Pacific Journal of Mathematics

Volume 299 No. 1 March 2019

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

EDITORS

Don Blasius (Managing Editor)
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
blasius@math.ucla.edu

Matthias Aschenbrenner
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
matthias@math.ucla.edu

Paul Balmer
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
balmer@math.ucla.edu

Vyjayanthi Chari
Department of Mathematics
University of California
Riverside, CA 92521-0135
chari@math.ucr.edu

Daryl Cooper
Department of Mathematics
University of California
Santa Barbara, CA 93106-3080
cooper@math.ucsb.edu

Wee Teck Gan
Mathematics Department
National University of Singapore
Singapore 119076
matgwt@nus.edu.sg

Kefeng Liu
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
liu@math.ucla.edu

Jiang-Hua Lu
Department of Mathematics
The University of Hong Kong
Pokfulam Rd., Hong Kong
jhlu@maths.hku.hk

Sorin Popa
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
popa@math.ucla.edu

Jie Qing
Department of Mathematics
University of California
Santa Cruz, CA 95064
qing@cats.ucsc.edu

Paul Yang
Department of Mathematics
Princeton University
Princeton NJ 08544-1000
yang@math.princeton.edu

PRODUCTION

Silvio Levy, Scientific Editor, production@msp.org

SUPPORTING INSTITUTIONS

ACADEMIA SINICA, TAIPEI
CALIFORNIA INST. OF TECHNOLOGY
INST. DE MATEMÁTICA PURA E APLICADA
KEIO UNIVERSITY
MATH. SCIENCES RESEARCH INSTITUTE
NEW MEXICO STATE UNIV.
OREGON STATE UNIV.
STANFORD UNIVERSITY
UNIV. OF BRITISH COLUMBIA
UNIV. OF CALIFORNIA, BERKELEY
UNIV. OF CALIFORNIA, DAVIS
UNIV. OF CALIFORNIA, LOS ANGELES
UNIV. OF CALIFORNIA, RIVERSIDE
UNIV. OF CALIFORNIA, SAN DIEGO
UNIV. OF CALIF., SANTA BARBARA
UNIV. OF CALIF., SANTA CRUZ
UNIV. OF MONTANA
UNIV. OF OREGON
UNIV. OF SOUTHERN CALIFORNIA
UNIV. OF UTAH
UNIV. OF WASHINGTON
WASHINGTON STATE UNIVERSITY

These supporting institutions contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its contents or policies.

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

The subscription price for 2019 is US $490/year for the electronic version, and $665/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/

© 2019 Mathematical Sciences Publishers

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 1, 2019
dx.doi.org/10.2140/pjm.2019.299.1

CALABI–YAU 4-FOLDS OF BORCEA–VOISIN TYPE FROM F-THEORY

Andrea Cattaneo, Alice Garbagnati and Matteo Penegini

We apply Borcea–Voisin's construction and give new examples of Calabi–Yau 4-folds Y, which admit an elliptic fibration onto a smooth 3-fold V, whose singular fibers of type I_5 lie above a del Pezzo surface $dP \subset V$. These are relevant models for F-theory according to Beasley et al. (2009a, 2009b). Moreover, we give the explicit equations of some of these Calabi–Yau 4-folds and their fibrations.

1. Introduction

New models of grand unified theory (GUT) have recently been developed using F-theory, a branch of string theory which provides a geometric realization of strongly coupled type IIB string theory backgrounds; see, e.g., [Beasley et al. 2009a; 2009b]. In particular, one can compactify F-theory on an elliptically fibered manifold, i.e., a fiber bundle whose general fiber is a torus.

We are interested in some of the mathematical questions posed by F-theory; above all, that of the construction of some of these models. For us, F-theory will be of the form $\mathbb{R}^{3,1} \times Y$, where Y is a Calabi–Yau 4-fold admitting an elliptic fibration with a section on a complex 3-fold V, namely:

$$\begin{tikzcd} E \arrow[r, hook] & Y \arrow[d, "\mathcal{E}"] \\ & V \end{tikzcd}$$

In general, the elliptic fibers E of $\mathcal{E}$ degenerate over a locus contained in a complex codimension one sublocus $\Delta(\mathcal{E})$ of V, the discriminant of $\mathcal{E}$. According to theoretical speculation in physics, $\Delta(\mathcal{E})$ should contain del Pezzo surfaces above which the general fiber is a singular fiber of type I_5 (Figure 1): see, for instance, [Beasley et al. 2009a; Bini and Penegini 2017].

MSC2010: 14J32, 14J35, 14J50.

Keywords: Calabi–Yau manifolds, elliptic fibrations, generalized Borcea–Voisin's construction, del Pezzo surfaces, K3 surfaces, F-theory.

The aim of this work is to investigate explicit examples of elliptically fibered Calabi–Yau 4-folds Y with this property by using a generalized Borcea–Voisin construction. The original Borcea–Voisin construction was described independently in [Borcea 1997] and [Voisin 1993], where the authors produced Calabi–Yau 3-folds starting from a K3 surface and an elliptic curve. Afterwards, generalization to higher dimensions was considered; see e.g., [Cynk and Hulek 2007; Dillies 2012]. There are two ways to construct 4-folds of Borcea–Voisin type, by using involutions, either starting from a pair of K3 surfaces, or considering a Calabi–Yau 3-fold and an elliptic curve. In this paper we will consider the former method. A first attempt to construct explicit examples of such Calabi–Yau 4-folds Y was made in [Bini and Penegini 2017], also using a generalized Borcea–Voisin's construction but applied to a product of a Calabi–Yau 3-fold and an elliptic curve. In that case the Calabi–Yau 3-fold was a complete intersection $(3, 3)$ in $\mathbb{P}^5$ containing a del Pezzo surface of degree 6; this construction was inspired by [Kapustka 2015].

In order to construct a Calabi–Yau 4-fold Y with the elliptic fibration $\mathcal{E}$ as required one needs both a map to a smooth 3-fold V whose generic fibers are genus 1 curves and a distinguished del Pezzo surface dP in V. A natural way to produce these data is to consider two K3 surfaces S_1 and S_2 such that S_1 is the double cover of dP and S_2 admits an elliptic fibration $\pi : S_2 \to \mathbb{P}^1$. In this way we will obtain $\mathcal{E} : Y \to V \simeq dP \times \mathbb{P}^1$. To get Y from S_1 and S_2 we need a nonsymplectic involution on each surface. Since S_1 is a double cover of dP, it clearly admits the cover involution, denoted by ι_1, while the involution ι_2 on S_2 is induced by the elliptic involution on each smooth fiber of π. Thus, $(S_1 \times S_2)/(\iota_1 \times \iota_2)$ is a singular Calabi–Yau 4-fold which admits a crepant resolution Y obtained blowing up the singular locus. It follows at once that there is a map $Y \to (S_1/\iota_1) \times \mathbb{P}^1 \simeq dP \times \mathbb{P}^1$ whose generic fiber is a smooth genus 1 curve and the singular fibers lie either on $dP \times \Delta(\pi)$ or on $C \times \mathbb{P}^1$ (where $C \subset dP$ is the branch curve of $S_1 \to dP$ and $\Delta(\pi)$ is the discriminant of π). The discriminant $\Delta(\pi)$ consists of a finite number of points and generically the fibers of $\mathcal{E}$ over $dP \times \Delta(\pi)$ are of the same type as the fiber of π over $\Delta(\pi)$. Therefore the requirements on the singular fibers of $\mathcal{E}$ needed in F-theory reduce to a requirement on the elliptic fibration $\pi : S_2 \to \mathbb{P}^1$.

Moreover, we show that the choice of S_1 as double cover of a del Pezzo surface and of S_2 as elliptic fibration with specific reducible fibers can be easily modified to obtain Calabi–Yau 4-folds with elliptic fibrations with a different basis (isomorphic to $S_1/\iota_1 \times \mathbb{P}^1$) and reducible fibers (over $S_1/\iota_1 \times \Delta(\pi)$).

Our first result, proven in Propositions 3.1 and 4.2 (see also Section 4C) is:

Theorem 1.1. *Let dP be a del Pezzo surface of degree $9 - n$ and $S_1 \to dP$ be a double cover with S_1 a K3 surface. Let $S_2 \to \mathbb{P}^1$ be an elliptic fibration on a K3 surface with singular fibers $mI_5 + (24 - 5m)I_1$. The blow up Y of $(S_1 \times S_2)/(\iota_1 \times \iota_2)$ along its singular locus is a crepant resolution. It is a Calabi–Yau 4-fold which* *admits an elliptic fibration $\mathcal{E}: Y \to dP \times \mathbb{P}^1$ whose discriminant contains m copies of dP above which the fibers are of type I_5. The Hodge numbers of Y depend only on n and m and are*

$$\begin{aligned} h^{1,1}(Y) &= 5+n+2m, & h^{2,1}(Y) &= 2(15-n-m), \\ h^{2,2}(Y) &= 4(138-9n-19m+2nm), & h^{3,1}(Y) &= 137-11n-22m+2nm. \end{aligned}$$

We also give more specific results on Y. Indeed, recalling that a del Pezzo surface is either $\mathbb{P}^1 \times \mathbb{P}^1$ or a blow up of $\mathbb{P}^2$ in n points $\beta : dP \to \mathbb{P}^2$, for $0 \leq n \leq 8$, we give a Weierstrass equation for the elliptic fibration $Y \to \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1$ or $Y \to \beta(dP) \times \mathbb{P}^1$, respectively, induced by $\mathcal{E}$; see (12) and (13). Moreover, for $n = 5, 6$ we provide the explicit Weierstrass equation of the fibration $\mathcal{E}: Y \to dP \times \mathbb{P}^1$; see (17) and (15).

For $m = 4$, there are two different choices for $\pi : S_2 \to \mathbb{P}^1$. One of them is characterized by the presence of a 5-torsion section for $\pi : S_2 \to \mathbb{P}^1$ and in this case the K3 surface S_2 is a $2:1$ cover of the rational surface with a level 5 structure; see [Balestrieri et al. 2018]. We observe that if $\pi : S_2 \to \mathbb{P}^1$ admits a 5-torsion section, the same is true for $\mathcal{E}$.

The particular construction of Y enables us to find two other distinguished fibrations (besides $\mathcal{E}$): one whose fibers are K3 surfaces and the other whose fibers are Calabi–Yau 3-folds of Borcea–Voisin type. So Y admits fibrations in Calabi–Yau manifolds of any possible dimension. Moreover, by the explicit description of these fibrations, we observe that $\mathcal{E}$ and the fibration in Calabi–Yau 3-folds are not isotrivial. So Y can be interpreted as a non (iso)trivial family of elliptic curves and of Calabi–Yau 3-folds.

The concrete geometric description and the explicit equation of Y are interesting in view of a possible application to F-theory and can be also used to specialize Y to some more specific Calabi–Yau 3-folds with extra symmetries. These specializations are intensively used in dimension 3 to construct Calabi–Yau 3-folds with prescribed Hodge numbers (see, e.g., [Constantin et al. 2017; Braun 2011]) and can be considered in higher dimensions.

The geometric description of the fibrations on Y and their projective realization is based on a detailed study of the linear systems of divisors on Y. In particular we consider divisors D_Y induced by divisors on S_1 and S_2. We relate the dimension of the spaces of sections of D_Y with the one of the associated divisors on S_1 and S_2. Thanks to this study we are also able to describe Y as a double cover of $\mathbb{P}^2 \times \mathbb{F}_4$ (where $\mathbb{F}_4$ is the Hirzebruch surface S_2/ι_2) and as an embedded variety in $\mathbb{P}^{59-n}$. The main results in this context are summarized in Propositions 6.1 and 6.2.

The paper is organized as follows. In Section 2, we recall the definition of Calabi–Yau manifold, K3 surface and del Pezzo surface. Moreover, we describe nonsymplectic involutions on K3 surfaces. Finally in Section 2E we introduce the Borcea–Voisin construction. Section 3 is devoted to presenting models Y for

F-theory described in the introduction. The Hodge numbers of Y are calculated in Section 4. Section 5 is devoted to the study of the linear systems on Y. The results are applied in Section 6 where several fibrations and projective models of Y are described. Finally, in Section 7 we provide the explicit equations for some of these models and fibrations.

Notation and conventions. We work over the field of complex numbers $\mathbb{C}$.

2. Preliminaries

Definition 2.1. A *Calabi–Yau* manifold X is a compact Kähler manifold with trivial canonical bundle such that $h^{i,0}(X) = 0$ if $0 < i < \dim X$.

A *K3* surface S is a Calabi–Yau manifold of dimension 2. The Hodge numbers of S are uniquely determined by these properties and are $h^{0,0}(S) = h^{2,0}(S) = 1$, $h^{1,0}(S) = 0$, and $h^{1,1}(S) = 20$.

2A. An involution ι on a K3 surface S can be either symplectic, i.e., it preserves the symplectic structure of the surface, or not, in which case we speak of nonsymplectic involution. In addition, an involution on a K3 surface is symplectic if and only if its fixed locus consists of isolated points; an involution on a K3 surface is nonsymplectic if and only if there are no isolated fixed points on S. These remarkable results depend on the possibility to linearize ι near the fixed locus. Moreover, the fixed locus of an involution on S is smooth. In particular, the fixed locus of a nonsymplectic involution on a K3 surface is either empty or consists of the disjoint union of smooth curves.

From now on we consider only nonsymplectic involutions ι on K3 surfaces S. As a consequence of the Hodge index theorem and of the adjunction formula, if the fixed locus contains at least one curve C of genus $g(C) := g \geq 2$, then all the other curves in the fixed locus are rational. On the other hand, if there is one curve of genus 1 in the fixed locus, then the other fixed curves are either rational curves or exactly one genus 1 curve.

So one obtains that the fixed locus of ι on S can be one of the following:

- Empty.
- The disjoint union of two smooth genus 1 curves E_1 and E_2.
- The disjoint union of k curves, such that $k - 1$ are surely rational, with the remaining curve having genus $g \geq 0$.

If we exclude the first two cases $\big(\mathrm{Fix}_\iota(S) = \varnothing,\ \mathrm{Fix}_\iota(S) = E_1 \coprod E_2\big)$, the fixed locus can be topologically described by the two integers (g, k).

There is another point of view in the description of the involution ι on S. Indeed, ι^* acts on the second cohomology group of S and its action is related to the moduli space of K3 surfaces admitting a prescribed involution; this is due to the construction of the moduli space of the lattice polarized K3 surfaces. So we are interested in the description of the lattice $H^2(S,\mathbb{Z})^{\iota^*}$. This coincides with the invariant part of the Néron–Severi group $NS(S)^{\iota_*}$ since the automorphism is nonsymplectic, and thus acts on $H^{2,0}(S)$ as $-\mathrm{id}_{H^{2,0}(S)}$; see [Nikulin 1979, Section 4, 2°]. The lattice $H^2(S,\mathbb{Z})^{\iota^*}$ of rank $r := \mathrm{rk}(H^2(S,\mathbb{Z})^{\iota^*})$ is known to be 2-elementary, i.e., its discriminant group is $(\mathbb{Z}/2\mathbb{Z})^a$. Hence one can attach to this lattice the two integers (r,a). A very deep and important result on the nonsymplectic involutions on K3 surfaces is that each admissible pair of integers (g,k) is uniquely associated to a pair of integers (r,a); see [Nikulin 1979, Theorem 4.2.2].

We observe that for several admissible choices of (r,a) this pair uniquely determines the lattice $H^2(S,\mathbb{Z})^{\iota^*}$, but there are some exceptions.

The relations between (g,k) and (r,a) are explicitly given by

$$g=\frac{22-r-a}{2},\quad k=\frac{r-a}{2}+1, \qquad r=10+k-g,\quad a=12-k-g. \tag{1}$$

2B. A surface dP is called a *del Pezzo surface* of degree d if the anticanonical bundle $-K_{dP}$ is ample and $K_{dP}^2=d$. Moreover,we say that dP is a *weak del Pezzo* surface if $-K_{dP}$ is big and nef.

The anticanonical map embeds dP in $\mathbb{P}^d$ as a surface of degree d. The del Pezzo surfaces are either $\mathbb{P}^1\times\mathbb{P}^1$ (which has degree 8) or a blow up of $\mathbb{P}^2$ in $9-d$ points in general position

$$\beta : dP \cong Bl_{9-d}(\mathbb{P}^2)\to\mathbb{P}^2; \tag{2}$$

see, e.g., [Dolgachev 2012].

2C. A double cover of a del Pezzo surface dP ramified along a smooth curve $C\in|-2K_{dP}|$ is a K3 surface S, endowed with the covering involution ι. Since dP is not a symplectic manifold, ι is nonsymplectic. For all the del Pezzo surfaces except $\mathbb{P}^1\times\mathbb{P}^1$, we can see S as the minimal resolution of a double cover of $\mathbb{P}^2$ branched along $\beta(C)$, which is a sextic with $9-d$ nodes. Let us denote by $\rho' : S\to\mathbb{P}^2$ the composition of the double cover with the minimal resolution. The ramification divisor of ρ' is a genus $1+d$ smooth curve, which is the fixed locus of ι.

If the del Pezzo surface is $\mathbb{P}^1\times\mathbb{P}^1$, then S is a double cover of $\mathbb{P}^1\times\mathbb{P}^1$ branched along a smooth curve of bidegree $(4,4)$ and we denote by $\rho' : S\to\mathbb{P}^1\times\mathbb{P}^1$ the double cover.

Definition 2.2. An *elliptic fibration* $\mathcal{E} : Y\to V$ is a surjective map with connected fibers between smooth manifolds such that: the general fiber of $\mathcal{E}$ is a smooth genus 1 curve; there is a rational map $O : V \dashrightarrow Y$ such that $\mathcal{E}\circ O=\mathrm{id}_V$. A *flat elliptic fibration* is an elliptic fibration with a flat map $\mathcal{E}$. In particular a flat elliptic fibration has equidimensional fibers.

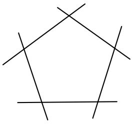

Figure 1. Fiber of type I_5.

2D. If Y is a surface then any elliptic fibration is flat. Moreover, on Y there is an involution ι which restricts to the elliptic involution on each smooth fiber. If Y is a K3 surface, then ι is a nonsymplectic involution.

2E. The generalized Borcea–Voisin construction. Let X_i, $i=1,2$, be a Calabi–Yau manifold endowed with an involution ι_i whose fixed locus has codimension 1. The quotient

$$(X_1\times X_2)/(\iota_1\times\iota_2)$$

admits a crepant resolution which is a Calabi–Yau manifold as well (see [Cynk and Hulek 2007]). We call *Borcea–Voisin* of X_1 and X_2 the Calabi–Yau $\mathrm{BV}(X_1, X_2)$ which is the blow up of $(X_1\times X_2)/(\iota_1\times\iota_2)$ in its singular locus.

2F. Let $b:\widetilde{X_1\times X_2}\to X_1\times X_2$ be the blow up of $X_1\times X_2$ in the fixed locus of $\iota_1\times\iota_2$. Let $\tilde{\iota}$ be the induced involution on $\widetilde{X_1\times X_2}$ and $q:\widetilde{X_1\times X_2}\to\widetilde{X_1\times X_2}/\tilde{\iota}=:Y$ its quotient. The commutative diagram

$$\begin{tikzcd}
\widetilde{X_1\times X_2} \arrow[r, "b"] \arrow[d, "q"'] & X_1\times X_2 \arrow[d] \\
BV(X_1, X_2)\cong Y \arrow[r] & (X_1\times X_2)/(\iota_1\times\iota_2)
\end{tikzcd}$$

exhibits the Borcea–Voisin manifold as a smooth quotient.

3. The construction

3A. In the following we apply the just-described Borcea–Voisin construction in order to get a Calabi–Yau 4-fold Y together with a fibration $\mathcal{E}:Y\to V$ onto a smooth 3-fold V, with the following properties: the general fiber of $\mathcal{E}$ is a smooth elliptic curve E, the discriminant locus of $\mathcal{E}$ contains a del Pezzo surface dP, and for a generic point $p\in dP$ the singular fiber $\mathcal{E}^{-1}(p)$ is of type I_5 (see Figure 1).

3B. Let S_1 and S_2 be two K3 surfaces with the following properties:

(1) S_1 admits either a $2:1$ covering $\rho':S_1\to\mathbb{P}^2$, branched along a curve C, which is a (possibly singular and possibly reducible) sextic curve in $\mathbb{P}^2$, or a $2:1$ covering $\rho':S_1\to\mathbb{P}^1\times\mathbb{P}^1$, branched along a curve C, which is a (possibly singular and possibly reducible) curve of bidegree $(4,4)$ on $\mathbb{P}^1\times\mathbb{P}^1$.

(2) S_2 admits an elliptic fibration $\pi:S_2\to\mathbb{P}^1$, with discriminant locus $\Delta(\pi)$.

The surface S_1 has the covering involution ι_1, which is a nonsymplectic involution. Moreover, if the branch curve $C \subset \mathbb{P}^2$ (resp. $C \subset \mathbb{P}^1 \times \mathbb{P}^1$) is singular, then the double cover of $\mathbb{P}^2$ (resp. $\mathbb{P}^1 \times \mathbb{P}^1$) branched along C is singular. In this case the K3 surface S_1 is the minimal resolution of this last singular surface. The fixed locus of ι_1 consists of the strict transform $\tilde{C}$ of the branch curve, and possibly of some other smooth rational curves, W_i (which arise from the resolution of the triple points of C). Moreover, notice that if we choose $C \subset \mathbb{P}^2$ to be a sextic with $n < 9$ nodes in general position then ρ' factors through

$$\rho : S_1 \xrightarrow{2:1} dP := \mathrm{Bl}_n\mathbb{P}^2,$$

where dP is a del Pezzo surface of degree $d = 9 - n$. If C is a smooth curve, then $\widetilde{C} = C$ and we put $\rho = \rho'$ so we still have $\rho : S_1 \xrightarrow{2:1} dP$.

The second K3 surface S_2 admits a nonsymplectic involution too, as in Section 2D. This is the elliptic involution ι_2, which acts on the smooth fibers of π as the elliptic involution of each elliptic curve. In particular it fixes the 2-torsion group on each fiber. Therefore, it fixes the zero section O, which is a rational curve, and the trisection T (not necessarily irreducible) passing through the 2-torsion points of the fibers.

3C. Applying the Borcea–Voisin construction (Section 2E) to (S_1, ι_1) and (S_2, ι_2), we obtain a smooth Calabi–Yau 4-fold Y. In particular, the singular locus of the quotient $X := (S_1 \times S_2)/(\iota_1 \times \iota_2)$ is the image of the fixed locus of the product involution $\iota_1 \times \iota_2$. As the involution acts componentwise, we have

$$\mathrm{Fix}_{S_1 \times S_2}(\iota_1 \times \iota_2) = \mathrm{Fix}_{S_1}\,\iota_1 \times \mathrm{Fix}_{S_2}\,\iota_2,$$

therefore the fix locus consists of the *disjoint union* of

(1) the surface $\tilde{C} \times O$, where $O \simeq \mathbb{P}^1$ is the section of π,

(2) the surface $\tilde{C} \times T$, where T is the trisection of π,

and, possibly,

(3) the surfaces $\tilde{C} \times E_i$ (where $E_i \simeq \mathbb{P}^1$ are the fixed components in the reducible fibers of π),

(4) the surfaces $W_i \times O$, $W_i \times T$ and $W_i \times E_j$ (where $W_i \simeq \mathbb{P}^1$ are the rational curves fixed by ι_1 on S_1).

As in Section 2E we have the following commutative diagram.

$$\begin{tikzcd}
\widetilde{S_1 \times S_2} \arrow[r, "b"] \arrow[d, "q"'] & S_1 \times S_2 \arrow[d] \\
Y \arrow[r] & X
\end{tikzcd} \tag{3}$$

3D. By construction, the smooth 4-fold Y comes with several fibrations. Let us analyze one of them and postpone the description of the others until Section 6.

Suppose that the del Pezzo surface is obtained by blowing up $\mathbb{P}^2$ in $9-d$ points in general position (the easier case $dP \simeq \mathbb{P}^1 \times \mathbb{P}^1$ can be studied in similar way). We have the fibration $Y \to \mathbb{P}^2 \times \mathbb{P}^1$ induced by the covering $\rho' : S_1 \to \mathbb{P}^2$ and the fibration $\pi : S_2 \to \mathbb{P}^1$. Recall from Section 3B that we can specialize the fibration if we require that ρ' is branched along a sextic with n nodes in general position. This further assumption yields

$$\begin{tikzcd} Y \arrow[d, "\varphi"] \\ dP \times \mathbb{P}^1 \end{tikzcd}$$

where dP is the del Pezzo surface obtained blowing up the nodes of the branch locus. The general fiber of φ is an elliptic curve. Indeed, let $(p, q) \in dP \times \mathbb{P}^1$ with $p \notin C$ and $q \notin \Delta(\pi)$. Then $(\varphi)^{-1}(p, q)$ is isomorphic to the smooth elliptic curve $\pi^{-1}(q)$. Hence the singular fibers lie on points $(p, q) \in dP \times \mathbb{P}^1$ of one of the following three types: $p \in C,\ q \notin \Delta(\pi)$; $p \notin C,\ q \in \Delta(\pi)$; $p \in C,\ q \in \Delta(\pi)$. We discuss these three cases separately.

Case 1: $(p, q) \in dP \times \mathbb{P}^1$ with $p \notin C$ and $q \in \Delta(\pi)$. Clearly $\pi^{-1}(q)$ is a singular curve, and since $p \notin C$, we get a singular fiber for φ,

$$\varphi^{-1}(p, q) \simeq \pi^{-1}(q). \tag{4}$$

Case 2: $(p, q) \in dP \times \mathbb{P}^1$ with $p \in C$ and $q \notin \Delta(\pi)$. Consider first $(\rho \times \pi)^{-1}(p, q)$ in $S_1 \times S_2$. This is a single copy of $\pi^{-1}(q)$, which is a smooth elliptic curve, over the point $p \in C \subseteq S_1$. In addition, this curve meets the fixed locus of $\iota_1 \times \iota_2$ in four distinct points: one of them corresponds to the intersection with $C \times O$ and the other three correspond to the intersections with $C \times T$. Notice that $\iota_1 \times \iota_2$ acts on $p \times \pi^{-1}(q)$ as the elliptic involution ι_2, hence the quotient curve is a rational curve. This discussion yields that $\varphi^{-1}(p, q)$ is a singular fiber of type I_0^*, where the central rational component is isomorphic to the quotient of $\pi^{-1}(q)/\iota_2$ and the other four rational curves are obtained by blowing up the intersection points described above.

Case 3: $(p, q) \in dP \times \mathbb{P}^1$ with $p \in C$ and $q \in \Delta(\pi)$. This time, $(\rho \times \pi)^{-1}(p, q)$ is the singular fiber $\pi^{-1}(q)$. Moreover, the quotient of this curve by ι_2 is determined by its singular fiber type. If ι_2 does not fix a component of $\pi^{-1}(q)$, then $(\rho \times \pi)^{-1}(p, q)$ meets the fixed locus of $\iota_1 \times \iota_2$ in a certain number of isolated points, depending on the fiber $\pi^{-1}(q)$ (which correspond to the intersection of the fiber with O and T). On the other hand, if ι_2 does fix a component of $\pi^{-1}(q)$, then there are curves in $(\rho \times \pi)^{-1}(p, q)$. In the latter case, $\phi^{-1}(p, q)$ contains a divisor.

In each of the previous cases, the fiber over (p, q) is not smooth and thus we obtain that the discriminant locus of φ is

$$\Delta(\varphi) = (C \times \mathbb{P}^1) \cup (dP \times \Delta(\pi)).$$

This discussion yields that the surface $dP \times \{q\} \subset \Delta(\varphi)$ for all $q \in \Delta(\pi)$ and for the generic point $p \in dP$ the fiber of φ over (p, q) is of the same type as the fiber of π over q. This implies the following proposition.

Proposition 3.1. *There exists a Calabi–Yau* 4*-fold with an elliptic fibration over* $dP \times \mathbb{P}^1$ *such that the discriminant locus contains a copy of dP. If, moreover, we assume that the generic fiber above it is reduced, i.e., is of type* I_n*, II, III, IV, then it is possible to construct this elliptic fibration to be flat.*

Proof. We show firstly that if the singular fibers of the elliptic fibration π are of type I_n, *II*, *III* or *IV*, then our method produces an equidimensional fibration on Y. By our analysis in Section 3D, Case 3, it suffices to show that the elliptic involution ι_2 on S_2 does not fix any irreducible component of such fibers. As we already observed, the fixed locus of a nonsymplectic involution on a K3 surface consists of the disjoint union of smooth curves, which readily rules out the irreducible singular fibers (i.e., those of type I_1 and *II*). Consider now the case of the I_n singular fibers: call Γ_i, $i \in \mathbb{Z}/n\mathbb{Z}$, its irreducible components, in such a way that the component meeting the section O is Γ_0 and Γ_i intersects Γ_{i+1}. Consider then Γ_0; since it meets the section O (which is a component of $\mathrm{Fix}_{S_2}\,\iota_2$) we deduce that it is invariant but not fixed for ι_2, hence this involution must switch the two points where Γ_0 meets Γ_1 and Γ_{n-1} and has another fixed point. As a result, ι_2 switches Γ_1 and Γ_{n-1} and consequently switches Γ_i with Γ_{n-i} for $1 \leq i \leq \left[\frac{n-1}{2}\right]$. In the end, either we have a fixed point at the intersection of $\Gamma_{(n-1)/2}$ and $\Gamma_{(n+1)/2}$ if n is odd, or we have two fixed points on $\Gamma_{n/2}$ if n is even (in this case, this curve is ι_2-invariant). Consider now a fiber of type *III*, and let Γ_0 be the component meeting the section O and Γ_1 the other component. On Γ_0 the point $\Gamma_0 \cap O$ is fixed, and so Γ_0 is an invariant curve which is not fixed by ι_2. As there is only one singular point in the fiber, this point must be fixed as well, and so there cannot be other fixed points on Γ_0. As a consequence, if the trisection T does not meet Γ_1, then it must meet Γ_0 and this would imply that Γ_0 is fixed. So T must meet Γ_1, which prevents Γ_1 from being a fixed curve. Consider finally a fiber of type *IV*, and let Γ_0 be the component meeting the section O and Γ_1, Γ_2 be the other components. The unique singular point is necessarily a fixed point. As before, on Γ_0 we have two fixed points, the intersection with the section O and the singular point of the fiber, and this component is globally invariant and can not meet the trisection T in a point different from the singular point. If T does not meet Γ_0 at all (and thus does not pass through the singular point), then it must meet one between Γ_1 and Γ_2 with multiplicity 2, which would imply that component is fixed (since it has at least three fixed points). As it meets the trisection (which is fixed), this is absurd, and so T must pass through the singular point. As a consequence, neither Γ_1 nor Γ_2 are fixed since they meet T at least in the singular point.

This is enough to claim that the elliptic fibration on Y is equidimensional. Finally, as we are dealing with morphisms between smooth varieties, by [Nowak 1997, criterion for flatness] we deduce that our fibration is also flat. □

3E. We shall now discuss a special case of the elliptic fibration φ. Apparently, a good phenomenological model for F-theory (see the introduction and references therein) is the one where the discriminant locus contains a del Pezzo surface over which there are I_5 singular fibers. Indeed, F-theory on an elliptically fibered Calabi–Yau 4-fold Y with base B is equivalent to Type IIB string theory on B with a dilaton-axion $\tau = C_0 + ie^{-\phi}$ varying over this base. At each point in B the complex number τ can be identified with the complex structure modulus of the elliptic fiber over this point. For Y to be a Calabi–Yau 4-fold this fiber has to degenerate over divisors D_i in B. These degeneration loci encode the location of space-time filling seven-branes of Type IIB compactified on B. In the case of an SU(5) gauge group theory, D_i should be del Pezzo surfaces and a singular fiber splits into an I_5 Kodaira singular fiber; see, e.g., [Braun et al. 2013]. The choice of an SU(5) gauge group theory lies on the fact that it is the smallest simple Lie group which contains the standard model, and upon which the first grand unified theory was based. Besides SU(5), another group which seems to be interesting for the grand unified theories is E_6: the corresponding fibrations will have singular fibers of type IV^* on the del Pezzo in the discriminant. Let us discuss the situation of SU(5).

Remark 3.2. By Proposition 3.1 it is possible to construct elliptic fibrations with fibers I_5. Nevertheless, it is not possible to obtain elliptic fibrations such that *all* the singular fibers are of type I_5. Indeed, there are two different obstructions:

(1) The fibers obtained in Case 2 of Section 3D are of type I_0^* and this does not depend on the choice of the properties of the elliptic fibration $S_2 \to \mathbb{P}^1$.

(2) The singular fibers as in Case 1 of Section 3D depend only on the singular fibers of $S_2 \to \mathbb{P}^1$ and these cannot be only of type I_5, indeed $24 = \chi(S_2)$ is not divisible by 5.

However, it is known that there exist elliptic K3 surfaces with m fibers of type I_5 and all other singular fibers of type I_1 for $m = 1, 2, 3, 4$; see [Shimada 2000]. In this case the number of fibers of type I_1 is $24 - 5m$.

4. The Hodge numbers of *Y*

The aim of this section is the computation of the Hodge numbers of the constructed 4-folds.

4A. By (3) the cohomology of Y is given by the part of the cohomology of $\widetilde{S_1 \times S_2}$ which is invariant under $(\iota_1 \times \iota_2)^*$. The cohomology of $\widetilde{S_1 \times S_2}$ is essentially obtained as the sum of two different contributions: the pullback by b^* of the cohomology of $S_1 \times S_2$ and the part of the cohomology introduced by the blow up of the fixed locus $\mathrm{Fix}_{\iota_1 \times \iota_2}(S_1 \times S_2)$. The fixed locus $\mathrm{Fix}_{\iota_1 \times \iota_2}(S_1 \times S_2) = \mathrm{Fix}_{\iota_1}(S_1) \times \mathrm{Fix}_{\iota_2}(S_2)$ consists of surfaces, which are products of curves. So $b : \widetilde{S_1 \times S_2} \to S_1 \times S_2$ introduces exceptional divisors which are $\mathbb{P}^1$-bundles over surfaces which are products of curves. The Hodge diamonds of these exceptional 3-folds depends only on the genus of the curves in $\mathrm{Fix}_{\iota_1}(S_1)$ and $\mathrm{Fix}_{\iota_2}(S_2)$.

Since, up to an appropriate shift of the indices, the Hodge diamond of $\widetilde{S_1 \times S_2}$ is just the sum of the Hodge diamond of $S_1 \times S_2$ and of all the Hodge diamonds of the exceptional divisors, the Hodge diamond of $\widetilde{S_1 \times S_2}$ depends only on the properties of the fixed locus of ι_1 on S_1 and of ι_2 on S_2. Denoted by (g_i, k_i), $i = 1, 2$, the pair of integers which describes the fixed locus of ι_i on S_i, we obtain that the Hodge diamond of $\widetilde{S_1 \times S_2}$ depends only on the four integers (g_1, k_1, g_2, k_2).

Now we consider the quotient 4-fold Y. Its cohomology is the invariant cohomology of $\widetilde{S_1 \times S_2}$ for the action of $(\iota_1 \times \iota_2)^*$. Since the automorphism induced by $\iota_1 \times \iota_2$ on $\widetilde{S_1 \times S_2}$ acts trivially on the exceptional divisors, one has only to compute the invariant part of the cohomology of $S_1 \times S_2$ for the action of $(\iota_1 \times \iota_2)^*$. But this depends of course only on the properties of the action of ι_i^* on the cohomology of S_i. We observe that ι_i^* acts trivially on $H^0(S_i, \mathbb{Z})$, and that $H^1(S_i, \mathbb{Z})$ is empty. Denote by (r_i, a_i), $i = 1, 2$, the invariants of the lattice $H^2(S_i, \mathbb{Z})^{\iota_i^*}$; these determine uniquely $H^*(S_1 \times S_2, \mathbb{Z})^{(\iota_1 \times \iota_2)^*}$.

Thus the Hodge diamond of Y depends only on (g_i, k_i) and (r_i, a_i), $i = 1, 2$. By (1), it is immediate that the Hodge diamond of Y depends only either on (g_1, k_1, g_2, k_2) or equivalently on (r_1, a_1, r_2, a_2).

This result is already known, due to J. Dillies who computed the Hodge numbers of the Borcea–Voisin of the product of two K3 surfaces by means of the invariants (r_1, a_1, r_2, a_2) in [Dillies 2012]:

Proposition 4.1 [Dillies 2012, Section 7.2.1]. *Let ι_i be a nonsymplectic involution on S_i, $i = 1, 2$, such that its fixed locus is nonempty and does not consist of two curves of genus* 1. *Let Y be the Borcea–Voisin* 4*-fold of S_1 and S_2. Then*

$$h^{1,1}(Y) = 1 + \frac{r_1 r_2}{4} - \frac{r_1 a_2}{4} - \frac{a_1 r_2}{4} + \frac{a_1 a_2}{4} + \frac{3r_1}{2} - \frac{a_1}{2} + \frac{3r_2}{2} - \frac{a_2}{2},$$

$$h^{2,1}(Y) = 22 - \frac{r_1 r_2}{2} + \frac{a_1 a_2}{2} + 5r_1 - 6a_1 + 5r_2 - 6a_2,$$

$$h^{2,2}(Y) = 648 + 3r_1 r_2 + a_1 a_2 - 30r_1 - 30r_2 - 12a_1 - 12a_2,$$

$$h^{3,1}(Y) = 161 + \frac{r_1 r_2}{4} + \frac{a_1 a_2}{4} + \frac{r_1 a_2}{4} + \frac{a_1 r_2}{4} - \frac{13r_1}{2} - \frac{13r_2}{2} - \frac{11a_1}{2} - \frac{11a_2}{2}.$$

4B. Now we apply these computations to our particular case: S_2 is an elliptic K3 surface with m fibers of type I_5 and S_1 is either the double cover of $\mathbb{P}^2$ branched along a sextic with n nodes or the double cover of $\mathbb{P}^1 \times \mathbb{P}^1$ branched along a smooth bidegree $(4, 4)$ curve. In this latter case we pose $n = 1$. We obtain the following proposition.

Proposition 4.2. *Let $m \geq 0$ be an integer, and suppose that $\pi : S_2 \to \mathbb{P}^1$ in an elliptic fibration with singular fibers of type $mI_5 + (24-5m)I_1$. Then*

$$\begin{aligned} h^{1,1}(Y) &= 5 + n + 2m, \\ h^{2,1}(Y) &= 2(15 - n - m), \\ h^{2,2}(Y) &= 4(138 - 9n - 19m + 2nm), \\ h^{3,1}(Y) &= 137 - 11n - 22m + 2nm. \end{aligned}$$

Proof. In order to deduce the Hodge numbers of Y by Proposition 4.1, we have to compute the invariants (g_i, k_i) of the action of ι_i on S_i in our context. If the surface S_1 is a $2:1$ cover of $\mathbb{P}^2$ branched on a sextic with n nodes and ι_1 is the cover involution, then the fixed locus of ι_1 is isomorphic to the branch curve and hence has genus $10-n$. So $(g_1, k_1) = (10-n, 1)$ and thus $r_1 = 1+n$ and $a_1 = 1+n$. If the surface S_1 is a $2:1$ cover of $\mathbb{P}^1 \times \mathbb{P}^1$ branched on a smooth bidegree $(4, 4)$ curve, and ι_1 is the cover involution, then the fixed locus of ι_1 is isomorphic to the branch curve and hence has genus $9 = 10 - n$. Also in this case $(g_1, k_1) = (10 - n, 1)$ and thus $r_1 = 1 + n$ and $a_1 = 1 + n$.

The involution ι_2 on S_2 is the elliptic involution, and hence fixes the section of the fibration, which is a rational curve, and the trisection passing through the 2-torsion points of the fibers. Moreover, ι_2 does not fix components of the reducible fibers. So $k_2 = 2$ and it remains to compute the genus of the trisection. The Weierstrass equation of the elliptic fibration S_2 is $y^2 = x^3 + A(t)x + B(t)$ and the equation of the trisection T is $x^3 + A(t)x + B(t) = 0$, which exhibits T as $3:1$ cover of $\mathbb{P}^1_t$ branched on the zero points of the discriminant $\Delta(t) = 4A(t)^3 + 27B(t)^2$. Under our assumptions, the discriminant has m roots of multiplicity 5 and $24 - 5m$ simple roots, so that T is a $3:1$ cover branched in $24 - 5m + m = 24 - 4m$ points with multiplicity 2. Therefore, by the Riemann–Hurwitz formula, one obtains $2g(T) - 2 = -6 + 24 - 4m$, i.e., $g(T) = 10 - 2m$. Hence $k_2 = 2,\ g_2 = 10 - 2m$ and so $r_2 = 2 + 2m$ and $a_2 = 2m$. □

4C. ***Proof of Theorem 1.1.*** Theorem 1.1 states the existence of an elliptic fibration $\mathcal{E}$ on a certain Calabi–Yau 4-fold Y and contains the Hodge numbers of Y. The construction of the Calabi–Yau 4-fold Y is contained in Section 3C, the existence of the elliptic fibration $\mathcal{E}$ is proved in Proposition 3.1 and the Hodge numbers of Y are given in Proposition 4.2. This concludes the proof of Theorem 1.1.

4D. By construction, Y is obtained as a (desingularization of a) quotient of $S_1 \times S_2$ by $\iota_1 \times \iota_2$, so each complex deformation of the pairs (S_i, ι_i), where S_i is a K3 surface admitting a prescribed nonsymplectic involution ι_i, induces a complex deformation of Y. Since Y is a Calabi–Yau 4-fold, the dimension of the space of its complex deformations is $h^{3,1}(Y)$. By [Nikulin 1979], the dimension of the space of complex deformations of (S_i, ι_i) is $20 - r_i$, $i = 1, 2$. So $h^{3,1}(Y) \geq (20 - r_1) + (20 - r_2)$ and the equality holds if and only if all the deformations of Y are induced by deformations of (S_i, ι_i) (compare with the definition of Borcea–Voisin maximal family in [Cattaneo and Garbagnati 2016], where similar concepts are discussed on Calabi–Yau 3-folds of Borcea–Voisin type). By Proposition 4.2, one has $h^{3,1}(Y) = 137 - 11n - 22m + 2nm$, $r_1 = 1 + n$, $r_2 = 2 + 2m$ and thus $(20-r_1)+(20-r_2) = 37-n-2m$. Hence, $h^{3,1}(Y)$ is strictly bigger than $37-n-2m$ and therefore a part of the deformations of Y are not induced by deformations of the pairs (S_i, ι_i).

Let us now fix the del Pezzo surface dP, and then the K3 surface S_1, with its involution ι_1. The moduli space of K3 surfaces with m fibers of type I_5 as in Proposition 4.2 has dimension $18 - 4m$ (because this is the space of the $(U \oplus A_4^{\oplus m})$-polarized K3 surfaces). So, given a K3 surface S_2 as in Proposition 4.2, the moduli of Y are $137 - 11n - 22m + 2nm$ and the moduli of the K3 surfaces S_2 admitting the prescribed elliptic fibration are $18 - 4m$. In particular, any complex deformation of S_2 which preserves the elliptic fibration induces a complex deformation of Y, but there are a lot of deformations of Y which are not induced by those of S_2.

5. Linear systems on *Y*

5A. Here we state some general results on linear systems on the product of varieties with trivial canonical bundle, which will be applied to $S_1 \times S_2$.

Let X_1 and X_2 be two smooth varieties with trivial canonical bundle, and $\mathcal{L}_{X_1}$ and $\mathcal{L}_{X_2}$ be two line bundles on X_1 and X_2, respectively. Observe that we have a natural injective homomorphism

$$H^0(X_1, \mathcal{L}_{X_1}) \otimes H^0(X_2, \mathcal{L}_{X_2}) \to H^0(X_1 \times X_2, \pi_1^*\mathcal{L}_{X_1} \otimes \pi_2^*\mathcal{L}_{X_2})$$

$$s \otimes t \mapsto \pi_1^* s \cdot \pi_2^* t,$$

where the π_i's are the two projections. We now want to determine some conditions which guarantee that this map is an isomorphism.

Using the Hirzebruch–Riemann–Roch theorem,

$$\chi(X_1 \times X_2, \pi_1^*\mathcal{L}_{X_1} \otimes \pi_2^*\mathcal{L}_{X_2}) = \chi(X_1, \mathcal{L}_{X_1}) \cdot \chi(X_2, \mathcal{L}_{X_2}).$$

If $\mathcal{L}_{X_1}$ and $\mathcal{L}_{X_2}$ are nef and big line bundles such that $\pi_1^*\mathcal{L}_{X_1} \otimes \pi_2^*\mathcal{L}_{X_2}$ is still nef and big, then the above formula and Kawamata–Viehweg vanishing theorem lead to

$$h^0(X_1 \times X_2, \pi_1^*\mathcal{L}_{X_1} \otimes \pi_2^*\mathcal{L}_{X_2}) = h^0(X_1, \mathcal{L}_{X_1}) \cdot h^0(X_2, \mathcal{L}_{X_2}).$$

However, we are interested also in divisors which are not big and nef, therefore we need the following result.

Proposition 5.1. *Let* X_1, X_2 *be two smooth varieties of dimension* n_1 *and* n_2, *respectively. Assume that they have trivial canonical bundle* $\omega_{X_i} = \mathcal{O}_{X_i}$ *and that* $h^{0,n_i-1}(X_i) = 0$. *Let* $D_i \subseteq X_i$ *be a smooth irreducible codimension* 1 *subvariety. Then the canonical map*

$$H^0(X_1, \mathcal{O}_{X_1}(D_1)) \otimes H^0(X_2, \mathcal{O}_{X_2}(D_2)) \xrightarrow{\psi} H^0(X_1 \times X_2, \pi_1^*\mathcal{O}_{X_1}(D_1) \otimes \pi_2^*\mathcal{O}_{X_2}(D_2))$$

is an isomorphism.

Proof. By Künneth's formula

$$\begin{aligned} h^{0,n-1}(X_1 \times X_2) = &\ h^{0,n_1-1}(X_1) \cdot h^{0,n_2}(X_2) + h^{0,n_1}(X_1) \cdot h^{0,n_2-1}(X_2) \\ = &\ h^{0,n_1-1}(X_1) + h^{0,n_2-1}(X_2) = 0, \end{aligned}$$

where $n = n_1 + n_2 = \dim(X_1 \times X_2)$.

As already remarked, the ψ map is injective, so it suffices to show that the source and target spaces have the same dimension.

We begin with the computation of $h^0(X_i, \mathcal{O}_{X_i}(D_i))$. From the exact sequence

$$0 \to \mathcal{O}_{X_i}(-D_i) \to \mathcal{O}_{X_i} \to \mathcal{O}_{D_i} \to 0,$$

we deduce the exact piece

$$H^{n_i-1}(X_i, \mathcal{O}_{X_i}) \to H^{n_i-1}(D_i, \mathcal{O}_{D_i}) \to H^{n_i}(X_i, \mathcal{O}_{X_i}(-D_i)) \to H^{n_i}(X_i, \mathcal{O}_{X_i}) \to 0.$$

Since $H^{n_i-1}(X_i, \mathcal{O}_{X_i}) = 0$ by hypothesis, we get by Serre duality that

$$h^0(X_i, \mathcal{O}_{X_i}(D_i)) = h^{n_i}(X_i, \mathcal{O}_{X_i}(-D_i)) = h^{n_i-1}(D_i, \mathcal{O}_{D_i}) + 1.$$

Now we pass to the computation of $h^0(X_1 \times X_2, \pi_1^*\mathcal{O}_{X_1}(D_1) \otimes \pi_2^*\mathcal{O}_{X_2}(D_2))$. Let $D = D_1 \times X_2 \cup X_1 \times D_2$, and observe that

$$\pi_1^*\mathcal{O}_{X_1}(D_1) \otimes \pi_2^*\mathcal{O}_{X_2}(D_2) = \mathcal{O}_{X_1 \times X_2}(D).$$

By the previous part of the proof, we have

$$h^0(X_1 \times X_2, \pi_1^*\mathcal{O}_{X_1}(D_1) \otimes \pi_2^*\mathcal{O}_{X_2}(D_2)) = h^{n-1}(D, \mathcal{O}_D) + 1,$$

so we need to compute $h^{n-1}(D, \mathcal{O}_D)$ in this situation. Consider the diagram of inclusions

$$\begin{tikzcd}
X_1 \times D_2 \arrow[r, hook, "i_1"] & D \\
D_1 \times D_2 \arrow[u, hook] \arrow[ur, hook, "i"] \arrow[r, hook] & D_1 \times X_2 \arrow[u, hook, "i_2"']
\end{tikzcd}$$

and the short exact sequence

$$0 \to \mathcal{O}_D \to i_{1*}\mathcal{O}_{X_1\times D_2} \oplus i_{2*}\mathcal{O}_{D_1\times X_2} \to i_*\mathcal{O}_{D_1\times D_2} \to 0,$$

where

$$\mathcal{O}_D \to i_{1*}\mathcal{O}_{X_1\times D_2} \oplus i_{2*}\mathcal{O}_{D_1\times X_2}, \qquad s \mapsto (s_{|X_1\times D_2}, s_{|D_1\times X_2})$$

and

$$i_{1*}\mathcal{O}_{X_1\times D_2} \oplus i_{2*}\mathcal{O}_{D_1\times X_2} \to i_*\mathcal{O}_{D_1\times D_2}, \qquad (s_1, s_2) \mapsto s_{1|_{D_1\times D_2}} - s_{2|_{D_1\times D_2}}.$$

This sequence induces the exact piece

$$\begin{aligned} &H^{n-2}(D_1\times D_2, \mathcal{O}_{D_1\times D_2}) \to H^{n-1}(D, \mathcal{O}_D)\\ &\qquad \to H^{n-1}(X_1\times D_2, \mathcal{O}_{X_1\times D_2}) \oplus H^{n-1}(D_1\times X_2, \mathcal{O}_{D_1\times X_2}) \to 0,\end{aligned}$$

from which we have that

$$\begin{aligned} h^{n-1}(D, \mathcal{O}_D) \le h^{n-1}(X_1\times D_2, \mathcal{O}_{X_1\times D_2}) + h^{n-1}(D_1\times X_2, \mathcal{O}_{D_1\times X_2})&\\ + h^{n-2}(D_1\times D_2, \mathcal{O}_{D_1\times D_2}).&\end{aligned}$$

These last numbers are easy to compute using Künneth's formula:

$$\begin{aligned} h^{n-1}(X_1\times D_2, \mathcal{O}_{X_1\times D_2}) &= \sum_{i=0}^{n-1} h^{0,i}(X_1)\cdot h^{0,n-1-i}(D_2)\\ &= h^{0,n_1}(X_1)\cdot h^{0,n_2-1}(D_2) = h^{0,n_2-1}(D_2);\\ h^{n-1}(D_1\times X_2, \mathcal{O}_{D_1\times X_2}) &= h^{0,n_1-1}(D_1);\\ h^{n-2}(D_1\times D_2, \mathcal{O}_{D_1\times D_2}) &= h^{0,n-2}(D_1\times D_2)\\ &= \sum_{i=0}^{n-2} h^{0,i}(D_1)\cdot h^{0,n-2-i}(D_2)\\ &= h^{0,n_1-1}(D_1)\cdot h^{0,n_2-1}(D_2),\end{aligned}$$

where we used the trivial observation that $h^{0,k}(D_i) = 0$ if $k \ge n_i$.

Finally, we have the following chain of inequalities:

$$\begin{aligned} &(h^{n_1-1}(D_1, \mathcal{O}_{D_1}) + 1)(h^{n_2-1}(D_2, \mathcal{O}_{D_2}) + 1)\\ &\qquad = h^0(X_1, \mathcal{O}_{X_1}(D_1))\cdot h^0(X_2, \mathcal{O}_{X_2}(D_2))\\ &\qquad \le h^0(X_1\times X_2, \mathcal{O}_{X_1\times X_2}(D)) = h^{n-1}(D, \mathcal{O}_D) + 1\\ &\qquad \le h^{0,n_1-1}(D_1) + h^{0,n_2-1}(D_2) + h^{0,n_1-1}(D_1)\cdot h^{0,n_2-1}(D_2) + 1\\ &\qquad = (h^{n_1-1}(D_1, \mathcal{O}_{D_1}) + 1)(h^{n_2-1}(D_2, \mathcal{O}_{D_2}) + 1),\end{aligned}$$

from which the proposition follows. □

Remark 5.2. Observe that this proposition can be deduced also from more general arguments; see, for instance, [Kashiwara and Schapira 1990, Exercise II.18] where a broader generalization of the Künneth formula is shown.

5B. In particular, this result applies when X_1 and X_2 are $K3$ surfaces or, more generally, when they are Calabi–Yau or hyperkähler manifolds.

By induction, it is easy to generalize this result to a finite number of factors. Notice that we require D_i to be smooth in order to use Künneth's formula. Indeed, there is a more general version of Proposition 5.1 for line bundles. Namely, if $\mathcal{L}_i$ are globally generated/base point free line bundles over X_i then their linear systems $|\mathcal{L}_i|$ have, by Bertini's theorem, a smooth irreducible member, and we can apply Proposition 5.1.

Let us denote $D_1+D_2 := \pi_1^*\mathcal{O}(D_1)+\pi_2^*\mathcal{O}(D_2)$. The linear system $|D_i|$ naturally defines the map $\varphi_{|D_i|} : X_i \to \mathbb{P}^{n_i}$. Denoting by $\sigma_{n_1,n_2} : \mathbb{P}^{n_1} \times \mathbb{P}^{n_2} \to \mathbb{P}^{n_1 n_2+n_1+n_2}$ the Segre embedding, Proposition 5.1 implies that $\varphi_{|D_1+D_2|}$ coincides with $\sigma_{n_1,n_2} \circ (\varphi_{|D_1|} \times \varphi_{|D_2|})$.

Corollary 5.3. *Let S_i, $i = 1, 2$, be two $K3$ surfaces and D_i be an irreducible smooth curve of genus g_i on S_i. Then $h^0(S_1 \times S_2, D_1 + D_2) = (g_1 + 1)(g_2 + 1)$.*

5C. Use the same notation as in Section 3 diagram (3). On $S_1 \times S_2$, let D be an invariant divisor (resp. an invariant line bundle $\mathcal{D}$) with respect to the $\iota_1 \times \iota_2$ action. Moreover, denote by D_Y the divisor on Y such that $q^*D_Y = b^*D$ (resp. $\mathcal{D}_Y$ is the line bundle such that $q^*\mathcal{D}_Y = b^*\mathcal{D}$).

Since q is a double cover branched along a codimension 1 subvariety B, it is uniquely defined by a line bundle $\mathcal{L}$ on Y such that $\mathcal{L}^{\otimes 2} = \mathcal{O}_Y(B)$ and we have

$$H^0(\widetilde{S_1 \times S_2}, q^*\mathcal{M}) = H^0(Y, \mathcal{M}) \oplus H^0(Y, \mathcal{M} \otimes \mathcal{L}^{\otimes -1})$$

for any line bundle $\mathcal{M}$ on Y.

The isomorphism $H^0(\widetilde{S_1 \times S_2}, b^*\mathcal{D}) \simeq H^0(S_1 \times S_2, \mathcal{D})$ yields

$$H^0(S_1 \times S_2, \mathcal{D}) \simeq H^0(\widetilde{S_1 \times S_2}, q^*\mathcal{D}_Y) \simeq H^0(Y, \mathcal{D}_Y) \oplus H^0(Y, \mathcal{D}_Y \otimes \mathcal{L}^{\otimes -1}).$$

As a consequence, one sees that the space $H^0(Y, \mathcal{D}_Y)$ corresponds to the invariant subspace of $H^0(S_1 \times S_2, \mathcal{D})$ for the ι^* action, while $H^0(Y, \mathcal{D}_Y \otimes \mathcal{L}^{-1})$ corresponds to the anti-invariant one. This yields at once the commutative diagram

$$\begin{tikzcd}
\widetilde{S_1 \times S_2} \arrow[r, "b"] \arrow[d, "q"'] & S_1 \times S_2 \arrow[r, "\varphi_{|\mathcal{D}|}"] \arrow[d] & \mathbb{P}(H^0(S_1 \times S_2, \mathcal{D})^\vee) \arrow[d] \\
Y \arrow[r] \arrow[rr, bend right, "\varphi_{|\mathcal{D}_Y|}"'] & X \arrow[r] & \mathbb{P}(H^0(Y, \mathcal{D}_Y)^\vee),
\end{tikzcd} \tag{5}$$

where the vertical arrow on the right is the projection on $\mathbb{P}(H^0(Y, \mathcal{D}_Y)^\vee)$ with center $\mathbb{P}(H^0(Y, \mathcal{D}_Y \otimes \mathcal{L}^{-1})^\vee)$ (observe that both these two spaces are pointwise fixed for the induced action of ι on $\mathbb{P}(H^0(S_1 \times S_2, \mathcal{D})^\vee)$).

In what follows we denote by D_Y and L the divisors such that $\mathcal{D}_Y = \mathcal{O}(D_Y)$ and $\mathcal{L} = \mathcal{O}(L)$, so L is half of the branch divisor.

5D. Let D_i be a smooth irreducible curve on S_i such that the divisor D_i is invariant for ι_i. Then ι_i^* acts on $H^0(S_i, D_i)^\vee$. Let us denote by $H^0(S_i, D_i)_{\pm 1}$ the eigenspace relative to the eigenvalue ± 1 for the action of ι_i on $H^0(S_i, D_i)$. Let h_i be the dimension of $\mathbb{P}(H^0(S_i, D_i)^\vee_{+1})$.

Corollary 5.4. *Let S_i, D_i, D_Y, L and h_i be as above. Then $\varphi_{|D_Y|} : Y \to \mathbb{P}^N$ where $N := (h_1 + 1)(h_2 + 1) + (g(D_1) - h_1)(g(D_2) - h_2) - 1$ and $\varphi_{|D_Y - L|} : Y \to \mathbb{P}^M$ where $M := (h_1 + 1)(g(D_2) - h_2) + (g(D_1) - h_1)(h_2 + 1) - 1$.*

Proof. By Corollary 5.3 the map $\varphi_{|D_1+D_2|}$ is a map from $S_1 \times S_2$ to the Segre embedding of $\mathbb{P}(H^0(S_1, D_1)^\vee)$ and $\mathbb{P}(H^0(S_2, D_2)^\vee)$. The action of the automorphism $\iota_1 \times \iota_2$ on $H^0(S_1 \times S_2, D_1 + D_2)$ is induced by the action of ι_i on $H^0(S_i, D_i)$ and in particular

$$\begin{aligned} &H^0(S_1 \times S_2, D_1 + D_2)_{+1} \\ &\qquad = H^0(S_1, D_1)_{+1} \otimes H^0(S_2, D_2)_{+1} \oplus H^0(S_1, D_1)_{-1} \otimes H^0(S_2, D_2)_{-1}, \end{aligned}$$

whose dimension is $(h_1+1)(h_2+1)+(g(D_1)-h_1)(g(D_2)-h_2)$. By Section 5C, the divisors D_Y and $D_Y - L$ define on Y two maps whose target space is the projection of $\mathbb{P}(H^0(S_1 \times S_2, D)^\vee)$ to the eigenspaces for the action of $\iota_1 \times \iota_2$ and the image is the projection of $\varphi_{|D|}(S_1 \times S_2)$. So the target space of $\varphi_{|D_Y|}$ is

$$\mathbb{P}(H^0(S_1 \times S_2, D_1 + D_2)^\vee_{+1}),$$

whose dimension is $(h_1+1)(h_2+1)+(g(D_1)-h_1)(g(D_2)-h_2)-1$. One concludes similarly for $\varphi_{|D_Y - L|}$. □

Lemma 5.5. *Let D_i be an effective divisor on S_i invariant for ι_i, and h_i be the dimension of $\mathbb{P}(H^0(S_i, D_i)^\vee_{+1})$ for $i = 1, 2$. Denote by δ_{D_i} the divisor on Y such that $q^*(\delta_{D_i}) = b^*(\pi_i^*(D_i))$. Then*

$$H^0(S_1 \times S_2, \pi_i^*(D_i)) \simeq H^0(S_i, D_i) \quad \textit{and} \quad \dim(\mathbb{P}(H^0(Y, \delta_{D_i}))) = h_i,$$

for $i = 1, 2$.

6. Projective models and fibrations

The aim of this section is to apply the general results of the previous sections to our specific situation. So, let (S_1, ι_1) and (S_2, ι_2) be as in Section 3B (i.e., S_1 is a double cover of *dP*, ι_1 is the cover involution, S_2 is an elliptic fibration and ι_2 is the elliptic involution). To simplify the notation, from now on we assume that the del Pezzo surface dP is a blow up of $\mathbb{P}^2$, so that S_1 is a double cover of $\mathbb{P}^2$. The results of this section can be easily generalized to the case $dP \simeq \mathbb{P}^1 \times \mathbb{P}^1$.

We now consider some interesting divisors on S_1 and S_2.

6A. Let $h \in \operatorname{Pic}(S_1)$ be the pullback of the hyperplane section of $\mathbb{P}^2$ by the generically $2:1$ map $\rho' : S_1 \to \mathbb{P}^2$. The divisor h is a nef and big divisor on S_1 and the map $\varphi_{|h|}$ is generically $2:1$ to the image (which is $\mathbb{P}^2$). The action of ι_1 is the identity on $H^0(S_1, h)^\vee$, since ι_1 is the cover involution.

We recall that the branch locus of ρ' is a sextic with n simple nodes in general position, for $0 \le n \le 8$. As explained in Section 3, in order to construct a smooth double cover we first blow up $\mathbb{P}^2$ at the n nodes of the sextic obtaining a del Pezzo surface dP. Thus on S_1 there are n rational curves, lying over these exceptional curves. We denote these curves by R_i, $i = 1, \dots, n$. We will denote by H the divisor $3h - \sum_{i=1}^n R_i$ if $n \ge 1$ or the divisor $3h$ if $n = 0$. Observe that H is the strict transform of the nodal sextic in $\mathbb{P}^2$.

For a generic choice of S_1 the Picard group of S_1 is generated by h and R_i. The divisor H is an ample divisor, because it has a positive intersection with all the effective -2 classes. Moreover, $H^2 = 18 - 2n > 2$, if $n \le 7$. By [Saint-Donat 1974], this divisor cannot be hyperelliptic and so the map $\varphi_{|H|}$ is $1:1$ onto its image in $\mathbb{P}^{10-n}$.

The divisor $\frac{1}{2}\rho_*(H)$ is the anticanonical divisor of the del Pezzo surface dP, which embeds dP in

$$\mathbb{P}^{9-n} = \mathbb{P}\big(H^0\big(dP, \tfrac{1}{2}\rho_*(H)\big)^\vee\big).$$

Since ι_1 is the cover involution of ρ, the action of ι_1^* on $H^0(S_1, H)^\vee$ has a $(10-n)$-dimensional eigenspace for the eigenvalue $+1$ and a 1-dimensional eigenspace for the eigenvalue -1. Observe that with this description, the projection

$$\mathbb{P}(H^0(S_1, H)^\vee) \to \mathbb{P}(H^0(S_1, H)^\vee_{+1})$$

from the point $\mathbb{P}(H^0(S_1, H)^\vee_{-1})$ coincides with the double cover ρ.

Notably, if $n = 6$, the del Pezzo surface dP is a cubic surface in $\mathbb{P}^3_{(x_0:x_1:x_2:x_3)}$, whose equation is $f_3(x_0 : x_1 : x_2 : x_3) = 0$. In this case the divisor H embeds the K3 surface S_1 in $\mathbb{P}^4$ as the complete intersection of a quadric with equation $x_4^2 = g_2(x_0 : x_1 : x_2 : x_3)$ and the cubic $f_3(x_0 : x_1 : x_2 : x_3) = 0$ and ι_1 acts multiplying x_4 by -1.

6B. Let S_2 be a K3 surface with an elliptic fibration. Generically $\operatorname{Pic}(S_2)$ is spanned by the divisors F and O, the class of the fiber and the class of the section, respectively. If S_2 has some other properties, for example some reducible fibers, then there are other divisors on S_2 linearly independent from F and O. In any case, it is still true that $\langle F, O\rangle$ is primitively embedded in $\mathrm{Pic}(S_2)$. We consider two divisors on S_2: F and $4F+2O$.

The divisor F is by definition the class of the fiber of the elliptic fibration on S_2, so that $\pi = \varphi_{|F|} : S_2 \to \mathbb{P}^1$ is the elliptic fibration on S_2. In particular F is a nef divisor, but it is not big, and it is invariant for ι_2 (since ι_2 preserves the fibration). Moreover, ι_2 preserves each fiber of the fibration, therefore ι_2^* acts as the identity on $H^0(S_2, F)^\vee$.

It is easy to see that the divisor $4F+2O$ is a nef and big divisor. The map $\varphi_{|4F+2O|}$ contracts the zero section and possibly the nontrivial components of the reducible fibers of the fibration. We see that

$$\varphi_{|4F+2O|} : S_2 \xrightarrow{2:1} \varphi_{|4F+2O|}(S_2)$$

is a double cover, where $\varphi_{|4F+2O|}(S_2)$ is the cone over a rational normal curve of degree 4 in $\mathbb{P}^5$. Blowing up the vertex of $\varphi_{|4F+2O|}(S_2)$ we obtain a surface isomorphic to the Hirzebruch surface $\mathbb{F}_4$. The involution ι_2 is the associated cover involution; this means that ι_2^* acts as the identity on $H^0(S_2, 4F+2O)^\vee$.

6C. We observe that the divisors h, H, F and $4F+2O$ are invariant for the action of ι_i for some i. So, by Corollary 5.4, we get the following:

Proposition 6.1. *Let Y and the divisors on Y be as above, then*:

(1) *The map*

$$\begin{tikzcd} \varphi_{|(h+F)_Y|}: & Y \arrow[rr] \arrow[dr] & & \mathbb{P}^5 \\ & & \mathbb{P}^2\times\mathbb{P}^1 \arrow[ur, "\sigma_{2,1}"'] & \end{tikzcd}$$

is an elliptic fibration on the image of $\mathbb{P}^2\times\mathbb{P}^1$ *by the Segre embedding.*

(2) *The map*

$$\begin{tikzcd} \varphi_{|(H+F)_Y|}: & Y \arrow[rr] \arrow[dr] & & \mathbb{P}^{19-2n} \\ & & \mathbb{P}^{9-n}\times\mathbb{P}^1 \arrow[ur, "\sigma_{9-n,1}"'] & \end{tikzcd}$$

is the same elliptic fibration as in (1) *with a different projective model of the basis, i.e., the image of $dP\times\mathbb{P}^1$ via $\sigma_{9-n,1}$.*

(3) *The map*

$$\begin{tikzcd} \varphi_{|(h+(4F+2O))_Y|}: & Y \arrow[rr] \arrow[dr] & & \mathbb{P}^{17} \\ & & \mathbb{P}^2\times\mathbb{P}^5 \arrow[ur, "\sigma_{2,5}"'] & \end{tikzcd}$$

is a generically $2:1$ *map onto its image contained in* $\sigma_{2,5}(\mathbb{P}^2\times\mathbb{P}^5)$.

(4) *The map*

$$\begin{tikzcd}
\varphi_{|(H+(4F+2O))_Y|}: & Y \arrow[rr] \arrow[dr] & & \mathbb{P}^{59-6n} \\
& & \mathbb{P}^{9-n}\times\mathbb{P}^5 \arrow[ur, "\sigma_{9-n,5}"'] &
\end{tikzcd}$$

is birational onto its image contained in $\sigma_{9-n,5}(\mathbb{P}^{9-n}\times\mathbb{P}^5)$.

Proof. The points (1) and (2) are proved in Section 6D. The points (3) and (4) are proved in Section 6E. □

Proposition 6.2. *Using the same notation as for* Lemma 5.5 *we have*:

(1) $\varphi_{|\delta_h|}: Y\to\mathbb{P}^2$ *is an isotrivial fibration in* $K3$ *surfaces whose generic fiber is isomorphic to* S_2.

(2) $\varphi_{|\delta_H|}: Y\to\mathbb{P}^{9-n}$ *is the same fibration as in* (1) *with a different projective model of the basis.*

(3) $\varphi_{|\delta_F|}: Y\to\mathbb{P}^1$ *is a fibration in Calabi–Yau* 3*-folds whose generic fiber is the Borcea–Voisin of the* $K3$ *surface* S_1 *and the elliptic fiber of the fibration* π.

(4) $\varphi_{|\delta_{4F+2O}|}: Y\to\mathbb{P}^5$ *is an isotrivial fibration in* $K3$ *surfaces whose generic fiber is isomorphic to* S_1.

Proof. The proof is explained in Section 6D, where all the previous maps are described in detail. □

6D. ***Fibrations on Y.*** As the natural map $\rho'\times\pi: S_1\times S_2\to\mathbb{P}^2\times\mathbb{P}^1$ satisfies $(\rho\times\pi)\circ\iota=\rho\times\pi$, we have an induced map $X\to\mathbb{P}^2\times\mathbb{P}^1$. The composition of this map with the resolution $Y\to X$ and with the two projections then gives

(1) an elliptic fibration $\mathcal{E}: Y\to\mathbb{P}^2\times\mathbb{P}^1$,

(2) a K3-fibration $\mathcal{G}: Y\to\mathbb{P}^2$,

(3) a fibration in elliptically fibered 3-folds $\mathcal{H}: Y\to\mathbb{P}^1$.

We describe these fibrations:

(1) The map $\mathcal{E}: Y\to\mathbb{P}^2\times\mathbb{P}^1$ is induced by the divisor $(h+F)_Y$ since $\varphi_{|h|}: S_1\to\mathbb{P}^2$ and $\varphi_{|F|}: S_2\to\mathbb{P}^1$. We already described the properties and the singular fibers for this fibration in Section 3D.

The composition of $\varphi_{|H|}(S_1)$ and the projection to the invariant subspace of $\mathbb{P}^{10-n}$ exhibits S_1 as the double cover of the del Pezzo surface dP anticanonically embedded in $\mathbb{P}^{9-n}$. The del Pezzo surface dP is the blow up of $\mathbb{P}^2$ in n points and the double cover $S_1\to dP$ corresponds (after the blow up) to the double cover $\varphi_{|h|}: S_1\to\mathbb{P}^2$ since $H=3h-\sum_{i=1}^n R_i$. Thus, the map $\varphi_{|(H+F)_Y|}$ is the same fibration as $\varphi_{|(h+F)_Y|}$, with a different model for the basis (which is now $dP\times\mathbb{P}^1$).

(2) The map $\mathcal{G} : Y \to \mathbb{P}^2$ is induced by δ_h. The fibers of these fibrations are isomorphic to S_2 since we have the commutative diagram

$$\begin{tikzcd}
S_1 \times S_2 \arrow[r] \arrow[d, "/\iota_1 \times \iota_2"] & S_1 \arrow[d, "\rho"] \arrow[dr, "\varphi_{|h|}"] & \\
X \arrow[r] & dP \arrow[r] & \mathbb{P}^2
\end{tikzcd}$$

The singular fibers of $\mathcal{G}$ lie over the branch curve $C \subset \mathbb{P}^2$ of the double cover $S_1 \to \mathbb{P}^2$. Let $P \in C$. It is easy to see that $(\rho' \times \pi)^{-1}(pr_{\mathbb{P}^2}^{-1}(P))$ is given by $P \times S_2$, and so in the quotient X we see a surface isomorphic to S_2/ι_2, which is a surface obtained from $\mathbb{F}_4$ by means of blow ups. Moreover, under the blow up $Y \to X$ we add a certain number of ruled surfaces: these are all disjoint from each other, and meet the blow up of $\mathbb{F}_4$ on the base curve of the rulings, i.e., on the section O, on the trisection T and possibly on the rational fixed components E_i (which are necessarily contained in reducible not-reduced fibers).

For the same reason as above, $\varphi_{|\delta_H|}$ is the fibration $\mathcal{G}$ with a different description of the basis.

(3) The fibration $\mathcal{H}$ is induced by δ_F. For every $t \in \mathbb{P}^1$, we denote by F_t the elliptic fiber of $S_2 \to \mathbb{P}^1$ over t. The inclusion $S_1 \times F_t \subset S_1 \times S_2$ induces

$$\begin{tikzcd}
 & S_1 \times F_t \arrow[r, hook] \arrow[d, "/\iota_1 \times (\iota_2)_{|F_t}"] & S_1 \times S_2 \arrow[d, "/\iota_1 \times \iota_2"] & \\
BV(S_1, F_t) \arrow[r] & (S_1 \times F_t)/(\iota_1 \times (\iota_2)_{|F_t}) \arrow[r, hook] & X & Y \arrow[l]
\end{tikzcd}$$

So the fibers of φ_{δ_F} are Borcea–Voisin Calabi–Yau 3-folds which are elliptically fibered by definition. The singular fibers lie on $\Delta(\pi)$.

(4) Moreover, there is another K3-fibration. Indeed, the map $\varphi_{|\delta_{4F+2O}|}$ gives an isotrivial fibration in K3 surfaces isomorphic to S_1 and with basis the cone over the rational normal curve in $\mathbb{P}^4$, by the diagram

$$\begin{tikzcd}
S_1 \times S_2 \arrow[r] \arrow[d, "/\iota_1 \times \iota_2"] & S_2 \arrow[d, "/\iota_2"] & \\
X \arrow[r] & (S_2/\iota_2) \arrow[r] & \mathbb{P}^5
\end{tikzcd}$$

6E. ***Projective models.*** By the diagram

$$\begin{tikzcd}
 & S_1 \times S_2 \arrow[r, "\varphi_{|h|} \times \varphi_{|4F+2O|}", "4:1"'] \arrow[d, "2:1"'] & \mathbb{P}^2 \times \mathbb{P}^5 \arrow[r, hook, "\sigma_{2,5}"'] & \mathbb{P}^{17} \\
Y \arrow[r] & X \arrow[ur, "2:1"'] & &
\end{tikzcd}$$

we can describe the map induced by the linear system $|(h+4F+2O)_Y|$ on Y as a double cover of the image (under the Segre embedding of the ambient spaces) of $\varphi_{|h|}(S_1)\times\varphi_{|4F+2O|}(S_2)$, which is the product of $\mathbb{P}^2$ with the cone over the rational normal curve of degree 4. This map is generically $2:1$, and its branch locus is given by the union of the product of the sextic curve in $\mathbb{P}^2$ with the vertex of the cone (the fiber over such points is a curve) and the product of the sextic with the trisection; the generic fiber is a single point, but there may be points where the fiber is a curve. The last case occurs only if the fibration $\pi : S_2 \to \mathbb{P}^1$ has reducible nonreduced fibers.

To describe the map induced by $|(H+4F+2O)_Y|$ we use the diagram

$$\begin{tikzcd}
 & S_1\times S_2 \arrow[r, "\varphi_{|H|}\times\varphi_{|4F+2O|}", "2:1"'] \arrow[d, "2:1"'] & \mathbb{P}^{10-n}\times\mathbb{P}^5 \arrow[r, hook, "\sigma_{10-n,5}"'] \arrow[d] & \mathbb{P}^{65-6n} \\
Y \arrow[r] & X \arrow[r, "\varphi_{|(H+4F+2O)_Y|}", "1:1"'] & \mathbb{P}^{9-n}\times\mathbb{P}^5 \arrow[r, hook, "\sigma_{9-n,5}"'] & \mathbb{P}^{59-6n}
\end{tikzcd}$$

where

$$\mathbb{P}^{10-n}\times\mathbb{P}^5 \to \mathbb{P}^{9-n}\times\mathbb{P}^5$$

is induced by the projection of $\mathbb{P}^{10-n}=\mathbb{P}(H^0(S_1, H)^\vee)$ to $\mathbb{P}(H^0(S_1, H)^\vee_{+1})$. Recall that H is an ample divisor on S_1 (indeed, it is very ample), so the image of $\varphi_{|H|}\times\varphi_{|4F+2O|}$ is the product of S_1 and the cone over the rational normal curve of degree 4. Observe that generically this map is $2:1$, and so it descends to a $1:1$ map on X and on Y. So $\varphi_{|(H+4F+2O)_Y|}$ maps Y onto the product of dP with the cone over the rational normal curve of degree 4.

7. Explicit equations of *Y*

The aim of this section is to give some explicit equations for the projective models described above, in terms of the corresponding equations for S_i.

With a slight abuse, in this section we will substitute $\mathbb{F}_4$ to its singular model as the cone on the rational normal curve of degree 4. In this way we will obtain better models for Y.

7A. If S_1 is the double cover of $\mathbb{P}^2_{(x_0:x_1:x_2)}$ we assume its equation to be

$$w^2=f_6(x_0:x_1:x_2) \tag{6}$$

so that the curve C is $V(f_6(x_0:x_1:x_2))$. We assume that C is irreducible, even if some of the following results can be easily generalized. The cover involution ι_1 acts as $(w;(x_0:x_1:x_2))\mapsto(-w;(x_0:x_1:x_2))$.

If S_1 is the double cover of $\mathbb{P}^1_{(x_0:x_1)}\times\mathbb{P}^1_{(x_2:x_3)}$ we assume its equation to be

$$w^2=f_{4,4}((x_0:x_1),(x_2:x_3)) \tag{7}$$

so that the curve C is $V(f_{4,4}((x_0:x_1),(x_2:x_3)))$.

In the following we give the details of our computations under the assumption that S_1 is a cover of $\mathbb{P}^2$, and we only state the main results in the case where S_1 is a cover of $\mathbb{P}^1 \times \mathbb{P}^1$.

7B. Before giving the description of S_2, we make a little digression on the Weierstrass equation of an elliptic fibration. In particular, let $Y \to V$ be an elliptic fibration and

$$y^2 = x^3 + Ax + B \tag{8}$$

be an equation for its Weierstrass model. The condition that Y is a Calabi–Yau variety is equivalent to

$$A \in H^0(V, -4K_V), \qquad B \in H^0(V, -6K_V).$$

The discriminant Δ is then an element of $H^0(V, -12K_V)$.

In particular if V is $\mathbb{P}^m$ (resp. $\mathbb{P}^n \times \mathbb{P}^m$), the functions A, B and Δ are homogeneous polynomials of degree $4m+4$, $6m+6$ and $12m+12$ (resp. of bidegree $(4n+4, 4m+4)$, $(6n+6, 6m+6)$ and $(12n+12, 12m+12)$).

We observe that, if V is $\mathbb{P}^m$ (resp. $\mathbb{P}^n \times \mathbb{P}^m$), requiring that all the singular fibers of the elliptic fibration (8) are of type I_5 implies that $m \equiv 4 \mod 5$ (resp. $n \equiv 4 \mod 5$ and $m \equiv 4 \mod 5$). When V is a 3-fold, this gives a stronger version of Remark 3.2.

7C. Let S_2 be the elliptic K3 surface whose Weierstrass equation is

$$y^2 = x^3 + A(t:s)x + B(t:s), \tag{9}$$

where (according to the previous section) $A(t:s)$ and $B(t:s)$ are homogeneous polynomials of degree 8 and 12, respectively. For generic choices of $A(t:s)$ and $B(t:s)$, the elliptic fibration (9) has 24 nodal curves as unique singular fibers. For specific choices one can obtain other singular and reducible fibers. The cover involution ι_2 acts as

$$(y, x; (t:s)) \mapsto (-y, x; (t:s)).$$

Equivalently S_2 is the double cover of the Hirzebruch surface $\mathbb{F}_4$ given by

$$u^2 = z(x^3 + A(t:s)xz^2 + B(t:s)z^3), \tag{10}$$

where the coordinates (t, s, x, z) are the homogeneous toric coordinates of $\mathbb{F}_4$; see, e.g., [Cattaneo and Garbagnati 2016, §2.3]. The action of ι_2 on these coordinates is $(u, t, s, x, z) \mapsto (-u, t, s, x, z)$. Observe that the curve on $\mathbb{F}_4$ defined by

$$z(x^3 + A(t:s)xz^2 + B(t:s)z^3) = 0$$

is linearly equivalent to $-2K_{\mathbb{F}_4}$.

7C1. The choice of particular polynomials in (9) is associated to the choice of particular fibers of the fibration. Indeed, this elliptic fibration has an I_5-fiber in $(\bar{t}:\bar{s})$ if and only if the following three conditions hold:

(1) $A(\bar{t}:\bar{s}) \neq 0$.

(2) $B(\bar{t}:\bar{s}) \neq 0$.

(3) Δ vanishes of order 5 in $(\bar{t}:\bar{s})$, where $\Delta := 4A^3+27B^2$.

Up to standard transformations one can assume that the fiber of type I_5 is over $t=0$ and

$$A(t:s) := t^8 + \sum_{i=1}^{7} a_i t^i s^{8-i} - 3s^8,$$

$$B(t:s) := b_{12}t^{12} + \sum_{i=5}^{11} b_i t^i s^{12-i} + \left(-a_4 + \frac{a_1^4}{1728} + \frac{a_3 a_1}{6} + \frac{a_2^2}{12} + \frac{a_2 a_1^2}{72}\right) t^4 s^8 + \left(-a_3 + \frac{a_2 a_1}{6} + \frac{a_1^3}{216}\right) t^3 s^9 + \left(-a_2 + \frac{a_1^2}{12}\right) t^2 s^{10} - a_1 t^1 s^{11} + 2s^{12}.$$

We observe that the polynomials $A(t:s)$ and $B(t:s)$ depend on 14 parameters and, indeed, 14 is exactly the dimension of the family of K3 surfaces whose generic member has an elliptic fibration with one fiber of type I_5.

We already noticed that an elliptic fibration on a K3 surface has at most four fibers of type I_5 and indeed there are two distinct families of K3 surfaces with this property: the Mordell–Weil group of the generic member of one of these families is trivial, the one of the other is $\mathbb{Z}/5\mathbb{Z}$, [Shimada 2000, Case 2345, Table 1].

The K3 surfaces of the latter family are known to be double covers of the extremal rational elliptic surface [1, 1, 5, 5] whose Mordell–Weil group is $\mathbb{Z}/5\mathbb{Z}$; see [Schütt and Shioda 2010, Section 9.1] for the definition of the extremal rational elliptic surface. By this property, it is easy to find the Weierstrass equation of the K3 surface (as described in [Balestrieri et al. 2018, Section 4.2.2]). Indeed, the equation of the rigid rational fibration over $\mathbb{P}^1_{(\mu)}$ is

$$y^2 = x^3 + A(\mu)x + B(\mu), \tag{11}$$

where

$$A(\mu) := -\tfrac{1}{48}\mu^4 - \tfrac{1}{4}\mu^3\lambda - \tfrac{7}{24}\mu^2\lambda^2 + \tfrac{1}{4}\mu\lambda^3 - \tfrac{1}{48}\lambda^4,$$
$$B(\mu) := \tfrac{1}{864}\mu^6 + \tfrac{1}{48}\mu^5\lambda + \tfrac{25}{288}\mu^4\lambda^2 + \tfrac{25}{288}\mu^2\lambda^4 - \tfrac{1}{48}\mu\lambda^5 + \tfrac{1}{864}\lambda^6.$$

In order to obtain the two-dimensional family of K3 surfaces we are looking for, it suffices to apply a base change of order two $f:\mathbb{P}^1_{(t:s)} \to \mathbb{P}^1_{(\mu:\lambda)}$ to the rational elliptic surface. In particular, if f branches over $(p_1:1)$ and $(p_2:1)$, the base change $\mu = p_1t^2+s^2$, $\lambda = t^2+s^2/p_2$ produces the required K3 surface if the fibers over $(p_1 : 1)$ and $(p_2 : 1)$ of the rational elliptic surface are smooth.

7D. ***The elliptic fibration $\mathcal{E}$.*** Let us now consider (6) for S_1 and (9) for S_2. The action of $\iota_1 \times \iota_2$ on $S_1 \times S_2$ leaves the functions $Y := yw^3$, $X := xw^2$, x_0, x_1, x_2, t, s invariant. Hence an equation for a birational model of Y expressed in these coordinates is

$$Y^2 = X^3 + A(t:s)f_6^2(x_0:x_1:x_2)X + B(t:s)f_6^3(x_0:x_1:x_2). \tag{12}$$

The previous equation is a Weierstrass form for the elliptic fibration

$$\mathcal{E} : Y \to \mathbb{P}^2_{(x_0:x_1:x_2)} \times \mathbb{P}^1_{(t:s)}.$$

Observe that the coefficient $A(t:s)f_6^2(x_0:x_1:x_2)$ and $B(t:s)f_6^3(x_0:x_1:x_2)$ are bihomogeneous on $\mathbb{P}^2 \times \mathbb{P}^1$ of bidegree (12, 8) and (18, 12) respectively, so by Section 7B we have another proof that the total space of the elliptic fibration $\mathcal{E}$ is indeed a Calabi–Yau variety.

One can check the properties of this fibration described in Section 3D directly by the computation of the discriminant of the Weierstrass equation (12), indeed

$$\Delta(\mathcal{E}) = f_6^6(x_0:x_1:x_2)(4A^3(t:s)+27B^2(t:s)) = f_6^6(x_0:x_1:x_2)\Delta(\pi).$$

In this birational model, the basis of the fibration is $\mathbb{P}^2 \times \mathbb{P}^1$ and the del Pezzo surface contained in the discriminant is the blow up of $\mathbb{P}^2$ in the singular points of $f_6(x_0 : x_1 : x_2)$. The singular fibers due to the factor $\Delta(\pi)$ in $\Delta(\mathcal{E})$ are not generically modified by the blow up of $\mathbb{P}^2$ in n points, so that over the generic point of $\mathbb{P}^2$ (and thus of the del Pezzo surface), the singular fibers of $\mathcal{E}$ correspond to singular fibers of π.

If the equation of S_1 is (7), the Weierstrass equation of $\mathcal{E}$ is

$$Y^2 = X^3 + A(t:s)f_{4,4}^2((x_0:x_1),(x_2:x_3))X + B(t:s)f_{4,4}^3((x_0:x_1),(x_2:x_3)). \tag{13}$$

In some special cases it is also possible to write more explicitly a Weierstrass form of this elliptic fibration with basis the product of the del Pezzo surface and $\mathbb{P}^1_{(t:s)}$, as we see in Sections 7D1 and 7D2.

Remark 7.1. A generalization of this construction produces 4-folds with Kodaira dimension equal to $-\infty$ (resp. > 0) with an elliptic fibration. Indeed, it suffices to consider S_2 which is no longer a K3 surface, but a surface with Kodaira dimension $-\infty$ (resp. > 0) admitting an elliptic fibration with basis $\mathbb{P}^1$. So the equation of S_2 is $y^2 = x^3 + A(t:s)x + B(t:s)$ with $\deg(A(t:s)) = 4m$ and $\deg(B(t:s)) = 6m$ for $m = 1$ (resp. $m > 2$). The surface S_2 admits the elliptic involution ι_2 and $(S_1 \times S_2)/\iota_1 \times \iota_2$ admits a Weierstrass equation analogous to (12) or to (13).

7D1. *The case $n=6$.* Let us assume that $C\subset\mathbb{P}^2$ has $n=6$ nodes in general position. In this case the del Pezzo surface dP has degree 3 and is canonically embedded as a cubic in $\mathbb{P}^3_{(y_0:y_1:y_2:y_3)}$. So it admits an equation of the form $g_3(y_0:y_1:y_2:y_3)=0$. The image of C under this embedding is the complete intersection of $g_3=0$ and a quadric $g_2(y_0:y_1:y_2:y_3)=0$ in $\mathbb{P}^3$.

The K3 surface S_1 is embedded by $\varphi_{|H|}$ in $\mathbb{P}^4(y_0:y_1:y_2:y_3:y_4)$ as the complete intersection of a cubic and a quadric, and since it is the double cover of dP, its equation is

$$\begin{cases} y_4^2=g_2(y_0:y_1:y_2:y_3),\\ 0=g_3(y_0:y_1:y_2:y_3).\end{cases} \tag{14}$$

The involution ι_1 acts on $\mathbb{P}^4$, changing only the sign of y_4.

With the same argument as before, this leads to the following equation for a birational model of Y:

$$\begin{cases} Y^2=X^3+A(t:s)g_2^2(y_0:y_1:y_2:y_3)X+B(t:s)g_2^3(y_0:y_1:y_2:y_3),\\ g_3(y_0:y_1:y_2:y_3)=0.\end{cases} \tag{15}$$

The first equation is the Weierstrass form of an elliptic fibration with basis $\mathbb{P}^3\times\mathbb{P}^1$ and the second equation corresponds to restricting this equation to the del Pezzo surface embedded in the first factor (i.e., in $\mathbb{P}^3$).

Corollary 7.2. *The equation*

$$\begin{cases} Y^2=X^3+\left(\sum\limits_{i=0}^{8}a_it^is^{8-i}\right)g_2^2(y_0:y_1:y_2:y_3)X+\left(\sum\limits_{i=0}^{12}b_it^is^{12-i}\right)g_2^3(y_0:y_1:y_2:y_3),\\ g_3(y_0:y_1:y_2:y_3)=0,\end{cases}$$

where g_i is a homogenous polynomial of degree i in $\mathbb{C}[y_0:y_1:y_2:y_3]$,

$$a_0=-3,\quad b_0=2,\quad b_1=-a_1,\quad b_2=-a_2+\frac{a_1^2}{12},$$
$$b_3=-a_3+\frac{a_2a_1}{6}+\frac{a_1^3}{216}\qquad b_4=-a_4+\frac{a_1^4}{1728}+\frac{a_3a_1}{6}+\frac{a_2^2}{12}+\frac{a_2a_1^2}{72},$$

describes a birational model of a Calabi–Yau 4-fold with an elliptic fibration such that the fibers over the del Pezzo surface $(g_3(y_0:y_1:y_2:y_3)=0)\times(t=0)\subset\mathbb{P}^3\times\mathbb{P}^1_t$ are generically of type I_5.

The other singular fibers are described by the zeros of the discriminant

$$g_2^6(y_0:y_1:y_2:y_3)\left(4\left(\sum_{i=0}^{8}a_it^is^{8-i}\right)^3+27\left(\sum_{i=0}^{12}b_it^is^{12-i}\right)^2\right).$$

Remark 7.3. With the same process one obtains the equation of an elliptic fibration over $dP \times \mathbb{P}^1$ such that there are $m \leq 4$ del Pezzo surfaces in $dP \times \mathbb{P}^1$ over each of which the general fiber is of type I_5. To do this it suffices to specialize the coefficients a_i and b_i according to the conditions described in Section 7C1. In the case $m = 4$ there are two different specializations; one of them is associated to the presence of a 5-torsion section and its equation is the given in Section 7C1.

7D2. *The case $n = 5$.* The treatment of the case $n = 5$ is similar to that for $n = 6$. So let us assume that $C \subset \mathbb{P}^2$ has $n = 5$ nodes in general position. In this case the del Pezzo surface dP has degree 4 and is canonically embedded in $\mathbb{P}^4_{(y_0:y_1:y_2:y_3:y_4)}$ as the complete intersection of two quadrics $q_2 = 0$ and $q_2' = 0$. The image of C under this embedding is the complete intersection of the del Pezzo with a quadric $q_2'' = 0$.

The K3 surface S_1 is embedded by $\varphi_{|H|}$ in $\mathbb{P}^5(y_0 : y_1 : y_2 : y_3 : y_4 : y_5)$ as the complete intersection of three quadrics, and since it is the double cover of dP, its equation is

$$\begin{cases} y_5^2 = q_2''(y_0 : y_1 : y_2 : y_3 : y_4), \\ 0 = q_2'(y_0 : y_1 : y_2 : y_3 : y_4), \\ 0 = q_2(y_0 : y_1 : y_2 : y_3 : y_4). \end{cases} \tag{16}$$

The involution ι_1 acts on $\mathbb{P}^5$ changing only the sign of y_5.

Hence a birational model of Y is:

$$\begin{cases} Y^2 = X^3 + A(t:s)q_2''^2(y_0:y_1:y_2:y_3:y_4)X + B(t:s)q_2''^3(y_0:y_1:y_2:y_3:y_4), \\ q_2'(y_0:y_1:y_2:y_3:y_4) = 0, \\ q_2(y_0:y_1:y_2:y_3:y_4) = 0. \end{cases} \tag{17}$$

The first equation is the Weierstrass form of an elliptic fibration with basis $\mathbb{P}^4 \times \mathbb{P}^1$ and the other two equations restrict this equation to the del Pezzo surface embedded in the first factor (i.e., in $\mathbb{P}^4$).

Remark 7.4. It is possible to obtain explicit equations for the elliptic fibrations with fiber(s) of type I_5 as in Corollary 7.2.

7E. ***The double cover $Y \to \mathbb{P}^2 \times \mathbb{F}_4$.*** Let us consider the equations (6) for S_1 and (10) for S_2. The functions

$$W := uw, \quad x_0, \quad x_1, \quad x_2, \quad t, \quad s, \quad x, \quad z$$

are invariant for $\iota_1 \times \iota_2$ and they satisfy the equation

$$W^2 = f_6(x_0 : x_1 : x_2) z (x^3 + A(t:s)xz^2 + B(t:s)z^3). \tag{18}$$

This equation exhibits a birational model of Y as a double cover of the rational

4-fold $\mathbb{P}^2 \times \mathbb{F}_4$ branched over a divisor in $|-2K_{\mathbb{P}^2\times\mathbb{F}_4}|$. In particular this is the equation associated to the linear system $|(h+4F+2O)_Y|$.

The projections of (18) give different descriptions of projective models: the one associated to the linear system $|\delta_h|$ is obtained by the projection to $\mathbb{P}^2$; the one associated to $|\delta_{4F+2O}|$ is obtained by the projection to $\mathbb{F}_4 \subset \mathbb{P}^5$; the one associated to the linear system $|\delta_F|$ is obtained by the projection to $\mathbb{P}^1_{(t:s)}$.

Consider first the composition with the projection on $\mathbb{P}^2$ to obtain an equation for $\mathcal{G}$. Fix a point $(\bar{x}_0 : \bar{x}_1 : \bar{x}_2) \in \mathbb{P}^2$ and assume that $f_6(\bar{x}_0 : \bar{x}_1 : \bar{x}_2) \neq 0$. Then the corresponding fiber has equation

$$W^2 = f_6(\bar{x}_0 : \bar{x}_1 : \bar{x}_2)z(x^3 + A(t:s)xz^2 + B(t:s)z^3),$$

which is easily seen to be isomorphic to S_2 (substitute W with $\sqrt{f_6(\bar{x}_0 : \bar{x}_1 : \bar{x}_2)}W$ to find an equation equivalent to (10)).

Consider now the composition with the projection on $\mathbb{F}_4$. Fix a point $(\bar{t}, \bar{s}, \bar{x}, \bar{z}) \in \mathbb{F}_4$ which does not lie on the negative curve nor on the trisection. Then the corresponding fiber is

$$W^2 = f_6(x_0 : x_1 : x_2)\bar{z}(\bar{x}^3 + A(\bar{t}:\bar{s})\bar{x}\bar{z}^2 + B(\bar{t}:\bar{s})\bar{z}^3),$$

which is a K3 surface isomorphic to S_1.

Finally we give an equation for $\mathcal{H}$. Let us put $z = 1$ in (18) and perform the change of coordinates $w \mapsto w/f_6$, $x \mapsto x/f_6$. Multiplying the resulting equation by f_6^2, we obtain

$$w^2 = x^3 + A(t:s)f_6^2(x_0 : x_1 : x_2)x + B(t:s)f_6^3(x_0 : x_1 : x_2).$$

For every fixed $(\bar{t} : \bar{s}) \in \mathbb{P}^1$, this is the equation of a Calabi–Yau 3-fold of Borcea–Voisin type obtained from the K3 surface $w^2 = f_6(x_0 : x_1 : x_2)$ and the elliptic curve $y^2 = x^3 + A(\bar{t}:\bar{s})x + B(\bar{t}:\bar{s})$; see [Cattaneo and Garbagnati 2016, Section 4.4].

7E1. We now want to describe what happens if the sextic curve in $\mathbb{P}^2$ has $n = 6$ or $n = 5$ nodes.

Assume first that $\rho' : S_1 \to \mathbb{P}^2$ is branched along a sextic with 6 nodes. Then we can use (14) and (10) to describe S_1 and S_2, respectively, and using the same argument as before (i.e., put $W = y_4u$) we obtain the equation

$$\begin{cases} W^2 = g_2(y_0 : y_1 : y_2 : y_3)z(x^3 + A(t:s)xz^2 + B(t:s)z^3), \\ 0 = g_3(y_0 : y_1 : y_2 : y_3), \end{cases}$$

which exhibits Y as the double cover of $dP \times \mathbb{F}^4$. Let us denote by $U \to \mathbb{P}^3 \times \mathbb{F}_4$ the double cover branched on $g_2(y_0 : y_1 : y_2 : y_3)z(x^3 + A(t:s)xz^2 + B(t:s)z^3)$. The branch divisor is $2H_{\mathbb{P}^3} - 2K_{\mathbb{F}_4}$ and so Y is a section of the anticanonical bundle of U.

With a further change of variables, where the only nonidentic transformations are $W' = g_2 W$ and $x' = g_2 x$, we then find the following equation for a birational model of Y (we drop the primes for simplicity of notation):

$$\begin{cases} W^2 = z(x^3 + A(t:s)g_2^2(y_0:y_1:y_2:y_3)xz^2 + B(t:s)g_2^3(y_0:y_1:y_2:y_3)z^3), \\ 0 = g_3(y_0:y_1:y_2:y_3). \end{cases}$$

Here the first equation gives an elliptic fibration over $\mathbb{P}^3 \times \mathbb{P}^1$ as a double cover, while the second restricts this fibration to $dP \times \mathbb{P}^1$.

Analogously, if $n = 5$, then S_1 and S_2 are described by (16) and (10), respectively, so that we have the following equation for Y:

$$\begin{cases} W^2 = q_2'' z(x^3 + Axz^2 + Bz^3), \\ 0 = q_2', \\ 0 = q_2, \end{cases}$$

with the same considerations as the case just treated.

7F. ***An involution on Y.*** By construction Y admits an involution ι induced by $\iota_1 \times \mathrm{id} \in \mathrm{Aut}(S_1 \times S_2)$ and acting as -1 on $H^{4,0}(Y)$. Since

$$\iota_1 \times \mathrm{id} = (\iota_1 \times \iota_2) \circ (\mathrm{id} \times \iota_2),$$

ι is equivalently induced by $\mathrm{id} \times \iota_2$. The involution ι has a clear geometric interpretation in several models described above. By Section 6E, Y is a $2:1$ cover of $\mathbb{P}^2 \times \mathbb{F}_4$ whose equation is given in (18). The involution ι is the cover involution, indeed it acts as -1 on the variable $W := uw$, and by (6) the map $\iota_1 \times \mathrm{id}$ acts as -1 on w.

By Section 6D, Y admits the elliptic fibration $\mathcal{E}$ whose equation is given in (12). The involution ι is the elliptic involution, indeed it acts as -1 on the variable $Y := yw^3$, and by (9) the map $\mathrm{id} \times \iota_2$ acts as -1 on y.

Hence Y/ι is birational to $\mathbb{P}^2 \times \mathbb{F}_4$ and admits a fibration in rational curves, whose fibers are the quotient of the fibers of the elliptic fibration $\mathcal{E}$.

Acknowledgements

The authors would like to thank Sergio Cacciatori and Gilberto Bini for suggesting this problem at the workshop “*Interazioni fra Geometria algebrica e Fisica teorica*”, Villa Grumello, Como, January 2016. They would also like to thank Lidia Stoppino and Matteo Bonfanti for useful conversations and suggestions.

The authors wish to thank the referees for many detailed comments and suggestions that considerably improved the presentation of these results.

Cattaneo gratefully acknowledges the Dipartimento di Scienza e Alta Tecnologia of Università degli Studi dell'Insubria, where the majority of this work was written. Garbagnati is partially supported by FIRB 2012 "*Moduli spaces and their applications*". Penegini is partially supported by Progetto MIUR di Rilevante Interesse Nazionale "*Geometria delle Varietà Algebriche e loro Spazi di Moduli*" PRIN 2015. The authors were also partially supported by GNSAGA of INdAM.

References

[Balestrieri et al. 2018] F. Balestrieri, J. Desjardins, A. Garbagnati, C. Maistret, C. Salgado, and I. Vogt, "Elliptic fibrations on covers of the elliptic modular surface of level 5", pp. 159–197 in *Women in numbers Europe, II* (Leiden, Netherlands, 2016), edited by I. Bouw et al., Association for Women in Math. **11**, Springer, 2018.

[Beasley et al. 2009a] C. Beasley, J. J. Heckman, and C. Vafa, "GUTs and exceptional branes in F-theory, I", *J. High Energy Phys.* **2009**:1 (2009), art. id. 058. MR Zbl

[Beasley et al. 2009b] C. Beasley, J. J. Heckman, and C. Vafa, "GUTs and exceptional branes in F-theory, II: Experimental predictions", *J. High Energy Phys.* **2009**:1 (2009), art. id. 059. MR Zbl

[Bini and Penegini 2017] G. Bini and M. Penegini, "New fourfolds from F-theory", *Math. Nachr.* **290**:5-6 (2017), 699–709. MR Zbl

[Borcea 1997] C. Borcea, "$K3$ surfaces with involution and mirror pairs of Calabi–Yau manifolds", pp. 717–743 in *Mirror symmetry, II*, edited by B. Greene and S.-T. Yau, AMS/IP Stud. Adv. Math. **1**, Amer. Math. Soc., Providence, RI, 1997. MR Zbl

[Braun 2011] V. Braun, "On free quotients of complete intersection Calabi–Yau manifolds", *J. High Energy Phys.* **2011**:4 (2011), art. id. 005. MR Zbl

[Braun et al. 2013] V. Braun, T. W. Grimm, and J. Keitel, "Geometric engineering in toric F-theory and GUTs with U(1) gauge factors", *J. High Energy Phys.* **2013**:12 (2013), art. id. 069.

[Cattaneo and Garbagnati 2016] A. Cattaneo and A. Garbagnati, "Calabi–Yau 3-folds of Borcea–Voisin type and elliptic fibrations", *Tohoku Math. J.* (2) **68**:4 (2016), 515–558. MR Zbl

[Constantin et al. 2017] A. Constantin, J. Gray, and A. Lukas, "Hodge numbers for all CICY quotients", *J. High Energy Phys.* **2017**:1 (2017), art. id. 001. MR Zbl

[Cynk and Hulek 2007] S. Cynk and K. Hulek, "Higher-dimensional modular Calabi–Yau manifolds", *Canad. Math. Bull.* **50**:4 (2007), 486–503. MR Zbl

[Dillies 2012] J. Dillies, "Generalized Borcea–Voisin construction", *Lett. Math. Phys.* **100**:1 (2012), 77–96. MR Zbl

[Dolgachev 2012] I. V. Dolgachev, *Classical algebraic geometry: a modern view*, Cambridge Univ. Press, 2012. MR Zbl

[Kapustka 2015] G. Kapustka, "Projections of del Pezzo surfaces and Calabi–Yau threefolds", *Adv. Geom.* **15**:2 (2015), 143–158. MR Zbl

[Kashiwara and Schapira 1990] M. Kashiwara and P. Schapira, *Sheaves on manifolds*, Grundlehren der Math. Wissenschaften **292**, Springer, 1990. MR Zbl

[Nikulin 1979] V. V. Nikulin, "Quotient-groups of groups of automorphisms of hyperbolic forms of subgroups generated by 2-reflections", *Dokl. Akad. Nauk SSSR* **248**:6 (1979), 1307–1309. In Russian; translated in *Soviet Math. Dokl.* **20**:5 (1979), 1156–1158. MR Zbl

[Nowak 1997] K. J. Nowak, "Flat morphisms between regular varieties", *Univ. Iagel. Acta Math.* **35** (1997), 243–246. MR Zbl

[Saint-Donat 1974] B. Saint-Donat, "Projective models of K-3 surfaces", *Amer. J. Math.* **96** (1974), 602–639. MR Zbl

[Schütt and Shioda 2010] M. Schütt and T. Shioda, "Elliptic surfaces", pp. 51–160 in *Algebraic geometry in East Asia* (Seoul, 2008), edited by J. Keum et al., Adv. Stud. Pure Math. **60**, Math. Soc. Japan, Tokyo, 2010. MR Zbl arXiv

[Shimada 2000] I. Shimada, "On elliptic $K3$ surfaces", *Michigan Math. J.* **47**:3 (2000), 423–446. MR Zbl

[Voisin 1993] C. Voisin, "Miroirs et involutions sur les surfaces $K3$", pp. 273–323 in *Journées de Géométrie Algébrique d'Orsay* (Orsay, 1992), Astérisque **218**, Société Mathématique de France, Paris, 1993. MR Zbl

Received July 13, 2017. Revised February 26, 2018.

ANDREA CATTANEO
UNIVERSITÉ CLAUDE BERNARD LYON 1
INSTITUT CAMILLE JORDAN
VILLEURBANNE
FRANCE

cattaneo@math.univ-lyon1.fr

ALICE GARBAGNATI
DIPARTIMENTO DI MATEMATICA FEDERIGO ENRIQUES
UNIVERSITÀ DEGLI STUDI DI MILANO
MILANO
ITALY

alice.garbagnati@unimi.it

MATTEO PENEGINI
DIPARTIMENTO DI MATEMATICA - DIMA
UNIVERSITÀ DEGLI STUDI DI GENOVA
GENOVA
ITALY

penegini@dima.unige.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 1, 2019
dx.doi.org/10.2140/pjm.2019.299.33

PARTIAL REGULARITY OF HARMONIC MAPS FROM A RIEMANNIAN MANIFOLD INTO A LORENTZIAN MANIFOLD

JIAYU LI AND LEI LIU

We study the partial regularity theorem for stationary harmonic maps from a Riemannian manifold into a Lorentzian manifold. For a weakly stationary harmonic map (u, v) from a smooth bounded open domain $\Omega \subset \mathbb{R}^m$ to a Lorentzian manifold with Dirichlet boundary condition, we prove that it is smooth outside a closed set whose $(m-2)$-dimensional Hausdorff measure is zero. Moreover, if the target manifold N does not admit any harmonic spheres S^l, $l = 2, \dots, m-1$, we show (u, v) is smooth.

1. Introduction

Suppose (M, g) and (N, h_N) are two compact Riemannian manifolds of dimensions m and n respectively. For a map $u \in C^1(M, N)$, the energy functional of u is defined as

$$E(u) = \frac{1}{2}\int_M |\nabla u|^2 \, d\,\mathrm{vol}_g\,. \tag{1-1}$$

A critical point of the energy functional E is called a harmonic map. By Nash's embedding theorem, we can embed N isometrically into some Euclidean space $\mathbb{R}^K$ and the corresponding Euler–Lagrange equation is

$$\Delta_g u = A(u)(\nabla u, \nabla u),$$

where Δ_g is the Laplace–Beltrami operator on M with respect to g and A is the second fundamental form of $N \subset \mathbb{R}^K$.

