). Hence, if $v \in V_{k,\infty}$, $v \notin V_D$, then $\pi'_v \cong J_{\mathbb{R}}(2, \theta)$ and, using Proposition 4.2, $\pi'_v \cong J'_{\mathbb{R}}(2, \theta')$ if $v \in V_D \cap V_{k,\infty}$. In the latter case, recall that $\theta' \cong \tau'_v$ where τ'_v is the representation of $\mathrm{GL}(h, \mathbb{H})$ which corresponds under the Jacquet–Langlands correspondence to τ_v. Again, like the representation $J_{\mathbb{R}}(2, \theta)$, the representation $J_{\mathbb{R}}(2, \theta') \cong J_{\mathbb{R}}(2, \tau'_v)$ is nontempered and it has nonvanishing continuous cohomology.

By construction, one sees that the cuspidal automorphic representations π' of $H'(\mathbb{A}_k)$ with $\Xi(\pi') = J(2, \tau)$ is a shadow of an Eisenstein series with cuspidal support in the parabolic k-subgroup $Q_d = Q_{\Delta \setminus \{\alpha_d\}}$ of $\mathrm{GL}(2d, k)$. Thus, it is a CAP-representation. $\square$

Remark 7.4. We remark that for the case $\mathrm{GL}(2, B)/\mathbb{Q}$ with B a quaternion division algebra of discriminant two over the field $\mathbb{Q}$ of rational numbers, [Muto et al. 2016] provides an explicit construction of cuspidal automorphic forms lifted from suitable Maass cusp forms, and thus, finally an explicit example of a CAP representation in this specific case.

8. Non-vanishing results for the square-integrable cohomology of GL(2, D)

In this section we prove various nonvanishing results for the square-integrable cohomology of $\mathrm{GL}(2, D)$ which are implied by the constructions of specific automorphic representations carried through in the previous two sections. We begin with a brief review of the cohomology groups in question.

8A. ***The cohomology groups $H^*(G, \mathbb{C})$.*** Let G be a reductive algebraic group over a totally real algebraic number field k, and suppose that G modulo its radical has k-rank greater than zero. We write G_∞ for the group $R_{k/\mathbb{Q}}(G)(\mathbb{R})$ of real points of the algebraic $\mathbb{Q}$-group $R_{k/\mathbb{Q}}(G)$ obtained from G by restriction of scalars, and K_∞ for a maximal compact subgroup of G_∞.

Let $J \subset Z(\mathfrak{g}_{\infty,\mathbb{C}})$ be the annihilator of the trivial representation in the center of the universal enveloping algebra $U(\mathfrak{g}_{\infty,\mathbb{C}})$ of the complexified Lie algebra of G_∞. Then J is an ideal of finite codimension in $Z(\mathfrak{g}_{\infty,\mathbb{C}})$. Let $V_{G,\mathrm{umg}} = C^\infty_{\mathrm{umg}}(G(k)\backslash G(\mathbb{A}_k))$ be the space of smooth complex-valued functions f of uniform moderate growth on $G(k)\backslash G(\mathbb{A}_k)$, in the sense of [Mœglin and Waldspurger 1994, I.2.3]. Define $\mathcal{A}(G) \subset V_{G,\mathrm{umg}}$ to be the subspace of functions $f \in V_{G,\mathrm{umg}}$ which are annihilated by a power of J and which are trivial on the identity component $A_{G,\infty}$ of the group $\mathrm{Res}_{k/\mathbb{Q}}(S)(\mathbb{R})$ with S the maximal split torus in the center Z of G. The space $\mathcal{A}(G)$ is naturally equipped with a $(\mathfrak{m}_G, K_\infty; G(\mathbb{A}_{k,f}))$-module structure where $\mathfrak{m}_G$ denotes the Lie algebra of $A_{G,\infty} \setminus G_\infty$. Thus, the $(\mathfrak{m}_G, K_\infty)$-cohomology $H^*(\mathfrak{m}_G, K_\infty; \mathcal{A}(G) \otimes \mathbb{C})$ is well-defined.

Following the work of Franke [1998], these cohomology groups present themselves as the automorphic interpretation of the cohomology groups given as the inductive limits

$$H^*(G, \mathbb{C}) := \mathrm{colim}_C\, H^*(X_C, \mathbb{C}) \tag{8-1}$$

over all sufficiently small open compact subgroups $C \subset G(\mathbb{A}_{k,f})$ of the de Rham cohomology groups $H^*(X_C, \mathbb{C})$ associated to the orbit space

$$X_C := G(k)A_{G,\infty} \setminus G(\mathbb{A}_k)/K_\infty C. \tag{8-2}$$

As proved by Rohlfs [1996, Corollary 2.12], the cohomology $H^*(G, \mathbb{C})$ is isomorphic (in a functorial way) to the cohomology of the projective limit $S := \lim_C X_C$, that is, we have

$$H^*(G, \mathbb{C}) = \mathrm{colim}_C\, H^*(X_C, \mathbb{C}) \cong H^*(\lim_C X_C, \mathbb{C}). \tag{8-3}$$

An analogous result is correct for the cohomology with compact supports, denoted by $H_c^*(-, \mathbb{C})$, that is,

$$H_c^*(G, \mathbb{C}) := \mathrm{colim}_C\, H_c^*(X_C, \mathbb{C}) \cong H_c^*(\lim_C X_C, \mathbb{C}). \tag{8-4}$$

We denote by $H_!^*(G, \mathbb{C})$ the image of the cohomology $H_c^*(G, \mathbb{C})$ with compact supports in $H^*(G, \mathbb{C})$, usually called the *interior cohomology*.[4]

8B. ***The square-integrable cohomology groups $H^*_{(\mathrm{sq})}(G, \mathbb{C})$.*** The space $\mathcal{A}(G)$ contains as a natural submodule the subspace $\mathcal{L}(G)$ consisting of all square-integrable automorphic forms in $\mathcal{A}(G)$. The inclusion $\mathcal{L}(G) \hookrightarrow \mathcal{A}(G)$ gives rise to a morphism in $(\mathfrak{m}_G, K_\infty)$-cohomology,

$$H^*(\mathfrak{m}_G, K_\infty; \mathcal{L}(G) \otimes \mathbb{C}) \to H^*(\mathfrak{m}_G, K_\infty; \mathcal{A}(G) \otimes \mathbb{C}). \tag{8-5}$$

We call the image of this map the *square-integrable* (*automorphic*) *cohomology of* G, to be denoted by $H^*_{(\mathrm{sq})}(G, \mathbb{C})$, whereas the right-hand side, usually denoted $H^*(G, \mathbb{C})$, presents the automorphic cohomology of G with trivial coefficients.

Let the submodule in $\mathcal{L}(G)$ with regard to the $(\mathfrak{m}_G, K_\infty; G(\mathbb{A}_{k,f}))$-module structure which is spanned by all irreducible submodules be denoted by $L^2_{\mathrm{disc},J}(G)$; it is called the discrete spectrum of G with regard to J. It contains the cuspidal spectrum $L^2_{\mathrm{cusp},J}(G)$ as a submodule. In fact, there is a decomposition

$$L^2_{\mathrm{disc},J}(G) = L^2_{\mathrm{cusp},J}(G) \oplus L^2_{\mathrm{res},J}(G) \tag{8-6}$$

as an $(\mathfrak{m}_G, K_\infty; G(\mathbb{A}_{k,f}))$-module where the complement $L^2_{\mathrm{res},J}(G)$ denotes the residual spectrum of G with regard to J. The two inclusions of $(\mathfrak{m}_G, K_\infty; G(\mathbb{A}_{k,f}))$-modules

$$j_{\mathrm{disc}} : L^2_{\mathrm{disc},J}(G) \to \mathcal{A}(G), \qquad j_{\mathrm{cusp}} : L^2_{\mathrm{cusp},J}(G) \to \mathcal{A}(G) \tag{8-7}$$

induce homomorphisms on the level of $(\mathfrak{m}_G, K_\infty)$-Lie algebra cohomology. The image of

$$j^*_{\mathrm{disc}} : H^*(\mathfrak{m}_G, K_\infty; L^2_{\mathrm{disc},J}(G) \otimes \mathbb{C}) \to H^*(\mathfrak{m}_G, K_\infty; \mathcal{A}(G) \otimes \mathbb{C}) \tag{8-8}$$

is equal to the square-integrable cohomology $H^*_{(\mathrm{sq})}(G, \mathbb{C})$ (see [Borel et al. 1996, Section 7] using [Mœglin and Waldspurger 1994, VI, 2.1]). In general, the

[4] These interior cohomology groups enjoy a natural interpretation in the framework of the Borel-Serre compactification, a manifold $\bar{S}$ with boundary $\partial\bar{S}$, of S. The interior cohomology consists of all those classes in $H^*(\bar{S}, \mathbb{C})$ which restrict trivially to the cohomology of the boundary $\partial\bar{S}$.

map j^*_{disc} need not be injective. Within de Rham theory, the cohomology space $H^*(\mathfrak{m}_G, K_\infty; L^2_{\mathrm{disc},J}(G) \otimes \mathbb{C})$ may be interpreted as the space of harmonic square-integrable differential forms on S, [Borel and Garland 1983, Proposition 5.6]. By [Borel et al. 1996, Section 5], the homomorphism

$$j^*_{\mathrm{cusp}} : H^*(\mathfrak{m}_G, K_\infty; L^2_{\mathrm{cusp},J}(G) \otimes \mathbb{C}) \to H^*(\mathfrak{m}_G, K_\infty; \mathcal{A}(G) \otimes \mathbb{C}) \tag{8-9}$$

is injective. We denote by $H^*_{\mathrm{cusp}}(G, \mathbb{C})$ its image, the cuspidal cohomology of G.

Note that, using Theorem 5.3 and its corollary in [Borel 1981], it is not difficult to show that the interior cohomology $H^*_!(G, \mathbb{C})$ contains the cuspidal cohomology $H^*_{\mathrm{cusp}}(G, \mathbb{C})$, thus, the topologically defined object $H^*_!(G, \mathbb{C})$ is sandwiched between two analytically defined cohomology groups, that is, we have

$$H^*_{\mathrm{cusp}}(G, \mathbb{C}) \subset H^*_!(G, \mathbb{C}) \subset H^*_{(\mathrm{sq})}(G, \mathbb{C}) \subset H^*(G, \mathbb{C}). \tag{8-10}$$

8C. *Construction of tempered nontrivial classes in $H^*_{\mathrm{cusp}}(\mathrm{GL}(2, D), \mathbb{C})$.* Let D be a central division algebra over k of degree d. The center Z of both of the two groups $H/k = \mathrm{GL}(2d)/k$ and $H'/k = \mathrm{GL}(2, D)/k$ is isomorphic to the group of ideles $\mathbb{I}_k$ via the isomorphism that assigns to an element $a \in \mathbb{I}_k$ the scalar matrix of the appropriate size with a on the diagonal. Thus, we may view a unitary character of $Z(k)\backslash Z(\mathbb{A}_k)$ as a unitary character of $k^\times\backslash\mathbb{I}_k$. We fix such a character ω. It is preserved by the global Jacquet–Langlands correspondence.

By Theorem 5.1, given a totally real number field k, and a finite set $S \subset V_f$ of finite places of k, there exists a cuspidal automorphic representation $\pi = \otimes'_{v\in V_k}\pi_v$ occurring nontrivially in $L^2_{\mathrm{cusp}}(\mathrm{GL}(2d, k)\backslash \mathrm{GL}(2d, \mathbb{A}_k))$ so that the local component $\pi_v,\ v \in S$, is the Steinberg representation of $\mathrm{GL}(2d, k_v)$ and so that the local components $\pi_v, v \in V_\infty$, of the representation π_∞ are (up to equivalence) the only irreducible tempered representation $\mathrm{Ind}(P_{\delta_{2d}}, \tau)$ of $\mathrm{GL}(2d, \mathbb{R})$ with nontrivial continuous cohomology $H^*_{\mathrm{ct}}(\mathrm{GL}(2d, \mathbb{R}), V_{\pi_\infty} \otimes \mathbb{C})$. By Proposition 4.1 the representation $\mathrm{Ind}(P_{\delta_{2d}}, \tau)$ of $\mathrm{GL}(2d, \mathbb{R})$ corresponds under the local Jacquet–Langlands correspondence to the irreducible tempered representation $\mathrm{Ind}(P'_{\delta_d}, \tau')$ of $\mathrm{GL}(d, \mathbb{H})$.

Theorem 8.1. *Let k be a totally real number field of degree ℓ, and let D be a central division algebra over k of degree d. Let V_D be the finite set of places of k at which D does not split. Let t denote the number of archimedean places of k at which D does not split. Then there exist cuspidal automorphic representations $\pi' = \otimes\pi'_v$ of $H'(\mathbb{A}_k)$ with $\Xi(\pi') =: \pi$ cuspidal under the Jacquet–Langlands correspondence Ξ, and so that the archimedean components π'_v, $v \in V_{\infty,k}$, have the following form: If $v \in V_D \cap V_{k,\infty}$, then $\pi'_v \cong \mathrm{Ind}(P'_{\delta_d}, \tau')$, and if $v \in V_{k,\infty}$, $v \notin V_D$, then $\pi'_v \cong \mathrm{Ind}(P_{\delta_{2d}}, \tau)$, that is, the archimedean components of π' are tempered representations of H'_v, $v \in V_{k,\infty}$. The representation π' represents a nontrivial class in $H^*_{\mathrm{cusp}}(\mathrm{GL}(2, D), \mathbb{C})$.*

Proof. We denote by S the finite set $V_D \cap V_{k,f}$. By Theorem 5.1 there exists a cuspidal automorphic representation $\pi = \otimes'_{v \in V_k} \pi_v$ occurring nontrivially in $L^2_{\mathrm{cusp}}(\mathrm{GL}(2d, k) \backslash \mathrm{GL}(2d, \mathbb{A}_k))$ so that the local component π_v, $v \in S$, is the Steinberg representation of $\mathrm{GL}(2d, k_v)$ and so that the local components π_v, $v \in V_{\infty,k}$, of the representation π_∞ are (up to equivalence) the irreducible representation $\mathrm{Ind}(P_{\delta_{2d}}, \tau)$ of $\mathrm{GL}(2d, \mathbb{R})$. The representation π is compatible with D, since at the places $v \in S$ the local component is square-integrable. Thus, the representation π is in the image of the injective map Ξ, that is, there exists an irreducible constituent π' of $L^2_{\mathrm{cusp}}(\mathrm{GL}(2, D)(\mathbb{A}_k))$ with $\Xi(\pi') = \pi$. Under the local correspondence, for $v \notin V_D$, clearly $\pi_v \cong \pi'_v$, and π_v corresponds to π'_v by the local Jacquet–Langlands correspondence at $v \in V_D$. In particular, let $v \in V_D \cap V_{k,\infty}$, then necessarily d is even and $D_v \cong M_{d/2}(\mathbb{H})$, thus $H'_v \cong \mathrm{GL}(d, \mathbb{H})$. By Proposition 4.1 the representation $\mathrm{Ind}(P_{\delta_{2d}}, \tau)$ of $\mathrm{GL}(2d, \mathbb{R})$ corresponds under the local Jacquet–Langlands correspondence to the irreducible tempered representation $\mathrm{Ind}(P'_{\delta_d}, \tau')$ of $\mathrm{GL}(d, \mathbb{H})$. Since the map

$$j^*_{\mathrm{cusp}} : H^*(\mathfrak{m}_{H'}, K'_\infty; L^2_{\mathrm{cusp},J}(H') \otimes \mathbb{C}) \to H^*(\mathfrak{m}_{H'}, K'_\infty; \mathcal{A}(H') \otimes \mathbb{C}) \tag{8-11}$$

induced by $j_{\mathrm{cusp}} : L^2_{\mathrm{cusp},J}(H') \to \mathcal{A}(H')$ is injective, we obtain nontrivial classes in the cuspidal cohomology $H^*_{\mathrm{cusp}}(\mathrm{GL}(2, D), \mathbb{C})$ whose archimedean components are tempered representations. □

8D. ***Construction of nontempered nontrivial classes in $H^*_{\mathrm{cusp}}(\mathrm{GL}(2, D), \mathbb{C})$.*** For a totally real number field k, let D be a finite-dimensional central division k-algebra of degree $d > 1$. We suppose that the set V_D of places of D at which D does not split contains at least one archimedean place.

