Pacific Journal of Mathematics

Volume 299 No. 2 April 2019

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

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

The subscription price for 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/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 2, 2019
dx.doi.org/10.2140/pjm.2019.299.257

HIERARCHICALLY HYPERBOLIC SPACES II: COMBINATION THEOREMS AND THE DISTANCE FORMULA

JASON BEHRSTOCK, MARK HAGEN AND ALESSANDRO SISTO

We introduce a number of tools for finding and studying *hierarchically hyperbolic spaces* (HHS), a rich class of spaces including mapping class groups of surfaces, Teichmüller space with either the Teichmüller or Weil–Petersson metrics, right-angled Artin groups, and the universal cover of any compact special cube complex. We begin by introducing a streamlined set of axioms defining an HHS. We prove that all HHS satisfy a Masur–Minsky-style distance formula, thereby obtaining a new proof of the distance formula in the mapping class group without relying on the Masur–Minsky hierarchy machinery. We then study examples of HHS; for instance, we prove that when M is a closed irreducible 3-manifold then $\pi_1 M$ is an HHS if and only if it is neither *Nil* nor *Sol*. We establish this by proving a general combination theorem for trees of HHS (and graphs of HH groups). We also introduce a notion of "hierarchical quasiconvexity", which in the study of HHS is analogous to the role played by quasiconvexity in the study of Gromov-hyperbolic spaces.

Behrstock was supported as a Simons Fellow. Hagen was supported by NSF Grant Number 1045119 and by EPSRC. Sisto was supported by the Swiss National Science Foundation project 144373.
MSC2010: 20F36, 20F65, 20F67.
Keywords: geometric group theory, hierarchically hyperbolic, mapping class group.

Introduction

One of the most remarkable aspects of the theory of mapping class groups of surfaces is that the coarse geometry of the mapping class group, $\mathcal{MCG}(S)$, can be fully reconstructed from its shadows on a collection of hyperbolic spaces — namely the curve graphs of subsurfaces of the underlying surface. Each subsurface of the surface S is equipped with a hyperbolic curve graph and a projection, the *subsurface projection*, to this graph from $\mathcal{MCG}(S)$; there are also projections between the various curve graphs. The powerful Masur–Minsky distance formula [2000] shows that the distance between points of $\mathcal{MCG}(S)$ is coarsely the sum over all subsurfaces of the distances between the projections of these points to the various curve graphs. Meanwhile, the consistency/realization theorem [Behrstock et al. 2012] tells us that tuples with coordinates in the different curve graphs that obey "consistency" conditions characteristic of images of actual points in $\mathcal{MCG}(S)$ are, coarsely, images of points in $\mathcal{MCG}(S)$. Finally, any two points in $\mathcal{MCG}(S)$ are joined by a uniform-quality quasigeodesic projecting to a uniform unparameterized quasigeodesic in each curve graph — a *hierarchy path* [Masur and Minsky 2000].

It is perhaps surprising that analogous behavior should appear in CAT(0) cube complexes, since the mapping class group cannot act properly on such complexes, cf. [Bridson 2010; Haglund 2007; Kapovich and Leeb 1996]. However, mapping class groups enjoy several properties reminiscent of nonpositively/negatively curved spaces, including: automaticity (and, thus, quadratic Dehn function) [Mosher 1995], having many quasimorphisms [Bestvina and Fujiwara 2002], super-linear divergence [Behrstock 2006], etc. Mapping class groups also exhibit coarse versions of some features of CAT(0) cube complexes, including coarse centroids/medians [Behrstock and Minsky 2011] and, more generally, a local coarse structure of a cube complex as made precise in [Bowditch 2013], applications to embeddings in trees, [Behrstock et al. 2011], etc. Accordingly, it is natural to seek a common thread joining these important classes of groups and spaces.

In [Hagen 2014], it was shown that, for an arbitrary CAT(0) cube complex $\mathcal{X}$, the intersection-graph of the hyperplane carriers — the *contact graph* — is hyperbolic, and in fact quasi-isometric to a tree. This object seems at first glance quite different from the curve graph (which records, after all, *non*-intersection), but there are a number of reasons this is quite natural, two of which we now mention. First, the curve graph can be realized as a coarse intersection graph of product regions in $\mathcal{MCG}$. Second, the contact graph is closely related to the intersection graph of the hyperplanes themselves; when $\mathcal{X}$ is the universal cover of the Salvetti complex of a right-angled Artin group, the latter graph records commutation of conjugates of generators, just as the curve graph records commutation of Dehn twists.

The cube complex $\mathcal{X}$ coarsely projects to its contact graph. Moreover, using disc diagram techniques, it is not hard to show that any two 0-cubes in a CAT(0) cube complex are joined by a combinatorial geodesic projecting to a geodesic in the contact graph; see our [Behrstock et al. 2017b], which we will henceforth abbreviate as [Part I]. The observation that CAT(0) cube complexes have “hierarchy paths” with very strong properties — motivated a search for an analogue of the theory of curve graphs and subsurface projections in the world of CAT(0) cube complexes. This was largely achieved in [Part I], where a theory completely analogous to the mapping class group theory was constructed for a wide class of CAT(0) cube complexes, with (a variant of) the contact graph playing the role of the curve graph. (Results of this type for right-angled Artin groups, using the *extension graph*, were obtained in [Kim and Koberda 2014]; see [Part I] for a comparison of the two approaches.)

These results motivated us to define a notion of “spaces with distance formulae”, which we did in [Part I], by introducing the class of *hierarchically hyperbolic spaces* (HHS) to provide a framework for studying many groups and spaces which arise naturally in geometric group theory, including mapping class groups and virtually special groups, and to provide a notion of “coarse nonpositive curvature” which is quasi-isometry invariant while still yielding some of those properties available via local geometry in the classical setting of nonpositively curved spaces.

As mentioned above, the three most salient features of hierarchically hyperbolic spaces are: the distance formula, the realization theorem, and the existence of hierarchy paths. In the treatment given in [Part I], these attributes are part of the definition of a hierarchically hyperbolic space. This is somewhat unsatisfactory since, in the mapping class group and cubical settings, proving these theorems requires serious work.

In this paper, we show that although the definition of hierarchically hyperbolic spaces previously introduced identifies the right class of spaces, there exists a streamlined set of axioms for that class of spaces which are much easier to verify in practice than those presented in Section 13 of [Part I] and which don’t require assuming a distance formula, realization theorem, or the existence of hierarchy paths. Thus, a significant portion of this paper is devoted to proving that those results can be derived from the simplified axioms we introduce here. Along the way, we obtain a new, simplified proof of the actual Masur–Minsky distance formula for the mapping class group. We then examine various geometric properties of hierarchically hyperbolic spaces and groups, including many reminiscent of the world of CAT(0) spaces and groups; for example, we show, using an argument due to Bowditch, that hierarchically hyperbolic groups have quadratic Dehn function. Finally, taking advantage of the simpler set of axioms, we prove combination theorems enabling the construction of new hierarchically hyperbolic spaces/groups from old.

The definition of a hierarchically hyperbolic space still has several parts, the details of which we postpone to Section 1. However, the idea is straightforward:

a hierarchically hyperbolic space is a pair $(\mathcal{X}, \mathfrak{S})$, where $\mathcal{X}$ is a metric space and $\mathfrak{S}$ indexes a set of δ-hyperbolic spaces with several features (for each $U \in \mathfrak{S}$ the associated space is denoted $\mathcal{C}U$). Most notably, $\mathfrak{S}$ is endowed with three mutually exclusive relations, *nesting, orthogonality*, and *transversality*, respectively generalizing nesting, disjointness, and overlapping of subsurfaces. For each $U \in \mathfrak{S}$, we have a coarsely Lipschitz projection $\pi_U : \mathcal{X} \to \mathcal{C}U$, and there are relative projections $\mathcal{C}U \to \mathcal{C}V$ when $U, V \in \mathfrak{S}$ are nonorthogonal. These projections are required to obey "consistency" conditions modeled on the inequality identified by Behrstock [2006], as well as a version of the bounded geodesic image theorem and large link lemma of [Masur and Minsky 2000], among other conditions. A finitely generated group G is *hierarchically hyperbolic* if it can be realized as a group of HHS automorphisms ("hieromorphisms", as defined in Section 1) so that the induced action on $\mathcal{X}$ by uniform quasi-isometries is geometric and the action on $\mathfrak{S}$ is cofinite. Hierarchically hyperbolic groups, endowed with word-metrics, are hierarchically hyperbolic spaces, but the converse does not appear to be true.

Combination theorems. One of the main contributions in this paper is to provide many new examples of hierarchically hyperbolic groups, thus showing that mapping class groups and various cubical complexes/groups are just two of many interesting families in this class of groups and spaces. We provide a number of combination theorems, which we will describe below. One consequence of these results is the following classification of exactly which 3-manifold groups are hierarchically hyperbolic:

Theorem 10.1 (which 3-manifolds are hierarchically hyperbolic). *Let M be a closed 3-manifold. Then $\pi_1(M)$ is a hierarchically hyperbolic space if and only if M does not have a Sol or Nil component in its prime decomposition.*

This result has a number of applications to the many fundamental groups of 3-manifolds which are HHS. For instance, in such cases, it follows from results in [Part I] that: except for $\mathbb{Z}^3$, the top dimension of a quasiflat in such a group is 2, and any such quasiflat is locally close to a "standard flat" (this generalizes one of the main results of [Kapovich and Leeb 1997, Theorem 4.10]); up to finite index, $\mathbb{Z}$ and $\mathbb{Z}^2$ are the only finitely generated nilpotent groups which admit quasi-isometric embeddings into $\pi_1(M)$; and, except in the degenerate case where $\pi_1(M)$ is virtually abelian, such groups are all acylindrically hyperbolic (as also shown in [Minasyan and Osin 2015]).

Remark (hierarchically hyperbolic spaces vs. hierarchically hyperbolic groups). There is an important distinction to be made between a *hierarchically hyperbolic space*, which is a metric space $\mathcal{X}$ equipped with a collection $\mathfrak{S}$ of hyperbolic spaces with certain properties, and a *hierarchically hyperbolic group*, which is a group acting geometrically on a hierarchically hyperbolic space in such a way that the induced action on $\mathfrak{S}$ is cofinite. The latter property is considerably stronger. For example, Theorem 10.1 shows that $\pi_1 M$, with any word-metric, is a hierarchically hyperbolic space, but, as we discuss in Remark 10.2, $\pi_1 M$ probably fails to be a hierarchically hyperbolic group in general; for instance we conjecture this is the case for those graph manifolds which can not be cocompactly cubulated.

In the course of proving Theorem 10.1, we establish several general combination theorems, including one about relative hyperbolicity and one about graphs of groups. The first is:

Theorem 9.1 (hyperbolicity relative to HHGs). *Let the group G be hyperbolic relative to a finite collection $\mathcal{P}$ of peripheral subgroups. If each $P \in \mathcal{P}$ is a hierarchically hyperbolic space, then G is a hierarchically hyperbolic space. Further, if each $P \in \mathcal{P}$ is a hierarchically hyperbolic group, then so is G.*

Another of our main results is a combination theorem, Theorem 8.6, establishing when a tree of hierarchically hyperbolic spaces is again a hierarchically hyperbolic space. In the statement below, *hierarchical quasiconvexity* is a natural generalization of both quasiconvexity in the hyperbolic setting and cubical convexity in the cubical setting, which we shall discuss in some detail shortly. The remaining conditions are technical and explained in Section 8, but are easily verified in practice.

Theorem 8.6 (combination theorem for HHS). *Let $\mathcal{T}$ be a tree of hierarchically hyperbolic spaces. Suppose that*

- *edge-spaces are uniformly* hierarchically quasiconvex *in incident vertex spaces*,
- *each edge-map is* full,
- *$\mathcal{T}$ has* bounded supports,
- *If e is an edge of T and S_e is the $\sqsubseteq$-maximal element of $\mathfrak{S}_e$, then for all $V \in \mathfrak{S}_{e^\pm}$, the elements V and $\phi^{\diamondsuit}_{e^\pm}(S_e)$ are not orthogonal in $\mathfrak{S}_{e^\pm}$. Moreover, there exists $K \geq 0$ such that for all vertices v of $\mathcal{T}$ and edges e incident to v, we have $\mathrm{d}_{\mathrm{Haus}}(\phi_v(\mathcal{X}_e)), \boldsymbol{F}_{\phi^{\diamondsuit}_v(S_e)} \times \{\star\}) \leq K$, where $S_e \in \mathfrak{S}_e$ is the unique maximal element and $\star \in \boldsymbol{E}_{\phi^{\diamondsuit}_v(S_e)}$.*

Then $\mathcal{X}(\mathcal{T})$ is hierarchically hyperbolic.

As a consequence, we obtain a set of sufficient conditions guaranteeing that a graph of hierarchically hyperbolic groups is a hierarchically hyperbolic group.

Corollary 8.24 (combination theorem for HHG). *Let $\mathcal{G} = (\Gamma, \{G_v\}, \{G_e\}, \{\phi^{\pm}_e\})$ be a finite graph of hierarchically hyperbolic groups. Suppose that $\mathcal{G}$ equivariantly satisfies the hypotheses of Theorem 8.6. Then the total group G of $\mathcal{G}$ is a hierarchically hyperbolic group.*

Finally, we prove that products of hierarchically hyperbolic spaces admit natural hierarchically hyperbolic structures.

As mentioned earlier, we will apply the combination theorems to fundamental groups of 3-manifolds, but their applicability is broader. For example, they can be applied to fundamental groups of higher dimensional manifolds such as the ones considered in [Frigerio et al. 2015].

The distance formula and realization. As defined in [Part I], the basic definition of a *hierarchically hyperbolic space* is modeled on the essential properties underlying the "hierarchy machinery" of mapping class groups. In this paper, we revisit the basic definition and provide a new, refined set of axioms; the main changes are the removal of the "distance formula" and "hierarchy path" axioms and the replacement of the "realization" axiom by a far simpler "partial realization". These new axioms are both more fundamental and more readily verified.

An important result in mapping class groups which provides a starting point for much recent research in the field is the celebrated "distance formula" of Masur and Minsky [2000] which provides a way to estimate distances in the mapping class group, up to uniformly bounded additive and multiplicative distortion, via distances in the curve graphs of subsurfaces. We give a new, elementary, proof of the distance formula in the mapping class group. The first step in doing so is verifying that mapping class groups satisfy the new axioms of a hierarchically hyperbolic space. We provide elementary, simple proofs of the axioms for which elementary proofs do not exist in the literature (most notably, the uniqueness axiom); this is done in Section 11. This then combines with our proof of the following result which states that any hierarchically hyperbolic space satisfies a "distance formula" (which in the case of the mapping class group provides a new proof of the Masur–Minsky distance formula):

Theorem 4.5 (distance formula for HHS). *Let $(X, \mathfrak{S})$ be hierarchically hyperbolic. Then there exists s_0 such that for all $s \geq s_0$ there exist constants K, C such that for all $x, y \in \mathcal{X}$,*

$$\mathsf{d}_{\mathcal{X}}(x, y) \asymp_{(K,C)} \sum_{W \in \mathfrak{S}} \{\!\{\mathsf{d}_W(x, y)\}\!\}_s,$$

where $\mathsf{d}_W(x, y)$ denotes the distance in the hyperbolic space $\mathcal{C}W$ between the projections of x, y and $\{\!\{A\}\!\}_B = A$ if $A \geq B$ and 0 otherwise.

Moreover, we show in Theorem 4.4 that any two points in $\mathcal{X}$ are joined by a uniform quasigeodesic γ projecting to a uniform unparameterized quasigeodesic in $\mathcal{C}U$ for each $U \in \mathfrak{S}$. The existence of such *hierarchy paths* was hypothesized as part of the definition of a hierarchically hyperbolic space in [Behrstock et al. 2017b], but now it is proven as a consequence of the other axioms.

The realization theorem for the mapping class group was established by Behrstock, Kleiner, Minsky and Mosher in [Behrstock et al. 2012, Theorem 4.3]. This theorem states that given a surface S and, for each subsurface $W \subseteq S$, a point in the curve complex of W, this sequence of points arises as the projection of a point in the mapping class group (up to bounded error), whenever the curve complex elements satisfy certain pairwise "consistency conditions." Thus the realization theorem provides another sense in which all of the quasi-isometry invariant geometry of the mapping class group is encoded by the projections onto the curve graphs of subsurfaces.[1] In Section 3 we show that an arbitrary hierarchically hyperbolic space satisfies a realization theorem. Given our elementary proof of the new axioms for mapping class groups in Section 11, we thus obtain a new proof of [Behrstock et al. 2012, Theorem 4.3].

Hulls and the coarse median property. Bowditch introduced a notion of *coarse median space* to generalize some results about median spaces to a more general setting, and, in particular, to the mapping class group [Bowditch 2013]. Bowditch [2018] observed that any hierarchically hyperbolic space is a coarse median space; for completeness we provide a short proof of this result in Theorem 7.3. Using Bowditch's results about coarse median spaces, we obtain a number of applications as corollaries. For instance, Corollary 7.9 is obtained from [Bowditch 2014a, Theorem 9.1] and says that any hierarchically hyperbolic space satisfies the rapid decay property and Corollary 7.5 is obtained from [Bowditch 2013, Corollary 8.3] to show that all hierarchically hyperbolic groups are finitely presented and have quadratic Dehn functions. This provides examples of groups that are not hierarchically hyperbolic, for example:

Corollary 7.6 ($\mathrm{Out}(F_n)$ is not an HHG). *For $n \geq 3$, the group $\mathrm{Out}(F_n)$ is not a hierarchically hyperbolic group.*

Indeed, $\mathrm{Out}(F_n)$ was shown in [Bridson and Vogtmann 1995; 2012; Handel and Mosher 2013b] to have exponential Dehn function. This result is interesting as a counter-point to the well-known and fairly robust analogy between $\mathrm{Out}(F_n)$ and the mapping class group of a surface; especially in light of the fact that $\mathrm{Out}(F_n)$ is known to have a number of properties reminiscent of the axioms for an HHS, cf. [Bestvina and Feighn 2014a; 2014b; Handel and Mosher 2013a; Sabalka and Savchuk 2012]. The coarse median property, via work of Bowditch, also implies that asymptotic cones of hierarchically hyperbolic spaces are contractible. Moreover, in Corollary 6.7, we bound the homological dimension of any asymptotic cone of a hierarchically hyperbolic space. This latter result relies on the use of *hulls* of finite sets of points in the HHS $\mathcal{X}$. This construction generalizes the Σ-*hull* of a finite set,

[1] In [Behrstock et al. 2012], the name consistency theorem is used to refer to the necessary and sufficient conditions for realization; since we find it useful to break up these two aspects, we refer to this half as the realization theorem, since anything that satisfies the consistency conditions is realized.

constructed in the mapping class group context in [Behrstock et al. 2012]. (It also generalizes a special case of the ordinary combinatorial convex hull in a CAT(0) cube complex.) A key feature of these hulls is that they are coarse retracts of $\mathcal{X}$ (see Proposition 6.3), and this plays an important role in the proof of the distance formula.

Hierarchical spaces. We also introduce the more general notion of a *hierarchical space* (HS). This is the same as a hierarchically hyperbolic space, except that we do not require the various associated spaces $\mathcal{C}U$, onto which we are projecting, to be hyperbolic. Although we mostly focus on HHS in this paper, a few things are worth noting. First, the realization theorem (Theorem 3.1) actually makes no use of hyperbolicity of the $\mathcal{C}U$, and therefore holds in the more general context of HS; see Section 3.

Second, an important subclass of the class of HS is the class of *relatively hierarchically hyperbolic spaces*, which we introduce in Section 6B. These are hierarchical spaces where the spaces $\mathcal{C}U$ are uniformly hyperbolic except when U is minimal with respect to the nesting relation. As their name suggests, this class includes all metrically relatively hyperbolic spaces; see Theorem 9.3. With an eye to future applications, in Section 6B we prove a distance formula analogous to Theorem 4.5 for relatively hierarchically hyperbolic spaces, and also establish the existence of hierarchy paths. The strategy is to build, for each pair of points x, y, in the relatively hierarchically hyperbolic space, a "hull" of x, y, which we show is hierarchically hyperbolic with uniform constants. We then apply Theorems 4.5 and 4.4.

Standard product regions and hierarchical quasiconvexity. In Section 5A, we introduce the notion of a *hierarchically quasiconvex* subspace of a hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$. In the case where $\mathcal{X}$ is hyperbolic, this notion coincides with the usual notion of quasiconvexity. The main technically useful features of hierarchically quasiconvex subspaces generalize key features of quasiconvexity: they inherit the property of being hierarchically hyperbolic (Proposition 5.6) and one can coarsely project onto them (Lemma 5.5).

Along with the hulls discussed above, the most important examples of hierarchically quasiconvex subspaces are *standard product regions*: for each $U \in \mathfrak{S}$, one can consider the set P_U of points $x \in \mathcal{X}$ whose projection to each $\mathcal{C}V$ is allowed to vary only if V is nested into or orthogonal to U; otherwise, x projects to the same place in $\mathcal{C}V$ as $\mathcal{C}U$ does under the relative projection. The space P_U coarsely decomposes as a product, with factors corresponding to the nested and orthogonal parts. Product regions play an important role in the study of boundaries and automorphisms of hierarchically hyperbolic spaces in [Durham et al. 2017], as well as in the study of quasiboxes and quasiflats in hierarchically hyperbolic spaces carried out in [Behrstock et al. 2017b].

Some questions and future directions. Before embarking on the discussion outlined above, we raise a few questions about hierarchically hyperbolic spaces and groups that we believe are of significant interest.

The first set of questions concern the scope of the theory, i.e., which groups and spaces are hierarchically hyperbolic and which operations preserve the class of HHS:

Question A (cubical groups). Let G act properly and cocompactly on a CAT(0) cube complex. Is G a hierarchically hyperbolic group? Conversely, suppose that $(G, \mathfrak{S})$ is a hierarchically hyperbolic group; are there conditions on the elements of $\mathfrak{S}$ which imply that G acts properly and cocompactly on a CAT(0) cube complex?[2]

Substantial evidence for this conjecture was provided in [Behrstock et al. 2017b] where we established that a CAT(0) cube complex $\mathcal{X}$ containing a *factor system* is a hierarchically hyperbolic space, and the associated hyperbolic spaces are all uniform quasitrees. (Roughly speaking, $\mathcal{X}$ contains a factor-system if the following collection of subcomplexes has finite multiplicity: the smallest collection of convex subcomplexes that contains all combinatorial hyperplanes and is closed under collecting images of closest-point projection maps between its elements.) The class of cube complexes that are HHS in this way contains all universal covers of special cube complexes with finitely many immersed hyperplanes, but the cube complexes containing factor systems have not been completely characterized. In [Durham et al. 2017], we show that the above question is closely related to a conjecture of Behrstock and Hagen on the simplicial boundary of cube complexes [2016, Conjecture 2.8].

More generally, we ask the following:

Question B (factor systems in median spaces). Is there a theory of factor systems in median spaces generalizing that in CAT(0) cube complexes, such that median spaces/groups admitting factor systems are hierarchically hyperbolic?

Presumably, a positive answer to Question B would involve the measured wallspace structure on median spaces discussed in [Chatterji et al. 2010]. One would have to develop an analogue of the contact graph of a cube complex to serve as the underlying hyperbolic space. One must be careful since, e.g., the Baumslag–Solitar group $\mathrm{BS}(1, 2)$ acts properly on a median space but has exponential Dehn function [Gersten 1992] and is thus not a hierarchically hyperbolic space, by Corollary 7.5. On the other hand, if the answer to Question B is positive, one might try to do the same thing for coarse median spaces.

There are a number of other groups and spaces where it is natural to inquire whether or not they are hierarchically hyperbolic. For example:

[2]The first question was partially answered positively in [Hagen and Susse 2016] after this paper was first posted.

Question C (handlebody group). Let H be a compact oriented 3-dimensional genus g handlebody, and let $G_g \leq \mathcal{MCG}(\partial H)$ be the group of isotopy classes of diffeomorphisms of H. Is G_g a hierarchically hyperbolic group?

Question D (graph products). Let G be a (finite) graph product of hierarchically hyperbolic groups. Is G hierarchically hyperbolic?

The answer to Question C is presumably no, while the answer to D is most likely yes. The positive answer to Question D would follow from a strengthened version of Theorem 8.6.

There are other candidate examples of hierarchically hyperbolic spaces. For example, it is natural to ask whether a right-angled Artin group with the syllable-length metric, introduced in [Kim and Koberda 2014], which is analogous to a Teichmüller space with the Weil–Petersson metric, is hierarchically hyperbolic.

As far as the difference between hierarchically hyperbolic *spaces* and *groups* is concerned, we conjecture that the following question has a positive answer:

Question E. Is it true that the fundamental group G of a nongeometric graph manifold is a hierarchically hyperbolic group if and only if G is virtually compact special?

It is known that G as above is virtually compact special if and only if it is chargeless in the sense of [Buyalo and Svetlov 2004]; see [Hagen and Przytycki 2015].

There remain a number of open questions about the geometry of hierarchically hyperbolic spaces in general. Theorem 7.3 ensures, via work of Bowditch [2013], that every asymptotic cone of a hierarchically hyperbolic space is a median space; further properties in this direction are established in Section 6. Motivated by combining the main result of [Sisto 2011] on 3-manifold groups with Theorem 10.1, we ask:

Question F. Are any two asymptotic cones of a given hierarchically hyperbolic space bi-Lipschitz equivalent?

The notion of hierarchical quasiconvexity of a subgroup of a hierarchically hyperbolic group $(G, \mathfrak{S})$ generalizes quasiconvexity in word-hyperbolic groups and cubical convex-cocompactness in groups acting geometrically on CAT(0) cube complexes with factor-systems. Another notion of quasiconvexity is *stability*, defined by Durham and Taylor [2015]. This is a quite different notion of quasiconvexity, since stable subgroups are necessarily hyperbolic. In [Durham and Taylor 2015], the authors characterize stable subgroups of the mapping class group; it is reasonable to ask for a generalization of their results to hierarchically hyperbolic groups.

Many hierarchically hyperbolic spaces admit multiple hierarchically hyperbolic structures. However, as discussed in [Behrstock et al. 2017b], a CAT(0) cube complex with a factor-system has a "minimal" factor-system, i.e., one that is contained in all other factor systems. In this direction, it is natural to ask whether a hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$ admits a hierarchically hyperbolic structure that is canonical in some way.

Recent developments. Since we posted the first version of this paper, there has been further progress on the theory of HHS and its applications.

More examples of HHS/HHG are now available, including a large class of CAT(0) cubical groups [Hagen and Susse 2016], "small-cancellation" quotients of HHGs [Behrstock et al. 2017a], and separating curve graphs of surfaces [Vokes 2017]. It was also recently shown by Spriano that hyperbolic spaces/groups admit alternate HHS structures that can be constructed from an arbitrary fixed collection of quasi-convex subspaces/subgroups [Spriano 2017]. Spriano [2018] has proven additional results on modifying hierarchically hyperbolic structures to include prespecified subgroups, under natural conditions. This allows him to prove that a large class of graphs of hierarchically hyperbolic groups are hierarchically hyperbolic. In the latter vein, Berlai and Robbio [2018] have generalized the combination theorem (Theorem 8.6) in this paper, and used this to show that the class of hierarchically hyperbolic groups is closed under taking graph products.

Further developments of the theory include finiteness of the asymptotic dimension (including a quadratic upper bound for mapping class groups) [Behrstock et al. 2017a]; a theory of boundaries generalizing the Gromov boundary of hyperbolic groups [Durham et al. 2017; Mousley 2017; 2018]; proof of the existence of largest acylindrical actions [Abbott et al. 2017]; and a theorem controlling quasiflats (new in both mapping class groups and cubical groups) with many applications including, for instance, a new proof of quasi-isometric rigidity for mapping class groups [Behrstock et al. 2017c]. Mousley and Russell [2018] have recently studied Morse boundaries of hierarchically hyperbolic groups, and Abbott and Behrstock [2018] have established a linear bound on conjugator lengths in hierarchically hyperbolic groups.

We stress that the present paper is foundational for almost all of the above developments; the results here are used as tools there.

Organization of the paper. Section 1 contains the full definition of a hierarchically hyperbolic space (and, more generally, a hierarchical space) and some discussion of background. Section 2 contains various basic consequences of the definition, and some tricks that are used repeatedly. In Section 3, we prove the realization theorem (Theorem 3.1). In Section 4 we establish the existence of hierarchy paths (Theorem 4.4) and the distance formula (Theorem 4.5). Section 5 is devoted to hierarchical quasiconvexity and product regions, and Section 6 to coarse convex hulls and relatively hierarchically hyperbolic spaces. The coarse median property and its consequences are detailed in Section 7. The combination theorems for trees of spaces, graphs of groups, and products are proved in Section 8, and groups hyperbolic relative to HHG are studied in Section 9. This is applied to 3-manifolds in Section 10. Finally, in Section 11, we prove that mapping class groups are hierarchically hyperbolic.

1. The main definition and background on hierarchically hyperbolic spaces

1A. *The axioms.* We begin by defining a hierarchically hyperbolic space. We will work in the context of a *q-quasigeodesic space*, $\mathcal{X}$, i.e., a metric space where any two points can be connected by a (q,q)-quasigeodesic. Obviously, if $\mathcal{X}$ is a geodesic space, then it is a quasigeodesic space. Most of the examples we are interested in are geodesic spaces, but in order to construct hierarchically hyperbolic structures on naturally occurring subspaces of hierarchically hyperbolic spaces, we must work in the slightly more general setting of quasigeodesic spaces.

Definition 1.1 (hierarchically hyperbolic space). The q-quasigeodesic space $(\mathcal{X},\mathsf{d}_{\mathcal{X}})$ is a *hierarchically hyperbolic space* if there exists $\delta \geq 0$, an index set $\mathfrak{S}$, and a set $\{\mathcal{C}W : W \in \mathfrak{S}\}$ of δ-hyperbolic spaces $(\mathcal{C}U, \mathsf{d}_U)$, such that the following conditions are satisfied:

(1) *Projections.* There is a set
$$\{\pi_W : \mathcal{X} \to 2^{\mathcal{C}W} \mid W \in \mathfrak{S}\}$$
of *projections* sending points in $\mathcal{X}$ to nonempty sets of diameter bounded by some $\xi \geq 0$ in the various $\mathcal{C}W \in \mathfrak{S}$. Moreover, there exists K so that for all $W \in \mathfrak{S}$, the coarse map π_W is (K,K)-coarsely Lipschitz and $\pi_W(\mathcal{X})$ is K-quasiconvex in $\mathcal{C}W$.

(2) *Nesting.* $\mathfrak{S}$ is equipped with a partial order $\sqsubseteq$, and either $\mathfrak{S} = \varnothing$ or $\mathfrak{S}$ contains a unique $\sqsubseteq$-maximal element; when $V \sqsubseteq W$, we say V is *nested* in W. (We emphasize that $W \sqsubseteq W$ for all $W \in \mathfrak{S}$.) For each $W \in \mathfrak{S}$, we denote by $\mathfrak{S}_W$ the set of $V \in \mathfrak{S}$ such that $V \sqsubseteq W$. Moreover, for all $V, W \in \mathfrak{S}$ with $V \sqsubsetneq W$ there is a specified nonempty subset $\rho^V_W \subset \mathcal{C}W$ with $\mathrm{diam}_{\mathcal{C}W}(\rho^V_W) \leq \xi$. There is also a *projection* $\rho^W_V : \mathcal{C}W \to 2^{\mathcal{C}V}$. (The similarity in notation is justified by viewing ρ^V_W as a coarsely constant map $\mathcal{C}V \to 2^{\mathcal{C}W}$.)

(3) *Orthogonality.* $\mathfrak{S}$ has a symmetric and antireflexive relation called *orthogonality*: we write $V \perp W$ when V, W are orthogonal. Also, whenever $V \sqsubseteq W$ and $W \perp U$, we require that $V \perp U$. We require that for each $T \in \mathfrak{S}$ and each $U \in \mathfrak{S}_T$ for which $\{V \in \mathfrak{S}_T \mid V \perp U\} \neq \varnothing$, there exists $W \in \mathfrak{S}_T - \{T\}$, so that whenever $V \perp U$ and $V \sqsubseteq T$, we have $V \sqsubseteq W$. Finally, if $V \perp W$, then V, W are not $\sqsubseteq$-comparable.

(4) *Transversality and consistency.* If $V, W \in \mathfrak{S}$ are not orthogonal and neither is nested in the other, then we say V, W are *transverse*, denoted $V \pitchfork W$. There

exists $\kappa_0 \geq 0$ such that if $V \pitchfork W$, then there are sets $\rho_W^V \subseteq \mathcal{C}W$ and $\rho_V^W \subseteq \mathcal{C}V$ each of diameter at most ξ and satisfying

$$\min\{\mathsf{d}_W(\pi_W(x), \rho_W^V), \mathsf{d}_V(\pi_V(x), \rho_V^W)\} \leq \kappa_0$$

for all $x \in \mathcal{X}$.

For $V, W \in \mathfrak{S}$ satisfying $V \sqsubseteq W$ and for all $x \in \mathcal{X}$, we have

$$\min\{\mathsf{d}_W(\pi_W(x), \rho_W^V), \mathsf{diam}_{\mathcal{C}V}(\pi_V(x) \cup \rho_V^W(\pi_W(x)))\} \leq \kappa_0.$$

The preceding two inequalities are the *consistency inequalities* for points in $\mathcal{X}$.

Finally, if $U \sqsubseteq V$, then $\mathsf{d}_W(\rho_W^U, \rho_W^V) \leq \kappa_0$ whenever $W \in \mathfrak{S}$ satisfies either $V \sqsubsetneq W$ or $V \pitchfork W$ and $W \not\perp U$.

(5) *Finite complexity.* There exists $n \geq 0$, the *complexity* of $\mathcal{X}$ (with respect to $\mathfrak{S}$), so that any set of pairwise-$\sqsubseteq$-comparable elements has cardinality at most n.

(6) *Large links.* There exist $\lambda \geq 1$ and $E \geq \max\{\xi, \kappa_0\}$ such that the following holds. Let $W \in \mathfrak{S}$ and let $x, x' \in \mathcal{X}$. Let $N = \lambda \mathsf{d}_{_W}(\pi_W(x), \pi_W(x')) + \lambda$. Then there exists $\{T_i\}_{i=1,\dots,\lfloor N \rfloor} \subseteq \mathfrak{S}_W - \{W\}$ such that for all $T \in \mathfrak{S}_W - \{W\}$, either $T \in \mathfrak{S}_{T_i}$ for some i, or $\mathsf{d}_T(\pi_T(x), \pi_T(x')) < E$. Also, $\mathsf{d}_W(\pi_W(x), \rho_W^{T_i}) \leq N$ for each i.

(7) *Bounded geodesic image.* There exists $E > 0$ such that for all $W \in \mathfrak{S}$, all $V \in \mathfrak{S}_W - \{W\}$, and all geodesics γ of $\mathcal{C}W$, either $\mathsf{diam}_{\mathcal{C}V}(\rho_V^W(\gamma)) \leq E$ or $\gamma \cap \mathcal{N}_E(\rho_W^V) \neq \varnothing$.

(8) *Partial realization.* There exists a constant α with the following property. Let $\{V_j\}$ be a family of pairwise orthogonal elements of $\mathfrak{S}$, and let $p_j \in \pi_{V_j}(\mathcal{X}) \subseteq \mathcal{C}V_j$. Then there exists $x \in \mathcal{X}$ so that
   - $\mathsf{d}_{V_j}(x, p_j) \leq \alpha$ for all j,
   - for each j and each $V \in \mathfrak{S}$ with $V_j \sqsubseteq V$, we have $\mathsf{d}_V(x, \rho_V^{V_j}) \leq \alpha$, and
   - if $W \pitchfork V_j$ for some j, then $\mathsf{d}_W(x, \rho_W^{V_j}) \leq \alpha$.

(9) *Uniqueness.* For each $\kappa \geq 0$, there exists $\theta_u = \theta_u(\kappa)$ such that if $x, y \in \mathcal{X}$ and $\mathsf{d}_{\mathcal{X}}(x, y) \geq \theta_u$, then there exists $V \in \mathfrak{S}$ such that $\mathsf{d}_V(x, y) \geq \kappa$.

We say that the q-quasigeodesic metric spaces $\{\mathcal{X}_i\}$ are *uniformly hierarchically hyperbolic* if each $\mathcal{X}_i$ satisfies the axioms above and all constants, including the complexities, can be chosen uniformly. We often refer to $\mathfrak{S}$, together with the nesting and orthogonality relations, and the projections as a *hierarchically hyperbolic structure* for the space $\mathcal{X}$. Observe that $\mathcal{X}$ is hierarchically hyperbolic with respect to $\mathfrak{S} = \varnothing$, i.e., hierarchically hyperbolic of complexity 0, if and only if $\mathcal{X}$ is bounded. Similarly, $\mathcal{X}$ is hierarchically hyperbolic of complexity 1 with respect to $\mathfrak{S} = \{\mathcal{X}\}$, if and only if $\mathcal{X}$ is hyperbolic.

Notation 1.2. Where it will not cause confusion, given $U \in \mathfrak{S}$, we will often suppress the projection map π_U when writing distances in $\mathcal{C}U$, i.e., given $x, y \in \mathcal{X}$ and $p \in \mathcal{C}U$ we write $\mathsf{d}_U(x, y)$ for $\mathsf{d}_U(\pi_U(x), \pi_U(y))$ and $\mathsf{d}_U(x, p)$ for $\mathsf{d}_U(\pi_U(x), p)$. Note that when we measure distance between a pair of sets (typically both of bounded diameter) we are taking the minimum distance between the two sets. Given $A \subset \mathcal{X}$ and $U \in \mathfrak{S}$ we let $\pi_U(A)$ denote $\cup_{a\in A}\pi_U(a)$.

Remark 1.3 (surjectivity of projections). In all of the motivating examples, and in most applications, the maps π_U are uniformly coarsely surjective.

One can always replace each $\mathcal{C}U$ with a thickening of $\pi_U(\mathcal{X})$, and hence make each π_U coarsely surjective. This is first discussed in [Durham et al. 2017], where this procedure gets used; the resulting spaces are termed *normalized* hierarchically hyperbolic spaces.

More precisely, since each $\pi_U(\mathcal{X})$ is K-quasiconvex, the subset $\mathcal{C}U_{\text{norm}}$ of $\mathcal{C}U$ consisting of all geodesics that start and end in $\pi_U(\mathcal{X})$ is uniformly quasiconvex, is a (uniformly) hyperbolic geodesic metric space, and uniformly coarsely coincides with $\pi_U(\mathcal{X})$. (This "quasiconvex hull" procedure is discussed in more detail in Section 6.) Hence we can endow $\mathcal{X}$ with a slightly different, normalized, hierarchically hyperbolic structure. Indeed, the index set is still $\mathfrak{S}$, each $\mathcal{C}U$ is replaced by $\mathcal{C}U_{\text{norm}}$, and the maps π_U remain unchanged (but are now coarsely surjective). Given $U, V \in \mathfrak{S}$ such that ρ^U_V is defined, we replace ρ^U_V (viewed as a coarse map $\mathcal{C}U \to \mathcal{C}V$) with the composition $p_V \circ \rho^U_V$, where $p_V : \mathcal{C}V \to \mathcal{C}U_{\text{norm}}$ is the coarse closest-point projection.

Remark 1.4 (surjectivity/quasiconvexity of projections in the extant applications). In the motivating examples (mapping class groups, Teichmüller space, virtually special groups, hyperbolic spaces, etc.), the projections π_U are uniformly coarsely surjective, but it is convenient to relax that requirement. As is evident from Theorem 3.1 and the key Lemma 2.6, the appropriate relaxation of coarse surjectivity is the requirement, from Definition 1.1.(1), that each $\pi_U(\mathcal{X})$ be uniformly quasiconvex in $\mathcal{C}U$.

In a few other places in the literature, this is not spelled out, but in each case where an issue arises, it does not affect the arguments in question. In the interest of clarity, we now summarize this as follows:

- In [Durham et al. 2017, p. 4, p. 19], the authors establish a standing assumption that they are working with normalized HHSs — each π_U is uniformly coarsely surjective. In view of Remark 1.3 (or [Durham et al. 2017, Proposition 1.16]), the results about normalized HHSs can be promoted to corresponding statements about general HHSs.
- In [Behrstock et al. 2017a], Remark 1.3 allows one to assume that the HHSs in question are normalized. However, there are three places where a new

HHS is constructed from an old one, and one must observe that in each of these cases, the new projections have quasiconvex image. In [Behrstock et al. 2017a, Proposition 2.4], this holds just because the projections used in the new HHS structure coincide with those used in the old HHS structure, so quasiconvexity persists. In Proposition 6.14 and Theorem 6.2 of [Behrstock et al. 2017a], the projections in the new HHS structures are of two types: they either coincide with projections from the old HHS structures, and thus have quasiconvex images, or they are surjective by construction.

Remark 1.5 (large link function). It appears as though there is no actual need to require in Definition 1.1.(6) that N depend linearly on $\mathsf{d}_W(x, x')$. Instead, we could have hypothesized that for any $C \geq 0$, there exists $N(C)$ so that the statement of the axiom holds with $N = N(C)$ whenever $\mathsf{d}_W(x, x') \leq C$. However, one could deduce from this and the rest of the axioms that $N(C)$ grows linearly in C, so we have elected to simply build linearity into the definition.

Remark 1.6 (summary of constants). Each hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$ is associated with a collection of constants often, as above, denoted δ, ξ, n, κ_0, E, θ_u, and K, where

(1) $\mathcal{C}U$ is δ-hyperbolic for each $U \in \mathfrak{S}$,

(2) each π_U has image of diameter at most ξ and each π_U is (K, K)-coarsely Lipschitz, $\pi_U(\mathcal{X})$ is K-quasiconvex in $\mathcal{C}U$, and each ρ_V^U has (image of) diameter at most ξ,

(3) for each $x \in \mathcal{X}$, the tuple $(\pi_U(x))_{U \in \mathfrak{S}}$ is κ_0-consistent,

(4) E is the larger of the constants from the bounded geodesic image axiom and the large link axiom.

Whenever working in a fixed hierarchically hyperbolic space, we use the above notation freely. We can, and shall, assume that $E \geq q$, $E \geq \delta$, $E \geq \xi$, $E \geq \kappa_0$, $E \geq K$, and $E \geq \alpha$.

Remark 1.7. We note that in Definition 1.1.(1), the assumption that the projections are Lipschitz can be replaced by the weaker assumption that there is a proper function of the projected distance which is a lower bound for the distance in the space $\mathcal{X}$. From this weaker assumption, the fact that the projections are actually coarsely Lipschitz then follows from the fact that we assume $\mathcal{X}$ to be quasi-geodesic. Since the Lipschitz hypothesis is cleaner to state and, in practice, fairly easy to verify, we just remark on this for those that might find this fact useful in proving that more exotic spaces are hierarchically hyperbolic.

1B. ***Comparison to the definition in [Behrstock et al. 2017b].*** Definition 1.1 is very similar to the definition of a hierarchically hyperbolic space given in [Behrstock et al. 2017b], with the following differences:

(1) The existence of *hierarchy paths* and the *distance formula* were stated as axioms in [Behrstock et al. 2017b]; below, we deduce them from the other axioms. Similarly, the below *realization theorem* was formerly an axiom, but has been replaced by the (weaker) partial realization axiom.

(2) We now require $\mathcal{X}$ to be a quasigeodesic space. In [Behrstock et al. 2017b], this follows from the existence of hierarchy paths, which was an axiom there.

(3) We now require the projections $\pi_U : \mathcal{X} \to \mathcal{C}U$ to be coarsely Lipschitz; although this requirement was not imposed explicitly in [Behrstock et al. 2017b], it follows from the distance formula, which was an axiom there.

(4) In [Behrstock et al. 2017b], there were five consistency inequalities; there are two in Definition 1.1.(4). The last three inequalities in the definition from [Behrstock et al. 2017b] follow from Proposition 1.8 below. (Essentially, the partial realization axiom has replaced part of the old consistency axiom.)

(5) In Definition 1.1.(4), we require that, if $U \sqsubseteq V$, then $\mathsf{d}_W(\rho^U_W, \rho^V_W) \leq \kappa_0$ whenever $W \in \mathfrak{S}$ satisfies either $V \sqsubsetneq W$ or $V \pitchfork W$ and $W \not\perp U$. In the context of [Behrstock et al. 2017b], this follows by considering the standard product regions constructed using realization (see [Behrstock et al. 2017b, Section 13.1] and Section 5B of the present paper).

Proposition 1.8 (ρ-consistency). *There exists κ_1 so that the following holds. Suppose that $U, V, W \in \mathfrak{S}$ satisfy both of the following conditions: $U \sqsubsetneq V$ or $U \pitchfork V$; and $U \sqsubsetneq W$ or $U \pitchfork W$. Then, if $V \pitchfork W$,*

$$\min\{\mathsf{d}_W(\rho^U_W, \rho^V_W), \mathsf{d}_V(\rho^U_V, \rho^W_V)\} \leq \kappa_1$$

and if $V \sqsubsetneq W$, then

$$\min\{\mathsf{d}_W(\rho^U_W, \rho^V_W), \mathrm{diam}_{\mathcal{C}V}(\rho^U_V \cup \rho^W_V(\rho^U_W))\} \leq \kappa_1.$$

Proof. Suppose that $U \sqsubsetneq V$ or $U \pitchfork V$ and $U \sqsubsetneq W$ or $U \pitchfork W$. Suppose that $V \pitchfork W$ or $V \sqsubseteq W$. Choose $p \in \pi_U(\mathcal{X})$. There is a uniform α so that partial realization (Definition 1.1.(8)) provides $x \in \mathcal{X}$ so that $\mathsf{d}_U(x, p) \leq \alpha$ and $\mathsf{d}_T(x, \rho^U_T) \leq \alpha$ whenever ρ^U_T is defined and coarsely constant. In particular, $\mathsf{d}_V(x, \rho^U_V) \leq \alpha$ and $\mathsf{d}_W(x, \rho^U_W) \leq \alpha$. The claim now follows from Definition 1.1.(4), with $\kappa_1 = \kappa_0 + \alpha$. $\square$

In view of the discussion above, we have:

Proposition 1.9. *The pair $(\mathcal{X}, \mathfrak{S})$ satisfies Definition 1.1 if and only if it is hierarchically hyperbolic in the sense of* [*Behrstock et al. 2017b*].

In particular, as observed in [Behrstock et al. 2017b]:

Proposition 1.10. *If $(\mathcal{X}, \mathfrak{S})$ is a hierarchically hyperbolic space, and $\mathcal{X}'$ is a quasigeodesic space quasi-isometric to $\mathcal{X}$, then there is a hierarchically hyperbolic space $(\mathcal{X}', \mathfrak{S})$.*

1C. ***A variant on the axioms.*** Here we introduce two slightly simpler versions of the HHS axioms and show that in the case, as in most situations which arise naturally, that the projections are coarsely surjective, it suffices to verify the simpler axioms.

The following is a subset of the nesting axiom; here we remove the definition of the projection map $\rho_V^W : \mathcal{C}W \to 2^{\mathcal{C}V}$ in the case $V \sqsubsetneq W$.

Definition 1.1.(2)$'$ (nesting variant). $\mathfrak{S}$ is equipped with a partial order $\sqsubseteq$, and either $\mathfrak{S} = \varnothing$ or $\mathfrak{S}$ contains a unique $\sqsubseteq$-maximal element; when $V \sqsubseteq W$, we say V is *nested* in W. We require that $W \sqsubseteq W$ for all $W \in \mathfrak{S}$. For each $W \in \mathfrak{S}$, we denote by $\mathfrak{S}_W$ the set of $V \in \mathfrak{S}$ such that $V \sqsubseteq W$. Moreover, for all $V, W \in \mathfrak{S}$ with $V \sqsubsetneq W$ there is a specified subset $\rho_W^V \subset \mathcal{C}W$ with $\mathsf{diam}_{\mathcal{C}W}(\rho_W^V) \le \xi$.

The following is a subset of the transversality and consistency axiom.

Definition 1.1.(4)$'$ (transversality). If $V, W \in \mathfrak{S}$ are not orthogonal and neither is nested in the other, then we say V, W are *transverse*, denoted $V \pitchfork W$. There exists $\kappa_0 \ge 0$ such that if $V \pitchfork W$, then there are sets $\rho_W^V \subseteq \mathcal{C}W$ and $\rho_V^W \subseteq \mathcal{C}V$ each of diameter at most ξ and satisfying

$$\min\{\mathsf{d}_W(\pi_W(x), \rho_W^V), \mathsf{d}_V(\pi_V(x), \rho_V^W)\} \le \kappa_0$$

for all $x \in \mathcal{X}$.

Finally, if $U \sqsubseteq V$, then $\mathsf{d}_W(\rho_W^U, \rho_W^V) \le \kappa_0$ whenever $W \in \mathfrak{S}$ satisfies either $V \sqsubsetneq W$ or $V \pitchfork W$ and $W \not\perp U$.

The following is a variant of the bounded geodesic image axiom:

Definition 1.1.(7)$'$ (bounded geodesic image variant). Suppose that $x, y \in X$ and $V \sqsubsetneq W$ have the property that there exists a geodesic from $\pi_W(x)$ to $\pi_W(y)$ which stays $(E + 2\delta)$-far from ρ_W^V. Then $\mathsf{d}_V(x, y) \le E$.

Proposition 1.11. *Given a quasigeodesic space $\mathcal{X}$ and an index set $\mathfrak{S}$, then $(\mathcal{X}, \mathfrak{S})$ is an HHS if it satisfies the axioms of Definition 1.1 with the following changes:*

- *Replace Definition 1.1.(2) by Definition 1.1.(2)$'$.*
- *Replace Definition 1.1.(4) by Definition 1.1.(4)$'$.*
- *Replace Definition 1.1.(7) by Definition 1.1.(7)$'$.*
- *Assume that for each $\mathcal{C}U$ the map π_U is uniformly coarsely surjective.*

Proof. To verify Definition 1.1.(2), for each $V, W \in \mathfrak{S}$ with $V \subsetneq W$, we define a map $\rho_V^W : \mathcal{C}W \to 2^{\mathcal{C}V}$ as follows. If $p \in \mathcal{C}W - \mathcal{N}_E(\rho_W^V)$, then let $\rho_V^W(p) = \pi_V(x)$ for some $x \in \mathcal{X}$ with $\pi_W(x)$ (uniformly) coarsely coinciding with p. Since p does not lie E-close to ρ_W^V, this definition is coarsely independent of x by Definition 1.1.(7)$'$. On $\mathcal{N}_E(\rho_W^V)$, we define ρ_V^W arbitrarily. By definition, the resulting map satisfies Definition 1.1.(4). Moreover, coarse surjectivity of π_W and Definition 1.1.(7)$'$ ensure that Definition 1.1.(7) holds. The rest of the axioms hold by hypothesis. $\square$

Remark 1.12. The definition of an HHS provided by Proposition 1.11 is convenient because it does not require one to define certain maps between hyperbolic spaces: Definition 1.1.(2)$'$ is strictly weaker than Definition 1.1.(2). On the other hand, it is often convenient to work with HHS in which some of the projections π_U are not coarsely surjective; for example, this simplifies the proof that hierarchically quasiconvex subspaces inherit HHS structures in Proposition 5.6. Hence we have included both definitions.

In practice, we almost always apply consistency and bounded geodesic image in concert, which involves applying bounded geodesic image to geodesics of $\mathcal{C}W$ joining points in $\pi_W(\mathcal{X})$. Accordingly, Definition 1.1.(7)$'$ is motivated by the following easy observation:

Proposition 1.13. *Let $(\mathcal{X}, \mathfrak{S})$ be an HHS. Then the conclusion of Definition 1.1.(7)$'$ holds for all $x, y \in \mathcal{X}$ and $V, W \in \mathfrak{S}$ with $V \subsetneq W$.*

1D. *Hierarchical spaces.* Although most of our focus in this paper is on hierarchically hyperbolic spaces, there are important contexts in which hyperbolicity of the spaces $\mathcal{C}U, U \in \mathfrak{S}$ is not used; notably, this is the case for the realization theorem (Theorem 3.1). Because of the utility of a more general definition in later applications, we now define the following more general notion of a *hierarchical space*; the reader interested only in the applications to the mapping class group, 3-manifolds, cube complexes, etc., may safely ignore this subsection.

Definition 1.14 (hierarchical space). A *hierarchical space* is a pair $(\mathcal{X}, \mathfrak{S})$ as in Definition 1.1, with $\mathcal{X}$ a quasigeodesic space and $\mathfrak{S}$ an index set, where to each $U \in \mathfrak{S}$ we associate a geodesic metric space $\mathcal{C}U$, which we *do not require to be hyperbolic*. As before, there are coarsely Lipschitz projections

$$\pi_U : \mathcal{X} \to \mathcal{C}U$$

and relative projections

$$\rho_V^U : \mathcal{C}U \to \mathcal{C}V$$

whenever U, V are nonorthogonal. We require all statements in the Definition 1.1 to hold, except for hyperbolicity of the $\mathcal{C}U$.

Remark 1.15. Let $\mathcal{X}$ be a quasigeodesic space that is hyperbolic relative to a collection $\mathcal{P}$ of subspaces. Then $\mathcal{X}$ has a hierarchical space structure: the associated spaces onto which we project are the various $\mathcal{P}$, together with the space $\widehat{\mathcal{X}}$ obtained by coning off the elements of $\mathcal{P}$ in $\mathcal{X}$. When the elements of $\mathcal{P}$ are themselves hierarchically hyperbolic, we obtain a hierarchically hyperbolic structure on $\mathcal{X}$ (see Section 9). Otherwise, the hierarchical structure need not be hierarchically hyperbolic since $\widehat{\mathcal{X}}$ is the only one of the elements of $\mathfrak{S}$ known to be hyperbolic.

Remark 1.16. Other than hierarchically hyperbolic spaces, we are mainly interested in hierarchical spaces $(\mathcal{X}, \mathfrak{S})$ where for all $U \in \mathfrak{S}$, except possibly when U is $\sqsubseteq$-minimal, we have that $\mathcal{C}U$ is hyperbolic. This is the case, for example, in relatively hyperbolic spaces.

1E. *Consistency and partial realization points.* The following definitions, which abstract the consistency inequalities from Definition 1.1.(4) and the partial realization axiom, Definition 1.1.(8), play important roles throughout our discussion. We will consider this topic in depth in Section 3.

Definition 1.17 (consistent). Fix $\kappa \geq 0$ and let $\vec{b} \in \prod_{U\in\mathfrak{S}} 2^{\mathcal{C}U}$ be a tuple such that for each $U \in \mathfrak{S}$, the coordinate b_U is a subset of $\mathcal{C}U$ with $\mathsf{diam}_{\mathcal{C}U}(b_U) \leq \kappa$. The tuple $\vec{b}$ is *κ-admissible* if $\mathsf{d}_U(b_U, \pi_U(\mathcal{X})) \leq \kappa$ for all $U \in \mathfrak{S}$. The κ-admissible tuple $\vec{b}$ is *κ-consistent* if, whenever $V \pitchfork W$,

$$\min\{\mathsf{d}_W(b_W, \rho_W^V), \mathsf{d}_V(b_V, \rho_V^W)\} \leq \kappa$$

and whenever $V \sqsubseteq W$,

$$\min\{\mathsf{d}_W(b_W, \rho_W^V), \mathsf{diam}_{\mathcal{C}V}(b_V \cup \rho_V^W(b_W))\} \leq \kappa.$$

In typical situations, where the maps π_U are uniformly coarsely surjective, up to a uniform enlargement of E, all tuples are admissible, so verifying consistency amounts to verifying the second condition.

Definition 1.18 (partial realization point). Given $\theta \geq 0$ and a κ-consistent tuple $\vec{b}$, we say that $x \in \mathcal{X}$ is a *θ-partial realization point* for $\{V_j\} \subseteq \mathfrak{S}$ if

(1) $\mathsf{d}_{V_j}(x, b_{V_j}) \leq \theta$ for all j,

(2) for all j, we have $\mathsf{d}_V(x, \rho_V^{V_j}) \leq \theta$ for any $V \in \mathfrak{S}$ with $V_j \sqsubseteq V$, and

(3) for all W such that $W \pitchfork V_j$ for some j, we have $\mathsf{d}_W(x, \rho_W^{V_j}) \leq \theta$.

Observe that if $\vec{b}$ is consistent and $\{V_j\}$ is a set of pairwise-orthogonal elements, then partial realization (Definition 1.1.(8)) provides a partial realization point, because of admissibility.

1F. *Levels.* The following definition is very useful for proving statements about hierarchically hyperbolic spaces inductively. Although it is natural, and sometimes useful, to induct on complexity, it is often better to induct on the level:

Definition 1.19 (level). Let $(X, \mathfrak{S})$ be hierarchically hyperbolic. The *level* ℓ_U of $U \in \mathfrak{S}$ is defined inductively as follows. If U is $\sqsubseteq$-minimal then we say that its level is 1. The element U has level $k+1$ if k is the maximal integer such that there exists $V \sqsubseteq U$ with $\ell_V = k$ and $V \neq U$. Given $U \in \mathfrak{S}$, for each $\ell \geq 0$, let $\mathfrak{S}_U^\ell$ be the set of $V \sqsubseteq U$ with $\ell_U - \ell_V \leq \ell$ and let $\mathfrak{T}_U^\ell = \mathfrak{S}_U^\ell - \mathfrak{S}_U^{\ell-1}$.

1G. *Maps between hierarchically hyperbolic spaces.*

Definition 1.20 (hieromorphism). Let $(\mathcal{X}, \mathfrak{S})$ and $(\mathcal{X}', \mathfrak{S}')$ be hierarchically hyperbolic structures on the spaces $\mathcal{X}$, $\mathcal{X}'$ respectively. A *hieromorphism*, consists of a map $f : \mathcal{X} \to \mathcal{X}'$, an injective map $f^\diamondsuit : \mathfrak{S} \to \mathfrak{S}'$ preserving nesting, transversality, and orthogonality, and, for each $U \in \mathfrak{S}$, a map $f^*(U) : \mathcal{C}U \to \mathcal{C}(f^\diamondsuit(U))$ which is a quasi-isometric embedding where the constants are uniform over all elements of $\mathfrak{S}$ and for which the following two diagrams coarsely commute (with uniform constants) for all $U, V \in \mathfrak{S}$ with $U \sqsubsetneq V$ or $U \pitchfork V$:

$$\begin{tikzcd}
\mathcal{X} \arrow[r, "f"] \arrow[d, "\pi_U"'] & \mathcal{X}' \arrow[d, "\pi_{f^\diamondsuit(U)}"] \\
\mathcal{C}(U) \arrow[r, "f^*(U)"] & \mathcal{C}(f^\diamondsuit(U))
\end{tikzcd}$$

and

$$\begin{tikzcd}
\mathcal{C}U \arrow[r, "f^*(U)"] \arrow[d, "\rho^U_V"'] & \mathcal{C}(f^\diamondsuit(U)) \arrow[d, "\rho^{f^\diamondsuit(U)}_{f^\diamondsuit(V)}"] \\
\mathcal{C}V \arrow[r, "f^*(V)"] & \mathcal{C}(f^\diamondsuit(V))
\end{tikzcd}$$

where $\rho^U_V : \mathcal{C}U \to \mathcal{C}V$ is the projection from Definition 1.1. As the functions f, $f^*(U)$, and $f^\diamondsuit$ all have distinct domains, it is often clear from the context which is the relevant map; in that case we periodically abuse notation slightly by dropping the superscripts and just calling all of the maps f.

Definition 1.21 (automorphism, hierarchically hyperbolic group). An *automorphism* of the hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$ is a hieromorphism $f : (\mathcal{X}, \mathfrak{S}) \to (\mathcal{X}, \mathfrak{S})$ such that $f^\diamondsuit$ is bijective and each $f^*(U)$ is an isometry; hence $f : \mathcal{X} \to \mathcal{X}$ is a uniform quasi-isometry by the distance formula (Theorem 4.5).

Note that the composition of two automorphisms is again an automorphism. We say that the automorphisms f, f' are *equivalent* if $f^\diamondsuit = (f')^\diamondsuit$ and $f^*(U) = (f')^*(U)$ for each $U \in \mathfrak{S}$. In particular, equivalent automorphisms give equivalent quasi-isometries. Given an automorphism f, any quasi-inverse $\bar{f}$ of f is an automorphism with $\bar{f}^\diamondsuit = (f^\diamondsuit)^{-1}$ and each $\bar{f}^*(U) = f^*(U)^{-1}$. Hence the set of equivalence classes of automorphisms forms a group, the *full automorphism group* of $(\mathcal{X}, \mathfrak{S})$, denoted $\mathrm{Aut}(\mathfrak{S})$.

The finitely generated group G is *hierarchically hyperbolic* if there exists a hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$ and an action $G \to \mathrm{Aut}(\mathfrak{S})$ so that the uniform quasiaction of G on $\mathcal{X}$ is metrically proper and cobounded and $\mathfrak{S}$ contains finitely many G-orbits. Note that if G is hierarchically hyperbolic by virtue of its action on the hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$, then $(G, \mathfrak{S})$ is a hierarchically hyperbolic structure with respect to any word-metric on G; for any $U \in \mathfrak{S}$ the projection is the composition of the projection $\mathcal{X} \to \mathcal{C}U$ with a G-equivariant quasi-isometry $G \to \mathcal{X}$. In this case, $(G, \mathfrak{S})$ (with the implicit hyperbolic spaces and projections) is a *hierarchically hyperbolic group structure*.

Definition 1.22 (equivariant hieromorphism). Let $(\mathcal{X}, \mathfrak{S})$ and $(\mathcal{X}', \mathfrak{S}')$ be hierarchically hyperbolic spaces and consider actions $G \to \mathrm{Aut}(\mathfrak{S})$ and $G' \to \mathrm{Aut}(\mathfrak{S}')$. For each $g \in G$, let $(f_g, f_g^\diamondsuit, \{f_g^*(U)\})$ denote its image in $\mathrm{Aut}(\mathfrak{S})$, and for each $g' \in G'$, let $(f_{g'}, f_{g'}^\diamondsuit, \{f_{g'}^*(U)\})$ denote its image in $\mathrm{Aut}(\mathfrak{S}')$. Let $\phi : G \to G'$ be a homomorphism. The hieromorphism $(f, f^\diamondsuit, \{f^*(U)\}) : (\mathcal{X}, \mathfrak{S}) \to (\mathcal{X}', \mathfrak{S}')$ is *ϕ-equivariant* if for all $g \in G$ and $U \in \mathfrak{S}$, we have $f^\diamondsuit(f_g^\diamondsuit(U)) = f_{\phi(g)}^\diamondsuit(f^\diamondsuit(U))$ and the following diagram (uniformly) coarsely commutes:

$$\begin{tikzcd}
\mathcal{C}U \arrow[r, "f^*(U)"] \arrow[d, "f_g^*(U)"'] & \mathcal{C}(f^\diamondsuit(U)) \arrow[d, "f_{\phi(g)}^*(U))"] \\
\mathcal{C}(f_g^\diamondsuit(U)) \arrow[r, "f^*(f_g^\diamondsuit(U))"] & \mathcal{C}(f^\diamondsuit(f_g^\diamondsuit(U)))
\end{tikzcd}$$

In this case, $f : \mathcal{X} \to \mathcal{X}'$ is (uniformly) coarsely ϕ-equivariant in the usual sense. Also, we note for the reader that $f_g^\diamondsuit : \mathfrak{S}\circlearrowleft$, while $f_{\phi(g)}^\diamondsuit : \mathfrak{S}'\circlearrowleft$, and $f^\diamondsuit : \mathfrak{S} \to \mathfrak{S}'$.

2. Tools for studying hierarchically hyperbolic spaces

We now collect some basic consequences of the axioms that are used repeatedly throughout the paper. However, this section need not all be read in advance. Indeed, the reader should feel free to skip this section on a first reading and return to it later when necessary. Throughout this section, we work in a hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$.

2A. *Handy basic consequences of the axioms.*

Lemma 2.1 ("finite dimension"). *Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space of complexity n and let $U_1, \dots, U_k \in \mathfrak{S}$ be pairwise-orthogonal. Then $k \le n$.*

Proof. By Definition 1.1.(3), there exists $W_1 \in \mathfrak{S}$, not $\sqsubseteq$-maximal, such that $U_2, \dots, U_k \sqsubseteq W_1$. Applying Definition 1.1.(3) inductively yields a sequence $W_{k-1} \sqsubseteq W_{k-2} \sqsubseteq \cdots \sqsubseteq W_1 \sqsubseteq S$ of distinct elements, where S is $\sqsubseteq$-maximal, so that $U_{i-1}, \dots, U_k \sqsubseteq W_i$ for $1 \le i \le k-1$. Hence $k \le n$ by Definition 1.1.(5). $\square$

Lemma 2.2. *There exists χ so that $|\mathfrak{S}'| \le \chi$ whenever $\mathfrak{S}' \subseteq \mathfrak{S}$ does not contain a pair of transverse elements.*

Proof. Let $\mathfrak{S}' \subseteq \mathfrak{S}$ be a collection of pairwise nontransverse elements, and let n be large enough that any collection of pairwise orthogonal (resp. pairwise $\sqsubseteq$-comparable) elements of $\mathfrak{S}$ has cardinality at most n; the complexity provides such an n, by Definition 1.1.(5) and Lemma 2.1. By Ramsey's theorem, there exists N so that if $|\mathfrak{S}'| > N$ then $\mathfrak{S}'$ contains a collection of elements, of cardinality at least $n+1$, whose elements are either pairwise orthogonal or pairwise $\sqsubseteq$-comparable. Hence, $|\mathfrak{S}'| \le N$. □

Lemma 2.3 (consistency for pairs of points). *Let $x, y \in \mathcal{X}$ and $V, W \in \mathfrak{S}$ satisfy $V \pitchfork W$ and $\mathsf{d}_V(x, y), \mathsf{d}_W(x, y) > 10E$. Then, up to exchanging V and W, we have $\mathsf{d}_V(x, \rho^W_V) \le E$ and $\mathsf{d}_W(y, \rho^V_W) \le E$.*

Proof. If $\mathsf{d}_V(x, \rho^W_V) > E$, then Definition 1.1.(4) implies $\mathsf{d}_W(x, \rho^V_W) \le E$. Then, either $\mathsf{d}_W(y, \rho^V_W) \le 9E$, in which case $\mathsf{d}_W(x, y) \le 10E$, which is a contradiction, or $\mathsf{d}_W(y, \rho^V_W) > E$, in which case consistency implies that $\mathsf{d}_V(y, \rho^W_V) \le E$. □

Corollary 2.4. *For x, y, V, W as in Lemma 2.3, and any $z \in \mathcal{X}$, there exists $U \in \{V, W\}$ such that $\mathsf{d}_U(z, \{x, y\}) \le 10E$.*

Proof. By Lemma 2.3, we may assume that $\mathsf{d}_V(x, \rho^W_V), \mathsf{d}_W(y, \rho^V_W) \le E$. Suppose that $\mathsf{d}_W(z, \{x, y\}) > 10E$. Then $\mathsf{d}_W(z, \rho^V_W) > 9E$, so that, by consistency, $\mathsf{d}_V(z, \rho^W_V) \le E$, whence $\mathsf{d}_V(z, x) \le 2E$. □

The following is needed for Theorem 3.1 and in [Durham et al. 2017].

Lemma 2.5 (passing large projections up the $\sqsubseteq$-lattice). *For every $C \ge 0$ there exists N with the following property. Let $V \in \mathfrak{S}$, let $x, y \in \mathcal{X}$, and let $\{S_i\}_{i=1}^N \subseteq \mathfrak{S}_V$ be distinct and satisfy $\mathsf{d}_{S_i}(x, y) \ge E$. Then there exists $S \in \mathfrak{S}_V$ and i so that $S_i \sqsubsetneq S$ and $\mathsf{d}_S(x, y) \ge C$.*

Proof. The proof is by induction on the level k of a $\sqsubseteq$-minimal $S \in \mathfrak{S}_V$ into which each S_i is nested. The base case $k = 1$ is empty.

Suppose that the statement holds for a given $N = N(k)$ when the level of S is at most k. Suppose further that $|\{S_i\}| \ge N(k+1)$ (where $N(k+1)$ is a constant much larger than $N(k)$ that will be determined shortly) and there exists a $\sqsubseteq$-minimal $S \in \mathfrak{S}_V$ of level $k+1$ into which each S_i is nested. There are two cases.

If $\mathsf{d}_{CS}(x, y) \ge C$, we are done. If not, then the large link axiom (Definition 1.1.(6)) yields $K = K(C)$ and $T_1, \dots, T_K$, each properly nested into S (and hence of level less than $k+1$), so that any given S_i is nested into some T_j. In particular, if $N(k+1) \ge KN(k)$, there exists a j so that at least $N(k)$ elements of $\{S_i\}$ are nested into T_j. By the induction hypothesis and the finite complexity axiom (Definition 1.1.(5)), we are done. □

The next lemma is used in the proof of Proposition 4.12, on which the existence of hierarchy paths (Theorem 4.4) relies. It is used again in Section 7 to construct coarse media.

Lemma 2.6 (centers are consistent). *There exists κ with the following property. Let $x, y, z \in \mathcal{X}$. Let $\vec{b} = (b_W)_{W\in\mathfrak{S}}$ be such that b_W is a point in $\mathcal{C}W$ with the property that there exists a geodesic triangle in $\mathcal{C}W$ with vertices in $\pi_W(x), \pi_W(y), \pi_W(z)$ each of whose sides contains a point within distance δ of b_W. Then $\vec{b}$ is κ-consistent.*

Proof. Recall that for $w \in \{x, y, z\}$ the tuple $(\pi_V(w))_{V\in\mathfrak{S}}$ is E-consistent. Let $U, V \in \mathfrak{S}$ be transverse. Then, by E-consistency, up to exchanging U and V and substituting z for one of x, y, we have $\mathsf{d}_V(x, \rho^U_V), \mathsf{d}_V(y, \rho^U_V) \le E$, so $\mathsf{d}_V(x, y) \le 3E$ (recall that the diameter of ρ^U_V is at most E). Since b_V lies at distance δ from the geodesic joining $\pi_V(x), \pi_V(y)$, we have $\mathsf{d}_V(b_V, \rho^U_V) \le 3E+\delta$, whence the lemma holds with $\kappa = 3E+\delta$.

Suppose now $U \subsetneq V$. If b_V is within distance $10E$ of ρ^U_V, then we are done. Otherwise, up to permuting x, y, z, any geodesic $[\pi_V(x), \pi_V(y)]$ is $5E$-far from ρ^U_V. By consistency of $(\pi_W(x))$, $(\pi_W(y))$ and bounded geodesic image (Definition 1.1.(7)), we have $\mathsf{diam}_U(\rho^V_U(\pi_V(y)) \cup \pi_U(y)) \le E$, $\mathsf{diam}_U(\rho^V_U(b_V \cup \pi_V(y))) \le 10E$ and $\mathsf{d}_U(x, y) \le 10E$. The first inequality and the definition of b_U imply $\mathsf{d}_U(b_U, y) \le 20E$, and taking into account the other inequalities, we get $\mathsf{diam}_U(\rho^V_U(b_V) \cup b_U) \le 100E$.

Moreover, since $\pi_W(\mathcal{X})$ is K-quasiconvex, and b_W lies δ-close to a geodesic starting and ending in $\pi_W(\mathcal{X})$, we see that b_W lies $(K+\delta)$-close to a point in $\pi_W(\mathcal{X})$. Hence, provided our initial choice of E was sufficiently large in terms of the constants from Definition 1.1, $\vec{b}$ is admissible. □

2B. ***Partially ordering sets of maximal relevant elements of $\mathfrak{S}$.*** In this subsection, we describe a construction used several times in this paper, including in the proof of realization (Theorem 3.1), in the construction of hierarchy paths (Theorem 4.4), and in the proof of the distance formula (Theorem 4.5). We expect that this construction will have numerous other applications, as is the case with the corresponding partial ordering in the case of the mapping class group, see for example [Behrstock et al. 2012; Behrstock and Minsky 2011; Clay et al. 2012].

Fix $x \in \mathcal{X}$ and a tuple $\vec{b} \in \prod_{U\in\mathfrak{S}} 2^{\mathrm{im}(\pi_U)}$, where the U-coordinate b_U is a set of diameter at most some fixed $\xi \ge 0$. For example, $\vec{b}$ could be the tuple $(\pi_U(y))$ for some $y \in \mathcal{X}$.

In the remainder of this section, we choose $\kappa \ge 0$ and require that $\vec{b}$ is κ-consistent. (Recall that if $\vec{b}$ is the tuple of projections of a point in $\mathcal{X}$, then $\vec{b}$ is E-consistent.)

Definition 2.7 (relevant). First, fix $\theta \ge 100\max\{\kappa, E\}$. Then $U \in \mathfrak{S}$ is *relevant* (with respect to $x, \vec{b}, \theta$) if $\mathsf{d}_U(x, b_U) > \theta$. Denote by $\mathrm{Rel}(x, \vec{b}, \theta)$ the set of relevant elements.

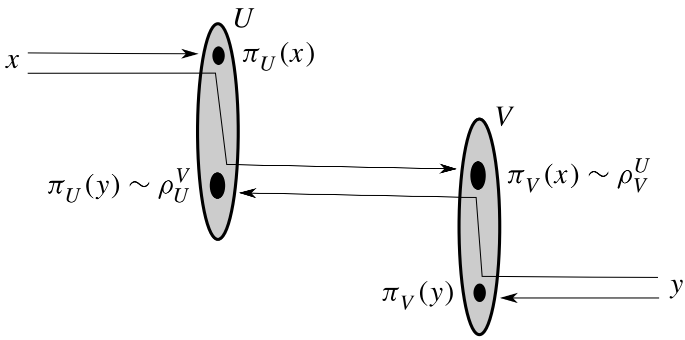

Figure 1. Heuristic picture of $U \prec V$ (for $\vec{b}$ the coordinates of $y \in \mathcal{X}$, for concreteness). The idea is that "on the way" from x to y one "first encounters" U and is forced to change the projection from $\pi_U(x)$ to $\pi_U(y) \sim \rho_U^V$. In doing so the projection to V is not affected.

Let $\mathrm{Rel}_{\max}(x, \vec{b}, \theta)$ be a subset of $\mathrm{Rel}(x, \vec{b}, \theta)$ whose elements are pairwise $\sqsubseteq$-incomparable (for example, they could all be $\sqsubseteq$-maximal in $\mathrm{Rel}(x, \vec{b}, \theta)$, or they could all have the same level). Define a relation $\preceq$ on $\mathrm{Rel}_{\max}(x, \vec{b}, \theta)$ as follows. Given $U, V \in \mathrm{Rel}_{\max}(x, \vec{b}, \theta)$, we have $U \preceq V$ if $U = V$ or if $U \pitchfork V$ and $\mathsf{d}_U(\rho_U^V, b_U) \leq \kappa$. Figure 1 illustrates $U \prec V$.

Proposition 2.8. *The relation $\preceq$ is a partial order. Moreover, either U, V are $\preceq$-comparable or $U \perp V$.*

Proof. Clearly $\preceq$ is reflexive. Antisymmetry follows from Lemma 2.9. Suppose that U, V are $\preceq$-incomparable. If $U \perp V$, we are done, and we cannot have $U \sqsubseteq V$ or $V \sqsubseteq U$, so suppose $U \pitchfork V$. Then, by $\preceq$-incomparability of U, V, we have $\mathsf{d}_U(\rho_U^V, b_U) > \kappa$ and $\mathsf{d}_V(\rho_V^U, b_V) > \kappa$, contradicting κ-consistency of $\vec{b}$. This proves the assertion that transverse elements of $\mathrm{Rel}_{\max}(x, \vec{b}, \theta)$ are $\preceq$-comparable. Finally, transitivity follows from Lemma 2.10. □

Lemma 2.9. *The relation $\preceq$ is antisymmetric.*

Proof. If $U \preceq V$ and $U \neq V$, then $\mathsf{d}_U(b_U, \rho_U^V) \leq \kappa$, so $\mathsf{d}_U(x, \rho_U^V) > \theta - \kappa \geq 99\kappa > E$. Then, $\mathsf{d}_V(x, \rho_V^U) \leq E$, by consistency. Thus $\mathsf{d}_V(b_V, \rho_V^U) > \kappa$, and so, by definition $V \not\preceq U$. □

Lemma 2.10. *The relation $\preceq$ is transitive.*

Proof. Suppose that $U \preceq V \preceq W$. If $U = V$ or $V = W$, then $U \preceq W$, and by Lemma 2.9, we cannot have $U = W$ unless $U = V = W$. Hence suppose $U \pitchfork V$ and $\mathsf{d}_U(\rho_U^V, b_U) \leq \kappa$, while $V \pitchfork W$ and $\mathsf{d}_V(\rho_V^W, b_V) \leq \kappa$. By the definition of $\mathrm{Rel}_{\max}(x, \vec{b}, \theta)$, we have $\mathsf{d}_T(x, b_T) > 100\kappa$ for $T \in \{U, V, W\}$.

We first claim $\mathsf{d}_V(\rho_V^U, \rho_V^W) > 10E$. Indeed, $\mathsf{d}_U(b_U, \rho_U^V) \leq \kappa$, so $\mathsf{d}_U(\rho_U^V, x) \geq 90\kappa$, whence $\mathsf{d}_V(\rho_V^U, x) \leq E \leq \kappa$ by E-consistency of the tuple $(\pi_T(x))_{T \in \mathfrak{S}}$. On the other hand, $\mathsf{d}_V(\rho_V^W, b_V) \le \kappa$, so $\mathsf{d}_V(\rho_V^U, \rho_V^W) > 10E$ as claimed. Hence, by Lemma 2.11, we have $U \pitchfork W$.

Since $\mathsf{diam}(\mathrm{im}(\pi_W)) > 100\kappa$ — and notably, therefore, $\mathsf{d}_W(x, b_W) > 100\kappa$ and $b_W \in \mathrm{im}(\pi_W(\mathcal{X}))$ — partial realization (Definition 1.1.(8)) provides $a \in \mathcal{X}$ satisfying $\mathsf{d}_W(a, \{\rho_W^U, \rho_W^V\}) \ge 10\kappa$.

We thus have $\mathsf{d}_U(a, \rho_U^W) \le E$ by E-consistency of $(\pi_T(a))_{T\in\mathfrak{S}}$, and the same is true if we replace U with V. Hence $\mathsf{d}_V(\rho_V^U, a) > E$, so consistency implies $\mathsf{d}_U(a, \rho_U^V) \le E$. Thus $\mathsf{d}_U(\rho_U^V, \rho_U^W) \le 2E$. Thus $\mathsf{d}_U(b_U, \rho_U^W) \le 2E + \kappa < 10\kappa$, whence $\mathsf{d}_U(x, \rho_U^W) > 50\kappa > E$, so $\mathsf{d}_W(x, \rho_W^U) \le E$ by consistency and the fact that $U \pitchfork W$. It follows that $\mathsf{d}_W(b_W, \rho_W^U) \ge 100\kappa - E > \kappa$, so, again by consistency, $\mathsf{d}_U(b_U, \rho_W^U) \le \kappa$, i.e., $U \preceq W$. □

Lemma 2.11. *Let $U, V, W \in \mathfrak{S}$ be such that all of* $\mathsf{diam}(\mathrm{im}(\pi_U))$, $\mathsf{diam}(\mathrm{im}(\pi_V))$, $\mathsf{diam}(\mathrm{im}(\pi_W))$, *and* $\mathsf{d}_V(\rho_V^U, \rho_V^W)$ *are greater than* $10E$, *and $U \pitchfork V$, $W \pitchfork V$. Suppose moreover that U and W are $\sqsubseteq$-incomparable. Then $U \pitchfork W$.*

Proof. If $U \perp W$, then by the partial realization axiom (Definition 1.1.(8)) and the lower bound on diameters, there exists an E-partial realization point x for $\{U, W\}$ so that

$$\mathsf{d}_U(\rho_U^V, x), \mathsf{d}_W(\rho_W^V, x) > E.$$

This contradicts consistency since $\mathsf{d}_V(\rho_V^U, \rho_V^W) > 10E$; indeed, by consistency $\mathsf{d}_V(\rho_V^U, x) \le E$, $\mathsf{d}_V(\rho_V^W, x) \le E$, i.e., $\mathsf{d}_V(\rho_V^U, \rho_V^W) \le 2E$. Hence $U \pitchfork W$. □

2C. ***Coloring relevant elements.*** In this subsection, the key result is Lemma 2.14, which we will apply in proving the existence of hierarchy paths in Section 4C.

Fix $x, y \in \mathcal{X}$. As above, let $\mathrm{Rel}(x, y, 100E)$ consist of those $V \in \mathfrak{S}$ for which $\mathsf{d}_V(x, y) > 100E$. Recall that, given $U \in \mathfrak{S}$, we denote by $\mathfrak{T}_U^\ell$ the set of $V \in \mathfrak{S}_U$ such that $\ell_U - \ell_V = \ell$. In particular, if $V, V' \in \mathfrak{T}_U^\ell$ and $V \sqsubseteq V'$, then $V = V'$. Let $\mathrm{Rel}_U^\ell(x, y, 100E) = \mathrm{Rel}(x, y, 100E) \cap \mathfrak{T}_U^\ell$, the set of $V \sqsubseteq U$ so that $\mathsf{d}_V(x, y) > 100E$ and $\ell_U - \ell_V = \ell$.

By Proposition 2.8, the relation $\preceq$ on $\mathrm{Rel}_U^\ell(x, y, 100E)$ defined as follows is a partial order: $V \preceq V'$ if either $V = V'$ or $\mathsf{d}_V(y, \rho_V^{V'}) \le E$.

Definition 2.12 (relevant graph). Denote by $\mathcal{G}$ the graph which has vertex-set $\mathrm{Rel}_U^\ell(x, y, 100E)$, with two vertices adjacent if and only if the corresponding elements of $\mathrm{Rel}_U^\ell(x, y, 100E)$ are orthogonal. Let $\mathcal{G}^c$ denote the complementary graph of $\mathcal{G}$, i.e., the graph with the same vertices and edges corresponding to $\preceq$-comparability.

The next lemma is an immediate consequence of Proposition 2.8:

Lemma 2.13. *Elements of $V, V' \in \mathrm{Rel}_U^\ell(x, y, 100E)$ are adjacent in $\mathcal{G}$ if and only if they are $\preceq$-incomparable.*

Lemma 2.14 (coloring relevant elements). *Let χ be the maximal cardinality of a set of pairwise orthogonal elements of $\mathfrak{T}_U^\ell$. Then there exists a χ-coloring of the set of relevant elements of $\mathfrak{T}_U^\ell$ such that nontransverse elements have different colors.*

Proof. Since each clique in $\mathcal{G}$ — i.e., each $\preceq$-antichain in $\mathrm{Rel}_U^\ell(x, y, 100E)$ — has cardinality at most χ, [Dilworth 1950, Theorem 1.1] implies that $\mathcal{G}$ can be colored with χ colors in such a way that $\preceq$-incomparable elements have different colors; hence nontransverse elements have different colors. □

Remark 2.15. The constant χ provided by Lemma 2.14 is bounded by the complexity of $(\mathcal{X}, \mathfrak{S})$, by Lemma 2.2.

3. Realization of consistent tuples

The goal of this section is to prove Theorem 3.1. In this section we will work with a fixed hierarchical space $(\mathcal{X}, \mathfrak{S})$. We will use the concepts of consistency and partial realization points; see Definitions 1.17 and 1.18.

Theorem 3.1 (realization of consistent tuples). *For each $\kappa \geq 1$ there exist $\theta_e, \theta_u \geq 0$ such that the following holds. Let $\vec{b} \in \prod_{W\in\mathfrak{S}} 2^{\mathcal{C}W}$ be κ-consistent; for each W, let b_W denote the $\mathcal{C}W$-coordinate of $\vec{b}$.*

*Then there exists $x \in \mathcal{X}$ so that $\mathsf{d}_W(b_W, \pi_W(x)) \leq \theta_e$ for all $\mathcal{C}W \in \mathfrak{S}$. Moreover, x is **coarsely unique** in the sense that the set of all x which satisfy $\mathsf{d}_W(b_W, \pi_W(x)) \leq \theta_e$ in each $\mathcal{C}W \in \mathfrak{S}$, has diameter at most θ_u.*

Remark 3.2. In typical cases, where the π_U are uniformly coarsely surjective, the admissibility part of the consistency hypothesis is satisfied automatically.

Proof of Theorem 3.1. The main task is to prove the following claim about a κ-consistent admissible tuple $\vec{b}$:

Claim 1. Let $\{V_j\}$ be a family of pairwise-orthogonal elements of $\mathfrak{S}$, all of level at most ℓ. Then there exists $\theta_e = \theta_e(\ell, \kappa) > 100E\kappa\alpha$ and pairwise-orthogonal $\{U_i\}$ such that

(1) each U_i is nested into some V_j,

(2) for each V_j there exists some U_i nested into it, and

(3) any E-partial realization point x for $\{U_i\}$ satisfies $\mathsf{d}_W(b_W, x) \leq \theta_e$ for each $W \in \mathfrak{S}$ for which there exists j with $W \sqsubseteq V_j$.

Applying Claim 1 when $\ell = \ell_S$, where $S \in \mathfrak{S}$ is the unique $\sqsubseteq$-maximal element, along with the partial realization axiom (Definition 1.1.(8)), completes the existence proof, giving us a constant θ_e. If x, y both have the desired property, then $\mathsf{d}_V(x, y) \leq 2\theta_e + \kappa$ for all $V \in \mathfrak{S}$, whence the uniqueness axiom (Definition 1.1.(9)) ensures that $\mathsf{d}(x, y) \leq \theta_u$, for an appropriate θ_u. Hence to prove the theorem it remains to prove Claim 1, which we do now.

The claim when $\ell = 1$ follows from admissibility and the partial realization axiom (Definition 1.1.(8)), so we assume that the claim holds for $\ell - 1 \geq 1$, with $\theta_e(\ell - 1, \kappa) = \theta'_e$, and prove it for level ℓ.

Reduction to the case $|\{V_j\}| = 1$. It suffices to prove the claim in the case where $\{V_j\}$ has a single element, V. To see this, note that once we prove the claim for each V_j separately, yielding a collection of pairwise-orthogonal sets $\{U_i^j \sqsubseteq V_j\}$ with the desired properties, then we take the union of these sets to obtain the claim for the collection $\{V_j\}$.

The case $\{V_j\} = \{V\}$. Fix $V \in \mathfrak{S}$ so that $\ell_V = \ell$. If for each $x \in \mathcal{X}$ that satisfies $\mathsf{d}_V(x, b_V) \leq E$ we have $\mathsf{d}_W(b_W, x) \leq 100E\kappa\alpha$ for $W \in \mathfrak{S}_V$, then the claim follows with $\{U_i\} = \{V\}$. Hence, we can suppose that this is not the case.

We are ready for the main argument, which is contained in Lemma 3.3 below. We will construct $\{U_i\}$ incrementally, using Lemma 3.3, which essentially says that either we are done at a certain stage or we can add new elements to $\{U_i\}$.

We will say that the collection $\mathfrak{U}$ of elements of $\mathfrak{S}_V$ is *totally orthogonal* if any pair of distinct elements of $\mathfrak{U}$ are orthogonal. Given a totally orthogonal family $\mathfrak{U}$ we say that $W \in \mathfrak{S}_V$ is *$\mathfrak{U}$-generic* if there exists $U \in \mathfrak{U}$ so that W is not orthogonal to U. Notice that no W is $\emptyset$-generic.

A totally orthogonal collection $\mathfrak{U} \subseteq \mathfrak{S}_V$ is *C-good* if any E-partial realization point x for $\mathfrak{U}$ has the property that for each $W \in \mathfrak{S}_V$ we have $\mathsf{d}_W(x, b_W) \leq C$. (Notice that our goal is to find such $\mathfrak{U}$.) A totally orthogonal collection $\mathfrak{U} \subseteq \mathfrak{S}_V$ is *C-generically good* if any E-partial realization point x for $\mathfrak{U}$ has the property that for each $\mathfrak{U}$-generic $W \in \mathfrak{S}_V$ we have $\mathsf{d}_W(x, b_W) \leq C$ (e.g., for $\mathfrak{U} = \emptyset$).

We can now quickly finish the proof of the claim using Lemma 3.3 about extending generically good sets, which we state and prove below. Start with $\mathfrak{U} = \emptyset$. If $\mathfrak{U}$ is C-good for $C = 100E\kappa\alpha$, then we are done. Otherwise we can apply Lemma 3.3 and get $\mathfrak{U}_1 = \mathfrak{U}'$ as in the lemma. Inductively, if $\mathfrak{U}_n$ is not $10^n C$-good, we can apply the lemma and extend $\mathfrak{U}_n$ to a new totally orthogonal set $\mathfrak{U}_{n+1}$. Since there is a bound on the cardinality of totally orthogonal sets by Lemma 2.1, in finitely many steps we necessarily get a good totally orthogonal set, and this concludes the proof of the claim, and hence of the theorem. □

Lemma 3.3. *For every $C \geq 100E\kappa\alpha$ the following holds. Let $\mathfrak{U} \subseteq \mathfrak{S}_V - \{V\}$ be totally orthogonal and C-generically good but not C-good. Then there exists a totally orthogonal, $10C$-generically good collection $\mathfrak{U}' \subseteq \mathfrak{S}_V$ with $\mathfrak{U} \subsetneq \mathfrak{U}'$.*

Proof. Let x_0 be an E-partial realization point for $\mathfrak{U}$ so that there exists some $W \sqsubseteq V$ for which $\mathsf{d}_W(b_W, x_0) > C$.

The idea is to try to "move towards" $\vec{b}$ starting from x_0, by looking at all relevant elements of $\mathfrak{S}_V$ that lie between them and finding out which ones are the "closest" to $\vec{b}$.

Let $\mathfrak{V}_{\max}$ be the set of all $W \sqsubseteq V$ for which

(1) $\mathsf{d}_W(b_W, x_0) > C$, and

(2) W is not properly nested into any element of $\mathfrak{S}_V$ satisfying the above inequality.

We now establish two facts about $\mathfrak{V}_{\max}$.

Applying Proposition 2.8 to partially order $\mathfrak{V}_{\max}$. For $U, U' \in \mathfrak{V}_{\max}$, write $U \preceq U'$ if either $U = U'$ or $U \pitchfork U'$ and $\mathsf{d}_U(\rho_U^{U'}, b_U) \le 10E\kappa$; this is a partial order by Proposition 2.8, which also implies that if $U, U' \in \mathfrak{V}_{\max}$ are transverse then they are $\preceq$-comparable. Hence any two $\preceq$-maximal elements of $\mathfrak{V}_{\max}$ are orthogonal, and we denote by $\mathfrak{V}'_{\max}$ the set of $\preceq$-maximal (hence pairwise-orthogonal) elements of $\mathfrak{V}_{\max}$.

Finiteness of $\mathfrak{V}_{\max}$. We now show that $|\mathfrak{V}_{\max}| < \infty$. By Lemma 2.2 and Ramsey's theorem, if $\mathfrak{V}_{\max}$ was infinite then it would contain an infinite subset of pairwise transverse elements, so, in order to conclude that $|\mathfrak{V}_{\max}| < \infty$, it suffices to bound the cardinality of a pairwise-transverse subset of $\mathfrak{V}_{\max}$.

Suppose that $W_1 \prec \cdots \prec W_s \in \mathfrak{V}_{\max}$ are pairwise transverse. By partial realization (Definition 1.1.(8)) and admissibility, there exists $z \in \mathcal{X}$ such that $\mathsf{d}_{W_s}(z, b_{W_s}) \le \alpha$ and $\mathsf{d}_{W_i}(\rho_{W_i}^{W_s}, z) \le \alpha$ for each $i \ne s$, and such that $\mathsf{d}_V(z, \rho_V^{W_s}) \le \alpha$. By consistency of $\vec{b}$ and bounded geodesic image, $\rho_V^{W_s}$ has to be within distance $10E\kappa$ of a geodesic in $\mathcal{C}V$ from x_0 to b_V. In particular $\mathsf{d}_V(x_0, z) \le \theta'_e + 100E\kappa\alpha + 10E\kappa$. Also, for each $i \ne s$,

$$\begin{aligned}\mathsf{d}_{W_i}(x_0, z) &\ge \mathsf{d}_{W_i}(x_0, b_{W_i}) - \mathsf{d}_{W_i}(b_{W_i}, \rho_{W_i}^{W_s}) - \mathsf{d}_{W_i}(\rho_{W_i}^{W_s}, z) \\ &\ge 100E\kappa\alpha - 10E\kappa - \alpha \ge 50E\kappa\alpha \ge 50E.\end{aligned}$$

Indeed, $\mathsf{d}_{W_i}(b_{W_i}, \rho_{W_i}^{W_s}) \le 10E\kappa$ since $W_i \prec W_s$, while $\mathsf{d}_{W_i}(\rho_{W_i}^{W_s}, z) \le \alpha$ by our choice of z. Lemma 2.5 now provides the required bound on s.

Choosing $\mathfrak{U}'$. Since $\ell_U < \ell_V$ for all $U \in \mathfrak{V}'_{\max}$, by induction there exists a totally orthogonal set $\{U_i\}$ so that any E-partial realization point x for $\{U_i\}$ satisfies $\mathsf{d}_T(b_T, x) \le \theta'_e$ for each $T \in \mathfrak{S}$ nested into some $U \in \mathfrak{V}'_{\max}$. Let $\mathfrak{U}' = \{U_i\} \cup \mathfrak{U}$.

Choose such a partial realization point x and let $W \sqsubseteq V$ be $\mathfrak{U}'$-generic. Our goal is to bound $\mathsf{d}_W(x, b_W)$, and we will consider four cases.

If there exists $U \in \mathfrak{U}$ that is not orthogonal to W, then we are done by hypothesis, since any E-partial realization point for $\mathfrak{U}'$ is also an E-partial realization point for $\mathfrak{U}$.

Hence, from now on, assume that W is orthogonal to each $U \in \mathfrak{U}$, i.e., W is not $\mathfrak{U}$-generic.

If $W \sqsubseteq U$ for some $U \in \mathfrak{V}'_{\max}$, then we are done by induction.

Suppose that $W \pitchfork U$ for some $U \in \mathfrak{V}'_{\max}$. For each $U_i \sqsubseteq U$ — and our induction hypothesis implies that there is at least one such U_i — we have $\mathsf{d}_W(x, \rho_W^{U_i}) \le E$

since x is a partial realization point for $\{U_i\}$ and either $U_i \sqsubseteq W$ or $U_i \pitchfork W$ (since W is $\mathfrak{U}'$-generic but not $\mathfrak{U}$-generic). The triangle inequality therefore yields

$$\mathsf{d}_W(x, b_W) \leq E + \mathsf{d}_W(\rho_W^{U_i}, \rho_W^U) + \mathsf{d}_W(b_W, \rho_W^U).$$

By Definition 1.1.(4), $\mathsf{d}_W(\rho_W^{U_i}, \rho_W^U) \leq E$, and we will show that $\mathsf{d}_W(b_W, \rho_W^U) \leq 2C$, so that $\mathsf{d}_W(x, b_W) \leq 2E + 2C$.

Suppose, for a contradiction, that $\mathsf{d}_W(b_W, \rho_W^U) > 2C$. If $\mathsf{d}_U(\rho_U^W, x_0) \leq E$, then

$$\mathsf{d}_U(\rho_U^W, b_U) \geq C - E > \kappa,$$

by consistency, whence $\mathsf{d}_W(\rho_W^U, b_W) \leq \kappa$, a contradiction.

On the other hand, if $\mathsf{d}_U(\rho_U^W, x_0) > E$, then $\mathsf{d}_W(x_0, \rho_W^U) \leq E$ by consistency. Hence $\mathsf{d}_W(x_0, b_W) \geq 2C - E$. Hence there exists a $\sqsubseteq$-maximal $W' \neq V$ with the property that $W \sqsubseteq W' \sqsubseteq V$ and $\mathsf{d}_{W'}(x_0, b_W) > C$ (possibly $W' = W$). Such a W' is in $\mathfrak{V}_{\max}$ by definition.

Since $W \pitchfork U$, and W' and U are $\sqsubseteq$-incomparable, $W' \pitchfork U$. Thus U and W' are $\preceq$-comparable, by Proposition 2.8. Since $W' \neq U$ and U is $\preceq$-maximal, we have $W' \preceq U$, i.e., $\mathsf{d}_{W'}(b_{W'}, \rho_{W'}^U) \leq 10E\kappa$. Since $\preceq$ is antisymmetric, by Lemma 2.9, we have $\mathsf{d}_U(b_U, \rho_U^{W'}) > 10E\kappa$. Since $\mathsf{d}_U(\rho_U^W, \rho_U^{W'}) \leq E$ (from Definition 1.1.(4)), we have $\mathsf{d}_U(b_U, \rho_U^W) > 10E\kappa - E > \kappa$, since $E \geq 1$, so, by consistency, $\mathsf{d}_W(b_W, \rho_W^U) \leq \kappa$, a contradiction.

Finally, suppose $U \sqsubsetneq W$ for some $U \in \mathfrak{V}'_{\max}$. Then, by $\sqsubseteq$-maximality of U, we have $\mathsf{d}_W(x_0, b_W) \leq C$. Also, $\mathsf{d}_W(x, \rho_W^{U_i}) \leq E$ for any $U_i \sqsubseteq U$ since x is a partial realization point, so that $\mathsf{d}_W(x, \rho_W^U) \leq 2E$, since $\mathsf{d}_W(\rho_W^U, \rho_W^{U_i}) \leq E$ by Definition 1.1.(4). If $\mathsf{d}_W(x, b_W) > 2C$, then we claim $\mathsf{d}_U(x_0, b_U) \leq 10E\kappa$, a contradiction. Indeed, any geodesic in $\mathcal{C}W$ from $\pi_W(x_0)$ to b_W does not enter the E-neighborhood of ρ_W^U. By bounded geodesic image, $\mathsf{diam}_U(\rho_U^W(\pi_W(x_0)) \cup \rho_U^W(b_W)) \leq E$ and by consistency, $\mathsf{diam}_U(\rho_U^W(\pi_W(x_0)) \cup \pi_U(x_0)) \leq E$ and $\mathsf{diam}_U(\rho_U^W(b_W) \cup b_U) \leq \kappa$, and we obtain the desired bound on $\mathsf{d}_U(x_0, b_U)$. This completes the proof of the lemma. □

4. Hierarchy paths and the distance formula

Throughout this section, fix a hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$.

4A. ***Definition of hierarchy paths and statement of main theorems.*** Our goal is to deduce the existence of hierarchy paths (Theorem 4.4) from the other axioms and to prove the distance formula (Theorem 4.5).

Definition 4.1 (quasigeodesic, unparameterized quasigeodesic). In the metric space M, a (D, D)-*quasigeodesic* is a (D, D)-quasi-isometric embedding $f : [0, \ell] \to M$; we allow f to be a coarse map, i.e., to send points in $[0, \ell]$ to uniformly bounded sets in M. A (coarse) map $f : [0, \ell] \to M$ is a (D, D)-*unparameterized quasigeodesic* if there exists a strictly increasing function $g : [0, L] \to [0, \ell]$ such that $g(0) = f(0)$, $g(L) = f(\ell)$, and $f \circ g : [0, L] \to M$ is a (D, D)-quasigeodesic and for each $j \in [0, L] \cap \mathbb{N}$, we have $\mathsf{diam}_M(f(g(j)) \cup f(g(j+1))) \le D$.

Definition 4.2 (hierarchy path). For $D \ge 1$, a (not necessarily continuous) path $\gamma : [0, \ell] \to \mathcal{X}$ is a *D-hierarchy path* if

(1) γ is a (D, D)-quasigeodesic,

(2) for each $W \in \mathfrak{S}$, the path $\pi_W \circ \gamma$ is an unparameterized (D, D)-quasigeodesic.

Notation 4.3. Given $A, B \in \mathbb{R}$, we denote by $\{\{A\}\}_B$ the quantity which is A if $A \ge B$ and 0 otherwise. Given C, D, we write $A \asymp_{C,D} B$ to mean $C^{-1}A - D \le B \le CA + D$.

Theorem 4.4 (existence of hierarchy paths). *Let $(\mathcal{X}, \mathfrak{S})$ be hierarchically hyperbolic. Then there exists D_0 such that any $x, y \in \mathcal{X}$ are joined by a D_0-hierarchy path.*

Theorem 4.5 (distance formula). *Let $(X, \mathfrak{S})$ be hierarchically hyperbolic. Then there exists s_0 such that for all $s \ge s_0$ there exist constants K, C such that for all $x, y \in \mathcal{X}$,*

$$\mathsf{d}_{\mathcal{X}}(x, y) \asymp_{(K,C)} \sum_{W \in \mathfrak{S}} \{\{\mathsf{d}_W(x, y)\}\}_s.$$

The proofs of the above two theorems are intertwined, and we give the proof immediately below. This relies on several lemmas, namely Lemma 4.11, proved in Section 4C, and Lemmas 4.19 and 4.18, proved in Section 4D.

Proof of Theorems 4.4 and 4.5. The lower bound demanded by Theorem 4.5 is given by Lemma 4.19 below. By Lemmas 4.11 and 4.18, there is a monotone path (see Definition 4.8) whose length realizes the upper bound on $\mathsf{d}_{\mathcal{X}}(x, y)$, and the same holds for any subpath of this path, which is therefore a hierarchy path, proving Theorem 4.4 and completing the proof of Theorem 4.5. □

4B. *Good and proper paths: definitions.*

We now define various types of (non-continuous) paths in $\mathcal{X}$ that will appear on the way to hierarchy paths.

Definition 4.6 (discrete path). A *K-discrete path* is a map $\gamma : I \to \mathcal{X}$, where I is an interval in $\mathbb{Z}$ and $\mathsf{d}_{\mathcal{X}}(\gamma(i), \gamma(i+1)) \le K$ whenever $i, i+1 \in I$. The *length* $|\alpha|$ of a discrete path α is $\max I - \min I$.

Definition 4.7 (efficient path). A discrete path α with endpoints x, y is *K-efficient* if $|\alpha| \le K\mathsf{d}_{\mathcal{X}}(x, y)$.

Definition 4.8 (monotone path). Given $U \in \mathfrak{S}$, a K-discrete path α and a constant L, we say that α is *L-monotone in U* if whenever $i \le j$ we have $\mathsf{d}_U(\alpha(0), \alpha(i)) \le \mathsf{d}_U(\alpha(0), \alpha(j)) + L$. A path which is L-monotone in U for all $U \in \mathfrak{S}$ is said to be *L-monotone*.

Definition 4.9 (good path). A K-discrete path that is L-monotone in U is said to be (K, L)-*good for* U. Given $\mathfrak{S}' \subseteq \mathfrak{S}$, a path α that is (K, L)-good for each $V \in \mathfrak{S}'$ is (K, L)-*good* for $\mathfrak{S}'$.

Definition 4.10 (proper path). A discrete path $\alpha : \{0, \dots, n\} \to \mathcal{X}$ is (r, K)-*proper* if for $0 \le i < n-1$, we have $\mathsf{d}_{\mathcal{X}}(\alpha(i), \alpha(i+1)) \in [r, r+K]$ and $\mathsf{d}_{\mathcal{X}}(\alpha(n-1), \alpha(n)) \le r + K$. Observe that (r, K)-properness is preserved by passing to subpaths.

4C. *Good and proper paths: existence.* Our goal in this subsection is to join points in $\mathcal{X}$ with proper paths, i.e., to prove Lemma 4.11. This relies on the much more complicated Proposition 4.12, which produces good paths (which are then easily made proper).

Lemma 4.11. *There exists K so that for any $r \ge 0$, any $x, y \in \mathcal{X}$ are joined by a K-monotone, (r, K)-proper discrete path.*

Proof. Let $\alpha_0 : \{0, \dots, n_0\} \to \mathcal{X}$ be a K-monotone, K-discrete path joining x, y, which exists by Proposition 4.12. We modify α_0 to obtain the desired path in the following way. Let $j_0 = 0$ and, proceeding inductively, let j_i be the minimal $j \le n$ such that either $\mathsf{d}_{\mathcal{X}}(\alpha_0(j_{i-1}), \alpha_0(j)) \in [r, r+K]$ or $j = n$. Let m be minimal so that $j_m = n$ and define $\alpha : \{0, \dots, m\} \to \mathcal{X}$ by $\alpha(j) = \alpha_0(i_j)$. The path α is (r, K)-proper by construction; it is easily checked that K-monotonicity is not affected by the above modification; the new path is again discrete, although for a larger discreteness constant. $\square$

It remains to establish the following proposition, whose proof is postponed until the end of this section, after several preliminary statements have been obtained.

Proposition 4.12. *There exists K so that any $x, y \in \mathcal{X}$ are joined by path that is (K, K)-good for each $U \in \mathfrak{S}$.*

Definition 4.13 (hull of a pair of points). For each $x, y \in \mathcal{X}$, $\theta \ge 0$, let $H_\theta(x, y)$ be the set of all $p \in \mathcal{X}$ so that, for each $W \in \mathfrak{S}$, the set $\pi_W(p)$ lies at distance at most θ from a geodesic in $\mathcal{C}W$ joining $\pi_W(x)$ to $\pi_W(y)$. Note that $x, y \in H_\theta(x, y)$.

Remark 4.14. The notion of a hull is generalized in Section 6 to hulls of arbitrary finite sets, but we require only the version for pairs of points in this section.

Lemma 4.15 (retraction onto hulls). *There exist $\theta, K \ge 0$ such that, for each $x, y \in \mathcal{X}$, there exists a (K, K)-coarsely Lipschitz map $r : \mathcal{X} \to H_\theta(x, y)$ that restricts to the identity on $H_\theta(x, y)$.*

Proof. Let κ be the constant from Lemma 2.6, let θ_e be chosen as in the realization theorem (Theorem 3.1), and let $p \in \mathcal{X} - H_{\theta_e}(x, y)$. Define a tuple $\vec{b} = (b_W^p) \in \prod_{W \in \mathfrak{S}} 2^{\mathcal{C}W}$ so that b_W^p is on a geodesic in $\mathcal{C}W$ from $\pi_W(x)$ to $\pi_W(y)$ and is within distance δ of the other two sides of a triangle with vertices in $\pi_W(x), \pi_W(y), \pi_W(p)$.

By Lemma 2.6, this is a consistent tuple. Hence, by the realization theorem (Theorem 3.1), there exists $r(p) \in H_{\theta_e}(x, y)$ so that $\mathsf{d}_W(\pi_W(r(p)), b_W^p) \le \theta_e$. For $p \in H_{\theta_e}(x, y)$, let $r(p) = p$.

To see that r is coarsely Lipschitz, it suffices to bound $\mathsf{d}_{\mathcal{X}}(r(p), r(q))$ when $p, q \in \mathcal{X}$ satisfy $\mathsf{d}_{\mathcal{X}}(p, q) \le 1$. For such p, q we have $\mathsf{d}_W(b_W^p, b_W^q) \le 100E$, so that Theorem 3.1 implies $\mathsf{d}_{\mathcal{X}}(r(p), r(q)) \le \theta_u(100E)$, as required. □

Corollary 4.16. *There exist $\theta, K \ge 0$ such that, for each $x, y \in \mathcal{X}$, there exists a K-discrete and K-efficient path that lies in $H_\theta(x, y)$ and joins x to y.*

Proof. We can assume that $\mathsf{d}_{\mathcal{X}}(x, y) \ge 1$. Since $\mathcal{X}$ is a quasigeodesic space, there exists $C = C(\mathcal{X}) \ge 1$ and a (C, C)-quasi-isometric embedding $\gamma : [0, L] \to \mathcal{X}$ with $\gamma(0) = x, \gamma(L) = y$. Let ρ be the path obtained by restricting $r \circ \gamma : [0, L] \to H_\theta(x, y)$ to $[0, L] \cap \mathbb{N}$, where r is the retraction obtained in Lemma 4.15. Then $\mathsf{d}_{\mathcal{X}}(\rho(i), \rho(i+1)) \le 10KC$ since r is (K, K)-coarsely Lipschitz and γ is (C, C)-coarsely Lipschitz, i.e., ρ is $10KC$-discrete. Finally, ρ is efficient because $L \le C\mathsf{d}_{\mathcal{X}}(x, y) + C \le 2C\mathsf{d}_{\mathcal{X}}(x, y)$. □

The efficiency part of the corollary is used in Lemma 4.19.

4C1. *Producing good paths.* We will need the following lemma, which is a special case of Proposition 6.4.(2). We give a proof in the interest of a self-contained exposition.

Lemma 4.17. *For any θ_0 there exists a constant θ such that for every $x, y \in \mathcal{X}$ and every $x', y' \in H_{\theta_0}(x, y)$, we have $H_{\theta_0}(x', y') \subseteq H_\theta(x, y)$.*

Proof. For any $z \in H_{\theta_0}(x', y')$ and $W \in \mathfrak{S}$ the projection $\pi_W(z)$ lies $2(\delta + \theta_0)$-close to a geodesic in $\mathcal{C}W$ from $\pi_W(x)$ to $\pi_W(y)$, by a thin quadrilateral argument. □

We now prove the main proposition of this subsection.

Proof of Proposition 4.12. Recall that, for $\ell \ge 0$ and $U \in \mathfrak{S}$, the set $\mathfrak{S}_U^\ell$ consists of those $V \in \mathfrak{S}_U$ with $\ell_U - \ell_V \le \ell$, and that $\mathfrak{T}_U^\ell$ consists of those $V \in \mathfrak{S}_U$ with $\ell_U - \ell_V = \ell$.

We prove by induction on ℓ that there exist θ, K such that for any $\ell \ge 0$, $x, y \in \mathcal{X}$ and $U \in \mathfrak{S}$, there is a path α in $H_\theta(x, y)$ connecting x to y such that α is (K, K)-good for $\mathfrak{S}_U^\ell$. It then follows that for any $x, y \in \mathcal{X}$, there exists a path α in $H_\theta(x, y)$ connecting x to y such that α is (K, K)-good for $\mathfrak{S}$; this latter statement directly implies the proposition.

