Central limit theorems for Fréchet means in the space of phylogenetic trees

Dennis Barden[*]   Huiling Le[†]   Megan Owen[‡]

Abstract

This paper studies the characterisation and limiting distributions of Fréchet means in the space of phylogenetic trees. This space is topologically stratified, as well as being a $CAT(0)$ space. We use a generalised version of the Delta method to demonstrate non-classical behaviour arising from the global topological structure of the space. In particular, we show that, for the space T_4 of trees with four leaves, although they are related to the Gaussian distribution, the forms taken by the limiting distributions depend on the co-dimensions of the strata in which the Fréchet means lie.

1 Introduction

It has become increasingly common in various research areas for statistical analysis to involve data that lies in non-Euclidean spaces, such as manifolds or even topologically stratified spaces. Two such examples are the statistical analysis of shape (cf. [6] & [15]) and the analysis of phylogenetic trees (cf. [4]). Consequently, many statistical concepts and techniques have been generalised and developed to adapt to such phenomena.

In this paper we focus on developing a central limit theorem on the space of phylogenetic trees, which is a topologically stratified space (cf. [14]). A phylogenetic tree represents the evolutionary history of a set of organisms, and as such, is one of the main data objects in evolutionary biology. Some methods have been developed for statistically evaluating phylogenetic trees (cf. [7] & [23]), however, these approaches often do not incorporate both the tree topology and edge lengths, which could represent mutation rate for example, in a holistic way. Addressing this deficiency was one of the goals of the construction of a space of phylogenetic trees [4] in which branch lengths, and hence tree topologies, vary continuously in a natural way. This space is a piecewise Euclidean metric space, and thus approaches from Euclidean statistics can be defined

---

[*]Girton College, University of Cambridge, UK. E-mail: `d.barden@dpmms.cam.ac.uk`
[†]University of Nottingham, UK. E-mail: `huiling.le@nottingham.ac.uk`
[‡]University of Waterloo, Canada. E-mail: `megan.owen@uwaterloo.ca`

and generalized in it. To date, some further statistical theory, including methods for non-parametric bootstrap and hypothesis testing, within this space has been developed by Holmes (cf. [10], [11] & [12]), while the Fréchet mean and variance within this space was defined in [17]. We continue with these theoretical investigations, which fit in with the larger research goal of developing rigorous statistical analyses for topologically stratified spaces that was initiated by the working group on sampling from stratified spaces of the Statistical and Applied Mathematical Sciences Institute (SAMSI) 2010-11 program on Analysis of Object Data.

The main results of this paper show that central limit theorems hold for Fréchet means in the space of phylogenetic trees with four leaves and that the limiting distributions of the sample Fréchet means are closely associated with multivariate Gaussian distributions. In particular, we prove that there is a central limit theorem regardless of whether the Fréchet mean is in a top dimensional stratum (Theorem 3.1), in a co-dimension one stratum (Theorem 4.4), or at the cone point of the space (Theorem 5.2). The central limit theorems describe the behaviour of the sample Fréchet means around the true Fréchet mean, as the sample size increases. Thus, our results have implications for the statistical analysis of phylogenetic trees. One example would be hypothesis testing: if we have samples from two potentially different distributions of trees, we may be able to reject the hypothesis that the two distributions are the same by computing the Fréchet means of the samples, and comparing the distance between them with that expected under the central limit theorem.

The concept of Fréchet means of random variables on a metric space is a generalisation of the least mean-square characterisation of Euclidean means: a point is a Fréchet mean of a probability measure μ on a metric space (M, d) if it minimises the Fréchet function for μ defined by

$$x \mapsto \frac{1}{2} \int_M d(x, x')^2 d\mu(x'). \tag{1.1}$$

Various aspects of Fréchet means have been studied for non-Euclidean spaces, including Riemannian manifolds and certain stratified spaces. Among other applications, the first use of Fréchet means to provide nonparametric statistical inference, such as confidence regions and two-sample tests for discriminating between two distributions, was carried out in [2] and [3] for both extrinsic and intrinsic inference applied to manifolds, while the earlier work [8] and [9] provided similar inference restricted to extrinsic means on regular submanifolds of Euclidean spaces. We first review some of the ideas and results in the literature which are relevant to our investigation of Fréchet means on tree space. When M is a Riemannian manifold with the distance function being that induced by its Riemannian metric, the results on central limit theorems for Fréchet means can be found in [3] and [16]. The result and the proof of the classical central limit theorem rely on the global linear structure of Euclidean space. Hence, a crucial step in both [3] and [16] is to find a relationship between the sample Fréchet means of a sequence of random variables in M and the sample Euclidean means of a sequence of appropriately defined random variables in Euclidean space. The way that [3] achieves this is to use an embedding of the support of the distribution of the random variables into a Euclidean space. Then, the chosen embedding maps the sequence of random variables in M to a sequence of random variables in Euclidean space. This allows the authors to apply known results in Euclidean spaces to the resulting sequence of random variables to obtain a central limit theorem where, as expected, the limiting distribution depends on the chosen embedding. Among others, [16] explores another relationship between Fréchet means and Euclidean means to study the limiting behaviour directly on the manifold itself. Since the gradient of the Fréchet function must be zero at a

Fréchet mean and since
$$\mathrm{grad}_1 d(x, x')^2 = 2 \exp_x^{-1}(x'),$$
where grad_1 denotes the gradient with respect to the first variable of d^2 and $\exp_x^{-1}$ is the inverse exponential, or logarithmic, map at a point x in $\boldsymbol{M}$, it follows that, if x_0 is a Fréchet mean of μ, then
$$\int_{\boldsymbol{M}} \exp_{x_0}^{-1}(x')\, d\mu(x') = 0.$$
Thus, when x_0 is a Fréchet mean of μ on the Riemannian manifold $\boldsymbol{M}$, the origin of the tangent space of $\boldsymbol{M}$ at x_0 is the Euclidean mean of the probability measure induced on that space by the log map $\exp_{x_0}^{-1}$ from μ. The difficulty arising from this method is that the log map varies with the reference point x_0. As the number of random variables increases, their sample Fréchet mean changes. This results in the sequence of the random variables in $\boldsymbol{M}$ being mapped to different sequences of random variables in the Euclidean space as their number increases. Hence, the classical central limit theorem cannot been applied directly. To deal with this, [16] uses the notion of parallel transport in Riemannian geometry. This gives intrinsic results, which show how the global geometry of the space influences the limiting probability measure defined on the tangent space at the Fréchet mean under consideration. On the other hand, the results in both papers imply that, since manifolds are locally homeomorphic to Euclidean spaces, the limiting distributions for sample Fréchet means on Riemannian manifolds are usually Gaussian, a phenomenon similar to that for Euclidean means.

The topological structure of spaces also plays a role in the limiting behaviour of sample Fréchet means, as studied in [1] and [13]. In [13], a non-classical phenomenon of central limit type theorems for Fréchet means is observed that does not occur in the case of Riemannian manifolds. This applies to metric spaces with an open book decomposition, which is isometric with a disjoint union of copies of a half Euclidean space, or 'pages', identified along their boundary hyperplanes to form the 'spine'. The features, termed 'sticky' and 'partly sticky', observed in [13] are that, under mild conditions, when the Fréchet mean of a probability measure on such a space lies in the spine, its *iid* sample Fréchet means will lie either on the spine or in one single half-space, for all sufficiently large sample sizes. This, in particular, implies that, in this case, the support of the limiting distribution is either the spine or on one page.

Open books are one of the simplest non-trivial topologically stratified spaces and any stratified space that is singular along a stratum of co-dimension one is locally homeomorphic to an open book along that stratum. This paper is a continuation of the investigation initiated in [13], in the direction of central limit type theorems on stratified spaces. It is also a first step towards the study of central limit type theorems for Fréchet means on the spaces of phylogenetic trees. A space of phylogenetic trees, or *tree space* for short, is formed from a disjoint union of Euclidean orthants of a given dimension with the identification of certain sets of faces of various co-dimensions. The simplest tree space is that for trees with three leaves and it comprises three half lines glued at their ends, so is a special case of an 'open book'. This paper concentrates on the space of phylogenetic trees with four leaves to investigate how other aspects of the global structure of the tree space influence the limiting behaviour of sample Fréchet means, including the so-called 'stickiness' feature. In the case of open books, the paper [13], by combining half-planes appropriately, in effect turns the problem into a Euclidean one as long as the Fréchet mean does not lie on the spine. Unfortunately, this feature of Fréchet means on 'open books' no longer holds for those in tree space. In particular, the Fréchet mean of a random variable in tree space has, in general, no closed form and the generalised log map for tree space is not linear with respect to the points at which it is defined. The reason for this is the existence of the 'umbral' set, that we shall

define, for each given point. Although the tree space is not a Riemannian manifold, the nature of our problem is similar to that of characterising the Fréchet mean of a random variable in a Riemannian manifold. This results in a different approach to that in [13] and in the difference in the nature of the results that we obtain. We characterise the Fréchet mean of a random variable in tree space in terms of the Euclidean mean of a certain function of that random variable. Due to the global structure of the tree space, the function obtained in this way also depends on the Fréchet mean of that random variable in tree space. Although the technique of parallel transport used in [16] is inapplicable here, we use the local Euclidean structure of the top strata of tree space to analyse this dependence explicitly. Then, the various properties of such functions enable us to derive the central limit theorems on tree space.

The paper is organised as follows. In Section 2, we introduce the space $\boldsymbol{Q}_5$, a subspace of $\mathbb{R}^3$ consisting of a cycle of five quadrants, and concentrate on the characterisation of, and the central limit theorem for, Fréchet means in $\boldsymbol{Q}_5$ that do not lie at the origin (Proposition 2.2). Although $\boldsymbol{Q}_5$ is globally flat away from the origin, the result and the methods used in this section make it clear that the central limit theorem for Fréchet means in $\boldsymbol{Q}_5$ takes a different form from its counterpart in Euclidean space. In addition to their intrinsic interest, the results and the approach in this section also form a basis for our investigations of $\boldsymbol{T}_4$, the space of trees with four leaves, in the following sections. The three remaining sections of the paper study the limiting distributions for Fréchet means in $\boldsymbol{T}_4$ for the three possible co-dimensions of the strata on which they can lie. In particular, we relate $\boldsymbol{T}_4$ to $\boldsymbol{Q}_5$ in such a way that the result for the top-dimensional strata is a direct consequence of that in $\boldsymbol{Q}_5$ (Section 3). Note that, although the main idea here follows closely that of general central limit theorems for M-estimators in statistics as in [3], the functions involved here are neither diffeomorphisms as in [3], nor second order differentiable like general M-estimators, except for the special case where the support of the distribution is diffeomorphic with a Euclidean space. The cases when Fréchet means lie either on co-dimension one strata (Section 4) or at the cone point (Section 5) require additional analysis and the results there take different forms. We show that, when a Fréchet mean lies on a co-dimension one stratum, the limiting distribution can take one of three possible forms, distinguished by the nature of its support. This support may be either the one-dimensional Euclidean space containing the co-dimension one stratum where the Fréchet mean lies, or a two-dimensional half Euclidean space whose boundary contains that co-dimension one stratum, or the union of two such half spaces. In contrast, when a Fréchet mean is at the cone point, the intersection of the support of the limiting distribution with any given quadrant can be either an empty set or a cone. In all these cases, the limiting distributions are linked closely with Gaussian distributions in Euclidean spaces. The Appendix contains a brief account of the underlying geometry of tree spaces.

2 Fréchet means in $\boldsymbol{Q}_5$

Let $\boldsymbol{Q}_5$ be the union of the five quadrants embedded in $\mathbb{R}^3$, with coordinates u, v, w as shown in Figure 1, and let d_Q denote the intrinsic metric on $\boldsymbol{Q}_5$. Without its cone point $(0, 0, 0)$, $\boldsymbol{Q}_5$ is a flat non-complete Riemannian surface. However the inclusion of the cone point, by allowing the realisation of geodesics through that point, makes $\boldsymbol{Q}_5$ into a geodesic metric space of non-positive curvature or a so-called $CAT(0)$ space (cf. [5]). Nevertheless, it still has the isometry that permutes the five quadrants cyclically in the obvious fashion, fixing the cone point. This implies, in particular, that the square of the distance from a fixed point in $\boldsymbol{Q}_5$ remains differentiable on each of the open semi-axes.

In this section, we assume that μ is a probability measure on $\boldsymbol{Q}_5$ such that its Fréchet
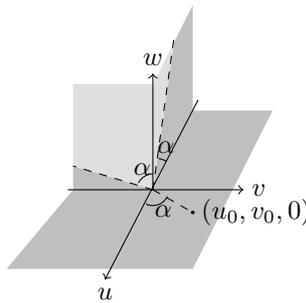
Figure 1: The space $\boldsymbol{Q}_5$.

function is finite and that $\{\xi_l\}$ is a sequence of *iid* random variables in $\boldsymbol{Q}_5$ with probability measure μ. Then, the fact that $\boldsymbol{Q}_5$ has non-positive curvature implies that the Fréchet mean $(\hat{u}, \hat{v}, \hat{w})$ of μ exists and is unique, and that, when it lies in the region where $\boldsymbol{Q}_5$ is a Riemannian manifold, i.e. when $(\hat{u}, \hat{v}, \hat{w}) \neq (0, 0, 0)$, it is characterised by

$$\int_{Q_5} \mathrm{grad}_1(d_Q(q, \xi)^2)\Big|_{q=(\hat{u}, \hat{v}, \hat{w})} d\mu(\xi) = 0 \tag{2.1}$$

(cf. [20] & [16] respectively). We shall exclude the case where the Fréchet mean is at the cone point. Then, the symmetry of the five quadrants of $\boldsymbol{Q}_5$ implies that, without loss of generality, we may restrict ourselves to the case where $(\hat{u}, \hat{v}, \hat{w})$ lies in the interior of the subset of $\boldsymbol{Q}_5$ determined by $w = 0$.