It follows that d is even. We write $d = 2h$. Let H' denote the algebraic k-group $\mathrm{GL}(2, D)$. In the following we construct cuspidal automorphic representations π' of $H(\mathbb{A}'_k)$ which contribute nontrivially to the cuspidal cohomology $H^*_{\mathrm{cusp}}(H', \mathbb{C})$ and which are CAP-representations.

Theorem 8.2. *Let k be a totally real number field, and let D be a finite-dimensional central division algebra over k of degree d. Suppose that the set $V_D \cap V_{k,\infty}$ is nonempty of places of D. Let H'/k denote the algebraic k-group $\mathrm{GL}(2, D)/k$. Then there exist cuspidal automorphic representations $\pi' = \otimes \pi'_v$ of $H'(\mathbb{A}_k)$ with $\Xi(\pi') =: \pi$ a residual representation of the group $H(\mathbb{A}_k)$ attached to the split group $H/k = \mathrm{GL}(2d)/k$ under the Jacquet–Langlands correspondence Ξ so that the archimedean components π'_v, $v \in V_{\infty,k}$, have the following form: If $v \in V_D \cap V_{k,\infty}$, that is, $H'_v \cong \mathrm{GL}(d, \mathbb{H})$, then $\pi'_v \cong J'_{\mathbb{R}}(2, \theta')$, and if $v \in V_{k,\infty}$, $v \notin V_D$, that is, $H'_v \cong \mathrm{GL}(2d, \mathbb{R})$, then $\pi'_v \cong J_{\mathbb{R}}(2, \theta)$. In both cases the archimedean component is a nontempered representation of H'_v. The representation π' represents a nontrivial class in $H^*_{\mathrm{cusp}}(\mathrm{GL}(2, D), \mathbb{C})$.*

Proof. By Theorem 7.3 there exist cuspidal automorphic representations π' of $H'(\mathbb{A}_k)$ with $\Xi(\pi') =: \pi$ a residual representation of the group $H(\mathbb{A}_k)$ attached to the split group $H/k = \mathrm{GL}(2d)/k$ under the Jacquet–Langlands correspondence Ξ so that the archimedean components π'_v, $v \in V_{\infty,k}$, have the following form: If $v \in V_D \cap V_{k,\infty}$, that is, $H'_v \cong \mathrm{GL}(d, \mathbb{H})$, then $\pi'_v \cong J'_{\mathbb{R}}(2, \theta')$, and if $v \in V_{k,\infty}$, $v \notin V_D$, that is, $H'_v \cong \mathrm{GL}(2d, \mathbb{R})$, then $\pi'_v \cong J_{\mathbb{R}}(2, \theta)$. In both cases the archimedean component is a nontempered representation of H'_v. Moreover, the continuous cohomology $H^*_{\mathrm{ct}}(H'_v, \pi'_v \otimes \mathbb{C})$ of H'_v with coefficients in π'_v, $v \in V_{k,\infty}$, does not vanish. If $\pi'_v \cong \pi_v \cong J_{\mathbb{R}}(2, \theta)$, this is proved in [Franke and Schwermer 1998, 5.6]. If $\pi'_v \cong J'_{\mathbb{R}}(2, \theta')$, we refer to Proposition 4.2 where one also finds the Poincaré polynomial of the cohomology space. Finally, as in the last step of the proof of Theorem 8.1, we see that we obtain nontrivial classes in the cuspidal cohomology $H^*_{\mathrm{cusp}}(\mathrm{GL}(2, D), \mathbb{C})$. □

Remark 8.3. As rounded off by Grbac [2008], the work of Badulescu [2008] gives a complete structural description of the discrete spectrum of $\mathrm{GL}(2, D)/k$. In particular, as a consequence of the description of the residual spectrum, Grbac [2008, A. 8] obtains a classification of the cuspidal spectrum. Using this result we observe that the cuspidal representations of $\mathrm{GL}(2, D)$ with cohomological archimedean components as constructed in Theorem 8.1 and Theorem 7.3 cover the only two possibilities to construct cuspidal cohomology classes for $\mathrm{GL}(2, D)/k$.

Remark 8.4. Grobner [2013] deals with the automorphic cohomology in the case $\mathrm{GL}(2, B)$ where B is a definite quaternion algebra over the field $\mathbb{Q}$. He also uses functoriality to construct residual and cuspidal cohomology classes in degree 1, the latter ones being CAP. However, his treatment of the cuspidal cohomology in degrees 2 and 3 is incomplete.

8E. *Existence of noncuspidal square-integrable cohomology classes for* GL(2, *D*). We can now formulate the implication of the construction of residual automorphic representations which occur nontrivially in the space $L^2_{\mathrm{res},J}(\mathrm{GL}(2, D))$ for the existence of square-integrable noncuspidal cohomology classes in $H^*_{(\mathrm{sq})}(\mathrm{GL}(2, D), \mathbb{C})$. For the sake of simplicity in the exposition we suppose that D does not split at all archimedean places. Taking into account some archimedean places where D splits is an easy matter; we refer to Remark 8.7.

Theorem 8.5. *Let k be a totally real algebraic number field of degree ℓ, and let D be a finite-dimensional central division algebra over k of degree $d > 1$. Let $V_D \subset V_k$ be the finite set of places of k at which D does not split. We suppose that $V_{k,\infty} \subset V_D$. Then there exists a nonvanishing cohomology class of degree $q = \ell \cdot \frac{1}{2} d(2d-3)$ in the square-integrable cohomology $H^*_{(\mathrm{sq})}(\mathrm{GL}(2, D))$. This class is noncuspidal, and it does not belong to the interior cohomology $H^*_!(\mathrm{GL}(2, D), \mathbb{C})$.*

Proof. As before, let H' denote the algebraic k-group $\mathrm{GL}(2, D)$, and let H denote the k-group $\mathrm{GL}(2d)/k$. By Section 2B, d is even, say $d = 2h$. Theorem 6.1 implies that there exist automorphic representations $\pi' = \otimes'_{v \in V_k} \pi_v$ of $H'(\mathbb{A}_k)$ which occur as irreducible constituents in the residual spectrum of $H'(\mathbb{A}_k)$ and whose component π_v at all archimedean places $v \in V_{k,\infty} \subset V_D$ is equivalent to the irreducible unitary representation $J'_{\mathbb{R}}(2, \theta')$ of $\mathrm{GL}(d, \mathbb{H})$. The continuous cohomology of $\mathrm{GL}(d, \mathbb{H})$ with coefficients in $J'_{\mathbb{R}}(2, \theta') \otimes \mathbb{C}$ is determined in Proposition 4.2. Inspecting the Poincaré polynomial as given in formula (4-16) tells us that the lowest possible degree in which this cohomology does not vanish is $q = \ell \cdot \frac{1}{2} d(2d - 3)$. Using [Rohlfs and Speh 2011, Theorem I.1–III.1], we conclude that the map

$$H^*_{\mathrm{ct}}(\mathrm{GL}(d, \mathbb{H}), J'_{\mathbb{R}}(2, \theta') \otimes \mathbb{C}) \to H^*_{(\mathrm{sq})}(H', \mathbb{C}) \tag{8-12}$$

induced by $\pi \hookrightarrow L^2_{\mathrm{disc}, J}(G) \to \mathcal{A}(G)$ is injective in the lowest degree in which the continuous cohomology $H^*_{\mathrm{ct}}(\mathrm{GL}(d, \mathbb{H}), J'_{\mathbb{R}}(2, \theta') \otimes \mathbb{C})$ is nonzero. Thus, there exists a nonvanishing cohomology class of degree $q = \ell \cdot \frac{1}{2} d(2d - 3)$ in the square-integrable cohomology $H^*_{(\mathrm{sq})}(\mathrm{GL}(2, D))$. By construction, this class is noncuspidal. Note that, as shown in [Rohlfs and Speh 2011], this nontrivial class represented by a residue of an Eisenstein series does not belong to the interior cohomology $H^*_!(\mathrm{GL}(2, D), \mathbb{C})$. Indeed, the restriction of this class to the cohomology of the boundary of the Borel–Serre compactification is nontrivial. □

Remark 8.6. The Poincaré polynomial of the representation $J'_{\mathbb{R}}(2, \theta')$ of $\mathrm{GL}(d, \mathbb{H})$ as determined in Proposition 4.2 gives precise information in which degrees the continuous cohomology $H^*_{\mathrm{ct}}(\mathrm{GL}(d, \mathbb{H}), J'_{\mathbb{R}}(2, \theta') \otimes \mathbb{C})$ is nonzero. Even in the case $d = 2$ this list contains more degrees than just the minimal degree q which matters in the assertion of Theorem 8.5. As proved in Theorem 7.3, this nontempered unitary representation $J'_{\mathbb{R}}(2, \theta')$ of $\mathrm{GL}(d, \mathbb{H})$ also appears as an archimedean component of a cuspidal automorphic representation of the adèle group $\mathrm{GL}(2, D)(\mathbb{A}_k)$ which contributes to the cuspidal cohomology In this case the contribution to the cuspidal cohomology is over the full range of degrees associated with Proposition 4.2.

In contrast, if the representation occurs as an archimedean component of a residual automorphic representation π' of $H'(\mathbb{A}_k)$, it is an important question to determine up to which degree or in which other degrees than q the cohomology attached to the automorphic representation π' contributes nontrivially to $H^*_{(\mathrm{sq})}(\mathrm{GL}(2, D))$.

In the case that the representation π is the trivial representation, obtained as the iterated residue of specific Eisenstein series attached to the constant functions on the Levi subgroup of proper parabolic subgroups of a semisimple group, a similar type of question is investigated in [Franke 2008].

Remark 8.7. Suppose that there exists a place $v \in V_{k,\infty}$ at which the central division algebra D over k splits, that is, $D_v \otimes M_d(\mathbb{R})$, thus, $H'_v \cong \mathrm{GL}(2d, \mathbb{R})$. By Remark 6.2, the corresponding local component of π as constructed is of the form $\pi_v \cong J_{\mathbb{R}}(2, \theta)$. The lowest possible degree in which the group $\mathrm{GL}(2d, \mathbb{R})$ has nontrivial continuous cohomology with coefficients is $\left(\frac{1}{2}\right)(2d^2 - d)$. To see this we proceed as in the proof of Proposition 4.2. The θ-stable parabolic subalgebra $\mathfrak{q}$ of the Lie algebra of $\mathrm{GL}(2d, \mathbb{R})$ which corresponds within the Vogan–Zuckerman classification to the irreducible representation $J_{\mathbb{R}}(2, \theta)$ has as real Lie subalgebra the algebra

$$\mathfrak{l} \cong \mathfrak{gl}(2, \mathbb{C}) \oplus \mathfrak{gl}(2, \mathbb{C}) \oplus \ldots \oplus \mathfrak{gl}(2, \mathbb{C}) \cong \mathfrak{gl}(2, \mathbb{C})^h, \tag{8-13}$$

where as before $d = 2h$. The lowest possible degree we look for is determined by the shift

$$R(\mathfrak{q}) = \dim(\mathfrak{u} \cap \mathfrak{p}_{\mathbb{C}}) = \left(\tfrac{1}{2}\right)(\dim X_{\mathrm{GL}(2d,\mathbb{R})} - \dim X_{L,u}). \tag{8-14}$$

We obtain

$$R(\mathfrak{q}) = \left(\tfrac{1}{2}\right)\left[\tfrac{1}{2}(2d(2d+1) - h \cdot 2^2\right] = \left(\tfrac{1}{2}\right)[2d^2 - d] \tag{8-15}$$

as claimed. Another way to determine this value is given in [Franke and Schwermer 1998, 5.6].

References

[Arthur and Clozel 1989] J. Arthur and L. Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*, Annals of Mathematics Studies **120**, Princeton University Press, 1989. MR Zbl

[Badulescu 2008] A. I. Badulescu, "Global Jacquet–Langlands correspondence, multiplicity one and classification of automorphic representations", *Invent. Math.* **172**:2 (2008), 383–438. MR Zbl

[Badulescu and Renard 2010] A. I. Badulescu and D. Renard, "Unitary dual of GL(n) at Archimedean places and global Jacquet–Langlands correspondence", *Compos. Math.* **146**:5 (2010), 1115–1164. MR Zbl

[Borel 1981] A. Borel, "Stable real cohomology of arithmetic groups, II", pp. 21–55 in *Manifolds and Lie groups* (Notre Dame, IN, 1980), edited by S. Murakami et al., Progr. Math. **14**, Birkhäuser, Boston, 1981. MR Zbl

[Borel and Garland 1983] A. Borel and H. Garland, "Laplacian and the discrete spectrum of an arithmetic group", *Amer. J. Math.* **105**:2 (1983), 309–335. MR Zbl

[Borel and Wallach 1980] A. Borel and N. R. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, Annals of Mathematics Studies **94**, Princeton University Press, 1980. MR Zbl

[Borel et al. 1996] A. Borel, J.-P. Labesse, and J. Schwermer, "On the cuspidal cohomology of S-arithmetic subgroups of reductive groups over number fields", *Compositio Math.* **102**:1 (1996), 1–40. MR Zbl

[Casselman 1995] W. Casselman, "Introduction to admissible representations of p-adic reductive groups", unpublished manuscript, 1995, Available at https://www.math.ubc.ca/~cass/research/pdf/p-adic-book.pdf.