Harmonic map is a very important notion in geometric analysis which has been widely studied in the past decades. Physically, harmonic maps come from the nonlinear sigma model, which plays an important role in quantum field and string theory. From the perspective of general relativity, it is natural to consider the targets of harmonic maps to be Lorentzian manifolds. Geometrically, the link between harmonic maps into S_1^4 and the conformal Gauss maps of Willmore surfaces in S^3

The research is supported by NSF in China No. 11426236, 11131007.
MSC2010: 53C43, 58E20.
Keywords: Lorentzian harmonic map, stationary, partial regularity, blow-up.

also leads to such harmonic maps [Bryant 1984]. The work on minimal surfaces in anti-de Sitter spaces and its applications in theoretical physics also shows the importance of such maps [Alday and Maldacena 2009]. In this paper, we shall focus on the interior partial regularity of stationary harmonic maps from a compact Riemannian manifold of dimension m (≥ 3) into a Lorentzian manifold.

We now proceed to introduce the model. Let $N \times \mathbb{R}$ be a Lorentzian manifold equipped with a warped product metric

$$h = h_N - \beta (d\theta)^2,$$

where $(\mathbb{R}, d\theta^2)$ is the standard 1-dimensional Euclidean space and β is a positive smooth function on (N, h_N). Since N is compact, there exist positive constants λ_1 and λ_2 such that

$$0 < \lambda_1 \leq \beta(y) \leq \lambda_2 < \infty \quad \text{and} \quad |\nabla \beta(y)| \leq \lambda_2 \quad \text{for all } y \in N.$$

Set

$$W^{1,2}(M, N \times \mathbb{R}) := \{u \in W^{1,2}(M, \mathbb{R}^K),\ v \in W^{1,2}(M, \mathbb{R}) \mid u(x) \in N \text{ for a.e. } x \in M\}.$$

For $(u, v) \in W^{1,2}(M, N \times \mathbb{R})$, we consider the functional

$$E_h(u, v; M) = \frac{1}{2} \int_M \{|\nabla u|^2 - \beta(u)|\nabla v|^2\} \, d\,\mathrm{vol}_g, \tag{1-2}$$

which is called the Lorentzian energy of the map (u, v) on M. A critical point (u, v) of the functional (1-2) is called a harmonic map from (M, g) into the Lorentzian manifold $(N \times \mathbb{R}, h)$.

When the target manifold is a Lorentzian manifold, the existence of geodesics was studied in [Benci et al. 1991] and Greco [1993; 1997] constructed a smooth harmonic map via some developed variational methods. Recently, Han, Jost, Liu and Zhao [Han et al. 2019] investigated a parabolic-elliptic system for maps and got a global existence result by assuming either some geometric conditions on the target manifold or small energy of the initial maps. The result implies the existence of a harmonic map in a given homotopy class. The blowup behavior for Lorentzian harmonic maps was studied in [Han et al. 2017b], and for approximate Lorentzian harmonic maps and Lorentzian harmonic maps, flow from a Riemann surface was studied in [Han et al. 2019; 2017a]. For the global weak solution of Lorentzian harmonic map flow, one can refer to [Han et al. 2018]. The regularity theory was studied in [Isobe 1998; Zhu 2013] for dimension 2 and in [Isobe 1997] for higher dimensions on some kinds of minimal type solutions.

Via direct calculations, Zhu [2013] derived the Euler–Lagrange equations for (1-2),

$$\begin{cases} -\Delta u = A(u)(\nabla u, \nabla u) - B^{\top}(u)|\nabla v|^2 & \text{in } M, \\ -\operatorname{div}(\beta(u)\nabla v) = 0, & \text{in } M, \end{cases} \tag{1-3}$$

where A is the second fundamental form of N in $\mathbb{R}^K$, $B(u) := (B^1, B^2, \dots, B^K)$ with

$$B^j := -\frac{1}{2}\frac{\partial\beta(u)}{\partial y^j}$$

and $B^\top$ is the tangential part of B along the map u.

Definition 1.1. We call $(u, v) \in W^{1,2}(\Omega, N \times \mathbb{R})$ a weakly Lorentzian harmonic map with Dirichlet boundary data

$$(u, v)|_{\partial\Omega} = (\phi, \psi)$$

if it is a weak solution of (1-3) with boundary data (ϕ, ψ).

Similar to harmonic maps, we introduce the notion of stationary Lorentzian harmonic maps.

Definition 1.2. A weakly Lorentzian harmonic map $(u, v) \in W^{1,2}(\Omega, N \times \mathbb{R})$ is called a stationary Lorentzian harmonic map if it is also a critical point of E_h with respect to the domain variations; i.e., for any $Y \in C_0^\infty(\Omega, \mathbb{R}^m)$, it holds

$$\frac{d}{dt}\Big|_{t=0}\int_\Omega \frac{1}{2}(|\nabla u_t|^2 - \beta(u_t)|\nabla v_t|^2)\, d\,\mathrm{vol}_g = 0,$$

where $u_t(x) = u(x + tY(x))$ and $v_t(x) = v(x + tY(x))$.

Our first main result is the following small-energy regularity theorem.

Theorem 1.3. *For $m \geq 2$ and any $\alpha \in (0, 1)$, there exists an $\epsilon_0 > 0$ depending only on m, α and (N, h_N) such that if $(u, v) \in W^{1,2}(\Omega, N \times \mathbb{R})$ is a weakly Lorentzian harmonic map satisfying*

$$\sup_{x\in B_{r_0}(x_0),\, 0<r\leq r_0} r^{2-m}\int_{B_r(x)} |\nabla u|^2\, d\,\mathrm{vol}_g \leq \epsilon_0^2, \tag{1-4}$$

then $(u, v) \in C^\infty(B_{r_0/2}(x_0))$. Moreover, it satisfies the estimate

$$\begin{aligned} r_0\|\nabla u\|_{L^\infty(B_{r_0/2}(x_0))} + r_0\|\nabla v\|_{L^\infty(B_{r_0/2}(x_0))} + r_0^{1+\alpha}\|\nabla u\|_{C^\alpha(B_{r_0/2}(x_0))} \\ + r_0^{1+\alpha}\|\nabla v\|_{C^\alpha(B_{r_0/2}(x_0))} \\ \leq C\big(r_0^{1-m/2}\|(\nabla u, \nabla v)\|_{L^2(B_{r_0}(x_0))} + r_0^{2-m}\|(\nabla u, \nabla v)\|^2_{L^2(B_{r_0}(x_0))} \\ + r_0^{4-2m}\|\nabla v\|^4_{L^2(B_{r_0}(x_0))}\big), \end{aligned} \tag{1-5}$$

where $C = C(m, \lambda_1, \lambda_2, \alpha, N)$ is a positive constant and

$$\|(\nabla u, \nabla v)\|^2_{L^2(B_{r_0}(x_0))} := \|\nabla u\|^2_{L^2(B_{r_0}(x_0))} + \|\nabla v\|^2_{L^2(B_{r_0}(x_0))}.$$

In this paper, we can get the following interior partial regularity theorem. For a similar result for harmonic maps, one can refer to [Bethuel 1993; Evans 1991; Li and Tian 1998]. For results on gauge theory, one can refer to [Tian 2000].

Theorem 1.4. *For $m \geq 2$, let $(u, v) \in W^{1,2}(\Omega, N \times \mathbb{R})$ be a stationary Lorentzian harmonic map with Dirichlet boundary data $(u, v)|_{\partial\Omega} = (\phi, \psi)$, where $\psi \in C^1(\partial\Omega)$. Then there exists a closed subset $S(u) \subset \Omega$, with $H^{m-2}(S(u)) = 0$, such that $(u, v) \in C^\infty(\Omega \setminus S(u))$.*

Remark 1.5. The boundary assumption $\psi \in C^1(\partial\Omega)$ is used to derive the estimate $\|v\|_{W^{1,p}(\Omega)}$ for some $p > m$. See Lemma 2.1. In fact, by the classical theory of the Laplace operator and the following proof in this paper, one may find that it is enough to assume that $\psi \in W^{1-1/p,p}(\partial\Omega)$ for some $p > m$.

Furthermore, we have:

Theorem 1.6. *Under the same assumption as the above theorem, if N does not admit harmonic spheres S^l, $l = 2, \ldots, m-1$, then (u, v) is smooth.*

To prove the partial regularity results, we first need to establish the monotonicity formula for stationary Lorentzian harmonic maps. Thanks to the elliptic estimates of the v-equation of divergence forms, we can control the additional terms (corresponding to harmonic maps) in the monotonicity formula. Secondly, we need to study the energy concentration set of a blow-up sequence of stationary Lorentzian harmonic maps. Here, we follow Lin's scheme [1999] to get the first bubble which is a nonconstant harmonic sphere. The proof is based on the analysis of defect measure using geometric measure theory.

The rest of paper is organized as follows. In Section 2, we establish the monotonicity formula for stationary Lorentzian harmonic maps which is crucial in the proof of our main theorems. In Section 3, we prove the small-energy regularity theorem, Theorem 1.3, and then the partial regularity theorem, Theorem 1.4, follows immediately from a standard monotonicity formula argument. Theorem 1.6 will be proved in Section 4.

2. Monotonicity formula

In this section, we firstly derive the monotonicity formula for stationary Lorentzian harmonic maps. Secondly, for reader's convenience, we recall a regularity theorem in [Sharp 2014] which will be used in the proof.

Thanks to the divergence structure of v-equation, we have the following estimate.

Lemma 2.1. *Let $(u, v) \in W^{1,2}(\Omega, N \times \mathbb{R})$ be a weakly Lorentzian harmonic map with Dirichlet boundary data (ϕ, ψ), where $\psi \in C^1(\partial\Omega)$. Then $v \in W^{1,p}(\Omega)$ for any $1 < p < \infty$ and*

$$\|\nabla v\|_{L^p} \leq C(p, \lambda_1, \lambda_2, \Omega)\|\psi\|_{C^1(\partial\Omega)}. \tag{2-1}$$

Proof. Let v be the unique smooth solution of the equation

$$\begin{cases} \Delta v = 0 & \text{in } \Omega, \\ v(x) = \psi & \text{on } \partial\Omega, \end{cases}$$

which satisfies

$$\|v\|_{C^1(\bar{\Omega})} \le C(\Omega)\|\psi\|_{C^1(\partial\Omega)}.$$

We call v an extension of ψ and for simplicity, we still denote it by $\psi \in C^1(\bar{\Omega})$. It is easy to see that $v - \psi \in W_0^{1,2}(\Omega)$ is a weak solution of

$$-\operatorname{div}(\beta(u)\nabla(v-\psi)) = \operatorname{div}(\beta(u)\nabla\psi).$$

By the standard theory of the second elliptic operator of divergence forms, see Theorem 1 in [Meyers 1963], we obtain that $v \in W^{1,p}$ for any $1 < p < \infty$ and satisfies

$$\|\nabla v\|_{L^p} \le C(p, \lambda_1, \lambda_2, \Omega)\|\nabla\psi\|_{L^p} \le C(p, \lambda_1, \lambda_2, \Omega)\|\psi\|_{C^1(\partial\Omega)}. \qquad \square$$

Next, we derive the stationary identity for stationary Lorentzian harmonic maps.

Lemma 2.2. *Let* $(u, v) \in W^{1,2}(\Omega, N \times \mathbb{R})$ *be a weakly Lorentzian harmonic map. Then* (u, v) *is stationary if and only if for any* $Y \in C_0^\infty(\Omega, \mathbb{R}^m)$, *there holds*

$$\int_\Omega \left(\left\langle \frac{\partial u}{\partial x^\alpha}, \frac{\partial u}{\partial x^\gamma} \right\rangle - \beta(u) \left\langle \frac{\partial v}{\partial x^\alpha}, \frac{\partial v}{\partial x^\gamma} \right\rangle - \frac{1}{2}(|\nabla u|^2 - \beta(u)|\nabla v|^2)\delta_{\alpha\gamma} \right) \frac{\partial Y^\gamma}{\partial x^\alpha}\, dx = 0. \tag{2-2}$$

Proof. For any $Y \in C_0^\infty(\Omega, \mathbb{R}^m)$, let $t \in \mathbb{R}$ small enough and $y = F_t(x) := x + tY(x)$ and $x = F_t^{-1}(y)$. By Definition 1.2, (u, v) is stationary if and only if

$$\frac{d}{dt}\Big|_{t=0} \int_\Omega \frac{1}{2}(|\nabla u_t|^2 - \beta(u_t)|\nabla v_t|^2)\, dx = 0,$$

where $u_t(x) = u(F_t(x))$ and $v_t(x) = v(F_t(x))$.

On the one hand, by a standard calculation, see, e.g., [Lin and Wang 2008], we have

$$\frac{d}{dt}\Big|_{t=0} \frac{1}{2} \int_\Omega |\nabla u_t|^2\, dx = \int_\Omega \left(\left\langle \frac{\partial u}{\partial x^\alpha}, \frac{\partial u}{\partial x^\gamma} \right\rangle - \frac{1}{2}|\nabla u|^2 \delta_{\alpha\gamma} \right) \frac{\partial Y^\gamma}{\partial x^\alpha}\, dx. \tag{2-3}$$

On the other hand, computing directly, we obtain

$$\begin{aligned} \frac{d}{dt}\Big|_{t=0} &\left(\frac{1}{2}\beta(u_t)|\nabla v_t|^2 \right) \\ &= \frac{1}{2}\frac{\partial \beta(u)}{\partial x^\alpha} Y^\alpha |\nabla v|^2 + \beta(u) \left\langle \frac{\partial v}{\partial x^\alpha}, \frac{\partial v}{\partial x^\gamma} \right\rangle \frac{\partial Y^\gamma}{\partial x^\alpha} + \beta(u) \left\langle \frac{\partial^2 v}{\partial x^\alpha \partial x^\gamma}, \frac{\partial v}{\partial x^\gamma} \right\rangle Y^\alpha \\ &= \frac{1}{2}\frac{\partial}{\partial x^\alpha}(\beta(u)|\nabla v|^2) Y^\alpha + \beta(u) \left\langle \frac{\partial v}{\partial x^\alpha}, \frac{\partial v}{\partial x^\gamma} \right\rangle \frac{\partial Y^\gamma}{\partial x^\alpha}. \end{aligned}$$

Thus,

$$\frac{d}{dt}\Big|_{t=0}\frac{1}{2}\int_{\Omega}\beta(u_t)|\nabla v_t|^2\,dx=\int_{\Omega}\beta(u)\left(\left\langle\frac{\partial v}{\partial x^{\alpha}},\frac{\partial v}{\partial x^{\gamma}}\right\rangle-\frac{1}{2}|\nabla v|^2\delta_{\alpha\gamma}\right)\frac{\partial Y^{\gamma}}{\partial x^{\alpha}}\,dx. \tag{2-4}$$

Combining (2-3) with (2-4), we will get the conclusion of the lemma. □

Now, we can derive the monotonicity formula for stationary Lorentzian harmonic maps.

Lemma 2.3. *Let $(u, v) \in W^{1,2}(\Omega, N \times \mathbb{R})$ be a stationary Lorentzian harmonic map. Then for any $x_0 \in \Omega$ and $0 < r_1 \le r_2 < \operatorname{dist}(x_0, \partial\Omega)$, there holds*

$$\begin{aligned} r_2^{2-m}\int_{B_{r_2}(x_0)}(|\nabla u|^2-\beta(u)|\nabla v|^2)\,dx-r_1^{2-m}\int_{B_{r_1}(x_0)}(|\nabla u|^2-\beta(u)|\nabla v|^2)\,dx \\ =2\int_{B_{r_2}(x_0)\setminus B_{r_1}(x_0)}|x-x_0|^{2-m}\left(\left|\frac{\partial u}{\partial r}\right|^2-\beta(u)\left|\frac{\partial v}{\partial r}\right|^2\right)dx, \end{aligned} \tag{2-5}$$

where $\partial_r = \partial/\partial r = \partial/\partial|x - x_0|$.

Proof. For simplicity, we assume $x_0 = 0 \in \Omega$. For any $\epsilon > 0$ and $0 < r < \operatorname{dist}(0, \partial\Omega)$, let $\varphi_\epsilon(x) = \varphi_\epsilon(|x|) \in C_0^\infty(B_r)$ be such that

$$0 \le \varphi_\epsilon(x) \le 1 \quad \text{and} \quad \varphi_\epsilon(x)|_{B_{(1-\epsilon)r}} = 1.$$

Taking $Y(x) = x\varphi_\epsilon(x)$ in the formula (2-2) and noting that

$$\frac{\partial Y^{\gamma}}{\partial x^{\alpha}}=\varphi_\epsilon(x)\delta_{\alpha,\gamma}+\frac{x^{\alpha}x^{\gamma}}{|x|}\varphi_\epsilon'(x),$$

we have

$$\begin{aligned}\left(1-\frac{m}{2}\right)\int_{B_r}(|\nabla u|^2-\beta(u)|\nabla v|^2)\varphi_\epsilon(x)\,dx \\ =\int_{B_r}\left(-\left|\frac{\partial u}{\partial r}\right|^2+\beta(u)\left|\frac{\partial v}{\partial r}\right|^2+\frac{1}{2}(|\nabla u|^2-\beta(u)|\nabla v|^2)\right)|x|\varphi_\epsilon'(x)\,dx.\end{aligned}$$

Letting $\epsilon \to 0$, we get

$$\begin{aligned}(2-m)\int_{B_r}(|\nabla u|^2-\beta(u)|\nabla v|^2)\,dx+r\int_{\partial B_r}(|\nabla u|^2-\beta(u)|\nabla v|^2) \\ =2r\int_{\partial B_r}\left(\left|\frac{\partial u}{\partial r}\right|^2-\beta(u)\left|\frac{\partial v}{\partial r}\right|^2\right),\end{aligned}$$

which yields

$$\frac{d}{dr}\left(r^{2-m}\int_{B_r}(|\nabla u|^2-\beta(u)|\nabla v|^2)\,dx\right)=r^{2-m}\int_{\partial B_r}\left(\left|\frac{\partial u}{\partial r}\right|^2-\beta(u)\left|\frac{\partial v}{\partial r}\right|^2\right).$$

The conclusion of the lemma follows by integrating r from r_1 to r_2. □

As a direct corollary of above monotonicity formula, we have:

Corollary 2.4. *Let* $(u, v) \in W^{1,2}(\Omega, N \times \mathbb{R})$ *be a stationary Lorentzian harmonic map with Dirichlet boundary data* (ϕ, ψ). *Then for any* $x_0 \in \Omega$ *and* $0 < r_1 \le r_2 < \operatorname{dist}(x_0, \partial\Omega)$, *there holds*

$$r_1^{2-m} \int_{B_{r_1}(x_0)} |\nabla u|^2 \, dx \\ \le r_2^{2-m} \int_{B_{r_2}(x_0)} |\nabla u|^2 d\, x + C(m, p, \lambda_1, \lambda_2, \Omega, \|\psi\|_{C^1(\partial\Omega)})(r_2)^{2-2m/p}.$$

Proof. By Lemma 2.3, we have

$$\begin{aligned} r_1^{2-m} \int_{B_{r_1}(z)} |\nabla u|^2 \, dx & \\ \le r_2^{2-m} \int_{B_{r_2}(x_0)} (|\nabla u|^2 - \beta(u)|\nabla v|^2) \, dx &+ r_1^{2-m} \int_{B_{r_1}(x_0)} \beta(u)|\nabla v|^2 \, dx \\ &+ 2 \int_{B_{r_2}(x_0)} |x - x_0|^{2-m} \beta(u) | \frac{\partial v}{\partial |x - x_0|} |^2 \, dx \\ \le r_2^{2-m} \int_{B_{r_2}(x_0)} |\nabla u|^2 \, dx &+ C(m, \lambda_2)(r_2)^{2-2m/p} \|\nabla v\|_{L^p}^2 \\ \le r_2^{2-m} \int_{B_{r_2}(x_0)} |\nabla u|^2 \, dx &+ C(m, p, \lambda_1, \lambda_2, \Omega, \|\psi\|_{C^1(\partial\Omega)})(r_2)^{2-2m/p}, \end{aligned}$$

where the second inequality follows from Young's inequality:

$$\tag{2-6} \int_{B_r} |x|^{2-m} |\nabla v|^2 \, dx \le \|\nabla v\|_{L^p}^2 \| |x|^{2-m} \|_{L^{p/(p-2)}(B_r)} \\ \le C(m, p, \lambda_1, \lambda_2, \Omega, \|\psi\|_{C^1(\partial\Omega)})(r)^{2-2m/p}. \qquad \square$$

In the end of this section, we want to recall a regularity theorem for a system of critical PDE in [Sharp 2014]. Systems of this form were introduced and studied by [Rivière and Struwe 2008]. For this, let us first recall the definition of Morrey spaces; see [Giaquinta 1983].

Definition 2.5. For $p \ge 1,\ 0 < \mu \le m$, and a domain $U \subset \mathbb{R}^m$, the Morrey space $M^{p,\mu}(U)$ is defined by

$$M^{p,\mu}(U) := \{ f \in L^p_{\mathrm{loc}}(U) \mid \|f\|_{M^{p,\mu}(U)} < \infty \},$$

where

$$\|f\|^p_{M^{p,\mu}(U)} := \sup_{B_r \subset U} r^{\mu - m} \int_{B_r} |f|^p.$$

Theorem 2.6 [Sharp 2014, Theorem 1.2]. *For every* $m \ge 2$ *and* $p \in \left(\frac{m}{2}, m\right)$, *there exists* $\epsilon = \epsilon(m, d, p) > 0$ *and* $C = C(m, d, p) > 0$ *with the following property.*

Suppose that $u \in W^{1,2}(B_1, \mathbb{R}^d)$, $\nabla u \in M^{2,2}(B_1, \mathbb{R}^d)$, $\Omega \in M^{2,2}(B_1, \mathrm{so}(d) \otimes \wedge^1 \mathbb{R}^m)$ *and* $f \in L^p(B_1, \mathbb{R}^d)$ *satisfy*

$$\Delta u = \Omega \cdot \nabla u + f \quad in\ x B_1 \tag{2-7}$$

weakly. If $\|\Omega\|_{M^{2,2}(B_1)} \le \epsilon$, *then*

$$\|\nabla^2 u\|_{M^{2p/m,2}(B_{1/2})} + \|\nabla u\|_{M^{2p/(m-p),2}(B_{1/2})} \le C(\|u\|_{L^1(B_1)} + \|f\|_{L^p(B_1)}).$$

3. Proofs of Theorems 1.3 and 1.4

Proof of Theorem 1.3. Without loss of generality, we may assume $r_0 = 1$ and

$$\frac{1}{|B_1|} \int_{B_1} v \, dx = 0.$$

Take a cut-off function $\eta \in C_0^\infty(B_1)$ such that $0 \le \eta \le 1$, $\eta|_{B_{7/8}} \equiv 1$ and $|\nabla \eta| \le C$. By a direct computation, we get

$$\operatorname{div}(\beta(u)\nabla(\eta v)) = \operatorname{div}(\beta(u)\nabla\eta v) + \beta(u)\nabla\eta\nabla v \quad \text{in } B_1.$$

Then according to the standard theory of the second elliptic operator of divergence forms, see Theorem 1 in [Meyers 1963], we have $v \in W^{1,2m/(m-2)}(B_{7/8})$ and

$$\begin{aligned}\|\nabla v\|_{L^{2m/(m-2)}(B_{7/8})} &\le C(m, \lambda_1, \lambda_2)(\|\nabla \eta v\|_{L^{2m/(m-2)}(B_1)} + \|\beta(u)\nabla\eta\nabla v\|_{L^2(B_1)}) \\ &\le C(m, \lambda_1, \lambda_2)\|\nabla v\|_{L^2(B_1)},\end{aligned}$$

where the last inequality follows from Sobolev's embedding $W^{1,2} \hookrightarrow L^{2m/(m-2)}$ and Poincaré's inequality

$$\|v\|_{L^2(B_1)} \le C(m)\|\nabla v\|_{L^2(B_1)}.$$

Using Theorem 1 in [Meyers 1963] and by a bootstrap argument, it is easy to see that $v \in W^{1,p}(B_{3/4})$ for any $1 < p < \infty$ and

$$\|\nabla v\|_{L^p(B_{3/4})} \le C(m, p, \lambda_1, \lambda_2)\|\nabla v\|_{L^2(B_1)}. \tag{3-1}$$

It is well known that the equation of u can be written in the form of (2-7) with

$$|\Omega| \le C(N)|\nabla u| \quad \text{and} \quad |f| \le C(\lambda_2, N)|\nabla v|^2.$$

By Theorem 2.6 and (3-1), taking $\epsilon_0 = \epsilon_0(m, p, N)$ sufficiently small, we know $u \in W^{1,p}(B_{5/8})$ for any $m < p < \infty$ and

$$\begin{aligned}\|\nabla u\|_{L^p(B_{5/8})} &\le C(m, p, \lambda_1, \lambda_2, N)(\|\nabla u\|_{L^2(B_1)} + \||\nabla v|^2\|_{L^{mp/(2+p)}(B_1)}) \\ &\le C(m, p, \lambda_1, \lambda_2, N)(\|\nabla u\|_{L^2(B_1)} + \|\nabla v\|^2_{L^2(B_1)}).\end{aligned}$$

Applying $W^{2,p}$ estimates of the Laplacian operator, we obtain

$$\begin{aligned}
&\|\nabla u\|_{W^{1,p}(B_{9/16})}\\
&\quad\le C(m,p,\lambda_2,N)(\|\nabla u\|^2_{L^{2p}(B_{5/8})}+\|\nabla v\|^2_{L^{2p}(B_{5/8})}+\|\nabla u\|_{L^2(B_{5/8})})\\
&\quad\le C(m,p,\lambda_1,\lambda_2,N)(\|\nabla u\|_{L^2(B_{5/8})}+\|\nabla u\|^2_{L^2(B_1)}+\|\nabla v\|^2_{L^2(B_1)}+\|\nabla v\|^4_{L^2(B_1)})
\end{aligned}$$

and

$$\begin{aligned}
\|\nabla v\|_{W^{1,p}(B_{9/16})}&\le C(m,p,\lambda_1,\lambda_2,N)(\||\nabla u||\nabla v|\|_{L^p(B_{5/8})}+\|\nabla v\|_{L^2(B_{5/8})})\\
&\le C(m,p,\lambda_1,\lambda_2,N)\|\nabla v\|_{L^2(B_1)}(1+\|\nabla u\|_{L^2(B_1)}+\|\nabla v\|^2_{L^2(B_1)}).
\end{aligned}$$

By Sobolev's embedding theorem, we see that $(\nabla u, \nabla v) \in C^\alpha(B_{9/16})$ for any $\alpha = 1-m/p \in (0,1)$ and the estimate (1-5) holds. Then the high regularity follows from the classical Schauder estimates of the Laplacian operator and a standard bootstrap argument. □

Now, we prove our main theorem, Theorem 1.4.

Proof of Theorem 1.4. Define

$$S(u) := \left\{ x\in\Omega \;\middle|\; \liminf_{r\searrow 0} r^{2-n}\int_{B_r(x)} |\nabla u|^2 \ge \frac{\epsilon_0^2}{2^m}\right\}, \tag{3-2}$$

where $\epsilon_0 > 0$ is the constant in Theorem 1.3. It is well known that $H^{n-2}(S(u)) = 0$. Next, we will show $S(u)$ is a closed set and $(u, v) \in C^\infty(\Omega\setminus S(\phi))$.

For any $x_0 \in \Omega \setminus S(u)$ and $\epsilon > 0$, there exists $0 < r_0 < \epsilon$ such that

$$(2r_0)^{2-m}\int_{B_{2r_0}(x_0)} |\nabla u|^2\,dx < \frac{\epsilon_0^2}{2^m}. \tag{3-3}$$

Therefore,

$$\sup_{z\in B_{r_0}(x_0)} r_0^{2-m}\int_{B_{r_0}(z)} |\nabla u|^2\,dx \le r_0^{2-m}\int_{B_{2r_0}(x_0)} |\nabla u|^2\,dx < \frac{2^{m-2}\epsilon_0^2}{2^m}. \tag{3-4}$$

By Corollary 2.4, we have

$$\begin{aligned}
&\sup_{z\in B_{r_0}(x_0),\,0<r\le r_0} r^{2-m}\int_{B_r(z)} |\nabla u|^2\,dx\\
&\qquad\le \sup_{z\in B_{r_0}(x_0)} r_0^{2-m}\int_{B_{r_0}(z)} |\nabla u|^2\,dx + C(m,p,\lambda_1,\lambda_2,\|\psi\|_{C^1(\partial\Omega)})(r_0)^{2-2m/p}\\
&\qquad\le \frac{2^{m-2}\epsilon_0^2}{2^m} + C_1(m,p,\lambda_1,\lambda_2,\|\psi\|_{C^1(\partial\Omega)})(r_0)^{2-2m/p}
\end{aligned} \tag{3-5}$$

for some $m < p < \infty$, where $C_1(m, p, \lambda_1, \lambda_2, \|\psi\|_{C^1(\partial\Omega)})$ is a positive constant.

Taking

$$\epsilon \le \left(\frac{\epsilon_0^2}{4C_1(m, p, \lambda_1, \lambda_2, \|\psi\|_{C^1(\partial\Omega)})} \right)^{2m/p-2},$$

we get

$$\sup_{z \in B_{r_0}(x_0),\, 0<r\le r_0} r^{2-m} \int_{B_r(z)} |\nabla u|^2 \, dx \le \frac{\epsilon_0^2}{2}. \tag{3-6}$$

Then Theorem 1.3 tells us that $(u, v) \in C^\infty(B_{r_0/2}(x_0))$, which implies $B_{r_0/4}(x_0) \subset \Omega \setminus S(u)$. □

4. Proof of Theorem 1.6

In this section, we will study the blow-up behavior of a sequence of stationary Lorentzian harmonic maps $\{(u_n, v_n)\}$ with Dirichlet boundary data (ϕ, ψ) and with bounded energy

$$E(u_n, v_n) = \frac{1}{2} \int_\Omega (|\nabla u_n|^2 + |\nabla v_n|^2) \, dx \le \Lambda.$$

Due to the weak compactness, we may assume $u_n \rightharpoonup u$ weakly in $W^{1,2}(\Omega, N)$ and

$$\mu_n := |\nabla u_n|^2 dx \to \mu := |\nabla u|^2 dx + \nu$$

in the sense of Radon measures, where ν is a nonnegative Radon measure by Fatou's lemma and is usually called the defect measure.

Without loss of generality, we assume $B_1(0) \subseteq \Omega$. Similar to harmonic maps [Lin 1999], we define the energy concentration set Σ as

$$\Sigma = \left\{ x \in B_1(0) \;\middle|\; \liminf_{r \searrow 0} \liminf_{n\to\infty} r^{2-n} \int_{B_r(x)} |\nabla u_n|^2 \, dx \ge \frac{\epsilon_0^2}{2^m} \right\}, \tag{4-1}$$

where ϵ_0 is the constant in Theorem 1.3.

Denoting by $\operatorname{spt}(\nu)$ the support set of ν and defining

$$\operatorname{sing}(u) := \{x \in B_1(0) \mid u \text{ is not smooth at } x\},$$

we have:

Lemma 4.1. *Suppose $\{(u_n, v_n)\}$ is a sequence of stationary Lorentzian harmonic maps with Dirichlet boundary data $(u_n, v_n)|_\Omega = (\phi, \psi)$ and bounded energy $E(u_n, v_n) \le \Lambda$; then the energy concentration set Σ is closed in B_1 and*

$$H^{m-2}(\Sigma) \le C(m, \epsilon_0, \Lambda).$$

Moreover, there holds

$$\Sigma = \operatorname{spt}(\nu) \cup \operatorname{sing}(u). \tag{4-2}$$

Proof. For $x_0 \in B_1 \setminus \Sigma$, by the definition of Σ, we know that for any positive constant

$$\epsilon \le \left(\frac{\epsilon_0^2}{4C_1(m, p, \lambda_1, \lambda_2, \|\psi\|_{C^1(\partial\Omega)})}\right)^{2m/p-2},$$

where $C_1(m, p, \lambda_1, \lambda_2, \|\psi\|_{C^1(\partial\Omega)})$ is the constant in (3-5), there exists a positive constant $r_0 < \epsilon$ and a subsequence of $\{n\}$ (also denoted by $\{n\}$), such that, for any n,

$$(2r_0)^{2-m} \int_{B_{2r_0}(x)} |\nabla u_n|^2 \, dx < \frac{\epsilon_0^2}{2^m},$$

which implies (similar to deriving (3-6))

$$\sup_{z \in B_{r_0}(x),\, 0<r\le r_0} r^{2-m} \int_{B_r(z)} |\nabla u_n|^2 \, dx < \frac{\epsilon_0^2}{2}.$$

By Theorem 1.4, we know

$$\|\nabla u_n\|_{L^\infty(B_{r_0/2}(x_0))} + \|\nabla v_n\|_{L^\infty(B_{r_0/2}(x_0))} \le C(m, \lambda_1, \lambda_2, \Lambda, N) r_0^{-m/2}. \tag{4-3}$$

Then, it is easy to see that there exists a small positive constant $r_1 = r_1(m, r_0, \lambda_1, \lambda_2, \Lambda, \epsilon_0, N)$, such that, whenever $r \le r_1$,

$$\sup_{x \in B_{r_0/4}(x_0)} r^{2-m} \int_{B_r(x)} |\nabla u_n|^2 \, dx < \frac{\epsilon_0^2}{2^{m+1}}.$$

Thus, $B_{r_0/4}(x_0) \subset B_1 \setminus \Sigma$. So, Σ is a closed set.

It is standard to get $H^{m-2}(\Sigma) \le C$ by a covering lemma; see [Lin 1999].

For (4-2), on the one hand, let $x_0 \in B_1 \setminus \Sigma$. Then (4-3) holds and by standard elliptic estimates of the Laplace operator, we have

$$\|u_n\|_{C^{1+\alpha}(B_{r_0/4}(x_0))} + \|v_n\|_{C^{1+\alpha}(B_{r_0/4}(x_0))} \le C \tag{4-4}$$

for some $0 < \alpha < 1$. Thus, up to a subsequence of $\{u_n, v_n\}$, $u_n \to u$ strongly in $W^{1,2}$ and $u \in C^\infty(B_{r_0/8}(x_0))$, which implies that $x_0 \notin \operatorname{sing}(u)$ and $x_0 \notin \operatorname{spt} \nu$ since $\nu \equiv 0$ on $B_{r_0/8}(x_0)$.

On the other hand, if $x_0 \in \Sigma$, by the definition, for any $r > 0$ sufficiently small, we have

$$\liminf_{n\to\infty} \frac{\mu_n(B_r(x_0))}{r^{m-2}} \ge \frac{\epsilon_0^2}{2^{m+1}},$$

which implies

$$\frac{\mu(B_r(x_0))}{r^{m-2}} \ge \frac{\epsilon_0^2}{2^{m+1}}$$

for a.e. $r > 0$. Suppose $x_0 \notin \operatorname{sing}(\phi)$; then

$$r^{2-m} \int_{B_r(x_0)} |\nabla u|^2 \, dx \le \frac{\epsilon_0^2}{2^{m+2}}$$

whenever $r > 0$ is small enough. Then we have

$$\frac{\nu(B_r(x_0))}{r^{m-2}} \geq \frac{\epsilon_0^2}{2^{m+2}}$$

for all small positive $r > 0$ and $x_0 \in \operatorname{spt} \nu$. □

Lemma 4.2. *Under the same assumption as above lemma, the limit*

$$\theta_\nu(x) := \lim_{r \to 0} \frac{\nu(B_r(x))}{r^{m-2}} \tag{4-5}$$

exists for H^{m-2}-a.e. $x \in \Sigma$. Moreover,

$$\frac{\epsilon_0^2}{2^m} \leq \theta_\nu(x) \leq C(m, \lambda_1, \lambda_2, \Lambda, N, \|\psi\|_{C^1(\partial\Omega)})\delta_0^{2-m},$$

where $\delta_0 := \operatorname{dist}(B_1(0), \partial\Omega)$.

Proof. Let $x \in \Omega$ and $s_i \to 0$, $t_i \to 0$ be two arbitrary positive sequences. By Corollary 2.4, we have

$$\frac{\mu_n(B_{s_i}(x))}{s_i^{m-2}} \leq \frac{\mu_n(B_{t_j}(x))}{t_j^{m-2}} + C(m, p, \lambda_1, \lambda_2, \Lambda, N, \|\psi\|_{C^1(\partial\Omega)})(t_j)^{2-2m/p} \tag{4-6}$$

for $s_i \leq t_j$ and some $m < p < \infty$. Letting firstly $i \to \infty$ and secondly $j \to \infty$, we get

$$\limsup_{r \to 0} \frac{\mu(B_r(x))}{r^{m-2}} \leq \liminf_{r \to 0} \frac{\mu(B_r(x))}{r^{m-2}}.$$

Thus,

$$\lim_{r \to 0} \frac{\mu(B_r(x))}{r^{m-2}}$$

exists. Noting that for H^{m-2}-a.e. $x \in \Omega$,

$$\lim_{r \to 0} r^{2-m} \int_{B_r(x)} |\nabla u|^2 \, dx = 0, \tag{4-7}$$

we have

$$\lim_{r \to 0} \frac{\nu(B_r(x))}{r^{m-2}} = \lim_{r \to 0} \frac{\mu(B_r(x))}{r^{m-2}}.$$

It is easy to see from (4-6) (taking $p = 2m$) that

$$\begin{aligned} r^{2-m}\mu(B_r(x)) &\leq C(\Lambda)\delta_0^{2-m} + C(m, \lambda_1, \lambda_2, \Lambda, N, \|\psi\|_{C^1(\partial\Omega)})\delta_0 \\ &\leq C(m, \lambda_1, \lambda_2, \Lambda, N, \|\psi\|_{C^1(\partial\Omega)})\delta_0^{2-m}, \end{aligned}$$

which implies $\mu \llcorner \Sigma$ is absolutely continuous with respect to $H^{m-2} \llcorner \Sigma$. By the Radon–Nikodym theorem, we know that there exists a measurable function $\theta(x)$

such that

$$\mu \lfloor \Sigma = \theta(x) H^{m-2} \lfloor \Sigma.$$

Noting that for H^{m-2}-a.e. $x \in \Sigma$,

$$2^{2-m} \leq \liminf_{r \to 0} \frac{H^{m-2}(\Sigma \cap B_r(x))}{r^{m-2}} \leq \limsup_{r \to 0} \frac{H^{m-2}(\Sigma \cap B_r(x))}{r^{m-2}} \leq 1$$

and by (4-7), we have

$$\nu \lfloor \Sigma = \theta(x) H^{m-2} \lfloor \Sigma$$

and

$$\frac{\epsilon_0^2}{2^m} \leq \theta_\nu(x) = \theta(x) \leq C(m, \lambda_1, \lambda_2, \Lambda, N, \|\psi\|_{C^1(\partial\Omega)}) \delta_0^{2-m}. \qquad \square$$

Since ν is absolutely continuous with respect to $H^{m-2} \lfloor \Sigma$ and $\nu = 0$ outside Σ, $\theta_\nu(x)$ is positive for ν-a.e. $x \in \Omega$. Hence by [Preiss 1987], we have:

Corollary 4.3. *The set of energy concentration points Σ is $(m-2)$-rectifiable.*

For any $y \in \Sigma$ and $\lambda > 0$, we define a scaled Radon measure $\mu_{y,\lambda}$ by

$$\mu_{y,\lambda}(A) = \lambda^{2-m} \mu(y + \lambda A).$$

A Radon measure μ_* is called the tangent measure of μ at y if

$$\mu_{y,\lambda} \rightarrow \mu_*$$

in the sense of Radon measures as $r \searrow 0$; see [Federer 1969; Simon 1983].

Lemma 4.4. *Suppose $H^{m-2}(\Sigma) > 0$. Then there exists a nonconstant harmonic sphere S^2 into N.*

Proof. Since Σ is $(m-2)$-rectifiable and $H^{m-2}(\Sigma) > 0$, we know there exists a point $x_0 \in \Sigma$ such that ν has a tangent measure ν_* at x_0 and

$$\nu_* = \theta_\nu(x_0) H^{m-2} \lfloor \Sigma_*,$$

where $\Sigma_* \subset \mathbb{R}^m$ is an $(m-2)$-dimensional linear subspace which is usually called the tangent space of Σ at x_0. Without loss of generality, we may assume $x_0 = 0$ and $\Sigma_* = \mathbb{R}^{m-2} \times \{(0, 0)\}$.

By a similar diagonal argument as that in [Lin 1999], there exists a sequence $r_n \to 0$ such that

$$\tilde{\mu}_n^1 := |\nabla \tilde{u}_n^1|^2 dx \to \nu_*$$

in the sense of Radon measures, where $\tilde{u}_n^1(x) := u_n(x_0 + r_n x)$.

Set $\tilde{v}_n^1(x) := v_n(x_0 + r_n x)$. It is easy to see that $(\tilde{u}_n^1, \tilde{v}_n^1)$ is also a stationary Lorentzian harmonic map. By Lemma 2.3, we have

$$\begin{aligned} r_2^{2-m} \int_{B_{r_2}(0)} (|\nabla \tilde{u}_n^1|^2 - \beta(\tilde{u}_n^1)|\nabla \tilde{v}_n^1|^2)\, dx \\ - r_1^{2-m} \int_{B_{r_1}(0)} (|\nabla \tilde{u}_n^1|^2 - \beta(\tilde{u}_n^1)|\nabla \tilde{v}_n^1|^2)\, dx \\ = 2 \int_{r_1}^{r_2} r^{2-m} \int_{\partial B_r(0)} \left(\left| \frac{\partial \tilde{u}_n^1}{\partial |x|} \right|^2 - \beta(\tilde{u}_n^1) \left| \frac{\partial \tilde{v}_n^1}{\partial |x|} \right|^2 \right) dH^{n-1}\, dr. \end{aligned} \tag{4-8}$$

By Young's inequality, there holds

$$\begin{aligned} \int_{B_r} |x|^{2-m} |\nabla \tilde{v}_n^1|^2\, dx &\le (r_n)^{2-2m/p} \|\nabla v_n\|^2_{L^p(B_{r_n r})} \| |x|^{2-m} \|_{L^{p/(p-2)}(B_r)} \\ &\le C(m, p, \lambda_1, \lambda_2, \Lambda, N, \|\psi\|_{C^1(\partial\Omega)}) (r_n r)^{2-2m/p}. \end{aligned} \tag{4-9}$$

Letting $n \to \infty$ in (4-8) and noting that

$$r_2^{2-m} \nu_*(B_{r_2}(0)) = r_1^{2-m} \nu_*(B_{r_1}(0)),$$

we get

$$\lim_{n\to\infty} \int_{B_2(0)} \left| \frac{\partial \tilde{u}_n^1}{\partial |x|} \right|^2 dx = 0. \tag{4-10}$$

Similarly, since $\nu_{*y,r} = \nu_*$ for any $y \in \Sigma_*$ and $r > 0$, we also have

$$\lim_{n\to\infty} \int_{B_2(0)} \left| \frac{\partial \tilde{u}_n^1}{\partial |x - y|} \right|^2 dx = 0 \quad \text{for } y \in \Sigma_* \cap B_2. \tag{4-11}$$

This implies

$$\lim_{n\to\infty} \sum_{k=1}^{m-2} \int_{B_2(0)} \left| \frac{\partial \tilde{u}_n^1}{\partial x^k} \right|^2 dx = 0. \tag{4-12}$$

Let $x' = (x_1, \dots, x_{m-2}),\ x'' = (x_{m-1}, x_m)$, and define $f_n : B_1^{m-2} \to \mathbb{R}$ by

$$f_n(x') := \sum_{k=1}^{m-2} \int_{B_1^2(0)} \left| \frac{\partial \tilde{u}_n^1}{\partial x_k} \right|^2 (x', x'')\, dx''.$$

Then, (4-12) tells us

$$\lim_{n\to\infty} \|f_n(x')\|_{L^1(B_1^{m-2}(0))} = 0.$$

Denote by $M(f_n)(x')$ the Hardy–Littlewood maximal function; i.e.,

$$M(f_n)(x) = \sup_{0<r<1/2} r^{2-m} \int_{B_r^{m-2}(x)} f_n(x')\, dx', \quad x \in B_{1/2}^{m-2}(0).$$

By a weak L^1-estimate, for any $\rho > 0$, we have

$$|\{x \in B^{m-2}_{1/2}(0) \mid M(f_n) > \rho\}| \leq \frac{C(m)}{\rho} \|f_n\|_{L^1(B^{m-2}_{1/2}(0))},$$

which implies

$$\left|\left\{x \in B^{m-2}_{1/2}(0) \;\middle|\; \limsup_{n\to\infty} M(f_n) > 0\right\}\right| = 0.$$

Combining this with Theorem 1.4, we know there exists a sequence of points $\{x'_n \in B^{m-2}_{1/2}(0)\}$ such that $(\tilde{u}^1_n, \tilde{v}^1_n)$ is smooth near (x'_n, x'') for all $x'' \in B^2_1(0)$ and

$$\lim_{n\to\infty} M(f_n)(x'_n) = 0. \tag{4-13}$$

By the blow-up argument in [Lin 1999], we can find sequences $\{\sigma_n\}$ and $\{x''_n\} \subset B^2_{1/2}(0)$ such that $\sigma_n \to 0$, $x''_n \to (0, 0)$ and

$$\max_{x'' \in B^2_{1/2}(0)} \sigma_n^{2-m} \int_{B^{m-2}_{\sigma_n}(x'_n)\times B^2_{\sigma_n}(x'')} |\nabla \tilde{u}^1_n|^2\, dx = \frac{\epsilon_0^2}{C_2(m)}, \tag{4-14}$$

where the maximum is achieved at the point x''_n and $C_2(m) > 2^m$ is a positive constant to be determined later.

In fact, define

$$g_n(\sigma) := \max_{x'' \in B^2_{1/2}(0)} \sigma^{2-m} \int_{B^{m-2}_{\sigma}(x'_n)\times B^2_{\sigma}(x'')} |\nabla \tilde{u}^1_n|^2\, dx.$$

On the one hand, noting that (u_n, v_n) is smooth near $x'_n \times B^2_1(0)$, we have

$$\lim_{\sigma\to 0} g_n(\sigma) = 0.$$

On the other hand, for any $\sigma > 0$, when n is big enough, it must hold that $g_n(\sigma) \geq \epsilon_0^2/2^m$, for otherwise, by Theorem 1.3, $\tilde{u}^1_n$ will converge strongly in $W^{1,2}$ to a constant map, which contradicts $\tilde{\mu}_n \to \nu_*$. Thus, there exists σ_n such that $g_n(\sigma_n) = \epsilon_0^2/C_2(m)$ and we may assume the maximum is achieved at x''_n. Next, we show $\sigma_n \to 0$ and $x''_n \to (0, 0)$.

If $\sigma_n \geq \delta > 0$, by Corollary 2.4, we have

$$\begin{aligned}\frac{\epsilon_0^2}{C_2(m)} &= \limsup_{n\to\infty} g_n(\sigma_n)\\ &\geq \limsup_{n\to\infty}\bigl(g_n(\delta) - C(m, p, \lambda_1, \lambda_2, \Omega, \|\psi\|_{C^1(\partial\Omega)})(r_n\delta)^{2-2m/p}\bigr) \geq \frac{\epsilon_0^2}{2^m},\end{aligned}$$

which is a contradiction.

If $x''_n \to x''_0 \in B^2_{1/2}(0)$ and $x''_0 \neq (0, 0)$, for any $\sigma < \frac{1}{2}|x''_0|$

$$\frac{\epsilon_0^2}{2^m} \leq \limsup_{n\to\infty} g_n(\sigma) \leq \sigma^{2-m} \nu_*(B^{m-2}_1(0) \times B^2_{2\sigma}(x''_0)) = 0.$$

This is also a contradiction.

Let $x_n = (x'_n, x''_n)$ and

$$(\tilde{u}_n^2(x), \tilde{v}_n^2(x)) := (\tilde{u}_n^1(x_n + \sigma_n x), \tilde{v}_n^1(x_n + \sigma_n x)).$$

Then $(\tilde{u}_n^2(x), \tilde{v}_n^2(x))$ is a stationary Lorentzian harmonic map defined on $B^{m-2}_{R_n}(0) \times B^2_{R_n}(0)$, where $R_n = 1/(4\sigma_n)$ which tends to infinity as $n \to \infty$.

By (4-13), we have

$$\begin{aligned} \text{(4-15)} \quad & \lim_{n\to\infty} \sup_{0<R<R_n} R^{2-m} \int_{B_R^{m-2}(0)\times B^2_{R_n}(0)} \sum_{k=1}^{m-2} \left| \frac{\partial \tilde{u}_n^2}{\partial x_k} \right|^2 dx \\ &= \lim_{n\to\infty} \sup_{0<R<R_n} (\sigma_n R)^{2-m} \int_{B^{m-2}_{\sigma_n R}(x'_n)\times B^2_{\sigma_n R_n}(x''_n)} \sum_{k=1}^{m-2} \left| \frac{\partial \tilde{u}_n^1}{\partial x_k} \right|^2 dx \\ &\le \lim_{n\to\infty} M(f_n)(x'_n) = 0. \end{aligned}$$

By (4-14), we get

$$\text{(4-16)} \quad \frac{\epsilon_0^2}{C_2(m)} = \int_{B_1^{m-2}(0)\times B_1^2(0)} |\nabla \tilde{u}_n^2|^2 \, dx = \max_{x''\in B^2_{R_n-1}(0)} \int_{B_1^{m-2}(0)\times B_1^2(x'')} |\nabla \tilde{u}_n^2|^2 \, dx.$$

By Corollary 2.4, for any $R > 0$, we obtain

$$\begin{aligned} \text{(4-17)} \quad \int_{B_R^{m-2}(0)\times B_R^2(0)} |\nabla \tilde{u}_n^2|^2 \, dx &= (\sigma_n)^{2-m} \int_{B^{m-2}_{\sigma_n R}(x'_n)\times B^2_{\sigma_n R}(x''_n)} |\nabla \tilde{u}_n^1|^2 \, dx \\ &\le C(m, \lambda_1, \lambda_2, \delta_0, \Lambda, \Omega, \|\psi\|_{C^1(\partial\Omega)}) R^{m-2}, \end{aligned}$$

when n is big enough.

Let $\zeta \in C_0^\infty(B_1^{m-2}(0))$ and $\eta \in C_0^\infty(B_1^2(0))$ be two cut-off functions such that $0 \le \zeta \le 1$, $\zeta|_{B^{m-2}_{1/2}(0)} \equiv 1$, $0 \le \zeta \le 1$, and $\eta|_{B^2_{1/2}(0)} \equiv 1$. Similar to [Lin 1999], for any $R > 0$, we define $F_n(a) : B_6^{m-2}(0) \times B_R^2(0) \to \mathbb{R}$ as

$$F_n(a) = \int_{B_1^{m-2}(0)\times B_1^2(0)} |\nabla \tilde{u}_n^2|^2 (a+x) \zeta(x') \eta(x'') \, dx.$$

Computing directly, one has

$$\begin{aligned} \frac{\partial F_n(a)}{\partial a_k} &= \int_{B_1^{m-2}(0)\times B_1^2(0)} \frac{\partial}{\partial x_k} |\nabla \tilde{u}_n^2|^2 (a+x) \zeta(x') \eta(x'') \, dx \\ &= 2 \int_{B_1^{m-2}(0)\times B_1^2(0)} \left\langle \frac{\partial \tilde{u}_n^2}{\partial x_l}, \frac{\partial^2 \tilde{u}_n^2}{\partial x_l \partial x_k} \right\rangle (a+x) \zeta(x') \eta(x'') \, dx \\ &= -2 \int_{B_1^{m-2}(0)\times B_1^2(0)} \left\langle \Delta \tilde{u}_n^2, \frac{\partial \tilde{u}_n^2}{\partial x_k} \right\rangle (a+x) \zeta(x') \eta(x'') \, dx \\ &\qquad - 2 \int_{B_1^{m-2}(0)\times B_1^2(0)} \left\langle \frac{\partial \tilde{u}_n^2}{\partial x_l}, \frac{\partial \tilde{u}_n^2}{\partial x_k} \right\rangle (a+x) \frac{\partial}{\partial x_l} (\zeta(x') \eta(x'')) \, dx. \end{aligned}$$

On the one hand, by (1-3), we have

$$\begin{aligned}
-2\int_{B_1^{m-2}(0)\times B_1^2(0)}&\Big\langle \Delta\tilde{u}_n^2,\, \frac{\partial \tilde{u}_n^2}{\partial x_k}\Big\rangle(a+x)\zeta(x')\eta(x'')\,dx\\
&= -2\int_{B_1^{m-2}(0)\times B_1^2(0)}\Big\langle B^\top(\tilde{u}_n^2)|\nabla\tilde{v}_n^2|^2,\, \frac{\partial \tilde{u}_n^2}{\partial x_k}\Big\rangle(a+x)\zeta(x')\eta(x'')\,dx\\
&\le C\Big(\int_{B_{R+1}^{m-2}(0)\times B_{R+1}^2(0)}|\nabla\tilde{v}_n^2|^4\,dx\Big)^{1/2}\Big(\int_{B_{R+1}^{m-2}(0)\times B_{R+1}^2(0)}\Big|\frac{\partial \tilde{u}_n^2}{\partial x_k}\Big|^2\,dx\Big)^{1/2}.
\end{aligned}$$

On the other hand, by Holder's inequality, we have

$$\begin{aligned}
-2\int_{B_1^{m-2}(0)\times B_1^2(0)}&\Big\langle \frac{\partial \tilde{u}_n^2}{\partial x_l},\, \frac{\partial \tilde{u}_n^2}{\partial x_k}\Big\rangle(a+x)\frac{\partial}{\partial x_l}(\zeta(x')\eta(x''))\,dx\\
&\le C\Big(\int_{B_{R+1}^{m-2}(0)\times B_{R+1}^2(0)}|\nabla\tilde{u}_n^2|^2\,dx\Big)^{1/2}\Big(\int_{B_{R+1}^{m-2}(0)\times B_{R+1}^2(0)}\Big|\frac{\partial \tilde{u}_n^2}{\partial x_k}\Big|^2\,dx\Big)^{1/2}.
\end{aligned}$$

Combining these together and letting $n\to\infty$, we obtain

$$\frac{\partial F_n(a)}{\partial a_k}\to 0,\quad k=1,\dots,m-2,$$

uniformly in $B_2^{m-2}(0)\times B_R^2(0)$ for any fixed $R>0$.

Thus, for any $a=(a',a'')=B_6^{m-2}(0)\times B_R^2(0)$,

$$\begin{aligned}
\int_{B_{1/2}^{m-2}(a')\times B_{1/2}^2(a'')}|\nabla\tilde{u}_n^2|^2\,dx \le F_n(a) &\le F_n((0,a''))+C(m)\sum_{k=1}^{m-2}\Big|\frac{\partial F_n(a)}{\partial a_k}\Big|\\
&\le \int_{B_1^{m-2}(0)\times B_1^2(a'')}|\nabla\tilde{u}_n^2|^2\,dx+C(m)\sum_{k=1}^{m-2}\Big|\frac{\partial F_n(a)}{\partial a_k}\Big|\\
&\le \frac{\epsilon_0^2}{C_2(m)}+C(m)\sum_{k=1}^{m-2}\Big|\frac{\partial F_n(a)}{\partial a_k}\Big|.
\end{aligned}$$

Therefore, when n is big enough, we have

$$6^{2-m}\int_{B_6^{m-2}(0)\times B_6^2(0)}|\nabla\tilde{u}_n^2|^2(x',x''+b)\,dx\le\frac{C(m)\epsilon_0^2}{C_2(m)}\quad\text{for all } b\in B_R^2(0). \tag{4-18}$$

Taking $C_2(m)\ge 2^m C(m)$, by Corollary 2.4, we have

$$\begin{aligned}
\sup_{x_0\in B_3(0),\,0<r\le 3} & r^{2-m}\int_{B_r(x_0)}|\nabla\tilde{u}_n^2|^2(x',x''+b)\,dx\\
&\le \sup_{x_0\in B_3(0)}3^{2-m}\int_{B_3(x_0)}|\nabla\tilde{u}_n^2|^2(x',x''+b)\,dx\\
&\qquad\qquad + C(m,p,\lambda_1,\lambda_2,\Omega,\|\psi\|_{C^1(\partial M)})(\sigma_n r_n)^{2-2m/p}
\end{aligned}$$

$$
\begin{aligned}
&\le 2^{m-2}6^{2-m}\int_{B_6^{m-2}(0)\times B_6^2(0)} |\nabla \tilde{u}_n^2|^2(x', x''+b)\,dx \\
&\qquad\qquad\qquad + C(m, p, \lambda_1, \lambda_2, \Omega, \|\psi\|_{C^1(\partial M)})(\sigma_n r_n)^{2-2m/p} \\
&\le \frac{2^{m-2}C(m)\epsilon_0^2}{C_2(m)} + C(m, p, \lambda_1, \lambda_2, \Omega, \|\psi\|_{C^1(\partial M)})(\sigma_n r_n)^{2-2m/p} \le \frac{\epsilon_0^2}{2}
\end{aligned}
$$

for some $m < p < \infty$, whenever n is large enough.

By Theorem 1.3, we know $(\tilde{u}_n^2, \tilde{v}_n^2)$ subconverges to a Lorentzian harmonic map $(\tilde{u}, \tilde{v})$ in $C^1_{\mathrm{loc}}(B_{3/2}^{m-2}(0) \times \mathbb{R}^2)$. Moreover, by (4-15)-(4-17), for any $R > 0$, we have

$$\int_{B_R(0)} \sum_{k=1}^{m-2} \left| \frac{\partial \tilde{u}}{\partial x_k} \right|^2 dx = 0,$$

and

$$
\begin{aligned}
&\int_{B_1(0)} |\nabla \tilde{u}|^2\,dx = \frac{\epsilon_0^2}{C_2(m)}, \\
&\int_{B_R(0)} |\nabla \tilde{u}|^2\,dx \le C(m, \lambda_1, \lambda_2, \delta_0, \Lambda, \Omega, \|\psi\|_{C^1(\partial\Omega)})R^{m-2}.
\end{aligned}
$$

Furthermore, since

$$
\begin{aligned}
\int_{B_R(0)} |\nabla \tilde{v}|^2\,dx &= \lim_{n\to\infty} \int_{B_R(0)} |\nabla \tilde{v}_n^2|^2\,dx \\
&\le \lim_{n\to\infty} (r_n \sigma_n)^{2-2m/p} R^{m(1-2/p)} \|\nabla v_n\|_{L^p}^2 = 0,
\end{aligned}
$$

we know $\tilde{v}$ is a constant and $\tilde{u} : \mathbb{R}^2 \to N$ is a nonconstant harmonic map with finite energy. By the conformal theory of harmonic maps in dimension 2, $\tilde{u}$ can be extended to a nonconstant harmonic sphere. □

Proof of Theorem 1.6. The conclusion of Theorem 1.6 standardly follows from Lemma 4.4 and the Federer dimensions reduction argument, which is similar to [Schoen and Uhlenbeck 1982] for minimizing harmonic maps. We omit the details here. □

References

[Alday and Maldacena 2009] L. F. Alday and J. Maldacena, "Null polygonal Wilson loops and minimal surfaces in anti-de-Sitter space", *J. High Energy Phys.* **2009**:11 (2009), art. id. 082. MR

[Benci et al. 1991] V. Benci, D. Fortunato, and F. Giannoni, "On the existence of multiple geodesics in static space-times", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **8**:1 (1991), 79–102. MR Zbl

[Bethuel 1993] F. Bethuel, "On the singular set of stationary harmonic maps", *Manuscripta Math.* **78**:4 (1993), 417–443. MR Zbl

[Bryant 1984] R. L. Bryant, "A duality theorem for Willmore surfaces", *J. Differential Geom.* **20**:1 (1984), 23–53. MR Zbl

[Evans 1991] L. C. Evans, "Partial regularity for stationary harmonic maps into spheres", *Arch. Rational Mech. Anal.* **116**:2 (1991), 101–113. MR Zbl

[Federer 1969] H. Federer, *Geometric measure theory*, Grundlehren der Math. Wissenschaften **153**, Springer, 1969. MR Zbl

[Giaquinta 1983] M. Giaquinta, *Multiple integrals in the calculus of variations and nonlinear elliptic systems*, Ann. Math. Studies **105**, Princeton Univ. Press, 1983. MR Zbl

[Greco 1993] C. Greco, "The Dirichlet problem for harmonic maps from the disk into a Lorentzian warped product", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **10**:2 (1993), 239–252. MR Zbl

[Greco 1997] C. Greco, "A multiplicity result for the Dirichlet problem for harmonic maps from the disk into a Lorentzian warped product", *Nonlinear Anal.* **28**:10 (1997), 1661–1668. MR Zbl

[Han et al. 2017a] X. Han, J. Jost, L. Liu, and L. Zhao, "Bubbling analysis for approximate Lorentzian harmonic maps from Riemann surfaces", *Calc. Var. Partial Differential Equations* **56**:6 (2017), art. id. 175. MR Zbl

[Han et al. 2017b] X. Han, L. Zhao, and M. Zhu, "Energy identity for harmonic maps into standard stationary Lorentzian manifolds", *J. Geom. Phys.* **114** (2017), 621–630. MR Zbl

[Han et al. 2018] X. Han, L. Liu, and Z. Liang, "A global weak solution to the Lorentzian harmonic map flow", *Sci. China Math.* (online publication April 2018).

[Han et al. 2019] X. Han, J. Jost, L. Liu, and L. Zhao, "Global existence of the harmonic map heat flow into Lorentzian manifolds", *J. Math. Pures Appl.* (9) (online publication January 2019).

[Isobe 1997] T. Isobe, "Optimal regularity of harmonic maps from a Riemannian manifold into a static Lorentzian manifold", *Pacific J. Math.* **178**:1 (1997), 71–93. MR Zbl

[Isobe 1998] T. Isobe, "Regularity of harmonic maps into a static Lorentzian manifold", *J. Geom. Anal.* **8**:3 (1998), 447–463. MR Zbl

[Li and Tian 1998] J. Li and G. Tian, "A blow-up formula for stationary harmonic maps", *Int. Math. Res. Not.* **1998**:14 (1998), 735–755. MR Zbl

[Lin 1999] F.-H. Lin, "Gradient estimates and blow-up analysis for stationary harmonic maps", *Ann. of Math.* (2) **149**:3 (1999), 785–829. MR Zbl

[Lin and Wang 2008] F. Lin and C. Wang, *The analysis of harmonic maps and their heat flows*, World Sci., Hackensack, NJ, 2008. MR Zbl

[Meyers 1963] N. G. Meyers, "An L^p-estimate for the gradient of solutions of second order elliptic divergence equations", *Ann. Scuola Norm. Sup. Pisa* (3) **17**:3 (1963), 189–206. MR Zbl

[Preiss 1987] D. Preiss, "Geometry of measures in $\mathbb{R}^n$: distribution, rectifiability, and densities", *Ann. of Math.* (2) **125**:3 (1987), 537–643. MR Zbl

[Rivière and Struwe 2008] T. Rivière and M. Struwe, "Partial regularity for harmonic maps and related problems", *Comm. Pure Appl. Math.* **61**:4 (2008), 451–463. MR Zbl

[Schoen and Uhlenbeck 1982] R. Schoen and K. Uhlenbeck, "A regularity theory for harmonic maps", *J. Differential Geom.* **17**:2 (1982), 307–335. Correction in **18**:2 (1983), 329. MR Zbl

[Sharp 2014] B. Sharp, "Higher integrability for solutions to a system of critical elliptic PDE", *Methods Appl. Anal.* **21**:2 (2014), 221–240. MR Zbl

[Simon 1983] L. Simon, *Lectures on geometric measure theory*, Proc. Centre for Math. Analysis **3**, Australian National Univ., Canberra, 1983. MR Zbl

[Tian 2000] G. Tian, "Gauge theory and calibrated geometry, I", *Ann. of Math.* (2) **151**:1 (2000), 193–268. MR Zbl

[Zhu 2013] M. Zhu, "Regularity for harmonic maps into certain pseudo-Riemannian manifolds", *J. Math. Pures Appl.* (9) **99**:1 (2013), 106–123. MR Zbl

Received August 4, 2017. Revised May 15, 2018.

JIAYU LI
SCHOOL OF MATHEMATICAL SCIENCES
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA
HEFEI
CHINA
jiayuli@ustc.edu.cn

LEI LIU
MAX PLANCK INSTITUTE FOR MATHEMATICS IN THE SCIENCES
LEIPZIG
GERMANY
leiliu@mis.mpg.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 1, 2019
dx.doi.org/10.2140/pjm.2019.299.53

SUR LES PAQUETS D'ARTHUR DES GROUPES UNITAIRES ET QUELQUES CONSÉQUENCES POUR LES GROUPES CLASSIQUES

Colette Mœglin et David Renard

Nous donnons une construction explicite des paquets d'Arthur des groupes unitaires réels par induction cohomologique et induction parabolique et en suivant une idée communiquée par P. Trapa, nous établissons la propriété de multiplicité un de ceux-ci. Nous montrons en particulier des résultats d'irréductibilité de certaines induites paraboliques pour les groupes unitaires, ce qui nous permet de compléter les démonstrations d'énoncés analogues annoncés dans nos travaux sur les paquets d'Arthur des groupes classiques.

We give an explicit construction of Arthur packets for real unitary groups by cohomological and parabolic induction and following an idea communicated to us by P. Trapa, we show that they satisfy the multiplicity one property. In particular, we show the irreducibility of some parabolically induced representations for unitary groups, and use this to give the proof of analogous statements made in our work on Arthur packets of classical groups.

1. Introduction

Le premier objet de cet article est de déterminer le plus explicitement possible les paquets d'Arthur des groupes unitaires réels, et d'établir un résultat de multiplicité un pour ceux-ci. Dans [Mœglin et Renard 2017] et les articles afférents [Mœglin et Renard 2018b, 2018a], des résultats analogues ont été établis pour les groupes classiques (i.e., spéciaux orthogonaux et symplectiques) réels. Nous complétons aussi nos résultats sur les groupes classiques en donnant les démonstrations d'énoncés de réduction aux paramètres de bonne parité et d'irréductibilité d'induites paraboliques dans cette réduction annoncés dans [Mœglin et Renard 2017]. Cette démonstration d'irréductibilité d'induites pour les groupes classiques utilise le résultat analogue pour les groupes unitaires démontré dans cet article, ce qui explique qu'elle apparaisse seulement ici.

MSC2010 : primary 20G05, 22E50; secondary 11F55.
Mots-clefs : Arthur packets, unitary groups, classical groups.

Les paquets d'Arthur des groupes classiques et unitaires sont déterminés par leurs propriétés, plus précisément par certaines identités endoscopiques [Arthur 2013; Mok 2015]. Nous renvoyons à [Mœglin et Renard 2017], §2 pour une discussion générale sur les paquets d'Arthur et §4 pour les énoncés de nos résultats pour les groupes classiques. Rappelons simplement ici quelques éléments. Soit $\boldsymbol{G}$ un groupe algébrique connexe réductif défini sur $\mathbb{R}$ et notons G le groupe de ses points réels. Soit

$$\psi_G : W_{\mathbb{R}} \times \mathbf{SL}_2(\mathbb{C}) \to {}^L G$$

un paramètre d'Arthur. Notons S_{ψ_G} le centralisateur de l'image de ψ_G dans $\widehat{G}$, $(S_{\psi_G})_0$ sa composante neutre, et posons $A(\psi_G) = S_{\psi_G}/(S_{\psi_G})_0$. Supposons pour simplifier que ces groupes soient abéliens (c'est une hypothèse qui porte sur $\boldsymbol{G}$ et qui est vérifiée si $\boldsymbol{G}$ est un groupe classique ou unitaire). D'autre part, supposons que $\boldsymbol{G}$ soit quasi-déployé, ou bien forme intérieure pure d'un groupe quasi-déployé. Les conjectures d'Arthur [1984; 1989] dans ce cadre reviennent alors à affirmer l'existence d'une certaine combinaison linéaire à coefficients complexes de représentations irréductibles unitaires de $G \times A(\psi_G)$, que nous notons $\pi^A(\psi_G)$, et qui doit vérifier certaines propriétés, notamment des identités de transfert endoscopique. Pour les groupes classiques quasi-déployés, ceci est établi dans [Arthur 2013], où $\pi^A(\psi_G)$ est caractérisée par les identités de transfert endoscopique attachées aux données endoscopiques elliptiques de $\boldsymbol{G}$ et par une identité de transfert endoscopique vers un groupe général linaire tordu. Il est démontré de plus que les coefficients complexes dans $\pi^A(\psi_G)$ sont en fait des entiers positifs. Ainsi on peut voir $\pi^A(\psi_G)$ comme une représentation unitaire de longueur finie de $G \times A(\psi_G)$. Décomposons cette représentation selon les représentations irréductibles unitaires de G en écrivant

$$\pi^A(\psi_G) = \bigoplus_{\pi \in \Pi(\psi_G)} \pi \boxtimes \rho_\pi .$$

Ici, $\Pi(\psi_G)$ est donc un ensemble fini de représentations irréductibles unitaires de G, le paquet d'Arthur attaché à ψ_G, et pour tout $\pi \in \Pi(\psi_G)$, ρ_π est une représentation unitaire de dimension finie de $A(\psi_G)$ (une somme directe de caractères car $A(\psi_G)$ est abélien). La dimension de ρ_π est la multiplicité de π dans le paquet $\Pi(\psi_G)$ (rappelons que les paquets d'Arthur ne sont pas disjoints). Ces résultats sont aussi démontrés pour les groupes classiques non quasi-déployés, c'est-à-dire les groupes spéciaux orthogonaux réels $\mathbf{SO}(p, q)$, dans [Mœglin et Renard 2018b] et pour les groupes unitaires dans [Mok 2015; Kaletha et al. 2014].

Le travail entrepris dans [Mœglin et Renard 2017] auquel nous renvoyons pour plus de détails, est de donner une construction explicite de $\pi^A(\psi_G)$ pour les groupes classiques. Un problème important qui nous occupe aussi est d'établir que les multiplicités sont 1, c'est-à-dire que ρ_π est un caractère de $A(\psi_G)$. Considérons la composition de ψ_G avec la représentation standard $\mathbf{Std}_{\boldsymbol{G}}$ du L-groupe de $\boldsymbol{G}$. On obtient un paramètre

$$\psi = \mathbf{Std}_{\boldsymbol{G}} \circ \psi_G : W_{\mathbb{R}} \times \mathbf{SL}_2(\mathbb{C}) \to \mathbf{GL}_N(\mathbb{C})$$

que l'on voit comme une représentation complètement réductible de dimension N de $W_{\mathbb{R}} \times \mathbf{SL}_2(\mathbb{C})$. On écrit une décomposition de ψ de la forme

$$\psi = \psi_{\mathrm{mp}} \oplus \psi_{\mathrm{bp}} = \psi_{\mathrm{mp}} \oplus \psi_{\mathrm{bp,disc}} \oplus \psi_{\mathrm{bp,u}}$$

où ψ_{mp} est la partie de mauvaise parité du paramètre, $\psi_{\mathrm{bp,disc}}$ la partie de bonne parité discrète, et $\psi_{\mathrm{bp,u}}$ la partie de bonne parité unipotente. Les constructions de [Mœglin et Renard 2017] se font en quatre étapes. La première étape est le cas $\psi = \psi_{\mathrm{bp,u}}$ des paramètres unipotents et de bonne parité. Dans ce cas, $\pi^A(\psi_G)$ est déterminée dans [Mœglin 2017] par des correspondance de Howe et la propriété de multiplicité un de tels paquets y est établie. Les représentations de G dans ces paquets sont faiblement unipotentes au sens de [Knapp et Vogan 1995, Chapter XII]. La deuxième étape est le cas où $\psi = \psi_{\mathrm{bp,disc}} \oplus \psi_{\mathrm{bp,u}}$, lorsque le paramètre $\psi_{\mathrm{bp,disc}}$ possède certaines propriétés de régularité. Les représentations dans le paquet $\Pi(\psi_G)$ sont alors obtenues par l'induction cohomologique de Vogan–Zuckerman à partir des représentations faiblement unipotentes dans le paquet $\Pi(\psi_{G_{\mathrm{u}}})$, où $\psi_{G_{\mathrm{u}}}$ est un paramètre d'Arthur unipotent pour un groupe $\boldsymbol{G}_{\mathrm{u}}$ de même type que $\boldsymbol{G}$ et de rang plus petit, qui après composition avec la représentation standard $\mathbf{Std}_{\boldsymbol{G}_{\mathrm{u}}}$ donne $\psi_{\mathrm{bp,u}}$, et de caractères de groupes unitaires associés à $\psi_{\mathrm{u,disc}}$. Sous l'hypothèse de régularité mentionnée, les inductions cohomologiques se font dans le « good range », et en particulier sont irréductibles, les paramètres de Langlands des induites se déduisent facilement de ceux des induisantes, et la propriété de multiplicité un des paquets est conservée. Les résultats sont démontrés dans [Mœglin et Renard 2017] et utilisent de manière cruciale les constructions d'Adams et Johnson [1987 ; Johnson 1984] qui sont reliées à celles d'Arthur dans [Arancibia et al. 2018]. La troisième étape établie dans [Mœglin et Renard 2018a] consiste à s'affranchir de l'hypothèse de régularité de $\psi_{\mathrm{bp,disc}}$, et l'on utilise pour cela les propriétés des foncteurs de translation [Knapp et Vogan 1995, Chapter VII]. Les représentations dans $\Pi(\psi_G)$ sont encore obtenues par induction cohomologique comme dans le cas régulier, mais celle-ci a maintenant lieu dans le « weakly fair range », où les résultats généraux sont moins fort. En particulier, l'irréductibilité n'est plus préservée, il peut y avoir des annulations, les paramètres de Langlands des induites deviennent très délicats à calculer car il n'y a pas de formule générale. En conséquence, on perd dans cette étape la conservation de la propriété de multiplicité un (mais nous conjecturons que celle-ci reste vraie, il faudrait l'établir en utilisant d'autres outils). Enfin, la quatrième étape consiste à passer des paquets de bonne parité aux paquets généraux. Les énoncés sont simples, cela se fait par une induction parabolique à partir des représentations du paquet $\Pi(\psi_{G_{\mathrm{bp}}})$, où $\psi_{G_{\mathrm{bp}}}$ est un paramètre d'Arthur de bonne parité pour un groupe $\boldsymbol{G}_{\mathrm{bp}}$ de même type que $\boldsymbol{G}$ et de rang plus petit, qui après composition avec la représentation standard $\mathbf{Std}_{\boldsymbol{G}_{\mathrm{bp}}}$ donne ψ_{bp} et d'une représentation d'un groupe linéaire attachée à ψ_{mp}. Ces résultats de réduction à la bonne parité ont été annoncés dans [Mœglin et Renard 2017] sans démonstrations, et nous donnons celles-ci ici (proposition 5.2 et théorème 5.4).

L'objet principal de cet article est donc d'établir des résultats analogues, mais pour les groupes unitaires. La stratégie suit les mêmes étapes, mais les résultats sont plus simples. Soit $\boldsymbol{G}$ un groupe unitaire de rang N défini sur $\mathbb{R}$, disons $G=\boldsymbol{U}(p,q)$, $p+q=N$, et $\psi_G : W_{\mathbb{R}}\times \mathbf{SL}_2(\mathbb{C}) \to {}^L G$ un paramètre d'Arthur pour G. Maintenant, on remplace la composition avec la représentation standard par la restriction du paramètre à $\mathbb{C}^\times \hookrightarrow W_{\mathbb{R}}$ (changement de base). On obtient donc un paramètre

$$\psi=\psi_{G|\mathbb{C}^\times} : \mathbb{C}^\times \times \mathbf{SL}_2(\mathbb{C}) = W_{\mathbb{C}}\times \mathbf{SL}_2(\mathbb{C}) \to \mathbf{GL}_N(\mathbb{C}),$$

que l'on voit comme un paramètre pour $\mathbf{GL}_N(\mathbb{C})$. Là encore, on décompose ψ en somme de représentations irréductibles de $\mathbb{C}^\times$ (donc de dimension 1) et l'on sépare les composantes de bonne et de mauvaise parité (voir section 2) :

$$\psi=\psi_{\mathrm{mp}}\oplus\psi_{\mathrm{bp}}.$$

La première simplification par rapport au cas des groupes classiques et qu'il n'y a pas à considérer de partie unipotente. La première étape consiste donc à établir les résultats dans le cas $\psi=\psi_{\mathrm{bp}}$, en faisant d'abord là aussi une hypothèse de régularité du paramètre. Écrivons

$$\psi=\psi_{\mathrm{bp}}=\bigoplus_{i=1}^{\ell}(\chi_{t_i}\boxtimes R[a_i])$$

où $R[a_i]$ est la représentation irréductible algébrique de dimension a_i de $\mathbf{SL}_2(\mathbb{C})$ et χ_{t_i}, $t_i\in\mathbb{Z}$, est le caractère de $\mathbb{C}^\times$ défini par $z\mapsto (z/\bar{z})^{t_i/2}$. La condition de bonne parité est que pour tout $i=1,\dots,\ell,\ t_i+a_i-N$ est pair. On suppose les t_i rangés dans l'ordre décroissant, ce qui est loisible. La condition de régularité est alors que pour tout $i=1,\dots,\ell-1$,

$$t_i-(a_i-1)>t_{i+1}+(a_{i+1}-1). \tag{1-1}$$

Remarquons que $N=\sum_{i=1}^{\ell}a_i$. On note $\mathcal{D}(\psi)$ l'ensemble des familles $\underline{d}=(p_i,q_i)_{i=1,\dots,\ell}$ de couples d'entiers tels que $\sum_{i=1}^{\ell}p_i=p$ et $\sum_{i=1}^{\ell}q_i=q$. La représentation $\pi^A(\psi)$ est construite par induction cohomologique. Notons $\mathfrak{g}$ l'algèbre de Lie complexifiée de $G=\boldsymbol{U}(p,q)$, K un sous-groupe compact maximal de G associé à une involution de Cartan θ, et $\mathfrak{k}$ la complexification de l'algèbre de Lie de K. À un élément $\underline{d}$ de $\mathcal{D}(\psi)$, on associe de manière explicite une sous-algèbre parabolique θ-stable $\mathfrak{q}_{\underline{d}}=\mathfrak{l}_{\underline{d}}\oplus\mathfrak{v}_{\underline{d}}$ de $\mathfrak{g}$. On pose $L_{\underline{d}}=\mathrm{Norm}_G(\mathfrak{q}_{\underline{d}})$. C'est un c-Levi de G, au sens de Shelstad [2015], isomorphe au produit $\prod_{i=1}^{\ell} \boldsymbol{U}(p_i, q_i)$, et la complexifiée de l'algèbre de Lie de ce groupe est $\mathfrak{l}_{\underline{d}}$. On introduit un caractère $\Lambda_{\underline{d}}$ de ce groupe, déterminé par les (t_i, a_i) (voir équation (4-2)), et l'on pose

$$\mathscr{A}_{\underline{d}}(\psi) = (\mathcal{R}^{\mathfrak{g},K}_{\mathfrak{q}_{\underline{d}},L_{\underline{d}}\cap K})^{\dim(\mathfrak{u}_{\underline{d}}\cap\mathfrak{k})}(\Lambda_{\underline{d}}) \tag{1-2}$$

où $(\mathcal{R}^{\mathfrak{g},K}_{\mathfrak{q}_{\underline{d}},L_{\underline{d}}\cap K})^k$ est le foncteur d'induction cohomologique de Vogan–Zuckerman en degré k (cf. [Knapp et Vogan 1995, Chapter V]). La condition de régularité (1-1) assure que cette induction cohomologique est dans le « good range ». De ceci, il découle que $(\mathcal{R}^{\mathfrak{g},K}_{\mathfrak{q}_{\underline{d}},L_{\underline{d}}\cap K})^k(\Lambda_{\underline{d}}) = 0$, si $k \neq \dim(\mathfrak{u}_{\underline{d}} \cap \mathfrak{k})$, et que $\mathscr{A}_{\underline{d}}(\psi)$ est un module unitaire et irréductible. De plus, les $\mathscr{A}_{\underline{d}}(\psi)$ lorsque $\underline{d}$ parcourt $\mathcal{D}(\psi)$ sont distincts.

Définissons maintenant pour tout $\underline{d} \in \mathcal{D}(\psi)$ un caractère $\epsilon_{\underline{d}}$ du groupe $A(\psi_G)$. On définit d'abord $\epsilon_{\underline{d}}$ comme une application de $[1, \ell]$ dans ± 1. Pour cela on pose pour tout entier $i \in [1, \ell]$, $a_{<i} = \sum_{j<i} a_j$ et

$$\epsilon_{\underline{d}}(i) = (-1)^{p_i a_{<i} + q_i(a_{<i}+1) + a_i(a_i-1)/2}. \tag{1-3}$$

Le groupe $A(\psi_G)$ s'identifie de manière naturelle à $(\pm 1)^{\ell}$ et $\epsilon_{\underline{d}}$ à un caractère de $A(\psi_G)$. Il découle alors essentiellement des résultats de [Adams et Johnson 1987; Johnson 1984; Arancibia et al. 2018] (voir aussi [Mœglin et Renard 2017]) que

$$\pi^A(\psi) = \sum_{\underline{d}\in\mathcal{D}(\psi)} \mathscr{A}_{\underline{d}}(\psi) \boxtimes \epsilon_{\underline{d}}. \tag{1-4}$$

D'après la remarque faite ci-dessus sur le fait que les $\mathscr{A}_{\underline{d}}(\psi)$ sont non isomorphes deux à deux, on en déduit la propriété de multiplicité un pour ces paquets.

Ensuite, on abandonne l'hypothèse de régularité (1-1) pour ne conserver que l'hypothèse de décroissance de la suite $(t_i)_{i=1,\ldots,\ell}$, et l'on définit $\mathscr{A}_{\underline{d}}(\psi)$ comme ci-dessus. L'induction cohomologique a alors lieu dans le « weakly fair range » et l'on a toujours $(\mathcal{R}^{\mathfrak{g},K}_{\mathfrak{q}_{\underline{d}},L_{\underline{d}}\cap K})^k(\Lambda_{\underline{d}}) = 0$, si $k \neq \dim(\mathfrak{u}_{\underline{d}} \cap \mathfrak{k})$. De plus, si $\mathscr{A}_{\underline{d}}(\psi)$ n'est pas nul, c'est un module unitaire et irréductible, cette dernière propriété étant propre aux groupes unitaires (cf. [Matumoto 1996; Trapa 2001]) et est due à Barbasch et Vogan. Le groupe $A(\psi_G)$ s'identifie maintenant à un quotient de $\{\pm 1\}^{\ell}$. Nous établissons que le caractère $\epsilon_{\underline{d}}$ défini ci-dessus ce factorise par ce quotient si $\mathscr{A}_{\underline{d}}(\psi)$ est non nul (proposition 4.4). Nous montrons que le terme de droite de (1-4), qui est donc encore bien défini comme représentation de $G \times A(\psi_G)$, est bien la représentation $\pi^A(\psi)$. On a donc (théorème 4.1 du texte) :

Théorème 1.1. *On suppose que ψ est de bonne parité. Alors la représentation associée à ψ est*

$$\pi^A(\psi) = \sum_{\underline{d}\in\mathcal{D}(\psi)} \mathscr{A}_{\underline{d}}(\psi) \boxtimes \epsilon_{\underline{d}}.$$

De plus, les représentations $\mathscr{A}_{\underline{d}}(\psi)$ non nulles sont non isomorphes deux à deux.

Comme dans le cas des groupes classiques on passe du cas régulier au cas général (de bonne parité) en utilisant les foncteurs de translation. Pour les groupes unitaires, on a donc en plus le fait que les $\mathscr{A}_{\underline{d}}(\psi)$ sont irréductibles ou nuls (on ne détermine pas quand ces modules sont nuls, voir la remarque 4.2). De plus, la seconde assertion du théorème montre que la propriété de multiplicité un est conservée.

Passons maintenant à un paramètre ψ_G général, avec $\psi = \psi_{\mathrm{mp}} \oplus \psi_{\mathrm{bp}}$. A la partie de bonne parité ψ_{bp} on attache donc un paquet $\Pi(\psi_{G_{\mathrm{bp}}})$ d'un groupe unitaire de rang N_{bp} plus petit par la construction que l'on vient de donner, à la partie de mauvaise parité on attache une représentation irréductible unitaire ρ d'un groupe $\mathbf{GL}_{N_\rho}(\mathbb{C})$, et l'on a $N = 2N_\rho + N_{\mathrm{bp}}$. Ceci détermine un sous-groupe parabolique standard $P = MN$ de G, du moins si $\inf(p,q) \geq N_\rho$, avec un facteur de Levi M isomorphe à $\mathbf{GL}_{N_\rho}(\mathbb{C}) \times G_{\mathrm{bp}}$ (et donc $G_{\mathrm{bp}} \simeq \boldsymbol{U}(p - N_\rho, q - N_\rho)$). On a alors (théorème 5.3 du texte), en remarquant que les groupes $A(\psi_G)$ et $A(\psi_{G_{\mathrm{bp}}})$ sont naturellement isomorphes :

Théorème 1.2. *La représentation $\pi^A(\psi_G)$ est obtenue à partir de $\pi^A(\psi_{G_{\mathrm{bp}}})$ par induction parabolique, c'est-à-dire*

$$\pi^A(\psi_G) = \bigoplus_{\pi_{\mathrm{bp}} \in \Pi(\psi_{G_{\mathrm{bp}}})} \mathrm{Ind}_P^G(\rho \boxtimes \pi_{\mathrm{bp}}) \boxtimes \rho_{\pi_{\mathrm{bp}}}.$$

De plus, les induites paraboliques dans le membre de droite sont irréductibles.

Dans cette étape, la conservation de la propriété de multiplicité un des paquets est évidente. On a donc

Théorème 1.3. *Les paquets d'Arthur des groupes unitaires réels ont la propriété de multiplicité un.*

2. Décomposition des A-paramètres

Supposons d'abord que $\boldsymbol{G}$ est un groupe classique. On note $\mathbf{Std}_{\mathrm{G}}$ la représentation standard du L-groupe de $\boldsymbol{G}$ dans $\mathbf{GL}_N(\mathbb{C})$ (voir [Mœglin et Renard 2017, §3.1]), par exemple si $\boldsymbol{G} = \mathbf{Sp}_{2n}(\mathbb{R})$, ${}^LG = \mathbf{SO}_{2n+1}(\mathbb{C}) \times W_{\mathbb{R}}$ et $\mathbf{Std}_{\mathrm{G}}$ est donné par l'inclusion de $\mathbf{SO}_{2n+1}(\mathbb{C})$ dans $\mathbf{GL}_{2n+1}(\mathbb{C})$.

Si $\psi_G : W_{\mathbb{R}} \times \mathbf{SL}_2(\mathbb{C}) \to {}^LG$ est un paramètre d'Arthur pour G, on pose $\psi = \mathbf{Std}_G \circ \psi_G$ et l'on voit ψ comme une représentation complètement réductible de $W_{\mathbb{R}} \times \mathbf{SL}_2(\mathbb{C})$. Dans [Mœglin et Renard 2017, §4.1], on a décomposé cette représentation en représentations irréductibles, et séparé ces représentations irréductibles selon leur *parité*, qui peut être *bonne* ou *mauvaise*, ce qui permet d'énoncer certains résultats de réduction.

Nous allons maintenant faire de même pour les groupes unitaires. Soit donc maintenant $\boldsymbol{G}$ un groupe unitaire réel de rang N, et soit

$$\psi_G : W_{\mathbb{R}} \times \mathbf{SL}_2(\mathbb{C}) \to {}^L G = \mathbf{GL}_N(\mathbb{C}) \rtimes W_{\mathbb{R}} \tag{2-1}$$

un paramètre d'Arthur pour G. On note ψ la restriction de ψ_G au sous-groupe $\mathbb{C}^\times \times \mathbf{SL}_2(\mathbb{C})$ de $W_{\mathbb{R}} \times \mathbf{SL}_2(\mathbb{C})$, que l'on voit comme une représentation de dimension N de $\mathbb{C}^\times \times \mathbf{SL}_2(\mathbb{C})$. Cette représentation est complètement réductible. Pour tout $a \in \mathbb{Z}_{>0}$, notons $R[a]$ la représentation algébrique de $\mathbf{SL}_2(\mathbb{C})$ de dimension a, et pour tout $(t, s) \in \mathbb{Z} \times i\mathbb{R}$, notons

$$\chi_{t,s} : z \mapsto (z/\bar{z})^{t/2}(z\bar{z})^{s/2} = z^{(t+s)/2}\bar{z}^{(-t+s)/2}. \tag{2-2}$$

C'est un caractère unitaire de $\mathbb{C}^\times$. On note pour tout $a \in \mathbb{Z}_{>0}$,

$$\chi_{t,s,a} = \chi_{t,s} \circ \det_a \tag{2-3}$$

où $\det_a$ est le déterminant de $\mathbf{GL}_a(\mathbb{C})$.

La forme générale de la décomposition de ψ en irréductibles (après une éventuelle conjugaison dans $\mathbf{GL}_N(\mathbb{C})$) est

$$\psi = \bigoplus_{(t,s,a)\in\mathcal{E}(\psi)} \chi_{t,s} \otimes R[a] \tag{2-4}$$

pour un certain ensemble avec multiplicités finies $\mathcal{E}(\psi)$ de triplets

$$(t, s, a) \in \mathbb{Z} \times i\mathbb{R} \times \mathbb{Z}_{>0}.$$

Définition 2.1. On dit que le triplet (t, s, a) est de bonne parité si $s = 0$, et si $(t + a - 1)/2 + (N - 1)/2 \in \mathbb{Z}$.

Remarque 2.2. Le fait que ψ provient d'un A-paramètre pour $G = \boldsymbol{U}(p, q)$ est équivalent à la propriété suivante : si $(t, s, a) \in \mathcal{E}(\psi)$ n'est pas de bonne parité, alors la multiplicité de ce triplet dans $\mathcal{E}(\psi)$ est égale à la multiplicité du triplet $(t, -s, a)$ dans le cas où $s \neq 0$, et si $s = 0$, la multiplicité de $(t, 0, a)$ est paire.

Cela revient à dire, en notant $\mathcal{E}(\psi)_{\mathrm{bp}}$ les triplets de $\mathcal{E}(\psi)$ ayant bonne parité, qu'il existe une décomposition de $\mathcal{E}(\psi)$ en l'union disjointe de trois sous-ensembles, $\mathcal{E}(\psi)_{\mathrm{bp}}$, $\mathcal{E}'(\psi)$, et $\mathcal{E}''(\psi)$ tel que si $(t, s, a) \in \mathcal{E}'(\psi)$ alors $(t, -s, a) \in \mathcal{E}''(\psi)$ avec la même multiplicité.

On note

$$\psi_{\mathrm{bp}} := \bigoplus_{(t,s,a)\in\mathcal{E}(\psi)_{\mathrm{bp}}} \chi_{t,s} \otimes R[a]. \tag{2-5}$$

Alors ψ_{bp} est un morphisme de $\mathbb{C}^\times \times \mathbf{SL}_2(\mathbb{C})$ dans $\mathbf{GL}_{N_{\mathrm{bp}}}(\mathbb{C})$ où

$$N_{\mathrm{bp}} = \sum_{(t,s,a)\in\mathcal{E}(\psi)_{\mathrm{bp}}} a, \tag{2-6}$$

qui provient d'un A-paramètre comme en (2-1), mais pour les groupes unitaires de rang N_{bp}.

3. Réalisation des groupes unitaires et de leur *c*-Levi. Induction cohomologique

3A. ***Paires paraboliques.*** Soit G le groupe des points réels d'un groupe algébrique connexe réductif défini sur $\mathbb{R}$. On fixe une involution de Cartan θ de G, et l'on note K le sous-groupe des points fixes de θ : c'est un sous-groupe compact maximal de G. On suppose que G et K sont de même rang ; autrement dit, G possède un sous-groupe de Cartan T inclus dans K et donc compact. On note $\mathfrak{t}_0$, $\mathfrak{k}_0$ et $\mathfrak{g}_0$ les algèbres de Lie respectives de T,K et G et $\mathfrak{t}$, $\mathfrak{k}$ et $\mathfrak{g}$ leur complexifiées. Les sous-algèbres paraboliques θ-stables de $\mathfrak{g}$ sont obtenues de la manière suivante. On fixe un élément $\nu \in \sqrt{-1}\mathfrak{t}_0^*$, et l'on pose :

$$\begin{aligned} &\mathfrak{l} = \mathfrak{g}^\nu = \mathfrak{t} \oplus \Bigg(\bigoplus_{\substack{\alpha\in\Delta(\mathfrak{g},\mathfrak{t}) \\ \langle\nu,\alpha\rangle=0}} \mathfrak{g}^\alpha \Bigg), \quad && \mathfrak{v} = \bigoplus_{\substack{\alpha\in\Delta(\mathfrak{g},\mathfrak{t}) \\ \langle\nu,\alpha\rangle>0}} \mathfrak{g}^\alpha, \\ &\mathfrak{q} = \mathfrak{l} \oplus \mathfrak{v}, && L = \mathrm{Norm}_G(\mathfrak{q}). \end{aligned} \tag{3-1}$$

Dans cet article, nous appellerons paire parabolique une paire $(\mathfrak{q}, L)$ obtenue comme ci-dessus, avec $\mathfrak{q}$ sous-algèbre parabolique θ-stable de $\mathfrak{g}$. Le sous-groupe L de G sera appelé c-Levi de G (terminologie de Shelstad [2015]).

On note $(\mathcal{R}_{\mathfrak{q},L\cap K}^{\mathfrak{g},K})^k$ le foncteur d'induction cohomologique de Vogan–Zuckerman (cf. [Vogan 1981, §6.3.1]) en degré k, de la catégorie des $(\mathfrak{l}, K\cap L)$-modules vers la catégorie des $(\mathfrak{g}, K)$-modules. Dans ce contexte, le degré qui nous intéresse particulièrement, et même exclusivement, est $S = \dim(\mathfrak{v}\cap\mathfrak{k})$, et dans l'article, nous écrirons $(\mathcal{R}_{\mathfrak{q},L\cap K}^{\mathfrak{g},K})^S$ sans préciser de nouveau ce qu'est S.

Si Λ est un caractère unitaire de L, on note λ sa différentielle, que l'on voit comme un élément de $i\mathfrak{t}_0^*$. On pose alors

$$A_{\mathfrak{q}}(\Lambda) = (\mathcal{R}_{\mathfrak{q},L\cap K}^{\mathfrak{g},K})^S(\Lambda).$$

Si le groupe L est connexe, λ détermine Λ et l'on note alors cette représentation $A_{\mathfrak{q}}(\lambda)$.

Sous certaines conditions sur le caractère infinitésimal de la représentation σ de L que l'on induit, on a des résultats d'annulation, d'irréductibilité et d'unitarité des modules $(\mathcal{R}_{\mathfrak{q},L\cap K}^{\mathfrak{g},K})^S(\sigma)$. Nous renvoyons à [Knapp et Vogan 1995] (p. 34, 35 et

793 respectivement) pour les définitions du (weakly) good range, du (weakly) fair range et des représentations faiblement unipotentes, pour lesquels on a les résultats suivants : dans le weakly good range $(\mathcal{R}_{\mathfrak{q},L\cap K}^{\mathfrak{g},K})^k(\sigma)$ est nul si $k \neq S = \dim(\mathfrak{v} \cap \mathfrak{k})$. Si σ est irréductible et dans le good range (resp. weakly good range), $\mathcal{R}_{\mathfrak{q},L,G}^{S}(\sigma)$ est irréductible (resp. irréductible ou nul). Si σ est unitaire et dans le weakly good range, $(\mathcal{R}_{\mathfrak{q},L\cap K}^{\mathfrak{g},K})^S(\sigma)$ est unitaire. Si σ est faiblement unipotente et dans le weakly fair range, alors $(\mathcal{R}_{\mathfrak{q},L\cap K}^{\mathfrak{g},K})^k(\sigma)$ est nul si $k \neq S$ et $(\mathcal{R}_{\mathfrak{q},L\cap K}^{\mathfrak{g},K})^S(\sigma)$ est unitaire si σ est de plus unitaire. En revanche, on n'a pas de résultat d'irréductibilité en général dans le weakly fair range, ni même le fair range.

3B. ***Le groupe symplectique.*** Une manière commode de réaliser les groupes unitaires $\boldsymbol{U}(p,q)$, $p+q=N$, est de les réaliser comme c-Levi d'un groupe symplectique $\mathbf{Sp}(2N,\mathbb{R})$, les c-Levi des $\boldsymbol{U}(p,q)$ s'identifient alors eux aussi à des c-Levi de $\mathbf{Sp}(2N,\mathbb{R})$. On suppose $\mathbb{R}^{2N}$ (identifié à $\mathcal{M}_{2N,1}(\mathbb{R})$, les matrices colonnes) muni de sa forme symplectique usuelle, c'est-à-dire, si $X, Y \in \mathbb{R}^{2N}$

$$(X|Y) = {}^tXJY \quad \text{où } J = \begin{pmatrix} 0_N & I_N \\ -I_N & 0_N \end{pmatrix}.$$

Soit $\mathscr{G} = \mathbf{Sp}(2N,\mathbb{R})$ le groupe des isomorphismes de $(\mathbb{R}^{2N}, (\cdot\,|\,\cdot))$, que l'on muni de l'involution de Cartan $\theta : g \mapsto {}^tg^{-1}$. Le sous-groupe de $\mathscr{G}$ des points fixes sous θ est un sous-groupe compact maximal de $\mathscr{G}$, que l'on note $\mathscr{K}$, et qui est isomorphe au groupe unitaire $\boldsymbol{U}(N)$. On note $\mathfrak{g}_0$ et $\mathfrak{k}_0$ les sous-algèbres de Lie respectives de $\mathscr{G}$ et $\mathscr{K}$, réalisées comme sous-algèbre de Lie de $\mathcal{M}_{2N}(\mathbb{R})$. Pour tout $(a_1,\dots,a_N) \in \mathbb{R}^N$, on pose :

$$t(a_1,\dots,a_N) = \left(\begin{array}{cccc|cccc} & & & & a_1 & & & \\ & & & & & a_2 & & \\ & & & & & & \ddots & \\ & & & & & & & a_N \\ \hline -a_1 & & & & & & & \\ & -a_2 & & & & & & \\ & & \ddots & & & & & \\ & & & -a_N & & & & \end{array}\right).$$

Alors

$$\mathfrak{t}_0 := \left\{ t(a_1,\dots,a_N),\ (a_1,\dots,a_N) \in \mathbb{R}^N \right\}$$

est une sous-algèbre de Cartan de $\mathfrak{k}_0$ et aussi de $\mathfrak{g}_0$.

Notons $\mathfrak{g}$, $\mathfrak{k}$, $\mathfrak{t}$ les complexifications des algèbres de Lie $\mathfrak{g}_0$, $\mathfrak{k}_0$, $\mathfrak{t}_0$, respectivement. Soient $\Delta(\mathfrak{g},\mathfrak{t})$, $\Delta(\mathfrak{k},\mathfrak{t})$ les systèmes de racines de $\mathfrak{g}$ et $\mathfrak{k}$ respectivement, relativement à la sous-algèbre de Cartan $\mathfrak{t}$. On a

$$\begin{aligned}\Delta(\mathfrak{g}, \mathfrak{t}) &= \{\pm(e_i \pm e_j),\ 1 \le i < j \le N\} \cup \{\pm 2e_i, 1 \le i \le N\},\\ \Delta(\mathfrak{k}, \mathfrak{t}) &= \{\pm(e_i - e_j),\ 1 \le i < j \le N\},\end{aligned}$$

où $e_i \in \sqrt{-1}\,\mathfrak{t}_0^* \subset \mathfrak{t}^*$ est la forme linéaire $t(a_1, \dots, a_N) \mapsto \sqrt{-1}\,a_i$. On fixe les systèmes de racines positives

$$\begin{aligned}\Delta^+(\mathfrak{g}, \mathfrak{t}) &= \{(e_i \pm e_j),\ 1 \le i < j \le N\} \cup \{2e_i, 1 \le i \le N\},\\ \Delta^+(\mathfrak{k}, \mathfrak{t}) &= \{(e_i - e_j),\ 1 \le i < j \le N\}.\end{aligned}$$

On identifie $\mathfrak{t}^*$ et $\mathbb{C}^N$ grâce à la base $(e_i)_{1 \le i \le N}$ de $\mathfrak{t}^*$, et de même pour $\mathfrak{t}$ grâce à la base duale.

3C. ***Paires paraboliques maximales et induction cohomologique.*** On continue avec les notations du paragraphe précédent, en particulier $\mathscr{G} = \mathbf{Sp}(2N, \mathbb{R})$. Soit $(\mathfrak{q}, L)$ une paire parabolique pour $\mathscr{G}$, et l'on suppose que la sous-algèbre parabolique θ-stable $\mathfrak{q} = \mathfrak{l} \oplus \mathfrak{v}$ est maximale. Une telle sous-algèbre est obtenue en prenant un élément de $\mathfrak{t}$ de la forme

$$t_{p,q} = \big(\underbrace{1, \dots, 1}_{p}, \underbrace{0, \dots, 0}_{N-p-q}, \underbrace{-1, \dots, -1}_{q}\big)$$

avec $p + q \le N$. On pose alors

$$\begin{aligned}&\mathfrak{l}_{p,q} = \mathfrak{g}^{t_{p,q}} = \mathfrak{t} \oplus \Bigg(\bigoplus_{\substack{\alpha \in \Delta(\mathfrak{g},\mathfrak{t})\\ \alpha(t) = 0}} \mathfrak{g}_\alpha\Bigg), && \mathfrak{v}'_{p,q} = \bigoplus_{\substack{\alpha \in \Delta(\mathfrak{g},\mathfrak{t})\\ \alpha(t) > 0}} \mathfrak{g}_\alpha,\\ &\mathfrak{q}'_{p,q} = \mathfrak{l}_{p,q} \oplus \mathfrak{v}'_{p,q}, && L_{p,q} = \mathrm{Norm}_{\mathscr{G}}(\mathfrak{q}_{p,q}).\end{aligned} \tag{3-2}$$

Dans le cas où $p + q = N$, $L_{p,q}$ est une réalisation du groupe unitaire $\boldsymbol{U}(p, q)$. Les racines de $\mathfrak{t}$ dans $\mathfrak{l}_{p,q}$ sont :

$$\begin{aligned}&\pm(e_i - e_j), && 1 \le i < j \le p \quad \text{ou} \quad p + 1 \le i < j \le N,\\ &\pm(e_i + e_j), && 1 \le i \le p < j \le N.\end{aligned}$$

On choisit comme système de racines positives :

$$\Delta^+_{p,q} = \Delta^+(\mathfrak{l}_{p,q}, \mathfrak{t}) = \begin{cases} e_i - e_j, & 1 \le i < j \le p,\\ -(e_i - e_j), & p + 1 \le i < j \le N\\ e_i + e_j, & 1 \le i \le p < j \le N.\end{cases}$$

On pose
(3-3)

$$\begin{aligned}\delta(\mathfrak{l}_{p,q}) &= \tfrac{1}{2}\sum_{\alpha\in\Delta^+_{p,q}}\alpha \\ &= \tfrac{1}{2}\big(\underbrace{N-1, N-3, \dots, q-p+1}_{p}, \underbrace{p-q+1, \dots, p-q+3, \dots, N-1}_{q}\big).\end{aligned}$$

On a $\mathscr{K} = L_{N,0}$ et $L_{p,q}\cap\mathscr{K} \simeq \boldsymbol{U}(p)\times\boldsymbol{U}(q)$.

3D. *c-Levi des $\boldsymbol{U(p,q)}$*. On continue avec les notations de la section précédente. Notons $\mathcal{D}(p,q)$ l'ensemble des familles $\underline{d} = (p_i, q_i)_{i=1,\dots,\ell}$ de couples d'entiers positifs ou nuls tels que

$$\sum_i p_i = p \quad \text{et} \quad \sum_i q_i = q \quad \text{et} \quad \mathcal{D}(N) = \coprod_{\substack{p,q\\ p+q=N}} \mathcal{D}(p,q).$$

Pour tout $\underline{d} = (p_i, q_i)_{i=1,\dots,\ell} \in \mathcal{D}(N)$, notons

$$t_{\underline{d}} = \big(\underbrace{\ell,\dots,\ell}_{p_1},\dots,\underbrace{2,\dots,2}_{p_{\ell-1}},\underbrace{1,\dots,1}_{p_\ell},\underbrace{-1,\dots,-1}_{q_\ell},\underbrace{-2,\dots,-2}_{q_{\ell-1}},\dots,\underbrace{-\ell,\dots,-\ell}_{q_1}\big).$$

On définit la sous-algèbre parabolique θ-stable $\mathfrak{q}'_{\underline{d}} = \mathfrak{l}_{\underline{d}} \oplus \mathfrak{v}'_{\underline{d}}$ et le c-Levi $L_{\underline{d}}$ associés à $t_{\underline{d}}$ comme en (3-1). Si $\underline{d} \in \mathcal{D}(p,q)$, alors $\mathfrak{l}_{\underline{d}} \subset \mathfrak{l}_{p,q}$ et $L_{\underline{d}}$ est un c-Levi de $L_{p,q}$, associé à la sous-algèbre parabolique θ-stable $\mathfrak{q}_{\underline{d}}$ de $\mathfrak{l}_{p,q}$, où

$$\mathfrak{q}_{\underline{d}} = \mathfrak{q}'_{\underline{d}} \cap \mathfrak{l}_{p,q} = \mathfrak{l}_{\underline{d}} \oplus (\mathfrak{l}_{p,q} \cap \mathfrak{v}'_{\underline{d}}) = \mathfrak{l}_{\underline{d}} \oplus \mathfrak{v}_{\underline{d}}. \tag{3-4}$$

On pose $\Delta^+(\mathfrak{l}_{\underline{d}}, \mathfrak{t}) = \Delta(\mathfrak{l}_{\underline{d}}, \mathfrak{t}) \cap \Delta^+(\mathfrak{l}_{p,q}, \mathfrak{t})$ et pour tout $i = 1, \dots, \ell,\ a_i = p_i + q_i$,

(3-5)

$$\begin{aligned}&\delta(\mathfrak{l}_{\underline{d}}) \\ &= \frac{1}{2}\sum_{\alpha\in\Delta^+(\mathfrak{l}_{\underline{d}},\mathfrak{t})}\alpha \\ &= \frac{1}{2}\Big(\underbrace{a_1-1, a_1-3, \dots, q_1-p_1+1}_{p_1}, \underbrace{a_2-1, a_2-3, \dots, q_2-p_2+1}_{p_2}, \dots, \\ &\qquad \underbrace{a_\ell-1, a_\ell-3, \dots, q_\ell-p_\ell+1}_{p_\ell}, \underbrace{p_\ell-q_\ell+1, \dots, p_\ell-q_\ell+3, \dots, a_\ell-1}_{q_\ell}, \dots, \\ &\qquad\qquad \underbrace{p_2-q_2+1, \dots, p_2-q_2+3, \dots, a_2-1}_{q_2}, \\ &\qquad\qquad\qquad \underbrace{p_1-q_1+1, \dots, p_1-q_1+3, \dots, a_1-1}_{q_1}\Big).\end{aligned}$$

4. Paramètres de bonne parité

Soit ψ un A-paramètre pour $\boldsymbol{U}(p,q)$ comme en (2-1). On suppose que ψ est de bonne parité, i.e., $\psi = \psi_{\mathrm{bp}}$. On a donc

$$\psi = \bigoplus_{i=1}^{\ell} (\chi_{t_i} \otimes R[a_i]). \tag{4-1}$$

On a supprimé les paramètres s_i des notations, puisque dans le cas de bonne parité, ils sont tous nuls. On ordonne les indices i pour que $t_1 \geq \cdots \geq t_\ell$ et si $t_i = t_{i+1}$ alors $a_i \geq a_{i+1}$.