Consider Figure 1 for a fixed $u_0 > 0$ and $v_0 \geqslant 0$. The geodesic in $\boldsymbol{Q}_5$ from the point $(u_0, v_0, 0)$ to a point in the (closed) dark shaded areas either is a straight linear segment in the full (u, v)-plane, or becomes a straight linear segment when the relevant quadrant is 'folded down' into the (u, v)-plane. The light grey shading in Figure 1 shows the set of points q in $\boldsymbol{Q}_5$ from which the geodesic to $(u_0, v_0, 0)$ is a bent line that is the union of two segments: one from $(u_0, v_0, 0)$ to the origin and the other from the origin to q. We denote the union of the darker (open) shaded regions by $\mathcal{I}^Q_{(u_0, v_0)}$ and the 'umbral' set, the union of the (closed) lighter shaded regions, by $\mathcal{U}^Q_{(u_0, v_0)}$. For example, in the extreme case when $v_0 = 0$, and so $\alpha = 0$ in Figure 1, $\mathcal{U}^Q_{(u_0, v_0)}$ is the back quadrant defined by $u \leqslant 0$, $v = 0$, $w \geqslant 0$ and $\mathcal{I}^Q_{(u_0, v_0)}$ is the union of the other four quadrants. By the symmetry, we can easily derive the forms of $\mathcal{I}^Q_{(u_0, v_0)}$ and $\mathcal{U}^Q_{(u_0, v_0)}$ for other possible (u_0, v_0) for which $(u_0, v_0, 0)$ lies in the interior of the subset of $\boldsymbol{Q}_5$ determined by $w = 0$.

For points $(u_0, v_0, 0)$ in $\boldsymbol{Q}_5$, we define a map $\Psi_{(u_0, v_0)}$ from $\boldsymbol{Q}_5$ to $\mathbb{R}^2$, that is an isometry on $\mathcal{I}^Q_{(u_0, v_0)}$ and collapses $\mathcal{U}^Q_{(u_0, v_0)}$ to the line $u_0 v = v_0 u$, isometrically on each relevant ray through the origin. We shall see that this map is closely related to the expression for the gradient of the squared distance function d_Q^2 at points with zero w-coordinate. It is defined by

$$\begin{aligned}
&\Psi_{(u_0, v_0)}(u, u, w) \\
&= \begin{cases} -\dfrac{\|(u, v, w)\|}{\|(u_0, v_0)\|}(u_0, v_0) & \text{if } (u, v, w) \in \mathcal{U}^Q_{(u_0, v_0)} \\ \varphi_{(u_0, v_0)}(u, v, w) & \text{if } (u, v, w) \in \mathcal{I}^Q_{(u_0, v_0)} \end{cases}
\end{aligned} \tag{2.2}$$

where $\|(u_0, v_0)\|$ is the standard norm in $\mathbb{R}^2$ and similarly $\|(u, v, w)\| =$

$\sqrt{u^2 + v^2 + w^2}$ is the distance of (u, v, w) from $(0, 0, 0)$ in $\boldsymbol{Q}_5$, and where

$$
\varphi_{(u_0,v_0)}(u,v,w) = \begin{cases}
(u,v) & \text{if } w = 0 \\
(u,-w) & \text{if } w > 0, v = 0 \text{ and } v_0 \geqslant 0 \\
(-w,-u) & \text{if } w > 0, v = 0, v_0 < 0 \text{ and } u_0 > 0 \\
(-w,v) & \text{if } w > 0, u = 0 \text{ and } u_0 \geqslant 0 \\
(-v,-w) & \text{if } w > 0, u = 0, u_0 < 0 \text{ and } v_0 > 0.
\end{cases}
$$

It can easily be checked that $\Psi_{(u_0,v_0)}(u_0,v_0,0) = (u_0,v_0)$ and that $\Psi_{s(u_0,v_0)} = \Psi_{(u_0,v_0)}$ for all $s > 0$. The squared distance function from $(u_0, v_0, 0)$ with, for example, $u_0 > 0$ and $v_0 \geqslant 0$ to any point (u, v, w) in $\boldsymbol{Q}_5$ can be expressed explicitly as

$$
d_Q((u_0,v_0,0),(u,v,w))^2
$$

$$
= \begin{cases}
(u_0 - u)^2 + (v_0 - v)^2 & \text{if } w = 0 \\
(u_0 - u)^2 + (v_0 + w)^2 & \text{if } w > 0, v = 0 \\
& \qquad \text{and } -w/u < v_0/u_0 \\
(u_0 + w)^2 + (v_0 - v)^2 & \text{if } w > 0, u = 0 \\
& \qquad \text{and } -v/w > v_0/u_0 \\
\{\|(u_0,v_0)\| + \|(u,v,w)\|\}^2 & \text{if } (u,v,w) \in \mathcal{U}^Q_{(u_0,v_0)}.
\end{cases}
\tag{2.3}
$$

From this, we deduce that

$$
\tfrac{1}{2}\mathrm{grad}_1 d_Q(q,(u,v,w))^2 \Big|_{q=(u_0,v_0,0)}
$$

$$
= \begin{cases}
(u_0 - u, v_0 - v) & \text{if } w = 0 \\
(u_0 - u, v_0 + w) & \text{if } w > 0, v = 0 \\
& \qquad \text{and } -w/u < v_0/u_0 \\
(u_0 + w, v_0 - v) & \text{if } w > 0, u = 0 \\
& \qquad \text{and } -v/w > v_0/u_0 \\
\left\{1 + \dfrac{\|(u,v,w)\|}{\|(u_0,v_0)\|}\right\}(u_0,v_0) & \text{if } (u,v,w) \in \mathcal{U}^Q_{(u_0,v_0)},
\end{cases}
$$

so that in particular

$$
-\frac{1}{2}\mathrm{grad}_1 d_Q(q,(u,v,w))^2 \Big|_{q=(u_0,v_0,0)} = \Psi_{(u_0,v_0)}(u,u,w) - (u_0,v_0).
\tag{2.4}
$$

It can be checked that (2.4) holds for all $(u_0, v_0, 0)$ in the interior of the subset of $\boldsymbol{Q}_5$ determined by $w = 0$. For such $(u_0, v_0, 0)$, $\mathbb{R}^2$ may be identified with the tangent space to $\boldsymbol{Q}_5$ at that point and so equation (2.4) implies that we may regard $\Psi_{(u_0,v_0)}(\cdot) - (u_0,v_0)$ as a generalised log map for $\boldsymbol{Q}_5$ at $(u_0, v_0, 0)$. In particular, $\Psi_{(u_0,v_0)}(u,v,w) - (u_0,v_0)$ is the tangent vector, at $(u_0, v_0, 0)$, to the geodesic from $(u_0, v_0, 0)$ to (u, v, w) in $\boldsymbol{Q}_5$, whose length is the same as the distance between those two points. However, although $\Psi_{(u_0,v_0)}$ is surjective, it is not injective: the exponential map itself is only defined on the subspace of the tangent space corresponding to $\mathcal{I}^Q_{(u_0,v_0)}$.

By (2.1), a direct consequence of (2.4) is the following characterisation of the Fréchet mean of μ, when it is away from the cone point, in terms of the Euclidean means of the random variables $\{\Psi_{(u,v)}(\xi) \mid u > 0, v \geqslant 0\}$, where ξ is a random variable in $\boldsymbol{Q}_5$ with probability μ.

Lemma 2.1. *Let μ be a probability measure on $\boldsymbol{Q}_5$ such that its Fréchet function is finite. Then, $(\hat{u}, \hat{v}, 0) \neq (0, 0, 0)$ is the Fréchet mean of μ if and only if*

$$
(\hat{u}, \hat{v}) = \int_{Q_5} \Psi_{(\hat{u},\hat{v})}(\xi)\, d\mu(\xi).
\tag{2.5}
$$

Recalling that $\Psi_{(u,v)}(\cdot) = \Psi_{s(u,v)}(\cdot)$ for all $s > 0$, this result also implies that $(\hat{u}, \hat{v})$ is the Fréchet mean of μ if and only if it is identical with the Euclidean mean of the random variable $\Psi_{s(\hat{u},\hat{v})}(\xi)$.

To study the fluctuations of the sample Fréchet means of $\{\xi_l\}$ around the true Fréchet mean $(\hat{u}, \hat{v}, 0)$, we first examine, for fixed $(u, v, w) \in \boldsymbol{Q}_5$, how the corresponding vector $\Psi_{(u_0,v_0)}(u, v, w)$ changes as (u_0, v_0) changes. For two points $(u_r, v_r, 0) \neq (0, 0, 0)$ in $\boldsymbol{Q}_5$, $r = 1, 2$,

$$
\begin{aligned}
&\left\{ \Psi_{(u_2,v_2)} - \Psi_{(u_1,v_1)} \right\}(u, v, w) \\
&= \begin{cases}
(0, 0) & \text{if } (u, v, w) \in \mathcal{I}^Q_{(u_1,v_1)} \cap \mathcal{I}^Q_{(u_2,v_2)} \\
\varphi_{(u_2,v_2)}(u, v, w) + \frac{\|(u,v,w)\|}{\|(u_1,v_1)\|}(u_1, v_1) & \text{if } (u, v, w) \in \mathcal{U}^Q_{(u_1,v_1)} \cap \mathcal{I}^Q_{(u_2,v_2)} \\
-\varphi_{(u_1,v_1)}(u, v, w) - \frac{\|(u,v,w)\|}{\|(u_2,v_2)\|}(u_2, v_2) & \text{if } (u, v, w) \in \mathcal{I}^Q_{(u_1,v_1)} \cap \mathcal{U}^Q_{(u_2,v_2)} \\
\|(u,v,w)\| \left\{ \frac{(u_1,v_1)}{\|(u_1,v_1)\|} - \frac{(u_2,v_2)}{\|(u_2,v_2)\|} \right\} & \text{if } (u, v, w) \in \mathcal{U}^Q_{(u_1,v_1)} \cap \mathcal{U}^Q_{(u_2,v_2)}.
\end{cases}
\end{aligned}
$$

When (u_2, v_2) is sufficiently close to (u_1, v_1), the second and third expressions above differ, on the relevant domains, from the fourth by terms which are $o(\|(u_2 - u_1, v_2 - v_1)\|)\|(u, v, w)\|$. For example, when $(u, v, w) \in \mathcal{U}^Q_{(u_1,v_1)} \cap \mathcal{I}^Q_{(u_2,v_2)}$, the summand $\varphi_{(u_2,v_2)}(u, v, w)$ lies in a wedge centred on the origin with edges determined by the vectors $-(u_1, v_1)$ and $-(u_2, v_2)$ and distant $\|(u, v, w)\|$ from the origin. On the other hand, the first order Taylor expansion of the vector-valued function $(u, v)/\|(u, v)\|$ results in

$$
\begin{aligned}
\frac{(u_2, v_2)}{\|(u_2, v_2)\|} &- \frac{(u_1, v_1)}{\|(u_1, v_1)\|} \\
&= (u_2 - u_1, \, v_2 - v_1) M_{(u_1,v_1)} + o(\|(u_2 - u_1, v_2 - v_1)\|),
\end{aligned}
$$

where the matrix $M_{(u,v)}$ is given by

$$
M_{(u,v)} = \frac{1}{\|(u,v)\|^3} \begin{pmatrix} -v \\ u \end{pmatrix} (-v, u). \tag{2.6}
$$

Note that $\|(u, v)\| M_{(u,v)}$ is the projection map to the line through the origin orthogonal to (u, v). Thus, for (u_2, v_2) sufficiently close to (u_1, v_1), we have

$$
\begin{aligned}
&\Psi_{(u_2,v_2)}(u, v, w) - \Psi_{(u_1,v_1)}(u, v, w) \\
&= \|(u,v,w)\| \left\{ -(u_2 - u_1, \, v_2 - v_1) M_{(u_1,v_1)} + o(\|(u_2 - u_1, v_2 - v_1)\|) \right\} \\
&\quad \times 1_{\{(u,v,w) \in \mathcal{U}^Q_{(u_1,v_1)} \cup \mathcal{U}^Q_{(u_2,v_2)}\}}.
\end{aligned} \tag{2.7}
$$

This analysis leads to the following central limit theorem for Fréchet means in $\boldsymbol{Q}_5$, away from its cone point.

Proposition 2.2. *Let μ be a probability measure on $\boldsymbol{Q}_5$ with finite Fréchet function and with Fréchet mean $(\hat{u}, \hat{v}, 0)$ lying in the interior of the subset of $\boldsymbol{Q}_5$ determined by $w = 0$. Also, let $\{\xi_l\}$ be a sequence of iid random variables in $\boldsymbol{Q}_5$ with probability measure μ and $(\hat{u}_n, \hat{v}_n, \hat{w}_n)$ be the sample Fréchet mean of $\xi_1, \cdots, \xi_n$. Then,*

$$
\sqrt{n}(\hat{u}_n - \hat{u}, \, \hat{v}_n - \hat{v}) \xrightarrow{d} N(0, A^\top V A), \qquad \text{as } n \to \infty,
$$

where V is the covariance matrix of the random variable $\Psi_{(\hat{u},\hat{v})}(\xi_1)$,

$$
A = \left\{ I + E\left[\|\xi_1\| 1_{\{\xi_1 \in \mathcal{U}^Q_{(\hat{u},\hat{v})}\}} \right] M_{(\hat{u},\hat{v})} \right\}^{-1}
$$

and $M_{(u,v)}$ is given by (2.6).