[Chevalley 1951] C. Chevalley, "Deux théorèmes d'arithmétique", *J. Math. Soc. Japan* **3** (1951), 36–44. MR Zbl

[Clozel 1987] L. Clozel, "On the cuspidal cohomology of arithmetic subgroups of SL($2n$) and the first Betti number of arithmetic 3-manifolds", *Duke Math. J.* **55**:2 (1987), 475–486. MR Zbl

[Franke 1998] J. Franke, "Harmonic analysis in weighted L_2-spaces", *Ann. Sci. École Norm. Sup.* (4) **31**:2 (1998), 181–279. MR Zbl

[Franke 2008] J. Franke, "A topological model for some summand of the Eisenstein cohomology of congruence subgroups", pp. 27–85 in *Eisenstein series and applications*, edited by W. T. Gan et al., Progr. Math. **258**, Birkhäuser, Boston, 2008. MR Zbl

[Franke and Schwermer 1998] J. Franke and J. Schwermer, "A decomposition of spaces of automorphic forms, and the Eisenstein cohomology of arithmetic groups", *Math. Ann.* **311**:4 (1998), 765–790. MR Zbl

[Gan 2008] W. T. Gan, "The Saito–Kurokawa space of PGSp_4 and its transfer to inner forms", pp. 87–123 in *Eisenstein series and applications*, edited by W. T. Gan et al., Progr. Math. **258**, Birkhäuser, Boston, 2008. MR Zbl

[Grbac 2008] N. Grbac, "The residual spectrum of GL_n over a division algebra", (2008). Appendix to A. I. Badulescu, "Global Jacquet–Langlands correspondence, multiplicity one and classification of automorphic representations", *Invent. Math.* **172**:2 (2008), 383–438. MR Zbl

[Greub et al. 1976] W. Greub, S. Halperin, and R. Vanstone, *Connections, curvature, and cohomology*, vol. III, Pure and Applied Mathematics **47-III**, Academic Press, New York, 1976. MR Zbl

[Grobner 2013] H. Grobner, "Automorphic forms, cohomology and CAP representations. The case GL_2 over a definite quaternion algebra", *J. Ramanujan Math. Soc.* **28**:1 (2013), 19–48. MR Zbl

[Henniart 2012] G. Henniart, "Induction automorphe globale pour les corps de nombres", *Bull. Soc. Math. France* **140**:1 (2012), 1–17. MR Zbl

[Henniart and Herb 1995] G. Henniart and R. Herb, "Automorphic induction for GL(n) (over local non-Archimedean fields)", *Duke Math. J.* **78**:1 (1995), 131–192. MR Zbl

[Jacquet and Langlands 1970] H. Jacquet and R. P. Langlands, *Automorphic forms on* GL(2), Lecture Notes in Mathematics **114**, Springer, 1970. MR Zbl

[Jacquet and Shalika 1981] H. Jacquet and J. A. Shalika, "On Euler products and the classification of automorphic forms, II", *Amer. J. Math.* **103**:4 (1981), 777–815. MR Zbl

[Jiang 2010] D. Jiang, "On some topics in automorphic representations", pp. 177–199 in *Proceedings of the Fourth International Congress of Chinese Mathematicians* (Hangzou, 2007), edited by L. Ji et al., AMS/IP Studies in Advanced Mathematics **48**, American Mathematical Society, Providence, RI, 2010. Zbl

[Labesse and Schwermer 1986] J.-P. Labesse and J. Schwermer, "On liftings and cusp cohomology of arithmetic groups", *Invent. Math.* **83**:2 (1986), 383–401. MR Zbl

[Labesse and Schwermer 2019] J.-P. Labesse and J. Schwermer, "Central morphisms and cuspidal automorphic representations", *J. Number Theory* **205** (2019), 170–193. MR Zbl

[Langlands 1989] R. P. Langlands, "On the classification of irreducible representations of real algebraic groups", pp. 101–170 in *Representation theory and harmonic analysis on semisimple Lie groups*, edited by P. J. Sally, Jr., Math. Surveys Monogr. **31**, American Mathematical Society, Providence, RI, 1989. MR Zbl

[Mœglin and Waldspurger 1989] C. Mœglin and J.-L. Waldspurger, "Le spectre résiduel de GL(n)", *Ann. Sci. École Norm. Sup.* (4) **22**:4 (1989), 605–674. MR Zbl

[Mœglin and Waldspurger 1994] C. Mœglin and J.-L. Waldspurger, *Décomposition spectrale et séries d'Eisenstein: une paraphrase de l'écriture*, Progress in Mathematics **113**, Birkhäuser, Basel, 1994. MR Zbl

[Muto et al. 2016] M. Muto, H.-A. Narita, and A. Pitale, "Lifting to GL(2) over a division quaternion algebra, and an explicit construction of CAP representations", *Nagoya Math. J.* **222**:1 (2016), 137–185. MR Zbl

[Rohlfs 1996] J. Rohlfs, "Projective limits of locally symmetric spaces and cohomology", *J. Reine Angew. Math.* **479** (1996), 149–182. MR Zbl

[Rohlfs and Speh 2011] J. Rohlfs and B. Speh, "Pseudo Eisenstein forms and the cohomology of arithmetic groups III: Residual cohomology classes", pp. 501–523 in *On certain L-functions*, edited by J. Arthur et al., Clay Math. Proc. **13**, American Mathematical Society, Providence, RI, 2011. MR Zbl

[Schwermer 1986] J. Schwermer, "Holomorphy of Eisenstein series at special points and cohomology of arithmetic subgroups of $\mathrm{SL}_n(\mathbf{Q})$", *J. Reine Angew. Math.* **364** (1986), 193–220. MR Zbl

[Schwermer 2010] J. Schwermer, "Geometric cycles, arithmetic groups and their cohomology", *Bull. Amer. Math. Soc.* (*N.S.*) **47**:2 (2010), 187–279. MR Zbl

[Schwermer and Waldner 2011] J. Schwermer and C. Waldner, "On the cohomology of uniform arithmetically defined subgroups in $SU^*(2n)$", *Math. Proc. Cambridge Philos. Soc.* **151**:3 (2011), 421–440. MR Zbl

[Shalika 1974] J. A. Shalika, "The multiplicity one theorem for GL_n", *Ann. of Math.* (2) **100** (1974), 171–193. MR Zbl

[Speh 1983] B. Speh, "Unitary representations of $\mathrm{Gl}(n, \mathbb{R})$ with nontrivial $(\mathfrak{g}, K)$-cohomology", *Invent. Math.* **71**:3 (1983), 443–465. MR Zbl

[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**, American Mathematical Society, Providence, RI, 1997. MR Zbl

[Vogan and Zuckerman 1984] D. A. Vogan, Jr. and G. J. Zuckerman, "Unitary representations with nonzero cohomology", *Compositio Math.* **53**:1 (1984), 51–90. MR Zbl

[Weil 1956] A. Weil, "On a certain type of characters of the idèle-class group of an algebraic number-field", pp. 1–7 in *Proceedings of the International Symposium on Algebraic Number Theory* (Tokyo & Nikko, 1955), Science Council of Japan, Tokyo, 1956. MR Zbl

Received October 19, 2018.

JOACHIM SCHWERMER
FACULTY OF MATHEMATICS
UNIVERSITY OF VIENNA
VIENNA
AUSTRIA

joachim.schwermer@univie.ac.at
Current address:
MAX PLANCK INSTITUTE FOR MATHEMATICS
BONN
GERMANY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.357

INVARIANT BANACH LIMITS AND APPLICATIONS TO NONCOMMUTATIVE GEOMETRY

EVGENII SEMENOV, FEDOR SUKOCHEV,
ALEXANDR USACHEV AND DMITRIY ZANIN

A linear functional B on the space of bounded sequences l_∞ is called a Banach limit if it is positive, normalised and invariant under the shift operator. There are Banach limits which possess additional invariance properties. We prove that every Banach limit invariant under the Cesaro operator is also invariant under all dilation operators. We also prove the "continuous version" of this result and apply it to the theory of singular traces.

1. Introduction

Let $B(H)$ be an algebra of all bounded linear operators on a separable Hilbert space H.

Define the Lorentz ideal $\mathcal{M}_{1,\infty}$ in $B(H)$ [Lord et al. 2013] as

$$\mathcal{M}_{1,\infty} := \left\{ A \in B(H) : \sup_{n\in\mathbb{N}} \frac{1}{\log(2+n)} \sum_{k=0}^{n} \mu(k, A) < \infty \right\},$$

where $\mu(A)$ is a sequence of singular values of an operator A.

Now we give a definition of Dixmier traces on $\mathcal{M}_{1,\infty}$. We give it in the form which is equivalent (see [Sukochev et al. 2013, Theorem 40] and [Lord et al. 2013]) to the original definitions from [Dixmier 1966].

Definition 1. For every extended limit ω on l_∞ which is dilation-invariant (that is, $\omega(x(1), x(2), \ldots) = \omega(x(1), x(1), x(2), x(2), \ldots)$), the weight

$$\text{(1)} \qquad \mathrm{Tr}_\omega(A) := \omega\left(\frac{1}{\log(2+n)} \sum_{k=0}^{n} \mu(k, A) \right), \quad 0 \le A \in \mathcal{M}_{1,\infty},$$

extends by linearity to a nonnormal trace on $\mathcal{M}_{1,\infty}$.

MSC2010: primary 46B45, 58B34; secondary 47B37.
Keywords: Banach limit, space of bounded sequences, Cesaro operator, dilation operator.

Recall that l_∞ is the space of all bounded real sequences $x = (x(1), x(2), \ldots)$ equipped with the norm

$$\|x\|_{l_\infty} = \sup_{n \in \mathbb{N}} |x(n)|,$$

and the usual partial order. Here $\mathbb{N}$ stands for the set of natural numbers. Also recall that the extended limit on l_∞ is a linear functional which coincides with the ordinary limit on the subspace of all convergent sequence. By the Hahn–Banach theorem, extended limits exist.

These nonnormal (or singular) traces were discovered by J. Dixmier [1966] and are termed Dixmier traces. Dixmier traces play a key role in Alain Connes' noncommutative geometry — they are used to define an integral in the noncommutative context. A computation of Dixmier traces is a highly nontrivial task. This is due to the fact that in geometrically or physically inspired examples one rarely has any information on singular values of an operator. The remedy to this is formulae relating Dixmier traces to the operator ζ-function and heat traces [Connes and Moscovici 1995; Carey et al. 2007; Lord et al. 2013; Sukochev et al. 2017]. These formulae have become central to noncommutative geometry. In particular, it was proved in [Benameur and Fack 2006, Theorem 2] that for every extended limit ω which is invariant under the logarithmic Cesaro operator M (see definition in Section 4) the formula

$$\mathrm{Tr}_\omega(A) = \lim_{t\to\infty} \frac{1}{t} \mathrm{Tr}(A^{1+\frac{1}{t}}), \quad 0 \le A \in \mathcal{M}_{1,\infty},$$

holds, provided that the limit on the right-hand side exists.

Essentially at the same time, Carey et al. proved the same formula without the assumption on the convergence of the right-hand side. However, it was proved under additional assumptions on ω. More precisely, in [Carey et al. 2003, Theorem 3.1] it was proved that

$$\mathrm{Tr}_\omega(A) = (\omega \circ \log)\left(\frac{1}{t} \mathrm{Tr}(A^{1+\frac{1}{t}})\right), \quad 0 \le A \in \mathcal{M}_{1,\infty},$$

for every extended limit ω which is M-invariant and also dilation and exponentiation-invariant (see definitions in Section 4).

A priori the latter formula is proved for a significantly smaller set of Dixmier traces than the former. Our initial motivation was to compare these two classes of traces. Surprisingly, it turned out that these two classes coincide. To tackle the problem about the classes of Dixmier traces we consider the corresponding classes of extended limits.

Definition 2. A linear functional $B \in l_\infty^*$ is said to be a Banach limit if

(1) $B \geqslant 0$, that is, $Bx \geqslant 0$ for every $x \geqslant 0$,

(2) $B\mathbb{1} = 1$, where $\mathbb{1} = (1, 1, \ldots)$,

(3) $B(Tx) = B(x)$ for every $x \in l_\infty$, where T is a translation operator, that is, $T(x(1), x(2), \ldots) = (0, x(1), x(2), \ldots)$.

We denote the set of all Banach limits by $\mathfrak{B}$. It is easy to see that $\mathfrak{B}$ is a closed convex set on the unit sphere of l_∞^*. It follows directly from the definition of Banach limit that

$$\liminf_{n\to\infty} x(n) \leqslant Bx \leqslant \limsup_{n\to\infty} x(n)$$

for every $x \in \ell_\infty,\ B \in \mathfrak{B}$. In particular, $Bx = \lim_{n\to\infty} x(n)$ for every convergent sequence $x \in l_\infty$. Thus, Banach limits are indeed the extended limits.

The existence of Banach limits was established by S. Mazur. It first appeared in the book of S. Banach [1932]. Since then Banach limits have proved to be a useful tool in functional analysis.

Since we are motivated by the problem involving Dixmier traces generated by extended limits with additional invariance properties, it is reasonable to consider subclasses of Banach limits which posses some additional invariance properties (besides the translation invariance). To the best of our knowledge the existence of such invariant Banach limits was firstly established by W. Eberlein [1950]. Recently, invariant Banach limits have been used to study the asymptotics of the Lebesgue constants of the Walsh system [Astashkin and Semenov 2016]. Let us introduce the operators of interest.

Let the Cesaro operator $C : l_\infty \to l_\infty$ be given by the formula

$$(Cx)(n) = \frac{1}{n}\sum_{k=1}^{n} x(k), \quad n \geq 1,\ x \in l_\infty.$$

A Banach limit B is said to be Cesaro-invariant if $B = B \circ C$. The set of all Cesaro-invariant Banach limits is denoted by $\mathfrak{B}(C)$. A thorough study of such Banach limits can be found in [Semenov and Sukochev 2010].

For every $m \in \mathbb{N}$, let the dilation operator $\sigma_m : l_\infty \to l_\infty$ be given by the formula

$$(\sigma_m x)(n) = x\left(\left\lceil \frac{n}{m} \right\rceil\right), \quad n \geq 1,\ x \in l_\infty.$$

Similarly, a Banach limit B is said to be dilation-invariant if $B = B \circ \sigma_m$ for some $m \in \mathbb{N}$. The set of all σ_m-invariant Banach limits is denoted by $\mathfrak{B}(\sigma_m)$.

Using the Markov–Kakutani theorem it was shown in [Dodds et al. 2003, Theorem 4.2] that there exists a Banach limit which is invariant under the Cesaro operator and all dilation operators $\sigma_m,\ m \in \mathbb{N}$. It is known that for every $m \in \mathbb{N},\ m \geq 2$ there exists $B \in \mathfrak{B}(\sigma_m)$ such that $B \notin \mathfrak{B}(C)$ [Alekhno et al. 2016]. However, the question of whether Cesaro-invariance of the Banach limit implies its dilation invariance was left open in [Alekhno et al. 2016]. Theorem 3 is the main result of the paper.

