CHAPTER

Statistics on Stratified Spaces 1

Aasa Feragen[*] **and Tom Nye**[**]

[*]*University of Copenhagen, Department of Computer Science, Universitetsparken 5, 2100 Copenhagen, Denmark* [**]*Newcastle University, School of Mathematics, Statistics and Physics, Newcastle upon Tyne, NE1 7RU, UK*

CHAPTER OUTLINE HEAD

ABSTRACT

While manifold statistics is an established tool for computational anatomy, a number of structures are not modelled well on manifolds. Examples include data with variable topological structure, such as trees and graphs, as well as objects that should be invariant with respect to groups that do not act freely on the space of measurements. Such data can often be represented more faithfully as residing on a *stratified space*, which consists of multiple manifold components, potentially with different dimensions, joined together in a controlled fashion. In this chapter we give a brief introduction to stratified spaces and geometric tools that are useful for performing statistics in them. We review existing least squares models in stratified spaces, along with some unexpected behavior that they exhibit, illustrated in simple stratified spaces. Next, we review two particular examples of stratified spaces given by two different tree-spaces. The first is the Billera-Holmes-Vogtmann space of phylogenetic trees, for which a number of statistical algorithms have been proposed. The second is the space of unlabelled trees, which models the more general attributed trees found in computational anatomy. This space has a more complicated geometry, and not much is known about its structure and statistics. We present a novel result connecting the two tree-spaces and their geodesics, along with a consequential theorem on uniqueness of geodesics. Finally, we discuss other, less studied applications of stratified spaces as statistical domains, including spaces of graphs, point sets and sequences, as well as quotient spaces.

Keywords: Stratified space statistics, tree-spaces, quotient spaces, variable topology

1.1 INTRODUCTION TO STRATIFIED GEOMETRY

The majority of statistical methodology is built on the premise that the data being analysed lie in a finite dimensional vector space equipped with the Euclidean L^2 inner product. As seen in the other chapters of this book, there are important applications for which data in fact lie in a smooth manifold, and for which work must be done to extend existing "linear" methodology to this new context. Instead, in this chapter, we consider a different class of data spaces for which the structure of a smooth manifold is not available everywhere, and for which the dimension of the space can vary from point to point. These *stratified spaces* have attracted interest from researchers in recent years, and examples of data lying in stratified spaces include trees [11, 25, 19], graphs [34], point sets such as persistence diagrams [56], objects invariant to a nontrivial group action (lying in a group quotient space) [43, 38] and positive semidefinite matrices [29, 55].

We will not give a general formal definition of a stratified space – more details can be found, for instance, in [51] – but instead illustrate the properties of such spaces and associated data analysis largely via examples. Indeed, existing statistical methods in stratified spaces generally make no use at all of the formal definition of a stratified space. This section describes three key toy examples before going on to survey definitions and results from metric geometry which are necessary to understand the geometry of stratified spaces more generally. Although simple, the toy examples are highly illustrative of the unusual properties of data analysis in stratified spaces.

Properties of least squares estimators are considered in more detail in Section 1.2, in particular by considering the examples introduced in this section. In Sections 1.3 and 1.4 we describe the geometry of two related stratified spaces: the space of evolutionary trees with leaves labelled by a fixed set of species, and a space of unlabelled trees. The final section goes beyond trees to illustrate how other types of data may also be modelled as residing in a stratified space. Examples include graphs, point sets, sequences and data invariant under nontrivial group actions. Some of these constructions are well known, while others are new, coming with associated open problems.

1.1.1 EXAMPLES

Statistical models and estimators on stratified spaces can display strikingly different behaviour than intuition suggests from working on linear Euclidean spaces. Many of these properties arise with the following fundamental simple examples.

Example 1.1 (Spiders)**.** *The k-spider consists of k copies of the positive real line $\mathbb{R}_{\geq 0}$ glued together at the origin. The metric is the Eucidean metric on each "leg" of the spider, extending in the obvious way to the whole space: given two points x, y on different legs, $d(x, y) = d(x, 0) + d(0, y)$. We use the notation $Spider_k$ to denote the k-spider. It is clear that for $k > 2$, the k-spider does not have the structure of a topological manifold: no chart can be defined at the origin. The set of tangent directions at the origin is not a vector space, but is in fact a copy of the space itself.*

Most of the examples we consider, like the k-spider, are formed by gluing pieces of Euclidean space or other manifolds together along their boundaries. We will be deliberately informal about the operation of gluing two topological spaces X_1, X_2, since the geometry of the resulting space does not depend on the technical details for all the examples we consider. However, formally we mean that two subsets of the spaces X_1 and X_2, respectively, are identified by a bijection (often an isometry when there are underlying metrics) and we then form the quotient of $X_1 \cup X_2$ where two points are equivalent if and only if they are identified under the bijection.

Example 1.2 (Open books)**.** *The open book of dimension $n + 1$ on k pages is $Book^n_k = \mathbb{R}^n \times Spider_k$. The spine of the book is the subset $\mathbb{R}^n \times \{0\}$, and each page of the book is a subset $\mathbb{R}^n \times \mathbb{R}_{\geq 0}$. The metric is the product metric, and so is just the Euclidean metric on each page $\mathbb{R}^n \times \mathbb{R}_{\geq 0}$. The simplest open book has $n = 1$ and $k = 3$, and consists of three half-planes joined along their shared edge. Open books are stratified in the following way: the spine is a n-dimensional manifold which forms the boundary of the k pages, each of which is a $n + 1$ dimensional manifold with boundary. Each piece in this decomposition is a* stratum *and the space consists of several strata glued together along lower dimensional sub-strata.*

The 3-spider parametrizes a certain set of trees and so forms a *tree space*. Consider the set of rooted trees with 3 leaves, which are labelled bijectively with the set $S = \{A, B, C\}$. There are three possible binary tree topologies, and one "star" tree

with no internal edges, shown in Figure 1.1. If we further assume that internal edges are *weighted*, that is, assigned weights or lengths in $\mathbb{R}_{\geq 0}$, then Spider_3 parametrizes the corresponding set of tree objects: each "leg" of the spider corresponds to a different binary tree, and the position along the leg determines the weight assigned to the single internal edge of the tree. If *all* the edges are weighted, then the space is $\mathbb{R}^3_{\geq 0} \times \text{Spider}_3 \subseteq \text{Book}_3^3$, and so open books parametrize certain sets of trees.