Proof. Since the sequence of the sample Fréchet means $(\hat{u}_n, \hat{v}_n, \hat{w}_n)$ converges a.s. to $(\hat{u}, \hat{v}, 0)$ (cf. [20] & [22]), it follows that $(\hat{u}_n, \hat{v}_n, \hat{w}_n)$ lies a.s. in the interior of the subset of $\boldsymbol{Q}_5$ determined by $w = 0$ for sufficiently large n. Applying Lemma 2.1 to discrete probability measures on $\boldsymbol{Q}_5$ it follows that, for n sufficiently large such that $\hat{w}_n = 0$,

$$(\hat{u}_n, \hat{v}_n) = \frac{1}{n} \sum_{l=1}^{n} \Psi_{(\hat{u}_n, \hat{v}_n)}(\xi_l).$$

Hence, for sufficiently large n,

$$
\begin{aligned}
&\sqrt{n}\left\{(\hat{u}_n, \hat{v}_n) - (\hat{u}, \hat{v})\right\} \\
={} & \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{\Psi_{(\hat{u}_n, \hat{v}_n)}(\xi_l) - (\hat{u}, \hat{v})\right\} \\
={} & \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{\Psi_{(\hat{u}, \hat{v})}(\xi_l) - (\hat{u}, \hat{v})\right\} + \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{\Psi_{(\hat{u}_n, \hat{v}_n)}(\xi_l) - \Psi_{(\hat{u}, \hat{v})}(\xi_l)\right\}.
\end{aligned}
\tag{2.8}
$$

By (2.7), the second summand on the right hand side of the above expression is

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{\Psi_{(\hat{u}_n, \hat{v}_n)}(\xi_l) - \Psi_{(\hat{u}, \hat{v})}(\xi_l)\right\} \\
={} & -\sqrt{n}(\hat{u}_n - \hat{u},\, \hat{v}_n - \hat{v}) M_{(\hat{u}, \hat{v})} \frac{1}{n} \sum_{l=1}^{n} \|\xi_l\| \, \mathbf{1}_{\{\xi_l \in \mathcal{U}_{(\hat{u}, \hat{v})}^Q \cup \mathcal{U}_{(\hat{u}_n, \hat{v}_n)}^Q\}} \\
& + o(\|(\hat{u}_n - \hat{u},\, \hat{v}_n - \hat{v})\|) \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \|\xi_l\| \, \mathbf{1}_{\{\xi_l \in \mathcal{U}_{(\hat{u}, \hat{v})}^Q \cup \mathcal{U}_{(\hat{u}_n, \hat{v}_n)}^Q\}}.
\end{aligned}
$$

Thus, we deduce from (2.8) that

$$
\begin{aligned}
& \sqrt{n}(\hat{u}_n - \hat{u},\, \hat{v}_n - \hat{v}) \\
& \quad \times \left\{1 + M_{(\hat{u}, \hat{v})} \frac{1}{n} \sum_{l=1}^{n} \|\xi_l\| \left(\mathbf{1}_{\{\xi_l \in \mathcal{U}_{(\hat{u}, \hat{v})}^Q\}} + \mathbf{1}_{\{\xi_l \in \mathcal{I}_{(\hat{u}, \hat{v})}^Q \cap \mathcal{U}_{(\hat{u}_n, \hat{v}_n)}^Q\}}\right)\right\} \\
={} & \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{\Psi_{(\hat{u}, \hat{v})}(\xi_l) - (\hat{u}, \hat{v})\right\} \\
& + o(\|(\hat{u}_n - \hat{u},\, \hat{v}_n - \hat{v})\|) \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \|\xi_l\| \left\{\mathbf{1}_{\{\xi_l \in \mathcal{U}_{(\hat{u}, \hat{v})}^Q\}} + \mathbf{1}_{\{\xi_l \in \mathcal{I}_{(\hat{u}, \hat{v})}^Q \cap \mathcal{U}_{(\hat{u}_n, \hat{v}_n)}^Q\}}\right\}.
\end{aligned}
\tag{2.9}
$$

From $E\left[\|\xi_1\|^2 \mathbf{1}_{\{\xi_1 \in \mathcal{U}_{(\hat{u}, \hat{v})}^Q\}}\right] \leqslant E[\|\xi_1\|^2] = \int d_Q((0,0,0), \xi)^2 d\mu(\xi)$ which is finite by assumption, it follows that $\frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{\|\xi_l\| \, \mathbf{1}_{\{\xi_l \in \mathcal{U}_{(\hat{u}, \hat{v})}^Q\}} - E\left[\|\xi_1\| \, \mathbf{1}_{\{\xi_1 \in \mathcal{U}_{(\hat{u}, \hat{v})}^Q\}}\right]\right\}$ converges in distribution. However, $\|(\hat{u}_n - \hat{u}, \hat{v}_n - \hat{v})\| \to 0$ a.s., so that

$$o(\|(\hat{u}_n - \hat{u},\, \hat{v}_n - \hat{v})\|) \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{\|\xi_l\| \, \mathbf{1}_{\{\xi_l \in \mathcal{U}_{(\hat{u}, \hat{v})}^Q\}} - E\left[\|\xi_1\| \, \mathbf{1}_{\{\xi_1 \in \mathcal{U}_{(\hat{u}, \hat{v})}^Q\}}\right]\right\} \xrightarrow{P} 0,$$

as $n \to \infty$. Since $\mathcal{I}_{(\hat{u}, \hat{v})}^Q \cap \mathcal{U}_{(\hat{u}_n, \hat{v}_n)}^Q$ converges to the empty set as $n \to \infty$, by replacing it for sufficiently large n with an appropriately defined cone D^ϵ where, for a given $\epsilon > 0$, D^ϵ has a sufficiently small angle that

$$E[\|\xi\| \mathbf{1}_{D^\epsilon}] < \epsilon \quad \text{and} \quad \text{var}[\|\xi\| \mathbf{1}_{D^\epsilon}] < \epsilon,$$

we also have that, in distribution and so in probability,

$$\frac{1}{\sqrt{n}} \sum_{l=1}^{n} \|\xi_l\| \, \mathbf{1}_{\{\xi_l \in \mathcal{I}_{(\hat{u}, \hat{v})}^Q \cap \mathcal{U}_{(\hat{u}_n, \hat{v}_n)}^Q\}} \to 0, \qquad \text{as } n \to \infty.$$
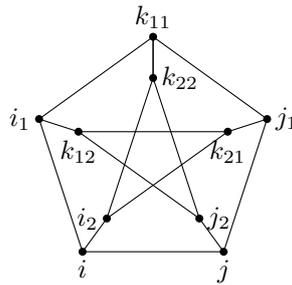
Figure 2: Labelled Peterson graph

On the other hand, $\{\Psi_{(\hat{u},\hat{v})}(\xi_l)\}$ is a sequence of *iid* random variables in $\mathbb{R}^2$ with the Euclidean mean $(\hat{u},\hat{v})$ by (2.5), so we can apply the classical central limit theorem to the first summand on the right hand side of (2.9) to get

$$\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\left\{\Psi_{(\hat{u},\hat{v})}(\xi_l)-(\hat{u},\hat{v})\right\} \xrightarrow{d} N(0,V), \qquad \text{as } n \to \infty.$$

The required result then follows from (2.9) above. □

3 Fréchet means in a top stratum of T_4

The tree space, T_n, is the moduli space of labelled n-trees, the moduli or parameters being the lengths of the internal edges. A labelled n-tree with unspecified edge lengths determines a topological type, modulo its leaves and root. Then T_n is a stratified space with a stratum for each topological type, such that a topological type with k internal edges determines a stratum with k positive parameters ranging over the points of an open k-dimensional orthant. It is known (cf. [4]) that T_n is a $CAT(0)$ space. A brief description of tree space is given in the Appendix and, for a more comprehensive study, we refer readers to papers such as [4], [18], [19] and [21].

The space T_4 can be visualised and described via the link $\bar{T}_4$ of its cone point, which is the tree whose internal edges all have zero length. The link $\bar{T}_4$ is the set of trees in T_4 whose edge lengths sum to 1. The entire space T_4 is the infinite cone on this link: for each point in the link there is a semi-infinite line in T_4 through this point to the cone point. This link is a finite graph, namely the Peterson graph, as illustrated in Figure 2: the quadrants of T_4 become edges of $\bar{T}_4$ and the semi-axes become vertices. As in the general case, each co-dimension one orthant, or semi-axis, is in the boundary of three quadrants so that the graph $\bar{T}_4$ is trivalent. A pentagon in $\bar{T}_4$ corresponds to a cycle of five quadrants in T_4. In $\bar{T}_4$, each edge lies in four pentagons and each vertex lies in six of them. The four pentagons containing a given edge are disjoint except for that edge and the neighbouring edges, and their union includes all the vertices and all but two edges of $\bar{T}_4$.

In view of the symmetries among the edges and vertices of the graph, we may take the vertices labelled i and j in Figure 2 to be the two vertices of an arbitrarily chosen edge; then i_r (j_r), $r = 1, 2$, will be the two vertices which share an edge with i (j); and finally k_{rs} will be the vertex together with which the vertices i, j, i_r, j_s form a pentagon in $\bar{T}_4$. Following the method of depiction in [4], the first four diagrams in Figure 3 illustrate the four cycles in T_4 of five quadrants that correspond to the four pentagons in the Peterson graph to which the (i, j)-edge belongs. In each of these representations,

we indicate the first four semi-axes, and hence the first three quadrants, as lying in a plane and the fifth semi-axis, and hence the remaining two quadrants, orthogonal to that plane. Then, the shaded quadrants of these four 5-cycles give 13 distinct quadrants of $\boldsymbol{T}_4$. The two remaining quadrants in $\boldsymbol{T}_4$ can be described using the further two 5-cycles in which the vertex i (or j) lies, as shown in the last two diagrams of Figure 3. Thus, these 15 shaded quadrants, any two of which have, apart from the cone point, at most a semi-axis in common, together form the entire space $\boldsymbol{T}_4$.

Each point in $\boldsymbol{T}_4$ can be specified by the quadrant in which it lies and its coordinate in that quadrant. Hence, it can be specified by a pair of non-negative numbers (x_r, x_s), where $r \neq s$ are the labels of the vertices of an edge in the Peterson graph. In general the ordering of r and s is unspecified so that both (x_r, x_s) and (x_s, x_r) represent the same point in $\boldsymbol{T}_4$. However, for our purposes, when working in any 5-cycle based on the (i, j)-quadrant as above, we shall take the axes in the cyclic order $i, j, j_r, k_{sr}, i_s, i$. This implies that the point $(x_i, 0)$ in the i-semi-axis of the (i, j)-quadrant will be represented as $(0, x_i)$ in the (i_r, i)-quadrants. We regard the j_r-semi-axis as the negative i-semi-axis and the i_s-semi-axis as the negative j-semi-axis so that, for example, the point (x_{i_r}, x_i) in the (i_r, i)-quadrant has coordinate $(x_i, -x_{i_r})$ with respect to the i- and j-semi-axes. We also regard the k_{rs}-semi-axis as the orthogonal semi-axis through the origin of the (i, j)-plane.

The metric d_T on $\boldsymbol{T}_4$ is obtained by identifying each quadrant with the principal quadrant in $\mathbb{R}^2$ thus inducing the Euclidean metric on it, and by defining the length of any rectifiable curve in $\boldsymbol{T}_4$ to be the sum of the lengths of the segments into which it is broken by the axes of the quadrants. Then, analogously to the dichotomy in $\boldsymbol{Q}_5$ we have, on the one hand, the union of the dark shaded areas in Figure 3, where the geodesic in $\boldsymbol{T}_4$ to (x_i, x_j) in the (closed) (i, j)-quadrant either is a straight linear segment or becomes one when the relevant quadrant is 'folded down' and, on the other hand, the union of the light shaded areas, the 'umbral' set $\mathcal{U}_{(x_i, x_j)}$, of points from which the geodesic in $\boldsymbol{T}_4$ to (x_i, x_j) passes through the cone point. Note that $\mathcal{U}_{(0,0)} = \emptyset$ and that $\mathcal{U}_{(sx_i, sx_j)} = \mathcal{U}_{(x_i, x_j)}$ for $s > 0$. Note also that, for each non-cone point (x_i, x_j), $\mathcal{U}_{(x_i, x_j)}$ accounts for 2/5 of the total area of $\boldsymbol{T}_4$.