For $a, b \in \mathcal{X}$, denote by $[a, b]_W$ a geodesic in $\mathcal{C}W$ from $\pi_W(a)$ to $\pi_W(b)$. Fix $U \in \mathfrak{S}$.

The case $\ell = 0$. In this case, $\mathfrak{S}_U^0 = \{U\}$. By Corollary 4.16, there exist θ_0, K and a K-discrete, K-efficient path $\alpha_0' : \{0, \dots, k\} \to H_{\theta_0}(x, y)$ joining x to y.

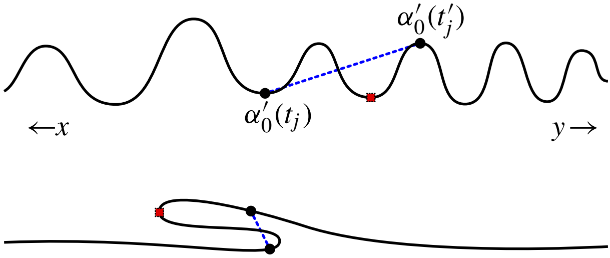

Figure 2. This shows part of α_0' in $\mathcal{X}$ (top) and its projection to U (bottom). The point t_j is an omen, as witnessed by the point marked with a square. Inserting the dashed path β_j, and deleting the corresponding subpath of α_0', makes t_j cease to be an omen.

Similarly, for each $x', y' \in H_{\theta_0}(x, y)$ there exists a K-discrete path β contained in $H_{\theta_0}(x', y')$, joining x' to y', and recall that $H_{\theta_0}(x', y')$ is contained in $H_\theta(x, y)$ for a suitable θ in view of Lemma 4.17.

We use the term *straight path* to refer to a path, such as β, which for each $V \in \mathfrak{S}$ projects uniformly close to a geodesic of $\mathcal{C}(V)$.

We now fix $U \in \mathfrak{S}$, and, using the observation in the last paragraph explain how to modify α_0' to obtain a K-discrete path α_0 in $H_\theta(x, y)$ that is K-monotone in U; the construction will rely on replacing problematic subpaths with straight paths.

A point $t \in \{0, \dots, k\}$ is a *U-omen* if there exists $t' > t$ so that $\mathsf{d}_U(\alpha_0'(0), \alpha_0'(t)) > \mathsf{d}_U(\alpha_0'(0), \alpha_0'(t')) + 5KE$. If α_0' has no U-omens, then we can take $\alpha_0 = \alpha_0'$, so suppose that there is a U-omen and let t_0 be the minimal U-omen, and let $t_0' > t_0$ be maximal so that $\mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_0)) > \mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_0'))$. Inductively define t_j to be the minimal U-omen with $t_j \geq t_{j-1}'$, if such t_j exists; and when t_j exists, we define t_j' to be maximal in $\{0, \dots, k\}$ satisfying $\mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_j)) > \mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_j'))$. For each $j \geq 0$, let $x_j' = \alpha_0'(t_j)$ and let $y_j' = \alpha_0'(t_j')$. See Figure 2.

For each j, there exists a K-discrete path β_j which lies in $H_{\theta_0}(x_j', y_j') \subseteq H_\theta(x, y)$ and is a straight path from x_j' to y_j'. Let α_0 be obtained from α_0' by replacing each $\alpha_0'([t_j, t_j'])$ with β_j. Clearly, α_0 connects x to y, is K-discrete, and is contained in $H_\theta(x, y)$. For each j we have that $\mathsf{diam}_{\mathcal{C}U}(\beta_j) \leq \mathsf{d}_U(x_j', y_j') + 2\theta_0$.

Notice that $\mathsf{d}_U(x_j', y_j') < 2KE + 10\theta_0$. In fact, since $\alpha_0'(0), \alpha_0'(t_j), \alpha_0'(t_j')$ lie θ_0-close to a common geodesic and $\mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_j)) \geq \mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_j'))$, we would otherwise have

$$\mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_j)) - \mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_j')) \geq \mathsf{d}_U(x_j', y_j') - 5\theta_0 \geq 2KE + \theta_0.$$

However, $\mathsf{d}_U(\alpha_0'(t_j), \alpha_0'(t_j + 1)) \leq 2KE$ because of K-discreteness and the projection map to $\mathcal{C}U$ being E-coarsely Lipschitz. Hence, the inequality above implies

$$\mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_j)) > \mathsf{d}_U(\alpha_0'(0), \alpha_0'(t_j')) + 2KE \geq \mathsf{d}_U(\alpha_0'(t_j), \alpha_0'(t_j + 1)),$$

which contradicts the maximality of t'_j. (Notice $t'_j \neq k$, and hence $t'_j + 1 \in \{0, \dots, k\}$ because $\mathsf{d}_U(\alpha'_0(0), \alpha'_0(t'_j)) + \theta_0 < \mathsf{d}_U(\alpha'_0(0), \alpha'_0(t_j)) \leq \mathsf{d}_U(\alpha'_0(0), \alpha'_0(k)) + \theta_0$.)

In particular, we get $\mathrm{diam}_{\mathcal{C}U}(\beta_j) \leq 2KE + 12\theta_0$, and it is then easy to check α_0 is $\max\{5KE, 2KE + 12\theta_0\}$-monotone in U. Replacing K with $\max\{5KE, 2KE + 12\theta_0\}$, we thus have a K-discrete path $\alpha_0 \subset H_\theta(x, y)$ that joins x, y and is K-monotone in U.

We now proceed to the inductive step. Specifically, we fix $\ell \geq 0$ and we assume there exist θ_{ind}, K such that there is a path α in $H_{\theta_{\mathrm{ind}}}(x, y)$ connecting x to y such that α is (K, K)-good for $\mathfrak{S}_U^{\ell-1}$.

The coloring. For short, we will say that $V \in \mathfrak{S}$ is A-relevant if $\mathsf{d}_U(x, y) \geq A$; see Definition 2.7. Notice that to prove that a path in $H_\theta(x, y)$ is monotone, it suffices to restrict our attention to only those $W \in \mathfrak{S}$ which are, say, $10KE$-relevant.

By Lemma 2.14, there exists $\chi \geq 0$, bounded by the complexity of $\mathcal{X}$, and a χ-coloring c of the $10KE$-relevant elements of $\mathfrak{T}_U^\ell$ such that $c(V) = c(V')$ only if $V \pitchfork V'$. In other words, the set of $10KE$-relevant elements of $\mathfrak{T}_U^\ell$ has the form $\bigsqcup_{i=0}^{\chi-1} c^{-1}(i)$, where $c^{-1}(i)$ is a set of pairwise-transverse relevant elements of $\mathfrak{T}_U^\ell$.

Induction hypothesis. Given $p < \chi - 1$, assume by induction (on ℓ and p) that there exist $\theta_p \geq \theta_{\mathrm{ind}}, K_p \geq K$, independent of x, y, U, and a path $\alpha_p : \{0, \dots, k\} \to H_{\theta_p}(x, y)$, joining x, y, that is (K_p, K_p)-good for $\bigsqcup_{i=0}^{p} c^{-1}(i)$ and good for $\mathfrak{S}_U^{\ell-1}$.

Resolving backtracks in the next color. Let θ_{p+1} be provided by Lemma 4.17 with input θ_p. We will modify α_p to construct a K_{p+1}-discrete path α_{p+1} in $H_{\theta_{p+1}}(x, y)$, for some $K_{p+1} \geq K_p$, that joins x, y and is (K_{p+1}, K_{p+1})-good in $\bigsqcup_{i=0}^{p+1} c^{-1}(i) \cup \mathfrak{S}_U^{\ell-1}$.

Notice that we can restrict our attention to the set $\mathcal{C}_{p+1}$ of $100(K_pE + \theta_p)$-relevant elements of $c^{-1}(p+1)$.

A point $t \in \{0, \dots, k\}$ is a $(p+1)$-*omen* if there exist $V \in \mathcal{C}_{p+1}$ and $t' > t$ so that $\mathsf{d}_V(\alpha_p(0), \alpha_p(t)) > \mathsf{d}_V(\alpha_p(0), \alpha_p(t')) + 5K_pE$. If α_p has no $(p+1)$-omens, then we can take $\alpha_{p+1} = \alpha_p$, since α_p is good in each V with $c(V) < p+1$. Therefore, suppose that there is a $(p+1)$-omen and let t_0 be the minimal $(p+1)$-omen, witnessed by $V_0 \in \mathcal{C}_{p+1}$. We can assume that t_0 satisfies $\mathsf{d}_{V_0}(\{x, y\}, \alpha_p(t_0)) > 10K_pE$. Let $t'_0 > t_0$ be maximal so that $\mathsf{d}_{V_0}(\alpha_p(0), \alpha_p(t_0)) > \mathsf{d}_{V_0}(\alpha_p(0), \alpha_p(t'_0))$. In particular $\mathsf{d}_{V_0}(y, \alpha_p(t'_0)) \geq 10E$.

Let $x'_0 = \alpha_0(t_0)$ and $y'_0 = \alpha_0(t'_0)$. Inductively, define t_j as the minimal $(p+1)$-omen, witnessed by $V_j \in \mathcal{C}_{p+1}$, with $t_j \geq t'_{j-1}$, if such t_j exists and let t'_j be maximal so that $\mathsf{d}_{V_j}(\alpha_p(0), \alpha_p(t_j)) > \mathsf{d}_{V_j}(\alpha_p(0), \alpha_p(t'_j))$ and $\mathsf{d}_{V_j}(y, \alpha_p(t'_j)) > 10E$. We can assume that t_j satisfies $\mathsf{d}_{V_j}(\{x, y\}, \alpha_p(t_j)) > 10K_pE$. Also, let $x'_j = \alpha_p(t_j)$, $y'_j = \alpha_p(t'_j)$.

Let β_j be a path in $H_{\theta_p}(x'_j, y'_j)$ joining x'_j to y'_j that is (K_p, K_p)-good for each

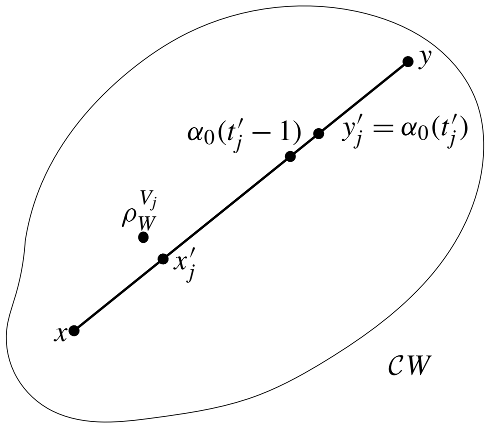

Figure 3. The situation in $\mathcal{C}W$.

relevant V with $c(V) \leq p$ and each relevant $V \in \mathfrak{S}_U^{\ell-1}$. Such paths can be constructed by induction. By Lemma 4.17, β_j lies in $H_{\theta_{p+1}}(x, y)$. Let α_{p+1} be obtained from α_p by replacing each $\alpha_p(\{t_j, \ldots, t'_j\})$ with β_j. Clearly, α_{p+1} connects x to y, is K_p-discrete, and is contained in $H_{\theta_{p+1}}(x, y)$.

We observe that the same argument as in the case $\ell = 0$ gives $\mathsf{d}_{V_j}(x'_j, y'_j) \leq 2K_pE + 10\theta_p$.

Verification that α_{p+1} is good for current colors We next check that each β_j is $10^3(K_pE + \theta_p)$-monotone in each $W \in \bigsqcup_{i=0}^{p+1} c^{-1}(i)$. We have to consider the following cases. (We can and shall assume below W is $100(K_pE + \theta_p)$-relevant.)

- If $W \sqsubseteq V_j$, then $W = V_j$, since $\ell_W = \ell_{V_j}$. Since the projections on $\mathcal{C}W$ of the endpoints of the straight path β_j coarsely coincide, β_j is $(2K_pE + 12\theta_p)$-monotone in W. (See the case $\ell = 0$.)

- Suppose $V_j \sqsubsetneq W$. We claim that the projections of the endpoints of β_j lie at a uniformly bounded distance in $\mathcal{C}W$.

  We claim that $\rho_W^{V_j}$ has to be E-close to either $[x, x'_j]_W$ or $[y'_j, y]_W$. In fact, if this was not the case, we would have

  $$\mathsf{d}_{V_j}(x, y) \leq \mathsf{d}_{V_j}(x, x'_j) + \mathsf{d}_{V_j}(x'_j, y'_j) + \mathsf{d}_{V_j}(y'_j, y) \leq 2E + 2K_pE + 10\theta_p,$$

  where we applied bounded geodesic image (Definition 1.1.(7)) to the first and last terms.

  This is a contradiction with V_j being $100(K_pE + \theta_p)$-relevant.

  Suppose for a contradiction that $\mathsf{d}_W(x'_j, y'_j) \geq 500(K_pE + \theta_p)$. Suppose first that $\rho_W^{V_j}$ is E-close to $[x, x'_j]_W$. Then, by monotonicity, $\rho_W^{V_j}$ is E-far from $[\alpha_p(t'_j), y]_W$. By the bounded geodesic image axiom, this contradicts $\mathsf{d}_{V_j}(y, \alpha_p(t'_j)) \geq 10E$. If instead $\rho_W^{V_j}$ is E-close to $[y'_j, y]_W$, then by bounded geodesic image we have $\mathsf{d}_{V_j}(x, \alpha_p(t_j)) \le E$, contradicting that t_j is an omen witnessed by V_j. See Figure 3.

Hence $\mathsf{d}_W(x_j', y_j') \le 500(K_p E + \theta_p)$ and β_j is $10^3(K_p E + \theta_p)$-monotone in W.

- Suppose $W \pitchfork V_j$. We again claim that the projections of the endpoints of β_j are uniformly close in $\mathcal{C}W$, by showing that they both coarsely coincide with $\rho_W^{V_j}$. Since V_j is relevant, either $\mathsf{d}_{V_j}(x, \rho_{V_j}^W) \ge E$ or $\mathsf{d}_{V_j}(y, \rho_{V_j}^W) \ge E$. Thus, by consistency, $\mathsf{d}_W(\rho_W^{V_j}, \{x, y\}) \le E$. Suppose for a contradiction, that $\mathsf{d}_W(x_j', y_j') > 100(K_p E + \theta_p)$. We consider separately the cases where $\mathsf{d}_W(x, \rho_W^{V_j}) \le E$ and $\mathsf{d}_W(y, \rho_W^{V_j}) \le E$.

  First, suppose that $\mathsf{d}_W(x, \rho_W^{V_j}) \le E$. Then $\mathsf{d}_W(y, \rho_W^{V_j}) \ge 10K_p E - E > E$, so by consistency, $\mathsf{d}_{V_j}(y, \rho_{V_j}^W) \le E$. If $\mathsf{d}_{V_j}(x, \{x_j', y_j'\}) > E$, then consistency implies $\mathsf{d}_W(x_j', \rho_W^{V_j}) \le E$ and $\mathsf{d}_W(y_j', \rho_W^{V_j}) \le E$, whence $\mathsf{d}_W(x_j', y_j') \le 2E$, a contradiction. If $\mathsf{d}_{V_j}(x, \{x_j', y_j'\}) \le E$, then since $\mathsf{d}_{V_j}(x_j', y_j') \le 2K_p E + 10\theta_p$, we have $\mathsf{d}_{V_j}(x, x_j') \le 5K_p E + 10\theta_p$; contradicting that, $\mathsf{d}_{V_j}(x, x_j') > 5K_p E$, since t_j was a $(p+1)$–omen witnessed by V_j.

  Second, suppose $\mathsf{d}_W(y, \rho_W^{V_j}) \le E$. Then by relevance of W and consistency, $\mathsf{d}_{V_j}(x, \rho_{V_j}^W) \le E$. As above, we have $\mathsf{d}_{V_j}(x_j', x) > 5K_p E + 10\theta_p$, so $\mathsf{d}_{V_j}(x, \{x_j', y_j'\}) > K_p E > 3E$ (since $\mathsf{d}_{V_j}(x_j', y_j') \le 2K_p E + 10\theta_p$ and we may assume $K_p > 3$), so $\mathsf{d}_{V_j}(\rho_{V_j}^W, \{x_j', y_j'\}) > E$. Thus, by consistency, $\pi_W(x_j'), \pi_W(y_j')$ both lie at distance at most E from $\rho_W^{V_j}$, so $\mathsf{d}_W(x_j', y_j') \le 3E$.
- Finally, suppose that $W \perp V_j$. Then either $c(W) < c(V_j)$ and β_j is K_p-monotone in W, or W is irrelevant.

Hence, each β_j is $10^3(K_p E + \theta_p)$-monotone in each $W \in c^{-1}(\{0, \dots, p+1\})$. Moreover, our above choice of β_j ensures that β_j is K_p-monotone in each $V \in \mathfrak{S}_U^{\ell-1}$.

Verification that α_{p+1} is monotone. Suppose that there exist t, t' such that $t < t'$ and $\mathsf{d}_V(\alpha_{p+1}(0), \alpha_{p+1}(t)) > \mathsf{d}_V(\alpha_{p+1}(0), \alpha_{p+1}(t')) + 10^4(K_p E + \theta_p)$ for some $V \in c^{-1}(\{0, \dots, p+1\}) \cup \mathfrak{S}_U^{\ell-1}$. We can assume $t, t' \notin \cup_i (t_i, t_i')$. Indeed, if $t \in (t_i, t_i')$ (respectively, $t' \in (t_j, t_j')$), then since all β_m are $10^3(K_p E + \theta_p)$-monotone, we can replace t with t_i' (respectively, t' with t_j). After such a replacement, we still have $\mathsf{d}_V(\alpha_{p+1}(0), \alpha_{p+1}(t)) > \mathsf{d}_V(\alpha_{p+1}(0), \alpha_{p+1}(t')) + 5K_p E$.

Let i be maximal so that $t_i' \le t$ (or let $i = -1$ if no such t_i' exists). By definition of t_{i+1}, we have $t_{i+1} \le t$, and hence $t_{i+1} = t$. But then $t_{i+1}' > t'$, which is not the case.

Conclusion. Continue this procedure as long as $p < \chi$, to produce a path α_χ which is (K, K)–good for $\mathfrak{S}_U^\ell$. In particular, when $U = S$ is $\sqsubseteq$-maximal and ℓ is the length of a maximal $\sqsubseteq$-chain, the proposition follows. □

4D. ***Upper and lower distance bounds.*** We now state and prove the remaining lemmas needed to complete the proof of Theorems 4.4 and 4.5.

Lemma 4.18 (upper bound). *For every K, s there exists r with the following property. Let $\alpha : \{0, \ldots, n\} \to \mathcal{X}$ be a K-monotone, (r, K)-proper discrete path connecting x to y. Then*

$$|\alpha| - 1 \leq \sum_{W \in \mathfrak{S}} \{\!\{\mathsf{d}_W(x, y)\}\!\}_s.$$

Proof. Let $r = r(K, E, s)$ be large enough that, for any $a, b \in \mathcal{X}$, if $\mathsf{d}_{\mathcal{X}}(a, b) \geq r$, then there exists $W \in \mathfrak{S}$ so that $\mathsf{d}_W(a, b) \geq 100KEs$. This r is provided by Definition 1.1.(9).

For $0 \leq j \leq n-1$, choose $V_j \in \mathfrak{S}$ so that $\mathsf{d}_{V_j}(\alpha(j), \alpha(j+1)) \geq 100KEs$. By monotonicity of α in V_j, for any $j' > j$ we have

$$\mathsf{d}_{V_j}(\alpha(0), \alpha(j')) \geq \mathsf{d}_{V_j}(\alpha(0), \alpha(j)) + 50KEs.$$

It follows by induction on $j \leq n$ that $\sum_{W \in \mathfrak{S}} \{\!\{\mathsf{d}_W(\alpha(0), \alpha(j))\}\!\}_s \geq \min\{j, n-1\}$. □

Lemma 4.19 (lower bound). *There exists s_0 such that for all $s \geq s_0$, there exists C with the following property.*

$$\mathsf{d}_{\mathcal{X}}(x, y) \geq \frac{1}{C} \sum_{W \in \mathfrak{S}} \{\!\{\mathsf{d}_W(x, y)\}\!\}_s.$$

Proof. From Corollary 4.16, we obtain a K-discrete path $\alpha : \{0, n\} \to \mathcal{X}$ joining x, y and having the property that the (coarse) path $\pi_V \circ \alpha : \{0, \ldots, n\} \to \mathcal{C}V$ lies in the K-neighborhood of a geodesic from $\pi_V(x)$ to $\pi_V(y)$. Moreover, α is K-efficient, by the same corollary.

Fix $s_0 \geq 10^3 KE$. A *checkpoint* for x, y in $V \in \mathfrak{S}$ is a ball Q in $\mathcal{C}V$ so that $\pi_V \circ \alpha$ intersects Q and $\mathsf{d}_V(\{x, y\}, Q) \geq 10KE + 1$. Note that any ball of radius $10KE$ centered on a geodesic from $\pi_V(x)$ to $\pi_V(y)$ is a checkpoint for x, y in V, provided it is sufficiently far from $\{x, y\}$.

For each $V \in \mathrm{Rel}(x, y, 10^3 KE)$, choose a set $\mathfrak{C}_V$ of $\lceil \mathsf{d}_V(x, y)/10 \rceil$ checkpoints for x, y in V, subject to the requirement that $\mathsf{d}_V(C_1, C_2) \geq 10KE$ for all distinct $C_1, C_2 \in \mathfrak{C}_V$. For each $V \in \mathrm{Rel}(x, y, 10^3 KE)$, we have $10|\mathfrak{C}_V| \geq \mathsf{d}_V(x, y)$, so

$$\sum_{V \in \mathfrak{S}} |\mathfrak{C}_V| \geq \frac{1}{10} \sum_{W \in \mathfrak{S}} \{\!\{\mathsf{d}_W(x, y)\}\!\}_{s_0}.$$

Each $j \in \{0, \ldots, n\}$ is a *door* if there exists $V \in \mathrm{Rel}(x, y, 10^3 KE)$ and $C \in \mathfrak{C}_V$ such that $\pi_V(\alpha(j)) \in C$ but $\pi_V(\alpha(j-1)) \notin C$. The *multiplicity* of a door j is the cardinality of the set $\mathcal{M}(j)$ of $V \in \mathrm{Rel}(x, y, 10^3 KE)$ for which there exists $C \in \mathfrak{C}_V$ with $\pi_V(\alpha(j)) \in C$ and $\pi_V(\alpha(j-1)) \notin C$. Since $\mathfrak{C}_V$ is a set of pairwise-disjoint checkpoints, j is a door for at most one element of $\mathfrak{C}_V$, for each V. Hence the multiplicity of j is precisely the total number of checkpoints in $\cup_{V \in \mathrm{Rel}(x,y,10^3 KE)} \mathfrak{C}_V$ for which j is a door.

We claim that the set $\mathcal{M}(j)$ does not contain a pair of transverse elements. Indeed, suppose that $U, V \in \mathcal{M}(j)$, satisfy $U \pitchfork V$. Let $Q_V \in \mathfrak{C}_V$, $Q_U \in \mathfrak{C}_U$ be the checkpoints containing $\pi_V(\alpha(j))$, $\pi_U(\alpha(j))$ respectively, so that

$$\mathsf{d}_U(\alpha(j), \{x, y\}), \mathsf{d}_V(\alpha(j), \{x, y\}) \geq 10KE + 1 > 10E,$$

contradicting Corollary 2.4. Lemma 2.2 thus gives $|\mathcal{M}_V| \leq \chi$. Now, $|\alpha|$ is at least the number of doors in $\{0, \ldots, n\}$, whence $|\alpha| \geq \frac{1}{\chi} \sum_{V \in \mathfrak{S}} |\mathfrak{C}_V|$. Since α is K-efficient, we obtain

$$\mathsf{d}_{\mathcal{X}}(x, y) \geq \frac{1}{10\chi K} \sum_{W \in \mathfrak{S}} \{\!\{\mathsf{d}_W(x, y)\}\!\}_{s_0}.$$

For $s \geq s_0$, $\sum_{W \in \mathfrak{S}} \{\!\{\mathsf{d}_W(x, y)\}\!\}_s \leq \sum_{W \in \mathfrak{S}} \{\!\{\mathsf{d}_W(x, y)\}\!\}_{s_0}$, so the claim follows. $\square$

5. Hierarchical quasiconvexity and gates

We now introduce the notion of hierarchical quasiconvexity, which is essential for the discussion of product regions, the combination theorem of Section 8, and in [Durham et al. 2017].

Definition 5.1 (hierarchical quasiconvexity). Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space. Then $Y \subseteq \mathcal{X}$ is *k-hierarchically quasiconvex* for some $k : [0, \infty) \to [0, \infty)$, if the following hold:

(1) For all $U \in \mathfrak{S}$, the projection $\pi_U(\mathcal{Y})$ is a $k(0)$-quasiconvex subspace of the δ-hyperbolic space $\mathcal{C}U$.

(2) For all $\kappa \geq 0$ and κ-consistent tuples $\vec{b} \in \prod_{U \in \mathfrak{S}} 2^{\mathcal{C}U}$ with $b_U \subseteq \pi_U(\mathcal{Y})$ for all $U \in \mathfrak{S}$, each point $x \in \mathcal{X}$ for which $\mathsf{d}_U(\pi_U(x), b_U) \leq \theta_e(\kappa)$ (where $\theta_e(\kappa)$ is as in Theorem 3.1) satisfies $\mathsf{d}(x, \mathcal{Y}) \leq k(\kappa)$.

Remark 5.2. Note that condition (2) in the above definition is equivalent to: For each $\kappa > 0$ and every $x \in \mathcal{X}$ for which $\mathsf{d}_U(\pi_U(x), \pi_U(\mathcal{Y})) \leq \kappa$ for all $U \in \mathfrak{S}$, the point x has the property that $\mathsf{d}(x, \mathcal{Y}) \leq k(\kappa)$.

Lemma 5.3. *For each Q there exists κ so that the following holds. Let $\mathcal{Y} \subseteq \mathcal{X}$ be such that $\pi_V(\mathcal{Y})$ is Q-quasiconvex for each $V \in \mathfrak{S}$. Let $x \in \mathcal{X}$ and, for each $V \in \mathfrak{S}$, let $p_V \in \pi_V(\mathcal{Y})$ satisfy $\mathsf{d}_V(x, p_V) \leq \mathsf{d}_V(x, \mathcal{Y}) + 1$. Then (p_V) is κ-consistent.*

Proof. For each V, choose $y_V \in \mathcal{Y}$ so that $\pi_V(y_V) = p_V$.

Suppose that $V \pitchfork W$ or $V \sqsubseteq W$. By Lemma 2.6 and Theorem 3.1, there exists $z \in \mathcal{X}$ so that for all $U \in \mathfrak{S}$, the projection $\pi_U(z)$ lies C-close to each of the geodesics $[\pi_U(x), \pi_U(y_V)]$, $[\pi_U(x), \pi_U(y_W)]$, and $[\pi_U(y_W), \pi_U(y_V)]$, where C depends on $\mathcal{X}$. Hence $\mathsf{d}_V(p_V, z)$ and $\mathsf{d}_W(p_W, z)$ are uniformly bounded, by quasiconvexity of $\pi_V(\mathcal{Y})$ and $\pi_W(\mathcal{Y})$.

Suppose that $V \pitchfork W$. Since the tuple $(\pi_U(z))$ is consistent, either y_V lies uniformly close in $\mathcal{C}V$ to ρ^W_V, or the same holds with V and W interchanged, as required. Suppose that $V \sqsubseteq W$. Suppose that $\mathsf{d}_W(p_W, \rho^V_W)$ is sufficiently large, so that we have to bound $\mathsf{diam}_V(\rho^W_V(p_W) \cup p_V)$. Since $\mathsf{d}_W(z, p_W)$ is uniformly bounded, $\mathsf{d}_W(z, \rho^V_W)$ is sufficiently large that consistency ensures that $\mathsf{diam}_V(\rho^W_V(\pi_W(z)) \cup \pi_V(z))$ is uniformly bounded. Since any geodesic from p_W to z lies far from ρ^V_W, the sets $\rho^W_V(\pi_W(z))$ and $\rho^W_V(p_V)$ coarsely coincide. Since $\pi_V(z)$ coarsely coincides with p_V by construction of z, we have the required bound. Hence the tuple with V-coordinate p_V is κ-consistent for uniform κ. □

Definition 5.4 (gate). A coarsely Lipschitz map $\mathfrak{g}_{\mathcal{Y}} : \mathcal{X} \to \mathcal{Y}$ is called a *gate map* if for each $x \in \mathcal{X}$ it satisfies: $\mathfrak{g}_{\mathcal{Y}}(x)$ is a point $y \in \mathcal{Y}$ such that for all $V \in \mathfrak{S}$, the set $\pi_V(y)$ (uniformly) coarsely coincides with the projection of $\pi_V(x)$ to the $k(0)$-quasiconvex set $\pi_V(\mathcal{Y})$. The point $\mathfrak{g}(x)$ is called the *gate of x in $\mathcal{Y}$*. The uniqueness axiom implies that when such a map exists it is coarsely well-defined.

We first establish that, as should be the case for a (quasi)convexity property, one can coarsely project to hierarchically quasiconvex subspaces. The next lemma shows that gates exist for k-hierarchically quasiconvex subsets.

Lemma 5.5 (existence of coarse gates). *If $\mathcal{Y} \subseteq \mathcal{X}$ is k-hierarchically quasiconvex and nonempty, then there exists a gate map for $\mathcal{Y}$, i.e., for each $x \in \mathcal{X}$ there exists $y \in \mathcal{Y}$ such that for all $V \in \mathfrak{S}$, the set $\pi_V(y)$ (uniformly) coarsely coincides with the projection of $\pi_V(x)$ to the $k(0)$-quasiconvex set $\pi_V(\mathcal{Y})$.*

Proof. For each $V \in \mathfrak{S}$, let $p_V \in \pi_V(\mathcal{Y})$ satisfy $\mathsf{d}_V(x, p_V) \leq \mathsf{d}_V(x, \mathcal{Y}) + 1$. Then (p_V) is κ-consistent for some κ independent of x by Lemma 5.3. (Note that (p_v) is admissible by construction.)

Theorem 3.1 and the definition of hierarchical quasiconvexity combine to supply $y' \in \mathcal{N}_{k(\kappa)}(\mathcal{Y})$ with the desired projections to all $V \in \mathfrak{S}$; this point lies at distance $k(\kappa)$ from some $y \in \mathcal{Y}$ with the desired property.

We now check that this map is coarsely Lipschitz. Let $x_0, x_n \in \mathcal{X}$ be joined by a uniform quasigeodesic γ. By sampling γ, we obtain a discrete path $\gamma' : \{0, \dots, n\} \to \mathcal{X}$ such that $\mathsf{d}_{\mathcal{X}}(\gamma'(i), \gamma'(i+1)) \leq K$ for $0 \leq i \leq n-1$, where K depends only on $\mathcal{X}$, and such that $\gamma'(0) = x_0$, $\gamma'(n) = x_n$. Observe that

$$\mathsf{d}_{\mathcal{X}}(\mathfrak{g}_{\mathcal{Y}}(x_0), \mathfrak{g}_{\mathcal{Y}}(x_n)) \leq \sum_{i=0}^{n-1} \mathsf{d}_{\mathcal{X}}(\mathfrak{g}_{\mathcal{Y}}(\gamma'(i)), \mathfrak{g}_{\mathcal{Y}}(\gamma'(i+1))),$$

so it suffices to exhibit C such that $\mathsf{d}_{\mathcal{X}}(\mathfrak{g}_{\mathcal{Y}}(x), \mathfrak{g}_{\mathcal{Y}}(x')) \leq C$ whenever $\mathsf{d}_{\mathcal{X}}(x, x') \leq K$. But if $\mathsf{d}_{\mathcal{X}}(x, x') \leq K$, then each $\mathsf{d}_U(x, x') \leq K'$ for some uniform K', by Definition 1.1.(1), whence the claim follows from the fact that each $\mathcal{C}U \to \pi_U(\mathcal{Y})$ is coarsely Lipschitz (with constant depending only on δ and $k(0)$) along with the uniqueness axiom (Definition 1.1.(9)). □

5A. ***Hierarchically quasiconvex subspaces are hierarchically hyperbolic.***

Proposition 5.6. *Let $\mathcal{Y} \subseteq \mathcal{X}$ be a hierarchically k-quasiconvex subset of the hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$. Then $(\mathcal{Y}, \mathsf{d})$ is a hierarchically hyperbolic space, where d is the metric inherited from $\mathcal{X}$.*

Proof. There exists K so that any two points in $\mathcal{Y}$ are joined by a uniform quasigeodesic. Indeed, any two points in $\mathcal{Y}$ are joined by a hierarchy path in $\mathcal{X}$, which must lie uniformly close to $\mathcal{Y}$.

We now define a hierarchically hyperbolic structure. For each U, let $r_U : \mathcal{C}U \to \pi_U(\mathcal{Y})$ be the coarse projection, which exists by quasiconvexity. The index set is $\mathfrak{S}$, and the associated hyperbolic spaces are the various $\mathcal{C}U$. For each U, define a projection $\pi'_U : \to \mathcal{C}U$ by $\pi'_U = r_U \circ \pi_U$, and for each nonorthogonal pair $U, V \in \mathfrak{S}$, the corresponding relative projection $\mathcal{C}U \to \mathcal{C}V$ is given by $r_V \circ \rho^U_V$. All of the requirements of Definition 1.1 involving only the various relations on $\mathfrak{S}$ are obviously satisfied, since we have only modified the projections. The consistency inequalities continue to hold since each r_U is uniformly coarsely Lipschitz. The same is true for bounded geodesic image and the large link lemma. Partial realization holds by applying the map $\mathfrak{g}_{\mathcal{Y}}$ to points constructed using partial realization in $(\mathcal{X}, \mathfrak{S})$. $\square$

Remark 5.7 (alternative hierarchically hyperbolic structures). In the above proof, one can replace each $\mathcal{C}U$ by a thickening $\mathcal{C}U_{\mathcal{Y}}$ of $\pi_U(\mathcal{Y})$ (this set is quasiconvex; the thickening is to make a hyperbolic geodesic space). This yields a hierarchically hyperbolic structure with coarsely surjective projections.

5B. ***Standard product regions.*** In this section, we describe a class of hierarchically quasiconvex subspaces called *standard product regions* that will be useful in future applications. We first recall a construction from [Behrstock et al. 2017b, Section 13].

Definition 5.8 (nested partial tuple). Recall $\mathfrak{S}_U = \{V \in \mathfrak{S} : V \sqsubseteq U\}$. Fix $\kappa \geq \kappa_0$ and let $\boldsymbol{F}_U$ be the set of κ-consistent tuples in $\prod_{V \in \mathfrak{S}_U} 2^{\mathcal{C}V}$.

Definition 5.9 (orthogonal partial tuple). Let $\mathfrak{S}^{\perp}_U = \{V \in \mathfrak{S} : V \perp U\} \cup \{A\}$, where A is a $\sqsubseteq$-minimal element A such that $V \sqsubseteq A$ for all $V \perp U$. Fix $\kappa \geq \kappa_0$, let $\boldsymbol{E}_U$ be the set of κ-consistent tuples in $\prod_{V \in \mathfrak{S}^{\perp}_U - \{A\}} 2^{\mathcal{C}V}$.

Construction 5.10 (product regions in $\mathcal{X}$). Given $\mathcal{X}$ and $U \in \mathfrak{S}$, there is a coarsely well-defined map $\phi_U : \boldsymbol{F}_U \times \boldsymbol{E}_U \to \mathcal{X}$, with hierarchically quasiconvex image, that restricts to coarsely well-defined maps $\phi^{\sqsubseteq} : \boldsymbol{F}_U \to \mathcal{X}$ and $\phi^{\perp} : \boldsymbol{E}_U \to \mathcal{X}$. Indeed, for each $(\vec{a}, \vec{b}) \in \boldsymbol{F}_U \times \boldsymbol{E}_U$, and each $V \in \mathfrak{S}$, define the coordinate $(\phi_U(\vec{a}, \vec{b}))_V$ as follows. If $V \sqsubseteq U$, then $(\phi_U(\vec{a}, \vec{b}))_V = a_V$. If $V \perp U$, then $(\phi_U(\vec{a}, \vec{b}))_V = b_V$. If $V \pitchfork U$, then $(\phi_U(\vec{a}, \vec{b}))_V = \rho^U_V$. Finally, if $U \sqsubseteq V$, and $U \neq V$, let $(\phi_U(\vec{a}, \vec{b}))_V = \rho^U_V$.

We now verify that the tuple $\phi_U(\vec{a}, \vec{b})$ is consistent. If $W, V \in \mathfrak{S}$, and either V or W is transverse to U, then the consistency inequality involving W and V is satisfied in view of Proposition 1.8. The same holds if $U \sqsubseteq W$ or $U \sqsubseteq V$. Hence, it remains to consider the cases where V and W are each either nested into or orthogonal to U: if $V, W \sqsubseteq U$ or $V, W \perp U$ then consistency holds by assumption; otherwise, up to reversing the roles of V and W we have $V \sqsubseteq U$ and $W \perp U$, in which case $V \perp W$ and there is nothing to check. Theorem 3.1 thus supplies the map $\phi_U : \boldsymbol{F}_U \times \boldsymbol{E}_U \to \mathcal{X}$. Fixing any $e \in \boldsymbol{E}_U$ yields a map $\phi^{\sqsubseteq} : \boldsymbol{F}_U \times \{e\} \to \mathcal{X}$, and $\phi^{\perp}$ is defined analogously. Note that these maps depend on choices of basepoints in $\boldsymbol{E}_U, \boldsymbol{F}_U$.

Where it will not introduce confusion (e.g., where the basepoints are understood or immaterial), we abuse notation and regard $\boldsymbol{F}_U, \boldsymbol{E}_U$ as subspaces of $\mathcal{X}$, i.e., $\boldsymbol{F}_U = \operatorname{im} \phi^{\sqsubseteq}$, $\boldsymbol{E}_U = \operatorname{im} \phi^{\perp}$.

Proposition 5.11. *When $\boldsymbol{E}_U, \boldsymbol{F}_U \subset \mathcal{X}$ are endowed with the subspace metric* d, *the spaces $(\boldsymbol{F}_U, \mathfrak{S}_U)$ and $(\boldsymbol{E}_U, \mathfrak{S}_U^{\perp})$ are hierarchically hyperbolic; if U is not $\sqsubseteq$-maximal, then their complexity is strictly less than that of $(\mathcal{X}, \mathfrak{S})$. Moreover, $\phi^{\sqsubseteq}$ and $\phi^{\perp}$ determine hieromorphisms $(\boldsymbol{F}_U, \mathfrak{S}_U), (\boldsymbol{E}_U, \mathfrak{S}_U^{\perp}) \to (\mathcal{X}, \mathfrak{S})$.*

Proof. For each $V \sqsubseteq U$ or $V \perp U$, the associated hyperbolic space $\mathcal{C}V$ is exactly the one used in the hierarchically hyperbolic structure $(\mathcal{X}, \mathfrak{S})$. For A, use an appropriate thickening $\mathcal{C}^*A$ of $\pi_A(\operatorname{im} \phi^{\perp})$ to a hyperbolic geodesic space. All of the projections $\boldsymbol{F}_U \to \mathcal{C}V$, $V \in \mathfrak{S}_U$ and $\boldsymbol{E}_U \to \mathcal{C}V$, $V \in \mathfrak{S}_U^{\perp}$ are as in $(\mathcal{X}, \mathfrak{S})$ (for A, compose with a quasi-isometry $\pi_A(\operatorname{im} \phi^{\perp}) \to \mathcal{C}^*A$). Observe that $(\boldsymbol{F}_U, \mathfrak{S}_U)$ and $(\boldsymbol{E}_U, \mathfrak{S}_U^{\perp})$ are hierarchically hyperbolic (this can be seen using a simple version of the proof of Proposition 5.6). If U is not $\sqsubseteq$-maximal in $\mathfrak{S}$, then neither is A, whence the claim about complexity.

The hieromorphisms are defined by the inclusions $\mathfrak{S}_U, \mathfrak{S}_U^{\perp} \to \mathfrak{S}$ and, for each $V \in \mathfrak{S}_U \cup \mathfrak{S}_U^{\perp}$, the identity $\mathcal{C}V \to \mathcal{C}V$, unless $V = A$, in which case we use $\mathcal{C}^*A \to \pi_A(\operatorname{im} \phi^{\perp}) \hookrightarrow \mathcal{C}A$. These give hieromorphisms by definition. □

Remark 5.12 (dependence on A). Note that A need not be the unique $\sqsubseteq$-minimal element of $\mathfrak{S}$ into which each $V \perp U$ is nested; the axioms don't require uniqueness of such $\sqsubseteq$-minimal elements. Observe that $\boldsymbol{E}_U$ (as a set and as a subspace of $\mathcal{X}$) is defined independently of the choice of A. It is the hierarchically hyperbolic structure from Proposition 5.11 that a priori depends on A. However, note that $A \not\sqsubseteq U$, since there exists $V \sqsubseteq A$ with $V \perp U$, and we cannot have $V \sqsubseteq U$ and $V \perp U$ simultaneously. Likewise, $A \not\perp U$ by definition. Finally, if $U \sqsubseteq A$, then the axioms guarantee the existence of B, properly nested into A, into which each $V \perp U$ is nested, contradicting $\sqsubseteq$-minimality of A. Hence $U \pitchfork A$. It follows that $\pi_A(\boldsymbol{E}_U)$ is bounded — it coarsely coincides with ρ^U_A. Thus the hierarchically hyperbolic structure on $\boldsymbol{E}_U$, and the hieromorphism structure of $\phi^{\perp}$, is actually essentially canonical: we can take the hyperbolic space associated to the $\sqsubseteq$-maximal element to be a point, whose image in each of the possible choices of A must coarsely coincide with ρ_A^U.

Remark 5.13 (orthogonality and product regions). If $U \perp V$, then we have $\boldsymbol{F}_U \subseteq \boldsymbol{E}_V$ and $\boldsymbol{F}_V \subseteq \boldsymbol{E}_U$, so there is a hierarchically quasiconvex map $\phi_U^{\sqsubseteq} \times \phi_V^{\sqsubseteq} : \boldsymbol{F}_U \times \boldsymbol{F}_V \to \mathcal{X}$ extending to $\phi_U^{\sqsubseteq} \times \phi_U^{\perp}$ and $\phi_V^{\perp} \times \phi_V^{\sqsubseteq}$.

Remark 5.14. Since $\boldsymbol{F}_U$, $\boldsymbol{E}_U$ are hierarchically quasiconvex spaces, Definition 5.4 provides coarse gates $\mathfrak{g}_{\boldsymbol{F}_U} : \mathcal{X} \to \boldsymbol{F}_U$ and $\mathfrak{g}_{\boldsymbol{E}_U} : \mathcal{X} \to \boldsymbol{E}_U$. These are coarsely the same as the following maps: given $x \in \mathcal{X}$, let $\vec{x}$ be the tuple defined by $x_W = \pi_W(x)$ when $W \sqsubseteq U$ and $x_W = \pi_W(x)$ when $W \perp U$ and ρ_W^U otherwise. Then $\vec{x}$ is consistent and coarsely equals $\mathfrak{g}_{\boldsymbol{F}_U \times \boldsymbol{E}_U}(x)$.

Definition 5.15 (standard product region). For each $U \in \mathfrak{S}$, let $P_U = \operatorname{im} \phi_U$, which is coarsely $\boldsymbol{F}_U \times \boldsymbol{E}_U$. We call this the *standard product region* in $\mathcal{X}$ associated to U.

The next proposition follows from the definition of the product regions and the fact that, if $U \sqsubseteq V$, then ρ_W^U, ρ_W^V coarsely coincide whenever $V \sqsubseteq W$ or $V \pitchfork W$ and $U \not\perp W$, which holds by Definition 1.1.(4).

Proposition 5.16 (parallel copies). *There exists $\nu \geq 0$ such that for all $U \in \mathfrak{S}$, all $V \in \mathfrak{S}_U$, and all $u \in \boldsymbol{E}_U$, there exists $v \in \boldsymbol{E}_V$ so that $\phi_V(\boldsymbol{F}_V \times \{v\}) \subseteq \mathcal{N}_\nu(\phi_U(\boldsymbol{F}_U \times \{u\}))$.*

5B1. *Hierarchy paths and product regions.* Recall that a *D-hierarchy path* γ in $\mathcal{X}$ is a (D, D)-quasigeodesic $\gamma : I \to \mathcal{X}$ such that $\pi_U \circ \gamma$ is an unparameterized (D, D)-quasigeodesic for each $U \in \mathfrak{S}$, and that Theorem 4.4 provides $D \geq 1$ so that any two points in $\mathcal{X}$ are joined by a D-hierarchy path. In this section, we describe how hierarchy paths interact with standard product regions.

In the next proposition and lemma, given $x, y \in \mathcal{X}$, we declare $V \in \mathfrak{S}$ to be *relevant* (*for* x, y) if $\mathsf{d}_V(x, y) \geq 200DE$.

Proposition 5.17 ("active" subpaths). *There exists $\nu \geq 0$ so that for all $x, y \in \mathcal{X}$, all $V \in \mathfrak{S}$ with V relevant for (x, y), and all D-hierarchy paths γ joining x to y, there is a subpath α of γ with the following properties*:

(1) $\alpha \subset \mathcal{N}_\nu(P_V)$.

(2) *$\pi_U|_\gamma$ is coarsely constant on $\gamma - \alpha$ for all $U \in \mathfrak{S}_V \cup \mathfrak{S}_V^{\perp}$.*

Proof. We may assume $\gamma : \{0, n\} \to \mathcal{X}$ is a $2D$-discrete path. Let $x_i = \gamma(i)$ for $0 \leq i \leq n$. Let $S \in \mathfrak{S}$ be the $\sqsubseteq$-maximal element. Since the proposition holds trivially for $V = S$, assume $V \sqsubsetneq S$.

First consider the case where V is $\sqsubseteq$-maximal among relevant elements of $\mathfrak{S}$. Lemma 5.18 provides $\nu'' \geq 0$, independent of x, y, and also provides $i \leq n$, such that $\mathsf{d}_S(x_i, \rho_S^V) \leq \nu''$. Let i be minimal with this property and let i' be maximal with this property. Observe that there exists $\nu' \geq \nu''$, depending only on ν'' and the (uniform) monotonicity of γ in $\mathcal{C}S$, such that $\mathsf{d}_S(x_j, \rho_S^V) \leq \nu'$ for $i \leq j \leq i'$.

For $j \in \{i, \dots, i'\}$, let $x_j' = \mathfrak{g}_{P_V}(x_j)$. Let $U \in \mathfrak{S}$. By definition, if $U \sqsubseteq V$ or $U \perp V$, then $\pi_U(x_j)$ coarsely coincides with $\pi_U(x_j')$, while $\pi_U(x_j')$ coarsely coincides with ρ_U^V if $V \sqsubseteq U$ or $V \pitchfork U$. We claim that there exist i_1, i_1' with $i \leq i_1 \leq i_1' \leq i'$ such that for $i_1 \leq j \leq i_1'$ and $U \in \mathfrak{S}$ with $V \sqsubseteq U$ or $U \pitchfork V$, the points $\pi_U(x_j)$ and $\pi_U(x_j')$ coarsely coincide; this amounts to claiming $\pi_U(x_j)$ coarsely coincides with ρ_U^V.

If $V \sqsubseteq U$ and some geodesic σ in $\mathcal{C}U$ from $\pi_U(x)$ to $\pi_U(y)$ fails to pass through the E-neighborhood of ρ_U^V, then bounded geodesic image shows that $\rho_V^U(\sigma)$ has diameter at most E. On the other hand, consistency shows that the endpoints of $\rho_V^U(\sigma)$ coarsely coincide with $\pi_V(x)$ and $\pi_V(y)$, contradicting that V is relevant. Thus σ passes through the E-neighborhood of ρ_U^V. Maximality of V implies that U is not relevant, so that $\pi_V(x)$, $\pi_V(y)$, and $\pi_V(x_j)$ all coarsely coincide, whence $\pi_V(x_j)$ coarsely coincides with ρ_U^V.

If $U \pitchfork V$ and U is not relevant, then $\pi_U(x_j)$ coarsely coincides with both $\pi_U(x)$ and $\pi_U(y)$, each of which coarsely coincides with ρ_U^V, for otherwise we would have $\mathsf{d}_V(x, y) \leq 2E$ by consistency and the triangle inequality, contradicting that V is relevant. If $U \pitchfork V$ and U is relevant, then, by consistency, we can assume that $\pi_U(y)$, ρ_U^V coarsely coincide, as do $\pi_V(x)$, ρ_V^U. Either $\pi_U(x_j)$ coarsely equals ρ_U^V, or $\pi_V(x_j)$ coarsely equals $\pi_V(x)$, again by consistency. If $\mathsf{d}_V(x, x_j) \leq 10E$ or $\mathsf{d}_V(y, x_j) \leq 10E$, discard x_j. Our discreteness assumption and the fact that V is relevant imply that there exist $i_1 \leq i_1'$ between i and i' so that x_j is not discarded for $i_1 \leq j \leq i_1'$. For such j, the distance formula now implies that $\mathsf{d}(x_j, x_j')$ is bounded by a constant ν independent of x, y.

We thus have i_1, i_1' such that $x_j \in \mathcal{N}_\nu(P_V)$ for $i \leq j \leq i'$ and $x_j \notin \mathcal{N}_\nu(P_V)$ for $j < i$ or $j > i'$, provided V is $\sqsubseteq$-maximal relevant. If $W \sqsubseteq V$ and W is relevant, and there is no relevant $W' \neq W$ with $W \sqsubseteq W' \sqsubseteq V$, then we may apply the above argument to $\gamma' = \mathfrak{g}_{P_V}(\gamma|_{i,\dots,i'})$ to produce a subpath of γ' lying ν-close to $P_W \subseteq P_V$, and hence a subpath of γ lying 2ν-close to P_W. Finiteness of the complexity (Definition 1.1.(5)) then yields assertion (1). Assertion (2) is immediate from our choice of i_1, i_1'. $\square$

Lemma 5.18. *There exists $\nu' \geq 0$ so that for all $x, y \in \mathcal{X}$, all relevant $V \in \mathfrak{S}$, and all D-hierarchy paths γ joining x to y, there exists $t \in \gamma$ so that $\mathsf{d}_S(t, \rho_S^V) \leq \nu'$.*

Proof. Let σ be a geodesic in $\mathcal{C}S$ joining the endpoints of $\pi_S \circ \gamma$. Since

$$\mathsf{d}_V(x, y) \geq 200DE,$$

the consistency and bounded geodesic image axioms (Definition 1.1.(4) and 1.1.(7)) imply that σ enters the E-neighborhood of ρ_S^V in $\mathcal{C}S$, whence $\pi_S \circ \gamma$ comes uniformly close to ρ_S^V. $\square$

6. Hulls

In this section we build "convex hulls" in hierarchically hyperbolic spaces. This construction is motivated by, and generalizes, the concept in the mapping class group called Σ-hull, as defined by Behrstock, Kleiner, Minsky and Mosher [Behrstock et al. 2012]. Recall that given a set A of points in a δ-hyperbolic space H, its *convex hull*, denoted $\mathrm{hull}_H(A)$, is the union of geodesics between pairs of points in this set. We will make use of the fact that the convex hull is 2δ-quasiconvex (since, if $p \in [x, y], q \in [x', y']$, then $[p, q] \subseteq \mathcal{N}_{2\delta}([p, x] \cup [x, x'] \cup [x', q]) \subseteq \mathcal{N}_{2\delta}([y, x] \cup [x, x'] \cup [x', y'])$).

The construction of these hulls is based on Proposition 6.3, which generalizes Lemma 4.15; indeed, the construction of hulls in this section generalizes the hulls of pairs of points used in Section 4 to prove the distance formula. The second part of Proposition 6.3 (which is not used in Section 4) relies on the distance formula.

Definition 6.1 (hull of a set). For each set $A \subset \mathcal{X}$ and $\theta \geq 0$, let $H_\theta(A)$ be the set of all $p \in \mathcal{X}$ so that, for each $W \in \mathfrak{S}$, the set $\pi_W(p)$ lies at distance at most θ from $\mathrm{hull}_{\mathcal{C}W}(A)$. Note that $A \subset H_\theta(A)$.

Lemma 6.2. *There exists θ_0 so that for each $\theta \geq \theta_0$ there exists $k : \mathbb{R}^+ \to \mathbb{R}^+$ such that for each $A \subseteq \mathcal{X}$, we have that $H_\theta(A)$ is k-hierarchically quasiconvex.*

Proof. For any θ and $U \in \mathfrak{S}$, due to δ-hyperbolicity we have that $\pi_U(H_\theta(A))$ is 2δ-quasiconvex, so we only have to check the condition on realization points.

Let A' be the union of all D_0-hierarchy paths joining points in A, where D_0 is the constant from Theorem 4.4. Then the Hausdorff distance between $\pi_U(A')$ and $\pi_U(A)$ is bounded by $C = C(\delta, D_0)$ for each $U \in \mathfrak{S}$. Also, $\pi_U(A')$ is $Q = Q(\delta, D_0)$-quasiconvex. Let κ be the constant from Lemma 5.3, and let $\theta_0 = \theta_e(\kappa)$ be as in Theorem 3.1.

Fix any $\theta \geq \theta_0$, and any $\kappa \geq 0$. Let (b_U) be a κ'-consistent tuple with $b_U \subseteq \mathcal{N}_\theta(\mathrm{hull}_{\mathcal{C}U}(A))$ for each $U \in \mathfrak{S}$. Let $x \in \mathcal{X}$ project $\theta_e(\kappa')$-close to each b_U. We have to find $y \in H_\theta(A)$ uniformly close to x. By Lemma 5.3, (p_U) is κ-consistent, where $p_U \in \mathrm{hull}_{\mathcal{C}W}(A)$ satisfies $\mathrm{d}_U(x, p_U) \leq \mathrm{d}_U(x, \mathrm{hull}_{\mathcal{C}W}(A)) + 1$. It is readily seen from the uniqueness axiom (Definition 1.1.(9)) that any $y \in \mathcal{X}$ projecting close to each p_U has the required property, and such a y exists by Theorem 3.1. To check admissibility, note that each p_U lies θ-close to $\mathrm{hull}_{\mathcal{C}U}(A)$, which in turn lies uniformly close to $\pi_U(\mathcal{X})$ by quasiconvexity of $\pi_U(\mathcal{X})$. □

We denote the Hausdorff distance in the metric space Y by $\mathrm{d}_{\mathrm{Haus},Y}(\cdot\,,\cdot)$. The next proposition directly generalizes [Behrstock et al. 2012, Proposition 5.2] from mapping class groups to general hierarchically hyperbolic spaces.

Proposition 6.3 (retraction onto hulls). *For each sufficiently large θ there exists $C \geq 1$ so that for each set $A \subset \mathcal{X}$ there is a (K, K)-coarsely Lipschitz map*

$r : \mathcal{X} \to H_\theta(A)$ restricting to the identity on $H_\theta(A)$. Moreover, if $A' \subset \mathcal{X}$ lies at finite Hausdorff distance from A, then $d_{\mathcal{X}}(r_A(x), r_{A'}(x))$ is C-coarsely Lipschitz in $\mathsf{d}_{\mathrm{Haus},\mathcal{X}}(A, A')$.

Proof. By Lemma 6.2, for all sufficiently large θ, $H_\theta(A)$ is hierarchically quasiconvex. Thus, by Lemma 5.5 there exists a map $r : \mathcal{X} \to H_\theta(A)$, which is coarsely Lipschitz and which is the identity on $H_\theta(A)$.

We now prove the "moreover" clause. By Definition 1.1.(1), for each W the projections π_W are each coarsely Lipschitz and thus $\mathsf{d}_{\mathrm{Haus},\mathcal{C}W}(\pi_W(A), \pi_W(A'))$ is bounded by a coarsely Lipschitz function of $\mathsf{d}_{\mathrm{Haus},\mathcal{X}}(A, A')$. It is then easy to conclude using the distance formula (Theorem 4.5) and the construction of gates (Definition 5.4) used to produce the map r. □

6A. *Homology of asymptotic cones.* In this subsection we make a digression to study homological properties of asymptotic cones of hierarchically hyperbolic spaces. This subsection is not needed for the proof of distance formula, and in fact we will use the distance formula in a proof.

Using Proposition 6.3, the identical proof as used in [Behrstock et al. 2012, Lemma 5.4] for mapping class groups, yields:

Proposition 6.4. *There exists $\theta_0 \geq 0$ depending only on the constants of the hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$ such that for all $\theta, \theta' \geq \theta_0$ there exist K, C, and θ'' such that given two sets $A, A' \subset \mathcal{X}$, then*:

(1) $\mathsf{diam}(H_\theta(A)) \leq K\ \mathsf{diam}(A) + C$.

(2) *If $A' \subset H_\theta(A)$ then $H_\theta(A') \subset H_{\theta''}(A)$.*

(3) $d_{\mathrm{Haus},\mathcal{X}}(H_\theta(A), H_\theta(A')) \leq K d_{\mathrm{Haus},\mathcal{X}}(A, A') + C$.

(4) $d_{\mathrm{Haus},\mathcal{X}}(H_\theta(A), H_{\theta'}(A)) \leq C$.

Remark 6.5. Proposition 6.4 is slightly stronger than the corresponding [Behrstock et al. 2012, Lemma 5.4], in which A, A' are finite sets and the constants depend on their cardinality. The source of the strengthening is just the observation that hulls in δ-hyperbolic spaces are 2δ-quasiconvex regardless of the cardinality of the set (see [Behrstock et al. 2012, Lemma 5.1]).

It is an easy observation that given a sequence $\boldsymbol{A}$ of sets $A_n \subset \mathcal{X}$ with bounded cardinality, the retractions to the corresponding hulls $H_\theta(A_n)$ converge in any asymptotic cone, $\mathcal{X}_\omega$, to a Lipschitz retraction from that asymptotic cone to the ultralimit of the hulls, $H(\boldsymbol{A})$. A general argument, see e.g., [Behrstock et al. 2012, Lemma 6.2] implies that the ultralimit of the hulls is then contractible. The proofs in [Behrstock et al. 2012, Section 6] then apply in the present context using the above proposition, with the only change needed being that the reference to the rank theorem for hierarchically hyperbolic spaces as proven in [Behrstock et al.

2017b, Theorem J] must replace the application of [Behrstock and Minsky 2008]. In particular, this yields the following two results:

Corollary 6.6. *Let $\mathcal{X}$ be a hierarchically hyperbolic space and $\mathcal{X}_\omega$ one of its asymptotic cones. Let $X \subset \mathcal{X}_\omega$ be an open subset and suppose that for any sequence, $\boldsymbol{A}$, of finite subsets of X we have $H(\boldsymbol{A}) \subset X$. Then X is acyclic.*

Corollary 6.7. *If (U, V) is an open pair in $\mathcal{X}_\omega$, then $H_k(U, V) = \{0\}$ for all k greater than the complexity of $\mathcal{X}$.*

6B. ***Relatively hierarchically hyperbolic spaces and the distance formula.*** In this section, we work in the following context:

Definition 6.8 (relatively hierarchically hyperbolic spaces). The hierarchical space $(\mathcal{X}, \mathfrak{S})$ is *relatively hierarchically hyperbolic* if there exists δ such that for all $U \in \mathfrak{S}$, either U is $\sqsubseteq$-minimal or $\mathcal{C}U$ is δ-hyperbolic. If U is $\sqsubseteq$-minimal and $\mathcal{C}U$ is not hyperbolic, then we insist that π_U is E-coarsely surjective.

Remark 6.9. One could, more generally, only insist that each $\pi_U(\mathcal{X})$ is a uniformly coarsely Lipschitz coarse retract. For hyperbolic $\mathcal{C}U$, this is equivalent to the uniform quasiconvexity from Definition 1.1, and is sufficient for our needs; for the present applications Definition 6.8 is sufficiently general, as well as for applications in [Behrstock et al. 2017a].

Our goal is to prove the following two theorems, which provide hierarchy paths and a distance formula in relatively hierarchically hyperbolic spaces. We will not use these theorems in the remainder of this paper, but they are required for future applications.

Theorem 6.10 (distance formula for relatively hierarchically hyperbolic spaces). *Let $(\mathcal{X}, \mathfrak{S})$ be a relatively hierarchically hyperbolic space. Then there exists s_0 such that for all $s \geq s_0$, there exist constants C, K such that for all $x, y \in \mathcal{X}$,*

$$\mathsf{d}_{\mathcal{X}}(x, y) \asymp_{K,C} \sum_{U \in \mathfrak{S}} \{\!\{\mathsf{d}_U(x, y)\}\!\}_s.$$

Proof. By Proposition 6.15 below, for some suitably chosen $\theta \geq 0$ and each $x, y \in \mathcal{X}$, there exists a subspace $M_\theta(x, y)$ of $\mathcal{X}$ (endowed with the induced metric) so that $(M_\theta(x, y), \mathfrak{S})$ is a hierarchically hyperbolic space (with the same nesting relations and projections from $(\mathcal{X}, \mathfrak{S})$, so that for all $U \in \mathfrak{S}$, we have that $\pi_U(M_\theta(x, y)) \subset \mathcal{N}_\theta(\gamma_U)$, where γ_U is an arbitrarily chosen geodesic in $\mathcal{C}U$ from $\pi_U(x)$ to $\pi_U(y)$. We emphasize that all of the constants from Definition 1.1 (for $M_\theta(x, y)$) are independent of x, y. The theorem now follows by applying the distance formula for hierarchically hyperbolic spaces (Theorem 4.5) to $(M_\theta(x, y), \mathfrak{S})$. $\square$

Theorem 6.11 (hierarchy paths in relatively hierarchically hyperbolic spaces). *Let $(\mathcal{X}, \mathfrak{S})$ be a relatively hierarchically hyperbolic space. Then there exists $D \geq 0$ such that for all $x, y \in \mathcal{X}$, there exists a (D, D)-quasigeodesic γ in $\mathcal{X}$ joining x, y so that $\pi_U(\gamma)$ is an unparameterized (D, D)-quasigeodesic.*

Proof. Proceed exactly as in Theorem 6.10, but apply Theorem 4.4 instead of Theorem 4.5. □

We now define hulls of pairs of points in the relatively hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$. Let θ be a constant to be chosen (it will be the output of the realization theorem for a consistency constant depending on the constants associated to $(\mathcal{X}, \mathfrak{S})$), and let $x, y \in \mathcal{X}$. For each $U \in \mathfrak{S}$, fix a geodesic γ_U in $\mathcal{C}U$ joining $\pi_U(x)$ to $\pi_U(y)$. Define maps $r_U : \mathcal{C}U \to \gamma_U$ as follows: if $\mathcal{C}U$ is hyperbolic, let r_U be the coarse closest-point projection map. Otherwise, if $\mathcal{C}U$ is not hyperbolic (so U is $\sqsubseteq$-minimal), define r_U as follows: parametrize γ_U by arc length with $\gamma_U(0) = x$, and for each $p \in \mathcal{C}U$, let $m(p) = \min\{\mathrm{d}_U(x, p), \mathrm{d}_U(x, y)\}$. Then $r_U(p) = \gamma_U(m(p))$. This r_U is easily seen to be an L-coarsely Lipschitz retraction, with L independent of U and x, y. (When U is minimal, r_U is 1-Lipschitz.)

Next, define the hull $M_\theta(x, y)$ to be the set of points $x \in \mathcal{X}$ such that $\mathrm{d}_U(x, \gamma_U) \leq \theta$ for all $U \in \mathfrak{S}$. In the next proposition, we show that $M_\theta(x, y)$ is a hierarchically hyperbolic space, with the following hierarchically hyperbolic structure:

(1) The index set is $\mathfrak{S}$.

(2) The nesting, orthogonality, and transversality relations on $\mathfrak{S}$ are the same as in $(\mathcal{X}, \mathfrak{S})$.

(3) For each $U \in \mathfrak{S}$, the associated hyperbolic space is γ_U.

(4) For each $U \in \mathfrak{S}$, the projection $\pi'_U : M_\theta(x, y) \to \gamma_U$ is given by $\pi'_U = r_U \circ \pi_U$.

(5) For each pair $U, V \in \mathfrak{S}$ of distinct nonorthogonal elements, the relative projection $\mathcal{C}U \to \mathcal{C}V$ is given by $r_V \circ \rho^U_V$.

Since there are now two sets of projections (those defined in the original hierarchical space $(\mathcal{X}, \mathfrak{S})$, denoted π_*, and the new projections π'_*), in the following proofs we will explicitly write all projections when writing distances in the various $\mathcal{C}U$.

Lemma 6.12 (gates in hulls). *Let $M_\theta(x, y)$ be as above. Then there exists a uniformly coarsely Lipschitz retraction $r : \mathcal{X} \to M_\theta(x, y)$ such that for each $U \in \mathfrak{S}$, we have, up to uniformly (independent of x, y) bounded error, $\pi_U \circ r = r_U \circ \pi_U$.*

Remark 6.13. It is crucial in the following proof that $\mathcal{C}U$ is δ-hyperbolic for each $U \in \mathfrak{S}$ that is not $\sqsubseteq$-minimal.

Proof of Lemma 6.12. Let $z \in \mathcal{X}$ and, for each U, let $t_U = r_U \circ \pi_U(z)$; this defines a tuple $(t_U) \in \prod_{U \in \mathfrak{S}} 2^{\mathcal{C}U}$ which we will check is κ-consistent for κ independent of x, y. The tuple (t_U) is admissible because of quasiconvexity of the images of projections to hyperbolic spaces, and coarse surjectivity of projections to nonhyperbolic ones.

Realization (Theorem 3.1) then yields $m \in \mathcal{X}$ such that $\mathsf{d}_U(\pi_U(m), t_U) \le \theta$ for all $U \in \mathfrak{S}$. By definition, $t_U \in \gamma_U$, so $m \in M_\theta(x, y)$ and we define $\mathfrak{g}_{x,y}(z) = m$. Note that up to perturbing slightly, we may take $\mathfrak{g}_{x,y}(z) = z$ when $z \in M_\theta$. Hence it suffices to check consistency of (t_U).

First let $U, V \in \mathfrak{S}$ satisfy $U \pitchfork V$. Then $\mathsf{d}_V(\pi_V(x), \pi_V(y)) \le 2E$ (up to exchanging U and V), and moreover each of $\pi_U(x), \pi_U(y)$ is E-close to ρ^U_V. Since t_V lies on γ_V, it follows that $\mathsf{d}_V(t_V, \rho^U_V) \le 2E$.

Next, let $U, V \in \mathfrak{S}$ satisfy $U \sqsubsetneq V$. Observe that in this case, $\mathcal{C}V$ is δ-hyperbolic because V is not $\sqsubseteq$-minimal. First suppose that $\mathsf{d}_V(\gamma_V, \rho^U_V) > 1$. Then by consistency and bounded geodesic image, $\mathsf{d}_U(x, y) \le 3E$, and $\mathsf{diam}_U(\rho^V_U(\gamma_V)) \le E$. It follows that $\mathsf{diam}_U(t_U \cup \rho^U_V(t_V)) \le 10E$.

Hence, suppose that $\mathsf{d}_V(\rho^U_V, \gamma_V) \le 10E$ but that $\mathsf{d}_V(t_V, \rho^U_V) > E$. Without loss of generality, ρ^U_V lies at distance $\le E$ from the subpath of γ_V joining t_V to $\pi_V(y)$. Let γ'_V be the subpath joining x to t_V. By consistency, bounded geodesic image, and the fact that $\mathcal{C}V$ is δ-hyperbolic and $t_V = r_V \circ \pi_V(z)$, the geodesic triangle between $\pi_V(x), \pi_V(z)$, and t_V projects under ρ^V_U to a set, of diameter bounded by some uniform ξ, containing $\pi_U(x), \pi_U(z)$, and $\rho^V_U(t_V)$. Hence, since $t_U = r_U \circ \pi_U(z)$, and $\pi_U(x) \in \gamma_U$, the triangle inequality yields a uniform bound on $\mathsf{diam}_U(t_U \cup \rho^V_U(t_V))$. Hence there exists a uniform κ, independent of x, y, so that (t_U) is κ-consistent. Finally, $\mathfrak{g}_{x,y}$ is coarsely Lipschitz by the uniqueness axiom (Definition 1.1.(9)), since each r_U is uniformly coarsely Lipschitz. □

Lemma 6.14. *Let $m, m' \in M_\theta(x, y)$. Then there exists $C \ge 0$ such that m, m' are joined by a (C, C)-quasigeodesic in $M_\theta(x, y)$.*

Proof. Since $\mathcal{X}$ is a quasigeodesic space, there exists $K \ge 0$ so that m, m' are joined by a K-discrete (K, K)-quasigeodesic $\sigma : [0, \ell] \to \mathcal{X}$ with $\sigma(0) = m, \sigma(\ell) = m'$. Note that $\mathfrak{g}_{x,y} \circ \sigma$ is a K'-discrete, efficient path for K' independent of x, y, since the gate map is uniformly coarsely Lipschitz. A minimal-length K'-discrete efficient path in $M_\theta(x, y)$ from x to y has the property that each subpath is K'-efficient, and is a uniform quasigeodesic, as needed. □

Proposition 6.15. *For all sufficiently large θ, (1)–(5) above make $(M_\theta(x, y), \mathfrak{S})$ a hierarchically hyperbolic space, where $M_\theta(x, y)$ inherits its metric as a subspace of $\mathcal{X}$. Moreover, the associated constants from Definition 1.1 are independent of x, y.*

Proof. By Lemma 6.14, $M_\theta(x, y)$ is a uniform quasigeodesic space. We now verify that the enumerated axioms from Definition 1.1 are satisfied. Each part of the definition involving only $\mathfrak{S}$ and the $\sqsubseteq, \perp, \pitchfork$ relations is obviously satisfied; this includes finite complexity. The consistency inequalities hold because they hold in $(\mathcal{X}, \mathfrak{S})$ and each r_U is L-coarsely Lipschitz. The same holds for bounded geodesic image and the large link lemma. We now verify the two remaining claims:

Uniqueness. Let $m, m' \in M_\theta(x, y)$, so that $\mathsf{d}_U(\pi_U(m), \gamma_U), \mathsf{d}_U(\pi_U(m'), \gamma_U) \leq \theta$ for all $U \in \mathfrak{S}$. The definition of r_U ensures that $\mathsf{d}_U(r_U \circ \pi_U(m), r_U \circ \pi_U(m')) \geq \mathsf{d}_U(\pi_U(m), \pi_U(m')) - 2\theta$, and uniqueness follows.

Partial realization. Let $\{U_i\}$ be a totally orthogonal subset of $\mathfrak{S}$ and choose, for each i, some $p_i \in \gamma_{U_i}$. By partial realization in $(\mathcal{X}, \mathfrak{S})$, there exists $z \in \mathcal{X}$ so that $\mathsf{d}_{U_i}(\pi_{U_i}(z), p_i) \leq E$ for each i and $\mathsf{d}_V(\pi_V(z), \rho_V^{U_i}) \leq E$ provided $U_i \subsetneq V$ or $U_i \pitchfork V$. Let $z' = \mathfrak{g}_{x,y}(z) \in M_\theta(x, y)$. Then, by the definition of the gate map and the fact that each r_U is L-coarsely Lipschitz, there exists α, independent of x, y, so that $\mathsf{d}_{U_i}(r_{U_i} \circ \pi_{U_i}(z'), p_i) \leq \alpha$, while $\mathsf{d}_V(r_V \circ \pi_V(z'), \rho_V^{U_i}) \leq \alpha$ whenever $U_i \pitchfork V$ or $U_i \sqsubseteq V$. Hence z' is the required partial realization point. This completes the proof that $(M_\theta(x, y), \mathfrak{S})$ is an HHS. $\square$

7. The coarse median property

In this section, we study the relationship between hierarchically hyperbolic spaces and spaces that are *coarse median* in the sense defined in [Bowditch 2013]. In particular, this discussion shows that $\mathrm{Out}(F_n)$ is not a hierarchically hyperbolic space, and hence not a hierarchically hyperbolic group, for $n \geq 3$.

Definition 7.1 (median graph). Let Γ be a graph with unit-length edges and path-metric d. Then Γ is a *median graph* if there is a map $\mathfrak{m} : \Gamma^3 \to \Gamma$ such that, for all $x, y, z \in \Gamma$, we have $\mathsf{d}(x, y) = \mathsf{d}(x, m) + \mathsf{d}(m, y)$, and likewise for the pairs x, z and y, z, where $m = \mathfrak{m}(x, y, z)$. Note that if $x = y$, then $\mathfrak{m}(x, y, z) = x$.

Chepoi [2000] established that Γ is a median graph precisely when Γ is the 1-skeleton of a CAT(0) cube complex.

Definition 7.2 (coarse median space). Let (M, d) be a metric space and let $\mathfrak{m} : M^3 \to M$ be a ternary operation satisfying the following:

(1) (triples) There exist constants $\kappa, h(0)$ such that for all $a, a', b, b', c, c' \in M$,

$$\mathsf{d}(\mathfrak{m}(a, b, c,), \mathfrak{m}(a', b', c')) \leq \kappa(\mathsf{d}(a, a') + \mathsf{d}(b, b') + \mathsf{d}(c, c')) + h(0).$$

(2) (tuples) There is a function $h : \mathbb{N} \cup \{0\} \to [0, \infty)$ such that for any $A \subseteq M$ with $1 \leq |A| = p < \infty$, there is a CAT(0) cube complex $\mathcal{F}_p$ and maps $\pi : A \to \mathcal{F}_p^{(0)}$ and $\lambda : \mathcal{F}_p^{(0)} \to M$ such that $\mathsf{d}(a, \lambda(\pi(a))) \leq h(p)$ for all $a \in A$ and such that

$$\mathsf{d}(\lambda(\mathfrak{m}_p(x, y, z)), \mathfrak{m}(\lambda(x), \lambda(y), \lambda(z))) \leq h(p)$$

for all $x, y, z \in \mathcal{F}_p$, where $\mathfrak{m}_p$ is the map that sends triples from $\mathcal{F}_p^{(0)}$ to their median.

Then $(M, \mathsf{d}, \mathfrak{m})$ is a *coarse median space*. The *rank* of $(M, \mathsf{d}, \mathfrak{m})$ is at most d if each $\mathcal{F}_p$ above can be chosen to satisfy $\dim \mathcal{F}_p \leq d$, and the rank of $(M, \mathsf{d}, \mathfrak{m})$ is exactly d if d is the minimal integer such that $(M, \mathsf{d}, \mathfrak{m})$ has rank at most d.

The next fact was observed by Bowditch [2018]; we include a proof for completeness.

Theorem 7.3 (hierarchically hyperbolic implies coarse median). *Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space. Then $\mathcal{X}$ is coarse median of rank at most the complexity of $(\mathcal{X}, \mathfrak{S})$.*

Proof. Since the spaces $\mathcal{C}U$, $U \in \mathfrak{S}$ are δ-hyperbolic for some δ independent of U, there exists for each U a ternary operation $\mathfrak{m}^U : \mathcal{C}U^3 \to \mathcal{C}U$ so that $(\mathcal{C}U, \mathsf{d}_U, \mathfrak{m}^U)$ is a coarse median space of rank 1, and the constant κ and function $h : \mathbb{N} \cup \{0\} \to [0, \infty)$ from Definition 7.2 can be chosen to depend only on δ (and not on U).

Definition of the median. Define a map $\mathfrak{m} : \mathcal{X}^3 \to \mathcal{X}$ as follows. Let $x, y, z \in \mathcal{X}$ and, for each $U \in \mathfrak{S}$, let $b_U = \mathfrak{m}^U(\pi_U(x), \pi_U(y), \pi_U(z))$. By Lemma 2.6, the tuple $\vec{b} \in \prod_{U \in \mathfrak{S}} 2^{\mathcal{C}U}$ whose U-coordinate is b_U is κ-consistent for an appropriate choice of κ. Hence, by the realization theorem (Theorem 3.1), there exists θ_e and $\mathfrak{m} = \mathfrak{m}(x, y, z) \in \mathcal{X}$ such that $\mathsf{d}_U(\mathfrak{m}, b_U) \leq \theta_u$ for all $U \in \mathfrak{S}$. Moreover, this is coarsely well defined (up to the constant θ_e from the realization theorem).

Application of [Bowditch 2013, Proposition 10.1]. Note, by Definition 1.1.(1), the projections $\pi_U : \mathcal{X} \to \mathcal{C}U$, $U \in \mathfrak{S}$ are uniformly coarsely Lipschitz. Moreover, for each $U \in \mathfrak{S}$, the projection $\pi_U : \mathcal{X} \to \mathcal{C}U$ is a "quasimorphism" in the sense of [Bowditch 2013, Section 10], i.e., $\mathsf{d}_U(\mathfrak{m}^U(\pi_U(x), \pi_U(y), \pi_U(z)), \pi_U(\mathfrak{m}(x, y, z)))$ is uniformly bounded, by construction, as U varies over $\mathfrak{S}$ and x, y, z vary in $\mathcal{X}$. Proposition 10.1 of [Bowditch 2013] then implies that $\mathfrak{m}$ is a coarse median on $\mathcal{X}$, since that the hypothesis $(P1)$ of that proposition holds in our situation by the distance formula. □

The following is a consequence of Theorem 7.3 and work of Bowditch [2013; 2014a]:

Corollary 7.4 (contractibility of asymptotic cones). *Let $\mathcal{X}$ be a hierarchically hyperbolic space. Then all the asymptotic cones of $\mathcal{X}$ are contractible, and in fact bi-Lipschitz equivalent to CAT(0) spaces.*

Corollary 7.5 (HHGs have quadratic Dehn function). *Let G be a finitely generated group that is a hierarchically hyperbolic space. Then G is finitely presented and has quadratic Dehn function. In particular, this conclusion holds when G is a hierarchically hyperbolic group.*

Proof. This follows from Theorem 7.3 and [Bowditch 2013, Corollary 8.3]. □

Corollary 7.6. *For $n \geq 3$, the group $\mathrm{Out}(F_n)$ is not a hierarchically hyperbolic space, and in particular is not a hierarchically hyperbolic group.*

Proof. This is an immediate consequence of Corollary 7.5 and the exponential lower bound on the Dehn function of $\mathrm{Out}(F_n)$ given by the combined results of [Bridson and Vogtmann 1995; 2012; Handel and Mosher 2013b]. □

We also recover a special case of Theorem I of [Behrstock et al. 2017b], using Corollary 7.5 and a theorem of Gersten, Holt and Riley [Gersten et al. 2003, Theorem A]:

Corollary 7.7. *Let N be a finitely generated virtually nilpotent group. Then G is quasi-isometric to a hierarchically hyperbolic space if and only if N is virtually abelian.*

Corollary 7.8. *Let S be a symmetric space of noncompact type, or a thick affine building. Suppose that the spherical type of S is not A_1^r. Then S is not hierarchically hyperbolic.*

Proof. This follows from Theorem 7.3 and Theorem A of [Haettel 2016]. □

Finally, Theorem 9.1 of [Bowditch 2014a] combines with Theorem 7.3 to yield:

Corollary 7.9 (rapid decay). *Let G be a group whose Cayley graph is a hierarchically hyperbolic space. Then G has the rapid decay property.*

7A. ***Coarse media and hierarchical quasiconvexity.*** The natural notion of quasiconvexity in the coarse median setting is related to hierarchical quasiconvexity.

Definition 7.10 (coarsely convex). Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space and let $\mathfrak{m} : \mathcal{X}^3 \to \mathcal{X}$ be the coarse median map constructed in the proof of Theorem 7.3. A closed subspace $\mathcal{Y} \subseteq \mathcal{X}$ is *μ-convex* if for all $y, y' \in \mathcal{Y}$ and $x \in \mathcal{X}$, we have $\mathfrak{m}(y, y', x) \in \mathcal{N}_\mu(\mathcal{Y})$.

Remark 7.11. We will not use μ-convexity in the remainder of the paper. However, it is of independent interest since it parallels a characterization of convexity in median spaces: a subspace $\mathcal{Y}$ of a median space is convex exactly when, for all $y, y' \in \mathcal{Y}$ and x in the ambient median space, the median of x, y, y' lies in $\mathcal{Y}$.

Proposition 7.12 (coarse convexity and hierarchical quasiconvexity). *Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space and let $\mathcal{Y} \subseteq \mathcal{X}$. If $\mathcal{Y}$ is hierarchically k-quasiconvex, then there exists $\mu \geq 0$, depending only on k and the constants from Definition 1.1, such that $\mathcal{Y}$ is μ-convex.*

Proof. Let $\mathcal{Y} \subseteq \mathcal{X}$ be k-hierarchically quasiconvex, let $y, y' \in \mathcal{Y}$ and $x \in \mathcal{X}$. Let $m = \mathfrak{m}(x, y, y')$. For any $U \in \mathfrak{S}$, the projection $\pi_U(\mathcal{Y})$ is by definition $k(0)$-quasiconvex, so that, for some $k' = k'(k(0), \delta)$, we have $\mathrm{d}_U(m_U, \pi_U(\mathcal{Y})) \leq k'$, where m_U is the coarse median of $\pi_U(x), \pi_U(y), \pi_U(y')$ coming from hyperbolicity of $\mathcal{C}U$.

The tuple $(m_U)_{U\in\mathfrak{S}}$ was shown above to be κ-consistent for appropriately chosen κ (Lemma 2.6), and $\mathsf{d}_U(m_U, \mathfrak{m}(x, y, y')) \le \theta_e(\kappa)$, so, by hierarchical quasiconvexity $\mathsf{d}_{\mathcal{X}}(\mathfrak{m}(x, y, y'), \mathcal{Y})$ is bounded by a constant depending on $k(\kappa)$ and k'. $\square$

8. Combination theorems for hierarchically hyperbolic spaces

The goal of this section is to prove Theorem 8.6, which enables the construction of new hierarchically hyperbolic spaces and groups from a tree of given ones. We postpone the statement of the theorem until after the relevant definitions.

Definition 8.1 (quasiconvex hieromorphism, full hieromorphism). Let

$$(f, f^{\diamond}, \{f^*(U)\}_{U\in\mathfrak{S}})$$

be a hieromorphism $(\mathcal{X}, \mathfrak{S}) \to (\mathcal{X}', \mathfrak{S}')$. We say f is *k-hierarchically quasiconvex* if its image is k-hierarchically quasiconvex and $f : \mathcal{X} \to \mathcal{X}'$ is a quasi-isometric embedding. The hieromorphism is *full* if

(1) there exists $\xi \ge 0$ such that each $f^*(U) : \mathcal{C}U \to \mathcal{C}(f^{\diamond}(U))$ is a (ξ, ξ)-quasi-isometry, and

(2) for each $U \in \mathfrak{S}$, if $V' \in \mathfrak{S}'$ satisfies $V' \sqsubseteq f^{\diamond}(U)$, then there exists $V \in \mathfrak{S}$ such that $V \sqsubseteq U$ and $f^{\diamond}(V) = V'$.

Remark 8.2. Observe that Definition 8.1.(2) holds automatically unless V' is bounded.

Definition 8.3 (tree of hierarchically hyperbolic spaces). Let $\mathcal{V}, \mathcal{E}$ denote the vertex and edge-sets, respectively, of the simplicial tree T. A *tree of hierarchically hyperbolic spaces* is a quadruple

$$\mathcal{T} = (T, \{\mathcal{X}_v\}, \{\mathcal{X}_e\}, \{\phi_{e_\pm} : v \in \mathcal{V}, e \in \mathcal{E}\})$$

satisfying:

(1) $\{\mathcal{X}_v\}$ and $\{\mathcal{X}_e\}$ are uniformly hierarchically hyperbolic: each $\mathcal{X}_v$ has index set $\mathfrak{S}_v$, and each $\mathcal{X}_e$ has index set $\mathfrak{S}_e$. In particular, there is a uniform bound on the complexities of the hierarchically hyperbolic structures on the $\mathcal{X}_v$ and $\mathcal{X}_e$.

(2) Fix an orientation on each $e \in \mathcal{E}$ and let e_+, e_- denote the initial and terminal vertices of e. Then, each $\phi_{e_\pm} : \mathcal{X}_e \to \mathcal{X}_{e_\pm}$ is a hieromorphism with all constants bounded by some uniform $\xi \ge 0$. (We adopt the hieromorphism notation from Definition 1.20. Hence we actually have maps $\phi_{e_\pm} : \mathcal{X}_e \to \mathcal{X}_{e_\pm}$, and maps $\phi^{\diamond}_{e_\pm} : \mathfrak{S}_e \to \mathfrak{S}_{e_\pm}$ preserving nesting, transversality, and orthogonality, and coarse ξ-Lipschitz maps $\phi^*_{e_\pm}(U) : \mathcal{C}U \to \mathcal{C}(\phi^{\diamond}_{e_\pm}(U))$ satisfying the conditions of that definition.)

Given a tree $\mathcal{T}$ of hierarchically hyperbolic spaces, denote by $\mathcal{X}(\mathcal{T})$ the metric space constructed from $\bigsqcup_{v\in\mathcal{V}} \mathcal{X}_v$ by adding edges of length 1 as follows: if $x \in \mathcal{X}_e$, we declare $\phi_{e_-}(x)$ to be joined by an edge to $\phi_{e_+}(x)$. Given $x, x' \in \mathcal{X}$ in the same vertex space $\mathcal{X}_v$, define $\mathsf{d}'(x, x')$ to be $\mathsf{d}_{\mathcal{X}_v}(x, x')$. Given $x, x' \in \mathcal{X}$ joined by an edge, define $\mathsf{d}'(x, x') = 1$. Given a sequence $x_0, x_1, \ldots, x_k \in \mathcal{X}$, with consecutive points either joined by an edge or in a common vertex space, define its length to be $\sum_{i=1}^{k-1} \mathsf{d}'(x_i, x_{i+1})$. Given $x, x' \in \mathcal{X}$, let $\mathsf{d}(x, x')$ be the infimum of the lengths of such sequences $x = x_0, \ldots, x_k = x'$.

Remark 8.4. Since the vertex spaces are (uniform) quasigeodesic spaces, $(\mathcal{X}, \mathsf{d})$ is a quasigeodesic space.

Definition 8.5 (equivalence, support, bounded support). Let $\mathcal{T}$ be a tree of hierarchically hyperbolic spaces. For each $e \in \mathcal{E}$, and each $W_{e_-} \in \mathfrak{S}_{e_-}$, $W_{e_+} \in \mathfrak{S}_{e_+}$, write $W_{e_-} \sim_d W_{e_+}$ if there exists $W_e \in \mathfrak{S}_e$ so that $\phi^{\diamond}_{e_\pm}(W_e) = W_{e_\pm}$. The transitive closure $\sim$ of $\sim_d$ is an equivalence relation on $\bigcup_v \mathfrak{S}_v$. The $\sim$-class of $W \in \bigcup_v \mathfrak{S}_v$ is denoted $[W]$.

The *support* of an equivalence class $[W]$ is the induced subgraph $T_{[W]}$ of T whose vertices are those $v \in T$ so that $\mathfrak{S}_v$ contains a representative of $[W]$. Observe that $T_{[W]}$ is connected. The tree $\mathcal{T}$ of hierarchically hyperbolic spaces has *bounded supports* if there exists $n \in \mathbb{N}$ such that each $\sim$-class has support of diameter at most n.

We can now state the main theorem of this section:

Theorem 8.6 (combination theorem for hierarchically hyperbolic spaces). *Let $\mathcal{T}$ be a tree of hierarchically hyperbolic spaces. Suppose that*:

(1) *There exists a function k so that each edge-hieromorphism is k-hierarchically quasiconvex.*

(2) *Each edge-hieromorphism is full.*

(3) *$\mathcal{T}$ has bounded supports of diameter at most n.*

(4) *If e is an edge of T and S_e is the $\sqsubseteq$-maximal element of $\mathfrak{S}_e$, then for all $V \in \mathfrak{S}_{e^\pm}$, the elements V and $\phi^{\diamond}_{e^\pm}(S_e)$ are not orthogonal in $\mathfrak{S}_{e^\pm}$. Moreover, there exists $K \geq 0$ such that for all vertices v of $\mathcal{T}$ and edges e incident to v, we have $\mathsf{d}_{\mathrm{Haus}}(\phi_v(\mathcal{X}_e)), \boldsymbol{F}_{\phi^{\diamond}_v(S_e)} \times \{\star\}) \leq K$, where $S_e \in \mathfrak{S}_e$ is the unique maximal element and $\star \in \boldsymbol{E}_{\phi^{\diamond}_v(S_e)}$.*

Then $\mathcal{X}(\mathcal{T})$ is hierarchically hyperbolic.

We postpone the proof until after the necessary lemmas and definitions. For the remainder of this section, fix a tree of hierarchically hyperbolic spaces $\mathcal{T} = (T, \{\mathcal{X}_v\}, \{\mathcal{X}_e\}, \{\phi_{e^\pm}\})$ satisfying the hypotheses of Theorem 8.6; let n be the constant implicit in (3).

Let $\mathfrak{S}^0 = \{T\} \cup \left(\bigcup_v \mathfrak{S}_v/\sim\right)$.

Definition 8.7 (nesting, orthogonality, transversality in $\mathfrak{S}^0$). For all $[W] \in \mathfrak{S}$, declare $[W] \sqsubseteq T$. If $[V], [W]$ are $\sim$-classes, then $[V] \sqsubseteq [W]$ if and only if there exists $v \in T$ such that $[V], [W]$ are respectively represented by $V_v, W_v \in \mathfrak{S}_v$ and $V_v \sqsubseteq W_v$; this relation is *nesting*. For convenience, for $A \in \mathfrak{S}$, we write $\mathfrak{S}_A$ to denote the set of $B \in \mathfrak{S}^0$ such that $B \sqsubseteq A$.

Likewise, $[V] \perp [W]$ if and only if there exists a vertex $v \in T$ such that $[V], [W]$ are respectively represented by $V_v, W_v \in \mathfrak{S}_v$ and $V_v \perp W_v$; this relation is *orthogonality*. If $[V], [W] \in \mathfrak{S}$ are not orthogonal and neither is nested into the other, then they are *transverse*, written $[V] \pitchfork [W]$. Equivalently, $[V] \pitchfork [W]$ if for all $v \in T_{[V]} \cap T_{[W]}$, the representatives $V_v, W_v \in \mathfrak{S}_v$ of $[V], [W]$ satisfy $V_v \pitchfork W_v$.

Fullness (Definition 8.1.(2)) was introduced to enable the following two lemmas:

Lemma 8.8. *Let $\mathcal{T}$ be a tree of hierarchically hyperbolic spaces, let v be a vertex of the underlying tree T, and let $U, U' \in \mathfrak{S}_v$ satisfy $U \sqsubseteq U'$. Then either $U = U'$ or $U \not\sim U'$.*

Proof. Suppose that $U \sim U'$, so that there is a closed path $v = v_0, v_1, \ldots, v_n = v$ in T and a sequence $U = U_0, U_1, \ldots, U_n = U'$ such that $U_i \in \mathfrak{S}_{v_i}$ and $U_i \sim_d U_{i+1}$ for all i. If $U \neq U'$, then condition (2) (fullness) from Definition 8.1 and the fact that hieromorphisms preserve nesting yields $U'' \in \mathfrak{S}_v$, different from U', such that $U'' \sim U$ and $U'' \sqsubsetneq U' \sqsubsetneq U$ (where $\sqsubsetneq$ denotes proper nesting). Repeating this argument contradicts finiteness of complexity. $\square$

Lemma 8.9. *The relation $\sqsubseteq$ is a partial order on $\mathfrak{S}^0$, and T is the unique $\sqsubseteq$-maximal element. Moreover, if $[V] \perp [W]$ and $[U] \sqsubseteq [V]$, then $[U] \perp [W]$ and $[V], [W]$ are not $\sqsubseteq$-comparable.*

Proof. Reflexivity is clear. Suppose that $[V_v] \sqsubseteq [U_u] \sqsubseteq [W_w]$. Then there are vertices $v_1, v_2 \in \mathcal{V}$ and representatives $V_{v_1} \in [V_v]$, $U_{v_1} \in [U_u]$, $U_{v_2} \in [U_u]$, $W_{v_2} \in [W_w]$ so that $V_{v_1} \sqsubseteq U_{v_1}$ and $U_{v_2} \sqsubseteq W_{v_2}$. Since edge-hieromorphisms are full, induction on $\mathrm{d}_T(v_1, v_2)$ yields $V_{v_2} \sqsubseteq U_{v_2}$ so that $V_{v_2} \sim V_{v_1}$. Transitivity of the nesting relation in $\mathfrak{S}_{v_2}$ implies that $V_{v_2} \sqsubseteq W_{v_2}$, whence $[V_v] \sqsubseteq [W_w]$.

Suppose that $[U_u] \sqsubseteq [V_v]$ and $[V_v] \sqsubseteq [U_u]$, and suppose by contradiction that $[U_u] \neq [V_v]$. Choose $v_1, v_2 \in \mathcal{V}$ and representatives $U_{v_1}, U_{v_2}, V_{v_1}, V_{v_2}$ so that $U_{v_1} \sqsubseteq V_{v_1}$ and $V_{v_2} \sqsubseteq U_{v_2}$. The definition of $\sim$ again yields $U_{v_2} \sim U_{v_1}$ with $U_{v_2} \sqsubseteq V_{v_2} \neq U_{v_2}$. This contradicts Lemma 8.8. Hence $\sqsubseteq$ is antisymmetric, whence it is a partial order. The underlying tree T is the unique $\sqsubseteq$-maximal element by definition.

Suppose that $[V] \perp [W]$ and $[U] \sqsubseteq [V]$. Then there are vertices v_1, v_2 and representatives $V_{v_1}, W_{v_1}, U_{v_2}, V_{v_2}$ such that $V_{v_1} \perp W_{v_1}$ and $U_{v_2} \perp V_{v_2}$. Again by fullness of the edge-hieromorphisms, there exists $U_{v_1} \sim U_{v_2}$ with $U_{v_1} \sqsubseteq V_{v_1}$, whence $U_{v_1} \perp W_{v_1}$.

Thus $[U]\perp[W]$ as required. Also, $\sqsubseteq$-incomparability of $[V]$, $[W]$ follows from fullness and the fact that edge-hieromorphisms preserve orthogonality and nesting. $\square$

Lemma 8.10. *Let $[W]\in\mathfrak{S}^0$ and let $[U]\sqsubseteq[W]$. Suppose moreover that*

$$\{[V]\in\mathfrak{S}_{[W]}:[V]\perp[U]\}\neq\varnothing.$$

Then there exists $[A]\in\mathfrak{S}_{[W]}-\{[W]\}$ such that $[V]\sqsubseteq[A]$ for all $[V]\in\mathfrak{S}_{[W]}$ with $[V]\perp[U]$.

Proof. Choose some $v\in\mathcal{V}$ so that there exist $V_v\in\mathfrak{S}_v$ and $U_v\in\mathfrak{S}_v$ with $[U_v]=[U]$ and $V_v\perp U_v$. Then by definition, there exists $A_v\in\mathfrak{S}_v$ so that $B_v\sqsubseteq A_v$ whenever $B_v\perp U_v$ and so that $[B_v]\sqsubseteq[W]$. It follows from the fact that the edge-hieromorphisms are full and preserve (non)orthogonality that $[B]\sqsubseteq[A_v]$ whenever $[B]\perp[U]$. $\square$

The set $\mathfrak{S}^0$ is not quite large enough to satisfy the orthogonality axiom, for the following reason: in Lemma 8.10, we needed $[W]$ to be a $\sim$-class, but since $T\in\mathfrak{S}^0$, we need to be able to satisfy the axiom with $[W]$ replaced by T. To this end, we add some new elements to $\mathfrak{S}^0$, and extend the $\sqsubseteq$, $\perp$, $\pitchfork$ relations, as follows.

Definition 8.11 (containers and $\mathfrak{S}$). We now define the index set $\mathfrak{S}$ for the HHS structure we will construct in order to prove Theorem 8.6. First, $\mathfrak{S}$ contains $\mathfrak{S}^0$. Next, for each $[W]$ for which there exists $[U]$ with $[U]\perp[W]$, let $\mathrm{K}_0^{\perp}([W])$ be a new element of $\mathfrak{S}$, which we call the *container* of $[W]$. We make the following declarations:

- $\mathrm{K}_0^{\perp}([W])\sqsubseteq T$.
- $[U]\sqsubseteq\mathrm{K}_0^{\perp}([W])$ if and only if $[U]\perp[W]$.
- $\mathrm{K}_0^{\perp}([W])\pitchfork\mathrm{K}_0^{\perp}([U])$ if $[U]\neq[W]$.
- $\mathrm{K}_0^{\perp}([W])\perp[V]$ if and only if $[V]\sqsubseteq[W]$.
- for all other $[U]$, we have $[U]\pitchfork\mathrm{K}_0^{\perp}([W])$.

Let $\mathcal{K}_0$ be the set of all $\mathrm{K}_0^{\perp}([W])$ as $[W]$ varies among those $\sim$-classes for which there is at least one orthogonal $\sim$-class.

Next, for each $\mathrm{K}_0^{\perp}([W])\in\mathcal{K}_0$, consider a $\sim$-class $[U]\sqsubseteq\mathrm{K}_0^{\perp}([W])$ such that $[U]\perp[V]$ for some other $[V]\sqsubseteq\mathrm{K}_0^{\perp}([W])$. Let $\mathrm{K}_1^{\perp}([W],[U])$ be a new element of $\mathfrak{S}$, and let $\mathcal{K}_1$ be the set of such containers, as $[W]$ varies and as $[U]$ varies over those $\sim$-classes nested in $\mathrm{K}_0^{\perp}([W])$ (i.e., orthogonal to $[W]$) that are orthogonal to some other $\sim$-class nested in $\mathrm{K}_0^{\perp}([W])$.

We now make the following declarations:

- $\mathrm{K}_1^{\perp}([W],[U])\sqsubseteq\mathrm{K}_0^{\perp}([W])$ and $\mathrm{K}_1^{\perp}([W],[U])$ is transverse to every other element of $\mathcal{K}_0\cup\mathcal{K}_1$.

- $\mathrm{K}_1^{\perp}([W],[U]) \sqsubseteq T$.
- $[V] \sqsubseteq \mathrm{K}_1^{\perp}([W],[U])$ if and only if $[V] \sqsubseteq \mathrm{K}_0^{\perp}([W])$ and $[V] \perp [U]$.
- $[V] \perp \mathrm{K}_1^{\perp}([W],[U])$ if and only if either $[V] \sqsubseteq [W]$ (i.e., $[V] \perp \mathrm{K}_0^{\perp}([W])$) or $[V] \sqsubseteq [U]$.
- If none of the two preceding conditions is satisfied by $[V]$, then $[V] \pitchfork \mathrm{K}_1^{\perp}([W],[U])$.

We now proceed as above to inductively construct sets $\mathcal{K}_\eta$, $\eta \geq 1$, of new "containers", where each $\mathrm{K}_\eta^{\perp}([W]_1,\dots,[W_\eta])$ is nested in $\mathrm{K}_i^{\perp}([W]_1,\dots,[W_i])$ for $i \leq \eta-1$, and also nested in T. Our inductive construction ensures that $[W_1],\dots,[W_\eta]$ are pairwise-orthogonal. The $\sim$-classes $[U]$ nested in $\mathrm{K}_\eta^{\perp}([W]_1,\dots,[W_\eta])$ are precisely those that are orthogonal to each of $[W_1],\dots,[W_\eta]$. The $\sim$-classes U orthogonal to $\mathrm{K}_\eta^{\perp}([W_1],\dots,[W_\eta])$ are precisely those $[U]$ nested into some $[W_i]$.

Let $\mathfrak{S} = \mathfrak{S}^0 \cup \bigcup_{\eta \geq 0} \mathcal{K}_\eta$.

Remark 8.12 (extension of $\sqsubseteq$, $\perp$, $\pitchfork$ satisfies the axioms). Lemma 8.9 shows that $\sqsubseteq$ is a partial order on $\mathfrak{S}^0$, and Definition 8.11 shows how to extend $\sqsubseteq$ to all of $\mathfrak{S}$. By construction, the extended $\sqsubseteq$ continues to be transitive. This follows from Lemma 8.9, the definition, and induction on the η in $\mathcal{K}_\eta$. By definition, T is still the unique $\sqsubseteq$-maximal element.

Now suppose that $[U] \sqsubseteq \mathrm{K}_\eta^{\perp}([W_1],\dots,[W_\eta])$ and $[V] \perp \mathrm{K}_\eta^{\perp}([W_1],\dots,[W_\eta])$. Then $[V] \sqsubseteq [W_i]$ for some i, and $[U] \perp [W_j]$ for all j. Lemma 8.9 implies $[U] \perp [V]$. On the other hand, $\mathrm{K}_\eta^{\perp}([W_1],\dots,[W_\eta])$ is never nested into any $\sim$-class or orthogonal to any element of $\bigcup_\eta \mathcal{K}_\eta$.

Lemma 8.13. *There exists $\chi \geq 0$ such that if $\{V_1,\dots,V_c\} \subset \mathfrak{S}$ consists of pairwise orthogonal or pairwise $\sqsubseteq$-comparable elements, then $c \leq \chi$. In particular, $\bigcup_{\eta \geq 0} \mathcal{K}_\eta = \bigcup_{\eta=0}^{(\chi-1)/2} \mathcal{K}_\eta$.*

Proof. For each $v \in T$, let χ_v be the complexity of $(\mathcal{X}_v, \mathfrak{S}_v)$ and let $\chi = 2\max_v \chi_v + 1$. Let $[V_1],\dots,[V_c] \in \mathfrak{S} - \{T\}$ be $\sim$-classes that are pairwise orthogonal or pairwise $\sqsubseteq$-comparable. The Helly property for trees yields a vertex v lying in the support of each $[V_i]$; let $V_i^v \in \mathfrak{S}_v$ represent $[V_i]$. Since edge-hieromorphisms preserve nesting, orthogonality, and transversality, $c \leq \chi_v$.

Any pairwise-orthogonal set in $\mathfrak{S}$ either has cardinality ≤ 1 or contains at most one element that is not a $\sim$-class, so the bound on pairwise-orthogonal sets is $\max_v \chi_v + 1$.

Hence it suffices to bound $\sqsubseteq$-chains in $\mathfrak{S}$. Any $\sqsubseteq$-chain $V_1 \sqsubseteq V_2 \sqsubseteq \cdots \sqsubseteq V_k$ has the property that, for some $0 \leq m \leq k$, the first m elements are $\sim$-classes, and the remaining elements lie in $\{T\} \cup \bigcup_\eta \mathcal{K}_\eta$. Hence it suffices to show that any $\sqsubseteq$-chain in $\bigcup_{\eta\geq 0}\mathcal{K}_\eta$ has length at most χ. But by definition, any such chain has the form

$$\mathrm{K}_0^\perp([W_0]) \sqsupset \mathrm{K}_1^\perp([W_0],[W_1]) \sqsupset \cdots \sqsupset \mathrm{K}_\eta^\perp([W_0],\ldots,[W_{\eta-1}]),$$

where the $[W_i]$ are pairwise orthogonal. Hence $\eta \leq \max_v \chi_v \leq (\chi-1)/2$, as required. This also proves the final assertion. □

Definition 8.14 (favorite representative, hyperbolic spaces associated to elements of $\mathfrak{S}$). Let $\mathcal{C}T = T$. For each $\sim$-class $[W]$, choose a *favorite vertex* v of $T_{[W]}$ and let $W_v \in \mathfrak{S}_{W_v}$ be the *favorite representative* of $[W]$. Let $\mathcal{C}[W] = \mathcal{C}W_v$. Note that each $\mathcal{C}[W]$ is δ-hyperbolic, where δ is the uniform hyperbolicity constant for $\mathcal{T}$.

Finally, for each $\mathrm{K}^\perp \in \bigcup_\eta \mathcal{K}_\eta$, let $\mathcal{C}\mathrm{K}^\perp$ be a single point.

Definition 8.15 (gates in vertex spaces). For each vertex v of T, define a *gate map* $\mathfrak{g}_v : \mathcal{X} \to \mathcal{X}_v$ as follows. Let $x \in \mathcal{X}_u$ for some vertex u of T. We define $\mathfrak{g}_v(x)$ inductively on $\mathsf{d}_T(u,v)$. If $u = v$, then set $\mathfrak{g}_v(x) = x$. Otherwise, $u = e^-$ for some edge e of T so that $d_T(e^+,v) = d_T(u,v)-1$. Then set $\mathfrak{g}_v(x) = \mathfrak{g}_v(\phi_{e^+}(\phi_{e^-}^{-1}(\mathfrak{g}_{\phi_{e^-}(\mathcal{X}_e)}(x))))$. We also have a map $\beta_{V_v} : \mathcal{X} \to \mathcal{C}V_v$, defined by $\beta_{V_v}(x) = \pi_{V_v}(\mathfrak{g}_v(x))$. (Here, $\mathfrak{g}_{\phi_{e^-}(\mathcal{X}_e)} : \mathcal{X}_{e^-} = \mathcal{X}_u \to \phi_{e^-}(\mathcal{X}_e)$ is the usual gate map to a hierarchically quasiconvex subspace, described in Definition 5.4, and $\phi_{e^\pm}^{-1}$ is a quasi-inverse for the edge-hieromorphism.)

Lemma 8.16. *There exists K, depending only on E and ξ, such that the following holds. Let e, f be edges of T and v a vertex so that $e^- = f^- = v$. Suppose for some $V \in \mathfrak{S}_v$ that there exist $x, y \in \phi_{e^-}(\mathcal{X}_e) \subseteq \mathcal{X}_v$ with $\mathsf{d}_V(\mathfrak{g}_{\phi_{f^-}(\mathcal{X}_f)}(x), \mathfrak{g}_{\phi_{f^-}(\mathcal{X}_f)}(y)) > 10K$. Then $V \in \phi_{e^-}^\diamondsuit(\mathfrak{S}_e) \cap \phi_{f^-}^\diamondsuit(\mathfrak{S}_f)$.*

Proof. Let $\mathcal{Y}_e = \phi_{e^-}(\mathcal{X}_e)$ and let $\mathcal{Y}_f = \phi_{f^-}(\mathcal{X}_f)$; these spaces are uniformly hierarchically quasiconvex in $\mathcal{X}_v$. Moreover, by fullness of the edge-hieromorphisms, we can choose $K \geq 100E$ so that the map $\pi_V : \mathcal{Y}_e \to \mathcal{C}V$ is K-coarsely surjective for each $V \in \phi_{e^-}^\diamondsuit(\mathfrak{S}_e)$, and likewise for $\phi_{f^-}^\diamondsuit(\mathfrak{S}_f)$ and $\mathcal{Y}_v$. If $V \in \mathfrak{S}_v - \phi_{f^-}^\diamondsuit(\mathfrak{S}_f)$, then π_V is K-coarsely constant on $\mathcal{Y}_f$, by the distance formula, since $\mathcal{X}_f$ is quasi-isometrically embedded. Likewise, π_V is coarsely constant on $\mathcal{Y}_e$ if $V \notin \phi_{e^-}^\diamondsuit(\mathfrak{S}_e)$. (This also follows from consistency when V is transverse to some unbounded element of $\phi_{e^-}^\diamondsuit(\mathfrak{S}_e)$ and from consistency and bounded geodesic image otherwise.)

Suppose that there exists $V \in \mathfrak{S}_v$ such that $\mathsf{d}_V(\mathfrak{g}_{\phi_{f^-}(\mathcal{X}_f)}(x), \mathfrak{g}_{\phi_{f^-}(\mathcal{X}_f)}(y)) > 10K$. Since $\mathfrak{g}_{\phi_{f^-}(\mathcal{X}_f)}(x), \mathfrak{g}_{\phi_{f^-}(\mathcal{X}_f)}(y) \in \mathcal{X}_f$, we therefore have that $V \in \phi_{f^-}^\diamondsuit(\mathfrak{S}_f)$. On the other hand, the definition of gates implies that $\mathsf{d}_V(x,y) > 8K$, so $V \in \phi_{e^-}^\diamondsuit(\mathfrak{S}_e)$. □

Lemma 8.17. *There exists a constant K' such that the following holds. Let e, f be edges of T and suppose that there do not exist $V_e \in \mathfrak{S}_e$, $V_f \in \mathfrak{S}_f$ for which $\phi_{e^-}^\diamondsuit(V_e) \sim \phi_{f^-}^\diamondsuit(V_f)$. Then $\mathfrak{g}_{e^-}(\mathcal{X}_f)$ has diameter at most K'. In particular, the conclusion holds if $\mathsf{d}_T(e,f) > n$, where n bounds the diameter of the supports.*

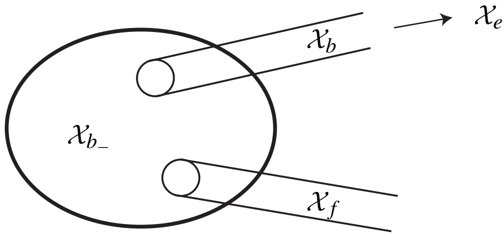

Figure 4. Schematic of the subset of $\mathcal{X}$ near $\mathcal{X}_{b^-}$.

Proof. The second assertion follows immediately from the first in light of how n was chosen.

We now prove the first assertion by induction on the number k of vertices on the geodesic in T from e to f. The base case, $k = 1$, follows from Lemma 8.16.

For $k \geq 1$, let $v_0, v_1, \ldots, v_k$ be the vertices on a geodesic from e to f, in the obvious order. Let b be the edge joining v_{k-1} to v_k, with $b^- = v_k$.