Example 1.3 (Cones)**.** *Let $\text{Cone}_{k\pi/2}$ denote the space formed by gluing together k copies of the positive quadrant $\mathbb{R}^2_{\geq 0}$ to form a cone, so that the origin is a common point in each quadrant and forms the point of the cone. An embedding of $\text{Cone}_{5\pi/2}$ in $\mathbb{R}^3$ is shown in Figure 1.2. On each quadrant of $\text{Cone}_{k\pi/2}$, the metric is Euclidean, and the distance between points in different quadrants is the length of the shortest path between them, where the path consists of straight line segments in each quadrant. In contrast to spiders and open books, each cone $\text{Cone}_{k\pi/2}$ has the structure of a topological manifold. As a stratified space, $\text{Cone}_{k\pi/2}$ can be thought of as containing two strata: the origin, and the complement of the origin. (An alternative stratification of $\text{Cone}_{5\pi/2}$ is by tree topology.) As we will see, a notion of curvature can be defined for certain metric spaces. For $k = 1, 2, 3$, $\text{Cone}_{k\pi/2}$ is non-negatively curved, while for $k > 4$ it is non-positively curved.*

The cone $\text{Cone}_{5\pi/2}$ parametrizes a certain set of trees, as shown in Figure 1.2. This is a subspace of the space of all leaf-labelled weighted rooted trees on 4 leaves. More details are given in Section 1.3, which describes *Billera-Holmes-Vogtmann* tree

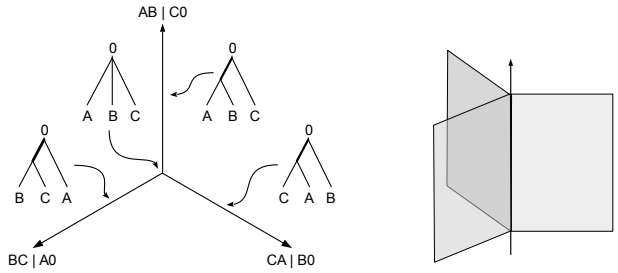

Figure 1.1: The space Spider_3 (left) consists of three copies of the positive real line joined together at the origin. It parametrizes the set of rooted trees with leaves A, B, C such that the internal edge has a positive weight or length. The position along the axis labelled with the bi-partition $AB|C0$, for example, determines the length of the highlighted edge on the corresponding tree. The origin corresponds to a tree obtained by contracting the internal edge to length zero. The open book Book_3^1 is shown on the right.

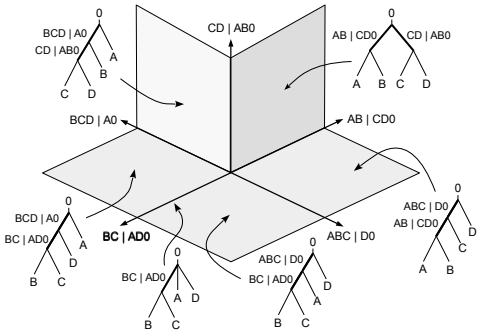

Figure 1.2: An embedding of $\text{Cone}_{5\pi/2}$ in $\mathbb{R}^3$, annotated to show how the cone parametrizes a certain set of trees with 4 leaves. The space consists of 5 quadrants glued along their edges, where each quadrant corresponds to a different binary tree. Each binary tree has two weighted internal edges, and the weights determine the position within the quadrant. The trees contain 5 different internal edges, labelled as the corresponding axes in $\mathbb{R}^3$. The quadrant boundaries correspond to trees where an edge has been contracted to have zero weight, as shown for the axis $BC|AD0$.

space [11], the space of edge-weighted trees on a fixed set of N labelled leaves. This stratified space has received the most attention to date in terms of the development of statistical methods, due to its importance in evolutionary biology and its attractive geometric properties.

All three above examples are metric spaces which fail to have the structure of a Riemannian manifold: spiders and open books are not topologial manifolds, while the metric on any cone is singular at the origin. How, then, can we analyse data in these spaces, or in more general stratified spaces? To answer this question we recall various definitions and results from metric geometry. The following overview gives the essential background geometry but a much fuller account is given in [12].

1.1.2 METRIC SPACES

Suppose X is the space in which we want to develop our statistical methodology, and let d be a metric on X. It is easily seen that an arbitrary metric d does not itself give enough structure on X in order to develop any useful statistics. For example, the metric defined by $d(x, y) = 1$ for all $x \neq y$ and $d(x, x) = 0$ only tells us whether two data points are the same or not, and so cannot be used to calculate any useful summary statistics. More structure on the space X is required, and so we consider paths in X and their associated lengths.

Definition 1.4 (Geodesics)**.** *A* geodesic *[12] in a metric space (X, d) is defined as a path $\gamma\colon [0,1] \to X$ such that for any $t, t' \in [0,1]$, we have $d(\gamma(t), \gamma(t')) = |t - t'| \cdot d(\gamma(0), \gamma(1))$. The image of a geodesic γ is called a* geodesic segment *in X.*

A path $\gamma\colon [0,1] \to X$ is *locally geodesic* if there exists $\varepsilon > 0$ such that $d(\gamma(t), \gamma(t')) = |t - t'| \cdot d(\gamma(0), \gamma(1))$ holds whenever $|t - t'| < \varepsilon$.