For a chosen (i, j)-quadrant in $\boldsymbol{T}_4$, labelling the remaining axes and quadrants as in Figure 3, we define the following map ψ_{ij} from $\boldsymbol{T}_4$ to $\boldsymbol{Q}_5$. The shaded areas in the first four 5-cycles in Figure 3 are mapped isometrically onto their corresponding quadrant in $\boldsymbol{Q}_5$, so that they overlay one another. The shaded areas of the two remaining quadrants in Figure 3 are mapped to the w-semi-axis by identifying each ray through the cone point isometrically with the w-semi-axis. More precisely, ψ_{ij} is given by

$$
\begin{aligned}
(x_i, x_j) &\mapsto (x_i, x_j, 0); \\
(x_j, x_{j_r}) &\mapsto (-x_{j_r}, x_j, 0), \quad r = 1, 2; \\
(x_{j_s}, x_{k_{rs}}) &\mapsto (-x_{j_s}, 0, x_{k_{rs}}), \quad r, s = 1, 2; \\
(x_{k_{rs}}, x_{i_r}) &\mapsto (0, -x_{i_r}, x_{k_{rs}}), \quad r, s = 1, 2; \\
(x_{i_r}, x_i) &\mapsto (x_i, -x_{i_r}, 0), \quad r = 1, 2; \\
(x_{k_{2r}}, x_{k_{1s}}) &\mapsto \left(0, 0, \sqrt{x_{k_{2r}}^2 + x_{k_{1s}}^2}\right), \quad r, s = 1, 2 \text{ and } r \neq s.
\end{aligned}
\tag{3.1}
$$

Note that ψ_{ij} is continuous, so measurable, and that ψ_{ij} depends on the chosen (i, j)-quadrant, but is independent of points in that chosen quadrant. In terms of the mapping ψ_{ij}, the metric d_T on $\boldsymbol{T}_4$ is related to the metric d_Q on $\boldsymbol{Q}_5$ by

$$
d_T(t_1, t_2) = d_Q(\psi_{ij}(t_1), \psi_{ij}(t_2)),
\tag{3.2}
$$

where t_1 is a point in the (i, j)-quadrant and t_2 is arbitrary. However, if t_1 is not in the (i, j)-quadrant, this relation does not necessarily hold.
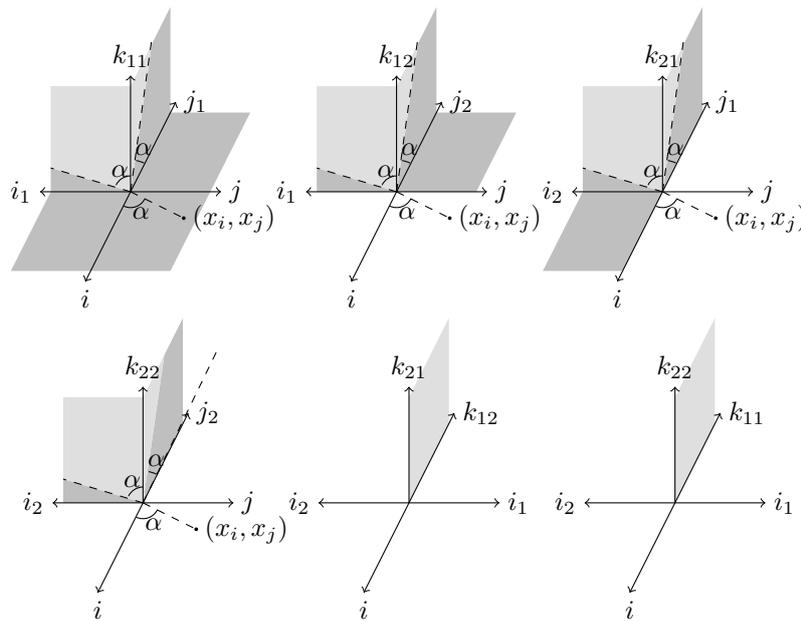
Figure 3: A decomposition of T_4 with respect to the (i,j)-quadrant: the fifteen shaded quadrants that form T_4; geodesics from (x_i, x_j) to the light shaded areas pass through the cone point.

We now consider central limit theorems for Fréchet means on the tree space T_4. Hence, for the remainder of the paper, we shall assume that μ is a probability measure on T_4 such that its Fréchet function is finite and that $\{\xi_l\}$ is a sequence of *iid* random variables in T_4 with probability measure μ.

In this section, we consider the case that the Fréchet mean $\hat{\xi}$ of μ lies in a top stratum. For this, by the symmetry of T_4, we may without loss of generality assume that $\hat{\xi}$ lies in the interior of the (i,j)-quadrant with coordinates $(\hat{x}_i, \hat{x}_j)$, so that both $\hat{x}_i$ and $\hat{x}_j$ are positive. Then, since the squared distance from a fixed point in T_4 is differentiable at $(\hat{x}_i, \hat{x}_j)$, it follows from a similar argument to that used in the previous section that $(\hat{x}_i, \hat{x}_j)$ is characterised by the analogous condition to that given by (2.1) in Q_5. Note that, similarly to the case for Q_5, although $\frac{1}{2}\mathrm{grad}_1 d_T(t, \cdot)^2\big|_{t=(\hat{x}_i, \hat{x}_j)}$ is a surjective map from T_4 to $\mathbb{R}^2$, it is not injective and we may regard it as a generalised log map for T_4 at $(\hat{x}_i, \hat{x}_j)$. Hence, by defining

$$\Phi_{(\hat{x}_i, \hat{x}_j)}(\cdot) = -\frac{1}{2}\mathrm{grad}_1 d_T(t, \cdot)^2\Big|_{t=(\hat{x}_i, \hat{x}_j)} + (\hat{x}_i, \hat{x}_j), \tag{3.3}$$

we have that $(\hat{x}_i, \hat{x}_j)$, where both $\hat{x}_i$ and $\hat{x}_j$ are positive, is the Fréchet mean of ξ_1 in T_4 if and only if

$$\int_{T_4} \Phi_{(\hat{x}_i, \hat{x}_j)}(\xi)\, d\mu(\xi) = (\hat{x}_i, \hat{x}_j). \tag{3.4}$$

This, as for Lemma 2.1 in the case of Q_5, gives the relationship between the Fréchet mean of ξ_1 in T_4, when that mean lies in a top stratum, and the Euclidean means of the random variables $\Phi_\cdot(\xi_1)$. The expression for d_T can be obtained from the expression (2.3) for d_Q using the map ψ_{ij} defined by (3.1) and the relationship between them given

by (3.2). In particular, for (x_i, x_j) in the (i, j)-quadrant of $\boldsymbol{T}_4$, we have $\psi_{ij}(x_i, x_j) = (x_i, x_j, 0)$ and, with a certain abuse of notation, we may identify $\psi_{ij}(x_i, x_j)$ with (x_i, x_j) when (x_i, x_j) is in the (i, j)-quadrant of $\boldsymbol{T}_4$. Then, a direct computation shows that $\Phi_{(\hat{x}_i, \hat{x}_j)}$ defined by (3.3) is the same as $\Psi_{(\hat{x}_i, \hat{x}_j)}$ composed with ψ_{ij}:

$$\Phi_{(\hat{x}_i, \hat{x}_j)} = \Psi_{\psi_{ij}(\hat{x}_i, \hat{x}_j)} \circ \psi_{ij}, \tag{3.5}$$

where Ψ is given by (2.2). When $\hat{\xi}$ lies in a top stratum of $\boldsymbol{T}_4$, this relationship between Φ and Ψ, together with Proposition 2.2, gives the central limit theorem for Fréchet means as follows. Since the sample Fréchet means $\hat{\xi}_n$ of $\{\xi_l\}$ converge to $\hat{\xi}$ a.s., $\hat{\xi}_n$ will lie in the interior of the (i, j)-quadrant when n is sufficiently large. Hence, in the following without loss of generality, we assume that that is the case for all n, so that the sample means $\hat{\xi}_n$ have coordinates $\hat{\xi}_n = (\hat{x}_i^n, \hat{x}_j^n)$ with both $\hat{x}_i^n$ and $\hat{x}_j^n$ positive.

Theorem 3.1. *Let μ be a probability measure on $\boldsymbol{T}_4$ with finite Fréchet function and with Fréchet mean $\hat{\xi} = (\hat{x}_i, \hat{x}_j)$ lying in the interior of the (i, j)-quadrant. Also, let $\{\xi_l\}$ be a sequence of iid random variables in $\boldsymbol{T}_4$ with probability measure μ and $\hat{\xi}_n$ be the sample Fréchet mean of $\xi_1, \cdots, \xi_n$. Then,*

$$\sqrt{n}(\hat{\xi}_n - \hat{\xi}) \equiv \sqrt{n}(\hat{x}_i^n - \hat{x}_i, \ \hat{x}_j^n - \hat{x}_j) \xrightarrow{d} N(0, A^\top V A), \qquad \text{as } n \to \infty,$$

where V is the covariance matrix of the random variable $\Phi_{(\hat{x}_i, \hat{x}_j)}(\xi_1)$,

$$A = \left\{ I + E\left[\|\xi_1\| \, 1_{\{\xi_1 \in \mathcal{U}_{(\hat{x}_i, \hat{x}_j)}\}} \right] M_{(\hat{x}_i, \hat{x}_j)} \right\}^{-1} \tag{3.6}$$

and $M_{(u,v)}$ is given by (2.6).

Proof. By (3.4), the sample Fréchet means satisfy

$$(\hat{x}_i^n, \hat{x}_j^n) = \frac{1}{n} \sum_{l=1}^{n} \Phi_{(\hat{x}_i^n, \hat{x}_j^n)}(\xi_l),$$

so that

$$\begin{aligned} &\sqrt{n}\{\hat{x}_i^n - \hat{x}_i, \hat{x}_j^n - \hat{x}_j\} \\ = \ & \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{ \Phi_{(\hat{x}_i^n, \hat{x}_j^n)}(\xi_l) - (\hat{x}_i, \hat{x}_j) \right\}. \end{aligned} \tag{3.7}$$

However, the point $\psi_{ij}(x_r, x_s)$ in the interior of the subset of $\boldsymbol{Q}_5$ determined by $w = 0$ is the Fréchet mean in $\boldsymbol{Q}_5$ of $\psi_{ij}(\xi_1)$ if and only if

$$\int \Psi_{\psi_{ij}(x_r, x_s)} \circ \psi_{ij}(\xi) \, d\mu(\xi) = \psi_{ij}(x_r, x_s)$$

by (2.5). Since $\psi_{ij}(\hat{x}_i, \hat{x}_j) = (\hat{x}_i, \hat{x}_j, 0)$, it follows from (3.4) and (3.5) that ξ_1 is a random variable in $\boldsymbol{T}_4$ with Fréchet mean $(\hat{x}_i, \hat{x}_j)$, where both $\hat{x}_i$ and $\hat{x}_j$ are positive, if and only if $(\hat{x}_i, \hat{x}_j, 0)$ is the Fréchet mean of $\psi_{ij}(\xi_1)$ in $\boldsymbol{Q}_5$. Hence, we can re-express (3.7) as

$$\begin{aligned} &\sqrt{n}\{\hat{x}_i^n - \hat{x}_i, \hat{x}_j^n - \hat{x}_j\} \\ = \ & \sqrt{n}\{\psi_{ij}(\hat{x}_i^n, \hat{x}_j^n) - \psi_{ij}(\hat{x}_i, \hat{x}_j)\} \\ = \ & \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{ \Psi_{\psi_{ij}(\hat{x}_i^n, \hat{x}_j^n)}(\psi_{ij}(\xi_l)) - \psi_{ij}(\hat{x}_i, \hat{x}_j) \right\}. \end{aligned}$$

Comparing the second equality with (2.8) and noting that $\mathcal{U}_{(\hat{x}_i, \hat{x}_j)}^Q = \psi_{ij}(\mathcal{U}_{(\hat{x}_i, \hat{x}_j)})$ and that $\|\xi_1\| = \|\psi_{ij}(\xi_1)\|$, the required result then follows from Proposition 2.2. $\qquad\square$

Recalling that $\mathcal{U}_{(\hat{x}_i,\hat{x}_j)}$ counts for 2/5 of the area of $\boldsymbol{T}_4$ and since $\Psi_{(\hat{x}_i,\hat{x}_j)}$ collapses $\mathcal{U}^Q_{(\hat{x}_i,\hat{x}_j)}$ to the line $\hat{x}_i v = \hat{x}_j u$, isometrically on each relevant ray through the origin, it follows that the distribution of $\Phi_{(\hat{x}_i,\hat{x}_j)}(\xi_1)$ has positive mass on the half line $\hat{x}_i v = \hat{x}_j u$ with $\text{sign}(u) = -\text{sign}(\hat{x}_i)$, as long as $\mu(\mathcal{U}_{(\hat{x}_i,\hat{x}_j)}) > 0$. In this case, the distribution of $\Phi_{(\hat{x}_i,\hat{x}_j)}(\xi_1)$ is singular. Note also the role played by $\mathcal{U}_{(\hat{x}_i,\hat{x}_j)}$ in the expression for the matrix A defined by (3.6).

Although Theorem 3.1 concerns the case that Fréchet means lie in the manifold part of $\boldsymbol{T}_4$ and, as noted earlier, $\Phi_{(x_i,x_j)}(\cdot) - (x_i, x_j)$ plays a similar role to that of the log maps for Riemannian manifolds, neither the result nor the proof of Theorem 3.1 is a special case of [16], as [16] deals with complete and simply connected Riemannian manifolds and as $\Phi_{(\hat{x}_i,\hat{x}_j)}$ is generally not a $\mathcal{C}^2$-injective map on the support of μ except in very special circumstances. For the case that $\Phi_{(\hat{x}_i,\hat{x}_j)}$ is an injective map on the support of μ, the support of μ would be so restricted that one would be able to deduce the result directly from the classical central limit theorem for random variables in Euclidean space. The result of Theorem 3.1 also differs from that of [3] since, in addition to the fact of its being neither $\mathcal{C}^2$ nor injective, the map $\Phi_{(x_i,x_j)}$ also depends on the point (x_i, x_j).