It follows from the definition of gates that $\mathfrak{g}_{e^-}(\mathcal{X}_f)$ has diameter (coarsely) bounded above by that of $\mathfrak{g}_{\phi_{b^-}(\mathcal{X}_b)}(\mathcal{X}_f)$ and that of $\mathfrak{g}_{e^-}(\mathcal{X}_b)$. Hence suppose that $\mathsf{diam}(\mathfrak{g}_{\phi_{b^-}(\mathcal{X}_b)}(\mathcal{X}_f)) > 10K$ and $\mathsf{diam}(\mathfrak{g}_{e^-}(\mathcal{X}_b)) > 10K$. Then, by induction and Lemma 8.16, we see that there exists $V_e \in \mathfrak{S}_e$, $V_f \in \mathfrak{S}_f$ for which $\phi^{\diamond}_{e^-}(V_e) \sim \phi^{\diamond}_{f^-}(V_f)$, a contradiction. $\square$

Lemma 8.18. *The map $\mathfrak{g}_v : \mathcal{X} \to \mathcal{X}_v$ is coarsely Lipschitz, with constants independent of v.*

Proof. Let $x, y \in \mathcal{X}$. If the projections of x, y to T lie in the ball of radius $2n+1$ about v, then this follows since $\mathfrak{g}_v$ is the composition of a bounded number of maps, each of which is uniformly coarsely Lipschitz by Lemma 5.5. Otherwise, by Remark 8.4, it suffices to consider x, y with $\mathsf{d}_{\mathcal{X}}(x, y) \leq C$, where C depends only on the metric d. In this case, let v_x, v_y be the vertices in T to which x, y project. Let v' be the median in T of v, v_x, v_y. Observe that there is a uniform bound on $\mathsf{d}_T(v_x, v')$ and $\mathsf{d}_T(v_y, v')$, so it suffices to bound $\mathsf{d}_v(\mathfrak{g}_v(\mathfrak{g}_{v'}(x)), \mathfrak{g}_v(\mathfrak{g}_{v'}(y)))$. Either $\mathsf{d}_T(v, v') \leq 2n+1$, and we are done, or Lemma 8.17 gives the desired bound, since equivalence classes have support of diameter at most n. $\square$

Definition 8.19 (projections). For each $[W] \in \mathfrak{S}$, define the *projection* $\pi_{[W]} : \mathcal{X} \to \mathcal{C}[W]$ by $\pi_{[W]}(x) = \beta_{W_v}(x)$, where W_v is the favorite representative of $[W]$. Note that these projections take points to uniformly bounded sets, since the collection of vertex spaces is uniformly hierarchically hyperbolic. Define $\pi_T : \mathcal{X} \to T$ to be the usual projection to T. Finally, for each $\mathrm{K}^{\perp} \in \bigcup_\eta \mathcal{K}_\eta$, just let $\pi_{\mathrm{K}}^{\perp} : \mathcal{X} \to \mathcal{C}\mathrm{K}^{\perp}$ be a constant map.

Lemma 8.20 (comparison maps). *There exists a uniform constant $\xi \geq 1$ such that* *for all $W_v \in \mathfrak{S}_v$, $W_w \in \mathfrak{S}_w$ with $W_v \sim W_w$, there exists a (ξ, ξ)-quasi-isometry $\mathfrak{c} : W_v \to W_w$ such that $\mathfrak{c} \circ \beta_v = \beta_w$ up to uniformly bounded error.*

Definition 8.21. A map $\mathfrak{c}$ as given by Lemma 8.20 is called a *comparison map*.

Proof of Lemma 8.20. We first clarify the situation by stating some consequences of the definitions. Let e^+, e^- be vertices of T joined by an edge e. Suppose that there exists $W^+ \in \mathfrak{S}_{e^+}$, $W^- \in \mathfrak{S}_{e^-}$ such that $W^+ \sim W^-$, so that there exists $W \in \mathfrak{S}_e$ with $\pi(\phi_{e^\pm})(W) = W^\pm$. Then the following diagram coarsely commutes (with uniform constants):

$$\begin{tikzcd}
 & \mathcal{X} \arrow[dl, "\mathfrak{g}_{e^-}"'] \arrow[dr, "\mathfrak{g}_{e^+}"] & \\
\mathcal{X}_{e^-} \arrow[d, "\pi_{W^-}"'] \arrow[dr, leftrightarrow] & & \mathcal{X}_{e^+} \arrow[d, "\pi_{W^-}"] \arrow[dl, leftrightarrow] \\
\mathcal{C}W^- \arrow[dr, leftrightarrow] & \mathcal{X}_e \arrow[d, "\pi_W"] & \mathcal{C}W^+ \arrow[dl, leftrightarrow] \\
 & \mathcal{C}W &
\end{tikzcd}$$

where $\mathcal{X}_e \to \mathcal{X}_{e^\pm}$ is the uniform quasi-isometry $\phi_{e^\pm}$, while $\mathcal{X}_{e^\pm} \to \mathcal{X}_e$ is the composition of a quasi-inverse for $\phi_{e^\pm}$ with the gate map $\mathcal{X}_{e^\pm} \to \phi_{e^\pm}(\mathcal{X}_e)$, and the maps $\mathcal{C}W \leftrightarrow \mathcal{C}W^\pm$ are the quasi-isometries implicit in the edge hieromorphism or their quasi-inverses. The proof essentially amounts to chaining together a sequence of these diagrams as e varies among the edges of a geodesic from v to w; an important ingredient is played by the fact that such a geodesic has length at most n.

Let $v = v_0, v_1, \dots, v_m, v_{m+1} = w$ be the geodesic sequence in T from v to w and let e^i be the edge joining v_i to v_{i+1}. For each i, choose $W_i \in \mathfrak{S}_{e^i}$ and $W_i^\pm \in \mathfrak{S}_{e^i_\pm}$ such that (say) $W_0^- = W_v$ and $W_m^+ = W_w$ and such that $\phi^\diamond_{e^i_\pm}(W_i) = W_i^\pm$ for all i. For each i, let $q_i^\pm : \mathcal{C}W_i \to \mathcal{C}W_i^\pm$ be $q_i = \phi^*_{e^i_\pm}(W_i)$, which is the (ξ', ξ')-quasi-isometry packaged in the edge-hieromorphism, and let $\bar{q}_i^\pm$ be a quasi-inverse; the constant ξ' is uniform by hypothesis, and $m \le n$ since $\mathcal{T}$ has bounded supports. The hypotheses on the edge-hieromorphisms ensure that the $W_i^\pm$ are uniquely determined by W_v, W_w, and we define $\mathfrak{c}$ by

$$\mathfrak{c} = q_m^{\epsilon_m} \bar{q}_m^{\epsilon'_m} \cdots q_1^{\epsilon_1} \bar{q}_1^{\epsilon'_1},$$

where $\epsilon_i, \epsilon'_i \in \{\pm\}$ depend on the orientation of e^i, and $\epsilon'_i = +$ if and only if $\epsilon_i = -$. This is a (ξ, ξ)-quasi-isometry, where $\xi = \xi(\xi'_n)$.

If $v = w$, then $\mathfrak{c}$ is the identity and $\mathfrak{c} \circ \beta_v = \beta_v$. Let $d \ge 1 = \mathsf{d}_T(v, w)$ and let w' be the penultimate vertex on the geodesic of T from v to w. Let $\mathfrak{c}' : \mathcal{C}W_v \to \mathcal{C}W_{w'}$ be a comparison map, so that, by induction, there exists $\lambda' \ge 0$ so that $\mathsf{d}_{\mathcal{C}W_{w'}}(\mathfrak{c}' \circ \beta_v(x), \beta_{w'}(x)) \le \lambda'$ for all $x \in \mathcal{X}$. Let $\mathfrak{c}'' = \bar{q}_k^+ q_k^- : \mathcal{C}W_{w'} \to \mathcal{C}W_w$

be the (ξ', ξ')-quasi-isometry packaged in the edge-hieromorphism, so that the following diagram coarsely commutes:

$$\begin{tikzcd} & \mathcal{X} \arrow[dl, "\mathfrak{g}_v"'] \arrow[d, "\mathfrak{g}_{w'}"] \arrow[dr, "\mathfrak{g}_w"] & \\ \mathcal{X}_v \arrow[r, "\mathfrak{g}_{w'}"'] \arrow[d, "\pi_{W_v}"'] & \mathcal{X}_{w'} \arrow[r, "\mathfrak{g}_w"'] \arrow[d, "\pi_{W_{w'}}"'] & \mathcal{X}_w \arrow[d, "\pi_{W_w}"'] \\ \mathcal{C}W_v \arrow[r, "\mathfrak{c}'"'] & \mathcal{C}W_{w'} \arrow[r, "\mathfrak{c}''"'] & \mathcal{C}W_w \end{tikzcd}$$

Since $\mathfrak{c} = \mathfrak{c}'' \circ \mathfrak{c}'$ and the constants implicit in the coarse commutativity of the diagram depend only on the constants of the hieromorphism and on $d \leq n$, the claim follows. $\square$

Lemma 8.22. *There exists K such that each $\pi_{[W]}$ is (K, K)-coarsely Lipschitz.*

Proof. For each vertex v of T and each $V \in \mathfrak{S}_v$, the projection $\pi_V : \mathcal{X}_v \to \mathcal{C}V$ is uniformly coarsely Lipschitz, by definition. By Lemma 8.18, each gate map $\mathfrak{g}_v : \mathcal{X} \to \mathcal{X}_v$ is uniformly coarsely Lipschitz. The lemma follows since $\pi_{[W]} = \pi_{W_v} \circ \mathfrak{g}_v$, where v is the favorite vertex carrying the favorite representative W_v of $[W]$. $\square$

Definition 8.23 (projections between hyperbolic spaces). If $[V] \sqsubseteq [W]$, then choose vertices $v, v', w \in \mathcal{V}$ so that V_v, W_w are respectively the favorite representatives of $[V], [W]$, while $V_{v'}, W_{v'}$ are respectively representatives of $[V], [W]$ with $V_{v'}, W_{v'} \in \mathfrak{S}_{v'}$ and $V_{v'} \sqsubseteq W_{v'}$. Let $\mathfrak{c}_V : \mathcal{C}V_{v'} \to \mathcal{C}V_v$ and $\mathfrak{c}_W : \mathcal{C}W_{v'} \to \mathcal{C}W_w$ be comparison maps. Then define

$$\rho^{[V]}_{[W]} = c_W(\rho^{V_{v'}}_{W_{v'}}),$$

which is a uniformly bounded set, and define $\rho^{[W]}_{[V]} : \mathcal{C}[W] \to \mathcal{C}[V]$ by

$$\rho^{[W]}_{[V]} = \mathfrak{c}_V \circ \rho^{W_{v'}}_{V_{v'}} \circ \bar{\mathfrak{c}}_W,$$

where $\bar{\mathfrak{c}}_W$ is a quasi-inverse of $\mathfrak{c}_W$ and $\rho^{W_{v'}}_{V_{v'}} : \mathcal{C}W_{v'} \to \mathcal{C}V_{v'}$ is the map provided by Definition 1.1.(2). Similarly, if $[V] \pitchfork [W]$, and there exists $w \in T$ so that $\mathfrak{S}_w$ contains representatives V_w, W_w of $[V], [W]$, then let

$$\rho^{[V]}_{[W]} = \mathfrak{c}_W(\rho^{V_w}_{W_w}).$$

Otherwise, choose a closest pair v, w so that $\mathfrak{S}_v$ (respectively, $\mathfrak{S}_w$) contains a representative of $[V]$ (respectively, $[W]$). Let e be the first edge of the geodesic in T joining v to w, so $v = e^-$ (say). Let S be the $\sqsubseteq$-maximal element of $\mathfrak{S}_e$, and let

$$\rho^{[W]}_{[V]} = \mathfrak{c}_V\Big(\rho^{\phi^\diamond_{e^-}(S)}_{V_v}\Big).$$

This is well-defined by hypothesis (4).

For each $\sim$-class $[W]$, let $\rho_T^{[W]}$ be the support of $[W]$ (a uniformly bounded set since $\mathcal{T}$ has bounded supports). Define $\rho_{[W]}^T : T \to \mathcal{C}W$ as follows: given $v \in T$ not in the support of $[W]$, let e be the unique edge with e^- (say) separating v from the support of $[W]$. Let $S \in \mathfrak{S}_e$ be $\sqsubseteq$-maximal. Then

$$\rho_{[W]}^T(v) = \rho_{[W]}^{[\phi_{e^-}^{\diamond}(S)]}.$$

If v is in the support of $[W]$, then let $\rho_{[W]}^T(v)$ be chosen arbitrarily.

Finally, let $\mathrm{K}^\perp, \mathrm{K}^{\perp\prime} \in \bigcup_\eta \mathcal{K}_\eta$ and let $[W]$ be a $\sim$-class. If $\mathrm{K}^\perp \pitchfork \mathrm{K}^{\perp\prime}$, then $\rho_{\mathrm{K}^\perp}^{\mathrm{K}^{\perp\prime}}$ is the single point $\mathcal{C}\mathrm{K}^\perp$. If $\mathrm{K}^\perp \sqsubseteq \mathrm{K}^{\perp\prime}$, then $\rho_{\mathrm{K}^\perp}^{\mathrm{K}^{\perp\prime}}$ is a constant map and $\rho_{\mathrm{K}^{\perp\prime}}^{\mathrm{K}^\perp}$ is the obvious single point. We never have $\mathrm{K}^\perp \sqsubseteq [W]$. If $[W] \sqsubseteq \mathrm{K}^\perp$, then, again, $\rho_{\mathrm{K}^\perp}^{[W]}$ is the obvious single point, and we can define $\rho_{[W]}^{\mathrm{K}^\perp} : \mathcal{C}\mathrm{K}^\perp \to \mathcal{C}[W]$ to be an arbitrary constant map. Finally, $\rho_{\mathrm{K}^\perp}^T : T \to \mathcal{C}\mathrm{K}^\perp$ is the constant map, and $\rho_T^{\mathrm{K}^\perp}$ is the bounded set defined as follows. By definition, there is a unique pairwise orthogonal set $[W_1], \dots, [W_\eta]$ so that $\mathrm{K}^\perp = \mathrm{K}_\eta^\perp([W_1], \dots, [W_\eta])$. By the proof of Lemma 8.13, the supports of the various $[W_i]$ all intersect in a subtree of T, which necessarily has diameter at most n, by the bounded supports hypothesis; we take this intersection to be $\rho_T^{\mathrm{K}^\perp}$.

We are now ready to complete the proof of the combination theorem.

Proof of Theorem 8.6. We claim that $(\mathcal{X}(\mathcal{T}), \mathfrak{S})$ is hierarchically hyperbolic. We take the nesting, orthogonality, and transversality relations for a tree of spaces given by Definitions 8.7 and 8.11. In Lemmas 8.9, 8.10, and Remark 8.12, it is shown that these relations satisfy all of the conditions (2) and (3) of Definition 1.1 not involving the projections. Moreover, the complexity of $(\mathcal{X}(\mathcal{T}), \mathcal{S})$ is finite, by Lemma 8.13, verifying Definition 1.1.(5). The set of δ-hyperbolic spaces $\{\mathcal{C}A : A \in \mathfrak{S}\}$ is provided by Definition 8.14, while the projections $\pi_{[W]} : \mathcal{X} \to \mathcal{C}[W]$ required by Definition 1.1.(1) are defined in Definition 8.19 and are uniformly coarsely Lipschitz by Lemma 8.22. Since $\pi_{[W]}(\mathcal{X})$ uniformly coarsely coincides with the image of an appropriately chosen vertex space $\mathcal{X}_v$, it is quasiconvex since $\pi_{W_v}(\mathcal{X}_v)$ is uniformly quasiconvex by Definition 1.1.(1). The projections $\rho_{[W]}^{[V]}$ when $[V], [W]$ are nonorthogonal are described in Definition 8.23. To complete the proof, it thus suffices to verify the consistency inequalities (Definition 1.1.(4)), the bounded geodesic image axiom (Definition 1.1.(7)), the large link axiom (Definition 1.1.(6)), partial realization (Definition 1.1.(8)), and uniqueness (Definition 1.1.(9)).

Consistency. Any consistency inequalities involving elements of $\bigcup_\eta \mathcal{K}_\eta$ hold trivially since in that case, at least one of the two associated hyperbolic spaces in question is a point. Suppose that $[U] \pitchfork [V]$ or $[U] \sqsubseteq [V]$ and let $x \in \mathcal{X}$. Choose representatives $U_u \in \mathfrak{S}_u$, $V_u \in \mathfrak{S}_v$ of $[U], [V]$ so that $\mathrm{d}_T(u, v)$ realizes the distance between the supports of $[U], [V]$. By composing the relevant maps in the remainder of the argument with comparison maps, we can assume that U_u, V_v are favorite representatives. Without loss of generality, there exists a vertex $w \in T$ so that $x \in \mathcal{X}_w$. If $u = v$, then consistency follows since it holds in each vertex space, so assume that u, v have disjoint supports and in particular $[U] \pitchfork [V]$.

If $w \notin [u, v]$, then (say) u separates w from v. Then $\pi_{[V]}(x) = \pi_{V_v}(\mathfrak{g}_v(x))$. Let e be the edge of the geodesic $[u, v]$ emanating from u, so that $\rho^{[V]}_{[U]} = \rho^{S}_{U_u}$, where S is the image in $\mathfrak{S}_u$ of the $\sqsubseteq$-maximal element of $\mathfrak{S}_e$. If

$$\mathsf{d}_{\mathcal{C}U_u}(\mathfrak{g}_u(x), \rho^{S}_{U_u}) \leq E,$$

then we are done. Otherwise, by consistency in $\mathfrak{S}_u$, we have

$$\mathsf{d}_{\mathcal{C}S}(\mathfrak{g}_u(x), \rho^{U_u}_{S}) \leq E,$$

from which consistency follows. Hence suppose that $w \in [u, v]$. Then without loss of generality, there is an edge e containing v and separating w from v. As before, projections to V factor through the $\sqsubseteq$-maximal element of $\mathfrak{S}_e$, from which consistency follows.

We verify consistency for $T, [W]$ for each $\sim$-class $[W]$. Choose $x \in \mathcal{X}_v$. If $\mathsf{d}_T(v, T_{[W]}) \geq n + 1$, then let e be the edge incident to $T_{[W]}$ separating it from v, so that (up to a comparison map) $\rho^{T}_{[W]}(v) = \rho^{S}_{W}$, where $W \in \mathfrak{S}_{e^+}$ represents W, and $e^+ \in T_{[W]}$, and S is the image in $\mathfrak{S}_{e^+}$ of the $\sqsubseteq$-maximal element of $\mathfrak{S}_e$. (Note that $W \pitchfork S$ by hypothesis (4) of the theorem and the choice of e). On the other hand (up to a comparison map) $\pi_{[W]}(x) = \pi_W(\mathfrak{g}_{e^+}(x)) \asymp \pi_W(\mathfrak{g}_{e^+}(\mathcal{X}_v)) \asymp \pi_W(\boldsymbol{F}_S) = \rho^{S}_{W}$, as desired. (The final coarse equality holds by hypothesis (4).)

Finally, suppose that $[U] \sqsubseteq [V]$ and that either $[V] \sqsubsetneq [W]$ or $[V] \pitchfork [W]$ and $[U] \not\perp [W]$. We claim that $\mathsf{d}_{[W]}(\rho^{[U]}_{[W]}, \rho^{[V]}_{[W]})$ is uniformly bounded. By definition, $T_{[U]} \cap T_{[V]} \neq \emptyset$ and we fix representatives $U_u \in \mathfrak{S}_u$, $V_u \in \mathfrak{S}_u$ of $[U], [V]$ with $U_u \sqsubseteq V_u$.

Next, suppose that $T_{[V]} \cap T_{[W]} \neq \emptyset$ and $T_{[U]} \cap T_{[W]} \neq \emptyset$, so that we can choose vertices v, w of T and representatives $V_w, W_w \in \mathfrak{S}_w$ so that $V_w \sqsubseteq W_w$ or $V_w \pitchfork W_w$ according to whether $[V] \sqsubseteq [W]$ or $[V] \pitchfork [W]$, and choose representatives $U_v, W_v \in \mathfrak{S}_v$ of $[U], [W]$ so that $U_v \sqsubseteq W_v$ or $U_v \pitchfork W_v$ according to whether $[U] \sqsubseteq [W]$ or $[U] \pitchfork [W]$. Let $m \in T$ be the median of u, v, w. Since u, w lie in the support of $[U], [W]$, so does m, since supports are connected. Likewise, m lies in the support of $[V]$. Let U_m, V_m, W_m be the representatives of $[U], [V], [W]$ in m. Since edge-maps are full hieromorphisms, we have $U_m \sqsubseteq V_m$ and $U_m \not\perp W_m$ and either $V_m \sqsubseteq W_m$ or $V_m \pitchfork W_m$. Hence Definition 1.1.(4) implies that $\mathsf{d}_{W_m}(\rho^{U_m}_{W_m}, \rho^{V_m}_{W_m})$ is uniformly bounded. Since the comparison maps are uniform quasi-isometries, it follows that $\mathsf{d}_{[W]}(\rho^{[U]}_{[W]}, \rho^{[V]}_{[W]})$ is uniformly bounded, as desired.

Next, suppose that $T_{[U]} \cap T_{[W]} = \emptyset$. Then $[U] \pitchfork [W]$. If there is an edge e separating $T_{[W]}$ from $T_{[U]} \cup T_{[V]}$, then $\rho^{[U]}_{[W]} = \rho^{[V]}_{[W]}$ by definition. Otherwise, $[V] \pitchfork [W]$ (by transitivity of $\sqsubseteq$ and the fact that $T_{[U]} \cap T_{[W]} = \emptyset$) but there exist some

$v \in T_{[V]} \cap T_{[W]}$ and representatives $V_v, W_v \in \mathfrak{S}_v$ of $[V], [W]$ with $V_v \pitchfork W_v$. But by fullness of the hieromorphism and induction on $\mathsf{d}_T(u, v)$, we find that $v \in T_{[U]}$, contradicting that $T_{[U]} \cap T_{[W]} = \varnothing$.

Bounded geodesic image and large link axiom in T and $\bigcup_\eta \mathcal{K}_\eta$. For bounded geodesic image, in the case where one of the two nested elements of $\mathfrak{S}$ in question is in $\bigcup_\eta \mathcal{K}_\eta$, the claim holds trivially since the associated hyperbolic space has diameter 0 and the associated map between hyperbolic spaces has either domain or codomain a single point.

Let γ be a geodesic in T and let $[W]$ be a $\sim$-class so that $\mathsf{d}_T(\gamma, \rho_T^{[W]}) > 1$, which is to say that γ does not contain vertices in the support of $[W]$. Let e be the terminal edge of the geodesic joining γ to the support of $[W]$. Then for all $u \in \gamma$, we have by definition that $\rho_{[W]}^T(u) = \rho_{[W]}^{[S]}$ for some fixed $\sim$-class $[S]$. This verifies the bounded geodesic image axiom for $T, [W]$.

By Lemma 8.17, there exists a constant K'' such that if $x, x' \in \mathcal{X}$ respectively project to vertices v, v', then any $[W] \in \mathfrak{S} - \{T\}$ with $\mathsf{d}_{[W]}(x, x') \geq K''$ is supported on a vertex $v_{[W]}$ on the geodesic $[v, v']$ and is hence nested into $[S_{v_W}]$, where S_{v_W} is maximal in $\mathfrak{S}_{v_{[W]}}$. Indeed, choose w in the support of $[W]$. Then either $\mathsf{d}_{[W]}(x, x')$ is smaller than some specified constant, or $\mathsf{d}_{\mathcal{X}_w}(\mathfrak{g}_w(x), \mathfrak{g}_w(x') > K'$. Thus $\mathfrak{g}_{\mathcal{X}_w}(\mathcal{X}_m)$ has diameter at least K', where m is the median of v, v', w. Hence m lies in the support of $[W]$, and $m \in [v, v']$, and $[W] \sqsubseteq [S]$, where S is $\sqsubseteq$-maximal in $\mathfrak{S}_m$. Finally, for each such S_{v_W}, it is clear that $\mathsf{d}_T(x, \rho_T^{[S_{v_W}]}) \leq \mathsf{d}_T(x, x')$, verifying the conclusion of the large link axiom for T.

Finally, we have to check that if $\mathrm{K}^\perp = \mathrm{K}^\perp([W_1], \ldots, [W_\eta])$ and $x, y \in \mathcal{X}$, then there exist a uniformly bounded number of elements $[U_j]$ so that for any $[V] \sqsubseteq \mathrm{K}^\perp$ with $\mathsf{d}_{[V]}(x, y) \geq E$, the class $[V]$ is nested into some U_j. We have shown above that any such $[V]$ is nested into the $\sqsubseteq$-maximal $S_v \in \mathfrak{S}_v$ for some v on the geodesic of T between $\pi_T(x)$ and $\pi_T(y)$. Now, since $[V] \sqsubseteq \mathrm{K}^\perp$, we have $[V] \perp [W_i]$ for all i, so that the support of $[V]$ uniformly coarsely coincides with the support of $[W_i]$ for each i. Hence v must be among the boundedly many vertices on $[\pi_T(x), \pi_T(y)]$ that lie in the intersection of the supports of the $[W_i]$. Thus we can take our set of U_j to be the set of such S_v, which has uniformly bounded cardinality (bounded in terms of n).

Bounded geodesic image and large link axiom in $W \sqsubset T$. Let $[W]$ be non-$\sqsubseteq$-maximal, let $[V] \sqsubseteq [W]$, and let γ be a geodesic in $\mathcal{C}[W]$. Then γ is a geodesic in $\mathcal{C}W_w$, by definition, where w is the favorite vertex of $[W]$ with corresponding representative W_w. Let V_w be the representative of $[V]$ supported on w, so that $\rho_{[W]}^{[V]} = \rho_{W_w}^{V_w}$, so that γ avoids the E-neighborhood of $\rho_{[W]}^{[V]}$ exactly when it avoids the E-neighborhood of $\rho_{W_w}^{V_w}$. The bounded geodesic image axiom now follows from bounded geodesic image in $\mathfrak{S}_w$, although the constant E has been changed according to the quasi-isometry constant of comparison maps.

Now suppose $x, x' \in \mathcal{X}_v, \mathcal{X}_{v'}$ and choose w to be the favorite vertex in the support of $[W]$. Suppose for some $[V] \sqsubseteq [W]$ that $\mathsf{d}_{[V]}(x, x') \geq E'$, where E' depends on E and the quasi-isometry constants of the edge-hieromorphisms. Then $\mathsf{d}_{V_w}(\mathfrak{g}_w(x), \mathfrak{g}_w(x')) \geq E$, for some representative $V_w \in \mathfrak{S}_w$ of $[V]$, by our choice of E'. Hence, by the large link axiom in $\mathfrak{S}_w$, we have that $V_w \sqsubseteq T_i$, where $\{T_i\}$ is a specified set of $N = \lfloor \mathsf{d}_{W_w}(\mathfrak{g}_w(x), \mathfrak{g}_w(x')) \rfloor = \lfloor \mathsf{d}_{[W]}(x, x') \rfloor$ elements of $\mathfrak{S}_w$, with each $T_i \sqsubset W_w$. Moreover, the large link axiom in $\mathfrak{S}_w$ implies that

$$\mathsf{d}_{[W]}(x, \rho_{[W]}^{[T_i]}) = \mathsf{d}_{W_w}(\mathfrak{g}_w(x), \rho_{W_w}^{T_i}) \leq N$$

for all i. This verifies the large link axiom for $(\mathcal{X}(\mathcal{T}), \mathfrak{S})$.

Partial realization. Let $[V_1], \dots, [V_k] \in \mathfrak{S}$ be pairwise-orthogonal, and, for each $i \leq k$, let $p_i \in \mathcal{C}[V_i]$. For each i, let $T_i \subseteq T$ be the induced subgraph spanned by the vertices w such that $[V_i]$ has a representative in $\mathfrak{S}_w$. The definition of the $\sim$-relation implies that each T_i is connected, so by the Helly property of trees, there exists a vertex $v \in T$ such that for each i, there exists $V_v^i \in \mathfrak{S}_v$ representing $[V_i]$. Moreover, we have $V_v^i \perp V_v^j$ for $i \neq j$, since the edge-hieromorphisms preserve the orthogonality relation. Applying the partial realization axiom (Definition 1.1.(8)) to $\{p_i' \in \mathcal{C}V_v^i\}$, where p_i' is the image of p_i under the appropriate comparison map, yields a point $x \in \mathcal{X}_v$ such that $\pi_{V_v^i}(x)$ is coarsely equal to p_i' for all i, whence $\mathsf{d}_{[V_i]}(x, p_i)$ is uniformly bounded. If $[V_i] \sqsubseteq [W]$, then W has a representative $W_v \in \mathfrak{S}_v$ such that $V_v^i \sqsubseteq W$, whence $\mathsf{d}_{[W]}(x, \rho_W^{[V_i]})$ is uniformly bounded since x is a partial realization point for $\{V_v^i\}$ in $\mathfrak{S}_v$. Finally, if $[W] \pitchfork [V_i]$, then either the subtrees of T supporting $[W]$ and $[V_i]$ are disjoint, in which case $\mathsf{d}_{[W]}(x, \rho_{[W]}^{[V_i]})$ is bounded, or $[W]$ has a representative in $\mathfrak{S}_v$ transverse to V_v^i, in which case the same inequality holds by our choice of x. There is nothing to check regarding projections onto $\mathcal{C}\mathrm{K}^\perp$ for $\mathrm{K}^\perp \in \bigcup_\eta \mathcal{K}_\eta$, since those spaces are single points.

It remains to consider pairwise orthogonal collections that include elements of $\bigcup_\eta \mathcal{K}_\eta$. Since no two of these elements can be orthogonal, we must consider $\mathrm{K}^\perp = \mathrm{K}_\eta^\perp([W_1], \dots, [W_\eta])$, which is orthogonal to $\sim$-classes $[V_1], \dots, [V_k]$, which themselves form an orthogonal collection. Let p be the unique point of $\mathcal{C}\mathrm{K}^\perp$, and for each $i \leq k$, let $p_i \in \mathcal{C}V_i$. Then the previous discussion provides a point x so that for any i, we have $\pi_{[V_i]})(x) \asymp p_i$, and $\pi_T(x) \asymp \rho_T^{[V_i]}$. Moreover, for any $[U]$ so that, for some i, we have $[U] \pitchfork [V_i]$ or $[V_i] \sqsubseteq [U]$, we have $\pi_{[U]}(x) \asymp \rho_{[U]}^{[V_i]}$. We claim that x also satisfies the conclusion of partial realization for the pairwise-orthogonal set $\mathrm{K}^\perp, [V_1], \dots, [V_k]$. Again, there is nothing to check regarding projections onto $\mathcal{C}\mathrm{K}^\perp$ for $\mathrm{K}^\perp \in \bigcup_\eta \mathcal{K}_\eta$, since those spaces are single points, and this includes the statement about $\pi_{\mathrm{K}^\perp}(x)$. So, it just remains to check that $\pi_T(x)$ uniformly coarsely coincides with $\rho_T^{\mathrm{K}^\perp}$. But $\pi_T(x)$ coarsely coincides with $\rho_T^{[V_i]}$ for any i, by the construction of x. Since $[V_i] \perp \mathrm{K}^\perp$, we have $[V_i] \sqsubseteq [W_j]$ for some j, so $\rho_T^{[V_i]}$ coarsely coincides with $\rho_T^{[W_j]}$. But $\rho_T^{[W_j]}$ coarsely coincides, by definition, with $\rho_T^{\mathrm{K}^\perp}$, as required.

Uniqueness of realization points. Suppose $x, y \in \mathcal{X}$ satisfy $\mathsf{d}_{[V]}(x, y) \le K$ for all $[V] \in \mathfrak{S}$. Then, for each vertex $v \in T$, applying the uniqueness axiom in $\mathcal{X}_v$ to $\mathfrak{g}_v(x), \mathfrak{g}_v(y)$ shows that $\mathsf{d}_{\mathcal{X}_v}(\mathfrak{g}_v(x), \mathfrak{g}_v(y)) \le \zeta = \zeta(K)$. Indeed, otherwise we would have $\mathsf{d}_V(\mathfrak{g}_v(x), \mathfrak{g}_v(y)) > \xi K + \xi$ for some $V \in \mathfrak{S}_v$, whence $\mathsf{d}_{[V]}(x, y) > K$. There exists $k \le K$ and a sequence $v_0, \dots, v_k$ of vertices in T so that $x \in \mathcal{X}_{v_0}$, $y \in \mathcal{X}_{v_k}$. For each j, let $x_j = \mathfrak{g}_{v_j}(x)$ and $y_j = \mathfrak{g}_{v_j}(y)$. Then $x = x_0, y_0, x_1, y_1, \dots, y_{j-1}, x_j, y_j, \dots, x_k, y_k = y$ is a path of uniformly bounded length joining x to y. Indeed, $\mathsf{d}_{\mathcal{X}_{v_j}}(x_j, y_j) \le \zeta$ and $k \le K$ by the preceding discussion, while x_j coarsely coincides with a point on the opposite side of an edge-space from y_{j-1} by the definition of the gate of an edge-space in a vertex-space and the fact that x_{j-1} and y_{j-1} coarsely coincide. This completes the proof. □

8A. ***Equivariant definition of*** $(\mathcal{X}(\mathcal{T}), \mathfrak{S})$**.** Let $\mathcal{T}$ denote the tree of hierarchically hyperbolic spaces $(\mathcal{T}, \{\mathcal{X}_v\}, \{\mathcal{X}_e\}, \{\pi_{e^\pm}\})$, and let $(\mathcal{X}(\mathcal{T}), \mathfrak{S})$ be the hierarchically hyperbolic structure defined in the proof of Theorem 8.6. Various arbitrary choices were made in defining the constituent hyperbolic spaces and projections in this hierarchically hyperbolic structure, and we now insist on a specific way of making these choices in order to describe automorphisms of $(\mathcal{X}(\mathcal{T}), \mathfrak{S})$.

Recall that an automorphism of $(\mathcal{X}(\mathcal{T}), \mathfrak{S})$ is determined by a bijection $g : \mathfrak{S} \to \mathfrak{S}$ and a set of isometries $g : \mathcal{C}Q \to \mathcal{C}gQ$, for $Q \in \mathfrak{S}$. Via the distance formula, this determines a uniform quasi-isometry $\mathcal{X}(\mathcal{T}) \to \mathcal{X}(\mathcal{T})$.

A bijection $g : \bigsqcup_{v \in \mathcal{V}} \mathfrak{S}_v \to \bigsqcup_{v \in \mathcal{V}} \mathfrak{S}_v$ is *T-coherent* if there is an induced isometry g of the underlying tree, T, so that $fg = gf$, where $f : \bigsqcup_{v \in \mathcal{V}} \mathfrak{S}_v \to T$ sends each $V \in \mathfrak{S}_v$ to v, for all $v \in \mathcal{V}$. The T-coherent bijection g is said to be *$\mathcal{T}$-coherent* if it also preserves the relation $\sim$. Noting that the composition of $\mathcal{T}$-coherent bijections is $\mathcal{T}$-coherent, denote by $\mathcal{P}_{\mathcal{T}}$ the group of $\mathcal{T}$-coherent bijections. For each $g \in \mathcal{P}_{\mathcal{T}}$, there is an induced bijection $g : \mathfrak{S}^0 \to \mathfrak{S}^0$.

Recall that the hierarchically hyperbolic structure $(\mathcal{X}(\mathcal{T}), \mathfrak{S})$ was completely determined except for the following three types of choices which were made arbitrarily.

(1) For $[V] \in \mathfrak{S}$, the stabiliser of $[V]$ fixes a point in the (bounded) support tree, which we can assume, by subdividing, to be a vertex v. This is the *favorite vertex* for $[V]$.

(2) we chose an arbitrary *favorite representative* $V_v \in \mathfrak{S}_v$ with $[V] = [V_v]$. (Note that if, as is often the case in practice, edge-hieromorphisms $\mathfrak{S}_e \to \mathfrak{S}_v$ are injective, then V_v is the unique representative of its $\sim$-class that lies in $\mathfrak{S}_v$, and hence our choice is uniquely determined.)

(3) For each $[W] \in \mathfrak{S}$, the point $\rho^T_{[W]}(v)$ is chosen arbitrarily in $\mathcal{C}W$, where W is the favorite representative of $[W]$ and v is a vertex in the support of $[W]$.

We now constrain these choices so that they are equivariant. For each $\mathcal{P}_{\mathcal{T}}$-orbit in $\mathfrak{S}$, choose a representative $[V]$ of that orbit, choose a favorite vertex v in its support, and choose a favorite representative $V_v \in \mathfrak{S}_v$ of $[V]$. Then declare $gV_v \in \mathfrak{S}_{gv}$ to be the favorite representative, and gv the favorite vertex, associated to $g[V]$, for each $g \in \mathcal{P}_{\mathcal{T}}$.

Recall that, for each $[W] \in \mathfrak{S}$, we defined $\mathcal{C}[W]$ to be $\mathcal{C}W$, where W is the favorite representative of $[W]$. Suppose that we have specified a subgroup $G \leq \mathcal{P}_{\mathcal{T}}$ and, for each $[W] \in \mathfrak{S}$ and $g \in \mathcal{P}_{\mathcal{T}}$, an isometry $g : \mathcal{C}[W] \to \mathcal{C}g[W]$. Then we choose $\rho^T_{[W]}$ in such a way that $\rho^T_{g[W]} = g\rho^T_{[W]}$ for each $[W] \in \mathfrak{S}$ and $g \in G$.

8B. ***Graphs of hierarchically hyperbolic groups.*** Recall that the finitely generated group G is *hierarchically hyperbolic* if there is a hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$ such that $G \leq \mathrm{Aut}(\mathfrak{S})$ and the action of G on $\mathcal{X}$ is metrically proper and cobounded and the action of G on $\mathfrak{S}$ is cofinite (this, together with the definition of an automorphism, implies that only finitely many isometry types of hyperbolic space are involved in the HHS structure). Endowing G with a word-metric, we see that $(G, \mathfrak{S})$ is a hierarchically hyperbolic space.

If $(G, \mathfrak{S})$ and $(G', \mathfrak{S}')$ are hierarchically hyperbolic groups, then a *homomorphism of hierarchically hyperbolic groups* $\phi : (G, \mathfrak{S}) \to (G', \mathfrak{S}')$ is a homomorphism $\phi : G \to G'$ that is also a ϕ-equivariant hieromorphism as in Definition 1.22.

Recall that a *graph of groups* $\mathcal{G}$ is a graph $\Gamma = (V, E)$ together with a set $\{G_v : v \in V\}$ of vertex groups, a set $\{G_e : e \in E\}$ of edge groups, and monomorphisms $\phi_e^{\pm} : G_e \to G_{e^{\pm}}$, where $e^{\pm}$ are the vertices incident to e. As usual, the *total group* G of $\mathcal{G}$ is the quotient of $(*_{v\in V} G_v) * F_E$, where F_E is the free group generated by E, obtained by imposing the following relations:

- $e = 1$ for all $e \in E$ belonging to some fixed spanning tree T of Γ.
- $\phi_e^+(g) = e\phi_e^-(g)e^{-1}$ for $e \in E$ and $g \in G_e$.

We are interested in the case where Γ is a finite graph and, for each $v \in V$, $e \in E$, we have sets $\mathfrak{S}_v$, $\mathfrak{S}_e$ so that $(G_v, \mathfrak{S}_v)$ and $(G_e, \mathfrak{S}_e)$ are hierarchically hyperbolic group structures for which $\phi_e^{\pm} : G_e \to G_{e^{\pm}}$ is a homomorphism of hierarchically hyperbolic groups. In this case, $\mathcal{G}$ is a *finite graph of hierarchically hyperbolic groups*. If in addition each $\phi_e^{\pm}$ has hierarchically quasiconvex image, then $\mathcal{G}$ has *quasiconvex edge groups*.

Letting $\widetilde{\Gamma}$ denote the Bass–Serre tree, observe that

$$\mathcal{T} = \widetilde{\mathcal{G}} = (\widetilde{\Gamma}, \{G_{\tilde{v}}\}, \{G_{\tilde{e}}\}, \{\phi^{\pm}_{\tilde{e}}\})$$

is a tree of hierarchically hyperbolic spaces, where $\tilde{v}$ ranges over the vertex set of $\widetilde{\Gamma}$, and each $G_{\tilde{v}}$ is a conjugate in the total group G to G_v, where $\tilde{v} \mapsto v$ under $\widetilde{\Gamma} \to \Gamma$, and an analogous statement holds for edge-groups. Each $\phi^{\pm}_{\tilde{e}}$ is conjugate to an edge-map in $\mathcal{G}$ in the obvious way. We say $\mathcal{G}$ has *bounded supports* if $\mathcal{T}$ does.

Corollary 8.24 (combination theorem for HHGs). *Let $\mathcal{G} = (\Gamma, \{G_v\}, \{G_e\}, \{\phi_e^{\pm}\})$ be a finite graph of hierarchically hyperbolic groups, with Bass–Serre tree $\widetilde{\Gamma}$. Suppose that:*

(1) *$\mathcal{G}$ has quasiconvex edge groups;*

(2) *each $\phi_e^{\pm}$, as a hieromorphism, is full;*

(3) *$\mathcal{G}$ has bounded supports;*

(4) *if e is an edge of Γ and S_e the $\sqsubseteq$-maximal element of $\mathfrak{S}_e$, then for all $V \in \mathfrak{S}_{e^{\pm}}$, the elements V and $\phi_{e^{\pm}}^{\diamond}(S_e)$ are not orthogonal in $\mathfrak{S}_{e^{\pm}}$;*

(5) *for each vertex v of Γ, there are finitely many G_v-orbits of subsets $\mathcal{U} \subset \mathfrak{S}_v$ for which the elements of $\mathcal{U}$ are pairwise-orthogonal;*

(6) *there exists $K \geq 0$ such that for all vertices v of $\widetilde{\Gamma}$ and edges e incident to v, we have $\mathsf{d}_{\mathrm{Haus}}(\phi_v(G_e)), \boldsymbol{F}_{\phi_v^{\diamond}(S_e)} \times \{\star\}) \leq K$, where $S_e \in \mathfrak{S}_e$ is the unique maximal element and $\star \in \boldsymbol{E}_{\phi_v^{\diamond}(S_e)}$.*

Then the total group G of $\mathcal{G}$ is a hierarchically hyperbolic group.

Remark 8.25. We have added hypothesis (5) because it is exactly what's required. In fact, it should follow from a stronger but more natural condition, namely that for each $V \in \mathfrak{S}_v$, the stabilizer in G_v of the standard product region $\boldsymbol{P}_V$ acts cocompactly on $\boldsymbol{P}_V$. This holds, for example, in the mapping class group. On the other hand, this stronger condition it is not a consequence of the definition of an HHG since, for example, one can put exotic HHG structures on a free group where this fails.

Proof of Corollary 8.24. By Theorem 8.6, $(G, \mathfrak{S})$ is a hierarchically hyperbolic space. Observe that $G \leq \mathcal{P}_{\mathcal{G}}$, since G acts on the Bass–Serre tree $\widetilde{\Gamma}$, and this action is induced by an action on $\bigcup_{v\in\mathcal{V}} \mathfrak{S}_v$ preserving the $\sim$-relation. Hence the hierarchically hyperbolic structure $(G, \mathfrak{S})$ can be chosen according to the constraints in Section 8A, whence it is easily checked that G acts on $\mathfrak{S}^0$ by HHS automorphisms. Moreover, for any $[V]$, there are finitely many $\mathrm{Stab}_G([V])$-orbits of $\sim$-classes nested in $[V]$.

The action on $\mathfrak{S}^0$ is cofinite since each G_v is a hierarchically hyperbolic group.

Moreover, since G preserves nesting and orthogonality in $\mathfrak{S}^0$, we have an induced action of G on $\bigcup_\eta \mathcal{K}_\eta$ defined by $g\mathrm{K}^{\perp}([W_1], \dots, [W_\eta]) = \mathrm{K}^{\perp}([gW_1], \dots, [gW_n])$. We must show that this action (and hence the action of G on $\mathfrak{S}$ obtained by combining this with the action on $\mathfrak{S}^0$) is cofinite.

Since each element of $\mathcal{K}_n$ corresponds to a n-element pairwise-orthogonal set in $\mathfrak{S}^0$, and this correspondence is injective, it suffices to show that there are only finitely many G-orbits of such sets. This follows from hypothesis (5).

Finally, the maps of the form $\pi_{\mathrm{K}^{\perp}} : G \to \mathcal{C}\mathrm{K}^{\perp}$ and $\rho^*_{\mathrm{K}^{\perp}}$ obviously satisfy the conditions required of an action by HHS automorphisms, since they are constant maps. Finally, for all $\mathrm{K}^\perp \in \bigcup_\eta \mathcal{K}_\eta$ and $[W]$, and $g \in G$, we can choose the arbitrary constant map $\rho^{\mathrm{K}^\perp}_{[W]}$ so that $g(\rho^{\mathrm{K}^\perp}_{[W]}) = \rho^{g\mathrm{K}^\perp}_{g[W]}$, where $g : \mathcal{C}[W] \to \mathcal{C}g[W]$ is the isometry from the automorphism action on $\mathfrak{S}^0$. The same holds with W replaced by T, since $\rho^{\mathrm{K}^\perp}_T$ was defined to be an intersection of support trees associated to $\mathrm{K}^\perp$, and $\rho^{g\mathrm{K}^\perp}_T$ is, by the definition of the G-action on $\bigcup_\eta \mathcal{K}_\eta$ and the G-equivariance of the assignment of each $\sim$-class to its support tree, the intersection of the g-translates of these support trees, i.e., $g\rho^{\mathrm{K}^\perp}_T$. This completes the proof. $\square$

Remark 8.26 (examples where the combination theorem does not apply). Examples where one cannot apply Theorem 8.6 or Corollary 8.24 are likely to yield examples of groups that are not hierarchically hyperbolic groups, or even hierarchically hyperbolic spaces.

(1) Let $\mathcal{G}$ be a finite graph of groups with $\mathbb{Z}^2$ vertex groups and $\mathbb{Z}$ edge groups, i.e., a *tubular group*. Wise [2014] completely characterized the tubular groups that act freely on CAT(0) cube complexes, and also characterized the (rare) tubular groups that admit cocompact such actions; Woodhouse [2016] gave a necessary and sufficient condition for the particular cube complex constructed in [Wise 2014] to be finite-dimensional. These results suggest that there is little hope of producing hierarchically hyperbolic structures for tubular groups via cubulation, except in particularly simple cases.

This is because the obstruction to cocompact cubulation is very similar to the obstruction to building a hierarchically hyperbolic structure using Theorem 8.6. Indeed, if some vertex-group $G_v \cong \mathbb{Z}^2$ has more than two independent incident edge-groups, then, if $\mathcal{G}$ satisfied the hypotheses of Theorem 8.6, the hierarchically hyperbolic structure on G_v would include three pairwise-orthogonal unbounded elements, contradicting partial realization. This shows that such a tubular group does not admit a hierarchically hyperbolic structure by virtue of the obvious splitting, and in fact shows that there is no hierarchically hyperbolic structure in which G_v and the incident edge-groups are hierarchically quasiconvex.

(2) Let $G = F \rtimes_\phi \mathbb{Z}$, where F is a finite-rank free group and $\phi : F \to F$ is an automorphism. When F is atoroidal, G is a hierarchically hyperbolic group simply by virtue of being hyperbolic [Bestvina and Feighn 1992; Brinkmann 2000]. There is also a more refined hierarchically hyperbolic structure in this case, in which all of the hyperbolic spaces involved are quasitrees. Indeed, by combining results in [Hagen and Wise 2015] and [Agol 2013], one finds that G acts freely, cocompactly, and hence virtually co-specially on a CAT(0) cube complex, which therefore contains a G-invariant *factor system* in the sense of [Behrstock et al. 2017b] and is hence a hierarchically hyperbolic group; the construction in [Behrstock et al. 2017b] ensures that the hierarchically

hyperbolic structure for such cube complexes always uses a collection of hyperbolic spaces uniformly quasi-isometric to trees. However, the situation is presumably quite different when G is not hyperbolic. In this case, it seems that G is rarely hierarchically hyperbolic.

8C. *Products.* In this short section, we briefly describe a hierarchically hyperbolic structure on products of hierarchically hyperbolic spaces.

Proposition 8.27 (product HHS). *Let $(\mathcal{X}_0, \mathfrak{S}_0)$ and $(\mathcal{X}_1, \mathfrak{S}_1)$ be hierarchically hyperbolic spaces. Then $\mathcal{X} = \mathcal{X}_0 \times \mathcal{X}_1$ admits a hierarchically hyperbolic structure $(\mathcal{X}, \mathfrak{S})$ such that for each of $i \in \{0, 1\}$ the inclusion map $\mathcal{X}_i \to \mathcal{X}$ induces a quasiconvex hieromorphism.*

Proof. Let $(\mathcal{X}_i, \mathfrak{S}_i)$ be hierarchically hyperbolic spaces for $i \in \{0, 1\}$. Let $\mathfrak{S}$ be a hierarchically hyperbolic structure consisting of the disjoint union of $\mathfrak{S}_0$ and $\mathfrak{S}_1$ (together with their intrinsic hyperbolic spaces, projections, and nesting, orthogonality, and transversality relations), along with the following domains whose associated hyperbolic spaces are points: S, into which everything will be nested; U_i, for $i \in \{0, 1\}$, into which everything in $\mathfrak{S}_i$ is nested; for each $U \in \mathfrak{S}_i$ a domain V_U, with $|\mathcal{C}V_U| = 1$, into which is nested everything in $\mathfrak{S}_{i+1}$ and everything in $\mathfrak{S}_i$ orthogonal to U. The elements V_U are all transverse to U_0 and U_1. Given U, U', the elements $V_U, V_{U'}$ are transverse unless $U \sqsubseteq U'$, in which case $V_U \sqsubseteq V_{U'}$. Projections $\pi_U : \mathcal{X}_0 \times \mathcal{X}_1 \to U \in \mathfrak{S}$ are defined in the obvious way when $U \notin \mathfrak{S}_0 \cup \mathfrak{S}_1$; otherwise, they are the compositions of the existing projections with projection to the relevant factor. Projections of the form ρ^U_V are either defined already, uniquely determined, or are chosen to coincide with the projection of some fixed basepoint (when $V \in \mathfrak{S}_0 \cup \mathfrak{S}_1$ and U is not). It is easy to check that this gives a hierarchically hyperbolic structure on $\mathcal{X}_1 \times \mathcal{X}_2$.

The hieromorphisms $(\mathcal{X}_i, \mathfrak{S}_i) \to (\mathcal{X}, \mathfrak{S})$ are inclusions on $\mathcal{X}_i$ and $\mathfrak{S}$; for each $U \in \mathfrak{S}_i$, the map $\mathfrak{S}_i \ni \mathcal{C}U \to \mathcal{C}U \in \mathfrak{S}$ is the identity. It follows immediately from the definitions that the diagrams from Definition 1.20 coarsely commute, so that these maps are indeed hieromorphisms. Hierarchical quasiconvexity likewise follows from the definition. □

Product HHS will be used in defining hierarchically hyperbolic structures on graph manifolds in Section 10. The next result follows directly from the proof of the previous proposition.

Corollary 8.28. *Let G_0 and G_1 be hierarchically hyperbolic groups. Then $G_0 \times G_1$ is a hierarchically hyperbolic group.*

9. Hyperbolicity relative to HHGs

Relatively hyperbolic groups possess natural hierarchically hyperbolic structures:

Theorem 9.1 (hyperbolicity relative to HHGs). *Let the group G be hyperbolic relative to a finite collection $\mathcal{P}$ of peripheral subgroups. If each $P \in \mathcal{P}$ is a hierarchically hyperbolic space, then G is a hierarchically hyperbolic space. Further, if each $P \in \mathcal{P}$ is a hierarchically hyperbolic group, then so is G.*

Proof. We prove the statement about hierarchically hyperbolic groups; the statement about spaces follows a fortiori.

For each $P \in \mathcal{P}$, let $(P, \mathfrak{S}_P)$ be a hierarchically hyperbolic group structure. For convenience, assume that the $P \in \mathcal{P}$ are pairwise nonconjugate (this will avoid conflicting hierarchically hyperbolic structures). For each P and each left coset gP, let $\mathfrak{S}_{gP}$ be a copy of $\mathfrak{S}_P$ (with associated hyperbolic spaces and projections), so that there is a hieromorphism $(P, \mathfrak{S}_P) \to (gP, \mathfrak{S}_{gP})$, equivariant with respect to the conjugation isomorphism $P \to P^g$.

Let $\widehat{G}$ be the usual hyperbolic space formed from G by coning off each left coset of each $P \in \mathcal{P}$. Let $\mathfrak{S} = \{\widehat{G}\} \cup \bigsqcup_{gP \in G\mathcal{P}} \mathfrak{S}_{gP}$. The nesting, orthogonality, and transversality relations on each $\mathfrak{S}_{gP}$ are as defined above; if $U, V \in \mathfrak{S}_{gP}, \mathfrak{S}_{g'P'}$ and $gP \neq g'P'$, then declare $U \pitchfork V$. Finally, for all $U \in \mathfrak{S}$, let $U \sqsubseteq \widehat{G}$. The hyperbolic space $\mathcal{C}\widehat{G}$ is $\widehat{G}$, while the hyperbolic space $\mathcal{C}U$ associated to each $U \in \mathfrak{S}_{gP}$ was defined above.

The projections are defined as follows: $\pi_{\widehat{G}} : G \to \widehat{G}$ is the inclusion, which is coarsely surjective and hence has quasiconvex image. For each $U \in \mathfrak{S}_{gP}$, let $\mathfrak{g}_{gP} : G \to gP$ be the closest-point projection onto gP and let $\pi_U = \pi_U \circ \mathfrak{g}_{gP}$, to extend the domain of π_U from gP to G. Since each π_U was coarsely Lipschitz on U with quasiconvex image, and the closest-point projection is uniformly coarsely Lipschitz, the projection π_U is uniformly coarsely Lipschitz and has quasiconvex image. For each $U, V \in \mathfrak{S}_{gP}$, the coarse maps ρ^V_U and ρ^U_V were already defined. If $U \in \mathfrak{S}_{gP}$ and $V \in \mathfrak{S}_{g'P'}$, then $\rho^U_V = \pi_V(\mathfrak{g}_{g'P'}(gP))$, which is a uniformly bounded set (here we use relative hyperbolicity, not just the weak relative hyperbolicity that is all we needed so far). Finally, for $U \neq \widehat{G}$, we define $\rho^U_{\widehat{G}}$ to be the cone-point over the unique gP with $U \in \mathfrak{S}_{gP}$, and $\rho^{\widehat{G}}_U : \widehat{G} \to \mathcal{C}U$ is defined as follows: for $x \in G$, let $\rho^{\widehat{G}}_U(x) = \pi_U(x)$. If $x \in \widehat{G}$ is a cone-point over $g'P' \neq gP$, let $\rho^{\widehat{G}}_U(x) = \rho^{S_{g'P'}}_U$, where $S_{g'P'} \in \mathfrak{S}_{g'P'}$ is $\sqsubseteq$-maximal. The cone-point over gP may be sent anywhere in U.

By construction, to verify that $(G, \mathfrak{S})$ is a hierarchically hyperbolic group structure, it suffices to verify that it satisfies the remaining axioms for a hierarchically hyperbolic space given in Definition 1.1, since the additional G-equivariance conditions hold by construction. Specifically, it remains to verify consistency, bounded geodesic image and large links, partial realization, and uniqueness.

Consistency. The nested consistency inequality holds automatically within each $\mathfrak{S}_{gP}$, so it remains to verify it only for $U \in \mathfrak{S}_{gP}$ versus $\widehat{G}$, but this follows directly from the definition: if $x \in G$ is far in $\widehat{G}$ from the cone-point over gP, then

$\rho^{\widehat{G}}_U(x) = \pi_U(x)$, by definition. To verify the transverse inequality, it suffices to consider $U \in \mathfrak{S}_{gP}$, $V \in \mathfrak{S}_{g'P'}$ with $gP \neq g'P'$. Let $x \in G$ and let $z = \mathfrak{g}_{g'P'}(x)$. Then, if $\mathsf{d}_U(x, z)$ is sufficiently large, then $\mathsf{d}_{gP}(x, z)$ is correspondingly large, so that by Lemma 1.15 of [Sisto 2013], $\mathfrak{g}_{g'P'}(x)$ and $\mathfrak{g}_{g'P'}(gP)$ coarsely coincide, as desired.

The last part of the consistency axiom, Definition 1.1.(4), holds as follows. If $U \sqsubseteq V$, then either $U = V$, and there is nothing to prove. Otherwise, if $U \sqsubseteq V$ and either $V \sqsubsetneq W$ or $W \pitchfork V$, then either $U, V \in \mathfrak{S}_{gP}$ for some g, P, or $U \in \mathfrak{S}_{gP}$ and $V = \widehat{G}$. The latter situation precludes the existence of W, so we must be in the former situation. If $W \in \mathfrak{S}_{gP}$, we are done since the axiom holds in $\mathfrak{S}_{gP}$. If $W = \widehat{G}$, then U, V both project to the cone-point over gP, so $\rho^U_W = \rho^V_W$. In the remaining case, $W \in \mathfrak{S}_{g'P'}$ for some $g'P' \neq gP$, in which case ρ^U_W, ρ^V_W both coincide with $\pi_W(\mathfrak{g}_{g'P'}(gP))$.

Bounded geodesic image. Bounded geodesic image holds within each $\mathfrak{S}_{gP}$ by construction, so it suffices to consider the case of $U \in \mathfrak{S}_{gP}$ nested into $\widehat{G}$. Let $\hat{\gamma}$ be a geodesic in $\widehat{G}$ avoiding gP and the cone on gP. Lemma 1.15 of [Sisto 2013] ensures that any lift of $\hat{\gamma}$ has uniformly bounded projection on gP, so $\rho^{\widehat{G}}_U \circ \hat{\gamma}$ is uniformly bounded.

Large links. The large link axiom (Definition 1.1.(6)) can be seen to hold in $(G, \mathfrak{S})$ by combining the large link axiom in each gP with malnormality of $\mathcal{P}$ and Lemma 1.15 of [Sisto 2013].

Partial realization. This follows immediately from partial realization within each $\mathfrak{S}_{gP}$ and the fact that no new orthogonality was introduced in defining $(G, \mathfrak{S})$, together with the definition of $\widehat{G}$ and the definition of projection between elements of $\mathfrak{S}_{gP}$ and $\mathfrak{S}_{g'P'}$ when $gP \neq g'P'$. More precisely, if $U \in \mathfrak{S}_{gP}$ and $p \in \mathcal{C}U$, then by partial realization within gP, there exists $x \in gP$ so that $\mathsf{d}_U(x, p) \leq \alpha$ for some fixed constant α and $\mathsf{d}_V(x, \rho^U_V) \leq \alpha$ for all $V \in \mathfrak{S}_{gP}$ with $U \sqsubseteq V$ or $U \pitchfork V$. Observe that $\mathsf{d}_{\widehat{G}}(x, \rho^U_{\widehat{G}}) = 1$, since $x \in gP$ and $\rho^U_{\widehat{G}}$ is the cone-point over gP. Finally, if $g'P' \neq gP$ and $V \in \mathfrak{S}_{g'P'}$, then $\mathsf{d}_V(x, \rho^U_V) = \mathsf{d}_V(\pi_V(\mathfrak{g}_{g'P'}(x)), \pi_V(\mathfrak{g}_{g'P'}(gP))) = 0$ since $x \in gP$.

Uniqueness. If x, y are uniformly close in $\widehat{G}$, then either they are uniformly close in G, or they are uniformly close to a common cone-point, over some gP, whence the claim follows from the uniqueness axiom in $\mathfrak{S}_{gP}$. □

Remark 9.2. Sisto [2013] established a characterization of relative hyperbolicity in terms of projections and, further, proved that, like for mapping class groups, there was a natural way to compute distances in relatively hyperbolic groups from certain related spaces, namely: if $(G, \mathcal{P})$ is relatively hyperbolic, then distances in G are coarsely obtained by summing the corresponding distance in the coned-off Cayley graph $\widehat{G}$ together with the distances between projections in the various

$P \in \mathcal{P}$ and their cosets. We recover a new proof of Sisto's formula as a consequence of Theorem 9.1 and Theorem 4.5.

Theorem 9.1 will be used in our analysis of 3-manifold groups in Section 10. However, there is a more general statement in the context of metrically relatively hyperbolic spaces (e.g., what Druțu and Sapir [2005] call asymptotically tree-graded, or spaces that satisfy the equivalent condition on projections formulated in [Sisto 2012]). For instance, arguing exactly as in the proof of Theorem 9.1 shows that if the space $\mathcal{X}$ is hyperbolic relative to a collection of uniformly hierarchically hyperbolic spaces, then $\mathcal{X}$ admits a hierarchically hyperbolic structure (in which each peripheral subspace embeds hieromorphically).

More generally, let the geodesic metric space $\mathcal{X}$ be hyperbolic relative to a collection $\mathcal{P}$ of subspaces, and let $\widehat{\mathcal{X}}$ be the hyperbolic space obtained from $\mathcal{X}$ by coning off each $P \in \mathcal{P}$. Then we can endow $\mathcal{X}$ with a hierarchical space structure as follows:

- The index-set $\mathfrak{S}$ consists of $\mathcal{P}$ together with an additional index S.
- For all $P, Q \in \mathcal{P}$, we have $P \pitchfork Q$, while $P \sqsubsetneq S$ for all $P \in \mathcal{P}$ (the orthogonality relation is empty and there is no other nesting).
- For each $P \in \mathcal{P}$, we let $\mathcal{C}P = P$.
- We declare $\mathcal{C}S = \widehat{\mathcal{X}}$.
- The projection $\pi_S : \mathcal{X} \to \widehat{\mathcal{X}}$ is the inclusion.
- For each $P \in \mathcal{P}$, let $\pi_P : \mathcal{X} \to P$ be the closest-point projection onto P (which is surjective).
- For each $P \in \mathcal{P}$, let ρ^P_S be the cone-point in $\widehat{\mathcal{X}}$ associated to P.
- For each $P \in \mathcal{P}$, let $\rho^S_P : \widehat{\mathcal{X}} \to P$ be defined by $\rho^S_P(x) = \pi_P(x)$ for $x \in \mathcal{X}$, while $\rho^S_P(x) = \pi_P(Q)$ whenever x lies in the cone on $Q \in \mathcal{P}$.
- For distinct $P, Q \in \mathcal{P}$, let $\rho^P_Q = \pi_Q(P)$ (which is uniformly bounded since $\mathcal{X}$ is hyperbolic relative to $\mathcal{P}$).

The above definition yields:

Theorem 9.3. *Let the geodesic metric space $\mathcal{X}$ be hyperbolic relative to the collection $\mathcal{P}$ of subspaces. Then, with $\mathfrak{S}$ as above, we have that $(\mathcal{X}, \mathfrak{S})$ is a hierarchical space, and is moreover relatively hierarchically hyperbolic.*

Proof. By definition, for each $U \in \mathfrak{S}$, we have that either $U = S$ and $\mathcal{C}S = \widehat{\mathcal{X}}$ is hyperbolic, or U is $\sqsubseteq$-minimal. The rest of the conditions of Definition 1.1 are verified as in the proof of Theorem 9.1. □

10. Hierarchical hyperbolicity of 3-manifold groups

In this section we show that fundamental groups of most 3-manifolds admit hierarchical hyperbolic structures. More precisely, we prove:

Theorem 10.1 (which 3-manifolds are hierarchically hyperbolic). *Let M be a closed 3-manifold. Then $\pi_1(M)$ is a hierarchically hyperbolic space if and only if M does not have a Sol or Nil component in its prime decomposition.*

Proof. It is well known that for a closed irreducible 3-manifold N the Dehn function of $\pi_1(N)$ is linear if N is hyperbolic, cubic if N is Nil, exponential if N is Sol, and quadratic in all other cases. Hence by Corollary 7.5, if $\pi_1(M)$ is a hierarchically hyperbolic space, then M does not contain Nil or Sol manifolds in its prime decomposition. It remains to prove the converse.

Since the fundamental group of any reducible 3-manifold is the free product of irreducible ones, the reducible case immediately follows from the irreducible case by Theorem 9.1.

When M is geometric and not Nil or Sol, then $\pi_1(M)$ is quasi-isometric to one of the following:

- $\mathbb{R}^3$ is hierarchically hyperbolic by Proposition 8.27.
- $\mathbb{H}^3$, $\mathbb{S}^3$, $\mathbb{S}^2 \times \mathbb{R}$ are (hierarchically) hyperbolic.
- $\mathbb{H}^2 \times \mathbb{R}$ and $PSL_2(\mathbb{R})$: the first is hierarchically hyperbolic by Proposition 8.27, whence the second is also since it is quasi-isometric to the first by [Rieffel 2001].

We may now assume M is not geometric. Our main step is to show that any irreducible nongeometric graph manifold group is a hierarchically hyperbolic space.

Let M be an irreducible nongeometric graph manifold. By [Kapovich and Leeb 1998, Theorem 2.3], by replacing M by a manifold whose fundamental group is quasi-isometric to that of M, we may assume that our manifold is a *flip graph manifold*, i.e., each Seifert fibered space component is a trivial circle bundle over a surface of genus at least 2 and each pair of adjacent Seifert fibered spaces are glued by flipping the base and fiber directions.

Let X be the universal cover of M. The decomposition of M into geometric components induces a decomposition of X into subspaces $\{S_v\}$, one for each vertex v of the Bass–Serre tree T of M. Each such subspace S_v is bi-Lipschitz homeomorphic to the product of a copy R_v of the real line with the universal cover Σ_v of a hyperbolic surface with totally geodesic boundary, and there are maps $\phi_v : S_v \to \Sigma_v$ and $\psi_v : S_v \to R_v$. Notice that Σ_v is hyperbolic, and in particular hierarchically hyperbolic. However, for later purposes, we endow Σ_v with the hierarchically hyperbolic structure originating from the fact that Σ_v is hyperbolic relative to its boundary components, see Theorem 9.1.

By Proposition 8.27 each S_v is a hierarchically hyperbolic space and thus we have a tree of hierarchically hyperbolic spaces. Each edge space is a product $\partial_0 \Sigma_v \times R_v$, where $\partial_0 \Sigma_v$ is a particular boundary component of Σ_v determined by the adjacent vertex. Further, since the graph manifold is flip, we also have that for any vertices v, w of the tree, the edge-hieromorphism between S_v and S_w sends $\partial_0 \Sigma_v$ to R_w and R_v to $\partial_0 \Sigma_w$.

We now verify the hypotheses of Theorem 8.6. The first hypothesis is that there exists k so that each edge-hieromorphism is k-hierarchically quasiconvex. This is easily seen since the edge-hieromorphisms have the simple form described above. The second hypothesis of Theorem 8.6, fullness of edge-hieromorphisms, also follows immediately from the explicit description of the edges here and the simple hierarchically hyperbolic structure of the edge spaces.

The third hypothesis of Theorem 8.6 requires that the tree has bounded supports. We can assume that the product regions S_v are maximal in the sense that each edge-hieromorphism sends the fiber direction R_v to $\partial_0 \Sigma_w$ in each adjacent S_w. It follows that the support of each $\sim$-class (in the language of Theorem 8.6) consists of at most 2 vertices. The last hypothesis of Theorem 8.6 is about nonorthogonality of maximal elements and again follows directly from the explicit hierarchically hyperbolic structure. Moreover, the part of the hypothesis about edge-spaces coinciding coarsely with standard product regions in vertex spaces follows from the explicit hierarchically hyperbolic structure.

All the hypotheses of Theorem 8.6 are satisfied, so $\pi_1 M$ (with any word metric) is a hierarchically hyperbolic space for all irreducible nongeometric graph manifolds M.

The general case that the fundamental group of any nongeometric 3-manifold is a hierarchically hyperbolic space now follows immediately by Theorem 9.1, since any 3-manifold group is hyperbolic relative to its maximal graph manifold subgroups. □

Remark 10.2 ((non)existence of HHG structures for 3–manifold groups). The proof of Theorem 10.1 shows that for many 3-manifolds M, the group $\pi_1 M$ is not merely a hierarchically hyperbolic space (when endowed with the word metric arising from a finite generating set), but is actually a hierarchically hyperbolic group. Specifically, if M is virtually compact special, then $\pi_1 M$ acts freely and cocompactly on a CAT(0) cube complex $\mathcal{X}$ that is the universal cover of a compact special cube complex. Hence $\mathcal{X}$ contains a $\pi_1 M$-invariant factor system (see [Behrstock et al. 2017b, Section 8]) consisting of a $\pi_1 M$-finite set of convex subcomplexes. This yields a hierarchically hyperbolic structure $(\mathcal{X}, \mathfrak{S})$ where $\pi_1 M \leq \mathrm{Aut}(\mathfrak{S})$ acts cofinitely on $\mathfrak{S}$ and geometrically on $\mathcal{X}$, i.e., $\pi_1 M$ is a hierarchically hyperbolic group.

The situation is quite different when $\pi_1 M$ is not virtually *compact* special. Indeed, when M is a nonpositively curved graph manifold, $\pi_1 M$ virtually acts freely, but not necessarily cocompactly, on a CAT(0) cube complex $\mathcal{X}$, and the quotient is virtually special; this is a result of Liu [2013] which was also shown to hold in the case where M has nonempty boundary by Przytycki and Wise [2014]. Moreover, $\pi_1 M$ acts with finitely many orbits of hyperplanes. Hence the $\pi_1 M$-invariant factor system on $\mathcal{X}$ from [Behrstock et al. 2017b] yields a $\pi_1 M$-equivariant HHS structure $(\mathcal{X}, \mathfrak{S})$ with $\mathfrak{S}$ $\pi_1 M$-finite. However, this yields an HHG structure on $\pi_1 M$ only if the action on $\mathcal{X}$ is cocompact. It was shown in [Hagen and Przytycki 2015] that $\pi_1 M$ virtually acts freely and cocompactly on a CAT(0) cube complex, with special quotient, only in the very particular situation where M is *chargeless*. This essentially asks whether the construction of the hierarchically hyperbolic structure on $\widetilde{M}$ from the proof of Theorem 10.1 can be done $\pi_1 M$-equivariantly. In general, this is impossible: recall that we passed from $\widetilde{M}$ to the universal cover of a flip manifold using a (nonequivariant) quasi-isometry. Motivated by this observation and the fact that the range of possible HHS structures on the universal cover of a JSJ torus is very limited, we conjecture that $\pi_1 M$ is a hierarchically hyperbolic group if and only if $\pi_1 M$ acts freely and cocompactly on a CAT(0) cube complex.

11. A new proof of the distance formula for mapping class groups

We now describe the hierarchically hyperbolic structure of mapping class groups. In [Behrstock et al. 2017b] we gave a proof of this result using several of the main results of [Behrstock 2006; 2012; Masur and Minsky 1999; 2000]. Here we give an elementary proof which is independent of the Masur–Minsky "hierarchy machinery." One consequence of this is a new and concise proof of the celebrated Masur–Minsky distance formula [2000, Theorem 6.12], which we obtain by combining Theorems 4.5 and 11.1.

(1) Let S be a closed connected oriented surface of finite type and let $\mathcal{M}(S)$ be its marking complex.

(2) Let $\mathfrak{S}$ be the collection of isotopy classes of essential nonpants subsurfaces of S, and for each $U \in \mathfrak{S}$ let $\mathcal{C}U$ be its curve complex. (We allow disconnected subsurfaces; the curve graph of a disconnected surface is the join of the curve graphs of its components.)

(3) The relation $\sqsubseteq$ is nesting, $\perp$ is disjointness and $\pitchfork$ is overlapping.

(4) For each $U \in \mathfrak{S}$, let $\pi_U : \mathcal{M}(S) \to \mathcal{C}U$ be the (usual) subsurface projection. For $U, V \in \mathfrak{S}$ satisfying either $U \sqsubseteq V$ or $U \pitchfork V$, denote $\rho_V^U = \pi_V(\partial U) \in \mathcal{C}V$, while for $V \sqsubseteq U$ let $\rho_V^U : \mathcal{C}U \to 2^{\mathcal{C}V}$ be the subsurface projection. When U is a component of the disconnected subsurface V, let ρ_V^U be the curve graph

of U, which is a subgraph of $\mathcal{C}V$ of bounded diameter. In general, if $U \subsetneq V$, then ρ^U_V is the union of the subsets $\rho^{U_i}_V$ where U_i varies over the components of U.

Theorem 11.1. *Let S be closed connected oriented surface of finite type. Then, $(\mathcal{M}(S), \mathfrak{S})$ is a hierarchically hyperbolic space, for $\mathfrak{S}$ as above. In particular the mapping class group $\mathcal{MCG}(S)$ is a hierarchically hyperbolic group.*

Proof. Hyperbolicity of curve graphs is the main result of [Masur and Minsky 1999]; more recent proofs of this were found in [Aougab 2013; Bowditch 2014b; Clay et al. 2014; Hensel et al. 2015; Przytycki and Sisto 2017], some of which are elementary.

Axioms (1), (2), (3) and (5) are clear (an elementary exposition of the Lipschitz condition for subsurface projections is provided in [Masur and Minsky 2000, Lemma 2.5], and the projections have quasiconvex image because they are coarsely surjective). Both parts of axiom (4) can be found in [Behrstock 2006]. The nesting part is elementary, and a short elementary proof in the overlapping case was obtained by Leininger and can be found in [Mangahas 2010].

Axiom (7) was proven in [Masur and Minsky 2000], and an elementary proof is available in [Webb 2015]. In fact, in the aforementioned papers it is proven that there exists a constant C so that for any subsurface W, markings x, y and geodesic from $\pi_W(x)$ to $\pi_W(y)$, the following holds. If $V \sqsubseteq W$ and $V \neq W$ satisfies $d_V(x, y) \geq C$ then some curve along the given geodesic does not intersect ∂V. This implies axiom (6), since we can take the T_i to be the complements of curves appearing along the aforementioned geodesic.

Axiom (8) follows easily from the following statement, which clearly holds: For any given collection of disjoint subsurfaces and curves on the given subsurfaces, there exists a marking on S that contains the given curves as base curves (or, up to bounded error, transversals in the case that the corresponding subsurface is an annulus).

Axiom (9) is hence the only delicate one. We are finished modulo this last axiom which we verify below in Proposition 11.2 (see also [Bestvina et al. 2015, Proposition 5.11]). □

Proposition 11.2. $(\mathcal{M}(S), \mathfrak{S})$ *satisfies the uniqueness axiom, i.e., for each $\kappa \geq 0$, there exists $\theta_u = \theta_u(\kappa)$ such that if $x, y \in \mathcal{M}(S)$ satisfy $\mathsf{d}_U(x, y) \leq \kappa$ for each $U \in \mathfrak{S}$ then $\mathsf{d}_{\mathcal{M}(S)}(x, y) \leq \theta_u$.*

Proof. Note that when the complexity (as measured by the quantity $3g + p - 3$ where g is the genus and p the number of punctures) is less than 2 then $\mathcal{M}(S)$ is hyperbolic and thus the axiom holds. We will proceed by inducting on complexity: thus we will fix S to have complexity at least 2 and assume that all the axioms for a hierarchically hyperbolic space, including the uniqueness axiom, hold for each proper subsurface of S.

Having fixed our surface S, the proof is by induction on $\mathsf{d}_{\mathcal{C}S}(base(x), base(y))$.

If $\mathsf{d}_{\mathcal{C}S}(base(x), base(y)) = 0$, then x and y share some nonempty multicurve $\sigma = c_1 \cup \cdots \cup c_k$. For x', y' the restrictions of x, y to $S - \sigma$ we have that, by induction, $\mathsf{d}_{\mathcal{M}(S-\sigma)}(x', y')$ is bounded in terms of κ. We then take the markings in a geodesic in $\mathcal{M}(S - \sigma)$ from x' to y' and extend these all in the same way to obtain markings in $\mathcal{M}(S)$ which yield a path in $\mathcal{M}(S)$ from x to $\hat{y}$ whose length is bounded in terms of κ, where $\hat{y}$ is the marking for which

- $\hat{y}$ has the same base curves as y,
- the transversal for each c_i is the same as the corresponding transversal for x, and
- the transversal for each curve in $base(y) - \{c_i\}$ is the same as the corresponding transversal for y.

Finally, it is readily seen that $\mathsf{d}_{\mathcal{M}(S)}(\hat{y}, y)$ is bounded in terms of κ because the transversals of each c_i in the markings x and y are within distance κ of each other. This completes the proof of the base case of the proposition.

Suppose now that the statement holds whenever $\mathsf{d}_{\mathcal{C}S}(base(x), base(y)) \leq n$, and let us prove it in the case $\mathsf{d}_{\mathcal{C}S}(base(x), base(y)) = n + 1$. Let $c_x \in base(x)$ and $c_y \in base(y)$ satisfy $\mathsf{d}_{\mathcal{C}S}(c_x, c_y) = n + 1$. Let $c_x = \sigma_0, \ldots, \sigma_{n+1} = c_y$ be a tight geodesic (hence, each σ_i is a multicurve). Let σ be the union of σ_0 and σ_1. Using the realization theorem in the subsurface $S - \sigma$ we can find a marking x' in $S - \sigma$ whose projections onto each $\mathcal{C}U$ for $U \subseteq S - \sigma$ coarsely coincide with $\pi_U(y)$. Let $\hat{x}$ be the marking for which

- $base(\hat{x})$ is the union of $base(x')$ and σ,
- the transversal in $\hat{x}$ of curves in $base(\hat{x}) \cap base(x')$ are the same as those in x',
- the transversal of c_x in $\hat{x}$ is the same as the one in x,
- the transversal in $\hat{x}$ of a curve c in σ_1 is $\pi_{A_c}(y)$, where A_c is an annulus around c.

Note that $\mathsf{d}_{\mathcal{C}S}(base(\hat{x}), base(y)) = n$. Hence, the following claims conclude the proof.

Claim 1. $\mathsf{d}_{\mathcal{M}(S)}(x, \hat{x})$ is bounded in terms of κ.

Proof. It suffices to bound $\mathsf{d}_{\mathcal{C}U}(x, \hat{x})$ in terms of κ for each $U \subseteq S - c_x$. In fact, once we do that, by induction on complexity we know that we can bound $\mathsf{d}_{\mathcal{M}(S-c_x)}(z, \hat{z})$, where $z, \hat{z}$ are the restrictions of $x, \hat{x}$ to $S - c_x$, whence the conclusion easily follows.

If U is contained in $S - \sigma$, then the required bound follows since $\pi_U(\hat{x})$ coarsely coincides with $\pi_U(x')$ in this case.

If instead ∂U intersects σ_1, then $\pi_U(\hat{x})$ coarsely coincides with $\pi_U(\sigma_1)$.

At this point, we only have to show that $\pi_U(\sigma_1)$ coarsely coincides with $\pi_U(y)$, and in order to do so we observe that we can apply the bounded geodesic image theorem to the geodesic $\sigma_1, \dots, \sigma_{n+1}$. In fact, σ_1 intersects ∂U by hypothesis and σ_i intersects ∂U for $i \geq 3$ because of the following estimate that holds for any given boundary component c of ∂U:

$$\mathsf{d}_{\mathcal{C}(S)}(\sigma_i, c) \geq \mathsf{d}_{\mathcal{C}(S)}(\sigma_i, \sigma_0) - \mathsf{d}_{\mathcal{C}(S)}(\sigma_0, c) \geq i - 1 > 1.$$

Lastly, σ_2 intersects ∂U because of the definition of tightness: ∂U intersects σ_1, so it must intersect $\sigma_0 \cup \sigma_2$, but, it does not intersect σ_0, so it intersects σ_2. □

Claim 2. There exists κ', depending on κ, so that for each subsurface U of S we have $\mathsf{d}_{\mathcal{C}U}(\hat{x}, y) \leq \kappa'$.

Proof. If σ_0 intersects ∂U, then $\pi_U(\hat{x})$ coarsely coincides with $\pi_U(\sigma_0)$. In turn, $\pi_U(\sigma_0)$ coarsely coincides with $\pi_U(x)$, which is κ-close to $\pi_U(y)$.

On the other hand, if U does not intersect σ, then we are done by the definition of x'.

Hence, we can assume that U is contained in $S - \sigma_0$ and that σ_1 intersects ∂U. In particular, $\pi_U(\hat{x})$ coarsely coincides with $\pi_U(\sigma_1)$. But we showed in the last paragraph of the proof of Claim 1 that $\pi_U(\sigma_1)$ coarsely coincides with $\pi_U(y)$, so we are done. □

As explained above, the proofs of the above two claims complete the proof. □

Acknowledgments

Hagen and Sisto thank the organizers of the Ventotene 2015 conference, and Behrstock and Hagen thank CIRM and the organizers of GAGTA 2015, where some of the work on this paper was completed. We thank Federico Berlai, Brian Bowditch, Bruno Robbio, Jacob Russell and Davide Spriano for numerous comments/corrections, and anonymous referees for further helpful comments.

References

[Abbott and Behrstock 2018] C. Abbott and J. Behrstock, "Conjugator lengths in hierarchically hyperbolic groups", preprint, 2018. arXiv

[Abbott et al. 2017] C. Abbott, J. Behrstock, and M. G. Durham, "Largest acylindrical actions and stability in hierarchically hyperbolic groups", preprint, 2017. arXiv

[Agol 2013] I. Agol, "The virtual Haken conjecture", *Doc. Math.* **18** (2013), 1045–1087. MR Zbl

[Aougab 2013] T. Aougab, "Uniform hyperbolicity of the graphs of curves", *Geom. Topol.* **17**:5 (2013), 2855–2875. MR Zbl

[Behrstock 2006] J. A. Behrstock, "Asymptotic geometry of the mapping class group and Teichmüller space", *Geom. Topol.* **10** (2006), 1523–1578. MR Zbl

[Behrstock and Hagen 2016] J. Behrstock and M. F. Hagen, "Cubulated groups: thickness, relative hyperbolicity, and simplicial boundaries", *Groups Geom. Dyn.* **10**:2 (2016), 649–707. MR Zbl

[Behrstock and Minsky 2008] J. A. Behrstock and Y. N. Minsky, "Dimension and rank for mapping class groups", *Ann. of Math.* (2) **167**:3 (2008), 1055–1077. MR Zbl

[Behrstock and Minsky 2011] J. A. Behrstock and Y. N. Minsky, "Centroids and the rapid decay property in mapping class groups", *J. Lond. Math. Soc.* (2) **84**:3 (2011), 765–784. MR Zbl

[Behrstock et al. 2011] J. Behrstock, C. Druțu, and M. Sapir, "Median structures on asymptotic cones and homomorphisms into mapping class groups", *Proc. Lond. Math. Soc.* (3) **102**:3 (2011), 503–554. MR Zbl

[Behrstock et al. 2012] J. Behrstock, B. Kleiner, Y. Minsky, and L. Mosher, "Geometry and rigidity of mapping class groups", *Geom. Topol.* **16**:2 (2012), 781–888. MR Zbl

[Behrstock et al. 2017a] J. Behrstock, M. F. Hagen, and A. Sisto, "Asymptotic dimension and small-cancellation for hierarchically hyperbolic spaces and groups", *Proc. Lond. Math. Soc.* (3) **114**:5 (2017), 890–926. MR Zbl

[Behrstock et al. 2017b] J. Behrstock, M. F. Hagen, and A. Sisto, "Hierarchically hyperbolic spaces, I: Curve complexes for cubical groups", *Geom. Topol.* **21**:3 (2017), 1731–1804. MR Zbl

[Behrstock et al. 2017c] J. Behrstock, M. F. Hagen, and A. Sisto, "Quasiflats in hierarchically hyperbolic spaces", preprint, 2017. arXiv

[Berlai and Robbio 2018] F. Berlai and B. Robbio, "A refined combination theorem for hierarchically hyperbolic groups", preprint, 2018. arXiv

[Bestvina and Feighn 1992] M. Bestvina and M. Feighn, "A combination theorem for negatively curved groups", *J. Differential Geom.* **35**:1 (1992), 85–101. MR Zbl

[Bestvina and Feighn 2014a] M. Bestvina and M. Feighn, "Hyperbolicity of the complex of free factors", *Adv. Math.* **256** (2014), 104–155. MR Zbl

[Bestvina and Feighn 2014b] M. Bestvina and M. Feighn, "Subfactor projections", *J. Topol.* **7**:3 (2014), 771–804. MR Zbl

[Bestvina and Fujiwara 2002] M. Bestvina and K. Fujiwara, "Bounded cohomology of subgroups of mapping class groups", *Geom. Topol.* **6** (2002), 69–89. MR Zbl

[Bestvina et al. 2015] M. Bestvina, K. Bromberg, and K. Fujiwara, "Constructing group actions on quasi-trees and applications to mapping class groups", *Publ. Math. Inst. Hautes Études Sci.* **122**:1 (2015), 1–64. MR Zbl

[Bowditch 2013] B. H. Bowditch, "Coarse median spaces and groups", *Pacific J. Math.* **261**:1 (2013), 53–93. MR Zbl

[Bowditch 2014a] B. H. Bowditch, "Embedding median algebras in products of trees", *Geom. Dedicata* **170**:1 (2014), 157–176. MR Zbl

[Bowditch 2014b] B. H. Bowditch, "Uniform hyperbolicity of the curve graphs", *Pacific J. Math.* **269**:2 (2014), 269–280. MR Zbl

[Bowditch 2018] B. H. Bowditch, "Large-scale rigidity properties of the mapping class groups", *Pacific J. Math.* **293**:1 (2018), 1–73. MR Zbl

[Bridson 2010] M. R. Bridson, "Semisimple actions of mapping class groups on CAT(0) spaces", pp. 1–14 in *Geometry of Riemann surfaces* (Crete, 2007), edited by F. P. Gardiner et al., London Math. Soc. Lecture Note Ser. **368**, Cambridge Univ. Press, 2010. MR Zbl

[Bridson and Vogtmann 1995] M. R. Bridson and K. Vogtmann, "On the geometry of the automorphism group of a free group", *Bull. London Math. Soc.* **27**:6 (1995), 544–552. MR Zbl

[Bridson and Vogtmann 2012] M. R. Bridson and K. Vogtmann, "The Dehn functions of $\mathrm{Out}(F_n)$ and $\mathrm{Aut}(F_n)$", *Ann. Inst. Fourier* (*Grenoble*) **62**:5 (2012), 1811–1817. MR Zbl

[Brinkmann 2000] P. Brinkmann, "Hyperbolic automorphisms of free groups", *Geom. Funct. Anal.* **10**:5 (2000), 1071–1089. MR Zbl

[Buyalo and Svetlov 2004] S. V. Buyalo and P. V. Svetlov, "Topological and geometric properties of graph manifolds", *Algebra i Analiz* **16**:2 (2004), 3–68. In Russian; translated in *St. Petersburg Math. J.* **16** (2005), 297–340. MR Zbl

[Chatterji et al. 2010] I. Chatterji, C. Druţu, and F. Haglund, "Kazhdan and Haagerup properties from the median viewpoint", *Adv. Math.* **225**:2 (2010), 882–921. MR Zbl

[Chepoi 2000] V. Chepoi, "Graphs of some CAT(0) complexes", *Adv. in Appl. Math.* **24**:2 (2000), 125–179. MR Zbl

[Clay et al. 2012] M. T. Clay, C. J. Leininger, and J. Mangahas, "The geometry of right-angled Artin subgroups of mapping class groups", *Groups Geom. Dyn.* **6**:2 (2012), 249–278. MR Zbl

[Clay et al. 2014] M. Clay, K. Rafi, and S. Schleimer, "Uniform hyperbolicity of the curve graph via surgery sequences", *Algebr. Geom. Topol.* **14**:6 (2014), 3325–3344. MR Zbl

[Dilworth 1950] R. P. Dilworth, "A decomposition theorem for partially ordered sets", *Ann. of Math.* (2) **51** (1950), 161–166. MR Zbl

[Druţu and Sapir 2005] C. Druţu and M. Sapir, "Tree-graded spaces and asymptotic cones of groups", *Topology* **44**:5 (2005), 959–1058. MR Zbl

[Durham and Taylor 2015] M. G. Durham and S. J. Taylor, "Convex cocompactness and stability in mapping class groups", *Algebr. Geom. Topol.* **15**:5 (2015), 2839–2859. MR Zbl

[Durham et al. 2017] M. G. Durham, M. F. Hagen, and A. Sisto, "Boundaries and automorphisms of hierarchically hyperbolic spaces", *Geom. Topol.* **21**:6 (2017), 3659–3758. MR Zbl

[Frigerio et al. 2015] R. Frigerio, J.-F. Lafont, and A. Sisto, *Rigidity of high dimensional graph manifolds*, Astérisque **372**, Soc. Math. France, Paris, 2015. MR Zbl arXiv

[Gersten 1992] S. M. Gersten, "Dehn functions and l_1-norms of finite presentations", pp. 195–224 in *Algorithms and classification in combinatorial group theory* (Berkeley, 1989), edited by G. Baumslag, Math. Sci. Res. Inst. Publ. **23**, Springer, 1992. MR Zbl

[Gersten et al. 2003] S. M. Gersten, D. F. Holt, and T. R. Riley, "Isoperimetric inequalities for nilpotent groups", *Geom. Funct. Anal.* **13**:4 (2003), 795–814. MR Zbl

[Haettel 2016] T. Haettel, "Higher rank lattices are not coarse median", *Algebr. Geom. Topol.* **16**:5 (2016), 2895–2910. MR Zbl

[Hagen 2014] M. F. Hagen, "Weak hyperbolicity of cube complexes and quasi-arboreal groups", *J. Topol.* **7**:2 (2014), 385–418. MR Zbl

[Hagen and Przytycki 2015] M. F. Hagen and P. Przytycki, "Cocompactly cubulated graph manifolds", *Israel J. Math.* **207**:1 (2015), 377–394. MR Zbl

[Hagen and Susse 2016] M. F. Hagen and T. Susse, "On hierarchical hyperbolicity of cubical groups", preprint, 2016. arXiv

[Hagen and Wise 2015] M. F. Hagen and D. T. Wise, "Cubulating hyperbolic free-by-cyclic groups: the general case", *Geom. Funct. Anal.* **25**:1 (2015), 134–179. MR Zbl

[Haglund 2007] F. Haglund, "Isometries of CAT(0) cube complexes are semi-simple", preprint, 2007. arXiv

[Handel and Mosher 2013a] M. Handel and L. Mosher, "The free splitting complex of a free group, I: Hyperbolicity", *Geom. Topol.* **17**:3 (2013), 1581–1672. MR Zbl

[Handel and Mosher 2013b] M. Handel and L. Mosher, "Lipschitz retraction and distortion for subgroups of Out(F_n)", *Geom. Topol.* **17**:3 (2013), 1535–1579. MR Zbl

[Hensel et al. 2015] S. Hensel, P. Przytycki, and R. C. H. Webb, "1-slim triangles and uniform hyperbolicity for arc graphs and curve graphs", *J. Eur. Math. Soc.* **17**:4 (2015), 755–762. MR Zbl

[Kapovich and Leeb 1996] M. Kapovich and B. Leeb, "Actions of discrete groups on nonpositively curved spaces", *Math. Ann.* **306**:2 (1996), 341–352. MR Zbl

[Kapovich and Leeb 1997] M. Kapovich and B. Leeb, "Quasi-isometries preserve the geometric decomposition of Haken manifolds", *Invent. Math.* **128**:2 (1997), 393–416. MR Zbl

[Kapovich and Leeb 1998] M. Kapovich and B. Leeb, "3-manifold groups and nonpositive curvature", *Geom. Funct. Anal.* **8**:5 (1998), 841–852. MR Zbl

[Kim and Koberda 2014] S.-H. Kim and T. Koberda, "The geometry of the curve graph of a right-angled Artin group", *Internat. J. Algebra Comput.* **24**:2 (2014), 121–169. MR Zbl

[Liu 2013] Y. Liu, "Virtual cubulation of nonpositively curved graph manifolds", *J. Topol.* **6**:4 (2013), 793–822. MR Zbl

[Mangahas 2010] J. Mangahas, "Uniform uniform exponential growth of subgroups of the mapping class group", *Geom. Funct. Anal.* **19**:5 (2010), 1468–1480. MR Zbl

[Masur and Minsky 1999] H. A. Masur and Y. N. Minsky, "Geometry of the complex of curves, I: Hyperbolicity", *Invent. Math.* **138**:1 (1999), 103–149. MR Zbl

[Masur and Minsky 2000] H. A. Masur and Y. N. Minsky, "Geometry of the complex of curves, II: Hierarchical structure", *Geom. Funct. Anal.* **10**:4 (2000), 902–974. MR Zbl

[Minasyan and Osin 2015] A. Minasyan and D. Osin, "Acylindrical hyperbolicity of groups acting on trees", *Math. Ann.* **362**:3-4 (2015), 1055–1105. MR Zbl

[Mosher 1995] L. Mosher, "Mapping class groups are automatic", *Ann. of Math.* (2) **142**:2 (1995), 303–384. MR Zbl

[Mousley 2017] S. C. Mousley, "Exotic limit sets of Teichmüller geodesics in the HHS boundary", preprint, 2017. arXiv

[Mousley 2018] S. C. Mousley, "Nonexistence of boundary maps for some hierarchically hyperbolic spaces", *Algebr. Geom. Topol.* **18**:1 (2018), 409–439. MR Zbl

[Mousley and Russell 2018] S. C. Mousley and J. Russell, "Hierarchically hyperbolic groups are determined by their Morse boundaries", preprint, 2018. arXiv

[Przytycki and Sisto 2017] P. Przytycki and A. Sisto, "A note on acylindrical hyperbolicity of mapping class groups", pp. 255–264 in *Hyperbolic geometry and geometric group theory* (Tokyo, 2014), edited by K. Fujiwara et al., Adv. Stud. Pure Math. **73**, Math. Soc. Japan, Tokyo, 2017. MR

[Przytycki and Wise 2014] P. Przytycki and D. T. Wise, "Graph manifolds with boundary are virtually special", *J. Topol.* **7**:2 (2014), 419–435. MR Zbl

[Rieffel 2001] E. G. Rieffel, "Groups quasi-isometric to $\mathbb{H}^2 \times \mathbb{R}$", *J. London Math. Soc.* (2) **64**:1 (2001), 44–60. MR Zbl

[Sabalka and Savchuk 2012] L. Sabalka and D. Savchuk, "Submanifold projection", preprint, 2012. arXiv

[Sisto 2011] A. Sisto, "3-manifold groups have unique asymptotic cones", preprint, 2011. arXiv

[Sisto 2012] A. Sisto, "On metric relative hyperbolicity", preprint, 2012. arXiv

[Sisto 2013] A. Sisto, "Projections and relative hyperbolicity", *Enseign. Math.* (2) **59**:1-2 (2013), 165–181. MR Zbl

[Spriano 2017] D. Spriano, "Hyperbolic HHS, I: Factor systems and quasi-convex subgroups", preprint, 2017. arXiv

[Spriano 2018] D. Spriano, "Hyperbolic HHS, II: Graphs of hierarchically hyperbolic groups", preprint, 2018. arXiv

[Vokes 2017] K. M. Vokes, "Hierarchical hyperbolicity of graphs of multicurves", preprint, 2017. arXiv

[Webb 2015] R. C. H. Webb, "Uniform bounds for bounded geodesic image theorems", *J. Reine Angew. Math.* **709** (2015), 219–228. MR Zbl

[Wise 2014] D. T. Wise, "Cubular tubular groups", *Trans. Amer. Math. Soc.* **366**:10 (2014), 5503–5521. MR Zbl

[Woodhouse 2016] D. J. Woodhouse, "Classifying finite dimensional cubulations of tubular groups", *Michigan Math. J.* **65**:3 (2016), 511–532. MR Zbl

Received January 25, 2018. Revised November 15, 2018.

JASON BEHRSTOCK
THE GRADUATE CENTER AND LEHMAN COLLEGE
CUNY
NEW YORK, NY
UNITED STATES
jason@math.columbia.edu

MARK HAGEN
SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
BRISTOL
UNITED KINGDOM
markfhagen@gmail.com

ALESSANDRO SISTO
DEPARTEMENT MATHEMATIK
ETH ZÜRICH
ZÜRICH
SWITZERLAND
sisto@math.ethz.ch

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 2, 2019
dx.doi.org/10.2140/pjm.2019.299.339

THE WEIGHTED σ_k-CURVATURE OF A SMOOTH METRIC MEASURE SPACE

JEFFREY S. CASE

We propose a definition of the weighted σ_k-curvature of a smooth metric measure space and justify it in two ways. First, we show that the weighted σ_k-curvature prescription problem is governed by a fully nonlinear second order elliptic PDE which is variational when $k = 1, 2$ or the smooth metric measure space is locally conformally flat in the weighted sense. Second, we show that, in the variational cases, quasi-Einstein metrics are stable with respect to the total weighted σ_k-curvature functional. We also discuss related conjectures for weighted Einstein manifolds.

1. Introduction

In Riemannian geometry, the σ_k-curvatures are scalar Riemannian invariants which have proven to be useful tools for studying geometric and analytic properties of Riemannian manifolds. For example, locally conformally flat Einstein metrics with nonzero scalar curvature locally extremize the total σ_k-curvature functional within their conformal class for all k; when $k \leq 2$, the same is true for all Einstein metrics with nonzero scalar curvature [Viaclovsky 2000]. This greatly expands the set of Riemannian functionals which one can use to study Einstein metrics and leads to new variational characterizations of such manifolds; see, e.g., [Guan et al. 2003; Gursky and Viaclovsky 2001]. Moreover, one can classify all critical points of

MSC2010: primary 53C21; secondary 35J60, 58E11.
Keywords: smooth metric measure space, σ_k-curvature, quasi-Einstein, weighted Einstein.

the total σ_k-curvature functional in the positive k-cone of the conformal class of the round metrics on the sphere [Chang et al. 2003; González 2006; Li and Li 2003; Viaclovsky 2000], providing an important first step to proving sharp fully nonlinear Sobolev inequalities [Guan and Wang 2004]. Since the σ_1-curvature is a dimensional multiple of the scalar curvature, these facts naturally generalize well-known properties of the Yamabe functional; cf. [Lee and Parker 1987].

Roughly speaking, smooth metric measure spaces are Riemannian manifolds equipped with a smooth measure. The geometric study of smooth metric measure spaces is based in large part on the m-Bakry–Émery Ricci tensor. This tensor generalizes the Ricci tensor and thereby leads to the notion of a "gradient Einstein-type manifold", obtained by requiring the m-Bakry–Émery Ricci tensor to be a multiple of the metric. Such manifolds include as special cases gradient Ricci solitons and static metrics in general relativity, and this framework provides a useful uniform approach to their study; cf. [Cao and Chen 2013; Catino et al. 2017; Chen and He 2013; Huang and Wei 2013]. Moreover, many of these manifolds admit a characterization as critical points of a generalization of the Yamabe functional [Case 2015a; 2015b]. In particular, the family of sharp Gagliardo–Nirenberg inequalities studied by Del Pino and Dolbeault [2002] can be understood via such functionals [Case 2013a; 2015b]. This family of sharp Gagliardo–Nirenberg inequalities has proven useful for studying certain fast diffusion equations [Carlen et al. 2010; Del Pino and Dolbeault 2002].

In this article we introduce the weighted σ_k-curvatures as appropriate generalizations of the σ_k-curvatures to smooth metric measure spaces. These curvatures form a family of scalar invariants on smooth metric measure spaces with properties which suitably generalize those algebraic and variational properties of the σ_k-curvatures which are important to the study of Einstein metrics. We expect that the weighted σ_k-curvatures will find further use in studying smooth metric measure spaces in general, and quasi-Einstein and weighted Einstein manifolds in particular. A more precise explanation of these results requires some notation and terminology.

A *smooth metric measure space* is a five-tuple

$$(M^n, g, v, m, \mu)$$

consisting of a Riemannian manifold (M^n, g), a positive function $v \in C^\infty(M; \mathbb{R}_+)$, a dimensional parameter $m \in \mathbb{R}$, and an auxiliary curvature parameter $\mu \in \mathbb{R}_+$. The metric g, the function v and the parameter m together determine the *weighted volume element* $d\nu := v^m \,\mathrm{dvol}_g$ on M. Roughly speaking, the dimensional parameter m indicates that we want to regard $(M^n, g, d\nu)$ as an $(m+n)$-dimensional metric measure space, in the sense that we consider the *m-Bakry–Émery Ricci tensor*

$$\mathrm{Ric}_\phi^m := \mathrm{Ric} + \nabla^2\phi - \frac{1}{m} d\phi \otimes d\phi$$

as the weighted analogue of the Ricci tensor, where $\phi = -m \ln v$. The auxiliary curvature parameter μ indicates that we want to regard $(M^n, g, d\nu)$ as the base of the warped product

$$(M^n \times F^m(\mu), g \oplus v^2 h), \tag{1-1}$$

where $(F^m(\mu), h)$ is the m-dimensional simply connected spaceform with constant sectional curvature μ. Most weighted invariants can be regarded as the restriction to M of Riemannian invariants on the warped product (1-1) when the latter makes sense. For example, $d\nu$ is the restriction of the Riemannian volume element of (1-1) and the *weighted scalar curvature*

$$R_\phi^m := R + 2\Delta\phi - \frac{m+1}{m}|\nabla\phi|^2 + m(m-1)\mu e^{\frac{2\phi}{m}}$$

is the scalar curvature of (1-1). The weighted σ_k-curvatures are defined in terms of the *weighted Schouten tensor*

$$P_\phi^m := \mathrm{Ric}_\phi^m - \frac{1}{m+n-2} J_\phi^m g,$$

where

$$J_\phi^m := \frac{m+n-2}{2(m+n-1)} R_\phi^m.$$

Note that all of the tensors just defined make sense in the limits $m = 0$ and $m = \infty$. The weighted σ_k-curvatures in the case $m = 0$ are the Riemannian σ_k-curvatures. The weighted σ_k-curvatures in the case $m = \infty$ have been previously considered by the author [Case 2016]. Since the results of this article are already known in the limiting cases $m = 0$ and $m = \infty$, we shall restrict our attention here to the cases $m \in \mathbb{R}_+$.

Informally, given nonnegative integers $k, n \in \mathbb{N}$ and a dimensional parameter $m \in \mathbb{R}_+$, the *m-weighted k-th elementary symmetric polynomial of n-variables* is the k-th elementary symmetric polynomial of $(m+n)$-variables; i.e.,

$$\sigma_k^m(\lambda; \lambda_1, \dots, \lambda_n) := \sigma_k\Big(\underbrace{\frac{\lambda}{m}, \dots, \frac{\lambda}{m}}_{m \text{ times}}, \lambda_1, \dots, \lambda_n\Big). \tag{1-2}$$

This can made precise by evaluating the right-hand side of (1-2) when $m \in \mathbb{N}_0$ and then extending the definition by treating $m \in \mathbb{R}_+$ as a formal variable; cf. Section 2. We extend this definition to symmetric matrices by considering the eigenvalues of the matrix; i.e., if $\lambda \in \mathbb{R}$ and if P is a symmetric n-by-n matrix, then we define

$$\sigma_k^m(\lambda; P) := \sigma_k^m(\lambda; \lambda_1, \dots, \lambda_n),$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of P.

Given $k \in \mathbb{N}$ and a smooth metric measure space (M^n, g, v, m, μ), the *weighted σ_k-curvature* is

$$\sigma^m_{k,\phi} := \sigma^m_k(Y^m_\phi; P^m_\phi),$$

where $Y^m_\phi := J^m_\phi - \mathrm{tr}_g P^m_\phi$. The most relevant cases are $k = 1, 2$, where we directly compute

$$\sigma^m_{1,\phi} = J^m_\phi, \qquad \sigma^m_{2,\phi} = \tfrac{1}{2}\Big((J^m_\phi)^2 - |P^m_\phi|^2 - \frac{1}{m}(Y^m_\phi)^2\Big).$$

Given $\kappa \in \mathbb{R}$ and a smooth metric measure space (M^n, g, v, m, μ), the *weighted σ_k-curvature with scale κ* is

$$\tilde{\sigma}^m_{k,\phi} := \sigma^m_k(Y^m_\phi + m\kappa v^{-1}; P^m_\phi).$$

With the goal of keeping our notation simple, we do not explicitly incorporate κ into our notation for the weighted σ_k-curvatures with scale κ. Instead, we always use tildes to denote quantities defined in terms of a possibly nonzero scale κ, and omit the tilde when we fix $\kappa = 0$; i.e., $\sigma^m_{k,\phi}$ is the weighted σ_k-curvature with scale $\kappa = 0$. The role of the scale κ, especially when the auxiliary curvature parameter μ vanishes, is to specify the "size" of the function v. Note that in the special cases $k = 1, 2$,

$$\begin{aligned}
\tilde{\sigma}^m_{1,\phi} &= \sigma^m_{1,\phi} + m\kappa v^{-1}, \\
\tilde{\sigma}^m_{2,\phi} &= \sigma^m_{2,\phi} + \frac{m(m+n-2)}{m+n-3}\kappa v^{-1}\sigma^{m-1}_{1,\phi} + \frac{m(m-1)}{2}\kappa^2 v^{-2},
\end{aligned}$$

where $\sigma^{m-1}_{1,\phi}$ is the weighted σ_1-curvature of $(M^n, g, v, m-1, \mu)$; see Lemma 4.4.

Our study of the weighted σ_k-curvatures is focused on their variational properties within a weighted conformal class. A *weighted manifold* is a triple (M^n, m, μ) of a smooth manifold M^n, a dimensional parameter $m \in \mathbb{R}_+$, and an auxiliary curvature parameter $\mu \in \mathbb{R}$. A *weighted conformal class* $\mathfrak{C}$ on a weighted manifold (M^n, m, μ) is an equivalence class with respect to the equivalence relation

$$(g, v) \sim (\hat{g}, \hat{v}) \text{ if and only if } (\hat{g}, \hat{v}) = (u^{-2}g, u^{-1}v) \text{ for some } u \in C^\infty(M; \mathbb{R}_+).$$

The weighted conformal class determined by (M^n, g, v, m, μ) is equivalent to the subset of the conformal class of the formal warped product (1-1) determined by restricting attention to conformal factors which depend only on the base M. We say that $\mathfrak{C}$ is *locally conformally flat in the weighted sense* if the formal warped product (1-1) is locally conformally flat; see Section 3 for an intrinsic characterization of this condition. We say that the weighted σ_k-curvature $\tilde{\sigma}^m_{k,\phi}$ is *variational in* $\mathfrak{C}$ if there is a functional $\mathcal{S} : \mathfrak{C} \to \mathbb{R}$ such that

$$\frac{d}{dt}\bigg|_{t=0} \mathcal{S}\Big(e^{-\frac{2t\psi}{m+n}}g, e^{-\frac{t\psi}{m+n}}v\Big) = \int_M \tilde{\sigma}^m_{k,\phi}\psi \, d\nu$$

for all $\psi \in C^\infty(M)$ and all $(g, v) \in \mathfrak{C}$. Our first main result is a characterization of when the weighted σ_k-curvatures are variational (cf. [Branson and Gover 2008]).

Theorem 1.1. *Fix $k \in \mathbb{N}$ and $\kappa \in \mathbb{R}$. Let $\mathfrak{C}$ be a weighted conformal class on a weighted manifold (M^n, m, μ). Then $\tilde{\sigma}^m_{k,\phi}$ is conformally variational in $\mathfrak{C}$ if and only if $k \leq 2$ or $\mathfrak{C}$ is locally conformally flat in the weighted sense.*

We prove Theorem 1.1 by showing that, as a function of $\mathfrak{C}$, the linearization of $\tilde{\sigma}^m_{k,\phi}$ is formally self-adjoint for every representative of $\mathfrak{C}$; for details, see Sections 4 and 5.

Our other main results concern the variational properties of the weighted σ_k-curvatures on weighted Einstein manifolds in the cases when $\tilde{\sigma}^m_{k,\phi}$ is variational. A *weighted Einstein manifold* is a smooth metric measure space (M^n, g, v, m, μ) for which there is a constant $\lambda \in \mathbb{R}$ such that $P^m_\phi = \lambda g$. For such manifolds, there is a scale $\kappa \in \mathbb{R}$ such that $\tilde{\sigma}^m_{1,\phi} = (m+n)\lambda$; see [Case 2015b, Lemma 9.1] or Lemma 3.8. We highlight two special classes of weighted Einstein manifolds. First, weighted Einstein manifolds with $\kappa = 0$ are equivalent to quasi-Einstein manifolds [Case et al. 2011]. Second, weighted Einstein manifolds for which the auxiliary curvature parameter μ vanishes are precisely the critical points of the weighted Yamabe functional [Case 2015b] through variations of the metric and the measure; cf. Theorem 8.11. Much more is known about quasi-Einstein manifolds than weighted Einstein manifolds, and for this reason we can prove more in the former setting.

Weighted Einstein manifolds can be understood in terms of the total weighted σ_k-curvature functionals. Our best such results are for quasi-Einstein manifolds, and are most naturally stated in terms of the set

$$\mathfrak{C}_1 := \left\{ (g, v) \in \mathfrak{C} \;\middle|\; \int_M dv = 1 \right\}$$

of representatives of $\mathfrak{C}$ with respect to which M has unit weighted volume.

Theorem 1.2. *Fix $k \in \mathbb{N}$. Let $\mathfrak{C}$ be a weighted conformal class on a closed weighted manifold (M^n, m, μ); if $k \geq 3$, assume additionally that $\mathfrak{C}$ is locally conformally flat in the weighted sense. Define the $\mathcal{F}_k$-functional $\mathcal{F}_k : \mathfrak{C}_1 \to \mathbb{R}$ by*

$$\mathcal{F}_k(g, v) := \int_M \sigma^m_{k,\phi}\, dv.$$

Suppose that $(g, v) \in \mathfrak{C}_1$ is such that $P^m_\phi = \lambda g$ and $\sigma^m_{1,\phi} = (m+n)\lambda$ for some constant $\lambda > 0$. Then (g, v) is a critical point of $\mathcal{F}_k$, and moreover

(1) *if $k < \frac{m+n}{2}$, then*

$$\left.\frac{d^2}{dt^2}\right|_{t=0} \mathcal{F}_k(\gamma(t)) > 0$$

for all $\gamma : \mathbb{R} \to \mathfrak{C}_1$ such that $\gamma(0) = (g, v)$ and $\gamma'(0) \neq 0$;

(2) *if* $\frac{m+n}{2} < k \leq m+n$, *then*

$$\left.\frac{d^2}{dt^2}\right|_{t=0} \mathcal{F}_k(\gamma(t)) < 0$$

for all $\gamma : \mathbb{R} \to \mathfrak{C}_1$ *such that* $\gamma(0) = (g, v)$ *and* $\gamma'(0) \neq 0$.