On reprend les notations de la section 3D, en particulier la réalisation de $\boldsymbol{U}(p,q)$ comme sous-groupe $L_{p,q}$ d'un groupe symplectique. Pour tout $i \in [1,\ell]$, on fixe une décomposition $a_i = p_i + q_i$ en somme de deux entiers éventuellement nuls. On note $\mathcal{D}(\psi)$ l'ensemble de ces décompositions qui vérifient de plus $p = \sum_i p_i$. On a donc $\mathcal{D}(\psi) \subset \mathcal{D}(p,q)$, ce dernier ensemble étant défini dans la section 3D, et en particulier un élément $\underline{d} \in \mathcal{D}(\psi)$ détermine un c-Levi $L_{\underline{d}}$ de $L_{p,q}$ isomorphe à $\prod_i \boldsymbol{U}(p_i,q_i)$ et une sous-algèbre parabolique θ-stable $\mathfrak{q}_{\underline{d}} = \mathfrak{l}_{\underline{d}} \oplus \mathfrak{v}_{\underline{d}}$ de $\mathfrak{l}_{p,q}$ dont la sous-algèbre de Levi est précisément l'algèbre de Lie complexifiée de $L_{\underline{d}}$. Pour tout $i \in [1,\ell]$, on pose $a_{<i} := \sum_{j<i} a_j$.

Via l'isomorphisme de $L_{\underline{d}}$ avec $\prod_i \boldsymbol{U}(p_i,q_i)$, on définit le caractère $\Lambda_{\underline{d}}$ du groupe $L_{\underline{d}}$ par

$$\Lambda_{\underline{d}} = \boxtimes_i \det{}^{(t_i+a_i-N)/2-a_{<i}}. \tag{4-2}$$

On peut aussi définir ce caractère en donnant sa différentielle comme un élément de $\mathfrak{t}^*$ dans le système de coordonnées de la section 3B. Posons $\lambda_i = (t_i+a_i-N)/2-a_{<i}$. On a alors

$$d\Lambda_{\underline{d}} = \big(\underbrace{\lambda_1,\ldots,\lambda_1}_{p_1},\ldots,\underbrace{\lambda_\ell,\ldots,\lambda_\ell}_{p_\ell},\underbrace{-\lambda_\ell,-\ldots,\lambda_\ell}_{q_\ell},\ldots,\underbrace{-\lambda_1,\ldots,-\lambda_1}_{q_1}\big)$$

Le foncteur d'induction cohomologique $(\mathcal{R}^{\mathfrak{g},K}_{\mathfrak{q}_{\underline{d}},L_{\underline{d}}\cap K})^{\dim(\mathfrak{v}_{\underline{d}}\cap\mathfrak{k})}$ permet de produire une représentation de $\boldsymbol{U}(p,q)$ à partir de ce caractère. On pose

$$\mathscr{A}_{\underline{d}}(\psi) = (\mathcal{R}^{\mathfrak{g},K}_{\mathfrak{q}_{\underline{d}},L_{\underline{d}}\cap K})^{\dim(\mathfrak{v}_{\underline{d}}\cap\mathfrak{k})}(\Lambda_{\underline{d}}). \tag{4-3}$$

La condition $t_1 \geq \cdots \geq t_\ell$ assure que cette induction cohomologique est dans le weakly fair range (cf. [Knapp et Vogan 1995, p. 35]). De ceci, il découle que $(\mathcal{R}^{\mathfrak{g},K}_{\mathfrak{q}_{\underline{d}},L_{\underline{d}}\cap K})^i(\Lambda_{\underline{d}}) = 0$, si $i \neq \dim(\mathfrak{v}_{\underline{d}}\cap\mathfrak{k})$, et que si $\mathscr{A}_{\underline{d}}(\psi)$ n'est pas nul, c'est un module unitaire et irréductible, cette dernière propriété étant propre aux groupes unitaires (cf. [Matumoto 1996; Trapa 2001]).

On a défini l'application $\epsilon_{\underline{d}}$ de $[1, \ell]$ dans ± 1 en (1-3). Nous allons expliquer plus loin comment $A(\psi_G)$ s'identifie à un quotient de $\{\pm 1\}^\ell$ et $\epsilon_{\underline{d}}$ à un caractère de $A(\psi_G)$.

Théorème 4.1. *On suppose que ψ est de bonne parité. Alors la représentation associée à ψ est*

$$\pi^A(\psi) := \sum_{\underline{d}} \mathscr{A}_{\underline{d}}(\psi) \boxtimes \epsilon_{\underline{d}}.$$

De plus, les représentations $\mathscr{A}_{\underline{d}}(\psi)$ non nulles sont non isomorphes deux à deux.

Remarque 4.2. On ne précise pas ici quand les $\mathscr{A}_{\underline{d}}(\psi)$ sont non nulles. Toutefois l'article [Trapa 2001] donne un algorithme pour résoudre cette question ; c'est un problème difficile. La deuxième assertion du théorème est une assertion de multiplicité un dans les paquets d'Arthur pour des paramètres de bonne parité. Les résultats de réduction à la bonne parité établis dans la section 5B montrent qu'il y a multiplicité un dans tous les paquets d'Arthur des groupes unitaires.

Démonstration du théorème 4.1. Le théorème est déjà établi sous la condition que les induites cohomologiques (4-3) soient réalisées dans le good range, ce qui est plus restrictif que le weakly fair range, la condition sur ψ étant (1-1). Dans ce cas, $A(\psi_G)$ s'identifie naturellement à $\{\pm 1\}^\ell$.

En effet, sous cette condition, le théorème est un cas particulier de [Mœglin et Renard 2017, théorème 9.3], qui résulte fondamentalement de [Adams et Johnson 1987; Johnson 1984; Arancibia et al. 2018]. De plus, dans ce cas, les $\mathscr{A}_{\underline{d}}(\psi)$ sont non nulles et non isomorphes deux à deux.

On se ramène à ce cas en utilisant les foncteurs de translation (cf. [Knapp et Vogan 1995, Chapter VII and Chapter VIII, §5]). On fixe des entiers $T_1 \gg \cdots \gg T_\ell \geq 0$ et on note ψ_+ le A-paramètre

$$\psi_+ = \bigoplus_i (\chi_{t_i+T_i} \otimes R[a_i]), \tag{4-4}$$

de sorte que ψ_+ vérifie les hypothèse de good range (1-1), et l'on a donc

$$\pi^A(\psi_+) = \bigoplus_{\underline{d} \in \mathcal{D}(\psi_+)} \mathscr{A}_{\underline{d}}(\psi_+) \boxtimes \epsilon_{\underline{d}},$$

où les modules $\mathscr{A}_{\underline{d}}(\psi_+)$ sont définis comme en (4-3) en remplaçant les t_i par les $t_i + T_i$ dans la définition de $\Lambda_{\underline{d}}$ en (4-2). Avec ces hypothèses, comme nous l'avons dit, $A(\psi_+)$ s'identifie à $\{\pm 1\}^{\underline{\ell}}$ et cette expression est alors bien définie.

On considère le foncteur de translation correspondant à la représentation de dimension finie $\mathscr{F}$ qui est la restriction à $\boldsymbol{U}(p, q)$ de la représentation de $\mathbf{GL}_N(\mathbb{C})$

de plus bas poids (dans les coordonnées usuelles pour ce groupe)

$$\big(\underbrace{-T_1/2, \cdots, -T_1/2}_{a_1}, \cdots, \underbrace{-T_\ell/2, \cdots, -T_\ell/2}_{a_\ell}\big).$$

Dans les coordonnées choisies ci-dessus pour $\mathfrak{t}^*$, ce plus bas poids est

$$\gamma = \Big(\underbrace{-T_1/2, \ldots, -T_1/2}_{p_1}, \ldots, \underbrace{-T_\ell/2, \ldots, -T_\ell/2}_{p_\ell}, \\ \underbrace{T_\ell/2, -\ldots, T_\ell/2}_{q_\ell}, \ldots, \underbrace{T_1/2, \ldots, T_1/2}_{q_1}\Big).$$

Posons $\delta(\mathfrak{v}_{\underline{d}}) = \delta(\mathfrak{l}_{p,q}) - \delta(\mathfrak{l}_{\underline{d}})$.

Remarque 4.3. On voit facilement avec les expressions (3-3) et (3-5) que l'on peut définir les T_i vérifiant les conditions ci-dessus en posant $\gamma = -2m\,\delta(\mathfrak{v}_{\underline{d}})$ pour un entier m assez grand.

On note $\mathcal{T}$ ce foncteur et on sait d'après [Mœglin et Renard 2018a, théorème 4.1], que l'on a

$$\mathcal{T}(\pi^A(\psi_+)) = \pi^A(\psi),$$

et d'autre part, d'après [Mœglin et Renard 2018a, théorème 4.3], en remarquant que $\mathcal{D}(\psi_+) = \mathcal{D}(\psi)$, on a $\mathcal{T}(\mathscr{A}_{\underline{d}}(\psi_+)) = \mathscr{A}_{\underline{d}}(\psi)$ pour tout $\underline{d} \in \mathcal{D}(\psi)$. Pour un tel $\underline{d}$ fixé, le caractère $\epsilon_{\underline{d}}$ se factorise en un caractère de $A(\psi_G)$ si et seulement si pour tout $i, j \in [1, \ell]$ tels que $t_i = t_j$ et $a_i = a_j$, on a $\epsilon_{\underline{d}}(i) = \epsilon_{\underline{d}}(j)$. Or on a d'après [Matumoto 1996, Theorem 3.3.4 ; Trapa 2001, Theorem 7.9, Lemma 9.3], que l'on commentera ci-dessous

Proposition 4.4. *La représentation $\mathcal{T}(\mathscr{A}_{\underline{d}}(\psi_+))$ est non nulle seulement si pour tout $i \in [1, \ell[$ tel que $t_{i+1} = t_i$ on a $p_i \geq q_{i+1}$ et $q_i \geq p_{i+1}$.*

Comme l'a remarqué Matumoto juste avant l'énoncé de son théorème, cette proposition est une conséquence du paragraphe 4.2 de [Barbasch et Vogan 1983]. On peut donc dire que cette proposition est essentiellement due à Barbasch et Vogan. On peut faire des inductions par étage et donc supposer que $\ell = 2$, $i = 1$ et $t_1 = t_2$. On rappelle que l'on a ordonné les couples (t_i, a_i) tel que si $t_i = t_{i+1}$ alors $a_i \geq a_{i+1}$ sans cette hypothèse la proposition serait fausse. Cette hypothèse est un oubli dans [Matumoto 1996] et les choix de cette référence pour l'induction cohomologique étant opposés aux nôtres les inégalités sont inversées.

Ainsi pour tout $\underline{d} \in \mathcal{D}(\psi)$, soit le caractère $\epsilon_{\underline{d}}$ se factorise en un caractère de $A(\psi_G)$, soit $\mathscr{A}_{\underline{d}}(\psi)$ est nul, et ceci donne un sens à la formule donnant $\pi^A(\psi)$ dans le théorème et établit la première partie de celui-ci.

Démontrons maintenant l'assertion de multiplicité un. La difficulté vient de l'utilisation du foncteur de translation permettant de passer du good range au weakly fair range. Avec les notations ci-dessus, on sait que si $\underline{d} \neq \underline{d}'$, alors $\mathscr{A}_{\underline{d}}(\psi_+)$ et $\mathscr{A}_{\underline{d}'}(\psi_+)$ sont non isomorphes, et l'on voudrait en déduire que si les représentations $\mathscr{A}_{\underline{d}}(\psi) = \mathcal{T}(\mathscr{A}_{\underline{d}}(\psi_+))$ et $\mathscr{A}_{\underline{d}'}(\psi) = \mathcal{T}(\mathscr{A}_{\underline{d}'}(\psi_+))$ sont non nulles, alors elles sont non isomorphes.

Pour cela, nous allons utiliser des résultats sur les foncteurs de translation disséminés dans [Knapp et Vogan 1995], et nous allons expliquer comment nous y ramener, ce qui nous oblige à quelques détours. Signalons aussi que les idées sont expliquées et mises en œuvre dans [Vogan 1988] dans un contexte différent, et que c'est P. Trapa qui nous a suggéré que la démonstration de l'irréductibilité de l'induction cohomologique convenablement comprise pouvait aussi donner l'énoncé de multiplicité un voulu. La première chose à faire est de se ramener à des foncteurs d'induction cohomologique où la sous-algèbre parabolique θ-stable $\mathfrak{q}$ est fixée et où ce sont les formes réelles fortes au sens de [Adams et al. 1992] qui vont varier. Expliquons le formalisme. On part du groupe compact $\boldsymbol{U}(N)$ et de sa réalisation usuelle comme sous-groupe des points fixes de l'involution $\sigma : g \mapsto {}^t\bar{g}^{-1}$ de $\mathbf{GL}_N(\mathbb{C})$. On note ici $\mathfrak{g}$ l'algèbre de Lie de $\mathbf{GL}_N(\mathbb{C})$. On choisit un tore maximal T de $\boldsymbol{U}(N)$, et l'on note $\mathfrak{t}$ la complexifiée de son algèbre de Lie. On fixe une sous-algèbre de Borel $\mathfrak{b}$ de $\mathfrak{g}$ contenant $\mathfrak{t}$. On identifie T à $\boldsymbol{U}(1)^N$, $\mathfrak{t}$ à $\mathbb{C}^N$ de sorte que les racines simples de $\mathfrak{t}$ dans $\mathfrak{b}$ soient les formes linéaires $e_i - e_{i+1}$, $i = 1, \ldots, N-1$, où $(e_i)_{i=1,\ldots,N}$ est la base canonique de $(\mathbb{C}^N)^*$. La partition $N = \sum_{i=1}^{\ell} a_i$ détermine alors une sous-algèbre parabolique $\mathfrak{q} = \mathfrak{l} \oplus \mathfrak{v}$ de $\mathfrak{g}$ contenant $\mathfrak{b}$. Notons $T[2]$ l'ensemble des éléments d'ordre 2 de T. Pour tout $t \in T[2]$, posons

$$\sigma_t = \mathrm{Ad}(t) \circ \sigma, \quad \theta_t = \mathrm{Ad}(t).$$

Alors σ_t est une forme réelle de $\mathbf{GL}_N(\mathbb{C})$ et nous notons $\boldsymbol{U}_t$ le groupe de ses points réels. L'involution θ_t est une involution de Cartan de $\boldsymbol{U}_t$, et l'on note K_t le sous-groupe de ses points fixes. C'est un sous-groupe compact maximal de $\boldsymbol{U}_t$. Les $\boldsymbol{U}_t$ sont des formes intérieures pures de $\boldsymbol{U}(N) = \boldsymbol{U}_{t=1}$ (on obtient plus généralement toutes les formes réelles fortes au sens de [Adams et al. 1992] en considérant plutôt que $T[2]$ l'ensemble des éléments de T dont le carré est dans le centre de $\boldsymbol{U}_N$, c'est-à-dire de la forme λI_N. Le choix d'une racine carré de λ identifie le groupe des points réels à un $\boldsymbol{U}(p,q)$). Si l'on écrit $t = (\eta_1, \ldots, \eta_N)$ par l'identification $T = \boldsymbol{U}(1)^N$, avec $\eta_i \in \{\pm 1\}$ pour tout $i = 1, \ldots, N$, alors $\boldsymbol{U}_t$ est isomorphe au groupe $\boldsymbol{U}(n_1(t), n_{-1}(t))$ où $n_1(t)$ est le nombre des η_i égaux à 1 et $n_{-1}(t)$ celui des η_i égaux à -1. Pour $t_1, t_2 \in T[2]$, les formes réelles $\boldsymbol{U}_{t_1}$ et $\boldsymbol{U}_{t_2}$ sont équivalentes si et seulement si $n_1(t_1) = n_1(t_2)$ et $n_{-1}(t_1) = n_{-1}(t_2)$. La sous-algèbre parabolique $\mathfrak{q}$ est θ_t-stable pour tout $t \in T[2]$. Posons $L_t = \mathrm{Norm}_{\boldsymbol{U}_t}(\mathfrak{q})$: c'est un c-Levi de $\boldsymbol{U}_t$ et la complexifiée de son algèbre de Lie est $\mathfrak{l}$. Ces groupes L_t, pour $t \in T[2]$, sont des formes réelles d'un groupe complexe $L_{\mathbb{C}}$ isomorphe à $\prod_i \mathbf{GL}_{a_i}(\mathbb{C})$. Le groupe L_t est isomorphe à $\prod_i \boldsymbol{U}(p_i, q_i)$, où p_1 (resp. q_1) est le nombre de 1 (resp. de -1)

dans les a_1 premières coordonnées de t, et ainsi de suite. On a donc pour chaque t des foncteurs d'induction cohomologique

$$(\mathcal{R}^{\mathfrak{g},K_t}_{\mathfrak{q},L_t\cap K_t})^k$$

que nous allons toujours considérer dans le bon degré, c'est-à-dire $S_t=\dim(\mathfrak{u}\cap\mathfrak{k}_t)$.

On revient maintenant à p et q fixés avec $p+q=N$ et pour tout élément $\underline{d}=(p_i,q_i)\in\mathcal{D}(\psi)$, on pose

$$t_{\underline{d}}=\big(\underbrace{1,\dots,1}_{p_1},\underbrace{-1,\dots,-1}_{q_1},\ \dots,\ \underbrace{1,\dots,1}_{p_\ell},\underbrace{-1,\dots,-1}_{q_\ell}\big).$$

On fixe un caractère Λ de $L_{\mathbb{C}}$: dans l'identification de $L_{\mathbb{C}}$ avec $\prod_i \mathbf{GL}_{a_i}(\mathbb{C})$, il est donné par

$$\Lambda=\boxtimes_i \det{}^{(t_i+a_i-N)/2-a_{<i}}\ .$$

On pose alors pour tout $\underline{d}=(p_i,q_i)\in\mathcal{D}(\psi)$,

$$\mathscr{A}'_{\underline{d}}(\psi)=(\mathcal{R}^{\mathfrak{g},K_{t_{\underline{d}}}}_{\mathfrak{q},L_{t_{\underline{d}}}\cap K_{t_{\underline{d}}}})^{S_{t_{\underline{d}}}}(\Lambda).$$

C'est un $(\mathfrak{g},K_{t_{\underline{d}}})$-module. Comme les formes réelles $\boldsymbol{U}_{t_{\underline{d}}}$ sont toutes équivalentes lorsque $\underline{d}$ décrit $\mathcal{D}(\psi)$, c'est-à-dire conjuguées deux à deux par un automorphisme intérieur de $\mathbf{GL}_N(\mathbb{C})$, on peut voir les classes d'équivalence de $(\mathfrak{g},K_{t_{\underline{d}}})$-modules comme des classes d'équivalence de modules de Harish-Chandra pour $\boldsymbol{U}(p,q)$, et les modules $\mathscr{A}'_{\underline{d}}(\psi)$ correspondent aux modules $\mathscr{A}_{\underline{d}}(\psi)$. Il s'agit donc de montrer que les modules $\mathscr{A}'_{\underline{d}}(\psi)$ non nuls sont inéquivalents deux à deux, vus comme modules de Harish-Chandra pour $\boldsymbol{U}(p,q)$. D'autre part, comme les groupes unitaires sont connexes, la catégorie des $(\mathfrak{g},K_{t_{\underline{d}}})$-modules est une sous-catégorie pleine de la catégorie des $\mathfrak{U}(\mathfrak{g})$-modules, où $\mathfrak{U}(\mathfrak{g})$ est l'algèbre enveloppante de $\mathfrak{g}$, et il s'agit donc de montrer que les modules $\mathscr{A}'_{\underline{d}}(\psi)$ non nuls sont inéquivalents deux à deux comme $\mathfrak{U}(\mathfrak{g})$-modules.

Bien sûr, on peut remplacer ψ par ψ_+ dans ces considérations, et l'on obtient de même des modules $\mathscr{A}'_{\underline{d}}(\psi_+)$, $\underline{d}\in\mathcal{D}(\psi_+)=\mathcal{D}(\psi)$ et l'on sait qu'ils sont inéquivalents deux à deux comme $\mathfrak{U}(\mathfrak{g})$-modules.

Rappelons maintenant quelques éléments sur le foncteur de translation tirés de [Knapp et Vogan 1995, Chapter VIII]. Celui-ci est donc défini par la représentation de dimension finie $\mathscr{F}$ de plus bas poids $\gamma=-2m\,\delta(\mathfrak{u})$, d'après la remarque 4.3, et nous sommes donc dans les hypothèses de la proposition 8.31 de [Knapp et Vogan 1995]. Le foncteur $\mathcal{T}$ nous fait passer des modules ayant comme caractère infinitésimal généralisé celui donné par le paramètre ψ_+, notons-le μ_+, à des modules ayant comme caractère infinitésimal généralisé celui donné par le paramètre ψ, notons-le $\mu=\mu_++\gamma=\mu_+-2m\,\delta(\mathfrak{u})$. Ici, on voit μ_+ et μ comme des éléments de t^* qui déterminent chacun un caractère du centre de l'algèbre enveloppante.

Reprenons des éléments de la démonstration de la proposition 8.31 de [Knapp et Vogan 1995], que l'on particularise au cas des groupes unitaires étudiés ici, en adaptant légèrement les notations. On y introduit pour tout $\lambda' \in \mathfrak{t}^*$, un module de Verma généralisé noté $M(\lambda')$, et l'on montre que si λ' vérifie les conditions du weakly fair range, alors $\mathcal{T}(M(\lambda'+2m\,\delta(\mathfrak{v})) = M(\lambda')$ [loc. cit., Lemma 8.35]. On en déduit [loc. cit., Lemma 8.39] que pour toute forme réelle G du groupe complexe ambiant muni d'une involution de Cartan θ ayant comme groupe des points fixes le sous-groupe compact maximal K de G, tel que $\mathfrak{q}$ soit une sous-algèbre parabolique θ-stable, avec de plus les conditions de weakly fair range sur λ', que l'on a

$$\mathcal{T}\big(A_{\mathfrak{q},L\cap K}^{\mathfrak{g},K}(\lambda'+2m\,\delta(\mathfrak{v}))\big) = A_{\mathfrak{q},L\cap K}^{\mathfrak{g},K}(\lambda').$$

En particulier, ceci s'applique aux formes réelles $\boldsymbol{U}_{t_{\underline{d}}}$ définies ci-dessus et a $\lambda'+\delta=\mu$ (δ désigne bien entendu la demi-somme des racines positive), et l'on a donc

$$\mathcal{T}(\mathscr{A}'_{\underline{d}}(\psi_+)) = \mathscr{A}'_{\underline{d}}(\psi),$$

pour tout $\underline{d} \in \mathcal{D}(\psi)$. On introduit $Q(\lambda') = \mathrm{End}(M(\lambda'))$, qui est un $\mathfrak{U}(\mathfrak{g})\otimes\mathfrak{U}(\mathfrak{g})$-module ayant pour caractère infinitésimal $(\lambda'+\delta,\,-(\lambda'+\delta))$ muni d'une application naturelle $\varphi:\mathfrak{U}(\mathfrak{g}) \to Q(\lambda')$ qui respecte les actions à gauche de $\mathfrak{U}(\mathfrak{g})$. Ensuite, on prend les élément $\boldsymbol{U}(N)$-finis de l'algèbre $Q(\lambda')$ en posant $R(\lambda')=\mathrm{End}(M(\lambda'))_{\boldsymbol{U}(N)}$. Ainsi, $R(\lambda')$ devient un $(\mathfrak{g}\oplus\mathfrak{g},\,\boldsymbol{U}(N)\times\boldsymbol{U}(N))$-module, et l'image de φ est à valeurs dans $R(\lambda')$. On pose alors $S=S(\lambda')=R(\lambda')\otimes\mathrm{End}(\mathscr{F})$, et cette algèbre admet une décomposition selon ses composantes primaires (à gauche et à droite), $S=\bigoplus_{\alpha,\beta} S_\alpha^\beta$.

Prenons maintenant $\lambda'=\mu_+-\delta$. La composante $S_\mu^{-\mu}$ est une sous-algèbre de S. Soit M un $\mathfrak{U}(\mathfrak{g})$-module ayant pour caractère infinitésimal μ_+, et supposons que M soit aussi un module à gauche unifère pour $R(\lambda')$ tel que les deux actions soient compatibles via φ. Alors la composante μ-primaire N_μ du S-module $N=M\otimes\mathscr{F}$ est naturellement un $S_\mu^{-\mu}$-module. Ce qui est fondamental pour nous ici est le résultat suivant : si M est un $R(\lambda')$-module simple, alors N_μ est $S_\mu^{-\mu}$-module simple ou bien 0, et de plus, si M^1, M^2 sont deux $R(\lambda')$-module simples non équivalents, et si N_μ^1 et N_μ^2 sont non nuls, alors ce sont deux $S_\mu^{-\mu}$-modules simples non équivalents. On trouve la démonstration à la page 524 de [Knapp et Vogan 1995], pour une algèbre S différente, mais la démonstration est formelle et se transpose sans changements. D'ailleurs un résultat formel général analogue (avec la même démonstration) dans le cadre des algèbres à idempotents se trouve dans [Renard 2010, proposition I.3.2].

Il y a un foncteur de translation pour les $\mathfrak{U}(\mathfrak{g})\otimes\mathfrak{U}(\mathfrak{g})$-modules construit avec $\mathrm{End}(\mathscr{F})$ allant des modules ayant pour caractère infinitésimal généralisé $(\mu_+,\,-\mu_+)$ vers les modules ayant pour caractère infinitésimal généralisé $(\mu,\,-\mu)$. Notons le $\mathcal{T}^2$. On a alors

$$S_\mu^{-\mu} = \mathcal{T}^2(R(\mu_+-\delta)) = R(\mu-\delta).$$

De plus, on peut munir les $\mathscr{A}'_{\underline{d}}(\psi_+)$, $\underline{d} \in \mathcal{D}(\psi)$, d'une structure de $R(\lambda')$-module compatibles avec l'action de $\mathfrak{U}(\mathfrak{g})$, ceci apparaît à la page 577 de [Knapp et Vogan 1995]. Il résulte de ceci que les $\mathscr{A}'_{\underline{d}}(\psi_+)$, $\underline{d} \in \mathcal{D}(\psi)$ sont des $R(\mu-\delta)$-modules nuls ou simples, les non nuls étant inéquivalents deux à deux. La proposition 8.31 de [Knapp et Vogan 1995] a en fait pour but d'énoncer un critère pour en déduire que ce sont des $\mathfrak{U}(\mathfrak{g})$-modules nuls ou simple, il suffit que l'application naturelle φ de $\mathfrak{U}(\mathfrak{g})$ dans $R(\mu-\delta)$ soit surjective. Or c'est le cas pour les groupes unitaires, cela vient du fait que les orbites nilpotentes en type A sont de Richardson avec une application moment birationnelle et sont d'adhérence normale [Kraft et Procesi 1979]. C'est ainsi que l'on montre que les $A_{\mathfrak{q}}(\lambda)$-modules dans le weakly fair range sont nuls ou irréductibles pour les groupes unitaires. Ce que nous venons de remarquer ici, c'est que l'inspection de la démonstration montre en plus que les modules $\mathscr{A}'_{\underline{d}}(\psi)$, $\underline{d} \in \mathcal{D}(\psi)$ non nuls ne sont pas équivalents en tant que $\mathfrak{U}(\mathfrak{g})$-modules et ceci termine la démonstration du théorème. □

Remarque 4.5. La définition des paquets d'Arthur par les identités de transfert endoscopiques suppose avoir choisi parmi les formes réelles fortes (au sens de [Adams et al. 1992]) $\boldsymbol{U}_t$, $t \in T[2]$, introduites ci-dessus, une forme quasi-déployée, et pour celle-ci, une donnée de Whittaker. Ici, la forme quasi-déployée choisie est donné par

$$t_* = \big(1, -1, 1, -1, \dots, (-1)^{N-1}\big)$$

et est donc isomorphe à $\boldsymbol{U}(N/2, N/2)$ si N est pair et $\boldsymbol{U}(\lfloor N/2\rfloor+1, \lfloor N/2\rfloor)$ si N est impair. D'autre part, on peut considérer un paramètre de séries discrètes ψ_G, pour les groupes unitaires de rang N, c'est-à-dire avec $\psi = \bigoplus_{i=1}^{N}(\chi_{t_i} \boxtimes R[1])$ avec les t_i distincts. Les constructions faites ci-dessus pour $\underline{d} \in \mathcal{D}(\psi) = ((1,0), (0,1), (1,0), \dots)$ déterminent une série discrète générique de ce groupe unitaire quasi-déployé, et l'on fixe la donnée de Whittaker pour que cette série discrète admette une fonctionnelle de Whittaker. Des choix différents mèneraient à une formule différente en (1-3) en tordant la paramétrisation par un caractère de $A(\psi_G)$. Voir [Mœglin et Renard 2017] pour une discussion analogue pour les groupes classiques.

5. Réduction au cas de bonne parité

5A. ***Mauvaise parité et induction parabolique.*** Dans cette section, nous démontrons des résultats énoncés sans démonstration dans [Mœglin et Renard 2017] ainsi que leurs analogues pour les groupes unitaires. Soient $\boldsymbol{G}$ un groupe classique ou unitaire, et ψ_G, ψ comme dans la section 2. Considérons une décomposition de ψ de la forme :

$$\psi = \rho \oplus \rho^* \oplus \psi' \tag{5-1}$$

où, dans ρ, il n'apparaît que des facteurs de mauvaise parité. Ici ρ^* désigne la représentation contragrédiente si $\boldsymbol{G}$ est un groupe classique, et la duale hermitienne si $\boldsymbol{G}$ est un groupe unitaire. Remarquons que toute la partie de mauvaise parité peut se mettre sous la forme $\rho \oplus \rho^*$.

Si $\boldsymbol{G}$ est classique, le paramètre ψ' se factorise par le L-groupe d'un groupe classique quasi-déployé $\boldsymbol{G}^\flat$ de même type que $\boldsymbol{G}$. Soit $\psi_{G^\flat}$ le paramètre d'Arthur pour le groupe $\boldsymbol{G}^\flat$ tel que $\psi' = \mathbf{Std}_{G'} \circ \psi_{G^\flat}$. De même, si $\boldsymbol{G}$ est unitaire, ψ' est la restriction à $\mathbb{C}^\times \times \mathbf{SL}_2(\mathbb{C})$ d'un paramètre $\psi_{G'}$ pour un groupe unitaire quasi-déployé $\boldsymbol{G}^\flat$ de rang plus petit.

Notons N_ρ la dimension de la représentation ρ. Si $\boldsymbol{G}$ est classique, c'est une représentation de $W_\mathbb{R} \times \mathbf{SL}_2(\mathbb{C})$, et l'on note $\Pi_\rho^{\mathbf{GL}}$ la représentation de $\mathbf{GL}_{N_\rho}(\mathbb{R})$ de paramètre d'Arthur ρ (cf. [Arancibia et al. 2018, §3.1]). Si $\boldsymbol{G}$ est unitaire, c'est une représentation de $\mathbb{C}^\times \times \mathbf{SL}_2(\mathbb{C})$, et l'on note $\Pi_\rho^{\mathbf{GL}}$ la représentation de $\mathbf{GL}_{N_\rho}(\mathbb{C})$ de paramètre d'Arthur ρ. Pour unifier les notations, on note simplement $\mathbf{GL}_{N_\rho}$ pour le groupe $\mathbf{GL}_{N_\rho}(\mathbb{R})$ si l'on est dans le cadre des groupes classiques, et $\mathbf{GL}_{N_\rho}(\mathbb{C})$ si l'on est dans le cadre des groupes unitaires.

Selon la forme intérieure G et la dimension N_ρ de ρ, le groupe G admet ou pas un sous-groupe de Levi maximal standard M isomorphe à $\mathbf{GL}_{N_\rho} \times G'$, où $\boldsymbol{G}'$ est une forme intérieure de $\boldsymbol{G}^\flat$. Par exemple, si $G = \boldsymbol{U}(p,q)$, la condition est que $\inf(p,q) \geq N_\rho$, et si c'est le cas, on a $G' = \boldsymbol{U}(p - N_\rho, q - N_\rho)$, et on a la même condition si $G = \mathbf{SO}(p,q)$. Si $\boldsymbol{G}$ est quasi-déployé, la condition est toujours vérifiée avec $\boldsymbol{G}' = \boldsymbol{G}^\flat$, et ceci fournit une injection

$$\iota : {}^L M = \bigl(\widehat{\mathbf{GL}_{N_\rho}} \times \widehat{G^\flat}\bigr) \rtimes W_\mathbb{R} \hookrightarrow {}^L G \tag{5-2}$$

de sorte que $\psi_G = \iota \circ \psi_M$ où $\psi_M : W_\mathbb{R} \to {}^L M$ est construit à partir de ρ et $\psi_{G^\flat}$. Ici $\widehat{\mathbf{GL}_{N_\rho}} = \mathbf{GL}_N(\mathbb{C})$ si $\mathbf{GL}_{N_\rho} = \mathbf{GL}_{N_\rho}(\mathbb{R})$ et $\widehat{\mathbf{GL}_{N_\rho}} = \mathbf{GL}_N(\mathbb{C}) \times \mathbf{GL}_N(\mathbb{C})$ si $\mathbf{GL}_{N_\rho} = \mathbf{GL}_{N_\rho}(\mathbb{C})$.

Comme le groupe $\boldsymbol{G}^\flat$ ne joue pas de rôle dans ce qui suit, on note plutôt $\psi_{G'}$ pour $\psi_{G^\flat}$ du moins si la condition d'existence de $\boldsymbol{G}'$ est satisfaite. On vérifie facilement l'énoncé suivant.

Remarque 5.1. Les groupes $A(\psi_G)$ et $A(\psi_{G'})$ sont naturellement isomorphes.

Reprenons les notations de l'introduction, où pour un paramètre d'Arthur ψ_G pour le groupe G, nous avons noté $\pi^A(\psi_G)$ la représentation unitaire de longueur finie de $G \times A(\psi_G)$ attachée à ψ_G. On la note aussi $\pi^A(\psi_G, G)$. On décompose maintenant cette représentation selon les caractères du groupe abélien fini $A(\psi_G)$:

$$\pi^A(\psi, G) = \bigoplus_{\eta \in \widehat{A(\psi_G)}} \pi(\psi_G, \eta, G) \boxtimes \eta \tag{5-3}$$

où les $\pi(\psi_G, \eta, G)$ sont maintenant des représentations unitaires de longueur finie de G.

Proposition 5.2. *Avec les notations ci-dessus, si la condition d'existence de la forme intérieure* $\boldsymbol{G}'$ *n'est pas vérifiée, on a* $\pi^A(\psi_G)=0$.

Si la condition d'existence de la forme intérieure $\boldsymbol{G}'$ *est vérifiée, soit* $\eta\in\widehat{A(\psi_G)}$ *et soient* $\pi(\psi_G,\eta,G)$ *et* $\pi(\psi_{G'},\eta,G')$ *les représentations semi-simples de* G *et* G' *respectivement attachées par Arthur* (*cf.* (5-3), *où pour* $\pi(\psi_{G'},\eta,G')$ *on tient compte de la remarque ci-dessus*). *On a alors*

$$\pi(\psi_G,\eta,G)=\operatorname{Ind}_P^G\big(\Pi_\rho^{\mathbf{GL}}\otimes\pi(\psi_{G'},\eta,G')\big),\tag{5-4}$$

où P *est un sous-groupe parabolique standard maximal de* G *de facteur de Levi* M *isomorphe à* $\mathbf{GL}_{N_\rho}\times G'$.

Pour les groupes classiques, c'est la proposition 4.3 de [Mœglin et Renard 2017], énoncée sans démonstration.

Démonstration. Notons $\pi^B(\psi_G,\eta,G)$ la représentation induite du membre de droite dans (5-4), et

$$\pi^B(\psi_G,G)=\bigoplus_{\eta\in\widehat{A(\psi_G)}}\pi^B(\psi_G,\eta,G)\boxtimes\eta.$$

Nous avons besoin de savoir que $\pi^B(\psi_G,G)$ est non nul si G est quasi-déployé avant de pouvoir démontrer que cette représentation est $\pi^A(\psi_G,G)$. Évidemment si G est quasi-déployé, on a remarqué que la condition d'existence de $\boldsymbol{G}'$ est toujours satisfaite, et que $\boldsymbol{G}'=\boldsymbol{G}^\flat$ est quasi-déployé. Ainsi $\pi^B(\psi_G,G)$ est non nul si et seulement si $\pi^A(\psi_{G'},G')$ est non nul. Or on sait que le paquet $\Pi^A(\psi_{G'},G')$ est non nul, car il contient au moins les représentations dans le paquet de Langlands associé au paquet d'Arthur. Ceci montre l'assertion voulue. Considérons alors la représentation virtuelle stable $\pi^A(\psi_{G'},G')(s_{\psi'})=[\pi(\psi_{G'},G')]$, où $s_{\psi'}=\psi'(1,-\mathrm{Id})$ (cf. ([Mœglin et Renard 2017, (2.3.3)]). Elle vérifie l'identité endoscopique tordue [Mœglin et Renard 2017, (3.2.4)]. Comme le transfert endoscopique tordu commute avec l'induction, on obtient que le transfert tordu de la représentation virtuelle $\pi^B(\psi_G,G)(s_\psi)$ est la trace tordue de l'induite pour le parabolique standard de $\mathbf{GL}_N$ de Levi standard $\mathbf{GL}_{N_\rho}\times\mathbf{GL}_{N'}$ de la représentation $\Pi_\rho^{\mathbf{GL}}\boxtimes\Pi_{\psi'}^{\mathbf{GL}}$. Or cette induite est $\Pi_\psi^{\mathbf{GL}}$, d'après la définition des paquets d'Arthur des groupes linéaires (voir [Arancibia et al. 2018, section 3.2]). D'autre part $\pi^A(\psi_G,G)(s_\psi)=[\pi(\psi_G,G)]$ vérifie aussi l'identité endoscopique tordue. On obtient donc que $\pi^A(\psi_G,G)(s_\psi)=\pi^B(\psi_G,G)(s_\psi)$, puisque ces deux représentations virtuelles stables ont même transfert endoscopique tordu.

Nous allons démontrer que $\pi^B(\psi_G,G)$ vérifie aussi les identités endoscopiques ordinaires [Mœglin et Renard 2017, (2.3.5)]. On ne suppose plus que G est quasi-déployé, mais l'on suppose que la condition d'existence de $\boldsymbol{G}'$ est satisfaite, car sinon, il est clair que $\pi^A(\psi_G,G)=0$. Soit $\boldsymbol{H}=(H,x,\xi:{}^LH\to{}^LG,\dots)$ une donnée endoscopique elliptique de G (cf. [Arthur 2013]) telle que ψ_G se factorise par le groupe dual de H et on fixe une telle factorisation $\psi_G=\xi\circ\psi_x$. En particulier l'élément $x \in {}^L G$ s'identifie à un élément du commutant de ψ_G. Il faut alors démontrer qu'il existe une donnée endoscopique elliptique $\boldsymbol{H}' = (H', x', \ldots)$ de G', tel que l'élément x' de cette donnée soit dans le centralisateur de $\psi_{G'}$ et tel que le transfert de la distribution stable associée à $\boldsymbol{H}$ et à la factorisation de ψ_G soit l'induite du produit tensoriel des données analogues pour $\psi_{G'}$ et $\boldsymbol{H}'$ et de la représentation Π_ρ^{GL}. Expliquons maintenant comment construire explicitement cette donnée endoscopique $\boldsymbol{H}'$. Comme il est loisible de le faire ici, on suppose que x vérifie $x^2 = 1$. On décompose alors ψ en $\psi_+ \oplus \psi_-$ suivant les valeurs propres de x. On remarque que l'on a aussi une décomposition analogue pour ψ' et pour ρ. On a alors

$$\psi_+ = \rho_+ \oplus \rho_+^* \oplus \psi'_+$$

et une décomposition analogue avec $+$ remplacé par $-$. C'est ici qu'a servi l'hypothèse sur la mauvaise parité des composantes de ρ, pour que le dual de ρ_+ apparaisse lui aussi dans l'espace propre de valeur propre $+1$. Notons $N_{\rho_\pm}$ les dimensions des représentations $\rho_\pm$, et $\Pi_{\rho_\pm}^{\mathbf{GL}}$ la représentation de $\mathbf{GL}_{N_{\rho_\pm}}$ associée à ce paramètre. On a bien sûr $N_\rho = N_{\rho_+} + N_{\rho_-}$ et $\Pi_\rho^{\mathbf{GL}}$ est l'induite parabolique de $\Pi_{\rho_+}^{\mathbf{GL}} \boxtimes \Pi_{\rho_-}^{\mathbf{GL}}$.

Ainsi il existe un sous-groupe de Levi

$$M_H \simeq \left(\mathbf{GL}_{N_{\rho_+}} \times M^+\right) \times \left(\mathbf{GL}_{N_{\rho_-}} \times M^-\right)$$

de H tel que ψ_x se factorise par le L-groupe de M_H et la représentation virtuelle stable $[\pi(\psi_x, H)]$ de H associée à ψ_x est une induite à partir de ce Levi. Notons H' le facteur $M^+ \times M^-$ de M : c'est un groupe endoscopique pour G', s'inscrivant dans une donnée endoscopique $\boldsymbol{H}' = (H', x', \xi', \ldots)$ de G' et le paramètre d'Arthur $\psi_{G'}$ se factorise en $\xi' \circ \psi_{x'}$. L'élément x' est dans le centralisateur de $\psi_{G'}$, on peut le prendre tel que $x'^2 = 1$ et $\psi' = \psi'^+ \oplus \psi'^-$ est la décomposition de ψ' selon les valeurs propres ± 1 de x'. Partons de la représentation stable $[\pi(\psi_{x'}, H')]$ associée à $\psi_{x'}$. On peut d'abord considérer son transfert endoscopique vers G', puis induire vers G avec $\Pi_\rho^{\mathbf{GL}}$:

$$\mathrm{Ind}_{P=MN}^{G}\big(\Pi_\rho^{\mathbf{GL}} \boxtimes \mathrm{Trans}_{H'}^{G'}([\pi(\psi_{x'}, H')])\big),$$

où $\mathrm{Trans}_{H'}^{G'}$ désigne le transfert endoscopique (spectral) du groupe endoscopique H' de G' vers G'. Or, ceci est égal à

$$\mathrm{Ind}_{P=MN}^{G}\big(\Pi_\rho^{\mathbf{GL}} \boxtimes \pi^A(\psi_{G'}, H')(s_{\psi'} x')\big) = \pi^B(\psi_G, G)(s_\psi x).$$

Le fait que le transfert commute à l'induction nous dit que l'on obtient le même résultat en prenant le produit tensoriel extérieur avec $\Pi_{\rho_+}^{\mathbf{GL}}$ et $\Pi_{\rho_-}^{\mathbf{GL}}$ pour obtenir une représentation virtuelle de M_H que l'on induit vers H, puis en prenant ensuite le transfert endoscopique de H vers G :

$$\mathrm{Trans}_H^G\big(\mathrm{Ind}_{P_h=M_H N_H}^{H}\big(\Pi_{\rho_+}^{\mathbf{GL}} \boxtimes \Pi_{\rho_-}^{\mathbf{GL}} \boxtimes [\pi(\psi_{x'}, G')]\big)\big).$$

Or ceci est égal à

$$\mathrm{Trans}_H^G([\pi(\psi_x, H)]) = \pi^A(\psi_G, G)(s_\psi x).$$

On obtient donc que $\pi^B(\psi_G, G)(s_\psi x) = \pi^A(\psi_G, G)(s_\psi x)$. Comme on a ceci pour tout $x \in A(\psi_G)$, que $\boldsymbol{G}$ soit ou non quasi-déployé, on en déduit par inversion de Fourier que $\pi^B(\psi_G, G) = \pi^A(\psi_G, G)$. □

5B. ***Irréductibilité de l'induite parabolique pour les groupes unitaires.*** Nous allons reformuler la proposition 5.2 de manière un peu plus explicite pour les groupes unitaires, en y ajoutant un résultat d'irréductibilité des induites paraboliques.

Théorème 5.3. *Soient ψ_G un A-paramètre pour $G = \boldsymbol{U}(p, q)$, ψ sa restriction à $\mathbb{C}^\times \times \mathbf{SL}_2(\mathbb{C})$ comme en (2-4) et ψ_{bp} la partie de bonne parité de ce paramètre. Soient $N = p+q$ et N_{bp} comme en (2-6). En particulier $N - N_{\mathrm{bp}}$ est pair, et l'on pose $a_{\mathrm{mp}} = (N - N_{\mathrm{bp}})/2$, c'est le cardinal de l'ensemble $\mathcal{E}'(\psi)$. On a alors :*

(i) *Si* $\inf(p, q) < a_{\mathrm{mp}}$, *alors* $\pi^A(\psi_G, G) = 0$.

(ii) *Si* $\inf(p, q) \geq a_{\mathrm{mp}}$ *on pose $p_{\mathrm{bp}} = p - a_{\mathrm{mp}}$, $q_{\mathrm{bp}} = q - a_{\mathrm{mp}}$ et ψ_{bp} est la restriction à $\mathbb{C}^\times \times \mathbf{SL}_2(\mathbb{C})$ d'un paramètre $\psi_{G_{\mathrm{bp}}}$ pour $G_{\mathrm{bp}} = \boldsymbol{U}(p_{\mathrm{bp}}, q_{\mathrm{bp}})$. On a donc une représentation unitaire $\pi^A(\psi_{G_{\mathrm{bp}}}, G_{\mathrm{bp}})$ de $\boldsymbol{U}(p_{\mathrm{bp}}, q_{\mathrm{bp}}) \times A(\psi_{G_{\mathrm{bp}}})$, qui s'écrit*

$$\pi^A(\psi_{G_{\mathrm{bp}}}, G_{\mathrm{bp}}) = \bigoplus_{\eta \in \widehat{A(\psi_{G_{\mathrm{bp}}})}} \pi(\psi_{G_{\mathrm{bp}}}, \eta, G_{\mathrm{bp}}) \boxtimes \eta.$$

Alors $\pi^A(\psi_G, G)$ s'écrit $\pi^A(\psi_G, G) = \bigoplus_{\eta \in \widehat{A(\psi)}} \pi(\psi_G, \eta, G) \boxtimes \eta$, avec pour tout $\eta \in A(\psi_G)$ (rappelons qu'en vertu de la remarque 5.1, on peut identifier $A(\psi_{G_{\mathrm{bp}}})$ et $A(\psi_G)$),

$$\pi(\psi_G, \eta, G) = \mathrm{Ind}_P^G\Bigg(\bigg(\boxtimes_{(t,s,a)\in\mathcal{E}'(\psi)} \chi_{t,s,a}\bigg) \boxtimes \pi(\psi_{G_{\mathrm{bp}}}, \eta, G_{\mathrm{bp}})\Bigg)$$

pour le sous-groupe parabolique standard P de $\boldsymbol{U}(p, q)$ dont le sous-groupe de Levi est $\prod_{(t,s,a)\in\mathcal{E}'(\psi)} \mathbf{GL}(a, \mathbb{C}) \times \boldsymbol{U}(p_{\mathrm{bp}}, q_{\mathrm{bp}})$.

De plus, si τ est une sous-représentation irréductible de $\pi(\psi_{G_{\mathrm{bp}}}, \eta, G_{\mathrm{bp}})$, alors $\mathrm{Ind}_P^G\big(\big(\boxtimes_{(t,s,a)\in\mathcal{E}'(\psi)} \chi_{t,s,a}\big) \boxtimes \tau\big)$ est irréductible.

Démonstration. Seule la dernière assertion est nouvelle par rapport à la proposition. Les représentations τ de la dernière assertion du théorème sont les représentations $\mathscr{A}_{\underline{d}}(\psi_{\mathrm{bp}})$ de la section 4 attachée à la partie de bonne parité du paramètre. Il est démontré dans [Matumoto 1996, Theorem 3.2.2(2)] que l'induite parabolique de

$$\bigg(\boxtimes_{(t,s,a)\in\mathcal{E}'(\psi)} \chi_{(t,s,a)}\bigg) \boxtimes \mathscr{A}_{\underline{d}}(\psi_{\mathrm{bp}})$$

est irréductible, ce qui est exactement l'assertion voulue. □

5C. ***Irréductibilité de l'induction parabolique pour les groupes classiques.*** Nous complétons maintenant la proposition 5.2 pour les groupes classiques en y ajoutant le fait que, comme pour les groupes unitaires, l'induction parabolique préserve l'irréductibilité. Ce résultat avait été énoncé sans démonstration dans [Mœglin et Renard 2017, théorème 4.4].

Théorème 5.4. *On se place dans les hypothèses de la proposition 5.2. On suppose que la condition d'existence de la forme intérieure* $\mathbf{G}'$ *est vérifiée. Soit* $\eta \in \widehat{A(\psi_G)}$. *Si* τ *est une sous-représentation irréductible de* $\pi(\psi_{G'}, \eta, G')$, *alors* $\mathrm{Ind}_P^G(\Pi_\rho^{\mathbf{GL}} \boxtimes \tau)$ *est irréductible.*

Rappelons que pour un groupe classique G, la mauvaise parité est : impaire si $\widehat{G}$ est un groupe spécial orthogonal, paire si $\widehat{G}$ est un groupe symplectique. On écrit la partie de mauvaise parité ρ (cf. [Mœglin et Renard 2017, §4.1]) sous la forme

$$\rho = \bigoplus_{i=1,\dots,\ell} \delta_{t_i,s_i} \boxtimes R[a_i] \oplus \bigoplus_{j=1,\dots,\ell'} \eta_{\epsilon_j,s_j} \boxtimes R[a'_j].$$

Dans la première somme, $t_i \in \mathbb{Z}_{>0}$, $s_i \in i\mathbb{R}$, et δ_{t_i,s_i} est le paramètre de Langlands de la série discrète de caractère infinitésimal $\big((t_i+s_i)/2, (-t_i+s_i)/2\big)$ de $\mathbf{GL}_2(\mathbb{R})$, et si $s_i = 0$, alors $t_i + a_i - 1$ est de mauvaise parité. Dans la seconde somme $\epsilon_j \in \{\pm 1\}$, $s_j \in i\mathbb{R}$, et η_{ϵ_j,s_j} est le paramètre de Langlands du caractère $x \mapsto \mathbf{sgn}(x)^{(1-\epsilon_j)/2}|x|^{s_j}$ de $\mathbf{GL}_1(\mathbb{R})$, et si $s_j = 0$, alors $a'_j - 1$ est de mauvaise parité. On note encore δ_{t_i,s_i} et η_{ϵ_j,s_j} les représentations de $\mathbf{GL}_2(\mathbb{R})$ et $\mathbf{GL}_1(\mathbb{R})$ dont ce sont les paramètres. Pour chaque indice i, on considère la représentation de Speh, notée $\mathbf{Speh}(\delta_{t_i,s_i}, a_i)$ de $\mathbf{GL}_{2a_i}(\mathbb{R})$ qui est irréductible et unitaire, et pour chaque indice j, le caractère unitaire $\eta_{\epsilon_j,s_j} \circ \det$ de $\mathbf{GL}_{a'_j}(\mathbb{R})$. La représentation $\Pi_\rho^{\mathbf{GL}}$ est alors obtenue par induction parabolique irréductible à partir de la représentation

$$\Big(\boxtimes_{i=1,\dots,\ell} \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Big) \boxtimes \Big(\boxtimes_{j=1,\dots,\ell} \eta_{\epsilon_j,s_j} \circ \det\Big)$$

du facteur de Levi

$$\Big(\prod_{i=1,\dots,\ell} \mathbf{GL}_{2a_i}(\mathbb{R})\Big) \times \Big(\prod_{j=1,\dots,\ell} \mathbf{GL}_{a'_j}(\mathbb{R})\Big)$$

de $\mathbf{GL}_{N_\rho}(\mathbb{R})$.

Soit G_0 le groupe de même type que G tel que $\mathbf{GL}_{N_\rho}(\mathbb{R}) \times G_0$ est un sous-groupe de Levi d'un parabolique P de G. On note τ_0 une représentation unitaire irréductible de G_0. On note N_0 son rang. On suppose que le caractère infinitésimal de τ_0 a bonne parité, c'est-à-dire qu'il est formé d'entiers si la demi-somme des racines positive de G_0 est formée d'entiers et est formé de demi-entiers non entiers sinon. Le fait que τ_0 soit unitaire n'est absolument pas nécessaire mais simplifie légèrement la preuve. Le théorème résulte alors de la proposition plus générale suivante.

Proposition 5.5. *La représentation induite* $\mathrm{Ind}_P^G(\Pi_\rho^{\mathrm{GL}} \boxtimes \tau_0)$ *est irréductible.*

Démonstration. On note $\pi = \mathrm{Ind}_P^G(\Pi_\rho^{\mathrm{GL}} \boxtimes \tau_0)$. Nous allons scinder la démonstration en plusieurs étapes.

Première étape. Pour tout $j \in [1, \ell']$, on remplace $\eta_{\epsilon_j, s_j} \boxtimes R[a'_j]$ dans la partie de mauvaise parité ρ par

$$\left(\eta_{\epsilon_j, s_j} \boxtimes R[a'_j]\right) \oplus \left(\eta_{-\epsilon_j, s_j} \boxtimes R[a'_j]\right).$$

On obtient ainsi un paramètre $\rho^\sharp$ de dimension plus grande, et toujours de mauvaise parité, et il est clair que la proposition est vraie pour ρ si elle l'est pour $\rho^\sharp$. Remarquons que l'on peut poser $\eta_{\epsilon_j, s_j} \oplus \eta_{-\epsilon_j, s_j} = \delta_{0, s_j}$ (la limite de séries discrètes δ_{0,s_j}de $\mathbf{GL}_2(\mathbb{R})$ est l'induite de $\eta_{\epsilon_j, s_j} \boxtimes \eta_{-\epsilon_j, s_j}$). Ainsi, on peut supposer que $\rho = \bigoplus_{i=1,\dots,\ell} \delta_{t_i, s_i} \boxtimes R[a_i]$, mais il est maintenant possible que certains t_i soient nuls. La dimension N_ρ de ρ est paire, et l'on pose $N'_\rho = N_\rho/2 = \sum_i a_i$. On change maintenant légèrement les notations, $P = MN$ désigne maintenant le sous-groupe parabolique standard de G de facteur de Levi isomorphe à $\left(\prod_i \mathbf{GL}_{2a_i}(\mathbb{R})\right) \times G_0$. La représentation π est donc avec ces notations

$$\pi = \mathrm{Ind}_P^G\left(\left(\boxtimes_i \mathbf{Speh}(\delta_{t_i, s_i}, a_i)\right) \boxtimes \tau_0\right).$$

Deuxième étape. Les représentations $\mathbf{Speh}(\delta_{t_i, s_i}, a_i)$ sont obtenues à partir du caractère χ_{t_i, s_i, a_i} de $\mathbf{GL}_{a_i}(\mathbb{C})$ par induction cohomologique. Ceci est bien connu, voir par exemple [Knapp et Vogan 1995, p. 586], et l'on utilise ici la version normalisée de l'induction cohomologique [loc. cit., (11.150b)]. Ainsi, π est obtenue en deux étapes, d'abord une induction cohomologique, puis une induction parabolique. Le théorème de transfert du chapitre 11 de [loc. cit.] permet d'échanger l'ordre de ces deux inductions. Une référence commode pour cela est [Matumoto 2004, Theorem 2.2.3], qui nous donne exactement l'énoncé voulu. Donnons les détails. Matumoto introduit la terminologie de $\sigma\theta$-paire pour un couple $(\mathfrak{p}, \mathfrak{q})$ de sous-algèbres paraboliques de G. La sous-algèbre $\mathfrak{p}$ est la complexifiée de l'algèbre de Lie d'un sous-groupe parabolique P de G, qui ici a été fixée à la fin de la première étape. La sous-algèbre $\mathfrak{q}$ est une sous-algèbre parabolique θ-stable. On pose $L = \mathrm{Norm}_G(\mathfrak{q})$. On choisit ici cette sous-algèbre de sorte que d'une part L soit isomorphe à $U(N'_\rho, N'_\rho) \times G_0$ et que d'autre part la condition S2 de la définition 2.1.1 de [Matumoto 2004] soit vérifiée, c'est-à-dire que $\mathfrak{p} \cap \mathfrak{q}$ contient une sous-algèbre de Cartan σ et θ-stable de $\mathfrak{g}$. Il est facile de vérifier que l'on trouve une telle sous-algèbre $\mathfrak{q}$, en partant d'un sous-groupe de Cartan isomorphe à $(\mathbb{C}^\times)^{N'_\rho} \times \boldsymbol{U}(1)^{N_0}$ de M. On adopte les notations de Matumoto, qui sont usuelles ($\mathfrak{p} = \mathfrak{m} \oplus \mathfrak{n},\ \mathfrak{q} = \mathfrak{l} \oplus \mathfrak{u}$, etc.). On a ainsi

$$\pi = \mathrm{Ind}_P^G\left(\left({}^n\mathcal{R}^{\mathfrak{m}, M\cap K}_{\mathfrak{q}\cap\mathfrak{m}, L\cap M\cap K}\right)^{\dim(\mathfrak{u}\cap\mathfrak{m}\cap\mathfrak{k})}\left(\left(\boxtimes_i \chi_{t_i, s_i, a_i}\right) \boxtimes \tau_0\right)\right).$$

Remarquons que ici $L \cap M$ est isomorphe à $\big(\prod_i \mathbf{GL}_{a_i}(\mathbb{C})\big) \times G_0$. Le résultat de Matumoto nous permet alors d'écrire, sous certaines conditions de positivité des paramètres,

$$\pi = (\mathcal{R}^{\mathfrak{g},K}_{\mathfrak{q},L\cap K})^{\dim(\mathfrak{u}\cap\mathfrak{k})}\Bigg(\mathrm{Ind}^{L}_{P\cap L}\bigg(\bigg(\Big(\boxtimes_{i} \chi_{t_i,s_i,a_i}\Big) \boxtimes \tau_0\bigg) \otimes \mathbb{C}_{-\delta(\mathfrak{u})}\bigg)\Bigg).$$

Ici, on a simplifié la formulation de Matumoto, en utilisant le fait que le groupe $L \cap M$ est connexe, et ainsi le caractère noté $\mathbb{C}_{\delta(\bar{\mathfrak{n}}\cap\mathfrak{u})'}$ par Matumoto venant de la normalisation subtile des foncteurs d'induction cohomologique dans [Knapp et Vogan 1995] coïncide ici avec $\mathbb{C}_{\delta(\bar{\mathfrak{n}}\cap\mathfrak{u})}$ (en général ces deux caractères sont seulement égaux sur la composante neutre de $L \cap M$). D'autre part, si l'on suppose les t_i suffisamment grand, les hypothèses de positivité dans le théorème de Matumoto sont vérifiées.

On veut montrer que π est irréductible. Or, avec les t_i suffisamment grand, l'induction cohomologique $(\mathcal{R}^{\mathfrak{g},K}_{\mathfrak{q},L\cap K})^{\dim(\mathfrak{u}\cap\mathfrak{k})}$ se fait dans le good range et préserve donc l'irréductibilité. Il suffit alors de démontrer que $\mathrm{Ind}^{L}_{P\cap L}\big(\big(\big(\boxtimes_i \chi_{t_i,s_i,a_i}\big) \boxtimes \tau_0\big) \otimes \mathbb{C}_{-\delta(\mathfrak{u})}\big)$ est irréductible. Or

$$\mathrm{Ind}^{L}_{P\cap L}\bigg(\bigg(\Big(\boxtimes_{i} \chi_{t_i,s_i,a_i}\Big)\boxtimes\tau_0\bigg)\otimes\mathbb{C}_{-\delta(\mathfrak{u})}\bigg) = \Bigg(\mathrm{Ind}^{\boldsymbol{U}(N'_\rho,N'_\rho)}_{P'}\bigg(\Big(\boxtimes_{i} \chi_{t_i,s_i,a_i}\Big)\otimes\mathbb{C}_{-\delta(\mathfrak{u})}\bigg)\Bigg)\boxtimes\tau_0$$

Ici P' est un sous-groupe parabolique de $\boldsymbol{U}(N'_\rho, N'_\rho)$ de facteur de Levi isomorphe à $\prod_i \mathbf{GL}_{a_i}(\mathbb{C})$ (ce dernier se plonge naturellement dans $\mathbf{GL}_{N_\rho}(\mathbb{C})$ et l'on voit $\mathbf{GL}_{N_\rho}(\mathbb{C})$ comme le Levi du parabolique de Siegel). On est ramené à montrer l'irréductibilité de $\mathrm{Ind}^{\boldsymbol{U}(N'_\rho,N'_\rho)}_{P'}\big(\big(\boxtimes_i \chi_{t_i,s_i,a_i}\big) \otimes \mathbb{C}_{-\delta(\mathfrak{u})}\big)$. Ceci découle de la dernière assertion du théorème 5.3 car l'hypothèse de mauvaise parité pour le groupe unitaire $\boldsymbol{U}(N'_\rho, N'_\rho)$ est vérifiée, comme on le montre ci-dessous.

On calcule le caractère de torsion $\mathbb{C}_{-\delta(\mathfrak{u})}$. Posons $\epsilon_G = 0$ si G est un groupe orthogonal pair, $\epsilon_G = 1$ si G est un groupe symplectique, et $\epsilon_G = \frac{1}{2}$ si G est un groupe orthogonal impair. Dans le système de coordonnées usuelles pour G, on a

$$\delta(\mathfrak{u}) = \big(\underbrace{N - N'_\rho - \tfrac{1}{2} + \epsilon_G, \ldots, N - N'_\rho - \tfrac{1}{2} + \epsilon_G}_{N_\rho}, \underbrace{0, \ldots, 0}_{N_0}\big).$$

L'hypothèse de mauvaise parité des (t_i, s_i, a_i) pour G est : soit $s_i \neq 0$, soit $s_i = 0$ et $(t_i + a_i - 1)/2 + \epsilon_G \in \frac{1}{2}\mathbb{Z} \setminus \mathbb{Z}$. Dans les deux cas, $\big(t_i + N - N'_\rho - \frac{1}{2} + \epsilon_G, s_i, a_i\big)$ est de mauvaise parité pour $\boldsymbol{U}(N'_\rho, N'_\rho)$, car dans le second cas

$$\tfrac{1}{2}(t_i + a_i - 1) + N - N'_\rho - \tfrac{1}{2} + \epsilon_G + \tfrac{1}{2}(2N'_\rho - 1) \in \tfrac{1}{2}\mathbb{Z} \setminus \mathbb{Z}.$$

Ceci termine la deuxième étape où l'on a établi le résultat voulu si la condition que les t_i soient suffisamment grands.

Troisième étape. On ne suppose plus ici que les t_i sont suffisamment grand, mais on choisit T suffisamment grand pour que la représentation π_T obtenue en remplaçant les t_i par t_i+T soit irréductible d'après la deuxième étape. C'est-à-dire que l'on pose

$$\begin{aligned}\pi_T &= \mathrm{Ind}_P^G\left(\left(\boxtimes_i \mathbf{Speh}(\delta_{t_i+T,s_i}, a_i)\right)\boxtimes\tau_0\right)\\ &= \mathrm{Ind}_P^G\left(\left({}^n\mathcal{R}^{\mathfrak{m},M\cap K}_{\mathfrak{q}\cap\mathfrak{m},L\cap M\cap K}\right)^{\dim(\mathfrak{u}\cap\mathfrak{m}\cap\mathfrak{k})}\left(\left(\boxtimes_i \chi_{t_i+T,s_i,a_i}\right)\boxtimes\tau_0\right)\right)\end{aligned}$$

On veut montrer que π s'obtient de π_T par un foncteur de translation [Knapp et Vogan 1995, Chapter 7], c'est-à-dire que l'on obtient π en tensorisant π_T par une représentation de dimension finie F de G, et l'on projette sur la composante de caractère infinitésimal généralisé le caractère infinitésimal de π. On note n_0 le rang de G_0. Le rang de G est donc $n_0+N_\rho = n_0+2N'_\rho$.

On choisit une sous-algèbre de Cartan $\mathfrak{h}$ de $\mathfrak{g}$ et un système de racines positives $\Delta^+(\mathfrak{g},\mathfrak{h})$ de $\mathfrak{h}$ dans $\mathfrak{g}$. On considère la représentation de dimension finie F, de G de plus haut poids

$$\big(\underbrace{T/2,\dots,T/2}_{2N'_\rho},\underbrace{0,\dots,0}_{n_0}\big).$$

On remarque que cette représentation admet le poids extrémal

$$\mu_0 := \big(\underbrace{-T/2,\dots,-T/2}_{N'_\rho},\underbrace{T/2,\dots,T/2}_{N'_\rho},\underbrace{0,\dots,0}_{n_0}\big).$$

On note $\mathfrak{p}'=\mathfrak{m}'\oplus\mathfrak{n}'$ la sous-algèbre parabolique de $\mathfrak{g}$ qui stabilise le sous-espace poids μ_0 de F (qui est de dimension 1 car μ_0 est extrémal). Son facteur de Levi $\mathfrak{m}'$ est isomorphe à $\mathfrak{gl}_{N'_\rho}(\mathbb{C})\times\mathfrak{gl}_{N'_\rho}(\mathbb{C})\times\mathfrak{g}_0$. On note $\mathfrak{q}'=\mathfrak{l}'\oplus\mathfrak{u}'$ la sous-algèbre parabolique de $\mathfrak{g}$ contenue dans $\mathfrak{p}'$ dont la sous-algèbre de Levi $\mathfrak{l}'$ est isomorphe à

$$\mathfrak{gl}_{a_1}(\mathbb{C})\times\cdots\times\mathfrak{gl}_{a_\ell}(\mathbb{C})\times\mathfrak{gl}_{a_\ell}(\mathbb{C})\times\cdots\times\mathfrak{gl}_{a_1}(\mathbb{C})\times\mathfrak{g}_0.$$

Ceci ne détermine pas $\mathfrak{q}'$: on a $\mathfrak{l}'\subset\mathfrak{m}'$, $\mathfrak{n}'\subset\mathfrak{u}'$, $\mathfrak{u}'=\mathfrak{n}'\oplus(\mathfrak{m}'\cap\mathfrak{u}')$ et l'on fixe $\mathfrak{m}'\cap\mathfrak{u}'$ en demandant que pour toute racine $\beta\in\Delta(\mathfrak{u}',\mathfrak{h})$,

$$\beta\big(\underbrace{t_1,\dots,t_1}_{a_1},\dots,\underbrace{t_\ell,\dots,t_\ell}_{a_\ell}\underbrace{-t_\ell,\dots,t_\ell}_{a_\ell},\dots,\underbrace{-t_1,\dots,-t_1}_{a_1}\big)\le 0. \tag{5-5}$$

On note $\tilde\lambda$ le caractère infinitésimal de π et $\tilde\lambda_T$ celui de π_T que l'on voit aussi comme éléments de $\mathfrak{h}^*$. On veut montrer que π s'obtient de π_T par le foncteur de translation défini par F, c'est-à-dire que

$$\pi = (\pi_T\otimes F)_{\tilde\lambda}$$

où $(\cdot)_{\tilde\lambda}$ dénote la projection d'un module $\mathfrak{Z}(\mathfrak{g})$-fini sur sa composante primaire pour le caractère infinitésimal généralisé défini par $\tilde\lambda$.

Pour cela, on suit [Vogan 1988, Proposition 4.7] en modifiant convenablement les hypothèses. On commence par montrer un analogue du lemme 4.8 de [Vogan 1988].

Lemme 5.6. *Soit $\mu \in \mathfrak{h}^*$ le plus haut poids d'une composante irréductible de la restriction de F à $\mathfrak{l}'$. On suppose que pour un certain w dans le groupe de Weyl de G, on a*

$$(\tilde{\lambda}_T + \mu) = w \cdot \tilde{\lambda},$$

Alors $\mu = \mu_0$.

Démonstration. On pose $\mathfrak{h}_\mathfrak{l} = \mathfrak{h} \cap [\mathfrak{l}', \mathfrak{l}']$ et l'on note $\mathfrak{z}_\mathfrak{l}$ le centre de $\mathfrak{l}'$. On a alors

$$\mathfrak{h} = \mathfrak{h}_\mathfrak{l} \oplus \mathfrak{z}_\mathfrak{l} \quad \text{et} \quad \mathfrak{h}^* = \mathfrak{h}_\mathfrak{l}^* \oplus \mathfrak{z}_\mathfrak{l}^*.$$

C'est une somme directe orthogonale. Si $\nu \in \mathfrak{h}^*$, on note $\nu = \nu_\mathfrak{l} \oplus \nu_\mathfrak{z}$ sa décomposition selon cette somme orthogonale. Le poids μ s'écrit $\mu_0 + \sum_{\beta \in \Delta(\mathfrak{n}', \mathfrak{h})} m_\beta \beta$, avec les coefficients m_β entiers négatifs. On a donc

$$(\tilde{\lambda}_T + \mu)_\mathfrak{z} = (\tilde{\lambda}_T + \mu_0)_\mathfrak{z} + \sum_{\beta \in \Delta(\mathfrak{n}', \mathfrak{h})} m_\beta \beta_{|\mathfrak{z}_\mathfrak{l}}$$

Dans le système de coordonnées choisi, $\tilde{\lambda}_T$ s'écrit

$$\frac{1}{2}\Big(\underbrace{t_1+T+(a_1-1), \dots, t_1+T-(a_1-1)}_{a_1}, \dots, \underbrace{t_\ell+T+(a_\ell-1), \dots, t_\ell+T+(a_\ell-1)}_{a_\ell},$$
$$\underbrace{-t_\ell-T+(a_\ell-1), \dots, -t_\ell-T-(a_\ell-1)}_{a_\ell},$$
$$\underbrace{-t_1-T+(a_1-1), \dots, -t_1-T-(a_1-1)}_{a_1}, *, \dots, *\Big)$$

où sur les dernière coordonnées, il apparaît le caractère infinitésimal de τ_0 que nous n'explicitons pas. Ainsi $\tilde{\lambda}_T + \mu_0$ s'écrit

$$\frac{1}{2}\Big(\underbrace{t_1+(a_1-1), \dots, t_1-(a_1-1)}_{a_1}, \dots, \underbrace{t_\ell+(a_\ell-1), \dots, t_\ell+(a_\ell-1)}_{a_\ell},$$
$$\underbrace{-t_\ell+(a_\ell-1), \dots, -t_\ell-(a_\ell-1)}_{a_\ell}, \underbrace{-t_1-+(a_1-1), \dots, -t_1-(a_1-1)}_{a_1}, *, \dots, *\Big)$$

et $\tilde{\lambda}_T + \mu_0 = \tilde{\lambda}$, d'où

$$(\tilde{\lambda}_T + \mu)_\mathfrak{z} = \tilde{\lambda}_\mathfrak{z} + \sum_{\beta \in \Delta(\mathfrak{n}', \mathfrak{h})} m_\beta \beta_{|\mathfrak{z}_\mathfrak{l}}$$

et en utilisant (5-5), on en conclut

$$\|(\tilde{\lambda}_T + \mu)_\mathfrak{z}\| \leq \|\tilde{\lambda}_\mathfrak{z}\| \tag{5-6}$$

avec égalité si tous les m_β sont nuls.

D'autre part $\tilde{\lambda}_T + \mu$ et $\tilde{\lambda} = \tilde{\lambda}_T + \mu_0$ ont même projection sur $\mathfrak{h}_\mathfrak{l}$ car $\mu - \mu_0 \in \mathfrak{h}_\mathfrak{z}^*$. Ainsi $\|(\tilde{\lambda}_T + \mu)_\mathfrak{l}\| = \|\tilde{\lambda}_\mathfrak{l}\|$. Or l'hypothèse force $\tilde{\lambda}_T + \mu$ et $\tilde{\lambda}$ à avoir même norme, et il y a donc égalité dans (5-6), d'où $\mu = \mu_0$. □

Pour en déduire le fait que la translation de π_T est bien π, on raisonne comme dans [Vogan 1988], où la proposition 4.7 est déduite du lemme 4.8. Ici, on utilise le fait que pour toute représentation τ de M et pour toute représentation de dimension finie F de G, $\mathrm{Ind}_P^G(\tau) \otimes F = \mathrm{Ind}_P^G(\tau \otimes F_{|M})$. On applique ceci à $\tau = ({}^n\mathcal{R}^{\mathfrak{m},M\cap K}_{\mathfrak{q}\cap\mathfrak{m},L\cap M\cap K})^{\dim(\mathfrak{u}\cap\mathfrak{m}\cap\mathfrak{k})}\big(\big(\boxtimes_i \chi_{t_i+T,s_i,a_i}\big) \boxtimes \tau_0\big)$ et à F comme ci-dessus, et l'on obtient

$$\pi_T \otimes F = \mathrm{Ind}_P^G\Bigg(\big({}^n\mathcal{R}^{\mathfrak{m},M\cap K}_{\mathfrak{q}\cap\mathfrak{m},L\cap M\cap K}\big)^{\dim(\mathfrak{u}\cap\mathfrak{m}\cap\mathfrak{k})}\Bigg(\Bigg(\boxtimes_i \chi_{t_i+T,s_i,a_i}\Bigg) \boxtimes \tau_0\Bigg) \otimes F_{|M}\Bigg).$$

Ensuite, on utilise le fait que l'induction cohomologique a lieu dans le weakly fair range, où il y a annulation des foncteurs d'induction cohomologique en degré différent de $\dim(\mathfrak{u}\cap\mathfrak{m}\cap\mathfrak{k})$, ce qui nous permet de remplacer le foncteur $\big({}^n\mathcal{R}^{\mathfrak{m},M\cap K}_{\mathfrak{q}\cap\mathfrak{m},L\cap M\cap K}\big)^{\dim(\mathfrak{u}\cap\mathfrak{m}\cap\mathfrak{k})}$ par $\mathcal{R} := \sum_i (-1)^i \big({}^n\mathcal{R}^{\mathfrak{m},M\cap K}_{\mathfrak{q}\cap\mathfrak{m},L\cap M\cap K}\big)^{\dim(\mathfrak{u}\cap\mathfrak{m}\cap\mathfrak{k})-i}$ (on obtient alors une égalité de représentations virtuelles). Ceci nous permet d'utiliser [Vogan 1981, Lemma 7.2.9(b)], et l'on a alors

$$\begin{aligned}\pi_T \otimes F &= \mathrm{Ind}_P^G\Bigg(\mathcal{R}\Bigg(\Bigg(\boxtimes_i \chi_{t_i+T,s_i,a_i}\Bigg) \boxtimes \tau_0\Bigg) \otimes F_{|M}\Bigg)\\ &= \mathrm{Ind}_P^G\Bigg(\mathcal{R}\Bigg(\Bigg(\Bigg(\boxtimes_i \chi_{t_i+T,s_i,a_i}\Bigg) \boxtimes \tau_0\Bigg) \otimes F_{|M\cap L}\Bigg)\Bigg).\end{aligned}$$

On remarque alors, avec les notations employées, que l'on a en fait $\mathfrak{l}' = \mathfrak{l} \cap \mathfrak{m}$ et $\mathfrak{q}' = \mathfrak{l}' \oplus \mathfrak{n}$. On conclut en remarquant que les contributions à la projection de $\pi_T \otimes F$ sur la composante primaire $\tilde{\lambda}$ proviennent des composantes de la restriction de F à $\mathfrak{l}'$ de plus haut poids μ vérifiant l'hypothèse du lemme, et que ceci donne alors le résultat voulu.

Quatrième étape. La translation préserve l'irréductibilité si l'image de l'algèbre enveloppante dans l'algèbre des endomorphismes G-finis de la représentation induite du caractère de $\mathfrak{p}$ est surjective. Un critère pour cela est que l'orbite de Richardson du parabolique P soit de fermeture normale et que l'application moment soit birationnelle. Les orbites décrites par Barbasch [1989, Proposition 14.5] vérifient ces critères. La description de Barbasch en termes d'induite de Richardson s'applique directement pour nous : dans le cas des groupes symplectiques, il suffit que G_0 soit trivial, et pour les groupes orthogonaux à l'inverse, il suffit que le rang de G_0 soit grand par rapport aux a_i, et que la représentation τ_0 soit de dimension finie. Comme on suppose τ_0 unitaire, si τ_0 n'est pas de dimension 1, cela veut dire que G_0 est un groupe compact, mais nous n'allons utiliser que le cas où τ_0 est la représentation triviale. Ainsi, on obtient l'irréductibilité de π à condition que τ_0 soit la représentation triviale du groupe trivial si G est symplectique, et la représentation triviale d'un groupe G_0 de rang suffisamment grand si G est orthogonal.

Cinquième étape. On revient à l'énoncé général de la proposition en utilisant toutefois la réduction effectuée dans la première étape. On a donc

$$\pi = \mathrm{Ind}_P^G\Bigg(\Bigg(\boxtimes_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Bigg) \boxtimes \tau_0\Bigg).$$

L'idée est de construire un opérateur d'entrelacement de π dans un module standard en position de Langlands négative, l'irréductibilité de π étant alors conséquence de l'injectivité de cet opérateur : rappelons que l'on a supposé τ_0 unitaire, et donc π est de longueur finie et unitaire, donc semi-simple, et bien sûr, le module standard en position de Langlands négative admet un unique sous-module irréductible.

Pour cela, nous allons avoir besoin de quelques considérations préliminaires sur les représentations induites des groupes généraux linéaires. Utilisons les notations usuelles pour les induites depuis les sous-groupes paraboliques standard dans les groupes généraux linéaires et classiques. Comme précédemment, notons $\delta_{t,s}$ la représentation de $\mathbf{GL}_2(\mathbb{R})$ essentiellement de carré intégrable de caractère infinitésimal $((-t+s)/2, (t+s)/2)$, où $t \in \mathbb{Z}_{>0}$ et $s \in \mathbb{C}$ (jusque là, nous n'avions introduit que les séries discrètes, c'est-à-dire les $\delta_{t,s}$ avec $s \in i\mathbb{R}$), et notons $\eta_{\epsilon,s}$ le caractère de $\mathbf{GL}_1(\mathbb{R})$ défini par $x \mapsto \mathbf{sgn}(x)^{(1-\epsilon)/2}|x|^s$, $\epsilon \in \pm 1$, $s \in \mathbb{C}$, de caractère infinitésimal s. Si δ est l'une de ces représentations $\delta_{t,s}$ ou $\eta_{\epsilon,s}$ de $\mathbf{GL}_2(\mathbb{R})$ ou $\mathbf{GL}_1(\mathbb{R})$, notons $\mathbb{Z}(\delta)$ l'ensemble $(t+s)/2+\mathbb{Z}$ dans le premier cas, et $s+\mathbb{Z}$ dans le second. D'après un résultat de B. Speh [1981], si δ_1 et δ_2 sont deux représentations de cette forme, alors la représentation induite $\delta_1 \times \delta_2$ est irréductible si $\mathbb{Z}(\delta_1) \neq \mathbb{Z}(\delta_2)$. En particulier, comme dans le groupe de Grothendieck on a $\delta_1 \times \delta_2 = \delta_2 \times \delta_1$ en toute généralité, on voit dans ce cas que $\delta_1 \times \delta_2$ et $\delta_2 \times \delta_1$ sont isomorphes. On a des familles d'opérateurs d'entrelacements

$$M(y_1, y_2) : \delta_1|\cdot|^{y_1} \times \delta_2|\cdot|^{y_2} \to \delta_2|\cdot|^{y_2} \times \delta_1|\cdot|^{y_1}$$

et

$$N(y_1, y_2) : \delta_2|\cdot|^{y_2} \times \delta_1|\cdot|^{y_1} \to \delta_1|\cdot|^{y_1} \times \delta_2|\cdot|^{y_2}$$

méromorphes en $(y_1, y_2) \in \mathbb{C}^2$. Les compositions

$$M(y_1, y_2) \times N(y_1, y_2) \quad \text{et} \quad N(y_1, y_2) \times N(y_1, y_2)$$

sont des opérateurs scalaires donnés par une même fonction méromorphe $\eta(y_1, y_2)$ à valeurs complexes. Cette fonction n'a pas de pôle en $(0, 0)$ (à cause de la condition sur $\mathbb{Z}(\delta_1)$ et $\mathbb{Z}(\delta_2)$) et l'un des opérateurs $M(y_1, y_2)$ ou $N(y_1, y_2)$ est holomorphe en $(0, 0)$ (celui pour lequel le terme de gauche de la flèche est en position de Langlands positive, et le terme de droite en position négative). Les opérateurs d'entrelacement

$M(0,0)$ et $N(0,0)$ sont donc définis et réalisent l'isomorphisme entre $\delta_1 \times \delta_2$ et $\delta_2 \times \delta_1$.

On tire de ceci le résultat suivant

Proposition 5.7. *Soient τ_1 et τ_2 des représentations irréductibles de $\mathbf{GL}_{n_1}(\mathbb{R})$ et $\mathbf{GL}_{n_2}(\mathbb{R})$ respectivement. Supposons que pour un $x_1 \in \mathbb{C}$, le caractère infinitésimal de τ_1 soit formé de nombres tous dans $x+\mathbb{Z}$, et supposons qu'aucune composante du caractère infinitésimal de τ_2 ne soit dans $x+\mathbb{Z}$. Alors les représentations induites $\tau_1 \times \tau_2$ et $\tau_2 \times \tau_1$ de $\mathbf{GL}_{n_1+n_2}(\mathbb{R})$ sont isomorphes.*

Démonstration. On réalise τ_1 et τ_2 comme sous-module de représentations standards en position de Langlands négative :

$$\tau_1 \hookrightarrow \delta_{1,1} \times \cdots \times \delta_{1,r_1}, \quad \tau_2 \hookrightarrow \delta_{2,1} \times \cdots \times \delta_{2,r_2}.$$

On a une famille méromorphe d'opérateurs d'entrelacement

$$\mathcal{M}(y_1, y_2) : \delta_{1,1}|\cdot|^{y_1} \times \cdots \times \delta_{1,r_1}|\cdot|^{y_1} \times \delta_{2,1}|\cdot|^{y_2} \times \cdots \times \delta_{2,r_2}|\cdot|^{y_2}$$
$$\to \delta_{2,1}|\cdot|^{y_2} \times \cdots \times \delta_{2,r_2}|\cdot|^{y_2} \times \delta_{1,1}|\cdot|^{y_1} \times \cdots \times \delta_{1,r_1}|\cdot|^{y_1}$$

qui se factorise en un produit de composition d'opérateurs d'entrelacement élémentaires de la forme considérée avant l'énoncé de la proposition et qui sont tous holomorphes bijectifs en $(0,0)$. L'opérateur $\mathcal{M}(y_1, y_2)$ est donc holomorphe bijectif en $(0,0)$. □

Revenons maintenant à notre but principal dans cette cinquième étape.

Pour tout indice i, la représentation de $\mathbf{Speh}(\delta_{t_i,s_i}, a_i)$ est réalisée comme unique sous-module de la représentation standard :

$$S_{t_i,s_i,a_i} = \delta_{t_i,s_i}|\cdot|^{-(a_i-1)/2} \times \delta_{t_i,s_i}|\cdot|^{-(a_i-3)/2} \times \ldots \times \delta_{t_i,s_i}|\cdot|^{(a_i-3)/2} \times \delta_{t_i,s_i}|\cdot|^{(a_i-1)/2}.$$

Réalisons aussi la représentation τ_0 comme sous-module de Langlands d'une représentation standard $\tilde{\tau}$ de G_0 en position négative. Écrivons $\tilde{\tau} = \tilde{\tau}_{--} \times \tilde{\tau}_{\text{temp}}$, où $\tilde{\tau}_{\text{temp}}$ est tempérée irréductible, et τ_{--} est en position de Langlands strictement négative (ici $\tilde{\tau}_{--}$ est donc une représentation d'un produit de $\mathbf{GL}_1(\mathbb{R})$ et $\mathbf{GL}_2(\mathbb{R})$, et $\tilde{\tau}_{\text{temp}}$ une représentation d'un produit de $\mathbf{GL}_1(\mathbb{R})$ et $\mathbf{GL}_2(\mathbb{R})$ et d'un groupe classique, le produit de ces deux facteurs formant un sous-groupe de Levi standard de G_0). On obtient donc un plongement

$$\pi \hookrightarrow \left(\prod_i S_{t_i,s_i,a_i}\right) \times \tilde{\tau}_{--} \times \tilde{\tau}_{\text{temp}}. \tag{5-7}$$

Formons maintenant une représentation standard en position négative de la manière suivante. La représentation $\prod_i S_{t_i,s_i,a_i} \times \tilde{\tau}_{--}$ s'écrit comme un produit de représentations de la forme $\delta|\cdot|^x$ où δ est une série discrète de $\mathbf{GL}_1(\mathbb{R})$ ou $\mathbf{GL}_2(\mathbb{R})$ et x est un demi-entier. Remplaçons dans ce produit les termes comme ci-dessus avec $x > 0$ par $\delta|\cdot|^{-x}$ et réordonnons les facteurs pour les mettre dans l'ordre des x croissants. Écrivons le produit obtenu comme étant $\Delta_{--} \times \Delta_{\text{temp}}$ où Δ_{--} est le produit des facteurs $\delta|\cdot|^{x}$ avec $x < 0$ et Δ_{temp} celui avec $x = 0$. Notons $\mathcal{N}$ l'opérateur d'entrelacement standard pour le groupe G qui envoie $\left(\prod_i S_{t_i,s_i,a_i}\right) \times \tau_{--} \times \tilde{\tau}_{\text{temp}}$ dans $\Delta_{--} \times \Delta_{\text{temp}} \times \tilde{\tau}_{\text{temp}}$. L'opérateur d'entrelacement $\mathcal{N}$ se factorise en opérateurs élémentaires, l'effet d'un opérateur élémentaire étant de remplacer un facteur $\delta|\cdot|^{x}$ avec $x > 0$ par $\delta|\cdot|^{-x}$ ou bien un produit de la forme $\delta_1|\cdot|^{x_1} \times \delta_2|\cdot|^{x_2}$ avec $0 \geq x_1 > x_2$ par $\delta_2|\cdot|^{x_2} \times \delta_1|\cdot|^{x_1}$ et ceux-ci sont bien définis (holomorphes) dans le domaine où on les considère. Notons encore $\mathcal{N}$ la composition de $\mathcal{N}$ avec (5-7) :

$$\mathcal{N} : \pi \hookrightarrow \left(\prod_i S_{t_i,s_i,a_i}\right) \times \tilde{\tau}_{--} \times \tilde{\tau}_{\text{temp}} \to \Delta_{--} \times \Delta_{\text{temp}} \times \tilde{\tau}_{\text{temp}}. \tag{5-8}$$

La représentation $\Delta_{\text{temp}} \times \tilde{\tau}_{\text{temp}}$ est une représentation tempérée d'un groupe classique, induite d'une tempérée irréductible. La théorie du R-groupe et l'hypothèse sur les parités de Δ_{temp} (mauvaise) et de $\tilde{\tau}_{\text{temp}}$ (bonne) nous dit que cette représentation est irréductible. D'autre part, Δ_{--} est en position de Langlands strictement négative. Le terme de droite est donc une représentation standard en position de Langlands négative, qui admet un unique sous-module irréductible.

Ainsi, on a bien construit un opérateur d'entrelacement de π vers un module standard en position de Langlands négative, et il reste à montrer son injectivité. Faisons tout d'abord quelques observations sur Δ_{--} et Δ_{temp}. La première est formée à partir de facteurs $\delta|\cdot|^{x}$ venant soit des S_{t_i,s_i,a_i}, soit de $\tilde{\tau}_{--}$. Mais un facteur $\delta_1|\cdot|^{x_1}$ venant d'un S_{t_i,s_i,a_i} et un facteur $\delta_1|\cdot|^{x_1}$ venant de $\tilde{\tau}_{--}$ commutent, car leur produit est irréductible d'après le résultat de Speh et les hypothèses sur les parités. On a donc en fait $\Delta_{--} = \Delta_{--}^{\text{Speh}} \times \tilde{\tau}_{--} = \tilde{\tau}_{--} \times \Delta_{--}^{\text{Speh}}$ où $\Delta_{--}^{\text{Speh}}$ est obtenue comme ci-dessus en changeant des exposants en leurs opposés et en remettant le tout dans l'ordre, mais seulement pour les facteurs provenant des représentations de Speh. Le terme Δ_{temp} est lui un produit de facteurs provenant des représentations de Speh. On peut donc noter $\Delta_{\text{temp}} = \Delta_{\text{temp}}^{\text{Speh}}$ pour insister sur ce fait. D'autre part, il commute avec $\tilde{\tau}_{--}$ en vertu du résultat de Speh invoqué ci-dessus. Ainsi (5-8) peut aussi s'écrire

$$\mathcal{N} : \pi \to \Delta_{--}^{\text{Speh}} \times \Delta_{\text{temp}}^{\text{Speh}} \times \tilde{\tau}_{--} \times \tilde{\tau}_{\text{temp}}. \tag{5-9}$$

Montrons maintenant que pour l'injectivité de $\mathcal{N}$, on se ramène au cas où τ_0 est une représentation d'un groupe compact. En effet, supposons que τ_0 soit sous-module d'une série principale $\left(\prod_j \gamma_j\right) \times \tau_0'$ où les γ_j sont des caractères de $\mathbf{GL}_1(\mathbb{R})$ et τ_0' est une représentation d'un groupe compact G_0' de même type que G (G_0' est la partie compacte du facteur de Levi d'un parabolique minimal de G_0). On a donc

$$\pi = \Bigl(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Bigr) \times \tau_0 \hookrightarrow \Bigl(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Bigr) \times \Bigl(\prod_j \gamma_j\Bigr) \times \tau'_0. \tag{5-10}$$

Grâce à la proposition ci-dessus,

$$\Bigl(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Bigr) \times \Bigl(\prod_j \gamma_j\Bigr) = \Bigl(\prod_j \gamma_j\Bigr) \times \Bigl(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Bigr)$$

et l'on peut réécrire le terme de droite en permutant les facteurs. On obtient donc un plongement

$$\begin{aligned}\pi = \Bigl(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Bigr) \times \tau_0 \\ \hookrightarrow \Bigl(\prod_j \gamma_j\Bigr) \times \Bigl(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Bigr) \times \tau'_0 = \Bigl(\prod_j \gamma_j\Bigr) \times \pi'.\end{aligned} \tag{5-11}$$

Admettons le résultat pour τ'_0, à savoir que l'opérateur d'entrelacement

$$\mathcal{N}' : \pi' = \Bigl(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Bigr) \times \tau'_0 \hookrightarrow \Delta'_{--} \times \Delta'_{\mathrm{temp}} \times \tilde{\tau}'_{\mathrm{temp}}$$

construit comme ci-dessus en partant de π' plutôt que de π, et avec les notations évidentes, est injectif. Remarquons que comme τ'_0 est une représentation irréductible d'un groupe compact, on a avec ces notations $\tilde{\tau}' = \tilde{\tau}'_{\mathrm{temp}} = \tau'_0$ et $\tilde{\tau}'_{--}$ est triviale et en particulier $\Delta'_{--} = \Delta^{\mathrm{Speh}}_{--}$ D'autre part $\Delta'_{\mathrm{temp}} = \Delta^{\mathrm{Speh}}_{\mathrm{temp}} = \Delta_{\mathrm{temp}}$. On a donc un opérateur injectif

$$\mathcal{N}' : \pi' = \Bigl(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\Bigr) \times \tau'_0 \hookrightarrow \Delta^{\mathrm{Speh}}_{--} \times \Delta^{\mathrm{Speh}}_{\mathrm{temp}} \times \tau'_0.$$

Par exactitude du foncteur d'induction parabolique, on en déduit un opérateur injectif,

$$\mathcal{N}' : \Bigl(\prod_j \gamma_j\Bigr) \times \pi' \hookrightarrow \Bigl(\prod_j \gamma_j\Bigr) \times \Delta^{\mathrm{Speh}}_{--} \times \Delta^{\mathrm{Speh}}_{\mathrm{temp}} \times \tau'_0.$$

On utilise à nouveau la proposition 5.7 pour écrire le terme de droite sous la forme $\Delta^{\mathrm{Speh}}_{--} \times \Delta^{\mathrm{Speh}}_{\mathrm{temp}} \times \bigl(\prod_j \gamma_j\bigr) \times \tau'_0$, et on compose avec le plongement (5-11) pour obtenir un morphisme injectif que l'on note encore $\mathcal{N}'$:

$$\mathcal{N}' : \pi \hookrightarrow \Delta^{\mathrm{Speh}}_{--} \times \Delta^{\mathrm{Speh}}_{\mathrm{temp}} \times \Bigl(\prod_j \gamma_j\Bigr) \times \tau'_0.$$

L'injection $\tau_0 \hookrightarrow \bigl(\prod_j \gamma_j\bigr) \times \tau'_0$ est obtenue en composant l'injection $\tau_0 \hookrightarrow \tilde{\tau}_{--} \times \tilde{\tau}_{\mathrm{temp}}$ et un morphisme $\tilde{\tau}_{--} \times \tilde{\tau}_{\mathrm{temp}} \to \bigl(\prod_j \gamma_j\bigr) \times \tau'_0$. Ainsi l'on voit que $\mathcal{N}'$ se factorise par $\mathcal{N}$. Ceci établit le fait que l'injectivité de $\mathcal{N}'$ implique celle de $\mathcal{N}$.

Pour les groupes symplectiques, qui sont déployés, τ_0' est la représentation triviale du groupe trivial. L'injectivité de $\mathcal{N}'$ provient alors de l'irréductibilité de π' établie à la quatrième étape, et du fait que $\mathcal{N}'$ est non nul.

Les groupes orthogonaux demandent encore un peu de travail pour conclure comme ci-dessus car G_0' peut être de rang trop petit pour pouvoir appliquer l'irréductibilité démontrée à la quatrième étape. On utilise les foncteurs de translation pour se ramener au cas où le caractère infinitésimal de τ_0' est celui de la représentation triviale, c'est-à-dire que τ_0' est la représentation triviale $\mathbf{Triv}_{G_0'}$ de G_0'.

Soit σ un facteur irréductible de $\pi' = \left(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\right) \times \mathbf{Triv}_{G_0'}$ (rappelons que π' est unitaire et donc semi-simple). Notons n_0 le plus grand entier entrant dans le caractère infinitésimal de $\mathbf{Triv}_{G_0'}$. Soit T un entier suffisamment grand. Notons G_0'' le groupe de même type que G_0' et de rang $T + \mathrm{rg}(G_0')$ tel que $\mathbf{GL}_1(\mathbb{R})^T \times G_0''$ soit un sous-groupe de Levi de G_0', et soit $\mathbf{Triv}_{G_0''}$ la représentation triviale de ce groupe. On suppose donc T assez grand pour que les hypothèses de la quatrième étape soient vérifiées pour $\pi'' = \left(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\right) \times \mathbf{Triv}_{G_0''}$ qui est donc irréductible. La représentation induite

$$|\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times \sigma \tag{5-12}$$

possède un seul quotient irréductible car elle est quotient de la représentation

$$|\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times S(\sigma)$$

où $S(\sigma)$ est une représentation standard en position de Langlands positive et dont σ est le quotient de Langlands. Grâce à la proposition 5.7, on peut mettre (5-12) en position de Langlands positive, et elle admet donc un unique quotient irréductible. Ce quotient irréductible est isomorphe à l'image de l'opérateur d'entrelacement standard

$$|\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times \sigma \to |\cdot|^{-n_0-T} \times |\cdot|^{-n_0-T+1} \times \cdots \times |\cdot|^{-n_0-1} \times \sigma$$

Donc les quotients irréductibles de

$$|\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times \pi' \tag{5-13}$$

sont isomorphes aux sous-modules irréductibles de l'image de l'opérateur d'entrelacement standard

$$\begin{multline} |\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times \pi' \\ \to |\cdot|^{-n_0-T} \times |\cdot|^{-n_0-T-1} \times \cdots \times |\cdot|^{-n_0-1} \times \pi' \end{multline} \tag{5-14}$$

et il y a bijection entre ces sous-modules irréductibles de l'image (et cette image est semi-simple) et les sous-modules irréductibles de π'.

L'opérateur d'entrelacement (5-14) se réécrit en utilisant la proposition 5.7 ci-dessus

$$\left(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\right) \times |\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times \mathbf{Triv}_{G_0'} \\ \to \left(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\right) \times |\cdot|^{-n_0-T} \times |\cdot|^{-n_0-T+1} \times \cdots \times |\cdot|^{-n_0-1} \times \mathbf{Triv}_{G_0'} \tag{5-15}$$

Or $|\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times \mathbf{Triv}_{G_0'}$, (resp. $|\cdot|^{-n_0-T} \times |\cdot|^{-n_0-T+1} \times \cdots \times |\cdot|^{-n_0-1} \times \mathbf{Triv}_{G_0'}$) est la représentation standard en position de Langlands positive (resp. négative) dont $\mathbf{Triv}_{G_0''}$ est l'unique quotient irréductible (resp. sous-module). Ainsi $\pi'' = \left(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\right) \times \mathbf{Triv}_{G_0''}$ apparaît comme image de (5-15), et cette image est irréductible. Mais (5-15) s'écrit comme la composition de l'isomorphisme

$$\begin{aligned} &\left(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\right) \times |\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times \mathbf{Triv}_{G_0'} \\ &\qquad \simeq |\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times \left(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\right) \times \mathbf{Triv}_{G_0'} \\ &\qquad = |\cdot|^{n_0+T} \times |\cdot|^{n_0+T-1} \times \cdots \times |\cdot|^{n_0+1} \times \pi', \end{aligned}$$

de (5-14), et de l'isomorphisme

$$\begin{aligned} &|\cdot|^{n_0+T} \times |\cdot|^{-n_0-T+1} \times \cdots \times |\cdot|^{-n_0-1} \times \pi' \\ &\qquad = |\cdot|^{-n_0-T} \times |\cdot|^{-n_0-T+1} \times \cdots \times |\cdot|^{-n_0-1} \times \left(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\right) \times \mathbf{Triv}_{G_0'} \\ &\qquad \simeq \left(\prod_i \mathbf{Speh}(\delta_{t_i,s_i}, a_i)\right) \times |\cdot|^{-n_0-T} \times |\cdot|^{-n_0-T+1} \times \cdots \times |\cdot|^{-n_0-1} \times \mathbf{Triv}_{G_0'} \end{aligned}$$

Ainsi l'image de (5-15) est d'une part irréductible et d'autre part a autant de composantes irréductibles que π' d'après la remarque faite après (5-14). Ceci montre que π' est irréductible.

Ceci termine la démonstration de la proposition 5.5. □

Remerciements

Nous remercions P. Trapa qui nous a indiqué que l'on pouvait déduire le résultat de multiplicité un du théorème 1.1 d'une lecture attentive des résultats de la littérature sur les propriétés des foncteurs de translation. Le deuxième auteur a bénéficié d'une aide de l'agence nationale de la recherche ANR-13-BS01-0012 FERPLAY.

Bibliographie

[Adams et Johnson 1987] J. Adams et J. F. Johnson, "Endoscopic groups and packets of non-tempered representations", *Compositio Math.* **64**:3 (1987), 271–309. MR Zbl

[Adams et al. 1992] J. Adams, D. Barbasch et D. A. Vogan, Jr., *The Langlands classification and irreducible characters for real reductive groups*, Progr. Math. **104**, Birkhäuser, Boston, 1992. MR Zbl

[Arancibia et al. 2018] N. Arancibia, C. Mœglin et D. Renard, "Paquets d'Arthur des groupes classiques et unitaires", *Ann. Fac. Sci. Toulouse Sér. 6* **27**:5 (2018), 1023–1105.

[Arthur 1984] J. Arthur, "On some problems suggested by the trace formula", pp. 1–49 dans *Lie group representations, II* (College Park, MD, 1982/1983), édité par R. Herb et al., Lecture Notes in Math. **1041**, Springer, 1984. MR Zbl

[Arthur 1989] J. Arthur, "Unipotent automorphic representations: conjectures", pp. 13–71 dans *Orbites unipotentes et représentations, II*, Astérisque **171-172**, Soc. Math. France, Paris, 1989. MR Zbl

[Arthur 2013] J. Arthur, *The endoscopic classification of representations: orthogonal and symplectic groups*, Amer. Math. Soc. Colloq. Publ. **61**, Amer. Math. Soc., Providence, RI, 2013. MR Zbl

[Barbasch 1989] D. Barbasch, "The unitary dual for complex classical Lie groups", *Invent. Math.* **96**:1 (1989), 103–176. MR Zbl

[Barbasch et Vogan 1983] D. Barbasch et D. Vogan, "Weyl group representations and nilpotent orbits", pp. 21–33 dans *Representation theory of reductive groups* (Park City, UT, 1982), édité par P. C. Trombi, Progr. Math. **40**, Birkhäuser, Boston, 1983. MR Zbl

[Johnson 1984] J. F. Johnson, "Lie algebra cohomology and the resolution of certain Harish-Chandra modules", *Math. Ann.* **267**:3 (1984), 377–393. MR Zbl

[Kaletha et al. 2014] T. Kaletha, A. Minguez, S. W. Shin et P.-J. White, "Endoscopic classification of representations : inner forms of unitary groups", preprint, 2014. arXiv

[Knapp et Vogan 1995] A. W. Knapp et D. A. Vogan, Jr., *Cohomological induction and unitary representations*, Princeton Math. Series **45**, Princeton Univ. Press, 1995. MR Zbl

[Kraft et Procesi 1979] H. Kraft et C. Procesi, "Closures of conjugacy classes of matrices are normal", *Invent. Math.* **53**:3 (1979), 227–247. MR Zbl

[Matumoto 1996] H. Matumoto, "On the representations of $\mathrm{U}(m, n)$ unitarily induced from derived functor modules", *Compositio Math.* **100**:1 (1996), 1–39. MR Zbl

[Matumoto 2004] H. Matumoto, "On the representations of $\mathrm{Sp}(p, q)$ and $\mathrm{SO}^*(2n)$ unitarily induced from derived functor modules", *Compositio Math.* **140**:4 (2004), 1059–1096. MR Zbl

[Mœglin 2017] C. Mœglin, "Paquets d'Arthur spéciaux unipotents aux places archimédiennes et correspondance de Howe", pp. 469–502 dans *Representation theory, number theory, and invariant theory* (New Haven, CT, 2015), édité par J. Cogdell et al., Progr. Math. **323**, Birkhäuser, Cham, 2017. MR

[Mœglin et Renard 2017] C. Mœglin et D. Renard, "Sur les paquets d'Arthur des groupes classiques réels", 2017. To appear in *J. Eur. Math. Soc.* arXiv

[Mœglin et Renard 2018a] C. Mœglin et D. Renard, "Sur les paquets d'Arthur aux places réelles, translation", pp. 299–320 dans *Geometric aspects of the trace formula*, édité par W. Müller et al., Springer, 2018.

[Mœglin et Renard 2018b] C. Mœglin et D. Renard, "Sur les paquets d'Arthur des groupes classiques et unitaires non quasi-déployés", pp. 341–361 dans *Relative aspects in representation theory, Langlands functoriality and automorphic forms*, édité par V. Heiermann et D. Prasad, Lecture Notes in Math. **2221**, Springer, 2018.

[Mok 2015] C. P. Mok, *Endoscopic classification of representations of quasi-split unitary groups*, Mem. Amer. Math. Soc. **1108**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Renard 2010] D. Renard, *Représentations des groupes réductifs p-adiques*, Cours Spécialisés **17**, Soc. Math. France, Paris, 2010. MR Zbl

[Shelstad 2015] D. Shelstad, "On elliptic factors in real endoscopic transfer, I", pp. 455–504 dans *Representations of reductive groups*, édité par M. Nevins et P. Trapa, Progr. Math. **312**, Birkhäuser, Cham, 2015. MR Zbl

[Speh 1981] B. Speh, "The unitary dual of $\mathrm{Gl}(3, \mathbb{R})$ and $\mathrm{Gl}(4, \mathbb{R})$", *Math. Ann.* **258**:2 (1981), 113–133. MR Zbl

[Trapa 2001] P. E. Trapa, "Annihilators and associated varieties of $A_{\mathfrak{q}}(\lambda)$ modules for $\mathrm{U}(p, q)$", *Compositio Math.* **129**:1 (2001), 1–45. MR Zbl

[Vogan 1981] D. A. Vogan, Jr., *Representations of real reductive Lie groups*, Progr. Math. **15**, Birkhäuser, Boston, 1981. MR Zbl

[Vogan 1988] D. A. Vogan, Jr., "Irreducibility of discrete series representations for semisimple symmetric spaces", pp. 191–221 dans *Representations of Lie groups* (Kyoto/Hiroshima, 1986), édité par K. Okamoto et T. Ōshima, Adv. Stud. Pure Math. **14**, Academic Press, Boston, 1988. MR Zbl

Received May 14, 2018.

COLETTE MŒGLIN
CNRS
INSTITUT MATHÉMATIQUES DE JUSSIEU
PARIS
FRANCE

colette.moeglin@imj-prg.fr

DAVID RENARD
CENTRE DE MATHÉMATIQUES LAURENT SCHWARTZ
ECOLE POLYTECHNIQUE
PALAISEAU
FRANCE

david.renard@polytechnique.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 1, 2019
dx.doi.org/10.2140/pjm.2019.299.89

TOPOLOGY AND DYNAMICS OF THE CONTRACTING BOUNDARY OF COCOMPACT CAT(0) SPACES

DEVIN MURRAY

Let X be a proper CAT(0) space and let G be a cocompact group of isometries of X which acts properly discontinuously. Charney and Sultan constructed a quasi-isometry invariant boundary for proper CAT(0) spaces which they called the contracting boundary. The contracting boundary imitates the Gromov boundary for δ-hyperbolic spaces. We will make this comparison more precise by establishing some well-known results for the Gromov boundary in the case of the contracting boundary. We show that the dynamics on the contracting boundary is very similar to that of a δ-hyperbolic group. In particular the action of G on $\partial_c X$ is minimal if G is not virtually cyclic. We also establish a uniform convergence result that is similar to the π-convergence of Papasoglu and Swenson and as a consequence we obtain a new North-South dynamics result on the contracting boundary. We additionally investigate the topological properties of the contracting boundary and we find necessary and sufficient conditions for G to be δ-hyperbolic. We prove that if the contracting boundary is compact, locally compact or metrizable, then G is δ-hyperbolic.