Theorem 3. $\mathfrak{B}(C) \subset \mathfrak{B}(\sigma_m)$ *for all* $m \in \mathbb{N}$.

This theorem tells us that every Cesaro-invariant Banach limit is in fact dilation-invariant. Having this result at hand, we prove in Proposition 13 that every M-invariant extended limit is in fact dilation- and exponentiation-invariant.

There is another important and surprising consequence of Theorem 3.

It was proved in [Sukochev et al. 2014] (and in [Gayral and Sukochev 2014] in a more general context) that if one considers P_α-invariant extended limits (instead of dilation-invariant) in the formula (1), then one obtains a proper subclass of Dixmier traces. This subclass is denoted by $\mathcal{D}_P$. The paper [Sukochev and Usachev 2015] has made the first attempt to provide a condition which guarantees that all traces from the class $\mathcal{D}_P$ take the same value (on an operator from the weak-$\mathcal{L}_1$, which is a proper subset of $\mathcal{M}_{1,\infty}$). However, that condition is somewhat cumbersome and is difficult to verify (see [Sukochev and Usachev 2015, Theorem 3.9]).

As a consequence of Theorem 3 we prove that the classes $\mathcal{D}_M$ (of all Dixmier traces generated by M-invariant extended limits) and $\mathcal{D}_P$ coincide (on the weak-$\mathcal{L}_1$). This fact enables a neat characterisation of the class $\mathcal{D}_P$ in terms of the Cesaro operator C (see Corollary 19).

The proof of Theorem 3 is based on the key estimate established in Proposition 4.

In Section 2, we prove Proposition 4. In Section 3, we finalise the proof of Theorem 3. In Section 4 we establish the "continuous version" of Theorem 3. In Section 5 we apply the results of Section 4 to the study of singular traces.

2. Proof of the key estimate

The idea behind the proof of Proposition 4 is to construct a bijection ϕ on the set of integers from 1 to ij, with a certain control of $\phi(l)$ in terms of $l \in \{1, \dots, ij\}$ (see Lemma 8).

For $x \in l_\infty$, we set

$$\alpha(x) = \limsup_{i\to\infty} \max_{0\le j\le i} |x(i+j) - x(i)|.$$

Define a function $h : [0, 1] \to \left[0, \frac{1}{2}\right]$ by setting

$$h(t) = \tfrac{1}{2}\sum_{k=1}^{\infty} \frac{\min\{kt, 1\}}{2^k}, \qquad t \in [0, 1]. \tag{2}$$

The following result contains the key estimate for the seminorm α on $l_\infty/\{c\mathbb{1}\}$.

Proposition 4. *If* $x \in l_\infty$ *is such that* $0 \le x \le 1$, *then*

$$\alpha(Cx) \le h(\alpha(x)).$$

Notation 5. For $i \ge 5$, let $r = 1 + \lfloor \log_2(i) \rfloor$.

(1) Let $P_{r+1} = \{1\}$. For $1 \le k \le r$, let

$$P_k = \left\{ m \in \{1, \dots, i\} : \left\lfloor \frac{i}{2^k} \right\rfloor + 2 \le m \le \left\lfloor \frac{i}{2^{k-1}} \right\rfloor + 1 \right\}.$$

(2) Define a function F on $\{1, \dots, i\}$ by setting

$$F(m) = 2^k m, \quad m \in P_k,\ 1 \le k \le r+1.$$

(3) For $1 \le j \le i$, set

$$a(l) = \left\lceil \frac{l}{j} \right\rceil, \quad b(l) = i + \left\lceil \frac{l}{i} \right\rceil, \quad 1 \le l \le ij.$$

(4) Here and throughout the paper symbols $\lfloor\ \rfloor$ and $\lceil\ \rceil$ stand for rounding a real number down and up, respectively.

We need another auxiliary result.

Lemma 6. *If* $G \subset \{1, \dots, i\}$, *then*

$$\max_{m \in G} F(m) \ge i + |G|.$$

Proof. Let

$$D = \{1 \le m \le i : F(m) \le l\}, \quad l = \max_{m \in G} F(m).$$

Since $F(m) \le l$ for every $m \in G$, it follows that $G \subset D$.

If $l \ge 2i$, then there is nothing to prove. Further, we assume that $l < 2i$. This implies $1 \notin D$. Indeed, if $1 \in D$, then $2^{r+1} = F(1) \le l$. However,

$$2^{r+1} = 2 \cdot 2^{1+\lfloor \log_2(i) \rfloor} \ge 2i.$$

We now write

$$|D| = \sum_{k=1}^{r} |D \cap P_k|.$$

If $m \in D \cap P_k,\ 1 \le k \le r$, then $2^k m \le l$ and, so, $m \le l/2^k$. Thus,

$$\left\lfloor \frac{i}{2^k} \right\rfloor + 2 \le m \le \left\lfloor \frac{l}{2^k} \right\rfloor.$$

Therefore,

$$|D \cap P_k| = \left| \left\{ \left\lfloor \frac{i}{2^k} \right\rfloor + 2, \dots, \left\lfloor \frac{l}{2^k} \right\rfloor \right\} \right| = \left\lfloor \frac{l}{2^k} \right\rfloor - \left\lfloor \frac{i}{2^k} \right\rfloor - 1, \quad 1 \le k \le r.$$

Hence,

$$|D| = \sum_{k=1}^{r} |D \cap P_k| = \sum_{k=1}^{r} \left(\left\lfloor \frac{l}{2^k} \right\rfloor - \left\lfloor \frac{i}{2^k} \right\rfloor - 1 \right)_+ \le \sum_{k=1}^{r} \frac{l-i}{2^k} \le l - i.$$

Thus,

$$|G| \le |D| \le l - i = \max_{m \in C} F(m) - i.$$

□

To proceed further we need the following result from graph theory [Berge 2001, Chapter 10].

Theorem 7 (marriage theorem). *Let Γ_1 and Γ_2 be finite sets such that $|\Gamma_1| = |\Gamma_2|$. For $A \subset \Gamma_1$, set*

$$M(A) = \{l' \in \Gamma_2 : l' \textit{ is adjacent to some } l \in A\}.$$

If, for all $A \subset \Gamma_1$, we have $|M(A)| \geq |A|$, then there exists a bijection $\phi : \Gamma_1 \to \Gamma_2$ such that every $l \in \Gamma_1$ is linked to $\phi(l)$.

The following result uses the above theorem to construct a bijection on the set of integers from 1 to ij.

Lemma 8. *For $1 \leq j \leq i$ and $r = 1 + \lfloor \log_2(i) \rfloor$, there exist*

(1) *a partition $\{A_k\}_{k=1}^{r+1}$ of the set $\{1, \ldots, ij\}$ such that*

$$|A_k| \leq \frac{ij}{2^k} + j \quad \textit{for all } 1 \leq k \leq r+1.$$

(2) *a bijection $\phi : \{1, \ldots, ij\} \to \{1, \ldots, ij\}$ such that*

$$a(l) \leq b(\phi(l)) \leq 2^k a(l) \quad \textit{for all } l \in A_k \textit{ for all } 1 \leq k \leq r+1.$$

Proof. **Step 1:** For $1 \leq k \leq r+1$, set

$$A_k = \{1 \leq l \leq ij : a(l) \in P_k\}.$$

For every $1 \leq k \leq r+1$, we have $|A_k| = j|P_k|$. For $1 \leq k \leq r$, we have

$$|A_k| = j|P_k| \leq j \cdot \left(\left\lfloor \frac{i}{2^{k-1}} \right\rfloor - \left\lfloor \frac{i}{2^k} \right\rfloor \right) \leq \frac{ij}{2^k} + j.$$

For $k = r+1$, we have $|A_{r+1}| = j$. This proves the first assertion.

Step 2: Let

$$\Gamma_1 = \Gamma_2 = \{1, \ldots, ij\}.$$

We connect $l \in \Gamma_1$ and $l' \in \Gamma_2$ if

$$b(l') \leq F(a(l)).$$

For a fixed $A \subset \Gamma_1$, let

$$G = \{m \in \{1, \ldots, i\} : m = a(l) \text{ for some } l \in A\}.$$

For every $m \in G$, there exist at most j elements $l \in A$ such that $a(l) = m$. In particular, we have $|A| \leq j|G|$. We have

$$M(A) = \{l' \in \Gamma_2 : b(l') \leq F(m) \text{ for some } m \in G\}.$$

In the sequence b, every value is repeated exactly i times. It is immediate that

$$|M(A)| = i \cdot |\{1 \le m' \le j : m' + i \le F(m) \text{ for some } m \in G\}|.$$

In other words, we have

$$|M(A)| = i \cdot \min\{j, \max_{m \in C} F(m) - i\}.$$

By Lemma 6, we have

$$|M(A)| \ge i \cdot \min\{j, |G|\}.$$

Since $ij \ge j|G|$ and $i|G| \ge j|G|$, it follows that

$$|M(A)| \ge j|G| \ge |A|.$$

By the marriage theorem, there exists a bijection $\phi : \Gamma_1 \to \Gamma_2$ such that every $l \in \Gamma_1$ is adjacent to $\phi(l) \in \Gamma_2$. That is, we have

$$b(\phi(l) \le F(a(l)), \quad 1 \le l \le ij.$$

If $l \in A_k$, then $a(l) \in P_k$ and

$$b(\phi(l)) \le 2^k a(l), \quad l \in A_k.$$

This proves the right-hand side inequality in the second assertion. Noting that $a(l) \le i \le b(\phi(l))$ for all $l \in \{1, \dots, ij\}$, we complete the proof. □

We need another auxiliary result, which roughly tells us what happens to the value of the functional α if one dilates the interval over which the maximum is taken.

Lemma 9. *Let $y \in l_\infty$ be such that*

$$\sup_{i \ge 1} \max_{0 \le j \le i} |y(i+j) - y(i)| \le \alpha.$$

If $l_1 \le l_2 \le 2^k l_1$, then

$$|y(l_1) - y(l_2)| \le k\alpha.$$

Proof. If $l_1 \le l_2 \le 2^k l_1$, then there exists $0 \le k' < k$ such that $2^{k'} l_1 \le l_2 \le 2^{k'+1} l_1$. We have

$$\begin{aligned} y(l_2) - y(l_1) &= (y(l_2) - y(2^{k'} l_1)) + (y(2^{k'} l_1) - y(l_1)) \\ &= (y(l_2) - y(2^{k'} l_1)) + \sum_{p=0}^{k'-1} (y(2^{p+1} l_1) - y(2^p l_1)). \end{aligned}$$

Therefore,

$$|y(l_1) - y(l_2)| \le |y(l_2) - y(2^{k'} l_1)| + \sum_{p=0}^{k'-1} |y(2^{p+1} l_1) - y(2^p l_1)|.$$

By assumption on y, we have

$$|y(l_2) - y(2^{k'} l_1)| \leq \alpha, \quad |y(2^{p+1} l_1) - y(2^p l_1)| \leq \alpha, \quad 0 \leq p < k'.$$

Therefore,

$$|y(l_1) - y(l_2)| \leq \alpha + \sum_{p=0}^{k'-1} \alpha = (k'+1)\alpha \leq k\alpha. \qquad \square$$

Proof of Proposition 4. Let $\alpha > \alpha(x)$. Fix $n \geq 0$ such that

$$\sup_{i \geq n} \max_{0 \leq j \leq i} |x(i+j) - x(i)| \leq \alpha.$$

Set $y(i) = x(i)$ for $i \geq n$ and $y(i) = x(n)$ for $i \leq n$. We have

$$\sup_{i \geq 1} \max_{0 \leq j \leq i} |y(i+j) - y(i)| \leq \alpha.$$

Note that we still have $0 \leq y \leq 1$.

If $0 \leq j \leq i$, then

$$\begin{aligned} (Cy)(i+j) - (Cy)(i) &= \frac{1}{i+j} \sum_{k=1}^{i+j} y(k) - \frac{1}{i} \sum_{k=1}^{i} y(k) \\ &= \frac{1}{i(i+j)} \left(i \sum_{k=1}^{i+j} y(k) - (i+j) \sum_{k=1}^{i} y(k) \right) \\ &= \frac{1}{i(i+j)} \left(i \sum_{k=i+1}^{i+j} y(k) - j \sum_{k=1}^{i} y(k) \right). \end{aligned}$$

It follows from the definition of a and b that

$$j \sum_{k=1}^{i} y(k) = \sum_{l=1}^{ij} y(a(l)) = \sum_{k=1}^{r+1} \sum_{l \in A_k} y(a(l)),$$

$$i \sum_{k=i+1}^{i+j} y(k) = \sum_{l=1}^{ij} y(b(l)) = \sum_{l=1}^{ij} y(b(\phi(l))) = \sum_{k=1}^{r+1} \sum_{l \in A_k} y(b(\phi(l))),$$

where the last equality is given by Lemma 8(1), and the second-last equality is given by Lemma 8(2). Therefore,

$$\left| i \sum_{k=i+1}^{i+j} y(k) - j \sum_{k=1}^{i} y(k) \right| \leq \sum_{k=1}^{r+1} \sum_{l \in A_k} |y(a(l)) - y(b(\phi(l)))|. \tag{3}$$

By Lemma 8(2) and Lemma 9, we have

$$|y(a(l)) - y(b(\phi(l)))| \leq k\alpha, \quad l \in A_k, \ 1 \leq k \leq r+1.$$

On the other hand, we have $0 \le y \le 1$. Hence,

$$|y(a(l)) - y(b(\phi(l)))| \le 1, \quad l \in A_k,\ 1 \le k \le r+1.$$

Combining these two estimates, we infer that

$$\sum_{l \in A_k} |y(a(l)) - y(b(\phi(l)))| \le |A_k| \cdot \min\{k\alpha, 1\}, \quad 1 \le k \le r+1.$$

Substituting this formula into (3), we obtain

$$\begin{aligned}\left| i \sum_{k=i+1}^{i+j} y(k) - j \sum_{k=1}^{i} y(k) \right| &\le \sum_{k=1}^{r+1} |A_k| \cdot \min\{k\alpha, 1\} \\ i &\le \sum_{k=1}^{r+1} \left(\frac{ij}{2^k} + j \right) \min\{k\alpha, 1\} \\ &\le ij \sum_{k=1}^{r+1} \frac{\min\{k\alpha, 1\}}{2^k} + j \sum_{k=1}^{r+1} 1 \\ &\le i(j+i) \cdot \tfrac{1}{2} \sum_{k=1}^{\infty} \frac{\min\{k\alpha, 1\}}{2^k} + 2i \log_2(i),\end{aligned}$$

where the second inequality comes from Lemma 8(1).

In other words, we have

$$|(Cy)(i) - (Cy)(i+j)| \le \frac{2i \log_2(i)}{i^2} + \tfrac{1}{2} \sum_{k=1}^{\infty} \frac{\min\{k\alpha, 1\}}{2^k}.$$

Therefore,

$$\alpha(Cx) = \alpha(Cy) \le \tfrac{1}{2} \sum_{k=1}^{\infty} \frac{\min\{k\alpha, 1\}}{2^k} = h(\alpha).$$

Since $\alpha > \alpha(x)$ is arbitrary, the assertion follows. □

3. Proof of the main result

The following lemma is the key technical component of the proof of Theorem 11.