In the case $k = \frac{m+n}{2}$, the functional $\mathcal{F}_k$ is constant. In the case $k > m+n$, the functional $\mathcal{F}_k$ is constant (and identically zero) if and only if $m \in \mathbb{N}$. When $k > m+n$ and m is not an integer, the sign of the second variation of $\mathcal{F}_k$ depends on the parity of the integer part of $m+n-k$; cf. (6-1).

The proof of Theorem 1.2 depends on two ingredients. First, one computes the first and second variations of the $\mathcal{F}_k$-functional. In particular, when $k \leq 2$ or $\mathfrak{C}$ is locally conformally flat in the weighted sense, $(g, v) \in \mathfrak{C}_1$ is a critical point of the $\mathcal{F}_k$-functional if and only if

$$\sigma^m_{k,\phi} = c \tag{1-3}$$

for some constant $c \in \mathbb{R}$. Second, one uses the weighted Lichnerowicz–Obata theorem [Bakry and Qian 2000] to conclude that under the assumptions of Theorem 1.2, the first eigenvalue λ_1 of the *weighted Laplacian* $-\Delta_\phi := -\Delta + \nabla\phi$ satisfies $\lambda_1 > \frac{2(m+n)}{m+n-2}\lambda$. Applying this to the second variation of the $\mathcal{F}_k$-functional yields the result.

For weighted Einstein manifolds with $\mu = 0$ and positive scale, one should instead consider the $\mathcal{Y}_k$-functional $\mathcal{Y}_k : \mathfrak{C} \times \mathbb{R}_+ \to \mathbb{R}$ defined by

$$\begin{aligned}
\mathcal{Y}_k(g, v, \kappa) &:= \mathcal{Z}_k(g, v, \kappa)\left(\int v^{-1}\, d\nu\right)^{-\frac{2mk}{(m+n)(2m+n-2)}}\left(\int d\nu\right)^{-\frac{m+n-2k}{m+n}}, \\
\mathcal{Z}_k(g, v, \kappa) &:= \kappa^{-\frac{2mk(m+n-1)}{(m+n)(2m+n-2)}} \int_M \tilde{\sigma}^m_{k,\phi}\, d\nu.
\end{aligned}$$

If $(M^n, g, v, m, 0)$ is a weighted Einstein manifold with scale $\kappa > 0$, then (g, v) is a critical point of the $\mathcal{Y}_k$-functional whenever $\tilde{\sigma}^m_{k,\phi}$ is variational. Indeed, when $\tilde{\sigma}^m_{k,\phi}$ is variational, a triple $(g, v, \kappa) \in \mathfrak{C} \times \mathbb{R}_+$ is a critical point of the $\mathcal{Y}_k$-functional if and only if

$$\tilde{\sigma}^m_{k,\phi} + \frac{m}{m+n-2k}\kappa v^{-1}\tilde{s}^m_{k-1,\phi} = \frac{\int \tilde{\sigma}^m_{k,\phi}\, d\nu}{\int d\nu} + \frac{m}{m+n-2k}\left(\frac{\int \tilde{s}^m_{k-1,\phi} v^{-1}\, d\nu}{\int v^{-1}\, d\nu}\right)\kappa v^{-1} \tag{1-4a}$$

and

$$\int_M \tilde{\sigma}^m_{k,\phi}\, d\nu = \frac{(m+n)(2m+n-2)}{2k(m+n-1)} \int_M \kappa v^{-1} \tilde{s}^m_{k-1,\phi}\, d\nu, \tag{1-4b}$$

where $\tilde{s}^m_{k-1,\phi}$ is defined by Definition 2.5 and (3-3) below. Note that $\tilde{\sigma}^m_{k,\phi}$ and $\tilde{s}^m_{k-1,\phi}$

are both constant for weighted Einstein metrics. However, such weighted Einstein manifolds need not have positive m-Bakry–Émery Ricci curvature (cf. Example 3.13). In particular, a new weighted Lichnerowicz–Obata theorem seems to be needed in order to prove that weighted Einstein manifolds with $\mu = 0$ and positive scale are local extrema of the $\mathcal{Y}_k$-functionals whenever $\tilde{\sigma}^m_{k,\phi}$ is variational; for further discussion, including a conjectural form of the required weighted Lichnerowicz–Obata theorem, see Section 6B.

In light of Theorem 1.2, one might hope that, up to scaling, quasi-Einstein manifolds are the only critical points of the $\mathcal{F}_k$-functionals which lie in the weighted elliptic k-cones. For quasi-Einstein manifolds which are locally conformally flat in the weighted sense, this is true.

Theorem 1.3. *Let (M^n, g, v, m, μ) be a closed smooth metric measure space which satisfies $P^m_\phi = \lambda g$ and $\sigma^m_{1,\phi} = (m+n)\lambda$ for some $\lambda \in \mathbb{R}$ and which is locally conformally flat in the weighted sense. Fix $k \in \mathbb{N}$, let $\mathfrak{C}$ be the weighted conformal class containing (g, v), and suppose that $(\hat{g}, \hat{v}) \in \mathfrak{C}$ is a critical point of the $\mathcal{F}_k$-functional such that $\sigma^m_{j,\phi}(\hat{g}, \hat{v}) > 0$ for all $1 \le j \le k$. Then $(\hat{g}, \hat{v}) = (c^2 g, cv)$ for some constant $c \in \mathbb{R}_+$.*

The proof of Theorem 1.3 is analogous to Obata's proof of the classification of conformally Einstein metrics with constant scalar curvature on closed manifolds [Obata 1971] and its generalization by Viaclovsky [2000] to the σ_k-curvatures; for details, see Section 7A. The condition $\sigma^m_{j,\phi}(\hat{g}, \hat{v}) > 0,\ 1 \le j \le k$, implies that the PDE (1-3) is elliptic at $(\hat{g}, \hat{v})$. By developing some additional integral estimates, we expect Theorem 1.3 to also hold for the round hemisphere $(S^n_+, d\theta^2, 1, m, 1)$; cf. [Chang et al. 2003; González 2006]. Indeed, we expect further study of the PDE $\sigma^m_{k,\phi} = f$ to lead to a sharp fully nonlinear Sobolev inequality in this setting; see Conjecture 7.8.

Many of the ideas in the proof of Theorem 1.3 can be used to study the analogous classification for weighted Einstein manifolds. More precisely, the proof of Obata's theorem [1971] begins by using the variational structure of the σ_k-curvatures to find a $(0, 2)$-tensor which is divergence-free for any metric g with σ_k-constant and by using the assumption that g is conformally Einstein to show that the trace-free part of the Schouten tensor is in the image of the adjoint of the divergence operator on trace-free $(0, 2)$-tensors. The proof ends by using the Newton inequalities and the specific form of the divergence-free $(0, 2)$-tensor to conclude that g is Einstein. The first step carries through for solutions of (1-3) (resp. (1-4)) which are conformally quasi-Einstein (resp. conformally weighted Einstein manifolds with $\mu = 0$), though with a more complicated stand-in for the divergence operator and its adjoint; see Section 7. At present, while Theorem 2.13 asserts the weighted Newton inequalities, we can only carry out the second step in the setting of quasi-Einstein manifolds. Indeed, carrying out the second step for weighted Einstein manifolds is even problematic in the case $k = 1$ (cf. [Case 2015b, Conjecture 1.5]). Nevertheless, we expect that the analogue of Theorem 1.3 for weighted Einstein manifolds is true. Inspired by the sharp Gagliardo–Nirenberg inequalities of Del Pino and Dolbeault [2002], we expect further study of the $\mathcal{Y}_k$-functionals to lead to sharp fully nonlinear Gagliardo–Nirenberg inequalities; see Conjecture 7.13.

The fact that quasi-Einstein manifolds and weighted Einstein manifolds are critical points of the $\mathcal{F}_k$- and $\mathcal{Y}_k$-functionals, respectively, within a weighted conformal class indicates that these functionals should be useful in studying such manifolds in a general variational context. Our final main result verifies this expectation, at least in the cases $k = 1, 2$. To be more precise, let

$$\mathfrak{M}(M^n, m, \mu) := \{(g, v) \mid g \in \mathrm{Met}(M),\, v \in C^\infty(M),\, v > 0\}$$

denote the space of *metric-measure structures* on (M^n, m, μ) and denote

$$\mathfrak{M}_1(M^n, m, \mu) := \left\{ (g, v) \in \mathfrak{M} \;\middle|\; \int_M v^m \,\mathrm{dvol} = 1 \right\}.$$

It is clear that we can extend the $\mathcal{F}_k$- and $\mathcal{Y}_k$-functionals to functionals on $\mathfrak{M}_1$ and $\mathfrak{M} \times \mathbb{R}_+$, respectively. Weighted Einstein manifolds are related to the critical points of these functionals in the following way:

Theorem 1.4. *Let (M^n, m, μ) be a weighted manifold.*

1. *$(g, v) \in \mathfrak{M}_1$ is a critical point of $\mathcal{F}_1 : \mathfrak{M}_1 \to \mathbb{R}$ if and only if $P^m_\phi = \lambda g$ and $\sigma^m_{1,\phi} = (m + n)\lambda$ for some $\lambda \in \mathbb{R}$.*
2. *If $(g, v) \in \mathfrak{M}_1$ satisfies $P^m_\phi = \lambda g$ and $\sigma^m_{1,\phi} = (m + n)\lambda$ for some $\lambda \in \mathbb{R}$, then (g, v) is a critical point of $\mathcal{F}_2 : \mathfrak{M}_1 \to \mathbb{R}$.*

Suppose additionally that $\mu = 0$.

3. *$(g, v, \kappa) \in \mathfrak{M} \times \mathbb{R}_+$ is a critical point of $\mathcal{Y}_1 : \mathfrak{M} \times \mathbb{R}_+ \to \mathbb{R}$ if and only if (g, v) is a weighted Einstein metric-measure structure with scale κ.*
4. *If $(g, v) \in \mathfrak{M}$ is a weighted Einstein manifold with scale $\kappa > 0$, then (g, v, κ) is a critical point of $\mathcal{Y}_2 : \mathfrak{M} \times \mathbb{R}_+ \to \mathbb{R}$.*

The final claim of Theorem 1.4 is the most noteworthy, as it explains the seemingly complicated definition of the $\mathcal{Y}_k$-functional and its Euler equation (1-4). More precisely, while Theorem 1.1 guarantees that there is a functional on $\mathfrak{C}$ for which its critical points are precisely those metric-measure structures with $\tilde{\sigma}^m_{2,\phi}$ constant, a computation involving the behavior of the weighted Bach tensor of a weighted Einstein manifold with $\mu = 0$ implies that such manifolds are only critical points of the $\mathcal{Y}_2$-functional; cf. Remark 8.13.

This article is organized as follows. In Section 2 we establish the key algebraic properties of the m-weighted elementary symmetric polynomials. In Section 3 we discuss the necessary background for smooth metric measure spaces, including some important facts about weighted Einstein manifolds. In Section 4 we set up a useful formalism for studying the space of metric-measure structures. In Section 5 we prove Theorem 1.1. In Section 6 we prove Theorem 1.2 and discuss its analogue for weighted Einstein manifolds. In Section 7 we prove the ellipticity of (1-3) and (1-4) within the appropriate elliptic cones, prove Theorem 1.3, and discuss our related conjectures for weighted Einstein manifolds and sharp fully nonlinear Sobolev inequalities. In Section 8 we prove Theorem 1.4.

2. Algebraic preliminaries

We begin our study with a discussion of the algebraic properties of the m-weighted k-th elementary symmetric polynomials of n-variables. Importantly, provided m is sufficiently large relative to k (cf. Theorem 2.13 below), these invariants possess all of the properties expected from the informal definition (1-2). This section makes this assertion precise. We begin with the formal definition of the weighted elementary symmetric polynomials.

Definition 2.1. Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. The *m-weighted k-th elementary symmetric polynomial σ_k^m of n variables* is the function $\sigma_k^m : \mathbb{R} \times \mathbb{R}^n$ defined recursively by

$$\sigma_0^m(\lambda; \Lambda) = 1, \quad \text{if } k = 0,$$

$$\sigma_k^m(\lambda; \Lambda) = \frac{1}{k}\sum_{j=0}^{k-1}(-1)^j \sigma_{k-1-j}^m(\lambda; \Lambda) N_{j+1}^m(\lambda; \Lambda), \quad \text{if } k \geq 1,$$

where $\Lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ and $N_k^m : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ is defined by

$$N_k^m(\lambda; \Lambda) = m\left(\frac{\lambda}{m}\right)^k + \sum_{j=1}^n \lambda_j^k.$$

One may also regard the weighted elementary symmetric polynomials as perturbations of the elementary symmetric polynomials through lower order terms. In other words, $\sigma_k^m(\lambda; \Lambda)$ and $\sigma_k(\Lambda)$ differ by an inhomogeneous polynomial in Λ of degree $k-1$. The precise relationship is as follows.

Lemma 2.2. *Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. Then*

$$\sigma_k^m(\lambda; \Lambda) = \sum_{j=0}^k \binom{m}{j}\left(\frac{\lambda}{m}\right)^j \sigma_{k-j}(\Lambda) \tag{2-1}$$

for all $(\lambda; \Lambda) \in \mathbb{R} \times \mathbb{R}^n$.

Proof. The proof is by induction. Clearly (2-1) holds if $k=0$. Suppose that (2-1) holds for some $k\in\mathbb{N}_0$. The definition of the weighted elementary symmetric polynomials and the inductive hypothesis imply that

$$(k+1)\sigma_{k+1}^m(\lambda;\Lambda)=\sum_{j=0}^{k}\sum_{\ell=0}^{k-j}(-1)^\ell\binom{m}{j}\left(\frac{\lambda}{m}\right)^j\sigma_{k-j-\ell}(\Lambda)N_{\ell+1}^m(\lambda;\Lambda). \tag{2-2}$$

Using the identity $\sum_{j=0}^k(-1)^{k-j}\binom{m}{j}=\binom{m-1}{k}$, we compute that

$$\sum_{j=0}^{k}\sum_{\ell=0}^{k-j}(-1)^\ell\binom{m}{j}\left(\frac{\lambda}{m}\right)^{j+\ell+1}\sigma_{k-j-\ell}(\Lambda)=\sum_{j=0}^{k}\binom{m-1}{j}\left(\frac{\lambda}{m}\right)^{j+1}\sigma_{k-j}(\Lambda). \tag{2-3}$$

Using the recursive definition of the elementary symmetric polynomials, we compute

$$\begin{aligned}\sum_{j=0}^{k}\sum_{\ell=0}^{k-j}(-1)^\ell\binom{m}{j}\left(\frac{\lambda}{m}\right)^{j}\sigma_{k-j-\ell}(\Lambda)\sum_{s=1}^n\lambda_s^{\ell+1}&\\ =\sum_{j=0}^{k}(k+1-j)\binom{m}{j}\left(\frac{\lambda}{m}\right)^{j}\sigma_{k+1-j}(\Lambda).&\end{aligned} \tag{2-4}$$

Combining (2-2), (2-3) and (2-4) yields the desired result. □

There is a similar relationship between weighted elementary symmetric polynomials when the value λ is changed.

Corollary 2.3. *Fix $m\in\mathbb{R}_+$ and $k,n\in\mathbb{N}_0$. Then*

$$\sigma_k^m(\lambda_1+\lambda_2;\Lambda)=\sum_{j=0}^{k}\binom{m}{j}\left(\frac{\lambda_1}{m}\right)^j\sigma_{k-j}^{m-j}\left(\frac{m-j}{m}\lambda_2;\Lambda\right)$$

for all $\mu_1,\lambda_2\in\mathbb{R}$ and $\Lambda\in\mathbb{R}^n$.

Proof. Lemma 2.2 and the binomial theorem yield

$$\sigma_k^m(\lambda_1+\lambda_2;\Lambda)=\sum_{j=0}^{k}\sum_{s=0}^{j}\binom{m}{j}\binom{j}{s}\left(\frac{\lambda_1}{m}\right)^{j-s}\left(\frac{\lambda_2}{m}\right)^{s}\sigma_{k-j}(\Lambda).$$

Combining this with (2-1) and the identity $\binom{m}{j}\binom{j}{s}=\binom{m}{j-s}\binom{m-j+s}{s}$ yields the desired result. □

A useful fact is that the relationship between an elementary symmetric polynomial of $\Lambda=(\lambda_1,\dots,\lambda_n)$ and the corresponding elementary symmetric polynomial of $\Lambda(i):=(\lambda_1,\dots,\lambda_{i-1},\lambda_{i+1},\dots,\lambda_n)$ persists to the weighted case.

Lemma 2.4. *Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. Then*

$$\sigma_k^m(\lambda; \Lambda) = \sigma_k^m(\lambda; \Lambda(i)) + \lambda_i \sigma_{k-1}^m(\lambda; \Lambda(i)). \tag{2-5}$$

for all $\lambda \in \mathbb{R}$, $\Lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$, and $1 \le i \le n$.

Proof. It is readily verified that (2-5) holds when $m = 0$. Lemma 2.2 then implies that (2-5) holds in general. □

Our interest is in considering the weighted elementary symmetric polynomials of a self-adjoint endomorphism of $\mathbb{R}^n$. These are defined in terms of the eigenvalues of the endomorphism. In this context, there are also natural weighted analogues of the Newton transforms.

Definition 2.5. Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. Let $\mathcal{M}_n$ denote the space of self-adjoint endomorphisms of $\mathbb{R}^n$ with its standard inner product.

(1) The *m-weighted k-th elementary symmetric function* $\sigma_k^m : \mathbb{R} \times \mathcal{M}_n \to \mathbb{R}$ is defined by $\sigma_k^m(\lambda; P) := \sigma_k^m(\lambda; \Lambda(P))$, where $\Lambda(P) := (\lambda_1, \dots, \lambda_n)$ is the list of the eigenvalues (with multiplicity) of P.

(2) The *m-weighted k-th Newton transform* $T_k^m : \mathbb{R} \times \mathcal{M}_n \to \mathcal{M}_n$ is defined by

$$T_k^m(\lambda; P) = \sum_{j=0}^{k} (-1)^j \sigma_{k-j}^m(\lambda; P)\, P^j.$$

(3) The *m-weighted k-th Newton scalar* $s_k^m : \mathbb{R} \times \mathcal{M}_n \to \mathbb{R}$ is defined by

$$s_k^m(\lambda; P) = \sum_{j=0}^{k} (-1)^j \left(\frac{\lambda}{m}\right)^j \sigma_{k-j}^m(\lambda; P).$$

Informally, the m-weighted k-th elementary symmetric function $\sigma_k^m(\lambda; P)$ is the k-th elementary symmetric function of the $(m+n) \times (m+n)$ block-diagonal matrix

$$P \oplus \frac{\lambda}{m} \operatorname{Id}_m, \tag{2-6}$$

where Id_m is the $m \times m$ identity matrix. The k-Newton transform T_k of (2-6) decomposes as

$$T_k = T_k^m \oplus s_k^m \operatorname{Id}_m .$$

The eigenvalues of the weighted Newton transforms are readily computed in terms of the weighted elementary symmetric functions.

Lemma 2.6. *Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. Given $\lambda \in \mathbb{R}$ and $P \in \mathcal{M}_n$, the eigenvalues of $T_k^m(\lambda; P)$ are $\sigma_k^m(\lambda; \Lambda(i))$ for $1 \le i \le n$, where $\Lambda = \Lambda(P)$ are the eigenvalues of P and $\Lambda(i)$ is as in Lemma 2.4.*

Proof. The proof is by induction. It is clear that the conclusion holds when $k=0$. Suppose that the eigenvalues of $T_k^m(\lambda;\, P)$ are $\sigma_k^m(\lambda;\, \Lambda(i))$ for $1 \le i \le n$. Note that

$$T_{k+1}^m(\lambda;\, P) = \sigma_{k+1}^m(\lambda;\, P)I - PT_k^m(\lambda;\, P) \tag{2-7}$$

for I the identity endomorphism. The conclusion follows readily from Lemma 2.4 and the inductive hypothesis. □

Remark 2.7. It follows immediately from Definition 2.5 that the analogue

$$s_{k+1}^m(\lambda;\, P) = \sigma_{k+1}^m(\lambda;\, P) - \frac{\lambda}{m} s_k^m(\lambda;\, P) \tag{2-8}$$

of (2-7) holds.

The values of the weighted Newton scalars are also readily computed from the weighted elementary symmetric functions.

Lemma 2.8. *Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. Then*

$$s_k^m(\lambda;\, P) = \sigma_k^{m-1}\left(\frac{m-1}{m}\lambda;\, P\right) \tag{2-9}$$

for all $\lambda \in \mathbb{R}$ and $P \in \mathcal{M}_n$.

Proof. The proof is by induction. It is clear that (2-9) holds when $k=0$. Suppose now that (2-9) holds for all $j \in \{0, \dots, k-1\}$. We thus compute that

$$\begin{aligned}
k\sigma_k^{m-1}\left(\frac{m-1}{m}\lambda;\, P\right) &= \sum_{j=0}^{k-1} (-1)^j s_{k-1-j}^m(\lambda;\, P)\left(N_{j+1}^m(\lambda;\, P) - \left(\frac{\lambda}{m}\right)^{j+1}\right) \\
&= \sum_{j=0}^{k-1}\sum_{\ell=0}^{k-1-j} (-1)^{j+\ell}\left(\frac{\lambda}{m}\right)^{\ell} \sigma_{k-1-j-\ell}^m(\lambda;\, P) N_{j+1}^m(\lambda;\, P) \\
&\qquad - \sum_{j=0}^{k-1}\sum_{\ell=0}^{k-1-j} (-1)^{j+\ell}\left(\frac{\lambda}{m}\right)^{j+\ell+1} \sigma_{k-1-j-\ell}^m(\lambda;\, P).
\end{aligned}$$

Switching the order of the first summation and computing the second summation by summing over ℓ and then $j+\ell$ yields the desired result. □

Lemma 2.8 yields another useful interpretation of the m-weighted Newton scalars.

Lemma 2.9. *Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. Let $\lambda \in \mathbb{R}$ and $P \in \mathcal{M}_n$. Consider the function $S_k^m : \mathbb{R} \to \mathbb{R}$ defined by $S_k^m(\kappa) := \sigma_k^m(\lambda + m\kappa;\, P)$. Then*

$$\frac{dS_k^m}{d\kappa} = m s_{k-1}^m(\lambda + m\kappa;\, P).$$

Proof. Expanding $S_k^m(\kappa)$ as a power series in κ via Corollary 2.3 and differentiating yields

$$\frac{dS_k^m}{d\kappa} = m \sum_{j=0}^{\infty} \binom{m-1}{j} \kappa^j \sigma_{k-1-j}^{m-1-j}\left(\frac{m-1-j}{m}\lambda;\, P\right).$$

Corollary 2.3 then implies that

$$\frac{dS_k^m}{d\kappa} = m\sigma_{k-1}^{m-1}\left(\frac{m-1}{m}(\lambda + m\kappa);\, P\right).$$

The final conclusion now follows from Lemma 2.8. □

One difference between the weighted and unweighted case is that the weighted elementary symmetric functions are not recovered by taking inner products between the endomorphism and the associated weighted Newton transforms. A specific relationship is as follows.

Lemma 2.10. *Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. Then*

$$\left\langle T_k^m(\lambda;\, P),\, P - \frac{\lambda}{m} I \right\rangle = (k+1)\sigma_{k+1}^m(\lambda;\, P) - (m+n-k)\frac{\lambda}{m}\sigma_k^m(\lambda;\, P)$$

for all $\lambda \in \mathbb{R}$ and $P \in \mathcal{M}_n$, where $I \in \mathcal{M}_n$ is the identity and $\langle A, B\rangle := \operatorname{tr}(AB)$ is the standard inner product on $\mathcal{M}_n$.

Proof. This follows immediately from the definitions of σ_k^m and T_k^m and the observation that

$$\left\langle P^j,\, P - \frac{\lambda}{m} I \right\rangle = N_{j+1}^m(\lambda;\, \Lambda(P)) - \frac{\lambda}{m} N_j^m(\lambda;\, \Lambda(P))$$

for all $j \in \mathbb{N}_0$. □

Nevertheless, there is a simple and useful relationship between inner products involving weighted Newton transforms and weighted elementary symmetric polynomials.

Corollary 2.11. *Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. Define $E_k^m : \mathbb{R} \times \mathcal{M}_n \to \mathcal{M}_n$ by*

$$E_k^m := T_k^m - \frac{m+n-k}{m+n}\sigma_k^m\, I \tag{2-10}$$

for $I \in \mathcal{M}_n$ the identity. Then

$$\left\langle E_k^m(\lambda;\, P),\, P - \frac{\lambda}{m} I \right\rangle = (k+1)\sigma_{k+1}^m(\lambda;\, P) - \frac{m+n-k}{m+n}\sigma_1^m(\lambda;\, P)\sigma_k^m(\lambda;\, P) \tag{2-11}$$

for all $\lambda \in \mathbb{R}$ and $P \in \mathcal{M}_n$.

The significance of Corollary 2.11 is contained in Corollary 2.17 below, which concludes that (2-11) has a sign under a natural condition on λ and P. The intuition motivating the consideration of (2-11) is as follows: one can regard (2-10) as defining the trace-free part of the weighted Newton transform and the left-hand side of (2-11) as the inner product of E_k^m with P. On the one hand, in the unweighted setting, the inner product of the trace-free part of a Newton transform with the underlying endomorphism is well-known to have a sign when the endomorphism lies in one of the Gårding cones (see [Viaclovsky 2000, Lemma 23]). On the other hand, if $P \in \operatorname{End}(\mathbb{R}^n)$, $\lambda \in \mathbb{R}$, and $m \in \mathbb{N}$, then the trace-free part of the k-th Newton transform of (2-6) is

$$E_k := E_k^m \oplus \left(-\frac{1}{m} \operatorname{tr} E_k^m\right) I_m.$$

It follows that

$$\left\langle E_k,\, P \oplus \frac{\lambda}{m} I_m \right\rangle = \left\langle E_k^m,\, P - \frac{\lambda}{m} I_n \right\rangle.$$

2A. ***The weighted Newton inequalities.*** In this subsection we show that the m-weighted elementary symmetric polynomials of n-variables satisfy the same Newton inequalities as the elementary symmetric polynomials of $(m+n)$-variables. Our proof of this fact is similar to the usual proof of the Newton inequalities (cf. [Hardy et al. 1934]). To that end, we use the following generating function for the weighted elementary symmetric polynomials.

Proposition 2.12. *Fix $m \in \mathbb{R}_+$, $k, n \in \mathbb{N}_0$, $\lambda \in \mathbb{R}$, and $\Lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$. Then*

$$\left(1 + \frac{\lambda t}{m}\right)^m \prod_{i=1}^n (1 + \lambda_i t) = \sum_{j=0}^\infty \sigma_j^m(\lambda;\, \Lambda) t^j$$

for all $t \in (-M, M)$, where

$$M = \begin{cases} \infty, & \text{if } \lambda = 0 \text{ or } m \in \mathbb{N}_0, \\ m/|\lambda|, & \text{otherwise}. \end{cases}$$

Proof. It is well known that

$$\prod_{i=1}^n (1 + \lambda_i t) = \sum_{j=0}^n \sigma_j(\Lambda) t^j, \qquad \left(1 + \frac{\lambda t}{m}\right)^m = \sum_{j=0}^\infty \binom{m}{j} \left(\frac{\lambda}{m}\right)^j t^j$$

for all $t \in (-M, M)$. Multiplying these expressions yields

$$\left(1 + \frac{\lambda t}{m}\right)^m \prod_{i=1}^n (1 + \lambda_i t) = \sum_{k=0}^\infty \sum_{j=0}^k \binom{m}{j} \left(\frac{\lambda}{m}\right)^j \sigma_{k-j}(\Lambda) t^k.$$

The conclusion follows from Lemma 2.2. $\square$

We are now ready to prove the weighted Newton inequalities.

Theorem 2.13. *Let $k \in \mathbb{N}$ and let $m \in [k-1, \infty)$. Then*

$$\sigma_{k-1}^m(\lambda;\Lambda)\sigma_{k+1}^m(\lambda;\Lambda) \leq \frac{k(m+n-k)}{(k+1)(m+n-k+1)}(\sigma_k^m(\lambda;\Lambda))^2 \tag{2-12}$$

for all $\lambda \in \mathbb{R}$ and $\Lambda \in \mathbb{R}^n$. Moreover, equality holds if and only if one of the following holds:

(1) $\Lambda = (\lambda/m, \dots, \lambda/m)$.

(2) $\lambda = 0$ *and at most $k-1$ components of Λ are nonzero.*

(3) $m = k-1$ *and* $\Lambda = (0, \dots, 0)$.

Remark 2.14. If $m \notin \mathbb{N}_0$, the assumption $m \geq k-1$ is necessary. This can be seen by computing both sides of (2-12) with $\Lambda = 0$ and $\lambda \in \mathbb{R}$ arbitrary.

Proof. Set $p_k^m := (m+n)^k \binom{m+n}{k}^{-1} \sigma_k^m$, so that, as a functional inequality on $\mathbb{R} \times \mathbb{R}^n$, (2-12) is equivalent to

$$p_{k-1}^m p_{k+1}^m \leq (p_k^m)^2. \tag{2-13}$$

If $m \in \mathbb{N}_0$, the conclusion follows from the usual Newton inequality [Hardy et al. 1934]. Suppose now that $m \notin \mathbb{N}_0$. We separate the proof into three cases.

First, suppose that $\lambda = 0$. Lemma 2.2 and the (unweighted) Newton inequalities imply that

$$\sigma_{k-1}^m \sigma_{k+1}^m \leq \frac{k(n-k)}{(k+1)(n-k+1)}(\sigma_k^m)^2 \leq \frac{k(m+n-k)}{(k+1)(m+n-k+1)}(\sigma_k^m)^2$$

with equality if and only if at most $k-1$ components of Λ are nonzero.

Second, suppose $k = 1$. Write $\Lambda = (\lambda_1, \dots, \lambda_n)$. We compute that

$$\begin{aligned} p_2^m(\lambda;\Lambda) &= \frac{m+n}{m+n-1}\left((p_1^m(\lambda;\Lambda))^2 - \frac{\lambda^2}{m} - \sum_{s=1}^n \lambda_s^2\right) \\ &\leq \frac{m+n}{m+n-1}\left((p_1^m(\lambda;\Lambda))^2 - \frac{\lambda^2}{m} - \frac{1}{n}\left(\sum_{s=1}^n \lambda_s\right)^2\right) \\ &= (p_1^m(\lambda;\Lambda))^2 - \frac{m}{n(m+n-1)}\left(\sum_{s=1}^n \lambda_s - \frac{n\lambda}{m}\right)^2 \end{aligned}$$

with equality if and only if $\lambda_1 = \cdots = \lambda_n$. The conclusion readily follows.

Third, suppose $\lambda \neq 0$ and $k \geq 2$. Set

$$P(t) = \left(1 + \frac{\lambda t}{m}\right)^m \prod_{j=1}^n (1 + \lambda_j t).$$

By Proposition 2.12,

$$P(t)=\sum_{j=0}^{\infty}\binom{m+n}{j}\left(\frac{t}{m+n}\right)^{j}p_j^m. \tag{2-14}$$

Write $\Lambda=(\lambda_1,\dots,\lambda_n)$ and let $\ell=\bigl|\{j\mid\lambda_j=0\}\bigr|$ denote the number of components of Λ which vanish. Up to reindexing the components of Λ, we see that

$$P(t)=\left(1+\frac{\lambda t}{m}\right)^{m}Q_0(t)$$

for Q_0 a polynomial of degree $n-\ell$ with $Q_0(0)=1$ and roots $r_j=-1/\lambda_j$ such that $r_1^{(0)}\leq\cdots\leq r_{n-\ell}^{(0)}$ are all nonzero.

For any $s\in\{0,\dots,k-1\}$, we compute from (2-14) that

$$\frac{d^sP}{dt^s}=\frac{(m+n)!}{(m+n)^s(m+n-s)!}\sum_{j=0}^{\infty}\binom{m+n-s}{j}\left(\frac{t}{m+n}\right)^{j}p_{j+s}^m. \tag{2-15}$$

On the other hand, by regarding P as an analytic function in $\mathbb{C}\setminus\{-\frac{m}{\lambda}+is\mid s\leq 0\}$ and applying Rolle's theorem along the real rays $x\geq-\frac{m}{\lambda}$ and $x\leq-\frac{m}{\lambda}$, we deduce that

$$\frac{d^sP}{dt^s}=\left(1+\frac{\lambda t}{m}\right)^{m-s}Q_s(t) \tag{2-16}$$

for Q_s a polynomial of degree $n-\ell$ with roots $r_1^{(s)}\leq\cdots\leq r_{n-\ell}^{(s)}$, at most one of which is zero.

We now consider two subcases. Suppose first that there is a $j\in\{1,\dots,n-\ell\}$ such that $r_j^{(k-1)}=0$. Then $Q_{k-1}(t)=ct+O(t^2)$ near $t=0$ for some $c\neq 0$. Comparing (2-15) and (2-16) yields $p_{k-1}^m=0$ and $p_k^m\neq 0$, from which the desired conclusion readily follows.

Suppose instead that $r_j^{(k-1)}\neq 0$ for all $j\in\{1,\dots,n-\ell\}$. Define

$$\lambda_j^{(k-1)}=\begin{cases}-1/r_j^{(k-1)}, & \text{if } j\leq n-\ell,\\ 0, & \text{otherwise.}\end{cases}$$

and set $\Lambda^{(k-1)}:=(\lambda_1^{(k-1)},\dots,\lambda_n^{(k-1)})$. Proposition 2.12 and comparison of (2-15) and (2-16) imply that

$$p_j^{m-k+1}\left(\frac{m-k+1}{m}\lambda;\Lambda^{(k-1)}\right)=\left(\frac{m+n-k+1}{m+n}\right)^{j}\frac{p_{k+j-1}^m(\lambda;\Lambda)}{p_{k-1}^m(\lambda;\Lambda)}$$

for all $j\in\mathbb{N}_0$. In particular, applying this to the cases $j\in\{0,1,2\}$ yields (2-13) with equality if and only if $\lambda_j^{(k-1)}=\frac{\lambda}{m}$ for all $j\in\{1,\dots,n\}$. Our applications of Rolle's theorem imply that the latter conclusion is equivalent to $\lambda_j=\frac{\lambda}{m}$ for all $j\in\{1,\dots,n\}$, as desired. $\square$

2B. ***Weighted elliptic cones.*** An important feature of the elementary symmetric polynomials is that they are monotone with respect to a single variable within the Gårding cones; this is closely related to the ellipticity of the σ_k-curvature prescription problem (cf. [Caffarelli et al. 1985; Gårding 1959; Viaclovsky 2000]). There are similar cones in the weighted case; we state our definition only in terms of the space $\mathcal{M}_n$ of self-adjoint endomorphisms of $\mathbb{R}^n$, from which one could easily formulate analogous definitions in terms of the weighted elementary symmetric polynomials.

Definition 2.15. Fix $m \in \mathbb{R}_+$ and $k, n \in \mathbb{N}_0$. The *positive* (*resp. negative*) *m-weighted elliptic k-cone* is the set $\Gamma_k^{m,+}$ (resp. $\Gamma_k^{m,-}$) defined by

$$\Gamma_k^{m,\pm} := \left\{(\lambda; P) \in \mathbb{R} \times \mathcal{M}_n \mid (\pm 1)^j \sigma_j^m(\lambda; P) > 0 \text{ for all } 1 \le j \le k\right\}.$$

One useful fact is that (2-11) has a sign in the appropriate weighted elliptic cone.

Proposition 2.16. *Let $k, n \in \mathbb{N}_0$ and $m \in [k-1, \infty)$. Then*

$$(\pm 1)^{k+1} \sigma_{k+1}^m(\lambda; P) \le (\pm 1)^{k+1} \frac{m+n-k}{(m+n)(k+1)} \sigma_1^m(\lambda; P) \sigma_k^m(\lambda; P) \tag{2-17}$$

for all $(\lambda; P) \in \Gamma_k^{m,\pm}$. Moreover, equality holds if and only if $P = \frac{\lambda}{m} I$.

Proof. The proof is by induction. Observe that, as a functional inequality on $\Gamma_k^{m,\pm}$, (2-17) is equivalent to

$$(\pm 1)^{k+1} p_{k+1}^m \le (\pm 1)^{k+1} p_1^m p_k^m. \tag{2-18}$$

Theorem 2.13 implies that (2-18) holds when $k = 1$. Suppose now that (2-13) holds in $\Gamma_k^{m,\pm}$. Since $\Gamma_{k+1}^{m,\pm} \subset \Gamma_k^{m,\pm}$, it follows from Theorem 2.13 and the inductive hypothesis that

$$p_k^m p_{k+2}^m \le ((\pm 1)^{k+1} p_{k+1}^m)^2 \le p_1^m p_k^m p_{k+1}^m$$

in $\Gamma_{k+1}^{m,\pm}$. Dividing both sides by $(\pm 1)^k p_k^m > 0$ yields the desired result. □

Corollary 2.17. *Let $k, n \in \mathbb{N}_0$ and $m \in [k-1, \infty)$. Then*

$$(\pm 1)^{k+1} \left\langle E_k^m(\lambda; P), P - \frac{\lambda}{m} I \right\rangle \le 0$$

for all $(\lambda; P) \in \Gamma_k^{m,\pm}$. Moreover, equality holds if and only if $P = \frac{\lambda}{m} I$.

Proof. This follows immediately from Corollary 2.11 and Proposition 2.16. □

Another useful fact is that the weighted Newton transform and the weighted Newton scalar have a sign in the corresponding weighted elliptic cone. This result encodes the relationship between the weighted elliptic cones and ellipticity of the weighted σ_k-curvatures; see Proposition 5.1 below.

Corollary 2.18. *Let $k \in \mathbb{N}$, $n \in \mathbb{N}_0$ and $m \in [k-1, \infty)$. Suppose $(\lambda; P) \in \Gamma_k^{m,\pm}$. Then*

$$(\pm 1)^{k-1} T_{k-1}^m(\lambda; P) > 0 \quad \textit{and} \quad (\pm 1)^{k-1} s_{k-1}^m(\lambda; P) > 0.$$

Proof. Let $\Lambda = \Lambda(P)$ denote the eigenvalues of P. By Lemma 2.6, the conclusion is equivalent to the assertions

$$(\pm 1)^{k-1} \sigma_{k-1}^m(\lambda; \Lambda(i)) > 0 \qquad \text{for all } 1 \le i \le n, \tag{2-19a}$$

$$(\pm 1)^{k-1} s_{k-1}^m(\lambda; P) > 0. \tag{2-19b}$$

We prove (2-19) by induction on k. Clearly (2-19) holds when $k = 1$. Suppose that (2-19) holds for all $(\lambda; P) \in \Gamma_k^{m,\pm}$. Let $(\lambda; P) \in \Gamma_{k+1}^{m,\pm}$. In particular, $(\pm 1)^{k+1} \sigma_{k+1}^m(\lambda; P) > 0$. Lemma 2.4 and (2-8), respectively, imply that

$$(\pm 1)^{k+1} \big[\sigma_{k+1}^m(\lambda; \Lambda(i)) + \lambda_i \sigma_k^m(\lambda; \Lambda(i))\big] > 0, \tag{2-20a}$$

$$(\pm 1)^{k+1} \left[s_{k+1}^m(\lambda; P) + \frac{\lambda}{m} s_k^m(\lambda; P) \right] > 0 \tag{2-20b}$$

for all $1 \le i \le n$. By the inductive hypothesis, $(\pm 1)^{k-1} \sigma_{k-1}^m(\lambda; \Lambda(i)) > 0$ and $(\pm 1)^{k-1} s_{k-1}^m(\lambda; P) > 0$. Multiplying these to both sides of (2-20a) and (2-20b), respectively, and then using Theorem 2.13 (and Lemma 2.8 for the inequality involving s_k^m) and Lemma 2.4 in succession yields

$$0 < \sigma_k^m(\lambda; \Lambda(i)) [\sigma_k^m(\lambda; \Lambda(i)) + \lambda_i \sigma_{k-1}^m(\lambda; \Lambda(i))] = \sigma_k^m(\lambda; \Lambda(i)) \sigma_k^m(\lambda; P),$$

$$0 < s_k^m(\lambda; P) \left(s_k^m(\lambda; P) + \frac{\lambda}{m} s_{k-1}^m(\lambda; P) \right) = s_k^m(\lambda; P) \sigma_k^m(\lambda; P).$$

The conclusion now follows from the assumption $(\lambda; P) \in \Gamma_{k+1}^{m,\pm}$. □

3. Smooth metric measure spaces

Recall that a smooth metric measure space is a five-tuple (M^n, g, v, m, μ) consisting of a Riemannian manifold (M^n, g), a positive function $v \in C^\infty(M; \mathbb{R}_+)$, a dimensional parameter $m \in \mathbb{R}_+$, and an auxiliary curvature parameter $\mu \in \mathbb{R}$. The geometric study of smooth metric measure spaces is based on *weighted local invariants* of smooth metric measure spaces; i.e., tensor-valued functions on the space $\mathfrak{M}(M, m, \mu)$ of metric-measure structures on (M^n, m, μ) which are invariant with respect to the natural action of the diffeomorphism group of M (see Section 4 for further discussion). Many weighted local invariants can be realized as local Riemannian invariants of the formal warped product (1-1). In this section, we intrinsically define and discuss the weighted local invariants which are important to our study of the weighted σ_k-curvatures, as well as the properties of these invariants for weighted Einstein manifolds.

Among the most familiar weighted local invariants of a smooth metric measure space (M^n, g, v, m, μ) are the *Bakry–Émery Ricci tensor*

$$\mathrm{Ric}_\phi^m := \mathrm{Ric} + \nabla^2\phi - \frac{1}{m} d\phi \otimes d\phi,$$

the *weighted Laplacian*

$$\Delta_\phi := \Delta - \nabla\phi,$$

and the *weighted volume element*

$$d\nu := v^m \,\mathrm{dvol}_g .$$

The Bakry–Émery Ricci tensor plays an important role in the comparison geometry of smooth metric measure spaces (cf. [Wei and Wylie 2009]). This is because of its appearance in the weighted Bochner formula for the weighted Laplacian on functions. Note also that the weighted Laplacian is the natural formally self-adjoint (rough) Laplacian on smooth metric measure spaces, in that $\Delta_\phi = -\nabla^*\nabla$, where ∇^* is the adjoint of the Levi-Civita connection ∇ of g with respect to the L^2-inner product induced by the weighted volume element.

Another important local invariant of a smooth metric measure space is the *weighted scalar curvature*

$$R_\phi^m := R + 2\Delta\phi - \frac{m+1}{m}|\nabla\phi|^2 + m(m-1)\mu e^{\frac{2}{m}\phi}.$$

Among the reasons that R_ϕ^m is the natural analogue of the scalar curvature are that it is the scalar curvature of the warped product (1-1), it plays the role of the scalar curvature in O'Neill's submersion theorem [Lott 2007] and in the weighted Weitzenböck formula for the Dirac operator on spinors [Perelman 2002], and the variational properties of the total weighted scalar curvature functional are closely related to sharp Sobolev inequalities [Case 2015b] and special Einstein-type structures [Case 2015a; 2015b]. Two smooth metric measure spaces (M^n, g, v, m, μ) and $(M^n, \hat{g}, \hat{v}, m, \mu)$ are *pointwise conformally equivalent* if there is a positive function $u \in C^\infty(M; \mathbb{R}_+)$ such that $\hat{g} = u^{-2}g$ and $\hat{v} = u^{-1}v$. Given (M^n, g, v, m, μ), we denote by $\mathfrak{C}$ the set of all $(\hat{g}, \hat{v})$ such that $(M^n, \hat{g}, \hat{v}, m, \mu)$ is pointwise conformally equivalent to (M^n, g, v, m, μ). A smooth metric measure space (M^n, g, v, m, μ), $m \neq 1$, is *locally conformally flat in the weighted sense* if around every point there is a neighborhood U such that the restriction $(U, g|_{TU}, v|_U, m, \mu)$ is conformally equivalent to $(B, g_{-\mu}, 1, m, \mu)$, where B is an open set in the simply connected spaceform $(X^n, g_{-\mu})$ with constant sectional curvature $-\mu$. A smooth metric measure space $(M^n, g, v, 1, \mu)$ is *locally conformally flat in the weighted sense* if around every point there is a neighborhood U such that the restriction $(U, g|_{TU}, v|_U, m, \mu)$ is conformally equivalent to $(B, g_c, 1, 1, \mu)$, where B is an open set in a simply connected spaceform (X^n, g_c).

In order to discuss the conformal properties of smooth metric measure spaces, it is convenient to consider the following modifications of the weighted scalar curvature, the Bakry–Émery Ricci tensor, and the Riemann curvature tensor Rm:

$$\begin{aligned} J_\phi^m &:= \frac{m+n-2}{2(m+n-1)} R_\phi^m, \\ P_\phi^m &:= \operatorname{Ric}_\phi^m - \frac{1}{m+n-2} J_\phi^m g, \\ A_\phi^m &:= \operatorname{Rm} - \frac{1}{m+n-2} P_\phi^m \wedge g. \end{aligned}$$

Here $\wedge : \Gamma(S^2T^*M) \times \Gamma(S^2T^*M) \to \Gamma(\Lambda^2 S^2 T^*M)$ denotes the Kulkarni–Nomizu product. We call P_ϕ^m the *weighted Schouten tensor* and A_ϕ^m the *weighted Weyl tensor*. By Lemma 3.3 below, the weighted Weyl tensor is a weighted conformal invariant of smooth metric measure spaces, so that, in a weighted conformal class, the Riemann curvature tensor is completely controlled by the weighted Schouten tensor. Indeed, it is straightforward to check that a smooth metric measure space with $n \geq 3$ and $m+n \neq 3$ is locally conformally flat in the weighted sense if and only if $A_\phi^m = 0$; see [Case 2012b, Lemma 6.6].

The scalar invariant J_ϕ^m should be regarded as the weighted analogue of the trace of the Schouten tensor. However, it is not the trace of the weighted Schouten tensor. The following formula for the difference $Y_\phi^m := J_\phi^m - \operatorname{tr} P_\phi^m$ will be useful.

Lemma 3.1. *Let (M^n, g, v, m, μ) be a smooth metric measure space. Then*

$$Y_\phi^m = \Delta_\phi \phi - \frac{m}{m+n-2} J_\phi^m + m(m-1)\mu v^{-2}.$$

Proof. The definitions of the weighted scalar curvature and the Bakry–Émery Ricci tensor yield

$$R_\phi^m = \operatorname{tr} \operatorname{Ric}_\phi^m + \Delta_\phi \phi + m(m-1)\mu v^{-2},$$

from which the result readily follows. □

We also need two tensors formed by taking certain derivatives of the weighted Schouten tensor. The *weighted Cotton tensor* $dP_\phi^m \in \Gamma(\Lambda^2 T^*M \otimes T^*M)$ is defined by

$$dP_\phi^m(x, y, z) := \nabla_x P_\phi^m(y, z) - \nabla_y P_\phi^m(x, z)$$

for all $x, y, z \in T_pM$ and all $p \in M$. The *weighted Bach tensor* $B_\phi^m \in \Gamma(S^2T^*M)$ is defined by

$$B_\phi^m(x, y) := (\delta_\phi dP_\phi^m)(x, y) - \frac{1}{m} d\phi(y) \operatorname{tr} dP_\phi^m(\cdot, x, \cdot) + \left\langle A_\phi^m(\cdot, x, \cdot, y), P_\phi^m - \frac{Y_\phi^m}{m} g \right\rangle$$

for all $x, y, z \in T_pM$ and all $p \in M$, where we define

$$(\delta_\phi dP_\phi^m)(x, y) := \sum_{i=1}^n \nabla_{e_i} dP_\phi^m(e_i, x, y) - dP_\phi^m(\nabla\phi, x, y)$$

for $\{e_i\}_1^n$ an orthonormal basis for T_pM; see [Case 2012b; 2013b] for further discussion. The following identities involving the weighted Schouten, weighted Cotton, and weighted Weyl tensors are useful; see [Case 2012b] for their derivations.

Lemma 3.2. *Let (M^n, g, v, m, μ) be a smooth metric measure space. Then*

$$\begin{aligned}
\operatorname{tr} dP_\phi^m &= P_\phi^m(\nabla\phi) + dY_\phi^m - \frac{1}{m} Y_\phi^m \, d\phi, \\
\operatorname{tr} A_\phi^m &= \frac{m}{m+n-2} P_\phi^m - \nabla^2\phi + \frac{1}{m} d\phi \otimes d\phi + \frac{1}{m+n-2} Y_\phi^m g, \\
\delta_\phi P_\phi^m &= dJ_\phi^m - \frac{1}{m} Y_\phi^m \, d\phi, \\
\delta_\phi A_\phi^m &= \frac{m+n-3}{m+n-2} dP_\phi^m - v^{-1} dv \wedge \operatorname{tr} A_\phi^m.
\end{aligned}$$

Note that Lemma 3.2 implies that dP_ϕ^m vanishes if (M^n, g, v, m, μ) is locally conformally flat in the weighted sense.

We also need to know the behavior of these weighted curvatures under pointwise conformal transformations; see [Case 2012b] for derivations.

Lemma 3.3. *Let (M^n, g, v, m, μ) and $(M^n, \hat{g}, \hat{v}, m, \mu)$ be pointwise conformally equivalent smooth metric measure spaces. Define $f \in C^\infty(M)$ by $\hat{g} = e^{-2f/(m+n-2)} g$. Then*

$$\begin{aligned}
e^{-\frac{2f}{m+n-2}} \widehat{J}_\phi^m &= J_\phi^m + \Delta_\phi f - \tfrac{1}{2}|\nabla f|^2, \\
\widehat{P}_\phi^m &= P_\phi^m + \nabla^2 f + \frac{1}{m+n-2} df \otimes df - \frac{1}{2(m+n-2)} |\nabla f|^2 g, \\
\widehat{A}_\phi^m &= e^{\frac{2f}{m+n-2}} A_\phi^m, \\
\widehat{dP_\phi^m} &= dP_\phi^m - A_\phi^m(\cdot, \cdot, \nabla f, \cdot).
\end{aligned}$$

3A. ***The weighted σ_k-curvatures.*** We are now prepared to define the weighted σ_k-curvatures of a smooth metric measure space.

Definition 3.4. Fix $k \in \mathbb{N}_0$ and $\kappa \in \mathbb{R}$. The *weighted σ_k-curvature* (*with scale κ*) of a smooth metric measure space (M^n, g, v, m, μ) is

(3-1) $$\tilde{\sigma}_{k,\phi}^m := \sigma_k^m(Y_\phi^m + m\kappa v^{-1}; P_\phi^m).$$

The *k-th weighted Newton tensor* (*with scale κ*) of (M^n, g, v, m, μ) is

(3-2) $$\widetilde{T}_{k,\phi}^m := T_k^m(Y_\phi^m + m\kappa v^{-1}; P_\phi^m).$$

The *k-th weighted Newton scalar* (*with scale κ*) of (M^n, g, v, m, μ) is

$$\tilde{s}^m_{k,\phi} := s^m_k(Y^m_\phi + m\kappa v^{-1};\, P^m_\phi). \tag{3-3}$$

We shall omit the tilde from our notation and denote by $\sigma^m_{k,\phi}$, $T^m_{k,\phi}$, and $s^m_{k,\phi}$ the weighted σ_k-curvature, the k-th weighted Newton tensor, and the k-th weighted Newton scalar, respectively, with scale $\kappa = 0$. In order to give more succinct derivations, given a smooth metric measure space (M^n, g, v, m, μ), a parameter $\kappa \in \mathbb{R}$, and a nonnegative integer k, we denote

$$\begin{aligned}
\widetilde{Y}^m_\phi &:= Y^m_\phi + m\kappa v^{-1}, \\
\widetilde{Z}^m_\phi &:= \frac{1}{m}\widetilde{Y}^m_\phi, \\
\widetilde{N}^m_{k,\phi} &:= \operatorname{tr}(P^m_\phi)^k + m\bigl(\widetilde{Z}^m_\phi\bigr)^k.
\end{aligned}$$

As discussed in Section 5 below, the variational properties of the weighted σ_k-curvatures are closely related to the properties of the weighted divergence of the k-th weighted Newton tensor. For example, the fact, from [Case 2015b], that the weighted σ_1-curvature is variational is closely related to the following formula for the divergence of the first weighted Newton tensor.

Lemma 3.5. *Let (M^n, g, v, m, μ) be a smooth metric measure space and fix $\kappa \in \mathbb{R}$. Then*

$$\delta_\phi \widetilde{T}^m_{1,\phi} = -\tilde{s}^m_{1,\phi}\, d\phi.$$

Proof. By definition (3-2), we have that

$$\widetilde{T}^m_{1,\phi} = (J^m_\phi + m\kappa v^{-1})g - P^m_\phi.$$

Lemma 3.2 implies that

$$\delta_\phi \widetilde{T}^m_{1,\phi} = -\left(J^m_\phi + (m-1)\kappa v^{-1} - \frac{Y^m_\phi}{m}\right)d\phi.$$

Comparing with (3-3) yields the desired result. □

Lemma 3.5 allows us to compute the weighted divergence of the weighted Newton tensors of all orders.

Proposition 3.6. *Let (M^n, g, v, m, μ) be a smooth metric measure space and fix $\kappa \in \mathbb{R}$. Let $k \in \mathbb{N}_0$. Then*

$$\delta_\phi \widetilde{T}^m_{k,\phi} = -\tilde{s}^m_{k,\phi}\, d\phi + \sum_{j=0}^{k-2}(-1)^j \widetilde{T}^m_{k-2-j,\phi}(dP^m_\phi \cdot \widetilde{Q}^m_{j+1,\phi}),$$

where

$$\widetilde{Q}^m_{\ell,\phi} := \bigl(P^m_\phi\bigr)^\ell - \bigl(\widetilde{Z}^m_\phi\bigr)^\ell g$$

and $\cdot : \Gamma(\otimes^3 T^*M) \times \Gamma(S^2T^*M) \to \Gamma(T^*M)$ *denotes the fiber-wise contraction of* *the second argument into the first and third components of the first argument*; *i.e.*, $(A\cdot T)_\gamma := A_{\alpha\gamma\beta}T^{\alpha\beta}$.

Proof. We first show by induction that

$$d\tilde{\sigma}^m_{k,\phi} = \sum_{j=0}^{k-1} \frac{(-1)^j}{j+1}\tilde{\sigma}^m_{k-1-j,\phi}\, d\widetilde{N}^m_{j+1,\phi} \tag{3-4}$$

for all $k\in\mathbb{N}_0$, with the convention that the empty sum equals zero. Clearly (3-4) holds for $k=0$. Suppose that (3-4) holds for some $k\in\mathbb{N}_0$. Using the definition

$$(k+1)\tilde{\sigma}^m_{k+1,\phi} = \sum_{j=0}^{k}(-1)^j\tilde{\sigma}^m_{k-j,\phi}\widetilde{N}^m_{j+1,\phi} \tag{3-5}$$

of the weighted σ_k-curvature, we compute that

$$\begin{aligned}
(k+1)&d\tilde{\sigma}^m_{k+1,\phi} \\
&= \sum_{j=0}^{k}(-1)^j\tilde{\sigma}^m_{k-j,\phi}\, d\widetilde{N}^m_{j+1,\phi} + \sum_{j=0}^{k}\sum_{\ell=0}^{k-1-j}\frac{(-1)^{j+\ell}}{\ell+1}\tilde{\sigma}^m_{k-1-j-\ell,\phi}\widetilde{N}^m_{j+1,\phi}\, d\widetilde{N}^m_{\ell+1,\phi} \\
&= (k+1)\sum_{j=0}^{k}\frac{(-1)^j}{j+1}\tilde{\sigma}^m_{k-j,\phi}\, d\widetilde{N}^m_{j+1,\phi},
\end{aligned}$$

as desired.

We next show that

$$\begin{aligned}
\delta_\phi(P^m_\phi)^k + \big(\widetilde{Z}^m_\phi\big)^k\, d\phi = \sum_{j=0}^{k-1}\frac{1}{j+1}(P^m_\phi)^{k-1-j}(\nabla\widetilde{N}^m_{j+1,\phi}) \\
+\sum_{j=0}^{k-2}(P^m_\phi)^{k-2-j}(dP^m_\phi\cdot\widetilde{Q}^m_{j+1,\phi})
\end{aligned} \tag{3-6}$$

for all $k\in\mathbb{N}_0$. By definition,

$$\delta_\phi(P^m_\phi)^k = (P^m_\phi)^{k-1}(\delta_\phi P^m_\phi) + \sum_{j=1}^{k-1}(P^m_\phi)^{k-1-j}((\nabla P^m_\phi)\cdot(P^m_\phi)^j). \tag{3-7}$$

Observe that

$$(\nabla P^m_\phi)\cdot(P^m_\phi)^j = \frac{1}{j+1}d\widetilde{N}^m_{j+1,\phi} - \big(\widetilde{Z}^m_\phi\big)^j\, d\widetilde{Y}^m_\phi + dP^m_\phi\cdot(P^m_\phi)^j. \tag{3-8}$$

Using Lemma 3.2 to evaluate $\operatorname{tr} dP^m_\phi := dP^m_\phi\cdot g$, using Lemma 3.5 to evaluate $\delta_\phi P^m_\phi$, and inserting (3-8) into (3-7) yields (3-6).

Finally, the definition (3-2) of the k-th weighted Newton tensor implies that

$$\delta_\phi \widetilde{T}^m_{k,\phi} = \sum_{j=0}^{k} (-1)^j [(P^m_\phi)^j (\nabla \tilde{\sigma}^m_{k-j,\phi}) + \tilde{\sigma}^m_{k-j,\phi} \delta_\phi (P^m_\phi)^j].$$

Using (3-4) and (3-6) to evaluate the right-hand side of the above display yields the desired result. □

3B. ***Weighted Einstein manifolds.*** We conclude this section with a brief discussion of weighted Einstein manifolds and their properties as needed in Theorems 1.2, 1.3, 1.4, and related discussions.

Definition 3.7. A *weighted Einstein manifold* is a smooth metric measure space (M^n, g, v, m, μ) such that $P^m_\phi = \lambda g$ for some $\lambda \in \mathbb{R}$.

The following lemma states that for every weighted Einstein manifold, there is a scale for which it has constant weighted σ_1-curvature (cf. [Kim and Kim 2003, Proposition 5]).

Lemma 3.8. *Let (M^n, g, v, m, μ) be such that $P^m_\phi = \lambda g$ for $\lambda \in \mathbb{R}$. Then there is a unique constant $\kappa \in \mathbb{R}$ such that $\tilde{\sigma}^m_{1,\phi} = (m+n)\lambda$.*

Definition 3.9. The *scale* of a weighted Einstein manifold (M^n, g, v, m, μ) satisfying $P^m_\phi = \lambda g$ is the constant $\kappa \in \mathbb{R}$ such that $\tilde{\sigma}^m_{1,\phi} = (m+n)\lambda$.

Proof of Lemma 3.8. We must find a $\kappa \in \mathbb{R}$ such that $J^m_\phi + m\kappa v^{-1} = (m+n)\lambda$. Lemma 3.2 implies that

$$\delta_\phi (P^m_\phi - \lambda g) - \frac{1}{m} \operatorname{tr}(P^m_\phi - \lambda g)\, d\phi = v^{-1} d((J^m_\phi - (m+n)\lambda) v),$$

from which the conclusion readily follows. □

A special case of weighted Einstein manifolds already studied in the literature are quasi-Einstein manifolds [Case et al. 2011], a class of manifolds which include static metrics in general relativity, the bases of Einstein warped product manifolds, and gradient Ricci solitons.

Definition 3.10. A *quasi-Einstein manifold* is a weighted Einstein manifold with scale $\kappa = 0$.

Remark 3.11. It is readily checked that if (M^n, g, v, m, μ) is a quasi-Einstein manifold in the sense of Definition 3.10, then $\mathrm{Ric}^m_\phi = \frac{2(m+n-1)}{m+n-2}\lambda g$; i.e., (M^n, g) is quasi-Einstein in the sense of [Case et al. 2011].

Conversely, if (M^n, g) is such that $\mathrm{Ric}^m_\phi = \lambda g$ for some $\phi \in C^\infty(M)$ and some constants $\lambda \in \mathbb{R}$ and $m \in \mathbb{R}_+$, it is known that there is a constant $\mu \in \mathbb{R}$ such that $(M^n, g, e^{-\phi/m}, m, \mu)$ satisfies $R^m_\phi = (m+n)\lambda$ [Kim and Kim 2003]. It follows that $(M^n, g, e^{-\phi/m}, m, \mu)$ is quasi-Einstein in the sense of Definition 3.10.

We next discuss the positively curved flat models of quasi-Einstein and weighted Einstein manifolds; i.e., the smooth metric measure spaces which are locally conformally flat in the weighted sense and are weighted Einstein manifolds with nonnegative scale and $\lambda > 0$. In what follows, we regard $S^n \subset \mathbb{R}^{n+1}$ as the set

$$S^n = \{x \in \mathbb{R}^{n+1} \mid |x|^2 = 1\}$$

and let $x_1, \dots, x_{n+1}$ denote both the standard coordinates on $\mathbb{R}^{n+1}$ and their restriction to S^{n+1}. The upper hemisphere S^n_+ is defined by

$$S^n_+ = \{x \in S^n \mid x_{n+1} \geq 0\}.$$

Note that in both of the examples below the function v is allowed to vanish on a set of measure zero. Thus these spaces are not examples of smooth metric measure spaces on closed manifolds as defined in this article.

First we discuss the model spaces for quasi-Einstein manifolds. Additional examples of quasi-Einstein manifolds are discussed, for example, in [Case 2012a; He et al. 2012].

Example 3.12. Fix $n \in \mathbb{N}$ and $m \in \mathbb{R}$. The *positive elliptic m-Gaussian* is the quasi-Einstein manifold $(S^n_+, d\theta^2, x_{n+1}, m, 1)$, where $d\theta^2$ is the round metric of constant sectional curvature 1 on the n-sphere. Indeed, using the well-known facts

$$\begin{aligned} \nabla^2 x_{n+1} &= -x_{n+1}\, d\theta^2, \\ |\nabla x_{n+1}|^2 &= 1 - x_{n+1}^2, \end{aligned} \tag{3-9}$$

we readily compute that the positive elliptic m-Gaussian satisfies

$$P^m_\phi = \frac{m+n-2}{2} g, \qquad J^m_\phi = \frac{(m+n)(m+n-2)}{2}.$$

Given $a \in \mathbb{R}$ and $\xi \in \mathbb{R}^{n+1}$ such that $|\xi|^2 < 1$ and $\xi_{n+1} = 0$, denote by u the function on S^n_+ given by

$$u(\zeta) = \frac{a}{\sqrt{1-|\xi|^2}}(1 + \xi \cdot \zeta). \tag{3-10}$$

Consider the metric-measure structure $(\hat{g}, \hat{v}) = (u^{-2} d\theta^2, u^{-1} x_{n+1})$. Using "hats" to denote weighted invariants determined by $(\hat{g}, \hat{v})$, it is straightforward to check that

$$\widehat{P}^m_\phi = \frac{m+n-2}{2} a^2 \hat{g}, \qquad \widehat{J}^m_\phi = \frac{(m+n)(m+n-2)}{2} a^2.$$

Thus $(\hat{g}, \hat{v})$ is a quasi-Einstein metric-measure structure which is pointwise conformally equivalent to the positive elliptic m-Gaussian. Indeed, all such quasi-Einstein metrics are given by conformally rescaling by a function of the form (3-10); see Proposition 7.2 below. Note that if we choose $\xi_{n+1} \neq 0$, then $\widehat{J}^m_\phi$ is not constant; more precisely, $(\hat{g}, \hat{v})$ is a weighted Einstein metric-measure structure with nonzero scale.

Next we discuss the model spaces for weighted Einstein manifolds with positive scale. Further examples of weighted Einstein manifolds will be discussed elsewhere.

Example 3.13. Fix $n \in \mathbb{N}$ and $m \in \mathbb{R}$. The *standard m-weighted n-sphere* is the weighted Einstein manifold $(S^n, d\theta^2, 1+x_{n+1}, m, 0)$, where $d\theta^2$ denotes the round metric of constant sectional curvature 1. Indeed, using the facts (3-9), we readily compute that the standard m-weighted n-sphere satisfies

$$P_\phi^m = \frac{m+n-2}{2}, \qquad J_\phi^m + m(m+n-2)v^{-1} = \frac{(m+n)(m+n-2)}{2}.$$

Given $a \in \mathbb{R}$ and $\xi \in \mathbb{R}^{n+1}$ such that $|\xi|^2 < 1$, denote by u the function on S^n given by

$$u(\zeta) = \frac{a}{\sqrt{1-|\xi|^2}}(1+\xi\cdot\zeta).$$

Consider the metric-measure structure $(\hat{g}, \hat{v}) = (u^{-2}d\theta^2, u^{-1}(1+x_{n+1}))$. Using "hats" to denote weighted invariants determined by $(\hat{g}, \hat{v})$, it is straightforward to check that

$$\begin{aligned} \widehat{P}_\phi^m &= \frac{m+n-2}{2}a^2\hat{g}, \\ \widehat{J}_\phi^m + m(m+n-2)\left(\frac{a(1-\xi_{n+1})}{\sqrt{1-|\xi|^2}}\right)\hat{v}^{-1} &= \frac{(m+n)(m+n-2)}{2}a^2. \end{aligned}$$

Thus $(\hat{g}, \hat{v})$ is a weighted Einstein metric-measure structure with positive scale which is pointwise conformally equivalent to the standard m-weighted n-sphere. Indeed, all such weighted Einstein metrics are given by a function of this form; see Proposition 7.2 below.

Remark 3.14. Note that the space of quasi-Einstein metrics in the weighted conformal class of the n-dimensional positive elliptic m-Gaussian is $(n+1)$-dimensional, while the space of weighted Einstein metrics in the weighted conformal class of the standard m-weighted n-sphere is $(n+2)$-dimensional. This and the relation between weighted Einstein manifolds and sharp Gagliardo–Nirenberg inequalities [Case 2013a] provide strong evidence that weighted Einstein metrics are, from the point of view of conformal geometry, the more natural Einstein-type structure to study on smooth metric measure spaces.

Lemma 3.8 asserts that a weighted Einstein manifold (M^n, g, v, m, μ) with scale κ has constant weighted σ_1-curvature. Such manifolds also have constant weighted σ_k-curvatures and constant weighted Newton tensors. However, the weighted Bach tensor is not necessarily constant, though it can be computed in terms of only the two-jets of g and v.

Proposition 3.15. *Let (M^n, g, v, m, μ) be a weighted Einstein manifold with scale $\kappa \in \mathbb{R}$ and such that $P_\phi^m = \lambda g$. Let $k \in \mathbb{N}_0$. Then*

$$\tilde{\sigma}_{k,\phi}^m = \binom{m+n}{k}\lambda^k, \tag{3-11}$$

$$\tilde{s}_{k,\phi}^m = \binom{m+n-1}{k}\lambda^k, \tag{3-12}$$

$$\widetilde{T}_{k,\phi}^m = \binom{m+n-1}{k}\lambda^k g, \tag{3-13}$$

$$B_\phi^m = m\kappa v^{-1} P_\phi^{m-1} - \frac{m(m+n-3)}{m+n-2}\lambda\kappa v^{-1} g, \tag{3-14}$$

where

$$P_\phi^{m-1} = P_\phi^m + v^{-1}\nabla^2 v + \frac{1}{m(m+n-2)} Y_\phi^m g \tag{3-15}$$

is the weighted Schouten tensor of $(M^n, g, v, m-1, \mu)$.

Proof. To begin, consider an arbitrary smooth metric measure space (M^n, g, v, m, μ). Lemma 3.1 implies that the weighted Schouten scalar J_ϕ^{m-1} of $(M^n, g, v, m-1, \mu)$ satisfies

$$J_\phi^{m-1} = \frac{m+n-3}{m+n-2}\left(J_\phi^m - \frac{1}{m}Y_\phi^m\right). \tag{3-16}$$

This readily yields (3-15).

Now suppose that (M^n, g, v, m, μ) is a weighted Einstein manifold with scale κ. Since $\tilde{\sigma}_{1,\phi}^m = \operatorname{tr} P_\phi^m + \widetilde{Y}_\phi^m$, we see that

$$m\lambda = \widetilde{Y}_\phi^m = Y_\phi^m + m\kappa v^{-1}. \tag{3-17}$$

This yields (3-11), (3-12), and (3-13). From the definition of the weighted Bach tensor, we see that $B_\phi^m = \kappa v^{-1} \operatorname{tr} A_\phi^m$. Combining this observation with Lemma 3.2 and (3-15) yields (3-14). □

In the special case of a closed weighted Einstein smooth metric measure space $(M^n, g, v, m, 0)$, Proposition 3.15 yields a formula for $\int v^{-1} d\nu$ in terms of only λ, κ, and the weighted volume.

Proposition 3.16. *Let $(M^n, g, v, m, 0)$ be a closed weighted Einstein manifold with scale $\kappa \in \mathbb{R}$ and such that $P_\phi^m = \lambda g$. Then*

$$\lambda \int_M d\nu = \frac{2m+n-2}{2(m+n-1)}\kappa \int_M v^{-1}\, d\nu. \tag{3-18}$$

Proof. Given any closed smooth metric measure space $(M^n, g, v, m, 0)$ and any $\kappa \in \mathbb{R}$, Lemma 3.1 implies that

$$(2m+n-2)m\kappa \int_M v^{-1}\,d\nu = \int_M (m\tilde{\sigma}_{1,\phi}^m + (m+n-2)\widetilde{Y}_\phi^m)\,d\nu. \tag{3-19}$$

The conclusion (3-18) then follows from (3-11) and (3-17). □

4. The space of metric-measure structures

In this section we briefly discuss a natural formalism for studying the space of metric-measure structures on weighted manifolds. Recall that the *space of metric-measure structures* on (M^n, m, μ) is

$$\mathfrak{M}(M, m, \mu) := \mathrm{Met}(M) \times C^\infty(M; \mathbb{R}_+),$$

where $\mathrm{Met}(M)$ is the space of Riemannian metrics on M and $C^\infty(M; \mathbb{R}_+)$ is the space of positive smooth functions on M. When the weighted manifold (M^n, m, μ) is clear by context, we denote by $\mathfrak{M}$ its space of metric-measure structures. Note that, as a set, $\mathfrak{M}(M, m, \mu)$ depends only on M. The role of the parameters m and μ is to determine the geometry of elements of $\mathfrak{M}$, especially in the definitions of weighted invariants. The fact that $\mathfrak{M}(M, m, \mu) = \mathfrak{M}(M, m', \mu)$ as sets for any $m, m' \in \mathbb{R}$ provides a useful way to relate metric-measure structures for different values of m.

It is clear that a weighted conformal class $\mathfrak{C}$ on a weighted manifold is a subset of $\mathfrak{M}$; moreover, any $(g, v) \in \mathfrak{M}$ uniquely determines a weighted conformal class $\mathfrak{C} = [g, v]$. Just as many geometric variational problems are most naturally posed under a volume constraint, we consider the sets

$$\mathfrak{M}_1 := \left\{ (g, v) \in \mathfrak{M} \;\middle|\; \int_M d\nu(g, v) = 1 \right\}, \qquad \mathfrak{C}_1 := \mathfrak{M}_1 \cap \mathfrak{C}$$

consisting of unit-volume metric-measure structures within $\mathfrak{M}$ and $\mathfrak{C}$, respectively.

A *weighted invariant* is a function I defined on $\mathfrak{M}(M, m, \mu)$ which is invariant with respect to the action of the diffeomorphism group $\mathrm{Diff}(M)$ of M. One special class of weighted invariants consists of *weighted functionals*, namely maps $\mathcal{S} : \mathfrak{M} \to \mathbb{R}$ such that $\mathcal{S}(f^*g, f^*v) = \mathcal{S}(g, v)$ for every $f \in \mathrm{Diff}(M)$ and every $(g, v) \in \mathfrak{M}$. Another special class consists of *weighted scalar invariants*, namely maps $I_\phi^m : \mathfrak{M} \to C^\infty(M)$ such that $I_\phi^m(f^*g, f^*v) = f^*(I_\phi^m(g, v))$ for every $f \in \mathrm{Diff}(M)$ and every $(g, v) \in \mathfrak{M}$. For example, the weighted volume element $d\nu$ is a volume-element-valued weighted invariant, the weighted σ_k-curvatures are weighted scalar invariants, and the total weighted σ_k-curvature functionals obtained by integrating the weighted σ_k-curvature with respect to $d\nu$ are weighted functionals.

Let $(g, v) \in \mathfrak{M}$ and consider the formal tangent space $T_{(g,v)}\mathfrak{M}$ consisting of derivatives $\gamma'(0)$ of smooth paths $\gamma : \mathbb{R} \to \mathfrak{M}$ such that $\gamma(0) = (g, v)$. We denote by

$$T\mathfrak{M} := \bigcup_{(g,v)\in\mathfrak{M}} T_{(g,v)}\mathfrak{M}$$

the formal tangent bundle of $\mathfrak{M}$. We identify $T_{(g,v)}\mathfrak{M} \cong \Gamma(S^2T^*M) \oplus C^\infty(M)$ via the bijection

$$\Psi^m_{(g,v)} : T_{(g,v)}\mathfrak{M} \to \Gamma(S^2T^*M) \oplus C^\infty(M)$$

defined as follows: Given $\gamma'(0) \in T_{(g,v)}\mathfrak{M}$, set

$$h = v^2 \frac{\partial}{\partial t}\bigg|_{t=0} (v_t^{-2} g_t), \tag{4-1a}$$

$$\psi = -\frac{\partial}{\partial t}\bigg|_{t=0} \frac{d\nu(g_t, v_t)}{d\nu(g, v)}, \tag{4-1b}$$

for $(g_t, v_t) := \gamma(t)$. Then set $\Psi^m_{(g,v)}(\gamma'(0)) := (h, \psi)$. The inverse is

$$\begin{aligned}\left(\Psi^m_{(g,v)}\right)^{-1}(h,\psi) &= \gamma'(0)\\ &:= \left(h - \frac{2}{m+n}\left(\psi + \tfrac{1}{2}\operatorname{tr}_g h\right)g, -\frac{1}{m+n}\left(\psi + \tfrac{1}{2}\operatorname{tr}_g h\right)v\right).\end{aligned} \tag{4-2}$$

From (4-1a) we observe that $h = 0$ if and only if $\left(\Psi^m_{(g,v)}\right)^{-1}(h, \psi)$ is tangent to a curve in the weighted conformal class $[g, v]$. When the weighted manifold (M^n, m, μ) and metric-measure structure $(g, v) \in \mathfrak{M}$ are clear from context, we omit the function $\Psi^m_{(g,v)}$ and simply identify $T_{(g,v)}\mathfrak{M}$ with $\Gamma(S^2T^*M) \oplus C^\infty(M)$.

The splitting $S^2T^*M = S^2_0T^*M \oplus \mathbb{R}g$ of $S^2T^*M \cong T_g \operatorname{Met}(M)$ into trace-free and pure trace parts is the Riemannian analogue of the decomposition (4-1), especially in that the $S^2_0T^*M$-component of a formal tangent vector $\gamma'(0) \in S^2T^*M$ vanishes if and only if $\gamma'(0)$ is tangent to a curve in the weighted conformal class of $\gamma(0)$. This motivates us to define a $C^\infty(M)$-linear map $\operatorname{tf}_\phi : \Gamma(S^2T^*M) \to \Gamma(S^2T^*M)$ on any smooth metric measure space (M^n, g, v, m, μ) which produces the "weighted trace-free component" of a given tensor field. While it is unclear how to define this map in general, it is clear on a case-by-case basis: We denote

$$\operatorname{tf}_\phi g := 0, \qquad \operatorname{tf}_\phi (P^m_\phi)^k := (P^m_\phi)^k - \frac{1}{m+n} N^m_{k,\phi} g,$$

$$\operatorname{tf}_\phi \nabla^2 u := \nabla^2 u - \frac{1}{m+n}\Delta_\phi u\, g, \qquad \operatorname{tf}_\phi B^m_\phi := B^m_\phi$$

for all $k \in \mathbb{N}_0$ and all $u \in C^\infty(M)$. These definitions are motivated by considering the trace-free part of the analogous Riemannian invariants in $(m+n)$-dimensions.

Formally, the first variation (or linearization) of a weighted invariant I is the exterior derivative DI. For example, if $I_\phi^m : \mathfrak{M} \to C^\infty(M)$ is a weighted scalar invariant, its first variation is the map $DI_\phi^m : T\mathfrak{M} \to C^\infty(M)$ defined by

$$DI_\phi^m[h, \psi] := \frac{\partial}{\partial t}\bigg|_{t=0} I_\phi^m(\gamma(t))$$

for all $(h, \psi) \in T_{(g,v)}\mathfrak{M}$ and all $(g, v) \in \mathfrak{M}$, where $\gamma : \mathbb{R} \to \mathfrak{M}$ is a smooth path such that $\gamma(0) = (g, v)$ and $\gamma'(0) = \left(\Psi_{(g,v)}^m\right)^{-1}(h, \psi)$. A special case of interest is the linearization of the restriction of a weighted scalar invariant I_ϕ^m to a weighted conformal class $\mathfrak{C}$, which we regard as a map $DI_\phi^m : T\mathfrak{C} \to C^\infty(M)$. Our convention (4-1) is such that for any $(g, v) \in \mathfrak{C}$, we may identify $T_{(g,v)}\mathfrak{C} \cong C^\infty(M)$ by

$$T_{(g,v)}\mathfrak{C} = (\Psi_{(g,v)}^m)^{-1}\{(0, \psi) \mid \psi \in C^\infty(M)\}.$$

Equivalently, we identify a function $\psi \in C^\infty(M)$ with the tangent vector to the curve

$$\gamma(t) = \left(e^{-\frac{2t\psi}{m+n}} g, e^{-\frac{t\psi}{m+n}} v\right)$$

at $t = 0$. Note also that the identification $T_{(g,v)}\mathfrak{C} \cong C^\infty(M)$ gives rise to the identification

$$T_{(g,v)}\mathfrak{C}_1 \cong \left\{\psi \in C^\infty(M) \;\middle|\; \int_M \psi \, d\nu = 0\right\}. \tag{4-3}$$

When (M^n, m, μ) and $(g, v) \in \mathfrak{C}$ are clear by context, we simply identify $T_{(g,v)}\mathfrak{C}$ with $C^\infty(M)$ and we identify $T_{(g,v)}\mathfrak{C}_1$ with mean-free elements of $C^\infty(M)$.

We say that DI_ϕ^m is *formally self-adjoint* if for each $(g, v) \in \mathfrak{C}$, the operator $DI_\phi^m : C^\infty(M) \to C^\infty(M)$ is formally self-adjoint with respect to the natural L^2-inner product on $C^\infty(M)$ induced by $d\nu(g, v)$.

Let $\mathfrak{C}$ be a weighted conformal class on a weighted manifold (M^n, m, μ). A weighted scalar invariant I_ϕ^m is *conformally variational* (*on* $\mathfrak{C}$) if there is a weighted functional $\mathcal{S}$ such that

$$D\mathcal{S}[\psi] = \int_M I_\phi^m \psi \, d\nu \tag{4-4}$$

for all $\psi \in T_{(g,v)}\mathfrak{C}$ and all $(g, v) \in \mathfrak{C}$. This property is equivalent to the self-adjointness of DI_ϕ^m (cf. [Bourguignon 1986; Branson and Gover 2008]).

Lemma 4.1. *Let $\mathfrak{C}$ be a weighted conformal class on (M^n, m, μ). A weighted scalar invariant I_ϕ^m is conformally variational if and only if its linearization DI_ϕ^m is formally self-adjoint. Moreover, if I_ϕ^m is conformally variational, then*

$$D\left(\int_M I_\phi^m \, d\nu\right)[\psi] = -\int_M (I_\phi^m - DI_\phi^m[1])\psi \, d\nu \tag{4-5}$$

for all $\psi \in T_{(g,v)}\mathfrak{C}$ *and all* $(g, v) \in \mathfrak{C}$.

Proof. Consider the one-form $\Omega : T\mathfrak{M} \to \mathbb{R}$ defined by

$$\Omega[\psi] := \int_M I_\phi^m \psi \, d\nu$$

for all $\psi \in T_{(g,v)}\mathfrak{C}$ and all $(g, v) \in \mathfrak{C}$. By definition, I_ϕ^m is conformally variational if and only if Ω is exact. Since $\mathfrak{C}$ is contractible, Ω is exact if and only if it is closed. A straightforward computation shows that the formal exterior derivative $D\Omega : \Lambda^2 T_{(g,v)}\mathfrak{C} \to \mathbb{R}$ of Ω is

$$D\Omega[\psi_1, \psi_2] = \int_M (\psi_2 DI_\phi^m[\psi_1] - \psi_1 DI_\phi^m[\psi_2]) \, d\nu$$

for all $\psi_1, \psi_2 \in T_{(g,v)}\mathfrak{C}$ and all $(g, v) \in \mathfrak{C}$. Thus Ω is closed if and only if DI_ϕ^m is formally self-adjoint, yielding the first assertion. The second assertion follows immediately from the self-adjointness of DI_ϕ^m and (4-1b). □

If we restrict our attention to conformally variational weighted scalar invariants I_ϕ^m which are homogeneous with respect to homotheties in $\mathfrak{C}$, then $DI_\phi^m[1]$ is a constant multiple of I_ϕ^m, and hence (4-5) generically identifies the functional $\mathcal{S}$ in the definition (4-4) of the conformally variational property. For this reason, one frequently restricts attention to homogeneous invariants in the Riemannian setting (cf. [Branson and Gover 2008]). However, due to our goal of producing weighted functionals which include weighted Einstein manifolds among their critical points, we do not impose this homogeneity requirement; see Section 8 for a detailed discussion.

When computing the linearizations of the total weighted scalar curvature functionals, it is useful to take advantage of the fact that $\mathfrak{M}(M, m, \mu) = \mathfrak{M}(M, m', \mu)$ as sets for all $m, m' \in \mathbb{R}$. Hence $T\mathfrak{M}(M, m, \mu) = T\mathfrak{M}(M, m', \mu)$ as sets for all $m, m' \in \mathbb{R}$. However, the identification $T_{(g,v)}\mathfrak{M}(M, m, \mu) \cong \Gamma(S^2T^*M) \oplus C^\infty(M)$ via the function $\Psi_{(g,v)}^m$ depends on m. These observations allow us to relate the maps $\Psi_{(g,v)}^m$ and $\Psi_{(g,v)}^{m'}$.

Lemma 4.2. *Fix* $m, m' \in \mathbb{R}_+$ *and* $\mu \in \mathbb{R}$. *Let* M^n *be a smooth manifold. Given a metric-measure structure* $(g, v) \in \mathfrak{M}(M, m, \mu) = \mathfrak{M}(M, m', \mu)$, *define*

$$\Phi_m^{m'} : \Gamma(S^2T^*M) \oplus C^\infty(M) \to \Gamma(S^2T^*M) \oplus C^\infty(M)$$

by

$$\Phi_m^{m'}(h, \psi) = \left(h, \frac{m'+n}{m+n}\psi - \frac{m-m'}{2(m+n)} \operatorname{tr}_g h \right).$$

Then $\left(\Psi_{(g,v)}^{m'}\right)^{-1} \circ \Phi_m^{m'} \circ \Psi_{(g,v)}^m : T_{(g,v)}\mathfrak{M}(M, m, \mu) \to T_{(g,v)}\mathfrak{M}(M, m', \mu)$ *is the identity map.*

Proof. Let $\gamma : \mathbb{R} \to \mathfrak{M}$ be a smooth curve such that $\gamma(0) = (g, v)$. Let $d\nu^{(m)}$ and $d\nu^{(m')}$ denote the weighted volume elements on $\mathfrak{M}(M, m, \mu)$ and $\mathfrak{M}(M, m', \mu)$, respectively. A straightforward computation yields

$$\frac{d\nu^{(m')}(\gamma(t))}{d\nu^{(m')}(\gamma(0))} = \left(\frac{d\nu^{(m)}(\gamma(t))}{d\nu^{(m)}(\gamma(0))}\right)^{\frac{m'+n}{m+n}} (\det((v^{-2}g)^{-1}v_t^{-2}g_t))^{\frac{m-m'}{2(m+n)}}.$$

for $(g_t, v_t) = \gamma(t)$. The conclusion readily follows from (4-1). $\square$

Lemma 4.2 can be reformulated as a statement about linearizations of weighted invariants:

Corollary 4.3. *Fix $m, m' \in \mathbb{R}_+$ and $\mu \in \mathbb{R}$. Let M^n be a smooth manifold. Given a metric-measure structure*

$$(g, v) \in \mathfrak{M}(M, m, \mu) = \mathfrak{M}(M, m', \mu)$$

and a weighted invariant I, denote by $D^{(m)}I$ and $D^{(m')}I$ the linearizations of I when regarded as functions of

$$\Gamma(S^2T^*M) \oplus C^\infty(M) \cong T_{(g,v)}\mathfrak{M}(M, m, \mu),$$
$$\Gamma(S^2T^*M) \oplus C^\infty(M) \cong T_{(g,v)}\mathfrak{M}(M, m', \mu),$$

respectively. Then

$$D^{(m)}I = D^{(m')}I \circ \Phi_m^{m'}.$$

Since $\mathfrak{C}(M, m, \mu) = \mathfrak{C}(M, m', \mu)$ as sets, Corollary 4.3 also applies to linearizations of weighted invariants within weighted conformal classes. This observation is also reflected in the fact that $\Phi_m^{m'}$ acts as the identity in its first component.

Corollary 4.3 is quite useful for computing the linearizations of the weighted σ_k-curvatures for arbitrary scales κ. The reason for this is that it is straightforward to compute the linearizations of the weighted σ_k-curvatures with scale $\kappa = 0$ (cf. Section 8), while one can also exhibit the weighted σ_k-curvatures for arbitrary scales as perturbations of the weighted σ_k-curvatures with scale κ through weighted σ_k-curvatures of lower degree when computed with respect to different values of m. We separate this latter observation into two results so as to highlight the role of weighted conformal classes which are locally conformally flat in the weighted sense for larger values of k.

Lemma 4.4. *Let (M^n, m, μ) be a weighted manifold and fix $\kappa \in \mathbb{R}$. Then*

$$\tilde{\sigma}_{1,\phi}^m = \sigma_{1,\phi}^m + m\kappa v^{-1}, \tag{4-6}$$

$$\tilde{\sigma}_{2,\phi}^m = \sigma_{2,\phi}^m + \frac{m(m+n-2)}{m+n-3}\kappa v^{-1}\sigma_{1,\phi}^{m-1} + \frac{m(m-1)}{2}\kappa^2 v^{-2} \tag{4-7}$$

for all $(g, v) \in \mathfrak{M}$.

Proof. (4-6) follows immediately from Corollary 2.3. Applying Corollary 2.3 and Lemma 2.8 implies that

$$\tilde{\sigma}_{2,\phi}^m = \sigma_{2,\phi}^m + m\kappa v^{-1} s_{1,\phi}^m + \binom{m}{2}\kappa^2 v^{-2}. \tag{4-8}$$

On the other hand, (3-16) implies that

$$s_{1,\phi}^m = \frac{m+n-2}{m+n-3}\sigma_{1,\phi}^{m-1}. \tag{4-9}$$

Inserting (4-9) into (4-8) yields (4-7). □

Lemma 4.5. *Let (M^n, m, μ) be a weighted manifold and fix a scale $\kappa \in \mathbb{R}$. Let $\mathfrak{C} \subset \mathfrak{M}$ be a weighted conformal class which is locally conformally flat in the weighted sense. Given an integer $k \leq m$, it holds that*

$$\tilde{\sigma}_{k,\phi}^m = \sum_{j=0}^{k} \left(\frac{m+n-2}{m+n-2-j}\right)^{k-j} \binom{m}{j} (\kappa v^{-1})^j \sigma_{k-j,\phi}^{m-j}$$

for all $(g, v) \in \mathfrak{C}$.

Proof. Note that Lemma 2.8 and (4-9) together imply that

$$\sigma_1^{m-1}\left(\frac{m-1}{m}Y_\phi^m;\, P_\phi^m\right) = \frac{m+n-2}{m+n-3}\sigma_1^{m-1}(Y_\phi^{m-1};\, P_\phi^{m-1}). \tag{4-10}$$

for all $(g, v) \in \mathfrak{M}$. On the other hand, since $\mathfrak{C}$ is locally conformally flat in the weighted sense, Lemma 3.2 and (3-15) imply that

$$P_\phi^m = \frac{m+n-2}{m+n-3}P_\phi^{m-1}.$$

It follows that

$$A_\phi^{m-1} := \mathrm{Rm} - \frac{1}{m+n-3}P_\phi^{m-1} \wedge g = 0;$$

i.e., $(M^n, g, v, m-1, \mu)$ is locally conformally flat in the weighted sense. Therefore

$$P_\phi^m = \frac{m+n-2}{m+n-2-k}P_\phi^{m-k} \tag{4-11}$$

for all integers $k \leq m$. Combining this with (4-10) and an obvious induction argument yields

$$Z_\phi^m = \frac{m+n-2}{m+n-2-k}Z_\phi^{m-k} \tag{4-12}$$

for all integers $k \leq m$. Inserting (4-11) and (4-12) into Corollary 2.3 yields the desired result. □

5. Variational status of the weighted σ_k-curvatures

By Lemma 4.1, answering the question of when the weighted σ_k-curvatures are conformally variational can be achieved by characterizing when the linearizations $D\tilde{\sigma}^m_{k,\phi}$ are formally self-adjoint. To that end, we compute $D\tilde{\sigma}^m_{k,\phi}$.

Proposition 5.1. *Let (M^n, g, v, m, μ) be a smooth metric measure space and fix $\kappa \in \mathbb{R}$ and $k \in \mathbb{N}_0$. Set $\mathfrak{C} = [g, v]$. The linearization*

$$D\tilde{\sigma}^m_{k,\phi} : T_{(g,v)}\mathfrak{C} \to C^\infty(M)$$

is

$$\begin{aligned} D\tilde{\sigma}^m_{k,\phi}[\psi] = \frac{1}{m+n}(2k\tilde{\sigma}^m_{k,\phi} - m\kappa v^{-1}\tilde{s}^m_{k-1,\phi})\psi + \frac{m+n-2}{m+n}\delta_\phi(\widetilde{T}^m_{k-1,\phi}(\nabla\psi)) \\ - \frac{m+n-2}{m+n}\sum_{\ell=0}^{k-3}(-1)^\ell \widetilde{T}^m_{k-3-\ell,\phi}(\nabla\psi, dP^m_\phi \cdot \widetilde{Q}^m_{\ell+1,\phi}), \end{aligned}$$

where $\widetilde{Q}^m_{\ell,\phi}$ and $\cdot$ are as in Proposition 3.6.

Proof. By Proposition 3.6, it suffices to prove that

$$\begin{aligned} D\tilde{\sigma}^m_{k,\phi}[\psi] = \frac{1}{m+n}(2k\tilde{\sigma}^m_{k,\phi} - m\kappa v^{-1}\tilde{s}^m_{k-1,\phi})\psi \\ + \frac{m+n-2}{m+n}(\langle \widetilde{T}^m_{k-1,\phi}, \nabla^2\psi\rangle - \langle \tilde{s}^m_{k-1,\phi}\nabla\phi, \nabla\psi\rangle) \end{aligned} \tag{5-1}$$

for all $k \in \mathbb{N}_0$. We prove (5-1) by induction.

Clearly (5-1) holds when $k = 0$. Suppose that (5-1) holds for some $k \in \mathbb{N}_0$. From (3-5), we note that

$$(k+1)D\tilde{\sigma}^m_{k+1,\phi} = \sum_{j=0}^{k}(-1)^j[\widetilde{N}^m_{j+1,\phi}D\tilde{\sigma}^m_{k-j,\phi} + \tilde{\sigma}^m_{k-j,\phi}D\widetilde{N}^m_{j+1,\phi}]. \tag{5-2}$$

Using Lemma 3.3, we observe that

$$\begin{aligned} DP^m_\phi[\psi] &= \frac{m+n-2}{m+n}\nabla^2\psi, \\ D\widetilde{Z}^m_\phi[\psi] &= \frac{1}{m+n}(2\widetilde{Z}^m_\phi - \kappa v^{-1})\psi - \frac{m+n-2}{m(m+n)}\langle\nabla\phi, \nabla\psi\rangle. \end{aligned}$$

In particular,

$$\begin{aligned} D\widetilde{N}^m_{k,\phi}[\psi] = \frac{k}{m+n}(2\widetilde{N}^m_{k,\phi} - m\kappa v^{-1}(\widetilde{Z}^m_\phi)^{k-1})\psi \\ + \frac{(m+n-2)k}{m+n}(\langle (P^m_\phi)^{k-1}, \nabla^2\psi\rangle - \langle (\widetilde{Z}^m_\phi)^{k-1}\nabla\phi, \nabla\psi\rangle). \end{aligned} \tag{5-3}$$

Inserting (5-3) and the inductive hypothesis (5-1) into (5-2) yields

$$D\tilde{\sigma}^m_{k+1,\phi}[\psi] = \frac{1}{m+n}(2(k+1)\tilde{\sigma}^m_{k+1,\phi} - m\kappa v^{-1}\tilde{s}^m_{k,\phi})\psi + \frac{m+n-2}{m+n}(\langle \widetilde{T}^m_{k,\phi}, \nabla^2\psi\rangle - \langle \tilde{s}^m_{k,\phi}\nabla\phi, \nabla\psi\rangle),$$

as desired. □

Our study of the formal self-adjointness of $D\tilde{\sigma}^m_{k,\phi}$ follows the Riemannian analogue carried out by Branson and Gover [2008].

Theorem 5.2. *Let $\mathfrak{C}$ be a weighted conformal class on (M^n, m, μ). Fix $\kappa \in \mathbb{R}$ and $k \in \mathbb{N}_0$. If $k \leq 2$ or $\mathfrak{C}$ is locally conformally flat in the weighted sense, then the weighted σ_k-curvature is conformally variational. If additionally $k \leq m+n$, then the converse holds.*

Proof. Denote

$$\widetilde{S}^m_{k,\phi} := \sum_{\ell=0}^{k-3}(-1)^\ell \widetilde{T}^m_{k-3-\ell,\phi}(dP^m_\phi \cdot \widetilde{Q}^m_{\ell+1,\phi}).$$

From Lemma 4.1 and Proposition 5.1, we see that $\tilde{\sigma}^m_{k,\phi}$ is conformally variational if and only if

$$\int_M \langle \widetilde{S}^m_{k,\phi}, \omega\nabla\eta - \eta\nabla\omega\rangle \, d\nu = 0 \tag{5-4}$$

for all $\eta, \omega \in C^\infty_0(M)$ and all representatives of $\mathfrak{C}$. Clearly $\widetilde{S}^m_{k,\phi} = 0$ is sufficient for (5-4) to hold. Suppose instead that (5-4) holds. Taking $\eta = 1$ and $\omega \in C^\infty_0(M)$ yields $\delta_\phi \widetilde{S}^m_{k,\phi} = 0$. Hence $\widetilde{S}^m_{k,\phi}$ is orthogonal to $\omega\nabla\eta$ in $L^2(d\nu)$ for all $\omega, \eta \in C^\infty_0(M)$. This implies that $\widetilde{S}^m_{k,\phi} = 0$. We conclude that $\tilde{\sigma}^m_{k,\phi}$ is conformally variational if and only if $\widetilde{S}^m_{k,\phi} = 0$ for all representatives of $\mathfrak{C}$.

Clearly $\widetilde{S}^m_{k,\phi} = 0$ if $k \leq 2$ or $\mathfrak{C}$ is locally conformally flat in the weighted sense. We show that if $3 \leq k \leq m+n$ and $\widetilde{S}^m_{k,\phi} = 0$ for all $(g, v) \in \mathfrak{C}$, then $\mathfrak{C}$ is locally conformally flat in the weighted sense. Fix a representative $(g, v) \in \mathfrak{C}$, a point $p \in M$, a vector $X \in T_pM$, and a tensor $\Omega \in S^2T^*_pM$. Let $f \in C^\infty(M)$ be such that $f(p) = 0$, $(\nabla f)_p = X$, and $(\nabla^2 f)_p = \Omega$. Set $(\hat{g}, \hat{v}) = (e^{-2f/(m+n-2)}g, e^{-f/(m+n-2)}v)$. By Lemma 3.3, at p it holds that

$$\begin{aligned}
(P^m_\phi)_{\hat{g}} &= P^m_\phi + \Omega + \frac{1}{m+n-2}X^\flat \otimes X^\flat - \frac{1}{2(m+n-2)}|X|^2 g, \\
(\widetilde{Z}^m_\phi)_{\hat{g}} &= \widetilde{Z}^m_\phi - \frac{1}{m}\langle X, \nabla\phi\rangle - \frac{1}{2(m+n-2)}|X|^2, \\
(dP^m_\phi)_{\hat{g}} &= dP^m_\phi - A^m_\phi(\cdot, \cdot, X, \cdot),
\end{aligned} \tag{5-5}$$

where the left-hand side (resp. right-hand side) of each equality is evaluated in terms of $(M^n, \hat{g}, \hat{v}, m, \mu)$ (resp. (M^n, g, v, m, μ)). We shall use (5-5) with two different choices of Ω and X.

First, choose $X_0 \in T_pM$ such that $|X_0|^2 = 2(m+n-2)$ and $\langle X_0, \nabla\phi\rangle = 0$. Let $t \in \mathbb{R}$, set $X = tX_0$, and choose

$$\Omega = -\frac{1}{m+n-2} X^\flat \otimes X^\flat.$$

It follows from (5-5) that $(\widetilde{Q}^m_{\ell+1,\phi})_{\hat{g}}$, $(dP^m_\phi)_{\hat{g}}$ and $(\widetilde{T}_{\ell,\phi})_{\hat{g}}$ are polynomial in t for all $\ell \in \mathbb{N}$; indeed,

$$\begin{aligned}
(dP^m_\phi)_{\hat{g}} &= -tA^m_\phi(\cdot,\cdot,X_0,\cdot) + dP^m_\phi, \\
(\widetilde{Q}^m_{\ell+1,\phi})_{\hat{g}} &= (-1)^\ell(\ell+1)t^{2\ell}\widetilde{Q}^m_{1,\phi} + O(t^{2\ell-2}), \\
(\widetilde{T}_{\ell,\phi})_{\hat{g}} &= (-1)^\ell\binom{m+n-1}{\ell}t^{2\ell}g + O(t^{2\ell-2}).
\end{aligned}$$

Using the identity $\sum_{\ell=0}^k(-1)^\ell(\ell+1)\binom{n}{k-\ell} = \binom{n-2}{k}$, we compute that

$$\bigl(\widetilde{S}^m_{k,\phi}\bigr)_{\hat{g}} = (-1)^{k-3}\binom{m+n-3}{k-3}t^{2k-6}(dP^m_\phi - tA^m_\phi(\cdot,\cdot,X_0,\cdot))\cdot\widetilde{Q}^m_{1,\phi} + O(t^{2k-7}).$$

Since t is arbitrary, $dP^m_\phi \cdot (P^m_\phi - \widetilde{Z}^m_\phi g) = 0$. Since the representative (g, v) is arbitrary, this holds for all representatives of $\mathfrak{C}$.

Second, choose $X = 0$ and let Ω be arbitrary. From (5-5) we see that, at p,

$$(dP^m_\phi \cdot (P^m_\phi - \widetilde{Z}^m_\phi g))_{\hat{g}} = dP^m_\phi \cdot (P^m_\phi - \widetilde{Z}^m_\phi g) + dP^m_\phi \cdot \Omega.$$

The conclusion of the previous paragraph implies that $dP^m_\phi \cdot \Omega = 0$ for all $(g, v) \in \mathfrak{C}$ and all $\Omega \in S^2T^*M$. Hence $dP^m_\phi = 0$ for all representatives $(g, v) \in \mathfrak{C}$. Lemma 3.3 then implies that $A^m_\phi = 0$, as desired. □

5A. ***The weighted σ_k-curvature functionals.*** Proposition 5.1 enables us to compute the linearization of the total weighted σ_k-curvature functionals. Recalling our goal that weighted Einstein metrics be stable with respect to these functionals, some care is needed in defining them. In the case of scale zero, the definition of the total weighted σ_k-curvature functionals is the expected one.

Definition 5.3. Let (M^n, m, μ) be a closed weighted manifold. Given $k \in \mathbb{N}$, the *$\mathcal{F}_k$-functional* $\mathcal{F}_k : \mathfrak{M} \to \mathbb{R}$ is defined by

$$\mathcal{F}_k(g, v) := \int_M \sigma^m_{k,\phi}\, dv.$$

From Proposition 5.1 we deduce that for generic values of m, the critical points of the restriction of the $\mathcal{F}_k$-functionals to $\mathfrak{C}_1$ have constant weighted σ_k-curvature.

Proposition 5.4. *Fix $k \in \mathbb{N}_0$ and let $\mathfrak{C}$ be a weighted conformal class on a closed weighted manifold (M^n, m, μ); if $k \geq 3$, assume that $\mathfrak{C}$ is locally conformally flat in the weighted sense. Then the first variation $D\mathcal{F}_k : T\mathfrak{C} \to \mathbb{R}$ of the $\mathcal{F}_k$-functional is*

$$D\mathcal{F}_k[\psi] = -\frac{m+n-2k}{m+n} \int_M \sigma_{k,\phi}^m \psi \, d\nu \tag{5-6}$$

for all $\psi \in T_{(g,v)}\mathfrak{C}$ and all $(g, v) \in \mathfrak{C}$. In particular, if $m+n \neq 2k$, then $(g, v) \in \mathfrak{C}_1$ is a critical point of the restriction $\mathcal{F}_k : \mathfrak{C}_1 \to \mathbb{R}$ if and only if (M^n, g, v, m, μ) is such that $\sigma_{k,\phi}^m$ is constant.

Proof. Proposition 5.1 and the proof of Theorem 5.2 imply that

$$D\sigma_{k,\phi}^m[\psi] = \frac{2k}{m+n}\sigma_{k,\phi}^m \psi + \frac{m+n-2}{m+n}\delta_\phi(T_{k-1,\phi}^m(\nabla\psi)).$$

The conclusion (5-6) follows from this and (4-5). The characterization of the critical points of $\mathcal{F}_k : \mathfrak{C}_1 \to \mathbb{R}$ follows from (4-3) and (5-6). □

In the case of positive scale, the total weighted σ_k-curvature functionals are the $\mathcal{Y}_k$-functionals defined below.

Definition 5.5. Let $(M^n, m, 0)$ be a closed weighted manifold. Given $k \in \mathbb{N}$, the *$\mathcal{Z}_k$-functional* $\mathcal{Z}_k : \mathfrak{M} \times \mathbb{R}_+ \to \mathbb{R}$ and the *$\mathcal{Y}_k$-functional* $\mathcal{Y}_k : \mathfrak{M} \times \mathbb{R}_+ \to \mathbb{R}$ are

$$\mathcal{Z}_k(g, v, \kappa) := \kappa^{-\frac{2mk(m+n-1)}{(m+n)(2m+n-2)}} \int_M \tilde{\sigma}_{k,\phi}^m \, d\nu,$$

$$\mathcal{Y}_k(g, v, \kappa) := \mathcal{Z}_k(g, v, \kappa)\left(\int_M v^{-1} \, d\nu\right)^{-\frac{2mk}{(m+n)(2m+n-2)}} \left(\int_M d\nu\right)^{-\frac{m+n-2k}{m+n}}$$

for all $(g, v) \in \mathfrak{M}$ and all $\kappa \in \mathbb{R}_+$.

The $\mathcal{Y}_k$-functionals are invariant with respect to both the natural homothetic rescalings in a weighted conformal class and the homothetic rescalings within a Riemannian conformal class.

Lemma 5.6. *Let $(M^n, m, 0)$ be a closed weighted manifold and let $k \in \mathbb{N}$. Then*

$$\mathcal{Y}_k(c^2 g, cv, c^{-1}\kappa) = \mathcal{Y}_k(g, v, \kappa),$$
$$\mathcal{Y}_k(c^2 g, v, c^{-2}\kappa) = \mathcal{Y}_k(g, v, \kappa)$$

for all $(g, v) \in \mathfrak{M}$ and all $c, \kappa \in \mathbb{R}_+$.

Proof. Observe that

$$\tilde{\sigma}_{k,\phi}^m(c^2 g, cv, c^{-1}\kappa) = c^{-2k}\tilde{\sigma}_{k,\phi}^m(g, v, \kappa), d\nu(c^2 g, cv, c^{-1}\kappa) = c^{m+n} d\nu(g, v, \kappa),$$
$$\tilde{\sigma}_{k,\phi}^m(c^2 g, v, c^{-2}\kappa) = c^{-2k}\tilde{\sigma}_{k,\phi}^m(g, v, \kappa), \; d\nu(c^2 g, v, c^{-2}\kappa) = c^n d\nu(g, v, \kappa).$$

The conclusion readily follows. □

In the remainder of this section, we begin to explore the variational properties of the $\mathcal{Y}_k$-functionals within a weighted conformal class $\mathfrak{C}$. Our interest is in the cases when the weighted σ_k-curvatures are variational, hence we assume that $\mathfrak{C}$ is locally conformally flat in the weighted sense if $k \geq 3$. The scale-invariance of the $\mathcal{Y}_k$-functionals implies that it suffices to consider the *$\mathcal{Y}_k$-functionals $\mathcal{Y}_k : \mathfrak{C} \to \mathbb{R}$ with scale $\kappa > 0$* defined by $\mathcal{Y}_k(g, v) := \mathcal{Y}_k(g, v, \kappa)$.

Remark 5.7. The scale-invariance of the $\mathcal{Y}_k$-functionals implies that minimizers of $\mathcal{Y}_k : \mathfrak{C} \times \mathbb{R}_+ \to \mathbb{R}$ are in one-to-one correspondence with minimizers of the functionals $\widetilde{\mathcal{Y}}_k : \mathfrak{C} \to \mathbb{R}$ defined by

$$\widetilde{\mathcal{Y}}_k(g, v) = \inf_{\kappa>0} \mathcal{Y}_k(g, v, \kappa).$$

Up to composition with a monotone function depending only on m and n, $\widetilde{\mathcal{Y}}_1$ is equivalent to the functional $\mathcal{Q}_1$ introduced by the author [Case 2015b] to study the weighted scalar curvature.

In order to compute the linearizations of the $\mathcal{Y}_k$-functionals, we first consider the linearizations of the $\mathcal{Z}_k$-functionals through variations of the scale κ.

Lemma 5.8. *Let $(M^n, g, v, m, 0)$ be a closed smooth metric measure space, fix $k \in \mathbb{N}$, and define $Z : \mathbb{R}_+ \to \mathbb{R}$ by $Z(\kappa) := \mathcal{Z}_k(g, v, \kappa)$. Then*

$$\kappa Z'(\kappa) = A(\kappa) \int_M \left[\tilde{\sigma}^m_{k,\phi} - \frac{(m+n)(2m+n-2)}{2k(m+n-1)} \kappa v^{-1} \tilde{s}^m_{k-1,\phi} \right] d\nu, \tag{5-7}$$

where $A(\kappa) = -2mk(m+n-1)/((m+n)(2m+n-2))\kappa^{-2mk(m+n-1)/((m+n)(2m+n-2))}$. In particular, if $(M^n, g, v, m, 0)$ is a weighted Einstein manifold with scale $\kappa > 0$, then $Z'(\kappa) = 0$.

Proof. Equation (5-7) follows immediately from Lemma 2.9. If $(M^n, g, v, m, 0)$ is a weighted Einstein manifold with scale κ, applying Propositions 3.15 and 3.16 yields $Z'(\kappa) = 0$. □

Remark 5.9. Let $p \in \mathbb{R}$ and let $(M^n, g, v, m, 0)$ be a weighted Einstein manifold with scale $\kappa > 0$ and nonvanishing weighted Schouten tensor. The same argument shows that the function $Z_p : \mathbb{R}_+ \to \mathbb{R}$ defined by $Z_p(\kappa) := \kappa^p \int \tilde{\sigma}^m_{k,\phi}\, d\nu$ satisfies $Z_p'(\kappa) = 0$ if and only if $p = -\frac{2mk(m+n-1)}{(m+n)(2m+n-2)}$.

