The Logarithm Map, its Limits
and Fréchet Means in Orthant Spaces

D. Barden[*]   H. Le[†]

Abstract

The first part of the paper studies the expression for, and the properties of, the logarithm map on an orthant space, which is a simple stratified space, with the aim of analysing Fréchet means of probability measures on such a space. In the second part, we use these results to characterise Fréchet means and to derive various of their properties, including the limiting distribution of sample Fréchet means.

Keywords: Fréchet mean; limiting distribution of sample Fréchet means; logarithm map; stratified space.

AMS MSC 2010: 60B05; 60B10.

1   Introduction

Several papers have recently appeared concerning probabilistic and statistical analysis of data on certain stratified spaces (cf. [5], [2], [10], [1] and [11]). One such example is the analysis of phylogenetic trees on the BHV space introduced in [5] (cf. [9], [19], [17], [3], [12], [15] and [18]). The BHV space $\boldsymbol{T}_{m+2}$ of metric trees with $m + 2$ leaves is a stratified CAT(0)-space with each stratum being isometric with a positive Euclidean orthant that is at most m-dimensional. It is already clear from these preliminary results that some fundamental statistics exhibit strikingly different features from the corresponding ones on Euclidean spaces or on manifolds and that one faces significant challenges in developing novel tools to analyse them, on account of the non-trivial topological structure of these spaces. It also becomes apparent that, although the topological and geometrical properties of stratified spaces have been extensively studied and are mostly well understood, many of the properties required for probabilistic and statistical analysis of data on these spaces have not been addressed.

This paper concentrates on the orthant spaces introduced in [15], a relatively simple type of stratified space but more general than the space $\boldsymbol{T}_{m+2}$ of phylogenetic trees. The latter has $(2m + 1)!!$ m-dimensional strata, together with their bounding strata, selected from among the $\binom{M}{m}$ positive orthants in

---

[*]Girton College, University of Cambridge, Cambridge, CB3 0JG, UK (d.barden@dpmms.cam.ac.uk).

[†]School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK (huiling.le@nottingham.ac.uk).

$\mathbb{R}^M$ where $M = 2^{m+2} - m - 4$. In particular, each co-dimension one stratum bounds exactly three top-dimensional strata. Thus not only are the relevant dimensions sparse, but the percentage of the positive orthants occupied by the tree space of each dimension declines exponentially. These constraints, such as the restrictions on the dimension and the number of orthants involved in the space, no longer hold in a general orthant space, although we do have to make one restriction to ensure that it is a CAT(0)-space. We shall recall, in the next section, the concept of an orthant space, introducing the subsidiary concepts and definitions we use to describe the structure of such spaces and, in particular, of their tangent cones at the various points.

A fundamental concept for statistical analysis of non-Euclidean data is that of the Fréchet mean, which generalises the concept of the mean of Euclidean data. A point $\boldsymbol{x}_0$ in a metric space $\boldsymbol{M}$ is a Fréchet mean of a probability measure μ on $\boldsymbol{M}$ if, at $\boldsymbol{x}_0$, the Fréchet function of μ defined by

$$\frac{1}{2} \int_{\boldsymbol{M}} d(\boldsymbol{x}, \boldsymbol{x}')^2 \, \mathrm{d}\mu(\boldsymbol{x}') \tag{1}$$

attains its global minimum. In order to characterise and locate Fréchet means, we need to take directional derivatives of the Fréchet function and hence, implicitly, of the distance function. The latter involves the logarithm map $\log_{\boldsymbol{x}^*}(\boldsymbol{x})$ which, analogous to the inverse of the exponential map on manifolds, is the initial tangent vector to the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$. This logarithm map is globally well-defined on CAT(0)-spaces and has been studied, for example, in [14] and [16]. However, these results do not cover all the properties required for our analysis, although naturally we do rely on some of their results. On the other hand, an algorithm for finding the geodesic between any two given trees in the tree space $\boldsymbol{T}_{m+2}$ was given in [19] and, using the analysis behind that algorithm, the expression for the logarithm map $\log_{\boldsymbol{x}^*}$ was obtained in [3] when $\boldsymbol{x}^*$ lies in a top-dimensional stratum. Although this expression for $\log_{\boldsymbol{x}^*}$ could be extended to more general orthant spaces, it is noted in [3] that these results are not adequate to provide a tool for analysing Fréchet means when they lie in any stratum of co-dimension at least two. The latter requires a better understanding of the behaviour of the logarithm map as the end points of the geodesics move within and between strata. To this end, we first re-examine geodesics directly from first principles in Section 3, in particular avoiding the implicit assumption that $\boldsymbol{x}^*$ lies in a top-dimensional stratum. This leads, in Section 4, to an explicit expression, given in Theorem 1, for a version of the logarithm map that we shall use, valid for any point in an orthant space. Since the form of this expression is determined by the carrier of the geodesic, we analyse possible changes in that carrier, focussing on the set, specified in Definition 11, of points $\boldsymbol{x}$ at which significant changes occur. This allows us, in Section 5, to derive the directional limits of the logarithm map as the reference point $\boldsymbol{x}'$ approaches $\boldsymbol{x}^*$ from a co-bounding stratum. We also study the projections of these limits, and the limits of the projections, onto the various strata related to the stratum in which $\boldsymbol{x}^*$ lies. This enables us to prove the existence of, and to identify, certain of their derivatives and directional derivatives.

With this understanding of the logarithm map, the second part of the paper turns its attention to the analysis of Fréchet means. In Section 6 we obtain, in Theorem 3, the necessary and sufficient conditions for a point $\boldsymbol{x}^*$ to be the Fréchet mean of a probability measure on the orthant space $\boldsymbol{X}^m$. Two special

sets arise in this analysis. Firstly, one of the criteria in Theorem 3 involves an inequality and the set Θ, specified in Definition 13, of unit vectors in the tangent cone to $\boldsymbol{X}^m$ at the Fréchet mean for which that is an equality is significant. Secondly, there is the set Σ given by Definition 14. This is related to a limit of the logarithm map and, in a certain sense, encapsulates the 'departure' of this limit from the analogous behaviour of the logarithm map on a Euclidean space. Both of these sets are related to the limiting distribution of sample Fréchet means, which we establish in the final Section 7. There, in particular, we relate the limiting distribution with Euclidean Gaussian random variables. The covariance matrices of these random variables are related to the derivative of the projection of the logarithm map and to projections of the limits of the logarithm map.

Although we do not make it explicit, in view of our previous results for $\boldsymbol{T}_{m+2}$ and the comments in [15], our interest in this paper is primarily in the case that $\boldsymbol{x}^*$ lies in a stratum of local co-dimension at least two. The results, when restricted to a locally top-dimensional or co-dimension one stratum, do generalise those for tree spaces in [3] although the approach here is necessarily more complex in order to encompass all cases.

2 Orthant spaces

Throughout this paper, we shall use the term 'positive' to mean strictly positive. By an open positive orthant in the Euclidean space $\mathbb{R}^M$ we shall mean, for some subset $E = \left(u_{l_1}, \cdots, u_{l_m}\right)$ of the standard ordered orthonormal basis $U = \left(u_1, \cdots, u_M\right)$ of $\mathbb{R}^M$, the relatively open set

$$\mathcal{O}(E) = \left\{ \sum_{i=1}^m \lambda_i u_{l_i} \mid \lambda_i > 0 \right\}.$$

We denote by $\mathbb{R}(E)$ the subspace spanned by E, and we shall refer to the $u_{l_i} \in E$ as the axes of $\mathbb{R}(E)$ or of $\mathcal{O}(E)$. Then, an orthant space is a union of open positive orthants in a common Euclidean space with certain natural constraints, as specified in the following definition, that ensure, for example, that such spaces are also CAT(0). Orthant spaces were first introduced in [15] as a generalisation of the tree spaces of [5].

Definition 1. *For two given integers $M \geqslant m$, an orthant space $\boldsymbol{X}^m$ of dimension m is a subspace of the Euclidean space $\mathbb{R}^M$ that is a union of open positive orthants, whose maximum dimension is m, and has the intrinsic metric induced from the Euclidean metric on $\mathbb{R}^M$. It satisfies the following conditions:*

(i) *for every orthant σ in $\boldsymbol{X}^m$, the orthants in the closure $\overline{\sigma}$ of σ are also included in $\boldsymbol{X}^m$;*

(ii) *if, for any positive orthant σ in $\mathbb{R}^M$, all the 2-dimensional orthants in its closure are in $\boldsymbol{X}^m$, then σ itself is in $\boldsymbol{X}^m$.*

The intrinsic metric on $\boldsymbol{X}^m$ is the length metric as defined in [6]. It is the metric d for which, for any two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $\boldsymbol{X}^m$, the distance $d(\boldsymbol{x}_1, \boldsymbol{x}_2)$

is the infimum of the lengths of piecewise linear paths in $\boldsymbol{X}^m$ joining $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$. In particular, a geodesic will also be piecewise linear and linear within each stratum.

Note that there is no loss of generality in restricting $\boldsymbol{X}^m$ to contain only positive orthants: given two orthants that differ only in having positive or negative coordinates with respect to one particular axis, the intrinsic metric will be the same as it would be if we replace, say the negative axis, by an axis orthogonal to $\mathbb{R}^M$. Thus, rather than considering $\boldsymbol{X}^m$ to be a union of arbitrary orthants in $\mathbb{R}^M$, we could consider it to be a union of positive orthants in $\mathbb{R}^{2M}$. Henceforth, we shall assume all our orthants to be open and positive, mentioning their closure explicitly where that is relevant.

The first condition in the above definition correlates with the constraints used in the definition for orthant space in [15] and the second one restricts attention to the 'non-positively curved' orthant spaces in [15] (Proposition 6.10). These two conditions were first used by the authors of [5] to ensure the CAT(0)-property for tree spaces.

Throughout the rest of the paper, $\boldsymbol{X}^m$ will denote an orthant space of fixed dimension m viewed as comprising strata that are orthants of a fixed Euclidean space $\mathbb{R}^M$, where M is not necessarily $2^{m+2} - m - 4$ as it would be for tree space. Also, whenever we specify an orthant by a union of subsets of the standard orthonormal basis U of $\mathbb{R}^M$, that will always be intended as a union of mutually disjoint subsets.

The orthant space $\boldsymbol{X}^m$ so defined is a Whitney stratified set in the sense of Thom, [21], the strata being the various orthants that comprise $\boldsymbol{X}^m$. Note that, since $\boldsymbol{X}^m$ is a union of orthants in a fixed Euclidean space $\mathbb{R}^M$, the number of strata in $\boldsymbol{X}^m$ is always finite. $\boldsymbol{X}^m$ has the structure of a cone with vertex, or 'cone point', the origin o in $\mathbb{R}^M$, since each orthant is such a cone without its vertex, but that vertex, the origin, is necessarily included in $\boldsymbol{X}^m$. In particular, $\{o\}$ is the unique zero-dimensional stratum in $\boldsymbol{X}^m$. Note however that our relatively open strata differ from those in [6].

The CAT(0)-property of the orthant space $\boldsymbol{X}^m$ results as follows, where all the references are to [6]. The intersection L of $\boldsymbol{X}^m$ with the unit sphere in $\mathbb{R}^M$ is a simplicial complex on account of condition (i) and, since the axes in $\mathbb{R}^M$ are orthogonal, it is an 'all-right spherical complex' (Section 7A.10) which, on account of condition (ii), is a 'flag complex'. Then, by a theorem of Gromov (Theorem 5.18), L is a CAT(1)-space. The metric on $\boldsymbol{X}^m$ implied by describing it as the 0-cone over L (Definition 5.6) is the intrinsic metric so that, by the theorem of Berestowski (Theorem 3.14), $\boldsymbol{X}^m$ is CAT(0).

In particular, by the Cartan-Hadamard theorem (cf. [6], p.193), there is a unique geodesic between any two points of the orthant space $\boldsymbol{X}^m$. It follows that each stratum is totally geodesic in the strong sense that, if a geodesic contains two points of a stratum, it must include the entire linear segment in that stratum determined by those two points. On the other hand, although the distance metric for the CAT(0)-structure is induced from the Euclidean metric, the angles along and between curves may differ for the two contexts. For example, a geodesic, defined as a shortest path between its endpoints in either context, will be a piecewise linear curve in $\mathbb{R}^M$, linear in each stratum, with angle $\pi/2$ in the Euclidean subspace metric where it changes stratum. However, for the CAT(0)-structure, that angle is defined to be π.

The properties of an orthant space are largely determined by the incidence relations between its various strata. The following definitions capture two such relationships that will be used frequently in the paper.

Definition 2. *For subsets E and F of the standard orthonormal basis $U = (u_1, \cdots, u_M)$ of $\mathbb{R}^M$, if $E \subseteq F$, then the orthant $\mathcal{O}(E)$ is said to bound $\mathcal{O}(F)$ and $\mathcal{O}(F)$ to co-bound $\mathcal{O}(E)$.*

Note that, unlike the case for tree spaces, strata of lower dimension than m need not bound any higher dimensional strata, in particular they need not bound m-dimensional strata.

Definition 3. *An orthant σ of dimension k in $\boldsymbol{X}^m$ is said to have co-dimension $m - k$ and, if $m'(\leqslant m)$ is the maximum dimension of orthants that σ co-bounds, then σ is said to have local co-dimension $m' - k$.*

The tangent cone

It is natural for our purposes to follow [6] and to define the tangent cone to $\boldsymbol{X}^m$ at a point $\boldsymbol{x}$ to consist of all initial tangent vectors to smooth curves starting from $\boldsymbol{x}$, the smoothness possibly only being one-sided at $\boldsymbol{x}$. Note, however, that this is not the same as the generalised tangent space of [8]. To describe the tangent cone in more detail we work in $\mathbb{R}^M$. Then, when $\boldsymbol{x}$ lies in a top-dimensional, or locally top-dimensional, stratum σ of dimension $m'(\leqslant m)$, the orthant space $\boldsymbol{X}^m$ is locally an m'-dimensional manifold so that a smooth curve can be extended on both sides of $\boldsymbol{x}$. Thus, the tangent cone will be the usual tangent space, a subspace of $\mathbb{R}^M$ isometric with $\mathbb{R}^{m'}$ and tangent to σ. However, if $\boldsymbol{x}$ lies in a stratum of locally positive co-dimension, then the orthant space $\boldsymbol{X}^m$ is no longer locally a manifold. Consequently, the tangent cone at $\boldsymbol{x}$ is no longer a Euclidean space. For example, if the stratum σ has co-dimension one and bounds top-dimensional strata, the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x}$ is an open book: it has a closed half space $\mathbb{H}^m$ for each top-dimensional stratum τ co-bounding σ, with all the boundary $(m-1)$-dimensional faces identified with each other and with the tangent space to σ at $\boldsymbol{x}$.

More generally, the tangent cone at a point $\boldsymbol{x}$, in a stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l(\geqslant 1)$, has a topology and stratification imitating that of $\boldsymbol{X}^m$ itself in the neighbourhood of $\boldsymbol{x}$: for each stratum $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l' < l$ that co-bounds σ, so that F comprises the basis vectors that have positive coordinates in τ but zero coordinates in σ, there is the closed stratum $\mathbb{R}(E) \times \overline{\mathcal{O}(F)}$ in the tangent cone. Then, the tangent cone at $\boldsymbol{x}$ has its stratification determined by identifying the various $\mathbb{R}(E) \times \{\boldsymbol{0}\}$ with each other as well as identifying any tangent axes shared by pairs of strata that co-bound σ. In particular, when no strata co-bound σ, the tangent cone is simply the Euclidean space $\mathbb{R}(E)$.

Definition 4. *Let $\sigma = \mathcal{O}(E)$ and $\tau = \mathcal{O}(E \cup F)$ be two strata in $\boldsymbol{X}^m$ with co-dimensions l and $l' < l$, respectively. The component $\mathbb{R}(E)$ common to all the strata in the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x} \in \sigma$ is referred to as the tangent space to σ at $\boldsymbol{x}$. Vectors in the (open) stratum $\mathbb{R}(E) \times \mathcal{O}(F)$ of the tangent cone at $\boldsymbol{x} \in \sigma$ with non-zero second component are referred to as vectors tangent to τ at $\boldsymbol{x}$.*

The set of unit vectors in $\mathbb{R}(E) \times \mathcal{O}(F)$ is denoted by $\mathcal{S}_{\tau,\sigma}^{m-l'}$ and the subset of those in $\{\mathbf{0}\} \times \mathcal{O}(F)$ by $\mathcal{S}_{\tau \backslash \sigma}^{l-l'}$.

The sets $\mathcal{S}_{\tau,\sigma}^{m-l'}$ and $\mathcal{S}_{\tau \backslash \sigma}^{l-l'}$ are open spherical segments of dimensions $m-l'-1$ and $l-l'-1$ respectively, the latter lying in the space $\mathbb{R}(F)$ orthogonal to $\mathbb{R}(E)$.

Note that the basis vectors in E do not generally precede those of F in the standard ordered basis U, and so writing the stratum as $\mathbb{R}(E) \times \mathcal{O}(F)$ implies an appropriate permutation of the coordinates.

Definition 5. *For any subset E of the standard ordered orthonormal basis U of $\mathbb{R}^M$, where E does not necessarily inherit its order from U, we denote by $\jmath : \mathbb{R}(E) \to \mathbb{R}^M$ the linear transformation permuting coordinates and positioning them appropriately as coordinates, with respect to U, of a vector in $\mathbb{R}^M$.*

Note that $\jmath$ depends on the set E which will be clear from the context.

We are mainly interested in the restriction of $\jmath$ to subspaces of $\mathbb{R}(E)$. For example, if $E = (u_1, u_4)$ and $F = (u_2, u_6)$, then a point $(\boldsymbol{x}, \boldsymbol{y})$ in $\mathbb{R}(E) \times \mathcal{O}(F)$ with coordinates $((x_1, x_2), (y_1, y_2))$ would have $\jmath(\boldsymbol{x}, \boldsymbol{y}) = (x_1, y_1, 0, x_2, 0, y_2, 0, \cdots, 0)$ in $\mathbb{R}^M$.

Inherited from the CAT(0)-structure of $\boldsymbol{X}^m$, the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x}$, since it is metrically complete, also has a CAT(0)-structure (cf. [6], Theorem 3.19). While the CAT(0)-metric on $\boldsymbol{X}^m$ is, by definition, the intrinsic metric, the CAT(0)-metric on the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x}$ is defined in terms of the Alexandrov angle. Recall that, for any three points $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2$ in $\boldsymbol{X}^m$, the comparison triangle of the geodesic triangle $\Delta(\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2)$ in $\boldsymbol{X}^m$ formed by $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2$ is the triangle $\bar{\Delta}(\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2)$ in the Euclidean plane with vertices $\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2$ such that the Euclidean distances $d(\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}}_1)$ etc. match the intrinsic distances $d(\boldsymbol{x}, \boldsymbol{x}_1)$ etc. in $\boldsymbol{X}^m$. Then, the Alexandrov angle $\angle_{\boldsymbol{x}}(\gamma_1, \gamma_2)$ between the geodesics γ_1 and γ_2 starting from $\boldsymbol{x}$ is defined to be

$$\angle_{\boldsymbol{x}}(\gamma_1, \gamma_2) = \lim_{t \to 0} \overline{\angle}_{\boldsymbol{x}}(\gamma_1(t), \gamma_2(t)),$$

where $\overline{\angle}_{\boldsymbol{x}}(\gamma_1(t), \gamma_2(t))$ is the Euclidean angle at $\bar{\boldsymbol{x}}$ of the comparison Euclidean triangle $\bar{\Delta}(\boldsymbol{x}, \gamma_1(t), \gamma_2(t))$ (cf. [6], Section 1.12). Note that, since geodesics in $\boldsymbol{X}^m$ are piecewise linear, the above limit is well-defined. Then, the inner product on the tangent cone of $\boldsymbol{X}^m$ at $\boldsymbol{x}$ is defined by

$$\ll \boldsymbol{w}_1, \boldsymbol{w}_2 \gg = \| \boldsymbol{w}_1 \| \| \boldsymbol{w}_2 \| \cos \angle_{\boldsymbol{x}}(\gamma_1, \gamma_2), \qquad (2)$$

where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are the initial tangent vectors of γ_1 and γ_2. By analogy with vectors in the tangent space to a manifold, the distance $\rho_{\boldsymbol{x}}(\boldsymbol{w}_1, \boldsymbol{w}_2)$ between vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ in the tangent cone at $\boldsymbol{x}$ is defined to be

$$\rho_{\boldsymbol{x}}(\boldsymbol{w}_1, \boldsymbol{w}_2) = \left\{ \| \boldsymbol{w}_1 \|^2 + \| \boldsymbol{w}_2 \|^2 - 2 \ll \boldsymbol{w}_1, \boldsymbol{w}_2 \gg \right\}^{1/2}$$

(cf. [16], p144). Note that, although in general $\ll , \gg$ differs from the usual Euclidean inner product $\langle , \rangle$, a geodesic triangle contained in the closure of a stratum of $\boldsymbol{X}^m$ is in fact a Euclidean geodesic triangle and its angles are the Euclidean ones. In particular, $\ll \boldsymbol{w}_1, \boldsymbol{w}_2 \gg = \langle \boldsymbol{w}_1, \boldsymbol{w}_2 \rangle$ for any $\boldsymbol{w}_1, \boldsymbol{w}_2$ in the closure of $\mathbb{R}(E) \times \mathcal{O}(F)$ and then $\rho_{\boldsymbol{x}}(\boldsymbol{w}_1, \boldsymbol{w}_2) = \| \boldsymbol{w}_1 - \boldsymbol{w}_2 \|$.

3 The carriers and supports of geodesics

In order to analyse the logarithm map, we first need to understand the geodesics. The intersection of a geodesic with a stratum, a Euclidean orthant, will be either a single point or a complete intersection of a Euclidean line with that orthant.

Definition 6. *The carrier of a geodesic is the sequence of strata each of whose intersection with the geodesic is a Euclidean line of positive length.*

This is essentially the terminology that was introduced in [22] in the context of tree spaces. The case of a single point intersection arises between successive strata of the carrier: between the (open) linear segment in one stratum and that in the next, there will be one point in the common bounding stratum of those two strata. This intermediate stratum is not listed in the carrier; it is in fact specified by the adjacent strata as the stratum of highest dimension in the intersection of their closures. Similarly, when a geodesic starts, or ends, in a stratum of positive co-dimension and does not remain in that stratum, but passes immediately to a co-bounding stratum, then the latter will be the first, or last, stratum in the carrier. In such a situation, we shall regard the point in the bounding stratum as having the same set of axes as the co-bounding stratum, albeit with the relevant coordinates zero. That is, we regard it as a point of the closure of the co-bounding stratum.

To describe the carrier of the geodesic from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ in more detail, as well as for later analysis, we require the following terminology.

Definition 7. (i) *The subsets E and F of U are said to be compatible in the orthant space $\boldsymbol{X}^m$ if the orthant $\mathcal{O}(E \cup F)$ is contained in $\boldsymbol{X}^m$.*

(ii) *For a subset E of the standard orthonormal basis U of $\mathbb{R}^M$, we denote the number of vectors in E by $|E|$.*

We first identify the set of axes common to the points along a geodesic where, for any $\boldsymbol{x} \in \boldsymbol{X}^m$, $E(\boldsymbol{x})$ denotes the set of axes in U with respect to which $\boldsymbol{x}$ has positive coordinates.

Proposition 1. *For any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{X}^m$, the set $E(\boldsymbol{x}_1, \boldsymbol{x}_2)$, defined by*

$$
\begin{aligned}
E(\boldsymbol{x}_1, \boldsymbol{x}_2) \quad = \quad & \{E(\boldsymbol{x}_1) \cap E(\boldsymbol{x}_2)\} \\
& \bigcup \{e \in E(\boldsymbol{x}_1) \mid e \text{ is compatible with } E(\boldsymbol{x}_2)\} \\
& \bigcup \{e \in E(\boldsymbol{x}_2) \mid e \text{ is compatible with } E(\boldsymbol{x}_1)\},
\end{aligned}
\tag{3}
$$

forms the set of axes common to all strata along the geodesic between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

Proof. Observe that, for a geodesic, each coordinate function must be linearly interpolated between any two values that are non-zero. It follows that once a particular coordinate, having been positive along the geodesic, becomes zero it must remain so or, having started at zero, once it becomes positive, it must continue monotonically to its final value. In particular, the only basis vectors that can occur with positive coordinate at any point along the geodesic from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ are those that belong to $\boldsymbol{x}_1$ or $\boldsymbol{x}_2$ or to both. Moreover, when the geodesic from $\boldsymbol{x}_1$ passes immediately to a co-bounding stratum, all the new axes in that stratum must have coordinate zero at $\boldsymbol{x}_1$ increasing linearly along the geodesic

to its value at $\boldsymbol{x}_2$. Any such additional axis e of the co-bounding stratum is in $E(\boldsymbol{x}_2)$ and is compatible with all the axes in $E(\boldsymbol{x}_1)$; and any such e must occur in this way. Thus, the set of axes common to all strata along the geodesic from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ is precisely the given set $E(\boldsymbol{x}_1, \boldsymbol{x}_2)$. $\qquad\square$

Note that, at one extreme, if $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ both lie in the closure of an orthant $\mathcal{O}(E)$ and not both in the same boundary component, then $E(\boldsymbol{x}_1, \boldsymbol{x}_2) = E$. At the other extreme, if the intersection $\overline{\mathcal{O}(E(\boldsymbol{x}_1))} \cap \overline{\mathcal{O}(E(\boldsymbol{x}_2))}$ comprises only the cone point, then $E(\boldsymbol{x}_1, \boldsymbol{x}_2) = \emptyset$. In general, $E(\boldsymbol{x}_1, \boldsymbol{x}_2)$ depends only on the orthants in which $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ lie, and is independent of their positions in those orthants.