Lemma 10. *Let* $h : [0, 1] \to [0, 1]$ *be as in* (2).

(1) *For all* $t \in (0, 1]$, *we have* $h(t) < t$.

(2) *For all* $t \in [0, 1]$, *we have* $h_n(t) \to 0$ *as* $n \to \infty$, *where* $h_0 = h$ *and* $h_n = h \circ h_{n-1}$ *for* $n = 1, 2, \ldots$.

Proof. For $k > \frac{1}{\alpha}$, we have

$$\min\left\{k, \frac{1}{\alpha}\right\} < k.$$

Therefore,

$$\sum_{k=1}^{\infty} \frac{\min\{k, \frac{1}{\alpha}\}}{2^k} < \sum_{k=1}^{\infty} \frac{k}{2^k} = 2.$$

Multiplying by $\frac{1}{2}\alpha$, we obtain the first part of the assertion.

By the first part and the definition of h_n, the sequence $\{h_n(t)\}_{n\geq 0}$ is monotone and, hence, convergent. Denote the limit by s. By continuity of h, we have

$$\begin{aligned} h(s) &= h(\lim_{n\to\infty} h_n(t)) = \lim_{n\to\infty} h(h_n(t)) \\ &= \lim_{n\to\infty} h_{n+1}(t) = s. \end{aligned}$$

By the first part, this inequality is impossible unless $s = 0$. This completes the second part of the assertion. □

It is possible to construct $x \in \{0, 1\}^{\mathbb{N}}$ such that $\liminf_{m\to\infty}(C^n x)(m) = 0$ and $\limsup_{m\to\infty}(C^n x)(m) = 1$ for every $n \in \mathbb{N}$. The following result describes a certain regularity property of the Cesaro operator.

Theorem 11. *If $x \in l_\infty$, then*

$$\alpha(C^n x) \to 0, \quad n \to \infty.$$

Proof. It follows directly from the definition of the functional α that $\alpha(x+c\mathbb{1}) = \alpha(x)$ and $\alpha(cx) = |c|\alpha(x)$ for every constant $c \in \mathbb{R}$.

Thus, by considering

$$y = \frac{x + \|x\|_\infty}{2\|x\|_\infty}$$

instead of x, we may assume without loss of generality that $0 \leq x \leq 1$.

By Proposition 4, we have

$$\alpha(Cx) \leq h(\alpha(x)).$$

Repeating this inequality n times, we obtain

$$\alpha(C^n x) \leq h_n(\alpha(x)), \quad n \geq 1.$$

By Lemma 10, we have

$$h_n(t) \to 0, \quad n \to \infty \text{ for every } t \in [0, 1].$$

This completes the proof. □

Proof of Theorem 3. Let $B = B \circ C$. We have

$$|B(x) - B(\sigma_m x)| = |B(C^n x) - B(C^n \sigma_m x)|.$$

Much as in [Semenov and Sukochev 2010, Lemma 16], it can be shown that for all $m, n \geq 1$ we have

$$C^n \sigma_m - \sigma_m C^n : l_\infty \to c_0.$$

Thus,

$$|B(x) - B(\sigma_m x)| = |(B \circ (1 - \sigma_m))(C^n x)|.$$

Therefore, we have

$$\begin{aligned} |B(x) - B(\sigma_m x)| &\leq \limsup_{k \to \infty} \Big| (C^n x)(k) - (C^n x)\Big(\Big\lceil \frac{k}{m} \Big\rceil\Big) \Big| \\ &\leq \limsup_{i \to \infty} \sup_{i \leq j \leq mi} |(C^n x)(i) - (C^n x)(j)| \\ &\leq \limsup_{i \to \infty} \sup_{i \leq j \leq 2^m i} |(C^n x)(i) - (C^n x)(j)|. \end{aligned}$$

Arguing as in Lemma 9, we have that

$$\limsup_{i \to \infty} \sup_{i \leq j \leq 2^m i} |(C^n x)(i) - (C^n x)(j)| \leq m \limsup_{i \to \infty} \sup_{i \leq j \leq 2i} |(C^n x)(i) - (C^n x)(j)|.$$

In this inequality, the right-hand side is simply $m\alpha(C^n x)$. It follows that

$$|B(x) - B(\sigma_m x)| \leq m\alpha(C^n x).$$

Passing $n \to \infty$ and using Theorem 11, we obtain that

$$B(x) = B(\sigma_m x).$$

Since $x \in l_\infty$ is arbitrary, it follows that $B = B \circ \sigma_m$. Since $m \geq 2$ is arbitrary,

$$B \in \bigcap_{m=2}^{\infty} \mathfrak{B}(\sigma_m)$$

and, thus,

$$\mathfrak{B}(C) \subset \bigcap_{m=2}^{\infty} \mathfrak{B}(\sigma_m). \qquad \square$$

It can be proved that the above inclusion is strict.

4. A continuous case

Let $L_\infty = L_\infty(0, \infty)$ be the space of all (classes of) Lebesgue measurable bounded functions on $(0, \infty)$ with the uniform norm. Let $C_0 \subset L_\infty$ be a subclass of functions vanishing at ∞.

For $l, \alpha > 0$ define operators T_l, σ_α, H on L_∞ as

$$(T_l f)(t) = f(t + l), \quad (\sigma_\alpha f)(t) = f\Big(\frac{t}{\alpha}\Big), \quad (Hf)(t) = \frac{1}{t}\int_0^t f(s)ds, \quad f \in L_\infty.$$

The notions of extended limits and their translation, dilation and H-invariance are defined in the same way as for discrete settings (see, e.g., [Carey et al. 2003; Sukochev et al. 2013]).

Let us also define the embedding operator $\pi : l_\infty \to L_\infty$ and the natural expectation operator $E : L_\infty \to l_\infty$ as follows:

$$(\pi x)(t) = x(k) \quad \text{for } t \in (k, k+1]i,\ k \geq 0;$$

$$(Ef)(n) = \int_n^{n+1} f(s)\,ds.$$

First, we prove the analogue of Theorem 3 for the continuous case.

Proposition 12. *Every H-invariant extended limit on L_∞ is σ_α-invariant for every $\alpha > 0$ and T_l-invariant for every $l > 0$.*

Proof. Let B be an H-invariant Banach limit on L_∞. It follows from [Semenov et al. 2015, Lemma 5.10] that the functional $B \circ \pi$ on l_∞ is a Cesaro-invariant Banach limit. Hence, by Theorem 3, $B \circ \pi$ is σ_m-invariant for every $m \in \mathbb{N}$.

Direct verification shows that the operator $H \circ \pi \circ E - H$ maps L_∞ into C_0 and $H \circ \sigma_\alpha = \sigma_\alpha \circ H$ for every $\alpha > 0$. Using these facts together with H-invariance of B we obtain

$$\begin{aligned} B \circ \sigma_m = B \circ H \circ \sigma_m &= B \circ \sigma_m \circ H \\ &= B \circ \sigma_m \circ H \circ \pi \circ E \\ &= B \circ H \circ \sigma_m \circ \pi \circ E. \end{aligned}$$

Using that $\sigma_m \circ \pi = \pi \circ \sigma_m$ for every $m \in \mathbb{N}$, H-invariance of B and σ_m-invariance of $B \circ \pi$ we obtain

$$\begin{aligned} B \circ \sigma_m &= B \circ \pi \circ \sigma_m \circ E \\ &= B \circ \pi \circ E \\ &= B \circ H \circ \pi \circ E \\ &= B \circ H = B. \end{aligned}$$

Hence, B is σ_m-invariant for every $m \in \mathbb{N}$.

Since $\sigma_m \circ \sigma_{1/m} = \mathrm{Id}$, it follows that

$$B \circ \sigma_{1/m} = B \circ \sigma_m \circ \sigma_{1/m} = B$$

and, so, B is $\sigma_{1/m}$-invariant for every $m \in \mathbb{N}$. Next, since $\sigma_m \circ \sigma_{1/n} = \sigma_{m/n}$, it follows that B is σ_q-invariant for every positive $q \in \mathbb{Q}$.

Further, for every $\alpha > 0$, positive $q \in \mathbb{Q}$ and every $f \in L_\infty$, using σ_q- and H-invariance of B, we obtain

$$\begin{aligned}|(B \circ \sigma_\alpha)(f) - B(f)| &= |(B \circ \sigma_\alpha)(f) - (B \circ \sigma_q)(f)| \\ &= |(B \circ (H \circ \sigma_\alpha - H \circ \sigma_q))(f)| \\ &\leq \|(H \circ \sigma_\alpha - H \circ \sigma_q)(f)\| \\ &= \sup_{t>0} \left| \frac{\alpha}{t} \int_0^{\frac{t}{\alpha}} f(s)ds - \frac{q}{t} \int_0^{\frac{t}{q}} f(s)ds \right| \\ &= \sup_{t>0} \left| \frac{\alpha - q}{t} \int_0^{\frac{t}{\alpha}} f(s)ds + \frac{q}{t} \int_{\frac{t}{\alpha}}^{\frac{t}{q}} f(s)ds \right| \\ &\leq 2\frac{|\alpha - q|}{\alpha}\|f\|.\end{aligned}$$

Letting $q \to \alpha,\ q \in \mathbb{Q}$, we obtain $|(B \circ \sigma_\alpha)(f) - B(f)| = 0$ for every $f \in L_\infty$. That is, B is σ_α-invariant for every $\alpha > 0$. This proves the first claim. The second claim follows directly from [Carey et al. 2003, Proposition 1.3(5)]. □

For $\alpha > 0$ define operators P_α and M on L_∞ as

$$(P_\alpha f)(t) = f(t^\alpha), \quad (Mf)(t) = \frac{1}{\log(1+t)} \int_0^t f(s)\frac{ds}{s}.$$

It was proved in [Carey et al. 2003, Theorem 1.5] that there exists an extended limit on L_∞ which is M-, σ_α- and P_α-invariant for every $\alpha > 0$. Proposition 13 and Corollary 15 below are refinements of this theorem.

Proposition 13. *Every M-invariant extended limit on L_∞ is P_α-invariant for every $\alpha > 0$ and σ_α-invariant for every $\alpha > 0$.*

Proof. It follows from [Carey et al. 2003, Proposition 1.4] that for every M-invariant extended limit γ, the extended limit $\gamma \circ \log$ defined on L_∞ by the formula

$$(\gamma \circ \log)(t \mapsto f(t)) := \gamma(t \mapsto f(\log t))$$

is H-invariant. Thus, by Proposition 12, $\gamma \circ \log$ is σ_α-invariant for every $\alpha > 0$. Again, by [Carey et al. 2003, Proposition 1.4] the functional γ is P_α-invariant for every $\alpha > 0$. The second claim follows from directly from [Carey et al. 2003, Proposition 1.3(6)]. □

The following results show that there are infinitely many M- and H-invariant extended limits.

Proposition 14. *The cardinality of the set of extreme points of the set of all M-invariant extended limits on L_∞ and that of all H-invariant extended limits on L_∞ is $2^{2^{\mathbb{N}}}$.*

Proof. It was proved in [Alekhno et al. 2018, Theorem 4.7] that the cardinality of the set of all extreme points of $\mathfrak{B}(C)$ equals $2^{2^{\mathbb{N}}}$. It is easy to see that for every $B \in \mathfrak{B}(C)$ the functional $B \circ P$ is an H-invariant extended limit on L_∞. Since the operator P is linear, it follows that if B is an extreme point of the set $\mathfrak{B}(C)$, then $B \circ P$ is an extreme point of the set of all H-invariant extended limits on L_∞. Thus, the cardinality of the set of all extreme points of the set of all H-invariant extended limit on L_∞ equals $2^{2^{\mathbb{N}}}$.

It is known (see, e.g., [Sukochev et al. 2013, Remark 31]) that if γ is an H-invariant extended limit on L_∞, then the extended limit $\gamma \circ \exp$ defined on L_∞ by the formula

$$(\gamma \circ \exp)(t \mapsto f(t)) := \gamma(t \mapsto f(e^t))$$

is M-invariant.

Since the mapping $\gamma \mapsto \gamma \circ \exp$ preserves the extreme points, we conclude that the cardinality of the set of extreme points of the set of all M-invariant extended limits on L_∞ is $2^{2^{\mathbb{N}}}$. □

The existence of an M-, σ_α- and P_α-invariant extended limit was proved in [Carey et al. 2003, Theorem 1.5]. The following corollary shows that this extended limit is not unique and, moreover, the set of such extended limits has the largest possible cardinality.

Corollary 15. *The cardinality of the set of all extended limits on L_∞ which are M-, σ_α- and P_α-invariant is $2^{2^{\mathbb{N}}}$.*

Proof. By [Sukochev et al. 2013, Lemma 7] the operator $M \circ \sigma_\alpha - M$ maps L_∞ into C_0. Thus, any M-invariant extended limit is automatically σ_α-invariant. Combining this with Proposition 13, we conclude that the set in the statement of the corollary contains the set of all M-invariant extended limits. Now, the result follows from Proposition 14. □

5. Applications

The result of Proposition 13 allows us to reduce M- and P_α-invariance to a single assumption of M-invariance.

The following theorem extends Theorem 2 in [Benameur and Fack 2006], relaxing the condition of the convergence of the right-hand side.

Theorem 16. *If the extended limit ω on L_∞ is such that $\omega = \omega \circ M$, then*

$$\mathrm{Tr}_\omega(A) = (\omega \circ \log)\left(\frac{1}{t}\mathrm{Tr}(A^{1+\frac{1}{t}})\right), \quad 0 \leq A \in \mathcal{M}_{1,\infty}.$$

Proof. By Proposition 13, we have that $\omega = \omega \circ P_\alpha$ for all $\alpha > 0$. Hence, the assertion follows from [Carey et al. 2003, Theorem 3.1]. □

Now we turn to the comparison of the subclasses $\mathcal{D}_M$ and $\mathcal{D}_P$ of Dixmier traces (defined at the end of the Introduction).

A straightforward consequence of Proposition 13 is the following result.

Corollary 17. $\mathcal{D}_M \subset \mathcal{D}_P$ *on* $\mathcal{M}_{1,\infty}$.

Besides the Lorentz ideal, the weak-trace class ideal

$$\mathcal{L}_{1,\infty} := \left\{ A \in B(H) : \sup_{n\in\mathbb{N}} (n+1)\mu(n, A) < \infty \right\}$$

is also widely used in noncommutative geometry. This ideal also admits Dixmier traces and they are also defined by the formula (1). Thus, every Dixmier trace on $\mathcal{L}_{1,\infty}$ is a restriction of that from $\mathcal{M}_{1,\infty}$.