Definition 1.5 (Geodesic spaces)**.** *(X, d) is called a* geodesic metric space *if there is at least one geodesic path between every pair of points in X. It is* uniquely geodesic *if there is exactly one geodesic between every pair of points.*

The existence of geodesics is really fundamental to the development of statistics on a metric space X, just as in the case of Riemannian manifolds. However, it is also useful to assign lengths to arbitrary paths in X.

Definition 1.6 (Path length)**.** *If $c : [0,1] \to X$ is a path in X then the* length *of c is*

$$\ell(c) = \sup_{a=t_0 \le t_1 \le \cdots \le t_n = b} \sum_{i=0}^{n-1} d(c(t_i), c(t_{i+1}))$$

where the supremum is taken over all possible n and partitions of the interval $[0,1]$. The length of c is taken to be infinite when this expression is unbounded.

The triangle inequality implies that $\ell(c) \ge d(c(a), c(b))$ for any path c. It follows from the definition of a geodesic γ on X that $\ell(\gamma) = d(\gamma(0), \gamma(1))$. Thus, a geodesic is a shortest path connecting its endpoints. Conversely, a shortest path can always be parametrized as a geodesic. Many spaces have pairs of points with no shortest connecting path: for example take $\mathbb{R}^2$ equipped with the Euclidean metric but with the origin removed. A point $\boldsymbol{x}$ cannot be joined to the antipodal point $-\boldsymbol{x}$ by a path of length $2\|\boldsymbol{x}\|$. A metric space (X, d) is called a *length space* if $d(x, y)$ is the infimum of lengths of paths connecting x, y for all $x, y \in X$. The Hopf-Rinow theorem states that any complete, locally compact length space (X, d) is a geodesic metric space. The example of $\mathbb{R}^2$ without the origin is a length space, but it fails the conditions of the Hopf-Rinow theorem as it is not complete.

1.1.3 CURVATURE IN METRIC SPACES

We next turn attention to the idea of curvature in a geodesic metric space. The idea is to look at whether triangles are "fat" or "thin" compared to triangles in Euclidean space. We will denote by $\Gamma(x, y)$ a choice of geodesic segment between $x, y \in X$. (Of course if X is uniquely geodesic then there is exactly one choice of segment.) Given $p, q, r \in X$ a geodesic triangle $\Delta(p, q, r) \subseteq X$ is a choice of geodesic segments $\Gamma(p, q), \Gamma(q, r), \Gamma(r, p)$. A corresponding flat Euclidean triangle is required in order to draw comparisons. A triangle $\Delta' = \Delta(p', q', r')$ in $\mathbb{R}^2$ is a comparison triangle if

$$d(p, q) = d(p', q'), \quad d(q, r) = d(q', r') \quad \text{and} \quad d(r, p) = d(r', p').$$

Such a triangle always exists in $\mathbb{R}^2$ (by applying the triangle inequality to Δ in X) and is unique up to isometries of $\mathbb{R}^2$. Given $x \in X$ on $\Gamma(p, q)$, a comparison point x' in Δ' is a point on $\Gamma(p', q')$ such that

$$d(x, p) = d(x', p') \quad \text{and} \quad d(x, q) = d(x', q').$$

We call (x, x') a *comparison pair* for the edge $\Gamma(p, q)$. This is illustrated in Figure 1.3.

Geodesic metric spaces with non-positive curvature play a very prominent role in the theory of statistics on stratified spaces. In order to have non-positive curvature, every geodesic triangle must be at least as "thin" as its Euclidean comparison triangle. This is made rigorous via the following definition of the so-called CAT(0) inequality.

Definition 1.7. *A geodesic triangle $\Delta(p, q, r)$ satisfies the* CAT(0) inequality *if $d(x, r) \leq d(x', r')$ for all comparison pairs (x, x') with $x \in \Gamma(p, q)$, $x' \in \Gamma(p', q')$, and similarly for all comparison pairs on the other two edges. The geodesic metric space X is a* CAT(0) space *if every geodesic triangle satisfies the CAT(0) inequality.*

CAT(0) spaces, and more generally spaces which are locally CAT(0), are often called non-positively curved spaces. They have a rich geometry, analogous to geometry on Riemannian manifolds, which lends them as very suitable spaces on which to develop statistical methods. There is an analogous definition of a CAT(κ) space for $\kappa \neq 0$, and these spaces can be thought of as having curvature $\leq \kappa$. Here, comparison triangles are constructed not in the plane, but in a model space M_κ. For $\kappa < 0$, M_κ is a scaled version of the hyperbolic plane; for $\kappa > 0$, M_κ is a scaled version of the sphere S^2.

The name CAT(κ) comes from the concatenated initials of Cartan, Alexandrov and Topogonov, all pioneers in defining and understanding the notion of curvature for metric spaces [12]. In contrast, there is a definition of non-negatively curved geodesic spaces, due to Alexandrov [2]: X is non-negatively curved if every geodesic triangle in the space is at least as "fat" as a comparison triangle in $\mathbb{R}^2$.

Referring back to our fundamental examples, it is straightforward to check that every triangle in a k-spider satisfies the CAT(0) inequality. Any product of two

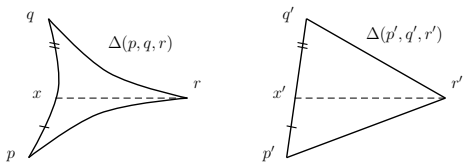

Figure 1.3: Comparing triangles in a geodesic metric space (left) and $\mathbb{R}^2$ (right).