1. Introduction

The Gromov boundary has been a very useful and powerful tool in understanding the structure of δ-hyperbolic groups. The boundary has a large array of nice topological, metric, and dynamical properties that can be used in probing everything from subgroups and splittings to algorithmic properties. It has also played an important role in proving various rigidity theorems.

For proper CAT(0) spaces, there is a nice visual boundary, but Croke and Kleiner [2000] showed that such a boundary is not a quasi-isometry invariant. They constructed two different CAT(0) spaces with nonhomeomorphic visual boundaries on which the same group acts geometrically. The visual boundary can still be used to study CAT(0) groups, for instance it can detect products [Bridson and Haefliger 1999, II.9.24], but the failure of quasi-isometry invariance is a serious blow. Charney

---

MSC2010: 20F65.

Keywords: geometric group theory, Morse boundary, rank-one isometries, CAT(0) geometry, contracting boundary,

and Sultan [2013] constructed a natural topological space associated to a CAT(0) space, called the contracting boundary, which is a quasi-isometry invariant.

One of the many properties that geodesics have in a δ-hyperbolic space is that there is a uniform bound on the diameter of the projection of a ball onto a geodesic disjoint from it. It turns out that this is a very powerful property and the existence of such geodesics in a space has significant consequences for the geometry [Algom-Kfir 2011; Bestvina and Fujiwara 2009; Hamenstädt 2009]. Such geodesics are called contracting geodesics.

The contracting boundary, $\partial_c X$, of a complete CAT(0) space X is the set of contracting rays in X up to asymptotic equivalence. It is homeomorphic to the Gromov boundary when X is also δ-hyperbolic and is designed to imitate the Gromov boundary for more general CAT(0) spaces. However, the contracting boundary for CAT(0) groups is still not very well understood, so we hope to help lay out the ground work for a program of study to better understand it and its implications for CAT(0) groups.

The rank-rigidity conjecture of Ballmann and Buyalo says that for sufficiently nice CAT(0) spaces, the nonexistence of a periodic contracting axis implies that the space is either a metric product, a symmetric space, or a Euclidean building [Ballmann and Buyalo 2008]. Rank-rigidity theorems have been proven for many different classes of spaces including Hadamard manifolds, CAT(0) cube complexes, and right-angled Artin groups, as well as some others [Ballmann 1985; Caprace and Sageev 2011; Behrstock and Charney 2012; Caprace and Fujiwara 2010]. In light of these results, the study of CAT(0) groups can often be reduced to the study of CAT(0) groups with a contracting axis. Building on the work of Ballmann and Buyalo, we show that a CAT(0) group has a contracting axis if and only if the contracting boundary, $\partial_c X$, is nonempty. Thus the contracting boundary is a promising tool for the study of CAT(0) spaces and groups.

Several of the rigidity theorems for hyperbolic groups can be proven through a careful study of the dynamics of the action of the group on its boundary [Freden 1995; Gabai 1992; Casson and Jungreis 1994]. These rigidity theorems become even more striking when further geometric structures are added, such as the Mostow rigidity of finite dimensional hyperbolic manifolds [Mostow 1968].

Our most promising results have been predominantly dynamical. While the topology of the contracting boundary tends to be rather pathological, many of the dynamical properties of the Gromov boundary are shared by the contracting boundary.

There are two main dynamics results that we obtain in this paper. The first says that the orbit of any contracting ray is dense.

Theorem 4.1. *Let G be a group acting geometrically on a proper CAT(0) space. Either G is virtually $\mathbb{Z}$ or the G orbit of every point in the contracting boundary is dense.*

The second dynamics result concerns a more powerful convergence-group-like property. This is similar to the π-convergence of a CAT(0) group on its visual boundary.

Theorem 4.2. *Let X be a proper $CAT(0)$ space and G a group acting geometrically on X. If g_i is a sequence of elements of G where $g_i x \to \gamma^+$ for some $x \in X$ and $\gamma^+ \in \partial_c X$, then there is a subsequence such that $g_i^{-1} x \to \gamma^-$ for $\gamma^- \in \partial_c X$ and for any open neighborhood U of γ^+ in $\partial_c X$ and any compact K in $\partial_c X - \gamma^-$ there is an n such that $g_i(K) \subseteq U$ for $i \geq n$.*

The normal version of π-convergence introduced by Papasoglu and Swenson [2009] and the North-South dynamics due to Hamenstädt [2009] both deal with the visual topology on the visual boundary. Because the topology on the contracting boundary is not the subspace topology, these theorems don't directly apply.

Both of these results are well known for the action of a hyperbolic group on its boundary. We will discuss them both in greater detail in Section 4.

The topology on the contracting boundary is defined as a direct limit of subspaces, $\partial_c^D X$, consisting of rays with contracting constant bounded by D. The topology is quite fine and is, perhaps, more pathological than one would expect from a bordification. While the subspaces $\partial_c^D X$ are compact and metrizable, we prove that the direct limit, $\partial_c X$, is not always compact (nor locally compact) for CAT(0) groups, though it is known to be σ-compact [Charney and Sultan 2013]. In Section 3 we define the topology and discuss some relevant basic topological facts.

One of the powerful tools that is available when studying the Gromov boundary is the family of metrics on it. In Section 5 we show that a number of topological properties, including the metrizability of the contracting boundary, characterize when the space is δ-hyperbolic.

Theorem 5.1. *Let X be a complete proper $CAT(0)$ space with a geometric group action. Then the following are equivalent*:

(i) *X is δ-hyperbolic.*

(ii) *$\partial_c X$ is compact.*

(iii) *$\partial_c X$ is locally compact.*

(iv) *$\partial_c X$ is metrizable.*

A generalization of the contracting boundary for proper geodesic metric spaces, called the Morse boundary, was introduced by Cordes [2017]. It would be interesting to see if any of these results hold true in that more general setting. In particular, it seems like many of the necessary pieces are already known for an analogue of Theorem 5.1 for the Morse boundary in some restricted cases [Cordes 2017; Fink 2015].

2. Some basics on contracting geodesics

For the entirety of this paper we will assume that a geodesic a is an isometric embedding of $\mathbb{R}$ or a segment of $\mathbb{R}$ into a metric space. For convenience we will often conflate the image of the embedding with the map itself. For the subsequent discussion we may assume that unless stated otherwise all metric spaces are proper and satisfy the CAT(0) inequality.

Notation. Recall that ∂X is the set of infinite geodesics, where two geodesics are considered equivalent if they are within a bounded neighborhood of one another. Throughout we will mostly consider the cone topology on this set; recall that a neighborhood system is described by the set $U_a(\epsilon, r)$ which are all geodesics, with b starting at $a(0)$ such that $d(b, a(r)) < \epsilon$. This is sometimes called the cone topology.

We will adopt the notation convention that $[x, y]$ represents the unique geodesic between the points $x, y \in X$. For a point $x \in X$ and a point $\alpha \in \partial X$ we will use $[x, \alpha)$ to denote the unique geodesic starting from x which is in the equivalence class α. When we write (α, β), we will mean a specific bi-infinite geodesic, c, such that $c|_{(-\infty,0]} \in \alpha$ and $c|_{[0,\infty)} \in \beta$. For a geodesic ray a in X we will use $a(\infty)$ to denote the equivalence class of a in ∂X.

Definition 2.1 (contracting geodesics). A geodesic a is said to be *A-contracting* for some constant A if for all $x, y \in X$,

$$d(x, y) < d(x, \pi_a(x)) \Longrightarrow d(\pi_a(x), \pi_a(y)) < A.$$

Note that this definition is sometimes called *strongly contracting* in the literature. Contracting geodesics can be thought of as detecting hyperbolic "directions" in a CAT(0) space. Another useful, and equivalent, property of hyperbolic-like geodesics is that of δ-slimness. This is much closer to the notion of Gromov hyperbolicity.

Definition 2.2 (slim geodesics). A geodesic a is said to be *δ-slim* if for all $y \notin a$ and all z on a there exists a point w on the geodesic $[y, z]$ such that $d(\pi_a(y), w) \leq \delta$.

It turns out that this property will be much more versatile for our purposes; luckily for us the two notions are equivalent in proper CAT(0) spaces. For a complete proof, see [Charney and Sultan 2013; Bestvina and Fujiwara 2009].

Lemma 2.3. *If a is a contracting geodesic with contracting constant A then a is δ_A-slim for some δ_A which depends only on A. The converse is also true*; *if a is δ-slim then it is $\Phi(\delta)$-contracting where $\Phi(\delta)$ depends linearly on δ.*

We will adopt the convention used in [Bestvina and Fujiwara 2009], that all constants will be denoted by $\Phi(\,\cdot\,)$. Typically the function $\Phi(\,\cdot\,)$ will be linear in its terms. When constants are referenced in later statements the relevant lemma and theorem number will be added as a subscript. Sometimes it will be expedient to drop the terms of Φ if they are clear from context, e.g., Lemma 2.3 says that if a is a δ-slim geodesic it is $\Phi_{2.3}$-contracting.

One of the most important facts about the contracting constant of a geodesic is that it is controlled by the contracting constants of nearby geodesics. This will very important in the sequel as it will allow us to push contracting geodesics around via isometries and give us fine-tuned control on the contracting constants of a target geodesic.

Lemma 2.4. *If we have geodesics* $[a, b]$ *and* $[a', b']$, *where* $[a, b]$ *is* A*-contracting*, $d(a, a') = D$, *and* $d(b, b') = D'$ *then* $[a', b']$ *is* $\Phi(A, D, D')$*-contracting. It suffices to take* $\Phi(A, D, D') = 16A + 28D + 7D' + 10$.

A proof for this is in [Bestvina and Fujiwara 2009, Lemma 3.8]. Though they do not write down the explicit $\Phi(A, D, D')$ in their paper, it is possible to recover the one above from their work.

Another important property of contracting geodesics is that subsegments of a contracting geodesic are contracting. So unless otherwise specified we may assume that if a is A-contracting, all subsegments are also A-contracting.

Lemma 2.5. *If* a *is a contracting ray with contracting constant* A *then a subsegment of it is* $\Phi(A)$*-contracting where* $\Phi(A) = A + 3$.

Bestvina and Fujiwara [2009] prove this in a slightly more general context. To understand why the contracting constant may have to increase, note that there are balls that don't intersect the subsegment but do intersect the original contracting ray. The increase can be thought of as making up for some possible differences in the local geometry of the subsegment compared to the original ray.

As a converse to the previous lemma, sometimes we will need to piece together two contracting geodesics into a longer geodesic. It is an easy warm-up exercise to show that this new geodesic is also contracting.

Lemma 2.6. *Let* a *and* b *be geodesics in a* $CAT(0)$ *space* X. *If* a *is* A*-contracting*, b *is* B*-contracting*, *and* $a(0) = b(0) = z$, *then the following hold*:

(i) *If the concatenation of* a *and* b *is a geodesic*, *then it is* $(A+B)$*-contracting.*

(ii) *For every point* $x \in a$ *and* $y \in b$, *the geodesic* $[x, y]$ *is a* $\Phi(A, B)$*-contracting geodesic where* $\Phi(A, B) = \Phi_{2.4}(A, 0, \delta_A) + \Phi_{2.4}(B, \delta_A, 0)$.

(iii) *If* X *is also a proper metric space*, *the geodesic* $[x, b(\infty))$ *and* $(a(\infty), b(\infty))$ *are* $\Phi(A, B)$*-contracting such that* $\Phi(A, B)$ *is as above.*

Proof. (i) Left as an exercise.

(ii) If the concatenation of a and b is a geodesic, this is obvious by part (i), so assume otherwise. By Lemma 2.3 we have that a is δ_A-slim and b is δ_B-slim. We may assume that $\delta_A \geq \delta_B$. By the definition of slimness there is a point w on $[x, y]$ which is within $2\delta_A$ of both of the other sides of the geodesic triangle $\Delta(x, y, z)$. By Lemma 2.4 the geodesic $[x, w]$ is $\Phi_{2.4}(A, 0, 2\delta_A)$-contracting and the geodesic $[w, y]$ is $\Phi_{2.4}(B, 2\delta_A, 0)$-contracting. So by the first part of this lemma, $[x, y]$ is contracting with $\Phi(A, B) = \Phi_{2.4}(A, 0, 2\delta_A) + \Phi_{2.4}(B, 2\delta_A, 0)$.

(iii) This follows easily from part (ii) by taking a sequence of points $y_i \to b(\infty)$. The uniqueness of infinite rays in proper CAT(0) spaces, Lemma 2.4, and the Arzelà–Ascoli theorem imply the statement. □

In contrast with geodesics in Euclidean flats, the diameter of the projection of any geodesic onto a contracting geodesic is finite. The proof can be found in [Charney and Sultan 2013].

Lemma 2.7. *If a is a contracting geodesic and b is any other infinite geodesic then the projection of b onto a is of bounded diameter D.*

In a δ-hyperbolic space, geodesics are coarsely determined by their endpoints on the boundary. For CAT(0) spaces which contain Euclidean flats, this is easily seen to be false; however, if the geodesic happens to be δ-slim, it is true.

Lemma 2.8. *Let a be a δ-slim bi-infinite geodesic. If b is a bi-infinite geodesic which stays a bounded distance from a then b will be in the 2δ-neighborhood of a.*

Proof. Left as an exercise to the reader. □

As a consequence of the bounded projection property for contracting geodesics, Charney and Sultan [2013, Proposition 3.7] proved that contracting geodesics have a strong visibility condition.

Lemma 2.9 (visibility). *If X is a CAT(0) space and a is a contracting geodesic then if b is any geodesic in X there is a bi-infinite geodesic from $b(\infty)$ to $a(\infty)$.*

Lemma 2.10. *If X is a proper CAT(0) space and a is a δ-slim geodesic in X then for any $x \in X$ the distance $d(\pi_a(x), [x, a(\infty)))$ is less than or equal to δ. This just extends the concept of δ-slimness.*

Proof. This is just an application of the δ-slim condition to the sequence of geodesics $[x, a(N)]$ and the Arzelà–Ascoli theorem □

The following lemma is simply Corollary 3.4 from [Bestvina and Fujiwara 2009].

Lemma 2.11 (thin rectangles). *Let w, x, y, z be points such that the geodesic $[x, y]$ is D-contracting and $\pi_{[x,y]}(w) = x$ and $\pi_{[x,y]}(z) = y$. Then there exists an $M > 0$ such that either $d(x, y) < M$ or $d([x, y], [w, z]) < M$, where M depends only on D.*

We are going to need a slightly beefier version of Lemma 2.11 in the following proofs. We are going to have to require tighter control of the entire geodesic and we will let one of our endpoints be a point in the boundary.

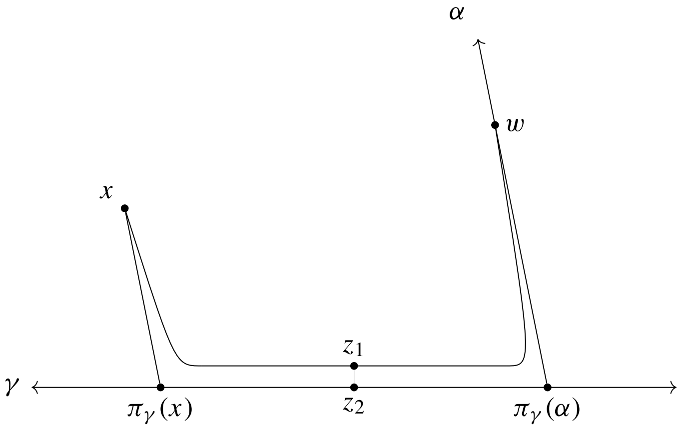

Figure 1. Bounding rectangles.

A remark on the notation in the following lemma: the projection of a point in the contracting boundary, α, onto a D-contracting geodesic, γ, is a well defined notion. For more details and a definition see Remark 3.5 and the discussion before it.

Lemma 2.12. *Let γ be a D-contracting geodesic, $\alpha \in \partial_c X$ and $x \in X$. If c is the geodesic from x to $\pi_\gamma(x)$, b is the geodesic from $\pi_\gamma(x)$ to $\pi_\gamma(\alpha)$ and a is the geodesic from $\pi_\gamma(\alpha)$ to α, then there exists an $M \geq 0$ such that either $d(\pi_\gamma(x), \pi_\gamma(\alpha)) < M$ or the geodesic $[x, \alpha)$ is in the M neighborhood of $c \cup b \cup a$ and vice versa.*

Proof. First, fix a w on the geodesic a. We will prove the lemma replacing α with w and that will suffice as you can take a sequence of w tending towards α and apply the Arzelà–Ascoli theorem to obtain the lemma.

Applying Lemma 2.11 to the points x, $\pi_\gamma(x)$, $\pi_\gamma(\alpha)$ and w we get an M' such that either there are points z_1 and z_2 on the geodesic $[x, w]$ such that $d(z_1, z_2) < M'$ or $d(\pi_\gamma(x), \pi_\gamma(\alpha)) < M'$ (see Figure 1). We may assume the former.

Since γ is D-contracting we may assume that $[z_2, \pi_\gamma(\alpha)]$ is also D-contracting, and thus the triangle z_2, w, $\pi_\gamma(\alpha)$ is δ_D-slim. Since $[z_1, w]$ is in the M' neighborhood of $[z_2, w]$ and $[z_2, w]$ is in the δ neighborhood of $[z_2, \pi_\gamma(\alpha)] \cup [\pi_\gamma(\alpha), w]$, (where δ might be a linear function of δ_D), we have that $[z_1, w]$ is in the $M'+\delta$ neighborhood of $[z_2, \pi_\gamma(\alpha)] \cup [\pi_\gamma(\alpha), w]$. Running the argument in the other direction gives that $[z_2, \pi_\gamma(\alpha)] \cup [\pi_\gamma(\alpha), w]$ is within the $M' + \delta$ neighborhood of $[z_1, w]$.

By repeating this argument with x, $\pi_\gamma(x)$, z_2 and z_1 and setting $M = M' + \delta$ we get the result. □

This next lemma gives us information about the global geometry when the equivalence class of a contracting ray is fixed by a cocompact group action. This lemma will allow us to rule out the existence of global fixed points in the contracting boundary later on.

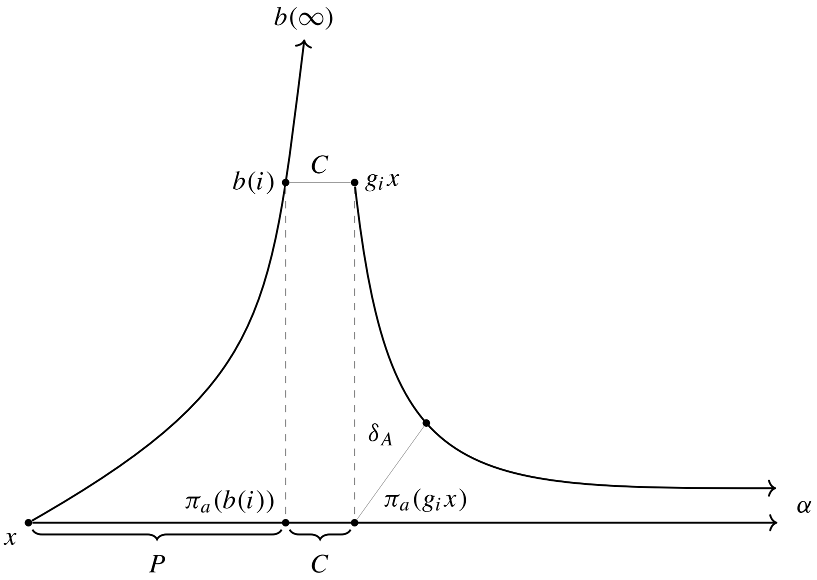

Figure 2. A globally fixed contracting geodesic.

Lemma 2.13. *Let G be some group acting cocompactly by isometries on a $CAT(0)$ space X. If there is some $\alpha \in \partial X$ such that G fixes α and some representative of α is contracting, then every ray in X is contracting.*

Proof. For the following argument, refer to Figure 2. Let b be some ray in X. Pick a representative a of α such that $a(0) = b(0) = x$. Note that because one of the representatives of α is contracting, all of them are, though the contracting constant will depend on x, so let A be such that a and all subsegments of a are A-contracting. By Lemma 2.7 the projection of b onto a is bounded, i.e., there is a P such that $d(x, \pi_a(b(i))) \leq P$ for all i. By cocompactness we also have a $C \geq 0$ and a collection $\{g_i\} \subseteq G$ with $d(g_ix, b(i)) \leq C$. This implies that

$$\begin{aligned} d(x, \pi_a(g_ix)) &\leq d(x, \pi_a(b(i))) + d(\pi_a(b(i)), \pi_a(g_ix)) \\ &\leq P + d(b(i), g_ix) \\ &\leq P + C, \end{aligned}$$

where the second inequality is by the definition of P and the fact that the projection function is nonincreasing.

Because the g_i leave α fixed we have that g_ia is the geodesic from g_ix to α. Since a is contracting, by Lemma 2.10 there is a δ_A such that $d(\pi_a(g_ix), g_ia) \leq \delta_A$ for all i. We can then derive the following inequality:

$$d(x, g_ia) \leq P + C + \delta_A.$$

Thus there is some N_i such that $d(x, g_ia(N_i)) \leq P + C + \delta_A$.

Because all subsegments of $g_i a$ are also A-contracting we have that $[b(0), b(i)]$ is close to an A-contracting geodesic and Lemma 2.4 then implies that $[b(0), b(i)]$ is $\Phi_{2.4}(A, C, P + C + \delta_A)$-contracting for all i. The geodesic b is then contracting since every initial segment is contracting with the same constant. □

Definition 2.14. Let X be a complete CAT(0) space. The angle $\angle(\alpha, \beta)$ between $\alpha, \beta \in \partial X$ is defined as

$$\sup_{x \in X} \angle_x(\alpha, \beta),$$

where $\angle_x(\alpha, \beta)$ is the Alexandrov angle between the two (unique) geodesics which start at x and are in the equivalence class of α and β. The function $\angle(\cdot, \cdot)$ defines a metric on ∂X making it a complete metric space. The associated length metric is called the *Tits metric* and is denoted $d_T(\alpha, \beta)$.

For further information on the Tits metric see [Bridson and Haefliger 1999, Chapter II.9].

The following is a result of Ballmann and Buyalo [2008, Proposition 1.10] and it supplies us with a rank-one isometry for all complete cocompact CAT(0) spaces which have a contracting ray.

Proposition 2.15. *Suppose X is a cocompact $CAT(0)$ space and ∂X is nonempty, then the following are equivalent.*

(1) *X contains a periodic rank-one geodesic.*

(2) *For each $\xi \in \partial X$ there is an $\eta \in \partial X$ with $d_T(\eta, \xi) > \pi$.*

Corollary 2.16. *Let X be a complete and proper $CAT(0)$ space and let G act on X geometrically. If X has a contracting ray then there is a rank-one isometry.*

Proof. By the strong visibility condition in Lemma 2.9, if X has a contracting ray a then it is visible from all points $\xi \in \partial X$. The geodesic between $a(\infty)$ and any $\xi_0 \in \partial X$ guaranteed by visibility tells us that the Alexandrov angle $\angle(a(\infty), \xi_0)$ is equal to π. To show that the Tits distance is larger than π from any point $\xi \in \partial X$, pick a geodesic from $a(\infty)$ to ξ *inside* the Tits boundary ∂X and call it c (note: if no such geodesic exists then $d_T(a(\infty), \xi) = \infty$). Now let ξ_0 be a point on c separate from $a(\infty)$ and ξ. Then we will have that $d_T(a(\infty), \xi) = \text{length}(c) \geq \angle(a(\infty), \xi_0) + \angle(\xi_0, \xi) \geq \pi + \epsilon$; note that $\epsilon > 0$ since $\xi_0 \neq \xi$. So $d_T(a(\infty), \xi) > \pi$ and we have that X has a rank-one periodic geodesic. □

We need the following technical fact about geodesics in metric spaces at several points in this paper; we include a proof for the sake of completeness.

Lemma 2.17. *Let γ be a geodesic in a metric space X and let x be a point in X such that $d(x, \gamma(0)) = t_0$. Then if the distance $d(x, \gamma) \leq D$, then $d(x, \gamma(t_0)) \leq 2D$.*

Proof. Since $d(x, \gamma) \leq D$, let ℓ be a point such that $d(x, \gamma(\ell)) \leq D$. There are two cases, $\ell \geq t_0$ or $\ell < t_0$.

In the first case, we consider the geodesic triangle defined by $\gamma(0)$, $\gamma(\ell)$, and x, but rewrite $\ell = t_0 + a$. The triangle inequality says that $t_0 + a \leq D + t_0$, i.e., $a \leq D$. Then, considering the triangle defined by the three points $\gamma(t_0 + a)$, $\gamma(t_0)$ and x, we get a new triangle inequality, $d(x, \gamma(t_0)) \leq a + D \leq 2D$.

In the second case, we again consider the geodesic triangle given by $\gamma(0)$, $\gamma(\ell)$ and x, but this time we write $\ell = t_0 - a$. The triangle inequality gives us $t_0 \leq D + (t_0 - a)$ or $a \leq D$. Considering the triangle defined by $\gamma(t_0 - a)$, $\gamma(t_0)$ and x, we get the triangle inequality $d(x, \gamma(t_0)) \leq D + a \leq 2D$. ☐

3. The topology of the contracting boundary

The topology of the contracting boundary is very different from any of the typical topologies put on the visual boundary. Later, in Section 5, we will show that the contracting boundary is not always a metric space. In fact, we show it to not even be first-countable. In anticipation of that we will prove some elementary topological facts about the contracting boundary (and direct limit spaces in general) to facilitate some of the later proofs.

First, let's define the contracting boundary and then we will talk about some of its basic topological properties.

Definition 3.1. Let X be a CAT(0) space. Let $\partial_c^D X_x$ be the set of infinite geodesic rays that start at x and are D-contracting; we call this the *D-component* of the contracting boundary. This is a subspace of the visual boundary of X, ∂X_x, and has the associated topology on it. If $D_0 \leq D_1$ then there is the natural continuous inclusion $\partial_c^{D_0} X_x \hookrightarrow \partial_c^{D_1} X_x$, so taking all nonnegative D we get a directed system.

The *contracting boundary*, denoted $\partial_c X_x$, is the union of all of the D-components with the direct limit topology.

The homeomorphism type (but not the contracting constants) of the contracting boundary is independent of the basepoint x, and so typically this will be suppressed when there is no danger of confusion [Charney and Sultan 2013].

One of the basic properties of a direct limit space is that a set in the space is open (respectively, closed) if and only if its intersection with each component is open (closed). In fact, this is often taken as the definition.

Because the topology of the contracting boundary is so dependent on the topology of the components, it will be useful to know how the subspace topology on the components sits inside the visual topology. The following is Lemma 3.3 in [Charney and Sultan 2013].

Lemma 3.2. *For all $D \geq 0$, the D-components of the contracting boundary are closed subsets of the visual boundary.*

Understanding compact sets in the contracting boundary will be important later in our investigation. It turns out that compact sets in the contracting boundary are closely related to the compact sets of the visual boundary, but are limited by their contracting constants.

Lemma 3.3. *A set K is compact in $\partial_c X$ if and only if $K = C \cap \partial_c^D X$ for some compact set $C \subset \partial X$ and some D.*

Proof. $\Leftarrow$ If $K = C \cap \partial_c^D X$ then because C is compact in ∂X and $\partial_c^D X$ is a closed set in ∂X by Lemma 3.2, K is a closed subset in C and therefore compact in ∂X. Now the topology on $\partial_c X$ is defined in such a way that each of the components $\partial_c^D X$ are topologically embedded into $\partial_c X$, i.e., compact subsets of $\partial_c^D X$ will also be compact in $\partial_c X$. So K is a compact set in $\partial_c X$.

$\Rightarrow$ Assume that K is a set in $\partial_c X$ but that K is not contained in $\partial_c^D X$ for any D. These assumptions guarantee that there is some sequence of geodesics $\{a_i\}$ in K where a_i is D_i-contracting and $D_i \to \infty$. By possibly passing to a subsequence we may assume that $D_i > D_{i-1}$ and that each a_i is not D_{i-1}-contracting. Let $A_n = \{a_i\}_{i \geq n+1}$. Note that for all n and all D, $A_n \cap \partial_c^D X$ is a finite set and therefore closed in each component, so A_n is closed in $\partial_c X$.

The collection $\mathcal{O} = \{\partial_c X \setminus A_n\}$ is an open cover of K, but each open set only contains finitely many of the a_i. Take any finite subcollection of $\mathcal{O}$; it will only cover finitely many of the a_i and so it is not a cover, therefore K is not compact. We can then conclude that if K is compact, it is contained in one of the components $\partial_c^D X$ for some D. Because the topology on $\partial_c X$ is finer than that of ∂X, any set which is compact in the contracting boundary is compact in the visual boundary. In other words, every compact set K in the contracting boundary is of the form $K = K \cap \partial_c^D X$ for some D. ☐

We will also want to know when sequences in the contracting boundary converge. It turns out that a sequence converges in the contracting boundary if and only if it converges in the visual boundary and its contracting constants are uniformly bounded above.

Lemma 3.4. *Let X be a proper CAT(0) metric space. A sequence a_i in $\partial_c X_x$ converges to a point $b \in \partial_c X_x$ if and only if the following two conditions hold:*

(1) *There is a uniform K such that a_i is K-contracting for all i.*

(2) *In the visual boundary, $a_i \to b$.*

Proof. $\Leftarrow$ Since the a_i are all K-contracting, $\{a_i, b\} \subseteq \partial_c^M X_x$ where M is the max of the contracting constant of b and K. The topology on this component is just the subspace topology and thus since $a_i \to b$ in the visual boundary, the convergence happens in this component as well. Because each of these components is topologically embedded, convergence occurs in $\partial_c X_x$ as well.

$\Rightarrow$ Note that the topology on $\partial_c X_x$ is finer than that of the subspace topology. In particular, if condition (2) fails then a_i will not converge to b in the contracting boundary.

Assume (1) fails, this means that for each $k \in \mathbb{N}$ there is an i_k such that a_{i_k} is not k-contracting. Consider the set $\{a_{i_k}\}$; this subsequence is in fact closed in $\partial_c X_x$ because only finitely many of its elements are in each $\partial_c^D X_x$ and are thus closed in the subspace topology. Thus $a_i \not\to b$ in $\partial_c X_x$. □

For a point in the contracting boundary, $\alpha \in \partial_c X$, and a C-contracting geodesic ray γ, whose forward endpoint is different from α, we can define $\pi_\gamma(\alpha)$, the projection of α onto γ. If we take a representative of α, say a geodesic a, the projection of a onto γ is a set of finite diameter by Lemma 2.7. There is then some unbounded sequence of t_i such that $\pi_\gamma(a(t_i))$ will converge to some point $\gamma(T)$. Note that the point $\gamma(T)$ depends, not only on the chosen representative of α, but also on the sequence of t_i. For a given representative a we have that a is eventually contained in the compliment of any bounded neighborhood of γ. Applying Lemma 2.11, given a large enough t_0, the geodesic $[a(t_0), a(\infty))$ is not contained in the M_C neighborhood of γ and so for all $t, t' > t_0$, we have $d(\pi_\gamma(a(t)), \pi_\gamma(a(t'))) < M_C$. Thus for any two sequences $a(t_i)$ and $a(t_i')$, since the projections of these two sequences are eventually within M_C of each other, the limits are as well. If another representative of α is chosen, say a', it is within a bounded distance of a, let us call that distance D. By picking t_0 large enough, the geodesics $[a(t_0), a(\infty))$ and $[a'(t_0), a'(\infty))$ are outside of the $D + M_C$ neighborhood of γ. Let $t > t_0$, then there is a t_1 such that $d(a'(t), a(t_1)) \leq D$, so applying Lemma 2.11, the projections of $a'(t)$ and $a(t_1)$ are within M_C of each other. Thus the projections of $[a'(t_0), a'(\infty))$ and $[a(t_0), a(\infty))$ onto γ are within $2M_C$ of each other.

Remark 3.5. A closed and bounded set in a proper CAT(0) metric space has a unique center. That is, a point which is the center of the smallest circle which contains the entire set exists and is unique [Bridson and Haefliger 1999, II.2.7]. In this paper that point will be referred to as the *barycenter*.

Definition 3.6. Given $\alpha \in \partial_c X$, and a C-contracting ray γ, by the previous discussion the set $E = \{\lim_{i\to\infty} \pi_\gamma a(t_i) \mid a \in \alpha \text{ and the limit exists}\}$ will be nonempty and have diameter at most $2M_C$. Let $\pi_\gamma(\alpha)$ be the barycenter of E (or the closure of E if necessary).

Remark 3.7. For any T the three points, $\gamma(T)$, α and $\pi_\gamma(\alpha)$ form an infinite δ-slim triangle where δ depends only on C.

We will topologize the set $\overline{X}_c = X \cup \partial_c X$. Recall that X can be redefined as the set of so-called “generalized” rays in X. A *generalized* ray is a map $a : [0, \infty) \to X$ such that an initial component $a|_{[0,t)}$ is an isometric embedding and the map $a|_{[t,\infty)}$

is constant (by setting $t = \infty$ we get back our infinite rays). Fixing a basepoint, x_0, such that all $a(0) = x_0$, each point $x \in X$ is represented by the unique generalized ray, a_x, whose initial component is the geodesic from x_0 to x, and is the constant function $a_x(k) = x$ for $k \geq d(x_0, x)$ otherwise. Let us denote such a representation of X by X_{x_0}. This defines a topological space $\overline{X} = X \cup \partial X$ which is endowed with the cone topology.

Definition 3.8 (topology of $\overline{X}_c$). Let x_0 be a basepoint in X. Define the following set of generalized rays:

$$\overline{X}_{x_0}^{D} = \{c \in \overline{X} \mid c(0) = x_0 \text{ and } c \text{ is at most } D\text{-contracting}\}.$$

Endowing these sets with the subspace topology from $\overline{X}$, the inclusions will form a directed system. This gives us our topology on $\overline{X}_c$ as the direct limit,

$$\overline{X}_{c,x_0} := \varinjlim \overline{X}_{x_0}^{D}.$$

Remark 3.9. For all $D < D'$ we have the following commutative diagrams, where all inclusions are topological embeddings:

$$\begin{tikzcd}
\partial_c^D X_{x_0} \arrow[r, hook] \arrow[d, hook] & \partial_c^{D'} X_{x_0} \arrow[d, hook] \\
\overline{X}_{x_0}^{D} \arrow[r, hook] & \overline{X}_{x_0}^{D'}
\end{tikzcd}$$

By the universal property of the direct limit topology, this implies that there is a continuous injection $\partial_c X_{x_0} \hookrightarrow \overline{X}_{c,x_0}$. Because all of the maps in the diagram are topological embeddings, it is immediate that this map is also a topological embedding.

X is also topologically embedded in $\overline{X}_{c,x_0}$. This is because, for each $x \in X$, every open ball is eventually contained in $\overline{X}_c^D$ for large enough D. This is just a consequence of Lemma 2.4. Furthermore, X is an open set in $\overline{X}_c$ and so $\partial_c X$ is a closed set.

Lemma 3.10. *Consider a sequence of points $x_i \in X$. Fixing a basepoint x, the sequence x_i converges to some $\alpha \in \partial_c X$ in the topology on $\overline{X}_c$ if and only if the geodesics $[x, x_i]$ converge to α in the cone topology on $\overline{X}$ and the contracting constants of the $[x, x_i]$ are uniformly bounded.*

Proof. The proof of this is the same as the proof of Lemma 3.4. ☐

4. The topological dynamics of the action on the boundary

The topological dynamics of a group action can be a powerful tool in understanding the global topology. In order to better gain an understanding of the contracting boundary of cocompact CAT(0) spaces we will attempt to exploit some well-known results from δ-hyperbolic spaces.

Most of the following results are well-known dynamical results for the visual boundary of a CAT(0) group which contains a rank-one isometry, and a result of Ballmann and Buyalo [2008] guarantees this is the case for the cocompact groups we are considering here. However, the definition of the topology as a direct limit of spaces gives the contracting boundary a much finer topology than the subspace topology would. Because of this different topology we are considering, it is necessary to reprove (and in some cases reword) these dynamics results as none of them will follow as immediate corollaries from the known theorems.

For nonelementary hyperbolic groups, the orbit of every point in the boundary is dense. This establishes a strong dichotomy: either the group is virtually $\mathbb{Z}$ or its boundary has no isolated points.

Our first theorem is establishing this result in the case of the contracting boundary, i.e., the contracting boundary either has no isolated points and has a countable dense subset, or the group is virtually cyclic.

Theorem 4.1. *Let X be a proper CAT(0) space such that G acts geometrically on X. If $\partial_c X \neq \varnothing$ and G is not virtually cyclic then the orbit of each point in $\partial_c X$ is dense.*

This is very similar to a result of Hamenstädt [2009] on the limit set when the group contains a rank-one isometry.

In the spirit of treating the contracting boundary as a replacement for the Gromov boundary for CAT(0) spaces, it is natural to ask whether axial isometries act with North-South Dynamics and if the group G acts as a convergence group action on $\partial_c X$. Recall that an axial isometry is an isometry that fixes a geodesic, called the axis of the isometry; these are also called loxodromic or hyperbolic isometries in the literature. Because the contracting boundary is not compact, the classical formulations of these dynamical properties will have to be reinterpreted somewhat.

Theorem 4.2. *Let X be a proper CAT(0) space on which G acts geometrically. Let g_i be a sequence of isometries in G such that $g_i x \to \gamma^+$ where $\gamma^+ \in \partial_c X$, then there is a subsequence of g_i's where $g_i^{-1} x \to \gamma^-$ for some $\gamma^- \in \partial_c X$ and for every open neighborhood U of γ^+ and every compact set $K \subseteq \partial_c X - \gamma^-$, we have uniform convergence of $g_i(K) \to \gamma^+$.*

This theorem is closer to Papasoglu and Swenson's π-convergence [2009] than it is to a true convergence action. A corollary of this theorem is that rank-one isometries act with a version of North-South dynamics on the contracting boundary.

Corollary 4.3. *Let X be a proper CAT(0) space and let G be a group acting geometrically on it. If g is a rank-one isometry in G, U is an open neighborhood of g^∞ and K is a compact set in $\partial_c X - g^{-\infty}$ then $g^n(K) \subseteq U$ for sufficiently large n.*

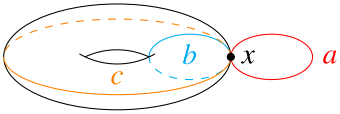

Figure 3. The Salvetti complex X of A_Γ.

Failure of classical North-South dynamics. By classical North-South dynamics, we mean the following theorem.

Theorem 4.4. *If G is a δ-hyperbolic group acting on its Cayley graph X and if g is an infinite-order element then for all open sets U and V with $g^\infty \in U$ and $g^{-\infty} \in V$ we have $g^n(V^c) \subseteq U$ for large enough n.*

It is a well-established fact that for CAT(0) groups, the classical version of North-South dynamics of axial isometries on the visual boundary fails. In particular, if the isometry is not rank-one, whole flats may be fixed by the isometry.

Unfortunately, even if g is a rank-one element of G, this classical version of North-South dynamics on $\partial_c X$ still fails. If a is an axis for g there are open sets U and V of $a(\infty)$ and $a(-\infty)$ such that $g^N(\partial_c X \setminus V) \nsubseteq U$ for any N.

Note: this is in direct contrast with the subspace topology on the set of contracting geodesics, $(\partial_c^{sub} X)$. In [Hamenstädt 2009] and [Ballmann 1995], it was proven that rank-one isometries act on the entire visual boundary with North-South dynamics and thus on any subspace containing the endpoints.

For an example of the failure of the classical North-South dynamics of rank-one isometries on the contracting boundary consider the RAAG, $A_\Gamma = \langle a, b, c \mid [b, c]\rangle$. This is the fundamental group of the Salvetti complex, X (see Figure 3), and its universal cover, $\widetilde{X}$, is a CAT(0) cube complex on which A_Γ acts geometrically [Charney and Davis 1995]. Let γ be an axis for the loxodromic element a. Let b_i be the geodesics following the words $a^{-i}b^i aaaa\cdots$. Note that the contracting geodesics b_i do *not* converge to $\gamma(-\infty)$ in the contracting boundary. This is because the intersection of the set $\{b_i\}$ with each of the contracting components $\partial_c^D X$ is a finite set and therefore closed in the subspace topology, and thus $\{b_i\}$ is closed in $\partial_c X$.

The set $V = (U_{\gamma^-}(r, \epsilon) \cap \partial_c X) \setminus \{b_i\}$ is then an open set around $\gamma(-\infty)$ but for all N we have $a^N b_N \notin U_{\gamma^+}(r', \epsilon')$ for all $\epsilon' < r'$.

Proof of Theorem 4.1. The first step in proving this theorem will be to prove an initially weaker result. We will prove that for a cocompact CAT(0) space, the orbit of a point in the contracting boundary is either a singleton or is dense. The proof relies on the observation that the orbit of a bi-infinite geodesic is easier to understand and contains more geometric information than the orbit of an infinite ray. We will take some contracting ray and one of its orbit points and connect the two with a bi-infinite geodesic. It is then reasonably easy to show that the orbit of this bi-infinite geodesic is dense in the contracting boundary.

Proposition 4.5. *If the action of G on X is cocompact and $\alpha^+ \in \partial_c X$ then α^+ is globally fixed by G or its orbit is dense in $\partial_c X$.*

Proof. First note that if there are only two points in $\partial_c X$ then the proposition is obvious. Either the orbit is a singleton or it is the entire boundary. So from now on we may assume that $|\partial_c X| > 2$ and that α^+ isn't globally fixed.

To show that the orbit is dense it suffices to show that for all $\beta \in \partial_c X$ there exists a sequence of $g_i \in G$ such that $g_i \alpha^+ \to \beta$.

If $\beta \in G\alpha^+$ then we are done since there is an h such that $\beta = h\alpha^+$ so the constant sequence $g_i = h$ will work.

If β is not in the orbit of α^+, pick a point distinct from α^+ in $G\alpha^+$ and call it α^-, i.e., $h\alpha^+ = \alpha^-$ for some $h \neq e$. By the visibility of $\partial_c X$ there is a geodesic connecting α^- to α^+. If we label this geodesic a and pick a basepoint $x = a(0)$ on it, there is also a representative, b, of β, such that $b(0) = x$.

Note: since α^+ and α^- are different elements of the contracting boundary, there are two different contracting constants for their representatives $a|_{[0,\infty)}$ and $a|_{(-\infty,0]}$, but by Lemma 2.6 we have a uniform contracting constant for all of a and we shall call it A. For the representative b of β, let B be its contracting constant. Since Lemma 2.3 guarantees that a and b are slim, we will denote δ_A and δ_B as their slimness constants, respectively. To make the following discussion simpler, we will assume that A and B are chosen so that all subsegments (finite or infinite) of either geodesic are also contracting with the same constant.

By the cocompactness of the action of G on X there is a uniform $C > 0$ such that for each $i \in \mathbb{N}$ there is a $g_i \in G$ such that $d(g_i x, b(i)) \leq C$. Since we've picked g_i so that the orbit of x travels up along b we'd like to say that the geodesic a follows suit, but first we need to pass to a subsequence.

Let $g_i a$ be the bi-infinite geodesic connecting $g_i\alpha^-$ to $g_i\alpha^+$ with basepoint $g_i x$. Now note that there is a t_i such that $g_i a(t_i)$ is the projection of x on $g_i a$ (see Figure 4).

Infinitely many of the t_i will be either positive or negative, so by passing to a subsequence we may assume that all the t_i have the same sign.

In the following argument we will consider the case when $t_i \leq 0$. In this case we will prove that $g_i\alpha^+ \to \beta$. If instead, $t_i > 0$, the following argument will go on to show, *mutatis mutandis*, that $g_i\alpha^- \to \beta$. Because $\alpha^- = h\alpha^+$, this tells us $g_i h\alpha^+ \to \beta$. Thus, in either case, the orbit of α^+ will accumulate on any $\beta \in \partial_c X$.

Consider the representatives of $g_i\alpha^+$ starting from the basepoint x and denote them k_i. To show that the sequence k_i converges in $\partial_c X$ to b, Lemma 3.4 says we only need the following two conditions:

(1) There is a uniform K such that for all i, k_i is K-contracting.

(2) k_i converges to b in the visual boundary ∂X.

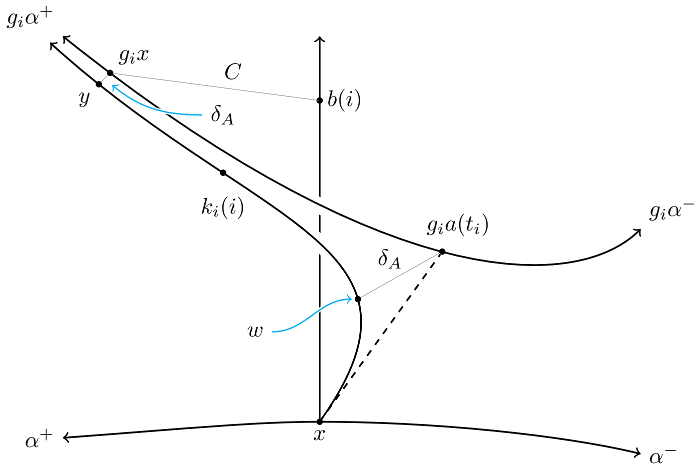

Figure 4. Convergence of α translates.

It turns out that these two ingredients are a direct consequence of the following lemma:

Lemma 4.6. *There is a constant C such that for each i the following holds*:

$$d(k_i(i), b(i)) \leq 2(\delta_A + C).$$

Proof of Lemma 4.6. For the following discussion see Figure 4. We only need to show that the distance from the point $b(i)$ to the geodesic k_i is $\delta_A + C$, then applying Lemma 2.17 we get the result.

Observe that g_ia is A-contracting and thus there is a point w on the geodesic k_i which is within δ_A of $\pi_{g_ia}(x)$ by Lemma 2.10. Recall that $\pi_{g_ia}(x) = g_ia(t_i)$ and that $t_i \leq 0$. By the convexity of the distance function, any point along the geodesic $g_ia|_{[t_i,\infty)}$ will also be within δ_A of k_i. In particular, since $g_ix = g_ia(0) \in g_ia|_{[t_i,\infty)}$, $d(g_ix, k_i) \leq \delta_A$.

Because of how the g_i were defined we also have that $d(g_ix, b(i)) \leq C$. This lets us conclude that $d(b(i), k_i) \leq \delta_A + C$. □

Lemma 4.6 is the key to establishing conditions (1) and (2).

Proof of (1). Since $d(k_i(i), b(i))$ is less than or equal to $2(\delta_A + C)$, which is independent of i, by Lemma 2.4 there exists a constant, $\Phi_{2.4}$, independent from i, such that the geodesic $k_i|_{[0,i]}$ is $\Phi_{2.4}$-contracting.

The cocompact constant gives us $d(b(i), g_ix) \leq C$, so together with Lemma 4.6 we have $d(k_i(i), g_ix) \leq 2\delta_A + 3C$.

Because $g_ia|_{[t_i,\infty)}$ is δ_A-slim, for large enough T the point $k_i(T)$ is within δ_A of $g_ia|_{[t_i,\infty)}$. You can apply Lemma 2.4 again to all subsegments $k_i|_{[i,T]}$ with large T, thus they are all $\Phi'_{2.4}$-contracting for some $\Phi'_{2.4}$ independent of i. This implies that the infinite ray $k_i|_{[i,\infty)}$ is contracting with the same contracting constant.

The concatenation of $k_i|_{[0,i]}$ and $k_i|_{[i,\infty)}$ gives us the entire geodesic k_i. Then Lemma 2.6 tells us that for all i, the k_i are $(\Phi_{2.4} + \Phi'_{2.4})$-contracting. ☐

Proof of (2). Recall that the sets

$$U_b(\varepsilon, r) = \{c \mid c(0) = x \text{ and } d(c(r), b(r)) < \varepsilon\}$$

form a local neighborhood basis for the visual boundary. So for each $U_b(\varepsilon, r)$ we need an N such that $k_i \in U_b(\varepsilon, r)$ for $i \geq N$.

$$N(\varepsilon, r) := \max\left\{r\ ,\ \frac{2r(\delta_A + C)}{\varepsilon}\right\}$$

is just such an N. When $i \geq N(\varepsilon, r)$ we get the following chain of inequalities:

$$d\big(k_i(r), b(r)\big) \leq \frac{r}{i} d\big(k_i(i), b(i)\big) \leq \frac{r}{i} 2(\delta_A + C) \leq \varepsilon.$$

The first inequality is just a restatement of the convexity of the distance function (and is the reason $N(\varepsilon, r)$ is chosen as a max), the second is a result of Lemma 4.6, and the final inequality is just a restatement of the definition of $N(\varepsilon, r)$. Thus we have that the sequence k_i converges to b in the visual boundary. ☐

Establishing conditions (1) and (2) tells us that $k_i \to b$ in $\partial_c X_x$. Because b was arbitrary this tells us that the orbit $G\alpha^+$ is dense in $\partial_c X_x$ and so the statement of the proposition is proven. ☐

The following corollary will come up later and so we will include it here. It states that the orbits of contracting rays which aren't globally fixed are dense in the *visual* boundary.

Corollary 4.7. *If G acts cocompactly on X and $\alpha^+ \in \partial_c X$ isn't globally fixed by G then its orbit is dense in ∂X.*

Proof. This is an immediate consequence of the proof of condition (2) above. At no point was the contracting constant of b used and so replacing it with a noncontracting geodesic gives the same result. (Note that, in this case, condition (1) fails). ☐

Proposition 4.5 is the major component of Theorem 4.1, but there remain a few loose ends. Here is an outline of what remains of the proof. We need to first show that there are enough contracting geodesics in any cocompact CAT(0) space, namely that if the contracting boundary is not empty it contains at least two points. Second, we need to show that if there are exactly two points in the contracting boundary the group is virtually cyclic. This will establish our dichotomy, that our group is virtually cyclic or there are strictly more than two points in our contracting boundary. Finally, it will be easy to then show that if there are more than two points in the contracting boundary, none of them are globally fixed.

Proposition 4.8. *If G acts geometrically on a proper CAT(0) space X, then $|\partial_c X| = 2$ if and only if G is virtually $\mathbb{Z}$.*

Proof. $\Rightarrow$ Let a be a contracting geodesic connecting the two points in $\partial_c X$. Recall that this implies that a is δ-slim for some δ. Because the action of G on X is cocompact there is some C such that for all points $x \in X$ there is some g_x such that $d(g_x a(0), x) \leq C$. Because the contracting boundary only contains two points, $g_x a$ is a bi-infinite geodesic which is asymptotic to the bi-infinite geodesic a. By Lemma 2.8 we have $d(g_x a, a) \leq 2\delta$, so the distance between x and a is bounded by $2\delta + C$. Thus a is a quasi-surjective quasi-isometric embedding of $\mathbb{R}$, i.e., X is quasi-isometric to the real line, and thus G is quasi-isometric to $\mathbb{Z}$. It is a standard exercise to show that a group which is quasi-isometric to $\mathbb{Z}$ is virtually cyclic. For a sketch of the proof, see [Ghys and de la Harpe 1990, p. 10, Exercise 1.16]

$\Leftarrow$ If G is virtually $\mathbb{Z}$, then it is quasi-isometric to $\mathbb{R}$. The contracting boundary of a CAT(0) space is a quasi-isometric invariant, so $\partial_c X$ is equal to $\partial_c \mathbb{R}$ which is two discrete points. ☐

Lemma 4.9. *If X is a proper CAT(0) space with a geometric action and $\partial_c X \neq \varnothing$ then $|\partial_c X| \geq 2$.*

Proof. Since the contracting boundary is nonempty we have at least one contracting ray a. Now look at the orbit of a; if it is not fixed we're done since the orbit of a contracting ray is contracting. If it is fixed, then by Lemma 2.13 every geodesic ray is contracting. So now the only way that we wouldn't have at least two points in the contracting boundary would be if all infinite geodesics were asymptotic. However, if a CAT(0) group is not finite, it contains an infinite order element which has an axis in X; for a proof, see [Swenson 1999]. ☐

Proposition 4.10 (the flat plane theorem). *If a group G is acting geometrically on a CAT(0) space, X, then X is δ-hyperbolic if and only if X contains no Euclidean flats $\mathbb{E}^2$.*

This is a standard result from [Bridson and Haefliger 1999, III.H.1.5].

Corollary 4.11. *Let G act geometrically on a proper CAT(0) space X with nonempty contracting boundary $\partial_c X$. If G fixes a point in $\partial_c X$ then G is virtually $\mathbb{Z}$.*

Proof. Let $\alpha \in \partial_c X$ be a fixed point. By Lemma 2.13 we have that every geodesic in X is contracting. In particular, we have that X cannot contain a Euclidean flat and thus by the flat plane theorem, X is δ-hyperbolic. The Švarc–Milnor lemma then tells us that G is a δ-hyperbolic group. Note that in this case the contracting boundary is the Gromov boundary.

Recall that if a δ-hyperbolic group is nonelementary, i.e., it is neither finite nor virtually cyclic, then it has no globally fixed points in its boundary. This is because it must contain an undistorted free group on two generators and the generators both act by North-South dynamics on the boundary with disjoint fixed points. For a proof of these facts see [Ghys and de la Harpe 1990, Chapter 8]. The group G is then virtually $\mathbb{Z}$ and so we are done. ☐

Proof of Theorem 4.2. We will prove Theorem 4.2 by proving the easier to state theorem below.

Theorem 4.12. *Let γ^+ and γ^- be points in the contracting boundary. If there is a sequence of isometries g_i such that $g_i x \to \gamma^+$ and $g_i^{-1} x \to \gamma^-$, then for any compact set K in $\partial_c X - \{\gamma^-\}$ and any open neighborhood, $U \subseteq \partial_c X$, of γ^+, $g_i(K) \subset U$ for large enough i.*

We can loosen the hypothesis that the g_i^{-1} converge to γ^- to obtain Theorem 4.2 from Theorem 4.12. By a result of Ballmann and Buyalo [2008], if $g_i x \to \gamma^+$, then (passing to a subsequence if necessary) the inverses converge to something in the boundary, let's call it γ^-. Because the contracting constants of the geodesic $[x, g_i^{-1} x]$ (by Lemma 3.10) are uniformly bounded above by some uniform constant B, you can bound the contracting constant of every finite subinterval of γ^- by $B + 1$ (and in fact by B with a little more work). Thus γ^- is contracting as well.

Because open sets in $\partial_c X$ can be much finer than in the visual boundary it is not a priori obvious that there will be any form of North-South dynamics on the contracting boundary. The important observation is that all open sets around γ^+ have a "B-contracting core" which contains the set of all B-contracting elements which are nearby to γ^+ in the visual topology. Because the action by g_i coarsely preserves the contracting constants in K, (and because they are already bounded) you can push the set K into the "core" of U with the dynamics of the visual boundary and establish that it is in fact a subset of U.

Note: I think this is *not* enough to use the ping-pong lemma, because compact sets and neighborhoods aren't complements of each other like they are with the visual topology. This makes me suspect that there is a decent chance this applies to the Morse boundary (where the ping-pong lemma fails in general see [Fink 2015]). Because of this I include a proof of a known dynamics result (Lemma 4.14) on the visual boundary of a CAT(0) space which I believe will be amenable to generalization onto the Morse boundary.

The proof will be broken up into two lemmas in order to simplify the discussion.

For the following we will assume that X is a proper CAT(0) space with nonempty contracting boundary and a group of isometries G acting geometrically.

Lemma 4.13. *Let V be an open set in the contracting boundary containing a point γ, then for each positive constant B there is an r and an ϵ, depending only on B, γ and V such that*

$$\partial_c^B X_x \cap U_\gamma(r, \epsilon) \subset V.$$

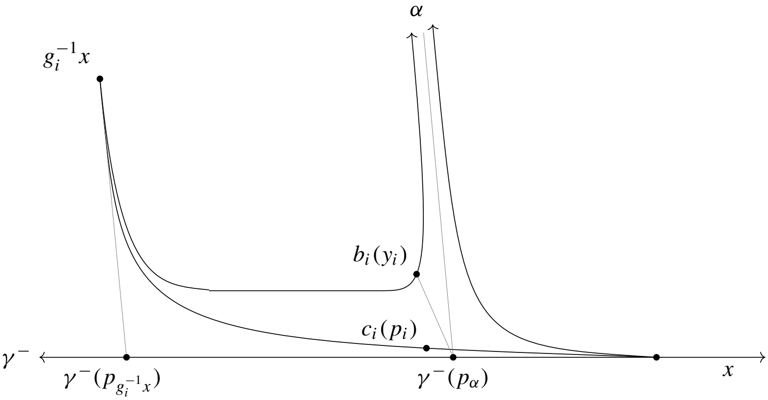

Figure 5. The length p_i is uniformly bounded.

Proof. We can prove this by contradiction. Assume that for some B no such r and ϵ exist. Then for each $n \in \mathbb{N}$ we could find an element of $\partial_c^B X_x \cap U_\gamma(n, 1)$ which is not in V. Thus we have a sequence of geodesics η_n such that $\eta_i \in U_\gamma(n, 1)$ for all $i \geq n$ which is no more than B-contracting. Because this is precisely the condition for convergence of a sequence in the contracting boundary laid out in Lemma 3.4, we have that $\eta_n \to \gamma$ but that the η_n are not in V. Because V is a neighborhood of γ this is a contradiction. □

The following lemma is a direct consequence of the π-convergence due to Papasoglu and Swenson [2009]. This lemma should be generalizable to the Morse boundary so we will provide a different proof which does not rely on the Tits metric and so is likely easier to generalize.

Lemma 4.14. *Let γ^+, γ^- be elements in $\partial_c X$ and g_i be a sequence of group elements such that $g_i x \to \gamma^+$ and $g_i^{-1}x \to \gamma^-$ in $\overline{X}_c$. For any neighborhoods of γ^- and γ^+ in ∂X of the form $U_{\gamma^-}(s, \varepsilon)$ and $U_{\gamma^+}(r, \epsilon)$, there is an N such that for all points α in the set $\partial_c X - U_{\gamma^-}(s, \varepsilon)$ we have $g_i(\alpha) \subset U_{\gamma^+}(r, \epsilon)$ for all $i \geq N$.*

Proof. For the sake of simplicity we can assume that the basepoint x is on a geodesic from γ^- to γ^+. Through an abuse of notation we will conflate the representatives of γ^- and γ^+ starting at x with the elements γ^- and γ^+. Let us denote the geodesic $[x, g_i^{-1}x]$ by c_i. Let a denote the parametrized geodesic from x to α and b_i be the geodesic from $g_i^{-1}x$ to α. For the following argument refer to Figure 5.

Denote by $\gamma^-(p_\alpha)$ the projection of α onto γ^-. This exists provided that $\gamma^+ \neq \alpha$; in the case where such a projection is unbounded, the geodesic b_i is asymptotic to γ^+ and in place of $\gamma^-(p_\alpha)$ a point sufficiently far along γ^+ will suffice since b_i is one leg of a slim ideal triangle. Because $\alpha \in \partial_c X - U_{\gamma^-}(s, \varepsilon)$, there is a uniform bound on $|p_\alpha|$ which depends only on γ^+, γ^-, s and ε. Similarly, denote the projection of $g_i^{-1}x$ onto γ by $\gamma^-(p_{g_i^{-1}x})$.

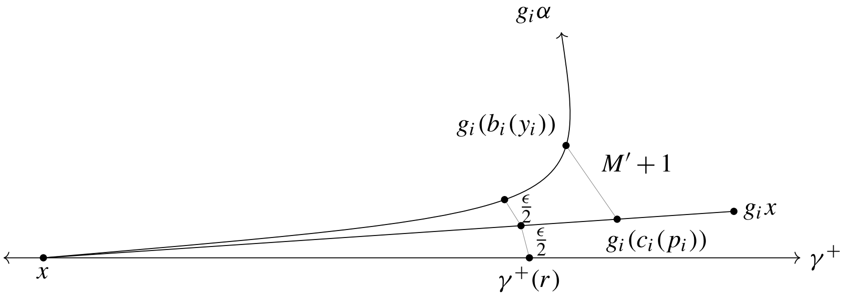

Figure 6. Convergence of the $g_i(b_i)$.

Because the c_i converge to γ^- in $\overline{X}_c$, they are uniformly contracting. Thus the triangle given by γ^- and $g^{-1}x,\ x$ is δ-slim for some δ only depending on the bound of the contracting constants of the c_i. A standard argument shows that given an M', for all sufficiently large i, we have the inequality $|p_{g^{-1}x} - p_\alpha| > M'$. Choosing the M' from Lemma 2.12 gives us that there is a y_i such that $d(\gamma^-(p_\alpha), b_i(y_i)) < M'$. For large enough i we also have a point on c_i, say $c_i(p_i)$, so that $d(c_i(p_i), \gamma^-(p_\alpha)) < 1$. This gives us a bound on the distance;

$$d(c_i(p_i), b_i(y_i)) \leq M' + 1.$$

Note that the length of $[g_i^{-1}x, b_i(y_i)]$ is no shorter than $|p_{g_i^{-1}x} - p_\alpha| - 2M'$, and so we can make this length larger than $\frac{1}{\epsilon}2(M'+1)r$ by picking yet larger i.

Shifting the picture by applying the isometry g_i gives us Figure 6. The previous estimation was arranged so that g_ia is in $U_{g_ic_i}(r, \frac{\epsilon}{2})$. Because the g_ic_i converge to γ^+ we can assume that g_ic_i is in $U_{\gamma^+}(r, \frac{\epsilon}{2})$. By setting N to be the largest of the previous i's we get that $g_ia \in U_{\gamma^+}(r, \epsilon)$. Note that none of the previous estimates depend on α (including the bound on $|p_\alpha|$). □

Proof of Theorem 4.2. We may assume that x is on the bi-infinite geodesic γ from γ^- to γ^+. Let K be a compact set in $\partial_c X - \gamma^-$ and U an open set containing γ^+ in $\partial_c X$. By Lemma 3.3 there is a uniform A such that all elements α in K (with basepoint x) are no more than A-contracting.

Because $[x, g_ix] \to \gamma^+$ by Lemma 3.10 they are no more than B-contracting where B depends only on the sequence of g_i. For all i and all $\alpha \in K$ by Lemma 2.6 the geodesic $[x, g_i\alpha)$ is $\Phi_{2.6}(A, B)$-contracting because the geodesics $[x, g_ix]$ are B-contracting and $[g_ix, g_i\alpha)$ is A-contracting. For notational convenience, set $C = \Phi_{2.6}(A, B)$.

By Lemma 4.13 for the contracting constant C there is an r and an ϵ such that $\partial_c^C X_x \cap U_{\gamma^+}(r, \epsilon) \subset U$. Because K is a compact set in the contracting boundary by Lemma 3.3 it is also a compact set in $\partial X - \{\gamma^-\}$, so there is an s and an ε such that $K \subseteq \partial X_x - U_{\gamma^-}(s, \varepsilon)$. So applying Lemma 4.14, for large enough i we have that $g_i(K) \subset U_{\gamma^+}(r, \epsilon)$, but we already know that $g_i(K) \subset \partial_c^C X_x$ and so $g_i(K) \subset U$. □

5. A characterization of δ-hyperbolicity

One of the ways in which the behavior of the contracting boundary diverges from that of the Gromov boundary is in its local topology. The Gromov boundary comes equipped with a family of visual metrics that induce the same topology on the boundary, making it a compact complete metric space. For the contracting boundary, this happens only in the rarest of circumstances. It is quite easy to cook up examples of spaces which have nonmetrizable contracting boundary. The following is one such example.

Consider again our favorite RAAG, $A_\Gamma = \langle a, b, c \mid [b, c]\rangle$, along with the universal cover of its Salvetti complex, $\widetilde{X}$. The infinite word $w = aaaa\cdots$ corresponds to a 0-contracting geodesic in $\widetilde{X}$ which starts at some lift of the natural basepoint x in X (see Figure 3). If we let $w_i^j = a^i b^j aaa\cdots$, this corresponds to an infinite geodesic starting at the lift of x which is exactly j-contracting (i.e., it is not B-contracting for any $B < j$). It is clear that for each fixed j, the sequences $\{w_i^j\}_{i\in\mathbb{N}}$ converge to w in the contracting boundary. Now if we construct a new sequence by picking an i for each j, i.e., we choose a function $f : \mathbb{N} \to \mathbb{N}$, then regardless of our choice of f the new sequence $\{w_{f(j)}^j\}_{j\in\mathbb{N}}$ will never converge to w. This is because the set $\{w_{f(j)}^j\}_{j\in\mathbb{N}}$ is closed in $\partial_c\widetilde{X}$, as its intersection with each component, $\partial_c^D\widetilde{X}_x$, is finite and therefore closed.

It is a general fact for all first-countable spaces that if you have a countable collection of sequences which all converge to the same point, it is always possible to pick a "diagonal" sequence which also converges. That is, if we have $\{x_i^j\}$ such that $\lim_i x_i^j = x$, there is always some function $f : \mathbb{N} \to \mathbb{N}$ such that $\lim_j x_{f(j)}^j = x$. The proof of this is an elementary exercise in point-set topology. Because this is impossible in the above example we can see that the contracting boundary of $\widetilde{X}$ cannot be metrizable.

Of course, for some CAT(0) spaces, the contracting boundary is metrizable; any CAT(-1) spaces, for instance. It turns out that this is completely generic, the metrizability of the contracting boundary completely characterizes δ-hyperbolicity of cocompact CAT(0) spaces.

Theorem 5.1. *Assume that there is a group G acting geometrically on a complete proper CAT(0) space X, with $|\partial_c X| > 2$, then the following are equivalent*:

(i) *X is δ-hyperbolic.*

(ii) *The contracting constants are bounded*, *i.e.*, $\partial_c X_{x_0} \subseteq \partial_c^D X_{x_0}$ *for some D.*

(iii) *The map* Id : $X \to X$ *induces a homeomorphism* $\partial X \cong \partial_c X$.

(iv) $\partial X \subseteq \partial_c X$, *i.e.*, *as sets the visual boundary and the contracting boundary are the same.*

(v) *$\partial_c X$ is compact.*

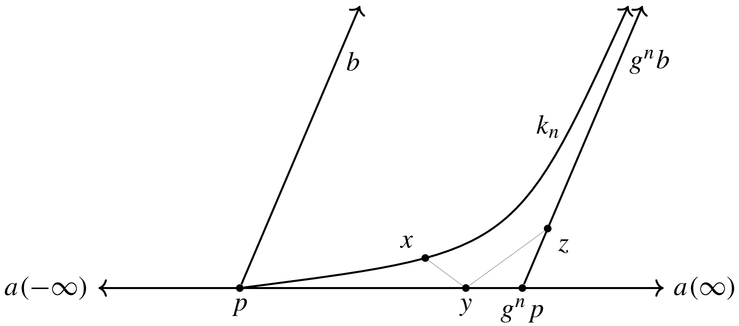

Figure 7. Periodic isometries coarsely fix contracting constants.

(vi) *$\partial_c X$ is locally compact.*

(vii) *$\partial_c X$ is first-countable, and in fact metrizable.*

In order to prove these equivalences we need a bit more fine control over how the contracting constants change under the group action. When there is a rank-one isometry you can say precisely how the contracting constants are changing as you act on a contracting ray. We will make that more precise below, but first we need some notation.

Notation. If b is a B-contracting geodesic in some CAT(0) space X then we will denote the minimum of all contracting constants $\widetilde{B} := \min\{B \mid b \text{ is } B\text{-contracting}\}$.

Lemma 5.2. *Let g be a rank-one isometry of a CAT(0) space X whose axis, a, is A-contracting. If b is a B-contracting geodesic with $b(0) = a(0) = p$, then $k_n = [p, g^n b(\infty))$ will be a K-contracting geodesic such that*

$$\Psi(\widetilde{B}, A) \leq K \leq \Phi(A, B),$$

where $\Psi(\widetilde{B}, A) = \frac{1}{16}(\widetilde{B} - 16A - 77\delta_A - 38) - 3$ and $\Phi(A, B)$ is as in Lemma 2.6.

Proof. Consider the geodesics $g^n b$ and k_n given in Figure 7. By Lemma 2.6, because a is A-contracting and b is B-contracting, the geodesic k_n is at most $\Phi_{2.6}(A, B)$-contracting.

Assume for the sake of contradiction that k_n is K-contracting with

$$K < \frac{\widetilde{B} - 16A - 77\delta_A - 38}{16} - 3.$$

In particular this gives us that $K + 3 < \frac{1}{16}(\widetilde{B} - 16A - 77\delta_A - 38)$.

Because a is δ_A-slim and by replacing δ_A with $2\delta_A$ if necessary, there are then an $x \in k_n$, $y \in a$ and $z \in g^n b$ with $d(x, y) \leq \delta_A$ and $d(y, z) \leq \delta_A$. Now because k_n is K-contracting, the subsegment $[x, k_n(\infty)]$ is $K{+}3$-contracting. The geodesic $[z, g^n b(\infty)]$ is within the $2\delta_A$ neighborhood of $[x, k_n(\infty)]$ so Lemma 2.4 gives us an explicit upper bound on the contracting constant for $[z, g^n b(\infty)]$. In particular we know that it is at worst $\Phi_{2.4}(K+3, 2\delta_A, 2\delta_A)$-contracting where

$$\Phi_{2.4}(K+3, 2\delta_A, 2\delta_A) = 16(K+3) + 70\delta_A + 10.$$

Similarly, we can see that $[g^n p, z]$ is $\Phi_{2.4}(A+3, 0, \delta_A) = (16A + 7\delta_A + 10)$-contracting. Now $g^n b$ is the concatenation of $[g^n p, z]$ and $[z, g^n b(\infty))$ and so we get that it is at most $B' = \Phi_{2.4}(K+3, 2\delta_A, 2\delta_A) + \Phi_{2.4}(A+3, 0, 2\delta_A)$-contracting.

Working everything out, the assumption that we made gives us the inequality

$$\begin{aligned} B' &= 16(K+3) + 16A + 77\delta_A + 38 \\ &< 16\Big(\frac{\widetilde{B} - 16A - 77\delta_A - 38}{16}\Big) + 16A + 77\delta_A + 38 = \widetilde{B}. \end{aligned}$$

But then $g^n b$ is B-contracting with $B < \widetilde{B}$ which is a contradiction. So we know that k_n is K-contracting where $K \geq \Psi(\widetilde{B}, A)$. □

Corollary 2.16 provides us with a rank-one axis whenever the contracting boundary is nonempty and so Lemma 5.2 gives us fine-tuned control over the contracting constants under the action of that rank-one isometry.

Remark 5.3. Suppose we have a sequence of contracting geodesics $\{k_n\}$ and another *noncontracting* geodesic b all with the same basepoint. If the endpoints $k_n(\infty)$ converge to $b(\infty)$ in the visual boundary, then the contracting constants for k_n are unbounded.

We now have all of the ingredients we needed in order to prove the main theorem.

Proof of Theorem 5.1. We will first prove the equivalence of (i) through (iv). The equivalence of (v), (vi), and (vii) with the others will then be easier to show.

(i) ⇒ (ii) The slim triangle condition for a δ-hyperbolic space is easily seen to imply the slim geodesic condition that we have been using; for an explicit proof see [Charney and Sultan 2013]. Because every geodesic is uniformly δ-slim by the hyperbolicity, condition they all have uniform contracting constants by Lemma 2.3.

(ii) ⇒ (iii) Note that because the contracting constants are bounded, the directed system stabilizes, i.e., the collection of contracting geodesics has the subspace topology induced from the visual boundary. Thus, if we can prove that every infinite ray is contracting we would be done, since the contracting boundary will then have the same topology as the entire visual boundary.

Let b be some geodesic ray in X and pick any $a \in \partial_c X$; by Corollary 4.7 there is a sequence of $\{g_i\}$ such that $g_i a \to b$. If b is not contracting then by Remark 5.3 the contracting constants of the representatives of $g_i a$ that start at $x = a(0)$ are growing without bound, which is a contradiction.

(iii) ⇒ (iv) This is obvious.

(iv) $\Rightarrow$ (i) This follows from the flat plane theorem, i.e., since $\partial X \subseteq \partial_c X$ we have that every geodesic in X is contracting, thus there are no noncontracting geodesics. In particular this implies that there are no Euclidean planes embedded in X, but by the above theorem this implies that X is δ-hyperbolic.

(iii) $\Rightarrow$ (v) For a proper CAT(0) space, ∂X is compact.

(v) $\Rightarrow$ (vi) This is trivial.

(vi) $\Rightarrow$ (ii) Assume (ii) is false, then we will show that $\partial_c X$ is not locally compact.

Let α be some element of $\partial_c X$ and let U be an arbitrary neighborhood of α. By Corollary 2.16 there is some rank-one isometry, g, and by Theorem 4.1 we may assume that the forward endpoint of g is in the interior of U. Let a be an axis of g and let A be its contracting constant.

By assumption, there is a subset $B = \{b_i\}$ such that the minimal contracting constant of each b_i is bounded below by $16(i + 16A + 77\delta_A + 38)$. Applying Theorem 4.1 with the g^n as the g_i and switching the roles of a and b_i we can see that $g^n b_i(\infty)$ converges to $a(\infty)$. In particular, for each i there is an n, say n_i, such that $g^{n_i} b_i(\infty)$ is in U. Let the geodesics $[a(0), g^{n_i} b_i(\infty))$ be denoted by c_i. Applying Lemma 5.2 we get that the c_i are at least i-contracting.

Let the collection $\mathcal{C} = \{C_i\}$ where $C_i = \{c_j\}_{j \geq i}$. Each set C_i is a closed subset of U and $\bigcap C_i = \varnothing$. The collection $\{U \setminus C_i\}$ will then be an open cover of U with no finite refinement and so U is not compact. Since α and U were arbitrary, $\partial_c X$ is not locally compact.

(i) + (iii) $\Rightarrow$ (vii) The Gromov boundary of a δ-hyperbolic group is metrizable and since the contracting boundary is homeomorphic to the visual boundary we are done.

(vii) $\Rightarrow$ (ii) Assume that (ii) is false, that there is no upper bound on the contracting constants of the contracting boundary. We will show that the contracting boundary is not first-countable (and thus not metrizable).

As with the example in the introduction to this section it is enough to exhibit a collection $\{\alpha_i^j\}$ and an α in the contracting boundary such that for each j, $\alpha_i^j \to \alpha$ as $i \to \infty$, but the α_i^j are at best j-contracting. In particular, this means that for any function $f : \mathbb{N} \to \mathbb{N}$, the sequence $\alpha_{f(j)}^j$ will not converge to α. This is because the intersection of $\{\alpha_{f(j)}^j\}$ with $\partial_c^D X_{x_0}$ will always be finite and thus the set $\{\alpha_{f(j)}^j\}$ is closed. We've already seen that the existence of such a sequence contradicts first-countability.

The construction of the α_i^j's isn't particularly hard in light of Lemma 5.2. Since $\partial_c X$ is nonempty we have, by Corollary 2.16, a rank-one isometry g with axis a. Now since a is rank-one it has a contracting constant A and is δ_A-slim. We are assuming that there is no upper bound on the contracting constants for geodesics so pick a geodesic b^j with a minimal contracting constant $\widetilde{B}_j$ of at least $16j+16A+77\delta_A+38$. By Lemma 5.2 the geodesics $k_i^j=[b(0),\, g^i b^j(\infty)]$ will be K-contracting where $j \leq \Psi(\widetilde{B}_j,\, A) \leq K \leq \Phi_{5.2}(A,\, \widetilde{B}_j)$.

So we have our collection of points in the contracting boundary $\{k_i^j(\infty)\}$. For each j, the geodesics k_i^j have a fixed upper bound on their contracting constants. To get convergence in the visual boundary recall that a rank-one isometry acts by North-South dynamics on the visual boundary [Hamenstädt 2009]. Thus $\lim_i k_i^j(\infty)=a(\infty)$ for each j in $\partial_c X$ and the contracting constants are bounded below by j. This gives us that $\partial_c X$ cannot be first-countable. □

Acknowledgements

The author would like to thank Ruth Charney and Matthew Cordes for always finding the time to answer the many questions that came up while he was working on this paper. The author would also like to thank the various referees for productive commentary which greatly improved the paper.

References

[Algom-Kfir 2011] Y. Algom-Kfir, "Strongly contracting geodesics in outer space", *Geom. Topol.* **15**:4 (2011), 2181–2233. MR Zbl

[Ballmann 1985] W. Ballmann, "Nonpositively curved manifolds of higher rank", *Ann. of Math.* (2) **122**:3 (1985), 597–609. MR Zbl

[Ballmann 1995] W. Ballmann, *Lectures on spaces of nonpositive curvature*, DMV Seminar **25**, Birkhäuser, Basel, 1995. MR Zbl

[Ballmann and Buyalo 2008] W. Ballmann and S. Buyalo, "Periodic rank one geodesics in Hadamard spaces", pp. 19–27 in *Geometric and probabilistic structures in dynamics*, edited by K. Burns et al., Contemp. Math. **469**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[Behrstock and Charney 2012] J. Behrstock and R. Charney, "Divergence and quasimorphisms of right-angled Artin groups", *Mathematische Annalen* **352**:2 (2012), 339–356. Zbl

[Bestvina and Fujiwara 2009] M. Bestvina and K. Fujiwara, "A characterization of higher rank symmetric spaces via bounded cohomology", *Geom. Funct. Anal.* **19**:1 (2009), 11–40. MR Zbl

[Bridson and Haefliger 1999] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **319**, Springer, 1999. MR Zbl

[Caprace and Fujiwara 2010] P.-E. Caprace and K. Fujiwara, "Rank-one isometries of buildings and quasi-morphisms of Kac–Moody groups", *Geom. Funct. Anal.* **19**:5 (2010), 1296–1319. MR Zbl

[Caprace and Sageev 2011] P.-E. Caprace and M. Sageev, "Rank rigidity for CAT(0) cube complexes", *Geom. Funct. Anal.* **21**:4 (2011), 851–891. MR Zbl

[Casson and Jungreis 1994] A. Casson and D. Jungreis, "Convergence groups and Seifert fibered 3-manifolds", *Invent. Math.* **118**:3 (1994), 441–456. MR Zbl

[Charney and Davis 1995] R. Charney and M. W. Davis, "Finite $K(\pi,\,1)$s for Artin groups", pp. 110–124 in *Prospects in topology* (Princeton, NJ, 1994), edited by F. Quinn, Ann. of Math. Stud. **138**, Princeton Univ. Press, 1995. MR Zbl

[Charney and Sultan 2013] R. Charney and H. Sultan, "Contracting boundaries of CAT(0) spaces", *J. Topol.* **8**:1 (2013), 93–117. MR Zbl

[Cordes 2017] M. Cordes, "Morse boundaries of proper geodesic metric spaces", *Groups Geom. Dyn.* **11**:4 (2017), 1281–1306. MR Zbl

[Croke and Kleiner 2000] C. B. Croke and B. Kleiner, "Spaces with nonpositive curvature and their ideal boundaries", *Topology* **39**:3 (2000), 549–556. MR Zbl

[Fink 2015] E. Fink, "Hyperbolicity via geodesic stability", 2015. arXiv

[Freden 1995] E. M. Freden, "Negatively curved groups have the convergence property, I", *Ann. Acad. Sci. Fenn. Ser. A I Math.* **20**:2 (1995), 333–348. MR Zbl

[Gabai 1992] D. Gabai, "Convergence groups are Fuchsian groups", *Ann. of Math.* (2) **136**:3 (1992), 447–510. MR Zbl

[Ghys and de la Harpe 1990] E. Ghys and P. de la Harpe, "Espaces métriques hyperboliques", pp. 27–45 in *Sur les groupes hyperboliques d'après Mikhael Gromov* (Bern, 1988), edited by E. Ghys and P. de la Harpe, Progr. Math. **83**, Birkhäuser, Boston, 1990. MR Zbl

[Hamenstädt 2009] U. Hamenstädt, "Rank-one isometries of proper CAT(0)-spaces", pp. 43–59 in *Discrete groups and geometric structures*, edited by K. Dekimpe et al., Contemp. Math. **501**, Amer. Math. Soc., Providence, RI, 2009. MR Zbl

[Mostow 1968] G. D. Mostow, "Quasi-conformal mappings in n-space and the rigidity of hyperbolic space forms", *Inst. Hautes Études Sci. Publ. Math.* 34 (1968), 53–104. MR Zbl

[Papasoglu and Swenson 2009] P. Papasoglu and E. Swenson, "Boundaries and JSJ decompositions of CAT(0)-groups", *Geom. Funct. Anal.* **19**:2 (2009), 559–590. MR Zbl

[Swenson 1999] E. L. Swenson, "A cut point theorem for CAT(0) groups", *J. Differential Geom.* 2 (1999), 327–358. MR Zbl

Received July 5, 2016. Revised February 1, 2018.

DEVIN MURRAY
DEPARTMENT OF MATHEMATICS
BRANDEIS UNIVERSITY
WALTHAM, MA
UNITED STATES

dmurray@brandeis.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 1, 2019
dx.doi.org/10.2140/pjm.2019.299.117

KMS CONDITIONS, STANDARD REAL SUBSPACES AND REFLECTION POSITIVITY ON THE CIRCLE GROUP

KARL-HERMANN NEEB AND GESTUR ÓLAFSSON

We continue our investigations of the representation theoretic side of reflection positivity by studying positive definite functions ψ on the additive group $(\mathbb{R}, +)$ satisfying a suitably defined KMS condition. These functions take values in the space $\mathrm{Bil}(V)$ of bilinear forms on a real vector space V. As in quantum statistical mechanics, the KMS condition is defined in terms of an analytic continuation of ψ to the strip

$$\{z \in \mathbb{C} : 0 \leq \mathrm{Im} z \leq \beta\}$$

with a coupling condition $\psi(i\beta + t) = \overline{\psi(t)}$ on the boundary. Our first main result consists of a characterization of these functions in terms of modular objects (Δ, J) (J an antilinear involution and $\Delta > 0$ selfadjoint with $J\Delta J = \Delta^{-1}$) and an integral representation.

Our second main result is the existence of a $\mathrm{Bil}(V)$-valued positive definite function f on the group $\mathbb{R}_\tau = \mathbb{R} \rtimes \{\mathrm{id}_{\mathbb{R}}, \tau\}$ with $\tau(t) = -t$ satisfying $f(t, \tau) = \psi(it)$ for $0 \leq t \leq \beta$. We thus obtain a 2β-periodic unitary one-parameter group on the GNS space $\mathcal{H}_f$ for which the one-parameter group on the GNS space $\mathcal{H}_\psi$ is obtained by Osterwalder–Schrader quantization.

Finally, we show that the building blocks of these representations arise from bundle-valued Sobolev spaces corresponding to the kernels

$$(\lambda^2 - d^2/dt^2)^{-1}$$

on the circle $\mathbb{R}/\beta\mathbb{Z}$ of length β.

MSC2010: primary 43A35, 43A65, 47L30; secondary 47L90, 81T05.
Keywords: KMS condition, reflection positivity, standard real subspace.

1. Introduction

In this note we continue our investigations of the mathematical foundations of *reflection positivity*, a basic concept in constructive quantum field theory [Glimm and Jaffe 1981; Klein and Landau 1983; Jorgensen and Ólafsson 1998; 2000; De Angelis et al. 1986; Jaffe and Ritter 2007]. Originally, reflection positivity, also called Osterwalder–Schrader positivity, arises as a requirement on the euclidean side to establish a duality between euclidean and relativistic quantum field theories [Osterwalder and Schrader 1973]. It is closely related to "Wick rotation" or "analytic continuation" in the time variable from the real to the imaginary axis.

The underlying fundamental concept is that of a *reflection positive Hilbert space*, introduced in [Neeb and Ólafsson 2014]. This is a triple $(\mathcal{E}, \mathcal{E}_+, \theta)$, where $\mathcal{E}$ is a Hilbert space, $\theta : \mathcal{E} \to \mathcal{E}$ is a unitary involution and $\mathcal{E}_+$ is a closed subspace of $\mathcal{E}$ which is θ-positive in the sense that $\langle \theta v, v \rangle \geq 0$ for $v \in \mathcal{E}_+$.

In [Neeb and Ólafsson 2014], we introduced the concept of a reflection positive cyclic representation $(\pi, \mathcal{E}, v)$, where $(\mathcal{E}, \mathcal{E}_+, \theta)$ is a reflection positive Hilbert space and $v \in \mathcal{E}$ a θ-fixed vector (or, more generally, a distribution vector). In the present paper we shall see that, to treat reflection positive representations of the circle group $G = \mathbb{T}$ corresponding to unitary representations of the dual group $G^c \cong \mathbb{R}$ arising from KMS states, or from their modular objects (Δ, J),[1] we are forced to work in a more general framework, where the representations are generated by the image of an $\mathbb{R}$-linear map $j : V \to \mathcal{E}$ from a real vector space V into the representation space $\mathcal{E}$ and where $j(V)$ does not consist of θ-fixed vectors.

To explain the corresponding concept of a *reflection positive representation*, we start with a *symmetric Lie group*, i.e., a pair (G, τ), where $\tau \in \mathrm{Aut}(G)$ is an involution. Then we form the extended group $G_\tau := G \rtimes \{\mathbf{1}, \tau\}$. Let $(U, \mathcal{E})$ be a unitary representation of G_τ and let $j : V \to \mathcal{E}$ be a linear map from the real vector space V to $\mathcal{E}$. Then $(U, \mathcal{E}, j, V)$ is called *reflection positive with respect to a subset* $G_+ \subseteq G$ if the closed subspace $\mathcal{E}_+$ generated by $U_{G_+}^{-1} j(V)$ defines a reflection positive Hilbert space $(\mathcal{E}, \mathcal{E}_+, U_\tau)$. Generalizing the well-known Gelfand–Naimark–Segal (GNS) construction leads to an encoding of representations generated by $j(V)$ in terms of form-valued positive definite functions $\psi(g)(v, w) := \langle j(v), U_g j(w) \rangle$ [Neeb and Ólafsson 2015b].

This paper continues the investigations started in [Neeb and Ólafsson 2015b], where we studied reflection positive representations of the circle group and their connections to KMS states, which was largely motivated by the work of Klein and Landau [1981] (see also [Cuniberti et al. 2001]). A long-term goal of this project is

---

[1] Recall that KMS stands for Kubo–Martin–Schwinger; see [Bratteli and Robinson 1981, §5.3.1] for more on KMS states and their interpretation in quantum statistical mechanics as thermal equilibrium states.

to combine our representation theoretic approach to reflection positivity with KMS states of operator algebras and Borchers triples corresponding to modular inclusions [Buchholz et al. 2011; Borchers 1992; Longo 2008; Schlingemann 1999].

A crucial step in this direction is the concept of a positive definite function satisfying a KMS condition that can be formulated as follows: First, let V be a real vector space and $\mathrm{Bil}(V)$ be the space of real bilinear maps $V \times V \to \mathbb{C}$. A function $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ is said to be *positive definite* if the kernel $\psi(t-s)(v, w)$ on $\mathbb{R} \times V$ is positive definite. For $\beta > 0$, we consider the open strip $\mathcal{S}_\beta := \{z \in \mathbb{C} : 0 < \mathrm{Im}\, z < \beta\}$. We say that a positive definite function $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ satisfies the *KMS condition* for $\beta > 0$ if ψ extends to a function $\overline{\mathcal{S}_\beta} \to \mathrm{Bil}(V)$ which is pointwise continuous and pointwise holomorphic on the interior $\mathcal{S}_\beta$, and satisfies

$$\psi(i\beta + t) = \overline{\psi(t)} \quad \text{for } t \in \mathbb{R}.$$

The central idea in the classification of positive definite functions satisfying a KMS condition is to relate them to *standard real subspaces* of a (complex) Hilbert space; these are closed real subspaces $V \subseteq \mathcal{H}$ for which $V \cap iV = \{0\}$ and $V + iV$ is dense (cf. Definition 2.4). Any such subspace determines a pair (Δ, J) of *modular objects*, where Δ is a positive selfadjoint operator and J is an antilinear involution satisfying $J\Delta J = \Delta^{-1}$. The connection is established by $V = \mathrm{Fix}(J\Delta^{1/2}) = \{v \in \mathcal{D}(\Delta^{1/2}) : J\Delta^{1/2}v = v\}$. Our first main result is the following characterization of the KMS condition in terms of standard real subspaces. Here we write $\mathrm{Bil}^+(V) \subseteq \mathrm{Bil}(V)$ for the convex cone of all those bilinear forms f for which the sesquilinear extension to $V_\mathbb{C} \times V_\mathbb{C}$ is positive semidefinite.

Theorem 2.6 (characterization of the KMS condition). *Let V be a real vector space and $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ be a pointwise continuous positive definite function. Then the following are equivalent*:

(i) *ψ satisfies the KMS condition for $\beta > 0$.*

(ii) *There exists a standard real subspace V_1 in a Hilbert space $\mathcal{H}$ and a linear map $j : V \to V_1$ such that*

$$\psi(t)(v, w) = \langle j(v), \Delta^{-it/\beta} j(w)\rangle \quad \textit{for } t \in \mathbb{R},\ v, w \in V. \tag{1}$$

(iii) *There exists a $\mathrm{Bil}^+(V)$-valued regular Borel measure μ on $\mathbb{R}$ satisfying*

$$\psi(t) = \int_\mathbb{R} e^{it\lambda}\, d\mu(\lambda), \quad \textit{where } d\mu(-\lambda) = e^{-\beta\lambda} d\overline{\mu}(\lambda).$$

If these conditions are satisfied, then the function $\psi : \overline{\mathcal{S}_\beta} \to \mathrm{Bil}(V)$ is pointwise bounded.

The equivalence of (i) and (ii) in Theorem 2.6 describes the tight connection between the KMS condition and the modular objects associated to a standard real subspace. Part (iii) provides an integral representation that can be viewed as a classification result.

For a function ψ satisfying the β-KMS condition, analytic continuation leads to the operator-valued function

$$\varphi : [0, \beta] \to B(V_{\mathbb{C}}), \qquad \langle v, \varphi(t)w\rangle = \psi(it)(v, w).$$

This function satisfies $\varphi(\beta) = \overline{\varphi(0)}$, and hence extends uniquely to a (weak operator) continuous function $\varphi : \mathbb{R} \to B(V_{\mathbb{C}})$ satisfying

$$\varphi(t + \beta) = \overline{\varphi(t)} \quad \text{for } t \in \mathbb{R}. \tag{2}$$

Recall the group $\mathbb{R}_\tau := \mathbb{R} \rtimes \{1, \tau\}$ with $\tau(t) = -t$. In Theorem 4.5 we show that there exists a positive definite function

$$f : \mathbb{R}_\tau \to \mathrm{Bil}(V) \quad \text{satisfying } f(t, \tau) = \varphi(t).$$

The function f is 2β-periodic, hence factors through a function on $\mathbb{T}_{2\beta,\tau} := \mathbb{R}_\tau/\mathbb{Z}2\beta \cong \mathrm{O}_2(\mathbb{R})$. This leads to a natural "euclidean" counterpart of the unitary one-parameter group $U_t = \Delta^{-it/\beta}$ associated to the KMS positive definite function ψ. To understand the structure of the positive definite functions which arise in this way, and the corresponding unitary representations of $\mathbb{T}_{2\beta,\tau}$, we write $f = f_+ + f_-$ with $f_+(\beta + t, \tau^\varepsilon) = f_+(t, \tau^\varepsilon)$ (the bosonic part) and $f_-(\beta + t, \tau^\varepsilon) = -f_-(t, \tau^\varepsilon)$ (the fermionic part). Then $f_\pm$ are both positive definite and combine to a matrix valued positive definite function

$$f^\sharp := \begin{pmatrix} f_+ & 0 \\ 0 & f_- \end{pmatrix} : \mathbb{R}_\tau \to M_2(B(V_{\mathbb{C}})) \cong B(V_{\mathbb{C}}^2)$$

(Lemma 4.12). Neglecting an additive summand which is constant, we can now define a unitary representation of the subgroup $P := (\mathbb{Z}\beta)_\tau$ on $V_{\mathbb{C}}^2$ by

$$\rho(\beta, \mathbf{1}) := \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \quad \text{and} \quad \rho(0, \tau) := \begin{pmatrix} \mathbf{1} & 0 \\ 0 & iI \end{pmatrix},$$

where I is a complex structure on V. Then we have the relation

$$f^\sharp(hg) = \rho(h) f^\sharp(g) \quad \text{for } h \in P,\ g \in \mathbb{R}_\tau,$$

which determines in particular how $f^\sharp$ is obtained from the function φ above. For the special case where the real representation corresponding to ψ is isotypic, or the associated modular operator Δ is a multiple of the identity, the GNS representation $(U^{f^\sharp}, \mathcal{H}_{f^\sharp})$ can be realized on the Hilbert space completion of

$$\Gamma_\rho := \{s \in C^\infty(\mathbb{R}_\tau, V_{\mathbb{C}}^2) : s(hg) = \rho(h)s(g) \text{ for all } g \in \mathbb{R}_\tau,\ h \in P\}$$

with respect to the scalar product

$$\langle s_1, s_2\rangle := \frac{1}{2\beta}\int_0^{2\beta} \langle s_1(t,\mathbf{1}), ((\lambda^2-\Delta)^{-1}s_2)(t,\mathbf{1})\rangle\,dt, \qquad \text{where } \Delta = \frac{d^2}{dt^2}.$$

On this space, $\mathbb{R}_\tau$ acts by right translation. This provides a natural "euclidean realization" of our representation on the Riemannian manifold $\mathbb{T}_\beta \cong \mathbb{S}^1$ in the spirit of [De Angelis et al. 1986; Dimock 2004; Jaffe and Ritter 2007]. The "periodicity in imaginary time" that we also observe here has been studied in detail from a physics perspective by Fulling and Ruijsenaars [1987].

We conclude this paper with a short Section 5, in which we prove a version of Theorem 2.6 for $\beta = \infty$ which connects naturally to our previous work on dilations of semigroups of contractions in [Neeb and Ólafsson 2015a]. In two appendices we provide some background material. Appendix A recalls some facts on positive definite kernels and discusses in particular the connection between complex and real-valued kernels. Appendix B discusses standard real subspaces in terms of skew-symmetric contractions on real Hilbert spaces. This perspective was crucial for the present paper, and we expect it to be useful in other contexts as well.

In a subsequent paper [Neeb and Ólafsson 2019], we extend the results obtained here for the group $G = \mathrm{O}_2(\mathbb{R}) = \mathrm{SO}_2(\mathbb{R})_\tau$ to more general groups such as $\mathrm{O}_{n+1}(\mathbb{R})$ (where reflection positivity refers to the sphere $\mathbb{S}^n$) and $\mathrm{O}_{1,n}(\mathbb{R})$ (where reflection positivity refers to the n-dimensional hyperbolic space $\mathbb{H}^n$). Eventually, we would like to see how our representation theoretic analysis can be blended with the existing work on relativistic KMS conditions [Bros and Buchholz 1994; Gérard and Jäkel 2007] and in particular with [Barata et al. 2013; 2016]. The close connection between modular objects (Δ, J) and standard real subspaces was first explored by Rieffel and van Daele [1977]. They also define a notion of a KMS condition for a unitary one-parameter group $(U_t)_{t\in\mathbb{R}}$ on a complex Hilbert space $\mathcal{H}$ with a real subspace $V \subseteq \mathcal{H}$. In our terms, their condition means that the function $\psi : \mathbb{R} \to \mathrm{Bil}(V)$, $\psi(t) = \langle v, U_t w\rangle$ satisfies the KMS condition for $\beta = -1$ (which refers to a function on the strip $\{-1 < \mathrm{Im}\, z < 0\}$). From [Rieffel and van Daele 1977, Proposition 3.7], one can easily derive the implication (ii) $\Rightarrow$ (i) of Theorem 2.6 (cf. also [Longo 2008, Proposition 3.7]). In this case, [Rieffel and van Daele 1977, Theorem 3.8] even implies that $U_t = \Delta^{-it/\beta}$ is the unique unitary one-parameter group satisfying the KMS condition for β. From [Rieffel and van Daele 1977, Theorem 3.9], one can also derive the implication (i) $\Rightarrow$ (ii). Instead of Δ, Rieffel and van Daele work with the bounded operator $R = 2(\mathbf{1}+\Delta)^{-1}$ which is the sum of the orthogonal projections of the real Hilbert space $\mathcal{H}$ onto the closed subspaces V and iV. In our context, this operator appears as $\mathbf{1} + i\widehat{C}$ for the skew-hermitian operator $\widehat{C} = i\frac{\Delta-\mathbf{1}}{\Delta+\mathbf{1}}$ (Lemma 4.2).

In the context of free fields, the interplay between standard real subspaces and von Neumann algebras of operators on Fock space has already been studied by Araki [1963] and Eckmann and Osterwalder [1973]. The connection between the KMS condition and the modular theory of von Neumann algebras has already been observed and studied in [Haag et al. 1967]. We refer to [Yngvason 1994] for some particularly interesting concrete subspaces corresponding to fields on light rays and to [Ramacher 2000] for descriptions of standard real subspaces in terms of boundary values of holomorphic functions. Numerical aspects of the KMS condition and rather general holomorphic extension aspects have recently been studied in [De Micheli and Viano 2012].

Notation. We follow the "physics convention" that the scalar product $\langle \cdot, \cdot \rangle$ on a complex Hilbert space is linear in the second argument.

For a real vector space V, we write $\mathrm{Bil}(V)$ for the complex vector space of complex-valued bilinear forms $V \times V \to \mathbb{C}$. For $h \in \mathrm{Bil}(V)$, we write $\bar{h}$ for the pointwise complex conjugate and put $h^\top(v, w) := h(w, v)$ and $h^* := \bar{h}^\top$. We say that h is *hermitian* if $\bar{h} = h^\top$, which means that $\mathrm{Re}\, h$ is symmetric and $\mathrm{Im}\, h$ is skew-symmetric. We write $\mathrm{Herm}(V) \subseteq \mathrm{Bil}(V)$ for the real subspace of hermitian forms.

Every $h \in \mathrm{Bil}(V)$ extends canonically to a sesquilinear form on $V_{\mathbb{C}}$ (linear in the second argument),

$$h_{\mathbb{C}}(v + iw, v' + iw') := h(v, v') - ih(w, v') + ih(v, w') + h(w, w').$$

We may therefore identify $\mathrm{Bil}(V)$ with the space $\mathrm{Sesq}(V_{\mathbb{C}})$ of sesquilinear forms on the complex vector space $V_{\mathbb{C}}$. We write $\mathrm{Bil}^+(V) \subseteq \mathrm{Bil}(V)$ for the convex cone of all those bilinear forms f for which the sesquilinear extension to $V_{\mathbb{C}} \times V_{\mathbb{C}}$ is positive semidefinite, i.e., for which h defines a positive definite kernel on V.

2. Positive definite functions and KMS conditions

Throughout this section V is an arbitrary real vector space. We recall from Definition A.3 that a function $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ is called *positive definite* if the kernel $K((t, v), (s, w)) := \psi(t - s)(v, w)$ on $\mathbb{R} \times V$ is positive definite. The main result of this section is Theorem 2.6. This result leads in particular to the analytic continuation of ψ to the strip $\mathcal{S}_\beta$. We also explain how the corresponding representation of $\mathbb{R}$ can be realized in a Hilbert space consisting of holomorphic functions on the strip $\mathcal{S}_{\beta/2}$ with continuous boundary values (Proposition 2.9).

We call a function $\psi : \overline{\mathcal{S}_\beta} \to \mathrm{Bil}(V)$ *pointwise continuous* if, for all $v, w \in V$, the function $\psi^{v,w}(z) := \psi(z)(v, w)$ is continuous. Moreover, we say that ψ is *pointwise holomorphic in* $\mathcal{S}_\beta$, if, for all $v, w \in V$, the function $\psi^{v,w}|_{\mathcal{S}_\beta}$ is holomorphic. By the Schwarz reflection principle, any pointwise continuous pointwise holomorphic function ψ is uniquely determined by its restriction to $\mathbb{R}$.

Definition 2.1. For $\beta > 0$, let $\mathcal{S}_\beta := \{z \in \mathbb{C} : 0 < \operatorname{Im} z < \beta\}$. For a real vector space V, we say that a positive definite function $\psi : \mathbb{R} \to \operatorname{Bil}(V)$ satisfies the *KMS condition* for $\beta > 0$ if ψ extends to a function $\psi : \overline{\mathcal{S}_\beta} \to \operatorname{Bil}(V)$ which is pointwise continuous, pointwise holomorphic on $\mathcal{S}_\beta$, and satisfies

$$\psi(i\beta + t) = \overline{\psi(t)} \quad \text{for } t \in \mathbb{R}. \tag{3}$$

Lemma 2.2. *Suppose that* $\psi : \mathbb{R} \to \operatorname{Bil}(V)$ *satisfies the KMS condition for* $\beta > 0$. *Then*

$$\psi(-\overline{z}) = \psi(z)^* \quad \textit{and} \quad \psi(i\beta + \overline{z}) = \overline{\psi(z)} \qquad \textit{for } z \in \overline{\mathcal{S}_\beta}. \tag{4}$$

The function $\varphi : [0, \beta] \to \operatorname{Bil}(V)$, $\varphi(t) := \psi(it)$ *has hermitian values and satisfies*

$$\varphi(\beta - t) = \overline{\varphi(t)} \quad \textit{for } 0 \leq t \leq \beta. \tag{5}$$

It extends to a unique pointwise continuous symmetric 2β*-periodic function* $\varphi : \mathbb{R} \to \operatorname{Herm}(V)$ *satisfying*

$$\varphi(\beta + t) = \overline{\varphi(t)} \quad \textit{for } t \in \mathbb{R}.$$

Proof. Note that $\psi(-t) = \psi(t)^*$ holds for every positive definite function $\psi : \mathbb{R} \to \operatorname{Bil}(V)$. By analytic continuation (and the Schwarz reflection principle), this leads to the first part of (4). Likewise, condition (3) leads to the second part of (4). This in turn implies (5), and the remainder is clear. ☐

Remark 2.3. Note that (4) implies in particular that $\psi(i\beta/2 + t)$ is real-valued for $t \in \mathbb{R}$ (cf. [Rieffel and van Daele 1977, Proposition 3.5]).

We now introduce standard real subspaces $V \subseteq \mathcal{H}$ and the associated modular objects (Δ, J).

Definition 2.4. A closed real subspace V of a complex Hilbert space $\mathcal{H}$ is said to be *standard* if

$$V \cap iV = \{0\} \quad \text{and} \quad \overline{V + iV} = \mathcal{H}.$$

For every standard real subspace $V \subseteq \mathcal{H}$, we define an unbounded antilinear operator

$$S : \mathcal{D}(S) = V + iV \to \mathcal{H}, \quad S(v + iw) := v - iw, \quad v, w \in V.$$

Then S is closed and has a polar decomposition $S = J\Delta^{1/2}$, where J is an antiunitary involution and Δ a positive selfadjoint operator (cf. [Neeb and Ólafsson 2015b, Lemma 4.2]; see also [Bratteli and Robinson 1979, Proposition 2.5.11; Longo 2008, Proposition 3.3]). We call (Δ, J) the *modular objects* of V.

Remark 2.5. (a) From $S^2 = \mathrm{id}$, it follows that the modular objects (Δ, J) of a standard real subspace satisfy the modular relation

$$J\Delta J = \Delta^{-1}. \tag{6}$$

If, conversely, (Δ, J) is a pair of a positive selfadjoint operator Δ and an antilinear involution J satisfying (6), then $S := J\Delta^{1/2}$ is an unbounded antilinear involution with $\mathcal{D}(S) = \mathcal{D}(\Delta^{1/2})$ whose fixed point space $\mathrm{Fix}(S)$ is a standard real subspace. Thus standard real subspaces are parametrized by pairs (Δ, J) satisfying (6) (cf. [Longo 2008, Proposition 3.2] and [Neeb and Ólafsson 2015b, Lemma 4.4]).

(b) As the unitary one-parameter group Δ^{it} commutes with J and Δ, it leaves the real subspace $V = \mathrm{Fix}(S)$ invariant.

We now come to the proof of Theorem 2.6.

Theorem 2.6 (Characterization of the KMS condition). *Let V be a real vector space and let $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ be a pointwise continuous positive definite function. Then the following are equivalent*:

(i) *ψ satisfies the KMS condition for $\beta > 0$.*

(ii) *There exists a standard real subspace V_1 in a Hilbert space $\mathcal{H}$ and a linear map $j : V \to V_1$ such that*

$$\psi(t)(v, w) = \langle j(v), \Delta^{-it/\beta} j(w)\rangle \quad \textit{for } t \in \mathbb{R}, v, w \in V. \tag{7}$$

(iii) *There exists a $\mathrm{Bil}^+(V)$-valued regular Borel measure μ on $\mathbb{R}$ satisfying*

$$d\mu(-\lambda) = e^{-\beta\lambda} d\overline{\mu}(\lambda),$$

such that

$$\psi(t) = \int_{\mathbb{R}} e^{it\lambda}\, d\mu(\lambda) = \widehat{\mu}(t).$$

If these conditions are satisfied, then the function

$$\psi : \overline{\mathcal{S}_\beta} \to \mathrm{Bil}(V)$$

is pointwise bounded.