4 Fréchet means in a co-dimension one stratum of $\boldsymbol{T}_4$

We now turn to consider the case where the Fréchet mean $\hat{\xi}$ of the probability measure μ on $\boldsymbol{T}_4$ lies in a co-dimension one stratum. Without loss of generality, we assume that it lies on the open i-semi-axis, the co-dimension one stratum corresponding to the i-vertex in the Peterson graph. Then, in terms of the coordinate system on $\boldsymbol{T}_4$ that we adopt, there is more than one way to represent $\hat{\xi}$: either as $(\hat{x}_i, 0)$ in the boundary of the (i, j)-quadrant or as $(0, \hat{x}_i)$ in the boundary of the (i_r, i)-quadrant for $r = 1, 2$, where $\hat{x}_i > 0$. To indicate explicitly the quadrant in which we are considering $\hat{\xi}$ to lie, we shall write $(\hat{x}_i, 0)$ as $(\hat{x}_i, 0_j)$ when it is to be regarded as a point in the (i, j)-quadrant and, similarly, write $(0, \hat{x}_i)$ as $(0_{i_r}, \hat{x}_i)$ when it is to be regarded as a point in the (i_r, i)-quadrant.

Note that, when $\hat{\xi} = (\hat{x}_i, \hat{x}_j)$ lies on the open i-semi-axis, we can regard the union of the three half planes that are bounded by the full i-axis and contain, respectively, the (i, j)-, (i_1, i)- and (i_2, i)-quadrants in which $\hat{\xi}$ lies as 'the tangent space' of $\boldsymbol{T}_4$ at $\hat{\xi}$. However, since it is no longer true that any neighbourhood of $\hat{\xi}$ is a manifold, the criterion equivalent to (2.1) for a point to be the Fréchet mean of μ cannot be applied. Instead, it may be characterised by requiring the non-negativity of the directional derivatives, along the j-, i_1- and i_2-semi-axes, of the Fréchet function for μ, as given by (1.1), together with its derivative along the i-semi-axis being zero, at $(\hat{x}_i, 0_j)$. By continuity, the directional derivative with respect to the first variable of $\frac{1}{2}d_T(t, (x_r, x_s))^2$ at $t = (\hat{x}_i, 0_j)$ along the j direction has the same expression as $\Psi_{\psi_{ij}(\hat{x}_i,0_j)} \circ \psi_{ij} - (\hat{x}_i, 0_j)$. Since (i, j) is arbitrary, there are similar expressions for the directional derivatives along the i_r, $r = 1, 2$, directions. Thus, using the relationship (3.5) to extend the definition of $\Phi_{(\hat{x}_i,\hat{x}_j)}$ to any point $(\hat{x}_i, \hat{x}_j)$ lying in the closed (i, j)-quadrant, the requirement that these three directional derivatives be non-negative may be re-expressed as

$$\text{the 2nd coordinate of } E[\Phi_{(\hat{x}_i,0_j)}(\xi)] \leqslant 0,$$
$$\text{the 1st coordinate of } E[\Phi_{(0_{i_r},\hat{x}_i)}(\xi)] \leqslant 0, \tag{4.1}$$

where $r = 1, 2$ and ξ is a random variable in $\boldsymbol{T}_4$ with probability measure μ. Writing

$$I_{i_1} = \left\{ \int_{(i_1,i)\text{-quadrant}} + \int_{(k_{11},i_1)\text{-quadrant}} + \int_{(k_{12},i_1)\text{-quadrant}} \right\} x_{i_1}\, d\mu$$

$$I_{i_2} = \left\{ \int_{(i_2,i)\text{-quadrant}} + \int_{(k_{21},i_2)\text{-quadrant}} + \int_{(k_{22},i_2)\text{-quadrant}} \right\} x_{i_2}\, d\mu$$

$$I_j = \left\{ \int_{(i,j)\text{-quadrant}} + \int_{(j,j_1)\text{-quadrant}} + \int_{(j,j_2)\text{-quadrant}} \right\} x_j\, d\mu$$

then, a direct computation using ψ_{ij} and (3.2) shows that the three inequalities in (4.1) are equivalent to

$$I_{i_1} + I_{i_2} \geqslant I_j, \quad I_{i_2} + I_j \geqslant I_{i_1} \text{ and } I_j + I_{i_1} \geqslant I_{i_2}. \tag{4.2}$$

Hence, we have the following result to characterise a point in the open i-semi-axis of $\boldsymbol{T}_4$ as the Fréchet mean of μ.

Lemma 4.1. *A point $\hat{\xi} = (\hat{x}_i, 0_j)$ on the open i-semi-axis is the Fréchet mean of μ if and only if the three inequalities in (4.2) are satisfied, together with $\hat{x}_i$ satisfying*

$$\begin{aligned}
\hat{x}_i = {} & \left\{ \int_{(i,j)-\text{quadrant}} + \sum_{r=1}^{2} \int_{(i_r,i)-\text{quadrant}} \right\} x_i\, d\mu \\
& - \sum_{r=1}^{2} \int_{(j,j_r)-\text{quadrant}} x_{j_r}\, d\mu - \sum_{r,s=1}^{2} \int_{(k_{rs},i_r)-\text{quadrant}} x_{k_{rs}}\, d\mu \\
& - \int_{\mathcal{U}_{\hat{\xi}}} \|(x_r, x_s)\|\, d\mu.
\end{aligned} \tag{4.3}$$

Note that $\mathcal{U}_{\hat{\xi}}$ now becomes the union of six quadrants: the (j_r, k_{sr})-, (k_{11}, k_{22})- and (k_{12}, k_{21})-quadrants, where $r, s = 1, 2$.

From the proof of Theorem 3.1 and the result of Lemma 4.1, we obtain the following relationship between the Fréchet means of μ in $\boldsymbol{T}_4$ and of $\mu \circ \psi_{ij}^{-1}$ in $\boldsymbol{Q}_5$. Note that, if ξ is a random variable in $\boldsymbol{T}_4$ with probability measure μ, then $\psi_{ij}(\xi)$ is a random variable in $\boldsymbol{Q}_5$ with probability measure $\mu \circ \psi_{ij}^{-1}$.

Corollary 4.2. *Assume that the Fréchet mean of μ lies in the (i, j)-quadrant.*

(a) *The Fréchet mean of μ is $(\hat{x}_i, \hat{x}_j)$ with $\hat{x}_i > 0$ and $\hat{x}_j > 0$ if and only if $(\hat{x}_i, \hat{x}_j, 0)$ is the Fréchet mean of $\mu \circ \psi_{ij}^{-1}$.*

(b) *If the Fréchet mean of μ is $(\hat{x}_i, 0_j)$ with $\hat{x}_i > 0$, then $\hat{x}_i$ is the first coordinate of the Fréchet mean of $\mu \circ \psi_{ij}^{-1}$.*

(c) *If the Fréchet mean of μ is $(\hat{x}_i, 0_j)$ with $\hat{x}_i > 0$ and if $I_j = 0$, then $(\hat{x}_i, 0, 0)$ is the Fréchet mean of $\mu \circ \psi_{ij}^{-1}$.*

In general, if $(\hat{x}_i, 0_j)$ with $\hat{x}_i > 0$ is the Fréchet mean of μ, $(\hat{x}_i, 0, 0)$ is not necessarily the Fréchet mean of $\mu \circ \psi_{ij}^{-1}$, since away from its cone point, $\boldsymbol{Q}_5$ is a Riemannian manifold so that the criterion (2.1) holds there.

The conditions (4.2) also have the following consequence for the behaviour of the sample Fréchet mean $\hat{\xi}_n$ of $\xi_1, \cdots, \xi_n$, analogous to the result of Theorem 4.3(1) in [13] for open book decompositions.

Lemma 4.3. *Let $\hat{\xi} = (\hat{x}_i, 0_j)$ on the open i-semi-axis be the Fréchet mean of μ and assume that the first of the inequalities (4.2) is strict. If $\{\xi_l\}$ is a sequence of iid random variables in T_4 with probability measure μ then, for all sufficiently large n, the sample Fréchet mean $\hat{\xi}_n$ cannot lie in the interior of the (i, j)-quadrant.*

Proof. The assumption that the first of the inequalities in (4.2) is strict implies that the second coordinate of $E[\Psi_{\psi_{ij}(\hat{x}_i, 0_j)}(\psi_{ij}(\xi_1))]$ is negative. Then, by the law of large numbers, the second coordinate of $\frac{1}{n} \sum_{l=1}^{n} \Psi_{\psi_{ij}(\hat{x}_i, 0_j)}(\psi_{ij}(\xi_l))$ is also negative when n is sufficiently large. On the other hand, since $\hat{\xi}_n$ converges to $\hat{\xi} = (\hat{x}_i, 0_j)$, the continuity of ψ_{ij} implies that $\psi_{ij}(\hat{\xi}_n)$ will be close to $(\hat{x}_i, 0, 0)$ for large n. In particular, for large n, the first coordinate of $\psi_{ij}(\hat{\xi}_n)$ is positive and the third zero. Thus, it follows from (2.7) that, for large n and for $1 \leqslant l \leqslant n$,

$$\text{the 2nd coordinate of } \left\{ \Psi_{\psi_{ij}(\hat{\xi}_n)}(\psi_{ij}(\xi_l)) - \Psi_{\psi_{ij}(\hat{x}_i, 0_j)}(\psi_{ij}(\xi_l)) \right\}$$
$$\approx -\text{ the 2nd coordinate of } \psi_{ij}(\hat{\xi}_n) \times \|\psi_{ij}(\xi_l)\| \frac{1}{\hat{x}_i} 1_{\{\psi_{ij}(\xi_l) \in \mathcal{U}^Q_{\psi_{ij}(\hat{x}_i, 0_j)} \cup \mathcal{U}^Q_{\psi_{ij}(\hat{\xi}_n)}\}}.$$

This implies that

$$
\begin{aligned}
&\text{the 2nd coordinate of } \frac{1}{n} \sum_{l=1}^{n} \Psi_{\psi_{ij}(\hat{\xi}_n)}(\psi_{ij}(\xi_l)) \\
\approx \ &\text{the 2nd coordinate of } \frac{1}{n} \sum_{l=1}^{n} \Psi_{\psi_{ij}(\hat{x}_i, 0_j)}(\psi_{ij}(\xi_l)) \\
&- \text{the 2nd coordinate of } \psi_{ij}(\hat{\xi}_n) \times \frac{1}{\hat{x}_i} \\
&\times \frac{1}{n} \sum_{l=1}^{n} \|\psi_{ij}(\xi_l)\| 1_{\{\psi_{ij}(\xi_l) \in \mathcal{U}^Q_{\psi_{ij}(\hat{x}_i, 0_j)} \cup \mathcal{U}^Q_{\psi_{ij}(\hat{\xi}_n)}\}}.
\end{aligned}
\tag{4.4}
$$

Suppose, if possible, that $\hat{\xi}_n$ lies in the interior of the (i, j)-quadrant with coordinates $(\hat{x}_i^n, \hat{x}_j^n)$, so that $\hat{x}_j^n > 0$. By Corollary 4.2(a) it would follow that $(\hat{\xi}_n, 0)$ is the sample Fréchet mean of $\psi_{ij}(\xi_1), \cdots, \psi_{ij}(\xi_n)$. Then, the left hand side of (4.4) would be equal to $\hat{x}_j^n$ which is positive. However, the right hand side of (4.4) is negative by the negativity of its first term on account of the given assumption. This contradiction shows that, for all sufficiently large n, $\hat{\xi}_n$ cannot lie in the interior of the (i, j)-quadrant. $\square$

Analogously, if the second, respectively third, inequality in (4.2) is strict then, for all sufficiently large n, $\hat{\xi}_n$ cannot lie in the interior of the (i_1, i)-quadrant, respectively the (i_2, i)-quadrant. Similar generalisations will hold for the following theorem on the form of the central limit theorem when $\hat{\xi}$ lies in a co-dimension one stratum, where Φ is the extension to the closed quadrant that occurs in (4.1).

Theorem 4.4. *Let μ be a probability measure on T_4 with finite Fréchet function and with Fréchet mean $\hat{\xi} = (\hat{x}_i, 0)$ lying on the open i-semi-axis. Also, let $\{\xi_l\}$ be a sequence of iid random variables in T_4 with probability measure μ and $\hat{\xi}_n$ be the sample Fréchet mean of $\xi_1, \cdots, \xi_n$.*

(a) *If all three inequalities in (4.2) are strict then, for all sufficiently large n, $\hat{\xi}_n$ will lie on the i-semi-axis and the sequence $\sqrt{n}\{\hat{x}_i^n - \hat{x}_i\}$ of the first coordinates of $\sqrt{n}\{\hat{\xi}_n - \hat{\xi}\}$ converges in distribution to $N(0, \sigma^2)$ as $n \to \infty$, where σ^2 is the variance of the first coordinate of the Euclidean random vector $\Phi_{(\hat{x}_i, 0_j)}(\xi_1)$.*

(b) *If the first inequality in (4.2) is an equality and the other two are strict then, for all sufficiently large n, $\hat{\xi}_n$ will lie in the (i,j)-quadrant and*

$$\sqrt{n}\{\hat{\xi}_n - \hat{\xi}\} \equiv \sqrt{n}\{(\hat{x}_i^n, \hat{x}_j^n) - (\hat{x}_i, 0_j)\} \xrightarrow{d} (\eta_1, \max\{0, \eta_2\}), \quad \text{as } n \to \infty,$$

where $(\eta_1, \eta_2) \sim N(0, A^\top V A)$, V is the covariance matrix of $\Phi_{(\hat{x}_i, 0_j)}(\xi_1)$ and A is as in (3.6) with $\hat{x}_j = 0$, and where '$\equiv$' is understood to hold for all sufficiently large n.