The following result shows the coincidence of $\mathcal{D}_M$ and $\mathcal{D}_P$ in the setting of the weak-$\mathcal{L}_1$ ideal.

Theorem 18. $\mathcal{D}_M = \mathcal{D}_P$ *on* $\mathcal{L}_{1,\infty}$.

Proof. One inclusion follows from Corollary 17 and the other one was established in [Sukochev and Usachev 2015, Theorem 3.10]. ☐

As a corollary we obtain the strengthening of [Sukochev and Usachev 2015, Theorem 3.9]. It provides the characterisation of the class $\mathcal{D}_P$.

Corollary 19. *All traces from the class* $\mathcal{D}_P$ *take the same value on the operator* $A \in \mathcal{L}_{1,\infty}$ *if and only if*

$$\lim_{m\to\infty} \liminf_{n\to\infty} (C^m x)(n) = \lim_{m\to\infty} \limsup_{n\to\infty} (C^m x)(n), \tag{4}$$

where $x := \left\{ \sum_{k=2^n-1}^{2^{n+1}-2} \mu(k, A) \right\}_{n\geq 0}$.

Proof. It was proved in [Semenov et al. 2015, Proposition 7.8] that all traces from $\mathcal{D}_M$ coincide on A if and only if the condition (4) holds. The assertion follows from Theorem 18. ☐

Acknowledgements

The authors thanks the anonymous reviewer for a careful reading of the paper and valuable comments. The work of Semenov and Usachev was supported by RNF (projects 16–11–10125 and 19–11–00197). The work of Usachev was partially supported by the Swedish Research Council Grant 2015-00137 and Marie Sklodowska Curie Actions, Cofund, Project INCA 600398. The work of Sukochev and Zanin was partially supported by the Australian Research Council.

References

[Alekhno et al. 2016] E. Alekhno, E. Semenov, F. Sukochev, and A. Usachev, "On the structure of invariant Banach limits", *C. R. Math. Acad. Sci. Paris* **354**:12 (2016), 1195–1199. MR Zbl

[Alekhno et al. 2018] E. Alekhno, E. Semenov, F. Sukochev, and A. Usachev, "Invariant Banach limits and their extreme points", *Studia Math.* **242**:1 (2018), 79–107. MR Zbl

[Astashkin and Semenov 2016] S. V. Astashkin and E. M. Semenov, "Lebesgue constants of a Walsh system and Banach limits", *Sibirsk. Mat. Zh.* **57**:3 (2016), 512–526. MR Zbl

[Banach 1932] S. Banach, *Théorie des opérations linéaires*, Monografie Matematyczne **1**, Seminarium Matematyczne Uniwersytetu Warszawskiego, Warsaw, 1932. MR Zbl JFM

[Benameur and Fack 2006] M.-T. Benameur and T. Fack, "Type II non-commutative geometry, I: Dixmier trace in von Neumann algebras", *Adv. Math.* **199**:1 (2006), 29–87. MR Zbl

[Berge 2001] C. Berge, *The theory of graphs*, Dover, Mineola, NY, 2001. MR Zbl

[Carey et al. 2003] A. Carey, J. Phillips, and F. Sukochev, "Spectral flow and Dixmier traces", *Adv. Math.* **173**:1 (2003), 68–113. MR Zbl

[Carey et al. 2007] A. L. Carey, A. Rennie, A. Sedaev, and F. Sukochev, "The Dixmier trace and asymptotics of zeta functions", *J. Funct. Anal.* **249**:2 (2007), 253–283. MR Zbl

[Connes and Moscovici 1995] A. Connes and H. Moscovici, "The local index formula in noncommutative geometry", *Geom. Funct. Anal.* **5**:2 (1995), 174–243. MR Zbl

[Dixmier 1966] J. Dixmier, "Existence de traces non normales", *C. R. Acad. Sci. Paris Sér. A-B* **262** (1966), A1107–A1108. MR Zbl

[Dodds et al. 2003] P. G. Dodds, B. de Pagter, A. A. Sedaev, E. M. Semenov, and F. A. Sukochev, "Singular symmetric functionals and Banach limits with additional invariance properties", *Izv. Ross. Akad. Nauk Ser. Mat.* **67**:6 (2003), 111–136. In Russian; translated in *Izv. Math.* **67**:6 (2003), 1187–1212. MR Zbl

[Eberlein 1950] W. F. Eberlein, "Banach–Hausdorff limits", *Proc. Amer. Math. Soc.* **1** (1950), 662–665. MR Zbl

[Gayral and Sukochev 2014] V. Gayral and F. Sukochev, "Dixmier traces and extrapolation description of noncommutative Lorentz spaces", *J. Funct. Anal.* **266**:10 (2014), 6256–6317. MR Zbl

[Lord et al. 2013] S. Lord, F. Sukochev, and D. Zanin, *Singular traces: theory and applications*, De Gruyter Studies in Mathematics **46**, De Gruyter, Berlin, 2013. MR Zbl

[Semenov and Sukochev 2010] E. M. Semenov and F. A. Sukochev, "Invariant Banach limits and applications", *J. Funct. Anal.* **259**:6 (2010), 1517–1541. MR Zbl

[Semenov et al. 2015] E. Semenov, F. Sukochev, A. Usachev, and D. Zanin, "Banach limits and traces on $\mathcal{L}_{1,\infty}$", *Adv. Math.* **285** (2015), 568–628. MR Zbl

[Sukochev and Usachev 2015] F. Sukochev and A. Usachev, "Characterization of singular traces on the weak trace class ideal generated by exponentiation-invariant extended limits", *Comment. Math.* **55**:2 (2015), 127–145. MR Zbl

[Sukochev et al. 2013] F. Sukochev, A. Usachev, and D. Zanin, "Generalized limits with additional invariance properties and their applications to noncommutative geometry", *Adv. Math.* **239** (2013), 164–189. MR Zbl

[Sukochev et al. 2014] F. Sukochev, A. Usachev, and D. Zanin, "Dixmier traces generated by exponentiation invariant generalised limits", *J. Noncommut. Geom.* **8**:2 (2014), 321–336. MR Zbl

[Sukochev et al. 2017] F. Sukochev, A. Usachev, and D. Zanin, "Singular traces and residues of the ζ-function", *Indiana Univ. Math. J.* **66**:4 (2017), 1107–1144. MR Zbl

Received March 22, 2018. Revised June 8, 2019.

EVGENII SEMENOV
VORONEZH STATE UNIVERSITY
VORONEZH
RUSSIA
nadezhka_ssm@geophys.vsu.ru

FEDOR SUKOCHEV
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEW SOUTH WALES
SYDNEY
AUSTRALIA
f.sukochev@unsw.edu.au

ALEXANDR USACHEV
DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
GOTHENBURG
SWEDEN
Current address:
SCHOOL OF MATHEMATICS AND STATISTICS
CENTRAL SOUTH UNIVERSITY
HUNAN
CHINA
a.usachev@csu.edu.cn

DMITRIY ZANIN
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEW SOUTH WALES
SYDNEY
AUSTRALIA
d.zanin@unsw.edu.au

PACIFIC JOURNAL OF MATHEMATICS
Vol. 306, No. 1, 2020
dx.doi.org/10.2140/pjm.2020.306.375

A COMBINATORIAL IDENTITY FOR THE JACOBIAN OF t-SHIFTED INVARIANTS

OKSANA YAKIMOVA

Let $\mathfrak{g}$ be a simple Lie algebra. There are classical formulas for the Jacobians of the generating invariants of the Weyl group of $\mathfrak{g}$ and of the images under the Harish-Chandra projection of the generators of the centre $\mathcal{Z}(\mathfrak{g}) \subset \mathcal{U}(\mathfrak{g})$. We present a modification of these formulas related to Takiff Lie algebras.

Introduction

Let $\mathfrak{g}$ be a simple complex Lie algebra, and $\mathfrak{h} \subset \mathfrak{g}$ be a Cartan subalgebra. Fix a triangular decomposition $\mathfrak{g} = \mathfrak{n}^- \oplus \mathfrak{h} \oplus \mathfrak{n}^+$. Let $\Delta \subset \mathfrak{h}^*$ be the corresponding root system with $\Delta^+ \subset \Delta$ being the subset of positive roots. Set $n = \operatorname{rk} \mathfrak{g} = \dim \mathfrak{h}$. Define $\rho = \frac{1}{2} \sum_{\alpha \in \Delta^+} \alpha$. For $\alpha \in \Delta^+$, let $\{f_\alpha, h_\alpha, e_\alpha\} \subset \mathfrak{g}$ be an $\mathfrak{sl}_2$-triple with $e_\alpha \in \mathfrak{g}_\alpha$. Let $W = W(\mathfrak{g}, \mathfrak{h})$ be the Weyl group of $\mathfrak{g}$.

In a basis $\{h_1, \dots, h_n\}$ for $\mathfrak{h}$, the Jacobian $\boldsymbol{J}$ of $P_1, \dots, P_n \in \mathcal{S}(\mathfrak{h}) \cong \mathbb{C}[\mathfrak{h}^*]$ is a polynomial with the following property:

$$dP_1 \wedge dP_2 \wedge \cdots \wedge dP_n = \boldsymbol{J} dh_1 \wedge \cdots \wedge dh_n.$$

Up to the sign, $\boldsymbol{J}$ is independent of the order of the elements P_i ; up to a nonzero scalar, $\boldsymbol{J}$ is independent of the choice of a basis for $\mathfrak{h}$. We set

$$J(\{P_i\}) = J(P_1, \dots, P_n) = \boldsymbol{J}.$$

Suppose $\{\hat{P}_1, \dots, \hat{P}_n\} \subset \mathcal{S}(\mathfrak{h})^W$ is a set of homogeneous generating invariants. Then

$$J(\{\hat{P}_i\}) = \prod_{\alpha \in \Delta^+} h_\alpha \quad \text{under a suitable normalisation of the elements } \hat{P}_i$$

by a classical argument, which is presented, for example, in [Humphreys 1990, Section 3.13].

Set $d_i := \deg \hat{P}_i$. Let $\mathcal{U}(\mathfrak{g}) = \bigcup_{d=0}^{\infty} \mathcal{U}_d(\mathfrak{g})$ be the canonical filtration on the enveloping algebra $\mathcal{U}(\mathfrak{g})$, let $\mathcal{Z}(\mathfrak{g})$ stand for the centre of $\mathcal{U}(\mathfrak{g})$. Then $\mathcal{Z}(\mathfrak{g})$ has a set $\{\mathcal{P}_i \mid 1 \leqslant i \leqslant n\}$ of algebraically independent generators such that $\mathcal{P}_i \in \mathcal{U}_{d_i}(\mathfrak{g})$.

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — project numbers 330450448, 404144169.
MSC2010: primary 17B20, 17B35, 17B70; secondary 22E47.
Keywords: Takiff Lie algebras, symmetric invariants, Jacobian, Harish-Chandra projection.

Let $P_i \in \mathcal{S}(\mathfrak{h}) \cong \mathbb{C}[\mathfrak{h}^*]$ be the image of $\mathcal{P}_i$ under the Harish-Chandra projection. Define $\hat{P}_i \in \mathbb{C}[\mathfrak{h}^*]$ by $\hat{P}_i(x) = P_i(x-\rho)$ for $x \in \mathfrak{h}^*$. Then $\hat{P}_i \in \mathcal{S}(\mathfrak{h})^W$; see, e.g., [Dixmier 1974, Section 7.4], and

$$J(\{P_i\}) = \prod_{\alpha\in\Delta^+} (h_\alpha + \rho(h_\alpha)) \text{ under a suitable normalisation of the elements } \mathcal{P}_i.$$

For any complex Lie algebra $\mathfrak{l}$, let $\varpi : \mathcal{S}(\mathfrak{l}) \to \mathcal{U}(\mathfrak{l})$ be the canonical symmetrisation map. Let $\mathcal{S}(\mathfrak{l})^{\mathfrak{l}}$ denote the ring of symmetric $\mathfrak{l}$-invariants. Since ϖ is an isomorphism of $\mathfrak{l}$-modules, it provides an isomorphism of vector spaces $\mathcal{S}(\mathfrak{l})^{\mathfrak{l}} \cong \mathcal{Z}(\mathfrak{l})$.

Suppose next that $\mathcal{P}_i = \varpi(H_i)$ is the symmetrisation of H_i and that $H_i \in \mathcal{S}(\mathfrak{g})^{\mathfrak{g}}$ is a homogeneous generator of degree d_i. Let $T : \mathfrak{g} \to \mathfrak{g}[t]$ be the $\mathbb{C}$-linear map sending each $x \in \mathfrak{g}$ to xt. Here $\mathfrak{g}[t]$ is the current algebra associated with $\mathfrak{g}$, where $[\xi t^a, \eta t^b] = [\xi, \eta]t^{a+b}$ for $\xi, \eta \in \mathfrak{g}$, $a, b \in \mathbb{Z}_{\geqslant 0}$. The map T extends uniquely to the commutative algebras homomorphism

$$T : \mathcal{S}(\mathfrak{g}) \to \mathcal{S}(\mathfrak{g}[t]).$$

Set $H_i^{[1]} = T(H_i)$ and $\mathcal{P}_i^{[1]} = \varpi(H_i^{[1]})$. Here $\mathcal{P}_i^{[1]} \in \mathcal{U}(t\mathfrak{g}[t])$.

The triangular decomposition of $\mathfrak{g}$ extends to $\mathfrak{g}[t]$ as $\mathfrak{g}[t] = \mathfrak{n}^-[t] \oplus \mathfrak{h}[t] \oplus \mathfrak{n}^+[t]$. Let $P_i^{[1]} \in \mathcal{S}(t\mathfrak{h}[t])$ be the image of $\mathcal{P}_i^{[1]}$ under the Harish-Chandra projection. For $h_j \in \mathfrak{h}$, let D_{h_j} be the unique derivation of $\mathcal{S}(t\mathfrak{h}[t])$ such that

$$D_{h_j}(yt^k) = kt^{k-1}\partial_{h_j} y$$

if $k \geqslant 1$ and $y \in \mathfrak{h}$. We define the Jacobian $J(\{P_i^{[1]}\}) = J(P_1^{[1]}, \dots, P_n^{[1]})$ by

$$J(\{P_i^{[1]}\}) = \det(D_{h_j} P_i^{[1]})|_{t=1}.$$

Note that $J(\{P_i^{[1]}\}) \in \mathcal{S}(\mathfrak{h})$.

Theorem 1. *Under a suitable normalisation of the elements H_i, we have the identity*

$$J(\{P_i^{[1]}\}) = \prod_{\alpha\in\Delta^+} (h_\alpha + \rho(h_\alpha) + 1).$$

Our proof of Theorem 1 interprets the zero set of $J(\{P_i^{[1]}\})$ in terms of the *Takiff Lie algebra* $\mathfrak{q} = \mathfrak{g}[u]/(u^2)$ and then uses the *extremal projector* associated with $\mathfrak{g}$; see Section 2A for the definition.