CAT(0) spaces is CAT(0), so since $\mathbb{R}^n$ is CAT(0) is follows that the open book Book_k^n is also CAT(0). More generally, a metric tree (X, d) is a tree where each edge has an isometry to an interval in $\mathbb{R}$, and such spaces are also CAT(0). On the other hand, when $k \leq 3$ the cone $\text{Cone}_{k\pi/2}$ is non-negatively curved. Triangles in $\text{Cone}_{k\pi/2}$ which do not contain the origin in the interior are easily seen to be Euclidean triangles. However, triangles which wind around the origin have interior angles which add up to $> \pi$ and are "fatter" than Euclidean triangles. The origin is repulsive: the only geodesics which pass through the origin have an end point at the origin. We will consider cones with $k \geq 5$ later.

CAT(0) spaces have many appealing properties which help the development of statistical methods within the spaces. First, they are uniquely geodesic, so every pair of points in a CAT(0) space is joined by a unique geodesic. The geodesic segment $\Gamma(x, y)$ between x and y varies continuously as a function of x, y. Moreover, any path which is locally geodesic is in fact a geodesic path.

In addition to these attractive properties of geodesics, there is a notion of projection onto closed sets in CAT(0) spaces. If X is a CAT(0) space then a function $f : X \to \mathbb{R}$ is convex if for any geodesic path $\gamma : I \to X$ parametrized proportional to length, the function $I \to \mathbb{R}$ defined by $t \mapsto f(\gamma(t))$ is convex. Given any $x \in X$, it can be shown that the distance function $d(\cdot, x) : X \to \mathbb{R}$ is convex. Similarly the function $d(\cdot, \cdot)$ is convex on the product space $X \times X$. Now suppose that $A \subseteq X$ is convex and complete in the induced metric. (A subset $A \subseteq X$ is convex if $\Gamma(x, y) \subseteq A$ for all $x, y \in A$.) Then given any $x \in X$ there is a unique point $\pi(x) \in A$ closest to x:

$$\pi(x) = \text{argmin}_{a \in A} d(a, x).$$

This is called the projection of x onto A. If A is closed but not convex, then a closest point in A to x exists, but it is not necessarily unique.

Cubical complexes are a rich source of examples of CAT(0) spaces, and are defined in the following way. Let $I^n \subset \mathbb{R}^n$ be the unit cube $[0, 1]^n$ equipped with the Euclidean metric. The codimension-k faces of I^n correspond to fixing k coordinates on I^n to be either 0 or 1. A cubical complex is a metric space obtained by gluing together cubes (potentially of different dimensions) along their faces: a dimension-k face in one cube can be glued isometrically to one or more dimension-k faces in other cubes. Cubical complexes are thus analogous to simplicial complexes, but each cell is a unit cube rather than a simplex. A cubical complex X can be given a metric as follows. On each cube, the metric is the Euclidean metric. More generally, the distance between $x, y \in X$ is defined to be the infimum of the lengths of paths between x and y which are straight line segments within each cube. When X is locally compact then it is a geodesic metric space by the Hopf-Rinow theorem. Several spaces of trees and networks [11, 28, 16] are examples of cubical complexes, although the space of networks in [16] is not CAT(0).

Gromov gave a combinatorial condition that specifies when a cubical complex is CAT(0) [30]. The condition is defined in terms of the *link* of each vertex in the complex. The link of a vertex v is the set $\{x \in X : d(x, v) = \varepsilon\}$ where $0 < \varepsilon < 1$ is

a fixed constant. The link of v can be regarded as an abstract simplicial complex, and Gromov's condition is expressed purely in terms of the combinatorics of this object. Rather than state the condition precisely, we will illustrate it using the example $\text{Cone}_{k\pi/2}$ for $k = 3$ and $k = 5$. The cones $\text{Cone}_{k\pi/2}$ can be constructed as cubical complexes by filling each quadrant with an infinite array of 2-cubes (unit squares). For $k = 3$ the link of the origin consists of three quarter-circular arcs forming a loop. In order to be CAT(0), Gromov's condition states that the link must contain the 2-simplex bounded by this loop – but it does not, so the condition fails and $\text{Cone}_{3\pi/2}$ is not CAT(0). On the other hand, when $k = 5$, the link of the origin consists of a loop formed from 5 quarter-circular arcs. Gromov's condition states that any simplex whose 1-dimensional faces (quarter-circular arcs) are in the link, must itself be in the link. Since the loop consists of 5 arcs rather than 3, it does not bound any 2-simplex, and so Gromov's condition is satisfied for the origin. It also holds for the other vertices in the cubical complex and so $\text{Cone}_{5\pi/2}$ is CAT(0). We will consider Gromov's condition again when describing evolutionary tree space, but we next turn attention to least squares estimators.

1.2 LEAST SQUARES MODELS

In Euclidean space, standard statistical methods such as computation of sample means, linear regression and principal component analysis, can be formulated as problems which minimize a least squares modelling error. Least squares errors generalize easily to metric spaces, and have therefore been popular for building statistical models both on manifolds and metric spaces. In stratified spaces, however, least squares statistics have surprising and potentially unwanted properties [32, 24].

1.2.1 LEAST SQUARES STATISTICS AND STICKINESS

In this section suppose X is a geodesic metric space. Given a finite dataset $\{x_1, \ldots, x_n\} \subset X$, its Fréchet mean is defined as the point minimizing the sum of squared distances to the data points [36]:

$$\bar{x} = \text{argmin}_{x \in X} \sum_{i=1}^{n} d(x_i, x)^2. \tag{1.8}$$

However, in stratified spaces, Fréchet means can be *sticky* [32]:

Definition 1.9 (Stickiness of Fréchet mean)**.** *The Fréchet mean $\bar{x}$ of a finite sample $\{x_1, \ldots, x_n\} \subset X$ is* sticky *if any sufficiently small perturbations $\{x'_1, \ldots, x'_n\}$ of the sample also have mean $\bar{x} = \bar{x}'$.*