Proof. (i) $\Rightarrow$ (ii): From the GNS construction (Proposition A.4), we obtain a continuous unitary representation $(U, \mathcal{H})$ and a linear map $j : V \to \mathcal{H}$ such that

$$\psi(t)(v, w) = \langle j(v), U_t j(w)\rangle \quad \text{for } t \in \mathbb{R},\ v, w \in V.$$

We further assume that the range of the map

$$\zeta : \mathbb{R} \times V \to \mathcal{H}, \quad \zeta(t, v) := U_t j(v)$$

spans a dense subspace. Using Stone's theorem, we write $U_t = e^{-itH}$ for a selfadjoint operator H on $\mathcal{H}$ and consider the positive selfadjoint operator

$$\Delta := e^{\beta H} \quad \text{satisfying } U_t = \Delta^{-it/\beta} \text{ for } t \in \mathbb{R}.$$

With the $B(\mathcal{H})$-valued spectral measure P on $\mathbb{R}$ with $H = \int_{\mathbb{R}} \lambda \, dP(\lambda)$, we thus obtain

$$\psi(t)(v, w) = \langle j(v), e^{-itH} j(w)\rangle = \int_{\mathbb{R}} e^{-it\lambda} \, dP^{j(v),j(w)}(\lambda),$$

where $P^{v,w} = \langle v, P(\,\cdot\,)w\rangle$. The KMS condition for ψ gives that, for each $v \in V$, the function $\psi(t)(v, v)$ extends holomorphically to $\overline{\mathcal{S}_\beta}$, which implies that the integral $\int_{\mathbb{R}} e^{\beta\lambda} \, dP^{j(v),j(v)}(\lambda)$ is finite, and hence that $j(V) \subseteq \mathcal{D}(\Delta^{1/2})$ [Neeb and Ólafsson 2015b, Lemma B.4]. The uniqueness of analytic continuation (Schwarz' principle) now implies

$$\begin{aligned} \psi(x+iy)(v, w) &= \int_{\mathbb{R}} e^{-i(x+iy)\lambda} \, dP^{j(v),j(w)}(\lambda) \\ &= \langle \Delta^{y/2\beta} j(v), \Delta^{-ix/\beta} \Delta^{y/2\beta} j(w)\rangle \end{aligned} \tag{8}$$

for $v, w \in V$ and $0 \le y \le \beta$. Since $\mathcal{D}(\Delta^{1/2})$ is U-invariant, we obtain from the KMS condition,

$$\begin{aligned} \langle \Delta^{1/2}\zeta(t, v), \Delta^{1/2}\zeta(s, w)\rangle = \langle \Delta^{1/2} j(v), \Delta^{1/2} U_{s-t} j(w)\rangle &= \psi(i\beta + s - t)(v, w) \\ = \overline{\psi(s-t)(v, w)} = \overline{\langle \zeta(t, v), \zeta(s, w)\rangle}. \end{aligned}$$

This implies the existence of a unique antilinear isometry $J : \mathcal{H} \to \mathcal{H}$ with

$$J\zeta(t, v) = \Delta^{1/2}\zeta(t, v) \quad \text{for all } t \in \mathbb{R},\ v \in V.$$

Then

$$U_s J\zeta(t, v) = \Delta^{1/2}\zeta(t+s, v) = J\zeta(t+s, v) = JU_s\zeta(t, v) \quad \text{for } t, s \in \mathbb{R},\ v \in V$$

shows that J commutes with every U_t. This implies that $J\Delta^{1/2}J^{-1} = \Delta^{-1/2}$, so

$$\zeta(t, v) = J^{-1}\Delta^{1/2}\zeta(t, v) = \Delta^{-1/2}J^{-1}\zeta(t, v),$$

which in turn implies

$$J\zeta(t, v) = \Delta^{1/2}\zeta(t, v) = J^{-1}\zeta(t, v) \quad \text{for } t \in \mathbb{R},\ v \in V.$$

Since the range of ζ is total, it follows that $J^{-1} = J$, so J is an anti-unitary involution. Therefore (Δ, J) is the modular object of the standard real subspace $V_1 := \mathrm{Fix}(S)$ for the unbounded antilinear involution $S := J\Delta^{1/2}$ (Remark 2.5).

For $v \in V$, we now have $j(v) \in \mathcal{D}(S) = \mathcal{D}(\Delta^{1/2})$ and $Sj(v) = J\Delta^{1/2}j(v) = J^2 j(v) = j(v)$, so that $j(V) \subseteq V_1$. This completes the proof of (ii).

(ii) $\Rightarrow$ (iii): For $v, w \in V$ we have

$$\psi(t)(v, w) = \langle j(v), \Delta^{-it/\beta} j(w)\rangle = \int_{\mathbb{R}} e^{it\lambda} \langle j(v), dP(\lambda) j(w)\rangle,$$

where P is the spectral measure of the selfadjoint operator $L := -\frac{1}{\beta}\log\Delta$ (the

Liouvillian). We therefore consider the $\mathrm{Bil}^+(V)$-valued measure defined by

$$\mu(\,\cdot\,)(v, w) := \langle j(v), P(\,\cdot\,)j(w)\rangle = P^{j(v),j(w)}.$$

It remains to show that $d\mu(-\lambda) = e^{-\beta\lambda} d\overline{\mu}(\lambda)$, which means that $r_*\mu = e_{-\beta}\overline{\mu}$ holds for $r(\lambda) = -\lambda$. To verify this relation, we first observe that $JLJ = -L$ implies that $JPJ = r_*P$. This leads to

$$\begin{aligned}\overline{\mu(\,\cdot\,)(v, w)} &= \langle P(\,\cdot\,)j(w), j(v)\rangle = \langle P(\,\cdot\,)Sj(w), Sj(v)\rangle \\ &= \langle P(\,\cdot\,)J\Delta^{1/2}j(w), J\Delta^{1/2}j(v)\rangle \\ &= \langle JP(\,\cdot\,)J\Delta^{1/2}j(v), \Delta^{1/2}j(w)\rangle = \langle (r_*P)(\,\cdot\,)\Delta^{1/2}j(v), \Delta^{1/2}j(w)\rangle \\ &= e_\beta \cdot \langle (r_*P)(\,\cdot\,)j(v), j(w)\rangle = e_\beta \cdot (r_*\mu)(\,\cdot\,)(v, w).\end{aligned}$$

This implies that $\overline{\mu} = e_\beta \cdot r_*\mu$.

(iii) $\Rightarrow$ (i): Condition (iii) implies that $\psi(0) = \mu(\mathbb{R})$ exists, so that μ is a pointwise finite measure. Further, the relation $r_*\mu = e_{-\beta}\overline{\mu}$ implies that the measure $e_{-\beta}\mu$ is also finite. Therefore the integral

$$\psi(z) := \int_{\mathbb{R}} e^{iz\lambda}\, d\mu(\lambda) \tag{9}$$

exists pointwise and extends ψ to $\overline{\mathcal{S}_\beta}$ in such a way that this extension is pointwise continuous on $\overline{\mathcal{S}_\beta}$ and pointwise holomorphic on the interior. The relation $r_*\mu = e_{-\beta}\overline{\mu}$ further leads to

$$\begin{aligned}\psi(i\beta + t) &= \int_{\mathbb{R}} e^{\lambda(-\beta+it)}\, d\mu(\lambda) = \int_{\mathbb{R}} e_{-\beta}(\lambda)e^{i\lambda t}\, d\mu(\lambda) \\ &= \int_{\mathbb{R}} e^{i\lambda t}\, d(r_*\overline{\mu})(\lambda) = \int_{\mathbb{R}} e^{-i\lambda t}\, d\overline{\mu}(\lambda) = \overline{\psi(t)}.\end{aligned}$$

Therefore ψ satisfies the KMS condition for β.

We finally assume that (i)–(iii) are satisfied and show that ψ is pointwise bounded on $\overline{\mathcal{S}_\beta}$. Since each $\psi(z)$ extends to a sesquilinear form $\psi(z)_{\mathbb{C}}$ on $V_{\mathbb{C}}$, in view of the polarization identity, it suffices to show the boundedness of the functions $z \mapsto \psi(z)_{\mathbb{C}}(v, v)$ for $v \in V_{\mathbb{C}}$. For the positive measure $\mu^{v,v}(E) := \mu(E)_{\mathbb{C}}(v, v)$, we obtain from (9) the estimate

$$|\psi(z)_{\mathbb{C}}(v, v)| \leq \int_{\mathbb{R}} |e^{-i\lambda z}|\, d\mu^{v,v}(\lambda) = \int_{\mathbb{R}} e^{\lambda\,\mathrm{Im}\, z}\, d\mu^{v,v}(\lambda).$$

The convexity of the function on the right, the Laplace transform of the finite positive measure $\mu^{v,v}$, and $\psi(\beta i)(v, v) = \|\Delta^{1/2}j(v)\|^2 < \infty$ now imply the boundedness of $\psi(z)_{\mathbb{C}}(v, v)$. $\square$

Remark 2.7. A special case worth noting arises from a C^*-dynamical system $(\mathcal{A}, \mathbb{R}, \alpha)$ for $V := \mathcal{A}_h := \{A \in \mathcal{A} : A^* = A\}$ and an invariant state ω on $\mathcal{A}$. Such a state is a β-KMS state if and only if

$$\psi : \mathbb{R} \to \mathrm{Bil}(\mathcal{A}_h), \quad \psi(t)(A, B) := \omega(A\alpha_t(B))$$

satisfies the KMS condition for $\beta > 0$ (cf. [Neeb and Ólafsson 2015b, Proposition 5.2; Rieffel and van Daele 1977, Theorem 4.10; Bratteli and Robinson 1981]). If $(\pi_\omega, U^\omega, \mathcal{H}_\omega, \Omega)$ is the corresponding covariant GNS representation of $(\mathcal{A}, \mathbb{R})$,

$$\omega(A) = \langle \Omega, \pi_\omega(A)\Omega\rangle \quad \text{for } A \in \mathcal{A} \qquad \text{and} \qquad U^\omega_t\Omega = \Omega \quad \text{for } t \in \mathbb{R}.$$

Therefore

$$\begin{aligned}\psi(t)(A, B) &= \omega(A\alpha_t(B)) = \langle \Omega, \pi_\omega(A\alpha_t(B))\Omega\rangle \\ &= \langle \Omega, \pi_\omega(A)U^\omega_t\pi_\omega(B)U^\omega_{-t}\Omega\rangle = \langle \pi_\omega(A)\Omega, U^\omega_t\pi_\omega(B)\Omega\rangle\end{aligned}$$

for $A, B \in \mathcal{A}_h$. The corresponding standard real subspace of $\mathcal{H}_\omega$ is $V_1 := \overline{\pi_\omega(\mathcal{A}_h)\Omega}$.

Corollary 2.8. *If $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ satisfies the β-KMS condition, then the kernel*

$$K : \overline{\mathcal{S}_{\beta/2}} \times \overline{\mathcal{S}_{\beta/2}} \to \mathrm{Bil}(V), \qquad K(z, w)(\xi, \eta) := \psi(z - \overline{w})(\xi, \eta) \tag{10}$$

is positive definite.

Proof. This follows immediately from the following relation that we derive from (8):

$$\begin{aligned}K(z, w)(\xi, \eta) &= \psi(z - \overline{w})(\xi, \eta) \\ &= \langle \Delta^{-\frac{i\overline{z}}{\beta}} j(\xi), \Delta^{-\frac{i\overline{w}}{\beta}} j(\eta)\rangle \quad \text{for } \xi, \eta \in V,\ z, w \in \overline{\mathcal{S}_{\beta/2}}.\end{aligned}$$ ☐

Now that we know from Corollary 2.8 that the kernel K in (10) is positive definite, we obtain a corresponding reproducing kernel Hilbert space consisting of functions on $\overline{\mathcal{S}_{\beta/2}} \times V$ which are linear in the second argument and holomorphic on $\mathcal{S}_{\beta/2}$ in the first. We may therefore think of these functions as having values in the algebraic dual space $V^* := \mathrm{Hom}(V, \mathbb{R})$ of V. We write $\mathcal{O}(\overline{\mathcal{S}_{\beta/2}}, V^*)$ for the space of those functions $f : \overline{\mathcal{S}_{\beta/2}} \to V^*$ with the property that, for every $\eta \in V$, the function $z \mapsto f(z)(\eta)$ is continuous on $\overline{\mathcal{S}_{\beta/2}}$ and holomorphic on the open strip $\mathcal{S}_{\beta/2}$.

Proposition 2.9 (Realization of $\mathcal{H}_\psi$ on $\mathcal{O}(\overline{\mathcal{S}_{\beta/2}}, V^*)$). *Assume that $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ satisfies the KMS condition for $\beta > 0$ and let $\psi : \overline{\mathcal{S}_\beta} \to \mathrm{Bil}(V)$ denote the corresponding extension and $\mathcal{H}_\psi \subseteq \mathcal{O}(\overline{\mathcal{S}_{\beta/2}}, V^*)$ denote the Hilbert space with reproducing kernel*

$$K(z, w)(\xi, \eta) := \psi(z - \overline{w})(\xi, \eta) \quad \textit{for } \xi, \eta \in V,$$

i.e.,

$$f(z)(\xi) = \langle K_{z,\xi}, f\rangle \quad \textit{for } f \in \mathcal{H}_\psi, \ \textit{where } K_{z,\xi}(w)(\eta) = \psi(w - \overline{z})(\eta, \xi).$$

Then

$$(U_t^\psi f)(z) := f(z+t), \qquad t \in \mathbb{R}, z \in \overline{\mathcal{S}_{\beta/2}},$$

defines a unitary one-parameter group on $\mathcal{H}_\psi$,

$$j : V \to \mathcal{H}_\psi, \quad j(\eta)(z) := \psi(z)(\,\cdot\,, \eta)$$

is a linear map with U^ψ*-cyclic range, and*

$$\psi(t)(\xi, \eta) = \langle j(\xi), U_t^\psi j(\eta)\rangle \quad \textit{for } t \in \mathbb{R}, \xi, \eta \in V.$$

The anti-unitary involution on $\mathcal{H}_\psi$ *corresponding to the standard real subspace* $V_1 \subseteq \mathcal{H}_\psi$ *from Theorem 2.6 is given by*

$$(J_1 f)(z) := \overline{f\left(\overline{z} + \frac{i\beta}{2}\right)}. \tag{11}$$

Proof. First we recall that the natural reproducing kernel Hilbert space $\mathcal{H}_\psi = \mathcal{H}_K$ is generated by the function $K_{(w,\eta)}$ satisfying

$$\begin{aligned} K_{(w,\eta)}(z)(\xi) = \langle K_{(z,\xi)}, K_{(w,\eta)}\rangle &= K(z,w)(\xi,\eta) \\ &= \psi(z-\overline{w})(\xi,\eta). \end{aligned}$$

As a function of z, the kernel K is continuous on $\overline{\mathcal{S}_{\beta/2}}$ and holomorphic on the interior. Therefore [Neeb 2000, Proposition I.1.9] implies that $\mathcal{H}_\psi$ is a subspace of $\mathcal{O}(\overline{\mathcal{S}_{\beta/2}}, V^*)$, where, for every $f \in \mathcal{H}_\psi$ and $\xi \in V$, we have

$$f(z)(\xi) = \langle K_{(z,\xi)}, f\rangle.$$

That the formula for U_t^ψ defines a unitary one-parameter group on $\mathcal{H}_\psi$ follows directly from the invariance of the kernel K under the action of $\mathbb{R}$ on $\overline{\mathcal{S}_\beta}$ by translation.

Next we observe that

$$\begin{aligned} \langle j(\xi), U_t^\psi j(\eta)\rangle &= \langle K_{(0,\xi)}, U_t^\psi K_{(0,\eta)}\rangle \\ &= \langle K_{(0,\xi)}, K_{(-t,\eta)}\rangle = \psi(t)(\xi,\eta). \end{aligned}$$

To see that $j(V)$ is U^ψ-cyclic, we have to show that the elements $U_t^\psi j(\eta) = K_{(-t,\eta)}$ form a total subset of $\mathcal{H}_\psi$. This means that any $f \in \mathcal{H}_\psi$ with

$$0 = \langle K_{(t,\eta)}, f\rangle = f(t)(\eta)$$

for every $t \in \mathbb{R}$ and $\eta \in V$ vanishes. As the function $t \mapsto f(t)(\eta)$ extends to a continuous function on $\overline{\mathcal{S}_{\beta/2}}$, holomorphic on the interior, it vanishes by the Schwarz reflection principle. Further, η was arbitrary, so $f = 0$ follows.

Now we turn to the involution J_1. As $K_{(w,\eta)}(z) = \psi(z-\overline{w})(\,\cdot\,,\eta)$, the operator J_1 on $\mathcal{O}(\overline{\mathcal{S}_{\beta/2}}, V^*)$, defined by the right hand side of (11) satisfies

$$(J_1 K_{(w,\eta)})(z) = \overline{K_{(w,\eta)}\Big(\overline{z}+\frac{i\beta}{2}\Big)} = \overline{\psi\Big(\overline{z}+\frac{i\beta}{2}-\overline{w}\Big)(\,\cdot\,,\eta)} \tag{12}$$

$$= \psi\Big(i\beta + z - \frac{i\beta}{2} - w\Big)(\,\cdot\,,\eta) = \psi\Big(z+\frac{i\beta}{2}-w\Big)(\,\cdot\,,\eta)$$

$$= K_{(\overline{w}+i\beta/2,\eta)}(z).$$

Here we have used that $\overline{\psi(z)} = \psi(i\beta+\overline{z})$ (Lemma 2.2). From

$$\langle K_{(\overline{w}+i\beta/2,\eta)}, K_{(\overline{z}+i\beta/2,\xi)}\rangle = \overline{K(\overline{z}+i\beta/2, \overline{w}+i\beta/2)(\xi,\eta)} = \overline{\psi(i\beta+\overline{z}-w)(\xi,\eta)}$$

$$= \psi(z-\overline{w})(\xi,\eta) = \langle K_{(z,\xi)}, K_{(w,\eta)}\rangle,$$

it now follows that the operator J_1 in (11) leaves the subspace $\mathcal{H}_\psi$ invariant and defines an antilinear isometry on this space. From the explicit formula it follows that J_1 is an involution. It is also clear that J_1 commutes with the unitary operators $(U_t f)(z) = f(z+t)$.

The relation $U_t K_{(w,\eta)} = K_{(w-t,\eta)}$ leads by analytic continuation to

$$J_1 K_{(0,\eta)} = K_{(i\beta/2,\eta)} = \Delta^{1/2} K_{(0,\eta)}.$$

In the proof of Theorem 2.6, we have seen that, for $\eta \in V$ and $t \in \mathbb{R}$, the anti-unitary involution J corresponding to the associated standard real subspace V_1 satisfies

$$Jj(\eta) = \Delta^{1/2} j(\eta).$$

As both J and J_1 commute with every U_t and the subset $\{U_t j(\eta) : t \in \mathbb{R}, \eta \in V\}$ is total in $\mathcal{H}_\psi$, we conclude that $J_1 = J$. ☐

3. Form-valued reflection positive functions

In this section we discuss reflection positivity on the level of form-valued positive definite functions. This is particularly well adapted to reflection positive Hilbert spaces $(\mathcal{E}, \mathcal{E}_+, \theta)$, for which $\mathcal{E}_+$ is generated by elements of the form $U_g^{-1} j(v)$, where g is contained in a certain subset $G_+ \subseteq G$ which is not necessarily a subsemigroup, and $j : V \to \mathcal{H}$ is a linear map for which $U_G j(V)$ spans a dense subspace of $\mathcal{E}$. In particular, we present a version of the GNS construction in this context (Proposition 3.9) and we briefly discuss it more specifically for the trivial group $G = \{\mathbf{1}\}$ (Section 3B) and the 2-element group (Section 3C). The latter case shows explicitly that the cone of reflection positive functions does not adapt naturally to the decomposition into even and odd functions. Put differently, if a reflection positive representation decomposes into two subrepresentations, the summands need not be reflection positive (see also [Neeb and Ólafsson 2014]).

3A. ***Reflection positivity and form-valued functions.*** Let (G, τ) be a *symmetric Lie group*, i.e., G is a Lie group and $\tau \in \operatorname{Aut}(G)$ with $\tau^2 = \mathrm{id}_G$. In the following we write $G_\tau := G \rtimes \{\mathbf{1}, \tau\}$ and $g^\sharp := \tau(g)^{-1}$ [Neeb and Ólafsson 2014]. In this section we introduce reflection positive functions on G_τ with values in $\operatorname{Bil}(V)$ for a real vector space V.

Definition 3.1. Let $\mathcal{E}$ be a Hilbert space and let $\theta \in \mathrm{U}(\mathcal{E})$ be an involution. A closed subspace $\mathcal{E}_+ \subseteq \mathcal{E}$ is called *θ-positive* if $\langle \theta v, v\rangle \geq 0$ for $v \in \mathcal{E}_+$. We then call the triple $(\mathcal{E}, \mathcal{E}_+, \theta)$ a *reflection positive Hilbert space*. For a reflection positive Hilbert space we put $\mathcal{N} := \{v \in \mathcal{E}_+ : \langle \theta v, v\rangle = 0\}$ and write $q : \mathcal{E}_+ \to \mathcal{E}_+/\mathcal{N}, v \mapsto \widehat{v} = q(v)$ for the quotient map and $\widehat{\mathcal{E}}$ for the Hilbert completion of $\mathcal{E}_+/\mathcal{N}$ with respect to the norm $\|\widehat{v}\|_{\widehat{\mathcal{E}}} := \|\widehat{v}\| := \sqrt{\langle \theta v, v\rangle}$.

Example 3.2. Suppose that $K \colon X \times X \to \mathbb{C}$ is a positive definite kernel on the set X and that $\tau \colon X \to X$ is an involution leaving K invariant. We further assume that $X_+ \subseteq X$ is a subset with the property that the kernel $K^\tau(x, y) := K(\tau x, y)$ is also positive definite on X_+.

Let $\mathcal{E} := \mathcal{H}_K \subseteq \mathbb{C}^X$ denote the corresponding reproducing kernel Hilbert space generated by elements $(K_x)_{x\in X}$ with $\langle K_x, K_y\rangle = K(x, y)$. Then the closed subspace $\mathcal{E}_+ \subseteq \mathcal{E}$ generated by $(K_x)_{x\in X_+}$ is θ-positive for $(\theta f)(x) := f(\tau x)$. We thus obtain a reflection positive Hilbert space $(\mathcal{E}, \mathcal{E}_+, \theta)$. We call such kernels K *reflection positive* with respect to (X, X_+, τ).

Definition 3.3. Let $G_+ \subseteq G$ be a subset. Let V be a real vector space and let $j : V \to \mathcal{H}$ be a linear map whose range is cyclic for the unitary representation $(U, \mathcal{E})$ of G_τ. Then we say that $(U, \mathcal{E}, j, V)$ is *reflection positive with respect to $G_+ \subseteq G$* if, for $\mathcal{E}_+ := \overline{\operatorname{span} U_{G_+}^{-1} j(V)}$, the triple $(\mathcal{E}, \mathcal{E}_+, U_\tau)$ is a reflection positive Hilbert space.

Definition 3.4. Let V be a real vector space. We call a function $\varphi : G_\tau \to \operatorname{Bil}(V)$ *reflection positive with respect to the subset G_+ of G* if

(RP1) φ is positive definite and

(RP2) the kernel $(s, t) \mapsto \varphi(st^\sharp\tau) = \varphi(s\tau t^{-1})$ is positive definite on G_+.

Remark 3.5. Let $\varphi : G_\tau \to \operatorname{Bil}(V)$ be a positive definite function, so that the kernel $K((x, v), (y, w)) := \varphi(xy^{-1})(v, w)$ on $G_\tau \times V$ is positive definite. The involution τ acts on $G_\tau \times V$ by $\tau.(g, v) := (g\tau, v)$ and the corresponding kernel $K^\tau((x, v), (y, w)) := K((x\tau, v), (y, w)) = \varphi(x\tau y^{-1})(v, w)$ is positive definite on $G_+ \times V$ if and only if φ is reflection positive in the sense of Example 3.2.

Positive definite functions on G extend canonically to G_τ if they are τ-invariant:

Lemma 3.6. *Let V be a real vector space and let (G, τ) be a symmetric Lie group. Then the following assertions hold*:

(i) *If* $\varphi : G \to \mathrm{Bil}(V)$ *is a positive definite function which is* τ*-invariant in the sense that* $\varphi \circ \tau = \varphi$, *then* $\widehat{\varphi}(g, \tau) := \varphi(g)$ *defines an extension to* G_τ *which is positive definite and* τ*-biinvariant.*

(ii) *Let* $(U, \mathcal{H})$ *be a unitary representation of* G_τ, *let* $\theta := U_\tau$, *let* $j : V \to \mathcal{H}$ *be a linear map, and let* $\varphi(g)(v, w) = \langle j(v), U_g j(w)\rangle$ *be the corresponding* $\mathrm{Bil}(V)$*-valued positive definite function. Then the following are equivalent*:

  (a) $\theta j(v) = j(v)$ *for every* $v \in V$.

  (b) φ *is* τ*-biinvariant.*

  (c) φ *is left* τ*-invariant.*

Proof. (i) From the GNS construction (Proposition A.4), we obtain a continuous unitary representation $(U, \mathcal{H})$ of G and a linear map $j : V \to \mathcal{H}$ such that

$$\varphi(g)(v, w) = \langle j(v), U_g j(w)\rangle \quad \text{for } g \in G,\ v, w \in V.$$

As $\varphi(g)(v, w) = \varphi(\tau(g))(v, w)$, the uniqueness in the GNS construction provide a unitary operator $\theta : \mathcal{H} \to \mathcal{H}$ with

$$\theta U_g j(v) = U_{\tau(g)} j(v) \quad \text{for } g \in G,\ v \in V.$$

Note that θ fixes each $j(v)$. Therefore $U_\tau := \theta$ defines an extension of G to a unitary representation of G_τ on $\mathcal{H}$. Hence $\psi(g)(v, w) = \langle j(v), U_g j(w)\rangle$ defines a positive definite $\mathrm{Bil}(V)$-valued function on G_τ which satisfies

$$\begin{aligned}\psi(g, \tau)(v, w) &= \langle \theta j(v), U_g j(w)\rangle \\ &= \langle j(v), U_g j(w)\rangle = \varphi(g)(v, w) \quad \text{for } g \in G, v, w \in V.\end{aligned}$$

(ii) Clearly, (a) $\Rightarrow$ (b) $\Rightarrow$ (c). It remains to show that (c) implies (a). So we assume that $\varphi(\tau g) = \varphi(g)$ for $g \in G_\tau$. This means that, for every $v, w \in V$, we have

$$\begin{aligned}\langle j(v), U_g j(w)\rangle &= \varphi(g)(v, w) = \varphi(\tau g)(v, w) \\ &= \langle j(v), \theta U_g j(w)\rangle = \langle \theta j(v), U_g j(w)\rangle.\end{aligned}$$

Since $U_{G_\tau} j(V)$ is total in $\mathcal{H}$, this implies that $\theta j(v) = j(v)$ for every $v \in V$. ☐

Remark 3.7. (a) As G_τ consists of the two cosets G and $G\tau = G \times \{\tau\}$, every function φ on G_τ is given by a pair of functions on G:

$$\varphi_\pm : G \to \mathrm{Bil}(V), \quad \varphi_+(g) := \varphi(g, \mathbf{1}), \qquad \varphi_-(g) := \varphi(g, \tau).$$

Then (RP2) is a condition on φ_- alone, and (RP1) is a condition on the pair (φ_+, φ_-).

(b) If φ is reflection positive, then its complex conjugate $\overline{\varphi}$ is also reflection positive because the convex cone of positive definite kernels on a set is stable under complex conjugation. This implies in particular that $\mathrm{Re}\,\varphi = \frac{1}{2}(\varphi + \overline{\varphi})$ is reflection positive (cf. Theorem A.13).

The following lemma provides a tool which is sometimes convenient to verify positive definiteness of a function on the extended group G_τ in terms of a kernel on the original group G.

Lemma 3.8. *Every function $\varphi : G_\tau \to B(V)$ leads to a $M_2(B(V))$-valued kernel*

$$Q : G \times G \to M_2(B(V)) \cong B(V \oplus V), \quad Q(g,h) = \begin{pmatrix} \varphi(gh^{-1}) & \varphi(g\tau h^{-1}) \\ \varphi(g\tau h^{-1}) & \varphi(gh^{-1}) \end{pmatrix},$$

and the function φ on G_τ is positive definite if and only if Q is positive definite.

Proof. That Q is positive definite is equivalent to the existence of a Hilbert space $\mathcal{H}$ and a map

$$\ell : G \to B(\mathcal{H}, V \oplus V) \cong B(\mathcal{H}, V)^{\oplus 2} \quad \text{with } Q(x,y) = \ell(x)\ell(y)^* \text{ for } x, y \in G$$

(cf. [Neeb 2000, Theorem I.1.4]). If ℓ is such a map, then it can be written as $\ell(x) = (\ell_1(x), \ell_2(x))$ with $\ell_j(x) \in B(\mathcal{H}, V)$. We thus obtain

$$Q(x,y) = \ell(x)\ell(y)^* = \begin{pmatrix} \ell_1(x)\ell_1(y)^* & \ell_1(x)\ell_2(y)^* \\ \ell_2(x)\ell_1(y)^* & \ell_2(x)\ell_2(y)^* \end{pmatrix}$$

and thus

$$\ell_1(x)\ell_1(y)^* = \ell_2(x)\ell_2(y)^* \quad \text{and} \quad \ell_1(x)\ell_2(y)^* = \ell_2(x)\ell_1(y)^*.$$

Therefore

$$j : G_\tau \to B(\mathcal{H}, V), \qquad j(x, \mathbf{1}) := \ell_1(x), \quad j(x,\tau) := \ell_2(x),$$

satisfies

$$j(x,\mathbf{1})j(y,\mathbf{1})^* = \ell_1(x)\ell_1(y)^* = \varphi(xy^{-1}),$$
$$j(x,\tau)j(y,\tau)^* = \ell_2(x)\ell_2(y)^* = \varphi(xy^{-1})$$

and

$$j(x,\mathbf{1})j(y,\tau)^* = \ell_1(x)\ell_2(y)^* = \varphi(x\tau y^{-1}),$$
$$j(x,\tau)j(y,\mathbf{1})^* = \ell_2(x)\ell_1(y)^* = \varphi(xy^{-1}).$$

We therefore have $\varphi(xy^{-1}) = j(x)j(y)^*$ for $x, y \in G_\tau$, and thus φ is positive definite.

If, conversely, φ is positive definite and $j : G_\tau \to B(\mathcal{H}, V)$ is such that $\varphi(x^{-1}y) = j(x)j(y)^*$ for $x, y \in G_\tau$, then $\ell(x) := (j(x,\mathbf{1}), j(x,\tau)) \in B(\mathcal{H}, V \oplus V)$ defines a map with $Q(x,y) = \ell(x)\ell(y)^*$ for $x, y \in G$. □

Proposition 3.9 (GNS construction for reflection positive functions). *Let V be a real vector space, let $(U, \mathcal{E})$ be a unitary representation of G_τ and put $\theta := U_\tau$. Then the following assertions hold:*

(i) *If* $(U, \mathcal{H}, j, V)$ *is reflection positive with respect to* G_+, *then*

$$\varphi(g)(v, w) := \langle j(v), U_g j(w)\rangle, \qquad g \in G_\tau,\ v, w \in V,$$

is a reflection positive $\mathrm{Bil}(V)$*-valued function.*

(ii) *If* $\varphi : G_\tau \to \mathrm{Bil}(V)$ *is a reflection positive function with respect to* G_+, *then the corresponding GNS representation* $(U^\varphi, \mathcal{H}_\varphi, j, V)$ *is a reflection positive representation, where* $\mathcal{H}_\varphi \subseteq \mathbb{C}^{G_\tau \times V}$ *is the Hilbert subspace with reproducing kernel* $K((x, v), (y, w)) := \varphi(xy^{-1})(v, w)$ *on which* G_τ *acts by*

$$(U^\varphi_g f)(x, v) := f(xg, v).$$

Proof. (i) For $s, t \in G_+$, we have

$$\begin{aligned}\varphi(s\tau t^{-1})(v, w) &= \langle j(v), U_{s\tau t^{-1}} j(w)\rangle = \langle U_{s^{-1}} j(v), U_\tau U_{t^{-1}} j(w)\rangle \\ &= \langle \theta U_{s^{-1}} j(v), U_{t^{-1}} j(w)\rangle,\end{aligned}$$

so that the kernel $(\varphi(s\tau t^{-1}))_{s,t\in G_+}$ is positive definite.

(ii) Recall the relation $\varphi(g)(v, w) = \langle j(v), U_g j(w)\rangle$ for $g \in G$, $v, w \in V$ from Proposition A.4. Moreover, $(\theta f)(x, v) = f(x\tau, v)$, and

$$\langle \theta U^\varphi_{s^{-1}} j(v), U^\varphi_{t^{-1}} j(w)\rangle = \langle j(v), U^\varphi_{s\tau t^{-1}} j(w)\rangle = \varphi(s\tau t^{-1})(v, w),$$

so the positive definiteness of the kernel $(\varphi(s\tau t^{-1}))_{s,t\in G_+}$ implies that we obtain, with $\mathcal{E} = \mathcal{H}_\varphi$ and $\mathcal{E}_+ := \overline{\mathrm{span}\,(U^\varphi_{G_+})^{-1} j(V)}$, a reflection positive Hilbert space $(\mathcal{E}, \mathcal{E}_+, \theta)$. □

3B. *Reflection positivity for the trivial group.* In this short section we discuss the case of the 2-element group $T = \{\mathbf{1}, \tau\}$ in some detail. It corresponds to G_τ where $G = \{\mathbf{1}\}$ is trivial, but it already demonstrates how the intricate structure of a reflection positive Hilbert space $(\mathcal{E}, \mathcal{E}_+, \theta)$ can be encoded in terms of positive definite functions on T.

A unitary representation $(U, \mathcal{E})$ of T is nothing but the specification of a unitary operator $\theta = U_\tau$ on $\mathcal{E}$. We write $\mathcal{E} = \mathcal{E}^1 \oplus \mathcal{E}^{-1}$ for the eigenspace decomposition of $\mathcal{E}$ under θ and $p^{\pm 1} : \mathcal{E} \to \mathcal{E}^{\pm 1}$ for the orthogonal projections.

Suppose, in addition, that V is a real or complex Hilbert space and that $j : V \to \mathcal{E}$ is a continuous linear map whose range generates $\mathcal{E}$ under the representation U, i.e., the projections $p^{\pm 1}(j(V)) \subseteq \mathcal{E}^{\pm 1}$ are dense subspaces. In view of the GNS construction, the data $(\mathcal{E}, U, j, V)$ is encoded in the operator-valued positive definite function

$$\varphi : T \to B(V), \quad \varphi(g) = j^* U_g j.$$

For a function $\varphi : T \to B(V)$, let $A := \varphi(\mathbf{1})$ and $B := \varphi(\tau)$. Then φ is positive definite if and only if $A = A^* \geq 0,\ B = B^*$, and the operator matrix

$$\begin{pmatrix} \varphi(\mathbf{1}) & \varphi(\tau) \\ \varphi(\tau) & \varphi(\mathbf{1}) \end{pmatrix} = \begin{pmatrix} A & B \\ B & A \end{pmatrix} \in M_2(B(V)) \cong B(V \oplus V)$$

defines a positive operator (Lemma 3.8 and [Neeb 2000, Remark I.1.3]). This is equivalent to

$$|\langle Bv, w\rangle|^2 \leq \langle Av, v\rangle\langle Aw, w\rangle \quad \text{for } v, w \in V \tag{13}$$

(cf. Corollary A.9). Note that (13) holds if $A = \mathbf{1}$ and $\|B\| \leq 1$. If, more generally, A is invertible, then (13) is equivalent to $\|A^{-1/2}BA^{-1/2}\| \leq \mathbf{1}$. Here $A = j^*j$ basically encodes how V is mapped into $\mathcal{E}$ and B encodes the unitary involution θ.

The function φ is reflection positive with respect to $G_+ = \{\mathbf{1}\}$ if and only if $B = \varphi(\tau) \geq 0$, which means that $j(V)$ is θ-positive. In this sense reflection positive functions on T encode reflection positive Hilbert spaces $(\mathcal{E}, \mathcal{E}_+, \theta)$ by $\theta = U_\tau$ and $\mathcal{E}_+ := \overline{j(V)}$. A pair (A, B) of hermitian operators on V corresponds to a reflection positive function $\varphi : T \to B(V)$ if and only if $0 \leq B \leq A$. By the Cauchy–Schwarz inequality, this is equivalent to (13) if A and B are positive operators. This shows that

$$\varphi = \varphi_0 + \varphi_1 \quad \text{with } \varphi_0(\mathbf{1}) = A - B,\ \varphi_0(\tau) = 0 \text{ and } \varphi_1(\mathbf{1}) = \varphi_1(\tau) = B,$$

where both functions φ_0 and φ_1 are reflection positive. The function φ_0 corresponds to the case where $\mathcal{E}_+ \perp \theta\mathcal{E}_+$, so that $\widehat{\mathcal{E}} = \{0\}$, and the constant function φ_1 corresponds to the trivial representation of T, and hence to $\theta = \mathbf{1}$, which means that $q : \mathcal{E}_+ \to \widehat{\mathcal{E}}$ is isometric.

Replacing V by $\mathcal{E}_+$, we see that reflection positive functions $\varphi : T \to B(\mathcal{E}_+)$ with $\varphi(\mathbf{1}) = \mathbf{1}$ encode reflection positive Hilbert spaces $(\mathcal{E}, \mathcal{E}_+, \theta)$ for which $p^{\pm 1}(\mathcal{E}_+)$ is dense in $\mathcal{E}^{\pm 1}$. By (13), these configurations are parametrized by the hermitian contractions $B = \varphi(\tau)$ on $\mathcal{E}_+$. For $v, w \in \mathcal{E}_+$, we then have

$$\langle v, \theta w\rangle = \langle v, Bw\rangle.$$

Therefore the 1-eigenspace $\ker(B - \mathbf{1})$ corresponds to the maximal subspace in $\mathcal{E}_+$ on which the map $q : \mathcal{E}_+ \to \widehat{\mathcal{E}}$ is isometric. We also observe that $\ker B = \ker q$. In this sense the operator B describes how $\widehat{\mathcal{E}}$ is obtained from the Hilbert space $\mathcal{E}_+$.

Remark 3.10. Suppose that θ is a unitary involution on $\mathcal{E}$ with the eigenspaces $\mathcal{E}^{\pm 1}$. If $\mathcal{K} \subseteq \mathcal{E}$ is a θ-positive subspace, then clearly $\mathcal{K} \cap \mathcal{E}^{-1} = \{0\}$ and this implies that $\mathcal{K}$ is the graph $\Gamma(Z)$ of the operator

$$Z : \mathcal{D}(Z) := \{v_+ \in \mathcal{E}^1 : (\exists v_- \in \mathcal{E}^{-1})\ (v_+, v_-) \in \mathcal{K}\} \to \mathcal{E}^{-1}, \quad v_+ \mapsto v_-.$$

That $\Gamma(Z)$ is a θ-positive subspace is equivalent to $\|Z\| \leq 1$. Therefore the closedness of $\mathcal{K}$ shows that $\mathcal{D}(Z)$ is a closed subspace of $\mathcal{E}^1$ (cf. [Jorgensen 2002, Lemma 5.1]). If $p^1(\mathcal{K}) = \mathcal{D}(Z)$ is dense in $\widehat{\mathcal{E}}$, the closedness of $\mathcal{D}(Z)$ implies that $Z \in B(\mathcal{E}^1, \mathcal{E}^{-1})$. The density of $p^{-1}(\mathcal{K}) = Z(\mathcal{E}^1)$ is equivalent to Z having dense range.

From this perspective, we can also generate the configuration $(\mathcal{E}, \mathcal{E}_+, \theta)$ in terms of $\mathcal{E}^1$. Then $j(v) = (v, Zv) \in \mathcal{E}^1 \oplus \mathcal{E}^{-1}$ defines a linear map $j : \mathcal{E}^1 \to \mathcal{E}$ whose range is $\mathcal{K}$. The corresponding $B(\mathcal{E}^1)$-valued positive definite function on T is given by

$$\psi(\mathbf{1}) = j^*j = \mathbf{1} + Z^*Z \quad \text{and} \quad \psi(\tau) = j^*\theta j = \mathbf{1} - Z^*Z.$$

The polar decomposition of $j : \mathcal{E}^1 \to \mathcal{K}$ takes the form

$$j = U\sqrt{j^*j} = U\sqrt{\mathbf{1} + Z^*Z},$$

where $U : \mathcal{E}^1 \to \mathcal{K}$ is unitary. Therefore the corresponding $B(\mathcal{K})$-valued positive definite function on T is given by

$$\varphi(\mathbf{1}) = \mathbf{1} \quad \text{and} \quad \varphi(\tau) = U\frac{\mathbf{1} - Z^*Z}{\mathbf{1} + Z^*Z}U^{-1}$$

because $j^*\varphi(\tau)j = j^*\theta j = \mathbf{1} - Z^*Z$ implies

$$\begin{aligned}\varphi(\tau) &= (j^*)^{-1}(\mathbf{1} - Z^*Z)j^{-1} = U(\mathbf{1} + Z^*Z)^{-1/2}(\mathbf{1} - Z^*Z)(\mathbf{1} + Z^*Z)^{-1/2}U^{-1} \\ &= U\frac{\mathbf{1} - Z^*Z}{\mathbf{1} + Z^*Z}U^{-1}.\end{aligned}$$

Relating this to the preceding discussion, we see that $U \ker Z \subseteq \mathcal{E}_+$ is the maximal subspace on which q is isometric and

$$U\{v \in \mathcal{E}^1 : \|Zv\| = \|v\|\} = U \ker(\mathbf{1} - Z^*Z) = \ker q.$$

In particular, q is injective if and only if Z is a strict contraction.

3C. *Reflection positivity for the 2-element group.* In this subsection, we take a closer look at the 2-element group $G = \{\mathbf{1}, \sigma\}$ because it nicely illustrates that if a reflection positive representation decomposes into two subrepresentations, then the summands need not be reflection positive (see also [Neeb and Ólafsson 2014]). On the level of positive definite functions, this is reflected in the fact that the cone of reflection positive functions does not adapt to the decomposition into even and odd functions.

We consider the 2-element group $G := \{\mathbf{1}, \sigma\}$, which leads to the Klein-4-group

$$G_\tau := G \rtimes \{\mathbf{1}, \tau\} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}.$$

We consider reflection positivity with respect to the subset $G_+ := \{\mathbf{1}\}$.

Any unitary representation $(U, \mathcal{E})$ of G_τ decomposes into four eigenspaces

$$\mathcal{E} = \mathcal{E}^{1,1} \oplus \mathcal{E}^{-1,1} \oplus \mathcal{E}^{1,-1} \oplus \mathcal{E}^{-1,-1}, \qquad \mathcal{E}^{\varepsilon_1,\varepsilon_2} = \{v \in \mathcal{E} : U_\sigma v = \varepsilon_1 v, U_\tau v = \varepsilon_2 v\},$$

and for $\theta := U_\tau$, the subspace $\mathcal{E}^\theta = \mathcal{E}^{1,1} \oplus \mathcal{E}^{-1,1}$ is U_σ-invariant. For $v = (a, b, c, d)$, we then have

$$U_\tau v = (a, b, -c, -d) \quad \text{and} \quad U_\sigma v = (a, -b, c, -d).$$

Assume $\mathcal{E}_+ = \mathbb{C}v$ for a single vector v. Then reflection positivity corresponds to

$$\langle v, \theta v\rangle = |a|^2 + |b|^2 - |c|^2 - |d|^2 \geq 0.$$

With respect to U_σ, we have

$$v = v_1 + v_{-1} = (a, 0, c, 0) + (0, b, 0, d)$$

and

$$\langle U_\sigma v, \theta U_\sigma v\rangle = \langle v, \theta v\rangle \geq 0 \quad \text{and} \quad \langle U_\sigma v, \theta v\rangle = |a|^2 - |b|^2 - |c|^2 + |d|^2.$$

Therefore the subspace $\mathbb{C}v + \mathbb{C}U_\sigma v$ is θ-positive if and only if

$$\pm(|a|^2 - |b|^2 - |c|^2 + |d|^2) \leq |a|^2 + |b|^2 - |c|^2 - |d|^2,$$

which is equivalent to

$$|d| \leq |b| \quad \text{and} \quad |c| \leq |a|.$$

Clearly, these two conditions are strictly stronger than the θ-positivity of $\mathbb{C}v$.

For the corresponding positive definite function $f(g) = \langle v, U_g v\rangle$ we have

$$f(\mathbf{1}) = |a|^2 + |b|^2 + |c|^2 + |d|^2, \qquad f(\tau) = |a|^2 + |b|^2 - |c|^2 - |d|^2,$$
$$f(\sigma) = |a|^2 - |b|^2 + |c|^2 - |d|^2, \qquad f(\sigma\tau) = |a|^2 - |b|^2 - |c|^2 + |d|^2.$$

Decomposing $f = f_1 + f_{-1}$ with respect to the left action of σ, we obtain

$$f_1(\mathbf{1}) = f_1(\sigma) = |a|^2 + |c|^2, \qquad f_1(\tau) = f_1(\sigma\tau) = |a|^2 - |c|^2$$

and

$$f_{-1}(\mathbf{1}) = -f_1(\sigma) = |b|^2 + |d|^2, \qquad f_{-1}(\tau) = -f_{-1}(\sigma\tau) = |b|^2 - |d|^2.$$

Both functions $f_{\pm 1}(g) = \langle v_{\pm 1}, U_g v_{\pm 1}\rangle$ are positive definite, but they are reflection positive if and only if $|c| \leq |a|$ and $|d| \leq |b|$.

Note that, even for $U_\sigma = \mathbf{1}$ and $U_\sigma = -\mathbf{1}$, there exist nontrivial reflection positive representations with $\langle v, \theta v\rangle > 0$.

4. Reflection positive functions and KMS conditions

In this section we build the bridge from positive definite functions $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ satisfying the KMS condition for $\beta > 0$ to reflection positive functions on the group $\mathbb{T}_{2\beta,\tau} \cong \mathrm{O}_2(\mathbb{R})$. We have already seen in Lemma 2.2 that analytic continuation leads to a 2β-periodic function $\varphi : \mathbb{R} \to \mathrm{Bil}(V)$ satisfying $\varphi(t + \beta) = \overline{\varphi(t)}$ for $t \in \mathbb{R}$

and $\varphi(t) = \psi(it)$ for $0 \leq t \leq \beta$. In this section we show the existence of a positive definite function $f : \mathbb{R}_\tau \to \mathrm{Bil}(V)$ with $f(t, \tau) = \varphi(t)$ for $t \in \mathbb{R}$. By construction, f is then reflection positive with respect to the interval $[0, \beta/2] = G_+ \subseteq G = \mathbb{R}$ in the sense of Definition 3.4.

Since we can build on Theorem 2.6, our first goal is to express, for a standard real subspace $V \subseteq \mathcal{H}$, the $\mathrm{Bil}(V)$-valued function

$$\begin{aligned} &\varphi : [0, \beta] \to \mathrm{Bil}(V), \\ &\varphi(t)(v, w) := \psi(it)(v, w) = \langle \Delta^{t/2\beta} v, \Delta^{t/2\beta} w\rangle \quad \text{for } v, w \in V, 0 \leq t \leq \beta \end{aligned} \tag{14}$$

from (8) in the proof of Theorem 2.6 as a $B(V_\mathbb{C})$-valued function. To this end, we shall need the description of V in terms of a skew-symmetric strict contraction C on V (Lemma B.9), and this leads to a quite explicit description of φ that we then use to prove our main theorem.

4A. ***From form-valued to operator-valued functions.*** In the following it will be more convenient to work with operator-valued functions instead of form-valued ones. The translation is achieved by the following lemma. For its formulation, we recall the polar decomposition of bounded skew-symmetric operators on real Hilbert spaces.

Remark 4.1. (polar decomposition of skew-symmetric operators) Let $D^\top = -D$ be an injective skew-symmetric operator on the real Hilbert space V and let $D = I|D|$ be its polar decomposition. Then $\mathrm{im}(D)$ is dense because D is injective, and therefore I defines an isometry $V \to V$. From

$$I|D| = D = -D^\top = -|D|I^{-1} = -I^{-1}(I|D|I^{-1})$$

it follows that $I^2 = -\mathbf{1}$, i.e., that I is a complex structure and that $|D|$ commutes with I.

Lemma 4.2. *Let $V \subseteq \mathcal{H}$ be a standard real subspace with modular objects (Δ, J), let $\widehat{C} := i\frac{\Delta - \mathbf{1}}{\Delta + \mathbf{1}}$, and let $C := \widehat{C}|_V \in B(V)$ be the skew-symmetric strict contraction from Lemma B.9. We assume that* $\ker C = \{0\}$, *so the polar decomposition $C = I|C|$ defines a complex structure I on V. Consider the skew-symmetric operator*

$$D := \log\left(\frac{\mathbf{1} - |C|}{\mathbf{1} + |C|}\right) I.$$

Then the function $\varphi(t)(v, w) = \langle \Delta^{t/2} v, \Delta^{t/2} w\rangle$ from (14) has the form

$$\varphi(t)(v, w) = \langle v, \widetilde{\varphi}(t) w\rangle_{V_\mathbb{C}} \quad \textit{for } t \in [0, 1],\ v, w \in V_\mathbb{C}, \tag{15}$$

where the function $\widetilde{\varphi} : [0, 1] \to B(V_\mathbb{C})$ is given by

$$\widetilde{\varphi}(t) = (\mathbf{1} + iC)^{1-t}(\mathbf{1} - iC)^t = \frac{e^{-t|D|} + e^{-(1-t)|D|}}{\mathbf{1} + e^{-|D|}} + iI\frac{e^{-t|D|} - e^{-(1-t)|D|}}{\mathbf{1} + e^{-|D|}}.$$

Note that $\widetilde{\varphi}(0) = \mathbf{1} + iC \neq \mathbf{1}$ if $C \neq 0$.

Proof. Since C is a skew-symmetric contraction on V, the operators $\mathbf{1} \pm iC$ on $V_{\mathbb{C}}$ are symmetric, so that we obtain a function

$$\widetilde{\varphi} : [0, 1] \to B(V_{\mathbb{C}}), \qquad \widetilde{\varphi}(t) := (\mathbf{1} + iC)^{1-t}(\mathbf{1} - iC)^t, \quad 0 \leq t \leq 1.$$

Therefore both sides of (15) are defined, and we have to show that

$$\langle v, \widetilde{\varphi}(t)w\rangle_{V_{\mathbb{C}}} = \langle \Delta^{t/2}v, \Delta^{t/2}w\rangle \quad \text{for } v, w \in V_{\mathbb{C}}. \tag{16}$$

For the skew-hermitian contraction $\widehat{C}$ on $\mathcal{H}$, we likewise obtain bounded operators

$$\widehat{\varphi}(t) := (\mathbf{1} + i\widehat{C})^{1-t}(\mathbf{1} - i\widehat{C})^t, \quad 0 \leq t \leq 1,$$

and the continuity of the inclusion $V_{\mathbb{C}} \hookrightarrow \mathcal{H}$ implies that

$$\widehat{\varphi}(t)|_{V_{\mathbb{C}}} = \widetilde{\varphi}(t) : V_{\mathbb{C}} \to V_{\mathbb{C}}.$$

From the relation

$$\Delta = \frac{\mathbf{1} - i\widehat{C}}{\mathbf{1} + i\widehat{C}},$$

we further obtain the identity

$$\widehat{\varphi}(t) = (\mathbf{1} + i\widehat{C})\Delta^t$$

of selfadjoint operators on $\mathcal{H}$. Let $V'_{\mathbb{C}}$ denote the domain of the (possibly) unbounded selfadjoint operator $\frac{\mathbf{1}-iC}{\mathbf{1}+iC}$ on $V_{\mathbb{C}}$. Then, for $0 \leq t \leq 1$, $V'_{\mathbb{C}}$ is a dense subspace which is contained in the domain of $\big(\frac{\mathbf{1}-iC}{\mathbf{1}+iC}\big)^t$. For $w \in V'_{\mathbb{C}}$, we have

$$\widetilde{\varphi}(t)w = (\mathbf{1} + iC)\left(\frac{\mathbf{1} - iC}{\mathbf{1} + iC}\right)^t w \quad \text{for } 0 \leq t \leq 1.$$

For $v \in V_{\mathbb{C}}$ and $\widetilde{w} := \big(\frac{\mathbf{1}-iC}{\mathbf{1}+iC}\big)^t w$ we now obtain with (39) from Lemma B.9 the relation

$$\begin{aligned}\langle v, \widetilde{\varphi}(t)w\rangle_{V_{\mathbb{C}}} &= \langle v, (\mathbf{1} + iC)\widetilde{w}\rangle_{V_{\mathbb{C}}} = \langle v, \widetilde{w}\rangle_{\mathcal{H}} = \left\langle v, \left(\frac{\mathbf{1} - iC}{\mathbf{1} + iC}\right)^t w\right\rangle_{\mathcal{H}} \\ &= \left\langle v, \left(\frac{\mathbf{1} - i\widehat{C}}{\mathbf{1} + i\widehat{C}}\right)^t w\right\rangle_{\mathcal{H}} = \langle v, \Delta^t w\rangle_{\mathcal{H}} = \langle \Delta^{t/2}v, \Delta^{t/2}w\rangle_{\mathcal{H}}.\end{aligned}$$

Since both sides of (16) define continuous hermitian forms on $V_{\mathbb{C}}$ and the preceding calculation shows that equality holds on a dense subspace, we obtain (16) for all $v, w \in V_{\mathbb{C}}$.

Next we observe that the polar decomposition of D is given by

$$D = -I|D| \quad \text{and} \quad |D| = \log\left(\frac{\mathbf{1} + |C|}{\mathbf{1} - |C|}\right).$$

The operator $|D|$ satisfies

$$e^{\mp|D|} = \frac{\mathbf{1} \mp |C|}{\mathbf{1} \pm |C|} \quad \text{and} \quad 1 + e^{\mp|D|} = \frac{2}{1 \pm |C|}. \tag{17}$$

Since iI is an involution with the two eigenvalues ± 1, comparing the action on both eigenspaces shows that, for $0 \leq t \leq 1$, we have

$$\left(\frac{\mathbf{1} - iC}{\mathbf{1} + iC}\right)^t = \left(\frac{\mathbf{1} - iI|C|}{\mathbf{1} + iI|C|}\right)^t = e^{-t|D|iI}.$$

The assertion of the lemma now follows from

$$\begin{aligned}
\widetilde{\varphi}(t) &= (\mathbf{1} + iC)\left(\frac{\mathbf{1} - iC}{\mathbf{1} + iC}\right)^t = (\mathbf{1} + iI|C|)e^{-t|D|iI} \\
&= (\mathbf{1} + iI|C|)\left(e^{t|D|}\frac{\mathbf{1} - iI}{2} + e^{-t|D|}\frac{\mathbf{1} + iI}{2}\right) \\
&= e^{-t|D|}(\mathbf{1} + |C|)\frac{\mathbf{1} + iI}{2} + e^{t|D|}(\mathbf{1} - |C|)\frac{\mathbf{1} - iI}{2} \\
&= (\mathbf{1} + |C|)\left(e^{-t|D|}\frac{\mathbf{1} + iI}{2} + e^{-(1-t)|D|}\frac{\mathbf{1} - iI}{2}\right) \\
&= (\mathbf{1} + e^{-|D|})^{-1}\big(e^{-t|D|}(\mathbf{1} + iI) + e^{-(1-t)|D|}(\mathbf{1} - iI)\big) \\
&= \frac{e^{-t|D|} + e^{-(1-t)|D|}}{\mathbf{1} + e^{-|D|}} + iI\frac{e^{-t|D|} - e^{-(1-t)|D|}}{\mathbf{1} + e^{-|D|}}.
\end{aligned}$$

□

Remark 4.3. (a) Since C is a strict contraction on V, $\mathbf{1} + iC$ is injective on $V_{\mathbb{C}}$, so

$$\widetilde{H} := \log\left(\frac{\mathbf{1} + iC}{\mathbf{1} - iC}\right) = \log\left(\frac{\mathbf{1} + iI|C|}{\mathbf{1} - iI|C|}\right) = iI|D| = -iD$$

also defines a selfadjoint operator on the complex Hilbert space $V_{\mathbb{C}}$.

Next we observe that $\widetilde{H}$ is a restriction of

$$L := \log\left(\frac{\mathbf{1} + i\widehat{C}}{\mathbf{1} - i\widehat{C}}\right) = -\log\Delta,$$

the infinitesimal generator of the one-parameter group $U_t = \Delta^{-it}$. For the orthogonal one-parameter group $U_t^V := U_t|_V \in \mathrm{O}(V)$, it follows that its infinitesimal generator is a skew-adjoint extension of the skew-adjoint operator D on V, and hence coincides with D. We therefore have

$$e^{tD} = \Delta^{-it}|_V \quad \text{for } t \in \mathbb{R}. \tag{18}$$

This provides an alternative characterization of the operator D in Lemma 4.2.

(b) Let $(V, (\cdot\,,\cdot))$ be a real Hilbert space and $(U_t)_{t\in\mathbb{R}}$ be an orthogonal strongly continuous one-parameter group with skew-symmetric infinitesimal generator D, i.e., $U_t = e^{tD}$ for $t \in \mathbb{R}$. Let us assume that $\ker D = \{0\}$, i.e., the subspace V^U of U-fixed points in V is trivial. Then the polar decomposition $D = I|D|$ can be used to define a skew-symmetric contraction

$$C := (-I)\frac{\mathbf{1} - e^{-|D|}}{\mathbf{1} + e^{-|D|}} \quad \text{with } |C| = \frac{\mathbf{1} - e^{-|D|}}{\mathbf{1} + e^{-|D|}}.$$

Then the hermitian form

$$h(v, w) := (v, w) + i(v, Cw)$$

defines a positive definite kernel on V (Lemma A.10). Let $\mathcal{H}$ denote the corresponding reproducing kernel space and let $j : V \to \mathcal{H}$ be the natural map. By construction, $|C|$ has no fixed points, so that $\mathbf{1} + C^2$ is injective, and therefore Lemma A.10(iii) implies that the complex linear extension $j_{\mathbb{C}} : V_{\mathbb{C}} \to \mathcal{H}$ is injective. As the real part of h is the original scalar product on V, the inclusion $V \hookrightarrow \mathcal{H}$ is isometric, so that $V \cong j(V)$ is a standard real subspace of $\mathcal{H}$. Since h is U-invariant, it defines a unitary one-parameter group $\widehat{U}$ on $\mathcal{H}$. Finally (18) implies that $\widehat{U}_t = \Delta^{-it}$ for $t \in \mathbb{R}$ and the modular operator Δ corresponding to $j(V)$. This shows that every orthogonal one-parameter group on a real Hilbert space is of the form (18) for a naturally defined embedding $V \hookrightarrow \mathcal{H}$ as a standard real subspace.

Before we turn to the associated reflection positive functions, we need the following technical lemma on Fourier expansions of certain operator-valued functions. In [Cuniberti et al. 2001], this is called the Matsubara formalism. (In view of [Dereziński and Gérard 2013, Definition 18.49], we have

$$u_B^+(t) = G_{E,\beta}(t) \cdot \frac{2B(\mathbf{1} - e^{-\beta B})}{\mathbf{1} + e^{-\beta B}},$$

where $G_{E,\beta}$ is the euclidean thermal Green's function associated to the positive operator $\varepsilon = B$.)

Lemma 4.4. *Let $B \geq 0$ be a selfadjoint operator on the complex Hilbert space $\mathcal{H}$ and let $\beta > 0$. We consider the operator-valued functions $u_B^{\pm} : \mathbb{R} \to B(\mathcal{H})$ satisfying*

$$u_B^{\pm}(t + \beta) = \pm u_B^{\pm}(t) \quad \textit{and} \quad u_B^{\pm}(t) = \frac{e^{-tB} \pm e^{-(\beta - t)B}}{\mathbf{1} + e^{-\beta B}} \quad \textit{for } 0 \leq t \leq \beta.$$

Then $u_B^{\pm}$ are weakly continuous symmetric 2β-periodic with the Fourier expansions

$$u_B^+(t) = \sum_{n\in\mathbb{Z}} c_{2n}^B e^{2n\pi it/\beta} \quad \textit{and} \quad u_B^-(t) = \sum_{n\in\mathbb{Z}} c_{2n+1}^B e^{(2n+1)\pi it/\beta}$$

with

$$c_n^B = c_{-n}^B = \frac{(\mathbf{1} - (-1)^n e^{-\beta B})}{\mathbf{1} + e^{-\beta B}} \cdot \frac{2\beta B}{(\beta B)^2 + (n\pi)^2} \quad \textit{for } n \in \mathbb{Z}.$$

Proof. (a) Every 2β-periodic continuous function $\xi : \mathbb{R} \to \mathbb{C}$ has a Fourier expansion

$$\xi(t) = \sum_{n\in\mathbb{Z}} c_n e^{\pi int/\beta} \quad \text{with} \quad c_n = \frac{1}{2\beta}\int_0^{2\beta} \xi(t)e^{-\pi int/\beta}\,dt.$$

For the β-periodic function with $u^+(t) = u^+_\lambda(t) := (e^{-t\lambda} + e^{-(\beta-t)\lambda})/(1+e^{-\beta\lambda})$ for $0 \leq t \leq \beta$ we have $u^+(t+\beta) = u^+(t)$, so that only even terms appear:

$$u^+(t) = \sum_{n\in\mathbb{Z}} c_{2n} e^{\pi i2nt/\beta}, \quad c_{2n} = \frac{1-e^{-\beta\lambda}}{1+e^{-\beta\lambda}} \frac{2\beta\lambda}{(\beta\lambda)^2 + (2\pi n)^2}.$$

To obtain this formula, we first calculate

$$\begin{aligned} a_{\lambda,n} &:= \frac{1}{\beta}\int_0^\beta e^{-t\lambda}e^{-\pi int/\beta}\,dt = \int_0^1 e^{-(\beta\lambda+\pi in)t}\,dt \\ &= \frac{1-e^{-(\beta\lambda+\pi in)}}{\beta\lambda + \pi i2n} = \frac{1-(-1)^n e^{-\beta\lambda}}{\beta\lambda+\pi i2n}. \end{aligned}$$

Therefore

$$\begin{aligned} (1+e^{-\lambda\beta})c_{2n} &= a_{\lambda,2n} + e^{-\beta\lambda}a_{-\lambda,2n} = \frac{1-e^{-\beta\lambda}}{\beta\lambda+2n\pi i} + e^{-\beta\lambda}\frac{1-e^{\beta\lambda}}{-\beta\lambda+2n\pi i} \\ &= \frac{1-e^{-\beta\lambda}}{\beta\lambda+2n\pi i} + \frac{1-e^{-\beta\lambda}}{\beta\lambda-2n\pi i} = \frac{(1-e^{-\beta\lambda})2\beta\lambda}{(\beta\lambda)^2+(2n\pi)^2} \end{aligned}$$

For the 2β-periodic function with $u^-(t) = u^-_\lambda(t) := (e^{-t\lambda} - e^{-(\beta-t)\lambda})/(1+e^{-\beta\lambda})$ for $0 \leq t \leq \beta$ and $u^-(t+\beta) = -u^-(t)$ only odd terms appear:

$$u^-(t) = \sum_{n\in\mathbb{Z}} c_{2n+1}e^{\pi i(2n+1)t/\beta}, \quad c_{2n+1} = \frac{2\beta\lambda}{(\beta\lambda)^2 + ((2n+1)\pi)^2}.$$

This follows from

$$\begin{aligned} c_{2n+1} &= \frac{a_{\lambda,2n+1} - e^{-\beta\lambda}a_{-\lambda,2n+1}}{1+e^{-\beta\lambda}} \\ &= \frac{1}{\beta\lambda+(2n+1)\pi i} - \frac{e^{-\beta\lambda}(1+e^{\beta\lambda})}{1+e^{-\beta\lambda}}\frac{1}{-\beta\lambda+(2n+1)\pi i} \\ &= \frac{1}{\beta\lambda+(2n+1)\pi i} + \frac{1}{\beta\lambda-(2n+1)\pi i} = \frac{2\beta\lambda}{(\beta\lambda)^2+((2n+1)\pi)^2}. \end{aligned}$$

Note that

$$c_n = c_{-n} = \frac{1-(-1)^n e^{-\beta\lambda}}{1+e^{-\beta\lambda}}\frac{2\beta\lambda}{(\beta\lambda)^2+(n\pi)^2} \quad \text{for } n \in \mathbb{Z}.$$

(b) If P denotes the spectral measure of B, we have for $v \in \mathcal{H}$ the relation

$$\langle v, Bv\rangle = \int_0^\infty x\, dP^{v,v}(x) \quad \text{with } P^{v,v} = \langle v, P(\cdot)v\rangle.$$

This leads for $0 \le t \le 2\beta$ to

$$\langle v, u_B^\pm(t)v\rangle = \int_0^\infty u_\lambda^\pm(t)\, dP^{v,v}(\lambda).$$

For the operator-valued Fourier coefficients, we thus obtain

$$\begin{aligned}\langle v, c_n^B v\rangle &= \int_{\mathbb{R}} c_n(\lambda)\, dP^{v,v}(\lambda) = \int_{\mathbb{R}} \frac{1-(-1)^n e^{-\beta\lambda}}{1+e^{-\beta\lambda}} \frac{2\beta\lambda}{(\beta\lambda)^2+(n\pi)^2}\, dP^{v,v}(\lambda)\\ &= \left\langle v, \frac{(\mathbf{1}-(-1)^n e^{-\beta B})}{\mathbf{1}+e^{-\beta B}} \frac{2\beta B}{(\beta B)^2+(n\pi)^2} v\right\rangle.\end{aligned}$$

This proves the assertion. □

4B. ***Existence of reflection positive extensions.*** We now come to one of our main results on reflection positive extensions. It shows that, for every positive definite function $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ satisfying the β-KMS condition, there exists a reflection positive function $f : G_\tau \to B(V_{\mathbb{C}})$ satisfying

$$\psi(it)(v, w) = \langle v, f(it, \tau)w\rangle$$

for $v, w \in V$, $0 \le t \le \beta$. Then the corresponding GNS representation $(U^f, \mathcal{H}_f)$ of the group $(\mathbb{T}_{2\beta})_\tau \cong \mathrm{O}_2(\mathbb{R})$ is a "euclidean realization" of the unitary one-parameter group $(\Delta^{-it/\beta})_{t\in\mathbb{R}}$ corresponding to ψ in the sense that it is obtained by Osterwalder–Schrader quantization from U^f (cf. [Neeb and Ólafsson 2014]). The following theorem generalizes the results of [Neeb and Ólafsson 2015b] dealing with the scalar-valued case.

Theorem 4.5 (Reflection positive extensions). *Let $V \subseteq \mathcal{H}$ be a standard real subspace and let $C = I|C|$ be the corresponding skew-symmetric strict contraction on V. We assume that* $\ker C = \{0\}$, *so that I defines a complex structure on V. We define a weakly continuous function $\widetilde{\varphi} : \mathbb{R} \to B(V_{\mathbb{C}})$ by*

$$\widetilde{\varphi}(t) = (\mathbf{1}+iC)^{1-t/\beta}(\mathbf{1}-iC)^{t/\beta} \quad \text{for } 0 \le t \le \beta \quad \text{and} \quad \widetilde{\varphi}(t+\beta) = \overline{\widetilde{\varphi}(t)} \quad \text{for } t \in \mathbb{R}.$$

Write

$$\widetilde{\varphi}(t) = u^+(t) + iIu^-(t) \quad \text{with } u^\pm(t) \in B(V),\ u^\pm(t+\beta) = \pm u^\pm(t).$$

Then

$$f : \mathbb{R}_\tau \to B(V_{\mathbb{C}}), \qquad f(t, \tau^\varepsilon) := u^+(t) + (iI)^\varepsilon u^-(t), \qquad t \in \mathbb{R}, \varepsilon \in \{0, 1\},$$

is a weak-operator continuous positive definite function satisfying $f(t, \tau) = \widetilde{\varphi}(t)$.

It is reflection positive with respect to the subset $[0, \beta/2] \subseteq \mathbb{R}$ *in the sense that the kernel*

$$f((t, \tau)(-s, \mathbf{1})) = f(t+s, \tau), \qquad 0 \leq s, t \leq \beta/2,$$

is positive definite.

Proof. We may, without loss of generality, assume that $\beta = 1$. Recall the operator D from Lemma 4.2. With this lemma, we write

$$\widetilde{\varphi}(t) = (\mathbf{1} + e^{-|D|})^{-1}\big(e^{-t|D|} + e^{-(1-t)|D|} + iI(e^{-t|D|} - e^{-(1-t)|D|})\big) \quad \text{for } 0 \leq t \leq 1.$$

Using Lemma 4.4 with $\beta = 1$ and $B = |D|$, we get

$$\widetilde{\varphi}(t) = u^{+}_{|D|}(t) + iIu^{-}_{|D|}(t) \quad \text{for } t \in \mathbb{R}.$$

(a) We define $f_1 : \mathbb{R}_\tau \to B(V_{\mathbb{C}})$ by $f_1(t, \tau^\varepsilon) := u^{+}_{|D|}(t)$ for $t \in \mathbb{R}$, $\varepsilon \in \{0, 1\}$. To see that f_1 is positive definite, it suffices to verify this for its restriction to $\mathbb{R}$ (Lemma 3.6), which follows from the positivity of the Fourier coefficients in the expansion

$$u^{+}_{|D|}(t) = \sum_{n \in \mathbb{Z}} c^{|D|}_{2n} e^{2n\pi it} \quad \text{with} \quad c^{|D|}_{2n} = \frac{\mathbf{1} - e^{-|D|}}{\mathbf{1} + e^{-|D|}} \frac{2|D|}{|D|^2 + (2n\pi)^2\mathbf{1}} \geq 0$$

(Lemma 4.4). Note that f_1 is 1-periodic.

(b) Likewise, the function $f_2 : \mathbb{R}_\tau \to B(V_{\mathbb{C}})$ defined by $f_2(t, \tau^\varepsilon) := u^{-}_{|D|}(t)$ for $t \in \mathbb{R}$, $\varepsilon \in \{0, 1\}$, is positive definite because the Fourier coefficients

$$c^{|D|}_{2n+1} = \frac{2|D|}{|D|^2 + ((2n+1)\pi)^2\mathbf{1}} \geq 0 \quad \text{for } n \in \mathbb{Z}$$

are positive. Note that $f_2(t+1, \tau^\varepsilon) = -f_2(t, \tau^\varepsilon)$ for $t \in \mathbb{R}$, $\varepsilon \in \{0, 1\}$.

(c) We now consider the function

$$\tilde{f}_2(g) := h(g) f_2(g) \quad \text{with } h(t, \tau^\varepsilon) = (iI)^\varepsilon \text{ for } t \in \mathbb{R},\ \varepsilon \in \{0, 1\}.$$

Since $h(g)$ commutes with $f_2(g')$ for $g, g' \in \mathbb{R}_\tau$, the function $\tilde{f}_2$ is positive definite if h is positive definite (Lemma A.6). As h is constant on the two $\mathbb{R}$-cosets and its restriction to the 2-element subgroup $\{\mathbf{1}, \tau\}$ is a unitary representation, h is positive definite. We conclude that the $B(V_{\mathbb{C}})$-valued function $f := f_1 + \tilde{f}_2$ on $\mathbb{R}_\tau$ is positive definite. □

Corollary 4.6. *Let* V *be a real vector space and let* $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ *be a continuous positive definite function satisfying the* β*-KMS condition. Then there exists a pointwise continuous function* $f : \mathbb{R}_\tau \to \mathrm{Bil}(V)$ *which is reflection positive with respect to the subset* $[0, \beta/2] \subseteq \mathbb{R}$ *and which satisfies*

$$f(t, \tau) = \psi(it) \quad \textit{for } 0 \leq t \leq \beta \qquad \textit{and} \qquad f(t+\beta, \tau) = \overline{f(t, \tau)} \quad \textit{for } t \in \mathbb{R}.$$

Remark 4.7. The function $\tilde{f}_2$ in the proof of Theorem 4.5 is not reflection positive because $\tilde{f}_2(\beta, \tau)$ is a negative operator. This also shows that the natural decomposition $f = f_1 + \tilde{f}_2$ into even and odd parts is not compatible with reflection positivity.

4C. ***Integral representation of reflection positive functions.*** We now describe an integral representation of the reflection positive function $f : \mathbb{R}_\tau \to \mathrm{Bil}(V)$ which corresponds to the decomposition of the corresponding unitary representation of $\mathbb{R}_\tau$. With

$$\widetilde{\varphi}(t) = (\mathbf{1} + iC)^{1-t/\beta}(\mathbf{1} - iC)^{t/\beta} \quad \text{for } 0 \le t \le \beta,$$

where $C \in B(V)$ is a skew-symmetric strict contraction, we first decompose V into $V_0 := \ker C$ and $V_1 := V_0^\perp = \overline{CV}$. Then the polar decomposition $C = I|C|$ yields a complex structure I on V_1. Accordingly, we write $\widetilde{\varphi} = \widetilde{\varphi}_0 + \widetilde{\varphi}_1$, where $\widetilde{\varphi}_0 = \mathbf{1}$ is constant. This component leads to the constant function $f_0(t, \tau) = \mathbf{1}$. We now assume that $V = V_1$, i.e., that C is injective. Then I is a complex structure on V.

Proposition 4.8. *If* $\ker C = \{0\}$ *and P denotes the spectral measure of the symmetric operator* $|D| = \frac{1}{\beta} \log \frac{\mathbf{1}+|C|}{\mathbf{1}-|C|}$ *on V, then we have the integral representation*

$$f(t, \tau^\varepsilon) = \int_{(0,\infty)} u_\lambda^+(t) + u_\lambda^-(t)(iI)^\varepsilon \, dP(\lambda), \tag{19}$$

where $u_\lambda^\pm : \mathbb{R} \to \mathbb{R}$ *are defined by* $u_\lambda^\pm(t + \beta) = \pm u_\lambda^\pm(t)$ *and*

$$u_\lambda^\pm(t) := \frac{e^{-t\lambda} \pm e^{-(\beta-t)\lambda}}{1 + e^{-\beta\lambda}} \quad \textit{for } 0 \le t \le \beta.$$

Proof. First we observe that $|D|$ is a positive symmetric operator with trivial kernel which commutes with I. We therefore have $|D| = \int_{(0,\infty)} \lambda \, dP(\lambda)$. With the notation from Lemma 4.4, we then have

$$f(t, \tau^\varepsilon) = u_{|D|}^+(t) + u_{|D|}^-(t)(iI)^\varepsilon \quad \text{for } t \in \mathbb{R}, \ \varepsilon \in \{0, 1\}.$$

From the integral representations $u_{|D|}^\pm(t) = \int_{(0,\infty)} u_\lambda^\pm(t) \, dP(\lambda)$, we obtain (19). ☐

Remark 4.9. (a) For $0 \le t \le \beta$, we have in particular

$$f(t, \tau^\varepsilon) = \int_{(0,\infty)} \frac{e^{-t\lambda} + e^{-(\beta-t)\lambda}}{1 + e^{-\beta\lambda}} \mathbf{1} + \frac{e^{-t\lambda} - e^{-(\beta-t)\lambda}}{1 + e^{-\beta\lambda}} (iI)^\varepsilon \, dP(\lambda).$$

(b) The most basic type is obtained for $D = \lambda\mathbf{1}$, $\lambda > 0$, which, for $0 \le t \le \beta$, leads to

$$f(t, \tau^\varepsilon) = \frac{(e^{-t\lambda} + e^{-(\beta-t)\lambda})\mathbf{1} + (e^{-t\lambda} - e^{-(\beta-t)\lambda})(iI)^\varepsilon}{1 + e^{-\beta\lambda}} = u_\lambda^+(t)\mathbf{1} + u_\lambda^-(t)(iI)^\varepsilon.$$

The simplest nontrivial example arises for $V = \mathbb{R}^2$ with $I = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

(c) Every Borel spectral measure P on $(0, \infty)$ which commutes with I defines a positive operator $|D| = \int_0^\infty \lambda \, dP(\lambda)$ and we may put $D := -I|D|$. Then $\ker |D| = 0$, so

$$|C| := \frac{e^{\beta |D|} - \mathbf{1}}{e^{\beta |D|} + \mathbf{1}} = \tanh\left(\frac{\beta |D|}{2}\right)$$

is a symmetric contraction with trivial kernel commuting with I, and therefore $C := I|C|$ is a skew-symmetric contraction with polar decomposition $I|C|$ and

$$|D| = \frac{1}{\beta} \log\left(\frac{\mathbf{1} + |C|}{\mathbf{1} - |C|}\right).$$

4D. ***Characterizing reflection positive extensions.*** In Theorem 4.5 we obtained positive definite extensions to all of $\mathbb{R}_\tau$ for certain functions on the coset $\mathbb{R} \rtimes \{\tau\}$. In this section we describe an intrinsic characterization of those weakly continuous reflection positive functions $f : \mathbb{R}_\tau \to B(V_{\mathbb{C}})$ arising from this construction. First we observe that we can recover ψ from f:

Lemma 4.10. *If $f : \mathbb{R}_\tau \to \mathrm{Bil}(V)$ is reflection positive and pointwise continuous, then there exists a unique β-KMS positive definite function $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ with*

$$f(t, \tau) = \psi(it) \quad \textit{for } 0 \leq t \leq \beta.$$

Proof. First we observe that the function $\varphi(t) := f(t, \tau)$ has values in $\mathrm{Herm}(V_{\mathbb{C}})$ and satisfies

$$\varphi(t + \beta) = \overline{\varphi(t)} \quad \text{for } t \in \mathbb{R}. \tag{20}$$

Reflection positivity implies that the kernel $\varphi\left(\frac{t+s}{2}\right)$ for $0 \leq t, s \leq \beta$ is positive definite. By [Neeb and Ólafsson 2015b, Theorem B.3], there exists a $\mathrm{Bil}^+(V)$-valued measure μ such that

$$\varphi(t) = \int_{\mathbb{R}} e^{-\lambda t} \, d\mu(\lambda) \quad \text{for } 0 < t < \beta. \tag{21}$$

The continuity of φ on $[0, \beta]$ actually implies that the integral representation also holds on the closed interval $[0, \beta]$ by the monotone convergence theorem. In particular, the measure μ is finite. Therefore its Fourier transform $\psi(t) := \int_{\mathbb{R}} e^{it\lambda} \, d\mu(\lambda)$ is a pointwise continuous $\mathrm{Bil}(V)$-valued positive definite function on $\mathbb{R}$. Further, (20) implies

$$e^{\beta\lambda} \, d\mu(-\lambda) = d\overline{\mu}(\lambda), \tag{22}$$

and $\varphi(t) = \psi(it)$ holds for the β-KMS function $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ by Theorem 4.5. ☐

Before we describe a realization of the GNS representation $(U^f, \mathcal{H}_f)$ in spaces of sections of a vector bundle, let us recall the general background for this.

Remark 4.11. For a $B(V)$-valued positive definite function $f : G \to B(V)$, the reproducing kernel Hilbert space with kernel $K(g, h) = \varphi(gh^{-1}) = K_g K_h^*$ is generated by the functions

$$K_{h,w} := K_h^* w \quad \text{with } K_{h,w}(g) = K_g K_h^* w = K(g, h)w = \varphi(gh^{-1})w.$$

The group G acts on this space by right translations

$$(U_g s)(h) := s(hg).$$

If $P \subseteq G$ is a subgroup and (ρ, V) is a unitary representation for which

$$f(hg) = \rho(h) f(g) \quad \text{for all } g \in G,\ h \in P,$$

then

$$\mathcal{H}_f \subseteq \mathcal{F}(G, V)_\rho := \{s : G \to V : s(hg) = \rho(h)s(g) \text{ for all } g \in G, h \in P\}.$$

Therefore $\mathcal{H}_f$ can be identified with a space of sections of the associated bundle

$$\mathbb{V} := (V \times_P G) = (V \times G)/P,$$

where P acts on the trivial vector bundle $V \times G$ over G by $h.(v, g) = (\rho(h)v, hg)$.

To derive a suitable characterization of the functions f arising in Theorem 4.5, we identify 2β-periodic function s on $\mathbb{R}$ with pairs of function (s_0, s_1) via $s = s_0 + s_1$, where s_0 is β-periodic and $s_1(\beta + t) = -s_1(t)$. Accordingly, any 2β-periodic function $s : \mathbb{R} \to V_\mathbb{C}$ defines a function

$$\tilde{s} : \mathbb{R} \to V_\mathbb{C}^2, \quad \tilde{s} = (s_1, s_2) \quad \text{with } \tilde{s}(\beta + t) = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \tilde{s}(t).$$

In this sense, $\tilde{s}$ is a section of the vector bundle over $\mathbb{T}_\beta$ with fiber $V_\mathbb{C}^2$ defined by the representation of $\beta\mathbb{Z}$, specified by

$$\rho(\beta) = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}.$$

Splitting the $B(V)$-valued positive definite function

$$f : \mathbb{R}_\tau \to B(V), \qquad f(t, \tau^\varepsilon) = u^+_{|D|}(t) + u^-_{|D|}(t)(iI)^\varepsilon \quad \text{for } t \in \mathbb{R}, \varepsilon \in \{0, 1\}$$

into even and odd parts with respect to the β-translation, we obtain:

Lemma 4.12. *For the subgroup* $P := (\mathbb{Z}\beta)_\tau \cong \mathbb{Z}\beta \rtimes \{\mathbf{1}, \tau\}$ *of* $G := \mathbb{R}_\tau$, *we consider the unitary representation* $\rho : P \to \mathrm{U}(V_\mathbb{C}^2)$ *defined by*

$$\rho(\beta, \mathbf{1}) := \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \quad \text{and} \quad \rho(0, \tau) := \begin{pmatrix} \mathbf{1} & 0 \\ 0 & iI \end{pmatrix},$$

where I is a complex structure on the real Hilbert space V commuting with the *positive operator $|D|$. Then*

$$f^\sharp : \mathbb{R}_\tau \to B(V^2) \cong M_2(B(V)), \qquad f^\sharp(t, \tau^\varepsilon) := \begin{pmatrix} u^+_{|D|}(t) & 0 \\ 0 & u^-_{|D|}(t)(iI)^\varepsilon \end{pmatrix}$$

is a positive definite function satisfying

$$f^\sharp(hg) = \rho(h) f^\sharp(g) \quad \text{for } h \in P, g \in G. \tag{23}$$

The corresponding GNS representation $(U^{f^\sharp}, \mathcal{H}_{f^\sharp})$ is equivalent to the GNS representation $(U^f, \mathcal{H}_f)$.

Proof. The first assertion follows from

$$\begin{aligned} f^\sharp((0,\tau)(t,\tau^\varepsilon)) = f^\sharp(-t, \tau^{\varepsilon+1}) &= \begin{pmatrix} u^+_{|D|}(-t) & 0 \\ 0 & u^-_{|D|}(-t)(iI)^{\varepsilon+1} \end{pmatrix} \\ &= \begin{pmatrix} u^+_{|D|}(t) & 0 \\ 0 & u^-_{|D|}(t)(iI)^{\varepsilon+1} \end{pmatrix} \end{aligned}$$

and

$$f^\sharp(\beta + t, \tau^\varepsilon) = \begin{pmatrix} u^+_{|D|}(t) & 0 \\ 0 & -u^-_{|D|}(t)(iI)^\varepsilon \end{pmatrix}.$$

As the GNS representation $(U^f, \mathcal{H}_f)$ decomposes under the involution U^f_β into ± 1-eigenspaces, this representation is equivalent to the GNS representation $(U^{f^\sharp}, \mathcal{H}_{f^\sharp})$ corresponding to $f^\sharp$. □

Remark 4.13. (a) The preceding lemma implies that, if the complex structure I on V is fixed, then the relation (23) determines $f^\sharp$ completely in terms of the function

$$[0, \beta] \to M_2(B(V)), \quad t \mapsto f^\sharp(t, \tau) = \begin{pmatrix} \operatorname{Re}\varphi(t) & 0 \\ 0 & i \operatorname{Im}\varphi(t) \end{pmatrix},$$

so that φ determines f in a natural way.

(b) This lemma also shows that we may identify the Hilbert space $\mathcal{H}_f \cong \mathcal{H}_{f^\sharp}$ as a space of section of a Hilbert bundle $V^2 \times_\rho G$ over the circle $\mathbb{T}_\beta \cong \mathbb{R}_\tau / H$ with fiber V^2.

(c) Every function $s : \mathbb{R}_\tau \to V^2$ satisfying $s(hg) = \rho(h)s(g)$ for $h \in (\beta\mathbb{Z})_\tau$ is determined by its restriction $\tilde{s}$ to the subgroup $\mathbb{R}$, which satisfies

$$\tilde{s}(\beta + t) = \rho(\beta, \mathbf{1})\tilde{s}(t) \quad \text{for } t \in \mathbb{R}.$$

The action of τ is in this picture given by

$$(\tau.\tilde{s})(t) := s(t, \tau) = s((0,\tau)(-t, \mathbf{1})) = \rho(\tau)\tilde{s}(-t). \tag{24}$$

Remark 4.14. (a) In view of (22), there exists a $\mathrm{Bil}^+(V)$-valued measure ν on $[0, \infty)$ for which we can write

$$d\mu(\lambda) = d\nu(\lambda) + e^{\beta\lambda} d\overline{\nu}(-\lambda). \tag{25}$$

This leads, for $0 \leq t \leq \beta$ and $\nu = \nu_1 + i\nu_2$, to

$$\varphi(t) = \int_0^\infty e^{-t\lambda} + e^{-(\beta-t)\lambda}\, d\nu_1(\lambda) + i \int_0^\infty e^{-t\lambda} - e^{-(\beta-t)\lambda}\, d\nu_2(\lambda). \tag{26}$$

In particular, the most elementary nontrivial examples correspond to the Dirac measures of the form $\nu = \delta_\lambda \cdot (\gamma + i\omega)$, where δ_λ is the Dirac measure in $\lambda > 0$:

$$\varphi(t) = (e^{-t\lambda} + e^{-(\beta-t)\lambda})\gamma + i(e^{-t\lambda} - e^{-(\beta-t)\lambda})\omega = e^{-t\lambda} h + e^{-(\beta-t)\lambda} \overline{h},$$

where $h := \gamma + i\omega \in \mathrm{Bil}^+(V)$. Writing $\omega(v, w) = \gamma(v, Cw)$ (Corollary A.9) and replacing V by the real Hilbert space defined by the positive semidefinite form γ on V, we obtain the $B(V_{\mathbb{C}})$-valued function

$$\widetilde{\varphi}(t) = (e^{-t\lambda} + e^{-(\beta-t)\lambda}) + iC(e^{-t\lambda} - e^{-(\beta-t)\lambda}) = e^{-t\lambda}(\mathbf{1} + iC) + e^{-(\beta-t)\lambda}(\mathbf{1} - iC)$$

for $0 \leq t \leq \beta$, which leads to

$$f(t, \tau^\varepsilon) = (1 + e^{-\beta\lambda})(u_\lambda^+(t)\mathbf{1} + u_\lambda^-(t)|C|(iI)^\varepsilon) \quad \text{for } t \in \mathbb{R}, \varepsilon \in \{0, 1\}.$$

(b) This can also be formulated in terms of forms. With $\gamma(v, w) = \langle v, w \rangle_V$ and

$$h(v, w) = \gamma(v, w) + i\omega(v, w) = \langle v, (\mathbf{1} + iC)w \rangle_{V_{\mathbb{C}}} = \langle v, (\mathbf{1} + iI|C|)w \rangle_{V_{\mathbb{C}}},$$

we get

$$f(t, \tau^\varepsilon)(v, w) = \langle v, (u_\lambda^+(t)\mathbf{1} + u_\lambda^-(t)|C|(iI)^\varepsilon)w \rangle.$$

4E. ***Realization by resolvents of the Laplacian.*** We have seen in the preceding subsection how to obtain a realization of the Hilbert space $\mathcal{H}_f$ as a space $\mathcal{H}_{f^\sharp}$ of sections of a Hilbert bundle $\mathbb{V}$ with fiber $V_{\mathbb{C}}^2$ over the circle $\mathbb{T}_\beta = \mathbb{R}/\beta\mathbb{Z}$. In this section we provide an analytic description of the scalar product on this space if $|D| = \lambda\mathbf{1}$ for some $\lambda > 0$. We shall see that it has a natural description in terms of the resolvent $(\lambda^2 - \Delta)^{-1}$ of the Laplacian of $\mathbb{T}_\beta$ acting on sections of the bundle $\mathbb{V}$.

On the circle group $\mathbb{T}_{2\beta}$, we consider the normalized Haar measure given by

$$\int_{\mathbb{T}_{2\beta}} h(t)\, d\mu_{\mathbb{T}_{2\beta}} = \frac{1}{2\beta} \int_0^{2\beta} h(t)\, dt,$$

where we identify functions h on $\mathbb{T}_{2\beta}$ with 2β-periodic functions on $\mathbb{R}$.

As in Lemma 4.12, we write

$$f^\sharp(t, \tau^\varepsilon) = \begin{pmatrix} u_\lambda^+(t)\mathbf{1} & 0 \\ 0 & u_\lambda^-(t)(iI)^\varepsilon \end{pmatrix} \in B(V_{\mathbb{C}}^2) \cong M_2(B(V_{\mathbb{C}})),$$

For $\chi_n(t) = e^{\pi int/\beta}$ we then have

$$u_\lambda^+ = \sum_{n\in\mathbb{Z}} c_{2n}^\lambda \chi_{2n} \quad \text{and} \quad u_\lambda^- = \sum_{n\in\mathbb{Z}} c_{2n+1}^\lambda \chi_{2n+1},$$

where

$$c_n^\lambda = c_{-n}^\lambda = \frac{1-(-1)^n e^{-\beta\lambda}}{1+e^{-\beta\lambda}} \cdot \frac{2\beta\lambda}{(\beta\lambda)^2+(n\pi)^2} = \frac{1-(-1)^n e^{-\beta\lambda}}{1+e^{-\beta\lambda}} \cdot \frac{2\lambda}{\beta} \cdot \frac{1}{\lambda^2+(n\pi/\beta)^2}$$

for $n \in \mathbb{Z}$ (the rightmost factors are called bosonic Matsubara coefficients if n is even and fermionic if n is odd [Dereziński and Gérard 2013, §18]). With

$$c_+^\lambda := \frac{1-e^{-\beta\lambda}}{1+e^{-\beta\lambda}} \frac{2\lambda}{\beta} = \tanh\left(\frac{\beta\lambda}{2}\right)\frac{2\lambda}{\beta} \quad \text{and} \quad c_-^\lambda := \frac{2\lambda}{\beta}, \tag{27}$$

we thus obtain

$$c_{2n}^\lambda = \frac{c_+^\lambda}{\lambda^2+(2n\pi/\beta)^2}, \qquad c_{2n+1}^\lambda = \frac{c_-^\lambda}{\lambda^2+((2n+1)\pi/\beta)^2}. \tag{28}$$

The following proposition shows that the positive operator $(\lambda^2-\Delta)^{-1}$ on the Hilbert space of L^2-section of $\mathbb{V}$ defines a unitary representation of $\mathbb{R}_\tau$ which is unitarily equivalent to the representation on $\mathcal{H}_f$ (cf. Lemma 4.12).

Proposition 4.15. *For $\lambda > 0$, let $\mathcal{H}_\lambda$ be the Hilbert space obtained by completing the space*

$$\Gamma_\rho := \{s \in C^\infty(\mathbb{R}_\tau, V_\mathbb{C}^2) : s(hg) = \rho(h)s(g) \textit{ for all } g \in \mathbb{R}_\tau,\ h \in (\mathbb{Z}\beta)_\tau)\}$$

with respect to

$$\langle s_1, s_2\rangle := \frac{1}{2\beta}\int_0^{2\beta} \langle s_1(t,\mathbf{1}), ((\lambda^2-\Delta)^{-1}s_2)(t,\mathbf{1})\rangle\, dt.$$

On $\mathcal{H}_\lambda$ we have a natural unitary representation U^λ of $\mathbb{R}_\tau$ by right translation which is unitarily equivalent to the GNS representation $(U^{f^\sharp}, \mathcal{H}_{f^\sharp})$. Here the corresponding j-map is given by

$$j : V \to \mathcal{H}_\lambda, \quad j\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \sqrt{c_+^\lambda}\sum_{n\in\mathbb{Z}} \chi_{2n}\begin{pmatrix} v_1 \\ 0 \end{pmatrix} + \sqrt{c_-^\lambda}\sum_{n\in\mathbb{Z}} \chi_{2n+1}\begin{pmatrix} 0 \\ v_2 \end{pmatrix}. \tag{29}$$

Proof. We identify Γ_ρ with the space

$$\{s \in C^\infty(\mathbb{R}, V_\mathbb{C}^2) : s(\beta+t) = \rho(\beta)s(t) \text{ for all } t \in \mathbb{R}\}$$

(Remark 4.13). Then $s = \binom{s_+}{s_-}$, where s_+ is β-periodic and s_- is β-antiperiodic. Accordingly, we have an orthogonal decomposition $\mathcal{H}_\lambda = \mathcal{H}_\lambda^+ \oplus \mathcal{H}_\lambda^-$, where $\mathcal{H}_\lambda^\pm =$

$\{s \in \mathcal{H}_\lambda : s(\beta + t) = \pm s(t) \text{ for all } t \in \mathbb{R}\}$. Then U^λ is given by

$$(U^\lambda_t s)(x) = s(t+x) \quad \text{for } t, x \in \mathbb{R} \quad \text{and} \quad (U^\lambda_\tau s)(x) = \begin{pmatrix} s_+(-x) \\ (iI)s_-(-x) \end{pmatrix}.$$

From the Fourier expansion $s = \sum_{n\in\mathbb{Z}} \chi_n s_n$ and the orthonormality of the χ_n, we then derive

$$\langle s_1, s_2\rangle_{\mathcal{H}_\lambda} = \sum_{n\in\mathbb{Z}} \frac{\langle s_{1,n}, s_{2,n}\rangle}{\lambda^2 + (n\pi/\beta)^2}. \tag{30}$$

For the map $j : V \to \mathcal{H}_\lambda$ in (29), the image is $U^\lambda_\mathbb{R}$-generating for $\mathcal{H}_\lambda$ because the projection onto each Fourier component generates the first or the second component of $V^2_\mathbb{C}$, according to parity. Therefore the unitary representation $(U^\lambda, \mathcal{H}_\lambda)$ is equivalent to the GNS representation of the positive definite function $\tilde{f} : \mathbb{R}_\tau \to B(V^2_\mathbb{C})$, given by

$$\langle v, \tilde{f}(g)w\rangle = \langle j(v), U^\tau_g j(w)\rangle_{\mathcal{H}_\lambda}.$$

From

$$U^\lambda_{(t,\tau^\varepsilon)} j(v) = \sqrt{c^\lambda_+} \sum_{n\in\mathbb{Z}} \chi_{2n}\chi_{2n}(t) \begin{pmatrix} v_1 \\ 0 \end{pmatrix} + \sqrt{c^\lambda_-} \sum_{n\in\mathbb{Z}} \chi_{2n+1}\chi_{2n+1}(t) \begin{pmatrix} 0 \\ (iI)^\varepsilon v_2 \end{pmatrix},$$

we derive with (28),

$$\begin{aligned} \langle v, \tilde{f}(t,\tau^\varepsilon)w\rangle &= c^\lambda_+ \sum_{n\in\mathbb{Z}} \frac{\chi_{2n}(t)}{\lambda^2 + (2n\pi/\beta)^2} \langle v_1, w_1\rangle \\ &\qquad + c^\lambda_- \sum_{n\in\mathbb{Z}} \frac{\chi_{2n+1}(t)}{\lambda^2 + ((2n+1)\pi/\beta)^2} \langle v_1, (iI)^\varepsilon w_2\rangle \\ &= \sum_{n\in\mathbb{Z}} \chi_{2n}(t) c^\lambda_{2n} \langle v_1, w_1\rangle + \sum_{n\in\mathbb{Z}} \chi_{2n+1}(t) c^\lambda_{2n+1} \langle v_2, (iI)^\varepsilon w_2\rangle \\ &= \langle v_1, u^+_\lambda(t) w_1\rangle + \langle v_2, u^-_\lambda(t)(iI)^\varepsilon w_2\rangle = \langle v, f^\sharp(t,\tau^\varepsilon)w\rangle. \end{aligned}$$

This shows that $\tilde{f} = f^\sharp$, which completes the proof. □

Remark 4.16. From $u^+_\lambda = \sum_{n\in\mathbb{Z}} c^\lambda_{2n}\chi_{2n}$, it follows that

$$(\lambda^2 - \Delta)u^+_\lambda = \sum_{n\in\mathbb{Z}} c^\lambda_{2n}\left(\lambda^2 + \frac{(2\pi n)^2}{\beta^2}\right)\chi_{2n} = c^\lambda_+ \sum_{n\in\mathbb{Z}} \chi_{2n} = c^\lambda_+ \delta_0,$$

where the latter relation means that

$$s_+(0) = \frac{1}{2\beta} \sum_{n\in\mathbb{Z}} \int_0^{2\beta} s_+(t)\chi_{2n}(t)\, dt$$

for every smooth β-periodic function s_+ on $\mathbb{R}$. This relation can also be written as

$$(\lambda^2 - \Delta)^{-1}\delta_0 = \frac{1}{c_+^\lambda} u_\lambda^+.$$

From $u_\lambda^- = \sum_{n\in\mathbb{Z}} c_{2n+1}^\lambda \chi_{2n+1}$, it follows that

$$(\lambda^2 - \Delta)u_\lambda^- = \sum_{n\in\mathbb{Z}} c_{2n+1}^\lambda \left(\lambda^2 + \frac{(2n+1)^2\pi^2}{\beta^2}\right)\chi_{2n+1} = c_-^\lambda \chi_1 \sum_{n\in\mathbb{Z}} \chi_{2n}.$$

As every smooth β-antiperiodic function s_- is of the form $s_- = \chi_{-1}s_+$, where s_+ is β-periodic, we obtain, in the sense of distributions,

$$\langle (\lambda^2 - \Delta)u_\lambda^-, s_-\rangle = c_-^\lambda s_+(0) = c_-^\lambda s_-(0) = \langle c_-^\lambda \delta_0, s_-\rangle,$$

and therefore

$$(\lambda^2 - \Delta)^{-1}\delta_0 = \frac{1}{c_-^\lambda} u_\lambda^-$$

on β-antiperiodic functions. Combining all this, we get

$$((\lambda^2 - \Delta) f^\sharp)(t, \tau^\varepsilon) = \begin{pmatrix} (\lambda^2 - \Delta)u_\lambda^+ \mathbf{1} & 0 \\ 0 & (\lambda^2 - \Delta)u_\lambda^-(iI)^\varepsilon \end{pmatrix} = \delta_0 \begin{pmatrix} c_+^\lambda \mathbf{1} & 0 \\ 0 & c_-^\lambda (iI)^\varepsilon \end{pmatrix}$$

as an operator-valued distribution on the space of smooth sections of $\mathbb{V}$ (cf. also the discussion of thermal euclidean Green's functions in [Dereziński and Gérard 2013, Definition 18.49]).

5. The case $\beta = \infty$

In the context of C^*-dynamical systems, it is well known that the positive energy condition for the unitary one-parameter group implementing the automorphisms of a C^*-algebra $\mathcal{A}$ in a representation can be viewed as a KMS condition for $\beta = \infty$ (cf. [Bratteli and Robinson 1981]). For reflection positive representations of $G = \mathbb{R}$, this case corresponds to $G_+ = \mathbb{R}_+$, which has been treated in [Neeb and Ólafsson 2014; 2015a] (cf. also the discussion of euclidean Green's functions in [Dereziński and Gérard 2013, Definition 18.48]). The following theorem makes this analogy also transparent in the context of our Theorem 2.6.

If $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ is a positive definite function satisfying the KMS condition for $\beta > 0$, then its extension to $\overline{\mathcal{S}_\beta}$ is pointwise bounded (Theorem 2.6). This observation explains the assumptions in the following theorem.

Theorem 5.1 (KMS condition for $\beta = \infty$). *Let V be a real vector space and let $\psi : \mathbb{R} \to \mathrm{Bil}(V)$ be a pointwise continuous positive definite function. Then the following are equivalent*:

(i) *ψ extends to a pointwise bounded function on the closed upper half plane which is pointwise holomorphic on $\mathbb{C}_+$.*

(ii) *There exists a* $\mathrm{Bil}^+(V)$*-valued regular Borel measure μ on* $[0,\infty)$ *satisfying*

$$\psi(t) = \int_0^\infty e^{it\lambda}\, d\mu(\lambda).$$

(iii) *The GNS representation* $(U^\psi, \mathcal{H}_\psi)$ *has spectrum contained in* $[0,\infty)$.

If this is the case, then the function

$$f(t, \tau^\varepsilon) := \psi(i|t|) \quad \textit{for } t \in \mathbb{R}, \varepsilon \in \{0,1\},$$

on $\mathbb{R}_\tau$ is reflection positive with respect to $\mathbb{R}_+ = [0,\infty)$.

Proof. (i) $\Rightarrow$ (ii): First we use [Neeb and Ólafsson 2015b, Proposition B.1] to write φ as the Fourier transform of a $\mathrm{Bil}^+(V)$-valued regular Borel measure μ on $\mathbb{R}$: $\psi(t) = \int_{\mathbb{R}} e^{it\lambda}\, d\mu(\lambda)$. Evaluating in $v \in V_{\mathbb{C}}$, we obtain for the positive measure $\mu^{v,v} := \mu(\,\cdot\,)(v,v)$ the relation

$$\psi(t)(v,v) = \int_{\mathbb{R}} e^{it\lambda}\, d\mu^{v,v}(\lambda).$$

This function extends to a bounded holomorphic function ψ on $\mathbb{C}_+$. In particular, the Laplace transform $\mathcal{L}(\mu^{v,v})(t) = \psi(it)(v,v)$ is bounded, which implies that $\mathrm{supp}(\mu^{v,v}) \subseteq [0,\infty)$ (cf. [Neeb 2000, Remark V.4.12]). This implies that μ is supported on $[0,\infty)$.

(ii) $\Rightarrow$ (iii): Write $U_t := U^\psi_t = e^{itH}$ with the selfadjoint generator H. We show that $H \geq 0$. Let E be the spectral measure of H, so that $H = \int_{\mathbb{R}} \lambda\, dE(\lambda)$ and $U_t = \int_{\mathbb{R}} e^{it\lambda}\, dE(\lambda)$. It suffices to show that, for every $f \in L^1(\mathbb{R})$ for which the Fourier transform $\hat{f}(\lambda) = \int_{\mathbb{R}} e^{i\lambda t} f(t)\, dt$ vanishes on $\mathbb{R}_+$, the operator

$$\begin{aligned} U_f = \int_{\mathbb{R}} f(t) e^{itH}\, dt &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(t) e^{it\lambda}\, dE(\lambda)\, dt \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(t) e^{it\lambda}\, dt\, dE(\lambda) = \int_{\mathbb{R}} \hat{f}(\lambda)\, dE(\lambda) = \hat{f}(H) \end{aligned}$$

vanishes. For $v, w \in V$, we obtain with (ii) that

$$\begin{aligned} \langle j(v), U_f j(w)\rangle &= \int_{\mathbb{R}} f(t) \langle j(v), U_t j(w)\rangle\, dt = \int_{\mathbb{R}} f(t) \int_0^\infty e^{it\lambda}\, d\mu^{v,w}(\lambda)\, dt \\ &= \int_0^\infty \int_{\mathbb{R}} f(t) e^{it\lambda}\, dt\, d\mu^{v,w}(\lambda) = \int_0^\infty \hat{f}(\lambda)\, d\mu^{v,w}(\lambda) = 0 \end{aligned}$$

if $\hat{f}$ vanishes on $\mathbb{R}_+$. This proves that $j(V) \subseteq \ker(U_f)$ and since U_f is an intertwining operator and the subspace $j(V) \subseteq \mathcal{H}_\psi$ is generating, it follows that $U_f = 0$. This implies that $H \geq 0$.

(iii) $\Rightarrow$ (i): Write $U_t := U_t^\psi = e^{itH}$ and assume that $H \geq 0$. The spectral calculus for selfadjoint operators now implies that $\widehat{U}_z := e^{izH}$, $\operatorname{Im} z \geq 0$ defines a strongly continuous representation on the upper half plane $\mathbb{C}_+$ which is holomorphic on the interior and whose range consists of contractions ([Neeb 2000, Chapter VI]). Then

$$\widehat{\psi}(z)(v, w) = \langle j(v), \widehat{U}_z j(w)\rangle = \langle j(v), e^{izH} j(w)\rangle, \qquad v, w \in V, \operatorname{Im} z \geq 0,$$

provides the bounded analytic extension of ψ to $\mathbb{C}_+$.

Now we assume that (i)–(iii) are satisfied. Writing $\psi(t)(v, w) = \langle j(v), U_t j(w)\rangle$ for a linear map $j : V \to \mathcal{H}$ and a unitary one-parameter group $U_t = e^{itH}$ on $\mathcal{H}$, we have $H \geq 0$ by (iii) and

$$f(t, \tau^\varepsilon) = \langle j(v), e^{-|t|H} j(w)\rangle,$$

so that the positive definiteness of f follows from the positive definiteness of the function $t \mapsto e^{-|t|H}$ on $\mathbb{R}$ [Neeb and Ólafsson 2014, Proposition 4.1]. $\square$

Appendix A. Some background on positive definite kernels

In this appendix we collect precise statements of some basic facts on positive definite kernels and functions to keep the paper more self-contained.

Form-valued positive definite kernels.

Definition A.1. Let X be a set and V be a real vector space. We write $\operatorname{Bil}(V) = \operatorname{Bil}(V, \mathbb{C})$ for the space of complex-valued bilinear forms on V. We call a map $K : X \times X \to \operatorname{Bil}(V)$ *a positive definite kernel* if the associated scalar-valued kernel

$$K^\flat : (X \times V) \times (X \times V) \to \mathbb{C}, \quad K^\flat((x, v), (y, w)) := K(x, y)(v, w)$$

is positive definite.[2] The corresponding reproducing kernel Hilbert space $\mathcal{H}_{K^\flat} \subseteq \mathbb{C}^{X \times V}$ is generated by the elements $K^\flat_{x,v}$, $x \in X, v \in V$, with the inner product

$$\langle K^\flat_{(x,v)}, K^\flat_{(y,w)}\rangle = K(x, y)(v, w) =: K^\flat((x, v), (y, w)) =: K^\flat_{y,w}(x, v),$$

so that, for all $f \in \mathcal{H}_{K^\flat}$, we have

$$f(x, v) = \langle K^\flat_{x,v}, f\rangle.$$

We identify $\mathcal{H}_{K^\flat}$ with a subspace of $(V^*)^X$ by identifying $f \in \mathcal{H}_{K^\flat}$ with the function $f^* : X \to V^*$, $f^*(x) := f(x, \cdot)$. We call

$$\mathcal{H}_K := \{f^* : f \in \mathcal{H}_{K^\flat}\} \subseteq (V^*)^X$$

---

[2] This definition is adapted to our convention that scalar products are linear in the second argument. Accordingly, a kernel $K : X \times X \to \operatorname{Bil}(V)$ is positive definite in the sense of Definition A.1 if and only if the kernel $(x, y) \mapsto K(x, y)^\top$ is positive definite in the sense of [Neeb and Ólafsson 2015b].

the (*vector-valued*) *reproducing kernel space associated to* K. The elements

$$K_{x,v} := (K^\flat_{x,v})^* \quad \text{with} \;\; K_{x,v}(y) = K(y,x)(\,\cdot\,, v) \;\; \text{for} \;\; x, y \in X,\, v, w \in V,$$

then form a dense subspace of $\mathcal{H}_K$ with

$$\langle K_{x,v}, K_{y,w}\rangle = K(x,y)(v,w). \tag{31}$$

Example A.2. If V is a complex Hilbert space, X is a set and $K : X \times X \to B(V)$ is an operator-valued kernel, then K is called positive definite if the corresponding kernel

$$\widetilde{K} : (X \times V) \times (X \times V) \to \mathbb{C}, \quad \widetilde{K}((x,v),(y,w)) := \langle v, K(x,y)w\rangle$$

is positive definite [Neeb 2000, Definition I.1.1], and this means that the kernel

$$K' : X \times X \to \mathrm{Sesq}(V) \subseteq \mathrm{Bil}(V), \quad K'(x,y)(v,w) := \langle v, K(x,y)w\rangle$$

is positive definite.