We now compute the linearizations of the $\mathcal{Y}_k$-functionals. Lemma 5.6 implies that (5-7) is proportional to the linearization of $\mathcal{Y}_k : \mathfrak{C} \to \mathbb{R}$ (with scale κ) when restricted to homotheties. The critical points of the $\mathcal{Y}_k$-functionals are characterized as follows:

Proposition 5.10. *Let $k \in \mathbb{N}$ and let $\mathfrak{C}$ be a weighted conformal class on a closed weighted manifold (M^n, m, μ); if $k \geq 3$, assume that $\mathfrak{C}$ is locally conformally flat* *in the weighted sense. Fix a scale* $\kappa > 0$. *Then* $(g, v) \in \mathfrak{C}$ *is a critical point of* $\mathcal{Y}_k : \mathfrak{C} \to \mathbb{R}$ *if and only if*

$$\text{(5-8)}\quad \tilde{\sigma}^m_{k,\phi} + \frac{m}{m+n-2k}\kappa v^{-1}\tilde{s}^m_{k-1,\phi} = \frac{\int \tilde{\sigma}^m_{k,\phi}\,d\nu}{\int d\nu} + \frac{m}{m+n-2k}\left(\frac{\int \tilde{s}^m_{k-1,\phi}v^{-1}\,d\nu}{\int v^{-1}\,d\nu}\right)\kappa v^{-1}$$

and

$$\text{(5-9)}\qquad \int_M \tilde{\sigma}^m_{k,\phi}\,d\nu = \frac{(m+n)(2m+n-2)}{2k(m+n-1)}\int_M \kappa v^{-1}\tilde{s}^m_{k-1,\phi}\,d\nu.$$

Proof. Using Proposition 5.1 and Theorem 5.2, we compute that

$$\text{(5-10)}\qquad D\mathcal{Z}_k[\psi] = -B(\kappa)\int_M\left(\frac{m+n-2k}{m+n}\tilde{\sigma}^m_{k,\phi} + \frac{m}{m+n}\kappa v^{-1}\tilde{s}^m_{k-1,\phi}\right)\psi\,d\nu$$

for all $\psi \in T_{(g,v)}\mathfrak{C}$ and all $(g, v) \in \mathfrak{C}$, where $B(\kappa) = \kappa^{-\frac{2mk(m+n-1)}{(m+n)(2m+n-2)}}$. It follows that

$$\left(\int_M v^{-1}\,d\nu\right)^{\frac{2mk}{(m+n)(2m+n-2)}}\left(\int_M d\nu\right)^{\frac{m+n-2k}{m+n}} D\mathcal{Y}_k[\psi] = D\mathcal{Z}_k[\psi] + \frac{2mk(m+n-1)}{(m+n)^2(2m+n-2)}\mathcal{Z}_k\frac{\int \psi v^{-1}\,d\nu}{\int v^{-1}\,d\nu} + \frac{m+n-2k}{m+n}\mathcal{Z}_k\frac{\int \psi\,d\nu}{\int d\nu}$$

for all $\psi \in T_{(g,v)}\mathfrak{C}$ and all $(g, v) \in \mathfrak{C}$. Combining these two observations, we see that $(g, v) \in \mathfrak{C}$ is a critical point of $\mathcal{Y}_k : \mathfrak{C} \to \mathbb{R}$ if and only if

$$\text{(5-11)}\quad \tilde{\sigma}^m_{k,\phi} + \frac{m}{m+n-2k}\kappa v^{-1}\tilde{s}^m_{k-1,\phi} = \left(\frac{1}{\int d\nu} + \frac{2mk(m+n-1)}{(m+n)(m+n-2k)(2m+n-2)}\frac{v^{-1}}{\int v^{-1}\,d\nu}\right)\int_M \tilde{\sigma}^m_{k,\phi}\,d\nu.$$

Integrating with respect to $d\nu$ yields that (5-11) is equivalent to (5-8) and (5-9). □

6. Stability results for the $\mathcal{F}$- and $\mathcal{Y}$-functionals

It is known that closed Einstein metrics and closed gradient Ricci solitons are stable with respect to the total σ_k-curvature functionals [Viaclovsky 2000] and the total weighted σ_k-curvature functionals [Case 2016], respectively. In this section we show that the same is true for quasi-Einstein metrics. Based on these results and their usual proofs via the Lichnerowicz–Obata theorem [Lichnerowicz 1958; Obata 1962], we conjecture a Poincaré-type inequality for weighted Einstein manifolds which would imply that such manifolds are stable with respect to the $\mathcal{Y}_k$-functionals.

6A. *Stability for quasi-Einstein manifolds.* We begin by computing the second variation of the $\mathcal{F}_k$-functional at a critical point in the cases when the weighted σ_k-curvature is variational.

Proposition 6.1. *Let $k \in \mathbb{N}$ and let $\mathfrak{C}$ be a weighted conformal class on a closed weighted manifold (M^n, m, μ); if $k \geq 3$, assume additionally that $\mathfrak{C}$ is locally conformally flat in the weighted sense. Suppose that $(g, v) \in \mathfrak{C}_1$ is a critical point of the $\mathcal{F}_k$-functional. Then the second variation $D^2\mathcal{F}_k : T_{(g,v)}\mathfrak{C}_1 \to \mathbb{R}$ is given by*

$$D^2\mathcal{F}_k[\psi] = \frac{(m+n-2)(m+n-2k)}{(m+n)^2}\int_M T^m_{k-1,\phi}(\nabla\psi, \nabla\psi)\, dv - \frac{2k(m+n-2k)}{(m+n)^2}\int_M \sigma^m_{k,\phi}\psi^2\, dv$$

for all $\psi \in T_{(g,v)}\mathfrak{C}_1$.

Proof. By Proposition 5.4, since (g, v) is a critical point of $\mathcal{F}_k : \mathfrak{C}_1 \to \mathbb{R}$, the weighted σ_k-curvature $\sigma^m_{k,\phi}$ is constant. Therefore

$$D^2\mathcal{F}_k[\psi] = -\frac{m+n-2k}{m+n}\int_M \psi\, D\sigma^m_{k,\phi}[\psi]\, dv$$

for all $\psi \in T_{(g,v)}\mathfrak{C}_1$. The conclusion now follows from Proposition 5.1. □

Applying Proposition 6.1 in the case of a closed quasi-Einstein manifold yields the desired stability result.

Theorem 6.2. *Let $k \in \mathbb{N}$ and let (M^n, g, v, m, μ) be a closed quasi-Einstein manifold such that $P^m_\phi = \lambda g > 0$; if $k \geq 3$, assume additionally that (g, v) is locally conformally flat in the weighted sense.*

(1) *If $k < \frac{m+n}{2}$, then $D^2\mathcal{F}_k : T_{(g,v)}\mathfrak{C}_1 \to \mathbb{R}$ is positive definite.*

(2) *If $\frac{m+n}{2} < k \leq m+n$, then $D^2\mathcal{F}_k T_{(g,v)}\mathfrak{C}_1 \to \mathbb{R}$ is negative definite.*

Proof. It follows readily from Propositions 3.15 and 6.1 that

$$D^2\mathcal{F}_k[\psi] = \binom{m+n-1}{k-1}\frac{(m+n-2)(m+n-2k)}{(m+n)^2}\lambda^{k-1}\int_M |\nabla\psi|^2\, dv - \binom{m+n}{k}\frac{2k(m+n-2k)}{(m+n)^2}\lambda^k\int_M \psi^2\, dv \tag{6-1}$$

for all $\psi \in T_{(g,v)}\mathfrak{C}_1$. Since $P^m_\phi = \lambda g > 0$, we see that $\mathrm{Ric}^m_\phi = \frac{2(m+n-1)}{m+n-2}\lambda g > 0$. The weighted Lichnerowicz theorem [Bakry and Qian 2000, Theorem 14] implies that $\lambda_1(-\Delta_\phi) > \frac{2(m+n)}{m+n-2}\lambda$. Inserting this into (6-1) yields the result. □

Remark 6.3. We only consider the case of quasi-Einstein manifolds with positive weighted Schouten tensor because any closed quasi-Einstein manifold with nonpositive weighted Schouten tensor is Einstein [Kim and Kim 2003].

6B. ***Stability for weighted Einstein metrics.*** Propositions 3.15 and 5.10 imply that weighted Einstein manifolds with $\mu = 0$ and scale κ are critical points of the $\mathcal{Y}_k$-functional. In this subsection we conjecture a Lichnerowicz–Obata-type result which, if true, implies that such manifolds are infinitesimal minimizers of the $\mathcal{Y}_k$ functionals. To that end, we compute the linearization of $\tilde{s}^m_{k-1,\phi}$ at a weighted Einstein manifold.

Lemma 6.4. *Let $k \in \mathbb{N}$ and let $(M^n, g, v, m, 0)$ be a weighted Einstein manifold with $P^m_\phi = \lambda g > 0$ and scale $\kappa > 0$; if $k \geq 3$, assume additionally that $\mathfrak{C} = [g, v]$ is locally conformally flat in the weighted sense. Then*

$$D\tilde{s}^m_{k-1,\phi} = \binom{m+n-2}{k-2}\lambda^{k-2} \times \left[\frac{2(m+n-1)}{m+n}\lambda\psi - \frac{m-1}{m+n}\kappa\psi v^{-1} + \frac{m+n-2}{m+n} v\delta_\phi(v^{-1}\nabla\psi)\right]. \tag{6-2}$$

Proof. By Lemma 2.8,

$$\tilde{s}^m_{k-1,\phi} = \sigma^{m-1}_{k-1}\left(\frac{m-1}{m}(Y^m_\phi + m\kappa v^{-1});\ P^m_\phi\right).$$

Set $\kappa^{(m-1)} := \frac{m+n-3}{m+n-2}\kappa$. Arguing as in the proof of Lemma 4.5, we observe that for any weighted conformal class $\mathfrak{C}$, assumed to be locally conformally flat in the weighted sense if $k \geq 3$,

$$\tilde{s}^m_{k-1,\phi} = \left(\frac{m+n-2}{m+n-3}\right)^{k-1} \tilde{\sigma}^{m-1}_{k-1,\phi}, \tag{6-3}$$

where $\tilde{\sigma}^{m-1}_{k-1,\phi}$ is defined in terms of the scale $\kappa^{(m-1)}$. It follows from Corollary 4.3 and Proposition 5.1 that the linearization $D\tilde{\sigma}^{m-1}_{k-1,\phi} : T_{(g,v)}\mathfrak{C}(M, m, 0) \to C^\infty(M)$ is

$$D\tilde{\sigma}^{m-1}_{k-1,\phi}[\psi] = \frac{1}{m+n}\big(2(k-1)\tilde{\sigma}^{m-1}_{k-1,\phi} - (m-1)\kappa^{(m-1)}v^{-1}\tilde{s}^{m-1}_{k-2,\phi}\big)\psi + \frac{m+n-3}{m+n} v\delta_\phi(v^{-1}\widetilde{T}^{m-1}_{k-2,\phi}(\nabla\psi)) \tag{6-4}$$

for

$$\begin{aligned} \tilde{s}^{m-1}_{k-2,\phi} &:= s^{m-1}_{k-2}\big(Y^{m-1}_\phi + (m-1)\kappa^{(m-1)}v^{-1};\ P^{m-1}_\phi\big), \\ \widetilde{T}^{m-1}_{k-2,\phi} &:= T^{m-1}_{k-2}\big(Y^{m-1}_\phi + (m-1)\kappa^{(m-1)}v^{-1};\ P^{m-1}_\phi\big). \end{aligned}$$

Inserting this into (6-3), using (3-12) and (6-3) to evaluate $\tilde{\sigma}^{m-1}_{k-1,\phi}$, and using (4-11) to evaluate $\tilde{s}^{m-1}_{k-2,\phi}$ and $\widetilde{T}^{m-1}_{k-2,\phi}$ when $k \geq 3$ yields (6-2). □

Lemma 6.4 enables us to compute the second variation of the $\mathcal{Y}_k$-functional at a weighted Einstein manifold.

Proposition 6.5. *Let $k \in \mathbb{N}$ and let $(M^n, g, v, m, 0)$ be a closed weighted Einstein manifold with $P_\phi^m = \lambda g > 0$ and scale $\kappa > 0$; if $k \geq 3$, assume additionally that $\mathfrak{C} = [g, v]$ is locally conformally flat in the weighted sense. Then (g, v) is a critical point of $\mathcal{Y}_k : \mathfrak{C} \to \mathbb{R}$ and the second variation*

$$D^2\mathcal{Y}_k : T_{(g,v)}\mathfrak{C} \to \mathbb{R}$$

is given by

$$D^2\mathcal{Y}_k[\psi] = \frac{(m+n-2)(m+n-2k)}{(m+n)^2}\binom{m+n-1}{k-1}\mathcal{V}_{-1}^{-a}\mathcal{V}_0^{-b}\lambda^{k-1}I_1[\psi] \\ + \frac{m(m+n-2)}{(m+n)^2}\binom{m+n-2}{k-2}\mathcal{V}_{-1}^{-a}\mathcal{V}_0^{-b}\lambda^{k-2}\kappa I_2[\psi]$$

for all $\psi \in T_{(g,v)}\mathfrak{C}$, where $\mathcal{V}_{-1}$ and $\mathcal{V}_0$ denote the functionals $\mathcal{V}_{-1} := \int v^{-1} d\nu$ and $\mathcal{V}_0 := \int d\nu$; a and b denote the constants $a := \frac{2mk}{(m+n)(2m+n-2)}$, $b := \frac{m+n-2k}{m+n}$; and

$$I_1[\psi] := B(\kappa)\int_M \Big[|\nabla\psi|^2 - \frac{2(m+n)}{m+n-2}\lambda(\psi-\overline{\psi})^2 \\ + \frac{m}{m+n-2}\Big(\psi^2 - 2\overline{\psi}\psi + \frac{(m+n-1)(2m+n)}{(m+n)(2m+n-2)} \\ \times \Big(\frac{\int \psi v^{-1}}{\int v^{-1}}\Big)\psi\Big)\kappa v^{-1}\Big] d\nu \tag{6-5a}$$

for $\overline{\psi} = \frac{\int \psi}{\int 1}$ and

$$I_2[\psi] := B(\kappa)\int_M \Big[|\nabla\psi|^2 - \frac{2(m+n-1)}{m+n-2}\lambda\psi^2 + \frac{m-1}{m+n-2}\kappa\psi^2 v^{-1} \\ + \frac{2(m+n-1)^2}{(m+n-2)(2m+n-2)}\Big(\frac{\int \psi v^{-1}}{\int v^{-1}}\Big)\lambda\psi\Big] v^{-1}\, d\nu. \tag{6-5b}$$

Proof. By rescaling if necessary, we may suppose that $\kappa = 1$. It follows immediately from Propositions 3.15, 3.16 and 5.10 that (g, v) is a critical point of the $\mathcal{Y}_k$-functional. In particular,

$$0 = D\mathcal{Y}_k = \frac{1}{\mathcal{V}_{-1}^a\mathcal{V}_0^b}D\mathcal{Z}_k - a\frac{\mathcal{Y}_k}{\mathcal{V}_{-1}}D\mathcal{V}_{-1} - b\frac{\mathcal{Y}_k}{\mathcal{V}_0}D\mathcal{V}_0. \tag{6-6}$$

It follows from (6-6) that

$$D^2\mathcal{Y}_k = \frac{1}{\mathcal{V}_{-1}^a\mathcal{V}_0^b}D^2\mathcal{Z}_k - a\frac{\mathcal{Y}_k}{\mathcal{V}_{-1}}D^2\mathcal{V}_{-1} - b\frac{\mathcal{Y}_k}{\mathcal{V}_0}D^2\mathcal{V}_0 - a(a-1)\frac{\mathcal{Y}_k}{\mathcal{V}_{-1}^2}(D\mathcal{V}_{-1})^2 \\ - 2ab\frac{\mathcal{Y}_k}{\mathcal{V}_{-1}\mathcal{V}_0}(D\mathcal{V}_{-1})(D\mathcal{V}_0) - b(b-1)\frac{\mathcal{Y}_k}{\mathcal{V}_0^2}(D\mathcal{V}_0)^2. \tag{6-7}$$

Next, observe that

$$\begin{aligned} D\mathcal{V}_{-1}[\psi] &= -\frac{m+n-1}{m+n}\int_M \psi v^{-1}\,d\nu, & D\mathcal{V}_0[\psi] &= -\int_M \psi\,d\nu, \\ D^2\mathcal{V}_{-1}[\psi] &= \left(\frac{m+n-1}{m+n}\right)^2\int_M \psi^2 v^{-1}\,d\nu, & D^2\mathcal{V}_0[\psi] &= \int_M \psi^2\,d\nu. \end{aligned} \tag{6-8}$$

From (5-10) we note that

$$\begin{aligned} D^2\mathcal{Z}_k[\psi] = -\int_M &\left(\frac{m+n-2k}{m+n}D\tilde{\sigma}^m_{k,\phi}[\psi] + \frac{m}{m+n}v^{-1}D\tilde{s}^m_{k-1,\phi}[\psi]\right)\psi\,d\nu \\ &+ \int_M \left(\frac{m+n-2k}{m+n}\tilde{\sigma}^m_{k,\phi} + \frac{m(m+n-1)}{(m+n)^2}v^{-1}\tilde{s}^m_{k-1,\phi}\right)\psi^2\,d\nu. \end{aligned} \tag{6-9}$$

Since $(M^n, g, v, m, 0)$ is weighted Einstein, Propositions 3.15 and 5.1 imply that

$$D\tilde{\sigma}^m_{k,\phi}[\psi] = \binom{m+n-1}{k-1}\left(2\lambda^k - \frac{m}{m+n}\lambda^{k-1}v^{-1} + \frac{m+n-2}{m+n}\Delta_\phi\right)\psi. \tag{6-10}$$

Inserting (6-2), (6-8), (6-9) and (6-10) into (6-7) yields the desired conclusion. □

Based on similar stability results for quasi-Einstein manifolds, we expect that weighted Einstein manifolds are stable with respect to the $\mathcal{Y}_k$-functionals in the cases when the weighted σ_k-curvatures are variational. Indeed, based on the proofs of those results, we expect the following Poincaré-type inequalities for weighted Einstein manifolds.

Conjecture 6.6. *Let $(M^n, g, v, m, 0)$ be a closed weighted Einstein manifold with $P^m_\phi = \lambda g > 0$ and scale $\kappa > 0$. Let $I_1, I_2 : C^\infty(M) \to \mathbb{R}$ be as in (6-5). Then*

$$\inf\left\{ I_j[\psi] \,\middle|\, \int_M \psi\,d\nu = 1 \right\} > 0$$

for $j \in \{1, 2\}$. In particular, $D^2\mathcal{Y}_k : T_{(g,v)}\mathfrak{C} \to \mathbb{R}$ is positive definite.

7. Ellipticity and some Obata-type theorems

The results of Section 6 prove that quasi-Einstein manifolds are infinitesimally rigid with respect to the $\mathcal{F}_k$-functionals within a volume-normalized weighted conformal class in the cases when the weighted σ_k-curvatures are variational. It is natural to ask if *global* rigidity holds. Global rigidity is known in the Riemannian [Obata 1971; Viaclovsky 2000] and infinite-dimensional [Case 2016] cases when $k = 1$ or within the weighted conformal class of the respective flat model. One expects similar results for general $m \in \mathbb{R}_+$. In this section we prove the analogous global rigidity result for quasi-Einstein metrics. We also prove a result which is expected to play a key role in establishing the analogous global rigidity result for weighted

Einstein metrics. These results hold within the positive weighted elliptic k-cones, so named because the Euler equations of the $\mathcal{F}_k$- and $\mathcal{Y}_k$-functions are elliptic within these cones. Based on these results, we conjecture certain sharp fully nonlinear Sobolev inequalities.

Our strategy is modeled on Obata's proof [1971] that on a compact manifold, every conformally Einstein constant scalar curvature metric is itself Einstein. There are two key ingredients in his proof. First, the variational structure of the scalar curvature yields a particular trace-free tensor which is divergence-free for every constant scalar curvature metric. Second, if a metric is conformally Einstein, then the trace-free part of the Schouten tensor is a positive multiple of an element of the range of the conformal Killing operator; i.e., the trace-free part of the Lie derivative on vector fields. In our setting, when the weighted σ_k-curvature is variational, Proposition 3.6 effectively identifies the desired weighted trace-free tensor which is divergence-free for critical points of the $\mathcal{F}_k$- and $\mathcal{Y}_k$-functionals (cf. Section 8). The analogous formula for the weighted Schouten tensor of a metric-measure structure which is conformal to a weighted Einstein metric is as follows (cf. [Case 2015a; 2015b]):

Lemma 7.1. *Let (M^n, g, v, m, μ) be a smooth metric measure space and fix a scale $\kappa \in \mathbb{R}$. Suppose that $u \in C^\infty(M; \mathbb{R}_+)$ is such that the smooth metric measure space $(M^n, \hat{g}, \hat{v}, m, \mu)$ with metric-measure structure $(\hat{g}, \hat{v}) := (u^{-2}g, u^{-1}v)$ is a weighted Einstein manifold with scale $\hat{\kappa} \in \mathbb{R}$. Then*

$$u(P^m_\phi - \widetilde{Z}^m_\phi g) = -(m+n-2)\left(\nabla^2 u + \frac{1}{m}\langle \nabla u, \nabla\phi\rangle\, g\right) + (\hat{\kappa} - \kappa u)v^{-1}\, g. \tag{7-1}$$

Proof. Because $(M^n, \hat{g}, \hat{v}, m, \mu)$ is weighted Einstein with scale $\hat{\kappa}$, it follows from (3-17) that

$$\widehat{P}^m_\phi - \widehat{Z}^m_\phi \hat{g} = \hat{\kappa}\hat{v}^{-1}\hat{g}.$$

Using Lemma 3.3 to evaluate $\widehat{P}^m_\phi$ and $\widehat{Z}^m_\phi$ in terms of (M^n, g, v, m, μ) and u yields the desired result. □

In the cases of interest to us, only the flat models discussed in Section 3 admit multiple quasi-Einstein or weighted Einstein metric-measure structures within the given weighted conformal class (cf. [Case 2015b, Proposition 9.5]).

Proposition 7.2. *Let $\mathfrak{C}$ be a weighted conformal class on a weighted manifold (M^n, m, μ). Suppose that $(g, v), (\hat{g}, \hat{v}) \in \mathfrak{C}$ are weighted Einstein metric-measure structures with scale $\kappa, \hat{\kappa} \in \mathbb{R}$, respectively. Set $u = v\hat{v}^{-1}$. Then either u is constant or (M^n, g, v, m, μ) splits isometrically as a warped product. In particular:*

(1) *If (M^n, g, v, m, μ) is closed and $\kappa = \hat{\kappa} = 0$, then u is constant.*

(2) *If* $(M^n, g, v, m, \mu) = (S^n_+, d\theta^2, x_{n+1}, m, 1)$ *and* $\hat{\kappa} = \kappa = 0$, *then there exist a constant* $c > 0$ *and a point* $\xi \in \mathbb{R}^n = x^{-1}_{n+1}(0)$ *such that*

$$u(\zeta) = c\big(\sqrt{1 + |\xi|^2} + \xi \cdot \zeta\big).$$

(3) *If* $(M^n, g, v, m, \mu) = (\mathbb{R}^n, dx^2, 1, m, 0)$ *and* $\hat{\kappa} > 0$, *then there exist a constant* $c > 0$ *and a point* $x_0 \in \mathbb{R}^n$ *such that*

$$u(x) = \frac{\hat{\kappa}}{2(m+n-2)}|x - x_0|^2 + c.$$

Remark 7.3. Recall that the assumption that (M^n, g, v, m, μ) is closed means that M^n is a closed manifold, $v \in C^\infty(M; \mathbb{R}_+)$ is a positive function on M, and $m \in \mathbb{R}_+$.

Proof. Suppose that u is nonconstant. Then (7-1) implies that $\nabla^2 u = \frac{1}{n}\Delta u\, g$. It is well known (see [Cheeger and Colding 1996]) that this condition implies that (M^n, g) splits as a warped product over a one-dimensional base and that u depends only on the base. Indeed, if $P^m_\phi = \frac{m+n-2}{2}\lambda g$ and $\widehat{P}^m_\phi = \frac{m+n-2}{2}\hat{\lambda}\hat{g}$, then Lemma 3.3 implies that

$$\nabla^2 u = \tfrac{1}{2}(\hat{\lambda} u^{-1} - \lambda u + u^{-1}|\nabla u|^2) g.$$

Integrating this yields a constant $c \in \mathbb{R}$ such that

$$(u')^2 = -\lambda u^2 + cu - \hat{\lambda} \tag{7-2}$$

(see [Case 2015b]).

Suppose now that (M^n, g, v, m, μ) is closed and $\kappa = \hat{\kappa} = 0$. Solving (7-2) implies that u is of the form $u(t) = a + b\cos t$ for $b \in \mathbb{R}$ and $a > |b|$. Hence (M^n, g) is homothetic to the round n-sphere. By rescaling and changing coordinates if necessary, we may thus suppose that $u(x) = a + bx_{n+1}$ for $a > b \geq 0$. It follows from (7-1) that v, and hence u, is constant.

Next, suppose that $\hat{\kappa} = 0$ and (M^n, g, v, m, μ) is the positive elliptic m-Gaussian. By homothetically scaling if necessary, we may suppose that $\widehat{P}^m_\phi = \frac{m+n-2}{2}\hat{g}$. From (7-2) we conclude that there is a point $\xi \in \mathbb{R}^{n+1}$ such that

$$u(\zeta) = \sqrt{1 + |\xi|^2} + \xi \cdot \zeta.$$

On the other hand, (7-1) implies that

$$\nabla^2 u = \frac{\langle \nabla u, \nabla x_{n+1}\rangle}{x_{n+1}}\, d\theta^2,$$

from which we conclude that $\xi_{n+1} = 0$.

Finally, suppose that $(M^n, g, v, m, \mu) = (\mathbb{R}^n, dx^2, 1, m, 0)$ and $\hat{\kappa} > 0$. From (7-1) we conclude that

$$\nabla^2 u = \frac{\hat{\kappa}}{m+n-2}\, dx^2.$$

Hence u is a quadratic polynomial on $\mathbb{R}^n$ with leading order term $\frac{\hat{\kappa}}{2(m+n-2)}|x|^2$. $\square$

7A. *An Obata-type theorem for quasi-Einstein manifolds.* When the scale κ vanishes, the tensor field $E^m_{k,\phi}$ defined below is the desired weighted analogue of the trace-free tensor which is divergence-free for Riemannian metrics with constant σ_k-curvature.

Lemma 7.4. *Let $k \in \mathbb{N}$ and let (M^n, g, v, m, μ) be a smooth metric measure space; if $k \geq 2$, assume additionally that (g, v) is locally conformally flat in the weighted sense. Define*

$$E^m_{k,\phi} := T^m_{k,\phi} - \frac{m+n-k}{m+n}\sigma^m_{k,\phi} g.$$

Then

$$\delta_\phi E^m_{k,\phi} - \frac{1}{m}\operatorname{tr} E^m_{k,\phi}\, d\phi = -\frac{m+n-k}{m+n} d\sigma^m_{k,\phi}.$$

Proof. A straightforward computation yields

$$\operatorname{tr} T^m_{k,\phi} = (m+n-k)\sigma^m_{k,\phi} - m s^m_{k,\phi}.$$

In particular, it holds that

$$\operatorname{tr} E^m_{k,\phi} = \frac{m(m+n-k)}{m+n}\sigma^m_{k,\phi} - m s^m_{k,\phi}.$$

The conclusion now follows from Proposition 3.6. □

We only expect global rigidity within the positive weighted elliptic k-cones.

Definition 7.5. Fix $k \in \mathbb{N}_0$. The *positive weighted elliptic k-cone* Γ^+_k on a weighted manifold (M^n, m, μ) is the set

$$\Gamma^+_k := \left\{(g, v) \in \mathfrak{M} \;\middle|\; (Y^m_\phi(p); P^m_\phi(p)) \in \Gamma^+_k \text{ for all } p \in M\right\}.$$

Note that Euler equation of the $\mathcal{F}_k$-functional is elliptic within the positive weighted elliptic k-cone.

Proposition 7.6. *Let $\mathfrak{C}$ be a weighted conformal class on (M^n, m, μ). Fix $k \in \mathbb{N}_0$ and a representative $(g, v) \in \mathfrak{C}$. Identify*

$$\mathfrak{C} \cap \Gamma^+_k = \left\{u \in C^\infty(M; \mathbb{R}_+) \mid (u^{-2}g, u^{-1}v) \in \Gamma^+_k\right\}. \tag{7-3}$$

Then the operator $D : \mathfrak{C} \cap \Gamma^+_k \to C^\infty(M)$ defined by $D(u) := \sigma^m_{k,\phi}(u^{-2}g, u^{-1}v)$ is elliptic.

Proof. Proposition 5.1 implies that the principal symbol of the linearization of D at $u \in \mathfrak{C} \cap \Gamma^+_k$ is $\frac{m+n-2}{m+n} T^m_{k-1,\phi}$, where $T^m_{k-1,\phi}$ is the $(k-1)$-th weighted Newton tensor of $(u^{-2}g, u^{-1}v)$. Corollary 2.18 then implies that D is elliptic at u. □

We now adapt Obata's argument [1971] to closed quasi-Einstein manifolds.

Theorem 7.7. *Let $k \in \mathbb{N}$ and let $(M^n, \hat{g}, \hat{v}, m, \mu)$ be a closed quasi-Einstein manifold such that $\int d\hat{v} = 1$ and $P_\phi^m > 0$; if $k \geq 2$, assume additionally that $\mathfrak{C} := [\hat{g}, \hat{v}]$ is locally conformally flat in the weighted sense. Then $(g, v) \in \mathfrak{C}_1 \cap \Gamma_k^+$ is a critical point of $\mathcal{F}_k : \mathfrak{C}_1 \to \mathbb{R}$ if and only if $(g, v) = (\hat{g}, \hat{v})$.*

Proof. First, suppose that $(g, v) = (\hat{g}, \hat{v})$. It follows from Propositions 3.15 and 5.4 that (g, v) is in the weighted elliptic cone Γ_k^+ and is a critical point of the $\mathcal{F}_k$-functional $\mathcal{F}_k : \mathfrak{C}_1 \to \mathbb{R}$.

Conversely, suppose that (g, v) is a critical point of the $\mathcal{F}_k$-functional. It follows from Proposition 5.4 that $\sigma_{k,\phi}^m$ is constant. Let $E_{k,\phi}^m$ be as in Lemma 7.4. Then

$$\delta_\phi E_{k,\phi}^m - \frac{1}{m} \operatorname{tr} E_{k,\phi}^m \, d\phi = 0. \tag{7-4}$$

Let $u = v\hat{v}^{-1}$. Using Lemma 7.1 and (7-4), we compute that

$$\begin{aligned} 0 &= \int_M \left\langle E_{k,\phi}^m, \nabla^2 u + \frac{1}{m} \langle \nabla u, \nabla \phi \rangle \, g \right\rangle d\nu \\ &= -\frac{1}{m+n-2} \int_M u \langle E_{k,\phi}^m, P_\phi^m - Z_\phi^m \, g \rangle \, d\nu. \end{aligned}$$

It follows from Corollary 2.17 that (g, v) is quasi-Einstein. Proposition 7.2 and the normalization $(g, v), (\hat{g}, \hat{v}) \in \mathfrak{C}_1$ then imply that $(g, v) = (\hat{g}, \hat{v})$. $\square$

We expect that the assumption that M is closed in Theorem 7.7 can be removed; i.e., that one can use the assumption that $d\nu$ is a finite measure to still carry out the integration by parts (cf. [Chang et al. 2003; González 2006]). We further expect that, with a lot of work, one can show that the $\mathcal{F}_k$-functional realizes its infimum under suitable geometric assumptions on the background smooth metric measure space (cf. [Guan and Wang 2003; Sheng et al. 2007]). These expectations motivate the following conjecture (cf. [Guan and Wang 2004]).

Conjecture 7.8. *Fix $k \in \mathbb{N}$ and let $\mathfrak{C}$ be the weighted conformal class of the weighted elliptic m-Gaussian $(S_+^n, d\theta^2, x_{n+1}, m, 1)$. Then,*

$$\int_M \sigma_{k,\phi}^m \, d\nu \geq C \left(\int_M d\nu \right)^{\frac{m+n-2k}{m+n}} \tag{7-5}$$

for all $(g, v) \in \mathfrak{C} \cap \Gamma_k^+$. Moreover, equality holds in (7-5) if and only if

$$(S_+^n, g, v, m, 1)$$

is homothetic to a weighted elliptic m-Gaussian.

7B. ***Towards an Obata theorem for the $\mathcal{Y}_k$-functional on S^n.*** For smooth metric measure spaces with positive scale, the analogue of Lemma 7.4 is as follows:

Proposition 7.9. *Fix $k \in \mathbb{N}$ and let (M^n, g, v, m, μ) be a smooth metric measure space; if $k \geq 2$, assume additionally that (g, v) is locally conformally flat in the weighted sense. Define*

$$\widetilde{E}^m_{k,\phi} := \widetilde{T}^m_{k,\phi} - \frac{m+n-k}{m+n}\tilde{\sigma}^m_{k,\phi} g, \tag{7-6}$$

$$\widetilde{U}^m_{k-1,\phi} := T^{m-1}_{k-1}\left(\frac{m-1}{m}\widetilde{Y}^m_\phi;\, P^m_\phi\right) - \frac{m+n-k}{m+n}\tilde{s}^m_{k-1,\phi} g, \tag{7-7}$$

$$\hat{\sigma}^m_{k,\phi} := \tilde{\sigma}^m_{k,\phi} + \frac{m}{m+n-2k}\left(\tilde{s}^m_{k-1,\phi} - \frac{\int \tilde{s}^m_{k-1,\phi} v^{-1}}{\int v^{-1}}\right)\kappa v^{-1}, \tag{7-8}$$

$$\begin{aligned}\widehat{E}^m_{k,\phi} := \widetilde{E}^m_{k,\phi} &+ \frac{m}{m+n-2k}\kappa v^{-1}\widetilde{U}^m_{k-1,\phi} \\ &- \frac{m(m+n-k)}{(m+n)(m+n-1)(m+n-2k)}\left(\frac{\int \tilde{s}^m_{k-1,\phi} v^{-1}}{\int v^{-1}}\right)\kappa v^{-1} g.\end{aligned} \tag{7-9}$$

Then

$$\delta_\phi \widehat{E}^m_{k,\phi} - \frac{1}{m}\operatorname{tr}\widehat{E}^m_{k,\phi}\, d\phi = -\frac{m+n-k}{m+n}d\hat{\sigma}^m_{k,\phi}.$$

Proof. A straightforward computation yields

$$\operatorname{tr}\widetilde{E}^m_{k,\phi} = \frac{m(m+n-k)}{m+n}\tilde{\sigma}^m_{k,\phi} - m\tilde{s}^m_{k,\phi}.$$

Combining this with Proposition 3.6 yields

$$\delta_\phi \widetilde{E}^m_{k,\phi} - \frac{1}{m}\operatorname{tr}\widetilde{E}^m_{k,\phi}\, d\phi = -\frac{m+n-k}{m+n}d\tilde{\sigma}^m_{k,\phi}. \tag{7-10}$$

Next we show that

$$\delta_\phi(v^{-1}\widetilde{U}^m_{k-1,\phi}) - \frac{1}{m}\operatorname{tr}(v^{-1}\widetilde{U}^m_{k-1,\phi})\, d\phi = -\frac{m+n-k}{m+n}d(v^{-1}\tilde{s}^m_{k-1,\phi}). \tag{7-11}$$

This is clear if $k = 1$, so suppose $k \geq 2$. Arguing as in the proof of Lemma 4.5, we see that

$$\widetilde{U}^m_{k-1,\phi} = \left(\frac{m+n-2}{m+n-3}\right)^{k-1}\left(\widetilde{E}^{m-1}_{k-1,\phi} + \frac{m+n-k}{(m+n)(m+n-1)}\tilde{\sigma}^{m-1}_{k-1,\phi} g\right) \tag{7-12}$$

(cf. Lemma 6.4), where $\widetilde{E}^{m-1}_{k-1,\phi}$ is defined by (7-6) in terms of $(M^n, g, v, m-1, \mu)$ and the scale $\kappa^{(m-1)} := (m+n-3)/(m+n-2)\kappa$. Let $\delta^{(m-1)}_\phi$ denote the divergence with respect to the weighted measure $d\nu^{(m-1)}$ of $(M^n, g, v, m-1, \mu)$. Note that $\delta^{(m-1)}_\phi = v \circ \delta^{(m)}_\phi \circ v^{-1}$, where v and v^{-1} act as multiplication operators. In particular, (7-10) and (7-12) together yield (7-11).

Finally, a simple calculation yields

$$\delta_\phi(v^{-1}g) - \frac{1}{m}\operatorname{tr}(v^{-1}g)\,d\phi = -(m+n-1)dv^{-1}. \tag{7-13}$$

Combining (7-10), (7-11) and (7-13) yields the conclusion. □

By Proposition 5.10, the Euler equation of the $\mathcal{Y}_k$-functional with scale κ is completely determined by $\hat{\sigma}^m_{k,\phi}$. The Euler equation is also elliptic within the positive weighted elliptic k-cone.

Proposition 7.10. *Fix $k \in \mathbb{N}_0$ and $\kappa \in \mathbb{R}_+$. Let $\mathfrak{C}$ be a weighted conformal class on $(M^n, m, 0)$ and fix a representative $(g, v) \in \mathfrak{C}$. If $k \geq 3$, assume additionally that $\mathfrak{C}$ is locally conformally flat in the weighted sense. In terms of (7-3), the operator $D : \mathfrak{C} \cap \Gamma^+_k \to C^\infty(M)$ defined by $D(u) := \hat{\sigma}^m_{k,\phi}(u^{-2}g, u^{-1}v)$ is elliptic.*

Proof. By Proposition 5.1 and (6-4), the principal symbol of the linearization of D at $u \in \mathfrak{C} \cap \Gamma^+_k$ is

$$\frac{m+n-2}{m+n}\left(\widetilde{T}^m_{k-1,\phi} + \frac{m}{m+n-2k}\kappa v^{-1}\widetilde{T}^{m-1}_{k-2,\phi}\right),$$

where $\widetilde{T}^m_{k-1,\phi}$ and $\widetilde{T}^m_{k-2,\phi}$ are defined in terms of $(M^n, u^{-2}g, u^{-1}v, m, 0)$ with scale κ and $(M^n, u^{-2}g, u^{-1}v, m-1, 0)$ with scale $\frac{m+n-3}{m+n-2}\kappa$, respectively. Corollary 2.18 implies that $\widetilde{T}^m_{k-1,\phi} > 0$. Lemma 2.8, Corollary 2.18, and an argument as in the proof of Lemma 4.5 imply that $\widetilde{T}^m_{k-2,\phi} > 0$. This yields the conclusion. □

The form of the tensor $\widehat{E}^m_{k,\phi}$ makes it difficult to deduce a general Obata-type theorem for conformally weighted Einstein manifolds $(M^n, g, v, m, 0)$ with scale $\kappa \in \mathbb{R}_+$ for which $\hat{\sigma}^m_{k,\phi}$ is constant. In this setting, it is still the case that an integral pairing with $\widehat{E}^m_{k,\phi}$ vanishes, but it is not apparent how to deduce that (g, v) is a weighted Einstein metric-measure structure. This difficulty is even apparent in the standard conformal class of the m-weighted n-sphere, as we illustrate below:

Corollary 7.11. *Fix $k \in \mathbb{N}$ and a scale $\kappa > 0$. Let $\mathfrak{C} = [g_0, v_0]$ be the standard weighted conformal class on the m-weighted n-sphere $(S^n, m, 0)$. Suppose that $(g, v) \in \mathfrak{C}$ is a critical point of $\mathcal{Y}_k : \mathfrak{C} \to \mathbb{R}$. Then*

$$\int_{S^n} \langle \widehat{E}^m_{k,\phi}, v(P^m_\phi - \widetilde{Z}^m_\phi g) + \kappa g\rangle\, d\nu = 0. \tag{7-14}$$

Remark 7.12. When $k = 1$, one can check that if $(g, v) \in \mathfrak{C} \cap \Gamma^+_1$, then (7-14) has a sign; see [Case 2015b, Proposition 9.7]. It is unclear whether the analogous statement holds for $k \geq 2$.

Proof. Since the compactification of the flat metric-measure structure $(dx^2, 1)$ on $(\mathbb{R}^n, m, 0)$ is an element of $\mathfrak{C}$, it holds that $v^{-2}g = dx^2$. In particular, Lemma 7.1

implies that

$$v(P_\phi^m - \widetilde{Z}_\phi^m g) + \kappa g = -(m+n-2)\left(\nabla^2 v + \frac{1}{m}\langle \nabla v, \nabla\phi\rangle\, g\right). \tag{7-15}$$

Since (g, v) is a critical point of the $\mathcal{Y}_k$-functional, Proposition 5.10 implies that $\hat{\sigma}_{k,\phi}^m$ is constant. In particular, Proposition 7.9 yields that

$$\delta_\phi \widehat{E}_{k,\phi}^m - \frac{1}{m}\operatorname{tr}\widehat{E}_{k,\phi}^m\, d\phi = 0. \tag{7-16}$$

Combining (7-15) and (7-16) yields (7-14). □

Motivated both by the fully nonlinear Sobolev-type inequality known in Riemannian geometry [Guan and Wang 2004] and our discussion surrounding Conjecture 7.8, we expect the following fully nonlinear Gagliardo–Nirenberg inequality:

Conjecture 7.13. *Fix $k \in \mathbb{N}$ and let $\mathfrak{C} = [dx^2, 1]$ be the standard weighted conformal class on the m-weighted Euclidean space $(\mathbb{R}^n, m, 0)$. Then,*

$$\kappa^{-\frac{2mk(m+n-1)}{(m+n)(2m+n-2)}} \int_{\mathbb{R}^n} \tilde{\sigma}_{k,\phi}^m \geq C\left(\int_{\mathbb{R}^n} v^{-1}\, d\nu\right)^{\frac{2mk}{(m+n)(2m+n-2)}} \left(\int_{\mathbb{R}^n} d\nu\right)^{\frac{m+n-2k}{m+n}} \tag{7-17}$$

for all $(g, v) \in \mathfrak{C} \cap \Gamma_k^+$ and all $\kappa > 0$, where $\tilde{\sigma}_{k,\phi}^m$ is defined in terms of the scale κ and

$$C = \mathcal{Y}_k(g_0, v_0)$$

is the $\mathcal{Y}_k$-functional evaluated at the standard m-weighted n-sphere with scale $\kappa = m+n-2$. Moreover, equality holds in (7-17) if and only if $(\mathbb{R}^n, g, v, m, 0)$ is homothetic to the standard m-weighted n-sphere with the point $(0, \dots, 0, 1)$ removed.

Note that Del Pino and Dolbeault [2002] have proven Conjecture 7.13 in the case $k = 1$ (see [Case 2015b]).

8. Critical points of the $\mathcal{Y}$-functional

In this section we compute the linearizations of the total weighted σ_1- and σ_2-curvatures over the space of metric-measure structures with the goal of showing that weighted Einstein manifolds are among their critical points. Specifically, in the case of scale zero, we show that quasi-Einstein manifolds are critical points of the restriction $\mathcal{F}_k : \mathfrak{M}_1 \to \mathbb{R}$ of the $\mathcal{F}_k$-functional to metric-measure structures of fixed volume when $k \in \{1, 2\}$. In the case when the scale κ is positive, we show that weighted Einstein manifolds with $\mu = 0$ and scale κ are among the critical points of the $\mathcal{Y}_k$-functional $\mathcal{Y}_k : \mathfrak{M} \to \mathbb{R}$ when $k \in \{1, 2\}$.

In order to achieve our goal, we compute the linearizations of the $\mathcal{F}_k$-functionals for $k \in \{0, 1, 2\}$. This is enough to compute the linearizations of the $\mathcal{Y}_k$-functionals for $k \in \{0, 1, 2\}$. Indeed, fix a scale $\kappa \in \mathbb{R}$ and define functionals $\widetilde{\mathcal{F}}_k : \mathfrak{M} \to \mathbb{R}$ by

$$\widetilde{\mathcal{F}}_k(g, v) := \int_M \tilde{\sigma}^m_{k,\phi}\, d\nu,$$

where $\tilde{\sigma}^m_{k,\phi}$ is determined by $(g, v) \in \mathfrak{M}$ and the scale κ. When $\kappa = 0$, it holds that $\widetilde{\mathcal{F}}_k = \mathcal{F}_k$. When $\kappa > 0$,

$$\mathcal{Y}_k(g, v) = \kappa^{-\frac{2mk(m+n-1)}{(m+n)(2m+n-2)}}\, \widetilde{\mathcal{F}}_k(g, v) \left(\int_M v^{-1}\, d\nu \right)^{-\frac{2mk}{(m+n)(2m+n-2)}} \mathcal{F}_0(g, v)^{-\frac{m+n-2k}{m+n}}.$$

Below we compute the linearizations of $\widetilde{\mathcal{F}}_k$ and of $\int_M v^{-1} d\nu$ in terms of the linearizations of $\mathcal{F}_j,\ j \in \{0, 1, 2\}$.

We compute the linearizations of the $\mathcal{F}_k$-functionals by first computing the linearizations of the weighted σ_k-curvatures as functions of $\mathfrak{M}$. While this level of generality is not necessary for our computations, we include it with the expectation that it will be useful for other purposes, such as computing the second variations of the weighted σ_k-curvature functionals or computing the linearizations of the weighted σ_k-curvature functionals for larger values of k.

8A. ***The first variation of J^m_ϕ.*** We begin by computing the first variation of the weighted scalar curvature. As we illustrate below, this readily yields the first variation of integrals of powers of J^m_ϕ.

Lemma 8.1. *Let (M^n, g, v, m, μ) be a smooth metric measure space. The first variation*

$$DJ^m_\phi : T_{(g,\psi)}\mathfrak{M} \to C^\infty(M)$$

is given by

$$DJ^m_\phi[h, \psi] = -\frac{m+n-2}{2(m+n-1)} \left[\langle \operatorname{tf}_\phi P^m_\phi, h \rangle - \delta^2_\phi h + \frac{1}{m+n} \Delta_\phi \operatorname{tr} h \right] + \frac{2}{m+n} J^m_\phi \psi + \frac{m+n-2}{m+n} \Delta_\phi \psi.$$

Proof. Let $\gamma : \mathbb{R} \to \mathfrak{M}$ be a smooth curve with $\gamma(0) = (g, \phi)$ and denote

$$\gamma'(0) = (\dot{g}, -\frac{1}{m} v\dot{\phi}).$$

By [Case 2012a, (4.8)],

$$\left.\frac{\partial R^m_\phi}{\partial t}\right|_{t=0} = -\langle \operatorname{Ric}^m_\phi, \dot{g} \rangle + \delta^2_\phi \dot{g} + 2\Delta_\phi\big(\dot{\phi} - \tfrac{1}{2} \operatorname{tr}_g \dot{g}\big) - \frac{2}{m} \langle \nabla\phi, \nabla\dot{\phi} \rangle + 2(m-1)\mu v^{-2} \dot{\phi}.$$

The result then follows by using Lemma 3.1 and (4-2) to write this in terms of h, ψ, the weighted Schouten tensor, and its trace. □

An immediate consequence of (4-1b) and Lemma 8.1 is the first variation of the total weighted scalar curvature functional.

Corollary 8.2. *Let (M^n, g, v, m, μ) be a closed smooth metric measure space. Then*

$$D\left(\int_M J_\phi^m \, dv\right)[h,\psi] = -\int_M \left[\frac{m+n-2}{2(m+n-1)}\langle \operatorname{tf}_\phi P_\phi^m, h\rangle + \frac{m+n-2}{m+n} J_\phi^m \psi\right] dv.$$

Lemma 8.1 also yields the first variation of $\int \left(J_\phi^m\right)^2$.

Corollary 8.3. *Let (M^n, g, v, m, μ) be a closed smooth metric measure space. Then*

$$\begin{aligned} D\left(\int_M (J_\phi^m)^2 \, dv\right)[h,\psi] &= \frac{m+n-2}{m+n-1}\int_M \langle \operatorname{tf}_\phi(\nabla^2 J_\phi^m - J_\phi^m P_\phi^m), h\rangle \, dv \\ &\quad + \int_M \left(\frac{2(m+n-2)}{m+n}\Delta_\phi J_\phi^m - \frac{m+n-4}{m+n}(J_\phi^m)^2\right)\psi \, dv. \end{aligned}$$

Proof. Recall that Δ_ϕ is formally self-adjoint with respect to dv and that δ_ϕ is the negative of the formal adjoint with respect to dv of the Levi-Civita connection. In particular,

$$\int_M u\delta_\phi^2 h \, dv = \int_M \langle \nabla^2 u, h\rangle \, dv$$

for all $u \in C^\infty(M)$ and all $h \in \Gamma(S^2T^*M)$. The conclusion follows readily from Lemma 8.1. $\square$

8B. ***The first variation of $N_{2,\phi}^m$.*** The first step in computing the first variation of $N_{2,\phi}^m$ is to compute the first variation $DP_\phi^m : T\mathfrak{M} \to \Gamma(S^2T^*M)$ of the weighted Schouten tensor. To that end, we require some additional notation.

Given sections $A \in \Gamma(S^2\Lambda^2T^*M)$ and $T \in \Gamma(S^2T^*M)$, define $A \cdot T \in \Gamma(S^2T^*M)$ by

$$(A \cdot T)(x,y) := \langle A(\cdot, x, \cdot, y), T\rangle$$

for all $x, y \in T_pM$ and all $p \in M$. Denote by $T \sharp$ the extension of the natural action of $g^{-1}T \in \Gamma(T^*M \otimes TM)$ on vector fields to a derivation on tensor fields. In particular, given $S \in \Gamma(S^2T^*M)$, the section $T \sharp S \in \Gamma(S^2T^*M)$ is given by

$$(T \sharp S)(x,y) := -S(T(x), y) - S(x, T(y)).$$

Denote by $dT \in \Gamma(\Lambda^2T^*M \otimes T^*M)$ the twisted exterior derivative

$$dT(x,y,z) := \nabla_x T(y,z) - \nabla_y T(x,z)$$

and denote by $\delta_\phi dT \in \Gamma(T^*M \otimes T^*M)$ the composition with the weighted divergence

$$(\delta_\phi dT)(x,y) := \sum_{i=1}^n \nabla_{e_i} dT(e_i, x, y) - dT(\nabla\phi, x, y),$$

where $\{e_i\}_{i=1}^n$ is an orthonormal basis for T_pM.

Lemma 8.4. *Let (M^n, g, v, m, μ) be a smooth metric measure space. Then*

$$\begin{aligned}DP^m_\phi[h,\psi] = &-\tfrac12\delta_\phi dh+\tfrac14 L_{\delta_\phi h}g-\frac{1}{2(m+n-1)}(\delta^2_\phi h-\Delta_\phi \operatorname{tr} h)g\\ &-\frac{1}{m+n}\nabla^2\operatorname{tr} h-\tfrac12 A^m_\phi\cdot h-\frac{1}{2(m+n-2)}J^m_\phi h\\ &+\frac{1}{2(m+n-1)(m+n-2)}\langle T^m_{1,\phi},h\rangle g-\frac{m+n}{4(m+n-2)}P^m_\phi\sharp h\\ &-\frac{1}{2(m+n-2)}(\operatorname{tr} h)P^m_\phi-\frac{1}{4m}(d\phi\otimes d\phi)\sharp h+\frac{m+n-2}{m+n}\nabla^2\psi.\end{aligned}$$

Proof. Let $\gamma:\mathbb{R}\to\mathfrak{M}$ be a smooth curve such that $\gamma(0)=(g,\phi)$ and denote $\gamma'(0)=(\dot g, -\frac1m v\dot\phi)$. It follows readily from well known variational formulae for the Ricci tensor and the Hessian (cf. [Besse 1987, Section 1.K]) that

$$\begin{aligned}\frac{\partial}{\partial t}\bigg|_{t=0}\operatorname{Ric}^m_\phi = -\tfrac12\Delta_\phi\dot g+\tfrac12 L_{\delta_\phi\dot g}g-\tfrac12\left(\operatorname{Ric}^m_\phi+\frac1m d\phi\otimes d\phi\right)\sharp\,\dot g-\operatorname{Rm}\cdot\dot g\\ -\frac2m d\phi\odot d\dot\phi+\nabla^2\big(\dot\phi-\tfrac12\operatorname{tr}\dot g\big).\end{aligned}$$

It follows from (4-2) that

$$\begin{aligned}D\operatorname{Ric}^m_\phi[h,\psi] = -\tfrac12\Delta_\phi h+\tfrac12 L_{\delta_\phi h}g-\frac{1}{m+n}\nabla^2\operatorname{tr} h+\frac{1}{2(m+n)}\Delta_\phi(\operatorname{tr} h)g\\ -\tfrac12\left(\operatorname{Ric}^m_\phi+\frac1m d\phi\otimes d\phi\right)\sharp\,h-\operatorname{Rm}\cdot h+\frac{m+n-2}{m+n}\nabla^2\psi+\frac{1}{m+n}\Delta_\phi\psi\, g.\end{aligned}$$

The final conclusion follows from Lemma 8.1 and the Weitzenböck formula

$$\Delta_\phi T=\delta_\phi dT+\tfrac12 L_{\delta_\phi T}g-\operatorname{Rm}\cdot T-\tfrac12\left(\operatorname{Ric}^m_\phi+\frac1m d\phi\otimes d\phi\right)\sharp\,T$$

which holds for all $T\in\Gamma(S^2T^*M)$ (cf. [Case 2016, Lemma 5.6]). □

This allows us to compute the linearization $DY^m_\phi: T\mathfrak{M}\to C^\infty(M)$.

Corollary 8.5. *Let (M^n, g, v, m, μ) be a smooth metric measure space. Then*

$$\begin{aligned}DY^m_\phi[h,\psi] = &-\frac{m}{2(m+n-1)}(\delta^2_\phi h-\Delta_\phi\operatorname{tr} h)-\tfrac12\delta_\phi(h(\nabla\phi))-\tfrac12\langle\delta_\phi h,\nabla\phi\rangle\\ &+\frac{1}{m+n}\langle\nabla\phi,\nabla\operatorname{tr} h\rangle+\tfrac12\langle\operatorname{tr} A^m_\phi,h\rangle-\frac{1}{2m}h(\nabla\phi,\nabla\phi)\\ &+\frac{m}{2(m+n-1)(m+n-2)}\langle T^m_{1,\phi},h\rangle+\frac{m+n-4}{2(m+n)(m+n-2)}Y^m_\phi\operatorname{tr} h\\ &+\frac{2}{m+n}\psi Y^m_\phi-\frac{m+n-2}{m+n}\langle\nabla\phi,\nabla\psi\rangle.\end{aligned}$$

Proof. It follows from (4-2) and the definition of Y_ϕ^m that

$$DY_\phi^m[h,\psi] = DJ_\phi^m[h,\psi] - \operatorname{tr} DP_\phi^m[h,\psi] + \left\langle P_\phi^m, h - \frac{2}{m+n}\bigl(\psi + \tfrac{1}{2}\operatorname{tr} h\bigr)g\right\rangle.$$

The final conclusion follows from Lemmas 8.1 and 8.4. ☐

Combining Lemma 8.4 and Corollary 8.5 yields a formula for the first variation

$$DN_{2,\phi}^m : T\mathfrak{M} \to C^\infty(M).$$

To that end, recall that given $A \in \Gamma(\Lambda^2 T^*M \otimes T^*M)$ and $T \in \Gamma(S^2T^*M)$, we denote by $A \cdot T \in \Gamma(T^*M)$ the contraction

$$(A\cdot T)(x) := \sum_{i=1}^n A(e_i, x, T(e_i)).$$

Lemma 8.6. *Let (M^n, g, v, m, μ) be a smooth metric measure space. Then*

$$DN_{2,\phi}^m[h,\psi] = -\langle \Psi_{2,\phi}^m, h\rangle + \Upsilon_{2,\phi}^m\psi + \delta_\phi(\Xi_{2,\phi}^m[h,\psi]),$$

where

$$\begin{aligned}
\Psi_{2,\phi}^m &= \operatorname{tf}_\phi\Bigl(B_\phi^m + \frac{m+n-4}{m+n-2}(P_\phi^m)^2 - \frac{m+n-2}{m+n-1}\nabla^2 J_\phi^m \\
&\qquad\qquad + \frac{m+n}{(m+n-1)(m+n-2)} J_\phi^m P_\phi^m\Bigr),\\
\Upsilon_{2,\phi}^m &= \frac{2(m+n-2)}{m+n}\Delta_\phi J_\phi^m + \frac{4}{m+n} N_{2,\phi}^m,\\
\Xi_{2,\phi}^m[h,\psi] &= dh\cdot P_\phi^m - dP_\phi^m\cdot h + P_\phi^m(\delta_\phi h) - \frac{2}{m+n}P_\phi^m(\nabla \operatorname{tr} h) - \frac{1}{m}Y_\phi^m h(\nabla\phi)\\
&\quad - \frac{m+n-2}{m+n-1}h(\nabla J_\phi^m) + \frac{m+n-2}{(m+n)(m+n-1)}(\operatorname{tr} h)dJ_\phi^m\\
&\quad + \frac{1}{m+n-1}(J_\phi^m(d\operatorname{tr} h - \delta_\phi h)) + \frac{2(m+n-2)}{m+n}(P_\phi^m(\nabla\psi) - \psi\, dJ_\phi^m).
\end{aligned}$$

Proof. From the definition of $N_{2,\phi}^m$ we see that

$$\begin{aligned}
DN_{2,\phi}^m[h,\psi] &= 2\langle P_\phi^m, DP_\phi^m[h,\psi]\rangle + \frac{2}{m}Y_\phi^m\, DY_\phi^m[h,\psi]\\
&\qquad - 2\left\langle (P_\phi^m)^2, h - \frac{2}{m+n}\bigl(\psi + \tfrac{1}{2}\operatorname{tr} h\bigr)g\right\rangle.
\end{aligned}$$

Lemma 3.2 implies that

$$\left\langle\left(P_\phi^m - \frac{Y_\phi^m}{m}g\right)(\nabla\phi), h(\nabla\phi)\right\rangle - Y_\phi^m\delta_\phi(h(\nabla\phi)) = \langle \operatorname{tr} dP_\phi^m \otimes d\phi, h\rangle - \delta_\phi(Y_\phi^m h(\nabla\phi)).$$

Lemma 3.2 also implies that

$$\begin{aligned}\tfrac{1}{2}\langle P_\phi^m, L_{\delta_\phi h} g\rangle - \frac{1}{m}\langle Y_\phi^m \nabla\phi, \delta_\phi h\rangle &= \langle \nabla^2 J_\phi^m, h\rangle + \delta_\phi(P_\phi^m(\delta_\phi h) - h(\nabla J_\phi^m)),\\ \langle P_\phi^m, \nabla^2 \operatorname{tr} h\rangle - \frac{1}{m}\langle Y_\phi^m\nabla\phi, \nabla \operatorname{tr} h\rangle &= (\operatorname{tr} h)\Delta_\phi J_\phi^m + \delta_\phi(P_\phi^m(\nabla \operatorname{tr} h) - (\operatorname{tr} h) dJ_\phi^m),\\ \langle P_\phi^m, \nabla^2\psi\rangle - \frac{1}{m}\langle Y_\phi^m\nabla\phi, \nabla\psi\rangle &= \psi\,\Delta_\phi J_\phi^m + \delta_\phi(P_\phi^m(\nabla\psi) - \psi\, dJ_\phi^m).\end{aligned}$$

Finally, straightforward computations yield

$$\begin{aligned}\langle \delta_\phi dh, P_\phi^m\rangle &= \langle \delta_\phi d P_\phi^m, h\rangle - \delta_\phi(dh\cdot P_\phi^m - dP_\phi^m\cdot h),\\ J_\phi^m(\delta_\phi^2 h - \Delta_\phi \operatorname{tr} h) &= \langle \nabla^2 J_\phi^m - \Delta_\phi J_\phi^m\, g, h\rangle\\ &\qquad + \delta_\phi\big(J_\phi^m(\delta_\phi h - d\operatorname{tr} h) + (\operatorname{tr} h) dJ_\phi^m - h(\nabla J_\phi^m)\big).\end{aligned}$$

Combined with Lemma 8.4 and Corollary 8.5, these yield the desired result. □

An immediate consequence of (4-1b) and Lemma 8.6 is the first variation of the total $N_{2,\phi}^m$-curvature functional.

Corollary 8.7. *Let (M^n, g, v, m, μ) be a smooth metric measure space. Then*

$$\begin{aligned} D\Big(\int_M N_{2,\phi}^m\, d\nu\Big)[h,\psi] = &-\int_M \langle \Psi_{2,\phi}^m, h\rangle\, d\nu\\ &+ \int_M \Big(\frac{2(m+n-2)}{m+n}\Delta_\phi J_\phi^m - \frac{m+n-4}{m+n} N_{2,\phi}^m\Big)\psi\, d\nu.\end{aligned}$$

An immediate consequence of Corollaries 8.3 and 8.7 is the first variation of the total σ_2-curvature functional $\mathcal{F}_2$.

Corollary 8.8. *Let (M^n, g, v, m, μ) be a smooth metric measure space. Then*

$$D\mathcal{F}_2[h,\psi] = \tfrac{1}{2}\int_M \langle E_{2,\phi}^m, h\rangle\, d\nu - \frac{m+n-4}{m+n}\int_M \sigma_{2,\phi}^m \psi\, d\nu,$$

where

$$E_{2,\phi}^m := \operatorname{tf}_\phi\Big(B_\phi^m + \frac{m+n-4}{m+n-2} T_{2,\phi}^m\Big).$$

8C. ***The first variation of the $\mathcal{Y}$-functionals.*** Corollaries 8.2 and 8.8 give formulae for the linearizations of the $\mathcal{F}_1$- and $\mathcal{F}_2$-functionals. In particular, combining these results with Proposition 3.15 immediately shows that quasi-Einstein manifolds are critical points of the volume-normalized $\mathcal{F}_k$-functionals for $k \in \{1, 2\}$.

Proposition 8.9. *Let (M^n, g, v, m, μ) be a closed quasi-Einstein manifold. Assume that $(g, v) \in \mathfrak{M}_1$. Then (g, v) is a critical point of $\mathcal{F}_k : \mathfrak{M}_1 \to \mathbb{R}$ for $k \in \{1, 2\}$. Moreover, if (g, v) is a critical point of $\mathcal{F}_1 : \mathfrak{M}_1 \to \mathbb{R}$, then it is a quasi-Einstein metric-measure structure.*

Corollaries 8.2 and 8.8 also provide key ingredients for proving that weighted Einstein manifolds with $\mu = 0$ and positive scale are critical points of the $\mathcal{Y}_k$-functionals for $k \in \{1, 2\}$. Indeed, combining these results with Corollary 4.3 and Lemma 4.4 allows us to compute the linearizations of the $\widetilde{\mathcal{F}}_k$-functionals. We begin by explaining the relevance of Corollary 4.3. Given a closed weighted manifold (M^n, m, μ) and an integer $k \leq m$, denote by $\mathcal{F}_0^{(m-k)}, \mathcal{F}_1^{(m-1)} : \mathfrak{M} \to \mathbb{R}$ the functionals

$$\mathcal{F}_0^{(m-k)}(g, v) := \int_M d\nu^{(m-k)},$$
$$\mathcal{F}_1^{(m-1)}(g, v) := \int_M \sigma_{1,\phi}^{m-1}\, d\nu^{(m-1)},$$

where $d\nu^{(m-k)}$ and $\sigma_{1,\phi}^{m-1}$ are the weighted volume element on $\mathfrak{M}(M, m-k, \mu)$ and the weighted scalar curvature on $\mathfrak{M}(M, m-1, \mu)$, respectively. The linearizations

$$D^{(m-k)}\mathcal{F}_0^{(m-k)} : T\mathfrak{M}(M, m-k, \mu) \to \mathbb{R},$$
$$D^{(m-1)}\mathcal{F}_1^{(m-1)} : T\mathfrak{M}(M, m-1, \mu) \to \mathbb{R}$$

are computed from (4-1) and Corollary 8.2, respectively. Applying Corollary 4.3 yields the following result.

Proposition 8.10. *Let (M^n, g, v, m, μ) be a closed smooth metric measure space and let $k \in \mathbb{R}$. Regard $\mathcal{F}_0^{(m-k)}$ and $\mathcal{F}_1^{(m-1)}$ as functionals on $\mathfrak{M}(M, m, \mu)$. Then*

$$D\mathcal{F}_0^{(m-k)}[h, \psi] = \frac{k}{2(m+n)} \int_M \langle v^{-k} g, h\rangle\, d\nu^{(m)} - \frac{m+n-k}{m+n} \int_M \psi v^{-k}\, d\nu^{(m)},$$

$$\begin{aligned} D\mathcal{F}_1^{(m-1)}[h, \psi] = \frac{m+n-3}{2(m+n-2)} \int_M v^{-1} \left\langle T_{1,\phi}^{m-1} - \frac{m+n-2}{m+n} \sigma_{1,\phi}^{m-1} g, h \right\rangle d\nu^{(m)} \\ - \frac{m+n-3}{m+n} \int_M \psi v^{-1} \sigma_{1,\phi}^{m-1}\, d\nu^{(m)}. \end{aligned}$$

Proof. Equation (4-1) implies that

$$D^{(m-k)}\mathcal{F}_0^{(m-k)}[h, \psi] = -\int_M \psi\, d\nu^{(m-k)}.$$

Corollary 4.3 then yields the formula for $D\mathcal{F}_0^{(m-k)}$. On the other hand, Corollary 8.2 implies that

$$\begin{aligned} D^{(m-1)}\mathcal{F}_1^{(m-1)}[h, \psi] = -\frac{m+n-3}{m+n-1} \int_M \sigma_{1,\phi}^{m-1} \psi\, d\nu^{(m-1)} \\ + \frac{m+n-3}{2(m+n-2)} \int_M \left\langle T_{1,\phi}^{m-1} - \frac{m+n-2}{m+n-1} \sigma_{1,\phi}^{m-1} g, h \right\rangle d\nu^{(m-1)}. \end{aligned}$$

Corollary 4.3 then yields the formula for $D\mathcal{F}_1^{(m-1)}$. □

Given a weighted manifold (M^n, m, μ) and a positive integer $j \leq m$, Proposition 8.10 motivates the definitions

$$\operatorname{tf}_\phi T^{m-j}_{k-j,\phi} := T^{m-j}_{k-j,\phi} - \frac{m+n-k}{m+n}\sigma^{m-j}_{k-j,\phi} g$$

on $\mathfrak{M}(M, m, \mu)$.

By Lemma 4.4 and Proposition 8.10, one can compute the linearization of $\widetilde{\mathcal{F}}_k$, $k \in \{0, 1, 2\}$, in terms of the linearizations of $\mathcal{F}^{(m-j)}_{(k-j)}$, $0 \leq j \leq k$. We begin by considering the $\widetilde{\mathcal{F}}_1$-functional, noting in particular that the critical points of the $\mathcal{Y}_1$-functional are exactly weighted Einstein manifolds.

Theorem 8.11. *Fix $\kappa \in \mathbb{R}_+$ and let (M^n, g, v, m, μ) be a smooth metric measure space. For every $(h, \psi) \in T_{(g,v)}\mathfrak{M}$,*

$$D\widetilde{\mathcal{F}}_1[h, \psi] = -\frac{m+n-2}{m+n}\int_M \left(\tilde{\sigma}^m_{1,\phi} + \frac{m}{m+n-2}\kappa v^{-1}\right)\psi \, d\nu + \frac{m+n-2}{2(m+n-1)}\int_M \left\langle \widetilde{E}^m_{1,\phi} + \frac{m}{m+n-2}\kappa v^{-1}\widetilde{U}^m_{0,\phi}, h\right\rangle d\nu, \tag{8-1}$$

where $U^m_{0,\phi}$ is as in (7-7). In particular, $(M^n, g, v, m, 0)$ is a critical point of $\mathcal{Y}_1 : \mathfrak{M} \to \mathbb{R}$ if and only if it is a weighted Einstein manifold with scale κ.

Proof. Lemma 4.4, Corollary 8.2, and Proposition 8.10 immediately yield (8-1). Combining Proposition 8.10 and (8-1), we see that $D\mathcal{Y}_1 : T_{(g,v)}\mathfrak{M} \to \mathbb{R}$ vanishes if and only if

$$\widetilde{E}^m_{1,\phi} + \frac{m}{m+n-2}\kappa v^{-1}\widetilde{U}^m_{0,\phi} = \frac{2m(m+n-1)}{(m+n)^2(m+n-2)(2m+n-2)}\left(\frac{\int \tilde{\sigma}^m_{1,\phi}}{\int v^{-1}}\right)v^{-1}g,$$

$$\tilde{\sigma}^m_{1,\phi} + \frac{m}{m+n-2}\kappa v^{-1} = \frac{2m(m+n-1)}{(m+n)(m+n-2)(2m+n-2)}\left(\frac{\int \tilde{\sigma}^m_{1,\phi}}{\int v^{-1}}\right)v^{-1} + \frac{\int \tilde{\sigma}^m_{1,\phi}}{\int 1}.$$

From the definitions of $\widetilde{E}^m_{1,\phi}$ and $\widetilde{U}^m_{0,\phi}$, we conclude that these two conditions are equivalent to

$$P^m_\phi = \frac{1}{m+n}\left(\frac{\int \tilde{\sigma}^m_{1,\phi}\, d\nu}{\int d\nu}\right)g, \tag{8-2}$$

$$\hat{\sigma}^m_{1,\phi} = \frac{\int \tilde{\sigma}^m_{1,\phi}\, d\nu}{\int d\nu}, \tag{8-3}$$

where $\hat{\sigma}^m_{1,\phi}$ is as in (7-8). Equations (8-2) and (8-3) are clearly equivalent to the condition that $(M^n, g, v, m, 0)$ is a weighted Einstein manifold with scale κ. $\square$

We next consider the $\widetilde{\mathcal{F}}_2$- and $\mathcal{Y}_2$-functionals. Here the relationship to weighted Einstein manifolds is more subtle. First, the Euler equation for the $\widetilde{\mathcal{F}}_2$-functional is fourth-order in the metric, so one cannot expect to characterize the critical points of the $\mathcal{Y}_2$-functional as weighted Einstein manifolds. Second, the fact that weighted Einstein manifolds with $\mu = 0$ and scale κ are critical points for the $\mathcal{Y}_2$-functional depends on the subtle cancellation (3-14) for the weighted Bach tensor of such manifolds. This latter point is discussed further in Remark 8.13.

Theorem 8.12. *Fix $\kappa \in \mathbb{R}_+$ and let (M^n, g, v, m, μ) be a smooth metric measure space. For every $(h, \psi) \in T_{(g,v)}\mathfrak{M}$,*

$$D\widetilde{\mathcal{F}}_2[h,\psi] = -\frac{m+n-4}{m+n}\int_M\left(\tilde{\sigma}_{2,\phi}^m + \frac{m}{m+n-4}\kappa v^{-1}\tilde{s}_{1,\phi}^m\right)\psi\,dv + \frac{m+n-4}{2(m+n-2)} \times \int_M\left\langle \widetilde{E}_{2,\phi}^m + \frac{m+n-2}{m+n-4}\widetilde{B}_\phi^m + \frac{m}{m+n-4}\kappa v^{-1}\widetilde{U}_{1,\phi}^m, h\right\rangle dv, \tag{8-4}$$

where $\widetilde{U}_{1,\phi}^m$ is as in (7-7) and $\widetilde{B}_\phi^m$ is the Bach tensor with scale κ,

$$\widetilde{B}_\phi^m := \delta_\phi d P_\phi^m - \frac{1}{m}(\operatorname{tr} dP_\phi^m)\otimes d\phi + A_\phi^m\cdot(P_\phi^m - \widetilde{Z}_\phi^m g).$$

In particular, if $(M^n, g, v, m, 0)$ is a weighted Einstein manifold with scale κ, then it is a critical point of $\mathcal{Y}_2 : \mathfrak{M} \to \mathbb{R}$.

Proof. Lemma 4.4, Corollary 8.8, and Proposition 8.10 immediately yield (8-4). Combining Proposition 8.10 and (8-4) implies that $D\mathcal{Y}_2 : T_{(g,v)}\mathfrak{M} \to \mathbb{R}$ vanishes if and only if

$$\widetilde{E}_{2,\phi}^m + \frac{m+n-2}{m+n-4}\widetilde{B}_\phi^m + \frac{m}{m+n-4}\kappa v^{-1}\widetilde{U}_{1,\phi}^m = \frac{4m(m+n-2)}{(m+n)^2(m+n-4)(2m+n-2)}\left(\frac{\int\tilde{\sigma}_{2,\phi}^m}{\int v^{-1}}\right)v^{-1}g$$

and

$$\tilde{\sigma}_{2,\phi}^m + \frac{m}{m+n-4}\kappa v^{-1}\tilde{s}_{1,\phi}^m = \frac{\int\tilde{\sigma}_{2,\phi}^m}{\int 1} + \frac{4m(m+n-1)}{(m+n)(m+n-4)(2m+n-2)}\left(\frac{\int\tilde{\sigma}_{2,\phi}^m}{\int v^{-1}}\right)v^{-1}.$$

It is clear that if $(M^n, g, v, m, 0)$ is a weighted Einstein manifold with scale κ, then $\widetilde{B}_\phi^m = 0$. It then follows from Propositions 3.15 and 3.16 that $D\mathcal{Y}_2 : T_{(g,v)}\mathfrak{M} \to \mathbb{R}$ vanishes for such a manifold. □

Remark 8.13. A key point in the proof of Theorem 8.12 is that, due to the identity

$$\widetilde{B}_\phi^m = B_\phi^m - m\kappa v^{-1}\left(P_\phi^{m-1} - \frac{m+n-3}{m+n-2}P_\phi^m\right),$$

the summands B_ϕ^m and P_ϕ^{m-1} in the formulae for $D\mathcal{F}_2$ and $D\mathcal{F}_1^{(m-1)}$, respectively, exactly combine to yield the weighted Bach tensor with scale κ. In particular, since P_ϕ^{m-1} need not be constant for a weighted Einstein manifold $(M^n, g, v, m, 0)$, if we add suitable multiples of $\mathcal{F}_1^{(m-1)}$ and $\mathcal{F}_0^{(m-2)}$ to $\mathcal{F}_2$ to obtain a functional with Euler equation given by $\tilde{\sigma}_{2,\phi}^m$, the resulting functional need not admit weighted Einstein manifolds among its critical points. This is the final justification for our focus on the $\widetilde{\mathcal{F}}_2$- and $\mathcal{Y}_2$-functionals.

Remark 8.14. It follows from Lemma 5.6 that

$$D\mathcal{Y}_2[g,0] = -\frac{2m+n}{2} D\mathcal{Y}_2[0,1].$$

Using (8-4), we conclude that if $(M^n, g, v, m, 0)$ is a closed smooth metric measure space such that (g, v) is locally conformally flat in the weighted sense and a critical point of $\mathcal{Y}_2 : \mathfrak{C} \to \mathbb{R}$, then

$$\int_M \operatorname{tr} \widehat{E}_{2,\phi}^m = 0.$$

This observation further simplifies (7-14) in the case $k = 2$. It also suggests that if $(M^n, g, v, m, 0)$ is a closed smooth metric measure space such that (g, v) is locally conformally flat in the weighted sense and a critical point of $\mathcal{Y}_k : \mathfrak{C} \to \mathbb{R}$, then $\int \operatorname{tr} \widehat{E}_{k,\phi}^m = 0$.

Acknowledgments

This work was begun during a month-long visit to the Centre de Recerca Matemàtica. The author would like to thank the CRM for providing a productive research environment. He was also partially supported by NSF Grant DMS-1004394.

References

[Bakry and Qian 2000] D. Bakry and Z. Qian, "Some new results on eigenvectors via dimension, diameter, and Ricci curvature", *Adv. Math.* **155**:1 (2000), 98–153. MR Zbl

[Besse 1987] A. L. Besse, *Einstein manifolds*, Ergebnisse der Mathematik (3) **10**, Springer, 1987. MR Zbl

[Bourguignon 1986] J.-P. Bourguignon, "Invariants intégraux fonctionnels pour des équations aux dérivées partielles d'origine géométrique", pp. 100–108 in *Differential geometry* (Peñíscola, Spain, 1985), edited by A. M. Naveira et al., Lecture Notes in Math. **1209**, Springer, 1986. MR Zbl

[Branson and Gover 2008] T. P. Branson and A. R. Gover, "Variational status of a class of fully nonlinear curvature prescription problems", *Calc. Var. Partial Differential Equations* **32**:2 (2008), 253–262. MR Zbl

[Caffarelli et al. 1985] L. Caffarelli, L. Nirenberg, and J. Spruck, "The Dirichlet problem for nonlinear second-order elliptic equations, III: Functions of the eigenvalues of the Hessian", *Acta Math.* **155**:3-4 (1985), 261–301. MR Zbl

[Cao and Chen 2013] H.-D. Cao and Q. Chen, "On Bach-flat gradient shrinking Ricci solitons", *Duke Math. J.* **162**:6 (2013), 1149–1169. MR Zbl

[Carlen et al. 2010] E. A. Carlen, J. A. Carrillo, and M. Loss, "Hardy–Littlewood–Sobolev inequalities via fast diffusion flows", *Proc. Natl. Acad. Sci. USA* **107**:46 (2010), 19696–19701. MR Zbl

[Case 2012a] J. S. Case, "Smooth metric measure spaces and quasi-Einstein metrics", *Int. J. Math.* **23**:10 (2012), art. id. 1250110. MR Zbl

[Case 2012b] J. S. Case, "Smooth metric measure spaces, quasi-Einstein metrics, and tractors", *Cent. Eur. J. Math.* **10**:5 (2012), 1733–1762. MR Zbl

[Case 2013a] J. S. Case, "Conformal invariants measuring the best constants for Gagliardo–Nirenberg–Sobolev inequalities", *Calc. Var. Partial Differential Equations* **48**:3-4 (2013), 507–526. MR Zbl

[Case 2013b] J. S. Case, "Sharp metric obstructions for quasi-Einstein metrics", *J. Geom. Phys.* **64** (2013), 12–30. MR Zbl

[Case 2015a] J. S. Case, "The energy of a smooth metric measure space and applications", *J. Geom. Anal.* **25**:1 (2015), 616–667. MR Zbl

[Case 2015b] J. S. Case, "A Yamabe-type problem on smooth metric measure spaces", *J. Differential Geom.* **101**:3 (2015), 467–505. MR Zbl

[Case 2016] J. S. Case, "A notion of the weighted σ_k-curvature for manifolds with density", *Adv. Math.* **295** (2016), 150–194. MR Zbl

[Case et al. 2011] J. Case, Y.-J. Shu, and G. Wei, "Rigidity of quasi-Einstein metrics", *Differential Geom. Appl.* **29**:1 (2011), 93–100. MR Zbl

[Catino et al. 2017] G. Catino, P. Mastrolia, D. D. Monticelli, and M. Rigoli, "On the geometry of gradient Einstein-type manifolds", *Pacific J. Math.* **286**:1 (2017), 39–67. MR Zbl

[Chang et al. 2003] S.-Y. A. Chang, M. J. Gursky, and P. C. Yang, "Entire solutions of a fully nonlinear equation", pp. 43–60 in *Lectures on partial differential equations*, edited by S.-Y. A. Chang et al., New Stud. Adv. Math. **2**, Int. Press, Somerville, MA, 2003. MR Zbl

[Cheeger and Colding 1996] J. Cheeger and T. H. Colding, "Lower bounds on Ricci curvature and the almost rigidity of warped products", *Ann. of Math.* (2) **144**:1 (1996), 189–237. MR Zbl

[Chen and He 2013] Q. Chen and C. He, "On Bach flat warped product Einstein manifolds", *Pacific J. Math.* **265**:2 (2013), 313–326. MR Zbl

[Del Pino and Dolbeault 2002] M. Del Pino and J. Dolbeault, "Best constants for Gagliardo–Nirenberg inequalities and applications to nonlinear diffusions", *J. Math. Pures Appl.* (9) **81**:9 (2002), 847–875. MR Zbl

[Gårding 1959] L. Gårding, "An inequality for hyperbolic polynomials", *J. Math. Mech.* **8** (1959), 957–965. MR Zbl

[González 2006] M. d. M. González, "Classification of singularities for a subcritical fully nonlinear problem", *Pacific J. Math.* **226**:1 (2006), 83–102. MR Zbl

[Guan and Wang 2003] P. Guan and G. Wang, "A fully nonlinear conformal flow on locally conformally flat manifolds", *J. Reine Angew. Math.* **557** (2003), 219–238. MR Zbl

[Guan and Wang 2004] P. Guan and G. Wang, "Geometric inequalities on locally conformally flat manifolds", *Duke Math. J.* **124**:1 (2004), 177–212. MR Zbl

[Guan et al. 2003] P. Guan, J. Viaclovsky, and G. Wang, "Some properties of the Schouten tensor and applications to conformal geometry", *Trans. Amer. Math. Soc.* **355**:3 (2003), 925–933. MR Zbl

[Gursky and Viaclovsky 2001] M. J. Gursky and J. A. Viaclovsky, "A new variational characterization of three-dimensional space forms", *Invent. Math.* **145**:2 (2001), 251–278. MR Zbl

[Hardy et al. 1934] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge Univ. Press, 1934.

[He et al. 2012] C. He, P. Petersen, and W. Wylie, "On the classification of warped product Einstein metrics", *Comm. Anal. Geom.* **20**:2 (2012), 271–311. MR Zbl

[Huang and Wei 2013] G. Huang and Y. Wei, "The classification of (m, ρ)-quasi-Einstein manifolds", *Ann. Global Anal. Geom.* **44**:3 (2013), 269–282. MR Zbl

[Kim and Kim 2003] D.-S. Kim and Y. H. Kim, "Compact Einstein warped product spaces with nonpositive scalar curvature", *Proc. Amer. Math. Soc.* **131**:8 (2003), 2573–2576. MR Zbl

[Lee and Parker 1987] J. M. Lee and T. H. Parker, "The Yamabe problem", *Bull. Amer. Math. Soc.* (*N.S.*) **17**:1 (1987), 37–91. MR Zbl

[Li and Li 2003] A. Li and Y. Li, "On some conformally invariant fully nonlinear equations", *Comm. Pure Appl. Math.* **56**:10 (2003), 1416–1464. MR Zbl

[Lichnerowicz 1958] A. Lichnerowicz, *Géométrie des groupes de transformations*, Travaux et Recherches Math. **3**, Dunod, Paris, 1958. MR Zbl

[Lott 2007] J. Lott, "Remark about scalar curvature and Riemannian submersions", *Proc. Amer. Math. Soc.* **135**:10 (2007), 3375–3381. MR Zbl

[Obata 1962] M. Obata, "Certain conditions for a Riemannian manifold to be isometric with a sphere", *J. Math. Soc. Japan* **14** (1962), 333–340. MR Zbl

[Obata 1971] M. Obata, "The conjectures on conformal transformations of Riemannian manifolds", *J. Differential Geom.* **6** (1971), 247–258. MR Zbl

[Perelman 2002] G. Perelman, "The entropy formula for the Ricci flow and its geometric applications", preprint, 2002. Zbl arXiv

[Sheng et al. 2007] W.-M. Sheng, N. S. Trudinger, and X.-J. Wang, "The Yamabe problem for higher order curvatures", *J. Differential Geom.* **77**:3 (2007), 515–553. MR Zbl

[Viaclovsky 2000] J. A. Viaclovsky, "Conformal geometry, contact geometry, and the calculus of variations", *Duke Math. J.* **101**:2 (2000), 283–316. MR Zbl

[Wei and Wylie 2009] G. Wei and W. Wylie, "Comparison geometry for the Bakry–Emery Ricci tensor", *J. Differential Geom.* **83**:2 (2009), 377–405. MR Zbl

Received January 12, 2017. Revised June 3, 2018.

JEFFREY S. CASE
MATHEMATICS DEPARTMENT
PENN STATE UNIVERSITY
UNIVERSITY PARK, PA
UNITED STATES
jscase@psu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 2, 2019
dx.doi.org/10.2140/pjm.2019.299.401

UNIQUENESS OF TANGENT CONES FOR BIHARMONIC MAPS WITH ISOLATED SINGULARITIES

YOUMIN CHEN AND HAO YIN

We study the problem of uniqueness of a tangent cone for minimizing extrinsic biharmonic maps. Following the celebrated result of Simon, we prove that if the target manifold is a compact analytic submanifold in $\mathbb{R}^p$ and if there is one tangent map whose singularity set consists of the origin only, then this tangent map is unique.

1. Introduction

In this paper, we prove the biharmonic map version of the celebrated result of Simon [1983]. Here we restrict ourselves to the case of extrinsic biharmonic maps. Let $B \subset \mathbb{R}^m$ be the unit ball around the origin and N be a closed Riemannian manifold isometrically embedded in $\mathbb{R}^p$. A map $u \in W^{2,2}(B, \mathbb{R}^p)$ into N is called a (extrinsic) biharmonic map if and only if it is the critical point of the energy

$$E(u) = \int_B |\Delta u|^2 \, dx. \tag{1}$$

It is called a minimizing (biharmonic) map if for any $B_r(x) \subset B$ and $W^{2,2}$ maps v with $v \equiv u$ on $B \setminus B_r(x)$, we have

$$E(v) \geq E(u).$$

Since the pioneering work of Chang, Wang, and Yang [Chang et al. 1999], many authors studied the regularity problem of biharmonic maps; see [Strzelecki 2003; Wang 2004a; 2004b; 2004c; Hong and Wang 2005; Lamm and Rivière 2008]. Roughly speaking, stationary biharmonic maps are regular away from a singularity set of codimension 4. For minimizing maps, one expects better regularity since it was proved by Schoen and Uhlenbeck [1982] that minimizing harmonic maps are regular away from a singularity of codimension 3. Moreover, Luckhaus [1988] proved the compactness of minimizing harmonic maps using a lemma which was later named after him. This compactness is crucial to the theory of singularity set

The research work of Yin is supported by NSFC 11471300.
MSC2010: primary 58E20; secondary 35J30.
Keywords: biharmonic map, tangent map.

of minimizing harmonic maps. We refer the readers to the book of Simon [1996] for a nice presentation of this deep theory. The limit of a sequence of minimizing biharmonic maps was studied by Scheven [2008]. Instead of proving the direct analogue of Luckhaus lemma, the author studied the defect measure after Lin [1999]. In particular, it was shown that the limit is a stationary biharmonic map, which implies that the singularity set of minimizing biharmonic maps is of codimension 5. The interesting problem of whether this limit is minimizing remains open.

Thanks to the result of Scheven, we may study the tangent map at a singular point of a minimizing biharmonic map. The problem of uniqueness of such tangent maps is usually very difficult. Simon [1983] set up a general framework to attack such problem under a set of assumptions. The argument has been adapted to many different problems, for example, to minimal submanifolds [Simon 1983], Yang–Mills fields [Yang 2003], and Einstein metrics [Colding and Minicozzi 2014; Cheeger and Tian 1994]. To the best of our knowledge, all such generalizations are about the isolated singularity of solutions to some second order partial differential equation. It is the purpose of this paper to show that this argument also works in the case of the fourth-order problem. More precisely, we prove

Theorem 1.1. *Suppose N is an analytic submanifold of the Euclidean space $\mathbb{R}^p$ and $u : B \to N$ is a minimizing biharmonic map (with finite energy), where $B \subset \mathbb{R}^m$ $(m \geq 5)$ is the unit ball. If 0 is a singularity of u and one of the tangent maps of u at 0 is of the form $\varphi(x/|x|)$ for some smooth $\varphi : S^{m-1} \to N$, then this tangent map is the unique tangent map at 0.*

Suppose that (r, θ) is the polar coordinates in B and that $t = -\log r$. Then the theorem claims that $\lim_{t\to\infty} u(t)$ exists (and therefore is unique). As is well known, this is related to an estimate on the speed of convergence of $\partial_t u$ to zero when $t \to +\infty$. It is not hard to derive from the monotonicity formula (see [Scheven 2008, (2.4)] and (38)) that

$$\int_1^{+\infty} \|\partial_t u\|^2_{L^2(S^{m-1})} < +\infty. \tag{2}$$

Here S^{m-1} is the unit sphere in $\mathbb{R}^m$. If we can show

$$\int_1^{+\infty} \|\partial_t u\|_{L^2(S^{m-1})} < +\infty, \tag{3}$$

then we know at least $u(t)$ converges to a unique limit in the sense of L^2 norm. However, in general, (3) does not follow from (2).

Simon [1983] observed that an infinite-dimensional version of the Lojasiewicz inequality is helpful here. In the case of a harmonic map, $u(t)$ is regarded as a family of maps from S^{m-1} into N evolving by some second-order (abstract) ODE:

$$u'' - (m-2)u' = \nabla \mathcal{E}_{S^{m-1}}(v) + R, \tag{4}$$

where $u' = \partial_t u$, $\mathcal{E}_{S^{m-1}}$ is the harmonic map energy on S^{m-1} and R is some small perturbation term. A stationary point of this ODE (i.e., a solution independent of t) is the smooth map φ in the assumptions of the theorem (in the harmonic map case). With the help of the Lojasiewicz inequality, he studied the dynamics of this second-order ODE in a small neighborhood of φ. More precisely, he proved [Simon 1996, Chapter 3]

$$\left(\int_{t+1}^{+\infty} \|\partial_t u\|^2_{L^2(S^{m-1})}\right)^{2-\alpha} \le C \int_{t-1}^{t+1} \|\partial_t u\|^2_{L^2(S^{m-1})} \tag{5}$$

for any t and some $\alpha \in (0,1)$. This amounts to (up to technical issue) an ordinary differential inequality of $h(t) = \int_t^{+\infty} \|\partial_t u\|^2_{L^2(S^{m-1})}$,

$$h(t)^{2-\alpha} \le C(-h'(t)).$$

From this inequality, it is easy to derive some decay estimate that implies (3).

To generalize this argument to the biharmonic map case, we found that the Lojasiewicz inequality is not a problem because it is a general property of analytic functions, and the Lyapunov–Schmidt reduction works as long as the gradient of the functional is elliptic. The difficulty is to find the correct counterpart of (5). We will eventually prove a discrete version of ordinary differential inequality with time delay (see (44)). Fortunately, we can still derive the decay estimate we need from it.

The paper is organized as follows. We recall some basic properties of biharmonic maps in Section 2. In particular, we prove an improved ε-regularity lemma of Schoen and Uhlenbeck type (see [Schoen and Uhlenbeck 1982, Proposition 4.5]). In Section 3, we prove the Lojasiewicz inequality (following [Simon 1996]). Section 4 is the most important part of this paper, which contains the derivation of our analogue of (5). Finally, we give the proof of Theorem 1.1 in Section 5 following the framework of Simon [1996].

2. Preliminaries on biharmonic maps

In this section, we collect a few results, mainly PDE estimates, that are needed for the proof of our main theorem.

We start by introducing the Euler–Lagrange equation for extrinsic biharmonic energy $E(u)$ (see [Wang 2004a, Proposition 2.2]),

$$\Delta^2 u = \Delta(A(u)(\nabla u, \nabla u)) + 2\nabla \cdot \langle \Delta u, \nabla(P(u))\rangle - \langle \Delta(P(u)), \Delta u\rangle. \tag{6}$$

Here A is the second fundamental form of N in $\mathbb{R}^p$ and $P(u)$ is the projection from $\mathbb{R}^p$ to $T_u N$. When u is a smooth (extrinsic) biharmonic map, this is equivalent to the statement that $\Delta^2 u$ is perpendicular to $T_u N$ in $\mathbb{R}^p$. Often in the following discussion, this simpler form is good enough.

An improved ε-regularity. The famous ε-regularity theorem for stationary harmonic maps requires that the (rescaled) energy is small on a ball. It has a biharmonic map analogue as follows:

Lemma 2.1 [Wang 2004a; Struwe 2008; Scheven 2008]. *There exist $\varepsilon_1 > 0$ and constants $C(k)$ only depending on N such that if u is a stationary* (*extrinsic*) *biharmonic map on $B_r(x) \subset \mathbb{R}^m (m \geq 5)$ satisfying*

$$r^{4-m} \int_{B_r(x)} (|\nabla^2 u|^2 + r^{-2} |\nabla u|^2)\, dx \leq \varepsilon_1, \tag{7}$$

then

$$\sup_{B_{r/2}(x)} r^k |\nabla^k u| \leq C(k) \quad \forall k \in \mathbb{N}.$$

Remark 2.2. Here and throughout the paper, $B_r(x)$ means the ball of radius r centered at x, which is usually omitted if $x = 0$. Also the subscript r is omitted if $r = 1$.

For minimizing harmonic maps, this result can be improved in the sense that a smallness condition on $⨍_{B_r(x)} |u - u^*|^2\, dx$ replaces (7), where u^* is the average of u on $B_r(x)$ (see [Schoen and Uhlenbeck 1982, Proposition 4.5]). The improved version plays an important role in the analysis of minimal tangent maps and the uniqueness of tangent cones of harmonic maps. Therefore, we also need a biharmonic map version of it.

Since the extension lemmas in [Schoen and Uhlenbeck 1982; Luckhaus 1988] are not available for biharmonic maps, the original proof in [Schoen and Uhlenbeck 1982] does not work here. Fortunately, Scheven [2008, Theorem 1.5] proved that if u_i is a sequence of minimizing biharmonic maps with bounded total energy, then there is a subsequence converging *strongly* to a stationary biharmonic map. More precisely, we have

Lemma 2.3 [Scheven 2008, Proposition 1.5]. *Suppose that $u_i : B_2 \to N$ is a sequence of minimizing biharmonic maps with bounded energy. Then there is a subsequence u_{i_j} that converges* strongly *to a stationary biharmonic map on B_1.*

Proof. Since we have assumed that N is compact, then u_i are uniformly bounded on B_2. The energy bound then implies that $\|u_i\|_{W^{2,2}(B_{3/2})}$ is bounded, so that we can use [Scheven 2008, Proposition 1.5] to get a subsequence converging to u strongly in $W^{2,2}$; u is stationary because the minimizers are stationary and the property of being stationary is preserved in strong limit. □

We then combine Lemma 2.3 and Lemma 2.1 to get the biharmonic version of [Schoen and Uhlenbeck 1982, Proposition 4.5].

Lemma 2.4 (biharmonic map version of [Schoen and Uhlenbeck 1982, Proposition 4.5]). *For $\Lambda > 0$ fixed, there is $\varepsilon_2 = \varepsilon_2(N, \Lambda) > 0$ such that the following holds: Suppose that $u : B_2 \to N$ is a minimizing biharmonic map with $E(u) < \Lambda$ and $B_r(x) \subset B_1$. If*

$$r^{-m} \int_{B_r(x)} |u - q|^2 \, dx \le \varepsilon_2$$

for some $q \in N$, then

$$\sup_{B_{r/4}(x)} r^k |\nabla^k u| \le C(k) \quad \forall k \in \mathbb{N}$$

for some $C(k) > 0$.

Proof. By Lemma 2.1, it suffices to show

$$\left(\tfrac{1}{2} r\right)^{4-m} \int_{B_{r/2}(x)} |\nabla^2 u|^2 + \left(\tfrac{1}{2} r\right)^{-2} |\nabla u|^2 \, dx \le \varepsilon_1.$$

If otherwise, we have a sequence of minimizing biharmonic maps $u_i : B_2 \to N$ with $E(u_i) < \Lambda$ such that for some $B_{r_i}(x_i) \subset B_1$, we have

$$r_i^{-m} \int_{B_{r_i}(x_i)} |u_i - q_i|^2 \, dx \to 0 \tag{8}$$

and

$$\left(\tfrac{1}{2} r_2\right)^{4-m} \int_{B_{r_i/2}(x)} |\nabla^2 u_i|^2 + \left(\tfrac{1}{2} r_2\right)^{-2} |\nabla u_i|^2 \, dx \ge \varepsilon_1. \tag{9}$$

Note that r_i may converge to zero. Let $v_i(x) = u_i(r_i x + x_i)$. The monotonicity formula (see [Wang 2004a, Lemma 5.3]) tells us that

$$\int_{B_2(0)} |\nabla^2 v_i|^2 + |\nabla v_i|^2 \, dx \le C(\Lambda).$$

By Lemma 2.3, taking subsequence if necessary, v_i converges to some stationary biharmonic map v strongly in $W^{2,2}(B_1)$, which must be the trivial map due to (8). Since the convergence is strong in $W^{2,2}$, we know that

$$\int_{B_1(0)} |\nabla^2 v_i|^2 + |\nabla v_i|^2 \, dx \to 0,$$

for i sufficiently large. This is a contradiction with (9) and therefore the lemma is proved. $\square$

Section of a tangent cone. Let u be the minimizing biharmonic map in Theorem 1.1. By the assumptions of the theorem, there is some sequence $r_i \to 0$ such that $u(r_i x)$ converges to a homogenous tangent map (which is biharmonic)

$$\tilde{\varphi} := \varphi\Big(\frac{x}{|x|}\Big) \tag{10}$$

and φ is a smooth map from S^{m-1} to N. It follows from Lemma 2.3 and Lemma 2.1 that this convergence is in fact smooth convergence away from the origin.

Recall that in the harmonic map case, $\tilde{\varphi}$ is a harmonic map if and only if so is φ. Here for the biharmonic maps, the situation is somewhat different and it is the purpose of this subsection to characterize φ that appears as the section of homogeneous biharmonic maps.

Let (r, θ) be the polar coordinates of $\mathbb{R}^m$. A direct computation shows

$$\Delta^2\tilde{\varphi} = \Big(\frac{\partial^2}{\partial r^2} + \frac{m-1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_{S^{m-1}}\Big)^2\tilde{\varphi} = r^{-4}(\Delta^2_{S^{m-1}}\varphi - (2m-8)\Delta_{S^{m-1}}\varphi).$$

If $\tilde{\varphi}$ is a biharmonic map, then $\Delta^2\tilde{\varphi} \perp T_{\tilde{\varphi}}N$, which is equivalent to

$$\Delta^2_{S^{m-1}}\varphi - (2m-8)\Delta_{S^{m-1}}\varphi \perp T_{\varphi}N. \tag{11}$$

Instead of working out the explicit formula of (11), it suffices for our purpose to note that it is the Euler–Lagrange equation of the energy functional

$$F(\varphi) := \int_{S^{m-1}} |\Delta_{S^{m-1}}\varphi|^2 + (2m-8)|\nabla_{S^{m-1}}\varphi|^2\, d\theta. \tag{12}$$

Here we write $d\theta$ for the volume element on S^{m-1} and φ is a map from S^{m-1} to N.

L^2 closeness implies C^5 closeness. Let φ be the smooth section in Theorem 1.1, which is a smooth critical map of F. We define

$$\mathcal{O}_{L^2}(\sigma) = \{\psi : S^{m-1} \to N \mid \|\psi - \varphi\|_{L^2(S^{m-1})} < \sigma\}$$

and

$$\mathcal{O}_{C^5}(\sigma) = \{\psi : S^{m-1} \to N \mid \|\psi - \varphi\|_{C^5(S^{m-1})} < \sigma\}.$$

Let u be a smooth biharmonic map defined on $B \setminus \{0\}$ and (t, θ) be the cylinder coordinates. In this paper, we often regard $u(t)$ as a family of maps from S^{m-1} to N. In the proof of our main theorem, these $u(t)$ are often close to φ in various sense. The next theorem roughly says that L^2-closeness (of $u(t)$) to φ on some t-interval implies C^5-closeness in a smaller t-interval.

Lemma 2.5. *For any $\sigma_1 > 0$, there is $\sigma_2 > 0$ (depending on σ_1, φ, and N) such that the following is true. Let $u(t, \theta)$ be as above. If*

$$u(s) \in \mathcal{O}_{L^2}(\sigma_2) \quad \forall\, s \in (t_0 - 2, t_0 + 2),$$

then

$$u(s) \in \mathcal{O}_{C^5}(\sigma_1) \quad \forall\, s \in (t_0 - 1, t_0 + 1). \tag{13}$$

Moreover, for some $C > 0$ (*depending on* σ_1, φ, *and* N),

$$\sum_{k=1,2,3,4} \sum_{j=0}^{4-k} |\partial_t^k \nabla_{S^{m-1}}^j u|(s, \theta) \le C \quad \forall\, s \in (t_0 - 1, t_0 + 1). \tag{14}$$

Remark 2.6. It is clear from the proof below that the lemma is still true for any C^k neighborhood of φ instead of C^5.

Proof. Although the lemma is stated in terms of (t, θ) coordinates, the proof is more clearly presented in the (r, θ) coordinates. By the scaling invariance of (6), we may assume that $t_0 = 2$ and study (6) on $B_1 \setminus B_{e^{-4}}$. By abuse of notation, we also write φ for the function

$$\varphi(r, \theta) = \varphi(\theta).$$

The assumption that $u(s) \in \mathcal{O}_{L^2}(\sigma_2)$ implies that there is a constant $C(\sigma_2)$ (satisfying $\lim_{\sigma_2 \to 0} C(\sigma_2) = 0$) such that

$$\int_{B_1 \setminus B_{e^{-4}}} |u - \varphi|^2 \, dx \le C(\sigma_2). \tag{15}$$

Since φ is smooth, there is some constant C_φ depending only on φ such that

$$|\varphi(x) - \varphi(y)| \le C_\varphi |x - y| \tag{16}$$

for any $x, y \in B_1 \setminus B_{e^{-4}}$. For some $y \in B_{e^{-1}} \setminus B_{e^{-3}}$, consider the ball $B_\sigma(y) \subset B_1 \setminus B_{e^{-4}}$ for some $\sigma > 0$ to be determined later. By (15) and (16), we have

$$\begin{aligned}
\sigma^{-m} \int_{B_\sigma(y)} & |u(x) - \varphi(y)|^2 \, dx \\
&\le 2\sigma^{-m} \int_{B_\sigma(y)} |u(x) - \varphi(x)|^2 \, dx + 2\sigma^{-m} \int_{B_\sigma(y)} |\varphi(x) - \varphi(y)|^2 \, dx \\
&\le 2\sigma^{-m} C(\sigma_2) + 2|B| C_\varphi^2 \sigma^2.
\end{aligned}$$

Here, $|B|$ is the volume of the unit ball in $\mathbb{R}^m$.

Let ε_2 be the constant in Lemma 2.4. We first take σ small with $2|B|C_\varphi^2\sigma^2 < \frac{1}{2}\varepsilon_2$ and then choose σ_2 sufficiently small so that $2\sigma^{-m}C(\sigma_2) < \frac{1}{2}\varepsilon_2$. Hence, Lemma 2.4 gives

$$\|u\|_{C^6(B_{e^{-1}} \setminus B_{e^{-3}})} \le C, \tag{17}$$

from which (14) follows. Equation (13) can be proved by interpolation between the C^6 bound (17) and the L^2 bound (15), if we choose σ_2 smaller. $\square$

Estimates of $\partial_t u$. Since being biharmonic is invariant under scaling and the group of scaling is generated by the vector field $r\partial_r = \partial_t$, if u is a biharmonic map, then $\partial_t u$ satisfies the linearization equation of (6), which is a homogeneous linear elliptic system whose coefficients depend on u. Using this equation, we can prove

Lemma 2.7. *If u satisfies (13) and (14) for $s \in (t_0 - 1, t_0 + 1)$, then we have*

$$\sum_{k=1,2,3,4} \sum_{j=0}^{4-k} |\partial_t^k \nabla_{S^{m-1}}^j u|^2(t_0, \theta) \le \widetilde{C} \int_{t_0-1}^{t_0+1} \int_{S^{m-1}} |\partial_t u|^2 \, d\theta \, dt, \tag{18}$$

for some constant $\widetilde{C}$ depending only on σ_1 (in (13)), C (in (14)), and the target manifold N.

Proof. The proof is an interior estimate of an elliptic system. By scaling invariance of (6), we may assume that $t_0 = 2$. Hence to show (18), it suffices to prove

$$\|\partial_r u\|_{C^3(B_{e^{-3/2}} \setminus B_{e^{-5/2}})} \le \widetilde{C} \|\partial_r u\|_{L^2(B_{e^{-1}} \setminus B_{e^{-3}})}.$$

The observation is that if we compute the homogeneous elliptic system of $r\partial_r u$ mentioned above, the Hölder norm of all coefficients are bounded due to (13) and (14). □

3. The Lojasiewicz–Simon inequality

The main purpose of this section is to prove the Lojasiewicz–Simon inequality for F defined by (12):

$$F(\psi) = \int_{S^{m-1}} |\Delta_{S^{m-1}} \psi|^2 + (2m-8)|\nabla_{S^{m-1}} \psi|^2 \, d\theta.$$

Lemma 3.1. *Let φ be a smooth critical point of $F(\psi)$. Then there are $\varepsilon > 0$, $\alpha \in (0, 1]$ and $C > 0$ depending on φ such that for all $\psi : S^{m-1} \to N$ with*

$$\|\psi - \varphi\|_{C^5(S^{m-1})} \le \varepsilon,$$

we have

$$|F(\psi) - F(\varphi)|^{1-\alpha/2} \le C \|\mathcal{M}_F(\psi)\|_{L^2(S^{m-1})}, \tag{19}$$

where $\mathcal{M}_F(\psi)$ is the Euler–Lagrange operator of F.

An equivalent form. Since φ is smooth, there is a natural correspondence between the maps that are close to φ (in C^5 topology) and the small (in C^5 norm) sections of the pull-back bundle $V := \varphi^* TN$. More precisely, we embed N isometrically as a submanifold in $\mathbb{R}^p$ and identify a section u of $\varphi^* TN$ with a map

$$u : S^{m-1} \to \mathbb{R}^p \quad \text{satisfiying} \quad u(\omega) \in T_{\varphi(\omega)} N \subset \mathbb{R}^p.$$

Via the nearest point projection Π defined in a tubular neighborhood of N, for any ψ close to φ, we define u by

$$\psi(\omega) = \Pi(\varphi(\omega) + u(\omega)). \tag{20}$$

This is well defined because for each ω, Π is a diffeomorphism between a neighborhood of $\varphi(\omega)$ in N and a neighborhood of 0 in $T_{\varphi(\omega)}N \subset \mathbb{R}^p$.

Since φ is a fixed smooth map, the $C^{k,\beta}$ norm of u as a section of V defined by the induced pull-back connection is equivalent to the $C^{k,\beta}$ norm of u as a map from S^{m-1} to $\mathbb{R}^p$ (with restrictions to the image). The same applies to the Sobolev norms as well. While the intrinsic role of u as a section is enough for the argument (the Lyapunov–Schmidt reduction), the extrinsic role of u as a map is important in using the analyticity assumption. (See Appendix A.)

With the above identification in mind, define

$$\widetilde{F}(u) = F(\psi) - F(\varphi).$$

Then $\widetilde{F}(0) = 0$ and $u = 0$ is a critical point of $\widetilde{F}$. Let $\mathcal{M}_{\widetilde{F}}(u)$ be the Euler–Lagrange operator of $\widetilde{F}$ at u. Since the L^2 inner product of V that we use to compute $\mathcal{M}_{\widetilde{F}}$ is not identical to the L^2 inner product used for the computation of $\mathcal{M}_F$, $\mathcal{M}_F(\psi)$ is not trivially the same as $\mathcal{M}_{\widetilde{F}}(u)$ with u and ψ related by (20). However, we have

Lemma 3.2. *Let $\widetilde{F}$ be defined as above. If ψ is a map from S^{m-1} to N with $\|\psi - \varphi\|_{C^{4,\beta}} < \delta$ for sufficiently small $\delta > 0$ and u is defined by (20), then*

$$(1 - C\delta)|\mathcal{M}_F(\psi)| \le |\mathcal{M}_{\widetilde{F}}(u)| \le |\mathcal{M}_F(\psi)|. \tag{21}$$

The proof follows trivially from the equation (whose derivation is given in Appendix A; see (47))

$$\mathcal{M}_{\widetilde{F}}(u) = P_\varphi \mathcal{M}_F(\psi)$$

and the fact that the tangent space $T_\psi N$ is close to $T_\varphi N$ since φ is close to ψ.

Given Lemma 3.2, Lemma 3.1 is reduced to

Lemma 3.3. *There are $\varepsilon > 0$, $\alpha \in (0, 1]$, and $C > 0$ depending on φ such that for all $u \in C^5(V)$ with*

$$\|u\|_{C^5(V)} \le \varepsilon,$$

we have

$$|\widetilde{F}(u)|^{1-\alpha/2} \le C\|\mathcal{M}_{\widetilde{F}}(u)\|_{L^2(V)}. \tag{22}$$

The Lyapunov–Schmidt reduction. The proof of Lemma 3.3 is an application of the Lyapunov–Schmidt reduction argument. The local behavior of $\widetilde{F}$ near $u = 0$ is related to an analytic function defined on the finite dimensional kernel of an elliptic operator. More precisely, let $\mathcal{L}_{\widetilde{F}}$ be the linearization of $\mathcal{M}_{\widetilde{F}}$ at $u = 0$, which is an elliptic operator from $C^{4,\beta}(V)$ to $C^{0,\beta}(V)$. By the theory of elliptic operators, the kernel of $\mathcal{L}_{\widetilde{F}}$ is a finite dimensional space, denoted by K. Let P_K be the orthogonal projection of $L^2(V)$ onto K.

Setting

$$\mathcal{N}(u) = P_K u + \mathcal{M}_{\widetilde{F}}(u),$$

we find that $\mathcal{N}(0) = 0$ and the linearization of $\mathcal{N}$ at $u = 0$ is given by

$$P_K + \mathcal{L}_{\widetilde{F}},$$

which is an isomorphism between $C^{4,\beta}(V)$ onto $C^{0,\beta}(V)$ because it is self-adjoint with trivial kernel. The inverse function theorem then gives an inverse $\Psi = \mathcal{N}^{-1}$ from a neighborhood of $0 \in C^{0,\beta}(V)$ to $C^{4,\beta}(V)$.

Remark 3.4. (1) For the ellipticity and self-adjointness of $\mathcal{L}_{\widetilde{F}}$, see Section A.

(2) The inverse function here actually appears as the real part of a complexified inverse function, which we need to justify the analyticity of f in (23) below.

Moreover, we have the following estimate for Ψ:

Lemma 3.5 (L^2 estimate). *There is a neighborhood W of 0 in $C^{0,\beta}(V)$ and a constant C, depending only on $\widetilde{F}$, such that*

$$\|\Psi(f_1) - \Psi(f_2)\|_{W^{4,2}(V)} \leq C\|f_1 - f_2\|_{L^2(V)} \quad \textit{for any} \quad f_1,\ f_2 \in W.$$

We refer to Appendix B for the proof.

With the help of Ψ, we define

$$f(\xi) = \widetilde{F}\left(\Psi\left(\sum_{j=1}^{l} \xi^j \varphi_j\right)\right) \tag{23}$$

for $|\xi|$ small, where $l = \dim K$ and $\{\varphi_j\}$ is a basis of K with respect to the L^2 inner product.

The key to the proof of Lemma 3.3 and hence Lemma 3.1 is the fact that f is real analytic in a neighborhood of 0. The proof relies on an analytic version of inverse function theorem for maps between complex Banach spaces and finally depends on the assumption about the analyticity of N in Theorem 1.1. It takes some effort to be precise in tracing the use of this assumption and the details of this argument are given in Appendix A.