(c) *If the first two inequalities in (4.2) are equalities and the third is strict then, for all sufficiently large n, $\hat{\xi}_n$ will lie either in the (i,j)-quadrant or in the (i_1, i)-quadrant and the limiting distribution of $\sqrt{n}\{\hat{\xi}_n - \hat{\xi}\}$, as $n \to \infty$, takes the same form as that given in Theorem 3.1 with $\hat{x}_j = 0$, where the coordinates of $\hat{\xi}_n$ are taken as $(\hat{x}_i^n, \hat{x}_j^n)$, respectively $(\hat{x}_i^n, -\hat{x}_{i_1}^n)$, if $\hat{\xi}_n$ is in the (i,j)-quadrant, respectively the (i_1, i)-quadrant.*

(d) *If all the equalities in (4.2) are actually equalities, then we have the same result as in (a).*

Proof. (a) By Lemma 4.3, when n is sufficiently large, $\hat{\xi}_n$ must lie on the i-semi-axis so that it has coordinates $(\hat{x}_i^n, 0_j)$ (equivalently $(0_{i_r}, \hat{x}_i^n)$). By Corollary 4.2(b), $\hat{x}_i^n$ is the first coordinate of the sample Fréchet mean of $\psi_{ij}(\xi_1), \cdots, \psi_{ij}(\xi_n)$ in $\boldsymbol{Q}_5$. Thus, by (2.7),

$$\text{the 1st coordinate of } \left\{ \Psi_{\psi_{ij}(\hat{\xi}_n)}(\psi_{ij}(\xi_l)) - \Psi_{\psi_{ij}(\hat{x}_i, 0_j)}(\psi_{ij}(\xi_l)) \right\}$$
$$\approx \quad \|\psi_{ij}(\xi_l)\| \, o(\|\hat{\xi}_n - \hat{\xi}\|) \, 1_{\{\psi_{ij}(\xi_l) \in \mathcal{U}^Q_{\psi_{ij}(\hat{x}_i, 0_j)} \cup \mathcal{U}^Q_{\psi_{ij}(\hat{\xi}_n)}\}}.$$

Then a modification of the argument of Section 2 to restrict it to the first coordinates of $\{\psi_{ij}(\xi_l)\}$ gives the required limiting distribution of $\sqrt{n}\{\hat{x}_i^n - \hat{x}_i\}$.

(b) By Corollary 4.2(c), $(\hat{x}_i, 0, 0)$ is the Fréchet mean of $\psi_{ij}(\xi_1)$. On the other hand, we deduce from the assumed strict inequalities and Lemma 4.3 that, when n is sufficiently large, $\hat{\xi}_n$ can only lie in the (closed) (i,j)-quadrant, so that it has coordinates $\hat{\xi}_n = (\hat{x}_i^n, \hat{x}_j^n)$ where we may assume that $\hat{x}_i^n > 0$. Then, by Corollary 4.2, $\hat{x}_i^n$ is the first coordinate of the sample Fréchet mean of $\psi_{ij}(\xi_1), \cdots, \psi_{ij}(\xi_n)$. However, $\hat{x}_j^n > 0$, if and only if both of the first two coordinates of $\psi_{ij}(\hat{\xi}_n)$ are positive and by Corollary 4.2(a), in that case, $(\hat{x}_i^n, \hat{x}_j^n, 0)$ is the sample Fréchet mean of $\psi_{ij}(\xi_1), \cdots, \psi_{ij}(\xi_n)$. Thus, $\hat{x}_j^n$ is zero if and only if the second coordinate of the sample Fréchet mean of $\psi_{ij}(\xi_1), \cdots, \psi_{ij}(\xi_n)$ is non-positive. Hence, Proposition 2.2, the central limit theorem for the sample Fréchet means of $\{\psi_{ij}(\xi_l)\}$, gives the central limit theorem for the sample Fréchet means of $\{\xi_l\}$.

(c) In this case, by Corollary 4.2(c), $(\hat{\xi}, 0)$ is the Fréchet mean both of $\psi_{ij}(\xi_1)$ and of $\psi_{i_1, i}(\xi_1)$ in $\boldsymbol{Q}_5$. So that

$$(\hat{x}_i, 0_j) = \int_{T_4} \Phi_{(\hat{x}_i, 0_j)}(\xi) \, d\mu(\xi) \text{ and } (0_{i_1}, \hat{x}_i) = \int_{T_4} \Phi_{(0_{i_1}, \hat{x}_i)}(\xi) \, d\mu(\xi).$$

Moreover, the integral I_{i_2} becomes zero and so, since the integrand is non-negative, $\mu(\mathcal{C}) = 0$ where $\mathcal{C}$, the domain of integration of I_{i_2}, is the union of the (i_2, i)-, (k_{21}, i_2)- and (k_{22}, i_2)-quadrants with the i-, k_{21}- and k_{22}-semi-axes removed. It is now more convenient to represent the union of the (i,j)- and (i_1, i)-quadrants by coordinates in the (x, y)-half-plane with $x \geqslant 0$. For this, we map:

$$(x_i, x_j) \mapsto (x_i, x_j) \text{ and } (x_{i_1}, x_i) \mapsto (x_i, -x_{i_1}) = (x_{i_1}, x_i)R, \tag{4.5}$$

where R is the rotation matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Similarly we define maps $\tilde{\Phi}_{(x,y)}$ to accord with this by $\tilde{\Phi}_{(x_i, -x_{i_1})}(\cdot) = \Phi_{(x_{i_1}, x_i)}(\cdot) R$, while $\tilde{\Phi}_{(x_i, x_j)} = \Phi_{(x_i, x_j)}$. Since $\mu(\mathcal{C}) = 0$ and since, restricted to $\mathcal{C}^c$, the complement of the set $\mathcal{C}$, $\Phi_{(x_i, 0_j)}(\cdot) = \Phi_{(0_{i_1}, x_i)}(\cdot) R$, the map $\tilde{\Phi}$ is indeed a.s. well defined for points $(x_i, 0)$. Under this new coordinate system we have, in particular, that $\hat{\xi} = (\hat{x}_i, 0_j) = (\hat{x}_i, 0_{i_1})$ and that

$$\hat{\xi} = (\hat{x}_i, 0) = \int_{T_4} \tilde{\Phi}_{(\hat{x}_i, 0_{i_1})}(\xi) \, d\mu(\xi). \tag{4.6}$$

By Lemma 4.3, the given assumption also implies that, for sufficiently large n, $\hat{\xi}_n$ will a.s. lie either in the (i,j)-quadrant or in the (i_1, i)-quadrant. If $\hat{\xi}_n$ lies in the interior of the (i,j)-quadrant with coordinates $\hat{\xi}_n = (\hat{x}_i^n, \hat{x}_j^n)$, then $\hat{x}_j^n > 0$ and

$$(\hat{x}_i^n, \hat{x}_j^n) = \frac{1}{n} \sum_{l=1}^{n} \Phi_{(\hat{x}_i^n, \hat{x}_j^n)}(\xi_l) = \frac{1}{n} \sum_{l=1}^{n} \tilde{\Phi}_{(\hat{x}_i^n, \hat{x}_j^n)}(\xi_l) \tag{4.7}$$

and, if $\hat{\xi}_n$ lies in the interior of the (i_1, i)-quadrant with (original) coordinates $\hat{\xi}_n = (\hat{x}_{i_1}^n, \hat{x}_i^n)$, then

$$(\hat{x}_{i_1}^n, \hat{x}_i^n) = \frac{1}{n} \sum_{l=1}^{n} \Phi_{(\hat{x}_{i_1}^n, \hat{x}_i^n)}(\xi_l),$$

i.e.

$$(\hat{x}_i^n, -\hat{x}_{i_1}^n) = \frac{1}{n} \sum_{l=1}^{n} \tilde{\Phi}_{(\hat{x}_i^n, -\hat{x}_{i_1}^n)}(\xi_l). \tag{4.8}$$

If $\hat{\xi}_n$ lies on the open i-semi-axis with coordinates $(\hat{x}_i^n, 0)$ then since, locally there, the support of μ is diffeomorphic with $\mathbb{R}^2$, we also have

$$(\hat{x}_{i_1}^n, 0) = \frac{1}{n} \sum_{l=1}^{n} \Phi_{(\hat{x}_{i_1}^n, 0)}(\xi_l) = \frac{1}{n} \sum_{l=1}^{n} \tilde{\Phi}_{(\hat{x}_i^n, 0)}(\xi_l) \quad \text{a.s..} \tag{4.9}$$

Recalling that, under the new coordinate system defined by (4.5),

$$\hat{\xi}_n \equiv \begin{cases} (\hat{x}_i^n, \hat{x}_j^n) & \text{if } \hat{\xi}_n \text{ is in the } (i,j)\text{-quadrant} \\ (\hat{x}_i^n, -\hat{x}_{i_1}^n) & \text{if } \hat{\xi}_n \text{ is in the } (i_1, i)\text{-quadrant} \end{cases}$$

we have by (4.6), (4.7), (4.8) and (4.9) that, in terms of the new coordinates,

$$
\begin{aligned}
\sqrt{n}\{\hat{\xi}_n - \hat{\xi}\} &= \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{ \Phi_{(\hat{x}_i, 0)}(\xi_l) - (\hat{x}_i, 0) \right\} \\
&+ \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \left\{ \tilde{\Phi}_{\hat{\xi}_n}(\xi_l) - \Phi_{(\hat{x}_i, 0)}(\xi_l) \right\}.
\end{aligned}
$$

Hence, a similar argument to that of the proof for Theorem 3.1, we see that the central limit theorem now takes the same form as in that theorem with $\hat{x}_j = 0$.

(d) Noting that all integrands in (4.2) are non-negative, the three equalities will together imply that μ must be concentrated on the union of the i-semi-axis and $\mathcal{U}_{(\hat{x}_i, 0_j)}$. Then, μ must have positive mass on the i-semi-axis. Otherwise, it would contradict (4.3), as its left hand side is positive by the assumption and its right hand side would become negative. This results in $\Phi_{(\hat{x}_i, 0_j)}(\xi_1)$ being a one-dimensional random variable

on $\mathbb{R}$ with mean $(\hat{x}_i, 0_j)$. Hence, the measure induced on $\mathbb{R}^2$ from μ by $\Phi_{(x_i, x_j)}$ for (x_i, x_j) in the (i,j)-quadrant has support contained in the half plane with non-positive second coordinate. Similarly, for $r = 1, 2$, the measure induced on $\mathbb{R}^2$ from μ by $\Phi_{(x_{i_r}, x_i)}$ for (x_{i_r}, x_i) in the (i_r, i)-quadrant has support contained in the half plane with non-positive first coordinate. This constraint on μ implies that $\hat{\xi}_n$ lies on the i-semi-axis for all sufficiently large n. Otherwise, if $\hat{\xi}_n = (\hat{x}_i^n, \hat{x}_j^n)$ lies in the interior of the (i,j)-quadrant, say, then, on the one hand, $\hat{x}_j^n > 0$ and, on the other hand, on account of the features of μ, we must have

$$\hat{x}_j^n = \text{the 2nd coordinate of } \frac{1}{n} \sum_{l=1}^{n} \Phi_{(\hat{x}_i^n, \hat{x}_j^n)}(\xi_l) \leqslant 0.$$

Thus, the argument for (a) implies that, when the inequalities in (4.2) are all equalities, the central limit theorem for the sample Fréchet means takes the same form as that when the three inequalities are all strict. □

5 Fréchet means at the cone point of T_4

The cone point o being the Fréchet mean of μ is equivalent to the fact that, for any i, j and any non-cone point (x_i, x_j) in the (i,j)-quadrant, we have

$$\int_{T_4} d_T((x_i, x_j), \xi)^2 d\mu(\xi) > \int_{T_4} d_T(o, \xi)^2 d\mu(\xi),$$

which is equivalent to all possible directional derivatives of the Fréchet function for μ being non-negative at the cone point. Recalling that $\Psi_{(u,v)} = \Psi_{s(u,v)}$, $s > 0$, for Ψ defined by (2.2), it then follows from the relationship between Φ and Ψ that $\Phi_{(x_i, x_j)} = \Phi_{s(x_i, x_j)}$ for any $s > 0$, Φ here being the extension to the closed (i,j)-quadrant defined at the beginning of the previous section. Using this invariance, it is more transparent to write $\Phi_{(x_i, x_j)} = \Phi_{\theta_{ij}}$ in studying the limiting behaviour of the sample Fréchet means $\hat{\xi}_n$ when $\hat{\xi}$ is at the cone point, where $\theta_{ij} \in [0, \pi/2]$ is determined by $\tan \theta_{ij} = x_j/x_i$. With this new notation, the above condition for the cone point o to be the Fréchet mean of μ is equivalent to the fact that, for any (i,j)-quadrant and any $\theta_{ij} \in [0, \pi/2]$,

$$\left\langle (\cos \theta_{ij}, \sin \theta_{ij}), \int_{T_4} \Phi_{\theta_{ij}}(\xi) \, d\mu(\xi) \right\rangle \leqslant 0. \tag{5.1}$$

For $\theta_{ij} \in (0, \pi/2)$, the condition (5.1) implies that, if both of its coordinates are non-negative, then $\int_{T_4} \Phi_{\theta_{ij}}(\xi) \, d\mu(\xi)$ must be at the origin. On the other hand, if at least one of the coordinates of $\int_{T_4} \Phi_{\theta_{ij}}(\xi) \, d\mu(\xi)$ is negative, it follows from (3.4) that, when n is sufficiently large, the sample Fréchet mean $\hat{\xi}_n$ of $\xi_1, \cdots, \xi_n$ will not be on the half line in the (i,j)-quadrant determined by θ_{ij}. Note that, when θ_{ij} varies, the distribution of $\Phi_{\theta_{ij}}(\xi)$ generally varies too.