If $X = G$ is a group and the kernel K is invariant under right translations, then it is of the form $K(g,h) = \varphi(gh^{-1})$ for a function $\varphi : G \to \mathrm{Bil}(V)$.

Definition A.3. Let G be a group and let V be a real vector space. A function $\varphi : G \to \mathrm{Bil}(V)$ is said to be *positive definite* if the $\mathrm{Bil}(V)$-valued kernel $K(g,h) := \varphi(gh^{-1})$ is positive definite.

The following proposition ([Neeb and Ólafsson 2015b, Proposition A.4]) generalizes the GNS construction to form-valued positive definite functions on groups.

Proposition A.4 (GNS-construction). *Let V be a real vector space.*

(a) *Let $\varphi : G \to \mathrm{Bil}(V)$ be a positive definite function. Then $(U^\varphi_g f)(h) := f(hg)$ defines a unitary representation of G on the reproducing kernel Hilbert space $\mathcal{H}_\varphi \subseteq (V^*)^G$ with kernel $K(g,h) = \varphi(gh^{-1})$ and the range of the map*

$$j : V \to \mathcal{H}_\varphi, \quad j(v)(g)(w) := \varphi(g)(w,v), \qquad j(v) = K^\flat_{\mathbf{1},v},$$

is a cyclic subspace, i.e., $U^\varphi_G j(V)$ spans a dense subspace of $\mathcal{H}$. We then have

$$\varphi(g)(v,w) = \langle j(v), U^\varphi_g j(w)\rangle \quad \textit{for} \;\; g \in G,\; v, w, \in V. \tag{32}$$

(b) *If, conversely, $(U, \mathcal{H})$ is a unitary representation of G and $j : V \to \mathcal{H}$ a linear map whose range is cyclic, then*

$$\varphi : G \to \mathrm{Bil}(V), \quad \varphi(g)(v,w) := \langle j(v), U_g j(w)\rangle$$

is a $\mathrm{Bil}(V)$*-valued positive definite function and $(U, \mathcal{H})$ is unitarily equivalent to $(U^\varphi, \mathcal{H}_\varphi)$.*

Remark A.5. If $\varphi : G \to \mathrm{Bil}(V)$ is a positive definite function, then (32) shows that, if $\widetilde{V} := \overline{j(V)}$, which is the real Hilbert space defined by completing V with respect to the positive semidefinite form $\varphi(\mathbf{1})$, then

$$\widetilde{\varphi}(g)(v, w) = \langle v, U_g w\rangle$$

defines a positive definite function

$$\widetilde{\varphi} : G \to \mathrm{Bil}(\widetilde{V}) \quad \text{with } \widetilde{\varphi}(g)(j(v), j(w)) = \varphi(g)(v, w) \text{ for } v, w \in V.$$

Therefore it often suffices to consider $\mathrm{Bil}(V)$-valued positive definite functions for a real Hilbert space V for which $\varphi(\mathbf{1})$ is a positive definite hermitian form on V whose real part is the scalar product on V. In terms of (32), this means that $j : V \to \mathcal{H}$ is an isometric embedding of the real Hilbert space V.

Products of operator-valued kernels.

Lemma A.6. *If $K_j : X \times X \to B(V)$, $j = 1, 2$, are two positive definite kernels with the property that*

$$K_1(x, y)K_2(x', y') = K_2(x', y')K_1(x, y) \quad \textit{for } x, x', y, y' \in X,$$

then the product kernel $K := K_1 \cdot K_2$ is also positive definite.

Proof. Let $x_1, \dots, x_k$. We have to show that the operator

$$C := (K_1(x_j, x_k)K_2(x_j, x_k))_{1\le j,k\le n} \in M_n(B(V)) \cong B(V^n)$$

is positive (cf. [Neeb 2000, Remark I.1.3]).

Let $\mathcal{A}_j \subseteq B(V)$ denote the von Neumann algebra generated by the values of K_j. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ commute. Further, the matrices

$$A^{(\ell)} := (K_\ell(x_j, x_k))_{1\le j,k\le n} \in M_n(\mathcal{A}_\ell), \qquad \ell = 1, 2,$$

are positive, so [Lance 1995, Lemma 4.3] implies that the matrix

$$D := (K_1(x_j, x_k) \otimes K_2(x_j, x_k)) \in M_n(\mathcal{A}_1 \otimes \mathcal{A}_2)$$

is positive. Since C is the image of D under the canonical representation of $M_n(\mathcal{A}_1 \otimes \mathcal{A}_2)$ on V^n, it follows that C is positive. □

From real to complex-valued kernels. In this section we take a brief look at the interplay between real and complex-valued positive definite kernels. Here Corollary A.9 is of central importance because it shows how the positive definiteness of a complex-valued form $h = \gamma + i\omega$ on a real vector space V leads to a skew-symmetric contraction on the real Hilbert space V_γ.

Lemma A.7. *Let $K : X \times X \to \mathbb{C}$ be a positive definite kernel. Then the corresponding Hilbert space $\mathcal{H}_K \subseteq \mathbb{C}^X$ is invariant under complex conjugation such that $\sigma(f) := \overline{f}$ defines an antilinear isometry on $\mathcal{H}_K$ if and only if K is real-valued.*

Proof. The invariance requirement implies the relation

$$\langle f, K_x \rangle = \overline{f(x)} = \langle K_x, \sigma(f) \rangle = \langle f, \sigma(K_x) \rangle \quad \text{for } f \in \mathcal{H}_K,$$

and therefore $\sigma(K_x) = K_x$, i.e., K is real-valued. If, conversely, K is real-valued, then $\mathcal{H}_K = \mathcal{H}_K^{\mathbb{R}} \oplus i\mathcal{H}_K^{\mathbb{R}}$ is an orthogonal sum of real Hilbert spaces, so that complex conjugation acts on $\mathcal{H}_K$ as an isometry. □

Proposition A.8. *Let $A, B : X \times X \to \mathbb{R}$ be real kernels on the set X. Then the kernel*

$$K = A + iB : X \times X \to \mathbb{C}$$

is positive definite if and only if

(a) *A is positive definite, and*

(b) *there exists a skew-symmetric contractive operator C on the real reproducing kernel Hilbert space $\mathcal{H}_A^{\mathbb{R}} \subseteq \mathbb{R}^X$ with*

$$B(x, y) = \langle A_x, CA_y \rangle = (CA_y)(x) \quad \textit{for } x, y \in X.$$

Proof. **Necessity:** If K is positive definite, then so is $\overline{K} = A - iB$, and this implies that $A = \frac{1}{2}(K + \overline{K})$ is positive definite. As $A - iB = 2A - K$ is positive definite, [Neeb 2000, Theorem I.2.8][3] implies the existence of a bounded operator $D \geq 0$ on the complex reproducing kernel Hilbert space $\mathcal{H}_A \subseteq \mathbb{C}^X$ with

$$K_y(x) = K(x, y) = \langle A_x, DA_y \rangle = (DA_y)(x) \quad \text{for } x, y \in X.$$

From Lemma A.7 we know that $\mathcal{H}_A = \mathcal{H}_A^{\mathbb{R}} \oplus i\mathcal{H}_A^{\mathbb{R}}$. From the relation $A_y + iB_y = DA_y$ for every $y \in X$ and the fact that B is real-valued it thus follows that $D = \mathbf{1} + iC$ for a bounded operator C on $\mathcal{H}_A^{\mathbb{R}}$ satisfying $CA_y = B_y$ for every $y \in X$. Now $D = D^* \geq 0$ implies that $C = -C^\top$ is a contraction and

$$B(x, y) = (CA_y)(x) = \langle A_x, CA_y \rangle \quad \text{for } x, y \in X.$$

Sufficiency: Suppose, conversely, that A is positive definite and that C is a skew-symmetric contraction on the real Hilbert space $\mathcal{H}_A^{\mathbb{R}}$. Then the hermitian operator $\mathbf{1} + iC$ on $\mathcal{H}_A^{\mathbb{C}}$ is nonnegative, and therefore its symbol

$$K(x, y) := ((\mathbf{1} + iC)A_y)(x) = A(x, y) + i(CA_y)(x)$$

is a positive definite kernel on X. □

---

[3] For two positive definite kernels K and Q on a set X, the relation $\mathcal{H}_K \subseteq \mathcal{H}_Q$ is equivalent to $\lambda Q - K$ being positive definite for some $\lambda > 0$, and this in turn is equivalent to the existence of a bounded positive operator B on $\mathcal{H}_Q$ with $\|B\| \leq \lambda$ satisfying $K(x, y) = \langle Q_x, BQ_y \rangle = (BQ_y)(x)$ for $x, y \in X$ [Neeb 2000, Theorem I.2.8].

Corollary A.9. *Let V be a real vector space, let $\gamma : V \times V \to \mathbb{R}$ be a symmetric and $\omega : V \times V \to \mathbb{R}$ be a skew-symmetric bilinear form and consider the corresponding hermitian form $h := \gamma + i\omega$. Then the following are equivalent*:

(i) *h is a positive definite kernel on V.*

(ii) *γ is positive semidefinite and there exists a skew-symmetric bounded operator C on the real Hilbert space V_γ obtained by completing $V/\{v \in V : \gamma(v,v) = 0\}$ such that $\omega(v,w) = \langle [v], C[w] \rangle_{V_\gamma}$, where $[v]$ denotes the image of v in V_γ.*

(iii) *γ is positive semidefinite and*

$$\omega(v,w)^2 \leq \gamma(v,v)\gamma(w,w) \quad \textit{for } v, w \in V. \tag{33}$$

Proof. (i) $\Leftrightarrow$ (ii): In view of Proposition A.8, the kernel h is positive definite if and only if the kernel γ is positive definite, i.e., γ is a positive semidefinite form, and the kernel ω can be written as

$$\omega(v,w) = \langle [v], C[w] \rangle_{V_\gamma} \quad \text{for } v, w \in V, \tag{34}$$

where C is a skew-symmetric contraction on the real Hilbert space V_γ.

(ii) $\Rightarrow$ (iii): (34) and $\|C\| \leq 1$ imply that

$$\omega(v,w)^2 \leq \|C\|^2 \|[v]\|^2 \|[w]\|^2 = \gamma(w,w)\gamma(v,v).$$

(iii) $\Rightarrow$ (ii): Suppose, conversely, that γ is positive semidefinite and that (33) is satisfied. Then ω defines a continuous bilinear form on the real Hilbert space V_γ with norm ≤ 1. Hence there exists a skew-symmetric contraction $C \in B(V_\gamma)$ satisfying (34). This proves the corollary. ☐

Lemma A.10. *Let $h = \gamma + i\omega$ be a positive definite kernel as in Corollary A.9, let $\mathcal{H}_h \subseteq \mathrm{Hom}(V, \mathbb{C})$ be the corresponding reproducing kernel Hilbert space and let $j : V \to \mathcal{H}_h$, $j(v) = h(\cdot, v)$ be the canonical map. The following assertions hold*:

(i) *j is injective if and only if γ is positive definite, i.e., defines an inner product on V.*

(ii) *The complex linear extension $j_{\mathbb{C}} : V_{\mathbb{C}} \to \mathcal{H}_h, v + iw \mapsto j(v) + i \cdot j(w)$ is injective if and only if*

$$\omega(v,w)^2 < \gamma(v,v)\gamma(w,w) \quad \textit{for } 0 \neq v, w \in V.$$

(iii) *Suppose that γ is positive definite, that (V, γ) is complete and that $\omega(v,w) = \langle [v], C[w] \rangle$ for an operator C on $\mathcal{H}_\gamma^{\mathbb{R}} \cong (V, \gamma)$. Then $j_{\mathbb{C}}$ is injective if and only if $\|Cv\| < \|v\|$ for every nonzero $v \in \mathcal{H}_\gamma^{\mathbb{R}}$.*

Proof. (i) In view of $\langle j(v), j(w) \rangle = \langle h(\cdot, v), h(\cdot, w) \rangle = h(v,w)$, we have $\|j(v)\|^2 = h(v,v) = \gamma(v,v)$, so that j is injective if and only if γ is positive definite.

(ii) First we calculate

$$\begin{aligned}\|j_{\mathbb{C}}(v+iw)\|^2 &= \|j(v)+i\cdot j(w)\|^2 = \gamma(v,v)+\gamma(w,w)+2\,\mathrm{Re}\langle j(v), i\cdot j(w)\rangle \\ &= \gamma(v,v)+\gamma(w,w)+2\,\mathrm{Re}\, ih(w,v) \\ &= \gamma(v,v)+\gamma(w,w)+2\omega(v,w).\end{aligned}$$

Writing $\omega(v,w)=\langle\gamma_w, C\gamma_v\rangle$ as in (34), it follows that $j_{\mathbb{C}}(v+iw)=0$ is equivalent to

$$2\langle\gamma_v, C\gamma_w\rangle = \langle\gamma_v,\gamma_v\rangle+\langle\gamma_w,\gamma_w\rangle. \tag{35}$$

Next we observe that $j(v)=-i\cdot j(w)$ implies $\gamma(v,v)=\|j(v)\|^2=\|j(w)\|^2=\gamma(w,w)$, which leads to

$$\langle\gamma_v, C\gamma_w\rangle = \|\gamma_v\|^2 = \|\gamma_w\|^2 = \|\gamma_v\|\cdot\|\gamma_w\|.$$

As C is a contraction, this is equivalent to $C\gamma_v=\gamma_w$ by the Cauchy–Schwarz inequality.

If, conversely, there exists a nonzero $v\in V$ with $C\gamma_v=\gamma_w$ and $\gamma(v,v)=\gamma(w,w)$, then $j_{\mathbb{C}}(v+iw)=0$ by (35). This proves (ii).

(iii) If (V,γ) is complete, $j(V)\cong(V,\gamma)$ is closed in $\mathcal{H}_h$. Therefore $Cj(V)\subseteq j(V)$, and (iii) follows from the preceding discussion. □

Remark A.11. If $V\subseteq\mathcal{H}$ is a standard real subspace (Definition 2.4), then the kernel $h(v,w):=\langle v,w\rangle$ on V has the property that the corresponding reproducing kernel Hilbert space is $\mathcal{H}$ and the inclusion is the corresponding map $j: V\to\mathcal{H}$. In particular, its complex linear extension is injective.

If, conversely, $h=\gamma+i\omega$ is a positive definite bilinear kernel on a real vector space V, then $j(V)$ is a standard real subspace of the corresponding complex Hilbert space $\mathcal{H}_h$ if and only if (V,γ) is complete (which is equivalent to the closedness of $j(V)$) and the complex linear extension $j_{\mathbb{C}}: V_{\mathbb{C}}\to\mathcal{H}_h$ is injective, which is equivalent to $j(V)\cap i\cdot j(V)=\{0\}$ (cf. Lemma A.10(iii)).

Example A.12. Consider the context of Proposition A.8, where $K=A+iB$ is a positive definite kernel and $C\in B(\mathcal{H}_A^{\mathbb{R}})$ is such that $B_y=CA_y$ for $y\in X$. Then

$$V:=(\mathbf{1}+iC)\mathcal{H}_A^{\mathbb{R}}\subseteq\mathcal{H}_A$$

is a real subspace. For the isometric antilinear involution defined on $\mathcal{H}_A$ by $\sigma(f)=\overline{f}$, we then have for every $f\in\mathcal{H}_A^{\mathbb{R}}$ the relation

$$\langle\sigma(\mathbf{1}+iC)f, (\mathbf{1}+iC)f\rangle = \|f\|^2-\|Cf\|^2\geq 0.$$

Therefore $(\mathcal{H}_A, V, \sigma)$ is a reflection positive real Hilbert space (Proposition B.3).

Real parts of positive definite functions. Let $\varphi : G \to \mathbb{C}$ be a positive definite function on the group G. Then $\overline{\varphi}$ is also positive definite, so that $\operatorname{Re}\varphi = \frac{1}{2}(\varphi + \overline{\varphi})$ is positive definite as well. From Lemma A.7(a) we know that a positive definite function φ on G is real-valued if and only if the corresponding reproducing kernel Hilbert space $\mathcal{H}_\varphi$ is invariant under conjugation with $\|\overline{f}\| = \|f\|$ for $f \in \mathcal{H}_\varphi$. Based on these observations, one would like to understand the set of all positive definite functions with a given real part. A natural description of this set in the spirit of the present paper is provided by the following theorem.

Theorem A.13 (Complex extensions of real positive definite functions). *Let $\varphi : G \to \mathbb{R}$ be a positive definite function and let $(U^\varphi, \mathcal{H}_\varphi^{\mathbb{R}})$ denote the corresponding orthogonal representation on the real reproducing kernel space $\mathcal{H}_\varphi^{\mathbb{R}} \subseteq \mathbb{R}^G$ by right translations*: $(U^\varphi(g)f)(h) := f(hg)$. *Then the following assertions hold*:

(a) *For each skew-symmetric contraction C on $\mathcal{H}_\varphi$ commuting with $U^\varphi(G)$, the function $\varphi_C := \varphi + iC\varphi \in \mathcal{H}_\varphi \subseteq \mathbb{C}^G$ is positive definite. Here we consider φ as an element of the real Hilbert space $\mathcal{H}_\varphi^{\mathbb{R}} \subseteq \mathbb{R}^G$.*

(b) *Each positive definite function $\widehat{\varphi}$ with* $\operatorname{Re}\widehat{\varphi} = \varphi$ *is of the form φ_C for a unique skew-symmetric contraction C on $\mathcal{H}_\varphi$ commuting with $U^\varphi(G)$.*

Proof. (a) Clearly $\mathcal{H}_\varphi = \mathcal{H}_\varphi^{\mathbb{R}} \oplus i\mathcal{H}_\varphi^{\mathbb{R}}$ is the Hilbert space complexification of $\mathcal{H}_\varphi^{\mathbb{R}}$ (Lemma A.7). On $\mathcal{H}_\varphi$ the operator $B := \mathbf{1} + iC$ is positive because it is hermitian and $\|C\| \leq 1$. Let $K(x, y) := \varphi(xy^{-1})$ be the kernel corresponding to φ which satisfies $K_y = U^\varphi(y)^{-1}\varphi$. Then the associated kernel

$$\begin{aligned} K^B(x, y) &:= \langle BK_y, K_x\rangle = \langle BU^\varphi(y)^{-1}\varphi, U^\varphi(x)^{-1}\varphi\rangle \\ &= \langle U^\varphi(y)^{-1}B\varphi, U^\varphi(x)^{-1}\varphi\rangle \\ &= \langle U^\varphi(xy^{-1})(\mathbf{1} + iC)\varphi, \varphi\rangle = ((\mathbf{1} + iC)\varphi)(xy^{-1}) \end{aligned}$$

is positive definite (cf. [Neeb 2000, Lemma I.2.4]), and this means that $\varphi + iC\varphi$ is a positive definite function.

(b) If $\widehat{\varphi} = \varphi + i\psi$ is positive definite with φ, ψ real-valued, then write $K = A + iB$ for the corresponding kernels:

$$K(x, y) = \widehat{\varphi}(xy^{-1}), \qquad A(x, y) = \varphi(xy^{-1}) \qquad \text{and} \qquad B(x, y) = \psi(xy^{-1}).$$

Then Proposition A.8 implies that φ is positive definite and that there exists a skew-symmetric contraction $C \in B(\mathcal{H}_\varphi^{\mathbb{R}})$ with

$$\psi(xy^{-1}) = (CA_y)(x) = \langle CU^\varphi(y)^{-1}\varphi, U^\varphi(x)^{-1}\varphi\rangle.$$

Since this kernel on $G \times G$ is invariant under right translations and $U^\varphi(G)\varphi$ is total in $\mathcal{H}_\varphi^{\mathbb{R}}$, it follows that C commutes with $U^\varphi(G)$. This in turn leads to

$$\psi(xy^{-1}) = \langle C\varphi, U^\varphi(yx^{-1})\varphi\rangle = (C\varphi)(xy^{-1})$$

and hence to $\psi = C\varphi$. □

Appendix B. Standard real subspaces via contractions

In this section we show how standard real subspaces can be parametrized in a very convenient way by skew-symmetric contractions in real Hilbert spaces. The survey article [Longo 2008] is an excellent source for the theory of standard real subspaces.

Skew symmetric contractions.

Lemma B.1. *Let C_V be a skew-symmetric contraction on the real Hilbert space E and $V := (\mathbf{1} + iC_V)E \subseteq E_{\mathbb{C}}$. For $0 \neq v \in E$, the following are equivalent*:

(i) $C_V^2 v = -v$.

(ii) $\|C_V v\| = \|v\|$.

(iii) *There exists* $0 \neq w \in V$ *with* $\langle C_V v, w\rangle = \|v\|\,\|w\|$.

(iv) $(\mathbf{1} + iC_V)v \in V \cap iV$.

Proof. (i) ⇔ (ii): First we observe that $\|v\|^2 - \|C_V v\|^2 = \langle(\mathbf{1} + C_V^2)v, v\rangle$. In view of the positivity of $\mathbf{1} + C_V^2$, the relation $\langle(\mathbf{1} + C_V^2)v, v\rangle = 0$ is equivalent to $(\mathbf{1} + C_V^2)v = 0$.

(ii) ⇔ (iii) follows from $\max\{\langle C_V v, w\rangle : w \in E, \|w\| \leq 1\} = \|C_V v\| \leq \|v\|$.

(iv) ⇔ (i): For $w \in E$, the condition $(\mathbf{1} + iC_V)v = i(\mathbf{1} + iC_V)w$ is equivalent to $C_V w = -v$ and $w = C_V v$. Such an element w exists if and only if $C_V^2 v = -v$. □

Lemma B.2. *For a skew-symmetric contraction C_V on the real Hilbert space E and $V := (\mathbf{1} + iC_V)E \subseteq E_{\mathbb{C}}$, the following are equivalent*:

(i) $C_V^2 + \mathbf{1}$ *is injective.*

(ii) $\|C_V v\| < \|v\|$ *for every nonzero* $v \in E$.

(iii) $\langle C_V v, w\rangle < \|v\|\,\|w\|$ *for nonzero elements* $v, w \in E$.

(iv) $V \cap iV = \{0\}$.

(v) *The operators* $\mathbf{1} \pm iC_V$ *on* $E_{\mathbb{C}}$ *are injective.*

(vi) $V + iV$ *is dense in* $E_{\mathbb{C}}$.

(vii) V *is a standard real subspace.*

Proof. The equivalence of (i)–(iv) follows immediately from Lemma B.1.

Further, (iv) can also be formulated as: $(\mathbf{1}+iC_V)(v+iw)=0$ for $v, w \in E$ implies $v+iw=0$, which in turn means that $\mathbf{1}+iC_V$ is injective. This in turn is equivalent to $\mathbf{1}-iC_V$ being injective. Therefore (iv) is equivalent to (v).

As $V+iV=(\mathbf{1}+iC_V)E_{\mathbb{C}}=\operatorname{im}(\mathbf{1}+iC_V)$, this complex subspace is dense if and only if the hermitian operator $\mathbf{1}+iC_V$ has dense range, and this is equivalent to $\mathbf{1}+iC_V$ being injective. Therefore (v) and (vi) are also equivalent.

Next we observe that V is closed because

$$\|(\mathbf{1}+iC_V)v\|^2=\|v\|^2+\|C_Vv\|^2\geq\|v\|^2 \quad \text{for } v\in E$$

shows that the range V of the operator $\mathbf{1}+iC_V : E\to E_{\mathbb{C}}$ is closed. Since (iv) and (vi) are equivalent, they are therefore equivalent to V being a standard real subspace. ☐

Proposition B.3. *Let E be a real Hilbert space, C_V be a skew-symmetric contraction on E, let $E_{\mathbb{C}}$ be the complexification of E and let $\sigma : E_{\mathbb{C}}\to E_{\mathbb{C}},\ a+ib\mapsto a-ib$ be complex conjugation on $E_{\mathbb{C}}$. Then the real subspace*

$$V:=(\mathbf{1}+iC_V)E\subseteq E_{\mathbb{C}}$$

has the following properties:

(i) *Let $E_0=\ker(C_V^2+\mathbf{1})$ and $E_1=E_0^{\perp}$, so that $E=E_0\oplus E_1$. Then $C_0:=C_V|_{E_0}$ is a complex structure on E_0 and $V_0:=(\mathbf{1}+iC_V)E_0\subseteq E_{\mathbb{C}}$ is the $(-i)$-eigenspace of C_V. It coincides with $V\cap iV$. In particular it is a complex subspace of $E_{\mathbb{C}}$. The subspace $V_1:=(\mathbf{1}+iC_V)E_1$ is a standard real subspace of $E_{1,\mathbb{C}}$.*

(ii) *If $V=V_1$, then the corresponding modular objects are given by*

$$(\Delta, J)=\left(\left(\frac{\mathbf{1}-iC_V}{\mathbf{1}+iC_V}\right)^2, \sigma\right).$$

Proof. (i) For $a, b\in E$, the relation $C_V(a+ib)=-i(a+ib)$ is equivalent to $C_Va=b$ and $C_Vb=-a$, i.e., to $a+ib\in V_0$. Therefore V_0 is the $(-i)$-eigenspace of C_V in $E_{\mathbb{C}}$. From Lemma B.1(iv) we further obtain $V\cap iV=V_0$. For $V_1:=(\mathbf{1}+iC_V)E_1$, we thus have $V_1\cap iV_1=\{0\}$, so that Lemma B.2(vii) implies that V_1 is a standard real subspace of $E_{1,\mathbb{C}}$.

(ii) If $V=V_1$, then

$$\Delta:=\left(\frac{\mathbf{1}-iC_V}{\mathbf{1}+iC_V}\right)^2 \tag{36}$$

is a positive selfadjoint operator on $E_{\mathbb{C}}$ with domain $(\mathbf{1}+iC_V)^2E_{\mathbb{C}}$. Further $\Delta^{1/2}=(\mathbf{1}-iC_V)(\mathbf{1}+iC_V)^{-1}$ has domain $V_{\mathbb{C}}$

Since $\sigma\Delta\sigma = \Delta^{-1}$ by (36), $S := \sigma\Delta^{1/2}$ is an unbounded antilinear involution with

$$\mathrm{Fix}(S) = \{\xi \in \mathcal{D}(\Delta^{1/2}) = V_{\mathbb{C}} : S\xi = \xi\}.$$

For $\xi = (\mathbf{1} + iC_V)v,\ v \in E_{\mathbb{C}}$, we have

$$S\xi = \sigma\Delta^{1/2}\xi = \sigma(\mathbf{1} - iC_V)v = (\mathbf{1} + iC_V)\sigma(v),$$

so $S\xi = \xi$ is equivalent to $v \in V$. We conclude that $\mathrm{Fix}(S) = V$. This proves (ii). □

Remark B.4. Let C be a skew-symmetric contraction on the real Hilbert space E. Then the selfadjoint operator $C^2 + \mathbf{1}$ is invertible if and only if $-\mathbf{1} \notin \mathrm{Spec}(C^2)$, which is equivalent to $\mathbf{1} \notin \mathrm{Spec}(iC)$, where iC is considered as a selfadjoint operator on the complex Hilbert space $E_{\mathbb{C}}$. This, in turn, is equivalent to the invertibility of $\mathbf{1} + iC$ and hence to the boundedness of $(\mathbf{1} - iC)(\mathbf{1} + iC)^{-1}$.

Real reflection positivity and standard subspaces. In this section we relate standard real subspaces to reflection positive real Hilbert spaces of the form $(E_{\mathbb{C}}, V, \sigma)$, where σ is the complex conjugation on the complexification $E_{\mathbb{C}}$ of a real Hilbert space. This sheds an interesting light on the close connection between standard real subspaces and reflection positivity.

Lemma B.5. *Let E be a real Hilbert space and $E_{\mathbb{C}}$ be its complexification. On $E_{\mathbb{C}}$ we consider the antilinear isometry defined by $\sigma(a + ib) := a - ib$. A real subspace $V \subseteq E_{\mathbb{C}}$ has the property that the form $(v, w) \mapsto \langle \sigma v, w\rangle$ is real-valued and positive semidefinite on V if and only if there exists a skew-symmetric contraction $C_V : \mathcal{D}(C_V) \to E$ with $V = (\mathbf{1} + iC_V)(\mathcal{D}(C_V))$. The subspace V is closed if and only if $\mathcal{D}(C_V)$ is closed.*

Proof. First, let $C_V : \mathcal{D}(C_V) \to E$ be a skew-symmetric contraction and put $V := (\mathbf{1} + iC_V)\mathcal{D}(C_V)$. For $v, w \in \mathcal{D}(C_V)$, we then have

$$\begin{aligned}\langle \sigma((\mathbf{1}+iC_V)v), (\mathbf{1}+iC_V)w\rangle &= \langle (\mathbf{1}-iC_V)v), (\mathbf{1}+iC_V)w\rangle\\ &= \langle v, w\rangle + \langle -iC_V v, w\rangle + \langle v, iC_V w\rangle - \langle C_V v, C_V w\rangle\\ &= \langle v, w\rangle - \langle C_V v, C_V w\rangle = \langle (\mathbf{1}+C_V^2)v, w\rangle \in \mathbb{R}.\end{aligned}$$

Moreover $\mathbf{1} + C_V^2 \geq 0$ implies that the form is positive semidefinite.

Conversely, let $V \subseteq E_{\mathbb{C}}$ be a real subspace which is σ-positive in the sense that the form $f(v, w) := \langle \sigma v, w\rangle$ is real-valued and positive semidefinite. This assumption implies that $V \cap iE = \{0\}$. Hence there exists a real linear operator $C_V : \mathcal{D}(C_V) \to E$ for which $V = (\mathbf{1} + iC_V)\mathcal{D}(C_V)$. Since

$$\begin{aligned}\langle \sigma(v + iC_V v), w + iC_V w\rangle &= \langle v - iC_V v, w + iC_V w\rangle\\ &= \langle v, w\rangle - \langle C_V v, C_V w\rangle + i(\langle C_V v, w\rangle + \langle C_V w, v\rangle)\end{aligned}$$

is supposed to be real-valued,

$$\langle C_V v, w\rangle + \langle v, C_V w\rangle = 0 \quad \text{for } v, w \in E.$$

This means that C_V is skew-symmetric on $\mathcal{D}(C_V)$. Further, the positivity assumption implies that $\|C_V v\| \leq \|v\|$ for $v \in E$.

The subspace V is closed if and only if the graph of C_V is closed, which is equivalent to the closedness of $\mathcal{D}(C_V)$ because C_V is a contraction. ☐

Proposition B.6. *Let E be a real Hilbert space, let C_V be a skew-symmetric contraction on E, let $E_{\mathbb{C}}$ be the complexification of E and let $\sigma : E_{\mathbb{C}} \to E_{\mathbb{C}}$, $a + ib \mapsto a - ib$ be complex conjugation on $E_{\mathbb{C}}$. Then the real subspace*

$$V := (\mathbf{1} + iC_V)E \subseteq E_{\mathbb{C}}$$

has the following properties:

- (i) *V is closed and σ-positive, so that $(E_{\mathbb{C}}, V, \sigma)$ is a reflection positive real Hilbert space.*
- (ii) *$V^{\perp} = i\sigma(V)$, i.e., the bilinear form $\gamma_\sigma(\xi, \eta) := \langle \sigma\xi, \eta\rangle$ on V is real-valued.*
- (iii) *The null space of the positive semidefinite form γ_σ on V coincides with the $(-i)$-eigenspace V_0 of C_V on $E_{\mathbb{C}}$. If $V_0 = \{0\}$, then the unbounded positive operator*

$$F := \sqrt{\frac{\mathbf{1} - iC_V}{\mathbf{1} + iC_V}} : V \to E_{\mathbb{C}}$$

satisfies $\|F\xi\|^2 = \langle \sigma\xi, \xi\rangle$ for $\xi \in V$, so that we can identify the real Hilbert space completion $\widehat{V}$ of V with respect to γ_σ with $\overline{F(V)}$. We further have $\sigma F \sigma = F^{-1}$.

Proof. (i) The subspace V is closed because

$$\|(\mathbf{1} + iC_V)v\|^2 = \|v\|^2 + \|C_V v\|^2 \geq \|v\|^2 \quad \text{for } v \in E$$

shows that the range of the operator $\mathbf{1} + iC_V : E \to V$ is closed.

For the complex conjugation σ on $E_{\mathbb{C}}$, we have for $v, w \in E$ the relation

$$\begin{aligned}\gamma_\sigma((\mathbf{1}+iC_V)v, (\mathbf{1}+iC_V)w) &= \langle \sigma(\mathbf{1}+iC_V)v, (\mathbf{1}+iC_V)w\rangle \\ &= \langle (\mathbf{1}-iC_V)v, (\mathbf{1}+iC_V)w\rangle \\ &= \langle (\mathbf{1}+iC_V)(\mathbf{1}-iC_V)v, w\rangle = \langle (\mathbf{1}+C_V^2)v, w\rangle \in \mathbb{R}\end{aligned}$$

and thus

$$\gamma_\sigma((\mathbf{1} + iC_V)v, (\mathbf{1} + iC_V)v) = \|v\|^2 - \|C_V v\|^2 \geq 0.$$

(ii) An element $a+ib \in E_{\mathbb{C}}$ $(a, b \in E)$ is orthogonal to V with respect to the real scalar product if and only if

$$0 = \mathrm{Re}\langle a+ib, v+iC_V v\rangle = \langle a, v\rangle + \langle b, C_V v\rangle = \langle a - C_V b, v\rangle$$

for every $v \in E$; this is equivalent to $C_V b = a$, i.e., to $a+ib = i(b - iC_V b) \in i\sigma(V)$.

(iii) An element $\xi := (\mathbf{1}+iC_V)v \in V$ satisfies $\langle \sigma\xi, \xi\rangle = 0$ if and only if $C_V^2 v = -v$, which is equivalent to

$$(\mathbf{1}-iC_V)\xi = (\mathbf{1}-iC_V)(\mathbf{1}+iC_V)v = (\mathbf{1}+C_V^2)v = 0,$$

i.e., to $C_V\xi = -i\xi$. This implies that $V_0 \subseteq V$ is the nullspace of γ_σ.

Now we assume that $V_0 = \{0\}$ and $V = V_1$. As $\mathbf{1} \pm iC_V$ are nonnegative hermitian operators on $E_{\mathbb{C}}$, they have a nonnegative square root and $(\mathbf{1}+iC_V)^{-1/2}$ is an unbounded operator whose domain is

$$\sqrt{\mathbf{1}+iC_V}E_{\mathbb{C}} \supseteq \sqrt{\mathbf{1}+iC_V}\sqrt{\mathbf{1}+iC_V}E_{\mathbb{C}} = (\mathbf{1}+iC_V)E_{\mathbb{C}}.$$

This leads to an unbounded symmetric operator

$$F := \sqrt{\frac{\mathbf{1}-iC_V}{\mathbf{1}+iC_V}} : V \to E_{\mathbb{C}}.$$

For $\xi = (\mathbf{1}+iC_V)v,\ v \in E$, we have

$$F\xi = \sqrt{(\mathbf{1}-iC_V)(\mathbf{1}+iC_V)}v = \sqrt{\mathbf{1}+C_V^2}v,$$

so $\|F\xi\|^2 = \langle(\mathbf{1}+C_V^2)v, v\rangle = \langle\sigma\xi, \xi\rangle$. Therefore $F : V \to \widehat{V} := \overline{F(V)} \subseteq E_{\mathbb{C}}$ is the canonical map of the reflection positive real Hilbert space $(E_{\mathbb{C}}, V, \sigma)$. It satisfies

$$\sigma F\sigma = \sqrt{\frac{\mathbf{1}+iC_V}{\mathbf{1}-iC_V}} = F^{-1}. \qquad \square$$

Remark B.7. Since $U_t = \Delta^{-it}$ acts on the reflection positive Hilbert space $(E_{\mathbb{C}}, V, \sigma)$ by automorphisms, it induces on the corresponding real Hilbert space $\widehat{V}$ an orthogonal representation. The natural map $\sqrt{\mathbf{1}+C_V^2} : E \to \widehat{V}$ in Proposition B.6 intertwines the orthogonal representations $U_t|_E$ and $U_t|_{\widehat{V}}$.

The following proposition asserts that all standard real subspaces are of the form described in Proposition B.3.

Proposition B.8. *Let $V \subseteq \mathcal{H}$ be a standard real subspace with modular objects (Δ, J). Then $E := \mathrm{Fix}(J)$ is a real Hilbert space with $\mathcal{H} \cong E_{\mathbb{C}}$ and there exists a skew-symmetric strict contraction $C_V : E \to E$ with $V = (\mathbf{1}+iC_V)E$. Then $\mathcal{D}(\Delta) \cap V$ is dense in V.*

Proof. First we observe that V is J-positive:

$$\langle J\xi, \xi\rangle = \langle JS\xi, \xi\rangle = \langle \Delta^{1/2}\xi, \xi\rangle \geq 0.$$

This implies the existence of a contraction $C_V : \mathcal{D}(C_V) \to E$ with

$$V = \Gamma(C_V) := (\mathbf{1} + iC_V)\mathcal{D}(C_V)$$

(Section 3B). That C_V is strict follows from Lemma B.2. From the real orthogonal decomposition $\mathcal{H} = V \oplus iJ(V)$ [Neeb and Ólafsson 2015b, Lemma 4.2(iv)], we now obtain

$$V^{\perp} = iJ(V) = i(\mathbf{1} - iC_V)\mathcal{D}(C_V) = i\Gamma(-C_V) = (C_V + i\mathbf{1})\mathcal{D}(C_V),$$

where $\perp$ refers to the real-valued scalar product $\operatorname{Re}\langle\,\cdot\,,\cdot\,\rangle$ on $\mathcal{H} \cong E \oplus iE$.

If $a \in E \cap \mathcal{D}(C_V)^{\perp}$, then $a \in V^{\perp} = iJ(V) = i\Gamma(-C_V)$ leads to $a = C_V 0 = 0$. Therefore $\mathcal{D}(C_V)$ is dense in E. As V is closed and $\mathbf{1} + iC_V : \mathcal{D}(C_V) \to V$ is a topological isomorphism, it follows that $\mathcal{D}(C_V)$ is closed, and thus $\mathcal{D}(C_V) = E$.

As $\gamma_J(\xi, \eta) := \langle J\xi, \eta\rangle$ is real-valued on V (recall $JV = (iV)^{\perp}$), we obtain for $v, w \in V$ the relation

$$\begin{aligned} 0 &= \operatorname{Im}\langle J(\mathbf{1} + iC_V)v, (\mathbf{1} + iC_V)w\rangle = \operatorname{Im}\langle(\mathbf{1} - iC_V)v, (\mathbf{1} + iC_V)w\rangle \\ &= \operatorname{Im}\langle(\mathbf{1} - iC_V^{\top})(\mathbf{1} - iC_V)v, w\rangle = -\langle(C_V^{\top} + C_V)v, w\rangle, \end{aligned}$$

so that $C_V^{\top} = -C_V$ (Lemma B.5).

It remains to show that $\mathcal{D}(\Delta) \cap V$ is dense in V. Since C_V is a strict contraction, the kernel of $\mathbf{1} + C_V^2$ is trivial, i.e., -1 is not an eigenvalue of C_V^2. Let $E_n \subseteq E$ be the spectral subspace of C_V^2 for the subset $[-1 + 1/n, 1]$. This subspace is C_V-invariant and the union of these subspace is dense in E because -1 is not an eigenvalue. As $(\mathbf{1} + iC_V)E_n \subseteq \mathcal{D}(\Delta)$, it follows that $\mathcal{D}(\Delta) \cap V$ is dense in V. ☐

Contractions and modular objects. The following lemma describes the complex-valued scalar product on a standard real subspace in terms of the corresponding modular objects (Δ, J).

Lemma B.9. *Let $V \subseteq \mathcal{H}$ be a standard real subspace, (Δ, J) be the corresponding modular objects and*

$$\langle v, w\rangle_{\mathcal{H}} = \gamma(v, w) + i\omega(v, w)$$

be the corresponding hermitian positive definite form on V; in particular $\langle v, w\rangle_V = \gamma(v, w)$. Then

$$\begin{aligned} \gamma(v, w) &= \frac{1}{2}\big(\langle v, w\rangle + \langle \Delta^{1/2}v, \Delta^{1/2}w\rangle\big), \\ \omega(v, w) &= \frac{1}{2i}\big(\langle v, w\rangle - \langle \Delta^{1/2}v, \Delta^{1/2}w\rangle\big). \end{aligned} \tag{37}$$

In particular, we have a strict contraction C on V satisfying

$$\omega(v,w)=\gamma(v,Cw) \quad \textit{and} \quad C=\widehat{C}|_V, \tag{38}$$

where

$$\widehat{C}=i\frac{\Delta-\mathbf{1}}{\Delta+\mathbf{1}}=i\frac{\Delta^{1/2}-\Delta^{-1/2}}{\Delta^{1/2}+\Delta^{-1/2}}=i\tanh\Big(\frac{\log\Delta}{2}\Big).$$

Moreover,

$$\langle v,w\rangle_{\mathcal{H}}=\langle v,(\mathbf{1}+iC)w\rangle_{V_{\mathbb{C}}} \quad \textit{for } v,w\in V_{\mathbb{C}}, \tag{39}$$

so that the map

$$\Phi:=\sqrt{\mathbf{1}+iC}:V_{\mathbb{C}}\to V_{\mathbb{C}}$$

extends to a unitary isomorphism $\mathcal{H}\hookrightarrow V_{\mathbb{C}}$.

Proof. As $V\subseteq\mathcal{D}(\Delta^{1/2})$ and $v=Sv=J\Delta^{1/2}v$ or $v\in V$ (Remark 2.5), we obtain

$$\langle\Delta^{1/2}v,\Delta^{1/2}w\rangle=\langle Jv,Jw\rangle=\langle w,v\rangle=\overline{\langle v,w\rangle} \quad \text{for } v,w\in V.$$

This implies (37). Next we note that

$$B:=\frac{\Delta-\mathbf{1}}{\Delta+\mathbf{1}}$$

is a bounded operator on $\mathcal{H}$ which can also be written as

$$B=\frac{\Delta^{1/2}-\Delta^{-1/2}}{\Delta^{1/2}+\Delta^{-1/2}}.$$

In this form we see that $JBJ=-B$. We also note that B commutes with Δ, and hence preserves $\mathcal{D}(\Delta^{1/2})$. This leads to

$$SB=J\Delta^{1/2}B=-BS,$$

and therefore to $BV=B\,\mathrm{Fix}(S)\subseteq i\,\mathrm{Fix}(S)=iV$. In particular, $\widehat{C}:=iB$ restricts to a bounded skew-symmetric operator $C:V\to V$. If v,w are contained in the dense subspace $V\cap\mathcal{D}(\Delta)$ of V (Proposition B.8), we obtain

$$\begin{aligned}\gamma(v,Cw)&=\tfrac{1}{2}\big(\langle v,Cw\rangle+\langle\Delta^{1/2}v,\Delta^{1/2}Cw\rangle\big)\\&=\tfrac{1}{2}\langle(\mathbf{1}+\Delta)v,Cw\rangle\\&=\tfrac{1}{2}\langle v,(\mathbf{1}+\Delta)\widehat{C}w\rangle\\&=\frac{1}{2i}\langle v,(\mathbf{1}-\Delta)w\rangle=\omega(v,w).\end{aligned}$$

Since ω and $\gamma(\,\cdot\,,C\cdot\,)$ are continuous on V, they coincide on all of V. By Lemma B.2, the operator C is a strict contraction. By (38), we have for $v,w\in V$ the relation (39), and since both sides are sesquilinear, it also holds for $v,w\in V_{\mathbb{C}}$. This implies the existence of an isometric extension $\Phi : \mathcal{H} \to V_{\mathbb{C}}$ of the operator $\sqrt{\mathbf{1} + iC}$ on $V_{\mathbb{C}}$. To see that Φ is unitary, we observe that

$$\mathrm{im}(\Phi)^{\perp} = ((\mathbf{1} + iC)^{1/2} V_{\mathbb{C}})^{\perp} = \ker(\mathbf{1} + iC)^{1/2} = \ker(\mathbf{1} + iC),$$

and this space is trivial by Lemma B.2. □

References

[Araki 1963] H. Araki, "A lattice of von Neumann algebras associated with the quantum theory of a free Bose field", *J. Math. Phys.* **4** (1963), 1343–1362. MR Zbl

[Barata et al. 2013] J. C. A. Barata, C. D. Jäkel, and J. Mund, "The $P(\varphi)_2$ model on the de Sitter space", preprint, 2013. arXiv

[Barata et al. 2016] J. C. A. Barata, C. D. Jäkel, and J. Mund, "Interacting quantum fields on de Sitter space", 2016. To appear in *Mem. Amer. Math. Soc.* arXiv

[Borchers 1992] H.-J. Borchers, "The CPT-theorem in two-dimensional theories of local observables", *Comm. Math. Phys.* **143**:2 (1992), 315–332. MR Zbl

[Bratteli and Robinson 1979] O. Bratteli and D. W. Robinson, *Operator algebras and quantum statistical mechanics, I*, Springer, 1979. MR Zbl

[Bratteli and Robinson 1981] O. Bratteli and D. W. Robinson, *Operator algebras and quantum statistical mechanics, II*, Springer, 1981. MR Zbl

[Bros and Buchholz 1994] J. Bros and D. Buchholz, "Towards a relativistic KMS-condition", *Nuclear Phys. B* **429**:2 (1994), 291–318. MR Zbl

[Buchholz et al. 2011] D. Buchholz, G. Lechner, and S. J. Summers, "Warped convolutions, Rieffel deformations and the construction of quantum field theories", *Comm. Math. Phys.* **304**:1 (2011), 95–123. MR Zbl

[Cuniberti et al. 2001] G. Cuniberti, E. De Micheli, and G. A. Viano, "Reconstructing the thermal Green functions at real times from those at imaginary times", *Comm. Math. Phys.* **216**:1 (2001), 59–83. MR Zbl

[De Angelis et al. 1986] G. F. De Angelis, D. de Falco, and G. Di Genova, "Random fields on Riemannian manifolds: a constructive approach", *Comm. Math. Phys.* **103**:2 (1986), 297–303. MR Zbl

[De Micheli and Viano 2012] E. De Micheli and G. A. Viano, "Holomorphic extensions associated with series expansions", *Forum Math.* **24**:6 (2012), 1269–1316. MR Zbl

[Dereziński and Gérard 2013] J. Dereziński and C. Gérard, *Mathematics of quantization and quantum fields*, Cambridge Univ. Press, 2013. MR Zbl

[Dimock 2004] J. Dimock, "Markov quantum fields on a manifold", *Rev. Math. Phys.* **16**:2 (2004), 243–255. MR Zbl

[Eckmann and Osterwalder 1973] J.-P. Eckmann and K. Osterwalder, "An application of Tomita's theory of modular Hilbert algebras: duality for free Bose fields", *J. Funct. Anal.* **13**:1 (1973), 1–12. MR Zbl

[Fulling and Ruijsenaars 1987] S. A. Fulling and S. N. M. Ruijsenaars, "Temperature, periodicity and horizons", *Phys. Rep.* **152**:3 (1987), 135–176. MR

[Gérard and Jäkel 2007] C. Gérard and C. D. Jäkel, "On the relativistic KMS condition for the $P(\varphi)_2$ model", pp. 125–140 in *Rigorous quantum field theory*, edited by A. Boutet de Monvel et al., Progr. Math. **251**, Birkhäuser, Basel, 2007. MR Zbl

[Glimm and Jaffe 1981] J. Glimm and A. Jaffe, *Quantum physics: a functional integral point of view*, Springer, 1981. MR Zbl

[Haag et al. 1967] R. Haag, N. M. Hugenholtz, and M. Winnink, "On the equilibrium states in quantum statistical mechanics", *Comm. Math. Phys.* **5** (1967), 215–236. MR Zbl

[Jaffe and Ritter 2007] A. Jaffe and G. Ritter, "Quantum field theory on curved backgrounds, I: The Euclidean functional integral", *Comm. Math. Phys.* **270**:2 (2007), 545–572. MR Zbl

[Jorgensen 2002] P. E. T. Jorgensen, "Diagonalizing operators with reflection symmetry", *J. Funct. Anal.* **190**:1 (2002), 93–132. MR Zbl

[Jorgensen and Ólafsson 1998] P. E. T. Jorgensen and G. Ólafsson, "Unitary representations of Lie groups with reflection symmetry", *J. Funct. Anal.* **158**:1 (1998), 26–88. MR Zbl

[Jorgensen and Ólafsson 2000] P. E. T. Jorgensen and G. Ólafsson, "Unitary representations and Osterwalder–Schrader duality", pp. 333–401 in *The mathematical legacy of Harish-Chandra* (Baltimore, 1998), edited by R. S. Doran and V. S. Varadarajan, Proc. Sympos. Pure Math. **68**, Amer. Math. Soc., Providence, RI, 2000. MR Zbl

[Klein and Landau 1981] A. Klein and L. J. Landau, "Periodic Gaussian Osterwalder–Schrader positive processes and the two-sided Markov property on the circle", *Pacific J. Math.* **94**:2 (1981), 341–367. MR Zbl

[Klein and Landau 1983] A. Klein and L. J. Landau, "From the Euclidean group to the Poincaré group via Osterwalder–Schrader positivity", *Comm. Math. Phys.* **87**:4 (1983), 469–484. MR Zbl

[Lance 1995] E. C. Lance, *Hilbert C^*-modules*, London Math. Soc. Lecture Note Series **210**, Cambridge Univ. Press, 1995. MR Zbl

[Longo 2008] R. Longo, "Real Hilbert subspaces, modular theory, $\mathrm{SL}(2,\mathbb{R})$ and CFT", pp. 33–91 in *Von Neumann algebras in Sibiu* (Sibiu, Romania, 2007), edited by K. Dykema and F. Rădulescu, Theta Ser. Adv. Math. **10**, Theta, Bucharest, 2008. MR Zbl

[Neeb 2000] K.-H. Neeb, *Holomorphy and convexity in Lie theory*, de Gruyter Expositions in Math. **28**, de Gruyter, Berlin, 2000. MR Zbl

[Neeb and Ólafsson 2014] K.-H. Neeb and G. Ólafsson, "Reflection positivity and conformal symmetry", *J. Funct. Anal.* **266**:4 (2014), 2174–2224. MR Zbl

[Neeb and Ólafsson 2015a] K.-H. Neeb and G. Ólafsson, "Reflection positive one-parameter groups and dilations", *Complex Anal. Oper. Theory* **9**:3 (2015), 653–721. MR Zbl

[Neeb and Ólafsson 2015b] K.-H. Neeb and G. Ólafsson, "Reflection positivity for the circle group", *J. Phys. Conf. Ser.* **597**:1 (2015), art. id. 012004.

[Neeb and Ólafsson 2019] K.-H. Neeb and G. Ólafsson, "Reflection positivity on spheres", (2019).

[Osterwalder and Schrader 1973] K. Osterwalder and R. Schrader, "Axioms for Euclidean Green's functions", *Comm. Math. Phys.* **31**:2 (1973), 83–112. MR Zbl

[Ramacher 2000] P. Ramacher, "Modular localization of elementary systems in the theory of Wigner", *J. Math. Phys.* **41**:9 (2000), 6079–6089. MR Zbl

[Rieffel and van Daele 1977] M. A. Rieffel and A. van Daele, "A bounded operator approach to Tomita–Takesaki theory", *Pacific J. Math.* **69**:1 (1977), 187–221. MR Zbl

[Schlingemann 1999] D. Schlingemann, "From Euclidean field theory to quantum field theory", *Rev. Math. Phys.* **11**:9 (1999), 1151–1178. MR Zbl

[Yngvason 1994] J. Yngvason, "A note on essential duality", *Lett. Math. Phys.* **31**:2 (1994), 127–141. MR Zbl

Received October 31, 2016.

KARL-HERMANN NEEB
DEPARTMENT MATHEMATIK
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
ERLANGEN
GERMANY

neeb@math.fau.de

GESTUR ÓLAFSSON
DEPARTMENT OF MATHEMATICS
LOUISIANA STATE UNIVERSITY
BATON ROUGE, LA
UNITED STATES

olafsson@math.lsu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 1, 2019
dx.doi.org/10.2140/pjm.2019.299.171

IMPROVED BUCKLEY'S THEOREM ON LOCALLY COMPACT ABELIAN GROUPS

VICTORIA PATERNOSTRO AND EZEQUIEL RELA

We present sharp quantitative weighted norm inequalities for the Hardy–Littlewood maximal function in the context of locally compact abelian groups, obtaining an improved version of the so-called Buckley's theorem. On the way, we prove a precise reverse Hölder inequality for Muckenhoupt A_∞ weights and provide a valid version of the "open property" for Muckenhoupt A_p weights.

1. Introduction and main results

The study of weighted norm inequalities for maximal type operators is one of the central topics in harmonic analysis that began with the celebrated theorem of Muckenhoupt [1972]. It states that the class of weights (nonnegative locally integrable functions) characterizing the boundedness of the Hardy–Littlewood maximal function M on the weighted space $L^p(\mathbb{R}^n, w dx)$ is the so-called Muckenhoupt A_p class (see below for the precise definitions). It is important to remark that Muckenhoupt's result is qualitative, that is, it does not provide any precise information on how the operator norm of M depends on the underlying weight in $w \in A_p$. The first quantitative result on the boundedness for the maximal function in $\mathbb{R}^n$ dates back to the 90s, is due to Buckley [1993], and gives the best possible power dependence on the A_p constant $[w]_{A_p}$. More precisely, Buckley proved

$$\|M\|_{L^p(\mathbb{R}^n, w dx) \to L^p(\mathbb{R}^n, w dx)} \le C[w]_{A_p}^{1/(p-1)}, \qquad 1 < p < \infty. \tag{1-1}$$

Recently a simpler and elegant proof was presented by Lerner [2008], who used a very clever argument composing weighted versions of the maximal function. Since then, finer improvements have been found. In particular, there is in [Hytönen et al. 2012] a sharp mixed bound valid in the context of spaces of homogeneous type.

Both authors are partially supported by grants UBACyT 20020130100403BA, CONICET-PIP 11220150100355 and PICT 2014-1480.

MSC2010: primary 42B25; secondary 43A70.

Keywords: locally compact abelian groups, reverse Hölder inequality, Muckenhoupt weights, maximal functions.

Our purpose here is to obtain sharp quantitative norm estimates in the context of locally compact abelian groups (LCA groups). The modern approach to this problem is to use a sharp version of the reverse Hölder inequality (RHI) with a precise quantitative expression for the exponent to derive a proper open property for the A_p classes. Then an interpolation type argument allows us to prove the desired bound.

In the rest of the introduction we first describe in detail the context where we will work in and then properly state the results that we will prove.

1A. ***Muckenhoupt weights and maximal functions on LCA groups.*** In the euclidean setting, the standard way to introduce A_p weights is by considering averages over cubes, balls or more general families of convex sets. In any case, the family is built using some specific metric. In our context of LCA groups we lack such a concept. However there are many LCA groups where we do have the possibility of considering a family of *base sets* satisfying the other fundamental property of the basis of cubes or balls: any point has a family of decreasing base sets shrinking to it and, in addition, the whole space can be covered by the increasing union of such a family.

In order to properly define the A_p classes, let us fix an LCA group G with a measure μ that is inner regular and such that $\mu(K) < \infty$ for every compact set $K \subset G$. Notice that μ does need to be the Haar measure because we do not assume μ to be translation invariant. The reader can find a comprehensive treatment of harmonic analysis on LCA groups in [Hewitt and Ross 1970; 1963; Rudin 1962]. The general assumption on the group will be that it admits a sequence of neighborhoods of 0 with certain properties that we described in the next definition (cf. [Edwards and Gaudry 1977, Section 2.1]).

Definition 1.1. A collection $\{U_i\}_{i\in\mathbb{Z}}$ is a covering family for G if

(1) $\{U_i\}_{i\in\mathbb{Z}}$ is an increasing base of relatively compact neighborhoods of 0, $\bigcup_{i\in\mathbb{Z}} U_i = G$ and $\bigcap_i U_i = \{0\}$.

(2) There exists a positive constant $D \geq 1$ and an increasing function $\theta : \mathbb{Z} \to \mathbb{Z}$ such that for any $i \in \mathbb{Z}$ and any $x \in G$,
   - $i \leq \theta(i)$,
   - $U_i - U_i \subset U_{\theta(i)}$,
   - $\mu(x + U_{\theta(i)}) \leq D\mu(x + U_i)$.

We will refer to the third condition as the *doubling* property of the measure μ with respect to θ and we will call D the doubling constant. In the case of $\mathbb{R}^n$ equipped with the natural metric and measure, we can consider the family of dyadic cubes of sidelength 2^i or the euclidean balls $B(x, 2^i)$ for $i \in \mathbb{Z}$. The doubling constant of the Lebesgue measure in this context is 2^n and the function θ can be taken to be $\theta(i) = i + 1$. Therefore, the intuition here is that the index i in the above definition can be seen as a sort of *radius* or *size* of the given set U_i.

For each $x \in G$, the set $x + U_i$ will be called a *base set* and the collection of all base sets will be denoted by

$$\mathcal{B} := \{x + U_i : x \in G,\ i \in \mathbb{Z}\}. \tag{1-2}$$

The notion of base sets allows us to define a direct analogue of the Hardy–Littlewood maximal function:

$$Mf(x) = \sup_{x \in U \in \mathcal{B}} \fint_U |f| \, d\mu := \sup_{x \in U \in \mathcal{B}} \frac{1}{\mu(U)} \int_U |f| \, d\mu, \tag{1-3}$$

where the supremum is taken over the sets $U \in \mathcal{B}$ with positive measure.

As we already mentioned, our purpose here is to prove sharp weighted norm inequalities for this operator in $L^p(G, w d\mu)$, where w is a weight on G. Firstly, recall that the celebrated Muckenhoupt's theorem asserts that the class of weights characterizing the boundedness of M on $L^p(\mathbb{R}^n, w dx)$, $p > 1$, is the Muckenhoupt A_p class defined in $\mathbb{R}^n$ by

$$[w]_{A_p(\mathbb{R}^n, dx)} := \sup_Q \left(\fint_Q w \, d\mu \right) \left(\fint_Q w^{1-p'} \, d\mu \right)^{p-1} < \infty. \tag{1-4}$$

Here p' denotes the conjugate exponent of p defined by the condition $\frac{1}{p} + \frac{1}{p'} = 1$. In the case of LCA groups the analogue of (1-4) is obtained by replacing the cubes by base sets. More precisely, a weight w is an $A_p = A_p(G, d\mu)$ weight if

$$[w]_{A_p} := \sup_{U \in \mathcal{B}} \left(\fint_U w \, d\mu \right) \left(\fint_U w^{1-p'} \, d\mu \right)^{p-1} < \infty. \tag{1-5}$$

The limiting case of (1-5), when $p = 1$, defines the class A_1; that is, the set of weights w such that

$$[w]_{A_1} := \sup_{U \in \mathcal{B}} \left(\fint_U w \, d\mu \right) \operatorname*{ess\,sup}_U (w^{-1}) < +\infty,$$

which is equivalent to w having the property

$$Mw(x) \le [w]_{A_1} w(x) \qquad \mu\text{-almost everywhere } x \in G.$$

As in the usual setting of $\mathbb{R}^n$ we will also often refer to $\sigma := w^{1-p'}$ as the *dual weight* for w. It is easy to verify that $w \in A_p$ if and only if $\sigma \in A_{p'}$.

The family of A_p classes is increasing and this motivates the definition of the larger class A_∞ as the union $A_\infty = \bigcup_{p \ge 1} A_p$. There are many characterizations of the class A_∞ (see [Duoandikoetxea et al. 2016] or the more classical reference [Grafakos 2004]). Some of them are given in terms of the finiteness of some A_∞

constant suitably defined. The classical definition consists in taking the limit on the A_p constant as p goes to infinity, namely:

$$(w)_{A_\infty} := \sup_{U\in\mathcal{B}} \left(\fint w \, d\mu \right) \exp \fint_U \log(w^{-1}) \, d\mu. \tag{1-6}$$

However, the modern tendency is to consider the so-called Fujii–Wilson constant implicitly introduced by Fujii [1977/78], and later rediscovered by Wilson [1987; 2008], and here we choose to follow this approach by defining the A_∞ constant as

$$[w]_{A_\infty} := \sup_{U\in\mathcal{B}} \frac{1}{w(U)} \int_U M(w\chi_U) \, d\mu, \tag{1-7}$$

where $w(U) = \int_U w \, d\mu$.

1B. *Our contribution.* As we have already seen, there is a proper — and natural — way to define the A_p and A_∞ classes on LCA groups having covering families. In contrast with the case $p < \infty$, it is not immediate that the weight w belongs to A_∞ when any of constants defined on (1-6) and (1-7) is finite. In fact, a weight w is in A_∞ (that is, in some A_p) if and only if it satisfies the reverse Hölder inequality, which says

$$\left(\fint_U w^r \, d\mu \right)^{1/r} \leq C \fint_{\widehat{U}} w \, d\mu$$

for some $r > 1$ and where $\widehat{U}$ is an open set defined in terms of U (in the euclidean case $\widehat{U} = U$ and in the case of spaces of homogeneous type, it is a dilation of U). This is a very well known result in the qualitative case. Concerning the quantitative aspect, a sharp result in terms of $[w]_{A_\infty}$ in the context of spaces of homogeneous type was proved recently in [Hytönen et al. 2012].

Our first result is the following version of the RHI. Note that, as in [Hytönen et al. 2012], we are able to precisely describe the exponent r in terms of the constant $[w]_{A_\infty}$.

Theorem 1.2 (sharp weak reverse Hölder inequality). *Let $w \in A_\infty$. Define the exponent $r(w)$ as*

$$r(w) = 1 + \frac{1}{4D^{10}[w]_{A_\infty} - 1},$$

where D is the doubling constant. Then, for a fixed $U = x_0 + U_{i_0} \in \mathcal{B}$, the following inequality holds:

$$\left(\fint_U w^{r(w)} \, d\mu \right)^{1/r(w)} \leq 2D^2 \fint_{\widehat{U}} w \, d\mu, \tag{1-8}$$

where $\widehat{U}$ is the union of the base sets $\{x + U_i : x \in U, \ i \leq i_0\}$.

Once we have proven such RHI, we are able to provide a quantitative open property for A_p classes. It is very well known that the A_p classes are open in the sense that if $w \in A_p$ for some $p > 1$, then w also belongs to some $A_{p-\varepsilon}$ for some $\varepsilon > 0$. But the best possible ε in this property is not completely characterized. Another related interesting and still open question (even in the euclidean setting) is to determine, given a weight $w \in A_\infty$, the smallest $p > 1$ such that $w \in A_p$. There are some estimates in [Hagelstein and Parissis 2016] but there is no proof of their sharpness.

Here we will deduce from Theorem 1.2 an open property for A_p classes in LCA groups with some control on the constants. More precisely, given $w \in A_p$ for $1 < p < \infty$ we will obtain that $w \in A_{p-\varepsilon}$ for $\varepsilon = (p-1)/(C[\sigma]_{A_\infty})$ with $C = 4D^{10}$. Further, $[w]_{A_{p-\varepsilon}} \le 2^{p-1}D^{4p-2}[w]_{A_p}$ (see Lemma 3.1).

In a recent article, Sauer [2015] proved a weighted bound for the maximal function for LCA groups following Lerner's approach. Additionally, he asked whether it is possible to obtain the sharp result from Buckley in this general setting. In our main theorem we answer this question in the affirmative and moreover, we provide a better mixed bound. By a mixed bound we understand a bound that depends on $[w]_{A_p}$ and $[w]_{A_\infty}$ of the form $\varphi([w]_{A_p}[w]_{A_\infty})$ where φ is some nonnegative function, typically a power function. Since $[w]_{A_\infty} \le [w]_{A_p}$ always, usually mixed type bounds are sharper than estimates involving only the A_p constant.

A result in this direction was obtained in [Hytönen et al. 2012] where the authors proved an improvement of Buckley's result (1-1) in terms of mixed bounds for spaces of homogeneous type, namely

$$\|M\|_{L^p_w \to L^p_w} \le C([w]_{A_p}[\sigma]_{A_\infty})^{1/p} \le C[w]_{A_p}^{1/(p-1)}.$$

Our main result provides an extension of the above estimate to the context of LCA groups and we will obtain it as a consequence of the RHI and the open property. We remark here that the lack of geometry in this setting constitutes a major obstacle to overcome.

Theorem 1.3. *Let M be the Hardy–Littlewood maximal function defined in (1-3) and let $1 < p < \infty$. Then there is a structural constant $C > 0$ such that*

$$\|Mf\|_{L^p_w(G)} \le C([w]_{A_p}[\sigma]_{A_\infty})^{1/p}\|f\|_{L^p_w(G)}. \tag{1-9}$$

In particular,

$$\|M\|_{L^p(w)} \le C[w]_{A_p}^{1/(p-1)}. \tag{1-10}$$

1C. *Outline.* The paper is organized as follows. In Section 2 we give some preliminary results. We prove the engulfing property in this context that will be used several times throughout the paper. We also define the local maximal function, prove a crucial covering lemma (Lemma 2.7) and show a localization property of the local maximal function. In Section 3 we give the proofs of the results described in Section 1B.

2. Preliminaries

In this section we provide some properties of covering families that we will use. Furthermore, we will introduce a local maximal function which will be crucial to proving the RHI.

As we already mentioned in the introduction, the family of dyadic cubes of sidelength 2^i or the euclidean balls $B(x, 2^i)$ for $i \in \mathbb{Z}$ are covering families for $G = \mathbb{R}$. Other examples are presented below.

Example 2.1. (1) When $G = \mathbb{T} = \{e^{2\pi it} : t \in [-\frac{1}{2}, \frac{1}{2})\}$ with the Haar measure, consider $U_k \subseteq G$ defined as $U_0 = \mathbb{T}$ and for $k \in \mathbb{N}$, $U_k = \{0\}$ and

$$U_{-k} = \left\{ e^{2\pi it} : |t| < \frac{1}{2^{k+1}} \right\}.$$

Then, $\{U_k\}_{k\in\mathbb{Z}}$ is a covering family for $\mathbb{T}$ with $\theta(k) = k + 1$ and $D = 2$.

(2) For $G = \mathbb{Z}$, take $U_i = \{k \in \mathbb{Z} : |k| \leq 2^{i-1}\}$ for $i \geq 1$ and $U_i = \{0\}$ otherwise. Then $\{U_i\}_{i\in\mathbb{Z}}$ is a covering family for $\mathbb{Z}$ with $\theta(i) = i + 1$ and $D = 2$.

(3) Let G be an LCA group with Haar measure μ and let H be a compact and open subgroup of G with $\mu(H) = 1$. Consider an expansive automorphism $A : G \to G$ with respect to H, which means that $H \subsetneq AH$ and $\bigcap_{i<0} A^i H = \{0\}$. If, additionally, $G = \bigcup_{i\in\mathbb{Z}} A^i H$, then $\{A^i H\}_{i\in\mathbb{Z}}$ is a covering family for G. Indeed, since $H \subsetneq AH$ and H is a group, $A^i H - A^i H = A^i H \subseteq A^{i+1} H$ so $\theta(i) = i + 1$. To see that the doubling property is satisfied, note that μ_A defined as $\mu_A(B) := \mu(AB)$ for $B \subseteq G$ a Borel set, is a Haar measure on G. Thus, there is a positive number α such that $\mu_A = \alpha\mu$. The constant α is the so-called *modulus of* A and is denoted by $\alpha = |A|$. Then, $\mu(A^{i+1} H) = \mu_A(A^i H) = |A|\mu(A^i H)$ for $i \in \mathbb{Z}$. Observe that G/H is discrete and AH/H is finite, so AH is the union of finitely many cosets of the quotient G/H, say $\{H + s_j\}_{j=1}^r$. Therefore, $|A| = |A|\mu(H) = \mu(AH) = r$, and $r \geq 2$ since $H \subsetneq AH$. Thus we can take $D = |A| \geq 2$. A structure of this type is considered in [Benedetto and Benedetto 2004] for constructing wavelets on LCA groups with open and compact subgroups.

For a concrete example of this situation, consider the p-adic group $G = \mathbb{Q}_p$ where $p \geq 2$ is a prime number consisting of all formal Laurent series in p with coefficients $\{0, 1, \ldots, p-1\}$, that is,

$$\mathbb{Q}_p = \left\{ \sum_{n \geq n_0} a_n p^n : n_0 \in \mathbb{Z},\ a_n \in \{0, 1, \ldots, p-1\} \right\}.$$

As a compact and open subgroup we can consider $H = \mathbb{Z}_p$ which is

$$\mathbb{Z}_p = \left\{ \sum_{n \geq 0} a_n p^n : a_n \in \{0, 1, \ldots, p-1\} \right\}.$$

Take $A : \mathbb{Q}_p \to \mathbb{Q}_p$ to be the automorphism defined as $A(x) = p^{-1}x$. Then, A is expansive with respect to $\mathbb{Z}_p$ and it can be easily checked that $\mathbb{Q}_p = \bigcup_{i\in\mathbb{Z}} A^i\mathbb{Z}_p$. Then, $\{A^i\mathbb{Z}_p\}_{i\in\mathbb{Z}}$ is a covering family for $\mathbb{Q}_p$ and in this case, $D = |A| = p$.

Let $\{U_i\}_{i\in\mathbb{Z}}$ be a fixed covering family for G. From now on, we assume the sets U_i to be symmetric. This is not a restriction at all because one can always consider the new family of base sets formed by the difference sets $U_i - U_i$ which increases the doubling constant from D to D^2. We denote $2U_i := U_i - U_i = U_i + U_i$.

Any covering family has the so-called *engulfing* property:

Lemma 2.2. *Let U, V be two base sets such that $U = x + U_i$ and $V = y + U_j$ with $i \le j$ and $x, y \in G$. If $U \cap V \neq \varnothing$, then $x + U_i \subset y + U_{\theta^2(j)}$.*

Proof. There are two points $u_i \in U_i$ and $u_j \in U_j$ such that $x + u_i = y + u_j$. Then $x = y + u_i - u_j \in y + U_j - U_j \subset y + U_{\theta(j)}$ and therefore

$$x + U_i \subset y + U_{\theta(j)} + U_{\theta(j)} \subset y + U_{\theta^2(j)}$$

(recall that we assume that the base sets are symmetric). □

Remark 2.3. For a given $V \in \mathcal{B}$, where $\mathcal{B}$ is the base of G defined as in (1-2), we will denote by $j(V) \in \mathbb{Z}$ the maximum integer such that $V = x + U_{j(V)}$ for some $x \in G$. To see that such a number exists, let us define $N(V) = \{j \in \mathbb{Z} : \exists\, x \in G,\ V = x + U_j\}$ and show that $\sup N(V) < \infty$. If $\sup N(V) = \infty$, we could find a sequence $\{x_n\}_{n\in\mathbb{N}}$ of points in G and a sequence of integer indices $\{i_n\}_{n\in\mathbb{N}}$ such that $i_n \to \infty$ as $n \to \infty$ and

$$V = x_n + U_{i_n} \qquad \text{for all } n \in \mathbb{N}.$$

By compactness of $\overline{V}$ we can assume (relabeling) that the sequence converges to some $x \in G$, which we can assume to be the origin. Now we claim that, for any $j \in \mathbb{N}$, there is some $m \in \mathbb{N}$ such that $U_j \subset x_m + U_{i_m}$ and from this fact would follow that $\mu(V) = \infty$, but this implies that $\infty = \mu(V) \le \mu(\overline{V}) < \infty$ which is a contradiction. To verify the claim, fix U_j and choose n_0 such that $x_n \in U_j$ and $i_n \ge j$ for all $n \ge n_0$. Then,

$$U_j \cap x_n + U_{i_n} \neq \varnothing$$

for all $n \ge n_0$. Furthermore, the above still holds if we replace x_n by any x_m with $m \ge n \ge n_0$ since $x_m \in U_j$ and $x_m \in x_m + U_{i_n}$. Therefore by the engulfing property (see e.g., Lemma 2.2) we obtain

$$U_j \subset x_m + U_{\theta^2(i_n)} \subset x_m + U_{i_m}$$

for any m such that $i_m \ge \theta^2(i_n)$.

In order to introduce the local maximal function, we first define a local base for a fixed base set U.

Definition 2.4. Let $U \in \mathcal{B}$ be a fixed base set and let $k := j(U)$. The local base $\mathcal{B}_U$ is defined as

$$\mathcal{B}_U := \{y + U_j : y \in U,\, j \leq k\}. \tag{2-1}$$

We also defined the *enlarged* set $\widehat{U}$ by the formula

$$\widehat{U} := \bigcup_{V \in \mathcal{B}_U} V. \tag{2-2}$$

Lemma 2.5. *Let $U = x + U_k$ be a fixed base set in $\mathcal{B}$ and set $k = j(U)$. We then have the following geometric properties*:

(a) *If $V \in \mathcal{B}_U$ then $V \subset x + U_{\theta(k)}$.*

(b) *For any $z \in U$,*

$$\widehat{U} \subset z + U_{\theta^2(k)},$$

where $\widehat{U}$ is as in (2-2). As a consequence of this last property, we obtain

$$\mu(\widehat{U}) \leq \mu(z + U_{\theta^2(k)}) \leq D^2 \mu(z + U_k)$$

for any $z \in U$. In particular, $\mu(\widehat{U}) \leq D^2 \mu(U)$, since $U = x + U_k$.

Proof. (a) Let $V = y + U_j$ with $j \leq k$ and take any $z \in V$. Then $z = y + u_j$ with $u_j \in U_j \subset U_k$. Since $y \in U$ we can write $y = x + u_k$, $u_k \in U_k$. Then we have

$$z = x + u_j + u_k \in x + U_k + U_k \subset x + U_{\theta(k)}.$$

(b) Let $V \in \mathcal{B}_U$, $V = y + U_j$ with $y \in U$, $j \leq k$. By Lemma 2.2, since $V \cap U \neq \varnothing$, $V \subset x + U_{\theta(k)}$. Take any $z \in U$, $z = x + u_k$, $u_k \in U_k$. Then,

$$V \subset x + U_{\theta(k)} = z - u_k + U_{\theta(k)} \subset z - U_k + U_{\theta(k)} \subset z - U_{\theta(k)} + U_{\theta(k)} \subset z + U_{\theta^2(k)}. \qquad \square$$

We now define the local maximal function as

$$M_U f(y) := \sup_{y \in V \in \mathcal{B}_U} \fint_V |f(z)|\, d\mu(z) \tag{2-3}$$

for any $y \in \widehat{U}$ and $M_U f(y) = 0$ otherwise.

Remark 2.6. (a) In [Hewitt and Ross 1970, Theorem 44.18], a version of the Lebesgue differentiation theorem is proven with respect to the Haar measure for LCA groups having D'-sequences (cf. [Hewitt and Ross 1970, Definition 44.10]). A careful reading of the proof of [Hewitt and Ross 1970, Theorem 44.18] reveals that the result is still true with the obvious changes for measures which are not translation invariant. Thus, since a covering family is in particular a D'-sequence, we have that the Lebesgue differentiation theorem holds in our context.

(*b*) As a consequence of the Lebesgue differentiation theorem, we have the elementary but important property of the local maximal function:

$$f(x) \leq M_U f(x) \qquad \mu\text{-almost everywhere } x \in U.$$

Consider now, for a fixed $U \in \mathcal{B}$, the level set for the local maximal function acting on a weight w at scale $\lambda > 0$:

$$\Omega_\lambda := \{x \in \widehat{U} : M_U w(x) > \lambda\}. \tag{2-4}$$

A key instrument will be a Calderón–Zygmund (C–Z) decomposition of Ω_λ. We will obtain it by using an adapted version of a covering lemma from [Edwards and Gaudry 1977, Lemma 2.2.1]. Although the proof follows standard arguments, we include it here for completeness. When w is a nonnegative and locally integrable function on G and $V \subseteq G$ is relatively compact, we denote the average of w on V as w_V; that is, $w_V = ⨍_V w \, d\mu$.

Lemma 2.7. *Let $U \in \mathcal{B}$ be a fixed base set in G and let w be a nonnegative and integrable function supported on $\widehat{U}$. For $\lambda > w_{\widehat{U}}$, define Ω_λ as in (2-4). If Ω_λ is nonempty, there exists a finite or countable index set Q and a family $\{y_i + U_{\alpha_i}\}_{i \in Q}$ of pairwise disjoint base sets from $\mathcal{B}_U$ such that*:

(a) *The sequence $\{\alpha_i\}_{i \in Q}$ is decreasing.*

(b) $\displaystyle\bigcup_{i \in Q} y_i + U_{\alpha_i} \subset \Omega_\lambda \subset \bigcup_{i \in Q} y_i + U_{\theta^2(\alpha_i)}$.

(c) *For any $i \in Q$,*

$$\lambda < ⨍_{y_i + U_{\alpha_i}} w \, d\mu.$$

(d) *Given $r > \alpha_i$ for some $i \in Q$,*

$$⨍_{y_i + U_r} w \, d\mu \leq D^2 \lambda. \tag{2-5}$$

Proof. Suppose that there is no finite sequence of points in Ω_λ such that the conclusion holds (in that case, there is nothing to prove). For $x \in \Omega_\lambda$, define

$$\alpha(x) = \max\left\{ j \in \mathbb{Z} : \exists\, V = y + U_j \in \mathcal{B}_U, x \in V, ⨍_V w \, d\mu > \lambda \right\}. \tag{2-6}$$

Since $V = y + U_j \in \mathcal{B}_U$ implies $j \leq j(U)$, we have that α is well defined. Consider now, for each $x \in \Omega_\lambda$, a base set $V_x \in \mathcal{B}_U$, $V_x := y_x + U_{\alpha(x)}$ such that $x \in V_x$. In other words, one of the base sets in $\mathcal{B}$ containing the point x where the map α attains its value. Observe that, in particular, $\alpha(y_x) \geq \alpha(x)$.

We start by picking x_1 as an extremal point for α, that is $\alpha(x_1) \geq \alpha(x)$ for all $x \in \Omega_\lambda$. Put $\alpha_1 = \alpha(x_1)$ and $y_1 := y_{x_1}$ such that $V_{x_1} = y_1 + U_{\alpha_1}$. Note that, since $\alpha_1 \leq$

$\alpha(y_1) \leq \alpha(x_1) = \alpha_1$, we also have $\alpha(y_1) = \alpha_1$. Now suppose that we have chosen the first n points $y_1, \dots, y_n$ and their respective base sets $U_{\alpha_1}, \dots, U_{\alpha_n}$ such that

- the sets $V_j := y_j + U_{\alpha_j},\ 1 \leq j \leq n$, are pairwise disjoint,
- $\alpha_j := \alpha(y_j) \geq \alpha(x)$ for all $x \in A_{j-1}$, where

$$A_j := \Omega_\lambda \setminus \bigcup_{\ell \leq j} y_\ell + U_{\theta^2(\alpha_\ell)}, \qquad 1 \leq j \leq n. \tag{2-7}$$

Since we are assuming that this procedure never ends, $A_j \neq \varnothing$ for all $1 \leq j \leq n$. Therefore we can choose $x_{n+1} \in A_n$ such that $\alpha_{n+1} := \alpha(x_{n+1}) \geq \alpha(x)$ for all $x \in A_n$. This means that there is a base set $V_{n+1} := y_{n+1} + U_{\alpha_{n+1}}$ and in particular $w_{V_{n+1}} > \lambda$ and $\alpha(y_{n+1}) = \alpha_{n+1}$. Note that this construction produces a decreasing sequence $\{\alpha_n\}_{n\in\mathbb{N}}$. Let's see that $V_{n+1} \cap V_j = \varnothing$ for all $1 \leq j \leq n$. Supposing that this is not the case, we could find $u \in U_{\alpha_{n+1}}$ and $v \in U_{\alpha_j}$ for some $j \leq n$ such that

$$y_{n+1} + u = y_j + v.$$

Since $x_{n+1} \in V_{n+1}$, we have that for some $z \in U_{\alpha_{n+1}}$,

$$x_{n+1} = y_{n+1} + z = y_j + v - u + z \in y_j + U_{\alpha_j} - U_{\alpha_{n+1}} + U_{\alpha_{n+1}}.$$

Since $U_{\alpha_{n+1}} \subset U_{\alpha_j}$ and trivially $U_{\alpha_j} \subset U_{\theta(\alpha_j)}$, we get

$$x_{n+1} \in y_j + U_{\theta^2(\alpha_j)},$$

which is a contradiction by the choice of x_{n+1}.

We are left to prove that this procedure exhausts the set Ω_λ. If not, there is a point $x \in A_n$ with $\alpha(x) \leq \alpha_n$ for all $n \geq 1$. Define the set S as

$$S := \{y_n : n \in \mathbb{N}\}.$$

Since

$$S \subset \{z \in \Omega_\lambda : \alpha(z) \geq \alpha(x)\} \subset \widehat{U}$$

and $\widehat{U}$ is contained in some base set (see item (b) in Lemma 2.5), we conclude that S is relatively compact.

By monotonicity of α, we have $U_{\alpha_n} \subset U_{\alpha_1}$. Therefore the set

$$F := \bigcup_n (y_n + U_{\alpha_n}) \subset S + U_{\alpha_1}$$

is also relatively compact and this implies $\mu(\overline{F}) < \infty$. Now consider $N \in \mathbb{Z}$ such that $\overline{S} \subset U_N$ and an integer $r > 0$ such that $\theta^r(\alpha(x)) \geq N$. Then for any $n \in \mathbb{N}$, $y_n \in S \subset U_N \subset U_{\theta^r(\alpha(x))}$ and thus $0 \in y_n + U_{\theta^r(\alpha(x))}$. Further, we get

$$U_N = 0 + U_N \subset y_n + U_{\theta^r(\alpha(x))} + U_N \subset y_n + 2U_{\theta^r(\alpha(x))} \subset y_n + U_{\theta^{r+1}(\alpha(x))}.$$

The doubling property shows

$$\mu(U_N) \le D^{r+1}\mu(y_n + U_{\alpha(x)})$$

and this implies

$$\mu(F) = \sum_n \mu(y_n + U_{\alpha_n}) \ge \sum_n \mu(y_n + U_{\alpha(x)}) \ge D^{-(r+1)} \sum_n \mu(U_N) = \infty.$$

This contradicts the condition $\mu(\bar{F}) < \infty$ and we conclude with the proof of items (a), (b) and (c) of the lemma.

We prove now item (d). To control the average on $y_i + U_r$, we consider two cases: first we consider $r \le k := j(U)$. Then $y_i + U_r \in \mathcal{B}_U$ and by maximality we have $\fint_{y_i+U_r} w\,d\mu \le \lambda$. Indeed, if not we would have $\alpha_i = \alpha(y_i) \ge r > \alpha_i$. Second, where $r > k$, we have $\theta^2(r) > \theta^2(k)$ and thus, by Lemma 2.5,

$$y_i + U_{\theta^2(r)} \supset y_i + U_{\theta^2(k)} \supset \widehat{U}.$$

Therefore, since $w = 0$ almost everywhere $\widehat{U}^c$,

$$\fint_{y_i+U_r} w\,d\mu \le \frac{\mu(\widehat{U})}{\mu(y_i + U_r)} \int_{\widehat{U}} w\,d\mu \le D^2\lambda.$$

The lemma is now completely proven. □

Now we present a localization argument for the local maximal function M_U. The idea is better understood when considering the usual dyadic maximal function M_Q^d localized on a cube Q in $\mathbb{R}^n$. Suppose that the level set $\Omega_\lambda = \{x \in Q : M_Q^d w(x) > \lambda\}$ for $\lambda > w_Q$ is decomposed into dyadic subcubes of Q such that $Q = \bigcup_i Q_i$ and the cubes Q_i are maximal with respect to the condition $w_{Q_i} > \lambda$. Then the conclusion is that for any $x \in Q_i$, the equality $M_Q^d w(x) = M_Q^d(w\chi_{Q_i})(x)$ holds. In this more general setting, the analogous result is contained in Lemma 2.8 which does not have a direct proof as in the dyadic case.

For simplicity in the exposition, we introduce the following notation. Given a base set of the form $V = y + U_j$, we denote by V^* the dilation of V by θ, i.e., $V^* = y + U_{\theta(j)}$. Further iterations of this operation are defined recursively, that is, $V^{**} = (V^*)^*$ and V^{n*} for n iterations of the dilation operation.

Lemma 2.8. *Let $U \in \mathcal{B}$ be a fixed base set and consider $w = w\chi_{\widehat{U}}$ a nonnegative and integrable function on $\widehat{U}$ where $\widehat{U}$ is as in (2-2). For $\lambda > w_{\widehat{U}}$, let Ω_λ be defined as above and let $\{V_i\}_{i\in Q} = \{y_i + U_{\alpha_i}\}_{i\in Q}$ be the C–Z decomposition of Ω_λ given by Lemma 2.7. Then, for $L = D^6$, any $i \in Q$ and any $x \in V_i^{**} \cap \Omega_{L\lambda}$,*

$$M_U w(x) \le M_U(w\chi_{V_i^{4*}})(x). \tag{2-8}$$

Proof. Let $x \in V_i^{**} \cap \Omega_{L\lambda}$. Then there exists $V \in \mathcal{B}_U$, $V = y + U_j$, with $y \in U$ and $j \le j(U)$ such that $x \in V$ and $w_V > L\lambda$. We claim $j \le \theta^2(\alpha_i)$. To see that this is in fact true, suppose towards a contradiction, that $j > \theta^2(\alpha_i)$. Then, $V \subset y_i + U_{\theta^2(j)}$. Indeed, if $z \in V$ then $z = y + w$ with $w \in U_j$. On the other hand, since $x \in V_i^{**} \cap V$, $x = y_i + u = y + v$ with $u \in U_{\theta^2(\alpha_i)}$ and $v \in U_j$. Then

$$z = y + v - v + w = x - v + w = y_i + u - v + w.$$

Since $U_{\theta^2(\alpha_i)} \subset U_j$, we get that $z \in y_i + U_j + U_{\theta(j)} \subset y_i + U_{\theta^2(j)}$. As a consequence,

$$\fint_V w \, d\mu \le \frac{\mu(y_i + U_{\theta^2(j)})}{\mu(V)} \fint_{y_i + U_{\theta^2(j)}} w \, d\mu.$$

We note that since $\theta^2(\alpha_i) < j$, $x \in V \cap V_i^{**} \subset V \cap (y_i + U_j)$ and then, by the engulfing property we have that $y_i + U_j \subset y + U_{\theta^2(j)}$. Thus, using the doubling property of the measure μ we obtain

$$\frac{\mu(y_i + U_{\theta^2(j)})}{\mu(y + U_j)} \le D^2 \frac{\mu(y_i + U_j)}{\mu(y + U_j)} \le D^2 \frac{\mu(y + U_{\theta^2(j)})}{\mu(y + U_j)} \le D^4.$$

Furthermore, since $\theta^2(j) \ge j > \theta^2(\alpha_i) \ge \alpha_i$, by item (4) in Lemma 2.7,

$$\fint_{y_i + U_{\theta^2(j)}} w \, d\mu \le D^2 \lambda$$

and we can conclude that

$$L\lambda < \fint_V w \, d\mu \le D^6 \lambda = L\lambda,$$

which gives a contradiction. Hence, the claim $j \le \theta^2(\alpha_i)$ holds.

Now, using Lemma 2.2 we have $V \subset V_i^{4*}$ and then

$$\fint_V w \, d\mu = \fint_V w \chi_{V_i^{4*}} \, d\mu \le M(w \chi_{V_i^{4*}})(x),$$

which proves inequality (2-8). □

3. Proof of the main results

We present here the proof of Theorem 1.2.

Proof of Theorem 1.2. **Step 1.** We start with the following estimate for the local maximal function. Let $U = x_0 + U_k$ be a fixed base set. We claim that, for $\varepsilon = 1/(4D^{10}[w]_{A_\infty} - 1)$,

$$\fint_{\widehat{U}} (M_U w)^{1+\varepsilon} \, d\mu \le 2[w]_{A_\infty} \left(\fint_{\widehat{U}} w \, d\mu \right)^{1+\varepsilon}. \tag{3-1}$$

Recall that we may assume that the weight w is supported on $\widehat{U}$. Let Ω_λ be defined as in (2-4). We write the norm using the layer cake formula as

$$\begin{aligned}\int_{\widehat{U}} (M_U w)^{1+\varepsilon}\, d\mu &= \int_0^\infty \varepsilon\lambda^{\varepsilon-1} M_U w(\Omega_\lambda)\, d\lambda\\ &= \int_0^{w_{\widehat{U}}} \varepsilon\lambda^{\varepsilon-1} M_U w(\Omega_\lambda)\, d\lambda + \int_{w_{\widehat{U}}}^\infty \varepsilon\lambda^{\varepsilon-1} M_U w(\Omega_\lambda)\, d\lambda\\ &= I + II.\end{aligned}$$

The first term is easily controlled by using the A_∞ constant of w (see (1-7)):

$$\begin{aligned} I \le M_U w(\widehat{U}) w_{\widehat{U}}^{\varepsilon} = w_{\widehat{U}}^{\varepsilon} \int_{\widehat{U}} M_U w\, d\mu &\\ &\le w_{\widehat{U}}^{\varepsilon} \int_{y+U_{\theta^2(k)}} M_U(w\chi_{y+U_{\theta^2(k)}})\, d\mu\\ &\le w_{\widehat{U}}^{\varepsilon} [w]_{A_\infty} w(y+U_{\theta^2(k)})\\ &= w_{\widehat{U}}^{\varepsilon} [w]_{A_\infty} w(\widehat{U}),\end{aligned}$$

where $y \in U$ and we used Lemma 2.5 and the definition of $[w]_{A_\infty}$.

Now, for each $\lambda > w_{\widehat{U}}$ we consider $\{V_i\}_{i\in Q}$ the C–Z decomposition of Ω_λ from Lemma 2.7 to control *II*. We have

$$M_U w(\Omega_\lambda) \le \sum_i M_U w(V_i^{**}).$$

For any $i \in Q$ we write $V_i^{**} = V_1 \cup V_2$ with $V_1 := V_i^{**} \cap \Omega_{L\lambda}$ and $V_2 := V_i^{**} \setminus \Omega_{L\lambda}$ where $L = D^6$. Thus, by Lemma 2.8 and the A_∞ property (1-7) we have

$$\begin{aligned} M_U w(V_i^{**}) &= \int_{V_1} M_U w\, d\mu + \int_{V_2} M_U w\, d\mu\\ &\le \int_{V_1} M_U(w\chi_{V_i^{4*}})(x)\, d\mu + L\lambda\mu(V_2)\\ &\le [w]_{A_\infty} w(V_i^{4*}) + L\lambda\mu(V_i^{4*}) = ([w]_{A_\infty} w_{V_i^{4*}} + L\lambda)\mu(V_i^{4*})\\ &\le ([w]_{A_\infty}\lambda D^2 + L\lambda) D^4 \mu(V_i) \le 2[w]_{A_\infty}\lambda D^{10}\mu(V_i),\end{aligned}$$

where in the last inequality we have used (2-5) and the doubling property of μ. This gives

$$\begin{aligned} M_U w(\Omega_\lambda) \le \sum_i M_U w(V_i^{**}) &\le 2[w]_{A_\infty}\lambda D^{10} \sum_i \mu(V_i)\\ &\le 2[w]_{A_\infty}\lambda D^{10}\mu(\Omega_\lambda).\end{aligned}$$

Thus,

$$\begin{aligned} II &\le 2[w]_{A_\infty} D^{10} \int_0^\infty \varepsilon \lambda^{\varepsilon} \mu(\Omega_\lambda)\, d\lambda \\ &= 2[w]_{A_\infty} D^{10} \frac{\varepsilon}{\varepsilon+1} \int_{\widehat{U}} M_U w^{1+\varepsilon}\, d\mu. \end{aligned}$$

Therefore, gathering all the estimations and averaging over $\widehat{U}$,

$$\Big(1 - 2[w]_{A_\infty} D^{10} \frac{\varepsilon}{\varepsilon+1}\Big) \fint_{\widehat{U}} M_U w^{1+\varepsilon}\, d\mu \le w_{\widehat{U}}^{1+\varepsilon}.$$

Choosing $\varepsilon \le 1/(4[w]_{A_\infty} D^{10} - 1)$ we get that $1 - 2[w]_{A_\infty} D^{10} \varepsilon/(\varepsilon+1) \ge \frac{1}{2}$ and we obtain the desired estimate (3-1).

Step 2. Now we proceed to prove the main estimate (1-8). By Remark 2.6, we have that $w(x) \le M_U w(x)$ holds on U. Then we obtain

$$\int_U w^{1+\varepsilon}\, d\mu \le \int_U (M_U w)^{\varepsilon} w\, d\mu \le \int_{\widehat{U}} (M_U w)^{\varepsilon} w\, d\mu.$$

Once again we use the layer cake formula combined with the C–Z decomposition of Ω_λ and proceeding much as above, we obtain

$$\begin{aligned} \int_{\widehat{U}} (M_U w)^{\varepsilon} w\, d\mu &= \int_0^\infty \varepsilon \lambda^{\varepsilon-1} w(\Omega_\lambda)\, d\lambda \\ &= \int_0^{w_{\widehat{U}}} \varepsilon \lambda^{\varepsilon-1} w(\Omega_\lambda)\, d\lambda + \int_{w_{\widehat{U}}}^\infty \varepsilon \lambda^{\varepsilon-1} w(\Omega_\lambda)\, d\lambda \\ &\le w(\widehat{U}) w_{\widehat{U}}^{\varepsilon} + \int_{w_{\widehat{U}}}^\infty \varepsilon \lambda^{\varepsilon-1} \sum_i w(V_i^{**})\, d\lambda \\ &\le w(\widehat{U}) w_{\widehat{U}}^{\varepsilon} + D^2 \int_{w_{\widehat{U}}}^\infty \varepsilon \lambda^{\varepsilon} \sum_i \mu(V_i^{**})\, d\lambda \\ &\le w(\widehat{U}) w_{\widehat{U}}^{\varepsilon} + D^4 \int_{w_{\widehat{U}}}^\infty \varepsilon \lambda^{\varepsilon} \sum_i \mu(V_i)\, d\lambda \\ &\le w(\widehat{U}) w_{\widehat{U}}^{\varepsilon} + D^4 \int_0^\infty \varepsilon \lambda^{\varepsilon} \mu(\Omega_\lambda)\, d\lambda \\ &\le w(\widehat{U}) w_{\widehat{U}}^{\varepsilon} + \frac{D^4 \varepsilon}{\varepsilon+1} \int_{\widehat{U}} (M_U w)^{1+\varepsilon}\, d\mu. \end{aligned}$$

Therefore, averaging over U, using $\mu(\widehat{U}) \le D^2 \mu(U)$ and (3-1), we have

$$\fint_U w^{1+\varepsilon}\, d\mu \le D^2 w_{\widehat{U}}^{\varepsilon+1} + \frac{2D^6 \varepsilon [w]_{A_\infty}}{\varepsilon+1} \Big(\fint_{\widehat{U}} w\, d\mu \Big)^{1+\varepsilon}.$$

By our previous choice of ε,

$$\frac{2D^6\varepsilon[w]_{A_\infty}}{\varepsilon+1} \le \frac{2D^{10}\varepsilon[w]_{A_\infty}}{\varepsilon+1} \le \frac{1}{2}$$

and we conclude that

$$\fint_U w^{1+\varepsilon}\,d\mu \le 2D^2\left(\fint_{\widehat{U}} w\,d\mu\right)^{1+\varepsilon}. \qquad \square$$

We present now some classical applications of the RHI to weighted norm inequalities for maximal functions. One crucial property of A_p classes is the well known open condition. In the next lemma we provide a quantitative version of it.

Lemma 3.1. *For* $1 < p < \infty$, *let* $w \in A_p$. *Then, for* $\varepsilon = (p-1)/(C[\sigma]_{A_\infty})$ *with* $C = 4D^{10}$ *and* $\sigma = w^{1-p'}$, *we have that* $w \in A_{p-\varepsilon}$. *Further,*

$$[w]_{A_{p-\varepsilon}} \le 2^{p-1}D^{4p-2}[w]_{A_p}.$$

Proof. Let $w \in A_p$. The $A_{p-\varepsilon}$ condition for w takes the form

$$\sup_{U\in\mathcal{B}}\left(\fint_U w\,d\mu\right)\left(\fint_U w^{1-(p-\varepsilon)'}\,d\mu\right)^{p-\varepsilon-1} < \infty.$$

Recall that the dual weight of w, $\sigma = w^{1-p'}$, is also in A_∞. Therefore it satisfies an RHI with exponent $r(\sigma)$ given by Theorem 1.2. Choose ε such that $1-(p-\varepsilon)' = (1-p')r(\sigma)$, namely $\varepsilon = (p-1)/(r(\sigma)')$ which is equivalent to the condition $r(\sigma) = (p-1)/(p-\varepsilon-1)$. Then we obtain

$$\begin{aligned}\left(\fint_U w^{1-(p-\varepsilon)'\,d\mu}\right)^{p-\varepsilon-1} &= \left(\fint_U \sigma^{(1-p')r(\sigma)}\,d\mu\right)^{(p-1)/(r(\sigma))}\\ &\le (2D^2\fint_{\widehat{U}}\sigma\,d\mu)^{p-1},\end{aligned}$$

for any $U \in \mathcal{B}$. Now, for $U = x + U_k \in \mathcal{B}$, recall that $U^{**} = x + U_{\theta^2(k)}$ and that $\widehat{U} \subset U^{**}$. Then,

$$\left(\fint_U w\,d\mu\right)\left(\fint_U w^{1-(p-\varepsilon)'}\,d\mu\right)^{p-\varepsilon-1} \le C\left(\fint_{U^{**}} w\,d\mu\right)\left(\fint_{U^{**}}\sigma\,d\mu\right)^{p-1}$$

with $C = 2^{p-1}D^{4p-2}$. We conclude that

$$[w]_{A_{p-\varepsilon}} \le 2^{p-1}D^{4p-2}[w]_{A_p}. \qquad \square$$

In what follows we will need the fact that the maximal function M maps $L^{q,\infty}_w(G)$ to itself with operator norm bounded by $C[w]^{1/q}_{A_q}$ for some $C > 0$. Without presenting any details on weak norms and Lorentz spaces, we include here a quantitative estimate on the size of level sets of the maximal function.

Lemma 3.2. *Let* $1 \le q < \infty$ *and let* M *be the maximal function defined in* (1-3). *Then*, *for any* $f \in L^q_w(G)$,

$$\sup_{\lambda>0} \lambda^q w(\{x \in G : Mf(x) > \lambda\}) \le D^{2q}[w]_{A_q} \|f\|^q_{L^q_w}. \tag{3-2}$$

Proof. For any locally integrable function f and any $\lambda > 0$, let Ω_λ be the level set $\Omega_\lambda := \{x \in G : Mf(x) > \lambda\}$. We also define some sort of *truncated* maximal operator as follows: for any $K \in \mathbb{Z}$, let M_K be the averaging operator given by

$$M_K f(x) = \sup_{V \in \mathcal{B}_K(x)} \fint_V |f(z)|\, d\mu, \tag{3-3}$$

where the supremum is taken over the subfamily $\mathcal{B}_K$ of $\mathcal{B}$ consisting of all base sets of the form $y + U_i$ with $y \in G$ and $i \le K$ containing the point x, i.e.,

$$\mathcal{B}_K(x) := \{V = y + U_i : x \in V, i \le K\}. \tag{3-4}$$

For each K we consider the corresponding level set $\Omega^K_\lambda := \{x \in G : M_K f(x) > \lambda\}$. We clearly have that the family $\{\Omega^K_\lambda\}$ is increasing in K and also $\Omega_\lambda = \bigcup_K \Omega^K_\lambda$. We therefore may compute the value of $w(\Omega_\lambda)$ as the limit of $w(\Omega^K_\lambda)$. In addition, we recall that the group G is σ-compact since $G = \bigcup_{r \in \mathbb{Z}} \overline{U}_r$. We will again use a limiting argument to compute $w(\Omega^K_\lambda)$ as the limit of $w(\Omega^K_\lambda \cap U_r)$ with $r \to +\infty$.

Now for $K \in \mathbb{Z}$ fixed, choose $r \in \mathbb{Z}$ such that $r \ge K$. A simple Vitali's covering lemma can be applied now to $\Omega^K_\lambda \cap U_r$. We want to select a countable subfamily of disjoint base sets whose dilates cover $\Omega^K_\lambda \cap U_r$. More precisely, suppose that the set $\Omega^K_\lambda \cap U_r$ is nonempty. For each $x \in \Omega^K_\lambda \cap U_r$, there exists a base set V_x of the form $V_x = y_x + U_{i_x}$ such that

$$\fint_{V_x} |f(z)|\, d\mu > \lambda. \tag{3-5}$$

Since $i_x \le K$ for all $x \in \Omega^K_\lambda \cap U_r$, there is some $i_1 = \max\{i_x\}$. We start the recursive selection procedure by picking one of these largest base sets as $V_1 = y_1 + U_{i_1}$. Now suppose that the first $V_1, V_2, \dots, V_k$ sets have been selected. We pick V_{k+1} verifying that $V_{k+1} = y_{k+1} + U_{i_{k+1}}$ where $i_{k+1} = \max\{i_x : y_x + U_{i_x} \cap V_j = \varnothing, j = 1, \dots, k\}$.

This process generates a sequence of disjoint base sets $\{V_k\}$. We note that the index sequence $\{i_k\}$ goes to $-\infty$ as k goes to infinity. If not, since it is decreasing, there would be some $i_0 = i_k$ for all $k \ge k_0$. Then we have that $V_k \cap U_r \ne \varnothing$ and $i_k \le K \le r$ and by the engulfing property, $V_k \subset U^{**}_r$ for all $k \ge k_0$. In particular, the set $S = \{y_k : k \ge k_0\} \subset U^{**}_r$ is relatively compact. Then, considering the set

$$F = \bigcup_{k \ge k_0} V_k \subset S + U_{i_0}$$

and proceeding as in Lemma 2.7 we get a contradiction.

We claim now that

$$\Omega_\lambda^K \cap U_r \subset \bigcup_{k\in\mathbb{N}} V_k^{**}.$$

To verify this, consider some $x \in \Omega_\lambda^K \cap U_r$ and the corresponding $V_x = y_x + U_{i_x}$. Suppose first that V_x intersects some V_k. Let k_0 be the smallest $k \in \mathbb{N}$ such that $V_x \cap V_k \neq \varnothing$. Then we have that $i_x \leq i_{k_0}$, since i_{k_0} was selected as the largest index among all the sets V_x disjoint from $V_1, \ldots, V_{k_0-1}$ (and by hypothesis V_x is one of them). Then the engulfing property yields

$$V_x = y_x + U_{i_x} \subset y_{k_0} + U_{\theta^2(i_{k_0})} = V_{k_0}^{**}.$$

We are left to consider the case when $V_x \cap V_k = \varnothing$ for all $k \in \mathbb{N}$. But in this case, we would have that $i_x \leq i_k$ for all k and this is a contradiction since we saw that $i_k \to -\infty$.

Summing up, we find a countable collection of base sets $\{V_k\}_k$ such that

$$⨍_{V_k} f\,d\mu > \lambda \quad \text{and} \quad \Omega_\lambda^K \cap U_r \subset \bigcup_k V_k^{**}.$$

Then we can compute

$$\begin{aligned}
\lambda^q w(\Omega_\lambda^K \cap U_r) &\leq \sum_k \lambda^q w(V_k^{**}) \leq \sum_k w(V_k^{**}) \left(⨍_{V_k} w^{-1/q} w^{1/q} |f| \right)^q \\
&\leq \sum_k \frac{w(V_k^{**})}{\mu(V_k)^q} \left(\int_{V_k} w^{1-q'}\,d\mu \right)^{q-1} \left(\int_{V_k} |f|^q w\,d\mu \right) \\
&\leq D^{2q} \sum_k \frac{w(V_k^{**})}{\mu(V_k^{**})^q} \left(\int_{V_k^{**}} w^{1-q'}\,d\mu \right)^{q-1} \left(\int_{V_k} |f|^q w\,d\mu \right) \\
&\leq D^{2q} [w]_{A_q} \sum_k \int_{V_k} |f|^q w\,d\mu \\
&\leq D^{2q} [w]_{A_q} \|f\|_{L_w^q}^q .
\end{aligned}$$

From this estimate we conclude that

$$\lambda^q w(\Omega_\lambda) \leq D^{2q} [w]_{A_q} \|f\|_{L_w^q}^q$$

for any $\lambda > 0$. □

Now we are able to present the proof of the sharp version of Buckley's theorem for the maximal function M on $L^p(G)$, $p > 1$.

Proof of Theorem 1.3. The idea is to use a sort of interpolation type argument, exploiting the sublinearity of the maximal operator M and the weak type estimate for M from Lemma 3.2. For any $f \in L^p_w(G)$ and any $t > 0$, define the truncation $f_t := f\chi_{\{|f|>t\}}$. Then, an easy computation of the averages defining M gives

$$\{x \in G : Mf(x) > 2t\} \subset \{x \in G : Mf_t(x) > t\}.$$

Now we compute the L^p_w norm as follows:

$$\begin{aligned}\|Mf\|^p_{L^p_w(G)} &= \int_0^\infty pt^{p-1} w(\{x \in G : Mf(x) > t\})\,dt \\ &= 2^p \int_0^\infty pt^{p-1} w(\{x \in G : Mf(x) > 2t\})\,dt \\ &\le 2^p \int_0^\infty pt^{p-1} w(\{x \in G : Mf_t(x) > t\})\,dt.\end{aligned}$$

We recall the open property for Muckenhoupt weights: any $w \in A_p$ also belongs to $A_{p-\varepsilon}$ for some explicit $\varepsilon > 0$ (see Lemma 3.1). Using the estimate of Lemma 3.2 for $q = p - \varepsilon$, we obtain

$$\begin{aligned}\|Mf\|^p_{L^p_w(G)} &\le 2^p p D^{2(p-\varepsilon)} [w]_{A_{p-\varepsilon}} \int_0^\infty t^{\varepsilon-1} \int_G f_t^{p-\varepsilon}(x) w(x)\,d\mu\,dt \\ &= \frac{2^p p D^{2(p-\varepsilon)} [w]_{A_{p-\varepsilon}}}{\varepsilon} \int_G |f(x)|^p w\,d\mu \\ &\le \frac{p 2^{2p-1} D^{6p-2} [w]_{A_p}}{\varepsilon} \|f\|^p_{L^p_w(G)},\end{aligned} \tag{3-6}$$

where in the last inequality we used Lemma 3.1. Noticing that in Lemma 3.1, $\varepsilon = (p-1)/(4D^{10}[\sigma]_{A_\infty})$, we finally conclude from (3-6) that

$$\|Mf\|_{L^p_w(G)} \le C([w]_{A_p}[\sigma]_{A_\infty})^{1/p} \|f\|_{L^p_w(G)}$$

and the proof of (1-9) is complete.