The number $k + 1$ of orthants in the carrier $\mathcal{C} = (\mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_k)$ of the geodesic from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ will, naturally, depend on both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. If $\boldsymbol{x}_1$ lies in a top dimensional stratum it will have m strictly positive coordinates, all of which, assuming that none are also positive in $\boldsymbol{x}_2$, must become zero somewhere along the geodesic and at least one must become zero on each change of stratum as they cannot vanish within a stratum of the carrier. Thus, there will be $m + 1$ strata in the carrier, that is $k = m$, if and only if they vanish one at a time. So, $k < m$ if and only if somewhere along the geodesic at least two coordinates become zero on passing from $\mathcal{O}_i$ to $\mathcal{O}_{i+1}$. When $|E(\boldsymbol{x}_1, \boldsymbol{x}_2)| = k_0$, the maximum value of k would now be $k' = m - k_0$. Similarly, if $\boldsymbol{x}_1$ were in a stratum of dimension m_0, this maximum would be $m_0 - k_0$.

From now on, for given $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, we shall denote the set $E(\boldsymbol{x}_1, \boldsymbol{x}_2)$ by both A_0 and B_0 to accord with the following notation. It follows from Proposition 1 that each member of the sequence of strata $\mathcal{C} = (\mathcal{O}_0, \mathcal{O}_1, \cdots, \mathcal{O}_k)$ that comprise the carrier of the geodesic γ from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ has $\mathcal{O}(A_0) = \mathcal{O}(B_0)$ as a factor. The carrier of γ determines further subsets of axes forming two sequences $(A_1, \cdots, A_k)$ and $(B_1, \cdots, B_k)$, where A_i is the set of all the axes whose coordinates become zero and B_i the set of all those whose coordinates become positive as the geodesic passes from $\mathcal{O}_{i-1}$ to $\mathcal{O}_i$. Thus, the stratum $\mathcal{O}_{i-1}$ is $\mathcal{O}(B_0 \cup B_1 \cup \cdots \cup B_{i-1} \cup A_i \cup \cdots \cup A_k)$ and $\mathcal{O}_i = \mathcal{O}(B_0 \cup B_1 \cup \cdots \cup B_i \cup A_{i+1} \cup \cdots \cup A_k)$, with $\mathcal{O}_0$ determined by $A_0 \cup A_1 \cup \cdots \cup A_k$. Clearly, the intermediate stratum between $\mathcal{O}_{i-1}$ and $\mathcal{O}_i$, their common boundary component, is

$$\mathcal{O}(B_0 \cup B_1 \cup \cdots \cup B_{i-1} \cup A_{i+1} \cup \cdots \cup A_k).$$

Thus, in particular,

(a) *the sets B_i and A_j of axes are non-empty for all positive i and j, and compatible in $\boldsymbol{X}^m$ for $0 \leqslant i < j$;*

(b) *γ passes successively with positive length through the orthants $\mathcal{O}_i$ except that it may meet at most one of $\mathcal{O}_0$ and $\mathcal{O}_k$ in a single point;*

(c) *$A_i \cap A_j = \emptyset$ and $B_i \cap B_j = \emptyset$ for all $i \neq j$.*

The property (c) follows from the facts that $A_1 \cup \cdots \cup A_k$ is disjoint from $B_1 \cup \cdots \cup B_k$ and that an axis once removed cannot be removed again, or once introduced cannot be introduced again.

Definition 8. *For any two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $\boldsymbol{X}^m$, the support of the geodesic γ from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ is defined to be the pair $(\mathcal{A}, \mathcal{B})$ of sequences of sets of axes,*

$$\mathcal{A} = (A_0, A_1, \cdots, A_k) \quad \text{and} \quad \mathcal{B} = (B_0, B_1, \cdots, B_k),$$

where γ passes successively through the orthants

$$\mathcal{O}_i = \mathcal{O}(A_0 \cup B_1 \cup \cdots \cup B_i \cup A_{i+1} \cup \cdots \cup A_k), \qquad i = 0, 1, \cdots, k, \qquad (4)$$

that form the carrier of γ.

In the context of tree spaces, the definition of the support of a geodesic given here is equivalent to that of the minimal support given in [15].

Example 1. For a geodesic passing successively through the orthants

$$
\begin{aligned}
\mathcal{O}_0 &= \mathcal{O}(e_0, e_1, e_2, e_3, e_4, e_5, e_6), \\
\mathcal{O}_1 &= \mathcal{O}(e_0, e_1, f_2, e_3, e_4, e_5, e_6), \\
\mathcal{O}_2 &= \mathcal{O}(e_0, e_1, f_2, f_3, e_5, e_6), \\
\mathcal{O}_3 &= \mathcal{O}(e_0, e_1, f_2, f_3, f_4, e_6), \\
\mathcal{O}_4 &= \mathcal{O}(e_0, e_1, f_2, f_3, f_4, f_5, f_6),
\end{aligned}
$$

the relevant sequences $\mathcal{A} = (A_0, A_1, \cdots, A_4)$ and $\mathcal{B} = (B_0, B_1, \cdots, B_4)$ forming the support would have members $A_0 = B_0 = \{e_0, e_1\}$, the basis vectors common to all five orthants; $A_1 = \{e_2\}$, $B_1 = \{f_2\}$; $A_2 = \{e_3, e_4\}$, $B_2 = \{f_3\}$; $A_3 = \{e_5\}$, $B_3 = \{f_4\}$; $A_4 = \{e_6\}$ and $B_4 = \{f_5, f_6\}$.

If both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ lie in the closure of the same orthant, then the geodesic between them is clearly the Euclidean line segment. To understand geodesics in general and, later, to describe and analyse various properties of the logarithm map, we require projections onto the various strata of $\boldsymbol{X}^m$.

Definition 9. *For $\boldsymbol{x} \in \boldsymbol{X}^m$ and $E \subset U$ such that the orthant $\sigma = \mathcal{O}(E)$ is contained in $\boldsymbol{X}^m$, $P_E(\boldsymbol{x})$ denotes the projection, orthogonal with respect to the Euclidean inner product on $\mathbb{R}^M$, of $\boldsymbol{x}$ onto $\mathcal{O}(E)$, that is the vector, or when relevant its coordinate vector, formed by the components of $\boldsymbol{x}$ in the directions of the unit vectors in E.*

In terms of projections, we have the following characterisation of the supports of geodesics when $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ do not lie in the closure of the same orthant.

Proposition 2. *Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be two given points in $\boldsymbol{X}^m$. Suppose that $\mathcal{A} = (A_0, A_1, \cdots, A_k)$ and $\mathcal{B} = (B_0, B_1, \cdots, B_k)$ are two sequences of sets of axes such that the $\mathcal{O}_i$ defined by (4) are all contained in $\boldsymbol{X}^m$, where $k > 0$ and where all subsets A_i and B_j are mutually disjoint and non-empty, except for $A_0 = B_0 = E(\boldsymbol{x}_1, \boldsymbol{x}_2)$ which may be empty. Then, $(\mathcal{A}, \mathcal{B})$ is the support of the geodesic from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ if and only if*

(i) *for $k > 1$ and for all $0 < i < k$,*

$$\frac{\|P_{A_i}(\boldsymbol{x}_1)\|}{\|P_{B_i}(\boldsymbol{x}_2)\|} < \frac{\|P_{A_{i+1}}(\boldsymbol{x}_1)\|}{\|P_{B_{i+1}}(\boldsymbol{x}_2)\|}; \qquad (5)$$

(ii) *for all $0 < i \leqslant k$ and all non-trivial partitions $C_{i1} \cup C_{i2}$ for A_i and $D_{i1} \cup D_{i2}$ for B_i, if the orthant*

$$\mathcal{O}' = \mathcal{O}(B_0 \cup B_1 \cup \cdots \cup B_{i-1} \cup D_{i1} \cup C_{i2} \cup A_{i+1} \cup \cdots \cup A_k) \qquad (6)$$

is contained in $\boldsymbol{X}^m$, then

$$\frac{\|P_{C_{i1}}(\boldsymbol{x}_1)\|}{\|P_{D_{i1}}(\boldsymbol{x}_2)\|} \geqslant \frac{\|P_{C_{i2}}(\boldsymbol{x}_1)\|}{\|P_{D_{i2}}(\boldsymbol{x}_2)\|}. \qquad (7)$$
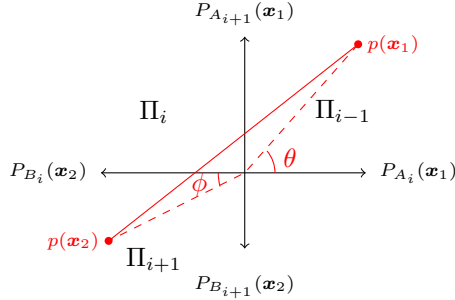
Figure 1: Projection of the geodesic

Compared with the result in [20] (Theorem 2.5) in the case of tree spaces, this result confirms the claim in Section 6 of [15] that the results on tree spaces also hold for orthant spaces. However, the condition (ii) above is necessarily stronger than that there. This is due to the fact that, in general orthant spaces, the condition that C_{i2} is compatible with D_{i1} does not necessarily guarantee that the orthant $\mathcal{O}'$ given by (6) is contained in $\boldsymbol{X}^m$.

Proof. Assuming that $(\mathcal{A}, \mathcal{B})$ is the support of the geodesic from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$, we focus on three consecutive strata of the carrier, $\mathcal{O}_{i-1}, \mathcal{O}_i, \mathcal{O}_{i+1}$, where $\mathcal{O}_j$ are as defined in (4), projecting the geodesic onto the subspace $\mathbb{R}(A_i \cup A_{i+1} \cup B_i \cup B_{i+1})$. As the geodesic passes from $\mathcal{O}_{i-1}$ to $\mathcal{O}_i$, the coordinates along the axes in A_i become zero and those in B_i start to grow. Then, on passing from $\mathcal{O}_i$ to $\mathcal{O}_{i+1}$, the coordinates of axes in A_{i+1} become zero and those in B_{i+1} grow. Consider the projection of the geodesic onto the three planar quadrants Π_{i-1} determined by the vectors $P_{A_i}(\boldsymbol{x}_1)$ and $P_{A_{i+1}}(\boldsymbol{x}_1)$, Π_i determined by $P_{A_{i+1}}(\boldsymbol{x}_1)$ and $P_{B_i}(\boldsymbol{x}_2)$ and Π_{i+1} determined by $P_{B_i}(\boldsymbol{x}_2)$ and $P_{B_{i+1}}(\boldsymbol{x}_2)$ as in Figure 1. This is an isometric representation of the relevant quadrants except that, in $\mathbb{R}^M$, all four vectors are mutually orthogonal. Then, $\mathcal{O}_i$ is in the carrier if and only if the projection of the geodesic passes through the interior of Π_i. That is if, and only if, the angle θ that the vector $p(\boldsymbol{x}_1) = P_{A_i \cup A_{i+1}}(\boldsymbol{x}_1)$ makes with $P_{A_i}(\boldsymbol{x}_1)$ in Π_{i-1} is greater than the angle ϕ that $p(\boldsymbol{x}_2) = P_{B_i \cup B_{i+1}}(\boldsymbol{x}_2)$ makes with $P_{B_i}(\boldsymbol{x}_2)$ in Π_{i+1}, as expressed by (5).

Similarly, if $\mathcal{O}'$ is contained in $\boldsymbol{X}^m$, the failure of (7) would ensure that the geodesic passed through $\mathcal{O}'$, with positive length, between $\mathcal{O}_{i-1}$ and $\mathcal{O}_i$.

To show that conditions (i) and (ii) determine the support of the geodesic from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$, we first note that, as seen above, (i) ensures that the geodesic must pass through the orthant $\mathcal{O}_i$ between $\mathcal{O}_{i-1}$ and $\mathcal{O}_{i+1}$. Since $\boldsymbol{X}^m$ is a cone, it is simply connected and any piecewise linear path from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ can be transformed by homotopy to a geodesic by a sequence of 'simple moves' whereby, for each move, two consecutive linear segments of the path are replaced by a single linear segment. Since the geodesic is linear within orthants, that can only occur between consecutive orthants and condition (ii) guarantees that there is no extra orthant in the carrier between $\mathcal{O}_{i-1}$ and $\mathcal{O}_i$. $\square$

As noted previously, if $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ both lie in the closure of an orthant, then $k = 0$ and the geodesic between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is always a Euclidean segment.

Then, when $\boldsymbol{x}_2$ varies within the orthant in which it lies, the support of the geodesic from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ remains the same. However, in general, the support may change. As will become clear in the next section, the changes in the supports of geodesics give rise to the major problems in adapting the theory for manifolds to our more general context. Nevertheless, the above characterisation of the support of a geodesic implies the following sufficient condition for the support to remain locally constant. This is related to the smoothness of the logarithm map to be studied in the following section.

Corollary 1. *Suppose that the hypotheses of Proposition 2 hold. If, for all $0 < i \leqslant k$ and for all relevant partitions of A_i and B_i as in (ii) of that proposition, the inequality (7) is strict then, for all $\boldsymbol{x}$ in a sufficiently small neighbourhood of $\boldsymbol{x}_2$ in its stratum, $(\mathcal{A}, \mathcal{B})$ remains the support for the geodesic from $\boldsymbol{x}_1$ to $\boldsymbol{x}$.*

Proof. Since $\boldsymbol{x}$ varies within the stratum in which $\boldsymbol{x}_2$ lies, the set $A_0 = B_0$ remains unchanged. For the other sets in the support, by continuity, the strict inequalities (5) and, we are assuming, (7) continue to hold for $\boldsymbol{x}$ in a sufficiently small neighbourhood of $\boldsymbol{x}_2$ within its stratum. Hence, the required result follows from Proposition 2. $\square$

4 The logarithm map

Analogous to an inverse of the exponential map on a Riemannian manifold, the logarithm map on $\boldsymbol{X}^m$ is defined as follows.

Definition 10. *The logarithm map at $\boldsymbol{x}^* \in \boldsymbol{X}^m$ is the map $\log_{\boldsymbol{x}^*}(\boldsymbol{x})$ from $\boldsymbol{X}^m$ to the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x}^*$, the image of $\boldsymbol{x}$ being the initial tangent vector, with norm $d(\boldsymbol{x}^*, \boldsymbol{x})$, to the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$.*

The logarithm map is globally well-defined since, as already mentioned, the Cartan-Hadamard theorem implies that there is a unique geodesic between any two points $\boldsymbol{x}^*$ and $\boldsymbol{x}$ of $\boldsymbol{X}^m$. If that geodesic has an initial segment in a stratum containing $\boldsymbol{x}^*$ it will certainly have an initial tangent vector. If it has only $\boldsymbol{x}^*$ in the initial stratum, it must then have an open segment $\gamma(0, \epsilon)$, with $\gamma(0) = \boldsymbol{x}^*$, in a co-bounding stratum. Then it will still have a one-sided derivative at $\boldsymbol{x}^*$ which suffices to define the logarithm map.

With the description of the carrier, as well as the results on the support, of a geodesic in the previous section, we are now in a position to derive and analyse its initial tangent vector, or equivalently $\log_{\boldsymbol{x}^*}(\boldsymbol{x})$ for $\boldsymbol{x}^* \in \sigma = \mathcal{O}(E)$. As in [2] and [3] for the space of trees, our analysis will mainly involve a modified version of the logarithm map. For this, since the tangent cones at various points in σ are all parallel, we may parallel translate them to the cone point o, the origin in $\mathbb{R}^M$, to produce a common isometric copy $\boldsymbol{C}_\sigma$. Then, since the coordinate vector of the point $\boldsymbol{x}^*$, which we also denote by $\boldsymbol{x}^*$, lies in the common factor $\mathbb{R}(E)$ of all the strata of $\boldsymbol{C}_\sigma$, it makes sense to add it to $\log_{\boldsymbol{x}^*}(\boldsymbol{x})$ and the result

$$\Phi(\boldsymbol{x}; \boldsymbol{x}^*) = \log_{\boldsymbol{x}^*}(\boldsymbol{x}) + \boldsymbol{x}^*$$

will also lie in $\boldsymbol{C}_\sigma$. We shall refer to Φ as the *translated logarithm map* to distinguish it from the logarithm map itself. All the vectors $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ being in the same space implies that the translated logarithm maps are directly comparable

as $\boldsymbol{x}^*$ varies within an orthant and such comparability will be necessary later. Moreover, the difference between the two maps is such that all our analysis of Φ can easily be translated to that of the logarithm map itself.

Note that, although the origin corresponds to the cone point o of the orthant space $\boldsymbol{X}^m$, $\boldsymbol{C}_\sigma$ is not the tangent cone to $\boldsymbol{X}^m$ at o, neither being contained in the other, unless $\sigma = \{o\}$. Note also that, when $\boldsymbol{X}^m$ is a tree space and $\boldsymbol{x}^*$ lies in a top-dimensional stratum, $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ was called the modified logarithm map and was denoted by $\Phi_{\boldsymbol{x}^*}(\boldsymbol{x})$ in [3], and the permutation map π there corresponds to the linear transformation $\jmath$ given by Definition 5.

The next theorem gives the expression for the translated logarithm map $\Phi(\,\cdot\,; \boldsymbol{x}^*)$ in terms of the projections, specified in Definition 9, onto various sets of axes appearing in the support of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$.

Theorem 1. *For any two points $\boldsymbol{x}^*$ and $\boldsymbol{x}$ in $\boldsymbol{X}^m$, let the sequences $\mathcal{A} = (A_0, \cdots, A_k)$ and $\mathcal{B} = (B_0, \cdots, B_k)$ of sets of axes form the support of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$. Then, the translated logarithm map $\Phi(\,\cdot\,; \boldsymbol{x}^*)$ at $\boldsymbol{x}^*$ is given by*

$$\Phi(\boldsymbol{x}; \boldsymbol{x}^*) = \jmath\left(P_{B_0}(\boldsymbol{x}), -\frac{\|P_{B_1}(\boldsymbol{x})\|}{\|P_{A_1}(\boldsymbol{x}^*)\|}P_{A_1}(\boldsymbol{x}^*), \cdots, -\frac{\|P_{B_k}(\boldsymbol{x})\|}{\|P_{A_k}(\boldsymbol{x}^*)\|}P_{A_k}(\boldsymbol{x}^*)\right), \quad (8)$$

where $\jmath$ is the linear transformation given by Definition 5.

In particular, $\Phi(\,\cdot\,; \lambda\boldsymbol{x}^*) = \Phi(\,\cdot\,; \boldsymbol{x}^*)$ for any constant $\lambda > 0$.

Recall that, if $k = 0$, then $\boldsymbol{x}$ and $\boldsymbol{x}^*$ lie the closure of an orthant and the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$ is a line segment in $\mathbb{R}^M$. In this case, $\Phi(\boldsymbol{x}; \boldsymbol{x}^*) = \jmath(\boldsymbol{x})$. If $k > 0$ and if $|A_i| = |B_i| = 1$ for $1 \leqslant i \leqslant k$, the form of the expression for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ is also similar to that of the corresponding translated logarithm map in a Euclidean space, after changing the axes B_i to $-A_i$.

Proof. The orthogonal projection of γ onto $\mathcal{O}(A_0)$ determines the component of the initial tangent vector to γ that is tangent to $\mathcal{O}(A_0)$, namely

$$v_0 = P_{B_0}(\boldsymbol{x}) - P_{A_0}(\boldsymbol{x}^*) \in \mathbb{R}(A_0). \quad (9)$$

For the remaining coordinates, since the sets A_i and B_j above are all mutually disjoint, it follows that, for each i, the subspace $\mathbb{R}(A_i \cup B_i)$ is orthogonal to all $\mathbb{R}(A_j)$ and $\mathbb{R}(B_j)$ for $j \neq i$, so that the coordinates of the geodesic γ that are positive with respect to the axes in $\mathbb{R}(A_i \cup B_i)$ are just those of the projection γ_i of γ onto that subspace. If s_i is the parameter such that $\gamma(s_i) \in \mathcal{O}_{i-1} \cap \mathcal{O}_i$, then $P_{A_i}(\gamma(s)) \in \mathcal{O}(A_i)$ declines linearly from $P_{A_i}(\gamma(0)) = P_{A_i}(\boldsymbol{x}^*)$ to $P_{A_i}(\gamma(s_i)) = \boldsymbol{0}$. Then, the coordinates $P_{B_i}(\gamma(s)) \in \mathcal{O}(B_i)$ increase linearly from zero at $\gamma(s_i)$ to $P_{B_i}(\gamma(1)) = P_{B_i}(\boldsymbol{x})$. Thus, the projected geodesic γ_i lies in the union of the orthogonal orthants $\mathcal{O}(A_i)$ and $\mathcal{O}(B_i)$ and hence has length $\|P_{A_i}(\boldsymbol{x}^*)\| + \|P_{B_i}(\boldsymbol{x})\|$. The initial tangent vector to γ_i is parallel to $-P_{A_i}(\boldsymbol{x})$ and so is

$$v_i = -\frac{\|P_{A_i}(\boldsymbol{x}^*)\| + \|P_{B_i}(\boldsymbol{x})\|}{\|P_{A_i}(\boldsymbol{x}^*)\|}P_{A_i}(\boldsymbol{x}^*).$$

Hence, the initial tangent vector to γ with norm $d(\boldsymbol{x}^*, \boldsymbol{x})$ is represented by $(v_0, v_1, \cdots, v_k)$. However, this ordering of the coordinates, with those in $\mathbb{R}(A_i)$ preceding those of $\mathbb{R}(A_{i+1})$ for each i, requires the linear transformation $\jmath$ to
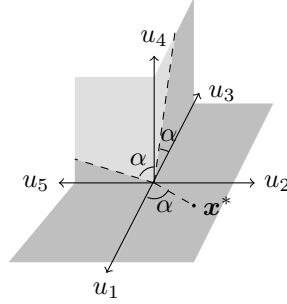
Figure 2: $\boldsymbol{X}^2$ in $\mathbb{R}^5$ consisting of five orthants

obtain its representation with respect to the standard basis in $\mathbb{R}^M$. Then, the logarithm map at $\boldsymbol{x}^*$ will be

$$\log_{\boldsymbol{x}^*} : \boldsymbol{x} \mapsto \jmath\left(v_0, v_1, \cdots, v_k\right),$$

so that equation (8) follows from the coordinates v_i since the coordinates of $\boldsymbol{x}^*$ are $\jmath\left(P_{A_0}(\boldsymbol{x}^*), P_{A_1}(\boldsymbol{x}^*), \cdots, P_{A_k}(\boldsymbol{x}^*)\right)$. $\qquad\square$

Clearly, the log, or translated log, map depends on the support $(\mathcal{A}, \mathcal{B})$ of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$, noting that the roles that $\mathcal{A}$ and $\mathcal{B}$ play are not symmetric. The final algebraic expression (8) for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$, which we shall refer to as its form, will also depend on the transformation $\jmath$. This form may change as $\boldsymbol{x}$ varies within an orthant. The following example illustrates this feature in the space $\boldsymbol{X}^2$ in $\mathbb{R}^5$, which was called Q_5 in [2], consisting of the five orthants depicted in Figure 2, where all five axes are mutually orthogonal. Although $\boldsymbol{x}$ lies in the same orthant in the second and third cases, the forms for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$, as a function of $\boldsymbol{x}$, differ in the two cases. However, along the boundary between the light and dark grey regions, the two forms give the same result.

Example 2. The tangent cone to $\boldsymbol{X}^2$ at $\boldsymbol{x}^* = (x_1^*, x_2^*, 0, 0, 0)$ indicated in Figure 2 is the (u_1, u_2)-plane and that at the cone point o is $\boldsymbol{X}^2$ itself. While $\Phi(\boldsymbol{x}; o) = \boldsymbol{x}$ for all $\boldsymbol{x}$, the expression for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ takes different forms depending the position of $\boldsymbol{x}$. For example, for any $\boldsymbol{x} = (0, x_2, x_3, 0, 0)$ in the orthant $\mathcal{O}(u_2, u_3)$,

$$\Phi(\boldsymbol{x}; \boldsymbol{x}^*) = (-x_3, x_2, 0, 0, 0);$$

for $\boldsymbol{x} = (0, 0, x_3, x_4, 0)$ in the dark grey region of $\mathcal{O}(u_3, u_4)$, i.e. if the coordinates of $\boldsymbol{x}$ satisfy $x_4/x_3 < \tan(\alpha) = x_2^*/x_1^*$, then

$$\Phi(\boldsymbol{x}, \boldsymbol{x}^*) = (-x_3, -x_4, 0, 0, 0);$$

However, if $\boldsymbol{x} = (0, 0, x_3, x_4, 0)$ lies in the light grey region of $\mathcal{O}(u_3, u_4)$, i.e. if the coordinates of $\boldsymbol{x}$ satisfy $x_4/x_3 > \tan(\alpha) = x_2^*/x_1^*$, then

$$\Phi(\boldsymbol{x}, \boldsymbol{x}^*) = -\frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}^*\|}(x_1^*, x_2^*, 0, 0, 0).$$

In particular, for all $\boldsymbol{x}$ in the light grey region of $\boldsymbol{X}^2$, the vectors $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ have the same direction $-\frac{1}{\|\boldsymbol{x}^*\|}(x_1^*, x_2^*, 0, 0, 0)$ and the only difference between them lies in the length of this vector.