Fix an arbitrary (i,j)-quadrant and let

$$\Theta_{ij} = \left\{ \theta_{ij} \in [0, \pi/2] \, \middle| \, \int_{T_4} \Phi_{\theta_{ij}}(\xi) \, d\mu(\xi) = 0 \text{ when } \theta_{ij} \neq 0, \pi/2; \text{ the 1st} \right.$$
$$\left. \text{(2nd) coordinate of } \int_{T_4} \Phi_{\theta_{ij}}(\xi) \, d\mu(\xi) = 0 \text{ when } \theta_{ij} = 0 \, (\pi/2) \right\}. \tag{5.2}$$

Lemma 5.1. *The restriction of Θ_{ij} to $(0, \pi/2)$ either is the empty set, or forms an interval, where the latter includes the case of a single point.*

Proof. If $\Theta_{ij} \cap (0, \pi/2)$ is not an empty set, let $0 < \theta_{ij}^{(1)} < \theta_{ij}^{(2)} < \pi/2$ be two distinguished angles within this set and let $(x_i^{(l)}, x_j^{(l)})$, $l = 1, 2$, be two points close to the cone point in the (i, j)-quadrant lying on the two corresponding half lines. Then, it follows from (2.7) that

$$\left\langle \int_{T_4} \left\{ \Phi_{\theta_{ij}^{(2)}}(\xi) - \Phi_{\theta_{ij}^{(1)}}(\xi) \right\} d\mu(\xi), \, (-\sin\theta_{ij}^{(1)}, \cos\theta_{ij}^{(1)}) \right\rangle$$
$$\approx \quad -\frac{1}{\|(x_i^{(1)}, x_j^{(1)})\|} \left\langle (-\sin\theta_{ij}^{(1)}, \cos\theta_{ij}^{(1)}), \, (x_i^{(2)} - x_i^{(1)}, x_j^{(2)} - x_j^{(1)}) \right\rangle$$
$$\times \int_{T_4} \|\xi\| \, 1_{\{\xi \in \mathcal{U}_{(x_i^{(1)}, x_j^{(1)})} \cup \mathcal{U}_{(x_i^{(2)}, x_j^{(2)})}\}} d\mu(\xi).$$

By the definition of Θ_{ij}, the left hand side of the above is equal to zero. Since

$$\left\langle (-\sin\theta_{ij}^{(1)}, \cos\theta_{ij}^{(1)}), \, (x_i^{(2)} - x_i^{(1)}, x_j^{(2)} - x_j^{(1)}) \right\rangle \neq 0,$$

Θ_{ij} contains two distinguished points $\theta_{ij}^{(l)}$, $l = 1, 2$, only if

$$\int_{T_4} \|\xi\| \, 1_{\{\xi \in \mathcal{U}_{(x_i^{(1)}, x_j^{(1)})} \cup \mathcal{U}_{(x_i^{(2)}, x_j^{(2)})}\}} d\mu(\xi) = 0. \tag{5.3}$$

However, for all $\theta_{ij} \in [\theta_{ij}^{(1)}, \theta_{ij}^{(2)}]$, $\mathcal{U}_{(\cos\theta_{ij}, \sin\theta_{ij})} \subseteq \mathcal{U}_{(x_i^{(1)}, x_j^{(1)})} \cup \mathcal{U}_{(x_i^{(2)}, x_j^{(2)})}$. Then (5.3) implies that the random variables $\Phi_{\theta_{ij}}(\xi)$ have the same distribution as that of $\Phi_{\theta_{ij}^{(1)}}(\xi)$, where ξ is a random variable in T_4 with probability measure μ. Hence, in particular, $\int_{T_4} \Phi_{\theta_{ij}}(\xi) d\mu(\xi) = 0$, so that $[\theta_{ij}^{(1)}, \theta_{ij}^{(2)}] \subset \Theta_{ij}$. $\qquad\square$

Theorem 5.2. *Assume that the Fréchet mean $\hat{\xi}$ of μ is at the cone point. Let $\{\xi_l\}$ be a sequence of iid random variables in T_4 with probability measure μ and $\hat{\xi}_n$ be the sample Fréchet mean of $\xi_1, \cdots, \xi_n$. With Θ_{ij} defined by (5.2):*

(a) *if $\Theta_{ij} = \emptyset$ then, for all sufficiently large n, the only position for $\hat{\xi}_n$ to be in the (i, j)-quadrant is at the cone point;*

(b) *if $\Theta_{ij} \cap (0, \pi/2) \neq \emptyset$ then, for any Borel set $B \subset \mathbb{R}^2$,*

$$\lim_{n\to\infty} P\left(\sqrt{n}\hat{\xi}_n 1_{\{\hat{\xi}_n \in \text{interior } (i, j)\text{-quadrant}\}} \in B \right) = P(Z \in B \cap C_{ij}),$$

where C_{ij} is the cone spanned by the cone point and the angles in $\Theta_{ij} \cap (0, \pi/2)$, $Z \sim N(0, V)$, V is the covariance matrix of $\Phi_{\theta_{ij}}(\xi_1)$ and $\theta_{ij} \in \Theta_{ij} \cap (0, \pi/2)$.

Proof. (a) Let

$$d_{ij} = \min\left\{ \left\| \int_{T_4} \Phi_{\theta_{ij}}(\xi) d\mu(\xi) \right\| \, \Big| \, \theta_{ij} \in [0, \pi/2] \right\}.$$

If $\Theta_{ij} = \emptyset$, so that $\left\| \int_{T_4} \Phi_{\theta_{ij}}(\xi) d\mu(\xi) \right\| > 0$ for all $\theta_{ij} \in [0, \pi/2]$, it follows from the continuity of $\left\| \int_{T_4} \Phi_{\theta_{ij}}(\xi) d\mu(\xi) \right\|$ in θ_{ij} that $d_{ij} > 0$.

Choose a finite set $\Theta_{ij}^* = \{\theta_{ij}^{(k)} \mid k = 1, \cdots, K_{ij}\} \subset [0, \pi/2]$ such that, for any $\theta_{ij} \in [0, \pi/2]$, there is a $\theta_{ij}^{(k)} \in \Theta_{ij}^*$ satisfying both

$$|\theta_{ij} - \theta_{ij}^{(k)}| < d_{ij}/(10 \, E[\|\xi_1\|]) \tag{5.4}$$

and

$$\|s(\cos\theta_{ij}, \sin\theta_{ij}) \pm d_{ij}(-\sin\theta_{ij}^{(k)}, \cos\theta_{ij}^{(k)})\| \geqslant d_{ij}/2, \qquad \forall s > 0. \tag{5.5}$$

By the classical central limit theorem there is an n_0 such that, for all $n > n_0$ and for all $\theta_{ij}^{(k)} \in \Theta_{ij}^*$,

$$\left\| \frac{1}{n} \sum_{l=1}^{n} \Phi_{\theta_{ij}^{(k)}}(\xi_l) - \int_{T_4} \Phi_{\theta_{ij}^{(k)}}(\xi)\, d\mu(\xi) \right\| < d_{ij}/4$$

and

$$\left| \frac{1}{n} \sum_{l=1}^{n} \|\xi_l\| - E[\|\xi_1\|] \right| < E[\|\xi_1\|]/4.$$

For any given $\theta_{ij} \in [0, \pi/2]$, if $\theta_{ij}^{(k)} \in \Theta_{ij}^*$ such that (5.4) and (5.5) hold then, for $n > n_0$, we have, by applying (2.7) and using the relationship between Ψ and Φ with $(x_i^{(1)}, x_j^{(1)}) = (\cos\theta_{ij}, \sin\theta_{ij})$ and $(x_i^{(2)}, x_j^{(2)}) = (\cos\theta_{ij}^*, \sin\theta_{ij}^*)$, that

$$
\begin{aligned}
&\left\| \frac{1}{n} \sum_{l=1}^{n} \Phi_{\theta_{ij}}(\xi_l) - \int_{T_4} \Phi_{\theta_{ij}^*}(\xi)\, d\mu(\xi) \right\| \\
\leqslant\quad & \left\| \frac{1}{n} \sum_{l=1}^{n} \Phi_{\theta_{ij}}(\xi_l) - \frac{1}{n} \sum_{l=1}^{n} \Phi_{\theta_{ij}^*}(\xi_l) \right\| \\
& + \left\| \frac{1}{n} \sum_{l=1}^{n} \Phi_{\theta_{ij}^*}(\xi_l) - \int_{T_4} \Phi_{\theta_{ij}^*}(\xi)\, d\mu(\xi) \right\| \\
\leqslant\quad & 2|\theta_{ij} - \theta_{ij}^*| \frac{1}{n} \sum_{l=1}^{n} \|\xi_l\| + \frac{1}{4} d_{ij} < \frac{d_{ij}}{2}.
\end{aligned}
\tag{5.6}
$$

However, if there were $s > 0$ and $\theta_{ij} \in (0, \pi/2)$ such that

$$\frac{1}{n} \sum_{l=1}^{n} \Phi_{\theta_{ij}}(\xi_l) = s(\cos\theta_{ij}, \sin\theta_{ij}),$$

then by (5.5) we would have

$$
\begin{aligned}
&\left\| \frac{1}{n} \sum_{l=1}^{n} \Phi_{\theta_{ij}}(\xi_l) - \int_{T_4} \Phi_{\theta_{ij}^*}(\xi)\, d\mu(\xi) \right\| \\
\geqslant\quad & \| s(\cos\theta_{ij}, \sin\theta_{ij}) \pm d_{ij}(-\sin\theta_{ij}^*, \cos\theta_{ij}^*) \| \geqslant d_{ij}/2,
\end{aligned}
$$

contradicting (5.6). Thus, it follows from (3.4) that, for $n > n_0$, $\hat{\xi}_n$ does not lie on the open half line determined by any θ_{ij} in $(0, \pi/2)$.

For $\theta_{ij} = 0$ (respectively $\pi/2$), the condition (5.1) implies that the first (respectively the second) coordinate of $\int_{T_4} \Phi_{\theta_{ij}}(\xi)\, d\mu(\xi)$ is non-positive and, since $\Theta_{ij} = \emptyset$, it must be negative. The result of Lemma 4.1 then implies that, when n is sufficiently large, $\hat{\xi}_n$ will not lie the boundary of the (i, j)-quadrant determined by $x_j = 0$ (respectively $x_i = 0$).

(b) If $\Theta_{ij} \cap (0, \pi/2) = [\theta_{ij}^{(1)}, \theta_{ij}^{(2)}]$, then the above argument shows that, for any $\epsilon > 0$, there is an n_0 such that when $n > n_0$ if $\hat{\xi}_n$ is in the (i, j)-quadrant, it must lie in the cone in spanned by the cone point and angles $[\theta_{ij}^{(1)} - \epsilon, \theta_{ij}^{(2)} + \epsilon] \cap [0, \pi/2]$. The arbitrariness of ϵ means that, given it lies in the (i, j)-quadrant, the probability that $\hat{\xi}_n$ lies in the cone C_{ij} tends to one as $n \to \infty$.

On the other hand, the proof of Lemma 5.1 shows that, for all $\theta_{ij} \in \Theta_{ij}$, the random variables $\Phi_{\theta_{ij}}(\xi_1)$ have the same distribution. Thus, as before, given that $\hat{\xi}_n$ is in C_{ij},

$$\hat{\xi}_n = \frac{1}{n} \sum_{l=1}^{n} \Phi_{\theta_{ij,n}}(\xi_l) \stackrel{d}{=} \frac{1}{n} \sum_{l=1}^{n} \Phi_{\theta_{ij}^{(1)}}(\xi_l),$$

where $\tan\theta_{ij,n} = \hat{x}_j^{(n)}/\hat{x}_i^{(n)}$. Moreover, the definition of Θ_{ij} implies that the Euclidean mean of $\Phi_{\theta_{ij}}(\xi_1)$ is at the origin for all $\theta_{ij} \in \Theta_{ij} \cap (0, \pi/2)$. Hence, the classical central limit theorem gives that, for any Borel set $B \subset \mathbb{R}^2$,

$$\lim_{n\to\infty} P\left(\sqrt{n}\hat{\xi}_n 1_{\{\hat{\xi}_n \in \text{ interior } (i,j)\text{-quadrant}\}} \in B\right)$$
$$= \lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\Phi_{\theta_{ij}^{(1)}}(\xi_l) \in B \cap C_{ij}\right).$$

The required result follows by noting that $\dfrac{1}{\sqrt{n}}\sum\limits_{l=1}^{n}\Phi_{\theta_{ij}^{(1)}}(\xi_l)$ tends, in distribution, to the 2-dimensional Gaussian $N(0, V)$, where V is the covariance matrix of $\Phi_{\theta_{ij}^{(1)}}(\xi)$. $\qquad\square$

Note that the above proof shows that, if $\Theta_{ij} \cap (0, \pi/2)$ contains only one single angle θ_{ij}, then $\Phi_{\theta_{ij}}(\xi)$ must be a one-dimensional random variable and the limiting distribution must be one-dimensional Gaussian. The following result is a direct consequence of that of Theorem 5.2(a).