An example [32] of a sticky Fréchet mean can be found on the 3-spider, as illustrated in Figure 1.4. Three unit point masses are positioned on the 3-spider, one on each leg of the spider and unit distance from the origin. These are shown as black

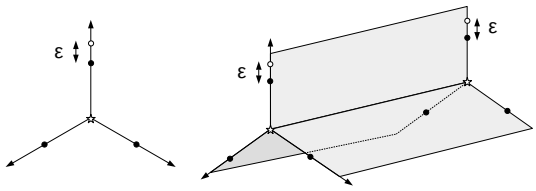

Figure 1.4: **Left:** The Fréchet mean (star-shaped point) of the dataset consisting of the black points on Spider_3 is sticky. **Right:** On the open book, the first principal component (the line connecting the two star-shaped endpoints) of the dataset consisting of the black points sticks to the spine of the book.

dots in the figure. When one of the point masses is moved by distance ε away from the origin, the Fréchet mean (shown by a star on the figure) remains at the origin until $\varepsilon = 1$, at which point it moves on to the upper leg in Figure 1.4 for $\varepsilon > 1$. In fact the mean remains at the origin for all sufficiently small perturbations of the point masses. For more general stratified spaces, stickiness implies that the Fréchet means of sampled data tend to be located at lower-dimensional strata where three or more strata are joined, just as for the 3-spider. In the case of tree spaces, such lower-dimensional strata correspond to trees where at least one node has degree ≥ 4.

A natural extension of the Fréchet mean is the first principal component. This can be defined provided there is a notion of projection onto geodesics or, more generally, onto closed sets. It is denoted $PC1$ and is defined [44, 46, 21] as the geodesic segment $\gamma_{a_0 b_0}$ minimizing the sum of squared residual distances $E(a, b)$:

$$a_0, b_0 = \text{argmin}_{a,b \in X} E(a, b), \quad \text{where} \quad E(a, b) = \sum_{i=1}^{n} d(x_i, \text{pr}_{\gamma_{ab}}(x_i))^2$$

and $\text{pr}_{\gamma_{ab}}(x_i)$ denotes the projection of x_i onto γ_{ab}. This definition is analogous to the definition of first principal component on manifolds due to Huckemann et al. [33], except for the restriction to geodesic segments, which is due to the problem of parametrizing geodesic rays in X [44, 46, 21].

Just like Fréchet means, first principal components can also be sticky:

Definition 1.10 (Stickiness of PC1)**.** *The first principal component $PC1$ for a finite sample in X* sticks *to a subset $S \subset X$ if the first principal component of any sufficiently small perturbation $\{x'_1, \ldots, x'_n\}$ of the sample also lies in S.*

An example of stickiness for $PC1$ on the open book $\mathbb{R} \times \text{Spider}_3$ is given in Figure 1.4, and it is a straightforward extension of the sticky mean example on the 3-spider. Here, two data points are positioned on each sheet of the book, several units apart parallel to the spine. It is clear that $PC1 \subset S$ where S is the spine of the open book, and the same in fact holds for all small perturbations of the sample.

Stickiness of $PC1$ indicates that, just as for Fréchet means, first principal components in stratified spaces will have a tendency to be contained in lower-dimensional strata, even when the data are contained in top-dimensional strata. This creates difficulties in building more advanced statistics: for instance, it is not clear how parallel transport along such principal components might be defined, which has consequences for extending techniques from manifold statistics to the open problem of defining second principal components.

1.2.2 THE PRINCIPAL COMPONENT AND THE MEAN

Stickiness is not the only surprising property of least squares statistics in stratified spaces. In Figure 1.5 we give an example on the open book, where the Fréchet mean does *not* lie on the first principal component. In the figure, the Fréchet mean lies on the spine In Euclidean space, the Fréchet mean always lies on the first principal component; on curved manifolds, this is known not to be the case [33]. This has consequences for the definition and interpretation of the fraction of variance captured by a principal component, which is frequently used to measure the success of dimensionality reduction via PCA in Euclidean space [46]. The definition of the fraction of variance relies on the Pythagorean theorem, and this breaks down in almost any non-Euclidean space.

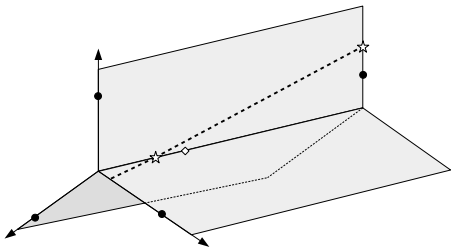

Figure 1.5: Let $\{x_1, x_2, x_3, x_4\}$ consist of the circular points in the figure. The Fréchet mean $\bar{x}$ is the open diamond, which sits on the spine. However, calculations show that $PC1$ is the dotted line segment connecting the two star-shaped points. It extends from the page of the book above the spine onto either of the lower two pages, and is therefore not unique. The dotted line can be shown to give a lower least squares error than the geodesic running along the spine.

1.3 BHV TREE-SPACE

While stratified spaces naturally appear in many applications, among the most investigated so far are spaces of *trees*. Tree spaces can be defined in different ways, leading to different models and geometries, and this chapter will visit two tree spaces in detail. The first, *BHV tree space* [11], assumes that all trees have the same, fixed set of labeled leaves. This is a strong modelling assumption, which makes sense for the study of evolutionary trees, where it was first defined – but which might be overly restrictive in other applications. The modelling cost does, however, come with strong computational advantages. The second tree space construction, which does *not* assume a fixed set of labeled leaves, will be discussed in Section 1.4.