Takiff [1971, Corollary 11.3] proved that $\mathcal{S}(\mathfrak{q})^{\mathfrak{q}}$ is a polynomial ring whose Krull dimension equals $2\,\mathrm{rk}\,\mathfrak{g}$. This has started a serious investigation of these Lie algebras and their generalisations [Chari and Greenstein 2009; 2011; Khare and Ridenour 2012; Greenstein and Mazorchuk 2017], see also [Panyushev and Yakimova 2020] and references therein. Verma modules and an analogue of the Harish-Chandra homomorphism for $\mathfrak{q}$ were defined and studied in [Geoffriau 1995; Wilson 2011]. We remark that $\mathfrak{q}$-modules appearing in this paper are essentially different.

1. Several combinatorial formulas

Keep the notation of the introduction. In particular, $H_i \in \mathcal{S}(\mathfrak{g})^{\mathfrak{g}}$ stands for a homogeneous generator of degree d_i, P_i is the image of $\mathcal{P}_i = \varpi(H_i)$ under the Harish-Chandra projection, $\hat{P}_i \in \mathcal{S}(\mathfrak{h})^W$ is the $(-\rho)$-shift of P_i, i.e., $\hat{P}_i(x) = P_i(x-\rho)$ for $x \in \mathfrak{h}^*$, and $P_i^{[1]}$ is the image of $\varpi(T(H_i))$ under the Harish-Chandra projection related to $\mathfrak{g}[t]$. Also let P_i° be the highest degree component of P_i. Then $P_i^\circ = H_i|_{\mathfrak{h}}$. By the Chevalley restriction theorem, the polynomials P_i° with $1 \leqslant i \leqslant n$ generate $\mathcal{S}(\mathfrak{h})^W$. It is clear that $J(\{P_i^\circ\}) = J(\{\hat{P}_i\})$. Let a choice of $H_1, \ldots, H_n$ be such that $J(\{P_i^\circ\}) = \prod_{\alpha\in\Delta^+} h_\alpha$.

Lemma 1.1. *The highest degree component of $J(\{P_i^{[1]}\})$ is equal to $\prod_{\alpha\in\Delta^+} h_\alpha$.*

Proof. The highest degree component of $P_i^{[1]}$ is $T(P_i^\circ) \in \mathcal{S}(\mathfrak{h}t)$. Each monomial of $T(P_i^\circ)$ is of the form $(x_1 t)\cdots(x_{d_i} t)$ with $x_j \in \mathfrak{h}$ for each j. By the construction, $D_{h_j} T(P_i^\circ)|_{t=1} = \partial_{h_j} P_i^\circ$. The result follows. $\square$

In order to prove the next lemma, we need a well-known equality, namely $\prod_{i=1}^n d_i = |W|$.

Lemma 1.2. *For the constant term of the Jacobian $J(\{P_i^{[1]}\})$, we have the formula $J(\{P_i^{[1]}\})(0) = \prod_{\alpha\in\Delta^+}(\rho(h_\alpha)+1)$.*

Proof. Clearly $P_i^{[1]}|_{t=1} = P_i$. Since H_i is a homogeneous polynomial of degree d_i, the linear in $\mathfrak{h}$ part of $P_i^{[1]}$ has degree d_i in t. It follows that

$$J(\{P_i^{[1]}\})(0) = \left(\prod_{i=1}^n d_i\right) J(\{P_i\})(0) = |W| \prod_{\alpha\in\Delta^+} \rho(h_\alpha).$$

We complete the proof with a formula of Kostant which states

$$\prod_{\alpha\in\Delta^+} \frac{\rho(h_\alpha)+1}{\rho(h_\alpha)} = |W^\vee| = |W|. \tag{1-1}$$

$\square$

Remark 1.3. The Kostant formula (1-1) is a particular case of another combinatorial statement. Let $W(t) = \sum_{w\in W} t^{l(w)}$ be the Poincaré polynomial of W. Then

$$W^\vee(t) = \prod_{\alpha\in\Delta^+} \frac{t^{(\rho,\alpha^\vee)+1}-1}{t^{(\rho,\alpha^\vee)}-1};$$

see equation (34) in [Humphreys 1990, Section 3.20]. Since $\rho(h_\alpha) = (\rho, \alpha^\vee)$, evaluating at $t=1$ one gets exactly (1-1).

Example 1.4. Take $\mathfrak{g} = \mathfrak{sl}_2$ with the usual basis $\{e, h, f\}$, where $[h, e] = 2e$. Then $H = H_1 = 2ef + \frac{1}{2}h^2$, $H^{[1]} = 2(et)(ft) + \frac{1}{2}(ht)^2$, and

$$\mathcal{P}^{[1]} = \varpi(H^{[1]}) = (et)(ft) + (ft)(et) + \tfrac{1}{2}(ht)^2 = \tfrac{1}{2}(ht)^2 + ht^2 + 2(ft)(et).$$

Therefore $P^{[1]} = P_1^{[1]} = \frac{1}{2}(ht)^2 + ht^2$. Computing the partial derivative and evaluating at $t = 1$, we obtain $J(\{P^{[1]}\}) = h + 2 = h + \rho(h) + 1$.

2. Takiff Lie algebras and branching

Theorem 1 can be interpreted as a statement in representation theory of Takiff Lie algebras

$$\mathfrak{q} = \mathfrak{g} \ltimes \mathfrak{g}^{\mathrm{ab}} \cong \mathfrak{g}[u]/(u^2).$$

The first factor, the nonabelian copy of $\mathfrak{g}$, acts on $\mathfrak{g}^{\mathrm{ab}} = \mathfrak{g}u$ as a subalgebra of $\mathfrak{gl}(\mathfrak{g})$. Therefore there is the canonical embedding $\mathfrak{q} \subset \mathfrak{gl}(\mathfrak{g}) \ltimes \mathfrak{g}^{\mathrm{ab}}$. Set $\ell = \dim \mathfrak{g} + 1$. In its turn, $\mathfrak{gl}(\mathfrak{g}) \ltimes \mathfrak{g}^{\mathrm{ab}}$ can be realised as a subalgebra of $\mathfrak{gl}(\mathfrak{g} \oplus \mathbb{C}) \cong \mathfrak{gl}_\ell(\mathbb{C})$. The Lie algebra $\mathfrak{gl}_\ell(\mathbb{C})$ is equipped with the standard triangular decomposition. Let $\mathfrak{b}_\ell \subset \mathfrak{gl}_\ell(\mathbb{C})$ be the corresponding positive Borel. Recall that we have chosen a triangular decomposition $\mathfrak{g} = \mathfrak{n}^- \oplus \mathfrak{h} \oplus \mathfrak{n}^+$. Set $\mathfrak{b} = \mathfrak{h} \oplus \mathfrak{n}^+$, $\mathfrak{b}^- = \mathfrak{h} \oplus \mathfrak{n}^-$. We fix an embedding $\mathfrak{gl}(\mathfrak{g}) \ltimes \mathfrak{g}^{\mathrm{ab}} \subset \mathfrak{gl}(\mathfrak{g} \oplus \mathbb{C})$ such that $\mathfrak{b}^- \ltimes \mathfrak{g}^{\mathrm{ab}}$ lies in the opposite Borel $\mathfrak{b}_\ell^-$ and $\mathfrak{b} \subset \mathfrak{b}_\ell$.

Let $\psi : \mathcal{S}(\mathfrak{g}) \to \mathcal{S}(\mathfrak{q})$ be a derivation defined uniquely by $\psi(\xi) = \xi u$ for any $\xi \in \mathfrak{g}$. Set $\mathcal{R}_i = \varpi(\psi(H_i))$. The elements $\mathcal{R}_1, \ldots, \mathcal{R}_n \in \mathcal{U}(\mathfrak{q})$ are not necessarily central. If we assume that $d_1 = 2$, then $\mathcal{R}_1 \in \mathcal{Z}(\mathfrak{q})$. However, since both maps, ψ and ϖ, are homomorphisms of $\mathfrak{g}$-modules, each $\mathcal{R}_i$ commutes with $\mathfrak{g}$. Note that the elements $\mathcal{R}_i$ have degree 1 in $\mathfrak{g}u$. They are crucial for further considerations and our next goal is to relate them to $\mathcal{P}_i^{[1]} \in \mathcal{U}(t\mathfrak{g}[t])$.

The map ψ is also well defined for the tensor algebra of $\mathfrak{g}$, but not for $\mathcal{U}(\mathfrak{g})$, because of the following obstacle:

$$\begin{aligned} \psi(\xi_1\xi_2 - \xi_2\xi_1) &= (\xi_1 u)\xi_2 + \xi_1(\xi_2 u) - (\xi_2 u)\xi_1 - \xi_2(\xi_1 u) \\ &= [\xi_1 u, \xi_2] + [\xi_1, \xi_2 u] = 2[\xi_1, \xi_2]u \neq [\xi_1, \xi_2]u. \end{aligned}$$

The remedy is to pass to the current algebras $\mathfrak{g}[t]$ and $\mathfrak{q}[t]$.

Lemma 2.1. *There is a well-defined $\mathbb{C}$-linear map $\mathcal{T} : \mathcal{U}(t\mathfrak{g}[t]) \to \mathcal{U}(\mathfrak{q}[t])$ such that*

- $\mathcal{T}(\xi t^k) = k(\xi u)t^{k-1}$ *for each* $\xi \in \mathfrak{g}$,
- $\mathcal{T}(ab) = \mathcal{T}(a)b + a\mathcal{T}(b)$ *for all* $a, b \in \mathcal{U}(t\mathfrak{g}[t])$, *i.e.,* $\mathcal{T}$ *is a derivation.*

Proof. Take $\xi, \eta \in \mathfrak{g}$ and $k, m \geqslant 1$. Then

$$\begin{aligned} &\mathcal{T}(\xi t^k \eta t^m - \eta t^m \xi t^k) \\ &\qquad = k(\xi u)t^{k-1}\eta t^m + m\xi t^k(\eta u)t^{m-1} - m(\eta u)t^{m-1}\xi t^k - k\eta t^m(\xi u)t^{k-1} \\ &\qquad = k[\xi u, \eta]t^{k-1+m} + m[\xi, \eta u]t^{k+m-1} \\ &\qquad = (k+m)([\xi, \eta]u)t^{k+m-1} = \mathcal{T}([\xi, \eta]t^{k+m}). \end{aligned}$$

□

For a monomial $\boldsymbol{\xi} = \xi_1 \cdots \xi_d$ with $\xi_i \in \mathfrak{g}$, we have $\mathcal{T} \circ \varpi \circ T(\boldsymbol{\xi})|_{t=1} = \varpi \circ \psi(\boldsymbol{\xi})$. Hence, by the construction, $\mathcal{R}_i = \mathcal{T}(\mathcal{P}_i^{[1]})|_{t=1}$.

A word of caution: in $\mathcal{U}(\mathfrak{b}^-)(\mathfrak{n}^+u)$ and similar expressions, $(\mathfrak{n}^+u)$ stands for the subspace $\mathfrak{n}^+u \subset \mathfrak{g}^{\mathrm{ab}}$ and *not* for an ideal generated by $\mathfrak{n}^+u$. The same applies to $(\mathfrak{b}u)$, $(\mathfrak{g}u)$, etc.

Lemma 2.2. *Let $M(\lambda) = \mathcal{U}(\mathfrak{b}_\ell^-)v_\lambda$ with $\lambda \in \mathbb{C}^\ell$ be a Verma module of $\mathfrak{gl}(\mathfrak{g}\oplus\mathbb{C})$. Set $\mu = \lambda|_{\mathfrak{h}}$. There exists a nontrivial linear combination $\mathcal{R} = \sum c_i\mathcal{R}_i$ such that $\mathcal{R}v_\lambda \in \mathfrak{n}^-\mathcal{U}(\mathfrak{b}^-)(\mathfrak{n}^+u)v_\lambda$ if and only if $J(\{P_i^{[1]}\})(\mu) = 0$.*

Proof. We have

$$\mathcal{P}_i^{[1]} \in P_i^{[1]} + \mathfrak{n}^-[t]\mathcal{U}(\mathfrak{g}[t])\mathfrak{n}^+[t].$$

Accordingly, $\mathcal{R}_i = \mathcal{T}(P_i^{[1]})|_{t=1} + \mathcal{X}$, where $\mathcal{X}$ is the image of the second summand of $\mathcal{P}_i^{[1]}$. Let $X = x_1 \cdots x_r$ be a monomial appearing in $\mathcal{X}$. If $x_r \in \mathfrak{n}^+$, then $Xv_\lambda = 0$. Assume that $Xv_\lambda \neq 0$. Then necessarily $x_r \in \mathfrak{n}^+u$ and $x_1, \ldots, x_{r-1} \in \mathfrak{g}$. If $x_i \in \mathfrak{n}^+$ for some $i \leqslant (r-1)$, then we replace X by $x_1 \cdots x_{i-1}[x_i, x_{i+1}, \ldots, x_r]$. Note that here $[x_i, x_r] \in \mathfrak{n}^+u$. Applying this procedure as often as possible one replaces X by an element of $\mathcal{U}(\mathfrak{b}^-)(\mathfrak{n}^+u)$ without altering Xv_λ. Since X is an invariant of $\mathfrak{h}$, the new element lies in $\mathfrak{n}^-\mathcal{U}(\mathfrak{b}^-)(\mathfrak{n}^+u)$. Summing up,

$$\mathcal{R}_iv_\lambda \in \mathcal{T}(P_i^{[1]})|_{t=1}v_\lambda + \mathfrak{n}^-\mathcal{U}(\mathfrak{b}^-)(\mathfrak{n}^+u)v_\lambda. \tag{2-1}$$

Further

$$\mathcal{T}(P_i^{[1]})|_{t=1} = \sum_{j=1}^{n}(D_{h_j}P_i^{[1]})|_{t=1}h_ju,$$

where D_{h_j} are the same as in the introduction. Recall that they have been used in order to define $J(\{P_i^{[1]}\})$. Hence $J(\{P_i^{[1]}\})(\mu) = 0$ if and only if there is a nonzero vector $\bar{c} = (c_1, \ldots, c_n)$ such that $\mathcal{A}(\mu) = \sum c_i\mathcal{T}(P_i^{[1]})|_{t=1,\mu} = 0$. This shows that if $J(\{P_i^{[1]}\})(\mu) = 0$, then $\mathcal{R}v_\lambda \in \mathfrak{n}^-\mathcal{U}(\mathfrak{b}^-)(\mathfrak{n}^+u)v_\lambda$.