13

The potential variation of the form of the expression (8) for the translated logarithm map, arising from the changes in the supports of the geodesics, is one of the main obstructions to generalising the theory for manifolds to orthant spaces, or more general stratified spaces. To study this variation, we first note the following result, which is a direct consequence of Corollary 1.

Corollary 2. *If the support of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$ satisfies the conditions of Corollary 1, then there is a neighbourhood $\mathcal{N}$ of $\boldsymbol{x}$ within its stratum such that, for any $\boldsymbol{x}' \in \mathcal{N}$, the form of the expression (8) for $\Phi(\boldsymbol{x}'; \boldsymbol{x}^*)$ takes the same form of that for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$.*

The singularities of the logarithm map

We now characterise, in terms of the two conditions on the support of a geodesic given in Proposition 2, changes in the form of the expression (8) for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ when $\boldsymbol{x}$ varies locally. Although the roles played by these two conditions in determining the support of a geodesic are different, to some extent, they play a similar role in the change of the form of that expression. Replacing the inequality (5) or (7) by equality determines a quadratic co-dimension one hypersurface. When two or more such hyper-surfaces meet, their normals are linearly independent so they intersect in surfaces of co-dimension at least two. Thus, it suffices to consider a point lying in a single such hyper-surface. Then, points on either side of the hyper-surface will have different supports for their geodesics from $\boldsymbol{x}^*$, but that will not always result in a change in the form the expression for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$.

Proposition 3. *Let $\boldsymbol{x}^*$ and $\boldsymbol{x}_0$ be two given points in $\boldsymbol{X}^m$, and let $(\mathcal{A}, \mathcal{B})$ be the support of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}_0$, where $\mathcal{A} = (A_0, A_1, \cdots, A_k)$ and $\mathcal{B} = (B_0, B_1, \cdots, B_k)$, and where $k > 1$. Assume that $\boldsymbol{x}$ moves from $\boldsymbol{x}_0$, within its stratum, to a first point $\boldsymbol{x}_1$ such that, for $i = i_0 > 0$, the inequality (5), with $\boldsymbol{x}_1, \boldsymbol{x}_2$ replaced by $\boldsymbol{x}^*, \boldsymbol{x}_1$ respectively, becomes an equality while all the other inequalities (5) and (7) remain strict. Then, the support $(\mathcal{A}', \mathcal{B}')$ of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}_1$ has*

$$\mathcal{A}' = (A_0, A_1, \cdots, A_{i_0-1}, A_{i_0} \cup A_{i_0+1}, A_{i_0+2}, \cdots, A_k)$$

and similarly for $\mathcal{B}'$.

  If the orthant

$$\mathcal{O}'' = \mathcal{O}(B_0 \cup B_1 \cup \cdots \cup B_{i_0-1} \cup B_{i_0+1} \cup A_{i_0} \cup A_{i_0+2} \cup \cdots \cup A_k)$$

is contained in $\boldsymbol{X}^m$, then there is a neighbourhood $\mathcal{N}$ of $\boldsymbol{x}_1$ within its stratum such that, for all $\boldsymbol{x} \in \mathcal{N}$, the form of the expression (8) for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ is identical with that for $\Phi(\boldsymbol{x}_0; \boldsymbol{x}^*)$.

  If $\mathcal{O}''$ is not an orthant of $\boldsymbol{X}^m$ then, in any neighbourhood $\mathcal{N}$ of $\boldsymbol{x}_1$ within its stratum, there are $\boldsymbol{x}'$ and $\boldsymbol{x}''$ such that the form for $\Phi(\boldsymbol{x}'; \boldsymbol{x}^*)$ is the same as that for $\Phi(\boldsymbol{x}_0; \boldsymbol{x}^*)$ and that for $\Phi(\boldsymbol{x}''; \boldsymbol{x}^*)$ is determined by the support $(\mathcal{A}', \mathcal{B}')$. When $\mathcal{N}$ is sufficiently small, there are no other possibilities.

Proof. By Corollary 2, the form of the expression (8) will remain constant, as long as the inequalities (5) and (7) remain strict. However, for $\boldsymbol{x} = \boldsymbol{x}_1$, on account of the equality (5) for i_0 at $\boldsymbol{x} = \boldsymbol{x}_1$, the angles θ and ϕ, in the projected
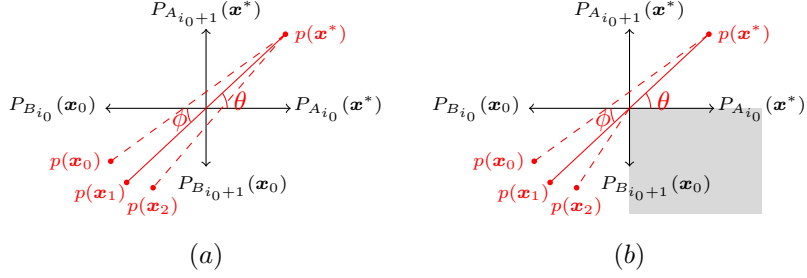
Figure 3: Change of carrier

diagram of Figure 3, will be equal where the projections are as specified in the proof of Proposition 2. Consequently at $\boldsymbol{x}_1$, $\mathcal{O}_{i_0}$ will drop out of the carrier, where $\mathcal{O}_i$ is defined by (4), and, by the continuity of geodesics, the support of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}_1$ will be $(\mathcal{A}', \mathcal{B}')$.

Now, let $\boldsymbol{x}$ continue to move past $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$, remaining sufficiently close to $\boldsymbol{x}_1$ and having projection $p(\boldsymbol{x}_2) = P_{B_{i_0} \cup B_{i_0+1}}(\boldsymbol{x}_2)$ in Figure 3 lying on the opposite side to $p(\boldsymbol{x}_0) = P_{B_{i_0} \cup B_{i_0+1}}(\boldsymbol{x}_0)$ of the ray from the origin to $p(\boldsymbol{x}_1) = P_{B_{i_0} \cup B_{i_0+1}}(\boldsymbol{x}_1)$. If $\mathcal{O}''$ is contained in $\boldsymbol{X}^m$, the projection of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}_2$ would be, as in Figure 3(a), the 'straight' line from $p(\boldsymbol{x}^*) = P_{A_{i_0} \cup A_{i_0+1}}(\boldsymbol{x}^*)$ to $p(\boldsymbol{x}_2)$ passing through the planar quadrant Π_0 determined by $\boldsymbol{P}_{A_{i_0}}(\boldsymbol{x}^*)$ and $P_{B_{i_0}}(\boldsymbol{x}_0)$. This would imply replacing $\mathcal{O}_{i_0}$ in the carrier by $\mathcal{O}''$ with the resulting support for the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}_2$ being $(\mathcal{A}'', \mathcal{B}'')$, where

$$\mathcal{A}'' = (A_0, A_1, \cdots, A_{i_0-1}, A_{i_0+1}, A_{i_0}, A_{i_0+2}, \cdots, A_k)$$

and similarly for $\mathcal{B}''$. In this case, the application of the linear transformation $\jmath$ in the expression for $\Phi(\,\cdot\,; \boldsymbol{x}^*)$ implies that, for such $\boldsymbol{x}_2$, the form of the expression (8) for $\Phi(\boldsymbol{x}_2; \boldsymbol{x}^*)$ is identical with that for $\Phi(\boldsymbol{x}_0; \boldsymbol{x}^*)$.

Assume now that $\mathcal{O}''$ is not an orthant of $\boldsymbol{X}^m$. There might still be an intermediate orthant between $\mathcal{O}_{i_0-1}$ and $\mathcal{O}_{i_0+1}$ arrived at by non-trivial partitions $A_{i_0} = C_1 \cup C_2$, $A_{i_0+1} = D_1 \cup D_2$, $B_{i_0} = E_1 \cup E_2$ and $B_{i_0+1} = F_1 \cup F_2$ such that the orthant

$$\widetilde{\mathcal{O}} = \mathcal{O}(B_0 \cup \cdots \cup B_{i_0-1} \cup E_1 \cup F_1 \cup C_2 \cup D_2 \cup A_{i_0+2} \cup \cdots \cup A_k)$$

is contained in $\boldsymbol{X}^m$ and provides a shorter path between $\mathcal{O}_{i_0-1}$ and $\mathcal{O}_{i_0+1}$. In which case, by Proposition 2(i), we must have

$$\frac{\|P_{C_1 \cup D_1}(\boldsymbol{x}^*)\|}{\|P_{E_1 \cup F_1}(\boldsymbol{x}_2)\|} < \frac{\|P_{C_2 \cup D_2}(\boldsymbol{x}^*)\|}{\|P_{E_2 \cup F_2}(\boldsymbol{x}_2)\|}.$$

This would result in

$$\frac{\|P_{C_1 \cup D_1}(\boldsymbol{x}^*)\|}{\|P_{E_1 \cup F_1}(\boldsymbol{x}_2)\|} < \frac{\|P_{A_{i_0} \cup A_{i_0+1}}(\boldsymbol{x}^*)\|}{\|P_{B_{i_0} \cup B_{i_0+1}}(\boldsymbol{x}_2)\|} < \frac{\|P_{C_2 \cup D_2}(\boldsymbol{x}^*)\|}{\|P_{E_2 \cup F_2}(\boldsymbol{x}_2)\|}$$

and, taking the limit as $\boldsymbol{x}_2 \to \boldsymbol{x}_1$,

$$\frac{\|P_{C_1 \cup D_1}(\boldsymbol{x}^*)\|}{\|P_{E_1 \cup F_1}(\boldsymbol{x}_1)\|} \leqslant \frac{\|P_{A_{i_0} \cup A_{i_0+1}}(\boldsymbol{x}^*)\|}{\|P_{B_{i_0} \cup B_{i_0+1}}(\boldsymbol{x}_1)\|} \leqslant \frac{\|P_{C_2 \cup D_2}(\boldsymbol{x}^*)\|}{\|P_{E_2 \cup F_2}(\boldsymbol{x}_1)\|}. \tag{10}$$

15

On the other hand, the closures of the orthants $\mathcal{O}_{i_0-1}$ and $\widetilde{\mathcal{O}}$ being in $\boldsymbol{X}^m$ ensure that all 2-dimensional orthants in the closure of

$$\widetilde{\mathcal{O}}^* = \mathcal{O}(B_0 \cup \cdots \cup B_{i_0-1} \cup E_1 \cup C_2 \cup A_{i_0+1} \cup \cdots \cup A_k)$$

are in $\boldsymbol{X}^m$ and hence, by Definition 1, so too is $\widetilde{\mathcal{O}}^*$ itself. Then, by the assumption of uniqueness of the equality at $\boldsymbol{x}_1$ of the proposition, we must have by Proposition 2(ii) that

$$\frac{\|P_{C_1}(\boldsymbol{x}^*)\|}{\|P_{E_1}(\boldsymbol{x}_1)\|} > \frac{\|P_{A_{i_0}}(\boldsymbol{x}^*)\|}{\|P_{B_{i_0}}(\boldsymbol{x}_1)\|} > \frac{\|P_{C_2}(\boldsymbol{x}^*)\|}{\|P_{E_2}(\boldsymbol{x}_1)\|}. \tag{11}$$

Similarly, by considering the orthant $\mathcal{O}(B_0 \cup \cdots \cup B_{i_0} \cup F_1 \cup D_2 \cup A_{i_0+2} \cup \cdots \cup A_k)$, we get

$$\frac{\|P_{D_1}(\boldsymbol{x}^*)\|}{\|P_{F_1}(\boldsymbol{x}_1)\|} > \frac{\|P_{A_{i_0+1}}(\boldsymbol{x}^*)\|}{\|P_{B_{i_0+1}}(\boldsymbol{x}_1)\|} > \frac{\|P_{D_2}(\boldsymbol{x}^*)\|}{\|P_{F_2}(\boldsymbol{x}_1)\|}. \tag{12}$$

Since, by assumption, $\mathcal{O}_{i_0}$ drops out of the carrier at $\boldsymbol{x}_1$, we also have

$$\frac{\|P_{A_{i_0}}(\boldsymbol{x}^*)\|}{\|P_{B_{i_0}}(\boldsymbol{x}_1)\|} = \frac{\|P_{A_{i_0+1}}(\boldsymbol{x}^*)\|}{\|P_{B_{i_0+1}}(\boldsymbol{x}_1)\|},$$

so that, combining (11) and (12), we have

$$\frac{\|P_{C_1 \cup D_1}(\boldsymbol{x}^*)\|}{\|P_{E_1 \cup F_1}(\boldsymbol{x}_1)\|} > \frac{\|P_{A_{i_0} \cup A_{i_0+1}}(\boldsymbol{x}^*)\|}{\|P_{B_{i_0} \cup B_{i_0+1}}(\boldsymbol{x}_1)\|} > \frac{\|P_{C_2 \cup D_2}(\boldsymbol{x}^*)\|}{\|P_{E_2 \cup F_2}(\boldsymbol{x}_1)\|},$$

contradicting (10).

Thus, if $\mathcal{O}''$ is not contained in $\boldsymbol{X}^m$, the projection of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}_2$ continues to pass through the origin, as shown in Figure 3(b), and the carrier remains as it was for $\boldsymbol{x}_1$, where the support is $(\mathcal{A}', \mathcal{B}')$ given above. In this case, the form of the expression (8) for $\Phi(\boldsymbol{x}_2; \boldsymbol{x}^*)$ clearly differs from that for $\Phi(\boldsymbol{x}_0; \boldsymbol{x}^*)$. $\qquad\square$

Note that the equality (5) for $i = i_0$ at $\boldsymbol{x}_1$ and the mutual orthogonality of all the axes together imply that

$$\left(\frac{\|P_{B_{i_0}}(\boldsymbol{x}_1)\|}{\|P_{A_{i_0}}(\boldsymbol{x}^*)\|} P_{A_{i_0}}(\boldsymbol{x}^*), \frac{\|P_{B_{i_0+1}}(\boldsymbol{x}_1)\|}{\|P_{A_{i_0+1}}(\boldsymbol{x}^*)\|} P_{A_{i_0+1}}(\boldsymbol{x}^*) \right)$$
$$= \frac{\|P_{B_{i_0} \cup B_{i_0+1}}(\boldsymbol{x}_1)\|}{\|P_{A_{i_0} \cup A_{i_0+1}}(\boldsymbol{x}^*)\|} P_{A_{i_0} \cup A_{i_0+1}}(\boldsymbol{x}^*).$$

This confirms that the form of the expression for $\Phi(\boldsymbol{x}_0; \boldsymbol{x}^*)$ is still valid for $\Phi(\boldsymbol{x}_1; \boldsymbol{x}^*)$, as expected by the continuity of geodesics. Similarly, the form of the expression for $\Phi(\boldsymbol{x}_2; \boldsymbol{x}^*)$ is still valid for $\Phi(\boldsymbol{x}_1; \boldsymbol{x}^*)$ whether or not the orthant $\mathcal{O}_{i_0}$ has been replaced by $\mathcal{O}''$.

A similar argument to that for the proof of Proposition 3 gives the following complementary result.

Proposition 4. *Let $\boldsymbol{x}^*$ and $\boldsymbol{x}_1$ be two given points in $\boldsymbol{X}^m$, and let $(\mathcal{A}, \mathcal{B})$ be the support of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}_1$, where $\mathcal{A} = (A_0, A_1, \cdots, A_k)$ and $\mathcal{B} = (B_0, B_1, \cdots, B_k)$, and where $k > 0$. Assume that all inequalities (5) and (7), with $\boldsymbol{x}_1, \boldsymbol{x}_2$ replaced by $\boldsymbol{x}^*, \boldsymbol{x}_1$ respectively, are strict except that, for $i = i_0 > 0$ and unique non-trivial partitions $C_{i_0 1} \cup C_{i_0 2}$ for A_{i_0} and $D_{i_0 1} \cup D_{i_0 2}$ for B_{i_0}, (7) is an equality and that the corresponding orthant $\mathcal{O}'$ given by (6) with $i = i_0$ is contained in $\boldsymbol{X}^m$.*

If the orthant

$$\mathcal{O}''' = \mathcal{O}(B_0 \cup \cdots \cup B_{i_0 - 1} \cup D_{i_0 2} \cup C_{i_0 1} \cup A_{i_0 + 1} \cup \cdots \cup A_k) \qquad (13)$$

is contained in $\boldsymbol{X}^m$, there is a neighbourhood $\mathcal{N}$ of $\boldsymbol{x}_1$ within its stratum such that the form of the expression for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ is the same, for all $\boldsymbol{x} \in \mathcal{N}$. Then, the common form of the expression for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ is determined by $(\mathcal{A}', \mathcal{B}')$, where

$$\mathcal{A}' = (A_0, \cdots, A_{i_0 - 1}, C_{i_0 1}, C_{i_0 2}, A_{i_0 + 1}, \cdots, A_k)$$

and similarly for $\mathcal{B}'$.

If $\mathcal{O}'''$ is not an orthant of $\boldsymbol{X}^m$ then, in any neighbourhood $\mathcal{N}$ of $\boldsymbol{x}_1$ within its stratum, there are $\boldsymbol{x}'$ and $\boldsymbol{x}''$ such that the form for $\Phi(\boldsymbol{x}'; \boldsymbol{x}^*)$ is the same as that for $\Phi(\boldsymbol{x}_1; \boldsymbol{x}^*)$ and that for $\Phi(\boldsymbol{x}''; \boldsymbol{x}^*)$ is determined by $(\mathcal{A}', \mathcal{B}')$. When $\mathcal{N}$ is sufficiently small, there are no other possibilities.

Note that the results in Propositions 3 and 4 where the changed support must be used to obtain the correct expression for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ are reflections of each other where an othant is removed or introduced, respectively, in the carrier.

The carrier of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$ will also change when $\boldsymbol{x}$ moves from one stratum to another which necessarily involves, as initial, final or intermediate stratum, a stratum of locally positive co-dimension. The set of all such strata, together with the quadratic hyper-surfaces determined by equalities in each of the relevant equations (5), form the defining boundaries for the *(pre)-vistal polyhedral subdivision*, with respect to $\boldsymbol{x}^*$, in [15]. The points in any component of the complement of these surfaces all have the same carrier. However, for our analysis, we shall only be concerned with changes in the forms of the expressions taken by $\log_{\boldsymbol{x}^*}(\boldsymbol{x})$, or equivalently by $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$, when $\boldsymbol{x}$ or $\boldsymbol{x}^*$ vary within their strata rather than when the underlying stratum changes. Thus, we may encapsulate as follows the hyper-surfaces for which, when crossed by a geodesic, it is necessary to take account of the change of support to obtain the correct value for the logarithm map. Note however that the support on either side of the hyper-surface will give the correct value at the surface.

Definition 11. *Given a point $\boldsymbol{x}^* \in \boldsymbol{X}^m$, $\mathcal{D}_{\boldsymbol{x}^*}$ denotes the set of the singularities of the log map $\log_{\boldsymbol{x}^*}$ that consists of all points $\boldsymbol{x} \in \boldsymbol{X}^m$ for which the support $(\mathcal{A}, \mathcal{B})$, where $\mathcal{A} = (A_0, \cdots, A_k)$ and $\mathcal{B} = (B_0, \cdots, B_k)$, of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$ has the property that, for one or more $i = i_0 > 0$, there are non-trivial partitions $A_{i_0} = C_{i_0 1} \cup C_{i_0 2}$ and $B_{i_0} = D_{i_0 1} \cup D_{i_0 2}$ with*

$$\frac{\|P_{C_{i_0 1}}(\boldsymbol{x}^*)\|}{\|P_{D_{i_0 1}}(\boldsymbol{x})\|} = \frac{\|P_{C_{i_0 2}}(\boldsymbol{x}^*)\|}{\|P_{D_{i_0 2}}(\boldsymbol{x})\|}, \qquad (14)$$

where the corresponding orthant $\mathcal{O}'$ of (6) is contained in $\boldsymbol{X}^m$, but $\mathcal{O}'''$ of (13) is not. $\mathcal{D}_{\boldsymbol{x}^*}$ will be referred to a type 1 singularity set.

In view of the symmetry that reverses the geodesics at the same time as it reverses the order of the strata and interchanges the roles of the sequences $\mathcal{A}$ and $\mathcal{B}$ of edge sets in the support, the definition is symmetric: $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}^*}$ if and only if $\boldsymbol{x}^* \in \mathcal{D}_{\boldsymbol{x}}$. Since each stratum is a Euclidean orthant, it is preserved under multiplication by $\lambda > 0$ in $\mathbb{R}^M$ which also multiplies the length of each curve by λ. Then, since the geodesic γ joining $\boldsymbol{x}^*$ to $\boldsymbol{x}$ is the shortest curve through the strata of $\boldsymbol{X}^m$ from $\boldsymbol{x}^*$ to $\boldsymbol{x}$, it follows that γ is mapped onto the geodesic from $\lambda \boldsymbol{x}^*$ to $\lambda \boldsymbol{x}$. In particular, these two geodesics have the same carrier. Thus, $\mathcal{D}_{\lambda \boldsymbol{x}^*} = \lambda \mathcal{D}_{\boldsymbol{x}^*}$ and, since the equations (14) are homogeneous, $\mathcal{D}_{\boldsymbol{x}^*} = \lambda \mathcal{D}_{\boldsymbol{x}^*}$.

The pseudo-partition of $\boldsymbol{X}^m$ with respect to $\boldsymbol{x}^*$ determined by $\mathcal{D}_{\boldsymbol{x}^*}$ gives rise to a polyhedral subdivision of each stratum by restriction. It is coarser than the (pre)-vistal subdivision of [15] and, if $\boldsymbol{X}^m$ is a tree space and if $\boldsymbol{x}^*$ lies in a top-dimensional stratum, it is equivalent to the polyhedral subdivision defined in [3].

5   Limits, projections and derivatives

We now turn to certain limits and projections of the translated logarithm map that, in particular, will enable us to calculate the directional derivatives we require.

Firstly, we obtain an expression for the limit of the translated logarithm map as the reference point $\boldsymbol{x}^*$ moves along a geodesic. For a vector $\boldsymbol{w}$ in the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x}^*$, write $\boldsymbol{x}^*(\lambda, \boldsymbol{w})$ for the point distant $\lambda \| \boldsymbol{w} \|$ along the geodesic γ starting at $\boldsymbol{x}^*$ with initial tangent vector $\boldsymbol{w}$. Then, we have the following result.

Theorem 2. *Let $\sigma = \mathcal{O}(E)$ be a stratum of $\boldsymbol{X}^m$, $\boldsymbol{x}^* \in \sigma$ and $\boldsymbol{x}$ be a fixed choice of point anywhere in $\boldsymbol{X}^m$.*

(i) If $\boldsymbol{w} \in \mathbb{R}(E)$ is tangent to σ at $\boldsymbol{x}^*$, then

$$\lim_{\lambda \to 0+} \Phi(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w})) = \Phi(\boldsymbol{x}; \boldsymbol{x}^*).$$

(ii) If σ bounds $\tau = \mathcal{O}(E \cup F)$ in $\boldsymbol{X}^m$ and $\boldsymbol{w}_\tau \in \mathbb{R}(E) \times \mathcal{O}(F)$ is tangent to τ at $\boldsymbol{x}^*$, then the limit

$$\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) = \lim_{\lambda \to 0+} \Phi(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)) \tag{15}$$

exists. Moreover, there exist $\epsilon > 0$ and sequences $\mathcal{A} = (A_0, A_1, \cdots, A_k)$ and $\mathcal{B} = (B_0, B_1, \cdots, B_k)$ of sets of axes such that, for each $\lambda \in (0, \epsilon)$, $(\mathcal{A}, \mathcal{B})$ forms the support of the geodesic from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)$ to $\boldsymbol{x}$. In terms of these $\mathcal{A}$ and $\mathcal{B}$,

$$\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) = \jmath\left(P_{B_0}(\boldsymbol{x}), -\frac{\|P_{B_1}(\boldsymbol{x})\|}{\|W_1\|} W_1, \cdots, -\frac{\|P_{B_k}(\boldsymbol{x})\|}{\|W_k\|} W_k \right), \tag{16}$$

where $W_i = P_{A_i \cap E}(\boldsymbol{x}^*)$, unless $P_{A_i \cap E}(\boldsymbol{x}^*) = 0$, in which case $W_i = P_{A_i \cap F}(\boldsymbol{w}_\tau)$, the projection of $\boldsymbol{w}_\tau$ on $\mathbb{R}(A_i)$, and $\jmath$ is the linear transformation given by Definition 5.

Thus, $\Psi(\boldsymbol{x}, \boldsymbol{w}; \boldsymbol{x}^*)$ defined by (15) is a 'boundary limit' of the translated logarithm map $\Phi(\boldsymbol{x}; \boldsymbol{x}')$, namely its limit approaching $\boldsymbol{x}^*$ in the boundary component σ of the orthant τ from the direction $\boldsymbol{w}$ tangent to τ at $\boldsymbol{x}^*$. In the following, we shall refer to $\Psi(\boldsymbol{x}, \boldsymbol{w}; \boldsymbol{x}^*)$ simply as the *boundary limit* of $\Phi(\boldsymbol{x}; \boldsymbol{x}')$ (with respect to $\boldsymbol{w}$).