Phylogenetic trees represent evolutionary relationships between a chosen set of biological species. The leaves of a phylogenetic tree, or *phylogeny*, represent present day species, while internal vertices represent speciation events when a population has differentiated into distinct subspecies. The edges in each tree are typically assigned a weight in $\mathbb{R}_{\geq 0}$, which represents the degree of evolutionary divergence along each edge. Trees can be either rooted or unrooted and we will describe the space of phylogenetic trees in both cases. Phylogenetic trees are usually estimated from incomplete noisy data (often genetic sequence data in present-day organisms) and so it is natural to study distributions on the space of all possible phylogenies relating a fixed set of species. A geometry for this space was first described by Billera, Holmes and Vogtmann [11], and the corresponding geodesic metric space has become known as BHV tree space. BHV tree space has been used to analyse sets of anatomical trees [21, 52], as well as evolutionary trees. This section explains the geometry of BHV tree space and reviews existing methods for analysing sets of phylogenetic trees via BHV geometry.

1.3.1 GEOMETRY IN BHV TREE SPACE

Definition 1.11 (Unrooted phylogenetic tree)**.** *Suppose $S = \{1, \ldots, N\}$ is a fixed set of labels. (We sometimes let S be any set with N elements.) An* unrooted phylogenetic tree *on S is an unrooted tree with N leaves which satisfies the following conditions.*

1. *The leaves are bijectively labelled with the elements of S.*
2. *The edges are weighted by values in $\mathbb{R}_{\geq 0}$.*
3. *There are no vertices with degree 2.*

Unrooted trees are important in evolutionary biology since it can be difficult to identify the position of the root, which represents a distant ancestor, with any certainty. Phylogenetic trees are often represented graphically with edge lengths drawn in proportion to their weights, and so edge weights are often referred to as *lengths*. Any unrooted phylogeny on S contains at most $2N - 3$ edges, with the upper bound being attained when every non-leaf vertex has degree 3. Such trees are called *resolved* or *bifurcating* trees. However, trees can contain $< 2N - 3$ edges, in which case one or more vertices has degree > 3. These trees are called *unresolved*. The

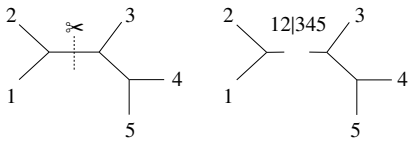

Figure 1.6: Cutting an edge on a leaf-labelled tree (left) creates a bipartition, or split, of the labels (right). The terms *edge* and *split* are therefore used interchangeably.

edges containing the leaves are called *pendant edges*. Conversely, *internal edges* and *internal vertices* are, respectively, edges which do not end in a leaf, and vertices with degree ≥ 3. The collection of all unrooted phylogenies on S is denoted $\mathcal{U}_N$.

A *split* is a bipartition of S, or in other words, a decomposition of S as a union of two disjoint subsets $S = A \cup A^c$. Splits are often written using the notation 12|345 to represent $\{1,2,3,4,5\} = \{1,2,3\} \cup \{4,5\}$ for example. Given a tree $x \in \mathcal{U}_N$ and an edge e in x, cutting e disconnects x and yields a bipartition of the leaves, and so every edge is associated with a unique split. In fact, on labelled trees, edges and splits are entirely equivalent, as shown in Figure 1.6, and so we will use the terms interchangeably in this section. The splits represented by a tree are called its *topology*. Equivalently, the topology of $x \in \mathcal{U}_N$ can be thought of as an unweighted *combinatorial tree*. Two splits $A|A^c$ and $B|B^c$ are called *compatible* if there exists at least one tree containing both splits. Examples of incompatible splits are easy to construct: 12|345 is not compatible with 13|245 since the corresponding edges cannot coexist in any tree. It can be shown that $A|A^c$ and $B|B^c$ are compatible if one of the sets $A \cap B$, $A \cap B^c$, $A^c \cap B$, $A^c \cap B^c$ is empty. Arbitrary sets of splits do not generally correspond to tree topologies due to incompatibility. The following theorem, due to Buneman [14], characterizes tree topologies.

Theorem 1.12 (Splits-Equivalence Theorem)**.** *Any set of pairwise compatible splits which contains the splits* $\{1\}|\{1\}^c, \ldots, \{N\}|\{N\}^c$ *determines an unweighted tree on* S.

To describe BHV tree space we first consider the pendant edges. These are present in all trees and so unrooted tree space can be written as a product

$$\mathcal{U}_N = \mathbb{R}^N_{\geq 0} \times \mathrm{BHV}_N$$

where BHV_N parametrizes the internal edge lengths and topologies of unrooted trees on S. At this stage it is convenient to consider rooted trees on S formally. *Rooted* phylogenetic trees are defined in the same way as unrooted phylogenies, except they contain a unique vertex labelled as the root. The root vertex has degree ≥ 2, with degree exactly 2 in fully resolved trees. By attaching an additional leaf labelled 0 to the root vertex via an unweighted edge, an unrooted phylogeny is obtained, but the leaf set is now labelled $\{0, 1, \ldots, N\}$. It follows that the collection of all rooted

phylogenies on $S = \{1, \ldots, N\}$, denoted $\mathcal{T}_N$, is given by

$$\mathcal{T}_N = \mathbb{R}^N_{\geq 0} \times \mathrm{BHV}_{N+1}.$$

The space BHV_N can be described either by an embedding into a high dimensional Euclidean space or, equivalently, by an intrinsic construction, and we consider both approaches. There are $M = 2^{N-1} - 1$ possible splits of S, of which $M - N$ correspond to internal edges on trees. To embed BHV_N in Euclidean space, we order these splits arbitrarily and then associate the i-th split σ_i with the standard basis vector e_i in $\mathbb{R}^{M-N}$ for $i = 1, \ldots, M - N$. Every point $x \in \mathrm{BHV}_N$ can be represented by its vector of internal edge weights $\sum_i \lambda_i(x) e_i$ where

$$\lambda_i(x) = \begin{cases} \text{weight of split } \sigma_i \text{ in } x & \text{if } x \text{ contains } \sigma_i, \text{ or} \\ 0 & \text{if } \sigma_i \text{ is not contained in } x. \end{cases}$$