For now, we take the analyticity of f near 0 for granted. Therefore, it follows from the classical Lojasiewicz inequality that there are constants $\alpha \in (0, 1]$, C, and $\sigma > 0$ such that

$$|f(\xi)|^{(1-\alpha/2)} \leq C|\nabla f(\xi)| \quad \text{for} \quad \xi \in B_\sigma(0). \tag{24}$$

For the proof of Lemma 3.3, we need the following:

Lemma 3.6. *When $\|u\|_{C^{4,\beta}(V)}$ is sufficiently small and hence $\xi^j = (u, \varphi_j)_{L^2}$ is small, we have*

$$|\widetilde{F}(u) - f(\xi)| \le C\|\mathcal{M}_{\widetilde{F}}(u)\|_{L^2}^2 \tag{25}$$

and

$$\tfrac{1}{2}|\nabla f|(\xi) \le \left\|\mathcal{M}_{\widetilde{F}}\left(\Psi\left(\sum_{j=1}^{l}\xi^j\varphi_j\right)\right)\right\|_{L^2} \le 2|\nabla f|(\xi). \tag{26}$$

Before the proof of Lemma 3.6, we show how Lemma 3.3 follows from it and (24).

In fact, by plugging (25) and (26) directly into (24), we get

$$\begin{aligned}|\widetilde{F}(u)|^{1-\alpha/2} &\le C\left(\left\|\mathcal{M}_{\widetilde{F}}\left(\Psi\left(\sum\xi^j\varphi_j\right)\right)\right\|_{L^2} + \|\mathcal{M}_{\widetilde{F}}(u)\|_{L^2}^{2-\alpha}\right)\\ &\le C\left(\left\|\mathcal{M}_{\widetilde{F}}\left(\Psi\left(\sum\xi^j\varphi_j\right)\right)\right\|_{L^2} + \|\mathcal{M}_{\widetilde{F}}(u)\|_{L^2}\right).\end{aligned} \tag{27}$$

Here in the last line above, we use the facts that $2-\alpha \ge 1$ and that $\|\mathcal{M}_{\widetilde{F}}(u)\|$ is bounded for u in the lemma. The first term in the right-hand side of (27) is dominated by the second, because

$$\begin{aligned}\left\|\mathcal{M}_{\widetilde{F}}\left(\Psi\left(\sum\xi^j\varphi_j\right)\right) - \mathcal{M}_{\widetilde{F}}(u)\right\|_{L^2} &\le C\left\|\Psi\left(\sum\xi^j\varphi_j\right) - u\right\|_{W^{4,2}}\\ &\le C\left\|\sum\xi^j\varphi_j - \Psi^{-1}u\right\|_{L^2}\\ &\le C\|\mathcal{M}_{\widetilde{F}}(u)\|_{L^2}.\end{aligned} \tag{28}$$

Here for the first line above, we noticed that $\mathcal{M}_{\widetilde{F}}$ is a (nonlinear) fourth-order differential operator (see (47)) and both the $C^{4,\beta}$ norms of $\Psi(\sum\xi^j\varphi_j)$ and u are bounded; for the third line above, we used Lemma 3.5; for the last line, we used the definition of $\mathcal{N} = \Psi^{-1}$ and $P_K u = \sum\xi^j\varphi_j$. Now, Lemma 3.3 is a consequence of (27) and (28).

The rest of this section is devoted to the proof of Lemma 3.6.

The proof of Lemma 3.6. By the definition of f (see (23)) and ξ (in the assumption of Lemma 3.6), $f(\xi) = \widetilde{F}(\Psi(P_K u))$. Hence, to prove (25), we compute

$$\begin{aligned}|\widetilde{F}(u) - \widetilde{F}(\Psi(P_K u))| &= \left|\int_0^1 \frac{d}{ds}\widetilde{F}\big(u + s(\Psi(P_K u) - u)\big)\,ds\right|\\ &= \left|\int_0^1 \big(\mathcal{M}_{\widetilde{F}}(u + s(\Psi(P_K u) - u)),\, \Psi(P_K u) - u\big)_{L^2}\,ds\right|.\end{aligned}$$

Again, by the facts that $\mathcal{M}_{\widetilde{F}}$ is a fourth-order operator and that $C^{4,\beta}$ norms of u and $u+s(\Psi(P_K u)-u)$ are bounded for any $s\in[0,1]$, we have

$$\|\mathcal{M}_{\widetilde{F}}(u+s(\Psi(P_K u)-u))-\mathcal{M}_{\widetilde{F}}(u)\|_{L^2}\le C\|\Psi(P_K u)-u\|_{W^{4,2}},$$

which implies that

$$\begin{aligned}|\widetilde{F}(u)-\widetilde{F}(\Psi(P_K u))| &\le C\|\Psi(P_K u)-u\|_{L^2}(\|\mathcal{M}_{\widetilde{F}}(u)\|_{L^2}+\|\Psi(P_K u)-u\|_{W^{4,2}})\\ &\le C\|\mathcal{M}_{\widetilde{F}}(u)\|_{L^2}^2.\end{aligned}$$

Here in the last line above, we used

$$\|\Psi(P_K u)-u\|_{W^{4,2}}\le C\|\mathcal{M}_{\widetilde{F}}(u)\|_{L^2},$$

which appeared as a part of (28) and was proved there. This concludes the proof of (25).

For the proof of (26), we compute using (23) and the chain rule to get

$$(\eta,\nabla f(\xi))_{\mathbb{R}^l}=\left(\mathcal{M}_{\widetilde{F}}\left(\Psi\left(\sum\xi^j\varphi_j\right)\right),d\Psi|_{\sum\xi^j\varphi_j}\left(\sum\eta^j\varphi_j\right)\right)_{L^2} \tag{29}$$

for some $\eta\in\mathbb{R}^l$ with $|\eta|=1$.

Notice that $d\Psi|_{\sum\xi^j\varphi_j}$ depends smoothly on ξ in a compact neighborhood of $\xi=0$, hence there is $C>0$ such that

$$\|d\Psi|_{\sum\xi^j\varphi_j}-d\Psi|_0\|\le C|\xi|\quad\text{for small } |\xi|. \tag{30}$$

Remark 3.7. (1) For the smooth dependence in ξ, we shall prove in Appendix A that Ψ has a complexification that is analytic (hence smooth by Theorem A.2).

(2) The norm in (30) should be the norm of bounded linear operator from $C^\beta(V)$ to $C^{4,\beta}(V)$, according to our discussion in the appendix. What we need here is the inequality

$$\left\|(d\Psi|_{\sum\xi^j\varphi_j}-d\Psi|_0)\left(\sum\eta^j\varphi_j\right)\right\|_{L^2}\le C|\xi|\left\|\sum\eta^j\varphi_j\right\|_{L^2}.$$

This is true because $\sum\eta^j\varphi_j$ lies in K and when restricted to the finite dimensional space K, L^2 norm is equivalent to C^β norm.

On the other hand,

$$d\Psi|_0\left(\sum\eta^j\varphi_j\right)=\sum\eta^j\varphi_j\quad\text{for any}\quad\eta\in\mathbb{R}^l, \tag{31}$$

because $d\Psi|_0=(d\mathcal{N}|_0)^{-1}=(P_K+\mathcal{L}_{\widetilde{F}})^{-1}$, and $\sum\eta^j\varphi_j$ is in K, the kernel of $\mathcal{L}_{\widetilde{F}}$.

By (30) and (31), (29) implies that

$$(32)\quad \left|(\eta, \nabla f(\xi))_{\mathbb{R}^l} - \left(\mathcal{M}_{\widetilde{F}}\left(\Psi\left(\sum \xi^j \varphi_j\right)\right), \sum \eta^j \varphi_j\right)_{L^2}\right| \\ \leq C|\xi| \|\mathcal{M}_{\widetilde{F}}\left(\Psi\left(\sum \xi^j \varphi_j\right)\right)\|_{L^2}.$$

Now, in (32), if we choose η parallel to $\nabla f(\xi)$ in $\mathbb{R}^l$, we obtain

$$|\nabla f| \leq (1 + C|\xi|)\left\|\mathcal{M}_{\widetilde{F}}\left(\Psi\left(\sum \xi^j \varphi_j\right)\right)\right\|_{L^2};$$

if we choose η so that $\sum \eta^j \varphi_j$ is parallel to $\mathcal{M}_{\widetilde{F}}(\Psi(\sum \xi^j \varphi_j))$ (which is in K), then we get

$$(1 - C|\xi|)\|\mathcal{M}_{\widetilde{F}}\left(\Psi\left(\sum \xi^j \varphi_j\right)\right)\|_{L^2} \leq |\nabla f|.$$

This finishes the proof of (26) and hence Lemma 3.6 if ξ is small.

4. Dynamics near a critical point of *F*

Let u be the minimizing biharmonic map given in Theorem 1.1. Recall that (r, θ) are the polar coordinates and that $t = -\log r$. By the assumptions of the theorem, there exists $t_i \to \infty$ such that $u(t_i, \theta)$ as maps on S^{m-1} converge smoothly to a critical point φ of F. (See the discussion in Section 2.)

Therefore, for i sufficiently large, $u(t_i, \theta)$ is very close in C^5 topology to the critical point φ of F. Since u is a biharmonic map, the biharmonic map equation determines how $u(t)$ should change as a map on S^{m-1}. In this section, we study these dynamics of $u(t)$ in a very small neighborhood of φ. More precisely, we are interested in the speed of decay of

$$\int_t^\infty \int_{S^{m-1}} |\partial_t u|^2 \, d\theta \, dt$$

as explained in the introduction. In fact, we shall control the decay of a larger quantity, namely,

$$(33)\quad G(t) = \int_t^\infty \int_{S^{m-1}} (2m-8)|\partial_t^2 u|^2 + (2m-8)|\partial_t \nabla_{S^{m-1}} u|^2 \\ + (2m-8)(m-2)|\partial_t u|^2 \, d\theta \, dt.$$

Lemma 4.1. *Suppose φ is a smooth critical point of F. There is some constant $\sigma > 0$ (depending on φ) such that if $u(t, \theta)$ (cylinder coordinates) is a smooth biharmonic map satisfying*

$$\|u(s) - \varphi\|_{C^5(S^{m-1})} \leq \sigma \quad \textit{for} \quad s \in [t-3, t+3],$$

then there exist $C' > 0$ *and* $\theta \in (0,1)$ *such that*

$$G(s-1)^{\theta} - G(s+1)^{\theta} \geq C'(G(s-1) - G(s+1))^{1/2}. \tag{34}$$

Before we start the proof, we rewrite $\Delta^2 u$ in (t,θ) coordinates and split it into two parts. Since

$$\Delta u = e^{2t}(\partial_t^2 - (m-2)\partial_t + \Delta_{S^{m-1}})u,$$

we have

$$\begin{aligned} \Delta^2 u &= e^{4t}(\partial_t^2 + \Delta_{S^{m-1}} - (m-6)\partial_t + (8-2m))(\partial_t^2 - (m-2)\partial_t + \Delta_{S^{m-1}})u \\ &:= e^{4t}(I_a + I_b), \end{aligned}$$

where

$$\begin{aligned} I_a = \partial_t^4 u + 2\partial_t^2 \Delta_{S^{m-1}} u - (2m-8)\partial_t^3 u - (2m-8)\partial_t \Delta_{S^{m-1}} u \\ + (m^2 - 10m + 20)\partial_t^2 u + (2m-8)(m-2)\partial_t u \end{aligned}$$

and

$$I_b = \Delta_{S^{m-1}}^2 u + (8-2m)\Delta_{S^{m-1}} u.$$

The idea behind this splitting is that we put everything involving ∂_t in I_a and the rest in I_b. An easy observation is that I_b is almost (up to a projection) the gradient of F discussed in Section 2, namely,

$$\begin{aligned} \partial_t F(u(t)) &= 2\int_{S^{m-1}} \Delta_{S^{m-1}} u \Delta_{S^{m-1}} \partial_t u + (2m-8)\nabla_{S^{m-1}} u \cdot \nabla_{S^{m-1}} \partial_t u \, d\theta \\ &= 2\int_{S^{m-1}} I_b \cdot \partial_t u \, d\theta. \end{aligned} \tag{35}$$

The way we use the biharmonic map equation has nothing to do with the right-hand side of (6). We multiply the equation by $\partial_t u$ and integrate over S^{m-1} to obtain

$$0 = \int_{S^{m-1}} \Delta^2 u \cdot \partial_t u \, d\theta = \int_{S^{m-1}} (I_a + I_b) \cdot \partial_t u \, d\theta. \tag{36}$$

While $\int_{S^{m-1}} I_b \cdot \partial_t u \, d\theta$ is known in (35), the structure of $\int_{S^{m-1}} I_a \cdot \partial_t u \, d\theta$ is still complicated. There is some positivity hidden in it. To reveal it, we use the elementary equalities

$$\partial_t^4 u \cdot \partial_t u = \partial_t\left(\partial_t^3 u \partial_t u - \tfrac{1}{2}|\partial_t^2 u|^2\right)$$

and

$$\partial_t^3 u \cdot \partial_t u = \partial_t(\partial_t^2 u \cdot \partial_t u) - |\partial_t^2 u|^2$$

to get

$$\begin{aligned}\int_{S^{m-1}} I_a \cdot \partial_t u \, d\theta \\ &= \partial_t \Big(\int_{S^{m-1}} \partial_t^3 u \partial_t u - \tfrac{1}{2} |\partial_t^2 u|^2 \\ &\qquad - |\partial_t \nabla_{S^{m-1}} u|^2 - (2m-8) \partial_t^2 u \partial_t u + \tfrac{1}{2}(m^2 - 10m + 20) |\partial_t u|^2 \, d\theta \Big) \\ &\quad + \Big(\int_{S^{m-1}} (2m-8) |\partial_t^2 u|^2 + (2m-8) |\partial_t \nabla_{S^{m-1}} u|^2 \\ &\qquad\qquad + (2m-8)(m-2) |\partial_t u|^2 \, d\theta \Big) \\ &:= \partial_t \Big(\int_{S^{m-1}} II_a \, d\theta \Big) + \int_{S^{m-1}} II_b \, d\theta.\end{aligned}$$

Notice that II_b is nonnegative and this is how we obtain the definition of $G(t)$ in (33), i.e.,

$$G(t) = \int_t^{\infty} \int_{S^{m-1}} II_b \, d\theta \, dt.$$

By (36) and (35), we have

$$\tfrac{1}{2} \partial_t F(u(t)) = - \int_{S^{m-1}} I_a \cdot \partial_t u \, d\theta. \tag{37}$$

Let t_i be the sequence mentioned in the beginning of this section such that $u(t_i)$ converges smoothly to the smooth section map φ. Moreover, $u(t + t_i)$ regarded as a map defined on $[-1, 1] \times S^{m-1}$ converges smoothly to $\tilde{\varphi}(t, \theta) = \varphi(\theta)$. This implies that

$$\lim_{i \to \infty} \int_{S^{m-1}} II_a(t_i) \, d\theta = 0,$$

so that if we integrate (37) from s to t_i and take the limit $i \to \infty$, we obtain

$$\tfrac{1}{2}(F(\varphi) - F(u(s))) = \int_{S^{m-1}} II_a(s) \, d\theta - \int_s^{+\infty} \int_{S^{m-1}} II_b \, d\theta. \tag{38}$$

As a by-product of the above computation, $G(t)$ is a finite number, which is the biharmonic counterpart of (2).

We may choose σ small so that for u in the lemma and $s \in [t-3, t+3]$, $\|u(s) - \varphi\|_{C^5(S^{m-1})}$ is small and hence $u(s)$ satisfies the assumption of Lemma 3.1. The Lojasiewicz–Simon inequality (in Lemma 3.1) and (38) imply that

$$- \int_{S^{m-1}} II_a(s) \, d\theta + \int_s^{+\infty} \int_{S^{m-1}} II_b \, d\theta \, dt \le C \|\mathcal{M}_F(u(s))\|_{L^2(S^{m-1})}^{2/(2-\alpha)} \tag{39}$$

for some $\alpha \in (0, 1]$.

Next, we show that the right-hand side and the first term in the left-hand side of (39) are controlled by $\int_{s-1}^{s+1} |II_b|^2 \, d\theta$. To see this, recall that by (35), $\mathcal{M}_F(u(s))$ is the projection of $I_b(s)$ onto the tangent bundle of TN at $u(s)$. If we denote this projection from $\mathbb{R}^p$ onto $T_u N$ by Π,

$$\mathcal{M}_F(u(s)) = 2\Pi(I_b(s)). \tag{40}$$

On the other hand, since u is an extrinsic biharmonic map, the Euler–Lagrange equation reads

$$\Pi(\Delta^2 u) = \Pi(I_a + I_b) = 0. \tag{41}$$

Combining (40) and (41), we get

$$\|\mathcal{M}_F(u(s))\|_{L^2(S^{m-1})} \le 2\|I_a(s)\|_{L^2(S^{m-1})}. \tag{42}$$

Notice that the integrands of both $I_a(s)$ and $II_a(s)$ involve $\partial_t u$ and its derivatives, which are estimated in Section 2. More precisely, by taking σ small, we may apply Lemma 2.5 first to get (14) and then Lemma 2.7 to see

$$\|I_a(s)\|^2_{L^2(S^{m-1})} + \int_{S^{m-1}} |II_a|(s) \, d\theta \le C \int_{s-1}^{s+1} \int_{S^{m-1}} |II_b|^2 \, d\theta \, dt. \tag{43}$$

By the definition of $G(t)$ in (33), equations (39), (42), and (43) imply

$$-C(G(s-1) - G(s+1)) + G(s) \le C(G(s-1) - G(s+1))^{1/(2-\alpha)}.$$

Since $G(s-1) - G(s+1)$ is bounded and $1/(2-\alpha) \le 1$, the first term can be absorbed into the left-hand side. In fact, in the proof that follows, we shall require $G(s)$ to be very small (see the definition of η in the next section). By the monotonicity of G, the above inequality is further simplified to

$$G(s+1) \le C(G(s-1) - G(s+1))^{1/(2-\alpha)}. \tag{44}$$

Here is a lemma similar to [Simon 1996, (9), §3.15].

Lemma 4.2. *Suppose that $\theta \in \left(0, \frac{1}{2}\right]$. If for some positive C and any $a, b \in (0, 1)$ satisfying $b < a$,*

$$b \le C(a-b)^{1/(2-2\theta)}, \tag{45}$$

then there is another C' depending only on C and θ such that

$$a^\theta - b^\theta \ge C'(a-b)^{1/2}.$$

Proof. The proof is an elementary discussion.

Case 1: $b < \frac{1}{2}a$. Noticing that $\theta \le \frac{1}{2}$ and $a < 1$, we have

$$a^\theta - b^\theta \ge \Big(1 - \frac{1}{2^\theta}\Big)a^\theta \ge \Big(1 - \frac{1}{2^\theta}\Big)a^{1/2} \ge \Big(1 - \frac{1}{2^\theta}\Big)(a-b)^{1/2}.$$

Case 2: $b \ge \frac{1}{2}a$. Equation (45) gives

$$\frac{a}{2C} \le (a-b)^{1/(2(1-\theta))},$$

which is

$$a^{1-\theta} \le (2C)^{1-\theta}(a-b)^{1/2}. \tag{46}$$

Therefore,

$$a^\theta - b^\theta \ge \theta a^{\theta-1}(a-b) \ge \frac{\theta}{(2C)^{1-\theta}}(a-b)^{1/2}.$$

Here in the above line we have used the mean value theorem for the first inequality and (46) for the second.

In either case, the lemma is proved by taking C' to be $\min\{1-1/2^\theta, \theta/(2C)^{1-\theta}\}$. □

5. A stability argument and the proof of Theorem 1.1

In this section, we prove Theorem 1.1 by using a routine stability argument. We shall define two neighborhoods of φ: a larger one (see $\mathcal{O}_{C^5}(\sigma_1)$ below) in which the results in Section 3 and Section 4 hold and a smaller one (see $\mathcal{O}_{L^2}(\eta)$ below) such that if $u(t_i)$ lies in the smaller neighborhood for sufficiently large i, then $u(t)$ will stay in the larger neighborhood forever and converge to the unique limit claimed in Theorem 1.1.

We choose σ_1 so that it is smaller than both the ε in Lemma 3.1 and the σ in Lemma 4.1. For some $\eta > 0$ small (to be determined later), by the definition of φ as the section of a tangent map, we can choose (and fix) t_i large such that

(1) for all $t \in (t_i - 3, t_i + 3)$, $u(t) \in \mathcal{O}_{C^5}(\sigma_1)$;

(2) $u(t_i) \in \mathcal{O}_{L^2}(\eta)$;

(3) $G(t_i) \le \eta^2$, because $G(t)$ is finite and decreases down to zero.

Set

$$T = \sup_t \{t \mid \text{for any } s \in [t_i, t),\ u(s) \in \mathcal{O}_{C_5}(\sigma_1)\}.$$

By (1) above, we know $T \ge t_i + 3$. Now we claim that T is infinity. If otherwise, we want to find a contradiction by showing $u(T) \in \mathcal{O}_{C^5}\big(\frac{1}{2}\sigma_1\big)$. Thanks to Lemma 2.5, there is $\sigma_2 > 0$ depending on $\frac{1}{2}\sigma_1$ such that it suffices to show for any $s \in (t_i, T+2)$, we have $u(s) \in \mathcal{O}_{L^2}(\sigma_2)$. Let k be the largest integer with $t_i + 2k \le s$. Hence,

$$\begin{aligned}\int_{t_i}^{s} &\|\partial_t u\|_{L^2(S^{m-1})}\\ &\le \sum_{j=1}^{k} \int_{t_i+2(j-1)}^{t_i+2j} \|\partial_t u\|_{L^2(S^{m-1})} + \int_{t_i+2k}^{s} \|\partial_t u\|_{L^2(S^{m-1})}\\ &\le C \sum_{j=1}^{k} \Big(\int_{t_i+2(j-1)}^{t_i+2j} \|\partial_t u\|^2_{L^2(S^{m-1})} \Big)^{1/2} + C \Big(\int_{t_i+2k}^{s} \|\partial_t u\|^2_{L^2(S^{m-1})} \Big)^{1/2}\\ &\le C \sum_{j=1}^{k} \Big(\int_{t_i+2(j-1)}^{t_i+2j} \|\partial_t u\|^2_{L^2(S^{m-1})} \Big)^{1/2} + C\eta.\end{aligned}$$

Here in the third line above, we used Hölder inequality and in the last line, we used (3).

By the definition of G, we have

$$\int_{t_i+2(j-1)}^{t_i+2j} \|\partial_t u\|^2_{L^2(S^{m-1})} \le G(t_i + 2(j-1)) - G(t_i + 2j).$$

We can apply Lemma 4.2 with $a = G(t_i + 2j)$ and $b = G(t_i + 2(j-1))$ to get

$$\begin{aligned}\int_{t_i}^{s} \|\partial_t u\|_{L^2(S^{m-1})} &\le C \sum_{j=1}^{k} (G(t_i + 2(j-1))^{\theta} - G(t_i + 2j)^{\theta}) + C\eta\\ &\le C \cdot G(t_i)^{\theta} + C\eta \le C\eta^{2\theta} + C\eta.\end{aligned}$$

If we choose η small, we can have for any $s \in (t_i, T+2)$,

$$\|u(s) - \varphi\|_{L^2(S^{m-1})} \le \|u(t_i) - \varphi\|_{L^2(S^{m-1})} + \int_{t_i}^{s} \|\partial_t u\|_{L^2(S^{m-1})} \le \tfrac{1}{2}\sigma_2.$$

Lemma 2.5 gives the contradiction and proves that $T = \infty$.

We can repeat the above computation with $k = \infty$ to get

$$\int_{t_i}^{+\infty} \|\partial_t u\|_{L^2(S^{m-1})} \le C\eta^{2\theta} + C\eta < \infty,$$

which shows that

$$\lim_{t\to\infty} \|u(t) - \varphi\|_{L^2(S^{m-1})} = 0.$$

As in Remark 2.6, we have u bounded in any $C^{k+1}(S^{m-1})$ norm. By interpolation, we know

$$\lim_{t\to\infty} \|u(t) - \varphi\|_{C^k(S^{m-1})} = 0.$$

Appendix A: The assumption of analyticity

The purpose of this section is to justify (see Lemma A.7) the use of the classical Łojasiewicz inequality to the function f (see (23)) that arises in the Lyapunov–Schmidt reduction in Section 3. Indeed, we shall show how the analyticity assumption of N in Theorem 1.1 carries on step by step to that of f. These arguments, independent from the rest of the proof, are technical and hence presented in the appendix.

Analytic function between Banach spaces. For completeness, we collect a few basic definitions and properties of analytic functions between abstract (complex) Banach spaces. We refer to [Taylor 1937] for proofs and more detailed discussions.

Let E, E' and E'' be complex Banach spaces.

Definition A.1. (1) Let $f(x)$ be a function on E to E', defined in the neighborhood of $x_0 \in E$. If for each $y \in E$, the limit

$$\lim_{\tau\to 0} \frac{f(x_0+\tau y)-f(x_0)}{\tau}$$

exists (for $\tau \in \mathbb{C}$), then it is called the *Gateaux* differential, denoted by $\delta f(x_0; y)$.

(2) A function $f(x)$ on a domain D of E to E' is said to be *analytic* in D if it is continuous and has a Gateaux differential at each point of D. A function is said to be analytic at a point x_0, if it is analytic in some neighborhood of x_0.

Recall that the *Fréchet* differential is defined to be the bounded linear map $Df(x_0)$ from E to E' such that

$$f(x_0+h) = f(x_0) + Df(x_0)h + o(\|h\|_E).$$

While the existence of the Fréchet differential is obviously stronger than the Gateaux differential, Taylor proved the following:

Theorem A.2 [Taylor 1937, Theorem 3, Theorem 12]. *If f is analytic at x_0, then it admits Fréchet differentials of all orders in the neighborhood of that point. Moreover, the Fréchet differential and the Gateaux differential are equal.*

With the equivalence in mind, we recall a version of the inverse function theorem, which follows from [Dieudonné 1960, (10.2.5)] (see also [Nirenberg 1974, §2.7]).

Theorem A.3. *Let E and E' be two complex Banach spaces, f an analytic function from a neighborhood V of $x_0 \in E$ to E'. If $Df(x_0)$ is a linear homeomorphism of E onto E', there exists an open neighborhood $U \subset V$ of x_0 such that the restriction of f to U is a homeomorphism of U onto an open neighborhood of $y_0 = f(x_0)$. Moreover, the inverse is analytic.*

Complexification and analyticity. In Section 3, we have defined the functional $\widetilde{F} : C^{4,\beta}(V) \to \mathbb{R}$ where V is the pullback bundle φ^*TN and a map $\mathcal{N}$ from $C^{4,\beta}(V)$ to $C^{0,\beta}(V)$. Instead of claiming the analyticity of $\widetilde{F}$ and $\mathcal{N}$ directly, we consider its complexification.

$C^{4,\beta}(V) \otimes \mathbb{C}$ is understood to be the set of $u + iv$, where $u, v \in C^{4,\beta}(V)$, with a naturally defined norm. The same applies to $C^{0,\beta}(V) \otimes \mathbb{C}$. Obviously, they are complex Banach spaces.

A complexification of a map f from a Banach space E_1 to another Banach space E_2 is some map $\tilde{f}$ from $E_1 \otimes \mathbb{C}$ to $E_2 \otimes \mathbb{C}$ such that f is the real part of $\tilde{f}$ when restricted to (some open set of) E_1. Such complexifications are usually not unique. We are interested in analytic ones, that we define below (making using of special properties of f).

The complexification of $\widetilde{F}$ and $\mathcal{N}$ relies on some particular form of the maps themselves. More precisely, we need the definition of $\widetilde{F}(u)$ and $\mathcal{N}(u)$ to be given by a converging series. For this purpose, we start with an extrinsic point of view of V.

Since N is embedded in $\mathbb{R}^p$, we regard T_yN as a subspace (not the affine space passing y) of $\mathbb{R}^p$. Hence, the pullback bundle V is the disjoint union of $V_\omega := T_{\varphi(\omega)}N$ and a section u of V is a map from S^{m-1} to $\mathbb{R}^p$ satisfying

$$u(\omega) \in T_{\varphi(\omega)}N \subset \mathbb{R}^p.$$

For a fixed smooth φ, the $C^{k,\beta}$ norm of u as a map into $\mathbb{R}^p$ agrees with the $C^{k,\beta}$ norm defined intrinsically using the pullback connection of φ^*TN. The same holds for various Sobolev norms.

For the complexification of $\widetilde{F}$, we regard it as the composition of

$$C^{2,\beta}(V) \xrightarrow{\mathcal{F}} C^{\beta}(S^{m-1}, \mathbb{R}) \xrightarrow{\mathcal{I}} \mathbb{R},$$

where

$$\mathcal{F}(u) = |\Delta_{S^{m-1}} \Pi(\varphi + u)|^2 + (2m-8)|\nabla_{S^{m-1}} \Pi(\varphi + u)|^2$$

and

$$\mathcal{I}(h) = \int_{S^{m-1}} h \, d\theta.$$

Recall that Π is the nearest-point-projection of N and the discussion works only for u with small C^0 norm.

We claim that there exists an analytic map $\widetilde{F}_C$ from $C^{2,\beta}(V) \otimes \mathbb{C}$ to $\mathbb{C}$ with $\widetilde{F}$ as its real part.

The proof of the claim is the combination of the following facts:

(F1) The $\Delta_{S^{m-1}}$ from $C^{2,\beta}(V)$ to $C^{0,\beta}(V)$, $\nabla_{S^{m-1}}$ from $C^{2,\beta}(V)$ to $C^{1,\beta}(V)$, and $\mathcal{I}$ are bounded linear maps. Their complexifications, obtained by linear extension, are naturally bounded linear maps and hence analytic.

(F2) Let $\mathcal{F}_1$ be the map from $C^{0,\beta}(S^{m-1}, \mathbb{R}^p)$ to $C^{0,\beta}(S^{m-1}, \mathbb{R})$ given by $u \mapsto |u|^2$. Its complexification $\mathcal{F}_{C,1}$ is given by

$$\mathcal{F}_{C,1}(u+iv) = (u+iv)\cdot(u+iv).$$

It is analytic.

(F3) If (as assumed in Theorem 1.1) $\Pi(\varphi + \cdot\,)$ is an analytic map from $B_r(0) \subset \mathbb{R}^p$ to $\mathbb{R}^p$, then the map

$$u \mapsto \Pi(\varphi + u)$$

has an analytic extension from $C^{2,\beta}(S^{m-1}, \mathbb{C}^p)$ to itself. To see this, one first expands $\Pi(\varphi + u)$ into converging power series of u and then replace u by $u + iv$. It is then an exercise to check that the map thus obtained are analytic in the sense of Definition A.1.

For the complexification of $\mathcal{N}$, it suffices to consider $\mathcal{M}_{\widetilde{F}}(u)$. For u and v in $C^4(V)$, setting $\psi = \Pi(\varphi + u)$, we compute

$$\begin{aligned}
\frac{d}{dt}\Big|_{t=0} &\widetilde{F}(u+tv) \\
&= \frac{d}{dt}\Big|_{t=0} \int_{S^{m-1}} |\Delta_{S^{m-1}}\Pi(\varphi+u+tv)|^2 \\
&\qquad\qquad\qquad + (2m-8)|\nabla_{S^{m-1}}\Pi(\varphi+u+tv)|^2\, d\theta \\
&= 2\int_{S^{m-1}} \Delta_{S^{m-1}}\psi\, \Delta_{S^{m-1}} D\Pi_{\varphi+u}v + (2m-8)\nabla_{S^{m-1}}\psi\, \nabla_{S^{m-1}} D\Pi_{\varphi+u} v\, d\theta \\
&= 2\int_{S^{m-1}} (\Delta^2_{S^{m-1}}\psi - (2m-8)\Delta_{S^{m-1}}\psi) D\Pi_{\varphi+u} v\, d\theta \\
&= 2\int_{S^{m-1}} P_\psi(\Delta^2_{S^{m-1}}\psi - (2m-8)\Delta_{S^{m-1}}\psi) v\, d\theta.
\end{aligned}$$

Here in the last line above, we used the fact that $D\Pi_{\varphi+u}v$ is nothing but the orthogonal projection from $\mathbb{R}^p$ onto $T_\psi N$, which we denote by P_ψ.

Similar to the (bi)harmonic map case, $P_\psi(\Delta^2_{S^{m-1}}\psi - (2m-8)\Delta_{S^{m-1}}\psi)$ is the Euler–Lagrange operator of $F(\psi)$, denoted by $\mathcal{M}_F(\psi)$. For each $\omega \in S^{m-1}$, $\mathcal{M}_F(\psi)(\omega)$ lies in $T_\psi N \subset \mathbb{R}^p$, while $v(\omega)$ is in $T_\varphi N$. Therefore,

$$\mathcal{M}_{\widetilde{F}}(u) = P_\varphi \mathcal{M}_F(\psi), \tag{47}$$

where $\psi = \Pi(\varphi + u)$.

Since the projection P_φ is a linear map that does not depend on u, the complexification of $\mathcal{M}_{\widetilde{F}}(u)$ is reduced to that of $\mathcal{M}_F(\Pi(\varphi+u))$, which we regard as the composition of the following:

(M1) the map $\Pi(\varphi + \cdot)$ from $C^{4,\beta}(V)$ to $C^{4,\beta}(S^{m-1}, \mathbb{R}^p)$, which has been discussed in (F3) above;

(M2) the map $\Delta^2_{S^{m-1}}\psi-(2m-8)\Delta_{S^{m-1}}\psi$ from $C^{4,\beta}(S^{m-1},\mathbb{R}^p)$ to $C^{0,\beta}(S^{m-1},\mathbb{R}^p)$, which has been discussed in (F1) above;

(M3) the projection P_ψ is a p by p matrix that depends analytically on ψ, since N is an analytic submanifold. Keeping in mind that $\psi=\Pi(\varphi+u)$ is known (see (M1) above) to be analytic map in u, the complexification of P_ψ is given by expanding the analytic (matrix-valued) map $P_\psi=P_{\Pi(\varphi+u)}$ as a converging power series of u and then replacing u by $u+iv$ as in (F3).

Properties of the complexification. Let's denote the complexification of $\mathcal{M}_{\widetilde{F}}$ by $\mathcal{M}_{\widetilde{F},C}$. In this section, we study the ellipticity of $\mathcal{M}_{\widetilde{F},C}$ and the self-adjointness of its linearization at 0. Please notice that although the ellipticity of $\mathcal{M}_{\widetilde{F}}$ is quite natural, the ellipticity of $\mathcal{M}_{\widetilde{F},C}$ as an operator between the complexified Banach spaces is not true in general. Fortunately, we have the following:

Lemma A.4. *The linearizations of both $\mathcal{M}_{\widetilde{F}}$ and $\mathcal{M}_{\widetilde{F},C}$ at $u=0$ are elliptic.*

Remark A.5. In fact, as the following proof shows, $\mathcal{M}_{\widetilde{F}}$ is elliptic for small u such that it is defined and $\mathcal{M}_{\widetilde{F},C}$ is elliptic at $u+iv\in C^{4,\beta}(V)\otimes\mathbb{C}$ if $v=0$.

Proof. Neglecting the lower order part, it suffices to compute the linearization of

$$P_\varphi P_\psi \Delta^2_{S^{m-1}}\Pi(\varphi+u),$$

where $\psi=\Pi(\varphi+u)$. If we do the computation at $u\in C^{4,\beta}(V)$ with infinitesimal increment h and neglect all lower order terms, we get

$$P_\varphi P_\psi \Delta^2_{S^{m-1}}h, \tag{48}$$

whose symbol is for any $\xi\in T^*_\omega S^{m-1}$,

$$\xi\mapsto P_\varphi P_\psi|\xi|^4 h. \tag{49}$$

If ξ is not zero, then this is clearly a linear isomorphism from the sections of V onto itself, because ψ is close to φ.

Now, for $\mathcal{M}_{\widetilde{F},C}$, we denote the complexification of P_ψ ($\Pi(\varphi+u)$) by $P_{\psi,C}$ ($\Pi_C(\varphi+u)$ respectively). Although we do not know any exact formula for them, it suffices for us to note that when computing (48),

(1) the contribution of Π_C goes to the lower order terms and does not matter;

(2) since we have assumed that $u\in C^{4,\beta}(V)$, by the definition of complexification, $P_{\psi,C}=P_\psi$. Therefore, we get the same symbol as in (49), which is now an isomorphism from the sections of complexified-V onto itself. □

If we denote the linearizations of $\mathcal{M}_{\widetilde{F}}$ and $\mathcal{M}_{\widetilde{F},C}$ at $u=0$ by $\mathcal{L}_{\widetilde{F}}$ and $\mathcal{L}_{\widetilde{F},C}$, then

Lemma A.6. *For any $u, v \in C^{4,\beta}(V)$,*

$$\mathcal{L}_{\widetilde{F},C}(u + iv) = \mathcal{L}_{\widetilde{F}}(u) + i\mathcal{L}_{\widetilde{F}}(v). \tag{50}$$

In particular, $\mathcal{L}_{\widetilde{F},C}$ is an elliptic and self-adjoint operator from $C^{4,\beta}(V) \otimes \mathbb{C}$ to $C^{0,\beta}(V) \otimes \mathbb{C}$.

Proof. By definition, $\mathcal{L}_{\widetilde{F},C}(u) = d/dt|_{t=0}\mathcal{M}_{\widetilde{F},C}(tu) = d/dt|_{t=0}\mathcal{M}_{\widetilde{F}}(tu) = \mathcal{L}_{\widetilde{F}}(u)$. Hence, it suffices to show

$$\widetilde{\mathcal{L}}_{\widetilde{F},C}(iv) = i\mathcal{L}_{\widetilde{F}}(v).$$

Since $\mathcal{M}_{\widetilde{F}}$ is a composition of P_φ, P_ψ, $\Delta_{S^{m-1}}$, $\nabla_{S^{m-1}}$, and $\Pi(\varphi + \cdot)$, it suffices to show that (50) holds for (the linearization of) each one of them. This is trivial for P_φ, $\Delta_{S^{m-1}}$, and $\nabla_{S^{m-1}}$, because they are linear operators and (50) is exactly how their complexification is defined.

For $\Pi(\varphi + \cdot)$, we recall that

$$\Pi_C(\varphi + (u + iv)) = \sum_k a_k (u + iv)^k$$

and the series converges for small u and v. Equation (50) then follows from direct computation. The same argument works for P_ψ.

The self-adjointness of $\mathcal{L}_{\widetilde{F}}$ follows from expanding the following identity:

$$\frac{d}{ds}\bigg|_{s=0}\frac{d}{dt}\bigg|_{t=0}\widetilde{F}(tu + sv) = \frac{d}{dt}\bigg|_{t=0}\frac{d}{ds}\bigg|_{s=0}\widetilde{F}(tu + sv).$$

The self-adjointness of $\mathcal{L}_{\widetilde{F},C}$ is then a consequence of (50). □

Now, we state the result that motivates the discussion in this section.

Lemma A.7. *For f defined (23), it is an analytic function of ξ in a neighborhood of $0 \in \mathbb{R}^l$.*

Proof. Let $\mathcal{N}_C$ be the complexification of $\mathcal{N}$ defined in Section A. Its linearization at $u = 0$ is given by

$$P_K + \widetilde{\mathcal{L}}.$$

By the results above, this $\widetilde{\mathcal{L}}$ is elliptic and self-adjoint with trivial kernel. Hence, the inverse function theorem (Theorem A.3) gives an inverse map Ψ_C, which is analytic, from a neighborhood of 0 in $C^{0,\beta}(V) \otimes \mathbb{C}$ to a neighborhood of 0 in $C^{4,\beta}(V) \otimes \mathbb{C}$. If $\widetilde{F}_C$ is the complexification of $\widetilde{F}$ given in Section A, then f in (23) is the restriction (to the real part of $(z_1, \dots, z_l)$) of

$$f_C(z_1, \dots, z_l) := \widetilde{F}_C\left(\Psi_C\left(\sum_{i=1}^{l} z_i\varphi_i\right)\right),$$

which is analytic in (a neighborhood of 0 in) $\mathbb{C}^l$. □

Appendix B: Proof of Lemma 3.5

For some $\delta > 0$ to be determined, we will take $W = \{f \mid \|f\|_{C^\beta(V)} < \delta\}$. For any f_1, f_2 in W, we have

$$\Psi(f_1) - \Psi(f_2) = \int_0^1 \frac{d}{dt}\Psi(tf_1 + (1-t)f_2)\,dt = \int_0^1 D\Psi|_{tf_1+(1-t)f_2}(f_1 - f_2)\,dt.$$

Hence, it suffices to show that for any $f \in W$, the linearization of Ψ at f, $D\Psi|_f$ is a uniformly bounded linear operator from $L^2(V)$ to $W^{4,2}(V)$. More precisely, we need to find $\delta > 0$ and $C > 0$ such that

$$\sup_{f\in W} \|D\Psi|_f\|_{L(L^2(V),W^{4,2}(V))} \le C.$$

Here $\|\cdot\|_{L(L^2(V),W^{4,2}(V))}$ is the norm of linear operators.

Since Ψ is the inverse of $\mathcal{N}$, it suffices to show that there exist $\delta' > 0$ and $C > 0$ such that if $W' = \{u \in C^{4,\beta}(V) \mid \|u\|_{C^{4,\beta}(V)} < \delta'\}$,

$$\inf_{u\in W'} \|D\mathcal{N}|_u\|_{L(W^{4,2},L^2)} \ge C > 0. \tag{51}$$

The proof of (51) consists of two steps. First, we show that

$$\|D\mathcal{N}|_0\|_{L(W^{4,2},L^2)} \ge C > 0. \tag{52}$$

Recall that $\mathcal{N} = P_K + \mathcal{M}_{\widetilde{F}}$, where K is the kernel of $D\mathcal{M}_{\widetilde{F}}|_0 = \mathcal{L}_{\widetilde{F}}$. For any $h \in W^{4,2}(V)$, we denote $h - P_K h$ by $h^\perp$. Since $\mathcal{L}_{\widetilde{F}}$ is an elliptic operator with trivial kernel in the compliment space of K, there is a constant depending only on φ such that

$$\|h^\perp\|_{W^{4,2}} \le C_0\|\mathcal{L}_{\widetilde{F}} h^\perp\|_{L^2}. \tag{53}$$

Since K is a finite-dimensional space, there is $C_1 > 0$ such that

$$\|P_K h\|_{W^{4,2}} \le C_1\|P_K h\|_{L^2}. \tag{54}$$

Combining (53) and (54) and noticing that the image of $\mathcal{L}_{\widetilde{F}}$ is normal to K in L^2, we get $C_2 > 0$ such that

$$\|h\|_{W^{4,2}} \le C_2\|D\mathcal{N}|_0 h\|_{L^2}, \tag{55}$$

which implies (52).

The second step is to show that for $u \in W'$,

$$\|(D\mathcal{N}|_u - D\mathcal{N}|_0)h\|_{L^2} \le C(\delta')\|h\|_{W^{4,2}} \tag{56}$$

for some $C(\delta')$ satisfying $\lim_{\delta'\to 0} C(\delta') = 0$. Before the proof of (56), we notice that if δ' is small, (56) and (55) imply that

$$\|h\|_{W^{4,2}} \le C\|D\mathcal{N}|_u h\|_{L^2},$$

which finishes the proof of (51) and hence the proof of Lemma 3.5.

For (56), we notice that the contribution of $P_K h$ cancels out and it suffices to bound

$$\|(D\mathcal{M}_{\widetilde{F}}|_u - D\mathcal{M}_{\widetilde{F}}|_0)h\|_{L^2}. \tag{57}$$

Recalling that $\mathcal{M}_{\widetilde{F}}(u) = P_\varphi P_\psi(\Delta^2_{S^{m-1}}\psi - (2m-8)\Delta_{S^{m-1}}\psi)$ with $\psi = \Pi(\varphi+u)$, we get

$$\begin{aligned} D\mathcal{M}_{\widetilde{F}}|_u h = P_\varphi (DP)_\psi(\Delta^2_{S^{m-1}}\psi - (2m-8)\Delta_{S^{m-1}}\psi)h \\ + P_\varphi P_\psi(\Delta^2_{S^{m-1}}h - (2m-8)\Delta_{S^{m-1}}h) \end{aligned} \tag{58}$$

and

$$\begin{aligned} D\mathcal{M}_{\widetilde{F}}|_0 h = P_\varphi (DP)_\varphi(\Delta^2_{S^{m-1}}\varphi - (2m-8)\Delta_{S^{m-1}}\varphi)h \\ + P_\varphi(\Delta^2_{S^{m-1}}h - (2m-8)\Delta_{S^{m-1}}h). \end{aligned} \tag{59}$$

Notice that (58) and (59) are fourth-order linear operators of h and if we subtract them, the difference of the corresponding coefficients are bounded by using

$$\|\psi - \varphi\|_{C^{4,\beta}} \le C\|u\|_{C^{4,\beta}(V)}.$$

References

[Chang et al. 1999] S.-Y. A. Chang, L. Wang, and P. C. Yang, "A regularity theory of biharmonic maps", *Comm. Pure Appl. Math.* **52**:9 (1999), 1113–1137. MR Zbl

[Cheeger and Tian 1994] J. Cheeger and G. Tian, "On the cone structure at infinity of Ricci flat manifolds with Euclidean volume growth and quadratic curvature decay", *Invent. Math.* **118**:3 (1994), 493–571. MR Zbl

[Colding and Minicozzi 2014] T. H. Colding and W. P. Minicozzi, II, "On uniqueness of tangent cones for Einstein manifolds", *Invent. Math.* **196**:3 (2014), 515–588. MR Zbl

[Dieudonné 1960] J. Dieudonné, *Foundations of modern analysis*, Pure and Applied Math. **10**, Academic Press, New York, 1960. MR Zbl

[Hong and Wang 2005] M.-C. Hong and C. Wang, "Regularity and relaxed problems of minimizing biharmonic maps into spheres", *Calc. Var. Partial Differential Equations* **23**:4 (2005), 425–450. MR Zbl

[Lamm and Rivière 2008] T. Lamm and T. Rivière, "Conservation laws for fourth order systems in four dimensions", *Comm. Partial Differential Equations* **33**:1-3 (2008), 245–262. 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

[Luckhaus 1988] S. Luckhaus, "Partial Hölder continuity for minima of certain energies among maps into a Riemannian manifold", *Indiana Univ. Math. J.* **37**:2 (1988), 349–367. MR Zbl

[Nirenberg 1974] L. Nirenberg, *Topics in nonlinear functional analysis*, Courant Inst. Math. Sci., New York, 1974. MR Zbl

[Scheven 2008] C. Scheven, "Dimension reduction for the singular set of biharmonic maps", *Adv. Calc. Var.* **1**:1 (2008), 53–91. 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

[Simon 1983] L. Simon, "Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems", *Ann. of Math.* (2) **118**:3 (1983), 525–571. MR Zbl

[Simon 1996] L. Simon, *Theorems on regularity and singularity of energy minimizing maps*, Birkhäuser, Basel, 1996. MR Zbl

[Struwe 2008] M. Struwe, "Partial regularity for biharmonic maps, revisited", *Calc. Var. Partial Differential Equations* **33**:2 (2008), 249–262. MR Zbl

[Strzelecki 2003] P. Strzelecki, "On biharmonic maps and their generalizations", *Calc. Var. Partial Differential Equations* **18**:4 (2003), 401–432. MR Zbl

[Taylor 1937] A. E. Taylor, "Analytic functions in general analysis", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (2) **6**:3-4 (1937), 277–292. MR Zbl

[Wang 2004a] C. Wang, "Stationary biharmonic maps from $\mathbb{R}^m$ into a Riemannian manifold", *Comm. Pure Appl. Math.* **57**:4 (2004), 419–444. MR Zbl

[Wang 2004b] C. Wang, "Biharmonic maps from $\mathbb{R}^4$ into a Riemannian manifold", *Math. Z.* **247**:1 (2004), 65–87. MR Zbl

[Wang 2004c] C. Wang, "Remarks on biharmonic maps into spheres", *Calc. Var. Partial Differential Equations* **21**:3 (2004), 221–242. MR Zbl

[Yang 2003] B. Yang, "The uniqueness of tangent cones for Yang–Mills connections with isolated singularities", *Adv. Math.* **180**:2 (2003), 648–691. MR Zbl

Received May 5, 2017. Revised May 15, 2018.

YOUMIN CHEN
SCHOOL OF MATHEMATICAL SCIENCES
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA
HEFEI
CHINA
chym189@mail.ustc.edu.cn

HAO YIN
SCHOOL OF MATHEMATICAL SCIENCES
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA
HEFEI
CHINA
haoyin@ustc.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 2, 2019
dx.doi.org/10.2140/pjm.2019.299.427

HAMILTONIAN UNKNOTTEDNESS OF CERTAIN MONOTONE LAGRANGIAN TORI IN $S^2 \times S^2$

KAI CIELIEBAK AND MARTIN SCHWINGENHEUER

We prove that a monotone Lagrangian torus in $S^2 \times S^2$ which suitably sits in a symplectic fibration with two sections in its complement is Hamiltonian isotopic to the Clifford torus.

1. Introduction

The classification of Lagrangian submanifolds in symplectic manifolds up to Lagrangian or Hamiltonian isotopy is an intriguing problem of symplectic topology. While there are many tools for distinguishing Lagrangian submanifolds, actual classification results have been very rare and restricted to special manifolds in dimension 4. The first circle of results concerns Lagrangian 2-planes and 2-spheres, in which case the two notions of isotopy coincide: up to Hamiltonian isotopy, there is a unique asymptotically linear Lagrangian 2-plane in $\mathbb{R}^4$ [Eliashberg and Polterovich 1996], and a unique Lagrangian 2-sphere in a given homology class in $S^2 \times S^2$ [Hind 2004], in T^*S^2 and some other Stein surfaces [Hind 2012], and in certain blow-ups of $\mathbb{C}P^2$ [Evans 2010; Li and Wu 2012]. See also [Borman, Li, and Wu 2014] for some uniqueness results up to global symplectomorphism. The second circle of results is due to [Dimitroglou Rizell, Goodman and Ivrii 2016] (building on A. Ivrii's Ph.D. thesis [2003]): they prove uniqueness up to Lagrangian isotopy of Lagrangian tori in $\mathbb{R}^4$, $S^2 \times S^2$ and $\mathbb{CP}^2$, and uniqueness up to Hamiltonian isotopy of exact Lagrangian tori in $T^*\mathbb{T}^2$. See [Dimitroglou Rizell, Goodman and Ivrii 2016] for an extensive discussion of the history of this problem.

MSC2010: primary 53D12; secondary 57R17.

Keywords: Hamiltonian isotopy, Lagrangian tori, symplectic fibrations.

Motivated by Ivrii's thesis, we address in this paper the question of *Hamiltonian* unknottedness of monotone Lagrangian tori in $S^2 \times S^2$. Recall that a Lagrangian torus is called *monotone* if its Maslov class is a positive multiple of its symplectic area class on relative π_2. The product of the equators in each S^2-factor in $S^2 \times S^2$ is called the *standard Lagrangian torus* L_{std}, or the *Clifford torus*. This torus is monotone for the standard split symplectic form $\omega_{\mathrm{std}} = \sigma_{\mathrm{std}} \oplus \sigma_{\mathrm{std}}$, where σ_{std} is the standard area form on S^2 normalised by $\int_{S^2} \sigma_{\mathrm{std}} = 1$. Motivated by Chekanov's construction [1996] of exotic monotone Lagrangian tori in $\mathbb{R}^{2n}$, there have been many constructions of monotone Lagrangian tori in $(S^2 \times S^2, \omega_{\mathrm{std}})$ that are not Hamiltonian isotopic to L_{std} due to [Eliashberg and Polterovich 1997; Chekanov and Schlenk 2010; Entov and Polterovich 2009; Biran and Cornea 2009; Fukaya, Oh, Ohta, and Ono 2009; Albers and Frauenfelder 2008]. All of these are Hamiltonian isotopic to each other [Gadbled 2013; Oakley and Usher 2016] and are collectively referred to as the *Chekanov torus*. Recently R. Vianna [2014; 2016; 2017] constructed infinitely many pairwise Hamiltonian nonisotopic monotone Lagrangian tori in $\mathbb{C}P^2$ and $(S^2 \times S^2, \omega_{\mathrm{std}})$, as well as in most other del Pezzo surfaces.

The following definition is implicit in [Dimitroglou Rizell, Goodman and Ivrii 2016]. Let us call a monotone Lagrangian torus L in $(S^2 \times S^2, \omega_{\mathrm{std}})$ *fibred* if there exists a foliation $\mathcal{F}$ of $S^2 \times S^2$ by symplectic 2-spheres in the homology class $[pt \times S^2]$ and a symplectic submanifold Σ in the class $[S^2 \times pt]$ with the following properties:

- Σ is transverse to the leaves of $\mathcal{F}$ and is disjoint from L.
- The leaves of $\mathcal{F}$ intersect L in circles (or not at all).

Note that each leaf of $\mathcal{F}$ which intersects the torus L is cut by L into two closed disks glued along L. The disks that intersect Σ form a solid torus T with $\partial T = L$.

Theorem D in [loc. cit.] asserts that *any monotone Lagrangian torus in* $S^2 \times S^2$ *is fibred*. In this paper, we prove:

Theorem 1.1 (main theorem). *Let $L \subset (S^2 \times S^2, \omega_{\mathrm{std}})$ be a monotone Lagrangian torus which is fibred by $\mathcal{F}$ and Σ. Assume in addition that there exists a second symplectic submanifold Σ' in the homology class $[S^2 \times pt]$ which is transverse to the leaves of $\mathcal{F}$, and which is disjoint from Σ and T. Then L is Hamiltonian isotopic to the standard torus L_{std}.*

Hence the presence of the second section Σ' characterises the standard torus among Vianna's infinitely many monotone Lagrangian tori in $(S^2 \times S^2, \omega_{\mathrm{std}})$. It suggests that the classification of these tori up to Hamiltonian isotopy may come within reach once we understand better the role of the second section Σ'. See [loc. cit.] for a homological reformulation of the presence of the section Σ', as well as an explicit presentation of the Chekanov torus as a fibred torus.

Remark 1.2. As explained in [Dimitroglou Rizell, Goodman and Ivrii 2016], Theorem 1.1 can be used to prove uniqueness up to Hamiltonian isotopy of exact Lagrangian tori in $T^*\mathbb{T}^2$: one realises $\left(-\frac{\pi}{4}, \frac{\pi}{4}\right)^2 \times T^2 \subset T^*\mathbb{T}^2$ as the complement of two sections Σ, Σ' and two fibres F, F' in $(S^2 \times S^2, \omega_{\mathrm{std}})$ and applies Theorem 1.1 to an exact Lagrangian torus in $\left(-\frac{\pi}{4}, \frac{\pi}{4}\right)^2 \times T^2$, noting that the Hamiltonian isotopy in Theorem 1.1 can be chosen to fix the sections Σ, Σ' and the fibres F, F'. See [loc. cit., Section 7] for more details.

Let us now outline the proof of the main theorem, and in particular explain where the second section is needed. By a *relative symplectic fibration* on $S^2 \times S^2$ we will mean a quintuple

$$\mathfrak{S} = (\mathcal{F}, \omega, L, \Sigma, \Sigma'),$$

as in Theorem 1.1, only with the standard form ω_{std} replaced by any symplectic form ω cohomologous to ω_{std}. We will prove (Corollary 4.13) that for every symplectic fibration $\mathfrak{S}$ with $\omega = \omega_{\mathrm{std}}$ there exists a homotopy of relative symplectic fibrations $\mathfrak{S}_t = (\mathcal{F}_t, \omega_{\mathrm{std}}, L_t, \Sigma_t, \Sigma_t')$ with fixed symplectic form ω_{std} such that $\mathfrak{S}_0 = \mathfrak{S}$ and $\mathfrak{S}_1 = \mathfrak{S}_{\mathrm{std}} := (\mathcal{F}_{\mathrm{std}}, \omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$, where $\mathcal{F}_{\mathrm{std}}$ denotes the standard foliation with leaves $\{z\} \times S^2$ and $S_0 = S^2 \times \{S\}$, $S_\infty = S^2 \times \{N\}$ are the standard sections at the south and north poles. Then L_t is an isotopy of monotone Lagrangian tori with respect to ω_{std} from L to L_{std}, which is Hamiltonian by Banyaga's isotopy extension theorem.

A relative symplectic fibration $\mathfrak{S}$ gives rise to a symplectic fibration $p : S^2 \times S^2 \to \Sigma$ by sending each leaf of $\mathcal{F}$ to its intersection point with Σ. It determines a symplectic connection whose horizontal subspaces are the symplectic orthogonal complements to the fibres. Parallel transport along closed paths $\gamma : [0, 1] \to \Sigma$ gives holonomy maps which are symplectomorphisms of the fibre $p^{-1}(\gamma(0))$ and measure the nonintegrability of the horizontal distribution. It is not hard to show that a symplectic fibration $\mathfrak{S}$ with trivial holonomy around all loops is diffeomorphic to $\mathfrak{S}_{\mathrm{std}}$, and a theorem of Gromov implies that they are actually homotopic with fixed symplectic form if they both have symplectic form ω_{std}.

Thus, most of the work will go into deforming a given relative symplectic fibration $\mathfrak{S}$ to one with trivial holonomy. After pulling back $\mathfrak{S}$ by a diffeomorphism, we may assume that $(\mathcal{F}, L, \Sigma, \Sigma') = (\mathcal{F}_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$ (but the symplectic form ω is nonstandard). In the first step, which takes up Section 3, we make the holonomy trivial near the two sections and near the fibres over the line of longitude m_0 through Greenwich in the base; see Figure 1.

In the second step, which takes up most of Section 4, we kill the holonomy along all circles of constant latitude C^λ. For this, let (λ, μ) be spherical coordinates on S^2, where λ denotes the latitude and μ denotes the longitude. After the first step, the holonomy maps ϕ^λ along C^λ give a loop in $\mathrm{Symp}(A, \partial A, \sigma_{\mathrm{std}})$, the group of

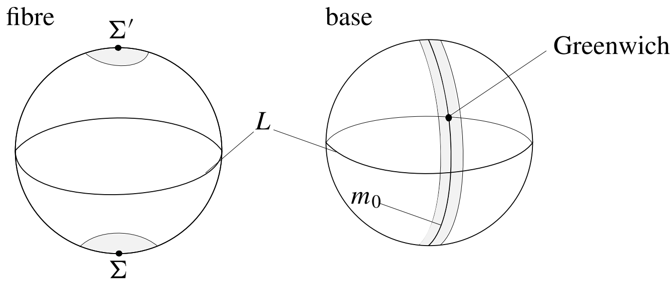

Figure 1. Where the holonomy is trivial after the first step.

symplectomorphisms of the annulus (the sphere minus two polar caps) which equal the identity near the boundary. Since the fundamental group of $\mathrm{Symp}(A, \partial A, \sigma_{\mathrm{std}})$ vanishes, we can contract the loop of inverses $\psi^\lambda = (\phi^\lambda)^{-1}$ and obtain a family of Hamiltonians H^λ_μ which generates the contraction. The closed 2-form

$$\Omega_H = \omega + d(H^\lambda_\mu d\mu)$$

then defines a symplectic connection with trivial holonomy around all C^λ. However, Ω_H need not be symplectic if $\partial H^\lambda_\mu / \partial \lambda$ is large. This can be remedied by the inflation procedure due to [Lalonde and McDuff 1996]. In this procedure, the symplectic form ω is deformed along a fibre and a section (and H suitably rescaled) in order to make Ω_H symplectic. However, this process will in general destroy monotonicity of L_{std}. In order to keep the Lagrangian torus monotone, we perform the inflation procedure along a fibre and the *two* sections S_0, S_∞ in a symmetric way. It is at this point of the proof that we need the existence of a second symplectic section.

Once the holonomy along circles of latitude is trivial, in the third and final step (at the end of Section 4) we deform the symplectic form to the standard form. This finishes the outline of the proof.

Remark 1.3. The idea to apply the results of Ivrii's thesis to the Hamiltonian classification of monotone tori in $S^2 \times S^2$ originated in 2003 in the first author's discussions with Y. Eliashberg. However, at the time we did not realise the necessity of a second symplectic section and were puzzled by the apparent contradiction between this result and the existence of an exotic monotone torus in $S^2 \times S^2$. This discrepancy was resolved in the second author's PhD thesis [Schwingenheuer 2010], of which this article is a shortened version.

2. Relative symplectic fibrations

2A. ***Symplectic connections and their holonomy.*** Consider a smooth fibration (by which we mean a fibre bundle) $p : M \to B$ and a closed 2-form ω on M whose restriction to each fibre $p^{-1}(b)$ is nondegenerate. We will refer to ω as a *symplectic* *connection* on M.[1] From the next subsection on we will assume ω to be symplectic, but for now this is not needed.

Parallel transport. Since ω is nondegenerate on the fibres, the ω-orthogonal complements

$$\mathcal{H}_x := (\ker d_x p)^{\omega}$$

to the tangent spaces of the fibres of p define a distribution of horizontal subspaces $\mathcal{H}$ such that

$$TM = \mathcal{H} \oplus \ker dp.$$

Horizontal lifts of a path $\gamma : [0, 1] \to B$ with given initial points in $p^{-1}(\gamma(0))$ give rise to the *parallel transport*

$$P_\gamma : p^{-1}(\gamma(0)) \to p^{-1}(\gamma(1))$$

along γ. Closedness of ω implies that P_γ is symplectic, i.e.,

$$P_\gamma^* \omega_{\gamma(1)} = \omega_{\gamma(0)},$$

where ω_b denotes the symplectic form $\omega|_{p^{-1}(b)}$.

Holonomy. The parallel transport $P_\gamma : p^{-1}(\gamma(0)) \to p^{-1}(\gamma(0))$ along a closed curve $\gamma : [0, 1] \to B$ is called the *holonomy* of ω along the loop γ. If $P_\gamma = \mathrm{id}$ for each loop γ, we say that ω has *trivial holonomy*. In this case, parallel transport along any (not necessarily closed) curve depends only on the end points, so we can use parallel transport to define local trivializations of $p : M \to B$.

Remark 2.1. There is a natural notion of *curvature* of a symplectic connection; see [McDuff and Salamon 1995]. This is a 2-form on the base with values in the functions on the fibres which measures the nonintegrability of the horizontal distribution. For simply connected base (which is the case of interest to us) the curvature and the holonomy carry the same information, so in this paper we will phrase everything in terms of holonomy.

From foliations to fibrations. More generally, we can consider a closed 2-form ω on M whose restriction to the leaves of a smooth foliation $\mathcal{F}$ of M is nondegenerate. If all leaves of $\mathcal{F}$ are compact, then the space of leaves is a smooth manifold B and the canonical projection $M \to B$ is a fibration, so we are back in the situation of a symplectic connection as above. Since in our case all leaves will be 2-spheres, we can switch freely between the terminologies of foliations and fibrations.

2B. *Fibered Lagrangian tori in $S^2 \times S^2$.* Suppose now that (M, ω) is a symplectic 4-manifold and $p : M \to B$ is a symplectic fibration over a surface B (i.e., the fibres are symplectic surfaces).

[1] This terminology differs slightly from the one in [McDuff and Salamon 1995].

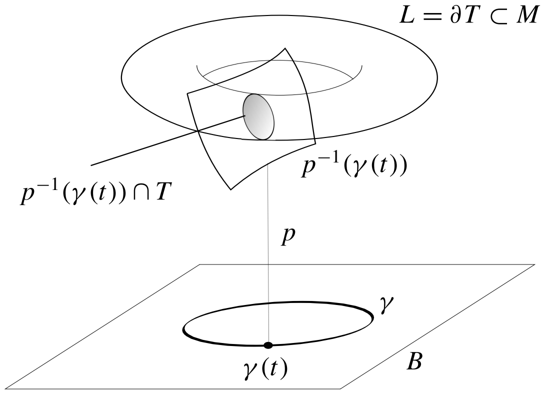

Figure 2. A fibred Lagrangian torus.

Definition 2.2. We say that an embedded 2-torus $L \subset M$ is *fibred by* p if (see Figure 2):

(i) $\gamma := p(L)$ is an immersed loop with transverse self-intersections which are at most double points.

(ii) $p^{-1}(\gamma(t)) \cap L$ is diffeomorphic to a circle if $\gamma(t)$ is not a double point, and to two disjoint circles if $\gamma(t)$ is a double point.

(iii) In each of the circles in $p^{-1}(\gamma(t)) \cap L$ we can fill in an embedded disk $D \subset p^{-1}(\gamma(t))$ in the fibre such that the two disks at a double point are disjoint and all the disks form a solid torus $T \cong S^1 \times D^2$ with L as its boundary.

Suppose now that L is in addition Lagrangian. The following two results are the basis for most of the sequel. The first one states that a fibred Lagrangian torus L is generated by parallel transport along γ of the circle in the fibre over a nondouble point; see Figure 3.

Proposition 2.3. *Let $L \subset M$ be an embedded Lagrangian torus which is fibred by the symplectic fibration $p : M \to B$. Then L is invariant under parallel transport along $\gamma = p(L)$ with respect to the symplectic connection ω.*

Proof. At a point $x \in L$ we have the ω-orthogonal splitting $T_x M = \mathcal{H}_x \oplus V_x$, where $V_x := \ker d_x p$ denotes the tangent space to the fibre. Since L is fibred by p, the subspace $T_x L + V_x \subset T_x M$ generated by $T_x L$ and V_x is 3-dimensional. The condition that L is Lagrangian implies $(T_x L)^\omega = T_x L$; thus $T_x L \cap \mathcal{H}_x = (T_x L)^\omega \cap (V_x)^\omega = (T_x L + V_x)^\omega$ is 1-dimensional. This 1-dimensional subspace therefore contains the horizontal lift of $\dot\gamma$ through x and the proposition follows. $\square$

Remark 2.4. Let $N := p^{-1}(\gamma)$ be the 3-dimensional submanifold of M formed by the fibres that meet the torus L. In the definition of being fibred by p we did

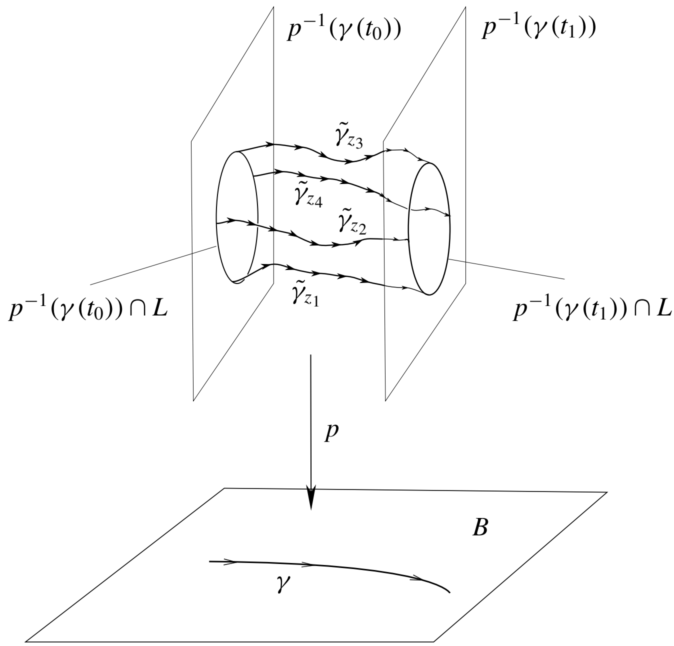

Figure 3. L is generated by parallel transport.

not require the torus to be transverse to the fibres of p in N. If, however, L is Lagrangian, then Proposition 2.3 shows that we get this property for free.

Monotone tori in $S^2 \times S^2$. From now on we assume that

$$M = S^2 \times S^2$$

and the symplectic form ω is cohomologous to the product form

$$\omega_{\text{std}} := \sigma_{\text{std}} \oplus \sigma_{\text{std}},$$

where σ_{std} is the standard area form on S^2 normalised by $\int_{S^2} \sigma_{\text{std}} = 1$. In other words, we require that

$$\int_{S^2 \times \text{pt}} \omega = \int_{\text{pt} \times S^2} \omega = 1.$$

Moreover, we assume that the Lagrangian torus L is *monotone*, i.e., its Maslov class μ (see [McDuff and Salamon 1995]) and its symplectic area satisfy

$$\mu(\sigma) = 4 \int_\sigma \omega \quad \text{for all } \sigma \in \pi_2(M, L).$$

Here the monotonicity constant must be equal to 4 because the class $A = [S^2 \times \text{pt}] \in \pi_2(M)$ has Maslov index $\mu(A) = 2\langle c_1(TM), A\rangle = 2\langle c_1(TS^2), [S^2]\rangle = 4$.

Lemma 2.5. *Let $L \subset (M = S^2 \times S^2, \omega)$ be a monotone Lagrangian torus with ω cohomologous to ω_{std}. Let $p : M \to B$ be a symplectic fibration over the surface B such that L is fibred by p. Then the loop $\gamma := p(L)$ is an embedded curve*; *i.e.*, *it has no double points.*

Proof. Since L is orientable, all its Maslov indices on $\pi_2(M, L)$ are even integers. In view of the monotonicity constant 4, this implies that the symplectic area of each embedded symplectic disk $D \subset M$ with boundary on L must be a positive multiple of $\frac{1}{2}$. If γ had a double point b, then the solid torus T from Definition 2.2 would intersect the fibre $p^{-1}(b)$ in two disjoint symplectic disks, which is impossible because the fibre has symplectic area 1. □

Remark 2.6. (a) For a smooth fibration $p : S^2 \times S^2 \to B$ over a surface B, both the fibres and the base are diffeomorphic to S^2. Indeed, denoting a fibre by F, the homotopy exact sequence $\pi_2(F) \to \pi_2(S^2 \times S^2) \to \pi_2(B)$ implies $\pi_2(F) \cong \pi_2(B) \cong \mathbb{Z}$, so F and B must be diffeomorphic to S^2 or $\mathbb{R}P^2$. Since by the product formula for the Euler characteristic $\chi(F)\chi(B) = \chi(S^2 \times S^2) = 4$, both F and B must be diffeomorphic to S^2.

(b) For a *monotone* Lagrangian torus L in $M = S^2 \times S^2$, the third condition in Definition 2.2 is actually a consequence of the first two. To see this, note first that in the proof of Lemma 2.5 we can rule out the double point b without reference to the solid torus T: by the Jordan curve theorem, the two circles in $L \cap p^{-1}(b)$ would bound two disjoint symplectic disks in the fibre $p^{-1}(b) \cong S^2$, each of area a positive multiple of $\frac{1}{2}$, which again gives the desired contradiction. Now an orientation of L and a parametrisation of the curve $\gamma \subset B$ induce via horizontal lifts of $\dot{\gamma}$ orientations of the circles $L_t := L \cap p^{-1}(\gamma(t))$, and we define T as the union of the disks $D_t \subset p^{-1}(\gamma(t))$ whose oriented boundary is L_t.

2C. *Relative symplectic fibrations of $S^2 \times S^2$*. We continue with the manifold $M = S^2 \times S^2$ and the generators

$$A = [S^2 \times \text{pt}], \; B = [\text{pt} \times S^2] \in H_2(M).$$

Now we define the main object of study for this paper.

Definition 2.7. A *relative symplectic fibration* on $M = S^2 \times S^2$ is a quintuple $\mathfrak{S} = (\mathcal{F}, \omega, L, \Sigma, \Sigma')$, where

- $\mathcal{F}$ is a smooth foliation of M by 2-spheres in homology class B;
- ω is a symplectic form on M making the leaves of $\mathcal{F}$ symplectic with $\omega(A) = \omega(B) = 1$;
- Σ, Σ' are disjoint symplectic submanifolds in class A which are transverse to all the leaves of $\mathcal{F}$, so in particular the projection $p : M \to \Sigma$ sending each leaf to its unique intersection point with Σ defines a symplectic fibration;

- $L \subset M$ is an embedded monotone Lagrangian torus fibred by p;
- Σ' is disjoint from the solid torus T with $\partial T = L$ in Definition 2.2;
- Σ intersects each fibre $p^{-1}(\gamma(t))$ in the interior of the disk $T \cap p^{-1}(\gamma(t))$.

Note that for a monotone Lagrangian torus L fibred by $p : M \to B$ there always exist disjoint smooth sections Σ, Σ' of p with Σ' disjoint from the solid torus T and $\Sigma \cap p^{-1}(\gamma)$ contained in the interior of T. The crucial condition in Definition 2.7 is that these sections can be chosen to be symplectic.

Definition 2.8. (a) A *homotopy* of relative symplectic fibrations is a smooth 1-parametric family

$$\mathfrak{S}_t = (\mathcal{F}_t, \omega_t, L_t, \Sigma_t, \Sigma'_t), \quad t \in [0, 1],$$

of relative symplectic fibrations.

(b) The group $\mathrm{Diff}_{\mathrm{id}}(M)$ of diffeomorphisms $\phi : M \to M$ inducing the identity on the second homology group $H_2(M)$ (and hence on all homology groups) acts on relative symplectic fibrations by pushforward

$$\phi(\mathfrak{S}) := (\phi(\mathcal{F}), \phi_*\omega, \phi(L), \phi(\Sigma), \phi(\Sigma')).$$

Two relative symplectic fibrations $\mathfrak{S}$ and $\widetilde{\mathfrak{S}}$ are called *diffeomorphic* if $\widetilde{\mathfrak{S}} = \phi(\mathfrak{S})$ for a diffeomorphism ϕ of M (which then necessarily belongs to $\mathrm{Diff}_{\mathrm{id}}(M)$).

(c) Two relative symplectic fibrations $\mathfrak{S}$ and $\widetilde{\mathfrak{S}}$ on M are called *deformation equivalent* if there exists a diffeomorphism $\phi \in \mathrm{Diff}_{\mathrm{id}}(M)$ such that $\phi(\mathfrak{S})$ is homotopic to $\widetilde{\mathfrak{S}}$.

Remark 2.9. (a) Note that a diffeomorphism $\phi \in \mathrm{Diff}_{\mathrm{id}}(M)$ intertwines the symplectic connections of $\mathfrak{S}$ and $\phi(\mathfrak{S})$ and their parallel transports. For example, $\mathfrak{S}$ has trivial holonomy if and only if $\phi(\mathfrak{S})$ does.

(b) It is easy to see that deformation equivalence is an equivalence relation. Moreover, $\mathfrak{S}$ is deformation equivalent to $\widetilde{\mathfrak{S}}$ if and only if there exists a homotopy $\mathfrak{S}_t$ such that $\mathfrak{S}_0 = \mathfrak{S}$ and $\mathfrak{S}_1$ is diffeomorphic to $\widetilde{\mathfrak{S}}$.

(c) Note that in the above definition nothing is said about the isotopy class of the diffeomorphism ϕ. In fact, it is an open problem whether every $\phi \in \mathrm{Diff}_{\mathrm{id}}(M)$ is isotopic to the identity, so we do not know whether diffeomorphic relative symplectic fibrations are homotopic in general. However, by a theorem of Gromov (see Theorem 4.12 below), two diffeomorphic relative symplectic fibrations *with the same symplectic form* ω_{std} are homotopic. This result will be crucial at the end of the proof of our main theorem.

2D. ***The standard relative symplectic fibration.*** The *standard relative symplectic fibration* $\mathfrak{S}_{\mathrm{std}} = (\mathcal{F}_{\mathrm{std}}, \omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$ of $S^2 \times S^2$ consists of the following data:

- $\mathcal{F}_{\mathrm{std}}$ is the foliation by the fibres $\{z\} \times S^2$ of the projection $p_1 : S^2 \times S^2 \to S^2$ onto the first factor.
- $\omega_{\mathrm{std}} = \sigma_{\mathrm{std}} \oplus \sigma_{\mathrm{std}}$ is the standard symplectic form.
- $S_0 = S^2 \times \{S\}$ and $S_\infty = S^2 \times \{N\}$, where $N, S \in S^2$ are the north and south poles.
- $L_{\mathrm{std}} = E \times E$ is the Clifford torus, i.e., the product of the equators in the base and fibre.
- $T_{\mathrm{std}} = E \times D_{\mathrm{lh}}$, where $D_{\mathrm{lh}} \subset S^2$ denotes the lower hemisphere, is the solid torus bounded by L_{std}.

The main goal of this paper will be to deform a given relative symplectic fibration to the standard one (see Theorem 4.11 below). For later use, let us record the relative homology and homotopy groups of the Clifford torus.

Lemma 2.10. *For the Clifford torus $L_{\mathrm{std}} \subset S^2 \times S^2$ the second relative homotopy/homology group*

$$\pi_2(S^2 \times S^2, L_{\mathrm{std}}) \cong H_2(S^2 \times S^2, L_{\mathrm{std}}) \cong H_2(S^2 \times S^2) \oplus H_1(T^2)$$

is free abelian generated by

$$[S^2 \times \mathrm{pt}], \quad [\mathrm{pt} \times S^2], \quad [D_{\mathrm{lh}} \times \mathrm{pt}], \quad [\mathrm{pt} \times D_{\mathrm{lh}}].$$

Proof. The long exact sequences of the pair $(M = S^2 \times S^2, L = L_{\mathrm{std}})$ and the Hurewicz maps yield the commuting diagram

$$\begin{tikzcd}
0 \arrow[r] \arrow[d] & \pi_2(M) \arrow[r] \arrow[d, "\cong"] & \pi_2(M, L) \arrow[r] \arrow[d] & \pi_1(L) \arrow[r] \arrow[d, "\cong"] & 0 \arrow[d] \\
H_2(L) \arrow[r, "0"] & H_2(M) \arrow[r] & H_2(M, L) \arrow[r] & H_1(L) \arrow[r] & 0
\end{tikzcd}$$

Here the first horizontal map in the lower row is zero because L bounds the solid torus $T_{\mathrm{std}} = E \times D_{\mathrm{lh}}$ in $S^2 \times S^2$, where $E \subset S^2$ denotes the equator. Now the middle vertical map is an isomorphism by the five lemma, and the generators of $H_2(M, L)$ are obtained from the generators $[S^2 \times \mathrm{pt}]$, $[\mathrm{pt} \times S^2]$ of $H_2(M)$ and $[E \times \mathrm{pt}]$, $[\mathrm{pt} \times E]$ of $H_1(L)$. □

3. Standardisations

In this section we show how to deform a relative symplectic fibration to make it split (in a sense defined below) near the symplectic sections Σ, Σ' and near one fibre F. In particular, the standardised fibration will have trivial holonomy in these regions. This provides a convenient setup for the discussion in Section 4.

3A. ***Pullback by diffeomorphisms.*** In this subsection, we show how to put a relative symplectic fibration $\mathfrak{S}$ into a nicer form via pullback by diffeomorphisms. Note that this is not really changing $\mathfrak{S}$ but just looking at it from a different angle. We will see that using pullbacks we can either standardise all data except the symplectic form ω, or all data except the foliation $\mathcal{F}$. So the nontriviality of a relative symplectic fibration only arises from the interplay of ω and $\mathcal{F}$, as measured by the holonomy of the corresponding symplectic connection.

In order to establish a clean picture of what can be achieved by pullbacks, we will prove some results in stronger versions than what we actually need in the sequel.

Fixing the fibration. We begin with a useful characterisation of diffeomorphisms that are trivial on homology. Recall the generators $A = [S^2 \times \mathrm{pt}]$ and $B = [\mathrm{pt} \times S^2]$ of $H_2(S^2 \times S^2)$.

Lemma 3.1. *A diffeomorphism ϕ of $S^2 \times S^2$ is trivial on homology if and only if it is orientation-preserving and satisfies $\phi_*(B) = B$.*

Proof. The "only if" is clear, so let us prove the "if". Let us write $\phi_*(A) = mA + nB$ for integers m, n. Since ϕ is orientation-preserving, it preserves the intersection form on $H_2(S^2 \times S^2)$ and we obtain

$$1 = \phi_*(A) \cdot \phi_*(B) = (mA + nB) \cdot B = m,$$
$$0 = \phi_*(A) \cdot \phi_*(A) = (A + nB) \cdot (A + nB) = 2n.$$

This shows that $\phi_* A = A$, so ϕ_* is the identity on $H_2(S^2 \times S^2)$. □

Our first normalisation result is:

Proposition 3.2 (fixing the fibration). *Let $\mathfrak{S} = (\mathcal{F}, \omega, L, \Sigma, \Sigma')$ be a relative symplectic fibration of $M = S^2 \times S^2$. Then there exists a diffeomorphism $\phi \in$ $\mathrm{Diff}_{\mathrm{id}}(M)$ such that $\phi^{-1}(\mathfrak{S}) = (\mathcal{F}_{\mathrm{std}}, \phi^*\omega, L_{\mathrm{std}}, S_0, S_\infty)$.*

Proof. Consider the fibration $p : M \to \Sigma$ defined by $\mathcal{F}$ and pick an orientation-preserving diffeomorphism $u : \Sigma \to S^2$. Then $u \circ p : M \to S^2$ is a fibration by 2-spheres. Since $\pi_1 \mathrm{Diff}_+(S^2)$ classifies S^2-bundles over S^2 and $\mathrm{Diff}_+(S^2)$ deformation retracts onto SO(3), there are up to bundle isomorphism precisely two S^2-bundles over S^2: the trivial one $p_1 : S^2 \times S^2 \to S^2$ and a nontrivial one $X \to S^2$. The total space X of the nontrivial bundle is the blow-up of $\mathbb{C}P^2$ at one point, which is not diffeomorphic to $S^2 \times S^2$ (e.g., their intersection forms differ). Thus the nontrivial bundle does not occur, and we conclude that there exists a diffeomorphism $\phi : M \to M$ such that the following diagram commutes:

$$\begin{tikzcd} S^2 \times S^2 \arrow[r, "\phi^{-1}"] \arrow[d, "p"] & S^2 \times S^2 \arrow[d, "p_1"] \\ \Sigma \arrow[r, "u"] & S^2 \end{tikzcd}$$

Moreover, the restriction of ϕ to each fibre is orientation-preserving, which implies the ϕ itself is orientation-preserving. Since the fibres of p and p_1 all represent the homology class B, it follows that $\phi_* B = B$ and thus $\phi \in \mathrm{Diff}_{\mathrm{id}}(M)$ by Lemma 3.1. After replacing $\mathfrak{S}$ by $\phi^{-1}(\mathfrak{S})$, we may hence assume that $\mathcal{F} = \mathcal{F}_{\mathrm{std}}$ and $p = p_1$.

The section Σ of $p_1 : S^2 \times S^2 \to S^2$ can be uniquely parametrised by $z \mapsto (z, f(z))$ for a smooth map $f : S^2 \to S^2$. After a preliminary isotopy we may assume that $f(z_0) = S$ equals the south pole S at a base point $z_0 \in S^2$. Then f represents a class in $\pi_2(S^2, S)$. Now by Hurewicz's theorem, $\pi_2(S^2, S) \cong H_2(S^2)$. Since $[\Sigma] = A = [S^2 \times \mathrm{pt}]$, the class of f is trivial in $H_2(S^2)$, and thus in $\pi_2(S^2, S)$, so that f is nullhomotopic. By smooth approximation, we find a smooth homotopy f_t from the constant map $f_0 \equiv S$ to $f_1 = f$. Now we use (a fibred version of) the isotopy extension theorem to extend the family of embeddings $S^2 \times \{S\} \hookrightarrow M$, $(z, S) \mapsto (z, f_t(z))$ to a family of fibre-preserving diffeomorphisms $\phi_t : M \to M$ with $\phi_0 = \mathrm{id}$. After replacing $\mathfrak{S}$ by $\phi_1^{-1}(\mathfrak{S})$, we may hence assume that $\Sigma = S^2 \times \{S\} = S_0$. Now we repeat the same argument with Σ' (this time it is even simpler because $S^2 \setminus \{S\}$ is contractible) to arrange $\Sigma' = S^2 \times \{N\} = S_\infty$.

Now the torus L is fibred by $p_1 : S^2 \times (S^2 \setminus \{N, S\}) \to S^2$. By an isotopy of the base S^2 we can move the embedded curve $p_1(L)$ to the equator $E \subset S^2$. Let $e(t)$, $t \in \mathbb{R}/\mathbb{Z}$, be a parametrisation of the equator E and consider the loop of embedded closed curves $\Lambda_t := L \cap p_1^{-1}(e(t))$ in the fibre S^2. After a further homotopy we may assume that $\Lambda_0 = E$. Pick a smooth family of diffeomorphisms $g_t : S^2 \to S^2$, $t \in [0, 1]$, such that $g_0 = \mathrm{id}$ and $g_t(E) = \Lambda_t$ for all t. Moreover, we can arrange that $g_t(N) = N$ and $g_t(S) = S$ for all t. Then g_1 satisfies $g_1(E) = E$ as well as $g_1(N) = N$ and $g_1(S) = S$. We can alter g_t so that g_1 fixes E pointwise. By [Smale 1959, Theorem B], the group $\mathrm{Diff}(D^2, \partial D^2)$ of diffeomorphisms of the disk that are the identity near the boundary is contractible. So we can alter g_t further (applying this to the upper and lower hemispheres) so that $g_1 = \mathrm{id}$. This may first destroy the conditions $g_t(N) = N$ and $g_t(S) = S$, but they can be reinstalled by a further alteration. Now we again use (a fibred version of) the isotopy extension theorem to extend the embedding $E \times S^2 \hookrightarrow M$, $(e(t), w) \mapsto (e(t), g_t(w))$, to a fibre-preserving diffeomorphism $\phi : M \to M$ isotopic to the identity. Then $\phi^{-1}(\mathfrak{S})$ has the desired properties and the proposition is proved. □

Proposition 3.2 has the following 1-parametric version. Let us fix closed neighbourhoods $U_F \times S^2$ of the fibre $F = \{z_0\} \times S^2$ and $S^2 \times (U_0 \amalg U_\infty \amalg U_E)$ of $S_0 \amalg S_\infty \amalg (S^2 \times E)$, where U_F, U_0, U_∞ are disks containing z_0, S, N and

$U_E \subset S^2$ is an annulus containing the equator E. For a relative symplectic fibration $\mathfrak{S} = (\mathcal{F}, \omega, L, \Sigma, \Sigma')$ we set $\mathfrak{S}^{\text{top}} := (\mathcal{F}, L, \Sigma, \Sigma')$.

Proposition 3.3 (fixing the fibration — parametric version). *Let $\mathfrak{S}_t = (\mathcal{F}_t, \omega_t, L_t, \Sigma_t, \Sigma_t')_{t\in[0,1]}$ be a homotopy of relative symplectic fibrations of $M = S^2 \times S^2$. Then there exists an isotopy of diffeomorphisms $\phi_t \in \mathrm{Diff}_{\mathrm{id}}(M)$ with $\phi_0 = \mathrm{id}$ such that $\phi_t^{-1}(\mathfrak{S}_t) = (\mathcal{F}_0, \phi_t^*\omega_t, L_0, \Sigma_0, \Sigma_0')$. Moreover, we can arrange the following properties*:

(i) *If $\mathfrak{S}_t^{\text{top}} = \mathfrak{S}_{\text{std}}^{\text{top}}$ near $U_F \times S^2$ for all t, then $\phi_t = \mathrm{id}$ on $U_F \times S^2$ for all t.*

(ii) *If $\mathfrak{S}_t^{\text{top}} = \mathfrak{S}_{\text{std}}^{\text{top}}$ near $S^2 \times U_E$ for all t, then $\phi_t = \mathrm{id}$ on $S^2 \times U_E$ for all t. If in addition $\mathfrak{S}_0^{\text{top}} = \mathfrak{S}_1^{\text{top}} = \mathfrak{S}_{\text{std}}^{\text{top}}$ near $S^2 \times (U_0 \amalg U_\infty)$, then $\phi_1(z, w) = (f_0(z), w)$ for $(z, w) \in S^2 \times U_0$ and $\phi_1(z, w) = (f_\infty(z), w)$ for $(z, w) \in S^2 \times U_\infty$ with $f_0, f_\infty \in \mathrm{Diff}_+(S^2)$.*

Proof. Pick a smooth family of sections $\widetilde{\Sigma}_t$ of $\mathcal{F}_t$ (e.g., $\widetilde{\Sigma}_t = \Sigma_t$) and orientation-preserving diffeomorphisms $u_t : \widetilde{\Sigma}_t \to S^2$. Composing the u_t with the projections $M \to \widetilde{\Sigma}_t$ along $\mathcal{F}_t$ yields a family of 2-sphere bundles $p_t : M \to S^2$ whose fibres are the leaves of $\mathcal{F}_t$. By the covering homotopy theorem [Steenrod 1951, Theorem 11.4], the 2-sphere bundle

$$P : [0, 1] \times M \to [0, 1] \times S^2, \quad (t, x) \mapsto (t, p_t(x)),$$

is equivalent to the pullback of the bundle $p_0 : M \to S^2$ under the projection $[0, 1] \times S^2 \to S^2$; i.e., there exists a diffeomorphism $\Phi : [0, 1] \times M \to [0, 1] \times M$ such that the following diagram commutes:

$$\begin{tikzcd}
{[0,1] \times M} \arrow[r, "\Phi^{-1}"] \arrow[d, "P"] & {[0,1] \times M} \arrow[d, "\mathrm{id}\times p_0"] \\
{[0,1] \times S^2} \arrow[r, "\mathrm{id}"] & {[0,1] \times S^2}
\end{tikzcd}$$

It follows that $\Phi(t, x) = (t, \phi_t(x))$ for diffeomorphisms $\phi_t : M \to M$ with $\phi_0 = \mathrm{id}$ and $p_t \circ \phi_t = p_0$; hence $\phi_t^{-1}(\mathcal{F}_t) = \mathcal{F}_0$. After pulling back $\mathfrak{S}_t$ by ϕ_t we may therefore assume that $\mathcal{F}_t = \mathcal{F}_0$ for all t. The data $(L_t, \Sigma_t, \Sigma_t')$ is dealt with as in the proof of Proposition 3.2 and can thus be pulled back to $(L_0, \Sigma_0, \Sigma_0')$. It remains to arrange the additional properties (i) and (ii).

For (i), suppose that $\mathfrak{S}_t^{\text{top}} = \mathfrak{S}_{\text{std}}^{\text{top}}$ on $\widetilde{U}_F \times S^2$ for some open neighbourhood $\widetilde{U}_F$ of U_F. Then we can choose the sections $\widetilde{\Sigma}_t$ such that $\widetilde{\Sigma}_t \cap (\widetilde{U}_F \times S^2) = \widetilde{U}_F \times \{w_0\}$ for some $w_0 \in S^2$, and the diffeomorphisms u_t to restrict to the identity map $\widetilde{U}_F \times \{w_0\} \to \widetilde{U}_F$. It follows that $p_t|_{\widetilde{U}_F \times S^2} : \widetilde{U}_F \times S^2 \to \widetilde{U}_F$ is the projection onto the first factor; thus $\phi_t(z, w) = (z, g_{t,z}(w))$ for $(z, w) \in \widetilde{U}_F \times S^2$ with diffeomorphisms $g_{t,z} : S^2 \to S^2$ satisfying $g_{0,z} = \mathrm{id}$. Replacing $g_{t,z}$ by $\tilde{g}_{t,z} := g_{\rho(z)t,z}$ with a smooth cutoff function $\rho : S^2 \to [0, 1]$ which equals 0 on U_F and 1 outside $\widetilde{U}_F$, we obtain new diffeomorphisms $\tilde{\phi}_t$ which equal the identity on $U_F \times S^2$ for all t.

For (ii), suppose that $\mathfrak{S}_t^{\mathrm{top}} = \mathfrak{S}_{\mathrm{std}}^{\mathrm{top}}$ on $S^2 \times U_E$. Then we choose the sections $\widetilde{\Sigma}_t := S^2 \times \{w_0\}$ for some $w_0 \in E$ and the diffeomorphisms u_t to be the identity map $S^2 \times \{w_0\} \to S^2$, so that $p_t(z, w) = z$ for $(z, w) \in S^1 \times U_E$. For each $(t, z) \in [0, 1] \times S^2$ define

$$\phi_{t,z} := \phi_t|_{p_0^{-1}(z)} : p_0^{-1}(z) \xrightarrow{\cong} p_t^{-1}(z).$$

By the smooth isotopy extension theorem, the embeddings

$$\phi_{t,z}^{-1}|_{\{z\}\times U_E} : \{z\} \times U_E \hookrightarrow p_0^{-1}(z)$$

extend to a smooth family of diffeomorphisms $\psi_{t,z} : p_0^{-1}(z) \to p_0^{-1}(z)$ with $\psi_{0,z} = \mathrm{id}$. It follows that the diffeomorphisms $\tilde{\phi}_{t,z} := \phi_{t,z} \circ \psi_{t,z}$ restrict to the identity on $\{z\} \times U_E$, so they fit together to a family of diffeomorphisms $\tilde{\phi}_t : M \to M$ satisfying $\tilde{\phi}_0 = \mathrm{id}$ and $\tilde{\phi}_t = \mathrm{id}$ on $S^2 \times U_E$. After renaming $\tilde{\phi}_t$ back to ϕ_t we have thus shown the first assertion in (ii).

Finally, suppose in addition that $\mathfrak{S}_0^{\mathrm{top}} = \mathfrak{S}_1^{\mathrm{top}} = \mathfrak{S}_{\mathrm{std}}^{\mathrm{top}}$ near $S^2 \times (U_0 \amalg U_\infty)$. This implies that $p_0(z, w) = g_0(z)$ and $p_1(z, w) = h_0(z)$ for $(z, w) \in S^2 \times U_0$, with $g_0, h_0 \in \mathrm{Diff}_+(S^2)$. For each $z \in S^2$ the diffeomorphism $\phi_{1,z} : p_0^{-1}(z) \xrightarrow{\cong} p_1^{-1}(z)$ constructed above restricts to the identity on $\{z\} \times U_E$. Hence the restrictions

$$\phi_{1,z}|_{\{g_0^{-1}(z)\}\times U_0} : \{g_0^{-1}(z)\} \times U_0 \hookrightarrow p_1^{-1}(z) \setminus (\{z\} \times U_E)$$

define an S^2-family of embeddings of the closed disk U_0 into the component of $S^2 \setminus U_E$ containing the south pole (which is an open disk). Since we can deform the embeddings to linear ones by shrinking the disks, and $\pi_2 \, \mathrm{GL}_+(2, \mathbb{R}) = \{0\}$, this family is smoothly homotopic to the constant family of inclusions $U_0 \hookrightarrow S^2 \setminus U_E$. So we can modify the family of diffeomorphisms $\phi_t : M \to M$ near $t = 1$ to arrange $\phi_1(g_0^{-1}(z), w) = (h_0^{-1}(z), w)$ for $(z, w) \in S^2 \times U_0$; hence $\phi_1(z, w) = (f_0(z), w)$ for $(z, w) \in S^2 \times U_0$ with $f_0 := h_0^{-1} \circ g_0 \in \mathrm{Diff}_+(S^2)$. The disk U_∞ is treated analogously. □

Fixing the symplectic form. Our next result is an easy consequence of Moser's and Banyaga's theorems. Since it will be used repeatedly in this paper, let us recall the latter [Banyaga 1978, Théorème II.2.1] for future reference.

Theorem 3.4 (Banyaga's isotopy extension theorem [1978]). *Let (M, ω) be a symplectic manifold and $\psi_t : M \to M$ a smooth isotopy with $\psi_0 = \mathrm{id}$ such that each ψ_t is symplectic on a neighbourhood of a compact subset $X \subset M$. Suppose that $\int_\sigma \psi_t^*\omega$ is constant in t for each $\sigma \in H_2(M, X)$. Then there exists a symplectic isotopy ϕ_t with $\phi_0 = \mathrm{id}$ and $\phi_t|_X = \psi_t|_X$.*

Proposition 3.5 (fixing the symplectic form — parametric version). *Let* $\mathfrak{S}_t = (\mathcal{F}_t, \omega_t, L_t, \Sigma_t, \Sigma'_t)_{t\in[0,1]}$ *be a homotopy of relative symplectic fibrations of* $M = S^2 \times S^2$. *Then there exists an isotopy of diffeomorphisms* $\phi_t \in \mathrm{Diff}_{\mathrm{id}}(M)$ *with* $\phi_0 = \mathrm{id}$ *such that* $\phi_t^{-1}(\mathfrak{S}_t) = (\widetilde{\mathcal{F}}_t, \omega_0, L_0, \Sigma_0, \Sigma'_0)$ *for some family of foliations* $\widetilde{\mathcal{F}}_t$.

Proof. First, Moser's theorem provides an isotopy of diffeomorphisms $\phi_t : M \to M$ with $\phi_0 = \mathrm{id}$ such that $\phi_t^*\omega_t = \omega_0$. After replacing $\mathfrak{S}_t$ by $\phi^{-1}(\mathfrak{S}_t)$, we may hence assume that $\omega_t = \omega_0$ for all t.

Next, consider the isotopy of submanifolds $X_t := L_t \amalg \Sigma_t \amalg \Sigma'_t$ of (M, ω_0). Let us write (using the smooth isotopy extension theorem) $X_t = \psi_t(X_0)$ for diffeomorphisms $\psi_t : M \to M$ with $\psi_0 = \mathrm{id}$. Since L_t is Lagrangian and $\Sigma_t \amalg \Sigma'_t$ is symplectic, the Lagrangian and symplectic neighbourhood theorems provide a modification of ψ_t which is symplectic on a neighbourhood of X_0.

We claim that the symplectic area $\int_{\sigma_t} \omega_0$ is constant in t for each $\sigma \in H_2(M, X_0)$, where we define $\sigma_t := (\psi_t)_*\sigma \in H_2(M, X_t)$. To see this, note that the map $H_2(M, L_0) \to H_2(M, X_0)$ is surjective because $H_1(\Sigma_0 \amalg \Sigma'_0) = 0$. So it suffices to prove the claim for classes $\sigma \in H_2(M, L_0) \cong \pi_2(M, L_0)$. Now recall that the Lagrangian tori L_t are monotone with respect to ω_0. Since the Maslov class $\mu(\sigma_t)$ of L_t is constant in t, so is the symplectic area $\int_{\sigma_t} \omega_0$ by monotonicity and the claim is proved.

In view of the claim, (X_0, ψ_t) satisfies the hypotheses of Banyaga's theorem, Theorem 3.4. It follows that the smooth isotopy ψ_t can be altered to a symplectic isotopy ϕ_t with $\phi_0 = \mathrm{id}$ and $\phi_t(X_0) = X_t$. This is the desired isotopy in the proposition. ☐

A nonparametric version of Proposition 3.5 is much more subtle and will be discussed in Section 4H.

3B. ***Standardisation near a fibre.*** Let us pick the point $z_0 := (1, 0, 0)$ on the equator E in the base, so that $F := p_1^{-1}(z_0)$ is a fibre of $\mathcal{F}_{\mathrm{std}}$ intersecting L_{std} in the equator E. The following proposition shows that we can deform every relative symplectic fibration to make the triple $(\mathcal{F}, \omega, L)$ standard near F.

Proposition 3.6 (standardisation near a fibre). *Every relative symplectic fibration* $\mathfrak{S} = (\mathcal{F}_{\mathrm{std}}, \omega, L_{\mathrm{std}}, \Sigma, \Sigma')$ *is homotopic to a fibration of the form* $\widetilde{\mathfrak{S}} = (\mathcal{F}_{\mathrm{std}}, \tilde{\omega}, L_{\mathrm{std}}, \widetilde{\Sigma}, \widetilde{\Sigma}')$ *such that* $\tilde{\omega} = \omega_{\mathrm{std}}$ *on a neighbourhood of the fibre* F.

The proof is given in [Ivrii 2003, Lemma 3.2.3]. For convenience, we recall the argument. It is based on two easy lemmas.

Lemma 3.7. *Let* E *denote the equator and* D_{lh} *the lower hemisphere in* $S^2 \subset \mathbb{R}^3$. *Let* σ *be a symplectic form on* S^2 *cohomologous to* σ_{std} *such that* $\int_{D_{\mathrm{lh}}} \sigma = \frac{1}{2}$. *Then there exists an isotopy of diffeomorphisms* $h_t : S^2 \to S^2$ *with* $h_0 = \mathrm{id}$ *such that* $h_t(E) = E$ *for all* t *and* $h_1^*\sigma = \sigma_{\mathrm{std}}$.

Proof. We apply Moser's theorem to $\sigma_t := (1-t)\sigma_{\rm std} + t\sigma$ to find an isotopy of diffeomorphisms $f_t : S^2 \to S^2$ with $f_0 = \mathrm{id}$ and $f_t^*\sigma_t = \sigma_{\rm std}$. Since $\int_{D_{\rm lh}} \sigma_t = \frac{1}{2}$ for all t, the $\sigma_{\rm std}$-Lagrangians $f_t^{-1}(E)$ all bound disks of $\sigma_{\rm std}$-area $\frac{1}{2}$. Hence Banyaga's theorem, Theorem 3.4, yields $\sigma_{\rm std}$-symplectomorphisms $g_t : S^2 \to S^2$ with $g_0 = \mathrm{id}$ and $g_t(E) = f_t^{-1}(E)$, so $h_t := f_t \circ g_t$ is the desired isotopy. □

Lemma 3.8. *Let ω be a symplectic form on $M = S^2 \times S^2$ compatible with the standard fibration $p_1 : M \to S^2$. Let $\delta \subset S^2$ be an embedded closed arc passing through z_0. Then every symplectomorphism $h : (F, \sigma_{\rm std}) \to (F, \omega|_F)$ extends to a diffeomorphism ψ between neighbourhoods of $p_1^{-1}(\delta)$ preserving the fibres over δ and such that $\psi^*\omega = \omega_{\rm std}$.*

Proof. Parallel transport in $N := p_1^{-1}(\delta)$ with respect to $\omega_{\rm std}$ from $p_1^{-1}(z)$ to F and then with respect to ω from F to $p_1^{-1}(z)$ yields a fibre-preserving diffeomorphism $\phi : N \to N$ extending h with $\phi^*(\omega|_N) = (\omega_{\rm std})|_N$. By the coisotropic neighbourhood theorem, ϕ extends to the desired diffeomorphism ψ. □

Proof of Proposition 3.6. By Lemma 3.7, there exists an isotopy of diffeomorphisms $h_t : F \to F$ with $h_0 = \mathrm{id}$, $h_t(E) = E$ and $h_1^*(\omega|_F) = \sigma_{\rm std}$. Let $\delta \subset E$ be an arc in the equator in the base passing through z_0. By Lemma 3.8, the diffeomorphisms h_t extend to diffeomorphisms ψ_t between neighbourhoods of $p_1^{-1}(\delta)$ preserving the fibres over δ such that $\psi_0 = \mathrm{id}$ and $\psi_1^*\omega = \omega_{\rm std}$. Thus the pullback fibrations $\psi_t^*\mathcal{F}_{\rm std}$ and the pullback tori $\psi_t^{-1}(L_{\rm std})$ coincide over δ with $\mathcal{F}_{\rm std}$ and $L_{\rm std}$, respectively (the latter holds because $\psi_t^{-1}(L_{\rm std})$ is obtained by parallel transport of $\{z_0\} \times E$ along δ). Therefore, we can restrict the ψ_t to a smaller neighbourhood $V = p_1^{-1}(V')$ of F and extend them from there to diffeomorphisms $\chi_t : M \to M$ preserving $L_{\rm std}$ (which equal the identity outside a larger neighbourhood of F) such that $\chi_0 = \mathrm{id}$. Then the pullbacks $\chi_t^{-1}(\mathfrak{S})$ satisfy $\chi_1^*\omega = \omega_{\rm std}$ on V and $\chi_t^{-1}(\mathcal{F}_{\rm std}) = \mathcal{F}_{\rm std}$ on $p_1^{-1}(\delta) \cap V$.

So far we have just put $\mathfrak{S}$ into a more convenient form by diffeomorphisms, but now we will modify it. Note that the fibres of $\mathcal{F}_{\rm std}$ near $p_1^{-1}(\delta) \cap V$ are C^1-close to those of $\chi_t^{-1}(\mathcal{F}_{\rm std})$ and therefore symplectic for $\chi_t^*\omega$. Hence we can deform the foliations $\chi_t^{-1}(\mathcal{F}_{\rm std})$ to foliations $\mathcal{F}_t$ with $\mathcal{F}_0 = \mathcal{F}_{\rm std}$, keeping them $\chi_t^*\omega$-symplectic and fixed on $p_1^{-1}(\delta) \cap V$ and outside V, such that $\mathcal{F}_t = \mathcal{F}_{\rm std}$ on a neighbourhood $U \subset V$ of F. This yields a homotopy of relative symplectic fibrations $\mathfrak{S}_t := (\mathcal{F}_t, \chi_t^*\omega, L_{\rm std}, \chi_t^{-1}(\Sigma), \chi_t^{-1}(\Sigma'))$ with $\mathfrak{S}_0 = \mathfrak{S}$ such that $\mathcal{F}_t = \mathcal{F}_{\rm std}$ on U for all t and $\chi_1^*\omega = \omega_{\rm std}$ on U. Finally, we apply Proposition 3.3 to this homotopy (ignoring the symplectic sections) to find an isotopy of diffeomorphisms $\phi_t : M \to M$ with $\phi_0 = \mathrm{id}$ such that $\phi_t^{-1}(\mathcal{F}_t) = \mathcal{F}_{\rm std}$ and $\phi_t^{-1}(L_{\rm std}) = L_{\rm std}$. Moreover, by Proposition 3.3(i) we can arrange that $\phi_t = \mathrm{id}$ near F. Then $\phi_1^*\chi_1^*\omega = \omega_{\rm std}$ near F, so the end point $\phi_1^{-1}(\mathfrak{S}_1)$ of the homotopy $\phi_t^{-1}(\mathfrak{S}_t)$ is the desired relative symplectic foliation $\widetilde{\mathfrak{S}}$. □

Remark 3.9. Even if in Proposition 3.6 the sections Σ, Σ' in $\mathfrak{S}$ are the standard sections S_0, S_∞, this will not be true for the sections in $\widetilde{\mathfrak{S}}$ unless the original sections were horizontal near F. This will be remedied in the following subsection.

3C. ***Standardisation near the sections.*** Consider a relative symplectic fibration of the form $\mathfrak{S} = (\mathcal{F}_{\mathrm{std}}, \omega, L_{\mathrm{std}}, S_0, S_\infty)$ with the projections $p_1, p_2 : S^2 \times S^2 \to S^2$ onto the two factors.

Definition 3.10. We say that ω is *split* on a neighbourhood

$$W = (U_F \times S^2) \cup (S^2 \times U_0) \cup (S^2 \times U_\infty)$$

of $F \cup S_0 \cup S_\infty$ if there exist symplectic forms σ_0, σ_∞ on S^2 such that

$$\begin{aligned} &\omega = p_1^*\sigma_0 + p_2^*\sigma_{\mathrm{std}} && \text{on the set } W_0 = (U_F \times S^2) \cup (S^2 \times U_0), \\ &\omega = p_1^*\sigma_\infty + p_2^*\sigma_{\mathrm{std}} && \text{on the set } W_\infty = (U_F \times S^2) \cup (S^2 \times U_\infty). \end{aligned} \tag{1}$$

Here the forms σ_0 and σ_∞ may differ, but they agree on U_F. Note that if ω is split, then in particular the sections S_∞ and S_0 are horizontal. Moreover, parallel transport of the symplectic connection defined by ω equals the identity on the region where ω is split.

The following is the main result of this section.

Proposition 3.11 (standardisation near a fibre and the sections). *Every relative symplectic fibration $\mathfrak{S} = (\mathcal{F}_{\mathrm{std}}, \omega, L_{\mathrm{std}}, S_0, S_\infty)$ is homotopic to one of the form $\widetilde{\mathfrak{S}} = (\mathcal{F}_{\mathrm{std}}, \tilde{\omega}, L_{\mathrm{std}}, S_0, S_\infty)$ such that $\tilde{\omega}$ is split on a neighbourhood W of $F \cup S_0 \cup S_\infty$.*

The proof of this proposition will occupy the remainder of this section. Standardisation near a symplectic section is more subtle than near a fibre because the section need not be horizontal, so it takes a large deformation to make it symplectically orthogonal to the fibres.