Comparing the expression (8) for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau))$ with the expression (16) for $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$, where $\lambda > 0$ is sufficiently small, $\boldsymbol{x}^* \in \sigma = \mathcal{O}(E)$, $\tau = \mathcal{O}(E \cup F)$ and $\boldsymbol{w}_\tau \in \mathbb{R}(E) \times \mathcal{O}(F)$, we see that the change from the former to the latter amounts to replacing each $P_{A_i}(x^*(\lambda, \boldsymbol{w}_\tau))$ by either $P_{A_i \cap E}(\boldsymbol{x}^*)$ or $P_{A_i \cap F}(\boldsymbol{x}^*)$ depending on whether $A_i \cap E$ is a non-empty set or not. Noting that $\boldsymbol{x}^*$ has zero coordinate with respect to any axes in F, the existence of the components corresponding to W_i in the expression for $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ implies that $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ generally differs from $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$. This is due to a 'sudden disappearance' of the axes in F when $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau) \in \tau$ moves towards $\boldsymbol{x}^* \in \sigma$.

Proof. (*i*) This follows from the uniform continuity of geodesics with respect to their end points (cf. [6], pp195-196) and also from a minor modification of the proof of (*ii*) below.

(*ii*) Note that, since $\boldsymbol{w}_\tau \in \mathbb{R}(E) \times \mathcal{O}(F)$, $\boldsymbol{x}^*$ and $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)$ lie in different strata. Writing γ_λ for the geodesic from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)$ to $\boldsymbol{x}$, as $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)$ moves along γ the support of γ_λ can only change when γ meets transversally one or more of the hyper-surfaces where the carrier of the geodesic to $\boldsymbol{x}$ changes. This can only happen at discrete points along γ so, for some $\epsilon > 0$ and $0 < \lambda \leqslant \epsilon$, the carriers of the geodesics γ_λ will be independent of λ. Let $(\mathcal{A}, \mathcal{B})$ be the support of γ_ϵ from $\boldsymbol{x}^*(\epsilon, \boldsymbol{w}_\tau)$ to $\boldsymbol{x}$, where $\mathcal{A} = (A_0, A_1, \cdots, A_k)$ and $\mathcal{B} = (B_0, B_1, \cdots, B_k)$. Then, $A_0 \cup A_1 \cup \cdots \cup A_k = E \cup F$ and, for $0 < \lambda \leqslant \epsilon$, the integer k and the support $(\mathcal{A}, \mathcal{B})$ will remain constant for the expression

$$
\begin{aligned}
\Phi(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)) \;=\; {}_{\jmath}\bigg(& P_{B_0}(\boldsymbol{x}), -\frac{\|P_{B_1}(\boldsymbol{x})\|}{\|P_{A_1}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau))\|} P_{A_1}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)), \cdots, \\
& -\frac{\|P_{B_k}(\boldsymbol{x})\|}{\|P_{A_k}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau))\|} P_{A_k}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)) \bigg),
\end{aligned}
$$

replacing $\boldsymbol{x}^*$ in (8) by $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)$. Then, since the $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)$ lie in τ for all sufficiently small positive λ, the vectors $\Phi(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau))$ all lie in $\mathbf{C}_\tau$ so that it makes sense to take the limit as $\lambda \to 0+$, where $\mathbf{C}_\tau$ is the common translated cone of the tangent cone at $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)$ as introduced in Section 2.

To evaluate it, we take the limit in the above expression for $\Phi(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau))$. Since $\boldsymbol{x}^* \in \mathcal{O}(E)$, $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau) = \boldsymbol{x}^* + \lambda \boldsymbol{w}_\tau$ for sufficiently small $\lambda > 0$ and it follows that $P_{A_i}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)) = P_{A_i \cap E}(\boldsymbol{x}^*) + \lambda P_{A_i}(\boldsymbol{w}_\tau)$. So the limit as $\lambda \to 0+$ of this term is $P_{A_i \cap E}(\boldsymbol{x}^*)$ if that is non-zero. If it is zero, then $A_i \cap E = \emptyset$ since $\|P_{\{e\}}(\boldsymbol{x}^*)\| > 0$ for all $e \in E$. Then $P_{A_i}(\boldsymbol{w}_\tau)$, the projection of $\boldsymbol{w}_\tau$ on $\mathbb{R}(A_i)$ is, in fact, $P_{A_i \cap F}(\boldsymbol{w}_\tau)$. $\qquad\square$

If σ has co-dimension l and τ co-dimension l' then, when $l - l' = 1$ and so $|F| = 1$, there is no $i > 0$ such that $|A_i| > 1$ and $P_{A_i \cap E}(\boldsymbol{x}^*) = 0$ as all the axes involved in the carrier that are not in $E \cup F$ are in $A_0 = B_0$. If further $l = 1$ and $l' = 0$, that is, σ is a stratum of local co-dimension one and τ co-bounding σ is a locally top-dimensional stratum, then $\Psi(\,\cdot\,, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ obtained here is identical with the map resulting from the 'folding map' composed with $\Phi(\,\cdot\,; \boldsymbol{x}^*)$ used

in [3] when $\boldsymbol{X}^m$ is a tree space, noting that $\boldsymbol{w}_\tau$ in this case is unique up to a positive scalar multiple.

Example 3. Consider the orthant space $\boldsymbol{X}^2$ in Example 2. Take $\sigma = \{o\}$ and $\tau = \mathcal{O}(u_1, u_2)$. Recall from Example 2 that the tangent cone to $\boldsymbol{X}^2$ at o is $\boldsymbol{X}^2$ itself. Take $\boldsymbol{w}_\tau = \boldsymbol{x}^*$, where $\boldsymbol{x}^* \in \tau$ is indicated in Figure 2. Then, $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; o) = \Phi(\boldsymbol{x}; \boldsymbol{x}^*)$ for any $\boldsymbol{x} \in \boldsymbol{X}^2$. Since the light grey region in Figure 2 may change if $\boldsymbol{x}^*$ changes, $\Phi(\cdot; \boldsymbol{x}^*)$ may change as a map when $\boldsymbol{x}^*$ changes. Hence, the boundary limit $\Psi(\cdot, \boldsymbol{w}_\tau; o)$ of $\Phi(\cdot; \lambda \boldsymbol{w}_\tau)$ from the direction $\boldsymbol{w}_\tau$ as $\lambda \to 0$, as a map, also depends on $\boldsymbol{w}_\tau$.

For $\boldsymbol{w}_\tau$ as given in Theorem $2(ii)$, write $\boldsymbol{w}_\tau^\perp$ for the component of $\boldsymbol{w}_\tau$ orthogonal to σ, that is, the component in $\{\boldsymbol{0}\} \times \mathcal{O}(F) \subset \mathbb{R}(E) \times \mathcal{O}(F)$. Then, the following consequences of Theorem 2 imply that, although the boundary limit $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ generally depends on $\boldsymbol{w}_\tau$, for given $\boldsymbol{x}$ and $\boldsymbol{x}^*$, as noted in Example 3 above, it remains constant in some circumstances. In particular, to consider the changes of $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ as $\boldsymbol{x}$ varies, it suffices to restrict attention to $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, recalling that $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is the open unit spherical segment of $\{\boldsymbol{0}\} \times \mathcal{O}(F)$ given by Definition 4.

Corollary 3. *With the notation and hypotheses of Theorem $2(ii)$,*

(i) $\Psi(\cdot, \lambda \boldsymbol{w}_\tau; \boldsymbol{x}^*) = \Psi(\cdot, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ *for all* $\lambda > 0$;

(ii) $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) = \Psi(\boldsymbol{x}, \boldsymbol{w}_\tau^\perp; \boldsymbol{x}^*)$.

Proof. (i) is obvious from the expression (16) and (ii) is immediate since $\sigma = \mathcal{O}(E)$, $\tau = \mathcal{O}(E \cup F)$ and only the F-coordinates of $\boldsymbol{w}_\tau$ are potentially involved in (16). □

When $\boldsymbol{x}^*$ lies in a stratum σ of positive co-dimension that is not locally top-dimensional, the vector $\log_{\boldsymbol{x}^*}(\boldsymbol{x})$, and so $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$, will usually have non-zero components both tangent to σ and orthogonal to it. In order to discuss the projections, onto these components, of the translated logarithm map and of its boundary limits, as well as to discuss their derivatives, we extend the notation P for projection maps on $\boldsymbol{X}^m$ given by Definition 9 to include projection maps on tangent cones, or their translated cones. However, since we are more interested in the orthant itself rather than the axes determining it, we shall use P_σ instead of P_E, where $\sigma = \mathcal{O}(E)$. In particular, for any stratum $\tau = \mathcal{O}(E \cup F)$ co-bounding σ in $\boldsymbol{X}^m$, P_σ and $P_{\tau \setminus \sigma}$ respectively are the projections onto the two factors of the corresponding stratum $\mathbb{R}(E) \times \mathcal{O}(F)$ in the common translated cone $\boldsymbol{C}_\sigma$, or equivalently in the tangent cone at a point of σ, depending on the context. Note that, on $\boldsymbol{C}_\sigma$, P_σ so defined is the tangential projection onto σ and $P_{\tau \setminus \sigma}$ is one of several possible normal projections. In particular, for $\boldsymbol{w}_\tau \in \mathbb{R}(E) \times \mathcal{O}(F)$, $\boldsymbol{w}_\tau^\perp = P_{\tau \setminus \sigma}(\boldsymbol{w}_\tau)$. We shall further extend the notation P_σ to include top-dimensional, or locally top-dimensional, strata by taking it to be the identity in that case.

Definition 12. *For* $\boldsymbol{x}^*$ *in* $\sigma = \mathcal{O}(E)$ *or in* $\tau = \mathcal{O}(E \cup F)$ *co-bounding* σ, $\log_{\boldsymbol{x}^*}^\sigma$ *is the map from* $\boldsymbol{X}^m$ *to the tangent cone to* $\boldsymbol{X}^m$ *at* $\boldsymbol{x}^*$ *defined by*

$$\log_{\boldsymbol{x}^*}^\sigma(\boldsymbol{x}) = P_\sigma(\log_{\boldsymbol{x}^*}(\boldsymbol{x})).$$

Similarly, $\Phi_\sigma(\,\cdot\,;\boldsymbol{x}^*)$ is the map from $\boldsymbol{X}^m$ to the translated tangent cone $\boldsymbol{C}_\sigma$ defined by

$$\Phi_\sigma(\boldsymbol{x};\boldsymbol{x}^*) = P_\sigma(\Phi(\boldsymbol{x};\boldsymbol{x}^*)).$$

In particular $\Phi_\sigma(\boldsymbol{x};\boldsymbol{x}^*) = \Phi(\boldsymbol{x};\boldsymbol{x}^*)$ if σ is a top-dimensional, or locally top-dimensional, stratum.

For $\boldsymbol{x}^*$ in σ of locally positive co-dimension, the non-zero components of $\log_{\boldsymbol{x}^*}(\boldsymbol{x})$ orthogonal to σ correspond to axes with respect to which $\boldsymbol{x}^*$ has zero coefficient and $\boldsymbol{x}$ has non-zero coefficient. Hence, these axes are in $A_0 = B_0 = E(\boldsymbol{x}^*,\boldsymbol{x})$, the set of axes common to all strata in the carrier of the geodesic between these two points, so that they correspond to components of v_0 in (9). This implies, in particular, that $\Phi_\sigma(\boldsymbol{x};\boldsymbol{x}^*)$ is given by (8) with $P_{B_0}(\boldsymbol{x})$ there replaced by $P_{B_0 \cap E}(\boldsymbol{x})$. Then, since the restriction to each stratum of the set $\mathcal{D}_{\boldsymbol{x}}$ given by Definition 11 is relatively closed, the form of the expression for $\Phi(\boldsymbol{x};\boldsymbol{x}')$ will remain constant for $\boldsymbol{x}'$ varying in a neighbourhood of $\boldsymbol{x}^*$ in σ when $\boldsymbol{x}^*$ is restricted to avoid $\mathcal{D}_{\boldsymbol{x}}$. Hence, the proof of Lemma 4 in [3] of the differentiability of $\Phi(\boldsymbol{x};\boldsymbol{x}^*)$ with respect to $\boldsymbol{x}^*$ for the case that $\boldsymbol{X}^m$ is a tree space and $\boldsymbol{x}^*$ lies in a top-dimensional stratum will give the following generalisation of that result to the derivative of $\Phi_\sigma(\boldsymbol{x};\boldsymbol{x}^*)$ with respect to $\boldsymbol{x}^*$. Since the proof is similar to that for Lemma 4 in [3], we omit it here.

Proposition 5. *Let $\boldsymbol{x}$ and $\boldsymbol{x}^*$ be fixed points in $\boldsymbol{X}^m$ with $\boldsymbol{x}^*$ in the stratum $\sigma = \mathcal{O}(E)$ and $\boldsymbol{x} \notin \mathcal{D}_{\boldsymbol{x}^*}$, where $\mathcal{D}_{\boldsymbol{x}^*}$ is the type 1 singularity set given by Definition 11. Then, the map*

$$\sigma \to \mathbb{R}(E); \quad x' \mapsto \Phi_\sigma(\boldsymbol{x};\boldsymbol{x}')$$

is differentiable with respect to $\boldsymbol{x}'$ at $\boldsymbol{x}^*$ with derivative given by

$$\begin{aligned}
&M_{\boldsymbol{x}^*}^\sigma(\boldsymbol{x}) \\
&= J^\top \mathrm{diag}\Big\{ \mathbf{0}_{|B_0 \cap E|}, -\|P_{B_1}(\boldsymbol{x})\| M_{P_{A_1}(\boldsymbol{x}^*)}^\dagger, \cdots, -\|P_{B_k}(\boldsymbol{x})\| M_{P_{A_k}(\boldsymbol{x}^*)}^\dagger \Big\} J
\end{aligned} \quad (17)$$

where the sequences $\mathcal{A} = (A_0,\cdots,A_k)$ and $\mathcal{B} = (B_0,\cdots,B_k)$ form the support of the geodesic from $\boldsymbol{x}^*$ to $\boldsymbol{x}$ and J is the matrix representation of the linear transformation $\jmath$ given by Definition 5, and where, for $\boldsymbol{y} = (y_1,\cdots,y_l) \neq 0$,

$$M_{\boldsymbol{y}}^\dagger = \frac{1}{\|\boldsymbol{y}\|} I_l - \frac{1}{\|\boldsymbol{y}\|^3} \boldsymbol{y}^\top \boldsymbol{y} \quad (18)$$

is the derivative of the map $\boldsymbol{y} \mapsto \frac{1}{\|\boldsymbol{y}\|}\boldsymbol{y}$.

Note that, if $l > 1$, $\|\boldsymbol{y}\| M_{\boldsymbol{y}}^\dagger$ is the projection onto the hyper-plane in $\mathbb{R}^l$ orthogonal to $\boldsymbol{y}$ and, when $l = 1$, $M_{y_1}^\dagger = 0$. Hence, if $k = 0$ or if $k > 0$ and $|A_i| = 1$ for all $1 \leqslant i \leqslant k$, then the derivative of $\Phi_\sigma(\boldsymbol{x};\boldsymbol{x}')$, with respect to $\boldsymbol{x}'$, at $\boldsymbol{x}' = \boldsymbol{x}^*$ is zero. Recall that the corresponding translated logarithm map in the Euclidean space is the identity map, independent of $\boldsymbol{x}'$, and so its derivative with respect to $\boldsymbol{x}'$ is identically zero. Hence, in a broad sense, Proposition 5 captures where and how the derivative of $\Phi_\sigma(\boldsymbol{x};\boldsymbol{x}')$ differs from that of the corresponding translated Euclidean logarithm map.

Returning to the boundary limit $\Psi(\boldsymbol{x},\boldsymbol{w}_\tau;\boldsymbol{x}^*)$ of $\Phi(\boldsymbol{x};\boldsymbol{x}')$ with $\boldsymbol{x}^* \in \sigma = \mathcal{O}(E)$, where $\tau = \mathcal{O}(E \cup F)$ co-bounds σ and $\boldsymbol{w}_\tau$ is in $\mathbb{R}(E) \times \mathcal{O}(F)$, since

$\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ is in $\boldsymbol{C}_\tau$, both projections $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) = P_\tau(\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*))$ and $P_\sigma(\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*))$ are well defined. In particular, $\Psi_\tau(\,\cdot\,, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ is a map from $\boldsymbol{X}^m$ onto $\mathbb{R}(E \cup F)$. Then, we also have the following consequences of Theorem 2, giving the relationships between the projections of the boundary limit of the translated logarithm map and the boundary limit of the projections of the translated logarithm map.

Corollary 4. *With the notation and hypotheses of Theorem 2(ii),*

(i) $\displaystyle\lim_{\lambda \to 0+} \Phi_\tau(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)) = \lim_{\lambda \to 0+} \Phi_\tau(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^\perp)) = \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*);$

(ii) $P_\sigma\left(\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)\right) = \Phi_\sigma(\boldsymbol{x}; \boldsymbol{x}^*).$

Proof. The equality of the extreme terms in (i) follows since the W_i in (16) are determined by the axes in $E \cup F$, so that it does not matter whether we project on $\mathcal{O}(E \cup F)$ before or after taking the limit, and the remaining term $P_{B_0 \cap (E \cup F)}(\boldsymbol{x})$ remains constant throughout the limiting process. The equality with the central term in (i) follows from Corollary 3(ii): $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) = \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^\perp; \boldsymbol{x}^*)$, which is $\displaystyle\lim_{\lambda \to 0+} \Phi_\tau(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^\perp))$ by the case already established.

Note that, since projection onto $\mathbb{R}(E) \subset \mathbb{R}(E \cup F)$ is unaffected by first projecting onto $\mathbb{R}(E \cup F)$, (ii) is equivalent to

$$P_\sigma(\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)) = \Phi_\sigma(\boldsymbol{x}; \boldsymbol{x}^*). \tag{19}$$

To establish (19), we need to allow for the fact that the geodesics γ_λ from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)$ to $\boldsymbol{x}$ and the geodesic γ_0 from $\boldsymbol{x}^*$ to $\boldsymbol{x}$ may have different carriers. We assume that λ is restricted to the range $0 < \lambda < \epsilon$ such that the initial segments of γ_λ all lie in $\zeta = \mathcal{O}(E \cup F \cup G)$, where possibly $G = \emptyset$ and so $\zeta = \tau$, and let K be the set of axes with respect to which the initial segment of γ_0 has positive coordinates. Then, $K \supseteq E \cup G$. Now, $e \in E \cup F \cup G$ if, and only if, for each λ and some maximal $\delta(\lambda) > 0$, $\|P_{\{e\}}(\gamma_\lambda(s))\| > 0$ for $s \in (0, \delta(\lambda))$. From the uniform continuity of geodesics with respect to their endpoints, it is clear that we must have $\delta(\lambda) \to \delta_0 \geqslant 0$ as $\lambda \to 0$. If $\delta_0 > 0$, then $\|P_{\{e\}}(\gamma_0(s))\| > 0$ for $s \in (0, \delta_0)$ and so $e \in K$. Conversely, $e \in K \cap (E \cup F \cup G)$ implies that $\|P_{\{e\}}(\gamma_0(s))\| > 0$ for $s \in (0, \delta(0))$ and we must have $\delta_0 = \delta(0)$.

Thus, for any axis e in $K \cap (E \cup F \cup G)$, the projections $P_{\{e\}}(\gamma_\lambda(s))$ and $P_{\{e\}}(\gamma_0(s))$ of the initial segments of these geodesics all lie in the closure of the stratum $\mathcal{O}(E \cup F \cup G)$. The uniform continuity of these geodesics, and so of their projections, with respect to their endpoints, together with their linearity within that closed stratum, implies that the components $P_{\{e\}}(\dot\gamma_\lambda(0))$ converge to $P_{\{e\}}(\dot\gamma_0(0))$ as $\lambda \to 0$. In particular, since $E \subseteq K$, this is valid for any axis e in E, which establishes (19). $\qquad\square$

The comments made prior to Proposition 5 regarding the form of the expression for $\Phi_\sigma(\boldsymbol{x}; \boldsymbol{x}^*)$ can be generalised to apply to $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$: using the notation in Theorem 2(ii) for $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ we have that

$$\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) = \jmath\left(P_{B_0 \cap (E \cup F)}(\boldsymbol{x}), -\frac{\|P_{B_1}(\boldsymbol{x})\|}{\|W_1\|} W_1, \cdots, -\frac{\|P_{B_k}(\boldsymbol{x})\|}{\|W_k\|} W_k\right). \tag{20}$$

Recall that $\mathcal{S}_{\tau \setminus \sigma}^{l - l'}$ denotes the set of unit vectors in $\{\boldsymbol{0}\} \times \mathcal{O}(F) \subset \mathbb{R}(E) \times \mathcal{O}(F)$ that comprises all unit vectors that are tangent to τ and orthogonal to σ. If

$l - l' = 1$, $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ comprises a single point. When $l - l' > 1$, for any fixed $\boldsymbol{x} \in \boldsymbol{X}^m$, the pseudo-partition of $\boldsymbol{X}^m$ determined by $\mathcal{D}_{\boldsymbol{x}}$ induces a polyhedral subdivision of $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ where, in each cell of the induced polyhedral subdivision, the form of the expression (16) for $\Psi(\boldsymbol{x}, \cdot \, ; \boldsymbol{x}^*)$, and so the form of the expression for $\Psi_\tau(\boldsymbol{x}, \cdot \, ; \boldsymbol{x}^*)$, remains the same. In particular, this implies that, for fixed $\boldsymbol{x}$, $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau ; \boldsymbol{x}^*)$ is a continuous function of $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$. In fact, the directional derivatives of $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau ; \boldsymbol{x}^*)$ with respect to $\boldsymbol{w}_\tau$ also exist in directions $\boldsymbol{v}$ in the tangent space to $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ at $\boldsymbol{w}_\tau$ that we denote by $\mathcal{T}_{\boldsymbol{w}_\tau}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'})$. These derivatives have the property given in the following proposition, where we note that $\mathbb{R}(E) \times \mathcal{O}(F) \subset \mathbb{R}(E) \times \mathbb{R}(F)$ so that, for fixed $\boldsymbol{x}$ and $\boldsymbol{x}^*$, $\boldsymbol{w}_\tau$ and $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau ; \boldsymbol{x}^*)$ lie in the same Euclidean space.

Proposition 6. *Let the stratum* $\sigma = \mathcal{O}(E)$ *of co-dimension* $l (\geqslant 2)$ *bound, in* $\boldsymbol{X}^m$, *the stratum* $\tau = \mathcal{O}(E \cup F)$ *of co-dimension* $l' (< l - 1)$. *Fix* $\boldsymbol{x}, \boldsymbol{x}^* \in \boldsymbol{X}^m$ *with* $\boldsymbol{x}^* \in \sigma$. *Then, as a function of* $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, *the directional derivative* D *of* $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau ; \boldsymbol{x}^*)$ *at* $\boldsymbol{w}_\tau$ *in the direction* $\boldsymbol{v} \in \mathcal{T}_{\boldsymbol{w}_\tau}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'})$ *exists and satisfies*

$$\langle \boldsymbol{w}_\tau, \, D\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau ; \boldsymbol{x}^*)(\boldsymbol{v}) \rangle = 0.$$

Proof. Without loss of generality, we may assume that $\| \boldsymbol{v} \| = 1$. Consider the geodesic on $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ given by $\alpha(s) = \boldsymbol{w}_\tau \cos s + \boldsymbol{v} \sin s$. Write w_1 for a vector whose coordinates comprise a subset of those of $\boldsymbol{w}_\tau$, and v_1, α_1 for the corresponding components of $\boldsymbol{v}$ and α respectively. Then, the initial tangent vector of the function $f(s) = \frac{\alpha_1(s)}{\| \alpha_1(s) \|}$ is $\dot{f}(0) = v_1 M_{w_1}^\dagger$, where $M_{\boldsymbol{y}}^\dagger$ is given by (18). Clearly, $\langle w_1, \dot{f}(0) \rangle = 0$, since the image of $M_{w_1}^\dagger$ is orthogonal to w_1.

On the other hand, it follows from the argument in the proof of Theorem 2 that, for all sufficiently small s, the expression for $\Psi_\tau(\boldsymbol{x}, \alpha(s); \boldsymbol{x}^*)$ all have the same form provided that, when $\boldsymbol{w}_\tau$ lies on the boundary of a cell of the induced polyhedral subdivision on $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, we use for $\boldsymbol{w}_\tau$ the expression valid for $s > 0$. Thus, we may use the expression for $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau ; \boldsymbol{x}^*)$ given by (20) to express $D\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau ; \boldsymbol{x}^*)(\boldsymbol{v})$ in the form $\boldsymbol{v} \, M_{\boldsymbol{x}^*, \boldsymbol{x}}(\boldsymbol{w}_\tau)$, where

$$M_{\boldsymbol{x}^*, \boldsymbol{x}}(\boldsymbol{w}_\tau) = J^\top \mathrm{diag} \left\{ \mathbf{0}, -\| P_{B_{l_1}}(\boldsymbol{x}) \| M_{W_{l_1}}^\dagger, \cdots, -\| P_{B_{l_j}}(\boldsymbol{x}) \| M_{W_{l_j}}^\dagger, \mathbf{0} \right\} J \quad (21)$$

and where, using the notation of Theorem 2, $W_{l_i} = P_{A_{l_i} \cap F}(\boldsymbol{w}_\tau)$ are just those components in the expression for $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau ; \boldsymbol{x}^*)$ for which $P_{A_{l_i} \cap E}(\boldsymbol{x}^*) = 0$ and $|A_{l_i} \cap F| > 1$. Since $\| y \| M_{\boldsymbol{y}}^\dagger$ is the projection onto the hyperplane orthogonal to y in the Euclidean space where y lies as noted after the statement of Proposition 5, the result follows. $\qquad \square$

The proof of Proposition 6 also shows that, if $\boldsymbol{w}_\tau$ lies in the interior of a single cell of the induced polyhedral subdivision of $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, then $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau'; \boldsymbol{x}^*)$ is differentiable with respect to $\boldsymbol{w}_\tau'$ at $\boldsymbol{w}_\tau$. However, if $\boldsymbol{w}_\tau$ lies in the boundary of a cell of the induced polyhedral subdivision, this no longer holds, although directional derivatives still exist.