Finally, since $[\sigma]_{A_\infty} \le [\sigma]_{A_{p'}} = [w]^{p'-1}_{A_p}$, (1-10) follows from (1-9). □

References

[Benedetto and Benedetto 2004] J. J. Benedetto and R. L. Benedetto, "A wavelet theory for local fields and related groups", *J. Geom. Anal.* **14**:3 (2004), 423–456. MR Zbl

[Buckley 1993] S. M. Buckley, "Estimates for operator norms on weighted spaces and reverse Jensen inequalities", *Trans. Amer. Math. Soc.* **340**:1 (1993), 253–272. MR Zbl

[Duoandikoetxea et al. 2016] J. Duoandikoetxea, F. J. Martín-Reyes, and S. Ombrosi, "On the A_∞ conditions for general bases", *Math. Z.* **282**:3-4 (2016), 955–972. MR Zbl

[Edwards and Gaudry 1977] R. E. Edwards and G. I. Gaudry, *Littlewood–Paley and multiplier theory*, Ergebnisse der Mathematik **90**, Springer, 1977. MR Zbl

[Fujii 1977/78] N. Fujii, "Weighted bounded mean oscillation and singular integrals", *Math. Japon.* **22**:5 (1977/78), 529–534. MR Zbl

[Grafakos 2004] L. Grafakos, *Classical and modern Fourier analysis*, Pearson, Upper Saddle River, NJ, 2004. MR Zbl

[Hagelstein and Parissis 2016] P. Hagelstein and I. Parissis, "Weighted Solyanik estimates for the Hardy–Littlewood maximal operator and embedding of A_∞ into A_p", *J. Geom. Anal.* **26**:2 (2016), 924–946. MR Zbl

[Hewitt and Ross 1963] E. Hewitt and K. A. Ross, *Abstract harmonic analysis, I: Structure of topological groups: integration theory, group representations*, Grundlehren der Math. Wissenschaften **115**, Springer, 1963. MR Zbl

[Hewitt and Ross 1970] E. Hewitt and K. A. Ross, *Abstract harmonic analysis, II: Structure and analysis for compact groups: analysis on locally compact Abelian groups*, Grundlehren der Math. Wissenschaften **152**, Springer, 1970. MR Zbl

[Hytönen et al. 2012] T. Hytönen, C. Pérez, and E. Rela, "Sharp reverse Hölder property for A_∞ weights on spaces of homogeneous type", *J. Funct. Anal.* **263**:12 (2012), 3883–3899. MR Zbl

[Lerner 2008] A. K. Lerner, "An elementary approach to several results on the Hardy–Littlewood maximal operator", *Proc. Amer. Math. Soc.* **136**:8 (2008), 2829–2833. MR Zbl

[Muckenhoupt 1972] B. Muckenhoupt, "Weighted norm inequalities for the Hardy maximal function", *Trans. Amer. Math. Soc.* **165** (1972), 207–226. MR Zbl

[Rudin 1962] W. Rudin, *Fourier analysis on groups*, Interscience Tracts in Pure and Applied Math. **12**, Interscience, New York, 1962. MR Zbl

[Sauer 2015] J. Sauer, "An extrapolation theorem in non-Euclidean geometries and its application to partial differential equations", *J. Elliptic Parabol. Equ.* **1** (2015), 403–418. MR Zbl

[Wilson 1987] J. M. Wilson, "Weighted inequalities for the dyadic square function without dyadic A_∞", *Duke Math. J.* **55**:1 (1987), 19–50. MR Zbl

[Wilson 2008] M. Wilson, *Weighted Littlewood–Paley theory and exponential-square integrability*, Lecture Notes in Math. **1924**, Springer, 2008. MR Zbl

Received November 14, 2017. Revised May 15, 2018.

VICTORIA PATERNOSTRO
DEPARTAMENTO DE MATEMÁTICA, FACULTAD DE CIENCIAS EXACTAS Y NATURALES
UNIVERSIDAD DE BUENOS AIRES
BUENOS AIRES
ARGENTINA

vpater@dm.uba.ar

EZEQUIEL RELA
DEPARTAMENTO DE MATEMÁTICA, FACULTAD DE CIENCIAS EXACTES Y NATURALES
UNIVERSIDAD DE BUENOS AIRES
BUENOS AIRES
ARGENTINA

erela@dm.uba.ar

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 1, 2019
dx.doi.org/10.2140/pjm.2019.299.191

ETERNAL FORCED MEAN CURVATURE FLOWS II: EXISTENCE

GRAHAM SMITH

We show that under suitable nondegeneracy conditions, complete gradient flow lines of the scalar curvature functional of a riemannian manifold perturb into eternal forced mean curvature flows with large forcing term.

1. Introduction

1.1. ***Background.*** Ye [1991] shows how nondegenerate critical points of the scalar curvature function of a riemannian manifold perturb into families of convex embedded spheres inside that manifold of arbitrary large constant mean curvature. While this result has been shown to have significant applications in the study of the isoperimetric problem (see, for example, [Brendle and Eichmair 2014; Eichmair and Metzger 2012; 2013a; 2013b; Nardulli 2009; 2014]), its applications to the study of the differential topologies of spaces of immersed and embedded submanifolds have been less exploited. However, in [Smith 2011], we show how — in heuristic terms — Ye's result implies that the Euler characteristic of the space of convex Alexandrov embedded spheres inside a given manifold is equal to (-1) times the Euler characteristic of that manifold. This has applications to the study of existence, and to some measure, uniqueness, of Alexandrov embedded spheres of constant curvature for many different notions of curvature.

However, if our aim is to prove existence, then the results of [Smith 2011] are unsatisfactory when the Euler characteristic of the ambient manifold vanishes. This happens, for example, when the ambient manifold is 3-dimensional, which is nonetheless one of the most interesting cases. Furthermore, even when these techniques can be successfully applied to prove existence (as in, for example, [Maximo et al. 2017; Rosenberg and Smith 2010; White 1991]), they still often fall short of optimal results, for there are good topological reasons to believe that — at least generically — there are far more solutions than those whose existence we have managed to prove.

With this in mind, in [Smith 2015], we initiated a programme for the study of the Morse homology of the spaces of immersed and embedded hypersurfaces, where

---

MSC2010: 58C30.
Keywords: Morse homology, mean curvature, forced mean curvature flow.

the natural Morse function to be studied is the area functional or, more generally, the "Area minus Volume" functional (defined below), which depends on a parameter h, and which we denote by $\mathcal{F}_h$. The critical points of $\mathcal{F}_h$, which define the chain groups of the Morse complex [Schwarz 1993], are then immersed hypersurfaces of constant mean curvature equal to h, and its complete gradient flows, which define its ∂ operator (see [Schwarz 1993], again), are then eternal forced mean curvature flows with forcing term h.

Within this context, Ye's result says that for large values of h, nondegenerate critical points of the scalar curvature function map to (in fact, nondegenerate) critical points of $\mathcal{F}_h$. In this paper, we prove the corresponding result for complete gradient flows of the scalar curvature function. That is, under suitable nondegeneracy conditions, we show that for sufficiently large values of h, these flows map to complete gradient flows of $\mathcal{F}_h$. Combined with a suitable converse (that is, a concentration result), which has been proven in Ye's case, but which we have not yet proven here, this would mean that for large values of h, the entire Morse complex of the scalar curvature functional maps to the Morse complex of $\mathcal{F}_h$. This would make the two isomorphic, thereby yielding an explicit description of the Morse homology of the space of Alexandrov embedded spheres. In particular, since the number of constant mean curvature immersed spheres should be bounded below by the sum of the Betti numbers of this homology, we should thereby obtain stronger existence results for such hypersurfaces than those that are currently known.

Finally, it is worth observing that the results of this paper are also of interest within the classical theory of mean curvature flows. Indeed, eternal mean curvature flows in $\mathbb{R}^{m+1}$, which arise as the blow-up limits of Type II singularities [Mantegazza 2011], are still not fully understood. For example, the class of all such flows trivially includes the class of mean curvature flow solitons — that is, complete hypersurfaces which evolve by translation under the mean curvature flow (see [Martín et al. 2015] for a good survey). Since the property of being eternal ought to be quite restrictive, it is reasonable to expect that there exist no others. However, at the time of writing, the problem of determining whether all eternal mean curvature flows in $\mathbb{R}^{m+1}$ are indeed solitons remains unsolved. With this in mind, the eternal forced mean curvature flows described in this paper have come as rather a surprise to experts in the field and we hope that they may shed some light on the above problem. Finally, after completion of this paper we were made aware of the work [Alikakos and Freire 2003] which bears some similarities to our own.

1.2. ***Notation, terminology and main result.*** Let $M := M^{m+1}$ be a complete $(m+1)$-dimensional riemannian manifold. Let S be its scalar curvature function, where, throughout the paper, we adopt the convention which normalises all curvature functions so that the unit sphere in Euclidean space always has positive unit curvature.

Let $\gamma : \mathbb{R} \to M$ solve the nonlinear ODE

$$\dot{\gamma} + \frac{m+1}{2(m+3)} \nabla S = 0, \tag{1}$$

so that γ is (up to reparametrisation) a *complete gradient flow line* of S. Consider the linearisation L of (1) about γ. This is a linear ordinary differential operator from $\Gamma(\gamma^* TM)$ to itself, having the following form (see the Appendix):

$$L = \frac{\partial}{\partial t} + \frac{m+1}{2(m+3)} \operatorname{Hess}(S). \tag{2}$$

We recall that S is said to be of *Morse* type whenever all of its critical points are nondegenerate. In this case, if γ has relatively compact image, then $\gamma(t)$ converges towards critical points of S as t tends to $\pm\infty$. Furthermore, by [Robbin and Salamon 1995], L defines a Fredholm map from the space of Hölder differentiable sections $\Gamma^{k+1,\alpha}(\gamma^* TM)$ into $\Gamma^{k,\alpha}(\gamma^* TM)$, and its Fredholm index is equal to the difference of the Morse indices of the two end-points of γ. We say that γ is *nondegenerate* whenever L is surjective, and we say that S is of *Morse–Smale* type whenever, in addition to all of its critical points being nondegenerate, all of its complete gradient flow lines which have relatively compact image are also nondegenerate. This is the property that we require for the Morse complex of S to be well-defined [Schwarz 1993]. There is no shortage of metrics whose scalar curvature function has this property. Indeed, they are generic (that is, in the second category in the sense of Baire) within any conformal class (see the Appendix).

Let B^{m+1} and S^m be respectively the closed unit ball and the unit sphere in $\mathbb{R}^{m+1}$.

Definition 1.2.1. Let $\hat{\mathcal{E}}$ denote the space of smooth immersions of B^{m+1} into M and let $\mathcal{E}$ denote the quotient of this space under the action of the group of smooth orientation preserving diffeomorphisms of B^{m+1} by reparameterization.

It is usual to identify an immersion in $\hat{\mathcal{E}}$ with its equivalence class in $\mathcal{E}$. By a slight abuse of terminology, for each $e \in \mathcal{E}$, we define $\operatorname{Vol}(e)$ and $\operatorname{Area}(e)$ to be respectively the volumes of B^{m+1} and S^m with respect to the metric $e^* g$.

Definition 1.2.2. For all $h > 0$, we define the “Area minus Volume” functional by

$$\mathcal{F}_h(e) := \operatorname{Area}(e) - h \operatorname{Vol}(e). \tag{3}$$

Many properties of the immersion e are actually determined by its restriction to S^m. Indeed, the restriction operator actually defines a local homeomorphism from $\mathcal{E}$ into the space of reparametrisation equivalence classes of immersions of S^m into M whose image is the space of *Alexandrov embeddings* of S^m into M. Furthermore, an embedding $e : B^{m+1} \to M$ is a critical point of $\mathcal{F}_h$ whenever its restriction to S^m has constant mean curvature equal to h. Likewise, the family

$e : \mathbb{R} \times B^{m+1} \to M$ is an L^2 gradient flow of $\mathcal{F}_h$ whenever its restriction to $\mathbb{R} \times S^m$ is an eternal forced mean curvature flow with forcing term h. That is, whenever this restriction satisfies

$$\left\langle \frac{\partial}{\partial t} e, N_t \right\rangle + H_t - h = 0, \tag{4}$$

where N_t and H_t are respectively the outward-pointing unit normal vector field and the mean curvature of the restriction of $e_t := e(t, \cdot\,)$ to $\mathbb{R} \times S^m$.

We now introduce the mechanism by which complete gradient flow lines of S perturb to eternal forced mean curvature flows. Let γ be a complete gradient flow line of S. Using parallel transport, we identify the bundle $\gamma^* TM$ with the trivial bundle $\mathbb{R} \times \mathbb{R}^{m+1}$, and we define $\text{Exp} : \mathbb{R} \times \mathbb{R}^{m+1} \to M$ such that, for all t, $\text{Exp}_t := \text{Exp}(t, \cdot\,)$ is the exponential map of M about the point $\gamma(t)$. Now, following [Ye 1991], for all $s > 0$, for all $Y : \mathbb{R} \to \mathbb{R}^{m+1}$ and for all $f : \mathbb{R} \times S^m \to]0, \infty[$, we define the function $e(s, Y, f) : \mathbb{R} \times S^m \to M$ by

$$e(s, Y, f)(t, x) = \text{Exp}_t(sY(t) + s(1 + s^2 f(t, x))x). \tag{5}$$

Heuristically, $e(s, Y, f)$ is a smooth family of immersed spheres in M whose centres move along γ with a small displacement given by Y.

Theorem 1.2.3. *If S is of Morse–Smale type, and if γ is a complete gradient flow line of S with relatively compact image then, for all sufficiently small s, there exist $Y : \mathbb{R} \to \mathbb{R}^{m+1}$ and $f : \mathbb{R} \times S^m \to]0, \infty[$ such that, up to reparametrisation in time, $e(s, Y, f)$ is an eternal forced mean curvature flow with forcing term $1/s$.*

Remark. A detailed formal statement of Theorem 1.2.3 is given in Theorem 4.7.2 below. In particular, not only do we obtain Hölder estimates for the pair (Y, f), but we also describe in Theorem 4.6.1, below, an iterative process for determining asymptotic expansions of these solutions up to arbitrary order.

1.3. ***Discussion.*** Like Ye's result, Theorem 1.2.3 is proven by first determining formal solutions in the form of asymptotic series, and then perturbing suitably high order partial sums of these series to yield exact solutions. There are, nonetheless, considerable differences between Theorem 1.2.3 and Ye's result, primarily because Theorem 1.2.3 is a parabolic, and not an elliptic, problem. On the one hand, since parabolic and elliptic operators are all hypoelliptic, the analytic tools that we use are barely different. However, on the other, the time-dependence introduces new — and rather confusing — phenomena as the scale parameter s tends to zero.

This is perhaps best illustrated by considering the first approximation $Y = 0$ and $f = 0$. Here, the mean curvature of the sphere $e(s, 0, 0)(t, \cdot\,)$ is equal to $1/s + O(s)$, so that the forced mean curvature flow with forcing term $1/s$ should move along the curve γ with speed approximately s, which trivially tends to 0. It is perhaps surprising that this scale dependence does not actually introduce any singularities as s tends to 0. However, a deeper study of the equations involved reveals the role played by operator

$$Q_s := s^4\frac{\partial}{\partial t} + \frac{1}{m}(m+\overline{\Delta}), \tag{6}$$

where $\overline{\Delta}$ is the standard Laplacian of the sphere S^m. In fact, the fourth power of s that appears in (6) can already be seen to arise from the coefficient of f in (5) together with the slowing of the flow by a factor of s. This term will affect our work in three different ways.

First, Theorem 1.2.3 becomes a genuine singular perturbation problem. In actual fact, Ye's result, although presented as a singular perturbation problem, transforms, after removal of the first few terms and division by a suitable factor, into a regular perturbation problem, which is then directly solved by the inverse function theorem. In the present case, however, when $s = 0$, the operator Q_s is no longer hypoelliptic, and the same simplification no longer applies.

Second, since the Green's operator of Q_s depends on s, the terms in the asymptotic series of the formal solution (determined in Theorem 4.6.1, below) actually also depend on s, so that more care is required in ensuring that the Hölder bounds we obtain are independent of s.

Third, the appropriate functional analytic framework for studying parabolic operators is that of inhomogeneous spaces (introduced here in Section 4.4, below). Furthermore, the s dependence of Q_s requires the use of weighted spaces (also defined in Section 4.4, below), where what appears to be the most appropriate weighting is in fact slightly counterintuitive (see the remarks following Equation (101)).

Finally, in order to develop a Morse homology theory for the space of convex Alexandrov embedded spheres, two further results are still required. Indeed, it would be necessary to show, first that the eternal flows obtained here are nondegenerate, and second, that for sufficiently large values of the forcing term, they are the only ones. However, we believe at this stage that it is more interesting to develop a more satisfactory compactness result than that obtained in [Smith 2015], and for this reason we postpone this study to later work.

1.4. ***Overview of paper.*** This paper is structured as follows. In Section 2, we develop a formalism for the succinct description of the Taylor series of various well-known geometric functions, and in Section 3, we extend this formalism in order to describe the functions used in the proof of Theorem 1.2.3. Our objective here is to understand the general terms of these series without having to resort to explicit calculations and, for the sake of completeness, we have studied this problem in far more depth than is actually necessary for our current applications. In Section 4, we then reformulate these results in the language of asymptotic series.

In particular, since the operation of composition by smooth functions yields smooth functionals between Hölder spaces, this immediately yields norm estimates for the functionals of interest to us without any further effort being required.

Having determined the asymptotic expansion of the forced mean curvature flow operator, the rest of Section 4 is devoted to constructing formal solutions and then perturbing these formal solutions into exact solutions. It is here that we introduce the required functional analytic framework, based on the Fredholm theory of parabolic operators over weighted inhomogeneous Hölder spaces [Krylov 2008]. In addition, using the theory of spherical harmonics, we improve our norm estimates for every term in the asymptotic series of our formal solutions. Although this is not strictly necessary, we believe it makes our reasoning a lot cleaner. Finally, once formal solutions have been constructed, a straightforward application of the inverse function theorem yields the desired result.

2. The Taylor series of geometric functions

2.1. ***Curvature tensors.*** Throughout this paper, Einstein's summation convention will be used. Let Ω be the unit ball in $\mathbb{R}^{m+1}$. Let g be a smooth metric over Ω with Levi-Civita covariant derivative ∇ and Riemann curvature tensor R. We suppose that

$$\nabla_{\partial_r}\partial_r = 0 \quad \text{and} \quad g(\partial_r, \partial_r) = 1, \tag{7}$$

where ∂_r here denotes the unit radial vector field. This simply means that (Ω, g) is an exponential chart of some riemannian manifold. Now denote

$$\delta_{ij} := g(0)_{ij} \tag{8}$$

and let δ^{ij} be its metric dual. By (7), δ_{ij} is simply the standard euclidean metric over $\mathbb{R}^{m+1}$. Finally, for convenience, we suppose that Ω is convex with respect to g in the sense that for all $x, y \in \Omega$, there exists a unique geodesic in Ω from x to y.

We say that a function defined over Ω is *geometric* when it only depends on the metric g. We are interested in the Taylor series about 0 of such functions and, in particular, how their coefficients depend on the Riemann curvature tensor. In order to describe this dependence, we introduce the following algebraic formalism. Consider the set of formal tensors $X := \{(R_{i_1 i_2 i_3}{}^{j}{}_{;i_4 \ldots i_{k+3}})_{k \in \mathbb{N}}\}$ where the subscript ; here denotes formal covariant differentiation. Observe that all elements of X are covariant of order 1 and contravariant of order at least 3. Given two formal tensors, $\rho^a_{bi_1 \ldots i_p}$ and $\rho^b_{j_1 \ldots j_q}$, which are both covariant of order 1, define their *matrix product* by $\rho^a_{bi_1 \ldots i_p}\rho^b_{j_1 \ldots j_q}$, and observe that this product is also covariant of order 1. Now let $\mathcal{R}$ be the vector space with basis the set of all finite formal combinations of elements of X obtained by permutation of indices and matrix multiplication. We call $\mathcal{R}$ the space of *curvature tensors*.

For all $k \in \mathbb{Z}$, let $\mathcal{R}^k$ be the subspace of $\mathcal{R}$ consisting of those elements which are contravariant of order $k+1$. When $\rho \in \mathcal{R}^k$, we say that it has *order-difference* k. Observe that order-difference is preserved by permutation of indices, and if ρ and ρ' have order-differences k and k' respectively, then their matrix product has order-difference $k+k'$. In particular, since every generator of $\mathcal{R}$ has order-difference at least 2, it follows that for $k<2$, $\mathcal{R}^k$ is trivial, and for $k \geq 2$, it is spanned by matrix products of those generators which have order-difference at most k. Considerations such as these make it relatively straightforward to determine $\mathcal{R}^k$ for all k. For example,

$$\begin{aligned}
&\mathcal{R}^2 = \langle (R_{i_{\sigma(1)}i_{\sigma(2)}i_{\sigma(3)}}{}^{j})_{\sigma\in\Sigma_3}\rangle,\\
&\mathcal{R}^3 = \langle (R_{i_{\sigma(1)}i_{\sigma(2)}i_{\sigma(3)}}{}^{j}{}_{;i_{\sigma(4)}})_{\sigma\in\Sigma_4}\rangle,\\
(9)\qquad &\mathcal{R}^4 = \big\langle (R_{i_{\sigma(1)}i_{\sigma(2)}i_{\sigma(3)}}{}^{j}{}_{;i_{\sigma(4)}i_{\sigma(5)}},\ R_{pi_{\sigma(1)}i_{\sigma(2)}}{}^{j}R_{i_{\sigma(3)}i_{\sigma(4)}i_{\sigma(5)}}{}^{p},\\
&\qquad\qquad\qquad\qquad\qquad\qquad R_{i_{\sigma(1)}i_{\sigma(2)}p}{}^{j}R_{i_{\sigma(3)}i_{\sigma(4)}i_{\sigma(5)}}{}^{p})_{\sigma\in\Sigma_5}\big\rangle,
\end{aligned}$$

and so on, where, for all k, Σ_k denotes the group of permutations of the set $\{1,\ldots,k\}$.

Identifying elements of $\mathcal{R}$ via the symmetries of the Riemann curvature tensor, we obtain

Proposition 2.1.1. *$\mathcal{R}$ is self-adjoint with respect to δ in the sense that if $\rho^{i}_{j_1\cdots j_k}$ is an element of $\mathcal{R}$, then $\delta^{ia}\delta_{j_lb}\rho^{b}_{j_1\cdots j_{l-1}aj_{l+1}\cdots j_k}$ identifies with a unique element of $\mathcal{R}$ for all $1\leq l\leq k$.*

Proof. It suffices to prove the result for each generator of $\mathcal{R}$. We thus show that $\delta^{ia}\delta_{j_lb}R_{j_1j_2j_3}{}^{b}{}_{;j_4\cdots j_{l-1}aj_{l+1}\cdots j_{k+3}}$ identifies with a unique element of $\mathcal{R}$ for all k and for all $1\leq l\leq k+3$. We achieve this by induction on k. Indeed, for $k=0$, the result follows directly from the symmetries of the Riemann curvature tensor. For $k=1$, it follows from these symmetries together with the second Bianchi identity. Now suppose that $k\geq 2$. Since the set of generators of $\mathcal{R}$ is closed under formal covariant differentiation, so too is $\mathcal{R}$, and we may therefore suppose that $l=k+3$. However,

$$\begin{aligned}
R_{j_1j_2j_3}{}^{i}{}_{;j_4\cdots j_{k+2}j_l} = R_{j_1j_2j_3}{}^{i}{}_{;j_4\cdots j_{k+1}j_lj_{k+2}} &+ R_{j_{k+2}j_lp}{}^{i}R_{j_1j_2j_3}{}^{p}{}_{;j_4\cdots j_{k+1}}\\
&- \sum_{b=1}^{k+1} R_{j_{k+2}j_lj_b}{}^{a}R_{j_1j_2j_3}{}^{i}{}_{;j_4\cdots j_{b-1}aj_{b+1}\cdots j_{k+1}},
\end{aligned}$$

and the result now follows by induction. □

The significance of Proposition 2.1.1 lies in the fact that although geometric functions are defined in terms of the metric, they can be approximated purely in terms of curvature tensors, as we will see presently.

Finally, denote

$$\overline{\mathcal{R}} := \mathcal{R} \oplus \langle \delta^i_j \rangle, \tag{10}$$

where δ^i_j here denotes the Kronecker delta tensor. For all k, define $\overline{\mathcal{R}}^k$ as before. We also call elements of $\overline{\mathcal{R}}$ *curvature tensors*. Observe that $\overline{\mathcal{R}}$ is also closed under matrix multiplication. Furthermore, $\overline{\mathcal{R}}^0 = \langle \delta^i_j \rangle$ and, for all $k \neq 0$, $\overline{\mathcal{R}}^k = \mathcal{R}^k$.

2.2. ***Curvature Polynomials.*** Let $\underline{X} := (X_1, \ldots, X_n)$ be a vector of formal variables each taking values in $\mathbb{R}^{m+1}$. For $\rho \in \mathcal{R}^k$ and for $0 \leq r_1 + \cdots + r_n \leq k+1$, define the formal polynomial

$$\begin{aligned}(\rho_{r_1,\ldots,r_n})^i_{j_{(r_1+\cdots+r_n)+1}\cdots j_{k+1}} \\ := \rho^i_{j_1\ldots j_{k+1}} X_1^{j_1} \ldots X_1^{j_{r_1}} \ldots X_n^{j_{(r_1+\ldots+r_{n-1})+1}} \ldots X_n^{(r_1+\ldots+r_n)},\end{aligned} \tag{11}$$

where X^i_j denotes the i-th component of the vector X_j. Abusing notation, let $\mathcal{R}[\underline{X}]$ be the vector space with basis the set of all such formal polynomials. We call $\mathcal{R}[\underline{X}]$ the space of *curvature polynomials*. Observe that $\mathcal{R}[\underline{X}]$ is closed under matrix multiplication, although it is not always possible to multiply two given elements (indeed, two elements which are both covariant of order 1 and contravariant of order 0 cannot be multiplied). Furthermore, since $\mathcal{R}$ is self-adjoint with respect to δ, so too is $\mathcal{R}[\underline{X}]$ in the sense that if $P^i_{j_1\ldots j_k}$ is an element of $\mathcal{R}[\underline{X}]$, then $\delta^{ia}\delta_{j_l b} P^b_{j_1\ldots j_{l-1} a j_{l+1}\ldots j_k}$ identifies with a unique element of $\mathcal{R}[\underline{X}]$ for all $1 \leq l \leq k$.

For $k \in \mathbb{Z}$ and for $\underline{r} := (r_1, \ldots, r_n) \in \mathbb{N}^n$, let $\mathcal{R}^k_{\underline{r}}[\underline{X}]$ denote the subspace of $\mathcal{R}[\underline{X}]$ consisting of those elements which are contravariant of order $k+1$ and homogeneous of degree r_i in X_i for each i. Likewise, denote

$$\mathcal{R}^k[\underline{X}] := \bigoplus_{\underline{r}} \mathcal{R}^k_{\underline{r}}[\underline{X}]. \tag{12}$$

When $P \in \mathcal{R}^k_{\underline{r}}[\underline{X}]$, we say that it has *order-difference* k and *degree* $\underline{r}$. As before, permutation of indices preserves order-difference, and if P and P' have order-differences k and k' respectively then their matrix product has order-difference $k + k'$.

Throughout most of this section, we will only be concerned with the case where $n = 1$ and we denote $r := r_1$. Here we have

Proposition 2.2.1. *If $r > k$ and if $\rho \in \mathcal{R}^k$, then $\rho_r = 0$. In particular, $\mathcal{R}^k[X]$ is nontrivial only if $k \geq 0$.*

Proof. It suffices to prove the result when ρ is a generator of $\mathcal{R}$. However, for each k, by symmetry, $R_{j_1 j_2 j_3}{}^{i}{}_{;j_4\dots j_k} X^{j_1}\dots X^{j_k} = 0$, and the result follows. □

Proposition 2.2.1 implies that every element of $\mathcal{R}^0[X]$ is a finite sum of matrix products of those generators of $\mathcal{R}[X]$ which are of order-difference 0, that is, formal polynomials of the form $R_{p_1 i p_2}{}^{j}{}_{;p_3\dots p_{k+2}} X^{p_1}\dots X^{p_{k+2}}$, where k varies over all nonnegative integers. By considerations such as these, we obtain, for example,

$$\begin{aligned} &\mathcal{R}^0_0[X] = 0, && \mathcal{R}^0_1[X] = 0,\\ &\mathcal{R}^0_2[X] = \langle R_{piq}{}^{j} X^p X^q\rangle, && \mathcal{R}^0_3[X] = \langle R_{piq}{}^{j}{}_{;r} X^p X^q X^r\rangle, \end{aligned} \tag{13}$$

and so on. Likewise, every element of $\mathcal{R}^1[X]$ is a finite sum of matrix products of generators all but one of which are elements of $\mathcal{R}^0[X]$ and the remaining one of which is an element of $\mathcal{R}^1[X]$, and we obtain,

$$\begin{aligned} &\mathcal{R}^1_0[X] = 0,\\ &\mathcal{R}^1_1[X] = \langle (R_{p i_{\sigma(1)} i_{\sigma(2)}}{}^{j} X^p,\ R_{i_{\sigma(1)} i_{\sigma(2)} p}{}^{j} X^p)_{\sigma\in\Sigma_2}\rangle,\\ &\mathcal{R}^1_2[X] = \big\langle \big(R_{p i_{\sigma(1)} q}{}^{j}{}_{;i_{\sigma(2)}} X^p X^q,\ R_{p i_{\sigma(1)} i_{\sigma(2)}}{}^{j}{}_{;q} X^p X^q,\\ &\qquad\qquad\qquad\qquad R_{i_{\sigma(1)} i_{\sigma(2)} p}{}^{j}{}_{;q} X^p X^q\big)_{\sigma\in\Sigma_2}\big\rangle, \end{aligned} \tag{14}$$

and so on. In summary, it is relatively straightforward to determine $\mathcal{R}^k_r[X]$ for all k and for all r.

For general n, since $\mathcal{R}^k$ is trivial for $k < 2$, $\mathcal{R}^{-1}_{\underline{r}}[\underline{X}]$ is trivial for $r_1 + \cdots + r_n \leq 2$ and $\mathcal{R}^0_{\underline{r}}[\underline{X}]$ is trivial for $r_1 + \cdots + r_n \leq 1$. This observation will play an important role in the sequel.

Finally, as before, denote

$$\overline{\mathcal{R}}[\underline{X}] = \mathcal{R}[\underline{X}] \oplus \langle \delta^i_j\rangle \oplus \langle X^i_j\rangle, \tag{15}$$

where X^i_j denotes the i-th component of the vector X_j. For all k, and for all $\underline{r}$, define $\overline{\mathcal{R}}^k_{\underline{r}}[\underline{X}]$ as before. We also call elements of $\overline{\mathcal{R}}[\underline{X}]$ *curvature polynomials*. Observe that $\overline{\mathcal{R}}[\underline{X}]$ is also closed under matrix multiplication. Furthermore,

$$\overline{\mathcal{R}}^{-1}[\underline{X}] = \mathcal{R}^{-1}[\underline{X}] \oplus \langle X_1,\dots,X_n\rangle, \quad \overline{\mathcal{R}}^0[\underline{X}] = \mathcal{R}^0[\underline{X}] \oplus \langle \delta^i_j\rangle, \tag{16}$$

and $\overline{\mathcal{R}}^k[\underline{X}] = \mathcal{R}^k[\underline{X}]$ for all other values of k.

2.3. ***General properties of Taylor series.*** As before, let $\underline{X} := (X_1,\dots,X_n)$ be a vector of formal variables taking values in $\mathbb{R}^{m+1}$. Abusing notation, let $A[\underline{X}]$ be an algebra of formal polynomials in $\underline{X}$, and let $A[\![\underline{X}]\!]$ be the algebra of formal power series in $\underline{X}$ all of whose partial sums are elements of $A[\underline{X}]$. For such a formal power series F and for every nonnegative integer k denote by $[F]_k$ its partial sum of order k.

Recall that for all real α, the binomial theorem furnishes a sequence $(a_{k,\alpha})$ of real numbers such that for $x \in]-1,1[$,

$$(1+x)^{\alpha} = \sum_{k=0}^{\infty} a_{k,\alpha} x^{k}. \tag{17}$$

Consequently, if the algebra $A[\underline{X}]$ contains an identity, which we always denote by I, then for all formal power series F in $A[\![\underline{X}]\!]$ with $F(0) = I$, and for any real exponent α we define

$$F^{\alpha} := \sum_{k=0}^{\infty} a_{k,\alpha}(F-I)^{k}. \tag{18}$$

Proposition 2.3.1. *Let F be a formal power series in X. If F belongs to $A[\![\underline{X}]\!]$, and if $F(0) = I$, then F^{α} also belongs to $A[\![\underline{X}]\!]$ for all real α.*

Proof. Denote $G := F - I$. For all k, $G^{k} \in A[\![\underline{X}]\!]$ and since $G(0) = 0$, $[G^{k}]_{l} = 0$ for all $l < k$. Thus, for all α and for all l,

$$[F^{\alpha}]_{l} = \left[\sum_{k=0}^{\infty} a_{k,\alpha} G^{k}\right]_{l} = \sum_{k=0}^{l} a_{k,\alpha}[G^{k}]_{l} \in A[\underline{X}],$$

and so $F^{\alpha} \in A[[\underline{X}]]$, as desired. □

Now let $\underline{T} := (T_1, \ldots, T_n)$ be a vector of formal variables taking values in $\mathbb{R}$, and let $A[\underline{X}][\![\underline{T}]\!]$ denote the algebra of formal power series in $\underline{T}$ all of whose coefficients are elements of $A[\underline{X}]$.

Proposition 2.3.2. *Let F be a formal power series in $\underline{X}$, and define $G(\underline{X}, \underline{T}) := F(T_1X_1, \ldots, T_nX_n)$. If G belongs to $A[\underline{X}][\![\underline{T}]\!]$, then F belongs to $A[\![\underline{X}]\!]$.*

Proof. By hypothesis,

$$G = \sum_{\underline{k}} \frac{1}{k_1! \ldots k_n!} T_1^{k_1} \ldots T_n^{k_n} P_{\underline{k}}(\underline{X}),$$

where, for all $\underline{k}$, the formal polynomial $P_{\underline{k}}$ belongs to $A[\underline{X}]$. Now consider the formal derivatives of F and G with respect to $\underline{X}$ and $\underline{T}$ respectively. By the chain rule,

$$\frac{\partial^{k_1} \ldots \partial^{k_n} F}{\partial X_1^{k_1} \ldots \partial X_n^{k_n}}(0)(X_1^{\otimes k_1} \otimes \cdots \otimes X_m^{\otimes k_m}) = \frac{\partial^{k_1} \ldots \partial^{k_n} G}{\partial T_1^{k_1} \ldots \partial T_n^{k_n}}(0, \underline{X}) = P_{\underline{k}}(\underline{X}) \in A[\underline{X}].$$

It follows that every partial sum of F belongs to $A[\underline{X}]$, and so F belongs to $A[\![\underline{X}]\!]$, as desired. □

2.4. ***Tensor-valued geometric functions.*** For all $p, q \in \mathbb{N}$, let $T^{p,q} := T^{p,q}(\mathbb{R}^{m+1})$ be the space of tensors over $\mathbb{R}^{m+1}$ which are covariant of order p and contravariant of order q. Consider a function $f : \Omega \to T^{1,k+1}$, and denote by $[f]$ its Taylor series. In the present context, the statement that $[f]$ belongs to $R^k[\![X]\!]$ means that the Taylor series of $[f]$ about 0 is given by

$$f(x) \sim \sum_{r=0}^{\infty} R_r(x),$$

where, for all r, R_r is a curvature polynomial of order-difference k and degree r.

Now observe that $T^{1,1}$ naturally identifies with $\mathrm{End}(\mathbb{R}^{m+1})$. In particular, since matrix multiplication coincides with the usual notion of matrix multiplication in this case, the space $\overline{\mathcal{R}}^0[\underline{X}]$ is also closed with respect to this product, and therefore constitutes an algebra.

Let $M : \Omega \to \mathrm{End}(\mathbb{R}^{m+1})$ be such that for all $x \in \Omega$ and for every vector U, $M(x)U$ is the parallel transport of U along the radial line from x to 0. The first few terms of the Taylor series of M are readily determined. Indeed,

Proposition 2.4.1.

$$M^i_j(x) \sim \delta^i_j + \tfrac{1}{6} R_{pjq}{}^i x^p x^q + \tfrac{1}{12} R_{pjq}{}^i{}_{;r} x^p x^q x^r + O(x^4). \tag{19}$$

Remark. Equations (19), (21) and (25) are all proven via the same classical Jacobi field argument [Chavel 2006]. For the reader's convenience, we provide a proof of (19) in order to illustrate the technique.

Proof. Fix a point $x_0 \in S^m$ and a vector $U_0 \in \mathbb{R}^{m+1}$. Let $x(y) := tx_0$ and $U(t) := tU_0$ so that x is a geodesic and U is a Jacobi field over x. We use a dot to denote both differentiation and covariant differentiation in the radial direction. By definition, $U(0) = 0$ and, since (Ω, g) is an exponential chart, $\dot{U}(0) = U_0$. Furthermore, by the Jacobi field equation,

$$\ddot{U} = R(\dot{x}, U, \dot{x}) := R_{\dot{x}U}\dot{x}.$$

Differentiating this two more times yields

$$\begin{aligned}
\nabla^3_{\dot{x}} U &= (\nabla R)(\dot{x}, U, \dot{x}; \dot{x}) + R(\dot{x}, \dot{U}, \dot{x}),\\
\nabla^4_{\dot{x}} U &= (\nabla^2 R)(\dot{x}, U, \dot{x}; \dot{x}, \dot{x}) + 2(\nabla R)(\dot{x}, \dot{U}, \dot{x}; \dot{x}) + R(\dot{x}, R(\dot{x}, U, \dot{x}), \dot{x}).
\end{aligned}$$

Upon evaluating at zero and applying Taylor's theorem, we obtain

$$M^i_j(x)U(t)^j := t\big(U^i_0 + \tfrac{1}{6} R_{pjq}{}^i x^p x^q U^j_0 + \tfrac{1}{12} R_{pjq}{}^i{}_{;r} x^p x^q x^r U^j_0 + O(x^4)\big),$$

where $x := tx_0$. The result now follows upon dividing each side by t. □

More generally, we have

Proposition 2.4.2.

$$[M] \in \overline{\mathcal{R}}^0[\![X]\!]. \tag{20}$$

Proof. Fix a point $x_0 \in \Omega$ and a vector $U_0 \in \mathbb{R}^{m+1}$. Let $x(t) := tx_0$ and $U(t) := tU_0$. Using a dot to denote differentiation and covariant differentiation in the radial direction, we obtain, as before

$$U(0) = 0, \quad \dot{U}(0) = U_0 \quad \text{and} \quad \ddot{U} = R_{\dot{x}U}\dot{x}.$$

We now claim that there exist sequences (P_k) and (Q_k) of polynomials over $\mathrm{End}(\mathbb{R}^{n+1})$ such that, for all k,

$$\nabla_{\dot{x}}^{k+2}U = P_k(R(x)(\dot{x}), \ldots, \nabla^k R(x)(\dot{x}))U + Q_k(R(x)(\dot{x}), \ldots, \nabla^{k-1}R(x)(\dot{x}))\dot{U},$$

where, for all l,

$$\nabla^l R(x)(\dot{x}) := R(x)_{p_1 j p_2}{}^{i}{}_{;p_3 \ldots p_{l+2}}\dot{x}^{p_1} \ldots \dot{x}^{p_{l+2}}.$$

This holds for $k = 0$ by the Jacobi field equation. For $k \geq 0$, using the inductive hypothesis and the fact that $\nabla_{\dot{x}}\dot{x} = 0$, we obtain

$$\begin{aligned}
\nabla_{\dot{x}}^{k+3}U &= \nabla_{\dot{x}}\big(P_k(R(x)(\dot{x}), \ldots, \nabla^k R(x)(\dot{x}))U \\
&\qquad\qquad + Q_k(R(x)(\dot{x}), \ldots, \nabla^{k-1}R(x)(\dot{x}))\dot{U}\big) \\
&= P_k(R(x)(\dot{x}), \ldots, \nabla^k R(x)(\dot{x}))\dot{U} + Q_k(R(x)(\dot{x}), \ldots, \nabla^{k-1}R(x)(\dot{x}))\ddot{U} \\
&\qquad + \sum_{l=0}^{k} P_{k,l}(R(x)(\dot{x}), \ldots, \nabla^k R(x)(\dot{x}), \nabla^{l+1}R(x)(\dot{x}))U \\
&\qquad + \sum_{l=0}^{k-1} Q_{k,l}(R(x)(\dot{x}), \ldots, \nabla^{k-1}R(x)(\dot{x}), \nabla^{l+1}R(x)(\dot{x}))\dot{U},
\end{aligned}$$

for suitable sequences of polynomials $(P_{k,l})$ and $(Q_{k,l})$. However, by the Jacobi field equation again, $\ddot{U} = R_{\dot{x}U}\dot{x}$, and the assertion follows by induction. Observe, furthermore, that for all k, the zeroth order terms of P_k and Q_k both vanish. Substituting $t = 0$ now yields

$$(\nabla_{x_0}^{k+2}U)(0) = Q_k(R(0)(x_0), \ldots, \nabla^{k-1}R(0)(x_0))U_0.$$

However, for all k,

$$\partial_t^k tM(tx_0)U_0|_{t=0} = (\nabla_{x_0}^k U)(0),$$

so that, by Taylor's theorem,

$$[M](TX) = \mathrm{Id} + \sum_{k=2}^{\infty} \frac{T^k}{(k+1)!} Q_{k-1}\big(R(0)(X), \ldots, \nabla^{k-2}R(0)(X)\big) \in \overline{\mathcal{R}}[X][\![T]\!],$$

and the result now follows by Proposition 2.3.2. □

Let $A, B : \Omega \to \mathrm{End}(\mathbb{R}^{m+1})$ be such that, for all x,

$$g_{ij}(x) = A_i^p(x)\delta_{pj} = \delta_{ip}A_j^p(x),$$
$$g^{ij}(x) = B_p^i(x)\delta^{pj} = \delta^{ip}B_p^j(x), \tag{21}$$

where $g^{ij}(x)$ denotes the metric inverse of $g_{ij}(x)$. Using the same Jacobi field techniques as before, we obtain

$$A_j^i(x) \sim \delta_j^i + \tfrac{1}{3}R_{pjq}{}^i x^p x^q + \tfrac{1}{6}R_{pjq}{}^i{}_{;r}x^p x^q x^r + O(x^4),$$
$$B_j^i(x) \sim \delta_j^i - \tfrac{1}{3}R_{pjq}{}^i x^p x^q - \tfrac{1}{6}R_{pjq}{}^i{}_{;r}x^p x^q x^r + O(x^4). \tag{22}$$

More generally,

Proposition 2.4.3.

$$[A], [B] \in \overline{\mathcal{R}}^0[\![X]\!]. \tag{23}$$

Proof. For every point x in Ω and for all vectors U and V in $\mathbb{R}^{m+1}$,

$$g(x)(U, V) = g(0)(M(x)U, M(x)V) = \langle M(x)U, M(x)V\rangle = \langle M^*(x)M(x)U, V\rangle.$$

Since U and V are arbitrary, it follows that $A = M^*M$. However, since $\overline{\mathcal{R}}[X]$ is self-adjoint with respect to δ, $[M^*]$ belongs to $\overline{\mathcal{R}}[\![X]\!]$ and therefore so too does $[A] = [M^*][M]$. Finally, since $[A](0) = A(0) = I$, by Proposition 2.3.1, $[B] = [A^{-1}] = [A]^{-1}$ also belongs to $\overline{\mathcal{R}}[\![X]\!]$, and this completes the proof. □

Let $\Gamma : \Omega \to T^{1,2}$ be the Christoffel symbol[1] of the Levi-Civita covariant derivative of g. That is,

$$\Gamma_{ij}^k(x)\partial_k := \nabla_{\partial_i}\partial_j - D_{\partial_i}\partial_j, \tag{24}$$

where D denotes the canonical differentiation operator over $\mathbb{R}^{m+1}$. Recall that Γ is symmetric in i and j. Furthermore, using the same Jacobi field techniques once again, we obtain

$$\Gamma_{ii}^k(x) \sim \tfrac{2}{3}R_{pii}{}^k x^p + \tfrac{5}{12}R_{pii}{}^k{}_{;q}x^p x^q + \tfrac{1}{12}R_{piq}{}^k{}_{;i}x^p x^q + O(x^3). \tag{25}$$

More generally,

Proposition 2.4.4.

$$[\Gamma] \in \overline{\mathcal{R}}^1[\![X]\!]. \tag{26}$$

---

[1] Of course, technically speaking, the Christoffel symbol is not actually a tensor, although this does not affect the following discussion. The reader uncomfortable with this may choose to view the Christoffel symbol instead as the difference between the covariant derivatives ∇ and D, in which case it is correctly a tensor.

Proof. By the Koszul formula, for all vectors U, V and W in $\mathbb{R}^{m+1}$ and for every point x in Ω,

$$\begin{aligned}(27)\quad & 2\langle A(x)\Gamma(x)(U, V), W\rangle \\ &\qquad = \langle DA(x; U)V, W\rangle + \langle DA(x; V)U, W\rangle - \langle DA(x; W)U, V\rangle.\end{aligned}$$

Since $[A]$ belongs to $\overline{\mathcal{R}}[\![X]\!]$, its formal derivative, $D[A] = [DA]$ also belongs to $\overline{\mathcal{R}}[\![X]\!]$. Now let $\Phi : \Omega \to T^{1,2}$ be such that

$$\langle \Phi(x)(U, V), W\rangle = \langle DA(x; W)U, V\rangle.$$

Since $\overline{\mathcal{R}}[\![X]\!]$ is self-adjoint with respect to δ, $[\Phi]$ also belongs to $\overline{\mathcal{R}}[\![X]\!]$, and therefore, by linearity, so too does $[A\Gamma]$. It follows that $[\Gamma] = [B][A][\Gamma] = [B][A\Gamma]$ belongs to $\overline{\mathcal{R}}[\![X]\!]$, and this completes the proof. ☐

2.5. ***The exponential map and parallel transport.*** Define $\Omega_2 \subseteq \mathbb{R}^{m+1} \times \mathbb{R}^{m+1}$ by

$$\Omega_2 := \{(x, y) \mid \|x\| + \|y\| < 1\}. \tag{28}$$

Let $\mathrm{Exp} : \Omega_2 \to \Omega$ be the exponential map of g. That is, for all (x, y), the curve $t \mapsto \mathrm{Exp}(x, ty)$ is the unique geodesic in Ω leaving the point x in the direction of the vector y.

Proposition 2.5.1.

$$[\mathrm{Exp}] \in \overline{\mathcal{R}}^{-1}[\![X, Y]\!], \tag{29}$$

and

$$[\mathrm{Exp}] = X + Y + O(\|X, Y\|^3). \tag{30}$$

Proof. For any function ϕ of s and t, and for all k, let $[\phi]_{\infty,k}$ denote its Taylor series up to order k in t. Likewise, for any formal series Φ in S and T, let $[\Phi]_{\infty,k}$ denote its partial sum up to order k in T. Now define $E(x, y, s, t) := \mathrm{Exp}(sx, ty)$. By definition, for every point $(x, y) \in \Omega_2$ and for all s,

$$\begin{aligned}E(x, y, s, 0) &= \mathrm{Exp}(sx, 0) = sx, \\ \partial_t E(x, y, s, 0) &= \partial_t \mathrm{Exp}(sx, ty)|_{t=0} = y,\end{aligned}$$

so that $[E]_{\infty,1} = SX + TY$, which belongs to $\overline{\mathcal{R}}[X, Y][\![S, T]\!]$. We now claim that the partial sum $[E]_{\infty,k}$ belongs to $\overline{\mathcal{R}}[X, Y][\![S, T]\!]$ for all k. Indeed, suppose that this holds for k. Observe that

$$[\Gamma(E)(\partial_t E, \partial_t E)]_{\infty,k-1} = [\Gamma([E]_{\infty,k-1})(\partial_T[E]_{\infty,k}, \partial_T[E]_{\infty,k})]_{\infty,k-1},$$

where ∂_T denotes formal partial differentiation with respect to T. Since $[\Gamma]$ belongs to $\overline{\mathcal{R}}[X]$, it follows by the inductive hypothesis that $[\Gamma(E)(\partial_t E, \partial_t E)]_{\infty,k-1}$ belongs to $\overline{\mathcal{R}}[X, Y][\![S, T]\!]$. However, by the geodesic equation,

$$\partial_T^2[E]_{\infty,k+1} = [\partial_t^2 E]_{\infty,k-1} = -[\Gamma(E)(\partial_t E, \partial_t E)]_{\infty,k-1} \in \overline{\mathcal{R}}[X, Y][\![S, T]\!],$$

and the claim now follows by induction. In particular $[E]$ belongs to $\overline{\mathcal{R}}[X, Y][\![S, T]\!]$ and the first assertion follows by Proposition 2.3.2. Finally, since $[\mathrm{Exp}] - (X + Y) \in \mathcal{R}^{-1}[\![X, Y]\!]$, its lowest degree term has degree at least 3 in X and Y, thus proving the second assertion. This completes the proof. □

Let $\mathrm{Tr} : \Omega_2 \times \mathbb{R}^{n+1} \to \mathrm{End}(\mathbb{R}^{n+1})$ be such that for all (x, y) and for every vector U, $\mathrm{Tr}(x, y)U$ is the parallel transport of U from the point x to the point $\mathrm{Exp}(x, y)$ along the geodesic $t \mapsto \mathrm{Exp}(x, ty)$.

Proposition 2.5.2.

$$[\mathrm{Tr}] \in \overline{\mathcal{R}}^0[\![X, Y]\!], \tag{31}$$

and

$$[\mathrm{Tr}] = I + O(\|X, Y\|^2). \tag{32}$$

Proof. As before, for any function ϕ of s and t, and for all k, let $[\phi]_{\infty,k}$ denote its Taylor series up to order k in t. Likewise, for any formal series Φ in S and T, let $[\Phi]_{\infty,k}$ denote its partial sum up to order k in T. Define $E(x, y, s, t) := \mathrm{Exp}(sx, ty)$ and $F(x, y, s, t) := \mathrm{Tr}(sx, ty)$. By definition, for every point $(x, y) \in \Omega_2$ and for all s,

$$F(x, y, s, 0) = \mathrm{Tr}(sx, 0) = I,$$

so that $[F]_{\infty,0} = I$, which belongs to $\overline{\mathcal{R}}[X, Y][\![S, T]\!]$. We now claim that the partial sum $[F]_{\infty,k}$ belongs to $\overline{\mathcal{R}}[X, Y][\![S, T]\!]$ for all k. Indeed, suppose that this holds for k. Observe that

$$[\Gamma(E)(\partial_t E, F)]_{\infty,k} = [\Gamma([E]_{\infty,k})(\partial_T[E]_{\infty,k+1}, [F]_{\infty,k})]_{\infty,k},$$

where ∂_T denotes formal partial differentiation with respect to T. Since $[\Gamma]$ belongs to $\overline{\mathcal{R}}[X]$ and since $[\mathrm{Exp}]$ belongs to $\overline{\mathcal{R}}[X, Y]$, it follows by the inductive hypothesis that $[\Gamma(E)(\partial_t E, F)]_{\infty,k}$ also belongs to $\overline{\mathcal{R}}[X, Y][\![S, T]\!]$. However, by the parallel transport equation

$$\partial_T[F]_{\infty,k+1} = [\partial_t F]_{\infty,k} = -[\Gamma(E)(\partial_t E, F)]_{\infty,k} \in \overline{\mathcal{R}}[X, Y][\![S, T]\!],$$

and the claim now follows by induction. In particular, F belongs to $\overline{\mathcal{R}}[X, Y][\![S, T]\!]$ and the first assertion follows by Proposition 2.3.2. Finally, since $[F] - I \in \mathcal{R}^0[\![X, Y]\!]$, its lowest degree term has degree at least 2 in X and Y, thus proving the second assertion. This completes the proof. □

Finally, we consider higher order iterates of the exponential map and the parallel transport. Thus, for all n, define $\Omega_{n+1} \subseteq (\mathbb{R}^{m+1})^{n+1}$ by

$$\Omega_{n+1} = \{(x_1, \ldots, x_{n+1}) \mid \|x_1\| + \cdots + \|x_{n+1}\| < 1\}, \tag{33}$$

and define the sequences of functions (Exp_n) and (Tr_n) such that

$$\mathrm{Exp}_1(x_1, x_2) := \mathrm{Exp}(x_1, x_2), \quad \mathrm{Tr}_0(x_1) := \mathrm{Id}, \tag{34}$$

and, for all n,

$$\begin{aligned} \mathrm{Exp}_n(x_1, \ldots, x_{n+1}) &:= \mathrm{Exp}(\mathrm{Exp}_{n-1}(x_1, \ldots, x_n), \mathrm{Tr}_{n-1}(x_1, \ldots, x_n)x_{n+1}), \\ \mathrm{Tr}_n(x_1, \ldots, x_{n+1})U &:= \mathrm{Tr}(\mathrm{Exp}_{n-1}(x_1, \ldots, x_n), \mathrm{Tr}_{n-1}(x_1, \ldots, x_{n-1})U). \end{aligned} \tag{35}$$

Proposition 2.5.3. *For all n,*

$$\begin{aligned} [\mathrm{Exp}_n] &\in \overline{\mathcal{R}}^{-1}[[X_1, \ldots, X_{n+1}]], \\ [\mathrm{Tr}_n] &\in \overline{\mathcal{R}}^{0}[[X_1, \ldots, X_{n+1}]], \end{aligned} \tag{36}$$

and

$$\begin{aligned} [\mathrm{Exp}_n] &= X_1 + \cdots + X_{n+1} + O(\|X_1, \ldots, X_{n+1}\|^3), \\ [\mathrm{Tr}_n] &= I + O(\|X_1, \ldots, X_{n+1}\|^2). \end{aligned} \tag{37}$$

Proof. This follows by induction using Propositions 2.5.1 and 2.5.2 and the recursive definitions of (Exp_n) and (Tr_n). □

3. Taylor series of functions derived from immersions

3.1. ***Graphs over spheres.*** Let S^m be the unit sphere in $\mathbb{R}^{m+1}$ and let $\overline{\nabla}$, $\overline{\mathrm{Hess}}$ and $\overline{\Delta}$ be respectively its gradient, Hessian and Laplace operators with respect to the standard euclidean metric. For $t \in]0, \infty[$, which we think of as a scale parameter, and for $f \in C^0(S^m)$, consider the function $e(t, f) : S^m \to \mathbb{R}^{m+1}$ given by

$$e(t, f)(x) := t(1 + t^2 f(x))x. \tag{38}$$

Heuristically, $e(t, f)$ is an immersed sphere of radius approximately t centred on the origin. For all k, let $J := J^k S^m$ denote the bundle of k-jets over S^m, and for a function $f \in C^k(S^m)$ and a point $x \in S^m$, denote by f_x its k-jet at x, where the order k of the jet should hopefully be clear from the context. Define the functions $N :]0, \infty[\times J^1 S^m \to S^m$ and $H :]0, \infty[\times J^2 S^m \to \mathbb{R}$ such that for all $t \in]0, \infty[$ and for all $f_x \in JS^m$, $N(t, f_x)$ and $H(t, f_x)$ are respectively the outward-pointing unit normal of the immersion $e(t, f)$ at the point $e(t, f)(x)$ and its mean curvature at that point, both with respect to the metric g. It is worth noting that both N and H are actually smooth functions defined over finite-dimensional domains and may both be expressed explicitly in terms of (rather complicated) formulae involving g.

We prefer to define these functions in the above manner in order to emphasise their clear geometric meanings.

We are interested in the Taylor series of $N(t, f_x)$ and $H(t, f_x)$ in t about 0. To this end, we first introduce the following auxiliary functions. Define $r : \mathbb{R}^{m+1} \to [0, \infty[$ and $x : \mathbb{R}^{m+1} \setminus \{0\} \to S^m$ by

$$r(y) := \|y\|, \quad x(y) := y/r. \tag{39}$$

Given $f \in C^1(S^m)$, define $\hat{f} :]0, \infty[\times (\mathbb{R}^{m+1} \setminus \{0\}) \to \mathbb{R}$ by

$$\hat{f}(t, y) := r - t(1 + t^2 f(x)). \tag{40}$$

Observe that the image of $e(t, f)$ coincides with the level set of $\hat{f}$ at height 0. Furthermore, for every point y in this level set, $\nabla \hat{f}(y)$ is orthogonal to this level set with respect to the metric g.

Proposition 3.1.1.

$$\nabla \hat{f}(t, y) = \frac{y}{r} - \frac{t}{r} B(y) t^2 \overline{\nabla} f(x). \tag{41}$$

Proof. The gradient of $\hat{f}$ with respect to the euclidean metric is

$$D\hat{f}(t, y) = \frac{y}{r} - \frac{t^3}{r} \overline{\nabla} f(x).$$

However, for all vectors U in $\mathbb{R}^{m+1}$,

$$d\hat{f}(t, y)(U) = \langle D\hat{f}(t, y), U\rangle = \langle A(y)B(y)D\hat{f}(t, y), U\rangle = g(B(y)D\hat{f}(t, y), U),$$

so that the gradient of $\hat{f}$ with respect to g is $\nabla \hat{f}(t, y) = B(y)D\hat{f}(t, y)$, and since $B(y)y = y$ for all y, the result follows. □

We now invert the situation and consider both r and y as functions of t and x, so that

$$r(t, x) := t(1 + t^2 f(x)), \quad y(t, x) := t(1 + t^2 f(x))x. \tag{42}$$

We define

$$\hat{N}(t, f_x) := \nabla \hat{f}(t, y) = \frac{y}{r} - \frac{t}{r} B(y) t^2 \overline{\nabla} f(x), \tag{43}$$

so that we obtain the following formula for N:

$$N(t, f_x) := \frac{1}{\|\hat{N}(t, x)\|_g} \hat{N}(t, f_x), \tag{44}$$

where $|\cdot|_g$ here denotes the norm with respect to the metric g.

It will also be necessary to extend e, N and H to allow for variations of the centre of the immersed sphere. Thus, for $t \in]0, \infty[$, for $y \in \mathbb{R}^{m+1}$, and for $f \in C^0(S^m)$, define $e(t, y, f) : S^m \to \mathbb{R}^{m+1}$ by

$$e(t, y, f)(x) := \mathrm{Exp}(ty, t(1 + t^2 f(x))x), \tag{45}$$

so that, heuristically, $e(t, y, f)$ is an immersed sphere of radius approximately t with centre displaced to the point y. Define $N :]0, \infty[\times \mathbb{R}^{m+1} \times J^1 S^m \to S^m$ and $H :]0, \infty[\times \mathbb{R}^{m+1} \times J^2 S^m \to \mathbb{R}$ as before. Observe, in particular, that $e(t, 0, f) = e(t, f)$, $N(t, 0, f_x) = N(t, f_x)$ and $H(t, 0, f_x) = H(t, f_x)$.

3.2. ***The Taylor series of the unit normal vector.*** We now study the Taylor series of the scale-dependent functions introduced in Section 3.1. In particular, we are interested in how the different terms in these series contribute to the exponent of t. To this end, we extend the formalism developed in Sections 2.1 and 2.2 as follows. For a vector $\underline{X} := (X_1, \dots, X_n)$ of formal variables taking values in $\mathbb{R}^{m+1}$, consider the set of formal polynomials

$$\{X^a_i \delta_{ab} P^b_{j_1 \dots j_k}(\underline{X}) \mid P \in \overline{\mathcal{R}}[\underline{X}]\}, \tag{46}$$

where X^i_j denotes the i-th component of the vector X_j. Let $\mathcal{Q}[\underline{X}]$ be the vector space with basis the set of all tensor products of elements of this set. We call $\mathcal{Q}[\underline{X}]$ the space of *curvature polynomials* of the second kind. For all $k \in \mathbb{N}$ and for all $\underline{r} := (r_1, \dots, r_n) \in \mathbb{N}^n$, denote by $\mathcal{Q}^k_{\underline{r}}[\underline{X}]$ the subspace consisting of those elements which are contravariant of order k and which are homogeneous of degree r_i in the variable X_i for all i. When $Q \in \mathcal{R}^k_{\underline{r}}[\underline{X}]$, we say that it has *order* k and *degree* $\underline{r}$. Finally, denote

$$\overline{\mathcal{Q}}[\underline{X}] := \mathcal{Q}[\underline{X}] \oplus \langle 1 \rangle. \tag{47}$$

We also call elements of $\overline{\mathcal{Q}}[\underline{X}]$ *curvature polynomials* of the second kind.

Now let A be an algebra graded by $\mathbb{N}^k$ for some k. Let $A[T]$ be the algebra of polynomials over $\mathbb{R}$ with coefficients in A. For a given weight $\underline{w} := (w_1, \dots, w_k) \in \mathbb{N}^k$, let $A[T]_{\underline{w}}$ be the subalgebra of $A[T]$ consisting of those polynomials whose coefficients of degree m are elements of $\bigoplus_{\langle \underline{w}, \underline{i} \rangle = m} A_{\underline{i}}$ for all m. Likewise, let $A\llbracket T \rrbracket$ be the algebra of formal power series over $\mathbb{R}$ with coefficients in A, and for $\underline{w} \in \mathbb{N}^k$, let $A\llbracket T \rrbracket_{\underline{w}}$ be the subalgebra of $A\llbracket T \rrbracket$ consisting of those formal power series all of whose partial sums are elements of $A[T]_{\underline{w}}$.

Now let $\mathbb{R}[F]$ be the algebra of formal polynomials in the variable F. Consider a smooth function $\phi : [0, \infty[\times J^k S^m \to \mathbb{R}$ which only depends on the metric g and the jet f_x. For such a function, the statement that $[\phi]$ belongs to $\mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}_{*,*}[X, \overline{\nabla} F]\llbracket T \rrbracket_{(2,1,2)}$, for example, means that its Taylor series in t about 0 takes the form

$$\phi(t,f_x)\sim\sum_{m=0}^{\infty}t^m\sum_{\langle\underline{i},(2,1,2)\rangle=m}\sum_{\alpha}P_{\underline{i},\alpha}(f(x))Q_{\underline{i},\alpha}(x,\overline{\nabla}f(x)),\tag{48}$$

where, for all $\underline{i}:=(i_1,i_2,i_3)$ and for all α, $P_{\underline{i},\alpha}$ is a polynomial of degree i_1 and $Q_{\underline{i},\alpha}$ is a curvature polynomial of the second kind of order 0 and degree (i_2,i_3). We leave the reader to interpret the meanings of other tensor products of spaces of formal polynomials. Importantly, this notation emphasises that all terms in X carry weight 1 in T whilst all terms in F and $\overline{\nabla}F$ carry weight 2. This behaviour will be common to all series studied in the sequel.

Proposition 3.2.1. *For all real* α,

$$[(r/t)^\alpha]\in\mathbb{R}_*[F][\![T]\!]_2.\tag{49}$$

Proof. By definition, $[(r/t)]=[1+t^2f]$ belongs to $\mathbb{R}_*[F][\![T]\!]_2$. Since $[(r/t)](0)=(r/t)(0)=1$, the result follows by Proposition 2.3.1. □

Proposition 3.2.2. *For all real* α,

$$[\|\hat{N}(t,f_x)\|^\alpha]\in\mathbb{R}_*[F]\otimes\overline{\mathcal{Q}}^0_{*,*}[X,\overline{\nabla}F][\![T]\!]_{(2,1,2)}\tag{50}$$

and

$$[\|\hat{N}(t,f_x)\|^\alpha]=1+O(T^4).\tag{51}$$

Proof. Using (7), (21) and (43), we obtain, for all t and for all x,

$$\|\hat{N}(t,f_x)\|_g^2=1+\left(\frac{t}{r}\right)^2\langle B(y)t^2\overline{\nabla}f,t^2\overline{\nabla}f\rangle.$$

However, by Propositions 2.4.3 and 3.2.1,

$$[B(y)]=\left[B\left(\frac{r}{t}(tx)\right)\right]$$

belongs to $\mathbb{R}_*[F]\otimes\overline{\mathcal{R}}_*[X][\![T]\!]_{(2,1)}$, so that $[\langle B(y)t^2\overline{\nabla}f,t^2\overline{\nabla}f\rangle]$ belongs to $\mathbb{R}_*[F]\otimes\overline{\mathcal{Q}}_{*,*}[X,\overline{\nabla}F][\![T]\!]_{(2,1,2)}$. It follows by Proposition 3.2.1 again that $\|\hat{N}(t,f_x)\|_g^2$ belongs to $\mathbb{R}_*[F]\otimes\overline{\mathcal{Q}}_{*,*}[X,\overline{\nabla}F][\![T]\!]_{(2,1,2)}$ and, since the first term in this series equals 1, the first assertion follows by Proposition 2.3.1. Finally, since $[(t/r)^2]$ has order 0 in T and since $[\langle B(y)t^2\overline{\nabla}f,t^2\overline{\nabla}f\rangle]$ has order 4 in T, we see that $\|\hat{N}(t,f_x)\|_g^2=1+O(T^4)$, and the second assertion follows by Proposition 2.3.1 again. This completes the proof. □

Proposition 3.2.3.

$$N(t,f_x)=\Phi_1(t,f_x)x+\Phi_2(t,f_x),\tag{52}$$

where

$$\begin{aligned}&[\Phi_1] \in \mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}^0_{*,*}[X, \overline{\nabla} F][\![T]\!]_{(2,1,2)},\\ &[\Phi_2] \in \mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}^0_{*,*}[X, \overline{\nabla} F] \otimes \overline{\mathcal{R}}^{-1}_{*,*}[X, \overline{\nabla} F][\![T]\!]_{(2,1,2,1,2)}.\end{aligned} \tag{53}$$

Furthermore

$$[N(t, y, f_x)] = X - T^2\overline{\nabla} F + O(T^4). \tag{54}$$

Proof. As in the above proof, $[B(y)]$ belongs to $\mathbb{R}_*[F] \otimes \overline{\mathcal{R}}_*[X][\![T]\!]_{(2,1)}$ and so $[B(y)t^2\overline{\nabla} f]$ belongs to $\mathbb{R}_*[F] \otimes \overline{\mathcal{R}}_{*,*}[X, \overline{\nabla} F][\![T]\!]_{(2,1,2)}$. By Proposition 3.2.1, the series $[(t/r)B(y)t^2\overline{\nabla} f]$ also belongs to $\mathbb{R}_*[F] \otimes \overline{\mathcal{R}}_{*,*}[X, \overline{\nabla} F][\![T]\!]_{(2,1,2)}$, and the result now follows by (43), (44) and Proposition 3.2.2. ☐

Proposition 3.2.4.

$$N(t, y, f_x) = \Phi_1(t, y, f_x)x + \Phi_2(t, y, f_x), \tag{55}$$

where

$$\begin{aligned}&[\Phi_1] \in \mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}^0_{*,*,*}[X, Y, \overline{\nabla} F][\![T]\!]_{(2,1,1,2)},\\ &[\Phi_2] \in \mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}^0_{*,*,*}[X, Y, \overline{\nabla} F] \otimes \overline{\mathcal{R}}^{-1}_{*,*,*}[X, Y, \overline{\nabla} F][\![T]\!]_{(2,1,1,2,1,1,2)}.\end{aligned} \tag{56}$$

Furthermore,

$$[N(t, f_x)] = X + O(T^2). \tag{57}$$

Proof. This Taylor series is obtained from Proposition 3.2.3 by substituting for every generator $R_{i_1i_2i_3}{}^j{}_{;i_4\dots i_{k+3}}$ of $\mathcal{R}$ its own Taylor series in t about 0:

$$[R_{i_1i_2i_3}{}^j{}_{;i_4\dots i_{k+3}}] = \sum_{m=0}^{\infty} \frac{1}{m!} T^m R_{i_1i_2i_3}{}^j{}_{;i_4\dots i_{k+3+m}} Y^{i_{k+3+1}} \dots Y^{i_{k+3+m}}.$$

The result follows. ☐

3.3. ***Normal variation of spheres.*** We extend e further in order to study variations of the base point, of the displacement of the centre, and of the immersion itself. Thus, for $t \in]0, \infty[$, for $y, z, w \in \mathbb{R}^{m+1}$ and for $f, g \in C^0(S^m)$, consider the function $e(t, y, z, w, f, g) : S^m \to \mathbb{R}^{m+1}$ given by

$$e(t, y, z, w, f, g)(x) := \mathrm{Exp}_2(z, t(y + w), t(1 + t^2(f(x) + g(x)))x), \tag{58}$$

and define $P, Q :]0, \infty[\times \mathbb{R}^{m+1} \times J^0S^m \to \mathrm{End}(\mathbb{R}^{m+1})$ and $R :]0, \infty[\times \mathbb{R}^{m+1} \times J^0S^m \to \mathbb{R}^{m+1}$ by

$$\begin{aligned}P(t, y, f_x) &:= \partial_z e(t, y, 0, 0, f, 0)(x),\\ Q(t, y, f_x) &:= \partial_w e(t, y, 0, 0, f, 0)(x),\\ R(t, y, f_x) &:= \partial_g e(t, y, 0, 0, f, 0)(x).\end{aligned} \tag{59}$$

Heuristically, for any given vectors U and V and for any given function g the vectors $P(t, y, f_x)U$, $Q(t, y, f_x)V$ and $R(t, y, f_x)g_x$ measure the respective infinitesimal variations of the immersion $e(t, y, f)$ at the point $e(t, y, f)(x)$ arising from infinitesimal perturbations of the base point of the displacement of the centre, and of the immersion itself in the directions of U, tV and t^3g respectively.

Now define $p, q :]0, \infty[\times\mathbb{R}^{m+1}\times J^0S^m \to \mathbb{R}^{m+1}$ and $r :]0, \infty[\times\mathbb{R}^{m+1}\times J^0S^m \to \mathbb{R}$ by

$$\begin{aligned}\langle p(t, y, f_x), U\rangle &:= \langle A(e(t, y, f_x))P(t, y, f_x)U, N(t, y, f_x)\rangle,\\ \langle q(t, y, f_x), V\rangle &:= \langle A(e(t, y, f_x))Q(t, y, f_x)V, N(t, y, f_x)\rangle,\\ r(t, y, f_x)g &:= \langle A(e(t, y, f_x))R(t, y, f_x)g, N(t, y, f_x)\rangle.\end{aligned}\tag{60}$$

Heuristically, p, q and r measure the normal components of the above infinitesimal variations.

Proposition 3.3.1.

$$p(t, y, f_x) = \Phi_1(t, y, f_x)x + \Phi_2(t, y, f_x),\tag{61}$$

where

$$\begin{aligned}&[\Phi_1] \in \mathbb{R}_*[F]\otimes\overline{\mathcal{Q}}^0_{*,*,*}[X, Y, \overline{\overline{\nabla}}F]\otimes\overline{\mathcal{R}}^0_{*,*,*}[X, Y, \overline{\overline{\nabla}}F][\![T]\!]_{(2,1,1,2,1,1,2)},\\ &[\Phi_2] \in \mathbb{R}_*[F]\otimes\overline{\mathcal{Q}}^0_{*,*,*}[X, Y, \overline{\overline{\nabla}}F]\otimes\overline{\mathcal{R}}^{-1}_{*,*,*}[X, Y, \overline{\overline{\nabla}}F][\![T]\!]_{(2,1,1,2,1,1,2)}.\end{aligned}\tag{62}$$

Furthermore,

$$[p] = \langle X, \cdot\rangle + O(T^2).\tag{63}$$

Proof. Let ∂_Z denote the formal partial derivative with respect to the variable Z. In particular, $[\partial_z \operatorname{Exp}_2(z, y, x)|_{z=0}] = \partial_Z[\operatorname{Exp}_2(z, y, x)]|_{Z=0}$. However, by Proposition 2.5.3,

$$\begin{aligned}&\partial_Z[\operatorname{Exp}_2(z, y, x)]|_{Z=0} \in \overline{\mathcal{R}}[X, Y],\\ &\partial_Z[\operatorname{Exp}_2(z, y, x)]|_{Z=0} = I + O(\|X, Y\|^2).\end{aligned}$$

Substituting ty and $t(1+t^2f(x))x$ for y and x respectively therefore yields

$$\begin{aligned}&[\partial_z e(z, ty, t(1+t^2f(x))x)|_{z=0}] \in \mathbb{R}_*[F]\otimes\overline{\mathcal{R}}_{*,*}[X, Y][\![T]\!]_{(2,1,1)},\\ &[\partial_z e(z, ty, t(1+t^2f(x))x)|_{z=0}] = I + O(T^2),\end{aligned}$$

so that

$$\begin{aligned}&[P] \in \mathbb{R}_*[F]\otimes\overline{\mathcal{R}}^0_{*,*}[X, Y][\![T]\!]_{(2,1,1)},\\ &[P] = I + O(T^2).\end{aligned}$$

The result now follows from the self-adjointness of $\mathcal{R}$ (Proposition 2.1.1) and Propositions 2.4.3 and 3.2.4. □

Proposition 3.3.2.

$$q(t, y, f_x) = t\Phi_1(t, y, f_x)x + t\Phi_2(t, y, f_x), \tag{64}$$

where

$$\begin{aligned}[\Phi_1] &\in \mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}^0_{*,*,*}[X, Y, \overline{\nabla} F] \otimes \overline{\mathcal{R}}^0_{*,*,*}[X, Y, \overline{\nabla} F][\![T]\!]_{(2,1,1,2,1,1,2)},\\ [\Phi_2] &\in \mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}^0_{*,*,*}[X, Y, \overline{\nabla} F] \otimes \overline{\mathcal{R}}^{-1}_{*,*,*}[X, Y, \overline{\nabla} F][\![T]\!]_{(2,1,1,2,1,1,2)}.\end{aligned} \tag{65}$$

Furthermore,

$$[q] = T\langle X, \cdot\rangle + O(T^3). \tag{66}$$

Proof. Let ∂_W denote the formal partial derivative with respect to the variable W. In particular, $[\partial_w \operatorname{Exp}(y + tw, x)|_{w=0}] = \partial_W[\operatorname{Exp}(y + tw, x)]|_{W=0}$. However, by Proposition 2.5.3,

$$\begin{aligned}\partial_W[\operatorname{Exp}(y + tw, x)]|_{W=0} &\in T\overline{\mathcal{R}}[X, Y],\\ \partial_W[\operatorname{Exp}(y + tw, x)]|_{W=0} &= TI + TO(\|X, Y\|^2).\end{aligned}$$

Substituting ty and $t(1 + t^2 f(x))x$ for y and x respectively therefore yields

$$\begin{aligned}[\partial_w e(z, ty, t(1 + t^2 f(x))x)|_{w=0}] &\in T\mathbb{R}_*[F] \otimes \overline{\mathcal{R}}_{*,*}[X, Y][\![T]\!]_{(2,1,1)},\\ [\partial_w e(z, ty, t(1 + t^2 f(x))x)|_{w=0}] &= TI + O(T^3),\end{aligned}$$

so that

$$\begin{aligned}[Q] &\in T\mathbb{R}_*[F] \otimes \overline{\mathcal{R}}^0_{*,*}[X, Y][\![T]\!]_{(2,1,1)},\\ [Q] &= TI + O(T^3).\end{aligned}$$

The result now follows from the self-adjointness of $\mathcal{R}$ (Proposition 2.1.1) and Propositions 2.4.3 and 3.2.4. □

Proposition 3.3.3.

$$[r] \in T^3\mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}^0_{*,*,*}[X, Y, \overline{\nabla} F][\![T]\!]_{(2,1,1,2)}, \tag{67}$$

and

$$[r] = T^3 + O(T^7). \tag{68}$$

Proof. Consider first the case where $Y = 0$ and observe that

$$e(t, 0, 0, 0, f, g) = \operatorname{Exp}\big(t(1 + t^2(f(x) + g(x)))x\big).$$

In particular, since Ω is an exponential chart,

$$\partial_g \operatorname{Exp}\big(t(1 + t^2(f(x) + g(x)))x\big)|_{g=0} = t^3 x,$$

so that $R(t, 0, f_x) = t^3 x$. Thus, by (43) and (44), since $A(y)x = x$ and since $\langle x, \overline{\overline{\nabla}} f(x)\rangle = 0$,

$$r(t, 0, f_x) = t^3 \langle x, N(t, 0, f_x)\rangle = t^3 \|\hat{N}(t, x)\|_g^{-1}.$$

The result for $Y = 0$ now follows by Proposition 3.2.2. The result for general Y follows by substituting for every generator $R_{i_1 i_2 i_3}{}^{j}{}_{;i_4 \dots i_{k+3}}$ of $\mathcal{R}$ its own Taylor series in Y about 0, as in the proof of Proposition 3.2.4. □

3.4. ***The Taylor series of the mean curvature.*** We end this section by determining the Taylor series of the mean curvature function. First recall (from [Smith 2011], for example) that

$$\begin{aligned}(69)\quad H(t, Y, f_x) \sim \frac{1}{t}\bigg(1 - \frac{t^2}{3}\operatorname{Ric}_{pq} x^p x^q - \frac{t^2}{n}(n + \overline{\Delta}) f \\ - \frac{t^3}{4}\operatorname{Ric}_{pq;r} x^p x^q x^r - \frac{t^3}{3}\operatorname{Ric}_{pq;r} x^p x^p Y^r \\ - \frac{t^4}{4}\operatorname{Ric}_{pq;rs} x^p x^q x^r Y^s + t^4 F(f_x) + O(t^5)\bigg),\end{aligned}$$

where F is a curvature polynomial. More generally,

Proposition 3.4.1.

$$\begin{aligned}(70)\quad & H(t, y, f_x) \\ &= \frac{1}{t}\operatorname{Tr}(\Phi_1) + \operatorname{Tr}(\Phi_2) + \frac{1}{t}\operatorname{Tr}(\Phi_3 t^2 \overline{\operatorname{Hess}}(f) \circ \pi) + \frac{1}{t}\langle \Phi_5, t^2 \overline{\operatorname{Hess}}(f) \circ \pi\rangle,\end{aligned}$$

where

$$(71)\quad \begin{aligned}\Phi_1, \Phi_2, \Phi_3 &\in \mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}^0_{*,*}[X, Y, \overline{\overline{\nabla}} F] \otimes \overline{\mathcal{R}}^0_{*,*}[X, Y, \overline{\overline{\nabla}} F][\![T]\!]_{(2,1,1,2,1,1,2)}, \\ \Phi_4 &\in \mathbb{R}_*[F] \otimes \overline{\mathcal{Q}}^2_{*,*}[X, Y, \overline{\overline{\nabla}} F][\![T]\!]_{(2,1,1,2)}.\end{aligned}$$

Proof. We first consider the case where $Y = 0$. Recall that

$$(72)\qquad H = \frac{1}{\|\nabla \hat{f}\|_g}\Delta \hat{f} - \frac{1}{\|\nabla \hat{f}\|_g^3} g(\nabla_{\nabla \hat{f}} \nabla \hat{f}, \nabla \hat{f}),$$

where Δ denotes the Laplace operator of the metric g. Furthermore, by (41),

$$(73)\qquad \nabla \hat{f} = \frac{1}{r}(y - tB(y)t^2 \overline{\overline{\nabla}} f(x)).$$

Now observe that, for all vectors U,

$$D_U \overline{\overline{\nabla}} f(x) = \frac{1}{r}\overline{\operatorname{Hess}}(f) \circ \pi(U) + \frac{1}{r}\langle U, \overline{\overline{\nabla}} f(x)\rangle \frac{y}{r},$$

where π is the orthogonal projection along x. Differentiating (73) therefore yields, for all U,

$$\begin{aligned}\nabla_U\nabla\hat{f} = \frac{1}{t}\left(\frac{t}{r}\right)U - \frac{1}{t}\left(\frac{t}{r}\right)\left\langle U, \frac{y}{r}\right\rangle\frac{y}{r}\\ - \left(\frac{t}{r}\right)DB(y;U)t^2\overline{\nabla}f - \frac{1}{t}\left(\frac{t}{r}\right)^2 B(y)t^2(\overline{\mathrm{Hess}}(f)\circ\pi)(U)\\ + \frac{1}{t}\left(\frac{t}{r}\right)^2\left\langle U, \frac{y}{r}\right\rangle B(y)t^2\overline{\nabla}f - \frac{1}{t}\left(\frac{t}{r}\right)^2\langle U, t^2\overline{\nabla}f\rangle B(y)\frac{y}{r} + \Gamma(U, \nabla\hat{f}),\end{aligned}$$

and, bearing in mind that $B(y)y = y$ and $\langle y, \overline{\nabla}f\rangle = 0$, we obtain

$$\begin{aligned}\Delta\hat{f} &= \frac{m}{t}\left(\frac{t}{r}\right) - \left(\frac{t}{r}\right)\mathrm{Tr}(DB(y;\cdot)t^2\overline{\nabla}f) - \frac{1}{t}\left(\frac{t}{r}\right)^2\mathrm{Tr}(B(y)t^2(\overline{\mathrm{Hess}}(f)\circ\pi))\\ &\qquad + \frac{1}{t}\mathrm{Tr}(\Gamma(\cdot, tx)) - \left(\frac{t}{r}\right)\mathrm{Tr}(\Gamma(\cdot, B(y)t^2\overline{\nabla}f))\\ &= \frac{1}{t}\mathrm{Tr}(\Phi_1) + \mathrm{Tr}(\Phi_2) + \frac{1}{t}\mathrm{Tr}(\Phi_3 t^2\overline{\mathrm{Hess}}(f)\circ\pi),\end{aligned}$$

where

$$\Phi_1, \Phi_2, \Phi_3 \in \mathbb{R}_*[F]\otimes\overline{\mathcal{Q}}^0_{*,*}[X, \overline{\nabla}F]\otimes\overline{\mathcal{R}}^0_{*,*}[X, \overline{\nabla}F][\![T]\!]_{(2,1,2,1,2)}.$$

Likewise,

$$\begin{aligned}g(\nabla_{\nabla\hat{f}}\nabla\hat{f}, \nabla\hat{f}) &= \frac{1}{t}\left(\frac{t}{r}\right)^3\langle B(y)t^2\overline{\nabla}f, t^2\overline{\nabla}f\rangle + \langle A(y)\Gamma(\nabla\hat{f}, \nabla\hat{f}), \nabla\hat{f}\rangle\\ &\quad - \left(\frac{t}{r}\right)\langle A(y)DB(y;\nabla\hat{f})t^2\overline{\nabla}f\nabla\hat{f}\rangle\\ &\quad - \frac{1}{t}\left(\frac{t}{r}\right)^4\langle t^2(\overline{\mathrm{Hess}}\circ\pi)B(y)t^2\overline{\nabla}f, B(y)t^2\overline{\nabla}f\rangle.\end{aligned}\tag{74}$$

However, for any symmetric bilinear form M_{ij} and for any vector V^i,

$$\begin{aligned}\delta_{ij}\delta^{ip}M_{pq}B^q_rV^rB^j_sV^s &= (\delta^{ip}\delta^{jq})M_{pq}(\delta_{jb}B^b_cV^c)(\delta_{ir}B^r_sV^s)\\ &= (\delta^{ip}\delta^{jq})M_{pq}(B^b_j\delta_{bc}V^c)(B^r_i\delta_{rs}V^s),\end{aligned}$$

so that

$$\langle MB(y)t^2\overline{\nabla}f, B(y)t^2\overline{\nabla}f\rangle = \langle M, \Psi\rangle,$$

for some $\Psi\in\mathbb{R}_*[F]\otimes\overline{\mathcal{Q}}^2_{*,*}[X, \overline{\nabla}F][\![T]\!]_{(2,1,2)}$.

Now, by (41), $\nabla\hat{f}$ contains a term in x that does not carry a factor of t. We need to show that this term in x is not repeated in any nontrivial component of (74). However, since $D_xx = \nabla_xx = 0$, we have $\Gamma(x, x) = 0$, so that, for all U,

$$\langle A(y)\Gamma(x, x), U\rangle = 0.$$

Next, since $g(y)(x,x)=1$ for all y, we obtain, for all vectors U,

$$0=g(y)(\nabla_U x,x)=g(y)(D_U x+\Gamma(U,x),x),$$

so that

$$\begin{aligned}\langle A(y)\Gamma(U,x),x\rangle&=\langle A(y)\Gamma(x,U),x\rangle\\&=-\langle A(y)D_U x,x\rangle=-\frac{1}{r}\langle A(y)\pi(U),x\rangle=0.\end{aligned}$$

Finally, since $D_x\overline{\nabla}f=0$, and since $\langle B(y)t^2\overline{\nabla}f,x\rangle=0$ for all y, we have

$$\langle DB(y,x)t^2\overline{\nabla}f,x\rangle=D_x\langle B(y)t^2\overline{\nabla}f,x\rangle=0,$$

and we conclude that the term in x is not repeated in any nontrivial component of (74), as desired. It follows that

$$g(\nabla_{\nabla\hat{f}}\nabla\hat{f},\nabla\hat{f})=\frac{1}{t}\Phi_4+\Phi_5+\langle\Phi_6,t^2(\overline{\mathrm{Hess}}\circ\pi)\rangle,$$

where

$$\begin{aligned}\Phi_4,\Phi_5&\in\mathbb{R}_*[F]\otimes\overline{\mathcal{Q}}^0_{*,*}[X,\overline{\nabla}F][\![T]\!]_{(2,1,2)},\\ \Phi_6&\in\mathbb{R}_*[F]\otimes\overline{\mathcal{Q}}^2_{*,*}[X,\overline{\nabla}F][\![T]\!]_{(2,1,2)}.\end{aligned}$$

The result for $Y=0$ now follows by Proposition 3.2.2. The general case follows by substituting for every generator $R_{i_1i_2i_3}{}^{j}{}_{;i_4\ldots i_{k+3}}$ of $\mathcal{R}$ its own Taylor series in Y about 0, as in the proof of Proposition 3.2.4. This completes the proof. □

4. Asymptotic expansions and formal solutions

4.1. ***Asymptotic expansions.*** In order to save on notation, which would otherwise quickly get out of hand, we shall no longer be so explicit about the definition of curvature polynomials, leaving the reader to infer how they are constructed in each case. We now reformulate the results of the previous sections in a manner that will allow us to construct formal solutions later on. To this end, we introduce the terminology of asymptotic expansions for functions defined near $t=0$ as follows. Let E be a finite-dimensional vector bundle over some finite-dimensional base B. Let $\phi:]0,\infty[\times E\to\mathbb{R}$ be a smooth function. Let (ϕ_k) be a sequence of smooth functions, where, for all k, $\phi_k:E^{\otimes k}\to\mathbb{R}$. For a formal power series $\xi_x(t)\sim\sum_{k=0}^{\infty}t^k\xi_{k,x}$ in E, we write

$$\phi(t,\xi_x)\sim\sum_{k=0}^{\infty}t^k\phi_k(\xi_{0,x},\ldots,\xi_{k,x})\tag{75}$$

to mean that for all $N \geq 0$, there exists a smooth function $R_N : [0, \infty[\times E^{\otimes(N+1)} \to \mathbb{R}$ such that

$$\phi\left(t, \sum_{k=0}^{N} t^k \xi_{x,k}\right) = \sum_{k=0}^{N} t^k \phi_k(\xi_{0,x}, \dots, \xi_{k,x}) + t^{N+1} R_N(t, \xi_{x,0}, \dots, \xi_{x,N}). \tag{76}$$

It is key to our definition that the remainder term R_N be smooth also at $t = 0$, as it would otherwise be of little use to us.

Proposition 4.1.1. *There exists a sequence (P_k) of curvature polynomials such that for all formal power series $Y \sim \sum_{k=0}^{\infty} t^k Y_k$ of vectors in $\mathbb{R}^{m+1}$ and $f_x \sim \sum_{k=0}^{\infty} t^k f_{k,x}$ of germs in $J^0 S^m$, and for all vectors U,*

$$t\langle p(t, y, f_x), U\rangle \sim t\langle x, U\rangle + \sum_{k=0}^{\infty} t^k \langle P_k(f_{0,x}, \dots, f_{k-3,x}, Y_0, \dots, Y_{k-3}), U\rangle, \tag{77}$$

and $P_k = 0$ for $k \leq 2$.

Proof. By Proposition 3.3.1, there exists a smooth function $\widetilde{P}$ such that

$$\langle p(t, Y, f_x), U\rangle = \langle x, U\rangle + t^2 \langle \widetilde{P}(t, Y, f_x), U\rangle.$$

Furthermore, since the coefficients of the Taylor series of $\widetilde{P}$ in t are all curvature polynomials, there exists a sequence $(\widetilde{P}_k)$ of curvature polynomials such that

$$\widetilde{P}(t, Y, f_x) \sim \sum_{k=0}^{\infty} t^k \widetilde{P}_k(f_{0,x}, \dots, f_{k,x}, Y_0, \dots, Y_k).$$

It follows that

$$t\langle p(t, Y, f_x), U\rangle \sim t\langle x, U\rangle + \sum_{k=3}^{\infty} t^k \langle \widetilde{P}_{k-3}(f_{0,x}, \dots, f_{k-3,x}, Y_0, \dots, Y_{k-3}), U\rangle. \qquad \square$$

Proposition 4.1.2. *There exists a sequence (Q_k) of curvature polynomials such that for all formal power series $Y \sim \sum_{k=0}^{\infty} t^k Y_k$ and $V \sim \sum_{k=0}^{\infty} t^k V_k$ of vectors in $\mathbb{R}^{m+1}$ and $f_x \sim \sum_{k=0}^{\infty} t^k f_{k,x}$ of germs in $J^0 S^m$,*

$$\begin{aligned} &t\langle q(t, Y, f_x), V\rangle \\ &\qquad \sim \sum_{k=0}^{\infty} t^k (\langle x, V_{k-2}\rangle + Q_k(f_{0,x}, \dots, f_{k-4,x}, Y_0, \dots, Y_{k-4}, V_0, \dots, V_{k-4})), \end{aligned} \tag{78}$$

where $Q_k = 0$ for $k \leq 3$.

Proof. By Proposition 3.3.2, there exists a smooth function $\widetilde{Q}$ such that

$$\langle q(t, Y, f_x), V\rangle = t\langle x, V\rangle + t^3 \langle \widetilde{Q}(t, Y, f_x), V\rangle.$$

Furthermore, since the coefficients of the Taylor series of $\widetilde{Q}$ in t are all curvature polynomials, there exists a sequence $(\widetilde{Q}_k)$ of curvature polynomials such that

$$\langle \widetilde{Q}(t, Y, f_x), V\rangle \sim \sum_{k=0}^{\infty} t^k \widetilde{Q}_k(f_{0,x}, \dots, f_{k,x}, Y_0, \dots, Y_k, V_0, \dots, V_k).$$

It follows that

$$\begin{aligned} &t\langle q(t, Y, f_x), V\rangle \\ &\sim \sum_{k=2}^{\infty} t^k \langle x, V_{k-2}\rangle + \sum_{k=4}^{\infty} \widetilde{Q}_{k-4}(f_{0,x}, \dots, f_{k-4,x}, Y_0, \dots, Y_{k-4}, V_0, \dots, V_{k-4}). \end{aligned}$$ □

Proposition 4.1.3. *There exists a sequence (R_k) of curvature polynomials such that for all formal power series $Y \sim \sum_{k=0}^{\infty} t^k Y_k$ of vectors in $\mathbb{R}^{m+1}$ and $f_x \sim \sum_{k=0}^{\infty} t^k f_{k,x}$ and $g_x \sim \sum_{k=0}^{\infty} t^k g_{k,x}$ of germs in $J^0 S^m$,*

$$\begin{aligned} &\mathrm{tr}(t, Y, f_x) g_x \\ &\sim \sum_{k=0}^{\infty} t^k (t^4 g_{k,x} + R_k(Y_0, \dots, Y_{k-4}, f_{0,x}, \dots, f_{k-4,x}, t^4 g_{0,x}, \dots, t^4 g_{k-4,x})), \end{aligned} \tag{79}$$

where $R_k = 0$ for $k \leq 4$.

Proof. By Proposition 3.3.3, there exists a smooth function $\widetilde{R}$ such that $r(t, Y, f_x) = t^3 + t^7 \widetilde{R}(t, Y, f_x)$. Furthermore, since the coefficients of the Taylor series of $\widetilde{R}$ in t are all curvature polynomials, there exists a sequence $(\widetilde{R}_k)$ of curvature polynomials such that

$$\widetilde{R}(t, Y, f_x) \sim \sum_{k=0}^{\infty} t^k \widetilde{R}_k(Y_0, \dots, Y_k, f_{0,x}, \dots, f_{k,x}).$$

Thus

$$t^4 \widetilde{R}(t, Y, f_x) g_x \sim \sum_{k=0}^{\infty} t^k \sum_{l=0}^{k} \widetilde{R}_l(Y_0, \dots, Y_l, f_{0,x}, \dots, f_{l,x})(t^4 g_{k-l,x}).$$

It follows that

$$\begin{aligned} &t R(t, Y, f_x) g_x \\ &\sim \sum_{k=0}^{\infty} t^k (t^4 g_{k,x}) + \sum_{k=4}^{\infty} t^k \sum_{l=0}^{k-4} \widetilde{R}_l(Y_0, \dots, Y_l, f_{0,x}, \dots, f_{l,x})(t^4 g_{k-l-4,x}). \end{aligned}$$ □

Proposition 4.1.4. *There exists a sequence (H_k) of curvature polynomials such that for all formal power series $Y \sim \sum_{k=0}^{\infty} t^k Y_k$ of vectors in $\mathbb{R}^{m+1}$ and $f_x \sim \sum_{k=0}^{\infty} t^k f_{k,x}$*

of germs in J^2S^m,

$$\begin{aligned}(80)\quad &\frac{1}{t}\Big(H(t,Y,f)-\frac{1}{t}\Big)\\ &\sim\sum_{k=0}^{\infty}t^k\Big(-\frac{1}{n}(n+\Delta)f_{k,x}-\tfrac{1}{4}\operatorname{Ric}_{pq;rs}x^px^qx^rY^s_{k-2}\\ &\qquad\qquad -\tfrac{1}{3}\operatorname{Ric}_{pq;r}x^px^qY^r_{k-1}+H_k(Y_0,\dots,Y_{k-3},f_{0,x},\dots,f_{k-2,x})\Big),\end{aligned}$$

where, *by convention*, $Y_k=0$ *for* $k<0$. *Furthermore*,

$$(81)\qquad \begin{aligned}H_0&=-\tfrac{1}{3}\operatorname{Ric}_{pq}x^px^q,\\ H_1&=-\tfrac{1}{4}\operatorname{Ric}_{pq;r}x^px^qx^r.\end{aligned}$$

Proof. Consider the formula (69) for H. Trivially,

$$\frac{1}{n}(n+\overline{\Delta})f_x\sim\sum_{k=0}^{\infty}t^k\frac{1}{n}(n+\overline{\Delta})f_{k,x},$$

$$\frac{t}{3}\operatorname{Ric}_{pq;r}x^px^qY^r\sim\sum_{k=1}^{\infty}\frac{t^k}{3}\operatorname{Ric}_{pq;r}x^px^qY^r_{k-1},$$

$$\frac{t^2}{4}\operatorname{Ric}_{pq;rs}x^px^qx^rY^s\sim\sum_{k=2}^{\infty}\frac{t^k}{4}\operatorname{Ric}_{pq;r}x^px^qx^rY^r_{k-2}.$$

Since F is a curvature polynomial, there exists a sequence (F_k) of curvature polynomials such that

$$F(f_x)\sim\sum_{k=0}^{\infty}t^kF_k(f_{0,x},\dots,f_{k,x}).$$

In particular,

$$t^2F(f_x)\sim\sum_{k=2}^{\infty}t^kF_{k-2}(f_{0,x},\dots,f_{k-2,x}).$$

Finally, denote the remainder term in (69) by $t^5G(t,Y,f)$. Since every coefficient in the Taylor series of G in t about 0 is a curvature polynomial, there exists a sequence (G_k) of curvature polynomials such that

$$G(t,Y,f)\sim\sum_{k=0}^{\infty}t^kG_k(Y_0,\dots,Y_k,f_{0,x},\dots,f_{k,x}).$$

In particular,

$$t^3G(t,Y,f)\sim\sum_{k=3}^{\infty}t^kG_{k-3}(Y_0,\dots,Y_{k-3},f_{0,x},\dots,f_{k-3,x}),$$

and the result follows upon combining these terms. □

4.2. ***Flows of surfaces.*** We now extend our framework to the time-dependent case. Thus, let M be an $(m+1)$-dimensional Riemannian manifold with metric g, let R be its Riemann curvature tensor, let S be its scalar curvature function, and suppose that S is of Morse–Smale type. Let $\gamma : \mathbb{R} \to M$ be a complete integral curve of $-\nabla S$ with relatively compact image. In particular (see [Schwarz 1993]), $\gamma(t)$ converges exponentially to critical points of S as t tends to $\pm\infty$, and its derivatives to all orders decay exponentially at infinity.

For convenience, we suppose that M has unit injectivity radius. We identify the bundle γ^*TM with the product bundle $\mathbb{R} \times \mathbb{R}^{m+1}$ via parallel transport. For all $t \in \mathbb{R}$, define the metric g_t over $\mathbb{R}^{m+1}$ by $g_t := \mathrm{Exp}^*_{\gamma(t)}\, g$, where $\mathrm{Exp}_{\gamma(t)}$ here denotes the exponential map of M about the point $\gamma(t)$. In particular, for all t, the metric g_t is of the type introduced in Section 2.1. Furthermore, the family (g_t) converges exponentially in the C^∞_{loc} sense to metrics $g_{\pm\infty}$ as t tends to $\pm\infty$ and its time derivatives to all orders also decay exponentially at infinity.

As in Section 2.5, for all $t \in \mathbb{R}$, let $\mathrm{Exp}_t : \Omega_2 \to \mathbb{R}^{m+1}$ be the exponential map of g_t. That is, for all $(x, y) \in \Omega_2$, the curve $s \mapsto \mathrm{Exp}_t(x, sy)$ is the unique geodesic with respect to g_t leaving the point x in the direction of the vector y. For $s > 0$, and for bounded functions $Y \in C^0(\mathbb{R}, \mathbb{R}^{m+1})$ and $f \in C^0(\mathbb{R} \times S^m)$, define $e(s, Y, f) : \mathbb{R} \times S^m \to \mathbb{R}^{m+1}$ by

$$e(s, Y, f)(t, x) := \mathrm{Exp}_t(sY(t), s(1 + s^2 f(t, x))x). \tag{82}$$

Heuristically, $e(s, Y, f)$ is a continuous family of immersed spheres all of radius approximately s, with centres displaced by the function Y. Composing with $\mathrm{Exp}_{\gamma(s)}$ then yields a continuous family of small immersed spheres in M which move along the geodesic γ. We will show that for sufficiently small s and for correct choices of Y and f, this family yields a forced mean curvature flow of immersed spheres in M with forcing term $1/s$.

For all k, let $J^k(\mathbb{R}, \mathbb{R}^{m+1})$ denote the bundle of k-jets over $\mathbb{R}$ taking values in $\mathbb{R}^{m+1}$. For all (k, l), let $J^{k,l}(\mathbb{R} \times S^m, \mathbb{R})$ denote the bundle of (k, l)-jets over $\mathbb{R} \times S^m$ taking values in $\mathbb{R}$, that is, the bundle of $\mathbb{R}$-valued jets that are of order at most k in $\mathbb{R}$ and at most l in S^m. Observe that $J^{k,l}(\mathbb{R} \times S^m, \mathbb{R})$ is actually also a bundle over $\mathbb{R}$ and we denote by $J := J^{k,l}$ its fibrewise cartesian product with $J^k(\mathbb{R}, \mathbb{R}^{m+1})$. In other words, an element of $J^{k,l}$ is a pair $(Y_t, f_{t,x})$ where Y_t is the jet of an $\mathbb{R}^{m+1}$-valued function over $\mathbb{R}$ at the point t, and $f_{t,x}$ is the jet of an $\mathbb{R}$-valued function over $\mathbb{R} \times S^m$ at the point (t, x).

Define the functions $N :]0, \infty[\times J \to S^m$ and $H :]0, \infty[\times J \to \mathbb{R}$ such that for all $s \in]0, \infty[$ and for all $(Y_t, f_{t,x}) \in J$, $N(s, Y_t, f_{t,x})$ and $H(s, Y_t, f_{t,x})$ are respectively the outward-pointing unit normal of the immersion $e(s, Y, f)(t, \cdot)$ at the point $e(s, Y, f)(t, x)$ and its mean curvature at that point, both with respect to the metric g_t. Define $V :]0, \infty[\times J \to \mathbb{R}^{m+1}$ by

$$V(s, (Y_t, f_{t,x})) := \partial_r \operatorname{Exp}^{-1}_{\gamma(t)}(\operatorname{Exp}_{\gamma(r)}(e(s, Y, f)(t+r, x)))|_{r=0}. \tag{83}$$

Heuristically, this vector field measures the variation of the immersion $e(s, Y, f)$ at the point $e(s, Y, f)(t, x)$ as we move along the flow. Finally, define $\Phi :]0, \infty[\times J \to \mathbb{R}$ by

$$\begin{aligned}\Phi(s, (Y_t, f_{t,x}))&\\ &:= \frac{1}{s}\Big(H(s, (Y_t, f_{t,x})) - \frac{1}{s}\Big) + s\langle V(s, (Y_t, f_{t,x})), N(s, (Y_t, f_{t,x}))\rangle.\end{aligned} \tag{84}$$

For all s, $\Phi(s, \cdot)$ is the *forced mean curvature flow operator* (with forcing term $1/s$). In particular, it is a quasilinear parabolic partial differential operator whose zeroes are (reparametrised) forced mean curvature flows with forcing term $1/s$.

Proposition 4.2.1. *There exists a sequence* (Φ_k) *of curvature polynomials such that for all formal power series* $(Y_t, f_{t,x}) \sim \sum_{k=0}^{\infty} s^k (Y_t, f_{t,x})$ *of germs in* J,

$$\begin{aligned}\Phi(s, Y_t, f_{t,x})&\\ \sim \sum_{k=0}^{\infty} s^k \Big[&\Big\langle\Big(\frac{\partial}{\partial t} + \frac{(m+1)}{2(m+3)}\operatorname{Hess}(S)\Big)Y_{k-2,t}, x\Big\rangle\\ &+ \Big(s^4\frac{\partial}{\partial t} + \frac{1}{m}(m + \overline{\Delta})\Big) f_{k,x,t}\\ &+ \Big(\tfrac{1}{4}\operatorname{Ric}_{t,ab;cd} x^a x^b x^c Y^d_{k-2,t} - \frac{(m+1)}{2(m+3)} S_{t,;ab} x^a Y^b_{k-2,t}\Big)\\ &- \tfrac{1}{3}\operatorname{Ric}_{t,ab;c} x^a x^b Y^c_{k-1,t}\\ &+ \Phi_k\big(f_{0,x,t}, \dots, f_{k-2,x,t}, s^4 \dot{f}_{0,x,t}, \dots, s^4 \dot{f}_{k-4,x,t},\\ &\qquad Y_{0,t}, \dots, Y_{k-3,t}, \dot{Y}_{0,t}, \dots, \dot{Y}_{k-4,t}\big)\Big],\end{aligned} \tag{85}$$

where Ric_t *and* S_t *denote respectively the Ricci and scalar curvatures of* M *at the point* $\gamma(t)$, *and*, *by convention*, $Y_k = 0$ *for* $k < 0$. *Furthermore*, *the curvature polynomials* Φ_0 *and* Φ_1 *are given by*

$$\Phi_0 = -\tfrac{1}{3}\operatorname{Ric}_{ab} x^a x^b, \quad \Phi_1 = -\tfrac{1}{4}\operatorname{Ric}_{ab;c} x^a x^b x^c + \frac{(m+1)}{2(m+3)} S_{;a} x^a. \tag{86}$$

Remark. Importantly, since they are curvature polynomials, the functions (Φ_k) vary with t only insofar as the curvature tensor itself, along with its derivatives, vary, and the same can also be said for the remainder terms in the asymptotic series. In particular, since the flow γ has relatively compact image in M, the derivatives of all these functions to all orders are uniformly bounded independent of s and t.

Remark. Observe that, as in Proposition 4.1.3, in every remainder term of this asymptotic series, the term $\dot{f}_k$ only ever appears accompanied by the factor s^4.

Proof. Indeed

$$V(t,(Y_t,f_{t,x}))=P(t,(Y_t,f_{t,x}))\dot{\gamma}+Q(t,(Y_t,f_{t,x}))\dot{Y}_t+R(t,(Y_t,f_{t,x}))\dot{f}_{t,x}.$$

Furthermore, since γ is a gradient flow of S,

$$\dot{\gamma}=-\frac{(m+1)}{2(m+3)}S_{;a}x^a.$$

The result now follows by Propositions 4.1.1, 4.1.2, 4.1.3 and 4.1.4. ☐

4.3. ***Parabolic operators I: the finite dimensional case.*** We first aim to determine formal solutions of the equation $\Phi(s,Y,f)=0$ for small values of s. To this end, we introduce the following functional analytic framework. For a finite-dimensional vector space E and for $\alpha\in]0,1]$, define the *Hölder seminorm* of order α over $C^0(\mathbb{R},E)$ by

$$[f]_\alpha:=\mathrm{Sup}_{0<|s-t|\leq 1}\frac{\|f(s)-f(t)\|}{|s-t|^\alpha}. \tag{87}$$

For all k and for all $\alpha\in]0,1]$, define the *Hölder norm* of order (k,α) over $C^k(\mathbb{R},E)$ by

$$\|f\|_{k,\alpha}:=\sum_{i=0}^{k}\|\partial_t^i f\|_0+[\partial_t^k f]_\alpha, \tag{88}$$

where $\|\cdot\|_0$ denotes the uniform norm. For all (k,α), define the *Hölder space* of order (k,α) by

$$C^{k,\alpha}(\mathbb{R},E):=\{f\in C^k(\mathbb{R},E)\mid \|f\|_{k,\alpha}<\infty\}. \tag{89}$$

Recall that $C^{k,\alpha}$ furnished with the norm $\|\cdot\|_{k,\alpha}$ constitutes a Banach space.

Define the operator $P:C^{1,\alpha}(\mathbb{R},\mathbb{R}^{m+1})\to C^{0,\alpha}(\mathbb{R},\mathbb{R}^{m+1})$ by

$$PY=\Big(\frac{\partial}{\partial t}+\frac{(m+1)}{2(m+3)}\,\mathrm{Hess}(S)\Big)Y. \tag{90}$$

Observe that this operator corresponds to the first summand in the asymptotic expansion (85) of Φ. Furthermore, since S is of Morse–Smale type, P is Fredholm and surjective. In addition, since every function in $\mathrm{Ker}(P)$ decays exponentially at infinity [Schwarz 1993], the L^2 orthogonal complement $\mathrm{Ker}(P)^\perp$ of $\mathrm{Ker}(P)$ in $C^{1,\alpha}(\mathbb{R},\mathbb{R}^{m+1})$ is well-defined. The restriction of P to $\mathrm{Ker}(P)^\perp$ is invertible, and we denote its inverse by G.