We first consider the local situation in $\mathbb{R}^4 \cong \mathbb{C}^2$ with the standard symplectic form $\Omega_0 = dx \wedge dy + du \wedge dv$ in coordinates $z = x + iy$, $w = u + iv$. Let $S = \{w = f(z)\}$ be the graph over the z-plane of a smooth function $f = g + ih : \mathbb{R}^2 \to \mathbb{R}^2$ with $f(0) = 0$. Orient S by projection onto the z-plane. The pullback of Ω_0 under the embedding $F(z) = (z, f(z))$ equals

$$F^*\Omega_0 = dx \wedge dy + dg \wedge dh = (1 + \det Df)\, dx \wedge dy.$$

Thus S is symplectic (with the given orientation) if and only if

$$1 + \det Df > 0.$$

For a smooth function $\phi : [0, \infty) \to \mathbb{R}$ consider the new function

$$\tilde{f}(z) := \phi(|z|) f(z).$$

We now derive the condition on ϕ such that the graph of $\tilde{f}$ is symplectic. We will see that it suffices to do this for linear maps f, so suppose that $f(z) = Az$ for a 2×2 matrix A. We compute for $r := |z| > 0$,

$$\begin{aligned} D\tilde{f}(z) &= \phi(r)Df(z) + \phi'(r)f(z)\left(\frac{z}{r}\right)^t = \phi(r)A + \frac{\phi'(r)}{r}Azz^t \\ &= A\left(\phi(r)\mathbb{1} + \frac{\phi'(r)}{r}zz^t\right). \end{aligned}$$

Since

$$\begin{aligned} \det\left(\phi(r)\mathbb{1} + \frac{\phi'(r)}{r}zz^t\right) &= \det\begin{pmatrix} \phi+(\phi'/r)x^2 & (\phi'/r)xy \\ (\phi'/r)xy & \phi+(\phi'/r)y^2 \end{pmatrix} \\ &= \phi^2 + \frac{\phi\phi'}{r}(x^2+y^2) = \phi^2 + r\phi\phi', \end{aligned}$$

we have $\det D\tilde{f} = (\phi^2 + r\phi\phi')\det A$. This proves:

Lemma 3.12. *Let $f(z) = Az$ be a linear function $\mathbb{R}^2 \to \mathbb{R}^2$ with $1 + \det A \geq \varepsilon > 0$. Let $\phi : [0, \infty) \to \mathbb{R}$ be a smooth function with $\phi(0) = \phi'(0) = 0$. Then the graph of $\tilde{f}(z) := \phi(|z|)f(z)$ is symplectic provided that for all $r > 0$,*

$$0 \leq \phi(r)^2 + r\phi(r)\phi'(r) < \frac{1}{1-\varepsilon}. \tag{2}$$

Lemma 3.13. *For every $0 < \varepsilon < 1$ and $\delta > 0$ there exists a smooth family of nondecreasing functions $\phi_s : [0, \infty) \to [0, 1]$, $s \in [0, 1]$, satisfying (2) such that $\phi_s(r) = s$ for $r \leq \delta$ and $\phi_s(r) = 1$ for $r \geq 2\delta/\sqrt{\varepsilon}$.*

Proof. For $r > 0$ define $\psi(r)$ by $\psi(r) := r^2\phi(r)^2$. Then $\psi' = 2r(\phi^2 + r\phi\phi')$, so (2) is equivalent to

$$\psi'(r) < \frac{2r}{1-\varepsilon}.$$

This will be satisfied if ψ solves the differential equation

$$\psi'(r) = \frac{2r}{1-\varepsilon/4}.$$

Then $\psi(r) = r^2/(1-\varepsilon/4) + c$ for some constant c and

$$\phi^2(r) = \frac{1}{1-\varepsilon/4} + \frac{c}{r^2}.$$

We fix the constant c by $\phi(\delta) = 0$ to $c = -\delta^2/(1-\varepsilon/4)$ and obtain

$$\phi^2(r) = \frac{1-\delta^2/r^2}{1-\varepsilon/4}.$$

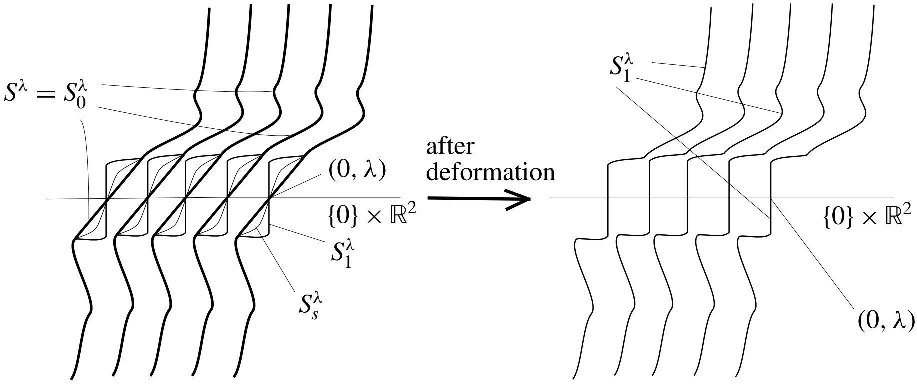

Figure 4. The family of foliations S_s^λ.

This is an increasing function with $\phi(\delta) = 0$ and $\phi(\gamma) = 1$ at the point $\gamma = 2\delta/\sqrt{\varepsilon}$. Now observe that if a solution of (2) satisfies $\phi(r_0) \geq 0$ and $\phi'(r_0) \geq 0$ for some $r_0 > 0$, then we can decrease the slope to 0 near r_0 and extend ϕ by $\phi(r) = \phi(r_0)$ for $r \geq r_0$ (or $r \leq r_0$) to a smooth solution of (2). Applying this procedure at $r_0 = \delta$ and $r_0 = \gamma$ yields the desired function ϕ_0 for $s = 0$. For $s > 0$, we obtain ϕ_s by smoothing the function $\max(s, \phi_0)$. □

Lemma 3.14. *Let $\Lambda \subset \mathbb{R}^2$ be compact and $(S^\lambda)_{\lambda\in\Lambda}$ be a smooth foliation of a region in $(\mathbb{R}^4, \Omega_0)$ by symplectic surfaces S^λ intersecting the symplectic plane $\{0\} \times \mathbb{R}^2$ transversely in $(0, \lambda)$. Then for every neighbourhood $W \subset \mathbb{R}^4$ of $\{0\} \times \Lambda$ there exists a neighbourhood $U \subset W$ of $\{0\} \times \Lambda$ and a family of foliations $(S_s^\lambda)_{s\in[0,1],\lambda\in\Lambda}$ with the following properties (see Figure 4):*

- (i) *$S_0^\lambda = S^\lambda$ and $S_s^\lambda = S^\lambda$ outside W.*
- (ii) *S_s^λ is symplectic and intersects $\{0\} \times \mathbb{R}^2$ transversely in $(0, \lambda)$.*
- (iii) *$S_1^\lambda = \mathbb{R}^2 \times \{\lambda\}$ in U.*

Moreover, for every λ with $S^\lambda = \mathbb{R}^2 \times \{\lambda\}$ in W we have $S_s^\lambda = S^\lambda$ for all s.

Proof. After shrinking W, we may assume that in W each surface can be written as a graph $S^\lambda = \{w = \lambda + f^\lambda(z)\}$ over the z-plane with $f^\lambda(0) = 0$. After a C^1-small perturbation of the surfaces in W (which keeps them symplectic) we may assume that the f^λ are linear functions $f^\lambda(z) = A^\lambda z$. Symplecticity implies $\det A^\lambda > -1$. Since Λ is compact, there exists an $\varepsilon > 0$ with $\det A^\lambda \geq -1 + \varepsilon$ in W for all λ. Moreover, we may assume that the ε-neighbourhood of Λ is contained in W. Pick $\delta > 0$ so small that $2\delta/\sqrt{\varepsilon} < \varepsilon$. Let $\phi_s : [0, \infty) \to [0, 1]$, $s \in [0, 1]$, be the functions of Lemma 3.13 and define $f_s^\lambda(z) := \phi_{1-s}(|z|) f^\lambda(z)$. By Lemma 3.12, the graph S_s^λ of f_s^λ satisfies conditions (i)–(iii) of the proposition, where U is the δ-neighbourhood of Λ. Note that if $S^\lambda = \mathbb{R}^2 \times \{\lambda\}$ for some λ, then $f^\lambda(z) \equiv 0$ and thus $S_s^\lambda = S^\lambda$ for all s.

It only remains to verify that the surfaces $(S_s^\lambda)_{\lambda\in\Lambda}$ form a foliation for each s, or equivalently, that the map $F_s : B^2(\varepsilon)\times\Lambda\to\mathbb{R}^4$,

$$F_s(z,\lambda) := (z, \lambda + f_s^\lambda(z)) = (z, \lambda + \phi_{1-s}(|z|)A^\lambda z),$$

is an embedding. For injectivity, suppose that $F_s(z,\lambda) = F_s(z',\lambda')$. Then $z = z'$ and $\lambda - \lambda' = -\phi_{1-s}(|z|)(A^\lambda - A^{\lambda'})z$. This implies

$$|\lambda-\lambda'| \le \|A^\lambda - A^{\lambda'}\|\,|z| \le \varepsilon\|A^\lambda - A^{\lambda'}\|.$$

Since A^λ depends smoothly on λ, there exists a constant C such that $\|A^\lambda - A^{\lambda'}\| \le C|\lambda-\lambda'|$. For $\varepsilon < 1/C$ it follows that $\lambda = \lambda'$. For the immersion property, consider the differential

$$DF_s(z,\lambda) = \begin{pmatrix} \mathbb{1} & 0 \\ D_z f_s^\lambda & \mathbb{1}+B_s \end{pmatrix}, \quad B_s = \frac{\partial f_s^\lambda}{\partial\lambda}.$$

This is invertible if and only if the matrix

$$\mathbb{1} + B_s = \mathbb{1} + \phi_{1-s}(|z|)\frac{\partial A^\lambda}{\partial\lambda}z$$

is invertible. By smoothness in λ, there exists a constant C with

$$\left\|\frac{\partial A^\lambda}{\partial\lambda}z\right\| \le C|z|.$$

Then for $\varepsilon < 1/C$ we get

$$\left\|\phi_{1-s}(|z|)\frac{\partial A^\lambda}{\partial\lambda}z\right\| \le C|z| \le C\varepsilon < 1,$$

which implies invertibility of $\mathbb{1}+B_s$. □

Proof of Proposition 3.11. We deform the given relative symplectic fibration $\mathfrak{S} = (\mathcal{F}_{\mathrm{std}}, \omega, L_{\mathrm{std}}, S_0, S_\infty)$ in four steps.

Step 1. By Proposition 3.6, $\mathfrak{S}$ is homotopic to $\widetilde{\mathfrak{S}} = (\mathcal{F}_{\mathrm{std}}, \tilde\omega, L_{\mathrm{std}}, \widetilde\Sigma, \widetilde\Sigma')$ such that $\tilde\omega = \omega_{\mathrm{std}}$ on a neighbourhood of the fibre $F = \{z_0\}\times S^2$. The sections $\widetilde\Sigma$, $\widetilde\Sigma'$ intersect the fibre in points (z_0, q) and (z_0, q'). After pulling back $\widetilde{\mathfrak{S}}$ by a symplectomorphism $(z,w)\mapsto(z,g(w))$, where $g: S^2\to S^2$ is a Hamiltonian diffeomorphism preserving the equator and mapping the south pole S to q and the north pole N to q', we may assume in addition that $\widetilde\Sigma\cap F = (z_0, S)$ and $\widetilde\Sigma'\cap F = (z_0, N)$. By Lemma 3.14 (with $\Lambda = \{0\}$, $S^0 = \widetilde\Sigma$ and $F = \{0\}\times\mathbb{R}^2$ in local coordinates), we can deform $\widetilde\Sigma$ such that it agrees with $S_0 = S^2\times\{S\}$ near $\widetilde\Sigma\cap F$. Since the section $\widetilde\Sigma$ is isotopic to S_0, there exists an $\mathcal{F}_{\mathrm{std}}$-preserving diffeomorphism isotopic to the identity and fixed near L_{std} and F mapping S_0 to $\widetilde\Sigma$. By pulling back everything by this diffeomorphism we arrange that the new symplectic section is $\widetilde\Sigma = S_0$. By the same arguments we arrange $\widetilde\Sigma' = S_\infty$. Note that the foliation at the end of this step is still $\mathcal{F}_{\mathrm{std}}$.

Step 2. As in the proof of Proposition 3.6, using the symplectic neighbourhood theorem, by pulling back by an isotopy of $S^2 \times S^2$ fixed near F we can arrange in addition that the new symplectic form $\hat{\omega}$ satisfies $\hat{\omega} = \omega_{\mathrm{std}}$ near $\widetilde{\Sigma} = S_0$ (but the foliation becomes nonstandard). The same arguments apply to the other section $\widetilde{\Sigma}'$. Thus $\widetilde{\mathfrak{S}}$ is homotopic to a relative symplectic fibration of the form $\widehat{\mathfrak{S}} = (\widehat{\mathcal{F}}, \hat{\omega}, L_{\mathrm{std}}, S_0, S_\infty)$ with the following properties: $\hat{\omega} = \omega_{\mathrm{std}}$ and $\widehat{\mathcal{F}} = \mathcal{F}_{\mathrm{std}}$ near the fibre $F = \{z_0\} \times S^2$, and $\hat{\omega} = \omega_{\mathrm{std}}$ near the symplectic sections S_0 and S_∞.

Step 3. Next, we adjust the foliation $\widehat{\mathcal{F}}$ near $S_0 \cup S_\infty$. Consider first S_0. Take a compact subset $\Lambda \subset S^2 \setminus \{z_0\}$ such that $\widehat{\mathcal{F}} = \mathcal{F}_{\mathrm{std}}$ on a neighbourhood of $(S^2 \setminus \operatorname{int} \Lambda) \times S^2$. We identify Λ with a subset of $(\mathbb{R}^2, dx \wedge dy)$, and a neighbourhood of $\Lambda \times \{S\}$ in $S^2 \times S^2$ with a neighbourhood W of $\{0\} \times \Lambda$ in $(\mathbb{R}^4, \Omega_0)$, by a symplectomorphism of the form $(z, w) \mapsto (f(w), g(z))$. Under this identification, $\widehat{\mathcal{F}}$ corresponds to a symplectic foliation of W transverse to $\{0\} \times \Lambda$ and standard near $\partial \Lambda \times \mathbb{R}^2$. By Lemma 3.14, $\widehat{\mathcal{F}}$ can be deformed in W, keeping it fixed near $\partial \Lambda \times \mathbb{R}^2$, to a symplectic foliation that is standard on a neighbourhood U of $\{0\} \times \Lambda$ in $\mathbb{R}^4$. Transferring back to $S^2 \times S^2$ and performing the same construction near S_∞, we have thus deformed $\widehat{\mathfrak{S}}$ to a relative symplectic fibration $\overline{\mathfrak{S}} = (\overline{\mathcal{F}}, \bar{\omega}, L_{\mathrm{std}}, S_0, S_\infty)$ satisfying $\bar{\omega} = \omega_{\mathrm{std}}$ and $\overline{\mathcal{F}} = \mathcal{F}_{\mathrm{std}}$ near the set $F \cup S_0 \cup S_\infty$. This was the main step. It only remains to deform $\overline{\mathcal{F}}$ back to $\mathcal{F}_{\mathrm{std}}$.

Step 4. Let $\mathfrak{S}_t$ be the homotopy of relative symplectic fibrations from $\mathfrak{S}_0 = \widetilde{\mathfrak{S}}$ to $\mathfrak{S}_1 = \overline{\mathfrak{S}}$ constructed in Steps 2 and 3. By construction, it satisfies the additional properties (i) and (ii) in Proposition 3.3 for disks U_F, U_0, U_∞ containing z_0, S, N and an annulus $U_E \subset S^2$ containing the equator E. Hence there exists an isotopy of diffeomorphisms $\phi_t \in \mathrm{Diff}_{\mathrm{id}}(M)$ with $\phi_0 = \mathrm{id}$ such that $\phi_t^{-1}(\mathfrak{S}_t) = (\mathcal{F}_{\mathrm{std}}, \phi_t^* \omega_t, L_{\mathrm{std}}, \widetilde{\Sigma}, \widetilde{\Sigma}')$. Moreover, $\phi_t = \mathrm{id}$ on $U_F \times S^2$ for all t, $\phi_1(z, w) = (f_0(z), w)$ for $(z, w) \in S^2 \times U_0$, and $\phi_1(z, w) = (f_\infty(z), w)$ for $(z, w) \in S^2 \times U_\infty$, with $f_0, f_\infty \in \mathrm{Diff}_+(S^2)$. It follows that

$$\begin{aligned} \phi_1^* \omega_1 &= p_1^* f_0^* \sigma_{\mathrm{std}} + p_2^* \sigma_{\mathrm{std}} \quad \text{on } W_0 = (U_F \times S^2) \cup (S^2 \times U_0), \\ \phi_1^* \omega_1 &= p_1^* f_\infty^* \sigma_{\mathrm{std}} + p_2^* \sigma_{\mathrm{std}} \quad \text{on } W_\infty = (U_F \times S^2) \cup (S^2 \times U_\infty), \end{aligned}$$

so $\phi_1^* \omega_1$ is split on $W = W_0 \cup W_\infty$ and $\phi_t^{-1}(\mathfrak{S}_t)$ is the desired homotopy. This concludes the proof of Proposition 3.11. □

Remark 3.15. Replacing Step 4 of the preceding proof by a more careful deformation of the foliation $\overline{\mathcal{F}}$ (not by diffeomorphisms but keeping it symplectic for $\bar{\omega}$), we could arrange $\tilde{\omega} = \omega_{\mathrm{std}}$ near $F \cup S_0 \cup S_\infty$ in Proposition 3.11. As the class of split forms is better suited for the modifications in the next section, we content ourselves with making $\tilde{\omega}$ split near $F \cup S_0 \cup S_\infty$.

4. Killing the holonomy

In this section we will deform a relative symplectic fibration to kill all the holonomy and conclude the proof of the main theorem. A crucial ingredient is the inflation procedure from [Lalonde and McDuff 1996].

4A. ***Setup.*** Recall that $\mathcal{F}_{\mathrm{std}}$ is the foliation on $S^2 \times S^2$ given by the fibres of the projection p_1 onto the first factor, $S_0 = S^2 \times \{S\}$ and $S_\infty = S^2 \times \{N\}$ are the standard sections, $F = p_1^{-1}(z_0)$ is the fibre over the point $z_0 = (1, 0, 0)$, and the Clifford torus $L_{\mathrm{std}} = E \times E$ is the product of the equators. In the following, we identify S_0 with the base S^2 of the projection p_1; i.e., we identify p_1 with the map $(z, w) \mapsto (z, S)$ sending each fibre to its intersection with S_0.

Our starting point is a relative symplectic fibration $\mathfrak{S} = (\mathcal{F}_{\mathrm{std}}, \omega, L_{\mathrm{std}}, S_0, S_\infty)$ as provided by Proposition 3.11 such that ω is split on a neighbourhood $W = (U_F \times S^2) \cup (S^2 \times U_0) \cup (S^2 \times U_\infty)$ of $F \cup S_0 \cup S_\infty$. In particular the sections S_0, S_∞ are horizontal for the symplectic connection. After pulling back $\mathfrak{S}$ by a diffeomorphism of the form $(z, w) \mapsto (\phi(z), w)$ (keeping the same notation), we may replace U_F by the ball

$$B := \Big\{(x, y, z) \in S^2 \;\Big|\; x \geq -\tfrac{1}{\sqrt{2}}\Big\},$$

so that ω now satisfies

$$\begin{aligned} \omega &= p_1^*\sigma_0 + p_2^*\sigma_{\mathrm{std}} && \text{on the set } W_0 = (B \times S^2) \cup (S^2 \times U_0), \\ \omega &= p_1^*\sigma_\infty + p_2^*\sigma_{\mathrm{std}} && \text{on the set } W_\infty = (B \times S^2) \cup (S^2 \times U_\infty). \end{aligned} \tag{3}$$

Consider the usual spherical coordinates $(\lambda, \mu) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [0, 2\pi]$ on the base S^2 centred at z_0. Thus λ denotes the latitude and μ the meridian, and z_0 lies at $(\lambda, \mu) = (0, 0)$; see Figure 5.

Denote by C^λ the circle of latitude λ in the base and by ϕ^λ the symplectic parallel transport around C^λ parametrised by $\mu \in [0, 2\pi]$. Since the starting and ending points of the parametrisation of C^λ are contained in B for all λ and the symplectic

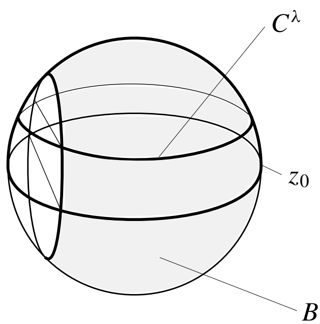

Figure 5. Circles of latitude and the set B.

form ω equals $p_1^*\sigma_0 + p_2^*\sigma_{\mathrm{std}}$ over B, we can regard ϕ^λ as living in $\mathrm{Symp}(S^2, \sigma_{\mathrm{std}})$ for all λ. Moreover, the maps ϕ^λ have the following two properties:

(i) Since $C^\lambda \subset B$ for all $|\lambda| \geq \frac{\pi}{4}$ and the form ω is split on $B \times S^2$, we have $\phi^\lambda = \mathrm{id}$ for $|\lambda| \geq \frac{\pi}{4}$.

(ii) Since ω is split on $S^2 \times (U_0 \cup U_\infty)$, each ϕ^λ restricts to the identity on $U_0 \cup U_\infty$.

Under stereographic projection $S^2 \setminus \{N\} \to \mathbb{C}$ from the north pole N, the standard symplectic form on S^2 corresponds to the form

$$\sigma_{\mathrm{std}} = \frac{r}{\pi(1+r^2)^2}\, dr \wedge d\theta$$

in polar coordinates on $\mathbb{C}$. We pick a closed annulus

$$A = \{z \in \mathbb{C} \mid a \leq |z| \leq b\} \subset \mathbb{C} \cong S^2 \setminus \{N\}$$

with $a > 0$ so small and $b > a$ so large that $\partial A \subset U_0 \cup U_\infty$. According to properties (i) and (ii) above, parallel transport along C^λ then defines maps $\phi^\lambda \in \mathrm{Symp}(A, \partial A, \sigma_{\mathrm{std}})$ (i.e., symplectomorphisms that equal the identity near ∂A; see the Appendix) that equal the identity for $|\lambda| \geq \frac{\pi}{4}$. In particular, $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \ni \lambda \mapsto \phi^\lambda$ defines a loop in the identity component $\mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$. Consider the loop of inverses

$$\psi^\lambda = (\phi^\lambda)^{-1}.$$

Since $L_{\mathrm{std}} = E \times E$ is invariant under parallel transport, the map ϕ^0, and thus ψ^0, preserves the equator E.

4B. ***A special contraction.*** According to Proposition A.4, the loop ψ^λ is contractible in $\mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$. However, in order for the inflation procedure below to work, we need a special contraction ψ_s^λ with the property that $\psi_s^0(E) = E$ for all $s \in [0, 1]$. Here we identify the equator E in S^2 via stereographic projection with the circle $E = \{|z| = 1\} \subset A$.

Proposition 4.1. *There exists a smooth contraction $\psi_s^\lambda \in \mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ of the loop ψ^λ, with $(s, \lambda) \in [0, 1] \times \left[\frac{-\pi}{2}, \frac{\pi}{2}\right]$, such that*

(i) $\psi_0^\lambda = \mathrm{id}$ *and* $\psi_1^\lambda = \psi^\lambda$ *for all* λ,

(ii) $\psi_s^\lambda = \mathrm{id}$ *for* $|\lambda| \geq \frac{\pi}{4}$ *and all* s,

(iii) ψ_s^λ *is constant in* s *near* $s = 0$ *and* $s = 1$,

(iv) $\psi_s^0(E) = E$ *for all* s.

Proof. Since the holonomy ψ^0 along the equator in the base preserves the equator E in the fibre, Lemma A.5 provides a path $\alpha(t) \in \mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ from the identity to ψ^0 which preserves E for all t. We split the loop ψ^λ into two paths $\delta_1 := \{\psi^\lambda\}_{\lambda\in[-\pi/2,0]}$ and $\delta_2 := \{\psi^\lambda\}_{\lambda\in[0,\pi/2]}$. Using these, we define two loops

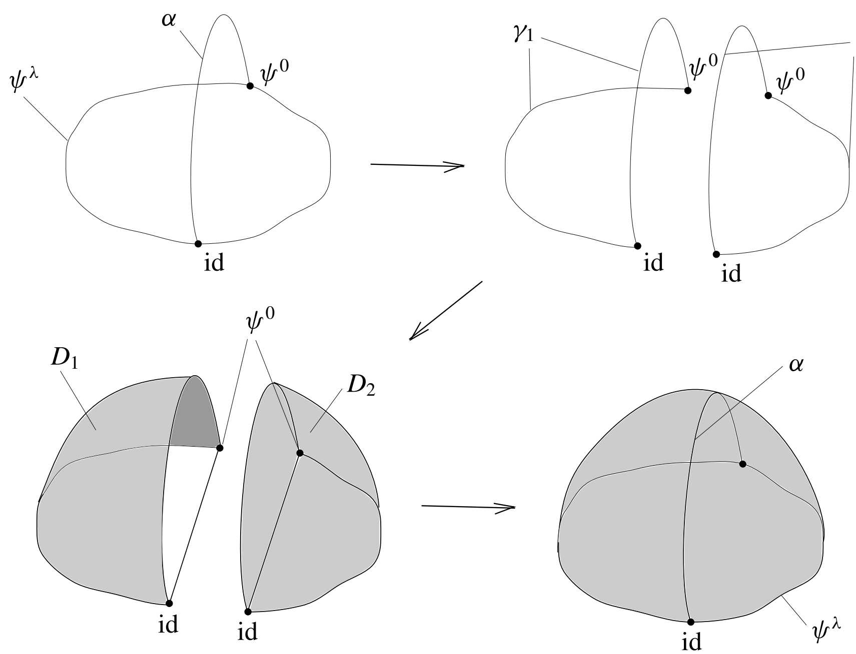

Figure 6. Construction of the special contraction ψ_s^λ.

$\gamma_1 := \delta_1 * \bar{\alpha}$ and $\gamma_2 := \alpha * \delta_2$, where $*$ means concatenation of paths and $\bar{\alpha}$ denotes the path α traversed in the opposite direction; see Figure 6. By Proposition A.4, these loops are contractible in $\mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$, so we can fill them by half-disks D_1, D_2 in $\mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$. Gluing these half-disks along α yields a map $\vartheta : D \to \mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ from the unit disk $D \subset \mathbb{C}$ which restricts to the loop ψ^λ on ∂D (starting and ending at $-i$) and to the path α on the imaginary axis. The composition of ϑ with the map

$$\eta : [0,1] \times \left[-\tfrac{\pi}{2}, \tfrac{\pi}{2}\right] \to D, \quad (s, \lambda) \mapsto (s-1)i + s e^{i(2\lambda + \pi/2)},$$

(see Figure 7) then has properties (i) and (iv) of the proposition. By smoothing and reparametrisation we finally arrange properties (ii) and (iii) to obtain the desired contraction ψ_s^λ. □

4C. ***A special Hamiltonian function.*** Now, we construct a family of time-dependent Hamiltonians generating the contraction ψ_s^λ of the previous subsection. We begin with a simple lemma.

Lemma 4.2. *Let $(M, \omega = d\lambda)$ be an exact symplectic manifold. Let $\phi_t : M \to M$ be a symplectic isotopy starting at $\phi_0 = \mathrm{id}$ generated by the time-dependent vector field X_t, i.e., $(d/dt)\phi_t = X_t \circ \phi_t$. Then $\iota_{X_t}\omega = dH_t$ for a smooth family of functions $H_t : M \to \mathbb{R}$ if and only if $\phi_t^*\lambda - \lambda = dF_t$ for a smooth family of functions $F_t : M \to \mathbb{R}$.*

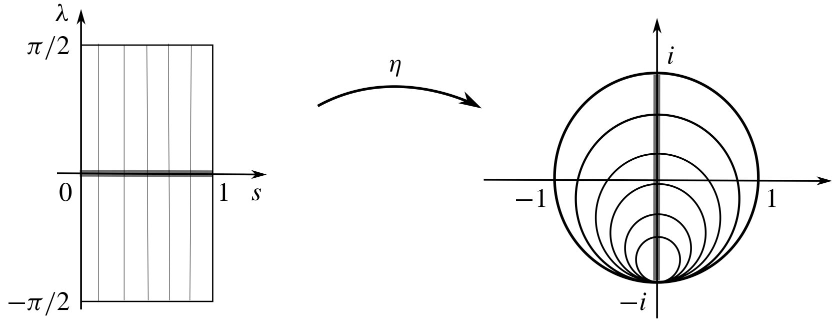

Figure 7. The reparametrisation η

Moreover, F_t *and* H_t *are related by the equations*

$$F_t = \int_0^t (H_s + \iota_{X_s}\lambda) \circ \phi_s \, ds, \qquad H_t = \dot{F}_t \circ \phi_t^{-1} - \iota_{X_t}\lambda.$$

Proof. Assume first that $\iota_{X_t}\omega = dH_t$. Then using Cartan's formula we find

$$\begin{aligned} \phi_t^*\lambda - \lambda = \int_0^t \frac{d}{ds}(\phi_s^*\lambda) \, ds &= \int_0^t \phi_s^*(L_{X_s}\lambda) \, ds \\ &= \int_0^t \phi_s^*(\iota_{X_s} d\lambda + d\iota_{X_s}\lambda) \, ds = d\int_0^t (H_s + \iota_{X_s}\lambda) \circ \phi_s \, ds, \end{aligned}$$

so $\phi_t^*\lambda - \lambda = dF_t$ holds for $F_t := \int_0^t (H_s + \iota_{X_s}\lambda) \circ \phi_s \, ds$. Conversely, if $\phi_t^*\lambda - \lambda = dF_t$, then we differentiate this equation to obtain

$$d\dot{F}_t = \frac{d}{dt}(\phi_t^*\lambda) = \phi_t^*(d\iota_{X_t}\lambda + \iota_{X_t} d\lambda),$$

which shows that $i_{X_t} d\lambda = dH_t$ holds for $H_t := \dot{F}_t \circ \phi_t^{-1} - \iota_{X_s}\lambda$. $\square$

Now let $\psi_s^\lambda \in \mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ be the special contraction from Proposition 4.1. Let

$$\lambda_{\mathrm{std}} = \frac{-1}{2(1+r^2)\pi} \, d\theta$$

be the standard primitive of σ_{std} (any other primitive would also do). Then for each (s, λ) the 1-form $\alpha_s^\lambda := (\psi_s^\lambda)^*\lambda_{\mathrm{std}} - \lambda_{\mathrm{std}}$ on A is closed and vanishes near ∂A. So by the relative Poincaré lemma,

$$(\psi_s^\lambda)^*\lambda_{\mathrm{std}} - \lambda_{\mathrm{std}} = dF_s^\lambda$$

for a unique smooth family of functions F_s^λ that vanish near the lower boundary component $\partial_- A = \{a\} \times S^1$ of A. (We can define $F_s^\lambda(w) := \int_{\gamma_w} \alpha_s^\lambda$ along any path γ_w from a base point on $\partial_- A$ to w, which does not depend on the path because every loop can be deformed into $\partial_- A$ where α_s^λ vanishes.) Note that F_s^λ will be constant near the upper boundary component $\partial_+ A = \{b\} \times S^1$, where the constant may depend on s and λ.

By Lemma 4.2, the family F_s^λ is related to a smooth family of Hamiltonians $\widetilde{H}_s^\lambda$ generating the isotopy ψ_s^λ (for fixed λ) by the formula

$$\widetilde{H}_s^\lambda = \frac{\partial F_s^\lambda}{\partial s} \circ (\psi_s^\lambda)^{-1} - \iota_{X_s^\lambda} \lambda_{\mathrm{std}},$$

where $(d/dt)\psi_t^\lambda = X_t^\lambda \circ \psi_t^\lambda$. By construction, $\widetilde{H}_s^\lambda$ vanishes near the lower boundary component $\partial_- A$ of A and it is constant near the upper boundary component $\partial_+ A$ (where the constant may vary with s and λ). Further, since ψ_s^λ is constant near its ends in both s and λ, we have $\widetilde{H}_s^\lambda = 0$ for $|\lambda| \geq \frac{\pi}{4}$ and for $s < 2\varepsilon$, $s > 1 - 2\varepsilon$ with some $\varepsilon > 0$.

Note that, since ψ_s^0 preserves the equator, the Hamiltonian vector field X_s^0 is tangent to E for all s. So the restriction $\widetilde{H}_s^0|_E$ is constant for all s and defines a function $\widetilde{H}_E(s)$. For reasons that will become clear in the next subsection, we wish to modify $\widetilde{H}$ to make this function vanish. For this, we pick a smooth cutoff function $\rho : \mathbb{R} \to [0, 1]$ with $\rho(0) = 1$ and support in $\left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ and define

$$H_s^\lambda := \widetilde{H}_s^\lambda - \rho(\lambda)\widetilde{H}_E(s).$$

Since H_s^λ differs from $\widetilde{H}_s^\lambda$ only by a function of s and λ, it still has the same Hamiltonian vector field and thus still generates the family ψ_s^λ. By construction, H_s^λ depends only on s and λ near the boundary ∂A (with possibly different functions at the two boundary components), $H_s^\lambda = 0$ for $|\lambda| \geq \frac{\pi}{4}$ and for $s < 2\varepsilon$, $s > 1 - 2\varepsilon$, and

$$H_s^0|_E = 0 \quad \text{for all } s.$$

We define in spherical coordinates on the base the squares

$$\begin{aligned} Q &:= \left\{(\mu, \lambda) \in S^2 \setminus \{N, S\} \;\middle|\; 2\varepsilon \leq \mu \leq 1 - 2\varepsilon,\ |\lambda| \leq \tfrac{\pi}{4}\right\}, \\ \widetilde{Q} &:= \left\{(\mu, \lambda) \in S^2 \setminus \{N, S\} \;\middle|\; \varepsilon \leq \mu \leq 1 - \varepsilon,\ |\lambda| \leq \tfrac{\pi}{3}\right\}. \end{aligned}$$

Note that $Q \subset \operatorname{int} \widetilde{Q}$ and $\widetilde{Q} \subset \operatorname{int} B$, where B is the region defined at the beginning of this section over which ω is split. The family H_s^λ constructed above gives rise to a smooth function

$$H : \widetilde{Q} \times A \to \mathbb{R}, \quad (\lambda, \mu, w) \mapsto H_\mu^\lambda(w).$$

Let us write the fibre sphere as

$$S^2 = \mathrm{Cap}_N \cup A \cup \mathrm{Cap}_S,$$

where Cap_N and Cap_S denote the northern and southern polar caps, respectively. Then we can extend H first to $\widetilde{Q} \times S^2$ by the corresponding functions of (λ, μ) on the

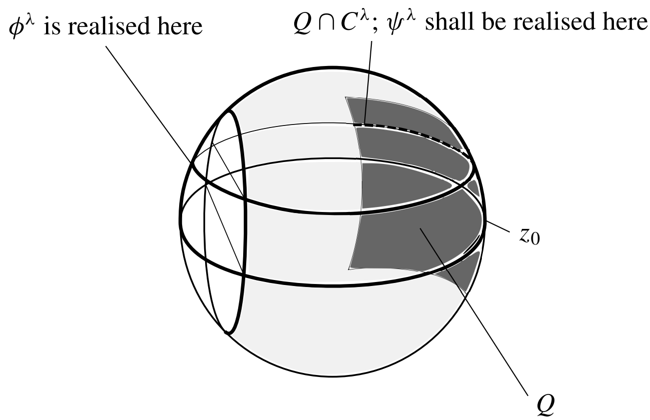

Figure 8. The path $Q \cap C^\lambda$ and $B \cap C^\lambda$.

southern and northern polar caps, and then to all of $S^2 \times S^2$ by zero outside $\widetilde{Q} \times S^2$. We still denote the resulting function by $H : S^2 \times S^2 \to \mathbb{R}$. By construction, H has support in $Q \times S^2$, it depends only on (λ, μ) outside $Q \times A$, and $H(0, \mu)|_E \equiv 0$ for all μ, where we set $H(\lambda, \mu) := H|_{p_1^{-1}(\lambda,\mu)}$.

4D. ***A special symplectic connection.*** Recall that we consider a relative symplectic fibration $(\mathcal{F}_{\mathrm{std}}, \omega, L_{\mathrm{std}}, S_0, S_\infty)$ such that the symplectic form ω is split on the set $(B \times S^2) \cup (S^2 \times (U_0 \cup U_\infty))$. Our current goal is to change the symplectic form ω, in its relative cohomology class in $H^2(S^2 \times S^2, L_{\mathrm{std}}; \mathbb{R})$, to a form ω' which has trivial holonomy around the circles of latitude. To explain the idea, consider a circle of latitude C^λ (see Figure 8). As the symplectic form is split over B, its parallel transport equals the identity along the part of C^λ lying within B, so the holonomy ϕ^λ is realised by travelling along the part of C^λ outside B.

The idea is now to modify ω to ω' such that the symplectic connection of ω' agrees with that of ω outside $Q \times S^2$ and realises the inverse holonomy ψ^λ along $C^\lambda \cap Q$ for all λ.

For the following computations, let us rename the coordinates (λ, μ) to

$$x := \mu \in [0, 1], \quad y := \lambda \in \left[-\tfrac{\pi}{3}, \tfrac{\pi}{3}\right].$$

Recall that the function $H : S^2 \times S^2 \to \mathbb{R}$ constructed in the previous subsection has support in $Q \times S^2$, where $Q = [2\varepsilon, 1 - 2\varepsilon] \times \left[-\tfrac{\pi}{4}, \tfrac{\pi}{4}\right]$ in the new coordinates (x, y). Consider the closed 2-form

$$\Omega_H = \omega + dH \wedge dx$$

on $Q \times A$, extended by ω to a form on all of $S^2 \times S^2$. Since Ω_H is vertically nondegenerate, the Ω_H-orthogonal complements to the tangent spaces of the fibres of p_1 induce a symplectic connection on $S^2 \times S^2$.

Lemma 4.3. (a) *The holonomy of Ω_H along each circle of latitude C^λ is trivial.*

(b) *The closed form Ω_H vanishes on L_{std} and is relatively cohomologous to ω.*

Proof. (a) Recall that H depends only on x and y outside the set $Q \times A$, so Ω_H and ω differ there by the pullback $dH \wedge dx = p_1^*\alpha$ of a 2-form α from the base. Since adding the pullback of a 2-form from the base does not change the symplectic connection (because $\iota_v(p_1^*\alpha) = 0$ for every vertical vector v), the induced connections of Ω_H and ω agree outside the set $Q \times A$. Within $Q \times A$ the form $\omega = p_1^*\sigma_0 + p_2^*\sigma_{\mathrm{std}}$ is split, so that its induced connection is flat. The horizontal spaces of the induced connection of Ω_H are spanned by the horizontal lifts of the coordinate vector fields ∂_x, ∂_y. These can be easily seen to be

$$\tilde{\partial}_x = \partial_x + X_{H_x^y}, \qquad \tilde{\partial}_y = \partial_y,$$

where $X_{H_x^y}$ is the Hamiltonian vector field of the Hamiltonian function $H_x^y(w) = H(x, y, w)$ on the annulus $(A, \sigma_{\mathrm{std}})$. To see this, let us write $\tilde{\partial}_x = \partial_x + v_x$ with a vertical vector v_x. This is horizontal if and only if

$$0 = \Omega_H(\tilde{\partial}_x, v) = \Omega_H(\partial_x, v) + \Omega_H(v_x, v) = -dH(v) + \sigma_{\mathrm{std}}(v_x, v)$$

for all vertical vectors v, which just means that v_x is the Hamiltonian vector field of H_x^y with respect to σ_{std}. A similar calculation shows that $\tilde{\partial}_y = \partial_y$.

It follows that the parallel transport of Ω_H along an interval of latitude $C^\lambda \cap Q \cong [2\varepsilon, 1 - 2\varepsilon] \times \{y\}$ is the time-1 map of the Hamiltonian flow of the time-dependent Hamiltonian H_s^λ. By construction of H_s^λ, this is the inverse ψ^λ of the holonomy of ω, and thus of Ω_H, along the interval $C^\lambda \setminus Q$. Hence the total holonomy of Ω_H along each circle of latitude C^λ is trivial.

(b) By construction, the horizontal vector field $\tilde{\partial}_x = \partial_x + X_{H_x^y}$ is tangent to L_{std}. Let v be the vertical vector field along L_{std} given by the positively oriented unit tangent vectors to the equators in the fibres. Since $\Omega_H(\tilde{\partial}_x, v) = 0$ by definition of horizontality, this shows that L_{std} is Lagrangian for Ω_H. Finally, let us compute the relative homology class of Ω_H in $H^2(S^2 \times S^2, L_{\mathrm{std}})$. For this, we evaluate Ω_H on the generators of $H_2(S^2 \times S^2, L_{\mathrm{std}})$ in Lemma 2.10:

$$\int_{S^2 \times \mathrm{pt}} \Omega_H = \int_{\mathrm{pt} \times S^2} \Omega_H = 1,$$

$$\int_{\mathrm{pt} \times D_{\mathrm{lh}}} \Omega_H = \int_{\mathrm{pt} \times D_{\mathrm{lh}}} \omega = \frac{1}{2},$$

$$\int_{D_{\mathrm{lh}} \times \mathrm{pt}} \Omega_H = \frac{1}{2} + \int_{(Q \cap \{y \leq 0\}) \times \{e\}} dH \wedge dx = \frac{1}{2} + \int_0^1 H(x, 0, e)\, dx = \frac{1}{2}.$$

Here in the last equation $e \in E$ is a base point on the equator in the fibre and we have used the normalisation condition $H(x, 0, e) = H(x, 0)|_E \equiv 0$ from the previous subsection. Since ω takes the same values on these classes by monotonicity of L_{std}, this shows that the relative cohomology classes of Ω_H and ω agree. $\square$

Let us analyse when the form Ω_H is symplectic. Since it is closed, this is equivalent to the form $\Omega_H \wedge \Omega_H$ being a volume form on $S^2 \times S^2$. This is clearly satisfied outside the set $Q \times S^2$ because there $H \equiv 0$. On the set $Q \times S^2$, the form ω is split of the form $\omega = p_1^*\sigma_0 + p_2^*\sigma_{\mathrm{std}}$. We work in the chosen coordinates x, y and write the form on the base as

$$\sigma_0 = f(x, y)\, dx \wedge dy,$$

with a positive function f. A short computation yields

$$\Omega_H \wedge \Omega_H = \left(1 - \frac{1}{f}\frac{\partial H}{\partial y}\right)\omega \wedge \omega.$$

So Ω_H will be symplectic if and only if

$$1 - \frac{1}{f}\frac{\partial H}{\partial y} > 0$$

everywhere. A priori, this need not be true for the given function H, but it can be remedied by the inflation procedure in the next subsection.

4E. ***Inflation.*** In this subsection we recall the inflation procedure of [Lalonde and McDuff 1996], suitably adapted to our situation. Let f_σ, $\bar{f}_\tau$ be two smooth nonnegative bump functions on S^2, where we think of f_σ as living on the fibre sphere and of $\bar{f}_\tau$ as living on the base sphere; see Figure 9. We require that

$$\operatorname{supp}(f_\sigma) \subset (U_0 \cup U_\infty) \setminus A = \mathrm{Cap}_S \amalg \mathrm{Cap}_N,$$

where U_0, U_∞ are the neighbourhoods of S, N over which ω is split and A is the annulus from the previous subsection, and that

$$\int_{\mathrm{Cap}_S} f_\sigma \sigma_{\mathrm{std}} = \int_{\mathrm{Cap}_N} f_\sigma \sigma_{\mathrm{std}} = \frac{1}{2}.$$

In particular, $\int_{S^2} f_\sigma \sigma_{\mathrm{std}} = 1$. The function $\bar{f}_\tau$ is required to have support in $\widetilde{Q}$ and satisfy

$$\bar{f}_\tau(x, y) = \bar{f}_\tau(x, -y)$$

as well as $\bar{f}_\tau|_Q \equiv 1$. We define

$$f_\tau := \frac{\bar{f}_\tau}{af}, \quad \text{with } a := \int_{\widetilde{Q}} \frac{\bar{f}_\tau}{f}\sigma_0 = \int_{\widetilde{Q}} \bar{f}_\tau\, dx \wedge dy,$$

where $\sigma_0 = f(x, y)\, dx \wedge dy$ as above. Then $\int_{\widetilde{Q}} f_\tau \sigma_0 = \frac{1}{a}\int_{\widetilde{Q}} \bar{f}_\tau\, dx \wedge dy = 1$, which by the symmetry of $\bar{f}_\tau$ implies

$$\int_{\widetilde{Q} \cap \{y \geq 0\}} f_\tau \sigma_0 = \frac{1}{a}\int_{\widetilde{Q} \cap \{y \geq 0\}} \bar{f}_\tau\, dx \wedge dy = \frac{1}{2}.$$

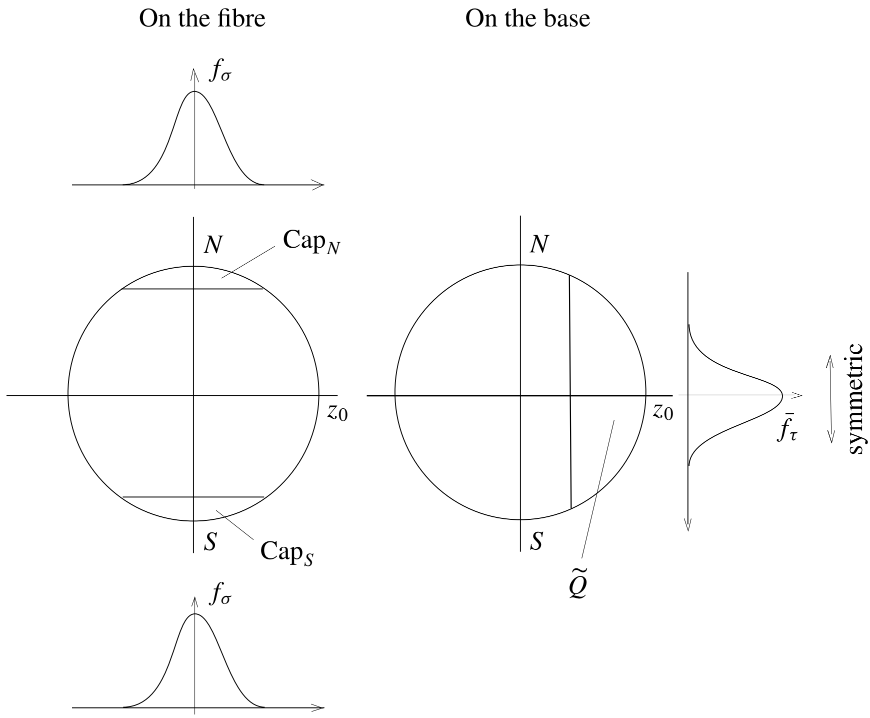

Figure 9. The functions f_σ and $\bar{f}_\tau$.

We define the two nonnegative 2-forms

$$\sigma := f_\sigma \sigma_{\mathrm{std}}, \quad \tau := f_\tau \sigma_0$$

on S^2 and consider the family of 2-forms on $S^2 \times S^2$

$$\omega_c := \frac{1}{c+1}(\omega + c p_1^* \tau + c p_2^* \sigma), \quad c \geq 0. \tag{4}$$

Lemma 4.4. *For each $c \geq 0$ the form ω_c has the following properties*:

(a) *ω_c is symplectic and L_{std} is Lagrangian for ω_c.*

(b) *ω_c is cohomologous to $\omega_0 = \omega$ in $H^2(S^2 \times S^2, L_{\mathrm{std}}; \mathbb{R})$.*

(c) *ω_c induces the same symplectic connection as ω.*

Proof. (a) First note that ω_c is closed for all $c \geq 0$ and

$$\omega_c = \frac{1}{c+1}\omega \quad \text{outside } W := W_0 \cup W_\infty,$$

where W_0, W_∞ are the sets from (3) on which ω is split. On the set W_0,

$$\omega_c = \frac{1}{c+1}\big((1 + c p_1^* f_\tau) p_1^* \sigma_0 + (1 + c p_2^* f_\sigma) p_2^* \sigma_{\mathrm{std}}\big), \tag{5}$$

and therefore

$$\omega_c \wedge \omega_c = \frac{1}{(c+1)^2}(1+cp_1^* f_\tau)(1+cp_2^* f_\sigma)\,\omega\wedge\omega > 0 \tag{6}$$

because c, f_τ, f_σ are nonnegative. For the set W_∞, we write $\sigma_\infty = g\sigma_0$ for a positive function g. Then on W_∞ we have

$$\omega_c = \frac{1}{c+1}\big((p_1^* g + cp_1^* f_\tau)p_1^*\sigma_0 + (1+cp_2^* f_\sigma)p_2^*\sigma_{\mathrm{std}}\big),$$

and again positivity of g and nonnegativity of c, f_τ, f_σ imply

$$\omega_c \wedge \omega_c = \frac{1}{(c+1)^2}(p_1^* g + cp_1^* f_\tau)(1+cp_2^* f_\sigma)\,\omega\wedge\omega > 0.$$

This proves that ω_c is symplectic. The torus L_{std} is Lagrangian for ω_c because all pullback forms from the base or the fibre vanish on L_{std}.

(b) To show that ω_c is relatively cohomologous to ω, we evaluate it on the basis of $H_2(S^2\times S^2, L_{\mathrm{std}})$ from Lemma 2.10. Using $\int_{S^2}\sigma = \int_{S^2}\tau = 1$, $\int_{D_{\mathrm{lh}}}\sigma = \int_{\mathrm{Cap}_S}\sigma = \frac{1}{2}$ and $\int_{D_{\mathrm{lh}}}\tau = \int_{\widetilde{Q}\cap\{y\le 0\}}\tau = \frac{1}{2}$, we compute with the point $z_0 \in E$ on the equator in the base or fibre:

$$\begin{aligned}
\int_{\mathrm{pt}\times S^2}\omega_c &= \frac{1}{c+1}\int_{\mathrm{pt}\times S^2}(\omega + cp_2^*\sigma) = \frac{1}{c+1}(1+c) = 1,\\
\int_{S^2\times\mathrm{pt}}\omega_c &= \frac{1}{c+1}\int_{S^2\times\mathrm{pt}}(\omega + cp_1^*\tau) = \frac{1}{c+1}(1+c) = 1,\\
\int_{\mathrm{pt}\times D_{\mathrm{lh}}}\omega_c &= \frac{1}{c+1}\int_{\{z_0\}\times D_{\mathrm{lh}}}(\omega + cp_2^*\sigma) = \frac{1}{c+1}\Big(\frac{1}{2}+\frac{c}{2}\Big) = \frac{1}{2},\\
\int_{D_{\mathrm{lh}}\times\mathrm{pt}}\omega_c &= \frac{1}{c+1}\int_{D_{\mathrm{lh}}\times\{z_0\}}(\omega + cp_1^*\tau) = \frac{1}{c+1}\Big(\frac{1}{2}+\frac{c}{2}\Big) = \frac{1}{2}.
\end{aligned}$$

By monotonicity of L_{std}, the form ω takes the same values on these classes, so $[\omega_c] = [\omega] \in H^2(S^2\times S^2, L_{\mathrm{std}}; \mathbb{R})$.

(c) On the set $W = W_0 \cup W_\infty$ the forms ω_c and ω are both split; hence both symplectic connections are flat and the horizontal subspaces are the tangent spaces to the other cartesian factor. Outside W we have $\omega_c = \omega/(c+1)$ and, since the symplectic complements to the fibres are not affected by scaling of the symplectic form, the symplectic connections of ω_c and ω agree here as well. □

A new symplectic connection. Now recall that the function H from the previous section is a pullback from the base outside the set $S^2\times A$. On the set $S^2\times A$, the function $p_2^* f_\sigma$ vanishes and thus

$$\omega_c = \frac{1}{c+1}\big((1+p_1^* f_\tau)p_1^*\sigma_0 + p_2^*\sigma_{\mathrm{std}}\big).$$

In particular, on this set the restriction of ω_c to the fibres it is just the standard form σ_{std} scaled by $1/(c+1)$. Now the fibrewise Hamiltonian vector field of the rescaled function $H/(c+1)$ with respect to $\sigma_{\mathrm{std}}/(c+1)$ equals the fibrewise Hamiltonian vector field $X_{H_x^y}$ of H with respect to σ_{std}. So the horizontal lift of ∂_x with respect to the closed 2-form

$$\Omega_H^c := \omega_c + \frac{1}{c+1} dH \wedge dx \tag{7}$$

agrees with its horizontal lift $\partial_x + X_{H_x^y}$ with respect to Ω_H (see the proof of Lemma 4.3), and since the horizontal lift of ∂_y is ∂_y in both cases, we see that Ω_H^c and Ω_H define the same symplectic connection for all $c \geq 0$. Moreover, the proof of Lemma 4.3(b) shows that Ω_H^c vanishes on L_{std} and is relatively cohomologous to ω_c, and thus to ω by Lemma 4.4.

Symplecticity. Again, let us analyse when the form Ω_H^c is symplectic. Outside $Q \times S^2$, the form Ω_H^c is just ω_c, which is symplectic by Lemma 4.4. On the set $Q \times S^2 \subset W_0$, using (5) and (6) we compute

$$\begin{aligned}
\omega_c \wedge \frac{1}{c+1} dH \wedge dx &= \frac{1}{(c+1)^2}\left(-\frac{\partial H}{\partial y}\right)(1 + c p_2^* f_\sigma)\, dx \wedge dy \wedge p_2^* \sigma_{\mathrm{std}} \\
&= \frac{1}{2f(c+1)^2}\left(-\frac{\partial H}{\partial y}\right)(1 + c p_2^* f_\sigma)\omega \wedge \omega,
\end{aligned}$$

and

$$\begin{aligned}
\Omega_H^c \wedge \Omega_H^c &= \omega_c \wedge \omega_c + 2\omega_c \wedge \frac{1}{c+1} dH \wedge dx \\
&= \frac{1}{(c+1)^2}\left(1 + c p_1^* f_\tau - \frac{1}{f}\frac{\partial H}{\partial y}\right)(1 + c p_2^* f_\sigma)\omega \wedge \omega.
\end{aligned}$$

Now $1 + c p_2^* f_\sigma \geq 1$ for all c by nonnegativity of c and f_σ. Moreover, by the choice of f_τ we have $p_1^* f_\tau = 1/(af)$ on $Q \times S^2$. Hence Ω_H^c is symplectic if and only if

$$1 + \frac{1}{f}\left(\frac{c}{a} - \frac{\partial H}{\partial y}\right) > 0$$

on $Q \times S^2$. But this is satisfied for

$$c \geq C := a \max_{Q \times S^2}\left|\frac{\partial H}{\partial y}\right|. \tag{8}$$

We summarise the preceding discussion in:

Lemma 4.5. *The closed 2-form Ω_H^c vanishes on L_{std}, is relatively cohomologous to ω_c (and thus ω), and has trivial holonomy along each circle of latitude C^λ for each $c \geq 0$. Moreover, Ω_H^c is symplectic for $c \geq C$ given by (8).*

4F. ***Killing the holonomy along circles of latitude.*** We denote the 0-meridian by $m_0 := \{(\lambda, \mu) \in S^2 \mid \mu = 0\}$ in spherical coordinates. Putting the previous subsections together, we can now prove:

Proposition 4.6. *Let $\mathfrak{S} = (\mathcal{F}_{\mathrm{std}}, \omega, L_{\mathrm{std}}, S_0, S_\infty)$ be a relative symplectic fibration such that ω is split on a neighbourhood of the fibre F and the sections S_0, S_∞. Then there exists a homotopy of relative symplectic fibrations $\mathfrak{S}_t = (\mathcal{F}_{\mathrm{std}}, \omega_t, L_{\mathrm{std}}, S_0, S_\infty)$ with $\mathfrak{S}_0 = \mathfrak{S}$ such that the holonomy of $\mathfrak{S}_1$ along the circles of latitude C^λ is the identity for all λ. Moreover, ω_1 is split near the set $(m_0 \times S^2) \cup S_0 \cup S_\infty$.*

Proof. As explained at the beginning of Section 4A, we may assume that ω is split on a set $(B \times S^2) \cup (S^2 \times (U_0 \cup U_\infty))$, where the ball $B \subset S^2$ contains the 0-meridian m_0. Let H be the Hamiltonian function constructed in Section 4C and let C be the constant defined in (8). For $c \in [0, C]$, let ω_c be the form defined in (4). By Lemma 4.4, $(\mathcal{F}_{\mathrm{std}}, \omega_c, L_{\mathrm{std}}, S_0, S_\infty)$ gives a homotopy of relative symplectic fibrations from $\mathfrak{S}$ to $(\mathcal{F}_{\mathrm{std}}, \omega_C, L_{\mathrm{std}}, S_0, S_\infty)$. For $t \in [0, 1]$, consider the forms

$$\Omega^C_{tH} = \omega_C + \frac{t}{C+1} dH \wedge dx = (1-t)\omega_C + t\,\Omega^C_H$$

as in (7) (with H replaced by tH and c by C). By Lemma 4.5 (applied to tH), $(\mathcal{F}_{\mathrm{std}}, \Omega^C_{tH}, L_{\mathrm{std}}, S_0, S_\infty)$ gives a homotopy of relative symplectic fibrations from $(\mathcal{F}_{\mathrm{std}}, \omega_C, L_{\mathrm{std}}, S_0, S_\infty)$ to $\mathfrak{S}_1 = (\mathcal{F}_{\mathrm{std}}, \Omega^C_H, L_{\mathrm{std}}, S_0, S_\infty)$. By the same lemma, $\mathfrak{S}_1$ has trivial holonomy along all circles of latitude. Hence the concatenation of the previous two homotopies gives the desired homotopy $\mathfrak{S}_t$. For the last assertion, simply observe that by construction all symplectic forms in this homotopy agree with the original split form ω near $(m_0 \times S^2) \cup S_0 \cup S_\infty$. $\square$

Remark 4.7. The point of departure for the preceding subsections was the standardisation provided by Proposition 3.11. If rather than making the symplectic form ω split we had made it equal to ω_{std} near $(m_0 \times S^2) \cup S_0 \cup S_\infty$ (as suggested in Remark 3.15), then the holonomies ϕ^λ would lie in the subgroup $\mathrm{Ham}(A, \partial A, \sigma_{\mathrm{std}}) \subset \mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ of symplectomorphisms generated by Hamiltonians with compact support in $A \setminus \partial A$ and the whole construction could be performed in that subgroup (which is also contractible). However, since we change the normalisation of the Hamiltonians H^λ_s anyway to make them vanish on the equator, we would gain nothing from working in this subgroup.

4G. ***Killing all the holonomy.*** Now we will further deform the relative symplectic fibration from the previous subsection to one which has trivial holonomy along *all* closed curves in the base. We begin with a simple lemma.

Lemma 4.8. *Let ω, ω' be linear symplectic forms on $\mathbb{R}^4$ which define the same orientation and agree on a real codimension one hyperplane H. Then $\omega_t := (1-t)\omega + t\omega'$ is symplectic for all $t \in [0, 1]$.*

Proof. Take a symplectic basis e_1, f_1, e_2, f_2 for ω such that e_1, f_1, e_2 is a basis of H. Take a vector $f_2' = a_1e_1 + b_1f_1 + a_2e_2 + b_2f_2$ such that e_1, f_1, e_2, f_2' is a symplectic basis for ω'. Since ω, ω' induce the same orientation, we have $b_2 > 0$, and therefore

$$\omega(e_2, f_2') = b_2 > 0, \quad \omega'(e_2, f_2) = \frac{1}{b_2} > 0.$$

For $\omega_t := (1-t)\omega + t\omega'$ we find

$$\omega_t \wedge \omega_t = (1-t)^2\omega\wedge\omega + 2t(1-t)\omega\wedge\omega' + t^2\omega'\wedge\omega',$$

and therefore

$$\begin{aligned}\omega_t\wedge\omega_t(e_1, f_1, e_2, f_2') &= 2(1-t)^2\omega(e_1, f_1)\omega(e_2, f_2') + 2t^2\omega'(e_1, f_1)\omega'(e_2, f_2') \\ &\quad + 2t(1-t)\big(\omega(e_1, f_1)\omega'(e_2, f_2') + \omega(e_2, f_2')\omega'(e_1, f_1)\big) \\ &> 0.\end{aligned}$$

□

Recall the definition of the 0-meridian m_0 from the previous subsection.

Proposition 4.9. *Let $\mathfrak{S} = (\mathcal{F}_{\rm std}, \omega, L_{\rm std}, S_0, S_\infty)$ be a relative symplectic fibration which is split near the set $S_0 \cup S_\infty \cup (m_0 \times S^2)$ and has trivial holonomy around all circles of latitude C^λ. Then there exists a homotopy of relative symplectic fibrations $\mathfrak{S}_t = (\mathcal{F}_{\rm std}, \omega_t, L_{\rm std}, S_0, S_\infty)$ with $\omega_0 = \omega$ and $\omega_1 = \omega_{\rm std}$.*

Proof. The idea of the proof is to use parallel transport along circles of latitude to define a fibre-preserving diffeomorphism ϕ of $S^2 \times S^2$ which pulls back the symplectic form ω to a form which agrees with the standard form $\omega_{\rm std}$ on $C^\lambda \times S^2$ for all λ, and then apply Lemma 4.8.

For each λ, μ let

$$P^\lambda_\mu : \{(\lambda, 0)\} \times S^2 \to \{(\lambda, \mu)\} \times S^2$$

be the parallel transport of (the symplectic connection on p_1 defined by) ω along the circle of latitude C^λ from $(\lambda, 0)$ to (λ, μ). Since ω has trivial holonomy along C^λ, this does not depend on the path in C^λ and is thus well-defined. Note that, due to the fact that ω is split near $S_0 \cup S_\infty$, the map P^λ_μ equals the identity near the north and south poles N, S in the fibre.

We define a fibre-preserving diffeomorphism ϕ of $S^2 \times S^2$ by parallel transport on the left sphere in Figure 10 with respect to the standard form $\omega_{\rm std}$ first going backwards along the circle of latitude until we hit the meridian m_0, then upwards along m_0 until we hit the north pole N, then by the identity to the fibre over the north pole of the right sphere, then by parallel transport with respect to ω along m_0, and finally along the circle of latitude to land in the fibre over the original point (λ, μ). Since parallel transport with respect to the symplectic connection $\omega_{\rm std}$ is the identity for all paths, as is parallel transport with respect to ω along paths in m_0 (since ω is

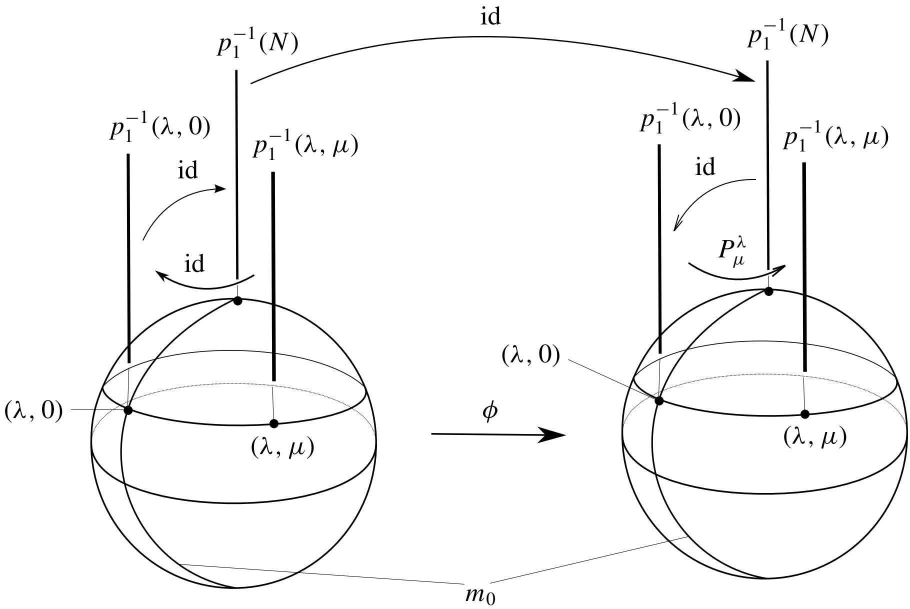

Figure 10. The maps P_μ^λ and the construction of ϕ.

split over m_0), we will not explicitly include these maps in the notation. Then we can write the preceding definition in formulas as

$$\phi : S^2 \times S^2 \to S^2 \times S^2, \quad ((\lambda, \mu), w) \mapsto ((\lambda, \mu), P_\mu^\lambda(w)).$$

Note that ϕ is smooth for (λ, μ) near the north and south poles because there $P_\mu^\lambda = \mathrm{id}$. For $z \in S^2$ let us denote by ω_z the restriction of ω to the fibre $F_z = \{z\} \times S^2$. We claim that ϕ has the following properties:

(a) ϕ restricts to symplectomorphisms $(F_z, \sigma_{\mathrm{std}}) \to (F_z, \omega_z)$ on all fibres.

(b) ϕ preserves the Clifford torus L_{std}.

(c) ϕ equals the identity near $S_0 \cup S_\infty \cup (m_0 \times S^2)$.

(d) ϕ is isotopic to the identity through fibre-preserving diffeomorphisms ϕ_t that preserve L_{std} and equal the identity near $S_0 \cup S_\infty \cup (m_0 \times S^2)$.

For (a), note that ω restricts to σ_{std} on the fibre F_N over the north pole (because ω is split there), so the identity defines a symplectomorphism $(F_N, \sigma_{\mathrm{std}}) \to (F_N, \omega_N)$. Now (a) follows because parallel transport is symplectic.

Property (b) follows from the fact that L_{std} is given by parallel transport of the equator in the fibre around the equator in the base; hence $P_\mu^0(E) = E$ and thus

$$\phi(\{(0, \mu)\} \times E) = \{(0, \mu)\} \times P_\mu^0(E) = \{(0, \mu)\} \times E.$$

Property (c) holds because ω is split near $S_\infty \cup S_0 \cup (m_0 \times S^2)$.

For (d), consider the map

$$P : R := \left[-\tfrac{\pi}{2}, \tfrac{\pi}{2}\right] \times [0, 2\pi] \to \mathrm{Diff}(A, \partial A), \quad (\lambda, \mu) \mapsto P^\lambda_\mu.$$

It maps the boundary ∂R to id and the interval $\{0\} \times [0, 2\pi]$ to the subspace $\mathrm{Diff}(A, \partial A; E) \subset \mathrm{Diff}(A, \partial A)$ of diffeomorphisms preserving the equator E (as a set, not pointwise). By Corollary A.3, the loop $[0, 2\pi] \ni \mu \mapsto P^0_\mu$ is contractible in $\mathrm{Diff}(A, \partial A; E)$. Using this and the fact (from Corollary A.2) that $\pi_2 \mathrm{Diff}(A, \partial A) = 0$, we find a contraction of P through maps $P_t : R \to \mathrm{Diff}(A, \partial A)$ sending ∂R to id and $\{0\} \times [0, 2\pi]$ to $\mathrm{Diff}(A, \partial A; E)$. (The argument is analogous to the proof of Proposition 4.1, with the 1-parametric family ψ^λ replaced by the 2-parametric family P^λ_μ.) Then $\phi_t((\lambda, \mu), w) := ((\lambda, \mu), P_t(\lambda, \mu)(w))$ is the desired isotopy and the claim is proved.

Now we construct the homotopy from ω to ω_{std} in two steps. For the first step, let ϕ_t be the isotopy in (d) from $\phi_0 = \mathrm{id}$ to $\phi_1 = \phi$. Then $\phi_t^{-1}(\mathfrak{S}) = (\mathcal{F}_{\mathrm{std}}, \phi_t^*\omega, L_{\mathrm{std}}, S_0, S_\infty)$ is a homotopy of relative symplectic fibrations from $\mathfrak{S}$ to $(\mathcal{F}_{\mathrm{std}}, \phi^*\omega, L_{\mathrm{std}}, S_0, S_\infty)$.

For the second step, note that $\phi^*\omega$ restricts to σ_{std} on every fibre by property (a). Moreover, since ϕ commutes with parallel transport along C^λ (with respect to ω_{std} and ω), the horizontal lifts of vectors tangent to circles of latitude with respect to ω_{std} and $\phi^*\omega$ agree. Accordingly, ω_{std} and $\phi^*\omega$ agree on the 3-dimensional subspaces $T_{((\lambda,\mu),w)}(C^\lambda \times S^2)$ in $T_{((\lambda,\mu),w)}(S^2 \times S^2)$ for all $((\lambda, \mu), w) \in S^2 \times S^2$. Thus, by Lemma 4.8, the linear interpolations $\omega_t := (1-t)\phi^*\omega + t\omega_{\mathrm{std}}$ are symplectic for all $t \in [0, 1]$.

We claim that $(\mathcal{F}_{\mathrm{std}}, \omega_t, L_{\mathrm{std}}, S_0, S_\infty)$ is a relative symplectic fibration for all $t \in [0, 1]$. For this, first note that L_{std} is monotone Lagrangian for both ω_{std} and $\phi^*\omega$: for ω_{std} this is clear, and for $\phi^*\omega$ it follows from $\phi(L_{\mathrm{std}}) = L_{\mathrm{std}}$ and monotonicity of L_{std} for ω. Hence L_{std} is monotone Lagrangian for ω_t for all t. Next, since ϕ preserves fibres as well as the sections S_0, S_∞, the form $\phi^*\omega$ and thus also the form ω_t is cohomologous to ω_{std} for all t. Finally, since ϕ preserves the sections S_0, S_∞, they are symplectic for $\phi^*\omega$ as well as ω_{std}, and hence for all ω_t.

The desired homotopy ω_t is the concatenation of the homotopies constructed in the two steps. This concludes the proof of Proposition 4.9. □

Remark 4.10. The two steps in the proof of Proposition 4.9 could have been performed in the opposite order: First homotope ω to the form $\phi_*\omega_{\mathrm{std}}$ which has trivial holonomy along all closed curves in the base, and then homotope $\phi_*\omega_{\mathrm{std}}$ to ω_{std}. The latter is then a special case of the more general fact that two symplectic fibrations with conjugate holonomy (e.g., both having trivial holonomy) are diffeomorphic.

4H. ***Proof of the main theorem and some consequences.*** We summarise the results of this and the previous section in:

Theorem 4.11 (classification of relative symplectic fibrations). *Every relative symplectic fibration* $\mathfrak{S} = (\mathcal{F}, \omega, L, \Sigma, \Sigma')$ *on* $S^2 \times S^2$ *is deformation equivalent to* $\mathfrak{S}_{\mathrm{std}} = (\mathcal{F}_{\mathrm{std}}, \omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$.

Proof. By Proposition 3.2, $\mathfrak{S}$ is diffeomorphic to a relative symplectic fibration of the form $\widetilde{\mathfrak{S}} = (\mathcal{F}_{\mathrm{std}}, \tilde{\omega}, L_{\mathrm{std}}, S_0, S_\infty)$ for some symplectic form $\tilde{\omega}$. Combining Propositions 3.11, 4.6 and 4.9, we find a homotopy from $\widetilde{\mathfrak{S}}$ to $\mathfrak{S}_{\mathrm{std}}$. □

The main theorem will be a consequence of Theorem 4.11 and the following theorem of Gromov.

Theorem 4.12 [Gromov 1985]. *Let* $\phi \in \mathrm{Symp}(S^2 \times S^2, \omega_{\mathrm{std}})$ *act trivially on homology. Then there exists a symplectic isotopy* $\phi_t \in \mathrm{Symp}(S^2 \times S^2, \omega_{\mathrm{std}})$ *with* $\phi_0 = \mathrm{id}$ *and* $\phi_1 = \phi$.

A first consequence is:

Corollary 4.13. *Let* $\mathfrak{S} = (\mathcal{F}, \omega_{\mathrm{std}}, L, \Sigma, \Sigma')$ *be a relative symplectic fibration of* $M = S^2 \times S^2$, *where* ω_{std} *is the standard symplectic form. Then there exists a homotopy of relative symplectic fibrations* $\mathfrak{S}_t = (\mathcal{F}_t, \omega_{\mathrm{std}}, L_t, \Sigma_t, \Sigma'_t)$ *with fixed symplectic form* ω_{std} *such that* $\mathfrak{S}_0 = \mathfrak{S}$ *and* $\mathfrak{S}_1 = (\mathcal{F}_{\mathrm{std}}, \omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$.

Proof. By Theorem 4.11, there exists a homotopy of relative symplectic fibrations $\mathfrak{S}_t = (\mathcal{F}_t, \omega_t, L_t, \Sigma_t, \Sigma'_t)$ with $\mathfrak{S}_1 = (\mathcal{F}_{\mathrm{std}}, \omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$ and a diffeomorphism ϕ of $S^2 \times S^2$ acting trivially on homology such that $\phi(\mathfrak{S}) = \mathfrak{S}_0$. After applying Proposition 3.5 and modifying ϕ accordingly (keeping the same notation), we may assume that $\omega_t = \omega_{\mathrm{std}}$ for all $t \in [0, 1]$. Then ϕ is a symplectomorphism with respect to ω_{std}, so by Gromov's theorem, Theorem 4.12, it can be connected to the identity by a family of symplectomorphisms ϕ_t. Now the concatenation of the homotopies $\phi_t(\mathfrak{S})_{t\in[0,1]}$, and $(\mathfrak{S}_t)_{t\in[0,1]}$ is the desired homotopy with fixed symplectic form ω_{std}. □

Proof of the main theorem, Theorem 1.1. The hypotheses of Theorem 1.1 just mean that $\mathfrak{S} = (\mathcal{F}, \omega_{\mathrm{std}}, L, \Sigma, \Sigma')$ is a relative symplectic fibration. By Corollary 4.13, $\mathfrak{S}$ can be connected to $(\mathcal{F}_{\mathrm{std}}, \omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$ by a homotopy of relative symplectic fibrations $\mathfrak{S}_t = (\mathcal{F}_t, \omega_{\mathrm{std}}, L_t, \Sigma_t, \Sigma'_t)$ with fixed symplectic form ω_{std}. In particular, L_t is an isotopy of monotone Lagrangian tori (with respect to ω_{std}) from $L_0 = L$ to $L_1 = L_{\mathrm{std}}$. By Banyaga's isotopy extension theorem, there exists a symplectic isotopy ϕ_t with $\phi_0 = \mathrm{id}$ and $\phi_t(L) = L_t$ for all t (see the proof of Proposition 3.5 for the argument, ignoring the symplectic sections). Since M is simply connected, the symplectic isotopy ϕ_t is actually Hamiltonian. □

Another consequence of Theorem 4.11 is the following result concerning standardisation by diffeomorphisms.

Corollary 4.14 (fixing the symplectic form). *Let $\mathfrak{S} = (\mathcal{F}, \omega, L, \Sigma, \Sigma')$ be a relative symplectic fibration of $M = S^2 \times S^2$. Then there exists a diffeomorphism ϕ of $S^2 \times S^2$ acting trivially on homology such that $\phi^{-1}(\mathfrak{S}) = (\widetilde{\mathcal{F}}, \omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$ for some foliation $\widetilde{\mathcal{F}}$.*

Proof. By Theorem 4.11, there exists a homotopy of relative symplectic fibrations $\mathfrak{S}_t = (\mathcal{F}_t, \omega_t, L_t, \Sigma_t, \Sigma'_t)$ with $\mathfrak{S}_1 = (\mathcal{F}_{\mathrm{std}}, \omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$ and a diffeomorphism ϕ of $S^2 \times S^2$ acting trivially on homology such that $\phi(\mathfrak{S}) = \mathfrak{S}_0$. After applying Proposition 3.5 and modifying ϕ accordingly (keeping the same notation), we may assume that $(\omega_t, L_t, \Sigma_t, \Sigma'_t) = (\omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$ for all $t \in [0, 1]$. Then ϕ maps $\mathfrak{S}$ to $(\mathcal{F}_0, \omega_{\mathrm{std}}, L_{\mathrm{std}}, S_0, S_\infty)$. □

In particular, Corollary 4.14 implies that every symplectic form ω on $S^2 \times S^2$ which is compatible with a relative symplectic fibration can be pulled back to ω_{std} by a diffeomorphism $\phi \in \mathrm{Diff}_{\mathrm{id}}(M)$. This is a special case of the deep result by [Lalonde and McDuff 1996] that every symplectic form ω on $S^2 \times S^2$ which is cohomologous to the standard form ω_{std} can be pulled back to ω_{std} by a diffeomorphism $\phi \in \mathrm{Diff}_{\mathrm{id}}(M)$. In fact, the hard part of the proof in [loc. cit.] (using Taubes' correspondence between Seiberg–Witten and Gromov invariants) consists in showing that any such ω is compatible with a symplectic fibration with a section.

Appendix: Homotopy groups of some diffeomorphism groups

In this appendix we collect some well-known facts about the diffeomorphism and symplectomorphism groups of the disk and annulus. We fix numbers $0 < a < b$ and let

$$D := \{z \in \mathbb{C} \mid |z| \le b\}, \qquad A := \{z \in \mathbb{C} \mid a \le |z| \le b\}$$

be equipped with the standard symplectic form

$$\sigma_{\mathrm{std}} = \frac{r}{\pi(1+r^2)^2}\, dr \wedge d\theta$$

in polar coordinates on $\mathbb{C}$ (the precise choice of the symplectic form does not matter because they are all isomorphic up to scaling by Moser's theorem). We define the following diffeomorphism groups, all equipped with the C^∞ topology:

- $\mathrm{Diff}(D, \partial D)$ is the group of diffeomorphisms of the closed disk D that are equal to the identity in some neighbourhood of the boundary.
- $\mathrm{Diff}(A, \partial A)$ is the group of diffeomorphisms of the closed annulus A that are equal to the identity in some neighbourhood of the boundary.
- $\mathrm{Symp}(A, \partial A, \sigma_{\mathrm{std}}) \subset \mathrm{Diff}(A, \partial A)$ is the subgroup of symplectomorphisms of $(A, \sigma_{\mathrm{std}})$.
- $\mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ is the identity component of $\mathrm{Symp}(A, \partial A, \sigma_{\mathrm{std}})$.

Diffeomorphisms. All results in this appendix are based on the following fundamental theorem of Smale.

Theorem A.1 [Smale 1959]. *The group* $\operatorname{Diff}(D, \partial D)$ *is contractible.*

With a nondecreasing cutoff function $\rho : \mathbb{R} \to [0, 1]$ which equals 0 near $(-\infty, a]$ and 1 near $[b, \infty)$, we define the *Dehn twist*

$$\phi^D : A \to A, \quad re^{i\theta} \mapsto re^{i(\theta+2\pi\rho(r))}.$$

Corollary A.2. *All homotopy groups* $\pi_i \operatorname{Diff}(A, \partial A)$ *vanish except for the group* $\pi_0 \operatorname{Diff}(A, \partial A) = \mathbb{Z}$, *which is generated by the Dehn twist* ϕ^D.

Proof. Restriction of elements in $\operatorname{Diff}(D, \partial D)$ to the smaller disk $D_a \subset D$ of radius a yields a Serre fibration

$$\operatorname{Diff}(A, \partial A) \to \operatorname{Diff}(D, \partial D) \to \operatorname{Diff}^+(D_a),$$

where Diff^+ denotes the orientation-preserving diffeomorphisms. In view of Smale's theorem, Theorem A.1, the long exact sequence of this fibration yields isomorphisms $\pi_i \operatorname{Diff}^+(D_a) \cong \pi_{i-1} \operatorname{Diff}(A, \partial A)$ for all $i \geq 1$. Again by Theorem A.1, the long exact sequence of the pair $(\operatorname{Diff}^+(D_a), \operatorname{Diff}^+(\partial D_a))$ yields isomorphisms $\pi_i \operatorname{Diff}^+(\partial D_a) \cong \pi_i \operatorname{Diff}^+(D_a)$ for all i. Since $\pi_i \operatorname{Diff}^+(\partial D_a) \cong \pi_i \operatorname{Diff}^+(S^1)$ equals $\mathbb{Z}$ for $i = 1$ and 0 otherwise, this proves the corollary. □

For the following slight refinement of Corollary A.2, let $E \subset A$ be a circle $\{c\} \times S^1$ for some $c \in (a, b)$.

Corollary A.3. *Every smooth loop* $(\phi_t)_{t\in[0,1]}$ *in* $\operatorname{Diff}(A, \partial A)$ *with* $\phi_0 = \phi_1 = \mathrm{id}$ *and* $\phi_t(E) = E$ *for all* t *can be contracted by a smooth family* $\phi_t^s \in \operatorname{Diff}(A, \partial A)$, $s, t \in [0, 1]$, *satisfying* $\phi_t^0 = \phi_0^s = \phi_1^s = \mathrm{id}$, $\phi_t^1 = \phi_t$ *and* $\phi_t^s(E) = E$ *for all* s, t.

Proof. For a point $e \in S^1$ the family of arcs $\phi_t([a, c] \times \{e\})$ starts and ends at $t = 0, 1$ with the arc $[a, c] \times \{e\}$. This shows that the loop $t \mapsto \phi_t(c, e)$ in E is contractible; hence so is the loop $\phi_t|_E$ in $\operatorname{Diff}(E)$. Thus we can find a family $\phi_t^s \in \operatorname{Diff}(A, \partial A)$, $s, t \in \left[\frac{1}{2}, 1\right]$, satisfying $\phi_0^s = \phi_1^s = \mathrm{id}$, $\phi_t^1 = \phi_t$, $\phi_t^s(E) = E$ for all s, t, and $\phi_t^{1/2}|_E = \mathrm{id}$ for all t. Now apply Corollary A.2 to contract the loops $\phi_t^{1/2}|_{[a,c]\times S^1}$ in $\operatorname{Diff}([a, c] \times S^1, \partial[a, c] \times S^1)$ and $\phi_t^{1/2}|_{[c,b]\times S^1}$ in $\operatorname{Diff}([c, b] \times S^1, \partial[c, b] \times S^1)$. □

Symplectomorphisms. The following is an immediate consequence of Corollary A.2 and Moser's theorem.

Proposition A.4. *The groups* $\operatorname{Diff}(A, \partial A)$ *and* $\operatorname{Symp}(A, \partial A, \sigma_{\mathrm{std}})$ *are weakly homotopy equivalent. Thus all homotopy groups of* $\operatorname{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ *vanish and* $\pi_0 \operatorname{Symp}(A, \partial A, \sigma_{\mathrm{std}}) = \mathbb{Z}$ *is generated by the (symplectic) Dehn twist* ϕ^D.

Finally, we need the following refinement of Proposition A.4. Again, let $E \subset A$ be a circle $\{c\} \times S^1$ for some $c \in (a, b)$.

Lemma A.5. *Each $\phi \in \mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ with $\phi(E) = E$ can be connected to the identity by a smooth path $\phi_t \in \mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ satisfying $\phi_t(E) = E$ for all $t \in [0, 1]$.*

Proof. After applying Moser's theorem and changing the values of a, b, c (viewing A as the cylinder $[a, b] \times S^1$), we may assume that $\sigma_{\mathrm{std}} = dr \wedge d\theta$ and $c = 0$. We connect ϕ to the identity in four steps.

Step 1. The restriction $f(\theta) := \phi(0, \theta)$ of ϕ to E defines an element in the group $\mathrm{Diff}^+(S^1)$ of orientation-preserving diffeomorphisms of the circle. Since this group is path connected, there exists a smooth family $f_t \in \mathrm{Diff}^+(S^1)$ with $f_0 = \mathrm{id}$ and $f_1 = f$. This family is generated by the time-dependent vector field ξ_t on the circle defined by $\xi_t(f_t(\theta)) := \dot{f}_t(\theta)$. Let $H_t : A \to \mathbb{R}$ be a smooth family of functions satisfying $H_t(r, \theta) = -r\xi_t(\theta)$ near E and $H_t = 0$ near ∂A. A short computation shows that the Hamiltonian vector field of H_t agrees with ξ_t on E. It follows that the Hamiltonian flow ψ_t of H_t satisfies $\psi_t|_E = f_t$; in particular $\psi_1|_E = f_1 = \phi|_E$. Thus $\phi_t := \phi \circ \psi_t^{-1}$ is a smooth path in $\mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ with $\phi_t(E) = E$ connecting ϕ to ϕ_1 satisfying $\phi_1|_E = \mathrm{id}$. After renaming ϕ_1 back to ϕ, we may hence assume that $\phi|_E = \mathrm{id}$.

Step 2. Let us write $\phi(r, \theta) = (P(r, \theta), Q(r, \theta)) \in \mathbb{R} \times S^1$. Since $\phi|_E = \mathrm{id}$ and ϕ is symplectic, the functions P, Q satisfy

$$P(0, \theta) = 0, \quad Q(0, \theta) = \theta, \quad \frac{\partial P}{\partial r}(0, \theta) = 1.$$

For $s \in (0, 1]$ consider the dilations $\tau_s(r, \theta) := (sr, \theta)$ on A. Since $\tau_s^*(dr \wedge d\theta) = s\, dr \wedge d\theta$, the maps $\psi_s := \tau_s^{-1} \circ \phi \circ \tau_s : A \to \mathbb{R} \times S^1$ are symplectic and preserve $E = \{0\} \times S^1$. Since

$$\psi_s(r, \theta) = \Big(\frac{1}{s} P(sr, \theta), Q(sr, \theta)\Big) \xrightarrow{s \to 0} \Big(r \frac{\partial P}{\partial r}(0, \theta), Q(0, \theta)\Big) = (r, \theta),$$

the family ψ_s extends smoothly to $s = 0$ by the identity (this is a fibred version of the Alexander trick). It follows that for a sufficiently small $\varepsilon > 0$ we have a smooth family of symplectic embeddings $\psi_s : A_\varepsilon := [-\varepsilon, \varepsilon] \times S^1 \hookrightarrow A$, $s \in [0, 1]$, with $\psi_s(E) = E$, $\psi_0 = \mathrm{id}$, and $\psi_1 = \phi|_{A_\varepsilon}$. We extend this family to smooth diffeomorphisms $\tilde{\psi}_s : A \to A$ with $\tilde{\psi}_s = \mathrm{id}$ near ∂A and $\tilde{\psi}_1 = \phi$. Since $\tilde{\psi}_s$ preserves the annuli $A^- := [a, 0] \times S^1$ and $A^+ := [0, b] \times S^1$, it satisfies $\int_{A^\pm} \tilde{\psi}_s^* \sigma_{\mathrm{std}} = \int_{A^\pm} \sigma_{\mathrm{std}}$ for all $s \in [0, 1]$. By Banyaga's theorem, Theorem 3.4, applied to the isotopy $t \mapsto \phi^{-1} \circ \tilde{\psi}_{1-t}$ and the set $X := [a, a+\varepsilon] \times S^1 \cup [b-\varepsilon, b] \times S^1 \cup A_\varepsilon$ for some possibly smaller $\varepsilon > 0$, there exists a symplectic isotopy $\phi_s : A \to A$, $s \in [0, 1]$, with $\phi_1 = \phi$ and $\phi_s|_X = \tilde{\psi}_s|_X$. In particular, $\phi_s \in \mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ preserves E and $\phi_0|_{A_\varepsilon} = \mathrm{id}$. After renaming ϕ_0 back to ϕ, we may hence assume that $\phi = \mathrm{id}$ on an annulus A_ε around E.

Step 3. Since $\phi|_{A_\varepsilon} = \mathrm{id}$, it restricts to maps $\phi|_{A^\pm} \in \mathrm{Symp}(A^\pm, \partial A^\pm, \sigma_{\mathrm{std}})$. By Proposition A.4, $\phi|_{A^\pm}$ can be connected in $\mathrm{Symp}(A^\pm, \partial A^\pm, \sigma_{\mathrm{std}})$ to a multiple $(\phi^D_\pm)^{k_\pm}$ of the Dehn twist on $A^\pm$. Since ϕ belongs to the identity component $\mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$, it follows that $k_+ = -k_-$. Hence we can simultaneously unwind the Dehn twists to connect the map ψ which equals $(\phi^D_\pm)^{k_\pm}$ on $A^\pm$ to the identity by a path ψ_t in $\mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ fixing E (but not restricting to the identity on E). Thus $\phi_t := \phi \circ \psi_t^{-1}$ is a path in $\mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ with $\phi_t(E) = E$ connecting ϕ to ϕ_1 such that $\phi_1|_{A^\pm}$ belongs to the identity component $\mathrm{Symp}_0(A^\pm, \partial A^\pm, \sigma_{\mathrm{std}})$. Again, we rename ϕ_1 back to ϕ.

Step 4. Finally, we apply Proposition A.4 on $A^\pm$ to connect $\phi|_{A^\pm}$ to the identity by a path $\phi_t^\pm$ in $\mathrm{Symp}_0(A^\pm, \partial A^\pm)$. The maps $\phi_t^\pm$ fit together to a path $\phi_t \in \mathrm{Symp}_0(A, \partial A, \sigma_{\mathrm{std}})$ fixing E that connects ϕ to the identity. □

Acknowledgement

We thank Y. Eliashberg for many fruitful discussions and his continued interest in this work.

References

[Albers and Frauenfelder 2008] P. Albers and U. Frauenfelder, "A nondisplaceable Lagrangian torus in T^*S^2", *Comm. Pure Appl. Math.* **61**:8 (2008), 1046–1051. MR Zbl

[Banyaga 1978] A. Banyaga, "Sur la structure du groupe des difféomorphismes qui préservent une forme symplectique", *Comment. Math. Helv.* **53**:2 (1978), 174–227. MR Zbl

[Biran and Cornea 2009] P. Biran and O. Cornea, "Rigidity and uniruling for Lagrangian submanifolds", *Geom. Topol.* **13**:5 (2009), 2881–2989. MR Zbl

[Borman, Li, and Wu 2014] M. S. Borman, T.-J. Li, and W. Wu, "Spherical Lagrangians via ball packings and symplectic cutting", *Selecta Math.* (*N.S.*) **20**:1 (2014), 261–283. MR Zbl

[Chekanov 1996] Y. V. Chekanov, "Lagrangian tori in a symplectic vector space and global symplectomorphisms", *Math. Z.* **223**:4 (1996), 547–559. MR Zbl

[Chekanov and Schlenk 2010] Y. Chekanov and F. Schlenk, "Notes on monotone Lagrangian twist tori", *Electron. Res. Announc. Math. Sci.* **17** (2010), 104–121. MR Zbl

[Dimitroglou Rizell, Goodman and Ivrii 2016] G. Dimitroglou Rizell, E. Goodman, and A. Ivrii, "Lagrangian isotopy of tori in $S^2 \times S^2$ and $\mathbb{C}P^2$", *Geom. Funct. Anal.* **26**:5 (2016), 1297–1358. MR Zbl

[Eliashberg and Polterovich 1996] Y. Eliashberg and L. Polterovich, "Local Lagrangian 2-knots are trivial", *Ann. of Math.* (2) **144**:1 (1996), 61–76. MR Zbl

[Eliashberg and Polterovich 1997] Y. Eliashberg and L. Polterovich, "The problem of Lagrangian knots in four-manifolds", pp. 313–327 in *Geometric topology* (Athens, GA, 1993), edited by W. H. Kazez, AMS/IP Stud. Adv. Math. **2.1**, Amer. Math. Soc., Providence, RI, 1997. MR Zbl

[Entov and Polterovich 2009] M. Entov and L. Polterovich, "Rigid subsets of symplectic manifolds", *Compos. Math.* **145**:3 (2009), 773–826. MR Zbl

[Evans 2010] J. D. Evans, "Lagrangian spheres in del Pezzo surfaces", *J. Topol.* **3**:1 (2010), 181–227. MR Zbl

[Fukaya, Oh, Ohta, and Ono 2009] K. Fukaya, Y.-G. Oh, H. Ohta, and K. Ono, *Lagrangian intersection Floer theory: anomaly and obstruction, II*, AMS/IP Stud. Adv. Math. **46.2**, Amer. Math. Soc., Providence, RI, 2009. MR Zbl

[Gadbled 2013] A. Gadbled, "On exotic monotone Lagrangian tori in $\mathbb{CP}^2$ and $\mathbb{S}^2\times\mathbb{S}^2$", *J. Symplectic Geom.* **11**:3 (2013), 343–361. MR Zbl

[Gromov 1985] M. Gromov, "Pseudo holomorphic curves in symplectic manifolds", *Invent. Math.* **82**:2 (1985), 307–347. MR Zbl

[Hind 2004] R. Hind, "Lagrangian spheres in $S^2 \times S^2$", *Geom. Funct. Anal.* **14**:2 (2004), 303–318. MR Zbl

[Hind 2012] R. Hind, "Lagrangian unknottedness in Stein surfaces", *Asian J. Math.* **16**:1 (2012), 1–36. MR Zbl

[Ivrii 2003] A. Ivrii, *Lagrangian unknottedness of tori in certain symplectic 4-manifolds*, Ph.D. thesis, Stanford University, 2003, available at https://search.proquest.com/docview/305300870.

[Lalonde and McDuff 1996] F. Lalonde and D. McDuff, "The classification of ruled symplectic 4-manifolds", *Math. Res. Lett.* **3**:6 (1996), 769–778. MR Zbl

[Li and Wu 2012] T.-J. Li and W. Wu, "Lagrangian spheres, symplectic surfaces and the symplectic mapping class group", *Geom. Topol.* **16**:2 (2012), 1121–1169. MR Zbl

[McDuff and Salamon 1995] D. McDuff and D. Salamon, *Introduction to symplectic topology*, Oxford Univ. Press, 1995. MR Zbl

[Oakley and Usher 2016] J. Oakley and M. Usher, "On certain Lagrangian submanifolds of $S^2 \times S^2$ and $\mathbb{CP}^n$", *Algebr. Geom. Topol.* **16**:1 (2016), 149–209. MR Zbl

[Schwingenheuer 2010] M. Schwingenheuer, *Hamiltonian unknottedness of certain monotone Lagrangian tori in $S2 \times S2$*, Ph.D. thesis, Ludwig-Maximilians-Universität München, 2010, available at https://search.proquest.com/docview/1964146694. Zbl

[Smale 1959] S. Smale, "Diffeomorphisms of the 2-sphere", *Proc. Amer. Math. Soc.* **10** (1959), 621–626. MR Zbl

[Steenrod 1951] N. Steenrod, *The topology of fibre bundles*, Princeton Math. Series **14**, Princeton Univ. Press, 1951. MR Zbl

[Vianna 2014] R. Vianna, "On exotic Lagrangian tori in $\mathbb{CP}^2$", *Geom. Topol.* **18**:4 (2014), 2419–2476. MR Zbl

[Vianna 2016] R. F. d. V. Vianna, "Infinitely many exotic monotone Lagrangian tori in $\mathbb{CP}^2$", *J. Topol.* **9**:2 (2016), 535–551. MR Zbl

[Vianna 2017] R. Vianna, "Infinitely many monotone Lagrangian tori in del Pezzo surfaces", *Selecta Math. (N.S.)* **23**:3 (2017), 1955–1996. MR Zbl

Received May 14, 2018.

KAI CIELIEBAK
UNIVERSITÄT AUGSBURG
AUGSBURG
GERMANY
kai.cieliebak@math.uni-augsburg.de

MARTIN SCHWINGENHEUER
BOS ROSENHEIM
ASCHAU
GERMANY
schwingenheuermartin@googlemail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 2, 2019
dx.doi.org/10.2140/pjm.2019.299.469

CHORD SHORTENING FLOW AND A THEOREM OF LUSTERNIK AND SCHNIRELMANN

MARTIN MAN-CHUN LI

We introduce a new geometric flow, called the chord shortening flow, which is the negative gradient flow for the length functional on the space of chords with end points lying on a fixed submanifold in Euclidean space. As an application, we give a simplified proof of a classical theorem of Lusternik and Schnirelmann (and a generalization by Riede and Hayashi) on the existence of multiple orthogonal geodesic chords. For a compact convex planar domain, we show that any convex chord not orthogonal to the boundary would shrink to a point in finite time under the flow.

1. Introduction

The existence of closed geodesics in a Riemannian manifold is one of the most fundamental questions in geometry that has been studied extensively since the time of Poincaré [1905]. The critical point theories developed by Morse and Lusternik–Schnirelmann have played an essential role in this problem in the early 20th century (see [Klingenberg 1978] for a detailed exposition). Although there do not exist closed geodesics in $\mathbb{R}^n$, it is natural to look for geodesics contained in a bounded domain $\Omega \subset \mathbb{R}^n$ which meets $\partial\Omega$ orthogonally at its end points. These are called *orthogonal geodesic chords* (see Definition 5.1 for a precise definition). Lusternik and Schnirelmann [1934] proved the following celebrated result:

Theorem 1.1 (Lusternik–Schnirelmann). *Any bounded domain in $\mathbb{R}^n$ with smooth convex boundary contains at least n distinct orthogonal geodesic chords.*

Kuiper [1964] showed that the same conclusion holds if the boundary is only $C^{1,1}$. For our convenience, we will assume that all the submanifolds and maps are C^∞. Recall that the boundary of a domain $\Omega \subset \mathbb{R}^n$ is said to be (locally) *convex* if the second fundamental form A of $\partial\Omega$ with respect to the unit normal ν (pointing into Ω) is positive semidefinite, i.e., for all $p \in \partial\Omega$, $u \in T_p\partial\Omega$, we have

$$A(u,u) := \langle D_u u, \nu\rangle \geq 0, \tag{1-1}$$

MSC2010: primary 53C22; secondary 58E10.
Keywords: geometric flows, orthogonal geodesic chords, Lusternik–Schnirelmann theory.

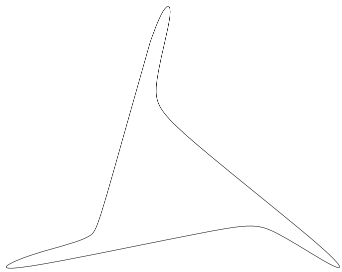

Figure 1. Bos's example of a nonconvex domain Ω in $\mathbb{R}^2$ which does not have any orthogonal geodesic chord contained in Ω.

where D is the standard flat connection in $\mathbb{R}^n$. Notice that Theorem 1.1 gives an optimal lower bound as seen in the example of the convex region bounded by the ellipsoid given by

$$\Omega := \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \sum_{i=1}^{n} \frac{x_i^2}{a_i^2} \leq 1 \right\},$$

where $a_1, \dots, a_n$ are distinct positive real numbers.

Bos [1963] generalized Lusternik–Schnirelmann's result to the setting of Riemannian (or even Finsler) manifolds.

Theorem 1.2 (Bos). *A compact Riemannian manifold (M^n, g) which is homeomorphic to the closed unit ball in $\mathbb{R}^n$ with locally convex boundary contains at least n orthogonal geodesic chords.*

Moreover, he showed that the convexity assumption cannot be dropped even in $\mathbb{R}^2$ (see Figure 1).

Nonetheless, one can still ask for the existence of orthogonal geodesic chords by allowing them to go *outside* the domain. This problem was first introduced by Riede [1968], where he studied the variational calculus of the space Γ consisting of piecewise smooth curves in a complete Riemannian manifold (M^n, g) with end points lying on a compact submanifold $\Sigma^k \subset M$. In particular, he estimated the minimum number of critical points, which are orthogonal geodesic chords, in terms of certain topological invariant called the "cup-length" of the equivariant cohomology of Γ with respect to the $\mathbb{Z}_2$-action reversing the orientation of a curve. Hayashi [1982] computed the cup-length when Σ is a compact submanifold in $\mathbb{R}^n$ and hence proved the following result.

Theorem 1.3 (Riede–Hayashi). *Any k-dimensional compact submanifold Σ in $\mathbb{R}^n$ admits at least $k+1$ orthogonal geodesic chords.*

Note that Theorem 1.3 generalizes Theorem 1.1 by taking Σ to be the boundary of a bounded convex domain. However, we emphasize that if $\Sigma = \partial\Omega$ is the boundary of a nonconvex domain $\Omega \subset \mathbb{R}^n$, then the orthogonal geodesic chords obtained in Theorem 1.3 are not necessarily contained in Ω (recall Figure 1).

The original proofs of Theorem 1.1, 1.2, and 1.3 all used a discrete curve shortening process similar to the one introduced by Birkhoff [1917] in the study of existence of closed geodesics in Riemannian manifolds. A description of the process can be found in [Gluck and Ziller 1983] (see also a modified version in [Zhou 2016]). The curve shortening process, denoted by Ψ, takes a piecewise smooth curve $c : [0, 1] \to M$ with end points lying on Σ to a piecewise geodesic curve $\Psi(c) : [0, 1] \to M$ which meets Σ orthogonally at its end points. The most important properties of Ψ are summarized below:

(1) $\text{Length}(\Psi(c)) \leq \text{Length}(c)$ and equality holds if and only if c is an orthogonal geodesic chord, in which case $\Psi(c) = c$.

(2) $\Psi(c)$ depends continuously on c, with respect to the C^0 topology.

(3) c and $\Psi(c)$ are homotopic in M relative to Σ, i.e., there exists a continuous family $c_t : [0, 1] \to M$, $t \in [0, 1]$, with end points on Σ such that $c_0 = c$ and $c_1 = \Psi(c)$. Moreover, the family c_t depends continuously on c.

The curve shortening process Ψ involves subdividing the curves and connecting points on the curve by minimizing geodesic segments (additional care has to be taken at the end points). The construction depends on some fixed parameter (which depends on the geometry of M, Σ, and $\text{Length}(c)$). However, it can be shown that for curves with uniformly bounded length, the parameters can be chosen uniformly to make (1)–(3) above hold. In fact (1) and (3) follow easily from the constructions, but (2) requires some convexity estimates (see [Zhou 2016, Lemma 3.2]). Using (1)–(3), it is not difficult to see that the sequence $\{\Psi^i(c)\}_{i=1}^{\infty}$ either converges to a point on Σ or has a subsequence converging to an orthogonal geodesic chord. Theorem 1.1, 1.2, and 1.3 then follow from the abstract Lusternik–Schnirelmann theory applied to families of curves with end points on Σ which represent a nontrivial homology class relative to point curves on Σ. Interested readers can refer to [Gluck and Ziller 1983; Giannoni and Majer 1997] for more details (for Theorem 1.1 there is a more elementary proof — see [Kuiper 1964] for example).

In this paper, we introduce a new curve shortening process called the *chord shortening flow* (see Definition 2.3), which evolves a geodesic chord according to the "contact angle" that the chord makes with Σ at its end points. It is the negative gradient flow for the length functional on the space of chords. We study the fundamental properties including the short-time existence and uniqueness and long-time convergence of the flow when the ambient space is $\mathbb{R}^n$. Note that the flow still makes sense in certain Riemannian manifolds but for simplicity we postpone the details to another forthcoming paper. The chord shortening flow, as a negative gradient flow, clearly satisfies all the properties (1)–(3) above; hence provide the most natural curve shortening process required in the proof of Theorem 1.1 and 1.3 (but not Theorem 1.2 in its full generality).

Remark 1.4. We would like to mention that Lusternik and Schnirelmann used the same ideas to prove the *theorem of three geodesics* which asserts that any Riemannian sphere (S^2, g) contains at least three *geometrically distinct* closed embedded geodesics. Unfortunately, the original proof by Lusternik and Schnirelmann [1934] contains a serious gap and various attempts have been made to fix it (see [Taĭmanov 1992]). The fundamental issue there is *multiplicity*: that one of the geodesics obtained may just be a multiple cover of another geodesic. It is extremely technical (and many false proofs were given) to rule out this situation by modifying the method of Lusternik–Schnirelmann. Grayson [1989] gave a rigorous proof of the theorem of three geodesics by a careful analysis of the curve shortening flow on Riemannian surfaces. He proved that under the curve shortening flow, any embedded curve remains embedded and would either converge to a point in finite time or an embedded closed geodesic as time goes to infinity. As a curve which is initially embedded stays embedded throughout the flow, this prevents the multiplicity problem encountered by Lusternik–Schnirelmann's approach using a discrete curve shortening process of Birkhoff [1917]. On the other hand, the situations in Theorem 1.1 and Theorem 1.3 are simpler as multiplicity cannot occur (see [Giannoni and Majer 1997, Remark 3.2]).

We show that the convergence behavior for the chord shortening flow is similar to that for the curve shortening flow on a closed Riemannian surface [Grayson 1989]. In particular, we prove that under the chord shortening flow, any chord would either converge to a point in finite time or to an orthogonal geodesic chord as time goes to infinity. Unlike [Grayson 1989], this dichotomy holds in any dimension and codimension, in contrast with the curve shortening flow where an embedded curve may develop self-intersections or singularities after some time when codimension is greater than one [Altschuler 1991]. In the special case that $\Sigma = \partial\Omega$, where $\Omega \subset \mathbb{R}^2$ is a compact convex planar domain, we give a sufficient condition for an initial chord to converge to a point in finite time. In fact, any "convex" chord in Ω which is not an orthogonal geodesic chord would converge to a point on $\partial\Omega$ in finite time. This can be compared to the famous result of Huisken [1984] which asserts that any compact embedded convex hypersurface in $\mathbb{R}^n$ converges to a point in finite time under the mean curvature flow.

The chord shortening flow is also of independent interest from the analytic point of view. Since any chord in $\mathbb{R}^n$ is determined uniquely by its end points, we can regard the chord shortening flow as an evolution equation for the two end points lying on Σ. As a result, the flow is a *nonlocal* evolution of a pair of points on Σ as it depends on the chord joining them. In fact, the chord shortening flow can be regarded as the heat equation for the half-Laplacian (or the *Dirichlet-to-Neumann map*).

The organization of this paper is as follows. In Section 2, we introduce the chord shortening flow, give a few examples, and prove the short time existence and uniqueness of the flow. In Section 3, we derive the evolution equations for some geometric quantities under the chord shortening flow. In Section 4, we prove the long-time existence to the flow provided that it does not shrink the chord to a point in finite time. In Section 5, we prove that an initial convex chord inside a compact convex domain in $\mathbb{R}^2$ would shrink to a point in finite time under the flow, provided that the initial chord is not an orthogonal geodesic chord.

Notation. Throughout this paper, we will denote $I := [0, 1]$ with $\partial I = \{0, 1\}$. The Euclidean space $\mathbb{R}^n$ is always equipped with the standard inner product $\langle \cdot , \cdot \rangle$ and norm $|\cdot|$. For any subset $S \subset \mathbb{R}^n$, we use $d(\cdot , S)$ to denote the distance function from S.

2. Chord shortening flow

In this section, we introduce a new geometric flow called the *chord shortening flow*. This flow has some similarities with the classical curve shortening flow. The main result in this section is the short-time existence and uniqueness theorem for the chord shortening flow (Proposition 2.7). We also study some basic examples of such a flow.

Let Σ be a k-dimensional smooth submanifold[1] in $\mathbb{R}^n$. Note that Σ can be disconnected in general. For any two points $p, q \in \Sigma$, we can consider the extrinsic chord distance between them in $\mathbb{R}^n$.

Definition 2.1. The *chord distance function* $d : \Sigma \times \Sigma \to \mathbb{R}_{\geq 0}$ is defined to be

$$d(p, q) := \mathrm{dist}_{\mathbb{R}^n}(p, q) = |p - q|.$$

Since any two distinct points in $\mathbb{R}^n$ are connected by a unique line segment realizing their distance, the chord distance function d is smooth away from the diagonal $\{(p, p) \in \Sigma \times \Sigma : p \in \Sigma\}$.

Definition 2.2. For any two distinct points p, q on Σ, we will use $C_{p,q}$ to denote the unique oriented chord from p to q. The outward unit conormal, denoted by η, is the unit vector at $\partial C_{p,q}$ tangent to $C_{p,q}$ pointing out of $C_{p,q}$. Note that $\eta(p) = -\eta(q)$. (see Figure 2)

[1]In fact all the following discussions make sense for *immersed* submanifolds. However, for simplicity, we will assume that all submanifolds are *embedded*.

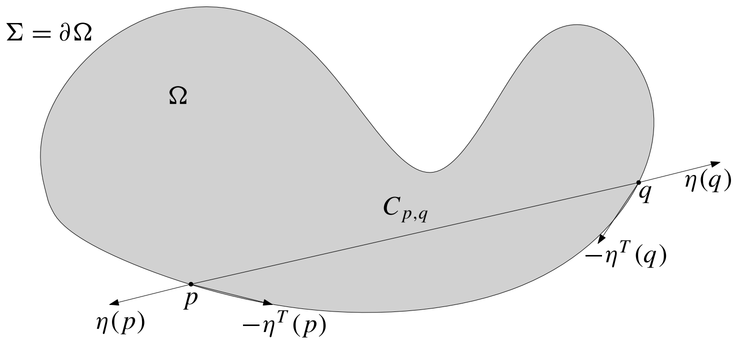

Figure 2. A chord $C_{p,q}$ joining p and q, the outward unit conormals η and their (negative) tangential components along $\Sigma = \partial\Omega$.

Let $C(t) = C_{p_t,q_t}$ be a smooth family of chords with distinct end points $p_t, q_t \in \Sigma$. If $\ell(t) = d(p_t, q_t)$ is the length of the chord $C(t)$, the first variation formula for arc length (see for example [Cheeger and Ebin 1975, (1.5)]) implies that

$$\frac{d\ell}{dt} = \left\langle \frac{dp_t}{dt}, \eta(p_t) \right\rangle + \left\langle \frac{dq_t}{dt}, \eta(q_t) \right\rangle. \tag{2-1}$$

Note that the interior integral term vanishes as $C(t)$ is a geodesic for every t. Since p_t and q_t lie on Σ for all t, both dp_t/dt and dq_t/dt are tangential to Σ. Therefore, we can express (2-1) as

$$\frac{d\ell}{dt} = \left\langle \frac{dp_t}{dt}, \eta^T(p_t) \right\rangle + \left\langle \frac{dq_t}{dt}, \eta^T(q_t) \right\rangle, \tag{2-2}$$

where $(\cdot)^T$ denotes the tangential component of a vector relative to Σ. More precisely, if $\pi_x : \mathbb{R}^n \to T_x\Sigma$ is the orthogonal projection onto the tangent space $T_x\Sigma \subset \mathbb{R}^n$, then $v^T = \pi_x(v)$ for any vector $v \in T_x\mathbb{R}^n \cong \mathbb{R}^n$.

It is natural to consider the (negative) gradient flow to the chord length functional, which leads to the following definition.

Definition 2.3 (chord shortening flow). A smooth family of curves

$$C(u, t) : I \times [0, T) \to \mathbb{R}^n$$

is a solution to the *chord shortening flow* (relative to Σ) if for all $t \in [0, T)$,

(a) $p_t := C(0, t)$ and $q_t := C(1, t)$ lie on Σ,

(b) $C(t) := C(\cdot, t) : I \to \mathbb{R}^n$ is a constant speed parametrization of C_{p_t,q_t}, and

(c) $\partial C/\partial t(0, t) = -\eta^T(C(0, t))$ and $\partial C/\partial t(1, t) = -\eta^T(C(1, t))$.

Let us begin with some basic examples of the chord shortening flow as defined in Definition 2.3.

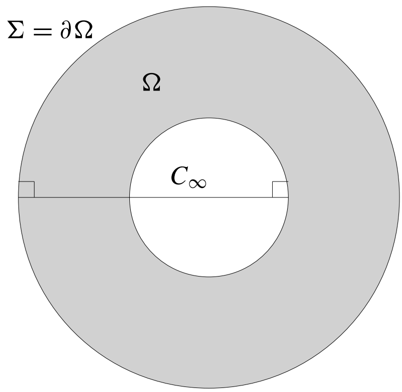

Figure 3. A limit chord C_∞ which meets $\partial\Omega$ orthogonally but not lying inside Ω.