The directional derivative of $\langle \boldsymbol{w}_\tau, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau ; \boldsymbol{x}^*) \rangle$, as a function of $\boldsymbol{w}_\tau$, now follows from Proposition 6.

Corollary 5. *Assume that all assumptions in Proposition 6 hold. Then, for any $\boldsymbol{v} \in \mathcal{T}_{\boldsymbol{w}_\tau}(\mathcal{S}_{\tau\backslash\sigma}^{l-l'})$, the derivative D in the direction $\boldsymbol{v}$ of $\langle \boldsymbol{w}_\tau,\, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \rangle$ at $\boldsymbol{w}_\tau$ is given by*

$$D\left(\langle \boldsymbol{w}_\tau,\, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \rangle\right)(\boldsymbol{v}) = \langle \boldsymbol{v}, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \rangle.$$

Proof. The second term in the expansion

$$
\begin{aligned}
& D\left(\langle \boldsymbol{w}_\tau,\, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \rangle\right)(\boldsymbol{v}) \\
= \;& \langle D\,\boldsymbol{w}_\tau(\boldsymbol{v}), \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \rangle + \langle \boldsymbol{w}_\tau,\, D\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)(\boldsymbol{v}) \rangle
\end{aligned}
$$

vanishes by Proposition 6. The result then follows since the directional derivative $D\,\boldsymbol{w}_\tau(\boldsymbol{v})$ is given by the derivative at $s = 0$ of the geodesic $\alpha(s) = \boldsymbol{w}_\tau \cos s + \boldsymbol{v} \sin s$. $\qquad\square$

6   Characterisation of Fréchet means

In the remainder of this paper, we use the knowledge obtained so far on the translated logarithm map to investigate Fréchet means of probability measures on $\boldsymbol{X}^m$. So, from now on we assume that μ is a probability measure on $\boldsymbol{X}^m$ and that its Fréchet function defined by (1), where $\boldsymbol{M} = \boldsymbol{X}^m$, is finite at one point. The latter ensures that the Fréchet function of μ is finite everywhere.

Since the squared distance on a CAT(0)-space is a convex function with respect to each of its variables, it follows that the Fréchet mean of μ is unique and that the condition for $\boldsymbol{x}^*$ to be the Fréchet mean of μ, that is, the condition for $\boldsymbol{x}^*$ to satisfy

$$\int_{\boldsymbol{X}^m} d(\boldsymbol{x}^*, \boldsymbol{x})^2 \, \mathrm{d}\mu(\boldsymbol{x}) < \int_{\boldsymbol{X}^m} d(\boldsymbol{x}', \boldsymbol{x})^2 \, \mathrm{d}\mu(\boldsymbol{x}), \qquad \text{for any } \boldsymbol{x}' \neq \boldsymbol{x}^*,$$

is equivalent to this inequality holding in any neighbourhood of $\boldsymbol{x}^*$. Then, since the Fréchet function of μ is differentiable at $\boldsymbol{x}^*$ if $\boldsymbol{x}^*$ lies in a top-dimensional, or locally top-dimensional, stratum, the above condition for such $\boldsymbol{x}^*$ to be the Fréchet mean of μ is equivalent to the condition that

$$\int_{\boldsymbol{X}^m} \log_{\boldsymbol{x}^*}(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) = 0, \tag{22}$$

similar to the condition for Fréchet means in Riemannian manifolds of non-positive curvature.

When $\boldsymbol{x}^*$ lies in a stratum σ of locally positive co-dimension, the squared distance $d(\boldsymbol{x}^*, \boldsymbol{x})^2$ is no longer differentiable at $\boldsymbol{x}^*$ for any fixed $\boldsymbol{x}$. Nevertheless, it has directional derivatives along all possible directions and then the above condition becomes that, at $\boldsymbol{x}^* \in \sigma$, the Fréchet function of μ has non-negative directional derivatives along all possible directions. The fact that $\boldsymbol{X}^m$ is a CAT(0)-space also implies that the derivative at $\boldsymbol{x}^*$ in the direction $\boldsymbol{w}$ of the distance function $d_{\boldsymbol{x}} = d(\cdot, \boldsymbol{x})$ can be expressed as

$$(Dd_{\boldsymbol{x}}(\boldsymbol{x}^*))(\boldsymbol{w}) = -\frac{1}{d_{\boldsymbol{x}}(\boldsymbol{x}^*)} \ll \boldsymbol{w}, \log_{\boldsymbol{x}^*}(\boldsymbol{x}) \gg,$$

where $\ll , \gg$ is defined by (2) (cf. [14], (2.5), p417). Thus, the criterion for a point $\boldsymbol{x}^*$ lying in a stratum σ of locally positive co-dimension to be the Fréchet mean of μ is equivalent to the condition that

$$\int_{\boldsymbol{X}^m} \ll \boldsymbol{w}, \log_{\boldsymbol{x}^*}(\boldsymbol{x}) \gg \mathrm{d}\mu(\boldsymbol{x}) \leqslant 0 \tag{23}$$

for all tangent vectors $\boldsymbol{w}$ at $\boldsymbol{x}^*$.

For any vector $\boldsymbol{w}$ at $\boldsymbol{x}^*$ which is tangent to σ, the fact that $-\boldsymbol{w}$ is also tangent to σ at $\boldsymbol{x}^*$ implies that the inequality (23) must be an equality for all such $\boldsymbol{w}$. From this it follows that

$$\int_{\boldsymbol{X}^m} \log^{\sigma}_{\boldsymbol{x}^*}(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) = 0, \tag{24}$$

analogous to the condition (22). On the other hand, for any given stratum τ co-bounding σ and any vector $\boldsymbol{w}$ at $\boldsymbol{x}^*$ tangent to τ, it is possible to link the derivative, at $\boldsymbol{x}^*$, of the Fréchet function in the direction $\boldsymbol{w}$ with $\Psi_{\tau}(\,\cdot\,, \boldsymbol{w}; \boldsymbol{x}^*)$, the projection of the boundary limit of $\Phi(\,\cdot\,; \boldsymbol{x}^*)$. To show this, we need the following limiting property of the directional derivatives on general Alexandov spaces of non-positive curvature which include CAT(0)-spaces.

Lemma 1. *Let X be an Alexandrov space of non-positive curvature, and let x_0 and x be two distinct fixed points in X. For some $\epsilon > 0$, assume that $\gamma : [0, \epsilon) \to X$ is a geodesic with $\gamma(0) = x$ and $\dot{\gamma}(0) = v_x$. Then, if $\{x_i : i \geqslant 1\}$ is a sequence of points along γ convergent to x, the derivative D at x in the direction v_x of the distance function $d_{x_0} = d(x_0, \cdot)$ has the property that*

$$Dd_{x_0}(v_x) = \lim_{i \to \infty} Dd_{x_0}(v_{x_i}),$$

where v_{x_i} denotes the tangent vector at x_i of the geodesic γ.

Proof. For $x, y, z \in X$, denote by $\angle_x(y, z)$ the Alexandrov angle at x between the geodesics from x to y and z respectively. Since $Dd_{x_0}(v_x) = - \ll v_x, \log_x(x_0) \gg / d_{x_0}(x) = -\|v_x\| \cos \angle_x(x', x_0)$ where x' is a point on the geodesic γ, it is sufficient to show that, for a fixed point x' chosen on γ, $\angle_x(x', x_0) = \lim_{i \to \infty} \angle_{x_i}(x', x_0)$.

For this, we write $\gamma_{a,b}$ for the (unique) geodesic segment joining a and b, for any two distinct points a and b in X. Then, given sequences of points $a_i \to a$, $b_i \to b$ and $c_i \to c$ in X, it follows from the Cartan-Hadamand theorem that the geodesic segments γ_{a_i, b_i} and γ_{a_i, c_i} converge uniformly, as maps, to $\gamma_{a,b}$ and $\gamma_{a,c}$ respectively. From this it follows that $\angle_a(b, c) \geqslant \limsup_{i \to \infty} \angle_{a_i}(b_i, c_i)$ (cf. [7], Theorem 4.3.11, p.119). Applying this to the sequence of geodesic triangles $\Delta(x' x_i x_0)$, we obtain

$$\angle_x(x', x_0) \geqslant \limsup_{i \to \infty} \angle_{x_i}(x', x_0). \tag{25}$$

On the other hand, using (4.3) p.124 of [7], we have

$$\limsup_{i \to \infty} \overline{\angle}_{x_i}(x, x_0) \leqslant \pi - \angle_x(x', x_0),$$

where, as in Section 2, $\overline{\angle}$ denotes the corresponding comparison angle in $\mathbb{R}^2$. Then, since $\overline{\angle}_{x_i}(x, x_0) \geqslant \angle_{x_i}(x, x_0)$, the above implies that

$$
\begin{aligned}
\angle_x(x', x_0) &\leqslant \pi - \limsup_{i \to \infty} \overline{\angle}_{x_i}(x, x_0) \\
&\leqslant \pi - \limsup_{i \to \infty} \angle_{x_i}(x, x_0) \\
&= \liminf_{i \to \infty} \left\{ \pi - \angle_{x_i}(x, x_0) \right\}.
\end{aligned}
$$

However, since X has non-positive curvature, if x_i lies between x and x' on the geodesic segment $\gamma_{x,x'}$, then $\angle_{x_i}(x', x_0) + \angle_{x_i}(x_0, x) \geqslant \pi$ (cf. [7], p117, line 5). Hence,

$$
\angle_x(x', x_0) \leqslant \liminf_{i \to \infty} \angle_{x_i}(x', x_0).
$$

This, together with (25), gives that

$$
\angle_x(x', x_0) = \lim_{i \to \infty} \angle_{x_i}(x', x_0),
$$

so that the required result follows. $\qquad\square$

Recalling that $\Phi(\boldsymbol{x}; \boldsymbol{x}^*) = \log_{\boldsymbol{x}^*}(\boldsymbol{x}) + \boldsymbol{x}^*$, the criteria (23) and (24) for a point $\boldsymbol{x}^*$ to be the Fréchet mean of μ may now be recast, the former in terms of the standard Euclidean inner product and $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}; \boldsymbol{x}^*)$, the projection of the boundary limit of $\Phi(\boldsymbol{x}; \boldsymbol{x}^*)$, when $\boldsymbol{x}^*$ lies in a stratum σ of positive co-dimension and $\boldsymbol{w}$ is tangent to a co-bounding stratum τ.

Theorem 3. *Let σ be a stratum in $\boldsymbol{X}^m$ of co-dimension $l(\geqslant 0)$. The necessary and sufficient conditions for a given point $\boldsymbol{x}^* \in \sigma$ to be the Fréchet mean of μ are*

(i)

$$
\boldsymbol{x}^* = \int_{\boldsymbol{X}^m} \Phi_\sigma(\boldsymbol{x}; \boldsymbol{x}^*) \, \mathrm{d}\mu(\boldsymbol{x}); \tag{26}
$$

(ii) for any stratum τ in $\boldsymbol{X}^m$ of co-dimension l', $0 \leqslant l' < l$, co-bounding σ and any $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau \backslash \sigma}^{l-l'}$,

$$
\left\langle \boldsymbol{w}_\tau, \int_{\boldsymbol{X}^m} \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \, \mathrm{d}\mu(\boldsymbol{x}) \right\rangle \leqslant 0, \tag{27}
$$

where $\mathcal{S}_{\tau \backslash \sigma}^{l-l'}$ is given by Definition 4.

Note that case *(ii)* may only occur if $l > 0$, but need not occur then. Note also that, if $\boldsymbol{X}^m$ is a tree space, the special case $l = 0$ of this result is the same as that of Lemma 3 of [3]; and the special case $l = 1$, so that $l' = 0$, is equivalent to that given by Lemma 5 of [3]: on the one hand, $\mathcal{S}_{\tau \backslash \sigma}^{l-l'}$ contains a single unit vector and, on the other hand, as we noted earlier, $\Psi_\tau(\,\cdot\,, \boldsymbol{w}_\tau; \boldsymbol{x}^*) = \Psi(\,\cdot\,, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ is identical with the composition of the 'folding map' with $\Phi(\,\cdot\,; \boldsymbol{x}^*)$ in [3].

Proof. Noting that (i) is precisely (24), it is sufficient to show that (ii) is equivalent to (23) for any tangent vector $\boldsymbol{w}$ that is not tangent to $\sigma = \mathcal{O}(E)$. For this, we fix any stratum $\tau = \mathcal{O}(E \cup F)$, of co-dimension l', co-bounding $\sigma = \mathcal{O}(E)$ and take $\boldsymbol{w} = \boldsymbol{w}_\tau \in \mathbb{R}(E) \times \mathcal{O}(F)$. Then, it follows from Lemma 1 that (23) is equivalent to

$$\lim_{\lambda \to 0+} \int_{\boldsymbol{X}^m} \ll \boldsymbol{w}_\tau, \log_{\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)}(\boldsymbol{x}) \gg \mathrm{d}\mu(\boldsymbol{x}) \leqslant 0, \tag{28}$$

where $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau) = \boldsymbol{x}^* + \lambda \boldsymbol{w}_\tau$, as defined prior to Theorem 2. Since $\boldsymbol{w}_\tau$ is tangent to τ at $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)$ for sufficiently small $\lambda > 0$ and, for any given $\boldsymbol{x}$, $\log_{\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)}(\boldsymbol{x})$ is tangent either to τ or to one of the strata that co-bound τ, we have

$$\ll \boldsymbol{w}_\tau, \log_{\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)}(\boldsymbol{x}) \gg = \langle \boldsymbol{w}_\tau, \log_{\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)}(\boldsymbol{x}) \rangle.$$

However,

$$\begin{aligned} \left\langle \boldsymbol{w}_\tau, \log_{\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)}(\boldsymbol{x}) \right\rangle &= \langle \boldsymbol{w}_\tau, \Phi(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)) - \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau) \rangle \\ &= \langle \boldsymbol{w}_\tau, \Phi_\tau(\boldsymbol{x}; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)) - \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau) \rangle . \end{aligned}$$

Hence, by Corollary 4(i) and then Corollary 3(ii), (28) is equivalent to

$$\int_{\boldsymbol{X}^m} \left\langle \boldsymbol{w}_\tau, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^\perp; \boldsymbol{x}^*) \right\rangle \mathrm{d}\mu(\boldsymbol{x}) \leqslant \langle \boldsymbol{w}_\tau, \boldsymbol{x}^* \rangle,$$

where $\boldsymbol{w}_\tau^\perp = P_{\tau \setminus \sigma}(\boldsymbol{w}_\tau)$. Decomposing $\boldsymbol{w}_\tau$ as $\boldsymbol{w}_\tau = \boldsymbol{w}_\sigma + \boldsymbol{w}_\tau^\perp$, where $\boldsymbol{w}_\sigma = P_\sigma(\boldsymbol{w}_\tau)$, leads to

$$\begin{aligned} \left\langle \boldsymbol{w}_\tau, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^\perp; \boldsymbol{x}^*) \right\rangle &= \left\langle \boldsymbol{w}_\sigma, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^\perp; \boldsymbol{x}^*) \right\rangle + \left\langle \boldsymbol{w}_\tau^\perp, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^\perp; \boldsymbol{x}^*) \right\rangle \\ &= \left\langle \boldsymbol{w}_\sigma, \Phi_\sigma(\boldsymbol{x}; \boldsymbol{x}^*) \right\rangle + \left\langle \boldsymbol{w}_\tau^\perp, \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^\perp; \boldsymbol{x}^*) \right\rangle , \end{aligned}$$

where the second equality follows from Corollary 4(ii). The required result now follows by noting (i), noting that $\langle \boldsymbol{w}_\tau, \boldsymbol{x}^* \rangle = \langle \boldsymbol{w}_\sigma, \boldsymbol{x}^* \rangle = 0$ and noting that, by applying the projection P_τ to the result of Corollary 3(i), $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^\perp; \boldsymbol{x}^*) = \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^\perp / \| \boldsymbol{w}_\tau^\perp \|; \boldsymbol{x}^*)$. $\square$

From now on, we assume that $\boldsymbol{\xi}$ is a random variable defined on a probability space $(\boldsymbol{\Omega}, \mathcal{F}, \mathbf{P})$ with values in $\boldsymbol{X}^m$ and that μ is the distribution (measure) of $\boldsymbol{\xi}$, i.e. $\mu(B) = \mathbf{P}(\boldsymbol{\xi}^{-1}(B))$ for any Borel set B in $\boldsymbol{X}^m$. When the stratum containing the Fréchet mean $\boldsymbol{x}^*$ of the probability measure μ on $\boldsymbol{X}^m$ is of locally positive co-dimension, (27) being an equality has a significant influence on the nature of the distributions of the Euclidean random variables $\Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$, which will be seen in Propositions 7, 8 and 9. We shall also see, in Proposition 10, its link with the long term behaviour of sample Fréchet means.

Definition 13. *For the stratum σ of co-dimension $l (\geqslant 1)$, in which the Fréchet mean $\boldsymbol{x}^*$ of μ lies, and the stratum τ, of co-dimension $l' (< l)$, co-bounding σ, the subset $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$ of $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is defined as*

$$\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu) = \left\{ \boldsymbol{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'} \mid \text{the inequality (27) for } \boldsymbol{w}_\tau \text{ is an equality} \right\}, \tag{29}$$

where $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is given by Definition 4.

The convexity of the directional derivative $D(d_{\boldsymbol{x}}^2)(\boldsymbol{w})$ in $\boldsymbol{w}$ (cf. [14], pp416-417) ensures that $\Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu)$ is a convex subset of $\mathcal{S}_{\tau\backslash\sigma}^{l-l'}$ and that

$$\Theta_\sigma(\boldsymbol{x}^*;\mu) = \bigcup_{\tau\supset\sigma}\Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu) \tag{30}$$

is a convex subset of $\bigcup_{\tau\supset\sigma}\mathcal{S}_{\tau\backslash\sigma}^{l-l'} \subseteq \boldsymbol{C}_\sigma$. If $l - l' = 1$, $\mathcal{S}_{\tau\backslash\sigma}^{l-l'}$ consists of a single unit vector so that $\Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu)$ is either $\mathcal{S}_{\tau\backslash\sigma}^{l-l'}$ itself or an empty set. In general, if the closure of $\Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu)$ is contained in $\mathcal{S}_{\tau\backslash\sigma}^{l-l'}$, the fact that $\langle \boldsymbol{w}_\tau, \Psi_\tau(\boldsymbol{x},\boldsymbol{w}_\tau;\boldsymbol{x}^*)\rangle$ is continuous in $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau\backslash\sigma}^{l-l'}$ implies that $\Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu)$ itself must be closed.

The following result gives a relationship between the Fréchet mean $\boldsymbol{x}^*$ of μ and the Euclidean mean of $\Psi_\tau(\boldsymbol{\xi},\boldsymbol{w}_\tau;\boldsymbol{x}^*)$. Here, and henceforth, by interior we intend the relative interior, that is, interior with respect to the subspace topology.

Proposition 7. *Let the stratum σ of co-dimension $l(\geqslant 2)$ bound, in $\boldsymbol{X}^m$, the stratum τ of co-dimension $l'(< l - 1)$. Assume that the Fréchet mean $\boldsymbol{x}^*$ of μ lies in σ and that $\mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu)) \neq \emptyset$. Then, for any $\boldsymbol{w}_\tau \in \Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu)$,*

$$\int_{\boldsymbol{X}^m}\{\Psi_\tau(\boldsymbol{x},\boldsymbol{w}_\tau;\boldsymbol{x}^*) - \Phi_\sigma(\boldsymbol{x};\boldsymbol{x}^*)\}\,\mathrm{d}\mu(\boldsymbol{x}) = 0. \tag{31}$$

Note that, if $l' = l - 1$, equality (31) holds automatically since its left hand side is a 1-dimensional vector so that the equality follows from the assumption that $\boldsymbol{w}_\tau \in \Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu)$.

Proof. By the continuity of $\Psi_\tau(\boldsymbol{x},\boldsymbol{w}_\tau;\boldsymbol{x}^*)$ in $\boldsymbol{w}_\tau$, we may assume that $\boldsymbol{w}_\tau \in \mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu))$. Then equality holds in (27) in a neighbourhood of $\boldsymbol{w}_\tau$, so that

$$D\left(\left\langle \boldsymbol{w}_\tau, \int_{\boldsymbol{X}^m}\Psi_\tau(\boldsymbol{x},\boldsymbol{w}_\tau;\boldsymbol{x}^*)\,\mathrm{d}\mu(\boldsymbol{x})\right\rangle\right)(\boldsymbol{v}) = 0, \quad \forall\,\boldsymbol{v}\in\mathcal{T}_{\boldsymbol{w}_\tau}(\mathcal{S}_{\tau\backslash\sigma}^{l-l'}).$$

By Corollary 5, this implies that

$$\left\langle \boldsymbol{v}, \int_{\boldsymbol{X}^m}\Psi_\tau(\boldsymbol{x},\boldsymbol{w}_\tau;\boldsymbol{x}^*)\,\mathrm{d}\mu(\boldsymbol{x})\right\rangle = 0, \quad \forall\,\boldsymbol{v}\in\mathcal{T}_{\boldsymbol{w}_\tau}(\mathcal{S}_{\tau\backslash\sigma}^{l-l'}).$$

On the other hand, it follows from $\int_{\boldsymbol{X}^m}\Phi_\sigma(\boldsymbol{x};\boldsymbol{x}^*)\,\mathrm{d}\mu(\boldsymbol{x}) = \boldsymbol{x}^*$ and $\langle\boldsymbol{w}_\tau,\boldsymbol{x}^*\rangle = 0$ that

$$\left\langle \boldsymbol{w}_\tau, \int_{\boldsymbol{X}^m}\Phi_\sigma(\boldsymbol{x};\boldsymbol{x}^*)\,\mathrm{d}\mu(\boldsymbol{x})\right\rangle = 0, \quad \forall\,\boldsymbol{w}_\tau\in\mathcal{S}_{\tau\backslash\sigma}^{l-l'}.$$

Hence, taking the directional derivative of the left hand side as a function of $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau\backslash\sigma}^{l-l'}$, we have

$$\left\langle \boldsymbol{v}, \int_{\boldsymbol{X}^m}\Phi_\sigma(\boldsymbol{x};\boldsymbol{x}^*)\,\mathrm{d}\mu(\boldsymbol{x})\right\rangle = 0$$

for all $\boldsymbol{v}\in\mathcal{T}_{\boldsymbol{w}_\tau}(\mathcal{S}_{\tau\backslash\sigma}^{l-l'})$. Noting that the left hand side of (31) is a vector lying in the $(l-l')$-dimensional Euclidean space containing $\mathcal{S}_{\tau\backslash\sigma}^{l-l'}$, the fact that $\boldsymbol{w}_\tau\in\Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu)$, together with the above, implies that the required result holds for any $\boldsymbol{w}_\tau\in\mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*;\mu))$. $\square$

One immediate consequence of Proposition 7 is the following.

Corollary 6. *Assume that the conditions given in Proposition 7 are satisfied. If $\sigma = \mathcal{O}(E)$ and $\tau = \mathcal{O}(E \cup F)$ then, for all $\boldsymbol{w}_\tau \in \Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$,*

$$\boldsymbol{x}^* = \int_{\boldsymbol{X}^m} \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \, \mathrm{d}\mu(\boldsymbol{x}). \tag{32}$$

That is, the point $\boldsymbol{x}^* \in \sigma$, as a point in $\mathbb{R}(E \cup F)$, is the Euclidean mean of each of the Euclidean random variables $\Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ for such $\boldsymbol{w}_\tau$.

If a stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l(\geqslant 1)$ bounds, in $\boldsymbol{X}^m$, the stratum τ of co-dimension $l'(< l)$ and if $\boldsymbol{x}^* \in \sigma$, then it follows from the proof of Proposition 6 that the maps $\Psi_\tau(\cdot, \boldsymbol{w}_\tau^1; \boldsymbol{x}^*)$ and $\Psi_\tau(\cdot, \boldsymbol{w}_\tau^2; \boldsymbol{x}^*)$ from $\boldsymbol{X}^m$ to $\mathbb{R}(E \cup F)$ are generally not identical for any given distinct $\boldsymbol{w}_\tau^i \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, $i = 1, 2$. With the insight obtained from that proof, to characterise the places where they differ we introduce the subset $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau)$ of $\boldsymbol{X}^m$ as follows. It will be clear later, in the proof of Proposition 9, that the set of $\boldsymbol{x} \in \boldsymbol{X}^m$ where $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau^1; \boldsymbol{x}^*) \neq \Psi(\boldsymbol{x}, \boldsymbol{w}_\tau^2; \boldsymbol{x}^*)$ is contained in the set $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^1) \cup \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^2) \cup \mathcal{D}_{\boldsymbol{x}^*}$. Thus, in particular, for $\boldsymbol{\xi}$ lying outside of the latter set, the Euclidean random variables $\Psi(\boldsymbol{\xi}, \boldsymbol{w}_\tau^1; \boldsymbol{x}^*)$ and $\Psi(\boldsymbol{\xi}, \boldsymbol{w}_\tau^2; \boldsymbol{x}^*)$ are identical. This fact will be used in the derivation of the limiting distribution of sample Fréchet means in the next section.