We will also be interested in families of constant coefficient parabolic operators over $C^1(\mathbb{R}, E)$. Thus, for an invertible linear map $A : E \to E$, which, for convenience, we take to be symmetric with respect to some fixed metric over E, and for $\epsilon > 0$, define $P_\epsilon : C^1(\mathbb{R}, E) \to C^0(\mathbb{R}, E)$ by

$$P_\epsilon f := (\epsilon\partial_t - A)f. \tag{91}$$

It follows from the invertibility of A that P_ϵ as also invertible. In fact, its Green's operator, which we denote by G_ϵ, is given by

$$G_\epsilon f(t) = -\frac{1}{\epsilon}\int_{-\infty}^{t} e^{-1/\epsilon(t-s)A^+} f(s)\,ds + \frac{1}{\epsilon}\int_{t}^{\infty} e^{-1/\epsilon(t-s)A^-} f(s)\,ds, \tag{92}$$

where A^+ (resp. A^-) denotes the composition of A with the orthogonal projection onto the direct sum of its eigenspaces of positive (resp. negative) eigenvalue. In order to obtain uniform estimates for the operator norm of G_ϵ, it is useful to introduce a weighting factor into the Hölder norm. Thus, for all (k, α) and for all $\epsilon > 0$, define the *weighted Hölder norm* of order (k, α) and weight ϵ by

$$\|f\|_{k,\alpha,\epsilon} := \sum_{i=0}^{k} \epsilon^i \|\partial_t^i f\|_0 + \epsilon^k [\partial_t^i f]_\alpha \tag{93}$$

Observe that, for all ϵ, the norm $\|\cdot\|_{k,\alpha,\epsilon}$ is uniformly equivalent to the norm $\|\cdot\|_{k,\alpha}$, so that $C^{k,\alpha}(\mathbb{R} \times E)$ is also a Banach space with respect to every weighted Hölder norm.

Proposition 4.3.1. *There exists* $B > 0$, *which only depends on the matrix* A, *such that for all* $\epsilon > 0$, *and for all* $f \in C^{0,\alpha}(\mathbb{R}, E)$,

$$\|G_\epsilon f\|_{1,\alpha,\epsilon} \leq B\|f\|_{0,\alpha}. \tag{94}$$

Proof. Since both P_ϵ and G_ϵ preserve the eigenspaces of A, we may suppose that $E = \mathbb{R}$ and that $A = \lambda > 0$. Thus,

$$G_\epsilon f(t) = -\frac{1}{\epsilon}\int_{-\infty}^{t} e^{-\lambda(t-s)/\epsilon} f(s)\,ds = -\frac{1}{\epsilon}\int_{0}^{\infty} e^{-\lambda s/\epsilon} f(t-s)\,ds.$$

Now fix $f \in C^{0,\alpha}(\mathbb{R}, \mathbb{R})$. For all t,

$$|G_\epsilon f(t)| \leq \frac{1}{\epsilon}\int_0^\infty e^{-\lambda/(\epsilon s)} |f(s)|\,ds \leq \frac{1}{\lambda}\|f\|_0,$$

and taking the supremum over all t yields $\|G_\epsilon f\|_0 \leq 1/\lambda\ \|f\|_0$. Likewise, for all $0 < |t - t'| \leq 1$,

$$|G_\epsilon f(t) - G_\epsilon f(t')| \leq \frac{1}{\epsilon}\int_0^\infty e^{-\lambda/\epsilon\, s} |f(t-s) - f(t'-s)| \leq \frac{1}{\lambda}|t - t'|^\alpha [f]_\alpha.$$

Dividing both sides by $|t-t'|^\alpha$, and taking the supremum over all t yields $[G_\epsilon f]_\alpha \leq 1/\lambda\,[f]_\alpha$. Combining these relations yields $\|G_\epsilon f\|_{0,\alpha} \leq \|A\|^{-1}\|f\|_{0,\alpha}$. Finally, by definition of G_ϵ, $\epsilon\partial_t G_\epsilon f = \lambda G_\epsilon f + f$, so that $\epsilon\|\partial_t G_\epsilon f\|_{0,\alpha} \leq \lambda\|G_\epsilon f\|_{0,\alpha} + \|f\|_{0,\alpha} \leq 2\|f\|_{0,\alpha}$. This completes the proof. □

4.4. ***Parabolic operators II: the infinite-dimensional case.*** For all $\alpha \in]0, 1]$, define the *Hölder seminorms* of order α over $C^0(\mathbb{R}\times\mathbb{S}^m)$ by

$$\begin{aligned}[f]_{x,\alpha} &:= \mathrm{Sup}_{t,x\neq y}\,\frac{|f(t,x)-f(t,y)|}{\|x-y\|^\alpha},\\ [f]_{t,\alpha} &:= \mathrm{Sup}_{x,0|t-s|\leq 1}\,\frac{|f(s,x)-f(t,x)|}{|s-t|^\alpha}.\end{aligned} \tag{95}$$

For all $k\in\mathbb{N}$, let $C^k_{\text{in}}(\mathbb{R}\times S^m)$ be the set of all functions $f:\mathbb{R}\times S^m\to\mathbb{R}$ which are continuously differentiable i times in the x direction and j times in the t direction for all $i+2j\leq 2k$. For all $k\in\mathbb{N}$ and for all $\alpha\in]0,1/2]$, define the *inhomogeneous Hölder norm* of order (k,α) over $C^k_{\text{in}}(\mathbb{R}\times S^m)$ by

$$\begin{aligned}&\|f\|_{k,\alpha,\text{in}}\\ &\quad:= \sum_{i+2j\leq 2k}\|D^i_x D^j_t f\|_0 + \sum_{i+2j=2k}[D^i_x D^j_t f]_{x,2\alpha} + \sum_{i+2j=2k}[D^i_x D^j_t f]_{t,\alpha}.\end{aligned} \tag{96}$$

For all k,α, define the *inhomogeneous Hölder space* of order (k,α) by

$$C^{k,\alpha}_{\text{in}}(\mathbb{R}\times S^m) := \{f\in C^k_{\text{in}}(\mathbb{R}\times S^m)\mid \|f\|_{k,\alpha,\text{in}}<\infty\}. \tag{97}$$

Recall that $C^{k,\alpha}_{\text{in}}(\mathbb{R}\times S^m)$ furnished with the norm $\|\cdot\|_{k,\alpha,\text{in}}$ constitutes a Banach space. More generally, for all (k,α) and for all $\epsilon>0$, define the *weighted inhomogeneous Hölder norm* of order (k,α) and weight ϵ over $C^k_{\text{in}}(\mathbb{R}\times S^m)$ by

$$\begin{aligned}\|f\|_{k,\alpha,\text{in},\epsilon} &:= \sum_{i+2j\leq 2k}\epsilon^j\|D^i_x D^j_t f\|_0\\ &\quad+ \sum_{i+2j=2k}\epsilon^j[D^i_x D^j_t f]_{x,2\alpha} + \sum_{i+2j=2k}\epsilon^j[D^i_x D^j_t f]_{t,\alpha}.\end{aligned} \tag{98}$$

For all $\epsilon>0$, the norm $\|\cdot\|_{k,\alpha,\text{in},\epsilon}$ is uniformly equivalent to the norm $\|\cdot\|_{k,\alpha,\text{in}}$ so that $C^{k,\alpha}_{\text{in}}(\mathbb{R}\times S^m)$ is also a Banach space with respect to every weighted inhomogeneous Hölder norm.

For all $s>0$, define the operator $Q_s : C^{1,\alpha}_{\text{in}}(\mathbb{R}\times S^m)\to C^{0,\alpha}_{\text{in}}(\mathbb{R}\times S^m)$ by

$$Q_s f := \Big(s^4\frac{\partial}{\partial t} + \frac{1}{m}(m+\overline{\Delta})\Big)f, \tag{99}$$

where, as in Section 3, $\overline{\Delta}$ denotes the Laplacian of the standard metric over S^m. Observe that this operator corresponds to the second summand in the asymptotic expansion (85) of Φ. Furthermore, the operator $(m+\overline{\Delta})$ defines a self-adjoint operator over $L^2(S^m)$ with kernel $\mathcal{H}_1$, the space of restrictions to S^m of linear functions over $\mathbb{R}^{m+1}$. In particular, $(m+\overline{\Delta})$ restricts to an invertible mapping of the orthogonal complement $\mathcal{H}_1^{\perp}$ to itself. With this in mind, for all k and for all α, we define

$$\hat{C}^{k,\alpha}_{\text{in}}(\mathbb{R}\times S^m) := \left\{ f\in C^{k,\alpha}_{\text{in}}(\mathbb{R}\times S^m) \mid \int_{S^m} f(t,x)x^i\,d\text{Vol}=0\ \forall 1\leq i\leq m+1\right\}. \tag{100}$$

It follows from the classical theory of parabolic operators that, for all s, Q_s restricts to an invertible mapping from $\hat{C}^{1,\alpha}_{\text{in}}(\mathbb{R}\times S^m)$ into $\hat{C}^{0,\alpha}_{\text{in}}(\mathbb{R}\times S^m)$. Uniform norm estimates for Green's operators in the infinite-dimensional setting differ significantly from those obtained in the finite-dimensional setting. Indeed,

Lemma 4.4.1. *There exists $B>0$ such that for all $s\leq 1$ and for all $f\in\hat{C}^{0,\alpha}_{\text{in}}(\mathbb{R}\times S^m)$*

$$\|H_s f\|_{1,\alpha,\text{in},s^4}\leq Bs^{-4\alpha}\|f\|_{0,\alpha,\text{in}}. \tag{101}$$

Remark. Although it may appear that this weaker estimate is merely a consequence of the naive approach to the proof, the study of solutions of the heat equation in euclidean space appears to indicate that it is probably optimal.

Remark. Alternatively, it may appear that this weaker estimate arises from the unusual definition (98) of the weighted inhomogeneous Hölder norm. Indeed, it would surely have made more sense to have multiplied the third summand of (98) by a factor of ϵ^{α}, thereby eliminating the factor of $s^{-4\alpha}$ from (101). However, we have chosen the above definition so that the operator $s^4\partial_t$ has unit norm with respect to the norms $\|\cdot\|_{1,\alpha,\text{in},s^4}$ and $\|\cdot\|_{0,\alpha,\text{in}}$, which ensures that other factors of $s^{-4\alpha}$ do not enter into our reasoning in places where they would present a greater technical nuisance.

Proof. For all $s>0$, define the isomorphism D_s of $C^{k,\alpha}_{\text{in}}(\mathbb{R}\times S^m)$ by $D_s f(t,x)=f(s^4t,x)$. For all $s\leq 1$, and for all $f\in C^{0,\alpha}_{\text{in}}(\mathbb{R}\times S^m)$, $\|D_s f\|_{0,\alpha,\text{in}}\leq\|f\|_{0,\alpha,\text{in}}$. On the other hand, for all $s\leq 1$ and for all $f\in C^{1,\alpha}_{\text{in}}(\mathbb{R}\times S^m)$, $\|D_s^{-1}f\|_{1,\alpha,\text{in},s^4}\leq t^{-4\alpha}\|f\|_{1,\alpha,\text{in}}$. However, for all s, $Q_s=D_s^{-1}Q_1D_s$. The result follows. ☐

Observe that $\mathcal{H}_1$ is really the space of eigenfunctions of $\overline{\Delta}$ of eigenvalue m. More generally, the decomposition of $L^2(S^m)$ into eigenspaces of $\overline{\Delta}$ actually yields better estimates for $\|H_s f\|_{1,\alpha,\text{in},s^4}$ in the case where $f(t,\cdot)$ is the restriction to S^m of an s-dependent polynomial function of bounded order. Indeed, for all l, let $\mathcal{H}_l\subseteq L^2(S^m)$ be the space of *spherical harmonics* of order l over S^m, that is, the space of eigenfunctions of the operator $\overline{\Delta}$ with eigenvalue $l(m+l-1)$. Recall that, for all l, $\mathcal{H}_l$ is the restriction to S^m of the space of homogeneous harmonic polynomials of order l over $\mathbb{R}^{m+1}$. In particular, any polynomial of order l over $\mathbb{R}^{m+1}$ restricts to an element of $\mathcal{H}_0 \oplus \cdots \oplus \mathcal{H}_l$ over S_m. Now define

$$\hat{\mathcal{H}}_l := \oplus_{i=0, i \neq 1}^{l} \mathcal{H}_i. \tag{102}$$

Observe that $\hat{\mathcal{H}}_l$ is contained in $\mathcal{H}_1^\perp$ for all l. Furthermore, for all l and for all (k, α), $C^{k,\alpha}(\mathbb{R}, \hat{\mathcal{H}}_l)$ naturally identifies with a subspace of $\hat{C}^{k,\alpha}_{\text{in}}(\mathbb{R} \times S^m)$. In particular, for all s, Q_s restricts to a mapping from $C^{1,\alpha}(\mathbb{R}, \hat{\mathcal{H}}_l)$ to $C^{0,\alpha}(\mathbb{R}, \hat{\mathcal{H}}_l)$. Furthermore, this restriction is invertible for all s, and Proposition 4.3.1 now yields

Proposition 4.4.2. *For all $l \in \mathbb{N}$, there exists $B_l > 0$ such that for all $f \in C^{0,\alpha}(\mathbb{R}, \hat{\mathcal{H}}_{\leq l})$ and for all ϵ,*

$$\|H_s f\|_{1,\alpha,\text{in},s^4} \leq B_l \|f\|_{0,\alpha,\text{in}}.$$

4.5. ***More on spherical harmonics.*** A tensor $T^{i_1 \dots i_k}$ is said to be *isotropic* whenever

$$A^{i_1}_{j_1} \dots A^{i_k}_{j_k} T^{j_1 \dots j_k} = T^{i_1 \dots i_k}, \tag{103}$$

for all $i_1, \dots, i_k$ and for every special-orthogonal matrix A. Given two symmetric tensors $T_1^{i_1 \dots i_k}$ and $T_2^{i_1 \dots i_l}$, their symmetric product is given by

$$(T_1 \odot T_2)^{i_1 \dots i_{k+l}} := \sum_{\sigma \in \widetilde{\Sigma}_{k,l}} T_1^{i_{\sigma(1)} \dots 1_{\sigma(k)}} T_2^{i_{\sigma(k+1)} \dots 1_{\sigma(k+l)}}, \tag{104}$$

where $\widetilde{\Sigma}_{k,l}$ denotes the set of permutations of the set $\{1, \dots, k+l\}$ such that $\sigma(1) < \cdots < \sigma(k)$ and $\sigma(k+1) < \cdots < \sigma(k+l)$. Let δ be as in Section 2.1. In particular, δ is symmetric and isotropic. Furthermore, for all k, its k-th symmetric power $\delta^{\odot k}$ is also a symmetric and isotropic tensor. In fact, up to rescaling, these are the only ones.

Lemma 4.5.1. *The space of symmetric, isotropic tensors of order k is 1-dimensional when k is even, and 0-dimensional when k is odd.*

Proof. Indeed, the space of symmetric tensors of order k is isomorphic to the space of homogeneous polynomials of the same order. However, since an $\mathrm{SO}(m+1)$-invariant polynomial is constant over every sphere centred on the origin, it is determined by its restriction to any straight line passing through the origin. When, in addition, this polynomial is homogeneous, it is determined by its value at a single point. This space thus has dimension at most 1. Now observe that the restriction of a homogeneous polynomial to a straight line through the origin is even when its order is even, and odd when its order is odd. However, by $\mathrm{SO}(m+1)$-invariance again, the restrictions of the polynomials considered here are always even. It follows that there are no nontrivial symmetric, isotropic tensors of odd order, and that every symmetric isotropic tensor of even order k is a scalar multiple of $\delta^{\odot k}$. This completes the proof. □

Given the tensor $T^{i_1\dots i_{k+2}}$, define the contraction $\delta\llcorner T$ by

$$(\delta\llcorner T)^{i_1\dots i_k}=\delta_{pq}T^{i_1\dots i_k pq}. \tag{105}$$

Lemma 4.5.2. *For any symmetric tensor T of order k,*

$$\delta\llcorner(\delta\odot T)=(m+2k+1)T+\delta\odot(\delta\llcorner T). \tag{106}$$

Proof. Observe that

$$(\delta\odot T)^{i_1\dots i_{k+2}}=\sum_{1\leq p<q\leq k+2}\delta^{i_pi_q}T^{i_1\dots i_{p-1}i_{p+1}\dots i_{q-1}i_{q+1}\dots i_{k+2}}.$$

Thus,

$$\begin{aligned}\delta_{i_{k+1}i_{k+2}}(\delta\odot T)^{i_1\dots i_{k+2}}=\delta_{i_{k+1}i_{k+2}}&\delta^{i_{k+1}i_{k+2}}T^{i_1\dots i_k}\\ &+\delta_{i_{k+1}i_{k+2}}\sum_{1\leq p\leq k}\delta^{i_pi_{k+2}}T^{i_1\dots i_{p-1}i_{p+1}\dots i_{k+1}}\\ &+\delta_{i_{k+1}i_{k+2}}\sum_{1\leq p\leq k}\delta^{i_pi_{k+1}}T^{i_1\dots i_{p-1}i_{p+1}\dots i_ki_{k+2}}\\ &+\sum_{1\leq p,q\leq k}\delta^{i_pi_q}(\delta_{i_{k+1}i_{k+2}}T^{i_1\dots i_{p-1}i_{p+1}\dots i_{q-1}i_{q+1}\dots i_{k+2}})\\ =[(m+1)T+2kT+\delta\odot(\delta\llcorner T)]^{i_1\dots i_k},&\end{aligned}$$

and the result follows. ☐

Lemma 4.5.3. *For all k,*

$$\delta\llcorner\delta^{\odot k}=k(m+2k-1)\delta^{\odot(k-1)}. \tag{107}$$

Proof. We proceed by induction. First observe that $\delta\llcorner\delta=(m+1)$. Next, suppose that it holds for k, then, by (106) and the inductive hypothesis,

$$\begin{aligned}\delta\llcorner\delta^{\odot(k+1)}&=\delta\llcorner(\delta\odot\delta^{\odot k})\\ &=(m+4k+1)\delta^{\odot k}+\delta\odot(\delta\llcorner\delta^{\odot k})\\ &=((m+4k+1)+k(m+2k-1))\delta^{\odot k}\\ &=(k+1)(m+2(k+1)-1)\delta^{\odot k},\end{aligned}$$

and the result follows. ☐

Lemma 4.5.4. *For all k,*

$$\begin{aligned}\int_{S^m}x^{i_1}\dots x^{i_{2k}}\,d\mathrm{Vol}&=\frac{\mathrm{Vol}(S^m)(m-1)!!}{k!(m+2k-1)!!}\delta^{\odot k},\\ \int_{S^m}x^{i_1}\dots x^{i_{2k+1}}\,d\mathrm{Vol}&=0.\end{aligned} \tag{108}$$

Proof. For all l, denote

$$M_l^{i_1\dots i_l} := \int_{S^m} x^{i_1}\dots x^{i_l}\, d\mathrm{Vol}.$$

Since M_l is symmetric and isotropic, it follows by Lemma 4.5.1 that M_l vanishes when l is odd, and when $l =: 2k$ is even $M_l = C_k\delta^{\odot k}$ for some constant C_k. It remains to show that

$$C_k = \frac{\mathrm{Vol}(S^m)(m-1)!!}{k!(m+2k-1)!!}$$

for all k. We prove this by induction on k. Indeed, $C_0 = \mathrm{Vol}(S^m)$. Now suppose that it holds for k. Since $\|x\|^2 = 1$ over S^m, for all k,

$$\begin{aligned}(\delta\llcorner M_{2(k+1)})^{i_1\dots i_{2k}} &= \delta_{i_{2k+1}i_{2k+2}}\int_{S^m} x^{i_1}\dots x^{i_{2k+2}}\, d\mathrm{Vol}\\ &= \int_{S^m} x^{i_1}\dots x^{i_{2k}}\, d\mathrm{Vol} = M_{2k}^{i_1\dots i_{2k}},\end{aligned}$$

so that, by (107) and the induction hypothesis,

$$C_{k+1} = \frac{1}{(k+1)(m+2k+1)}C_k = \frac{\mathrm{Vol}(S^m)(m-1)!!}{(k+1)!(m+2k+1)!!},$$

and the result follows. □

Proposition 4.5.5. *The functions* $(x^i)_{1\leq i\leq m+1}$ *constitute an orthogonal basis of* $\mathcal{H}_1$ *with respect to the* L^2 *inner product over* S^m.

Proof. These functions trivially constitute a basis of $\mathcal{H}_1$. Further, by Lemma 4.5.4, for all $1\leq i, j\leq m+1$,

$$\int_{S^m} x^i x^j\, d\mathrm{Vol} = \frac{\mathrm{Vol}(S^m)}{(m+1)}\delta^{ij},$$

and orthogonality follows. □

Let $\Pi : L^2(S^m)\to\mathcal{H}_1$ be the orthogonal projection.

Proposition 4.5.6.

$$\Pi\Big(\tfrac{1}{4}\,\mathrm{Ric}_{ab;c}\,x^a x^b x^c - \frac{(m+1)}{2(m+3)}S_{;a}x^a\Big) = 0, \tag{109}$$

and, for any fixed vector V,

$$\Pi\Big(\tfrac{1}{4}\,\mathrm{Ric}_{ab;cd}\,x^a x^b x^c V^d - \frac{(m+1)}{2(m+3)}S_{;ab}x^a V^b\Big) = 0. \tag{110}$$

Remark. Observe that (110) corresponds to the third summand in the asymptotic series (85) of Φ.

Proof. Indeed, bearing in mind Lemma 4.5.4, for all $1 \leq i \leq m+1$,

$$\int_{S^m}\Big(\tfrac{1}{4}\,\mathrm{Ric}_{ab;c}\,x^a x^b x^c - \frac{m+1}{2(m+3)}S_{;a}x^a\Big)x^i\,d\mathrm{Vol}$$
$$= \frac{\mathrm{Vol}(S_m)}{4(m+1)(m+3)}\,\mathrm{Ric}_{ab;c}(\delta^{ab}\delta^{ci}+\delta^{ac}\delta^{bi}+\delta^{ai}\delta^{bc}) - \frac{\mathrm{Vol}(S_m)}{2(m+3)}S_{;a}\delta^{ai}.$$

The first relation now follows by the second Bianchi identity and the second follows upon taking its formal derivative. ☐

Proposition 4.5.7.

$$\Pi\big(\tfrac{1}{3}\,\mathrm{Ric}_{ab}\,x^a x^b\big) = 0, \tag{111}$$

and, for any fixed vector V,

$$\Pi\big(\tfrac{1}{3}\,\mathrm{Ric}_{ab;c}\,x^a x^b V^c\big) = 0. \tag{112}$$

Remark. Observe that (112) corresponds to the fourth summand in the asymptotic series (85) of Φ.

Proof. The first relation follows directly from Lemma 4.5.4 and the second relation follows upon taking the formal derivative. ☐

4.6. *Formal solutions.*

Theorem 4.6.1. *There exist increasing sequences (C_k) of positive constants and (n_k) of positive integers with the property that, for all s, there exist canonical sequences $(Y_{k,s}) \in C^{1,\alpha}(\mathbb{R}, \mathbb{R}^{m+1})$ and $(f_{k,s}) \in \hat{C}^{1,\alpha}_{\mathrm{in}}(\mathbb{R}\times S^m)$ such that, for all k,*

$$f_{k,s} \in C^{1,\alpha}(\mathbb{R}, \hat{\mathcal{H}}_{n_k}),\quad \|f_{k,s}\|_{1,\alpha,\mathrm{in},s^4} \leq C_k,\quad \|Y_{k,s}\|_{1,\alpha} \leq C_k, \tag{113}$$

and, for all N,

$$\Big\|\Phi\Big(s, \sum_{k=0}^{N-1} s^k Y_{k,s}, \sum_{k=0}^{N} s^k f_{k,s}\Big)\Big\|_{0,\alpha,\mathrm{in}} \leq C_k s^{N+1}. \tag{114}$$

Proof. We prove this by induction. First define the projection $\Pi : C^{k,\alpha}_{\mathrm{in}}(\mathbb{R}\times S^m) \to \hat{C}^{k,\alpha}(\mathbb{R}, \mathcal{H}_1)$ by

$$\Pi(f)(t,x) := \sum_{i=0}^{m+1}\frac{(m+1)}{\mathrm{Vol}(S^m)}\int_{S^m} f(t,x)x^i\,d\mathrm{Vol}x^i.$$

That is, for each t, $\Pi(f)(t,\cdot)$ is the L^2-orthogonal projection of the function $f(t,\cdot)$ onto $\mathcal{H}_1$. Observe that for all l, for all $f \in C^{1,\alpha}(\mathbb{R}, \hat{\mathcal{H}}_l)$ and for all s,

$$\Pi Q_s f = 0, \tag{115}$$

so that, by Proposition 4.5.7, the terms up to order k in the asymptotic expansion of $\Pi\Phi$ only depend on the asymptotic expansions of f and Y up to order $k-2$. Finally, define $\Pi^\perp := \mathrm{Id} - \Pi$.

Fix $s>0$, and define $f_{0,s} := -H_s\Phi_0$. By Proposition 4.5.7, $\Phi_0 \in C^{0,\alpha}(\mathbb{R}, \hat{\mathcal{H}}_2)$, and since the restriction of Q_sH_s to this space equals the identity, it follows that with $f_{0,s}$ so defined, the term of order 0 in the asymptotic expansion (85) of Φ vanishes. Furthermore, by Proposition 4.4.2, there exists $C_0>0$ such that $\|f_{0,s}\|_{1,\alpha,\mathrm{in},s^4} \leq C_0$ for all s. Finally, by Propositions 4.5.6 and 4.5.7, the terms of order 0 and 1 in the asymptotic expansion of $\Pi\Phi$ both vanish.

Now suppose that we have defined $C_0, \dots, C_k$, $n_0, \dots, n_k$, $f_{0,s}, \dots, f_{k,s}$, $Y_{0,s}$, $\dots$, $Y_{k-1,s}$ such that the terms up to order k and $k+1$ in the asymptotic expansions of Φ and $\Pi\Phi$ respectively all vanish, for all s, and for all $0 \leq l \leq k$:

$$f_{l,s} \in C^{1,\alpha}(\mathbb{R}, \hat{\mathcal{H}}_{n_l}), \quad \|f_{l,s}\|_{1,\alpha,\mathrm{in},s^4} \leq C_l,$$

and for all $0 \leq l \leq k-1$,

$$\|Y_{l,s}\|_{1,\alpha} \leq C_l.$$

Define

$$\begin{aligned} Y_{k,s} := -G \circ \Pi \circ \Phi_{k+2}(f_{0,s}, \dots, f_{k,s}, s^4\dot{f}_{0,s}, \dots, s^4\dot{f}_{k-2,s}, \\ Y_{0,s}, \dots, Y_{k-1,s}, \dot{Y}_{0,s}, \dots, \dot{Y}_{k-2,s}), \end{aligned}$$

and define $f_{k+1,s} := -H_s\Pi^\perp\Psi_{k+1,s}$, where

$$\begin{aligned} \Psi_{k+1,s} = \Big(\tfrac{1}{4}\,\mathrm{Ric}_{pq;rs}\,x^px^qx^rY^s_{k-1,s} - \frac{(m+1)}{2(m+3)}S_{;pq}x^pY^q_{k-1,s}\Big) - \tfrac{1}{3}\,\mathrm{Ric}_{pq;r}\,x^px^qY^r_{k,s} \\ + \Phi_{k+1}(f_{0,s}, \dots, f_{k-1,s}, s^4\dot{f}_{0,s}, \dots, s^4\dot{f}_{k-3,s}, \\ Y_{0,s}, \dots, Y_{k-2,2}, \dot{Y}_{0,s}, \dots, \dot{Y}_{k-3,s}). \end{aligned}$$

Since Φ_{k+1} is a curvature polynomial, and since f_l takes values in $\hat{\mathcal{H}}_{n_l}$ for all $0 \leq l \leq k$, there exists $n_{k+1} \geq n_k$ such that $\Pi^\perp\Psi_{k+1,s}(t,\cdot)$ is an element of $\hat{\mathcal{H}}_{n_{k+1}}$ for all s and for all t. By hypothesis, the term of order $k+1$ in the asymptotic expansion of $\Pi\Phi$ vanishes, and so, since the restriction of Q_sH_s to $C^{0,\alpha}(\mathbb{R}, \hat{\mathcal{H}}_{n_{k+1}})$ equals the identity, with $f_{k+1,s}$ so defined, the term of order $k+1$ in the asymptotic expansion of Φ vanishes. Finally, observe that the function $\Phi_{k+1}(\cdot, \dots, \cdot)$ is bounded, and since its derivatives are uniformly bounded in t, it is uniformly Lipschitz. There therefore exists $B>0$ such that, for all s,

$$\|\Phi_{k+1,s}\|_{0,\alpha,\mathrm{in}} \leq B,$$

and by Proposition 4.4.2, there exists $C_{k+1} > C_k$ such that, for all s,

$$\|f_{k+1,s}\|_{1,\alpha,\mathrm{in},s^4} \leq C_{k+1}.$$

In like manner, since PG equals the identity, by Propositions 4.5.6 and 4.5.7, with $Y_{k,s}$ so defined, the term of order $k+2$ in the asymptotic expansion of $\Pi\Phi$ vanishes. Furthermore, upon increasing C_k if necessary, we may suppose that, for all s,

$$\|Y_{k,s}\|_{1,\alpha} \leq C_k.$$

We have therefore constructed sequences (C_k), (n_k), $(Y_{k,s})$ and $(f_{k,s})$ satisfying the conclusions of the theorem such that

$$\Phi\left(s, \sum_{k=0}^{\infty} s^k Y_{k,s}, \sum_{k=0}^{\infty} s^k f_{k,s}\right) \sim 0.$$

Observe that the partial sum of Φ up to order N only involves terms up to order $N-1$ in Y. Furthermore, the time-derivative of f only ever appears together with a factor of s^4. Thus, for all $N \geq 0$, there exists a smooth function R_N with uniformly bounded derivatives such that for all s and for all (t, x),

$$\begin{aligned}\Phi\left(s, \sum_{k=0}^{N-1} s^k Y_{k,s,t}, \sum_{k=0}^{N} s^k f_{k,s,t,x}\right)\\ = s^{N+1} R_N(s, Y_{0,s,t}, \ldots, Y_{N-1,s,t}, \dot{Y}_{0,s,t}, \ldots, \dot{Y}_{N-1,s,t},\\ f_{N,s,t,x}, \ldots, f_{N,s,t,x}, s^4 \dot{f}_{0,s,t,x}, \ldots, s^4 \dot{f}_{N,s,t,x}).\end{aligned}$$

The function R_N is bounded, and since its derivatives are uniformly bounded in t, it is uniformly Lipschitz. Thus, upon increasing C_k if necessary, it follows as before that

$$\left\|\Phi\left(s, \sum_{k=0}^{N-1} s^k Y_{k,s,t}, \sum_{k=0}^{N} s^k f_{k,s,t,x}\right)\right\|_{0,\alpha,\text{in}} \leq C_k s^{N+1}. \qquad \square$$

4.7. ***Exact solutions.*** We recall the classical inverse function theorem (see [Rudin 1976], for example).

Theorem 4.7.1 (inverse function theorem). *Let E and F be Banach spaces. Let Ω be a neighbourhood of 0 in E. Let $\Phi : \Omega \to F$ be a C^2 mapping. Suppose that there exists A, $B > 0$ such that*

$$\|D\Phi(0)^{-1}\| \leq A, \quad \|D^2\Phi(x)\| \leq B\ \forall\, x \in \Omega.$$

If $\epsilon := \|\Phi(0)\| < \frac{1}{4}A^2 B$, and if $B_{2A\epsilon}(0) \subseteq \Omega$, then there exists a unique point $x \in B_{2A\epsilon}(0)$ such that $\Phi(x) = 0$.

We now obtain existence.

Theorem 4.7.2. *For all sufficiently small s, there exist canonical functions $Y_s \in C^{1,\alpha}(\mathbb{R}, \mathbb{R}^{m+1})$ and $f_s \in C^{1,\alpha}_{\text{in}}(\mathbb{R} \times S^n)$ such that $\Phi(s, f_s, Y_s) = 0$. Furthermore,* *there exists a sequence* (C_k) *of positive numbers such that if* $(Y_{k,s})$ *and* $(f_{k,s})$ *are as in Theorem 4.6.1, then, for all* N,

$$\left\| Y_s - \sum_{k=0}^{N} s^k Y_{k,s} \right\|_{1,\alpha}, \left\| f_s - \sum_{k=0}^{N} s^k f_{k,s} \right\|_{1,\alpha,\text{in},s^4} \leq C_N s^{N+1}.$$

Remark. In particular, as observed following (84) of Φ, it follows that $e(s, Y, f)$ is, up to reparametrisation, an eternal forced mean curvature flow with forcing term $1/s$. Observe, however, that the definition (82) of $e(s, Y, f)$ differs slightly from (5) given in the introduction. It is nonetheless straightforward to see that, for sufficiently small s, the two are equivalent.

Proof. Let Π and $\Pi^\perp$ be as in the proof of Theorem 4.6.1. Define the mapping $\Psi :]0, \infty[\times C^{1,\alpha}(\mathbb{R}, \mathbb{R}^{m+1}) \times \hat{C}^{1,\alpha}_{\text{in}}(\mathbb{R} \times S^m) \to C^{0,\alpha}(\mathbb{R}, \mathbb{R}^{m+1}) \times \hat{C}^{0,\alpha}_{\text{in}}(\mathbb{R} \times S^m)$ by

$$\Psi(s, Y, f) := (s^{-2}\Pi \circ \Phi(s, Y, f), \Pi^\perp \circ \Phi(s, Y, f)).$$

Consider the asymptotic series (85) for Φ up to order 2 in s. Substituting $f_{0,s} = f$, $f_{1,s} = f_{2,s} = 0$, $Y_{0,s} = Y$ and $Y_{1,s} = 0$, yields

$$\begin{aligned} s^{-2}(\Pi \circ \Phi)(s, Y, f) &= PY + (\Pi \circ R_1)(f, s^4 \dot{f}) + s(\Pi \circ R_2)(s, f, s^4 \dot{f}, \dot{Y}), \\ (\Pi^\perp \circ \Phi)(s, Y, f) &= Q_s f - \tfrac{1}{3} \operatorname{Ric}_{pq} x^p x^q + s R_3(s, f, s^4 \dot{f}, Y, \dot{Y}), \end{aligned}$$

for functions R_1, R_2 and R_3 which are smooth at $s = 0$. Differentiating with respect to Y and f, it follows that

$$D\Psi(s, Y, f) = \begin{pmatrix} P & A(s, Y, f) \\ 0 & Q_s \end{pmatrix} + sB(s, Y, f), \tag{116}$$

where, for all $R > 0$, there exists ϵ, $C > 0$ such that if $s < \epsilon$ and if $\|Y\|_{1,\alpha} + \|f\|_{1,\alpha,\text{in}} \leq R$, then $\|A(s, Y, f)\|_{0,\alpha,\text{in}}$, $\|B(s, Y, f)\|_{0,\alpha,\text{in}} \leq C$. In particular, by (101), we may suppose that $D\Psi(s, Y, f)$ is invertible with $\|D\Psi(s, Y, f)\| \leq Cs^{-\alpha}$. Furthermore, we may likewise suppose that for all such s, Y and f, $D^2\Psi(s, Y, f) \leq C$.

Let (C_k), $(Y_{k,s})$ and $(f_{k,s})$ be as in Theorem 4.6.1. Upon reducing ϵ if necessary, we may suppose that, for all $s < \epsilon$,

$$\Phi(s, Y_0, f_0 + sf_1) \leq \frac{s^{2\alpha}}{4C^3},$$

and it follows by the inverse function theorem that for all such s, there exists a unique pair (Y, f) such that $\|Y_s\|_{1,\alpha} + \|f_s\|_{1,\alpha,\text{in},s^4} < s^\alpha / 2C^2$ and $\Phi(s, Y, f) = 0$. Now fix $N > 0$. Upon reducing ϵ further if necessary, we may suppose that for all $s < \epsilon$,

$$\Phi\left(s, \sum_{k=0}^{N-1} s^k Y_k, \sum_{k=0}^{N} s^k f_k\right) \leq Cs^{N+1} < \frac{s^{2\alpha}}{4C^3},$$

and it follows by the inverse function theorem again there for all such s, there exists a unique pair (Y', f') such that $\|Y'_s\|_{1,\alpha} + \|f'_s\|_{1,\alpha,\mathrm{in},t^4} < 2C^2 s^{N+1-\alpha} < s^\alpha/2C^2$ and $\Phi(s, Y', f') = 0$. By uniqueness, $Y' = Y$ and $f' = f$. It follows that for all $N > 0$, there exists $\epsilon, C > 0$ such that for $s < \epsilon$,

$$\left\| Y_s - \sum_{k=0}^{N_1} s^k Y_k \right\|_{1,\alpha}, \left\| f_s - \sum_{k=0}^{N} s^k f_k \right\|_{1,\alpha,\mathrm{in},s^4} \leq C s^{N+1-\alpha}.$$

The result follows. □

Appendix: Genericity

For the reader's convenience, we derive the linearisation of the gradient flow operator and sketch the prove that the space of riemannian metrics with scalar curvature of Morse–Smale type is dense in every conformal class.

Proposition A.1.1. *The linearisation of the gradient flow operator* (1) *is given by*

$$L = \frac{\partial}{\partial t} + \frac{m+1}{2(m+3)} \operatorname{Hess}(S). \tag{117}$$

Proof. Let $\phi :]-\epsilon, \epsilon[\times \mathbb{R} \to \mathbb{R}$ and define the section X of $\phi^* TM$ by

$$X(s, t) := \phi_* \partial_t + \frac{m+1}{2(m+3)} \phi^* \nabla S,$$

where $\phi^* \nabla S$ denotes $\nabla S \circ \phi$. Taking the covariant derivative in the s direction yields

$$\begin{aligned} (\phi^*\nabla)_{\partial_s} X &= (\phi^*\nabla)_{\partial_s}(\phi_*\partial_t) + \frac{m+1}{2(m+3)} (\phi^*\nabla)_{\partial_s}(\phi^*\nabla S) \\ &= (\phi^*\nabla)_{\partial_t}(\phi_*\partial_s) + \frac{m+1}{2(m+3)} \phi^*(\operatorname{Hess}(S))(\phi_*\partial_s) \\ &= L(\phi_*\partial_s). \end{aligned}$$

□

Theorem A.1.2. *Let M be a manifold. The set of complete metrics over M with scalar curvature of Morse–Smale type is generic* (*that is*, *of the second category in the sense of Baire*) *in every conformal class.*

Sketch of proof. Let g be a riemannian metric over M. For convenience, we shall suppose that M is compact and that we have already shown that the scalar curvature of g is of Morse type. Now, for a given smooth function ϕ over g, define

$$\bar{g} := e^{2\phi} g.$$

The scalar curvature of $\bar{g}$ is given by

$$\bar{S} = e^{-2\phi} S - \frac{e^{-2\phi}}{n} (2\Delta\phi + (n-2)\|\nabla\phi\|^2),$$

where S denotes the scalar curvature of g. The linearisation of the scalar curvature operator about $\phi = 0$ is thus

$$L_1\phi := -\frac{2}{n}(\Delta\phi + nS\phi).$$

Let $\gamma : \mathbb{R} \to M$ be a gradient flow line of S. For all k, let $H^k(\gamma^*TM)$ denote the Sobolev space of sections of the bundle γ^*TM whose distributional derivatives up to and including order k are square integrable over $\mathbb{R}$. Let $L_0 : H^1(\gamma^*TM) \to H^0(\gamma^*TM)$ be the operator given by (1). By a standard transversality argument [Guillemin and Pollack 1974], it suffices to show that the operator

$$L(X,\phi) := L_0X + \frac{m+1}{2(m+3)}\nabla L_1\phi - \frac{m+1}{2(m+3)}\phi\nabla S \tag{118}$$

defines a surjective map from $C^\infty(M) \times H^1(\gamma^*TM)$ into $H^0(\gamma^*TM)$. To this end, recall first that the formal dual of L_0 is given by

$$L_0^* := -\frac{\partial}{\partial t} + \frac{m+1}{2(m+3)}\,\mathrm{Hess}(S).$$

Since S is of Morse type, L_0 and L_0^* are both Fredholm and the image of L_0 is the orthogonal complement of $\mathrm{Ker}(L_0^*)$ in $L^2(\gamma^*TM)$ (see [Schwarz 1993]). It is thus sufficient to show that $\mathrm{Ker}(L_0^*)$ lies in the image of L or, equivalently, that no nontrivial element of $\mathrm{Ker}(L_0^*)$ is orthogonal to every element of $\mathrm{Im}(L)$.

Let Y be a nontrivial element of $\mathrm{Ker}(L_0^*)$. Let $\psi \in C^\infty(M)$ be any function such that

$$\int_{-\infty}^{+\infty} \langle (\nabla\psi)\circ\gamma, Y\rangle \, dt \neq 0.$$

Let Ω be an open subset of M that does not intersect γ. By the classical theory of elliptic operators, there exists $\phi \in C^\infty(M)$ such that

$$(L_1\phi - \psi)|_{M\setminus\Omega} = 0,$$

so that

$$\int_{-\infty}^{+\infty} \langle (\nabla L_1\phi)\circ\gamma, Y\rangle \, dt \neq 0.$$

It remains the address the third term in (118). To this end, denote $X := \dot\gamma$. Since γ is a gradient flow of S, we have

$$X = -\frac{m+1}{2(m+3)}(\nabla S)\circ\gamma.$$

Furthermore, upon differentiating the gradient flow equation, we see that $L_0X = 0$. Next, since S is of Morse type (see [Schwarz 1993]), there exists $C > 0$ such that, for all $t \in \mathbb{R}$,

$$\|X(t)\|, \|Y(t)\| \leq Ce^{-|t|/C}. \tag{119}$$

Now choose $B > 0$ and let $g \in C_0^\infty(\mathbb{R})$ be such that

$$(\dot{g} + \phi \circ \gamma)|_{[-B,B]} = 0 \quad \text{and} \quad \|\dot{g}\|_{L^\infty} \leq \|\phi\|_{L^\infty}.$$

Since

$$L_0(gX) = \dot{g}X,$$

by (119), we have

$$\left|\int_{-\infty}^{\infty}\left\langle L_0(gX) - \frac{m+1}{2(m+3)}(\phi\nabla S)\circ\gamma, Y\right\rangle dt\right| \leq 4C^2\|\phi\|_{L^\infty}e^{-B/C}.$$

Since this may be made arbitrarily small by choosing B sufficiently large, it follows that there exists $g \in C_0^\infty(\mathbb{R})$ such that

$$\int_{-\infty}^{+\infty}\langle L(gX, \phi), Y\rangle dt \neq 0. \qquad \square$$

Acknowledgements

The results of this paper were established when the author was benefitting from a Marie Curie postdoctoral fellowship in the Centre de Recerca Matemàtica, Barcelona. The author is grateful to Andrew Clarke for helpful conversations. The author is also grateful to the anonymous referee for helpful and encouraging comments.

References

[Alikakos and Freire 2003] N. D. Alikakos and A. Freire, "The normalized mean curvature flow for a small bubble in a Riemannian manifold", *J. Differential Geom.* **64**:2 (2003), 247–303. MR

[Brendle and Eichmair 2014] S. Brendle and M. Eichmair, "Large outlying stable constant mean curvature spheres in initial data sets", *Invent. Math.* **197**:3 (2014), 663–682. MR Zbl

[Chavel 2006] I. Chavel, *Riemannian geometry*, 2nd ed., Cambridge Studies in Advanced Mathematics **98**, Cambridge University Press, 2006. MR Zbl

[Eichmair and Metzger 2012] M. Eichmair and J. Metzger, "On large volume preserving stable CMC surfaces in initial data sets", *J. Differential Geom.* **91**:1 (2012), 81–102. MR Zbl

[Eichmair and Metzger 2013a] M. Eichmair and J. Metzger, "Large isoperimetric surfaces in initial data sets", *J. Differential Geom.* **94**:1 (2013), 159–186. MR Zbl

[Eichmair and Metzger 2013b] M. Eichmair and J. Metzger, "Unique isoperimetric foliations of asymptotically flat manifolds in all dimensions", *Invent. Math.* **194**:3 (2013), 591–630. MR Zbl

[Guillemin and Pollack 1974] V. Guillemin and A. Pollack, *Differential topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974. MR Zbl

[Krylov 2008] N. V. Krylov, *Lectures on elliptic and parabolic equations in Sobolev spaces*, Graduate Studies in Mathematics **96**, American Mathematical Society, Providence, RI, 2008. MR Zbl

[Mantegazza 2011] C. Mantegazza, *Lecture notes on mean curvature flow*, Progress in Mathematics **290**, Springer, Basel, Switzerland, 2011. MR Zbl

[Martín et al. 2015] F. Martín, A. Savas-Halilaj, and K. Smoczyk, "On the topology of translating solitons of the mean curvature flow", *Calc. Var. Partial Differential Equations* **54**:3 (2015), 2853–2882. MR Zbl

[Maximo et al. 2017] D. Maximo, I. Nunes, and G. Smith, "Free boundary minimal annuli in convex three-manifolds", *J. Differential Geom.* **106**:1 (2017), 139–186. MR Zbl

[Nardulli 2009] S. Nardulli, "The isoperimetric profile of a smooth Riemannian manifold for small volumes", *Ann. Global Anal. Geom.* **36**:2 (2009), 111–131. MR Zbl

[Nardulli 2014] S. Nardulli, "The isoperimetric profile of a noncompact Riemannian manifold for small volumes", *Calc. Var. Partial Differential Equations* **49**:1-2 (2014), 173–195. MR Zbl

[Robbin and Salamon 1995] J. Robbin and D. Salamon, "The spectral flow and the Maslov index", *Bull. London Math. Soc.* **27**:1 (1995), 1–33. MR Zbl

[Rosenberg and Smith 2010] H. Rosenberg and G. Smith, "Degree theory of immersed hypersurfaces", 2010. arXiv

[Rudin 1976] W. Rudin, *Principles of mathematical analysis*, 3rd ed., McGraw-Hill, New York, 1976. MR Zbl

[Schwarz 1993] M. Schwarz, *Morse homology*, Progress in Mathematics **111**, Birkhäuser Verlag, Basel, Switzerland, 1993. MR Zbl

[Smith 2011] G. Smith, "Constant curvature hyperspheres and the Euler Characteristic", 2011. arXiv

[Smith 2015] G. Smith, "Eternal forced mean curvature flows I: a compactness result", *Geom. Dedicata* **176** (2015), 11–29. MR Zbl

[White 1991] B. White, "The space of minimal submanifolds for varying Riemannian metrics", *Indiana Univ. Math. J.* **40**:1 (1991), 161–200. MR Zbl

[Ye 1991] R. Ye, "Foliation by constant mean curvature spheres", *Pacific J. Math.* **147**:2 (1991), 381–396. MR Zbl

Received July 30, 2016. Revised January 8, 2018.

GRAHAM SMITH
INSTITUTO DE MATEMÁTICA
UFRJ
AV. ATHOS DA SILVEIRA RAMOS 149
CENTRO DE TECNOLOGIA - BLOCO C
CIDADE UNIVERSITÁRIA ILHA DO FUNDÃO
68530, 21941-909 RIO DE JANEIRO-
BRAZIL

grahamandrewsmith@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 1, 2019
dx.doi.org/10.2140/pjm.2019.299.237

SYMMETRY AND NONEXISTENCE OF SOLUTIONS FOR A FULLY NONLINEAR NONLOCAL SYSTEM

BIRAN ZHANG AND ZHONGXUE LÜ

We study the system involving fully nonlinear nonlocal operators:

$$\begin{aligned} F_\alpha(u(x)) &= C_{n,\alpha}\,\mathrm{PV}\int_{\mathbb{R}^n}\frac{G(u(x)-u(y))}{|x-y|^{n+\alpha}}\,dy = f(u(x),v(x)),\\ F_\beta(v(x)) &= C_{n,\beta}\,\mathrm{PV}\int_{\mathbb{R}^n}\frac{G(v(x)-v(y))}{|x-y|^{n+\beta}}\,dy = g(u(x),v(x)). \end{aligned}$$

We will prove the symmetry and monotonicity for positive solutions to the nonlinear system in whole space by using the method of moving planes. To achieve it, a narrow region principle and a decay at infinity are established. Further more, nonexistence of positive solutions to the nonlinear system on a half space is derived. In addition, the symmetry and monotonicity in whole space for positive solutions to a fully nonlinear nonlocal system

$$F_\alpha(u(x)) = -u^p(x)+v^q(x), \quad F_\beta(v(x)) = -v^p(x)+u^q(x)$$

can be derived.

1. Introduction

We are interested in the general nonlinear system involving fully nonlinear nonlocal operators:

$$F_\alpha(u(x)) = f(u(x),v(x)), \quad F_\beta(v(x)) = g(u(x),v(x))$$

with

$$F_\alpha(u(x)) = C_{n,\alpha}\,\mathrm{PV}\int_{\mathbb{R}^n}\frac{G(u(x)-u(y))}{|x-y|^{n+\alpha}}\,dy,$$

where the PV stands for the Cauchy principal value, G is a nonlinear operator and is at least local Lipschitz continuous with $G(0)=0$ and $0<\alpha,\beta<2$. The operator F_α was introduced by Caffarelli and Silvestre [2009].

The work was partially supported by NSFC(No.11271166), sponsored by Qing Lan Project.
MSC2010: 35B06, 35B09, 35B50, 35B53.
Keywords: fully nonlinear nonlocal operator, narrow region principle, decay at infinity, method of moving planes.

In order to make the integral significative, we require

$$u(x) \in C^{1,1}_{\mathrm{loc}} \cap L_\alpha \quad \text{and} \quad v(x) \in C^{1,1}_{\mathrm{loc}} \cap L_\beta$$

with

$$L_\alpha = \left\{ u : \mathbb{R}^n \to \mathbb{R} \;\middle|\; \int_{\mathbb{R}^n} \frac{|u(x)|}{1 + |x|^{n+\alpha}} \, dx < \infty \right\},$$

and L_β is defined similarly.

The special case is that when $G(\cdot)$ is an identity map, F_α becomes the usual fractional Laplacian $(-\Delta)^{\alpha/2}$. It is the nonlocal nature of fractional operators that makes them difficult to study. To circumvent this, Caffarelli and Silvestre [2007] introduced the *extension method*, which turns the nonlocal problem involving the fractional Laplacian into a local one in higher dimensions. A series of fruitful results show that this method has been applied successfully to treat equations involving the fractional Laplacian (see [Brändle et al. 2013; Chen and Zhu 2016; Gilbarg and Trudinger 1977]). Another way is using *the integral equations method*, such as *the method of moving planes in integral forms* (see [Cao and Chen 2013; Cao and Dai 2013; Li and Zhuo 2010; Lu and Zhu 2012; Zhuo et al. 2016]) and *regularity lifting* to investigate equations involving fractional Laplacian by showing that they are equivalent to corresponding integral equations (see [Chen et al. 2005; 2006; 2015]). For more articles concerning the method of moving planes for nonlocal equations and for integral equations, see [Gilbarg and Trudinger 1977; Hang et al. 2009; 2012; Hang 2007; Lei et al. 2012; Li 2017; Li and Ma 2017; Lu and Zhu 2011; 2012; Ma and Chen 2006; Ma and Zhao 2010; Wang and Niu 2017].

Chen, Li, and Li [Chen et al. 2017b] developed a systematic approach to carry out the method of moving planes for equations involving fractional Laplacian directly. Subsequently, by using this direct method, many authors investigated different equations involving fractional Laplace; see, for example, [Cheng et al. 2017a; 2017b; Cheng 2017; Li and Ma 2017; Liu and Ma 2016; Zhang et al. 2017].

However, for the fully nonlinear nonlocal equations, so far as we know, there is neither any corresponding *extension method* nor equivalent integral equations that one can begin to work. Chen, Li, and Li [Chen et al. 2017a], developed a new method that can deal with the fully nonlinear nonlocal equations directly. Then with the help of the direct method of moving planes, Wang and Yu [2017] studied a fully nonlinear nonlocal system where $u(x)$ and $v(x)$ belong to different nonhomogeneous terms. Wang and Niu [2017] studied a fully nonlinear nonlocal system with special nonhomogeneous terms which have $u(x)$ and $v(x)$ simultaneously while $u(x)$ and $v(x)$ have positive coefficients.

In this paper, we extend the direct method in [Chen et al. 2017a] to more general fully nonlinear nonlocal systems:

$$\begin{aligned} F_\alpha(u(x)) &= f(u(x), v(x)), \quad x \in \mathbb{R}^n, \\ F_\beta(v(x)) &= g(u(x), v(x)), \quad x \in \mathbb{R}^n, \\ u(x) > 0, \quad v(x) &> 0, \quad x \in \mathbb{R}^n, \end{aligned} \tag{1-1}$$

and

$$\begin{aligned} F_\alpha(u(x)) &= f(u(x), v(x)), \quad x \in \mathbb{R}^n_+, \\ F_\beta(v(x)) &= g(u(x), v(x)), \quad x \in \mathbb{R}^n_+, \\ u(x) \equiv 0, \quad v(x) &\equiv 0, \quad x \notin \mathbb{R}^n_+, \end{aligned} \tag{1-2}$$

where f, g are continuous functions. It is worth mentioning that (1-1) is more general than the system in [Wang and Yu 2017] and is different from the system in [Wang and Niu 2017]. Because (1-1) can be allowed, $u(x)$ or $v(x)$ have negative coefficients.

We first establish the *narrow region principle* and *decay at infinity* for the systems and they will play important roles in carrying out the method of moving planes.

To state them, let

$$T_\lambda = \{x \in \mathbb{R}^n \mid x_1 = \lambda, \lambda \in \mathbb{R}\}$$

be the moving plane, and denote by

$$\Sigma_\lambda = \{x \in \mathbb{R}^n \mid x_1 < \lambda\}$$

the left region of the plane T_λ, by

$$x^\lambda = (2\lambda - x_1, x_2, \dots, x_n)$$

the reflection of x about T_λ, and let

$$U_\lambda(x) = u_\lambda(x) - u(x), \quad V_\lambda(x) = v_\lambda(x) - v(x)$$

with

$$u_\lambda(x) = u(x^\lambda), \quad v_\lambda(x) = v(x^\lambda).$$

For simplicity of notation, in the following, we denote $U_\lambda(x)$ by $U(x)$ and $V_\lambda(x)$ by $V(x)$. Throughout this paper, we assume that

$$G \in C^1(\mathbb{R}), \quad G(0) = 0, \quad \text{and} \quad G'(t) \geq c_0 > 0 \quad \forall\, t \in \mathbb{R}. \tag{1-3}$$

Theorem 1.1 (narrow region principle). *Let $\Omega \subset \Sigma_\lambda$ be a bounded narrow region contained in the strip*

$$\{x \mid \lambda - l < x_1 < \lambda\}$$

with small $l > 0$. Suppose that $U(x) \in L_\alpha \cap C^{1,1}_{\mathrm{loc}}(\Omega)$, $V(x) \in L_\beta \cap C^{1,1}_{\mathrm{loc}}(\Omega)$, $U(x)$ and $V(x)$ are both lower semicontinuous on $\overline{\Omega}$, and satisfy

$$\begin{aligned} &F_\alpha(u_\lambda(x)) - F_\alpha(u(x)) + c_{11}(x)U(x) + c_{12}(x)V(x) \geq 0, && x \in \Omega, \\ &F_\beta(v_\lambda(x)) - F_\beta(v(x)) + c_{21}(x)U(x) + c_{22}(x)V(x) \geq 0, && x \in \Omega, \\ &\qquad\qquad U(x),\, V(x) \geq 0, && x \in \Sigma_\lambda \setminus \Omega, \\ &U(x^\lambda) = -U(x), \quad V(x^\lambda) = -V(x), && x \in \Sigma_\lambda, \end{aligned} \tag{1-4}$$

if $c_{12}(x) < 0$, $c_{21}(x) < 0$, and $c_{ij}(x)$, $i, j = 1, 2$, are bounded from below in Ω, then for sufficiently small l, we have

$$U(x) \geq 0, \quad V(x) \geq 0 \quad in \quad \Omega; \tag{1-5}$$

furthermore, if $U(x)$ or *$V(x)$ equals 0 at some point in Ω, then*

$$U(x) = V(x) \equiv 0, \quad x \in \mathbb{R}^n. \tag{1-6}$$

These conclusions hold for the unbounded region Ω if we further assume that

$$\varliminf_{|x|\to\infty} U(x) \geq 0, \quad \varliminf_{|x|\to\infty} V(x) \geq 0. \tag{1-7}$$

Theorem 1.2 (decay at infinity). *Let Ω be an unbounded region in Σ_λ. Assume that $U(x) \in C^{1,1}_{\mathrm{loc}}(\Omega) \cap L_\alpha(\mathbb{R}^n)$, $V(x) \in C^{1,1}_{\mathrm{loc}}(\Omega) \cap L_\beta(\mathbb{R}^n)$ are both lower semicontinuous and satisfy*

$$\begin{aligned} &F_\alpha(u_\lambda(x)) - F_\alpha(u(x)) + c_{11}(x)U(x) + c_{12}(x)V(x) \geq 0, && x \in \Omega, \\ &F_\beta(v_\lambda(x)) - F_\beta(v(x)) + c_{21}(x)U(x) + c_{22}(x)V(x) \geq 0, && x \in \Omega, \\ &\qquad\qquad U(x),\, V(x) \geq 0, && x \in \Sigma_\lambda \setminus \Omega, \\ &U(x^\lambda) = -U(x), \quad V(x^\lambda) = -V(x), && x \in \Sigma_\lambda, \end{aligned} \tag{1-8}$$

with

$$c_{11}(x), c_{12}(x) \sim o\left(\frac{1}{|x|^\alpha}\right), \quad c_{21}(x), c_{22}(x) \sim o\left(\frac{1}{|x|^\beta}\right) \quad for\ |x|\ large \tag{1-9}$$

and $c_{12}(x), c_{21}(x) < 0$, then there exists a constant $R_0 > 0$ depending only on $c_{ij}(x)$ such that if

$$U(\tilde{x}) = \min_\Omega U(x) < 0 \quad and \quad V(\bar{x}) = \min_\Omega V(x) < 0,$$

then

$$|\tilde{x}| \leq R_0 \quad or \quad |\bar{x}| \leq R_0. \tag{1-10}$$

Based on Theorems 1.1 and 1.2, we apply the *method of moving planes* to obtain symmetry and monotonicity of positive solutions to (1-1) in $\mathbb{R}^n$, as well as nonexistence of positive solutions to (1-2) on the half space $\mathbb{R}^n_+$.

Theorem 1.3. *Assume that* $u(x) \in L_\alpha \cap C^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$, $v(x) \in L_\beta \cap C^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$ *are positive solutions of system* (1-1). *Suppose that for some* $\gamma_1, \gamma_2 > 0$,

$$u(x) = o\left(\frac{1}{|x|^{\gamma_1}}\right), \quad v(x) = o\left(\frac{1}{|x|^{\gamma_2}}\right) \quad \textit{as } |x| \to \infty \tag{1-11}$$

and f, g *are continuous functions satisfying*

$$\begin{aligned} &\text{(i)} \quad \textit{for fixed } t: \quad -C_1 s^p \le f_1'(s,t) < 0, \quad 0 < g_1'(s,t) \le C_2 s^q; \\ &\text{(ii)} \quad \textit{for fixed } s: \quad 0 < f_2'(s,t) \le C_3 t^p, \quad -C_4 t^q \le g_2'(s,t) < 0; \end{aligned} \tag{1-12}$$

with $\min\{p\gamma_1, p\gamma_2\} \ge \alpha$, $\min\{q\gamma_1, q\gamma_2\} \ge \beta$, *and* $C_i > 0$, $i = 1, 2, 3, 4$.

Then $u(x)$ *and* $v(x)$ *must be radially symmetric and monotone decreasing about some point in* $\mathbb{R}^n$.

Theorem 1.4. *Assume that* $u(x) \in L_\alpha \cap C^{1,1}_{\mathrm{loc}}(\mathbb{R}^n_+)$, $v(x) \in L_\beta \cap C^{1,1}_{\mathrm{loc}}(\mathbb{R}^n_+)$ *are nonnegative solutions to system* (1-2) *where* f, g *are nonnegative continuous functions and* u, v *are lower semicontinuous on* $\overline{\mathbb{R}^n_+}$. *Suppose*

$$\lim_{|x|\to\infty} u(x) = 0, \quad \lim_{|x|\to\infty} v(x) = 0, \tag{1-13}$$

then $u(x) \equiv 0$, $v(x) \equiv 0$.

In Section 2, we prove Theorems 1.1 and 1.2 with a key inequality (2-2) below. Sections 3 and 4 are devoted to the proofs of Theorems 1.3 and 1.4, respectively, by using the previous results and the method of moving planes. In Section 5, we will consider the fully nonlinear nonlocal system

$$\begin{aligned} F_\alpha(u(x)) &= -u^p(x) + v^q(x), & x &\in \mathbb{R}^n, \\ F_\beta(v(x)) &= -v^p(x) + u^q(x), & x &\in \mathbb{R}^n, \\ u(x), v(x) &> 0, & x &\in \mathbb{R}^n, \end{aligned}$$

where $p, q > 0$. And it is a specific case of (1-1).

2. Proofs of Theorems 1.1 and 1.2

Let

$$F_\alpha(u(x)) = C_{n,\alpha}\, \mathrm{PV} \int_{\mathbb{R}^n} \frac{G(u(x) - u(y))}{|x-y|^{n+\alpha}}\, dy \\ = C_{n,\alpha} \lim_{\epsilon \to 0} \int_{\mathbb{R}^n \setminus B_\epsilon(x)} \frac{G(u(x) - u(y))}{|x-y|^{n+\alpha}}\, dy,$$

and use c and C for general various positive constants that are usually different in different contexts.

We first introduce a lemma which is often called the strong maximum principle to F_α.

Lemma 2.1. *Let Ω be a bounded domain in $\mathbb{R}^n$. Assume that $u(x) \in C^{1,1}_{\text{loc}} \cap L_\alpha(\mathbb{R}^n)$ is lower semicontinuous on $\overline{\Omega}$ and satisfies*

$$\begin{aligned} F_\alpha(u(x)) &\geq 0, \quad x \in \Omega, \\ u(x) &\geq 0, \quad x \in \Omega^c. \end{aligned} \tag{2-1}$$

If $u(x)$ attains 0 *somewhere in Ω, then*

$$u(x) \equiv 0, \quad x \in \mathbb{R}^n.$$

The proof of this lemma was completed in [Wang and Yu 2017], we omit the details here.

Proof of Theorem 1.1. If (1-5) does not hold, without loss of generality, we assume $U(x) < 0$ at some point in Ω. By the lower semicontinuity of $U(x)$ on $\overline{\Omega}$, we know that there exists some $\tilde{x} \in \Omega$ such that

$$U(\tilde{x}) = \min_{\Omega} U(x) < 0.$$

It follows from (1-4) that $\tilde{x}$ must be in the interior of Ω. Then we have

$$\begin{aligned} &F_\alpha(u_\lambda(\tilde{x})) - F_\alpha(u(\tilde{x})) \\ &\quad = C_{n,\alpha}\,\mathrm{PV}\int_{\mathbb{R}^n} \frac{G(u_\lambda(\tilde{x}) - u_\lambda(y)) - G(u(\tilde{x}) - u(y))}{|\tilde{x} - y|^{n+\alpha}}\,dy \\ &\quad = C_{n,\alpha}\,\mathrm{PV}\int_{\Sigma_\lambda} \frac{G(u_\lambda(\tilde{x}) - u_\lambda(y)) - G(u(\tilde{x}) - u(y))}{|\tilde{x} - y|^{n+\alpha}}\,dy \\ &\qquad\quad + C_{n,\alpha}\,\mathrm{PV}\int_{\Sigma_\lambda} \frac{G(u_\lambda(\tilde{x}) - u(y)) - G(u(\tilde{x}) - u_\lambda(y))}{|\tilde{x} - y^\lambda|^{n+\alpha}}\,dy \\ &\quad \leq C_{n,\alpha}\,\mathrm{PV}\int_{\Sigma_\lambda} \frac{G(u_\lambda(\tilde{x}) - u_\lambda(y)) - G(u(\tilde{x}) - u(y))}{|\tilde{x} - y^\lambda|^{n+\alpha}}\,dy \\ &\qquad\quad + C_{n,\alpha}\,\mathrm{PV}\int_{\Sigma_\lambda} \frac{G(u_\lambda(\tilde{x}) - u(y)) - G(u(\tilde{x}) - u_\lambda(y))}{|\tilde{x} - y^\lambda|^{n+\alpha}}\,dy \\ &\quad = C_{n,\alpha}\,\mathrm{PV}\int_{\Sigma_\lambda} \frac{2G'(\cdot)U(\tilde{x})}{|\tilde{x} - y^\lambda|^{n+\alpha}}\,dy \\ &\quad \leq 2C_{n,\alpha}U(\tilde{x})\int_{\Sigma_\lambda} \frac{1}{|\tilde{x} - y^\lambda|^{n+\alpha}}\,dy. \end{aligned} \tag{2-2}$$

Let $D = \{y \mid l < y_1 - \tilde{x}_1 < 1, |y' - \tilde{x}'| < 1\}$, $s = y_1 - \tilde{x}_1$, $\tau = |y' - \tilde{x}'|$, and ω_{n-2} be the area of an $(n-2)$-dimensional unit sphere. Here we write $x = (x_1, x')$. Then we have

$$\begin{aligned}
\text{(2-3)}\quad \int_{\Sigma_\lambda} \frac{1}{|\tilde{x} - y^\lambda|^{n+\alpha}}\,dy &\geq \int_D \frac{1}{|\tilde{x} - y|^{n+\alpha}}\,dy = \int_l^1 \int_0^1 \frac{\omega_{n-2}\tau^{n-2}}{(s^2+\tau^2)^{(n+\alpha)/2}}\,d\tau\,ds \\
&= \int_l^1 \int_0^{1/s} \frac{\omega_{n-2}(st)^{n-2}s}{s^{n+\alpha}(1+t^2)^{(n+\alpha)/2}}\,dt\,ds \\
&= \int_l^1 \frac{1}{s^{1+\alpha}} \int_0^{1/s} \frac{\omega_{n-2}t^{n-2}}{(1+t^2)^{(n+\alpha)/2}}\,dt\,ds \\
&\geq \int_l^1 \frac{1}{s^{1+\alpha}} \int_0^{1} \frac{\omega_{n-2}t^{n-2}}{(1+t^2)^{(n+\alpha)/2}}\,dt\,ds \\
&\geq C\int_l^1 \frac{1}{s^{1+\alpha}}\,ds \geq \frac{C}{l^\alpha}.
\end{aligned}$$

Thus from (2-2) and the fact that $c_{11}(x)$ is bounded from below in Ω,

$$\text{(2-4)}\qquad F_\alpha(u_\lambda(\tilde{x})) - F_\alpha(u(\tilde{x})) + c_{11}(\tilde{x})u(\tilde{x}) \leq \frac{C}{l^\alpha}U(\tilde{x}) < 0.$$

Together (2-4) with (1-4), we have

$$\text{(2-5)}\qquad U(\tilde{x}) \geq -cl^\alpha c_{12}(\tilde{x})V(\tilde{x}).$$

From (2-5) and the lower semicontinuity of $V(x)$ on $\overline{\Omega}$, we know that there exists $\bar{x}$ in Ω such that

$$V(\bar{x}) = \min_\Omega V(x) < 0.$$

Similar to (2-4), it is easy to see that

$$F_\beta(v_\lambda(\bar{x})) - F_\beta(v(\bar{x})) + c_{22}(\bar{x})V(\bar{x}) \leq \frac{C}{l^\beta}V(\bar{x}) < 0.$$

Together with (2-5), for l sufficiently small, we have

$$\begin{aligned}
0 &\leq F_\beta(v_\lambda(\bar{x})) - F_\beta(v(\bar{x})) + c_{21}(\bar{x})U(\bar{x}) + c_{22}(\bar{x})V(\bar{x}) \\
&\leq \frac{C}{l^\beta}V(\bar{x}) + c_{21}(\bar{x})U(\tilde{x}) \\
&\leq \frac{C}{l^\beta}V(\bar{x}) - cc_{21}(\bar{x})l^\alpha c_{12}(\tilde{x})V(\tilde{x}) \\
&\leq \frac{C}{l^\beta}V(\bar{x}) - cc_{21}(\bar{x})l^\alpha c_{12}(\tilde{x})V(\bar{x}) \\
&\leq \frac{C}{l^\beta}V(\bar{x})(1 - c_{21}(\bar{x})c_{12}(\tilde{x})l^{\alpha+\beta}) < 0.
\end{aligned}$$

This contradiction shows that (1-5) must be true. If Ω is unbounded, then (1-5) is easily obtained by using (1-7).

To prove (1-6), without loss of generality, we suppose that there exists $\eta \in \Omega$ such that

$$U(\eta) = 0.$$

Combining the fact

$$\frac{1}{|x-y|} > \frac{1}{|x-y^\lambda|} \quad \forall\, x, y \in \Sigma_\lambda,$$

we have

$$\begin{aligned}
&F_\alpha(u_\lambda(\eta)) - F_\alpha(u(\eta)) \\
&\quad = C_{n,\alpha}\,\mathrm{PV}\!\int_{\mathbb{R}^n} \frac{G(u_\lambda(\eta)-u_\lambda(y)) - G(u(\eta)-u(y))}{|\eta-y|^{n+\alpha}}\,dy \\
&\quad = C_{n,\alpha}\,\mathrm{PV}\!\int_{\Sigma_\lambda} \frac{G(u_\lambda(\eta)-u_\lambda(y)) - G(u(\eta)-u(y))}{|\eta-y|^{n+\alpha}}\,dy \\
&\qquad\qquad + C_{n,\alpha}\,\mathrm{PV}\!\int_{\Sigma_\lambda} \frac{G(u_\lambda(\eta)-u(y)) - G(u(\eta)-u_\lambda(y))}{|\eta-y^\lambda|^{n+\alpha}}\,dy \\
&\quad = C_{n,\alpha}\,\mathrm{PV}\!\int_{\Sigma_\lambda} [G(u_\lambda(\eta)-u_\lambda(y)) - G(u(\eta)-u(y))]\left(\frac{1}{|\eta-y|^{n+\alpha}} - \frac{1}{|\eta-y^\lambda|^{n+\alpha}}\right)dy \\
&\qquad\qquad + C_{n,\alpha}\,\mathrm{PV}\!\int_{\Sigma_\lambda} \frac{1}{|\eta-y^\lambda|^{n+\alpha}}\big(G(u_\lambda(\eta)-u(y)) - G(u(\eta)-u_\lambda(y)) \\
&\qquad\qquad\qquad\qquad + G(u_\lambda(\eta)-u_\lambda(y)) - G(u(\eta)-u(y))\big)\,dy \\
&\quad = C_{n,\alpha}\,\mathrm{PV}\!\int_{\Sigma_\lambda} [G(u_\lambda(\eta)-u_\lambda(y)) - G(u(\eta)-u(y))]\left(\frac{1}{|\eta-y|^{n+\alpha}} - \frac{1}{|\eta-y^\lambda|^{n+\alpha}}\right)dy \\
&\qquad\qquad + C_{n,\alpha}\,\mathrm{PV}\!\int_{\Sigma_\lambda} \frac{1}{|\eta-y^\lambda|^{n+\alpha}}\big(G(u_\lambda(\eta)-u(y)) - G(u(\eta)-u(y)) \\
&\qquad\qquad\qquad\qquad + G(u_\lambda(\eta)-u_\lambda(y)) - G(u(\eta)-u_\lambda(y))\big)\,dy \\
&\quad = C_{n,\alpha} G'(\cdot)\int_{\Sigma_\lambda} (U(\eta)-U(y))\left(\frac{1}{|\eta-y|^{n+\alpha}} - \frac{1}{|\eta-y^\lambda|^{n+\alpha}}\right)dy \\
&\qquad\qquad + C_{n,\alpha}\int_{\Sigma_\lambda} \frac{G'(\cdot)U(\eta) + G'(\cdot)U(\eta)}{|\eta-y^\lambda|^{n+\alpha}}\,dy \\
&\quad \le -C_{n,\alpha} c_0 \int_{\Sigma_\lambda} U(y)\left(\frac{1}{|\eta-y|^{n+\alpha}} - \frac{1}{|\eta-y^\lambda|^{n+\alpha}}\right)dy.
\end{aligned}$$

That is,

$$F_\alpha(u_\lambda(\eta)) - F_\alpha(u(\eta)) + c_{11}(\eta)U(\eta) \le -C_{n,\alpha} c \int_{\Sigma_\lambda} U(y)\left(\frac{1}{|\eta-y|^{n+\alpha}} - \frac{1}{|\eta-y^\lambda|^{n+\alpha}}\right)dy. \tag{2-6}$$

If $U(x) \not\equiv 0$, $x \in \Sigma_\lambda$, then (2-6) implies

$$F_\alpha(u_\lambda(\eta)) - F_\alpha(u(\eta)) + c_{11}(\eta)U(\eta) < 0.$$

Together with (1-4), it is easy to see that $V(\eta) < 0$. This contradicts with (1-5). Hence $U(x)$ must be identically 0 in Σ_λ. Since

$$U(x^\lambda) = -U(x), \quad x \in \Sigma_\lambda,$$

it gives

$$U(x) \equiv 0, \quad x \in \mathbb{R}^n.$$

Together with the first equation in (1-4), we see

$$V(x) \le 0, \quad x \in \Sigma_\lambda.$$

Noting we already have

$$V(x) \ge 0, \quad x \in \Sigma_\lambda,$$

it must hold

$$V(x) = 0, \quad x \in \Sigma_\lambda.$$

Recalling $V(x^\lambda) = -V(x)$, we deduce

$$V(x) \equiv 0, \quad x \in \mathbb{R}^n.$$

Similarly, one can show that if $V(x)$ attains 0 at some point in Ω, then both $U(x)$ and $V(x)$ are identically 0 in $\mathbb{R}^n$. This completes the proof. □

Proof of Theorem 1.2. Assume that there exists $\tilde{x} \in \Omega$ such that

$$U(\tilde{x}) = \min_{\Omega} U(x) < 0.$$

Using the key inequality (2-2), we have

$$F_\alpha(u_\lambda(\tilde{x})) - F_\alpha(u(\tilde{x})) \le 2C_{n,\alpha}c_0 U(\tilde{x}) \int_{\Sigma_\lambda} \frac{1}{|\tilde{x} - y^\lambda|^{n+\alpha}}\, dy.$$

For each fixed $\lambda \in \mathbb{R}$, there exists $C > 0$ such that for $\tilde{x} \in \Sigma_\lambda$ and $|\tilde{x}|$ sufficiently large,

$$\int_{\Sigma_\lambda} \frac{1}{|\tilde{x} - y^\lambda|^{n+\alpha}}\, dy \ge \int_{B_{3|\tilde{x}|}(\tilde{x}) \setminus B_{2|\tilde{x}|}(\tilde{x})} \frac{1}{|\tilde{x} - y|}\, dy \sim \frac{C}{|\tilde{x}|^\alpha}. \tag{2-7}$$

Hence, from (2-7) and (1-9), we have

$$F_\alpha(u_\lambda(\tilde{x})) - F_\alpha(u(\tilde{x})) + c_{11}(\tilde{x})U(\tilde{x}) \le \frac{C}{|\tilde{x}|^\alpha} U(\tilde{x}) < 0. \tag{2-8}$$

Together (2-8) with (1-8), it is easy to know

$$V(\tilde{x}) < 0, \tag{2-9}$$

and

$$U(\tilde{x}) \geq -cc_{12}(\tilde{x})|\tilde{x}|^{\alpha} V(\tilde{x}). \tag{2-10}$$

From (2-9) and the lower semicontinuity of $V(x)$ on $\overline{\Omega}$, there exists $\bar{x}$ such that

$$V(\bar{x}) = \min_{\Omega} V(x) < 0.$$

Similarly to (2-8), we can derive

$$F_{\beta}(v_{\lambda}(\bar{x})) - F_{\beta}(v(\bar{x})) + c_{22}(\bar{x})V(\bar{x}) \leq \frac{C}{|\bar{x}|^{\beta}} V(\bar{x}) < 0. \tag{2-11}$$

Combining (1-8), (1-10), and (2-11), for λ sufficiently negative, it follows that

$$\begin{aligned} 0 &\leq F_{\beta}(v_{\lambda}(\bar{x})) - F_{\beta}(v(\bar{x})) + c_{21}(\bar{x})U(\bar{x}) + c_{22}(\bar{x})V(\bar{x}) \\ &\leq \frac{C}{|\bar{x}|^{\beta}} V(\bar{x}) + c_{21}(\bar{x})U(\tilde{x}) \\ &\leq \frac{C}{|\bar{x}|^{\beta}} V(\bar{x}) - cc_{21}(\bar{x})c_{12}(\tilde{x})|\tilde{x}|^{\alpha} V(\tilde{x}) \\ &\leq \frac{C}{|\bar{x}|^{\beta}} V(\bar{x}) - cc_{21}(\bar{x})c_{12}(\tilde{x})|\tilde{x}|^{\alpha} V(\bar{x}) \\ &\leq \frac{C}{l^{\beta}} V(\bar{x})(1 - c_{12}(\tilde{x})|\tilde{x}|^{\alpha} c_{21}(\bar{x})|\bar{x}|^{\beta}) < 0. \end{aligned} \tag{2-12}$$

The last inequality follows from assumption (1-9). This contradiction shows that (1-10) must be true. □

3. Symmetry of solutions in the whole space $\mathbb{R}^n$

Proof of Theorem 1.3. Choose an arbitrary direction for the x_1-axis. Let

$$T_{\lambda} = \{x \in \mathbb{R}^n \mid x_1 = \lambda, \lambda \in \mathbb{R}\}, \qquad \Sigma_{\lambda} = \{x \in \mathbb{R}^n \mid x_1 < \lambda\},$$
$$x^{\lambda} = (2\lambda - x_1, x'), \quad u_{\lambda}(x) = u(x^{\lambda}),$$
$$U_{\lambda}(x) = u_{\lambda}(x) - u(x), \quad V_{\lambda}(x) = v_{\lambda}(x) - v(x).$$

Step 1: Starting moving the plane T_{λ} from $-\infty$ to the right along the x_1-axis. We need to show that for λ sufficiently negative,

$$U_{\lambda}(x) \geq 0, \quad V_{\lambda}(x) \geq 0, \quad x \in \Sigma_{\lambda}. \tag{3-1}$$

By the assumption (1-11), for fixed λ and $x \in \Sigma_{\lambda}$, we know that

$$u(x) \to 0 \text{ as } |x| \to +\infty.$$

Since $|x^\lambda| \to +\infty$, as $|x| \to +\infty$, we have

$$u_\lambda(x) = u(x^\lambda) \to 0.$$

Hence for $x \in \Sigma_\lambda$,

$$U_\lambda(x) \to 0 \text{ as } |x| \to +\infty. \tag{3-2}$$

Similarly, one can show that for $x \in \Sigma_\lambda$,

$$V_\lambda(x) \to 0 \text{ as } |x| \to +\infty.$$

If

$$\Sigma_\lambda^- = \{x \in \Sigma_\lambda \mid U_\lambda(x) < 0\} \neq \varnothing,$$

then by the lower semicontinuity of $U_\lambda(x)$, there must exist some $\tilde{x} \in \Sigma_\lambda$ such that

$$U_\lambda(\tilde{x}) = \min_{\Sigma_\lambda} U(x) < 0.$$

Let

$$\begin{aligned} I &= f(u_\lambda(\tilde{x}), v_\lambda(\tilde{x})) - f(u(\tilde{x}), v_\lambda(\tilde{x})), \\ J &= f(u(\tilde{x}), v_\lambda(\tilde{x})) - f(u(\tilde{x}), v(\tilde{x})). \end{aligned}$$

Then

$$\begin{aligned} I + J &= f(u_\lambda(\tilde{x}), v_\lambda(\tilde{x})) - f(u(\tilde{x}), v(\tilde{x})) \\ &= F_\alpha(u_\lambda(\tilde{x})) - F_\alpha(u(\tilde{x})) \\ &\leq 2C_{n,\alpha} c_0 U_\lambda(\tilde{x}) \int_{\Sigma_\lambda} \frac{1}{|\tilde{x} - y^\lambda|^{n+\alpha}}\, dy \\ &< 0. \end{aligned} \tag{3-3}$$

By the mean value theorem and the assumption (1-12), we have

$$I = f_1'(\xi_\lambda(\tilde{x}), v_\lambda(\tilde{x}))U_\lambda(\tilde{x}) > 0 \quad \text{and} \quad J = f_2'(u(\tilde{x}), \eta_\lambda(\tilde{x}))V_\lambda(\tilde{x}), \tag{3-4}$$

where $\xi_\lambda(\tilde{x})$ is between $u_\lambda(\tilde{x})$ and $u(\tilde{x})$; $\eta_\lambda(\tilde{x})$ is between $v_\lambda(\tilde{x})$ and $v(\tilde{x})$. Together (3-3) with (3-4) and (1-12), it is easy to see that

$$V_\lambda(\tilde{x}) < 0.$$

This implies that there exists some $\bar{x} \in \Sigma_\lambda$ such that

$$V_\lambda(\bar{x}) = \min_{\Sigma_\lambda} V(x) < 0.$$

By the mean value theorem again, we have

$$\begin{aligned} F_\alpha(u_\lambda(\tilde{x})) - F_\alpha(u(\tilde{x})) &= I + J \\ &\geq f_1'(\xi_\lambda(\tilde{x}), v_\lambda(\tilde{x}))U_\lambda(\tilde{x}) + f_2'(u(\tilde{x}), \eta_\lambda(\tilde{x}))V_\lambda(\tilde{x}). \end{aligned}$$

By the decay assumptions (1-11) and (1-12), we deduce that

$$f_1'(\xi_\lambda(\tilde{x}), v_\lambda(\tilde{x})),\ f_2'(u(\tilde{x}), \eta_\lambda(\tilde{x})) \sim o\left(\frac{1}{|\tilde{x}|^\alpha}\right).$$

Hence

$$F_\alpha(u_\lambda(\tilde{x})) - F_\alpha(u(\tilde{x})) + c_{11}(\tilde{x})U_\lambda(\tilde{x}) + c_{12}(\tilde{x})V_\lambda(\tilde{x}) \geq 0,$$

where

$$c_{11}(\tilde{x}) = -f_1'(\xi_\lambda(\tilde{x}), v_\lambda(\tilde{x})) \quad \text{and} \quad c_{12}(\tilde{x}) = -f_2'(u(\tilde{x}), \eta_\lambda(\tilde{x})).$$

Similarly, we have

$$F_\beta(v_\lambda(\bar{x})) - F_\beta(v(\bar{x})) + c_{21}(\bar{x})U_\lambda(\bar{x}) + c_{22}(\bar{x})V_\lambda(\bar{x}) \geq 0,$$

where

$$c_{21}(\bar{x}) = -g_1'(\hat{\xi}_\lambda(\bar{x}), v_\lambda(\bar{x})) \quad \text{and} \quad c_{22}(\bar{x}) = -g_2'(u(\bar{x}), \hat{\eta}_\lambda(\bar{x}))$$

with

$$c_{21}(\bar{x}), c_{22}(\bar{x}) \sim o\left(\frac{1}{|\bar{x}|^\beta}\right).$$

Consequently, there exists $R_0 > 0$, such that if $\tilde{x}$ and $\bar{x}$ are negative minima of $U_\lambda(x)$ and $V_\lambda(x)$ in Σ_λ respectively, then by (1-2) we know that

$$|\tilde{x}| \leq R_0 \quad \text{or} \quad |\bar{x}| \leq R_0. \tag{3-5}$$

Without loss of generality, we may assume

$$|\tilde{x}| \leq R_0. \tag{3-6}$$

Combining (3-2) with the fact that $U_\lambda(x) = 0$, $x \in T_\lambda$, it is easy to see if $U_\lambda(x) < 0$ at some point in Σ_λ, then $U_\lambda(x)$ must have a negative minimum in Σ_λ. For λ sufficiently negative, it contradicts (3-6). Hence we have for λ sufficiently negative,

$$U_\lambda(x) \geq 0. \tag{3-7}$$

It follows that $U_\lambda(x) \geq 0$ in Σ_λ. Otherwise, there exists $\bar{x}$ in Σ_λ such that

$$V_\lambda(\bar{x}) = \min_{\Sigma_\lambda} V(x) < 0.$$

From (2-11), we have

$$F_\beta(v_\lambda(\bar{x})) - F_\beta(v(\bar{x})) + c_{22}(\bar{x})V_\lambda(\bar{x}) < 0. \tag{3-8}$$

Combining (1-8) with (3-7), however, we have

$$F_\beta(v_\lambda(\bar{x})) - F_\beta(v(\bar{x})) + c_{22}(\bar{x})V_\lambda(\bar{x}) \geq 0.$$

This is a contradiction with (3-8) and $V_\lambda(x)$ cannot attain its negative value in Σ_λ. It follows that (3-1) must be true. This completes the preparation for the moving planes.

Step 2: Keep moving the plane to the limiting position T_{λ_0} as long as (3-1) holds.

Let

$$\lambda_0 = \sup\{\lambda \mid U_\mu(x), V_\mu(x) \geq 0,\ x \in \Sigma_\mu, \mu \leq \lambda\}.$$

Obviously,

$$\lambda_0 < \infty. \tag{3-9}$$

Otherwise, if $\lambda_0 = \infty$, then for any $\lambda > 0$,

$$\begin{aligned} &u(0^\lambda) > u(0) > 0, \qquad v(0^\lambda) > v(0) > 0,\\ &u(0^\lambda) \sim o\left(\frac{1}{|0^\lambda|^{\gamma_1}}\right), \quad v(0^\lambda) \sim o\left(\frac{1}{|0^\lambda|^{\gamma_2}}\right), \quad \lambda \to \infty. \end{aligned}$$

This is a contradiction and (3-9) is true.

Now we point out that

$$U_{\lambda_0}(x) \equiv 0, \quad V_{\lambda_0}(x) \equiv 0, \quad x \in \Sigma_{\lambda_0}. \tag{3-10}$$

If (3-10) is not true, then from the proof of Theorem 1.1, we only have the case that $U_{\lambda_0}(x) \geq 0$ and $V_{\lambda_0}(x) \geq 0$ but $U_{\lambda_0}(x) \not\equiv 0$ and $V_{\lambda_0}(x) \not\equiv 0$.

In what follows, we will show that the plane T_λ can be moved further to the right. More rigorously, there exists some $\epsilon > 0$, such that for any $\lambda \in [\lambda_0, \lambda_0 + \epsilon)$,

$$U_\lambda(x) \geq 0, \quad V_\lambda(x) \geq 0, \quad x \in \Sigma_\lambda. \tag{3-11}$$

This contradicts the definition of λ_0 and hence (3-10) must be true.

Now we prove (3-11) by using Theorems 1.1 and 1.2. From Theorem 1.1, we have

$$U_{\lambda_0}(x) > 0, \quad V_{\lambda_0}(x) > 0, \quad x \in \Sigma_{\lambda_0}.$$

Let R_0 be the constant determined in Theorem 1.2. It follows that for any $\delta > 0$,

$$U_{\lambda_0}(x) \geq c_0 > 0, \quad V_{\lambda_0}(x) \geq c_0 > 0, \quad x \in \overline{\Sigma_{\lambda_0-\delta} \cap B_{R_0}(0)}.$$

Together with the continuity of $U_\lambda(x)$ and $V_\lambda(x)$ with respect to λ, there exists $\epsilon > 0$, such that for all $\lambda \in [\lambda_0, \lambda_0 + \epsilon)$, we have

$$U_\lambda(x) \geq 0, \quad V_\lambda(x) \geq 0, \quad x \in \overline{\Sigma_{\lambda_0-\delta} \cap B_{R_0}(0)}. \tag{3-12}$$

Suppose that (3-11) is not true. By the proofs of Theorems 1.1 and 1.2, we know that if one of $U_\lambda(x)$ and $V_\lambda(x)$ becomes the negative minimum value at some point in Σ_λ, then there exist $\tilde{x}$ and $\bar{x}$ which are the negative minima of $U_\lambda(x)$ and

$V_\lambda(x)$ in Σ_λ respectively. Additionally, by Theorem 1.2, at least one of them lies in $(\Sigma_\lambda \setminus \Sigma_{\lambda_0-\delta}) \cap B_{R_0}(0)$. Here we consider two possibilities.

Case 1: One of the negative minima of $U_\lambda(x)$ and $V_\lambda(x)$ lies in $B_{R_0}(0)$, i.e., in the narrow region $\Sigma_{\lambda_0+\epsilon} \backslash \Sigma_{\lambda_0-\delta}$, and the other is outside of $B_{R_0}(0)$. Without loss of generality, we assume the negative minimum of $U_\lambda(x)$ lies in $B_{R_0}(0)$. From (2-5), we have

$$U_\lambda(\tilde{x}) \geq -cl^\alpha c_{12}(\tilde{x}) V_\lambda(\tilde{x}) \tag{3-13}$$

and

$$\begin{aligned} 0 &\leq F_\beta(v_\lambda(\bar{x})) - F_\beta(v(\bar{x})) + c_{21}(\bar{x}) U_\lambda(\bar{x}) + c_{22}(\bar{x}) V_\lambda(\bar{x}) \\ &\leq \frac{C}{|\bar{x}|^\beta} V_\lambda(\bar{x}) + c_{21}(\bar{x}) U_\lambda(\bar{x}) \\ &\leq \frac{C}{|\bar{x}|^\beta} V_\lambda(\bar{x}) - cc_{21}(\bar{x}) c_{12}(\tilde{x}) l^\alpha V_\lambda(\tilde{x}) \\ &\leq \frac{C}{|\bar{x}|^\beta} V_\lambda(\bar{x}) - cc_{21}(\bar{x}) c_{12}(\tilde{x}) l^\alpha V_\lambda(\bar{x}) \\ &\leq \frac{C}{|\bar{x}|^\beta} V_\lambda(\bar{x})(1 - c_{12}(\tilde{x}) l^\alpha c_{21}(\bar{x}) |\bar{x}|^\beta). \end{aligned}$$

Hence,

$$1 \leq c_{12}(\tilde{x}) l^\alpha c_{21}(\bar{x}) |\bar{x}|^\beta. \tag{3-14}$$

By (1-9), we know that $c_{21}(\bar{x})|\bar{x}|^\beta$ is small for $|\bar{x}|$ sufficiently large. Since $l = \epsilon + \delta$ is very narrow and $c_{12}(\tilde{x})$ is bounded from below in $\Sigma_{\lambda_0+\epsilon} \backslash \Sigma_{\lambda_0-\delta}$, it is easy to see that $c_{12}(\tilde{x}) l^\alpha$ can be small. Consequently,

$$c_{12}(\tilde{x}) l^\alpha c_{21}(\bar{x}) |\bar{x}|^\beta < 1.$$

This is a contradiction with (3-14) and (3-11) is proved.

Case 2: Both of the negative minima of $U_\lambda(x)$ and $V_\lambda(x)$ lie in $B_{R_0}(0)$, i.e., they are all in the narrow region $\Sigma_{\lambda_0+\epsilon} \backslash \Sigma_{\lambda_0-\delta}$.

Recalling (2-4),

$$F_\alpha(u_\lambda(\tilde{x})) - F_\alpha(u(\tilde{x})) + c_{11}(\tilde{x}) U_\lambda(\tilde{x}) \leq \frac{C}{l^\alpha} U_\lambda(\tilde{x}) < 0, \tag{3-15}$$

where $l = \delta + \epsilon$. Together with (1-4), it implies

$$U_\lambda(\tilde{x}) \geq -cc_{12}(\tilde{x}) l^\alpha V_\lambda(\tilde{x}). \tag{3-16}$$

Similarly to (3-15), we have

$$F_\beta(v_\lambda(\bar{x})) - F_\beta(v(\bar{x})) + c_{22}(\bar{x}) V_\lambda(\bar{x}) \leq \frac{C}{l^\beta} V_\lambda(\bar{x}) < 0.$$

Noting (3-16), for l sufficiently small, it gives

$$\begin{aligned}0 &\le F_\beta(v_\lambda(\bar{x})) - F_\beta(v(\bar{x})) + c_{21}(\bar{x})U_\lambda(\bar{x}) + c_{22}(\bar{x})V_\lambda(\bar{x}) \\ &\le \frac{C}{l^\beta} V_\lambda(\bar{x}) + c_{21}(\bar{x})U_\lambda(\tilde{x}) \\ &\le \frac{C}{l^\beta} V_\lambda(\bar{x}) - cc_{21}(\bar{x})c_{12}(\tilde{x})l^\alpha V_\lambda(\tilde{x}) \\ &\le \frac{C}{l^\beta} V_\lambda(\bar{x}) - cc_{21}(\bar{x})c_{12}(\tilde{x})l^\alpha V_\lambda(\bar{x}) \\ &\le \frac{C}{l^\beta} V_\lambda(\bar{x})(1 - c_{12}(\tilde{x})c_{21}(\bar{x})l^{\alpha+\beta}) < 0.\end{aligned}$$

This contradiction shows that (3-11) has to be true.

Now we have proved that $U_{\lambda_0}(x) \equiv 0$, $V_{\lambda_0}(x) \equiv 0$, $x \in \Sigma_{\lambda_0}$. Since the x_1-direction can be chosen arbitrarily, we actually indicate that $u(x)$ and $v(x)$ are radically symmetric about some point x^0. Also the monotonicity follows easily from the argument. This completes the proof of Theorem 1.3. □

4. Nonexistence of positive solutions on a half space $\mathbb{R}^n_+$

In this section, we investigate the system (1-2).

Proof of Theorem 1.4. Based on (1-3), from the proof of Lemma 2.1 in [Wang and Yu 2017], one can see that either

$$u(x) > 0,\ v(x) > 0 \quad \text{or} \quad u(x) \equiv 0,\ v(x) \equiv 0 \quad \text{for } x \in \mathbb{R}^n_+,$$

where $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_n > 0\}$. In fact, assume $u(x) \not\equiv 0$, and there exists $x^0 \in \mathbb{R}^n_+$ such that $u(x^0) = 0$, then

$$\begin{aligned}F_\alpha(u(x^0)) &= \int_{\mathbb{R}^n} \frac{G(u(x^0) - u(y))}{|x^0 - y|^{n+\alpha}}\,dy = \int_{\mathbb{R}^n} \frac{G(u(x^0) - u(y)) - G(0)}{|x^0 - y|^{n+\alpha}}\,dy \\ &= \int_{\mathbb{R}^n} \frac{G'(\cdot)(u(x^0) - u(y))}{|x^0 - y|^{n+\alpha}}\,dy \le c_0 \int_{\mathbb{R}^n} \frac{-u(y)}{|x^0 - y|^{n+\alpha}}\,dy < 0,\end{aligned}$$

i.e., $0 \le f(u(x), v(x)) = F_\alpha(u(x)) < 0$, which is impossible. Hence if $u(x)$ or $v(x)$ attains 0 somewhere in $\mathbb{R}^n_+$, then $u(x) = v(x) \equiv 0$, $x \in \mathbb{R}^n$.

Now we always assume that $u(x) > 0$ and $v(x) > 0$ in $\mathbb{R}^n_+$. Let us carry on the method of moving planes to the solution u along the x_n-direction.

Denote

$$T_\lambda = \{x \in \mathbb{R}^n_+ \mid x_n = \lambda, \lambda > 0\}, \qquad \Sigma_\lambda = \{x \in \mathbb{R}^n_+ \mid 0 < x_n < \lambda\}.$$

Let

$$x^\lambda = (x_1, x_2, \dots, 2\lambda - x_n)$$

be the reflection of x about the plane T_λ, and

$$U_\lambda(x) = u_\lambda(x) - u(x), \quad V_\lambda(x) = v_\lambda(x) - v(x).$$

Using the key inequality (2-2) obtained in the proof of Theorem 1.1, we only need to take $\Sigma = \Sigma_\lambda \cup \mathbb{R}^n_-$, where $\mathbb{R}^n_- = \{x \in \mathbb{R}^n \mid x_n \le 0\}$.

Step 1: It is obvious that, for $\lambda \le 0$, we have

$$U_\lambda(x) \ge 0, \quad V_\lambda(x) \ge 0, \quad x \in \mathbb{R}^n_-. \tag{4-1}$$

For $\lambda > 0$ sufficiently small, Σ_λ is a narrow region, we have immediately

$$U_\lambda(x) \ge 0, \quad V_\lambda(x) \ge 0, \quad x \in \Sigma_\lambda. \tag{4-2}$$

Step 2: Since (4-2) provides a starting point, we move the plane T_λ upward as long as (4-2) holds. Define

$$\lambda_0 = \sup\{\lambda \ge 0 \mid U_\mu(x) \ge 0, V_\mu(x) \ge 0, x \in \Sigma_\mu, \mu \le \lambda\}.$$

We show that

$$\lambda_0 = \infty. \tag{4-3}$$

Otherwise, if $\lambda_0 < \infty$, we show that the plane T_λ can be moved further up. To be more rigorous, there exists some $\epsilon > 0$, such that, for any $\lambda \in (\lambda_0, \lambda_0 + \epsilon)$,

$$U_\lambda(x) \ge 0, \quad V_\lambda(x) \ge 0, \quad x \in \Sigma_\lambda.$$

This is a contradiction with the definition of λ_0. Hence, (4-3) holds.

By using Theorem 1.1, Theorem 1.2, and similar arguments as in Section 3, we can prove that

$$U_{\lambda_0} \equiv 0, \quad V_{\lambda_0} \equiv 0, \quad x \in \Sigma_{\lambda_0}, \quad \lambda_0 = \infty,$$

which implies

$$u(x_1, \dots, x_{n-1}, 2\lambda_0) = u(x_1, \dots, x_{n-1}, 0) = 0,$$
$$v(x_1, \dots, x_{n-1}, 2\lambda_0) = v(x_1, \dots, x_{n-1}, 0) = 0.$$

This is impossible, because we have assumed that $u(x), v(x) > 0$ in $\mathbb{R}^n_+$.

Therefore, (4-3) must be valid and the solutions $u(x)$, $v(x)$ are increasing with respect to x_n. This contradicts (1-13) and completes the proof of Theorem 1.4. □

5. Application to fully nonlinear nonlocal system

In this section, we consider

$$\begin{aligned} F_\alpha(u(x)) &= -u^p(x) + v^q(x), && x \in \mathbb{R}^n, \\ F_\beta(v(x)) &= -v^p(x) + u^q(x), && x \in \mathbb{R}^n, \\ u(x), v(x) &> 0, && x \in \mathbb{R}^n. \end{aligned} \tag{5-1}$$

Obviously, (5-1) is a specific case of (1-1) and we have the similar conclusion here.

Theorem 5.1. *Assume that $u(x) \in L_\alpha \cap C^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$, $v(x) \in L_\beta \cap C^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$ are positive solutions of system (5-1). Suppose that for some $\gamma_1, \gamma_2 > 0$, $u(x), v(x)$ satisfy the assumption (1-11) and*

$$\min\{(p-1)\gamma_1, (q-1)\gamma_1\} > \alpha, \quad \min\{(p-1)\gamma_2, (q-1)\gamma_2\} > \beta.$$

Then $u(x), v(x)$ must be radially symmetric and monotone decreasing about some point in $\mathbb{R}^n$.

By using Theorem 1.3, we can prove Theorem 5.1 directly. Notice that, if we let $f(u(x), v(x)) = -u^p(x) + v^q(x)$ and $g(u(x), v(x)) = -v^p(x) + u^q(x)$, it is easy to see that f, g satisfy the assumption (1-12). For convenience, we omit the proof of Theorem 5.1 here.

References

[Brändle et al. 2013] C. Brändle, E. Colorado, A. de Pablo, and U. Sánchez, "A concave-convex elliptic problem involving the fractional Laplacian", *Proc. Roy. Soc. Edinburgh Sect. A* **143**:1 (2013), 39–71. MR Zbl

[Caffarelli and Silvestre 2007] L. Caffarelli and L. Silvestre, "An extension problem related to the fractional Laplacian", *Comm. Partial Differential Equations* **32**:7-9 (2007), 1245–1260. MR Zbl

[Caffarelli and Silvestre 2009] L. Caffarelli and L. Silvestre, "Regularity theory for fully nonlinear integro-differential equations", *Comm. Pure Appl. Math.* **62**:5 (2009), 597–638. MR Zbl

[Cao and Chen 2013] L. Cao and W. Chen, "Liouville type theorems for poly-harmonic Navier problems", *Discrete Contin. Dyn. Syst.* **33**:9 (2013), 3937–3955. MR Zbl

[Cao and Dai 2013] L. Cao and Z. Dai, "A Liouville-type theorem for an integral system on a half-space R^n_+", *J. Inequal. Appl.* **2013** (2013), art. id. 37. MR Zbl

[Chen and Zhu 2016] W. Chen and J. Zhu, "Indefinite fractional elliptic problem and Liouville theorems", *J. Differential Equations* **260**:5 (2016), 4758–4785. MR Zbl

[Chen et al. 2005] W. Chen, C. Li, and B. Ou, "Qualitative properties of solutions for an integral equation", *Discrete Contin. Dyn. Syst.* **12**:2 (2005), 347–354. MR Zbl

[Chen et al. 2006] W. Chen, C. Li, and B. Ou, "Classification of solutions for an integral equation", *Comm. Pure Appl. Math.* **59**:3 (2006), 330–343. MR Zbl

[Chen et al. 2015] W. Chen, Y. Fang, and R. Yang, "Liouville theorems involving the fractional Laplacian on a half space", *Adv. Math.* **274** (2015), 167–198. MR Zbl

[Chen et al. 2017a] W. Chen, C. Li, and G. Li, "Maximum principles for a fully nonlinear fractional order equation and symmetry of solutions", *Calc. Var. Partial Differential Equations* **56**:2 (2017), art. id. 29. MR Zbl

[Chen et al. 2017b] W. Chen, C. Li, and Y. Li, "A direct method of moving planes for the fractional Laplacian", *Adv. Math.* **308** (2017), 404–437. MR Zbl

[Cheng 2017] T. Cheng, "Monotonicity and symmetry of solutions to fractional Laplacian equation", *Discrete Contin. Dyn. Syst.* **37**:7 (2017), 3587–3599. MR Zbl

[Cheng et al. 2017a] C. Cheng, Z. Lü, and Y. Lü, "A direct method of moving planes for the system of the fractional Laplacian", *Pacific J. Math.* **290**:2 (2017), 301–320. MR Zbl

[Cheng et al. 2017b] T. Cheng, G. Huang, and C. Li, "The maximum principles for fractional Laplacian equations and their applications", *Commun. Contemp. Math.* **19**:6 (2017), art. id. 1750018. MR Zbl

[Gilbarg and Trudinger 1977] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Grundlehren der Math. Wissenschaften **224**, Springer, 1977. MR Zbl

[Han et al. 2012] X. Han, G. Lu, and J. Zhu, "Characterization of balls in terms of Bessel-potential integral equation", *J. Differential Equations* **252**:2 (2012), 1589–1602. MR Zbl

[Hang 2007] F. Hang, "On the integral systems related to Hardy–Littlewood–Sobolev inequality", *Math. Res. Lett.* **14**:3 (2007), 373–383. MR Zbl

[Hang et al. 2009] F. Hang, X. Wang, and X. Yan, "An integral equation in conformal geometry", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **26**:1 (2009), 1–21. MR Zbl

[Lei et al. 2012] Y. Lei, C. Li, and C. Ma, "Asymptotic radial symmetry and growth estimates of positive solutions to weighted Hardy–Littlewood–Sobolev system of integral equations", *Calc. Var. Partial Differential Equations* **45**:1-2 (2012), 43–61. MR Zbl

[Li 2017] J. Li, "Monotonicity and symmetry of fractional Lane–Emden-type equation in the parabolic domain", *Complex Var. Elliptic Equ.* **62**:1 (2017), 135–147. MR Zbl

[Li and Ma 2017] Y. Li and P. Ma, "Symmetry of solutions for a fractional system", *Sci. China Math.* **60**:10 (2017), 1805–1824. MR Zbl

[Li and Zhuo 2010] D. Li and R. Zhuo, "An integral equation on half space", *Proc. Amer. Math. Soc.* **138**:8 (2010), 2779–2791. MR Zbl

[Liu and Ma 2016] B. Liu and L. Ma, "Radial symmetry results for fractional Laplacian systems", *Nonlinear Anal.* **146** (2016), 120–135. MR Zbl

[Lu and Zhu 2011] G. Lu and J. Zhu, "Symmetry and regularity of extremals of an integral equation related to the Hardy–Sobolev inequality", *Calc. Var. Partial Differential Equations* **42**:3-4 (2011), 563–577. MR Zbl

[Lu and Zhu 2012] G. Lu and J. Zhu, "An overdetermined problem in Riesz-potential and fractional Laplacian", *Nonlinear Anal.* **75**:6 (2012), 3036–3048. MR Zbl

[Ma and Chen 2006] L. Ma and D. Chen, "A Liouville type theorem for an integral system", *Commun. Pure Appl. Anal.* **5**:4 (2006), 855–859. MR Zbl

[Ma and Zhao 2010] L. Ma and L. Zhao, "Classification of positive solitary solutions of the nonlinear Choquard equation", *Arch. Ration. Mech. Anal.* **195**:2 (2010), 455–467. MR Zbl

[Wang and Niu 2017] P. Wang and P. Niu, "A direct method of moving planes for a fully nonlinear nonlocal system", *Commun. Pure Appl. Anal.* **16**:5 (2017), 1707–1718. MR Zbl

[Wang and Yu 2017] P. Wang and M. Yu, "Solutions of fully nonlinear nonlocal systems", *J. Math. Anal. Appl.* **450**:2 (2017), 982–995. MR Zbl

[Zhang et al. 2017] L. Zhang, M. Yu, and J. He, "A Liouville theorem for a class of fractional systems in $\mathbb{R}^n_+$", *J. Differential Equations* **263**:9 (2017), 6025–6065. MR Zbl

[Zhuo et al. 2016] R. Zhuo, W. Chen, X. Cui, and Z. Yuan, "Symmetry and non-existence of solutions for a nonlinear system involving the fractional Laplacian", *Discrete Contin. Dyn. Syst.* **36**:2 (2016), 1125–1141. MR Zbl

Received December 22, 2017. Revised May 16, 2018.

BIRAN ZHANG
SCHOOL OF MATHEMATICS AND STATISTICS
JIANGSU NORMAL UNIVERSITY
XU ZHOU
CHINA
443792411@qq.com

ZHONGXUE LÜ
SCHOOL OF MATHEMATICS AND STATISTICS
JIANGSU NORMAL UNIVERSITY
XUZHOU
CHINA
lvzx1@tom.com

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 299 No. 1 March 2019