Suppose now that $\mathcal{R}v_\lambda \in \mathfrak{n}^-\mathcal{U}(\mathfrak{b}^-)(\mathfrak{n}^+u)v_\lambda \subset \mathcal{U}(\mathfrak{n}^-)(\mathfrak{n}^+u)v_\lambda$. Then

$$\mathcal{A}(\mu)v_\lambda \in \mathcal{U}(\mathfrak{n}^-)(\mathfrak{n}^+u)v_\lambda.$$

We are working with a Verma module of $\mathfrak{gl}_\ell(\mathbb{C})$; the subspaces $\mathfrak{n}^+u$ and $\mathfrak{h}u$ are contained in $\mathfrak{n}_\ell^-$, and $\mathcal{A}(\mu) \in \mathfrak{h}u$. Hence $\mathcal{A}(\mu)$ must be an element of $\mathcal{U}(\mathfrak{n}^-)(\mathfrak{n}^+u)$. At the same time, $\mathfrak{h}u \cap \mathfrak{n}^+u = 0$. Therefore $\sum_{i=1}^{n} c_i(D_{h_j}P_i^{[1]})|_{t=1,\mu} = 0$ for each j and thus $J(\{P_i^{[1]}\})(\mu) = 0$. □

For $\gamma \in \mathfrak{h}^*$, let $M(\lambda|_{\mathfrak{q}})_\gamma$ be the corresponding weight subspace of $\mathcal{U}(\mathfrak{q})v_\lambda \subset M(\lambda)$. Since $\mathfrak{h}u \subset \mathfrak{n}_\ell^-$, either $M(\lambda|_{\mathfrak{q}})_\gamma = 0$ or $\dim M(\lambda|_{\mathfrak{q}})_\gamma = \infty$. We also have $(\mathfrak{h}u)v_\lambda \neq 0$. Because of these facts, the $\mathfrak{q}$-modules $M(\lambda)$ and $\mathcal{U}(\mathfrak{q})v_\lambda$ do not fit in the framework of the highest weight theory developed in [Geoffriau 1995; Wilson 2011]. Nevertheless, they may have some nice features.

Lemma 2.2 relates $J(\{P_i^{[1]}\})$ to a property of the branching $\mathfrak{q} \downarrow \mathfrak{g}$ in a particular case of the $\mathfrak{q}$-module $\mathcal{U}(\mathfrak{q})v_\lambda$. In order to get a better understanding of this branching problem, we employ a certain projector introduced by Asherova, Smirnov, and Tolstoy in [Asherova et al. 1971].

2A. ***The extremal projector.*** Recall that $\{f_\alpha, h_\alpha, e_\alpha\} \subset \mathfrak{g}$ is the $\mathfrak{sl}_2$-triple corresponding to $\alpha \in \Delta^+$. Set $N = |\Delta^+|$ and choose a numbering of positive roots, $\alpha_1, \dots, \alpha_N$. Define

$$p_\alpha = 1 + \sum_{k=1}^{\infty} f_\alpha^k e_\alpha^k \frac{(-1)^k}{k!(h_\alpha + \rho(h_\alpha) + 1) \cdots (h_\alpha + \rho(h_\alpha) + k)}.$$

Each p_α, as well as any product of finitely many of them, is regarded as an element of the algebra of formal series of monomials

$$f_{\alpha_1}^{r_1} \cdots f_{\alpha_N}^{r_N} e_{\alpha_N}^{k_N} \cdots e_{\alpha_1}^{k_1} \text{ such that } (k_1 - r_1)\alpha_1 + \cdots + (k_N - r_N)\alpha_N = 0$$

with coefficients in the field of fractions of the commutative algebra $\mathcal{U}(\mathfrak{h})$. A total order on Δ^+ is said to be *normal* (or *convex*) if either $\alpha < \alpha+\beta < \beta$ or $\beta < \alpha+\beta < \alpha$ for each pair of positive roots α, β such that $\alpha + \beta \in \Delta$. There is a bijection between the normal orders and the reduced decompositions of the longest element of W.

Choose now a normal order $\alpha_1 < \alpha_2 < \cdots < \alpha_N$, and define

$$p = p_{\alpha_1} \cdots p_{\alpha_N}$$

accordingly. The element p is known as the *extremal projector* [Asherova et al. 1971]. It is independent of the choice of a normal order and $p^2 = p$. For proofs and more details on this operator, see, e.g., [Molev 2007, Section 9.1]. Most importantly, it has the property that

$$e_\alpha p = p f_\alpha = 0 \tag{2-2}$$

for each $\alpha \in \Delta^+$.

The nilpotent radical $\mathfrak{n}_\ell \subset \mathfrak{b}_\ell$ acts on $M(\lambda)$ locally nilpotently. Recall that $\mathfrak{n}^+ \subset \mathfrak{n}_\ell$. Let $v \in M(\lambda)$ be an eigenvector of $\mathfrak{h}$ of weight $\gamma \in \mathfrak{h}^*$. The element pv is a finite sum of vectors $q_j v_j$, where $q_j \in \operatorname{Quot}\mathcal{U}(\mathfrak{h}) \cong \mathbb{C}(\mathfrak{h}^*)$ and $v_j \in M(\lambda)$. If all the appearing denominators are nonzero at γ, then pv is a well-defined vector of $M(\lambda)$ of the same weight γ.

3. Proof of Theorem 1

Let λ, μ, and $M(\lambda)$ be as in Lemma 2.2. Keep in mind that λ and μ are arbitrary elements of $\mathbb{C}^\ell$ and $\mathbb{C}^n$. Since each $\mathcal{R}_i$ commutes with $\mathfrak{g}$, each $\mathcal{R}_i v_\lambda$ is a highest weight vector of $\mathfrak{g}$.

We use the extremal projector p associated with $\mathfrak{g}$. If p can be applied to a highest weight vector v, then $pv = v$. Suppose that p is defined at μ. Then, in view of (2-1) and (2-2),

$$\mathcal{R}_i v_\lambda = p\mathcal{R}_i v_\lambda = p\mathcal{T}(P_i^{[1]})|_{t=1} v_\lambda.$$

Assume that $J(\{P_i^{[1]}\})(\mu) = 0$. Then there is a nontrivial linear combination $\mathcal{R} = \sum c_i \mathcal{R}_i$ such that

$$\mathcal{R} v_\lambda \in \mathfrak{n}^- \mathcal{U}(\mathfrak{b}^-)(\mathfrak{n}^+ u) v_\lambda;$$

see Lemma 2.2. Here $p\mathcal{R} v_\lambda = 0$ and hence $\mathcal{R} v_\lambda = 0$ as well.

Since we are considering a Verma module of $\mathfrak{gl}_\ell(\mathbb{C})$, this implies that

$$\mathcal{R} \in \mathcal{U}(\mathfrak{gl}_\ell(\mathbb{C}))\mathfrak{b}_\ell \cap \mathcal{U}(\mathfrak{q}) = \mathcal{U}(\mathfrak{q})\mathfrak{b}.$$

Hence the symbol $\operatorname{gr}(\mathcal{R})$ of $\mathcal{R}$ lies in the ideal of $\mathcal{S}(\mathfrak{q})$ generated by $\mathfrak{b}$.

The decomposition $\mathfrak{g} = \mathfrak{n}^- \oplus \mathfrak{b}$ defines a bigrading on $\mathcal{S}(\mathfrak{g})$. Let $H_i^\bullet$ be the bihomogeneous component of H_i having the highest degree with respect to $\mathfrak{n}^-$. According to [Panyushev and Yakimova 2012, Section 3], $H_i^\bullet \in \mathfrak{b}\mathcal{S}^{d_i-1}(\mathfrak{n}^-)$ and the polynomials $H_1^\bullet, \dots, H_n^\bullet$ are algebraically independent. We have

$$\psi(H_i^\bullet) \in (\mathfrak{b}u)\mathcal{S}^{d_i-1}(\mathfrak{n}^-) \oplus \mathfrak{b}(\mathfrak{n}^- u)\mathcal{S}^{d_i-2}(\mathfrak{n}^-).$$

Write this as $\psi(H_i^\bullet) \in H_{i,1} + \mathfrak{b}(\mathfrak{n}^- u)\mathcal{S}^{d_i-2}(\mathfrak{n}^-)$. Then the polynomials $H_{i,1}$ with $1 \leqslant i \leqslant n$ are still algebraically independent. As can be easily seen, $\psi(H_i) \in H_{i,1} + \mathfrak{b}\mathcal{S}(\mathfrak{q})$.

Set $d = \max_{i:\, c_i \neq 0} d_i$. Then

$$\operatorname{gr}(\mathcal{R}) = \sum_{i:\, d_i = d} c_i \psi(H_i)$$

and it lies in $(\mathfrak{b}) \lhd \mathcal{S}(\mathfrak{q})$ if and only if $\sum_{i:\, d_i=d} c_i H_{i,1} = 0$. Since at least one c_i in this linear combination is nonzero, we get a contradiction. The following is settled: if p is defined at μ, then $J(\{P_i^{[1]}\})(\mu) \neq 0$.

Now we know that the zero set of $J(\{P_i^{[1]}\})$ lies in the union of hyperplanes $h_\alpha + \rho(h_\alpha) = -k$ with $k \geqslant 1$. At the same time, this zero set is an affine subvariety of $\mathbb{C}^n$ of codimension one. By Lemma 1.1, the highest degree component of $J(\{P_i^{[1]}\})$ is equal to $\prod_{\alpha \in \Delta^+} h_\alpha$. Therefore the zero set of $J(\{P_i^{[1]}\})$ is the union of N hyperplanes and $J(\{P_i^{[1]}\})$ is the product of N linear factors of the form $(h_\alpha + \rho(h_\alpha) + k_\alpha)$. Moreover, each $\alpha \in \Delta^+$ must appear in exactly one linear factor of $J(\{P_i^{[1]}\})$.

By Lemma 1.2, $J(\{P_i^{[1]}\})(0) = \prod_{\alpha \in \Delta^+}(\rho(h_\alpha) + 1)$. Hence we have

$$\prod_{\alpha \in \Delta^+} \frac{\rho(h_\alpha) + k_\alpha}{\rho(h_\alpha) + 1} = 1.$$

Since $\rho(h_\alpha), k_\alpha \geqslant 1$, each k_α is equal to 1. □

4. Conclusion

The elements $\mathcal{R}_i$ are rather natural $\mathfrak{g}$-invariants in $\mathcal{U}(\mathfrak{q})$ of degree one in $\mathfrak{g}u$. Note that because $\mathfrak{g}u$ is an abelian ideal of $\mathfrak{q}$, there is no ambiguity in defining the degree in $\mathfrak{g}u$.

The involvement of these elements in the branching rules $\mathfrak{q} \downarrow \mathfrak{g}$ remains unclear. However, combining Lemma 2.2 with Theorem 1, we obtain the following statement.

Corollary 4.1. *In the notation of Lemma 2.2, there is a nontrivial linear combination $\mathcal{R} = \sum c_i \mathcal{R}_i$ such that $\mathcal{R}v_\lambda \in \mathfrak{n}^- \mathcal{U}(\mathfrak{q})v_\lambda$ if and only if $\mu(h_\alpha) = -\rho(h_\alpha) - 1$ for some $\alpha \in \Delta^+$.* □

As the theory of finite-dimensional representations suggests, it is unusual for a highest weight vector of $\mathfrak{g}$ to belong to the image of $\mathfrak{n}^-$. The proof of Theorem 1 shows that $\mathcal{R}v_\lambda \neq 0$ for the linear combination of Corollary 4.1.

Remark 4.2. The subspace $\mathcal{V}[1] = (\mathcal{U}(\mathfrak{g})(\mathfrak{g}u))^{\mathfrak{g}} \subset \mathcal{U}(\mathfrak{q})$ is a $\mathcal{Z}(\mathfrak{g})$-module. From a well-known description of $(\mathfrak{g} \otimes \mathcal{S}(\mathfrak{g}))^{\mathfrak{g}}$, one can deduce that $\mathcal{V}[1]$ is freely generated by $\mathcal{R}_1, \dots, \mathcal{R}_n$ as a $\mathcal{Z}(\mathfrak{g})$-module. There are other choices of generators in $\mathcal{V}[1]$ and it is not clear, whether one can get nice formulas for the corresponding Jacobians.

References

[Asherova et al. 1971] R. M. Asherova, Y. F. Smirnov, and V. N. Tolstoy, "Projection operators for simple Lie groups", *Teoret. Mat. Fiz.* **8**:2 (1971), 255–271. In Russian; translated in *Theor. Math. Phys.* **8**:2 (1971), 813–825. MR Zbl

[Chari and Greenstein 2009] V. Chari and J. Greenstein, "A family of Koszul algebras arising from finite-dimensional representations of simple Lie algebras", *Adv. Math.* **220**:4 (2009), 1193–1221. MR Zbl

[Chari and Greenstein 2011] V. Chari and J. Greenstein, "Minimal affinizations as projective objects", *J. Geom. Phys.* **61**:3 (2011), 594–609. MR Zbl

[Dixmier 1974] J. Dixmier, *Algèbres enveloppantes*, Cahiers Sci. **37**, Gauthier-Villars, Paris, 1974. MR Zbl

[Geoffriau 1995] F. Geoffriau, "Homomorphisme de Harish-Chandra pour les algèbres de Takiff généralisées", *J. Algebra* **171**:2 (1995), 444–456. MR Zbl

[Greenstein and Mazorchuk 2017] J. Greenstein and V. Mazorchuk, "Koszul duality for semidirect products and generalized Takiff algebras", *Algebr. Represent. Theory* **20**:3 (2017), 675–694. MR Zbl

[Humphreys 1990] J. E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge Stud. Adv. Math. **29**, Cambridge Univ. Press, 1990. MR Zbl

[Khare and Ridenour 2012] A. Khare and T. Ridenour, "Faces of weight polytopes and a generalization of a theorem of Vinberg", *Algebr. Represent. Theory* **15**:3 (2012), 593–611. MR Zbl

[Molev 2007] A. Molev, *Yangians and classical Lie algebras*, Math. Surv. Monogr. **143**, Amer. Math. Soc., Providence, RI, 2007. MR Zbl

[Panyushev and Yakimova 2012] D. I. Panyushev and O. S. Yakimova, "A remarkable contraction of semisimple Lie algebras", *Ann. Inst. Fourier* (*Grenoble*) **62**:6 (2012), 2053–2068. MR Zbl

[Panyushev and Yakimova 2020] D. I. Panyushev and O. Yakimova, "Takiff algebras with polynomial rings of symmetric invariants", *Transform. Groups* **25**:2 (2020), 609–624.

[Takiff 1971] S. J. Takiff, "Rings of invariant polynomials for a class of Lie algebras", *Trans. Amer. Math. Soc.* **160** (1971), 249–262. MR Zbl

[Wilson 2011] B. J. Wilson, "Highest-weight theory for truncated current Lie algebras", *J. Algebra* **336** (2011), 1–27. MR Zbl

Received December 7, 2018. Revised December 13, 2019.

OKSANA YAKIMOVA
UNIVERSITÄT ZU KÖLN, MATHEMATISCHES INSTITUT
KÖLN
DEUTSCHLAND
Current address:
INSTITUT FÜR MATHEMATIK
FRIEDRICH SCHILLER UNIVERSITÄT JENA
JENA
GERMANY
oksana.yakimova@uni-jena.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 pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 306 No. 1 May 2020