Definition 14. *Let the stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l(\geqslant 1)$ bound, in $\boldsymbol{X}^m$, the stratum $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l'(< l)$. For $\boldsymbol{x}^* \in \sigma$ and $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, a point $\boldsymbol{x} \in \boldsymbol{X}^m$ is called singular with respect to $(\boldsymbol{x}^*, \boldsymbol{w}_\tau)$, if at least one A_i with $A_i \cap E = \emptyset$ has $|A_i \cap F| > 1$, where $i \geqslant 1$ and the sequences $\mathcal{A} = (A_0, A_1, \cdots, A_k)$ and $\mathcal{B} = (B_0, B_1, \cdots, B_k)$ form the support of the geodesics from $\boldsymbol{x}^* + \lambda \boldsymbol{w}_\tau$ to $\boldsymbol{x}$ for all sufficiently small $\lambda > 0$. The set $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau)$ consists of all points $\boldsymbol{x}$ that are singular with respect to $(\boldsymbol{x}^*, \boldsymbol{w}_\tau)$, and we refer to it as type 2 singularity set.*

For example, in the orthant space $\boldsymbol{X}^2$ of Example 3, using the notation there, $\Sigma_{\tau,\{o\}}(o; \boldsymbol{w}_\tau)$ is the closure of the light grey region in Figure 2. It follows from comparison of the corresponding expressions (16) and (20) that the singularity of $\boldsymbol{x}$ with respect to $(\boldsymbol{x}^*, \boldsymbol{w}_\tau)$ has the same effect on $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ as it does on $\Psi(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$. In particular, in terms of the matrix $M_{\boldsymbol{x}^*, \boldsymbol{x}}(\boldsymbol{w})$ given by (21), we can express the type 2 singularity set $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau)$ defined above as

$$\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau) \setminus \mathcal{D}_{x^*} = \{\boldsymbol{x} \in \boldsymbol{X}^m \mid M_{\boldsymbol{x}^*, \boldsymbol{x}}(\boldsymbol{w}_\tau) \neq 0\}.$$

Note that $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau) = \emptyset$ if $l - l' = 1$, since then $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ contains a single unit vector $\boldsymbol{w}_\tau$ which leads to the impossibility that $|A_i \cap F| > 1$. Generally, if $l - l' > 1$, which implies that $l \geqslant 2$, $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau)$ could be relatively substantial. Nevertheless, we have the following result on the measure of $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau)$.

Proposition 8. *Let the stratum σ of co-dimension $l(\geqslant 2)$ bound, in $\boldsymbol{X}^m$, the stratum τ of co-dimension $l'(< l - 1)$. Assume that the Fréchet mean $\boldsymbol{x}^*$ of μ lies in σ and that $\boldsymbol{w}_\tau \in \mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$, where $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$ is defined by (29). Then, the type 2 singularity set $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau)$ has μ-measure zero, i.e. $\mu(\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau)) = 0$.*

Proof. Let $\alpha(s)$ be a unit speed geodesic in $\mathcal{S}_{\tau\setminus\sigma}^{l-l'}$, write $\boldsymbol{v}(s) = \dot{\alpha}(s)$ and define $h(s) = \langle \boldsymbol{v}(s), \int_{\boldsymbol{X}^m} \Psi_\tau(\boldsymbol{x}, \alpha(s); \boldsymbol{x}^*) \, \mathrm{d}\mu(\boldsymbol{x}) \rangle$. Since $\mathcal{S}_{\tau\setminus\sigma}^{l-l'}$ is an open subset of a Euclidean sphere, we have $\dot{\boldsymbol{v}}(s) = -\alpha(s)$, $\ddot{\alpha}(s) = -\alpha(s)$ and so, by Proposition 6 and its proof,

$$
\begin{aligned}
\dot{h}(s) &= \left\langle \boldsymbol{v}(s), \int_{\boldsymbol{X}^m} D\Psi_\tau(\boldsymbol{x}, \alpha(s); \boldsymbol{x}^*)(\boldsymbol{v}(s)) \mathrm{d}\mu(\boldsymbol{x}) \right\rangle \\
&= \left\langle \boldsymbol{v}(s), \int_{\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \alpha(s))} \boldsymbol{v}(s) M_{\boldsymbol{x}^*, \boldsymbol{x}}(\alpha(s)) \mathrm{d}\mu(\boldsymbol{x}) \right\rangle,
\end{aligned}
$$

where $M_{\boldsymbol{x}^*, \boldsymbol{x}}(\boldsymbol{w})$ is given by (21). The expression for $M_{\boldsymbol{x}^*, \boldsymbol{x}}(\boldsymbol{w})$ implies that, for $\boldsymbol{w} \in \mathcal{S}_{\tau\setminus\sigma}^{l-l'}$ and any fixed $\boldsymbol{x} \in \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w})$, $\langle \boldsymbol{v}, \boldsymbol{v} \, M_{\boldsymbol{x}^*, \boldsymbol{x}}(\boldsymbol{w}) \rangle$ can be written in the form

$$
-\sum_{i=1}^{j} \frac{\|P_{B_{l_i}}(\boldsymbol{x})\|}{\|W_{l_i}\|} \left\{ \|v_{l_i}\|^2 - \left\langle v_i, \frac{W_{l_i}}{\|W_{l_i}\|} \right\rangle^2 \right\}
$$

for some $1 \leqslant j \leqslant k$, where W_{l_i} and $P_{B_{l_i}}(\boldsymbol{x})$ are those required for the expression (21) for $M_{\boldsymbol{x}^*, \boldsymbol{x}}(\boldsymbol{w})$ in the proof of Proposition 6. This implies that $\dot{h}(0)$ must be non-positive. Moreover, for any open or closed subset $\mathcal{E} \subseteq \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \alpha(0))$ such that $\Psi_\tau(\boldsymbol{x}, \alpha(0); \boldsymbol{x}^*)$ has the same expression for all $\boldsymbol{x} \in \mathcal{E}$, there is a vector $\boldsymbol{v}(0) \in \mathcal{T}_{\alpha(0)}(\mathcal{S}_{\tau\setminus\sigma}^{l-l'})$ such that $\langle \boldsymbol{v}(0), \boldsymbol{v}(0) \, M_{\boldsymbol{x}^*, \boldsymbol{x}}(\alpha(0)) \rangle < 0$ for all $\boldsymbol{x} \in \mathcal{E}$. Then, if $\mu(\mathcal{E}) \neq 0$, the corresponding h satisfies

$$
\dot{h}(0) \leqslant \left\langle \boldsymbol{v}(0), \int_{\mathcal{E}} \boldsymbol{v}(0) M_{\boldsymbol{x}^*, \boldsymbol{x}}(\alpha(0)) \, \mathrm{d}\mu(\boldsymbol{x}) \right\rangle < 0.
$$

Clearly, $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \alpha(0))$ can be decomposed as a finite disjoint union of such sets $\mathcal{E}$.

If $\boldsymbol{w}_\tau = \alpha(0) \in \mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$ then, for any $v(0) \in \mathcal{T}_{\boldsymbol{w}_\tau}(\mathcal{S}_{\tau\setminus\sigma}^{l-l'})$, the corresponding geodesic $\alpha(s)$ lies in $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$ for all sufficiently small $s \geqslant 0$. Using a similar argument to that for the proof of Proposition 7, the corresponding $h(s)$ must be identically zero for all sufficiently small $s \geqslant 0$, which implies that $\dot{h}(0) = 0$. Hence, we must have $\mu(\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau)) = 0$. $\qquad\square$

If a stratum σ bounds τ in $\boldsymbol{X}^m$, $\boldsymbol{x}^* \in \sigma$ and $\boldsymbol{w}_\tau^1$, $\boldsymbol{w}_\tau^2$ are two different vectors at $\boldsymbol{x}^*$ tangent to τ, then it follows from the map $\Psi_\tau(\,\cdot\,, \boldsymbol{w}_\tau^1; \boldsymbol{x}^*)$ generally differing from $\Psi_\tau(\,\cdot\,, \boldsymbol{w}_\tau^2; \boldsymbol{x}^*)$ that the distribution of the Euclidean random variable $\Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau^1; \boldsymbol{x}^*)$ generally differs from that of $\Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau^2; \boldsymbol{x}^*)$. Nevertheless, under the conditions in Proposition 8, the $\Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ are in fact a.s. identical for $\boldsymbol{w}_\tau \in \mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$.

Proposition 9. *Assume that $\boldsymbol{\xi}$ is a random variable on $\boldsymbol{X}^m$ with distribution measure μ having Fréchet mean $\boldsymbol{x}^*$. Assume further that $\mu(\mathcal{D}_{\boldsymbol{x}^*}) = 0$ and that $\boldsymbol{x}^*$ lies in the stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l(\geqslant 2)$. Let the stratum τ of co-dimension $l'(< l - 1)$ co-bound σ, in $\boldsymbol{X}^m$. If $\mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)) \neq \emptyset$, then the distributions of the Euclidean random variables $\Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ are independent of $\boldsymbol{w}_\tau \in \mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$, where the set $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$ is defined by (29).*

Note that the example in the next section makes it clear that the condition $\boldsymbol{w}_\tau \in \mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$ in the statement of Proposition 9 cannot be relaxed to $\boldsymbol{w}_\tau \in \Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$.

Proof. First, we show that, for any given distinct $\boldsymbol{w}_\tau^j \in \mathcal{S}_{\tau \backslash \sigma}^{l-l'}$, $j = 1, 2$, and for $\boldsymbol{x} \notin \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^1) \bigcup \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^2) \bigcup \mathcal{D}_{\boldsymbol{x}^*}$, $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^1; \boldsymbol{x}^*) = \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^2; \boldsymbol{x}^*)$. Then, it follows from the assumption and Proposition 8 that $\Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau^1; \boldsymbol{x}^*) = \Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau^2; \boldsymbol{x}^*)$ a.s. Recall from the proof of Theorem 2(*ii*) that, for fixed $\boldsymbol{x} \in \boldsymbol{X}^m$, $\boldsymbol{x}^* \in \sigma$ and $\boldsymbol{w}_\tau \in S_{\tau \backslash \sigma}^{l-l'}$, the supports of the geodesics from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau) = \boldsymbol{x}^* + \lambda \boldsymbol{w}_\tau$ to $\boldsymbol{x}$ are the same, for all sufficiently small $\lambda > 0$, and that the expression for $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ is determined by this common support. Thus, $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^j; \boldsymbol{x}^*)$ is identical if the geodesics from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^j)$ to $\boldsymbol{x}$ have the same support when $\lambda > 0$ is sufficiently small.

Suppose now that the supports $(\mathcal{A}^j, \mathcal{B}^j)$, $j = 1, 2$, of the geodesics from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^1)$ and $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^2)$ respectively to $\boldsymbol{x}$ are different, for all sufficiently small $\lambda > 0$. Then, the geodesic γ_λ between $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^1)$ and $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^2)$ must meet at least one hyper-surface in $\mathcal{D}_{\boldsymbol{x}}$. If there are more than one, but necessarily finitely many, such hyper-surfaces, by introducing a point on γ_λ between each pair of consecutive such hyper-surfaces, the change of the supports of the geodesics from points of γ_λ to $\boldsymbol{x}$ can be considered inductively to reduce the case to where γ_λ meets only one such hyper-surface.

Hence, without loss of generality, we assume that γ_λ only meets $\mathcal{D}_{\boldsymbol{x}}$ at a point $\boldsymbol{x}_\lambda$ on H, one of the hyper-surfaces in $\mathcal{D}_{\boldsymbol{x}}$. That is, $\boldsymbol{x}_\lambda$ satisfies (14) for a particular i_0 with $\boldsymbol{x}^*$ being replaced by $\boldsymbol{x}_\lambda$ and all the other relevant inequalities in Proposition 2, with $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ replaced by $\boldsymbol{x}_\lambda$ and $\boldsymbol{x}$, are strict. If $\boldsymbol{x} \notin \mathcal{D}_{\boldsymbol{x}^*}$ so that $\boldsymbol{x}^* \notin \mathcal{D}_{\boldsymbol{x}}$, we may assume that the points $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^j)$ lie on the opposite sides of H for all sufficiently small $\lambda > 0$. Then, by Proposition 4, as γ_λ moves through $\boldsymbol{x}_\lambda$, the supports of the geodesics from γ_λ to $\boldsymbol{x}$ change, with the relevant subset $A_{i_0}^1 = C_{i_0 1} \cup C_{i_0 2}$ of the sequence $\mathcal{A}^1 = (A_0^1, \cdots, A_k^1)$ in the support $(\mathcal{A}^1, \mathcal{B}^1)$ on the one side splitting, say, into two subsets $C_{i_0 1}, C_{i_0 2}$ on the other, and similarly for $B_{i_0}^1$ in $\mathcal{B}^1$. That is, the support $(\mathcal{A}^2, \mathcal{B}^2)$ of the geodesics from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^2)$ to $\boldsymbol{x}$ is related to $(\mathcal{A}^1, \mathcal{B}^1)$ by $\mathcal{A}^2 = (A_0^1, \cdots, A_{i_0-1}^1, C_{i_0 1}, C_{i_0 2}, A_{i_0+1}^1, \cdots, A_k^1)$, and similarly $\mathcal{B}^2$ to $\mathcal{B}^1$. We show now that neither of these subsets $C_{i_0 1}$ and $C_{i_0 2}$ can meet E. If only one of these two sets meets E, say $C_{i_0 1}$, then since $P_{C_{i_0 2}}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^1)) \to 0$ as $\lambda \to 0$, it is impossible that there are $\boldsymbol{x}_\lambda$ such that the corresponding equality (14) holds for all sufficiently small $\lambda > 0$. Similarly, if both of these sets meet E, then the proof of Corollary 4(*ii*) shows that $P_{C_{i_0 s}}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^j)) \to P_{C_{i_0 s}}(\boldsymbol{x}^*)$, as $\lambda \to 0$, for $j = 1, 2$. This implies that, for $j = 1$, the corresponding strict inequality (7) holds for $\boldsymbol{x}^*$ while, for $j = 2$, it is reversed. Hence, that is also impossible.

Thus, in the case when the supports $(\mathcal{A}^i, \mathcal{B}^i)$ are different, we still have $A_j^1 = A_j^2$ for all $j > 0$ such that $A_j^1 \cap E \neq \emptyset$.

If further $\boldsymbol{x} \notin \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^1) \bigcup \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^2)$, then the change of the support described above cannot happen when both $C_{i_0 1} \cap E$ and $C_{i_0 2} \cap E$ are empty, as then $|A_{i_0}^1 \cap F| > 1$, and so we would have $\boldsymbol{x} \in \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^1)$. Since $A_0^1 = A_0^2$, the above implies that we must have $(\mathcal{A}^i, \mathcal{B}^i)$ identical for $i = 1, 2$ and so, for such $\boldsymbol{x}$, $\Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^2; \boldsymbol{x}^*) = \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau^1; \boldsymbol{x}^*)$.

Next, assume that the two $\boldsymbol{w}_\tau^j$ are chosen to be sufficiently close that, for any given $\boldsymbol{x}$ and all sufficiently small $\lambda > 0$, the geodesics from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^j)$ to $\boldsymbol{x}$ have the same support. Then, if $\boldsymbol{w}_\tau(\alpha)$, $\alpha \in [0, 1]$, is the geodesic between $\boldsymbol{w}_\tau^1$ and $\boldsymbol{w}_\tau^2$, an elementary argument on the relevant parameters in the inequalities (5) and (7) that determine the carrier will show that these parameters are monotonic

in α along the geodesic. So, the geodesic from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau(\alpha))$ to $\boldsymbol{x}$ will have the same support as that for the geodesics from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau^j)$ to $\boldsymbol{x}$. This implies that $\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau(\alpha)) \subseteq \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^1) \bigcup \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^2)$, so that $\Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau(\alpha); \boldsymbol{x}^*)$ are a.s. independent of $\alpha \in [0, 1]$.

Finally, since $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$ is convex, there is a sequence $\{\boldsymbol{w}_\tau^n \mid n \geqslant 1\} \subset \text{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$ such that

$$\text{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)) = \lim_{n \to \infty} C_n, \tag{33}$$

where C_n is the convex hull in $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ of $\{\boldsymbol{w}_\tau^1, \cdots, \boldsymbol{w}_\tau^n\}$. The above argument implies that, without loss of generality, we may also assume that $\{\boldsymbol{w}_\tau^n \mid n \geqslant 1\}$ have the property that, for any $\boldsymbol{w}_\tau \in C_n$,

$$\Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau) \subseteq \bigcup_{i=1}^n \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^i).$$

This shows that

$$\mu\left(\bigcup_{\boldsymbol{w}_\tau \in C_n} \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau) \right) = \mu\left(\bigcup_{i=1}^n \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau^i) \right) = 0,$$

so that $\Psi_\tau(\boldsymbol{\xi}, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ are a.s. independent of $\boldsymbol{w}_\tau \in C_n$. Hence, it follows from (33) that

$$\mu\left(\bigcup_{\boldsymbol{w}_\tau \in \text{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))} \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau) \right) = \lim_{n \to \infty} \mu\left(\bigcup_{\boldsymbol{w}_\tau \in C_n} \Sigma_{\tau,\sigma}(\boldsymbol{x}^*; \boldsymbol{w}_\tau) \right) = 0,$$

which gives the required result. $\qquad\square$

7 The limiting distribution of sample Fréchet means

In this section, we assume that $\{\boldsymbol{\xi}_i : i \geqslant 1\}$ is a sequence of *i.i.d.* random variables defined on a common probability space $(\boldsymbol{\Omega}, \mathcal{F}, \mathbf{P})$ with values in $\boldsymbol{X}^m$; that μ is the distribution measure of $\boldsymbol{\xi}_1$; and that $\hat{\boldsymbol{\xi}}_n$ is the sample Fréchet mean of $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_n$. Then, $\hat{\boldsymbol{\xi}}_n$ converges to the Fréchet mean $\boldsymbol{x}^*$ of μ almost surely as n tends to infinity (cf. [23]).

7.1 On the support of the limiting distribution

If $\boldsymbol{x}^*$ lies in a top-dimensional stratum, $\boldsymbol{X}^m$ is locally an m-dimensional manifold. One would expect that the limiting behaviour of sample Fréchet means $\hat{\boldsymbol{\xi}}_n$ is similar, to some extent, to that of sample Fréchet means in a Riemannian manifold as obtained in [4] and [13]. In particular, the support of the limiting distribution of $\sqrt{n} \log_{x^*}(\hat{\boldsymbol{\xi}}_n) = \sqrt{n}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*)$ is the entire tangent space to $\boldsymbol{X}^m$ at $\boldsymbol{x}^*$, as long as $\text{cov}(\Phi(\boldsymbol{\xi}_1; \boldsymbol{x}^*))$ has rank m. This fact was proved for the case of open books in [10] and for the case of tree spaces in [2] and [3]. We shall

see in the following that the argument used in [3] can be generalised to $\boldsymbol{X}^m$, so that the corresponding conclusion is also valid for orthant spaces.

However, when $\boldsymbol{x}^*$ lies in a stratum of locally positive co-dimension, the limiting behaviour of sample Fréchet means is generally very different. In the case that $\boldsymbol{X}^m$ is an open book or a tree space and that the stratum containing $\boldsymbol{x}^*$ is of the co-dimension one, this phenomenon was observed and studied in [10], [2] and [3]. Similarly, for general orthant spaces, the strictness or otherwise of the inequality (27) affects the limiting behaviour of $\hat{\boldsymbol{\xi}}_n$. In particular, when (27) is strict, there is a constraint on the support of the limiting distribution. To describe this, we recall that, for $\sigma = \mathcal{O}(E)$ of co-dimension l and $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l' < l$ co-bounding σ, we are denoting the set of unit vectors in $\mathbb{R}(E) \times \mathcal{O}(F)$ by $\mathcal{S}_{\tau,\sigma}^{m-l'}$ and those in $\{\boldsymbol{0}\} \times \mathcal{O}(F)$ by $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$. Then, for $\boldsymbol{w}_\tau$ in the latter, denote by $\mathcal{H}_{\boldsymbol{w}_\tau}$ the intersection of the half hyper-plane $\mathbb{R}(E) \times \{c\,\boldsymbol{w}_\tau \mid c > 0\}$ with $\mathcal{S}_{\tau,\sigma}^{m-l'}$, namely

$$\mathcal{H}_{\boldsymbol{w}_\tau} = \left\{ \boldsymbol{w}_\sigma + c\,\boldsymbol{w}_\tau \in \mathcal{S}_{\tau,\sigma}^{m-l'} \;\middle|\; c > 0 \text{ and } \boldsymbol{w}_\sigma \in \mathbb{R}(E) \times \{\boldsymbol{0}\} \subset \mathbb{R}(E) \times \mathcal{O}(F) \right\}.$$

Also, for such a $\boldsymbol{w}_\tau$, denote by $\Omega_n^k(\boldsymbol{w}_\tau)$ the subset of Ω that consists of all ω such that (i) the corresponding sample Fréchet means $\hat{\boldsymbol{\xi}}_n$ lie in the stratum τ, and (ii) the distance to the set $\mathcal{H}_{\boldsymbol{w}_\tau}$ of the initial direction of the geodesic from $\hat{\boldsymbol{\xi}}_n$ to the Fréchet mean $\boldsymbol{x}^*$ of μ is at most $1/k$, i.e.

$$\Omega_n^k(\boldsymbol{w}_\tau) = \left\{ \omega \in \boldsymbol{\Omega} \mid \hat{\boldsymbol{\xi}}_n(\omega) \in \tau \text{ and } d\left(\frac{\hat{\boldsymbol{\xi}}_n(\omega) - \boldsymbol{x}^*}{\|\hat{\boldsymbol{\xi}}_n(\omega) - \boldsymbol{x}^*\|}, \mathcal{H}_{\boldsymbol{w}_\tau} \right) \leqslant \frac{1}{k} \right\}.$$

Proposition 10. *Let the stratum* $\sigma = \mathcal{O}(E)$ *of co-dimension* $l(\geqslant 1)$ *bound, in* $\boldsymbol{X}^m$, *the stratum* $\tau = \mathcal{O}(E \cup F)$ *of co-dimension* $l'(< l)$. *Assume that the Fréchet mean* $\boldsymbol{x}^*$ *of* μ *lies in* σ *and that* $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'} \setminus \Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$, *where* $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ *and* $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$ *are given by Definitions 4 and 13 respectively. Then,*

$$\lim_{k \to \infty} \mathbf{P}\left(\limsup_n \Omega_n^k(\boldsymbol{w}_\tau) \right) = 0.$$

Proof. For $\boldsymbol{w}_\tau$ as given in the proposition, let

$$\Omega_{\boldsymbol{w}_\tau} = \bigcap_{k \geqslant 1} \limsup_n \Omega_n^k(\boldsymbol{w}_\tau) = \bigcap_{k \geqslant 1} \bigcap_{l \geqslant 1} \bigcup_{n \geqslant l} \Omega_n^k(\boldsymbol{w}_\tau).$$

Then, the set $\Omega_{\boldsymbol{w}_\tau}$ consists of points with the property that, for arbitrary $\epsilon > 0$, there exist arbitrarily large n such that $\hat{\boldsymbol{\xi}}_n$ lies in τ and $(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*)/\|\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*\|$ is within a distance ϵ of $\mathcal{H}_{\boldsymbol{w}_\tau}$. Since $\Omega_n^k(\boldsymbol{w}_\tau) \supseteq \Omega_n^{k+1}(\boldsymbol{w}_\tau)$, the required result is equivalent to showing that $\mathbf{P}(\Omega_{\boldsymbol{w}_\tau}) = 0$.

Without loss of generality, we may assume that, restricted to $\Omega_{\boldsymbol{w}_\tau}$, $\hat{\boldsymbol{\xi}}_n$ lie in τ for all n and $\boldsymbol{w}_n = (\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*)/\|\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*\| \to \boldsymbol{w}$ as $n \to \infty$ for some (random) unit vector $\boldsymbol{w} \in \mathcal{H}_{\boldsymbol{w}_\tau}$.

Recall that, for given $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'} \setminus \Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$, each $\Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ is a Euclidean random variable on $\mathbb{R}(E \cup F)$. Then, let

$$\hat{\boldsymbol{\xi}}_n^{\boldsymbol{w}_\tau} = \frac{1}{n} \sum_{i=1}^n \Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$$

and write Ω_0 for the subset of $\boldsymbol{\Omega}$ consisting of points such that $\hat{\boldsymbol{\xi}}_n^{\boldsymbol{w}_\tau}$ converges to

$$\int_{\boldsymbol{X}^m} \Psi_\tau(\boldsymbol{x}, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \, \mathrm{d}\mu(\boldsymbol{x}).$$

It follows from the classical Law of Large Numbers that $\mathbf{P}(\Omega_0) = 1$. Hence, restricted to $\Omega_{\boldsymbol{w}_\tau} \cap \Omega_0$, the assumption on $\boldsymbol{w}_\tau$ implies that, for some constant $c < 0$, there is an n_0 such that, for $n > n_0$, $\langle \boldsymbol{w}_\tau, \hat{\boldsymbol{\xi}}_n^{\boldsymbol{w}_\tau} \rangle < c$. However, the assumption that $\boldsymbol{w}_n \to \boldsymbol{w} \in \mathcal{H}_{\boldsymbol{w}_\tau}$ implies that $\langle \boldsymbol{w}_\tau, \hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^* \rangle > 0$ for all sufficiently large n. Putting these two conclusions together, we have that, restricted to $\Omega_{\boldsymbol{w}_\tau} \cap \Omega_0$,

$$\left\langle \boldsymbol{w}_\tau, \hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\xi}}_n^{\boldsymbol{w}_\tau} \right\rangle = \left\langle \boldsymbol{w}_\tau, \hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^* \right\rangle - \left\langle \boldsymbol{w}_\tau, \hat{\boldsymbol{\xi}}_n^{\boldsymbol{w}_\tau} \right\rangle > -c > 0 \tag{34}$$

as $\langle \boldsymbol{w}_\tau, \boldsymbol{x}^* \rangle = 0$.