Arbitrary vectors in $\mathbb{R}^{M-N}$ do not generally correspond to trees, as arbitrary collections of splits do not give valid tree topologies. In fact there are $(2N-5)!! = 1 \times 3 \times 5 \times \cdots \times (2N-5)$ fully resolved unrooted topologies on N leaves, so the fully resolved trees occupy $(2N-5)!!$ copies of $\mathbb{R}^{N-3}_{\geq 0}$ in $\mathbb{R}^{M-N}$. Hence, the number of topologies and the dimension of the ambient space grow exponentially in N, while the local dimension of tree space grows linearly. The space BHV_4, corresponding to the unrooted trees on 4 leaves, or (equivalently) rooted trees on 3 leaves, has $M = 7$ different splits and the embedding into $\mathbb{R}^{M-N} = \mathbb{R}^3$ consists of the three positive orthogonal axes. In other words, $\mathrm{BHV}_4 = \mathrm{Spider}_3$ as illustrated in Figure 1.1.

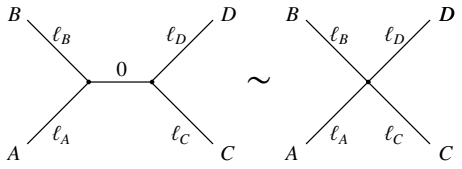

Figure 1.7: Illustration of the equivalence relation in equation (1.13). Here A, B, C, D are subtrees, and $\ell_A, \ell_B, \ell_C, \ell_D$ are the associated edge lengths. The internal edge in the tree on the left has length zero. The equivalence relation similarly applies when there are arbitrarily many subtrees either side of an internal edge.

While the embedding into $\mathbb{R}^{M-N}$ can be used to give a complete description of the geometry of BHV_N, it is not useful computationally due to the sparsity of vectors representing trees. Instead, BHV_N can be constructed intrinsically in terms of an equivalence relation in the following way [42]. This description differs in some

ways from the original BHV paper [11], but the explicit use of an equivalence relation and quotient makes the construction comparable to that for unlabelled trees in Section 1.4. The collection of trees O_T with some fixed fully resolved topology T can be parametrized by $\mathbb{R}_{\geq 0}^{N-3}$ by associating each internal edge in the topology with a coordinate axis. In fact we equip O_T with the induced Euclidean metric so that $O_T \cong \mathbb{R}_{\geq 0}^{N-3}$ is an isometry. The same notation is used when T is unresolved: $O_T \cong R_{\geq 0}^{k}$ when T contains k internal edges. Each set of trees O_T is called an *orthant*, and if T is fully resolved then O_T is a *maximal orthant*. BHV tree space is constructed by taking the disjoint union of all orthants, and quotienting by an equivalence relation:

$$\mathrm{BHV}_N = \bigcup_T O_T \Big/ \sim, \tag{1.13}$$

where the union is taken over all possible topologies. The equivalence relation $\sim$ is defined in Figure 1.7. Under the relation, trees are identified if and only if they are identical modulo the presence of splits with zero weight. Thus, when an edge is contracted to length zero, it can equivalently be removed from the tree.

The equivalence relation glues orthants together to form BHV_N. Orthants corresponding to unresolved topologies are contained in the equivalence classes of elements contained in the boundary of maximal orthants. Maximal orthants are glued at their codimension-1 boundaries in a relatively simple way. If a single internal edge in a tree with fully resolved topology T is contracted to length zero and removed from the tree, the result is a vertex of degree 4. There are then three possible ways to add in an extra edge to give a fully resolved topology, including the original topology T, so each codimension-1 face of O_T is glued to two other maximal orthants at their boundaries. It follows that near codimension-1 boundaries, BHV_N locally resembles Book_3^{N-4}. On the other hand, the tree containing no internal edges, called the *star tree*, corresponds to the origin in every set O_T.

For $N = 5$, the embedding of BHV_5 into $\mathbb{R}^{M-N} = \mathbb{R}^{10}$ is difficult to visualize, but the intrinsic construction is more accessible. There are 15 different maximal orthants defined on 10 different internal splits. The graph representing attachments between codimension-1 faces of maximal orthants is a 3-valent graph with 10 vertices (one for each split) and 15 edges (one for each orthant) called the Petersen graph, illustrated in Figure 1.8. Assuming the graph is equipped with unit edge lengths, points on the graph are in 1-to-1 correspondence with the points in BHV_5 whose two internal edge weights sum to a fixed non-zero constant. In entirety, BHV_5 is the cone of the Petersen graph, in other words the set of rays (copies of $\mathbb{R}_{\geq 0}$) joined at their common origin and in 1-to-1 correspondence with points on the graph. The rays passing through a fixed edge of the Petersen graph form the quadrant associated to that edge. As shown by Figure 1.8, the graph contains various cycles of length 5, each of which corresponds to an arrangement of 5 quadrants resembling Figure 1.2. Similarly, in neighbourhoods of vertices of the Petersen graph, the cone resembles Book_3^1.

It is straightforward to see that BHV_N is a cubical complex: each maximal orthant is an infinite array of unit $(N-3)$-cubes, and the structure of the complex is

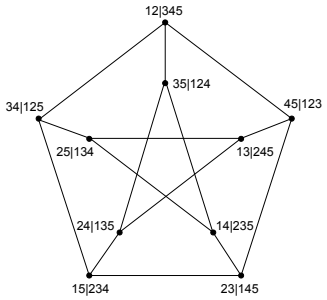

Figure 1.8: Vertices are drawn as dots and other edge crossings do not correspond to graph vertices. Each vertex is labelled with a split and corresponds to the codimension-1 boundary between three quadrants in BHV_5. Each edge corresponds to the quadrant in BHV_5 comprising trees whose two internal edges are determined by the two splits at either end of the edge.

determined by the way cubes are glued together at their boundaries according to the unresolved trees they represent. Since each cube is glued to a finite number of other cubes, the space is locally compact and the Hopf-Rinow theorem implies that BHV_N is a geodesic metric space. Billera, Holmes and Vogtmann [11] proved that BHV_N is CAT(0) by showing that Gromov's condition for cubical complexes holds, as discussed in Section 1.1.3. Gromov's condition corresponds exactly to the condition that pairwise compatible collections of splits determine valid tree topologies, as established in theorem 1.12. Since BHV_N is CAT(0), the product spaces $\mathcal{T}_N$ and $\mathcal{U}_N$ of rooted and unrooted trees are also CAT(0).