Example 2.4. Let Σ be an affine k-dimensional subspace in $\mathbb{R}^n$. The chord shortening flow with respect to Σ will contract any initial chord $C(0) = C_{p,q}$ to a point in finite time. The end points would move towards each other with unit speed along the chord $C(0)$ until they meet at the midpoint of $C(0)$ at the "blow-up" time $T = \frac{1}{2}d(p, q)$.

Example 2.5. Let Σ be a union of two disjoint circles in $\mathbb{R}^2$ (see Figure 3). We will see (from Theorem 4.1) that any chord joining two distinct connected components of Σ would evolve under the chord shortening flow to a limit chord C_∞ orthogonal to Σ as $t \to \infty$. The same phenomenon holds for any $\Sigma \subset \mathbb{R}^n$ which is disconnected.

Example 2.6. Let Σ be the ellipse $\{(x, y) \in \mathbb{R}^2 : x^2 + 4y^2 = 1\}$ in $\mathbb{R}^2$. By symmetry it is not difficult to see that for any initial chord passing through the origin (with the exception of the major axis), it would evolve under the chord shortening flow to the minor axis of the ellipse, which is a chord orthogonal to Σ and contained inside the region enclosed by the ellipse. See Figure 4. This example shows that the number of distinct orthogonal chords guaranteed by the Lusternik–Schnirelmann theorem is optimal. If we start with an initial chord that lies completely on one side of the major or minor axis, then the chord will shrink to a point in finite time (by Theorem 6.5).

We end this section with a fundamental result on the short-time existence and uniqueness for the chord shortening flow.

Proposition 2.7 (short-time existence and uniqueness). *For any initial chord $C_0 : I \to \mathbb{R}^n$ with $C_0(\partial I) \subset \Sigma$, there exists an $\epsilon > 0$ and a smooth solution $C(u, t) : I \times [0, \epsilon) \to \mathbb{R}^n$ to the chord shortening flow relative to Σ as in Definition 2.3 with initial condition $C(\cdot, 0) = C_0$. Moreover, the solution is unique.*

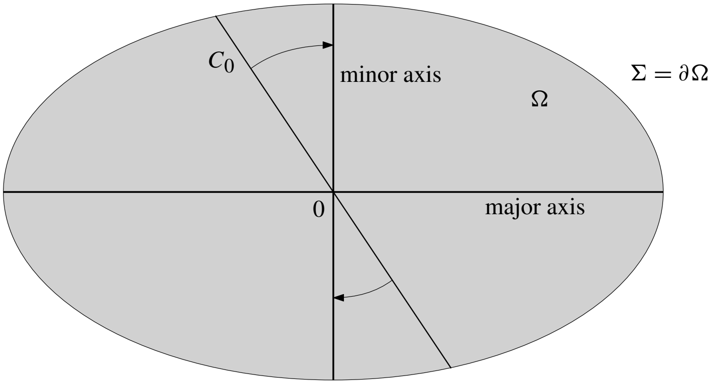

Figure 4. Any initial chord C_0 through the origin (other than the major axis) would converge under the chord shortening flow to the minor axis.

Proof. Note that for any given $p \neq q \in \Sigma$, the outward unit conormal η at the end points p, q of the chord $C_{p,q}$ is given by

$$\eta(p) = \frac{p-q}{|p-q|} = -\eta(q).$$

Therefore, Definition 2.3(c) is equivalent to the following system of nonlinear first-order ODEs:

$$\frac{dp}{dt} = -\frac{\pi_p(p-q)}{|p-q|}, \qquad \frac{dq}{dt} = -\frac{\pi_q(q-p)}{|q-p|}, \tag{2-3}$$

where $\pi_x : \mathbb{R}^n \to \mathbb{R}^n$ is the orthogonal projection onto $T_x\Sigma$ (which depends smoothly on x). Note that the right-hand side of (2-3) is a Lipschitz function in p and q as long as $|p-q|$ is bounded away from 0. Therefore, the existence and uniqueness of the initial value problem follows from the fundamental local existence and uniqueness theorem for first-order ODE systems (see, for example, [Taylor 1996, Theorem 2.1]). Hence, (2-3) is uniquely solvable on some interval $t \in [0, \epsilon)$ for any initial data $p(0) = p_0$ and $q(0) = q_0$ such that $p_0 \neq q_0 \in \Sigma$. Finally we get a solution $C(u, t) : I \times [0, \epsilon) \to \mathbb{R}^n$ to the chord shortening flow by defining $C(\cdot, t) : I \to \mathbb{R}^n$ to be the constant speed parametrization of the chord C_{p_t,q_t}. ☐

3. Evolution equations

In this section, we derive the evolution of some geometric quantities under the chord shortening flow relative to any k-dimensional submanifold Σ in $\mathbb{R}^n$.

Definition 3.1. Let $C : I = [0, 1] \to \mathbb{R}^n$ be a chord joining p to q. For any (vector-valued) function $f : \partial I = \{0, 1\} \to \mathbb{R}^m$, we define the L^2-*norm* $\|f\|_{L^2}$ and the *sum*

$\bar{f}$ of f to be

$$\|f\|_{L^2} := (|f(0)|^2 + |f(1)|^2)^{1/2} \quad \text{and} \quad \bar{f} := f(0) + f(1). \tag{3-1}$$

Also, we define the $\frac{1}{2}$-Laplacian of f relative to the chord C to be the vector-valued function $\Delta^{1/2} f : \partial I = \{0, 1\} \to \mathbb{R}^m$ defined by

$$(\Delta^{1/2} f)(0) = \frac{f(0) - f(1)}{\ell} = -(\Delta^{1/2} f)(1), \tag{3-2}$$

where $\ell = |p - q|$ is the length of the chord C.

Lemma 3.2. *Given any $f : \partial I \to \mathbb{R}^m$, we have $\overline{\Delta^{1/2} f} = 0$ and $\overline{\langle f, \Delta^{1/2} f\rangle} = \frac{\ell}{2}\|\Delta^{1/2} f\|^2_{L^2} \leq \frac{2}{\ell}\|f\|^2_{L^2}$.*

Proof. It follows directly from (3-1) and (3-2). □

Definition 3.3. Let $C = C_{p,q} : I \to \mathbb{R}^n$ be a chord joining two distinct points p, q on Σ. We define the tangential outward conormal $\eta^T : \partial I = \{0, 1\} \to \mathbb{R}^n$ to be the tangential component (relative to Σ) of the outward unit conormal of C, i.e., (recall (2-2) and Definition 2.2)

$$\eta^T(u) = \pi_{C(u)}\eta \quad \text{for } u = 0, 1. \tag{3-3}$$

Lemma 3.4 (evolution of chord length). *Suppose $C(u, t) : I \times [0, T) \to \mathbb{R}^n$ is a solution to the chord shortening flow relative to Σ as in Definition 2.3. If we denote the length of the chord $C(t)$ at time t by*

$$\ell(t) := d(C(0, t), C(1, t)),$$

then ℓ is a nonincreasing function of t and (recall (3-1) and (3-3))

$$\frac{d\ell}{dt} = -\|\eta^T\|^2_{L^2} \leq 0. \tag{3-4}$$

Proof. It follows directly from the first variation formula (2-2). □

Theorem 3.5. *Suppose $C(u, t) : I \times [0, T) \to \mathbb{R}^n$ is a solution to the chord shortening flow relative to Σ as in Definition 2.3. Then the tangential outward conormal η^T of the chord $C(t)$ satisfies the following evolution equation:*

$$\begin{aligned}\frac{\partial}{\partial t}\eta^T = -\Delta^{1/2}\eta^T + \tfrac{1}{\ell}\|\eta^T\|^2_{L^2}\eta^T \\ - \sum_{i=1}^{k}\langle A(\eta^T, e_i), \eta^N\rangle e_i - \tfrac{1}{\ell}(\overline{\eta^T} - \eta^T)^N - A(\eta^T, \eta^T),\end{aligned} \tag{3-5}$$

where $\{e_i\}_{i=1}^k$ is an orthonormal basis of $T\Sigma$ at the end points of $C(t)$. Here, $(\cdot)^N$ denotes the normal component of a vector relative to Σ and $A : T\Sigma \times T\Sigma \to N\Sigma$ is the second fundamental form of Σ defined by $A(u, v) := (D_u v)^N$.

Proof. Let $C(u,t) : I \times [0,T) \to \mathbb{R}^n$ be a solution to the chord shortening flow relative to Σ. Since $C(t) = C(\cdot\,,t) : I \to \mathbb{R}^n$ is a family of chords which are parametrized proportional to arc length, $\partial/\partial t$ is a Jacobi field (not necessarily normal) along each chord which can be explicitly expressed as

$$\frac{\partial}{\partial t} = -(1-u)\,\eta^T(0) - u\,\eta^T(1),$$

where η is the outward unit conormal for $C(t)$. Since $[\partial/\partial u, \partial/\partial t] = 0$, we have

$$D_{\partial/\partial t}\frac{\partial}{\partial u} = D_{\partial/\partial u}\frac{\partial}{\partial t} = \eta^T(0) - \eta^T(1). \tag{3-6}$$

Moreover, as $C(t)$ is parametrized with constant speed, we have $\|\partial/\partial u\| = \ell$, thus

$$-\eta(0) = \frac{1}{\ell}\frac{\partial}{\partial u}\Big|_{u=0} \quad \text{and} \quad \eta(1) = \frac{1}{\ell}\frac{\partial}{\partial u}\Big|_{u=1}.$$

Fix $u = 0$. Let $p = C(0,t) \in \Sigma$ and $\{e_1, \dots, e_k\}$ be an orthonormal basis of $T_p\Sigma$ such that $(D_{e_i}e_j(p))^T = 0$ for $i, j = 1, \dots, k$. Therefore, we have

$$D_{\partial/\partial t}e_i = -A(\eta^T, e_i). \tag{3-7}$$

Using Lemma 3.4, (3-6), and (3-7), we have

$$\begin{aligned}
\frac{\partial \eta^T}{\partial t} &= \frac{\partial}{\partial t}\left(-\tfrac{1}{\ell}\right)\sum_{i=1}^k \left\langle \frac{\partial}{\partial u}, e_i\right\rangle e_i - \tfrac{1}{\ell}\sum_{i=1}^k \frac{\partial}{\partial t}\left(\left\langle \frac{\partial}{\partial u}, e_i\right\rangle e_i\right) \\
&= \tfrac{1}{\ell}\|\eta^T\|_{L^2}^2\eta^T \\
&\qquad - \tfrac{1}{\ell}\sum_{i=1}^k\left(\left\langle D_{\partial/\partial u}\frac{\partial}{\partial t}, e_i\right\rangle e_i + \left\langle \frac{\partial}{\partial u}, D_{\partial/\partial t}e_i\right\rangle e_i + \left\langle \frac{\partial}{\partial u}, e_i\right\rangle D_{\partial/\partial t}e_i\right) \\
&= \tfrac{1}{\ell}\|\eta^T\|_{L^2}^2\eta^T - \frac{\eta^T}{\ell} - A(\eta^T, \eta^T) \\
&\qquad\qquad - \tfrac{1}{\ell}\sum_{i=1}^k(\langle -\eta^T(1), e_i\rangle e_i + \ell\langle \eta^N, A(\eta^T, e_i)\rangle e_i) \\
&= -\Delta^{1/2}\eta^T + \tfrac{1}{\ell}\|\eta^T\|_{L^2}^2\eta^T \\
&\qquad\qquad - \sum_{i=1}^k\langle A(\eta^T, e_i), \eta^N\rangle e_i - \tfrac{1}{\ell}(\overline{\eta^T} - \eta^T)^N - A(\eta^T, \eta^T).
\end{aligned}$$

A similar calculation yields (3-5) at $u = 1$. This proves the proposition. □

Remark 3.6. When Σ is an embedded planar curve (i.e., $k = 1 = n - 1$), one can give a simpler formula of (3-5) since (after introducing an orientation of the curve Σ) η^T is completely described by the "boundary angle" Θ between η and Γ (see Definition 6.2). As a result, (3-5) reduces to the evolution of Θ, which is a scalar quantity instead of a vector quantity η^T as in (3-5) (see Proposition 6.6).

Corollary 3.7. *Under the same assumptions as in Theorem 3.5, we have*

$$\frac{1}{2}\frac{d}{dt}\|\eta^T\|^2_{L^2} = -\tfrac{\ell}{2}\|\Delta^{1/2}\eta^T\|^2_{L^2} + \tfrac{1}{\ell}\|\eta^T\|^4_{L^2} - \overline{\langle A(\eta^T,\eta^T),\eta\rangle}. \tag{3-8}$$

Proof. Using (3-5) and Lemma 3.2, noting that the last two terms of (3-5) are normal to Σ, we have

$$\frac{1}{2}\frac{d}{dt}\|\eta^T\|^2_{L^2} = \overline{\left\langle \eta^T, \frac{\partial \eta^T}{\partial t}\right\rangle} = -\tfrac{\ell}{2}\|\Delta^{1/2}\eta^T\|^2_{L^2} + \tfrac{1}{\ell}\|\eta^T\|^4_{L^2} - \overline{\langle A(\eta^T,\eta^T),\eta^N\rangle}. \quad \square$$

Example 3.8. In the case of Example 2.4, we have $\eta^T(0) = -\eta^T(1)$ equals a constant unit vector independent of t and hence both sides are identically zero in (3-5) and (3-8).

Example 3.9. Consider the vertical strip $\Omega := \{(x, y) \in \mathbb{R}^2 : 0 \le x \le 1\}$ with boundary $\Sigma = \partial\Omega$ consisting of two parallel vertical lines. Let $p_0 = \left(0, -\frac{1}{2}h\right)$ and $q_0 = \left(1, \frac{1}{2}h\right)$ for some $h > 0$. It is easy to check that the solution to the chord shortening flow with initial chord C_{p_0,q_0} is given by $p_t = \left(0, -\frac{1}{2}h(t)\right)$, $q_t = \left(1, \frac{1}{2}h(t)\right)$, where $h(t)$ is the unique solution to the ODE

$$h'(t) = -\frac{2h(t)}{\sqrt{1+h^2(t)}}$$

with initial condition $h(0) = h$. From this we can see that the solution $h(t)$ exists for all $t \ge 0$. Moreover, $-h'(t) \le 2h(t)$ implies $h(t) \le he^{-2t}$ and thus $h(t) \to 0$ exponentially as $t \to +\infty$. Therefore, the chord converges to a chord meeting $\partial\Omega$ orthogonally. In this case, we have

$$-\eta^T(0) = \frac{1}{\sqrt{1+h^2(t)}}(0, h(t)) = \eta^T(1),$$

which satisfies the evolution equation (3-5) and $\eta^T \to 0$ as $t \to +\infty$. See Figure 5.

4. Long-time existence

In this section, we prove our main convergence result which says that the only two possible convergence scenarios are given in Example 3.8 and 3.9. One should compare this convergence result with a similar result of Grayson [1989, Theorem 0.1] for curve shortening flow on surfaces. For simplicity, we assume that Σ is compact. However, the same result holds for noncompact Σ which satisfies some convexity condition at infinity as in [Grayson 1989].

Theorem 4.1 (long-time convergence). *Let $\Sigma \subset \mathbb{R}^n$ be a compact k-dimensional smooth submanifold without boundary. Suppose $C(0) : I \to \mathbb{R}^n$ is a chord with distinct end points on Σ. Then there exists a maximally defined smooth family of chords $C(t) : I \to \mathbb{R}^n$ for $t \in [0, T)$ with distinct end points on Σ, and $C(t) = C(\cdot, t)$*

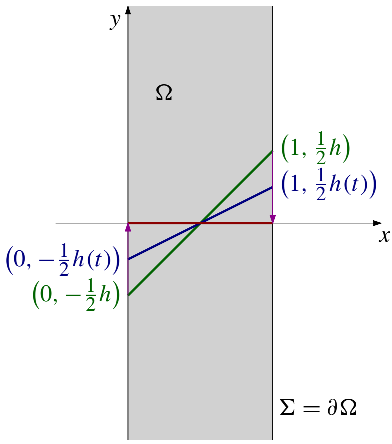

Figure 5. A chord converging to a limit chord orthogonal to $\partial\Omega$.

where $C(u, t) : I \times [0, T) \to \mathbb{R}^n$ is the unique solution to the chord shortening flow (relative to Σ) as in Definition 2.3.

Moreover, if $T < +\infty$, then $C(t)$ converges to a point on Σ as $t \to T$. If T is infinite, then $C(t)$ converges to an orthogonal geodesic chord with end points on Σ as $t \to \infty$.

By the short time existence and uniqueness theorem (Proposition 2.7), the chord shortening flow continues to exist and is unique as long as $\ell > 0$. Therefore, $C(t)$ is uniquely defined for $t \in [0, T)$ where either $T < +\infty$ or $T = +\infty$.

Lemma 4.2. *Let $C(t)$, $t \in [0, T)$, be a maximally defined chord shortening flow. Then one of the following holds:*

(a) *$T < +\infty$ and $C(t)$ shrinks to a point on Σ as $t \to T$.*

(b) *$T = +\infty$ and $\ell(t) \to \ell_\infty > 0$ as $t \to +\infty$.*

Proof. As $\ell(t)$ is a nonincreasing function of t by Lemma 3.4, it either converges to 0 or to some positive number $\ell_\infty > 0$ as $t \to T$. By short-time existence (Proposition 2.7), it cannot converge to $\ell_\infty > 0$ in finite time. So when $T < +\infty$, $C(t)$ must converge to a point on Σ by compactness of Σ. It remains to show that $\ell(t)$ cannot converge to 0 if $T = +\infty$. We will prove this by a contradiction argument. Suppose, on the contrary, that $T = +\infty$ and $\ell(t) \to 0$ as $t \to +\infty$. Since Σ is compact, there exists some constant $\epsilon_0 > 0$ such that for any two points $p, q \in \Sigma$ with $d(p, q) < \epsilon_0$, the chord $C_{p,q}$ joining them has $\|\eta^T\|^2_{L^2}$ be bounded from below by a universal positive constant (see, for example, [Colding and Minicozzi 2011, Lemma 5.2]). By Lemma 3.4, $\ell(t)$ must decrease to zero in finite time, which is a contradiction. $\square$

Next, we claim that if the flow exists for all time (i.e., $T = +\infty$), then it must converge to an orthogonal geodesic chord to Σ as $t \to \infty$. Since $|\eta^T| \le \|\eta^T\|_{L^2}$, it suffices to prove the following lemma. Theorem 4.1 clearly follows from Lemma 4.2 and 4.3.

Lemma 4.3. *Under the same assumption as Lemma 4.2 and suppose $T = +\infty$, then $\|\eta^T\|_{L^2} \to 0$ as $t \to +\infty$.*

Proof. Write $\ell_t = \ell(t)$ for $t \in [0, +\infty]$. By Lemma 3.4 and 4.2, we have

$$\ell_0 \ge \ell_t \ge \ell_\infty > 0 \quad \text{for all } t. \tag{4-1}$$

Moreover, integrating the inequality in Lemma 3.4 we obtain

$$\ell_t - \ell_\infty = \int_t^\infty \|\eta^T\|_{L^2}^2 \, d\tau \ge 0.$$

As a result,

$$\int_t^\infty \|\eta^T\|_{L^2}^2 \, d\tau \to 0 \quad \text{as } t \to \infty.$$

In other words, $\|\eta^T\|_{L^2}^2$ is L^2-integrable on $t \in [0, +\infty)$. If we can control the time derivative of $\|\eta^T\|_{L^2}^2$, then we can conclude that $\|\eta^T\|_{L^2} \to 0$ as $t \to \infty$. Using (3-8), (4-1), Lemma 3.2, and $\|\eta^T\|_{L^2}^2 \le 2$, we have the following differential inequality

$$\frac{1}{2}\frac{d}{dt}\|\eta^T\|_{L^2}^2 \le \left(C + \frac{4}{\ell_\infty}\right)\|\eta^T\|_{L^2}^2, \tag{4-2}$$

where $C = \sup_\Sigma |A| > 0$ is a constant depending only on the compact submanifold Σ. We now combine (4-2) with the fact that $\int_t^\infty \|\eta^T\|_{L^2}^2 \, d\tau \to 0$ as $t \to \infty$ to conclude that $\|\eta^T\|_{L^2}^2 \to 0$ as $t \to \infty$.

To simplify notation, let $f(t) := \|\eta^T\|_{L^2}^2$ and $c := C + 4/\ell_\infty$. Then $\int_t^\infty f \to 0$ as $t \to \infty$ and $f' \le cf$. We argue that $f(t) \to 0$ as $t \to \infty$. Suppose not, then there exists an increasing sequence $t_n \to +\infty$ such that

$$f(t_n) > \frac{1}{n} \quad \text{and} \quad \int_{t_n/2}^\infty f \le \frac{1}{n^3}. \tag{4-3}$$

We claim that there exists $t_n^* \in (t_n - 1/n, t_n + 1/n)$ such that $f(t_n^*) \le 1/n^2$. If not, then by (4-3),

$$\frac{2}{n^3} \le \int_{t_n-1/n}^{t_n+1/n} f \le \int_{t_n/2}^\infty f \le \frac{1}{n^3},$$

which is a contradiction. Using that $f' \le cf$, we see that by (4-3),

$$\frac{1}{n} < f(t_n) \le f(t_n^*)e^{c/n} \le \frac{1}{n^2}e^{c/n}.$$

As a result, there is a contradiction when n is sufficiently large. We have thus proved that $f(t) \to 0$ as $t \to \infty$, as claimed. $\square$

5. Existence of orthogonal geodesic chords

In this section, we give several geometric applications of the chord shortening flow concerning the existence of multiple orthogonal geodesic chords. We first give the precise definition.

Definition 5.1. Let $\Sigma \subset \mathbb{R}^n$ be a smooth k-dimensional submanifold without boundary. An *orthogonal geodesic chord for* Σ is a geodesic $c : [0, 1] \to \mathbb{R}^n$ with endpoint $c(0)$ and $c(1)$ lying on Σ such that $c'(0)$ and $c'(1)$ are normal to Σ at $c(0)$ and $c(1)$ respectively.

An *orthogonal geodesic chord* is also called a *free boundary geodesic* [Zhou 2016] or a *double normal* [Kuiper 1964] in the literature. Note that in the case where $\Sigma \subset \mathbb{R}^n$ is an embedded hypersurface which bounds a domain Ω in $\mathbb{R}^n$, our definition of orthogonal geodesic chords does not require the chord be contained inside $\overline{\Omega}$ as in [Giambò et al. 2014]. The problem of the existence of multiple orthogonal geodesic chords for submanifolds in $\mathbb{R}^n$ was first treated by Riede [1968] as follows. Let $\mathcal{C}_\Sigma$ be the space of all piecewise smooth curves $c : [0, 1] \to \mathbb{R}^n$ with end points on Σ, endowed with the compact open topology. There exists a $\mathbb{Z}_2$-action on $\mathcal{C}_\Sigma$ by $c(t) \mapsto c(1-t)$ whose fixed point set is denoted by Δ'. Denote by $H_*^{\mathbb{Z}_2}(\mathcal{C}_\Sigma, \Delta')$ and $H^*_{\mathbb{Z}_2}(\mathcal{C}_\Sigma)$ the $\mathbb{Z}_2$-equivariant homology groups (relative to Δ') and cohomology groups respectively. All the (co)homology groups in this paper are considered with a $\mathbb{Z}_2$ coefficient. The following result is taken from [Riede 1968, Satz (5.5)].

Lemma 5.2. *If there exists* $\beta \in H_*^{\mathbb{Z}_2}(\mathcal{C}_\Sigma, \Delta')$ *and* $\alpha_1, \dots, \alpha_s \in H^*_{\mathbb{Z}_2}(\mathcal{C}_\Sigma)$ (*not necessarily distinct*) *with* $\deg \alpha_i > 0$ *for all* i *such that* $(\alpha_1 \cup \cdots \cup \alpha_s) \cap \beta \neq 0$, *then there exists at least* $s+1$ *orthogonal geodesic chords for* Σ.

The largest possible integer s such that the hypothesis holds in Lemma 5.2 is often called the *cup length* of $\mathcal{C}_\Sigma$. The proof of Lemma 5.2 in [Riede 1968] involves a discrete curve shortening process Ψ on $\mathcal{C}_\Sigma$ which satisfies properties (1)–(3) as described in the introduction. As we have pointed out, it is no easy task to verify the continuity of Ψ with respect to the initial curve. For our problem at hand, one can in fact reduce it to a much simpler situation as follows. Since any curve $c \in \mathcal{C}_\Sigma$ can be continuously deformed into the unique chord joining the same end points, we can restrict $\mathcal{C}_\Sigma$ to the subset $\mathcal{C}^0_\Sigma$ consisting of all the chords with end points on Σ. The chord shortening flow is then a curve shortening process on $\mathcal{C}^0_\Sigma$ which satisfies all the required properties. Moreover, the space of chords $\mathcal{C}^0_\Sigma$ can also be described as the orbit space of $\Sigma \times \Sigma$ under the $\mathbb{Z}_2$-action $(p, q) \mapsto (q, p)$. As before, if we let $\Delta \subset \Sigma \times \Sigma$ be the fixed point set of the $\mathbb{Z}_2$-action, and $H_*^{\mathbb{Z}_2}(\Sigma \times \Sigma, \Delta)$, $H^*_{\mathbb{Z}_2}(\Sigma \times \Sigma)$ be the $\mathbb{Z}_2$-equivariant homology and cohomology respectively, we have by naturality

$$H_*^{\mathbb{Z}_2}(\Sigma \times \Sigma, \Delta) \cong H_*^{\mathbb{Z}_2}(\mathcal{C}_\Sigma, \Delta') \quad \text{and} \quad H^*_{\mathbb{Z}_2}(\Sigma \times \Sigma) \cong H^*_{\mathbb{Z}_2}(\mathcal{C}_\Sigma). \tag{5-1}$$

Hayashi [1982] studied the equivariant (co)homology of $\Sigma \times \Sigma$ and obtained the following result.

Lemma 5.3 [Hayashi 1982, Theorem 2]. *There exists* $\beta \in H_{2k}^{\mathbb{Z}_2}(\Sigma \times \Sigma, \Delta)$ *and* $\alpha \in H^1_{\mathbb{Z}_2}(\Sigma \times \Sigma)$ *such that* $\alpha^k \cap \beta \neq 0$ *in* $H_k^{\mathbb{Z}_2}(\Sigma \times \Sigma, \Delta)$, *where* $\alpha^k = \alpha \cup \cdots \cup \alpha$ *is the k-th power of cup products of* α *and* $k = \dim \Sigma$.

We then obtained Theorem 1.3, which clearly implies Lusternik–Schnirelmann's theorem (Theorem 1.1) as a special case since the orthogonal geodesic chords must be contained inside the convex domain by convexity of the domain $\Omega \subset \mathbb{R}^n$, by combining Lemma 5.3, 5.2, and (5-1). For the sake of completeness, we provide below some details of the min-max arguments.

Denote by $\Lambda = \mathcal{C}^0_\Sigma$ the space of chords with end points on Σ and for each $\ell \in [0, +\infty)$,

$$\Lambda^\ell := \{c \in \Lambda : \text{Length}(c) \le \ell\}.$$

Let α and β be given as in Lemma 5.3 and under the identification (5-1) one defines the homology classes $h_j \in H_j(\Lambda, \Lambda^0)$ where

$$h_j := \alpha^{k-j} \cap \beta \quad \text{for } j = 0, \ldots, k.$$

For each of the homology class h_j (which is nonzero by Lemma 5.3) above, one can define κ_j to be the infimum over all cycles representing h_j of the length of the longest chord in the cycle. Since the h_j are pairwise *subordinate* (see, e.g., [Klingenberg 1978] for a precise definition) to each other, we have the inequalities

$$\kappa_0 \le \kappa_1 \le \kappa_2 \le \cdots \le \kappa_k.$$

By similar arguments in the proof of Lemma 4.2, there exists $\epsilon_0 > 0$ such that Λ^0 is a deformation retract of Λ^{ϵ_0}. Since $h_1 \neq 0$, we must have $\kappa_0 \ge \epsilon_0 > 0$.

Next, we claim that each κ_j arises as the length of some orthogonal geodesic chord. Let $\Psi_t : \Lambda \to \Lambda$, $t \in [0, +\infty)$, be the chord shortening flow and for each ℓ, define the critical set

$$K_\ell := \{c \in \Lambda : c \text{ is an orthogonal geodesic chord of length } \ell\}.$$

Our main theorem (Theorem 4.1) implies the following "deformation lemma": Let $U \subset \Lambda$ be any open neighborhood of K_ℓ; there exists some small $\epsilon > 0$ such that for any $c \in \Lambda^{\ell+\epsilon}$, one can find a neighborhood U_c of c and $t_c \ge 0$ such that $\Psi_t(U_c) \subset U \cup \Lambda^{\ell-\epsilon}$ for each $t \ge t_c$. Standard arguments as in [Klingenberg 1978] then imply that K_{ℓ_j} is not empty, hence proving our claim.

Finally, it remains to show that if $\kappa_{j-1} = \kappa_j = \kappa$ for some $j = 1, \ldots, k$, then there exist infinitely many distinct orthogonal geodesic chords with length κ. We argue by contradiction. Suppose there are only finitely many orthogonal geodesic chords with length κ, i.e., $K_\kappa = \{c_1, \ldots, c_m\}$. Choosing pairwise disjoint contractible neighborhoods $U_1, \ldots, U_m$ in $\Lambda \setminus \Lambda^0$ for $c_1, \ldots, c_m$ respectively, we have $H^1(U_1 \cup \cdots \cup U_m) = 0$. Fix $\epsilon > 0$ for the neighborhood W of K_κ as in the deformation lemma above. There exists a cycle representing h_j such that all the chords in the cycle have length at most $\kappa + \epsilon$. By the deformation lemma, we can apply the chord shortening flow to every chord in the cycle for some fixed positive time so that every chord lies in $W \cup \Lambda^{\kappa-\epsilon}$. This gives a contradiction as in [Klingenberg 1978, Theorem 2.1.10] and thus our proof is completed.

6. Shrinking convex chord to a point

In this section, we study the evolution of chords inside a convex connected planar domain in $\mathbb{R}^2$. In particular, we prove that if an initial chord is *convex*, then it will shrink to a point in finite time under the chord shortening flow. In order to make precise the concept of *convexity*, we need to be consistent with the orientation of a curve in $\mathbb{R}^2$. For this reason, we restrict our attention to plane curves which bounds a domain in $\mathbb{R}^2$.

Definition 6.1 (boundary orientation). For any smooth domain $\Omega \subset \mathbb{R}^2$, we always orient the boundary $\partial\Omega$ as the boundary of Ω with the standard orientation inherited from $\mathbb{R}^2$. The orientation determines uniquely a global unit tangent vector field, called the *orientation field*, $\xi : \partial\Omega \to T(\partial\Omega)$ such that $\nu := J\xi$ is the inward pointing normal of $\partial\Omega$ relative to Ω. Here, $J : \mathbb{R}^2 \to \mathbb{R}^2$ is the counterclockwise rotation by $\frac{1}{2}\pi$.

Using Definition 6.1, we can define the *boundary angle* Θ which measures the contact angle between a chord C and the boundary $\partial\Omega$.

Definition 6.2 (boundary angle). For any (oriented) chord $C_{p,q}$ joining p to q with $p \neq q \in \partial\Omega$, we define the *boundary angle* $\Theta : \{p, q\} \to \mathbb{R}$ by

$$\Theta(p) := \langle \eta(p), \xi(p) \rangle \quad \text{and} \quad \Theta(q) := -\langle \eta(q), \xi(q) \rangle,$$

where ξ is the orientation field on $\partial\Omega$ as in Definition 6.1.

Definition 6.3. An oriented chord $C_{p,q}$ is *convex* if $\Theta \geq 0$ at both end points.

Remark 6.4. If we change the orientation of the chord from $C_{p,q}$ to $C_{q,p}$, the boundary angle Θ changes sign. Since the orientation field ξ is always tangent to $\partial\Omega$, we have $\Theta(p) = \Theta(q) = 0$ if and only if $C_{p,q}$ meets $\partial\Omega$ orthogonally at its end points p and q.

If we define the "unit normal" N of $\partial C_{p,q} = \{p, q\}$ inside $\partial\Omega$ by setting

$$N(p) = -\xi(p) \quad \text{and} \quad N(q) = \xi(q),$$

then a solution to the chord shortening flow (2-3) can be consider as a smooth 1-parameter family of point pairs on $\partial\Omega$ given by $\gamma : \{0, 1\} \times [0, T) \to \partial\Omega$ such that

$$\frac{\partial \gamma}{dt}(u, t) = \Theta(\gamma(u, t)) N(\gamma(u, t)), \tag{6-1}$$

where Θ is the boundary angle for the oriented chord from $\gamma(0, t)$ to $\gamma(1, t)$. Since the value of Θ at $u = 0$ depends also on the other end point $\gamma(1, t)$, this is a nonlocal function. Therefore, the chord shortening flow can be thought of as a nonlocal curve shortening flow driven by the boundary angle Θ.

We are now ready to state the main theorem of this section. The readers can compare Theorem 6.5 with the famous result of Huisken [1984] which says that any compact embedded convex hypersurface in $\mathbb{R}^n$ would contract to a point in finite time under the mean curvature flow.

Theorem 6.5. *Let $\Omega \subset \mathbb{R}^2$ be a compact connected domain with smooth convex boundary. Any convex chord which is not an orthogonal geodesic chord would converge to a point in finite time under the chord shortening flow.*

To prove Theorem 6.5 we need to establish a few propositions, which are of geometric interest. We first state the evolution of the boundary angle Θ under the chord shortening flow. Note that we always have $|\Theta| \leq 1$ by definition.

Proposition 6.6 (evolution of boundary angle). *Suppose $C(u, t) : I \times [0, T) \to \mathbb{R}^2$ is a solution to the chord shortening flow as in Definition 2.3. Then, the boundary angle $\Theta(u, t) : \{0, 1\} \times [0, T) \to \mathbb{R}$ satisfies the following equation (recall (3-1) and (3-2)):*

$$\frac{\partial}{\partial t}\Theta = -\Delta^{1/2}\Theta + \tfrac{1}{\ell}(\|\Theta\|^2_{L^2} + \ell k\langle -\eta, \nu\rangle)\Theta + \tfrac{1}{\ell}(1 + \langle \xi(p), \xi(q)\rangle)(\Theta - \overline{\Theta}), \tag{6-2}$$

where $k := \langle \nabla_\xi \xi, \nu\rangle$ is the curvature of $\partial\Omega$ with respect to ν (recall Definition 6.1), $\ell = \ell(t)$ is the length of the chord $C(\,\cdot\,, t) : I \to \mathbb{R}^2$ with outward unit conormal η.

Proof. It follows directly from Theorem 3.5 and Definition 6.2 ☐

Using (6-2), we immediately have the following evolution equations.

Corollary 6.7. *Under the same hypothesis as Proposition 6.6, we have*

$$\frac{d}{dt}\overline{\Theta} = \tfrac{1}{\ell}\big(\|\Theta\|^2_{L^2} - 1 - \langle \xi(p), \xi(q)\rangle\big)\overline{\Theta} + \overline{k\langle -\eta, \nu\rangle\Theta}, \tag{6-3}$$

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}\|\Theta\|^2_{L^2} = \tfrac{\ell}{2}\langle \xi(p), \xi(q)\rangle \|\Delta^{1/2}\Theta\|^2_{L^2} + \overline{k\langle -\eta, \nu\rangle\Theta^2} \\ + \tfrac{1}{\ell}\big(\|\Theta\|^2_{L^2} - 1 - \langle \xi(p), \xi(q)\rangle\big)\|\Theta\|^2_{L^2}. \end{aligned} \tag{6-4}$$

Proof. Both equation follows from (6-2) and Lemma 3.2. □

Our first lemma is that convexity is preserved under the chord shortening flow. From now on, we will use $C(t)$ to denote the unique solution to the chord shortening flow with initial chord $C(0)$ defined on the maximal time interval $t \in [0, T)$ (where T could be infinite).

Lemma 6.8. *Let $C(0)$ be a convex chord inside a compact domain $\Omega \subset \mathbb{R}^2$ with convex boundary $\partial\Omega$. Then, $C(t)$ remains convex for all $t \in [0, T)$.*

Proof. Let $\Theta_{\min}$ and $\Theta_{\max}$ be the minimum and maximum of Θ, both of which are Lipschitz functions of t. By (6-2), we have the following equality:

$$\frac{d}{dt}\Theta_{\min} = \tfrac{1}{\ell}\big((\|\Theta\|_{L^2}^2 - 1)\Theta_{\min} + \ell k\langle -\eta, \nu\rangle\Theta_{\min} - \langle \xi(p), \xi(q)\rangle\Theta_{\max}\big). \tag{6-5}$$

As $\partial\Omega$ is convex, we have $k \geq 0$ and $\langle -\eta, \nu\rangle \geq 0$. Moreover, if the chord is convex, then $\Theta_{\min} \geq 0$. Therefore, (6-5) implies the following differential inequality:

$$\frac{d}{dt}\Theta_{\min} \geq \tfrac{1}{\ell}\big((\|\Theta\|_{L^2}^2 - 1)\Theta_{\min} - \langle \xi(p), \xi(q)\rangle\Theta_{\max}\big). \tag{6-6}$$

By elementary geometry (see Figure 6), we can express the term involving the orientation field as

$$\langle \xi(p), \xi(q)\rangle = \Theta_p\Theta_q - \sqrt{(1-\Theta_p^2)(1-\Theta_q^2)}. \tag{6-7}$$

Combining (6-6) with (3-4), noting that $\|\eta^T\|_{L^2}^2 = \|\Theta\|_{L^2}^2$, and using (6-7),

$$\begin{aligned}\frac{d}{dt}\left(\frac{\Theta_{\min}}{\ell}\right) &\geq \frac{1}{\ell^2}\big((2\|\Theta\|_{L^2}^2 - 1)\Theta_{\min} - \langle \xi(p), \xi(q)\rangle\Theta_{\max}\big)\\ &= \frac{1}{\ell^2}\big(2\Theta_{\min}^3 - (1-\Theta_{\max}^2)\Theta_{\min} + \sqrt{(1-\Theta_{\min}^2)(1-\Theta_{\max}^2)}\Theta_{\max}\big)\\ &\geq \frac{1}{\ell^2}\big(2\Theta_{\min}^3 + (1-\Theta_{\max}^2)(\Theta_{\max} - \Theta_{\min})\big) \geq 0.\end{aligned}$$

Therefore, if $\Theta_{\min} \geq 0$ at $t = 0$, then $\Theta_{\min}/\ell$ is a nondecreasing function of t, hence is nonnegative for all $t \in [0, T)$. This proves that $C(t)$ remains convex for all $t \in [0, T)$. □

We are now ready to prove the main result of this section.

Proof of Theorem 6.5. By Theorem 4.1, it suffices to show that the chord shortening flow $C(t)$ exists only on a maximal time interval $t \in [0, T)$ with $T < +\infty$. First of all, $\Theta \geq 0$ for all $t \in [0, T)$ by Lemma 6.8. Using (6-3) and (3-4), noticing that $2\|\Theta\|_{L^2}^2 \geq \overline{\Theta}^2$, a similar argument as in the proof of Lemma 6.8 gives

$$\frac{d}{dt}\left(\frac{\overline{\Theta}}{\ell}\right) \geq \frac{1}{\ell^2}(\overline{\Theta}^2 - 1 - \langle \xi, \xi\rangle)\overline{\Theta} \geq \frac{1}{\ell^2}(\Theta_{\min}^2 + \Theta_{\min}\Theta_{\max})\overline{\Theta} \geq 0.$$

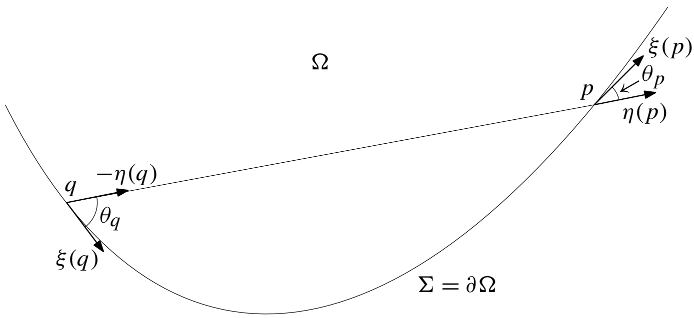

Figure 6. The convex region cut out by a convex chord in Ω. Note that $\langle \xi(p), \xi(q) \rangle = \cos(\theta_p + \theta_q)$.

Therefore, $\bar{\Theta}/\ell$ is a nondecreasing function of t. Since $\bar{\Theta}/\ell > 0$ at $t = 0$, it remains bounded away from zero for all $t \in [0, T)$. Therefore, if $T = +\infty$, by Theorem 4.1 we must have that $C(t)$ converges to an orthogonal geodesic chord and thus $\bar{\Theta}/\ell \to 0$, which is a contradiction. □

Acknowledgements

The author would like to express his gratitude to Prof. Richard Schoen for his interest in this work. He also want to thank Mario Micallef and Luen-Fai Tam for helpful comments and discussions. This work is partially supported by a research grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (project no. CUHK 14323516) and CUHK Direct Grant (project code 3132705). The author would also like to thank the anonymous referee for various useful suggestions which help improve the exposition of this paper.

References

[Altschuler 1991] S. J. Altschuler, "Singularities of the curve shrinking flow for space curves", *J. Differential Geom.* **34**:2 (1991), 491–514. MR Zbl

[Birkhoff 1917] G. D. Birkhoff, "Dynamical systems with two degrees of freedom", *Trans. Amer. Math. Soc.* **18**:2 (1917), 199–300. MR Zbl

[Bos 1963] W. Bos, "Kritische Sehnen auf Riemannschen Elementarraumstücken", *Math. Ann.* **151** (1963), 431–451. MR Zbl

[Cheeger and Ebin 1975] J. Cheeger and D. G. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland Math. Library **9**, North-Holland, Amsterdam, 1975. MR Zbl

[Colding and Minicozzi 2011] T. H. Colding and W. P. Minicozzi, II, *A course in minimal surfaces*, Graduate Studies in Math. **121**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[Giambò et al. 2014] R. Giambò, F. Giannoni, and P. Piccione, "Multiplicity results for orthogonal geodesic chords and applications", *J. Fixed Point Theory Appl.* **16**:1-2 (2014), 259–272. MR Zbl

[Giannoni and Majer 1997] F. Giannoni and P. Majer, "On the effect of the domain on the number of orthogonal geodesic chords", *Differential Geom. Appl.* **7**:4 (1997), 341–364. MR Zbl

[Gluck and Ziller 1983] H. Gluck and W. Ziller, "Existence of periodic motions of conservative systems", pp. 65–98 in *Seminar on minimal submanifolds*, edited by E. Bombieri, Ann. of Math. Stud. **103**, Princeton Univ. Press, 1983. MR Zbl

[Grayson 1989] M. A. Grayson, "Shortening embedded curves", *Ann. of Math.* (2) **129**:1 (1989), 71–111. MR Zbl

[Hayashi 1982] K. Hayashi, "Double normals of a compact submanifold", *Tokyo J. Math.* **5**:2 (1982), 419–425. MR Zbl

[Huisken 1984] G. Huisken, "Flow by mean curvature of convex surfaces into spheres", *J. Differential Geom.* **20**:1 (1984), 237–266. MR Zbl

[Klingenberg 1978] W. Klingenberg, *Lectures on closed geodesics*, Grundlehren der Math. Wissenschaften **230**, Springer, 1978. MR Zbl

[Kuiper 1964] N. H. Kuiper, "Double normals of convex bodies", *Israel J. Math.* **2** (1964), 71–80. MR Zbl

[Lusternik and Schnirelmann 1934] L. Lusternik and L. Schnirelmann, *Méthodes topologiques dans les problèmes variationnels*, Hermann, Paris, 1934. Zbl

[Poincaré 1905] H. Poincaré, "Sur les lignes géodésiques des surfaces convexes", *Trans. Amer. Math. Soc.* **6**:3 (1905), 237–274. MR JFM

[Riede 1968] A. Riede, "Lotgeodätische: Morse-Theorie für Probleme mit kompakten Randbedingungen", *Arch. Math.* (*Basel*) **19** (1968), 103–112. MR Zbl

[Taĭmanov 1992] I. A. Taĭmanov, "On the existence of three nonintersecting closed geodesics on manifolds that are homeomorphic to the two-dimensional sphere", *Izv. Ross. Akad. Nauk Ser. Mat.* **56**:3 (1992), 605–635. In Russian; translated in *Russian Acad. Sci. Izv. Math.* **40**:3 (1993), 565–590. MR

[Taylor 1996] M. E. Taylor, *Partial differential equations, I: Basic theory*, Applied Math. Sciences **115**, Springer, 1996. MR Zbl

[Zhou 2016] X. Zhou, "On the free boundary min-max geodesics", *Int. Math. Res. Not.* **2016**:5 (2016), 1447–1466. MR Zbl

Received February 6, 2018. Revised June 17, 2018.

MARTIN MAN-CHUN LI
DEPARTMENT OF MATHEMATICS
THE CHINESE UNIVERSITY OF HONG KONG
HONG KONG
CHINA

martinli@math.cuhk.edu.hk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 299, No. 2, 2019
dx.doi.org/10.2140/pjm.2019.299.489

RIGIDITY THEOREMS OF HYPERSURFACES WITH FREE BOUNDARY IN A WEDGE IN A SPACE FORM

JUNCHEOL PYO

Dedicated to Professor Jaigyoung Choe in honor of his 65th birthday.

This paper presents some rigidity results about compact hypersurfaces with free boundary in a wedge in a space form. First, we prove that every compact immersed stable constant mean curvature hypersurface with free boundary in a wedge is part of an intrinsic sphere centered at a point of the edge of the wedge. Second, we show that the same rigidity result holds for a compact embedded constant higher-order mean curvature hypersurface with free boundary in a wedge. Finally, we extend this result to a compact immersed hypersurface with free boundary in a wedge that has the additional property that the ratio of two higher-order mean curvatures is constant.

The same conclusions hold for a compact hypersurface with free boundary that lies in a half-space in a space form.

1. Introduction

The set of all points at a given positive intrinsic distance from a fixed point in a manifold will be called an *intrinsic sphere*. Intrinsic spheres in space forms have been characterized in a number of different ways. Among all hypersurfaces of a given volume bounding a domain in a space form, an intrinsic sphere has the least area; that is, it is the boundary of an isoperimetric domain in a space form. Every smooth boundary of an isoperimetric domain is a stable constant mean curvature (CMC) hypersurface. Barbosa and do Carmo [1984] proved that an intrinsic sphere in Euclidean space is the only closed stable immersed CMC hypersurface; Barbosa, do Carmo, and Eschenburg [Barbosa et al. 1988] extended this result to other space forms.

We call a hypersurface a *totally geodesic hypersurface* if all of its intrinsic geodesics are also geodesic curves in the ambient manifold. Totally geodesic hypersurfaces and intrinsic spheres are the only totally umbilic hypersurfaces. The

This research was supported in part by the National Research Foundation of Korea (NRF-2017R1E1A1A03070495 and NRF-2017R1A5A1015722).
MSC2010: 53C24, 49Q10.
Keywords: intrinsic spheres, surfaces with free boundary, higher-order mean curvature.

mean curvature as well as the higher-order mean curvature are all constant on such a surface. Alexandrov [1962] proved that a closed embedded CMC hypersurface in Euclidean space must be an intrinsic sphere. This result has also been generalized to hyperbolic space and the open hemisphere. Ros [1988] generalized Alexandrov's results to a closed embedded hypersurface in Euclidean space of constant scalar curvature. Using the so-called Alexandrov reflection method, Korevaar [1988] gave another proof of Ros' result and extended it to the hyperbolic space and the open hemisphere. Ros [1987] generalized Alexandrov's result to hypersurfaces of constant higher-order mean curvature in Euclidean space; Montiel and Ros [1991] settled Alexandrov's result for other space forms.

Koh and Lee [2001] characterized intrinsic spheres in a space form in terms of the ratio of two higher-order mean curvatures. They proved that a closed hypersurface in a space form is an intrinsic sphere if it has constant ratio H_r/H_l, where $l < r$, and nonvanishing H_l, where H_r is the r-th order mean curvature of the hypersurface.

It is natural to extend the above results for closed surfaces to compact surfaces with nonempty boundary in a domain. When the domain is a ball, Nitsche [1985] showed that an immersed disk-type CMC surface in a ball which intersects the boundary sphere orthogonally is part of a sphere, and Souam [1997] extended Nitsche's result to other space forms. Presently, only partial results are known for higher-dimensional stable CMC hypersurfaces in a ball [Ros and Vergasta 1995; Souam 1997].

Recently, when the domain is a wedge in Euclidean space, López [2014] showed that a compact connected CMC surface orthogonally meeting the boundary of the wedge in $\mathbb{R}^3$ is part of sphere if it is either stable or embedded. In this paper, we extend López's results both to other space forms and to a higher-dimensional case. We now establish some notation that will be used throughout the remainder of this paper.

Let $\overline{M}^{n+1}(k)$ be the $(n+1)$-dimensional simply connected space form of constant sectional curvature k. By changing the metric conformally we may assume that $k=0$ or $k=\pm 1$; that is, $\overline{M}^{n+1}(0)=\mathbb{R}^{n+1}$, $\overline{M}^{n+1}(-1)=\mathbb{H}^{n+1}$, and $\overline{M}^{n+1}(1)=\mathbb{S}^{n+1}$. When $k=1$, we consider the open hemisphere $\mathbb{S}^{n+1}_+$ rather than the whole sphere.

Let Π_1 and Π_2 be two totally geodesics in $\overline{M}^{n+1}(k)$ which intersect. By Π_1 and Π_2, $\overline{M}^{n+1}(k)$ is divided into four connected domains. Choosing any of the four domains and then taking closure of the domain, we have a wedge-shaped closed connected domain $W \subset \overline{M}^{n+1}(k)$. For simplicity, we refer to W as a *wedge*. Let M^n be an n-dimensional compact connected orientable Riemannian manifold with nonempty boundary ∂M. Let $\psi : M \to \overline{M}^{n+1}(k)$ be an isometric immersion, and we identify M with $\psi(M)$. In this paper, we consider a hypersurface M in a wedge W, which means that there exists an immersion $\psi : M \to W$ such that $\psi(\mathrm{int}(M)) \subset \mathrm{int}(W)$ and $\psi(\partial M) \subset \partial W$, where ∂W is the boundary of W and $\mathrm{int}(A)$ denotes the interior of a set A. The $(n-1)$-dimensional totally geodesic

$E = \Pi_1 \cap \Pi_2$ is called the *edge* of W. Since we consider the open hemisphere $\mathbb{S}^{n+1}_+$ rather than the whole sphere, for $k = 1$, the edge is a connected totally geodesic. In the other cases ($k = 0, -1$), clearly the edge is a connected one. Throughout this paper, we assume that $\partial M \cap (\Pi_1 \setminus E) \neq \varnothing \neq \partial M \cap (\Pi_2 \setminus E)$ and all hypersurfaces are connected. We call M a *hypersurface with free boundary* in W when M intersects ∂W orthogonally along ∂M.

First, in Section 3, we prove:

> *A compact immersed stable CMC hypersurface with free boundary in a wedge in a space form is part of an intrinsic sphere centered at a point of the edge of the wedge.*

This is a generalization of Barbosa, do Carmo, and Eschenburg's results [Barbosa and do Carmo 1984; Barbosa et al. 1988] for hypersurfaces with free boundary in a wedge. A CMC hypersurface M in a wedge W is called a *capillary* hypersurface if M meets the boundary of W with a constant angle along ∂M. McCuan [1997] and Park [2005] showed that a capillary surface in a wedge that is topologically an annulus is part of a sphere. Recently, Choe and Koiso [2016] proved that a compact capillary hypersurface in a wedge that is disjoint from the edge is part of an intrinsic sphere if the boundary of the capillary surface is embedded for the surface case, or if the boundary is convex for the higher-dimensional case. More results and more physical explanation for capillary surfaces can be found in [Concus and Finn 1998; Concus et al. 2001; Finn 1986].

Ros [1987] obtained an interesting inequality for closed hypersurfaces of positive mean curvature. When the mean curvature is a positive constant, a linear isoperimetric inequality for closed hypersurfaces of nonvanishing mean curvature is satisfied. This inequality was extended to other space forms by Brendle [2013] and Qiu and Xia [2015] in different ways. In Section 4, we extend these results to compact hypersurfaces with nonempty boundary. Besides Reilly's formula, somewhat surprisingly, many geometric and rigidity results can be deduced from the so-called Minkowski formula; see, for example, [Koh 1998; Montiel and Ros 1991; Ros 1987]. Montiel and Ros [1991] extended the Minkowski formula in space forms. In Section 5, we extend the Minkowski formula for closed hypersurfaces to hypersurfaces with free boundary in a wedge. Choe and Park [2011] extended the Minkowski formula for hypersurfaces with free boundary in a cone in Euclidean space. Second, in Section 6, we extend the result of [Montiel and Ros 1991] to hypersurfaces with free boundary in a wedge using a Ros-type inequality and a Minkowski-type formula for compact hypersurfaces with free boundary:

> *A compact embedded constant higher-order mean curvature hypersurface with free boundary in a wedge in a space form is part of an intrinsic sphere centered at a point of the edge of the wedge.*

In the last section, we extend the results of [Koh and Lee 2001] to hypersurfaces with free boundary. In this case, the same rigidity holds when the hypersurface lies in the wedge near the boundary. More precisely, we prove:

> *If a compact immersed hypersurface has nonempty boundary such that near the boundary the hypersurface meets the boundary of a wedge orthogonally along the boundary, then it is part of an intrinsic sphere if H_l does not vanish and the ratio H_r/H_l is constant for $0 < l < r$.*

Note that there are no a priori restrictions on the topology of the hypersurface M; that is, it may have some genus or boundary components. Also note that the proof works in the case that the boundary ∂M lies in a hyperplane, that is, we obtain similar results for M in a half-space in a space form.

2. Preliminaries

For $\overline{M}^{n+1}(k)$, when $k = -1$, use the hyperboloid model and when $k = 1$, take the usual embedding to $\mathbb{R}^{n+2}$. More precisely, let $\mathbb{L}^{n+2}$ be the $(n+2)$-dimensional Lorentz–Minkowski space with the Lorentzian metric

$$\langle x, y\rangle = x_1 y_1 + \cdots + x_{n+1} y_{n+1} - x_{n+2} y_{n+2}.$$

Then, $\overline{M}^{n+1}(-1) \subset \mathbb{L}^{n+2}$ is defined as

$$\{x \in \mathbb{L}^{n+2} \mid |x|^2 = -1, x_{n+2} \geq 1\}.$$

Let $\psi : M \to \overline{M}^{n+1}(k)$ be an immersion. If $k = -1$, we regard this immersion as $\psi : M \to \mathbb{L}^{n+2}$, and if $k = 1$ we regard it as $\psi : M \to \mathbb{R}^{n+2}$.

Denote by $\bar{\nabla}$, $\bar{\Delta}$, and $\bar{\nabla}^2$ the gradient, the Laplacian, and the Hessian on $\overline{M}^{n+1}(k)$, respectively, and denote by ∇, Δ, N, σ, and H the gradient, the Laplacian, the unit outward normal vector field whenever this makes sense, the second fundamental form, and the normalized mean curvature on $M \subset \overline{M}^{n+1}(k)$, respectively. Let dV, dA, and ds be canonical measures of $\overline{M}^{n+1}(k)$, M, and ∂M, respectively.

We recall the formal definition of stability of CMC hypersurfaces; see [Barbosa et al. 1988; Ros and Vergasta 1995; Souam 1997] for further details. Let $W \subset \overline{M}^{n+1}(k)$ be a wedge. A CMC hypersurface with free boundary in W arises from a critical point of the area functional for all volume-preserving variations in W. More precisely, let $\psi : M \to W \subset \overline{M}^{n+1}(k)$ be an immersion such that $\psi(\mathrm{int}(M)) \subset \mathrm{int}(W)$ and $\psi(\partial M) \subset \partial W$. A variation of ψ is a smooth family of proper hypersurfaces in W given by a 1-parameter family of immersions $\Psi_t : M \times (-\epsilon, \epsilon) \to W$ with $\Psi_0 = \psi$.

The area function is defined by

$$A(t) = \int_M dA_t,$$

where dA_t is the volume form of $\Psi_t(M)$. The volume function enclosing the space between $\psi(M)$ and $\Psi_t(M)$ is defined by

$$V(t) = \int_{M\times[0,t]} \Psi^* \, dV,$$

where dV is the volume form of $\overline{M}^{n+1}(k)$. The variation is said to be *volume-preserving* if $V(t) = V(0)$ for all t.

With the associated variational vector field $Y = (\partial\Psi/\partial t)|_{t=0}$, the first variation formulas of the area and the volume are

$$A'(0) = -n \int_M Hf \, dA + \int_{\partial M} \langle Y, \nu\rangle \, ds, \tag{1}$$

$$V'(0) = \int_M f \, dA, \tag{2}$$

where $f = \langle Y, N\rangle$. The variation is called *normal* if $Y = fN$, and *admissible* if $\Psi_t(\mathrm{int}(M)) \subset \mathrm{int}(W)$ and $\Psi_t(\partial M) \subset \partial W$ for all t.

From (1) and (2), ψ is a critical point of the area functional $A(t)$ for all volume-preserving and admissible variations if and only if $\psi(M) \subset W$ is a CMC hypersurface with free boundary.

By a standard computation, the second variation formula of any admissible volume-preserving normal variation is

$$A''(0) = -\int_M (f\,\Delta f + (|\sigma|^2 + kn) f^2) \, dA + \int_{\partial M} \left(f \frac{\partial f}{\partial \nu} - II(N, N) f^2 \right) ds,$$

where II is the second fundamental form of ∂W in $\overline{M}^{n+1}(k)$.

A stationary immersion $\psi : M \to W$ is called *stable* if $A''(0) \geq 0$ for any admissible volume-preserving normal variation of ψ. Let $\mathcal{F} = \left\{ f \in H^1(M) \,\middle|\, \int_M f \, dA = 0 \right\}$, where $H^1(M)$ denotes the first Sobolev space of M, and we define the *index form* $\mathcal{I}$ of ψ as the symmetric bilinear form on $H^1(M)$ given by

$$\mathcal{I}(f, g) = \int_M (\langle \nabla f, \nabla g\rangle - (|\sigma|^2 + kn) fg) \, dA - \int_{\partial M} II(N, N) fg \, ds.$$

It follows that the stationary immersion ψ is stable if and only if $\mathcal{I}(f, f) \geq 0$ for any $f \in \mathcal{F}$.

3. Stable CMC surfaces with free boundary in a wedge

Theorem 1. *Let W be a wedge in $\overline{M}^{n+1}(k)$. If M is a compact immersed stable CMC ($H \neq 0$) hypersurface with free boundary in W, then it is part of an intrinsic sphere centered at a point of the edge of W.*

Proof. Suppose the wedge W is determined by Π_1 and Π_2 and the edge E is given by $E = \Pi_1 \cap \Pi_2$. By an isometry in $\overline{M}^{n+1}(k)$, we assume that E contains the origin of $\mathbb{R}^{n+1}$ for $k=0$ case, the north pole $(0,\dots,0,1)\in\mathbb{R}^{n+2}$ for $k=1$ case and the point $(0,\dots,0,1)\in\mathbb{L}^{n+2}$ for $k=-1$ case. Let η_i, $i=1,2$, be the unit normal vector of Π_i, $i=1,2$, outward-pointing with respect to W. Let N be the unit normal vector field of M. Let ν be the outward unit conormal vector field along ∂M which means that ν is tangential to M and normal to ∂M. The free boundary condition implies that $\nu=\eta_i$ on $\partial M\cap\Pi_i$, $i=1,2$.

In $\mathbb{R}^{n+1}$, $\mathbb{H}^{n+1}$, and $\mathbb{S}^{n+1}_+$, the only totally umbilic hypersurfaces are the geodesic hypersurfaces and the intrinsic spheres; see Chapter 7 of [Spivak 1975]. Since M is assumed to satisfy $\partial M\cap\Pi_1\neq\varnothing\neq\partial M\cap\Pi_2$, the only possibility to be a compact surface with free boundary in a wedge is that M is part of an intrinsic sphere centered at a point on the edge $E\subset W$. So, we claim that M is totally umbilic. Proving this claim completes the proof.

Case: $k=0$. Let $h=\langle\psi,N\rangle$ be the support function of ψ. From a direct computation in [Barbosa and do Carmo 1984, Lemma 3.5], we have

$$\Delta h = nH - |\sigma|^2 h. \tag{3}$$

Since Π_1 and Π_2 are totally geodesics and M intersects ∂W orthogonally along ∂M, ν is a principal direction of ψ along ∂M. More precisely, it follows that, for any tangent vector field X of ∂M, we have

$$\langle\bar{\nabla}_\nu N, X\rangle = -\langle\sigma(X,\nu),N\rangle = -\langle\bar{\nabla}_X\nu, N\rangle = 0,$$

where the last equality follows from the fact that ν is constant on ∂M. Hence, for a function λ,

$$\frac{\partial h}{\partial\nu} = \langle\nu,N\rangle + \langle\psi,\bar{\nabla}_\nu N\rangle = \langle\nu,N\rangle + \langle\psi,\lambda\nu\rangle = 0. \tag{4}$$

Integrating (3) on M and applying Stokes' theorem, by (4) we have

$$\int_M nH - |\sigma|^2 h\, dA = 0. \tag{5}$$

From a direct computation, we have

$$\Delta|\psi|^2 = 2n(1 - H\langle\psi,N\rangle);$$

by integrating on M and applying Stokes' theorem, we obtain

$$\int_M (1 - H\langle\psi,N\rangle)\, dA = \frac{1}{n}\int_{\partial M}\langle\psi,\nu\rangle\, ds.$$

Since E contains the origin of $\mathbb{R}^{n+1}$ and $\nu=\eta_i$ on Π_i, $i=1,2$, respectively, $\langle\psi,\nu\rangle=0$ on ∂M. So, for a hypersurface with free boundary, we also have the Minkowski-type formula

$$\int_M (1 - H\langle\psi,N\rangle)\, dA = 0.$$

Set $u = 1 - Hh,\ u \in \mathcal{F}$. Since $\partial u/\partial \nu = 0$ by (4) and $II \equiv 0$ on ∂W, the second variation formula becomes

$$A''(0) = -\int_M (u\,\Delta u + |\sigma|^2 u^2)\,dA.$$

From a direct computation using (3), $u\,\Delta u + |\sigma|^2 u^2 = u(|\sigma|^2 - nH^2)$. Thus

$$\begin{aligned}
0 \le \mathcal{I}(u,u) &= -\int_M (u\,\Delta u + |\sigma|^2 u^2)\,dA && \text{(by stability condition)}\\
&= -\int_M ((1-Hh)(|\sigma|^2 - nH^2))\,dA\\
&= -\int_M (|\sigma|^2 - nH^2)\,dA + \int_M nH^2\,dA - \int_M nH^3 h\,dA && \text{(by (5))}\\
&= -\int_M (|\sigma|^2 - nH^2)\,dA + nH^2\int_M u\,dA\\
&= -\int_M (|\sigma|^2 - nH^2)\,dA \le 0 && \text{(by } nH^2 \le |\sigma|^2\text{)}.
\end{aligned}$$

It follows that $|\sigma|^2 = nH^2$ on M; that is, all points of M are umbilic.

From now on we consider the case $k \neq 0$. We first recall the following identities [Barbosa et al. 1988, Lemma 3.3]:

$$\Delta\psi = -nHN - kn\psi, \tag{6}$$

$$\Delta N = -|\sigma|^2 N - knH\psi. \tag{7}$$

<u>Case: $k = 1$</u>. Let $\bar{\psi} = \int_M \psi\,dA$ and $\bar{N} = \int_M N\,dA$. We claim that $\bar{N}$ belongs to the vector space spanned by $\{\bar{\psi}, \eta_1, \eta_2\}$.

Integrating (6) and applying Stokes' theorem, we obtain

$$\begin{aligned}
-nH\int_M N\,dA &= n\int_M \psi\,dA + \int_M \Delta\psi\,dA\\
&= n\bar{\psi} + \mathrm{Vol}(\partial M \cap \Pi_1)\eta_1 + \mathrm{Vol}(\partial M \cap \Pi_2)\eta_2.
\end{aligned}$$

Therefore $\bar{N}$ is spanned by $\{\bar{\psi}, \eta_1, \eta_2\}$, completing the claim.

Now, choose $n-1$ vectors $\{v_1, \dots, v_{n-1}\}$ in $\mathbb{R}^{n+2}$ such that

$$\langle \bar{\psi}, v_i\rangle = \langle \eta_1, v_i\rangle = \langle \eta_2, v_i\rangle = 0, \quad i = 1, \dots, n-1.$$

Clearly, $\langle \bar{N}, v_i\rangle = 0$ for $i = 1, \dots, n-1$.

For each $i = 1, \dots, n-1$, define $f_i = \langle \psi, v_i\rangle$ and $g_i = \langle N, v_i\rangle$. Since $\langle \bar{\psi}, v_i\rangle = \langle \bar{N}, v_i\rangle = 0,\ i = 1, \dots, n-1$, we have $\int_M f_i\,dA = \int_M g_i\,dA = 0$.

From (6) and (7), for each $i = 1, \dots, n-1$, we deduce

$$\Delta f_i + nf_i = -nHg_i, \tag{8}$$

$$\Delta g_i + |\sigma|^2 g_i = -nHf_i. \tag{9}$$

Recall ν is the outward unit conormal vector field along ∂M and $\eta_i = \nu$ on $\partial M \cap \Pi_i$ for $i = 1, 2$. Along the boundary ∂M,

$$\frac{\partial f_i}{\partial \nu} = \frac{\partial}{\partial \nu}\langle \psi, v_i\rangle = \langle \nu, v_i\rangle = 0, \quad i = 1, \dots, n-1. \tag{10}$$

Since ν is a principal direction of ψ along ∂M, for a function λ,

$$\frac{\partial g_i}{\partial \nu} = \frac{\partial}{\partial \nu}\langle N, v_i\rangle = \langle \bar{\nabla}_\nu N, v_i\rangle = \langle \lambda\nu, v_i\rangle = 0, \quad i = 1, \dots, n-1. \tag{11}$$

By combining (8)–(11), for each $i = 1, \dots, n-1$, the index form is

$$\mathcal{I}(f_i, f_i) = n\int_M H f_i g_i \, dA - \int_M |\sigma|^2 f_i^2 \, dA, \tag{12}$$

$$\mathcal{I}(g_i, g_i) = n\int_M H f_i g_i \, dA - n\int_M g_i^2 \, dA, \tag{13}$$

and summing up,

$$\begin{aligned}
0 &\le \sum_{i=1}^{n-1} \mathcal{I}(f_i, f_i) + \mathcal{I}(g_i, g_i) && \text{(by stability condition)} \\
&= -\sum_{i=1}^{n-1}\Bigl(\int_M (|\sigma|^2 f_i^2 - 2nH f_i g_i + n g_i^2)\, dA\Bigr) \\
&\le -n\sum_{i=1}^{n-1}\Bigl(\int_M (H^2 f_i^2 - 2H f_i g_i + g_i^2)\, dA\Bigr) && \text{(by } nH^2 \le |\sigma|^2) \qquad (14)\\
&= -n\sum_{i=1}^{n-1}\int_M (H f_i - g_i)^2 \, dA \le 0.
\end{aligned}$$

Since inequality (14) turns to equality, we have

$$\sum_{i=1}^{n-1}\int_M (|\sigma|^2 - nH^2) f_i^2 \, dA = 0, \tag{15}$$

and by $nH^2 \le |\sigma|^2$ again, we obtain

$$(|\sigma|^2 - nH^2)\Bigl(\sum_{i=1}^{n-1} f_i^2\Bigr) = 0 \quad \text{on } M. \tag{16}$$

For each $i = 1, \dots, n-1$, the zero set of f_i in M is the set of points that belong to M and the hyperplane $\{x \in \mathbb{R}^{n+2} \mid \langle x, v_i\rangle = 0\}$, so, the zero set of $\sum_{i=1}^{n-1} f_i^2$ in M is the set of points that belong to M and the three-dimensional subspace which is orthogonal to $\{v_i \mid i = 1, \dots, n-1\}$.

For $n \ge 3$, the zero set of $\sum_{i=1}^{n-1} f_i^2$ in M has measure zero in M. From (15), $|\sigma|^2 = nH^2$ on M; that is, M is totally umbilic. For the surfaces case ($n = 2$), we use the reductio ad absurdum argument. Suppose M is not totally umbilic, then the umbilic points are isolated by the holomorphic Hopf differential. By (15), $f_1 \equiv 0$ on M, and it follows that M is a surface in a three-dimensional subspace that is orthogonal to v_1 and hence it is totally geodesic, a contradiction. Therefore, M is totally umbilic. This completes the claim when $k = 1$.

Case: $k = -1$. Let $v = (0, \dots, 1) \in E$. Define $f = \langle \psi, v \rangle$ and $g = \langle N, v \rangle$. By (6) and (7), a direct computation yields

$$\tfrac{1}{2}\Delta f^2 = f\,\Delta f + |\nabla f|^2 = -nHfg + nf^2 + |\nabla f|^2, \tag{17}$$

$$\tfrac{1}{2}\Delta g^2 = g\,\Delta g + |\nabla g|^2 = -|\sigma|^2 g^2 + nHfg + |\nabla g|^2, \tag{18}$$

$$\begin{aligned}\Delta(fg) &= f\,\Delta g + g\,\Delta f + 2\langle \nabla f, \nabla g\rangle \\ &= -|\sigma|^2 fg + nHf^2 - nHg^2 + nfg + 2\langle \nabla f, \nabla g\rangle,\end{aligned} \tag{19}$$

and

$$\tfrac{1}{2}H^2\,\Delta f^2 - H\,\Delta(fg) + \tfrac{1}{2}\Delta g^2 = |H\nabla f - \nabla g|^2 - (|\sigma|^2 - nH^2)(g^2 - Hfg).$$

Recall $\eta_i = \nu$ on $\partial M \cap \Pi_i$, $i = 1, 2$. Along the boundary ∂M,

$$\frac{\partial f}{\partial \nu} = \frac{\partial}{\partial \nu}\langle \psi, v\rangle = \langle \nu, v\rangle = 0. \tag{20}$$

Similar to the case when $k = 0$, ν is a principal direction of ψ along ∂M. Hence,

$$\frac{\partial g}{\partial \nu} = \frac{\partial}{\partial \nu}\langle N, v\rangle = \langle \bar{\nabla}_\nu N, v\rangle = 0. \tag{21}$$

From (20) and (21), we get $\int_M \frac{1}{2}H^2\,\Delta f^2 - H\,\Delta(fg) + \frac{1}{2}\Delta g^2\,dA = 0$, and hence

$$\int_M (|\sigma|^2 - nH^2)(g^2 - Hfg)\,dA = \int_M |H\nabla f - \nabla g|^2\,dA. \tag{22}$$

Define $u = Hg - f$. Since $u = Hg - f = -\frac{1}{n}\Delta f$ and by Stokes' theorem,

$$\int_M u\,dA = -\frac{1}{n}\int_{\partial M}\langle \nu, v\rangle\,ds = 0. \tag{23}$$

From (22) and (23),

$$\begin{aligned}\mathcal{I}(u,u) &= \int_M (|\sigma|^2 - nH^2)(Hfg - f^2)\,dA \\ &= \int_M (|\sigma|^2 - nH^2)(g^2 - f^2)\,dA - \int_M |H\nabla f - \nabla g|^2\,dA.\end{aligned} \tag{24}$$

To simplify computations, we choose an orthonormal frame $\{e_A \mid A = 0, \dots, n+1\}$ around a point $\psi(p)$, $p \in M$, such that $e_0 = \psi$, $e_{n+1} = N$ and $e_1, \dots, e_n$ are tangential to $\psi(M)$.

With this frame,

$$v = -\langle \psi, v\rangle \psi + \langle N, v\rangle N + \sum_{i=1}^{n} \langle e_i, v\rangle e_i,$$

and

$$\nabla f = \sum_{i=1}^{n} \langle e_i, v\rangle e_i.$$

It follows that

$$\begin{aligned} -1 = \langle v, v\rangle &= -\langle \psi, v\rangle^2 + \langle N, v\rangle^2 + \sum_{i=1}^{n} \langle e_i, v\rangle^2 \\ &= -f^2 + g^2 + |\nabla f|^2. \end{aligned}$$

Hence,

$$g^2 - f^2 = -(1 + |\nabla f|^2). \tag{25}$$

By (24) and (25),

$$\begin{aligned} 0 &\le \mathcal{I}(u, u) && \text{(by stability condition)} \\ &= -\int_M (|\sigma|^2 - nH^2)(1 + |\nabla f|^2)\, dA - \int_M |H\nabla f - \nabla g|^2\, dA \\ &\le 0 && \text{(from } nH^2 \le |\sigma|^2\text{)}, \end{aligned}$$

that is, all points of M are umbilic, and hence, the conclusion for the case $k = -1$ follows. □

Observe that the proof also holds when the boundary ∂M lies in a hyperplane. This gives rise to the following theorem.

Theorem 2. *Let $\mathcal{H}$ be a half-space in $\overline{M}^{n+1}(k)$ determined by a hyperplane P. Let M be a compact immersed stable CMC hypersurface with free boundary in $\mathcal{H}$. Then M is an intrinsic hemisphere centered at a point of P.*

4. Ros-type inequality

In this section, we extend the Ros-type inequality for closed hypersurfaces to compact hypersurfaces with free boundary in a wedge.

Theorem 3. *Let $W \subset \overline{M}^{n+1}(k)$ be a wedge, and E be the edge of W. Let M be a compact embedded hypersurface with free boundary in W. Let Ω be the compact domain enclosed by M and ∂W. Defining $r(x) = \mathrm{dist}_{\overline{M}^{n+1}(k)}(x, v)$ for a fixed point $v \in E$,*

$$V_k(x) = \begin{cases} 1 & \text{if } k = 0, \\ \cos r(x) & \text{if } k = 1, \\ \cosh r(x) & \text{if } k = -1. \end{cases}$$

If the mean curvature H is positive on M, then

$$\int_M \frac{V_k}{H}\,dA \geq (n+1)\int_\Omega V_k\,dV, \tag{26}$$

and equality holds if and only if M is part of an intrinsic sphere.

Proof. Take $\Omega_\epsilon \subset \Omega$ to be a domain with a smooth boundary obtained from Ω by rounding off the singular part of $\partial\Omega$ in a small distance $\epsilon > 0$. Let N be the outward unit normal vector field of $\partial\Omega$; it is the same one on M as in the previous section.

From a direct computation, $\bar{\nabla}^2 V_k = -kV_k g$, where g is the metric of $\overline{M}^{n+1}(k)$. For any smooth function f on Ω_ϵ, the Reilly-type formula is given by

$$\begin{aligned}\int_{\Omega_\epsilon} V_k\Big((\bar{\Delta} f + k(n+1)f)^2 - |\bar{\nabla}^2 f + kfg|^2\Big)\,dV \\ = \int_{\partial\Omega_\epsilon} V_k(2u\,\Delta z + nHu^2 + \sigma(\nabla z, \nabla z) + 2nkuz)\,dA \\ + \int_{\partial\Omega_\epsilon} \bar{\nabla}_N V_k(|\nabla z|^2 - nkz^2)\,dA,\end{aligned} \tag{27}$$

where $z = f|_M$ and $u = \bar{\nabla}_N f$. Equation (27) is a particular case of the general Reilly-type formula in a Riemannian manifold; see [Qiu and Xia 2015, Theorem 1.1].

Case: $k = 0$. Let $f : \Omega_\epsilon \to \mathbb{R}$ be the solution to the mixed boundary value problem

$$\begin{cases} \bar{\Delta} f = 1 & \text{in } \Omega_\epsilon, \\ f = 0 & \text{on } \partial\Omega_\epsilon \setminus \partial W, \\ u = \partial f/\partial N = 0 & \text{on } \partial\Omega_\epsilon \cap \partial W. \end{cases}$$

Equation (27) becomes the classical Reilly formula

$$\int_{\Omega_\epsilon} ((\bar{\Delta} f)^2 - |\bar{\nabla}^2 f|^2)\,dV = \int_{\partial\Omega_\epsilon\setminus\partial W} nHu^2\,dA + \int_{\partial\Omega_\epsilon\cap\partial W} \sigma(\nabla z, \nabla z)\,dA. \tag{28}$$

Since ∂W is composed of part of a totally geodesic, $\sigma \equiv 0$ on ∂W. From the Cauchy–Schwarz inequality, (28) becomes

$$\frac{\mathrm{Vol}(\Omega_\epsilon)}{n+1} \geq \int_{\partial\Omega_\epsilon\setminus\partial W} Hu^2\,dA. \tag{29}$$

On the other hand,

$$\begin{aligned}(\mathrm{Vol}(\Omega_\epsilon))^2 = \Big(\int_{\Omega_\epsilon} \bar{\Delta} f\,dV\Big)^2 = \Big(\int_{\partial\Omega_\epsilon\setminus\partial W} u\,dA\Big)^2 \\ \leq \int_{\partial\Omega_\epsilon\setminus\partial W} Hu^2\,dA \int_{\partial\Omega_\epsilon\setminus\partial W} \frac{1}{H}\,dA \leq \frac{\mathrm{Vol}(\Omega_\epsilon)}{n+1}\int_{\partial\Omega_\epsilon\setminus\partial W} \frac{1}{H}\,dA,\end{aligned} \tag{30}$$

where the first inequality comes from the Hölder inequality and the second inequality is a consequence of (29). Therefore, letting $\epsilon \to 0$, we obtain (26).

When equality occurs, the Cauchy–Schwarz inequality implies that the Hessian $\bar{\nabla}^2 f$ is proportional to the identity matrix. Because $\bar{\Delta} f = 1$ on Ω, $\bar{\nabla}^2 f = \frac{1}{n} g$ in Ω. With $f = 0$ on M, the conclusion follows from the Obata-type result that M is part of an intrinsic sphere. This completes the proof when $k = 0$; see [Reilly 1980, Theorem B].

Case: $k \neq 0$. Let $f : \Omega_\epsilon \to \mathbb{R}$ be the solution to the mixed boundary value problem

$$\begin{cases} \bar{\Delta} f + k(n+1) f = 1 & \text{in } \Omega_\epsilon, \\ f = 0 & \text{on } \partial\Omega_\epsilon \setminus \partial W, \\ u = \partial f / \partial N = 0 & \text{on } \partial\Omega_\epsilon \cap \partial W. \end{cases} \tag{31}$$

From the Cauchy–Schwarz inequality,

$$\begin{aligned} \frac{n}{n+1} \int_{\Omega_\epsilon} & V_k (\bar{\Delta} f + k(n+1) f)^2 \, dV \\ & \geq \int_{\Omega_\epsilon} V_k ((\bar{\Delta} f + k(n+1) f)^2 - |\bar{\nabla}^2 f + kfg|^2) \, dV. \end{aligned} \tag{32}$$

We deal with $\partial\Omega_\epsilon$ in two parts, $\partial\Omega_\epsilon \setminus \partial W$ and $\partial\Omega_\epsilon \cap \partial W$.

On $\partial\Omega_\epsilon \setminus \partial W$, $z = f|_{\partial\Omega_\epsilon \setminus \partial W} = 0$, and

$$\begin{aligned} \int_{\partial\Omega_\epsilon \setminus \partial W} & V_k (2u \, \Delta z + nHu^2 + \sigma(\nabla z, \nabla z) + 2nkuz) \, dA \\ & + \int_{\partial\Omega_\epsilon \setminus \partial W} \bar{\nabla}_N V_k (|\nabla z|^2 - nkz^2) \, dA = \int_{\partial\Omega_\epsilon \setminus \partial W} n V_k H u^2 \, dA. \end{aligned} \tag{33}$$

On $\partial\Omega_\epsilon \cap \partial W$, $u = 0$. Since ∂W is part of a totally geodesic, $\sigma(\nabla z, \nabla z) = 0$. Since $N = \eta_i$ on Π_i, $i = 1, 2$, and $\bar{\nabla} r(x) \subset \partial W$, we have $V_k(x) = \cos r(x)$ and $\bar{\nabla}_N V_k = -\sin r(x) g(\bar{\nabla} r(x), N) = 0$ or $V_k(x) = \cosh r(x)$ and $\bar{\nabla}_N V_k = -\sinh r(x) g(\bar{\nabla} r(x), N) = 0$ on ∂W. Then, we obtain

$$\begin{aligned} \int_{\partial\Omega_\epsilon \cap \partial W} & V_k (2u \, \Delta z + nHu^2 + \sigma(\nabla z, \nabla z) + 2nkuz) \, dA \\ & + \int_{\partial\Omega_\epsilon \cap \partial W} \bar{\nabla}_N V_k (|\nabla z|^2 - nkz^2) \, dA = 0. \end{aligned} \tag{34}$$

Then, from (31)–(34), we have

$$\frac{1}{n+1} \int_{\Omega_\epsilon} V_k \, dV \geq \int_{\partial\Omega_\epsilon \setminus \partial W} V_k H u^2 \, dA. \tag{35}$$

Because $\bar{\Delta} V_k = -(n+1) k V_k$ and $\bar{\nabla}_N V_k = 0$ on $\partial\Omega_\epsilon \cap \partial W$, the Green's formula implies

$$\int_{\Omega_\epsilon} V_k \, dV = \int_{\partial\Omega_\epsilon \setminus \partial W} V_k u \, dA. \tag{36}$$

On the other hand,

$$\left(\int_{\Omega_\epsilon} V_k\,dV\right)^2 = \left(\int_{\partial\Omega_\epsilon\setminus\partial W} V_k u\,dA\right)^2 \le \int_{\partial\Omega_\epsilon\setminus\partial W} V_k H u^2\,dA \int_{\partial\Omega_\epsilon\setminus\partial W} \frac{V_k}{H}\,dA \le \frac{1}{n+1}\int_{\Omega_\epsilon} V_k\,dV \int_{\partial\Omega_\epsilon\setminus\partial W} \frac{V_k}{H}\,dA, \tag{37}$$

where the first inequality follows from the Hölder inequality and the second inequality follows from (35).

Therefore, letting $\epsilon \to 0$ we obtain (26).

Combining (32) and (35)–(37) and the equality in (26),

$$|\bar\nabla^2 f + kfg|^2 = \frac{1}{n+1}(\bar\Delta f + k(n+1)f)^2.$$

Since $\bar\Delta f + k(n+1)f = 1$, we have

$$\bar\nabla^2\Big(f+\frac{1}{n+1}\Big) = -k\Big(f+\frac{1}{n+1}\Big)g \quad \text{in } \Omega.$$

With $f + 1/(n+1) = 1/(n+1)$ on M, the conclusion follows from the Obata-type result [Reilly 1980, Theorem B] that M is part of an intrinsic sphere. □

The above result is counterpart of the Ros-type inequality for closed hypersurfaces in [Brendle 2013, Theorem 3.5]. Qiu and Xia [2015] also gave another proof of a Ros-type inequality for closed hypersurfaces in manifolds which include space forms.

If the boundary of the compact hypersurface lies in a hyperplane of $\overline{M}^{n+1}(k)$ we conclude an analogous result:

Theorem 4. *Let $\mathcal{H}$ be a half-space in $\overline{M}^{n+1}(k)$ determined by a hyperplane P. Let M be a compact embedded hypersurface with free boundary in $\mathcal{H}$. Let Ω be the compact domain enclosed by M and P. Defining $r(x) = \mathrm{dist}_{\overline{M}^{n+1}(k)}(x, v)$ for a fixed point $v \in P$,*

$$V_k(x) = \begin{cases} 1 & \text{if } k = 0,\\ \cos r(x) & \text{if } k = 1,\\ \cosh r(x) & \text{if } k = -1.\end{cases}$$

If the mean curvature H is positive on M, then

$$\int_M \frac{V_k}{H}\,dA \ge (n+1)\int_\Omega V_k\,dV,$$

and equality holds if and only if M is an intrinsic hemisphere centered at a point of P.

5. Minkowski-type formula

With the unit normal vector field N of M, we denote by κ_i, $i = 1, \dots, n$, the principal curvatures of M. For any $r = 1, \dots, n$, the *mean curvature of order* r, H_r, is defined by the identity

$$P_n(t) := (1 + \kappa_1 t) \cdots (1 + \kappa_n t) = 1 + \binom{n}{1} H_1 t + \cdots + \binom{n}{n} H_n t^n \tag{38}$$

for any real number t. Note that H_1 is the normalized mean curvature of M, H_2 is the scalar curvature of M up to a constant, and H_n is the Gauss–Kronecker curvature of M. For convenience, we define $H_0 = 1$.

For higher-order mean curvatures, the following inequalities hold:

Lemma 5. *If there is a point of M where all the principal curvatures are positive and $H_r > 0$, $r = 1, \dots, n$, on M, then*:

(i) $H_l > 0$ *if* $l < r$.

(ii) $H_r / H_l \le H_{r-1} / H_{l-1}$ *for any* $l < r$.

(iii) $H_s^{(s-1)/s} \le H_{s-1}$ *and* $H_s^{1/s} \le H_1 = H$, *where equality holds only at umbilic points if* $s > 1$.

Proof. For (i) and (iii), see, for example, Lemma 1 of [Montiel and Ros 1991]. For (ii), see, for example, Section 12 of [Beckenbach and Bellman 1961]. □

Besides Reilly's formula, somewhat surprisingly, many geometric and rigidity results can be deduced from the so-called Minkowski formula; see, for example, [Montiel and Ros 1991; Ros 1987]. Montiel and Ros [1991] extended the Minkowski formula in space forms and gave another characterization of an intrinsic sphere. We now extend the Minkowski formula for closed hypersurfaces to hypersurfaces with free boundary in a wedge.

We include the proof of the Minkowski formula for closed hypersurfaces in space forms for the reader's convenience (see [Montiel and Ros 1991] for further details), and then generalize it to hypersurfaces with free boundary.

<u>Case: $k = 0$</u>. From a direct computation, we have

$$\Delta |\psi|^2 = 2n(1 - H\langle \psi, N \rangle). \tag{39}$$

For a real number t close enough to 0, the parallel hypersurface is given by

$$\psi_t = \exp_\psi tN = \psi + tN$$

and this is also an immersion.

If dA and $\kappa_1, \dots, \kappa_n$ denote the volume form and the principal curvatures of $\psi(M)$, respectively, then the volume form of $\psi_t(M) = M_t$ is given by

$$dA_t = (1 + \kappa_1 t) \cdots (1 + \kappa_n t)\, dA = P_n(t)\, dA,$$

where P_n is as in (38). From a direct computation, the mean curvature $H(t)$ of M_t is

$$H(t) = \frac{1}{n}\sum_i \frac{\kappa_i}{1+\kappa_i t} = \frac{1}{n}\frac{P_n'(t)}{P_n(t)}. \tag{40}$$

Integrating (39) on M_t gives,

$$\begin{aligned} 0 &= \int_M (1 - H(t)\langle \psi + tN, N\rangle)\, dA_t \\ &= \int_M \Big(P_n(t) - \frac{t}{n}P_n'(t) - \frac{1}{n}P_n'(t)\langle \psi, N\rangle\Big)\, dA, \end{aligned} \tag{41}$$

where the second equality follows from (38) and (40). Because (41) holds for any real variable t, all of its coefficients vanish. As a result, we obtain the Minkowski-type identity

$$\int_M H_{r-1} - H_r\langle \psi, N\rangle\, dA = 0, \quad r = 1, \dots, n. \tag{42}$$

<u>Case: $k \neq 0$</u>. Because of the similarity between $\overline{M}^{n+1}(-1)$ and $\overline{M}^{n+1}(1)$, we focus on $k = -1$. From a direct computation, for any $v \in \mathbb{L}^{n+2}$, we have

$$\Delta\langle \psi, v\rangle = n(\langle \psi, v\rangle - H\langle N, v\rangle), \tag{43}$$

and then, integration on M and applying the Stokes' theorem yield

$$\int_M (\langle \psi, v\rangle - H\langle N, v\rangle)\, dA = 0. \tag{44}$$

For a real number t close enough to 0, the parallel hypersurface is given by

$$\psi_t = \exp_\psi(tN) = \cosh t\,\psi + \sinh t\, N$$

and this is also an immersion.

If dA and $\kappa_1, \dots, \kappa_n$ denote the volume form and the principal curvatures of $\psi(M)$, respectively, then the volume form of M_t is given by

$$\begin{aligned} dA_t &= (\cosh t + \kappa_1 \sinh t)\cdots(\cosh t + \kappa_n \sinh t)\, dA \\ &= \cosh^n t\, P_n(\tanh t)\, dA, \end{aligned}$$

where P_n is as in (38). From a direct computation, the mean curvature $H(t)$ of ψ_t is

$$H(t) = \frac{n\cosh t \sinh t\, P_n(\tanh t) + P_n'(\tanh t)}{n\cosh^2 t\, P_n(\tanh t)}. \tag{45}$$

Integrating (44) on M_t and using (38) and (45), we have

$$\int_M (nP_n(\tanh t) - \tanh t\, P_n'(\tanh t))\langle \psi, v\rangle - P_n'(\tanh t)\langle N, v\rangle\, dA = 0. \tag{46}$$

Equation (46) holds for any variable $\tanh t$. By comparing its coefficients, we obtain the Minkowski-type identity

$$\int_M H_{r-1}\langle\psi, v\rangle - H_r\langle N, v\rangle\, dA = 0, \quad r = 1, \dots, n.$$

Similarly for the case $k = 1$, we have the following identities:

Minkowski-type identity [Montiel and Ros 1991]. *Let $\psi : M \to \overline{M}^{n+1}(k)$ be a closed orientable immersed hypersurface. For any $r = 1, \dots, n$, the following hold*:

(a) *If $k = 0$, then $\int_M H_{r-1} - H_r\langle\psi, N\rangle\, dA = 0$.*

(b) *If $k = -1$, then $\int_M H_{r-1}\langle\psi, v\rangle - H_r\langle N, v\rangle\, dA = 0$ for any $v \in \mathbb{L}^{n+2}$.*

(c) *If $k = 1$, then $\int_M H_{r-1}\langle\psi, v\rangle + H_r\langle N, v\rangle\, dA = 0$ for any $v \in \mathbb{R}^{n+2}$.*

We extend the Minkowski-type identity to immersed hypersurfaces with free boundary in a wedge in a space form.

Proposition 6. *Let $W \subset \overline{M}^{n+1}(k)$ be a wedge and E be the edge of W. Let M be a compact immersed hypersurface in $\overline{M}^{n+1}(k)$ with $\partial M \subset \partial W$ such that near ∂M, M lies inside of W and perpendicular to ∂W. Then, for any $r = 1, \dots, n$ we obtain*:

(a) *If $k = 0$, then $\int_M H_{r-1} - H_r\langle\psi, N\rangle\, dA = 0$.*

(b) *If $k = -1$, then $\int_M H_{r-1}\langle\psi, v\rangle - H_r\langle N, v\rangle\, dA = 0$ for any $v \in E$.*

(c) *If $k = 1$, then $\int_M H_{r-1}\langle\psi, v\rangle + H_r\langle N, v\rangle\, dA = 0$ for any $v \in E$.*

Proof. By an isometry in $\mathbb{R}^{n+1}$, we assume that E contains the origin of $\mathbb{R}^{n+1}$. For sufficiently small t, the parallel hypersurface $\psi_t(M) = M_t$ is an immersed hypersurface. Since W is a wedge and M is a hypersurface with free boundary, ∂M_t lies on ∂W and M_t intersects ∂W orthogonally along ∂M_t. Integrating (39) on M_t and applying Stokes' theorem, we have

$$\int_M \Big(P_n(t) - \frac{t}{n}P_n'(t) - \frac{1}{n}P_n'(t)\langle\psi, N\rangle\Big)\, dA = \frac{1}{2n}\int_{\partial M_t} \frac{\partial|\psi + tN|^2}{\partial\nu_t}\, ds, \tag{47}$$

where ν_t is the outward unit conormal vector field to ∂M_t. Since ∂M_t lies on ∂W and M_t intersects ∂W orthogonally along ∂M_t, $\partial|\psi + tN|^2/\partial\nu_t = 0$ on ∂M_t. Then (47) is the same as (41). The conclusion follows the same argument as that of the closed case.

Because of the similarity between the two cases ($k = \pm 1$), we consider only the case $k = -1$.

Recall η_i, $i = 1, 2$, is the unit normal vector of Π_i, $i = 1, 2$. By an isometry in $\overline{M}^{n+1}(-1)$, we assume that $v = (0, \dots, 0, 1) \in E$ and $\langle v, \eta_i\rangle = 0$, $i = 1, 2$. For sufficiently small t, the parallel hypersurface $\psi_t(M) = M_t$ is an immersed hypersurface. Since W is a wedge and M is a hypersurface with free boundary, ∂M_t lies on ∂W and M_t intersects ∂W orthogonally along ∂M_t.

Integrating (43) on M_t and applying Stokes' theorem give

$$\int_{M_t} (\langle \psi_t, v\rangle - H(t)\langle N_t, v\rangle)\, dA_t - \frac{1}{n}\int_{\partial M_t} \langle \nu_t, v\rangle\, ds = 0, \tag{48}$$

where ν_t is the outward unit conormal vector field to ∂M_t.

Since M_t intersects ∂W orthogonally along ∂M_t, $\nu_t = \eta_i$ on $\partial M_t \cap \Pi_i$, $i = 1, 2$, and then, $\langle \nu_t, v\rangle \equiv 0$ on ∂M_t. Then (48) is the same as (44). The conclusion follows the same argument as that of the closed case. □

Using the same argument, a similar result holds if the boundary of a hypersurface with free boundary lies in a hyperplane of $\overline{M}^{n+1}(k)$.

6. Constant-H_r embedded hypersurfaces with free boundary

Theorem 7. *Let $W \subset \overline{M}^{n+1}(k)$ be a wedge. Let $M \subset W$ be a compact embedded constant-H_r $(r = 1, \ldots, n)$ hypersurface with free boundary. Then M is part of an intrinsic sphere centered at a point of the edge of W.*

Proof. Denote by Ω the compact domain enclosed by M and ∂W.

For the case $k = 0$, by an isometry in $\mathbb{R}^{n+1}$, we assume that E contains the origin of $\mathbb{R}^{n+1}$. Because the unit normal vector to $\partial\Omega \cap \partial W$ is perpendicular to the position vector ψ,

$$\mathrm{Vol}(\Omega) = \frac{1}{n+1}\int_M \langle \psi, N\rangle\, dA, \tag{49}$$

where $\mathrm{Vol}(\Omega)$ is the volume of Ω and N is the outward unit vector field of M.

From (a) of Proposition 6 and (49), we have

$$\int_M H_{r-1}\, dA = H_r \int_M \langle \psi, N\rangle\, dA = (n+1) H_r \,\mathrm{Vol}(\Omega).$$

Denote by $S(r)$ the intrinsic sphere of radius r centered at the origin. For sufficiently large r, M is contained inside of $S(r)$. Decreasing $r \searrow 0$, we can find $r_0 > 0$ such that $S(r) \cap M = \varnothing$ for $r > r_0$ but $S(r_0) \cap M \neq \varnothing$. That is, $S(r_0)$ is the first touching to M at a point $q \in S(r_0) \cap M$. At the touching point q, all the principal curvatures of M and H_1 are positive by comparison with $S(r_0)$. This argument also holds in $\overline{M}^{n+1}(-1)$ without any change. When $\mathbb{S}^{n+1}_+$, if r close enough to $\frac{\pi}{2}$, M is contained inside of $S(r)$; thus, the Euclidean argument also holds.

From (iii) of Lemma 5,

$$\int_M H_{r-1}\, dA \geq \int_M H_r^{(r-1)/r}\, dA,$$

and then,

$$(n+1)\,\mathrm{Vol}(\Omega) \geq \int_M H_r^{-1/r}\,dA \geq \int_M \frac{1}{H}\,dA. \tag{50}$$

Comparing (26) and (50), M is part of an intrinsic sphere by Theorem 3.

Now, we consider $k \neq 0$ case. From a direct computation, we have $\bar{\Delta}\langle\psi, v\rangle = -k(n+1)\langle\psi, v\rangle$ for any $v \in E$. Integrating on Ω and using Stokes' theorem, we have

$$-k(n+1)\int_\Omega \langle\psi, v\rangle\,dV = \int_M \langle N, v\rangle\,dA + \int_{\partial\Omega\cap\partial W} \langle\nu, v\rangle\,dA,$$

where N and ν are the outward unit normal vector fields of M and $\partial\Omega\cap\partial W$, respectively.

By an isometry of $\overline{M}^{n+1}(k)$, we assume $v = (0, \dots, 0, 1) \in E$ and $\langle\eta_i, v\rangle = 0$, $i = 1, 2$. With v, $\langle\nu, v\rangle \equiv 0$ on $\partial\Omega\cap\partial W$, that is,

$$-k(n+1)\int_\Omega \langle\psi, v\rangle\,dV = \int_M \langle N, v\rangle\,dA. \tag{51}$$

Let $r(x) = \mathrm{dist}(x, v)$ be the distance function from v to x in $\overline{M}^{n+1}(k)$. If $k = -1$, then $\langle\psi, v\rangle = -\cosh r(\psi)$ and if $k = 1$, then $\langle\psi, v\rangle = \cos r(\psi)$; that is, $k\langle\psi, v\rangle = V_k(\psi)$ in $\overline{M}^{n+1}(k)$.

From (b) of Proposition 6, we have

$$\int_M H_{r-1}V_k(\psi) + H_r\langle N, v\rangle\,dA = 0.$$

Since H_r is constant and (51), $(n+1)H_r\int_\Omega V_k\,dV = \int_M H_{r-1}V_k\,dA$. By the same argument for the $k = 0$ case, there exists a point in M such that all the principal curvatures are positive. From (iii) of Lemma 5,

$$(n+1)H_r\int_\Omega V_k\,dV = \int_M H_{r-1}V_k\,dA \geq \int_M H_r^{(r-1)/r}V_k\,dA,$$

and then,

$$(n+1)\int_\Omega V_k\,dV \geq \int_M H_r^{-1/r}V_k\,dA \geq \int_M \frac{V_k}{H}\,dA. \tag{52}$$

Comparing (26) and (52) and using the results of Theorem 3, we conclude that M is part of an intrinsic sphere. □

As before, when ∂M lies in a hyperplane, the following conclusion holds.

Theorem 8. *Let $\mathcal{H}$ be a half-space in $\overline{M}^{n+1}(k)$ determined by a hyperplane P. Let M be a compact embedded constant-H_r ($r = 1, \dots, n$) hypersurface with free boundary in $\mathcal{H}$. Then M is an intrinsic hemisphere centered at a point of P.*

7. Constant-H_r/H_l immersed hypersurfaces with free boundary

Using the Minkowski formula and the inequalities for higher-order mean curvatures (Lemma 5), Koh and Lee [2001] gave characterizations of an intrinsic sphere in space forms. In Proposition 6, the Minkowski formula is extended to hypersurfaces with free boundary in space forms; then, Koh and Lee's results are naturally extended for hypersurfaces with free boundary. For the reader's convenience, we give the proof in detail.

Theorem 9. *Let $W \subset \overline{M}^{n+1}(k)$ be a wedge. Let M be a compact immersed hypersurface in $\overline{M}^{n+1}(k)$ with $\partial M \subset \partial W$ such that near ∂M, M lies inside of W and meets ∂W perpendicularly along ∂M. If, for $r, l = 1, \dots, n$ and $r > l$, the ratio H_r/H_l is constant and H_l does not vanish on M, then it is part of an intrinsic sphere centered at a point of the edge of W.*

Proof. For the case $k = 0$, by an isometry in $\mathbb{R}^{n+1}$, we assume that E contains the origin of $\mathbb{R}^{n+1}$. By the same argument as the proof of Theorem 7, there is an elliptic point q in M; that is, all the principal curvatures are positive, and clearly, both H_r and H_l are positive at q. Because $\alpha = H_r/H_l$ is constant and H_l does not vanish on M, the curvatures H_r, H_l are positive on M and $\alpha > 0$. By (i) of Lemma 5, $H_s > 0$ if $s < r$. By (ii) of Lemma 5,

$$0 < \alpha = \frac{H_r}{H_l} \le \frac{H_{r-1}}{H_{l-1}}. \tag{53}$$

Because $H_r = \alpha H_l$ and by (a) of Proposition 6,

$$\int_M H_{r-1} - \alpha H_l \langle \psi, N \rangle \, dA = 0. \tag{54}$$

Because $\alpha > 0$ is constant and by (a) of Proposition 6,

$$\int_M \alpha (H_{l-1} - H_l \langle \psi, N \rangle) \, dA = 0. \tag{55}$$

Combining (54) and (55) yields

$$\int_M (H_{r-1} - \alpha H_{l-1}) \, dA = 0.$$

From (53),

$$\frac{H_r}{H_l} = \frac{H_{r-1}}{H_{l-1}} = \alpha \quad \text{on } M.$$

Proceeding inductively, and defining $p = r - l$, we obtain

$$\frac{H_{p+1}}{H_1} = \frac{H_p}{H_0} = H_p \quad \text{on } M; \tag{56}$$

that is, $H_{p+1}/H_p = H_1$.

On the other hand, by (ii) of Lemma 5,

$$H_{p+1}/H_p \le H_p/H_{p-1} \le \cdots \le H_1. \tag{57}$$

Combining (56) and (57) gives,

$$H_{p+1}/H_p = H_p/H_{p-1} = \cdots = H_1,$$

and therefore,

$$H_r = H_1^r, \quad r = 1, 2, \ldots, p+1.$$

By (iii) of Lemma 5, M is part of an intrinsic sphere.

By an isometry in $\overline{M}^{n+1}(-1)$, we assume that E contains $v = (0, \ldots, 0, 1) \in \mathbb{L}^{n+2}$. As before there exists a point q such that all the principal curvatures are positive, and clearly, both H_r and H_l are positive at q. Because $\alpha = H_r/H_l$ is constant and H_l does not vanish on M, the curvatures H_r, H_l are positive on M and $\alpha > 0$.

By (ii) of Lemma 5,

$$0 < \alpha = \frac{H_r}{H_l} \le \frac{H_{r-1}}{H_{l-1}}. \tag{58}$$

Because $H_r = \alpha H_l$ and by Proposition 6,

$$\int_M H_{r-1}\langle \psi, v\rangle - \alpha H_l \langle N, v\rangle \, dA = 0. \tag{59}$$

Because $\alpha > 0$ is constant and by Proposition 6,

$$\int_M \alpha(H_{l-1}\langle \psi, v\rangle - H_l\langle N, v\rangle)\, dA = 0. \tag{60}$$

Combining (59) and (60) yields,

$$\int_M (H_{r-1} - \alpha H_{l-1})\langle \psi, v\rangle \, dA = 0.$$

Because $\langle \psi, v\rangle \le -1$ on M and by (58),

$$\frac{H_r}{H_l} = \frac{H_{r-1}}{H_{l-1}} = \alpha \quad \text{on } M.$$

Proceeding inductively, and defining $p = r - l$, we obtain

$$\frac{H_{p+1}}{H_1} = \frac{H_p}{H_0} = H_p \quad \text{on } M; \tag{61}$$

that is, $H_{p+1}/H_p = H_1$.

On the other hand, by (ii) of Lemma 5,

$$H_{p+1}/H_p \le H_p/H_{p-1} \le \cdots \le H_1. \tag{62}$$

Combining (61) and (62) gives,

$$H_{p+1}/H_p = H_p/H_{p-1} = \cdots = H_1,$$

and therefore,

$$H_r = H_1^r, \quad r = 1, 2, \ldots, p+1.$$

By (iii) of Lemma 5, M is part of an intrinsic sphere.

For the case $k = 1$, we assume that the edge E contains $v = (0, 0, \ldots, 1) \in \mathbb{R}^{n+2}$. Because $\psi : M \to \mathbb{S}_+^{n+1}$, we have $\langle \psi, v \rangle > 0$. By the same argument for $k = -1$, the conclusion follows as for the $k = 1$ case. □

Theorem 10. *Let P be a hyperplane in $\overline{M}^{n+1}(k)$. Let M be a compact immersed hypersurface in $\overline{M}^{n+1}(k)$ with $\partial M \subset P$ such that near ∂M, M lies on one side of P and meets P perpendicularly along ∂M. If, for $r, l = 1, \ldots, n$ and $r > l$, the ratio H_r/H_l is constant and H_l does not vanish on M, then it is an intrinsic hemisphere centered at a point of P.*

Acknowledgement

We warmly thank the referees for their careful reading of the paper and their valuable suggestions to improve the paper.

References

[Alexandrov 1962] A. D. Alexandrov, "A characteristic property of spheres", *Ann. Mat. Pura Appl.* (4) **58** (1962), 303–315. MR Zbl

[Barbosa and do Carmo 1984] J. L. Barbosa and M. do Carmo, "Stability of hypersurfaces with constant mean curvature", *Math. Z.* **185**:3 (1984), 339–353. MR Zbl

[Barbosa et al. 1988] J. L. Barbosa, M. do Carmo, and J. Eschenburg, "Stability of hypersurfaces of constant mean curvature in Riemannian manifolds", *Math. Z.* **197**:1 (1988), 123–138. MR Zbl

[Beckenbach and Bellman 1961] E. F. Beckenbach and R. Bellman, *Inequalities*, Ergebnisse der Mathematik **30**, Springer, 1961. MR Zbl

[Brendle 2013] S. Brendle, "Constant mean curvature surfaces in warped product manifolds", *Publ. Math. Inst. Hautes Études Sci.* **117** (2013), 247–269. MR Zbl

[Choe and Koiso 2016] J. Choe and M. Koiso, "Stable capillary hypersurfaces in a wedge", *Pacific J. Math.* **280**:1 (2016), 1–15. MR Zbl

[Choe and Park 2011] J. Choe and S.-H. Park, "Capillary surfaces in a convex cone", *Math. Z.* **267**:3-4 (2011), 875–886. MR Zbl

[Concus and Finn 1998] P. Concus and R. Finn, "Discontinuous behavior of liquids between parallel and tilted plates", *Phys. Fluids* **10**:1 (1998), 39–43. MR Zbl

[Concus et al. 2001] P. Concus, R. Finn, and J. McCuan, "Liquid bridges, edge blobs, and Scherk-type capillary surfaces", *Indiana Univ. Math. J.* **50**:1 (2001), 411–441. MR Zbl

[Finn 1986] R. Finn, *Equilibrium capillary surfaces*, Grundlehren der Math. Wissenschaften **284**, Springer, 1986. MR Zbl

[Koh 1998] S.-E. Koh, "A characterization of round spheres", *Proc. Amer. Math. Soc.* **126**:12 (1998), 3657–3660. MR Zbl

[Koh and Lee 2001] S.-E. Koh and S.-W. Lee, "Addendum to the paper: 'Sphere theorem by means of the ratio of mean curvature functions"', *Glasg. Math. J.* **43**:2 (2001), 275–276. Addendum to S.-E. Koh, **42**:1 (2000), 91–95. MR Zbl

[Korevaar 1988] N. J. Korevaar, "Sphere theorems via Alexandrov for constant Weingarten curvature hypersurfaces", (1988), 221–223. Appendix to A. Ros, "Compact hypersurfaces with constant scalar curvature and a congruence theorem", *J. Differential Geom.* **27**:2 (1988), 215–220. MR Zbl

[López 2014] R. López, "Capillary surfaces with free boundary in a wedge", *Adv. Math.* **262** (2014), 476–483. MR Zbl

[McCuan 1997] J. McCuan, "Symmetry via spherical reflection and spanning drops in a wedge", *Pacific J. Math.* **180**:2 (1997), 291–323. MR Zbl

[Montiel and Ros 1991] S. Montiel and A. Ros, "Compact hypersurfaces: the Alexandrov theorem for higher order mean curvatures", pp. 279–296 in *Differential geometry*, edited by B. Lawson and K. Tenenblat, Pitman Monogr. Surveys Pure Appl. Math. **52**, Longman Sci. Tech., Harlow, UK, 1991. MR Zbl

[Nitsche 1985] J. C. C. Nitsche, "Stationary partitioning of convex bodies", *Arch. Rational Mech. Anal.* **89**:1 (1985), 1–19. MR Zbl

[Park 2005] S.-h. Park, "Every ring type spanner in a wedge is spherical", *Math. Ann.* **332**:3 (2005), 475–482. MR Zbl

[Qiu and Xia 2015] G. Qiu and C. Xia, "A generalization of Reilly's formula and its applications to a new Heintze–Karcher type inequality", *Int. Math. Res. Not.* **2015**:17 (2015), 7608–7619. MR Zbl

[Reilly 1980] R. C. Reilly, "Geometric applications of the solvability of Neumann problems on a Riemannian manifold", *Arch. Rational Mech. Anal.* **75**:1 (1980), 23–29. MR Zbl

[Ros 1987] A. Ros, "Compact hypersurfaces with constant higher order mean curvatures", *Rev. Mat. Iberoamericana* **3**:3-4 (1987), 447–453. MR Zbl

[Ros 1988] A. Ros, "Compact hypersurfaces with constant scalar curvature and a congruence theorem", *J. Differential Geom.* **27**:2 (1988), 215–220. MR Zbl

[Ros and Vergasta 1995] A. Ros and E. Vergasta, "Stability for hypersurfaces of constant mean curvature with free boundary", *Geom. Dedicata* **56**:1 (1995), 19–33. MR Zbl

[Souam 1997] R. Souam, "On stability of stationary hypersurfaces for the partitioning problem for balls in space forms", *Math. Z.* **224**:2 (1997), 195–208. MR Zbl

[Spivak 1975] M. Spivak, *A comprehensive introduction to differential geometry, IV*, Publish or Perish, Boston, 1975. MR Zbl

Received December 25, 2016. Revised March 19, 2018.

JUNCHEOL PYO
jcpyo@pusan.ac.kr
DEPARTMENT OF MATHEMATICS
PUSAN NATIONAL UNIVERSITY
BUSAN
SOUTH KOREA

and

KOREA INSTITUTE FOR ADVANCED STUDY
SEOUL
SOUTH KOREA

CONTENTS

Volume 299, no. 1 and no. 2

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal’s preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 299 No. 2 April 2019