Corollary 5.3. *Assume that the Fréchet mean $\hat{\xi}$ of μ is at the cone point. If $\Theta_{ij} = \emptyset$ for all (i,j)-quadrants of $\boldsymbol{T}_4$, then the sample Fréchet mean $\hat{\xi}_n$ will be at the cone point for all sufficiently large n.*

We note finally that the proof of the central limit theorem when the Fréchet mean of a probability measure on $\boldsymbol{T}_4$ is at the cone point can easily be simplified to obtain similar results for the central limit theorem when the Fréchet mean of a probability measure on $\boldsymbol{Q}_5$ is at the cone point of $\boldsymbol{Q}_5$.

Appendix: The tree space

A tree is a contractible graph, that is, a connected graph with no circuits. An *n-tree* has $n + 1$ vertices of degree 1, one of which is distinguished as its *root*, the others being termed *leaves*. We are interested in labelled trees for which the names, generally $a, b, c, \cdots$, of the leaves matter. However, since all our trees are labelled, we shall generally omit that adjective. The remaining *internal* vertices all have degree at least 3. If both vertices of an edge are internal, then that edge is also called *internal*. A tree in which all internal vertices have degree 3 is called a *binary tree* and such an n-tree has $n - 1$ internal vertices and $n - 2$ internal edges, giving a total of $2n$ vertices and $2n - 1$ edges. Note that it does not matter how the tree is oriented in the plane.

The tree space, $\boldsymbol{T}_n$, is the moduli space of labelled n-trees, the moduli or parameters being the lengths of the internal edges. A labelled n-tree with unspecified edge lengths determines a topological type, modulo its leaves and root, that we shall refer to as an *n-tree-type*, or just *n-type*. Then $\boldsymbol{T}_n$ is a stratified space with a stratum for each n-type, such that an n-type with k internal edges determines a stratum with k positive parameters ranging over the points of an open k-dimensional orthant. The top dimensional strata correspond to the binary tree-types and, if β_n is the number of binary labelled n-tree-types, then any edge, internal or not, of a binary labelled $(n-1)$-tree may be replaced by three edges, an internal vertex and a new leaf n, to give a well-defined binary labelled n-tree:
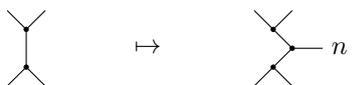

The procedure being reversible, we see that $\beta_n = (2n - 3)\beta_{n-1}$. Thus, $\beta_n = (2n - 3)!! = (2n - 3)(2n - 5)\cdots 5 \cdot 3$, since $\beta_2 = 1$. These $(2n - 3)!!$ top dimensional strata all

have dimension $n - 2$. If we also specify the lengths of the non-internal edges, those with leaves or the root attached, we obtain the 'full tree space', which is the product $\boldsymbol{T}_n \times (\mathbb{R}_+)^n$. In this paper, we shall only consider $\boldsymbol{T}_n$ since a central limit theorem on the full tree space may be derived from the corresponding one on $\boldsymbol{T}_n$.

Such an n-tree-type can be represented as a particular product of the leaves that, assuming commutativity, is determined by an appropriate sequence of associations. Thus,

$$\text{(trees)} \quad = \quad \text{(trees)} \quad \sim \ a(bc) \ = \ (cb)a \ \neq \ (ab)c \ \sim \quad \text{(tree)} \quad \text{etc.}$$
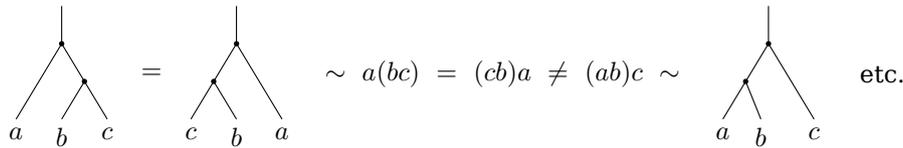
The boundary relation in the stratification of $\boldsymbol{T}_n$ corresponds to a parameter becoming zero. That is equivalent to removing a bracket from the associative pattern and corresponds to the two vertices of an edge in the corresponding n-type coalescing to form a single vertex. For the top stratum, that new vertex will have degree 4. Such a vertex may be resolved in three ways to re-establish a binary tree:

$$\text{(tree diagrams with } X, A, B, C)$$
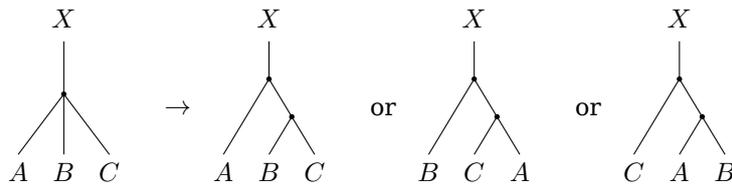
where A, B, C are subtrees, possibly just leaves, and X is a subtree containing the root and any leaves that are not involved in A, B or C. Thus each such stratum of co-dimension one lies on the boundary of three top dimensional strata. On the other hand each top dimensional stratum has $n - 2$, the number of internal edges of the binary n-type, such co-dimension one strata forming its boundary. Thus there are $\frac{n-2}{3}(2n-3)!!$ co-dimension one strata. At the other extreme, each $\boldsymbol{T}_n$ has a single zero dimensional stratum, a vertex representing the tree with no internal edges and a single internal vertex of degree $n + 1$. The one dimensional strata correspond to just one bracket enclosing two or more, but not all, leaves so that there are $2^n - n - 2$ of them. Note that two such strata may belong to the same tree-type if, and only if, the corresponding brackets are not linked: either one includes the other or their contents are disjoint. We shall refer to the top dimensional strata as *cells*, specifically $(n-2)$-cells, and the co-dimension one strata as their *faces*. We shall often refer to the one dimensional strata as *semi-axes*, since that is indeed what they represent in $\boldsymbol{T}_n$. Note that $\boldsymbol{T}_3$ is special in that the 1-cells are the three one-dimensional strata and the only face is the unique zero-dimensional stratum.

An alternative notation for a tree-type arises from the observation that an internal edge partitions the full set of leaves, including the root, into two subsets and a tree-type may be specified by the set of such partitions, or *splits*, that correspond to its internal edges. This notation is more apposite when one is not distinguishing the root. The correlation is that the bracket that determines a semi-axis in the former notation contains the members of the subset that does not contain the root in the split notation.

Following [4], we introduce a metric d_T on $\boldsymbol{T}_n$ by identifying each $(n-2)$-cell with the principal orthant in $\mathbb{R}^{n-2}$ with the Euclidean metric, and defining the length of any rectifiable curve in $\boldsymbol{T}_n$ to be the sum of the lengths of the segments into which it is broken by the faces of the cells. Since each orthant is a cone with vertex at the origin,
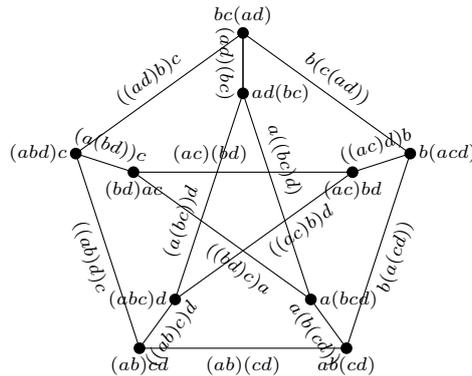
Figure 4: The structure of T_4 represented on the Peterson graph

this makes T_n into an open infinite metric cone with its vertex at the origin which, to avoid confusion with vertices in trees, we shall refer to as the *cone point*. Given trees t_1, t_2, there is always a path from t_1 to t_2 comprising the ray from t_1 to the cone point followed by the ray from the cone point to t_2. In a good number of cases – 40% in T_4 – this will be the geodesic from t_1 to t_2. Obviously, however, this is not the geodesic when t_1 and t_2 lie in the same cell. Moreover, if they lie in $(n-2)$-cells σ_1 and σ_2 that share a common face f_{12}, then the union of these cells may be isometrically embedded in $\mathbb{R}^{n-2}$ in the obvious way and the geodesic in T_n will correspond to the straight line in $\mathbb{R}^{n-2}$. Similar considerations give rise to further examples in T_n.

When we restrict attention to T_4, just one more feature occurs as follows. In T_4 the 2-cells are now plane quadrants and their faces are the two semi-axes. Given a chain of three quadrants with consecutive quadrants sharing a semi-axis, their union may be embedded in $\mathbb{R}^2$ by omitting, say, its $+-$ quadrant. If the geodesic starts in the interior of the first quadrant, then it can be realised as a straight linear segment to some points in the third quadrant, but must pass through the cone point for the rest. Any such sequence of three quadrants lies in a unique cycle of five quadrants in T_4. Thus, to describe the geodesics in T_4 in general, it will be convenient to embed such 5-cycles in $\mathbb{R}^3$ as the space Q_5.

We now describe these 5-cycles, and how they all fit together in T_4. A quadrant in T_4 will correspond to a 4-type with bracket notation of the form $((ab)c)d$ or $(ab)(cd)$. The quadrant $((ab)c)d$ has the two semi-axes $(ab)cd$ and $(abc)d$. This first semi-axis is shared by the quadrants $((ab)d)c$ and $(ab)(cd)$, while the second is shared by $(a(bc))d$ and $((ac)b)d$. Choosing the sequence $((ab)d)c$, $((ab)c)d$, $(a(bc))d$, the 'free' semi-axes at each end, which are not part of the quadrant $((ab)c)d$, are $(abd)c$ and $(bc)ad$. The only edge with which these are both compatible is $(ad)bc$, giving the other two quadrants in the 5-cycle. The same is true for any sequence of three contiguous quadrants in T_4: there is a unique choice of two further quadrants to give a cycle of five quadrants in which cyclically consecutive members share a semi-axis. However, there are two independent choices for each of the quadrants cobounding, i.e. sharing a semi-axis with, the central of the three initial quadrants. This implies that each quadrant will lie in four such 5-cycles. It is convenient to visualise and describe these details of the structure of T_4 via the Peterson graph which is the link $\bar{T}_4$ of its cone point, as illustrated in Figure 4. The labelling given there is not unique, but all possible labellings are equivalent.

References

[1] Basrak, B.: Limit theorems for the inductive mean on metric trees. *J. Appl. Prob.* **47** (2010), 1136–1149. MR-2752884

[2] Bhattacharya, R. and Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds. *Ann. Statist.* **31** (2003), 1–29. MR-1962498

[3] Bhattacharya, R. and Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds-II. *Ann. Statist.* **33** (2005), 1225–1259. MR-2195634

[4] Billera, L. J., Holmes, S. P. and Vogtmann, K.: Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** (2001), 733–767. MR-1867931

[5] Bridson, M. and Haefliger, A.: Metric spaces of non-positive curvature. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 319. *Springer-Verlag, Berlin,* 1999. xxii+643 pp. ISBN: 3-540-64324-9 MR-1744486

[6] Dryden, I. L. and Mardia, K. V.: Statistical shape analysis. Wiley Series in Probability and Statistics: Probability and Statistics. *John Wiley & Sons, Ltd., Chichester,* 1998. xx+347 pp. ISBN: 0-471-95816-6 MR-1646114

[7] Felsenstein, J.: Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39** (1985), 783–791.

[8] Hendriks, H. and Landsman, Z.: Asymptotic behaviour of sample mean location for manifolds. *Statist. Probab. Lett.* **26** (1996), 169–178. MR-1659156

[9] Hendriks, H. and Landsman, Z.: Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *J. Multivariate Anal.* **67** (1998), 227–243. MR-1659156

[10] Holmes, S.: Bootstrapping trees: theory and methods. *Statist. Sci.* **18** (2003), 241–255. MR-2026083

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

[12] Holmes, S.: Statistical approach to tests involving phylogenies. In: Mathematics of Evolution and Phylogeny, O. Gascuel Editor. *Oxford University Press*, USA, 2007.

[13] Hotz, T., Huckemann, S., Le, H., Marron, J. S., Mattingly, J. C., Miller, E., Nolen, J., Owen, M., Patrangenaru, V. and Skwerer, S.: Sticky central limit theorems on open books. *Ann. Appl. Probab.* to appear, (2012). arXiv:1202.4267.

[14] Goresky, M. and MacPherson, R.: Stratified Morse theory. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 14. *Springer-Verlag, Berlin,* 1988. xiv+272 pp. ISBN: 3-540-17300-5 MR-0932724

[15] Kendall, D. G., Barden, D., Carne, T. K. and Le, H.: Shape and shape theory. Wiley Series in Probability and Statistics. *John Wiley & Sons, Ltd., Chichester,* 1999. xii+306 pp. ISBN: 0-471-96823-4 MR-1891212

[16] Kendall, W. S. and Le, H.: Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Braz. J. Probab. Stat.* **25** (2011), 323–352. MR-2832889

[17] Miller, E., Owen, M. and Provan, S.: Averaging metric phylogenetic trees. In preparation, (2012).

[18] Owen, M.: Computing geodesic distances in tree space. *SIAM J. Discrete Math.* **25** (2011), 1506–1529. MR-2873201

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

[20] Sturm, K. T.: Probability measures on metric spaces of nonpositive curvature. Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002), 357–390, Contemp. Math., 338, *Amer. Math. Soc., Providence, RI,* 2003. MR-2039961

[21] Vogtmann, K. (2007) Geodesics in the space of trees, (2007), `http://www.math.cornell.edu/~vogtmann/papers/TreeGeodesicss/index.html`

[22] H. Ziezold, H.: On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974),* Vol. A, pp. 591–602. *Reidel, Dordrecht,* 1977. MR–0501230

[23] Z. Yang, Z. and Rannala, B.: (1997) Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Molecular Biology and Evolution* **14** (1997), 717-724.