On the other hand, restricted to $\Omega_{\boldsymbol{w}_\tau} \cap \Omega_0$, $\hat{\boldsymbol{\xi}}_n$ is in τ by the assumption made earlier. Then, it follows from (26) that

$$\hat{\boldsymbol{\xi}}_n = \frac{1}{n} \sum_{i=1}^n \Phi_\tau(\boldsymbol{\xi}_i; \hat{\boldsymbol{\xi}}_n).$$

Thus, we can express the difference $\hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\xi}}_n^{\boldsymbol{w}_\tau}$ as

$$\hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\xi}}_n^{\boldsymbol{w}_\tau} = \frac{1}{n} \sum_{i=1}^n \left\{ \Phi_\tau(\boldsymbol{\xi}_i; \hat{\boldsymbol{\xi}}_n) - \Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \right\}. \tag{35}$$

Decompose $\boldsymbol{w}_n = (\boldsymbol{w}_n)_\sigma + (\boldsymbol{w}_n)^\perp$, where $(\boldsymbol{w}_n)_\sigma = P_\sigma(\boldsymbol{w}_n)$ and $(\boldsymbol{w}_n)^\perp = P_{\tau \setminus \sigma}(\boldsymbol{w}_n)$. Then, by Corollary 3(ii), for each $1 \leqslant i \leqslant n$,

$$\begin{aligned} &\Phi_\tau(\boldsymbol{\xi}_i; \hat{\boldsymbol{\xi}}_n) - \Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \\ =\ &\Phi_\tau(\boldsymbol{\xi}_i; \hat{\boldsymbol{\xi}}_n) - \Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_n; \boldsymbol{x}^*) + \Psi_\tau(\boldsymbol{\xi}_i, (\boldsymbol{w}_n)_\tau; \boldsymbol{x}^*) - \Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_\tau; \boldsymbol{x}^*), \end{aligned} \tag{36}$$

where $(\boldsymbol{w}_n)_\tau = (\boldsymbol{w}_n)^\perp / \|(\boldsymbol{w}_n)^\perp\| \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$. Without loss of generality, we assume that the carriers of the geodesics from $\hat{\boldsymbol{\xi}}_n$ to $\boldsymbol{\xi}_i$ remain constant. The general case follows from a similar inductive argument to that outlined in the beginning of the proof of Proposition 9 and from the fact that $\hat{\boldsymbol{\xi}}_n$ converges to $\boldsymbol{x}^*$ a.s. Then, if $(\mathcal{A}, \mathcal{B})$ is the common support of the geodesics, where $\mathcal{A} = (A_0, \cdots, A_k)$ and $\mathcal{B} = (B_0, \cdots, B_k)$, and, for $0 < j \leqslant k$, writing W_j for $P_{A_j \cap E}(\boldsymbol{x}^*)$ if $A_j \cap E \neq \emptyset$ and otherwise $P_{A_j \cap F}(\boldsymbol{w}_n)$, Theorem 1 tells us that the jth set of components of $(\jmath^{-1})(\Phi(\boldsymbol{\xi}_i, \hat{\boldsymbol{\xi}}_n))$ is the vector $-\frac{\|P_{B_j}(\boldsymbol{\xi}_i)\|}{\|P_{A_j}(\hat{\boldsymbol{\xi}}_n)\|} P_{A_j}(\hat{\boldsymbol{\xi}}_n)$ and, for $(\jmath^{-1})(\Psi(\boldsymbol{\xi}_i, \boldsymbol{w}_n; \boldsymbol{x}^*))$, Theorem 2(ii) tells us that the corresponding vector is $-\frac{\|P_{B_j}(\boldsymbol{\xi}_i)\|}{\|W_j\|} W_j$. Hence, the proof of Theorem 2(ii) shows that, when $A_j \cap E = \emptyset$, these two vectors are identical, since $P_{A_j}(\hat{\boldsymbol{\xi}}_n) = P_{A_j \cap F}(\hat{\boldsymbol{\xi}}_n) = P_{A_j \cap F}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*)$. While, if $A_j \cap E \neq \emptyset$, the difference between these two vectors is of the same order as $\frac{P_{A_j \cap F}(\hat{\boldsymbol{\xi}}_n)}{\|P_{A_j}(\hat{\boldsymbol{\xi}}_n)\|}$ whose limit, as $n \to \infty$, is zero since $\|P_{A_j}(\hat{\boldsymbol{\xi}}_n)\| \geqslant \|P_{A_j \cap E}(\hat{\boldsymbol{\xi}}_n)\| \to \|P_{A_j \cap E}(\boldsymbol{x}^*)\| > 0$ but $P_{A_j \cap F}(\hat{\boldsymbol{\xi}}_n) \to 0$ a.s. It follows that, as $n \to \infty$,

$$\left\langle \boldsymbol{w}_\tau, \Phi_\tau(\boldsymbol{\xi}_i; \hat{\boldsymbol{\xi}}_n) - \Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_n; \boldsymbol{x}^*) \right\rangle \longrightarrow 0 \quad \text{a.s.} \tag{37}$$

34

Moreover, since $\boldsymbol{w}^\perp = P_{\tau \setminus \sigma}(\boldsymbol{w}) \neq 0$, $\boldsymbol{w}_n \to \boldsymbol{w}$ implies that $(\boldsymbol{w}_n)_\tau \to \frac{\boldsymbol{w}^\perp}{\|\boldsymbol{w}^\perp\|} = \boldsymbol{w}_\tau$. Then, it follows from a similar argument to that of the proof of Proposition 6 that, for sufficiently large n,

$$\langle \boldsymbol{w}_\tau,\, \Psi_\tau(\boldsymbol{\xi}_i, (\boldsymbol{w}_n)_\tau; \boldsymbol{x}^*) - \Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \rangle$$
$$\approx \left\langle \boldsymbol{w}_\tau,\, (D\Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_\tau; \boldsymbol{x}^*)) \left(\arccos(\langle (\boldsymbol{w}_n)_\tau, \boldsymbol{w}_\tau \rangle) \frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|} \right) \right\rangle,$$

where $\boldsymbol{v}_n$ is the component of $(\boldsymbol{w}_n)_\tau - \boldsymbol{w}_\tau$ orthogonal to $\boldsymbol{w}_\tau$, so that as $n \to \infty$,

$$\langle \boldsymbol{w}_\tau,\, \Psi_\tau(\boldsymbol{\xi}_i, (\boldsymbol{w}_n)_\tau; \boldsymbol{x}^*) - \Psi_\tau(\boldsymbol{\xi}_i, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \rangle \longrightarrow 0 \quad \text{a.s.}, \tag{38}$$

by Proposition 6. Then, (35), (36), (37) and (38) together imply that, when it is restricted to $\Omega_{\boldsymbol{w}_\tau} \cap \Omega_0$, $\langle \boldsymbol{w}_\tau, \hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\xi}}_n^{\boldsymbol{w}_\tau} \rangle \to 0$ a.s., as $n \to \infty$, contradicting (34). Hence, $\mathbf{P}(\Omega_{\boldsymbol{w}_\tau} \cap \Omega_0) = 0$, so that $\mathbf{P}(\Omega_{\boldsymbol{w}_\tau}) = 0$ as required. $\qquad\square$

When $l - l' = 1$, $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ contains a single unit vector, so that we have the following special case: under certain conditions, the sample Fréchet mean cannot be in any co-bounding stratum when the sample size is sufficiently large. In particular, taking $l = 1$ and so $l' = 0$ recovers the result of Lemma 6 in [3] for the case of co-dimension one when $\boldsymbol{X}^m$ is a tree space.

Corollary 7. *Let the stratum σ of co-dimension $l(\geqslant 1)$ bound, in $\boldsymbol{X}^m$, the stratum τ of co-dimension $l' = l - 1$ co-bounding σ. Assume that the Fréchet mean $\boldsymbol{x}^*$ of μ lies in σ. If the inequality (27) corresponding to the unique $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is strict then, for all sufficiently large n, $\hat{\boldsymbol{\xi}}_n$ cannot lie in τ.*

Thus, when $l - l' = 1$, the support of the limiting distribution of any appropriately scaled difference $\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*$ intersects the stratum $\mathbb{R}(E) \times \mathcal{O}(F)$ in the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x}^*$ only if the inequality (27) corresponding to the unique $\boldsymbol{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is an equality.

Similar to the case where $l - l' = 1$, Proposition 10 has the following consequence on the support of the limiting distribution when $l - l' > 1$, where $\mathcal{C}(\Theta)$ denotes the Euclidean cone on Θ.

Corollary 8. *Let the stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l(\geqslant 2)$ bound, in $\boldsymbol{X}^m$, the stratum $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l' \leqslant l - 2$. Assume that $\boldsymbol{x}^* \in \sigma$ is the Fréchet mean of μ. Then the support of the limiting distribution of an appropriately scaled difference $\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*$, if it meets the stratum $\mathbb{R}(E) \times \mathcal{O}(F)$ in the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x}^*$, must be contained in $\mathbb{R}(E) \times \mathcal{C}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$, where $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$ is defined by (29).*

Hence, the support of the limiting distribution of an appropriately scaled difference $\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*$ is contained in $\mathcal{K}_\mu$ where, for the closed sets

$$\mathcal{K}_{\mu,\tau} = \mathbb{R}(E) \times \mathcal{C}\left(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)\right) \tag{39}$$

in the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x}^*$,

$$\mathcal{K}_\mu = \bigcup_{\tau \text{ co-bounds } \sigma} \mathcal{K}_{\mu,\tau} \tag{40}$$

and where we regard σ as co-bounding itself. Nevertheless, the following example shows that
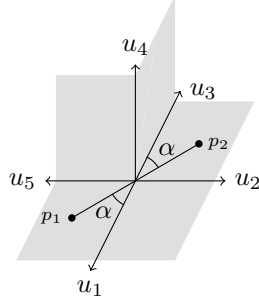
Figure 4: Probability measure μ on $\boldsymbol{X}^2$ has mass $1/2$ at p_1 and p_2.

(i) if it is non-empty, $\mathbb{R}(E) \times \mathcal{C}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$ is not necessarily an entire stratum $\mathbb{R}(E) \times \mathcal{O}(F)$;

(ii) even if it is the entire stratum, the support of the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*)$ does not necessarily intersect that stratum; and

(iii) it is possible that the support of the limiting distribution, when restricted to the stratum, is only a subset of $\mathbb{R}(E) \times \mathcal{C}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$.

Example 4. Consider the orthant space $\boldsymbol{X}^2$ of Example 2. Let μ have mass $1/2$ at the two points p_1 and p_2 equidistant from the cone point o along a geodesic through that point as illustrated in Figure 4. Then its Fréchet mean is at the cone point and the sample Fréchet means always lie on this geodesic segment. This, in particular, implies that the support of the limiting distribution of $\sqrt{n}\{\hat{\boldsymbol{\xi}}_n - o\}$ is the union of the cone point with two half lines, one in $\{o\} \times \tau_{1,5}$ and other in $\{o\} \times \tau_{2,3}$, each extending the relevant geodesic segment, where $\tau_{i,j} = \mathcal{O}(u_i, u_j)$.

(a) For any direction $\boldsymbol{w}_{\tau_{1,2}} \in \mathcal{S}^2_{\tau_{1,2}\setminus\{o\}}$, $\Psi_{\tau_{1,2}}(p, \boldsymbol{w}_{\tau_{1,2}}; o)$ lies in the plane spanned by the orthant $\tau_{1,2}$, for any p, and, by identifying u_3 and u_5 with $-u_1$ and $-u_2$ respectively, $\Psi_{\tau_{1,2}}(p_i, \boldsymbol{w}_{\tau_{1,2}}; o) = p_i$ for $i = 1, 2$. Thus, $\int_{\boldsymbol{X}^2} \Psi_{\tau_{1,2}}(p, \boldsymbol{w}_{\tau_{1,2}}; o) \, d\mu(p) = 0$ and so $\Theta_{\tau_{1,2},\{o\}}(o; \mu) = \mathcal{S}^2_{\tau_{1,2}\setminus\{o\}}$. Since, in this case, the support of the limiting distribution does not intersect $\{0\} \times \tau_{1,2}$, this illustrates (ii) above with $\sigma = \{o\}$ and $\tau = \tau_{1,2}$.

(b) For any direction $\boldsymbol{w}_{\tau_{1,5}} \in \mathcal{S}^2_{\tau_{1,5}\setminus\{o\}}$ such that the angle between $\boldsymbol{w}_{\tau_{1,5}}$ and u_1-axis is less than or equal α, a similar argument shows that

$$\int_{\boldsymbol{X}^2} \Psi_{\tau_{1,5}}(p, \boldsymbol{w}_{\tau_{1,5}}; o) \, d\mu(p) = 0.$$

Hence, such $\boldsymbol{w}_{\tau_{1,5}}$ are always contained in $\Theta_{\tau_{1,5},\{o\}}(o; \mu)$, i.e.

$$\Theta_{\tau_{1,5},\{o\}}(o; \mu) \supseteq \{\theta \in \mathcal{S}^2_{\tau_{1,5}\setminus\{o\}} \mid \theta \leqslant \alpha\}$$

where $\theta \in \mathcal{S}^2_{\tau_{1,5}} \setminus \{o\}$ is measured from the u_1-axis.

(c) However, for any direction $\boldsymbol{w}_{\tau_{1,5}} \in \mathcal{S}^2_{\tau_{1,5}\backslash\{o\}}$ such that the angle between $\boldsymbol{w}_{\tau_{1,5}}$ and the u_1-axis is greater than α, the vector $\Psi_{\tau_{1,5}}(p_1, \boldsymbol{w}_{\tau_{1,5}}; o) = p_1$, but the vector $\Psi_{\tau_{1,5}}(p_2, \boldsymbol{w}_{\tau_{1,5}}; o)$ lies on the line spanned by the unit vector $\boldsymbol{w}_{\tau_{1,5}}$ in (u_1, u_5)-plane. Hence, these two vectors do not lie on the same line in the (u_1, u_5)-plane through the origin. This gives

$$\langle \boldsymbol{w}_{\tau_{1,5}}, \int_{\boldsymbol{X}^2} \Psi_{\tau_{1,5}}(p, \boldsymbol{w}_{\tau_{1,5}}; o)\, \mathrm{d}\mu(p) \rangle < 0.$$

Hence, if the angle between $\boldsymbol{w}_{\tau_{1,5}}$ and the u_1-axis is greater than α, then $\boldsymbol{w}_{\tau_{1,5}} \notin \Theta_{\tau_{1,5},\{o\}}(o; \mu)$. Combining this with the conclusion (b) shows that $\Theta_{\tau_{1,5},\{o\}}(o; \mu) = \{\theta \in \mathcal{S}^2_{\tau_{1,5}\backslash\{o\}} \mid \theta \leqslant \alpha\}$, illustrating (i) and (iii) above with $\sigma = \{o\}$ and $\tau = \tau_{1,5}$.

7.2 The limiting distribution

To describe the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*)$, where the Fréchet mean $\boldsymbol{x}^*$ of μ lies in a stratum $\sigma = \mathcal{O}(E)$ of local co-dimension $l \geqslant 0$, we continue to regard σ as co-bounding itself so that, in this case, the set F of additional axes in the 'co-bounding' stratum is empty. Moreover, we shall relate the form of the limiting distribution in the set (39) for each τ co-bounding σ to a limiting distribution of the Euclidean means of various Euclidean random variables depending on τ:

(i) for $\tau = \sigma$, corresponding to the set $\mathbb{R}(E) \times \{\boldsymbol{0}\}$ in (39), the relevant Euclidean random variable is $\Phi_\sigma(\boldsymbol{\xi}_1; \boldsymbol{x}^*)$;

(ii) for $\tau \neq \sigma$ the relevant Euclidean random variable is $\Psi_\tau(\boldsymbol{\xi}_1, \boldsymbol{w}_\tau; \boldsymbol{x}^*)$ where, if $l - l' > 1$, $\boldsymbol{w}_\tau$ is any chosen vector in $\mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu))$ if this set is not empty and, if $l - l' = 1$ with $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu) \neq \emptyset$, $\boldsymbol{w}_\tau$ is its unique element;

(iii) we take the zero random variable otherwise.

Note that, by Proposition 9, different choices of $\boldsymbol{w}_\tau$ in the case $l - l' > 1$ of (ii) give random variables that are a.s. equal. Note also that, by Corollary 8, the random variables in the case (iii) play no role in the description of the limiting distribution so that they can be replaced by any other random variables. For simplicity, we denote the relevant random variable above in each case by $\tilde{\Psi}_\tau(\boldsymbol{\xi}_1; \boldsymbol{x}^*)$. With this context and notation write, for each τ co-bounding σ,

$$A_{\sigma,\tau}^{-1} = U_\tau^\top \{I_M - E[M_{\boldsymbol{x}^*}^\sigma(\boldsymbol{\xi}_1)]\} U_\tau, \tag{41}$$

where $M_{\boldsymbol{x}^*}^\sigma(x)$ is defined by (17), U_τ is the $M \times (m-l')$ matrix whose entries are all zero except for those at (l_i, i) being one, and $u_{l_1}, \cdots, u_{l_{m-l'}}$ are the ordered axes that span $\jmath(\mathbb{R}(E \cup F))$. Note that, since $M_{\boldsymbol{x}^*}^\sigma(\boldsymbol{x})$ is negative semi-definite, the above inverse is well defined when $E[M_{\boldsymbol{x}^*}^\sigma(\boldsymbol{\xi}_1)]$ exists. Then, letting Z_τ be a random variable in $\mathbb{R}(E \cup F)$ with normal distribution $N(0, A_{\sigma,\tau}^\top V_\tau A_{\sigma,\tau})$, where $V_\tau = \mathrm{cov}(\tilde{\Psi}_\tau(\boldsymbol{\xi}_1; \boldsymbol{x}^*))$, we have the following result.

Theorem 4. *Let $\sigma = \mathcal{O}(E)$ be a stratum in $\boldsymbol{X}^m$ of co-dimension $l(\geqslant 0)$. Assume that*

(i) *the Fréchet mean $\boldsymbol{x}^*$ of μ lies in σ;*

(ii) *$\mu(\mathcal{D}_{\boldsymbol{x}^*}) = 0$, where $\mathcal{D}_{\boldsymbol{x}^*}$ is given by Definition 11;*

(iii) *$E\left[M^\sigma_{\boldsymbol{x}^*}(\boldsymbol{\xi}_1)\right]$ exists, where $M^\sigma_{\boldsymbol{x}^*}(\boldsymbol{x})$ is given by (17);*

(iv) *for any stratum τ in $\boldsymbol{X}^m$ which co-bounds σ and has co-dimension $l' \leqslant l - 2$, if $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu) \neq \emptyset$ then $\mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)) \neq \emptyset$.*

Then, if there exists a random variable η on the tangent cone at $\boldsymbol{x}^*$ such that

$$\sqrt{n}\{\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*\} \xrightarrow{d} \eta, \qquad as \ n \to \infty,$$

then η has the following property: for any stratum $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l'(\leqslant l)$ co-bounding σ, if $\mathbf{P}\left(\eta \in \mathbb{R}(E) \times \mathcal{O}(F)\right) > 0$ then, for Z_τ defined as above and $\mathcal{K}_\mu$ by (40),

$$\mathbf{P}(\eta \in B) = \mathbf{P}(Z_\tau \in B)$$

for any Borel set B contained in

$$\begin{cases} (\mathrm{int}\,(\mathbb{R}(E) \times \mathcal{C}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)))) \setminus \partial\mathcal{K}_\mu & \textit{if } l' \leqslant l - 2 \\ (\mathbb{R}(E) \times \mathcal{O}(F)) \setminus \partial\mathcal{K}_\mu & \textit{if } l' = l - 1 \textit{ or } l. \end{cases}$$

Intuitively, the result of the theorem says that, under certain technical conditions, the distribution of the limiting random variable η on the tangent cone to $\boldsymbol{X}^m$ at the Fréchet mean $\boldsymbol{x}^*$ of μ is closely related to normal distributions. More precisely, if $\boldsymbol{x}^*$ lies in a stratum σ and if, for a given co-bounding stratum τ, the distribution of η has positive mass on the corresponding stratum in the tangent cone then, when restricted to the interior of the support of the distribution of η, η behaves like a normal random variable.

Proof. We assume that $l' \leqslant l - 2$. The case for $l' = l - 1$ can be similarly derived by noting Corollary 7, whereas for $l' = l$ the result can be derived directly by simplifying the following arguments.

Write $\Xi_\tau = \mathcal{O}(E) \times \mathcal{C}(\mathrm{int}(\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)))$. By Corollary 7, given $\hat{\boldsymbol{\xi}}_n \in \tau$, $\hat{\boldsymbol{\xi}}_n \in \Xi_\tau$ for sufficiently large n and, by Theorem 3, we also have

$$\sqrt{n}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \Phi_\tau(\boldsymbol{\xi}_i; \hat{\boldsymbol{\xi}}_n) - \tilde{\Psi}_\tau(\boldsymbol{\xi}_i; \boldsymbol{x}^*) \right\}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \tilde{\Psi}_\tau(\boldsymbol{\xi}_i; \boldsymbol{x}^*) - \boldsymbol{x}^* \right\}.$$

For any $\boldsymbol{x}' \in \tau$ and $\boldsymbol{x} \in \boldsymbol{X}^m$, denote the projection $P_{\tau \setminus \sigma}\left(\Phi_\tau(\boldsymbol{x}; \boldsymbol{x}')\right)$ of $\Phi_\tau(\boldsymbol{x}; \boldsymbol{x}')$ by $\Phi_{\tau \setminus \sigma}(\boldsymbol{x}; \boldsymbol{x}')$. Define $\tilde{\Psi}_{\tau \setminus \sigma}(\boldsymbol{x}; \boldsymbol{x}^*)$ similarly. Then, $\tilde{\Psi}_{\tau \setminus \sigma}(\boldsymbol{x}; \boldsymbol{x}^*) = \tilde{\Psi}_\tau(\boldsymbol{x}; \boldsymbol{x}^*) - \Phi_\sigma(\boldsymbol{x}; \boldsymbol{x}^*)$ by Corollary 4(ii). Since $\hat{\boldsymbol{\xi}}_n$ is in Ξ_τ and converges to $\boldsymbol{x}^*$ a.s., the result of Proposition 9 and the argument for the proof of Theorem 2 together imply that, for any given $\boldsymbol{x}$ and all sufficiently large n, $\Phi_{\tau \setminus \sigma}(\boldsymbol{x}; \hat{\boldsymbol{\xi}}_n) = \tilde{\Psi}_{\tau \setminus \sigma}(\boldsymbol{x}; \boldsymbol{x}^*)$ a.s.. Hence, in particular, for sufficiently large n,

$$P_{\tau \setminus \sigma}\left(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*\right) = \frac{1}{n} \sum_{i=1}^{n} \Phi_{\tau \setminus \sigma}(\boldsymbol{\xi}_i; \hat{\boldsymbol{\xi}}_n)$$

is a.s. the Euclidean mean of $\tilde{\Psi}_{\tau\backslash\sigma}(\boldsymbol{\xi}_1; \boldsymbol{x}^*), \cdots, \tilde{\Psi}_{\tau\backslash\sigma}(\boldsymbol{\xi}_n; \boldsymbol{x}^*)$, so that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \Phi_{\tau\backslash\sigma}(\boldsymbol{\xi}_i; \hat{\boldsymbol{\xi}}_n) - \tilde{\Psi}_{\tau\backslash\sigma}(\boldsymbol{\xi}_i; \boldsymbol{x}^*) \right\} 1_{\Xi_\tau}(\hat{\boldsymbol{\xi}}_n) \xrightarrow{\mathbf{P}} 0.$$

Thus, the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*) 1_{\Xi_\tau}(\hat{\boldsymbol{\xi}}_n)$ is the same as that of

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \Phi_\sigma(\boldsymbol{\xi}_i; \hat{\boldsymbol{\xi}}_n) - \tilde{\Psi}_\sigma(\boldsymbol{\xi}_i; \boldsymbol{x}^*) \right\} 1_{\Xi_\tau}(\hat{\boldsymbol{\xi}}_n) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \tilde{\Psi}_\tau(\boldsymbol{\xi}_i; \boldsymbol{x}^*) - \boldsymbol{x}^* \right\} 1_{\Xi_\tau}(\hat{\boldsymbol{\xi}}_n).$$

Since $\tilde{\Psi}_\sigma(\boldsymbol{\xi}_i; \boldsymbol{x}^*) = \Phi_\sigma(\boldsymbol{\xi}_i; \boldsymbol{x}^*)$, Proposition 5 implies that the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*) 1_{\Xi_\tau}(\hat{\boldsymbol{\xi}}_n)$ is equal to that of

$$\sqrt{n} P_\sigma(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*) 1_{\Xi_\tau}(\hat{\boldsymbol{\xi}}_n) \frac{1}{n} \sum_{i=1}^{n} M_{\boldsymbol{x}^*}^\sigma(\boldsymbol{\xi}_i) + 1_{\Xi_\tau}(\hat{\boldsymbol{\xi}}_n) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \tilde{\Psi}_\tau(\boldsymbol{\xi}_i; \boldsymbol{x}^*) - \boldsymbol{x}^* \right\}.$$

Hence, by (32), the required result follows from a similar argument to that used in [2] and [3]. $\qquad\square$

As for $\Theta_\sigma(\boldsymbol{x}^*; \mu)$ defined by (30), the convexity in $\boldsymbol{w}$ of the directional derivative $D(d_{\boldsymbol{x}}^2)(\boldsymbol{w})$ implies that $\mathcal{K}_\mu$ is a convex subset of the tangent cone to $\boldsymbol{X}^m$ at $\boldsymbol{x}^*$. This, together with the structure of an orthant space, implies that the result of Theorem 4 refers to the behaviour of the limiting distribution only within the interior of $\mathcal{K}_\mu$. Its behaviour at the boundaries will depend on how these sets relate to each other and on the shape of the boundary $\partial\mathcal{K}_\mu$.

The assumption in Theorem 4 that $\mu(\mathcal{D}_{\boldsymbol{x}^*}) = 0$ ensures that we are able to employ the so-called delta method for the approximate probability distribution of a function of an asymptotically normal statistical estimator. In principle, it is possible to relax this assumption by using directional derivatives and combining that with the use of the law of the total probability. However, it is clear from the definition of $\mathcal{D}_{\boldsymbol{x}^*}$ that its structure, although conceptually straightforward, is generally more complex than will admit a simple algebraic representation, and the ensuing results will consequently depend heavily on the behaviour of μ on $\mathcal{D}_{\boldsymbol{x}^*}$.

To observe special cases of Theorem 4, let $\sigma = \mathcal{O}(E)$ be a stratum in $\boldsymbol{X}^m$ of co-dimension $l(\geqslant 0)$ in which the Fréchet mean $\boldsymbol{x}^*$ of μ lies, assume that the conditions of Theorem 4 are satisfied and write

$$l(\mu) = \inf\{l' \mid l' = \text{co-dimension of } \tau, \text{ where } \Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu) \neq \emptyset\},$$

where we assume that $l(\mu) = l$ if there is no τ with co-dimension $l' < l$ which satisfies the above required condition. We assume further that, for $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l(\mu)$ co-bounding σ and, if $l(\mu) < l$, with $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu) \neq \emptyset$, $V_\tau = \text{cov}(\tilde{\Psi}_\tau(\boldsymbol{\xi}_1; \boldsymbol{x}^*))$ is of full rank $m - l(\mu)$. Then, it is clear from the proof of Theorem 4 that $\mathbf{P}(\eta \in \mathbb{R}(E) \times \mathcal{O}(F)) > 0$. We shall call $l(\mu)$ the *co-dimension of the support* of the limiting distribution of sample Fréchet means of μ.

Case $l(\mu) = l$: in this case, $\mathcal{K}_\mu = \mathbb{R}(E)$ and the support of the distribution of η is contained in the tangent space of σ. Then, Theorem 4 says

39

that η is a normal random variable with mean zero and covariance matrix $A_{\sigma,\sigma}^{\top} \text{cov}(\Phi_\sigma(\boldsymbol{\xi}_1; \boldsymbol{x}^*)) A_{\sigma,\tau}$, where $A_{\sigma,\tau}$ is defined by (41). This generalises the limiting distribution of $\sqrt{n}\{\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*\}$ when $\boldsymbol{x}^*$ lies in a top-dimensional stratum of a tree space obtained in [3].

Case $l(\mu) = l - 1$ so that $l \geqslant 1$: if $\tau = \mathcal{O}(E \cup F)$ is a stratum of co-dimension $l' = l - 1$ such that $\Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu) \neq \emptyset$, then F contains only one axis. By taking the Borel set $B = \mathbb{R}(E) \times \mathcal{O}(F)$, we see that $\mathbf{P}\left(\eta \in \mathbb{R}(E) \times \mathcal{O}(F)\right) = 1/2$ since the corresponding Z_τ is a normal random variable in $\mathbb{R}^{m-l+1}$ with mean zero. Hence, there are at most two strata of co-dimension $l(\mu)$ co-bounding σ on which infinitely many $\hat{\boldsymbol{\xi}}_n$ lie. Moreover, in the case of there being only one such a stratum, $\mathbf{P}(\eta \in \sigma) = 1/2$ and, in case of two such strata, $\mathbf{P}(\eta \in \sigma) = 0$.

Case that $0 \leqslant l(\mu) < l$, that there is a single $\tau_0 = \mathcal{O}(E \cup F_0)$ such that the co-dimension of τ_0 is $l(\mu)$ and that $\Theta_{\tau_0,\sigma}(\boldsymbol{x}^*; \mu) = \mathcal{S}_{\tau_0 \setminus \sigma}^{l-l(\mu)}$: in this case, we have the following full description of the distribution of η in terms of ϕ_{τ_0}, the probability density function of the random variable Z_{τ_0} defined prior to Theorem 4. We first note that, since $\mathcal{K}_\mu$ defined by (40) is convex and closed, the result of Proposition 9 implies that, in this case,

$$\mathcal{K}_\mu = \bigcup_{F \subseteq F_0} \mathbb{R}(E) \times \mathcal{O}(F) = \mathbb{R}(E) \times \overline{\mathcal{O}(F)}.$$

Then, we extend the projection map P to $\mathbb{R}(E \cup F_0)$ in an obvious fashion and, for any $\tau = \mathcal{O}(E \cup F)$, where $F \subseteq F_0$, and any $\boldsymbol{z} \in \mathbb{R}(E \cup F_0)$, write $\boldsymbol{z}_\tau = P_\tau(\boldsymbol{z})$ and $\boldsymbol{z}_{\tau_0 \setminus \tau} = P_{\tau_0 \setminus \tau}(\boldsymbol{z}) = \boldsymbol{z} - \boldsymbol{z}_\tau$.

Proposition 11. *Under the above assumptions and notation, the limiting distribution of $\sqrt{n}\{\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*\}$ is given as follows: for any $\tau = \mathcal{O}(E \cup F)$, where $F \subseteq F_0$, and any Borel subset $B \subseteq \mathbb{R}(E) \times \mathcal{O}(F)$,*

$$\mathbf{P}(\eta \in B) = \int_B \psi_F(\boldsymbol{z}_\tau) \, \mathrm{d}\boldsymbol{z}_\tau,$$

where

$$\psi_F(\boldsymbol{z}_\tau) = \int_{-\infty}^0 \phi_{\tau_0}(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z}_{\tau_0 \setminus \tau}.$$

The special case that $l(\mu) = l - 1$ of this Proposition, together with the comments in the previous two paragraphs, generalises the limiting distribution of $\sqrt{n}\{\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*\}$ when $\boldsymbol{X}^m$ is a tree space and $\boldsymbol{x}^*$ lies in a stratum of co-dimension one obtained in [3].

Proof. By Theorem 4, we only need to consider the case where $F \neq F_0$. Assume that $\tau = \mathcal{O}(E \cup F)$ has co-dimension l' and fix $\boldsymbol{w}_\tau \in S_{\tau \setminus \sigma}^{l-l'}$. We first show that

$$P_\tau\left(\tilde{\Psi}_{\tau_0}(\boldsymbol{\xi}_1; \boldsymbol{x}^*)\right) = \tilde{\Psi}_\tau(\boldsymbol{\xi}_1; \boldsymbol{x}^*) \qquad \text{a.s.} \tag{42}$$

Recall, from the proof of Theorem 2, that the geodesics from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_{\tau_0}) = \boldsymbol{x}^* + \lambda \boldsymbol{w}_{\tau_0}$ to $\boldsymbol{x}$ have the same support for all $\boldsymbol{w}_{\tau_0} \in S_{\tau_0 \setminus \sigma}^{l-l(\mu)}$ sufficiently close to $\boldsymbol{w}_\tau$ and all sufficiently small $\lambda > 0$. For such $\boldsymbol{w}_{\tau_0}$ and λ, by Definition 14,

Corollary 4(i), Propositions 8 and 9, the sequence $\mathcal{A} = (A_0, \cdots, A_k)$ in the support $(\mathcal{A}, \mathcal{B})$ of the geodesics from $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_{\tau_0}) = \boldsymbol{x}^* + \lambda \boldsymbol{w}_{\tau_0}$ to $\boldsymbol{\xi}_1$ has the property that, if $i > 0$ and if $A_i \cap E = \emptyset$, then A_i consists of a single axis in F_0 a.s., so that the $P_{A_i}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_{\tau_0}))/\|P_{A_i}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_{\tau_0}))\|$ is independent of the value of λ a.s. This, together with the fact implied by Corollary 4(ii) that, if $A_i \cap E \neq \emptyset$, $P_{A_i}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_{\tau_0})) \to P_{A_i}(\boldsymbol{x}^*)$ as $\lambda \to 0$, shows that, with probability one, each $P_{A_i}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_{\tau_0}))/\|P_{A_i}(\boldsymbol{x}^*(\lambda, \boldsymbol{w}_{\tau_0}))\|$ in the expression (8) for $\Phi_{\tau_0}(\boldsymbol{\xi}_1; \boldsymbol{x}^* + \lambda \boldsymbol{w}_{\tau_0})$ is a continuous function at $\boldsymbol{x}^*$ in the corresponding Euclidean space. It follows that

$$\lim_{\substack{\boldsymbol{w}_{\tau_0} \to \boldsymbol{w}_\tau \\ \lambda \to 0+}} \Phi_{\tau_0}(\boldsymbol{\xi}_1; \boldsymbol{x}^* + \lambda \boldsymbol{w}_{\tau_0})$$

exists a.s. and so, in particular,

$$\lim_{\boldsymbol{w}_{\tau_0} \to \boldsymbol{w}_\tau} \lim_{\lambda \to 0+} \Phi_{\tau_0}(\boldsymbol{\xi}_1; \boldsymbol{x}^* + \lambda \boldsymbol{w}_{\tau_0}) = \lim_{\lambda \to 0+} \lim_{\boldsymbol{w}_{\tau_0} \to \boldsymbol{w}_\tau} \Phi_{\tau_0}(\boldsymbol{\xi}_1; \boldsymbol{x}^* + \lambda \boldsymbol{w}_{\tau_0}) \qquad a.s.$$

Thus, the definition of Ψ gives

$$\lim_{\boldsymbol{w}_{\tau_0} \to \boldsymbol{w}_\tau} \Psi_{\tau_0}(\boldsymbol{\xi}_1, \boldsymbol{w}_{\tau_0}; \boldsymbol{x}^*) = \lim_{\lambda \to 0+} \lim_{\boldsymbol{w}_{\tau_0} \to \boldsymbol{w}_\tau} \Phi_{\tau_0}(\boldsymbol{\xi}_1; \boldsymbol{x}^* + \lambda \boldsymbol{w}_{\tau_0}) \qquad a.s.$$

Since the limit on the right hand side exists, to find it, we take a particular path for $\boldsymbol{w}_{\tau_0}$ to approach $\boldsymbol{w}_\tau$: $\boldsymbol{w}_{\tau_0} = \sin \alpha \, \boldsymbol{w}^\perp + \cos \alpha \, \boldsymbol{w}_\tau$, where $\langle \boldsymbol{w}^\perp, \boldsymbol{w}_\tau \rangle = 0$ and $\|\boldsymbol{w}^\perp\| = 1$. Then, writing $\beta = \lambda \sin \alpha$, we have

$$\lim_{\boldsymbol{w}_{\tau_0} \to \boldsymbol{w}_\tau} \Phi_{\tau_0}(\boldsymbol{\xi}_1, \boldsymbol{x}^* + \lambda \boldsymbol{w}_{\tau_0})$$
$$= \lim_{\alpha \to 0+} \Phi_{\tau_0}\left(\boldsymbol{\xi}_1, \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau) + \lambda(\sin \alpha \, \boldsymbol{w}^\perp + (\cos \alpha - 1) \, \boldsymbol{w}_\tau)\right)$$
$$= \lim_{\beta \to 0+} \Phi_{\tau_0}\left(\boldsymbol{\xi}_1, \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau) + \beta \, \boldsymbol{w}^\perp\right)$$
$$= \Psi_{\tau_0}(\boldsymbol{\xi}_1, \boldsymbol{w}^\perp; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)) \qquad a.s.,$$

where the second equality follows from Corollary 3(ii). Hence, it follows from Corollary 4(ii) that

$$P_\tau\left(\lim_{\boldsymbol{w}_{\tau_0} \to \boldsymbol{w}_\tau} \Psi_{\tau_0}(\boldsymbol{\xi}_1, \boldsymbol{w}_{\tau_0}; \boldsymbol{x}^*)\right) = \lim_{\lambda \to 0+} \Phi_\tau(\boldsymbol{\xi}_1; \boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau)) = \Psi_\tau(\boldsymbol{\xi}_1, \boldsymbol{w}_\tau; \boldsymbol{x}^*) \quad a.s.$$

as $\boldsymbol{x}^*(\lambda, \boldsymbol{w}_\tau) \in \tau$. Hence, (42) follows.

Since $\hat{\boldsymbol{\xi}}_n$ will lie in $\mathcal{K}_\mu$ for sufficiently large n a.s. by Corollary 7, without loss of generality, we assume that it is true for all n. Let $\hat{\boldsymbol{\xi}}_n^\tau$ denote the sample Euclidean mean of $\tilde{\Psi}_\tau(\boldsymbol{\xi}_1; \boldsymbol{x}^*), \cdots, \tilde{\Psi}_\tau(\boldsymbol{\xi}_1; \boldsymbol{x}^*)$. Then, $\hat{\boldsymbol{\xi}}_n^\tau \in \mathbb{R}(E \cup F)$ and, by Corollary 6, $\hat{\boldsymbol{\xi}}_n^\tau \to \boldsymbol{x}^*$ a.s. Also, application of (42) gives $P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n^\tau\right) = P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n^{\tau_0}\right)$. On the other hand, the argument for the proof of Theorem 4 implies that, for all sufficiently large n,

$$1_\tau(\hat{\boldsymbol{\xi}}_n) P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n\right) = 1_\tau(\hat{\boldsymbol{\xi}}_n) P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n - \boldsymbol{x}^*\right) = 1_\tau(\hat{\boldsymbol{\xi}}_n) P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n^\tau\right)$$

so that, for all sufficiently large n,

$$1_\tau(\hat{\boldsymbol{\xi}}_n) P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n\right) = 1_\tau(\hat{\boldsymbol{\xi}}_n) P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n^{\tau_0}\right).$$
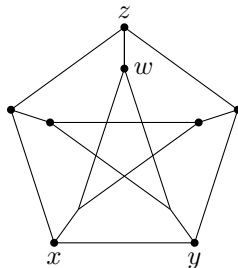
41

Figure 5: The Petersen graph, representing an intersection of $\boldsymbol{T}_4$ with unit sphere centred on the cone point.

However, given that $\hat{\boldsymbol{\xi}}_n$ is in $\mathcal{K}_\mu$, since $P_\sigma(\hat{\boldsymbol{\xi}}_n) = P_\sigma(\hat{\boldsymbol{\xi}}^{\tau_0})$ by Corollary 4(ii) and Corollary 6, the fact that $\hat{\boldsymbol{\xi}}_n \in \tau$ is equivalent to the fact that $P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n^{\tau_0}\right)$ lies in $\mathcal{O}(F)$ and $-P_{\tau_0 \backslash \tau}\left(\hat{\boldsymbol{\xi}}_n^{\tau_0}\right) \in \mathcal{O}(F_0 \backslash F)$. Hence, we can re-express the above equality as

$$1_\tau(\hat{\boldsymbol{\xi}}_n) P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n\right) = 1_{\mathcal{O}(F)}\left(P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n^{\tau_0}\right)\right) 1_{\mathcal{O}(F_0 \backslash F)}\left(-P_{\tau_0 \backslash \tau}\left(\hat{\boldsymbol{\xi}}_n^{\tau_0}\right)\right) P_{\tau \backslash \sigma}\left(\hat{\boldsymbol{\xi}}_n^{\tau_0}\right).$$

The required result then follows by a slight modification to the proof of Theorem 4. $\qquad\square$

In fact, the argument for the proof of Proposition 11, in particular (42), also shows that, if $\tau = \mathcal{O}(E \cup F)$ has co-dimension greater than $l(\mu)$ and if $\mathbb{R} \times \Theta_{\tau,\sigma}(\boldsymbol{x}^*; \mu)$ is contained in the interior of $\mathcal{K}_\mu$, then $\mathbf{P}(\eta \in \mathbb{R} \times \mathcal{O}(F)) = 0$.

We illustrate Proposition 11 with the following example.

Example 5. Consider the probability measure μ, on the BHV phylogenetic tree space $\boldsymbol{T}_4$ of four leaves, with support on the closure of the union of the one-dimensional stratum τ_z and the top-dimensional stratum $\tau_{x,y}$ whose one-dimensional co-bounding strata are the one-dimensional strata τ_x and τ_y, where τ_x, τ_y and τ_z are as represented by the vertices x, y and z in the Petersen graph shown in Figure 5. The density function of μ, with respect to $\mathrm{d}z$ and $\mathrm{d}x\,\mathrm{d}y$ respectively, is given as follows:

$$
\begin{aligned}
g(z) &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, & z > 0, & \qquad \text{and} \\
f(x,y) &= \frac{2}{\pi\sqrt{2\pi}} \frac{1}{\sqrt{x^2+y^2}} e^{-(x^2+y^2)/2}, & x, y > 0.
\end{aligned}
$$

By the uniqueness of Fréchet means and by the symmetry of the space and the symmetry of the distribution, if the Fréchet mean of μ were not at the cone point, it would lie either in one of the top-dimensional strata $\tau_{x,y}$ or $\tau_{z,w}$, or in one of the one-dimensional strata τ_z or τ_w, where $\tau_{x,y}$ is determined by the x- and y-axes etcetera. However,

$$\int_0^\infty \int_0^\infty \sqrt{x^2+y^2}\, f(x,y)\, \mathrm{d}x\,\mathrm{d}y = \int_0^\infty z\, g(z)\, \mathrm{d}z.$$

42

This implies that $\int_{\boldsymbol{T}_4} \Phi_{\tau_z}(\boldsymbol{x}; \boldsymbol{x}')\,\mathrm{d}\mu(\boldsymbol{x}) = 0$ for any $\boldsymbol{x}' \in \tau_z$. Hence, using the expression obtained in [2] on the expression for $\Phi(\,\cdot\,; \boldsymbol{x}')$, Theorem 3 implies that the Fréchet mean of μ does not lie in τ_z. Moreover, it follows from

$$\int_0^\infty \int_0^\infty x\, f(x, y)\,\mathrm{d}x\,\mathrm{d}y = \frac{2}{\pi\sqrt{\pi}}$$

that the expression of $\Phi(\,\cdot\,; \boldsymbol{x}')$ in [2] for $\boldsymbol{x}' \in \tau_{x,y}$ gives

$$
\int_{\boldsymbol{T}_4} \Psi(\boldsymbol{x}, (\cos\theta, \sin\theta); o)\,\mathrm{d}\mu(\boldsymbol{x})
$$
$$
= \int_{\boldsymbol{T}_4} \Phi(\boldsymbol{x}; \boldsymbol{x}')\,\mathrm{d}\mu(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}}\left(\frac{2}{\pi} - \cos\theta, \frac{2}{\pi} - \sin\theta\right),
$$

where $\theta \in (0, \pi/2)$ and $(\cos\theta, \sin\theta) = \boldsymbol{x}'/\|\boldsymbol{x}'\|$ using its (x, y) coordinates. It is clear that at least one of the two coordinates of $\int_{\boldsymbol{T}_4} \Phi(\boldsymbol{x}; \boldsymbol{x}')\,\mathrm{d}\mu(\boldsymbol{x})$ is negative for any $\boldsymbol{x}' \in \tau_{x,y}$. Hence, the Fréchet mean of μ, as well as the corresponding sample Fréchet mean $\hat{\boldsymbol{\xi}}_n$ for sufficiently large n, can not lie in $\tau_{x,y}$. Similarly, the Fréchet mean of μ can lie in neither $\tau_{z,w}$ nor τ_w. Thus, the Fréchet mean of μ must be at the cone point.

Moreover, using the expression for Φ and Proposition 10, it can be checked that, for sufficiently large n, the corresponding sample Fréchet mean $\hat{\boldsymbol{\xi}}_n$ cannot lie in any stratum other than τ_z or the cone point. Since

$$\int_{\boldsymbol{T}_4} \Psi_{\tau_z}(\boldsymbol{x}, e_z; o)\,\mathrm{d}\mu(\boldsymbol{x}) = \int_{\boldsymbol{T}_4} \Phi_{\tau_z}(\boldsymbol{x}; \boldsymbol{x}')\,\mathrm{d}\mu(\boldsymbol{x})(= 0)$$

where e_z is the axis of τ_z and $\boldsymbol{x}' \in \tau_z$, there must be infinitely many $\hat{\boldsymbol{\xi}}_n$ lying in τ_z. So that there are infinitely many $\hat{\boldsymbol{\xi}}_n$ lying in each of these two strata.

Finally, since $\mathrm{var}\left(\Psi_{\tau_z}(\boldsymbol{\xi}, e_z; o)\right) = 1$ where $\boldsymbol{\xi}$ is a random variable on $\boldsymbol{T}_4$ with distribution μ, by Proposition 11, the limiting random variable η has distribution given by $\mathbf{P}(\eta = 0) = 1/2$ and, for any Borel set $B \in \tau_z$

$$\mathbf{P}(\eta \in B) = \frac{1}{\sqrt{2\pi}} \int_B e^{-z^2/2}\,\mathrm{d}z.$$

References

[1] M. Bacak (2014). Computing medians and means in Hadamard spaces, *SIAM J. Optimiz.* **24**, 1542-1566.

[2] D. Barden, H. Le and M. Owen (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees, *Electron. J. Probab.* **18**, no. 25.

[3] D. Barden, H. Le and M. Owen (2018). Limiting behaviour of Fréchet means in the space of phylogenetic trees, *Ann. Inst. Stat. Math.* **70**, 99-129.

[4] R. Bhattacharya and V. Patrangenaru (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds-II, *Ann. Statist.* **33**, 1225-1259.

[5] L.J. Billera, S.P. Holmes and K. Vogtmann (2001). Geometry of the space of phylogenetic trees, *Advances in Applied Mathematics* **27**, 733-767.

[6] M.R. Bridson and A. Haefliger (1999). *Metric Spaces of Non-positive Curvature.* Springer-Verlag, Berlin/New York.

[7] B. Burago, Y. Burago and S. Ivanov (2001). *A Course in Metric Geometry.* American Mathematical Society.

[8] M. Goresky and R. MacPherson (1980). *Stratified Morse Theory.* Springer-Verlag, Berlin/New York.

[9] S. Holmes (2003). Statistics for phylogenetic trees, *Theoretical Population Biology* **63**, 17-32.

[10] T. Hotz, S. Huckemann, H. Le, J.S. Marron, J.C. Mattingly, E. Miller, J. Nolen, M. Owen, V. Patrangenaru and S. Skwerer (2013). Sticky central limit theorems on open books, *Ann. Appl. Probab.* **23**, 2238-2258.

[11] S. Huckemann, J.C. Mattingly, E. Miller and J. Nolen (2015). Sticky central limit theorems at isolated hyperbolic planar singularities, *Electron. J. Probab.* **20**, no. 78.

[12] M. Kendall and C. Colijn (2015). A tree metric using structure and length to capture distinct phylogenetic signals, arXiv:1507.05211 [q-bio.PE].

[13] W.S. Kendall and H. Le (2011). Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables, *Brazilian Journal of Probability and Statistics* **25**, 323-352.

[14] B. Kleiner (1999). The local structure of length spaces with curvature bounded above, *Math. Z.* **231**, 409-456.

[15] E. Miller, M. Owen and S. Provan (2015). Polyhedral computational geometry for averaging metric phylogenetic trees, *Advances in Applied Mathematics* **68**, 51-91.

[16] I. Nikolaev (1995). The tangent cone of an Aleksandrov space of curvature $\leqslant K$, *Manuscripta Math.* **86**, 137-147.

[17] T.M.W. Nye (2011). Principal components analysis in the space of phylogenetic trees, *Ann. Statist.* **39**, 2716-2739.

[18] T.M.W. Nye (2015). Convergence of random walks to Brownian motion in phylogenetic tree-space, arXiv:1508.02906 [q-bio.PE].

[19] M. Owen (2011). Computing geodesic distances in tree space, *SIAM J. Discrete Math.* **25**, 1506-1529.

[20] M. Owen and J.S. Provan (2011). A fast algorithm for computing geodesic distances in tree space, *ACM/IEEE Transactions on Computational Biology and Bioinformatics* **8**, 2-13.

[21] R.Thom (1969). Ensembles et morphismes stratifiés, *Bull. Amer. Math. Soc.* **75**, 240-284.

[22] K. Vogtmann (2007). Geodesics in the space of trees, available at www.math.cornell.edu/∼vogtmann/papers/TreeGeodesicss/index.html.

[23] H. Ziezold (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces, *Trans. 7th Prague Conf. Info. Theory, Stat. Dec. Functions, Rand. Proc.* **A**, 591-602.