By definition, if two points $x, y \in \mathrm{BHV}_N$ lie in the same orthant, then the geodesic segment $\Gamma(x, y)$ between them is simply the straight line segment within the orthant. When the points x, y lie in different orthants, the geodesic comprises straight line segments in different orthants joining x to y. Along each geodesic, x is continuously deformed into y by contracting and expanding various edges. One possibility for the geodesic between points in different orthants is that it consists of the straight line segment from x to the origin and then the straight line segment from the origin to y. This is the path given by contracting all internal edges in x to length zero, to give the star tree, followed by expanding out all the edges in y. This is called the *cone path* between x and y. Cone paths are geodesics for certain points $x, y \in \mathrm{BHV}_N$, and the length of the cone path provides an upper bound on the length of the geodesic segment $\Gamma(x, y)$. Geodesics on $\mathcal{T}_N$ and $\mathcal{U}_N$ are the obvious trivial product between the geodesics in $\mathbb{R}^N_{\geq 0}$ for the pendant edges and the geodesics in BHV_N.

Figure 1.9 shows three geodesics in $\mathcal{T}_4$. All three occupy orthants on a single

copy of $\text{Cone}_{5\pi/2}$, and the figure shows the different types on geodesic which can occur on $\mathcal{T}_4$. Depending on the end-points, some geodesics are cone paths, in which case there is a "kink" at the origin. Examples like this on $\mathcal{T}_4$ and $\mathcal{U}_5$ give the impression that unresolved topologies are isolated points along geodesics. Figure 1.10

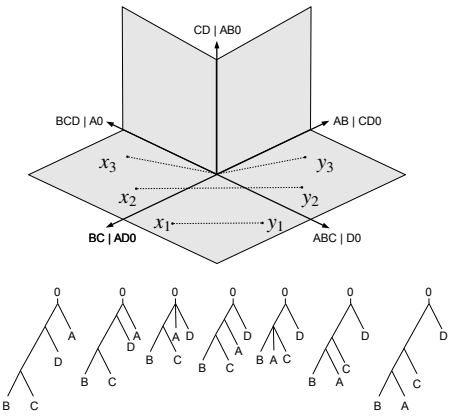

Figure 1.9: Three geodesics drawn in a copy of $\text{Cone}_{5\pi/2}$ within tree space (dashed lines). The geodesic $\Gamma(x_1, y_1)$ is the line segment in one orthant. $\Gamma(x_2, y_2)$ traverses three orthants, by expanding and contracting edges: representative trees along the geodesic are shown below. The geodesic $\Gamma(x_3, y_3)$ is a cone path: both internal edges are contracted to length zero with alternative edges then expanding out.

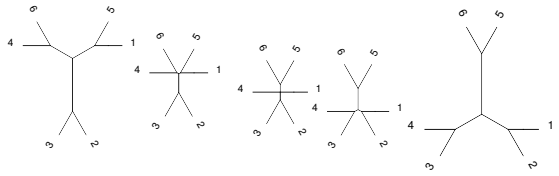

Figure 1.10: The unresolved topology displayed by the central tree is not an isolated point along the geodesic: the section of the geodesic in the corresponding non-maximal orthant has length strictly greater than zero.

gives an example in $\mathcal{U}_6$ where this is not the case. Geodesics on $\text{Cone}_{5\pi/2}$ can be thought of via a physical analogy. Suppose you construct an object like the squares in Figure 1.9 out of some rigid board. A piece of elastic string can be stretched between any two points on this object. The string acts to minimize its length and gives an approximate geodesic. As you move the end-points of the string around, it is easy to see it will tend to catch on the origin, so that the geodesic is a cone-path. In fact, this is a characteristic of tree space caused by the non-positive curvature: geodesics have a tendancy to move through high-codimension regions.

While the existence and uniqueness of geodesics on tree space follow as a result of BHV_N being CAT(0), it took a number of years following the original paper by Billera-Holmes-Vogtmann for a computationally efficient algorithm for constructing geodesics to emerge. Given the exponential number of orthants in tree space, computing geodesics could potentially be non-polynomial. However, Owen and Provan [47] developed a remarkable $O(N^4)$ algorithm for constructing geodesics which forms the basis of most of the methods described below. It operates by finding a maximum flow on a certain bipartite graph whose two vertex sets correspond to the splits in the two trees being connected. It can be thought of via the physical analogy of "tightening the string" between two points, like the example on the cone above, where the imaginary string is initiated as a cone path.

Before turning attention to existing statistical methodology in BHV tree space, we mention generalizations and related spaces. In BHV tree space, edges are assigned positive weights. However, arbitrary values in $\mathbb{R}$ or even elements of some vector space can also be used as the set of possible edge weights [26]. In this case, orthants are replaced with products of vector spaces, glued together at points where vectors are zero.

Retaining the assumption that edge weights are positive reals, three spaces related to BHV tree space have been studied. First is the subspace of trees for which the sum of all edge lengths is some fixed constant [61]. This space is also CAT(0), but it is not a cubical complex; as yet, there is no exact algorithm for computing geodesics. Another space of trees of interest to biologists comprises equidistance trees, namely rooted phylogenies for which all leaves are the same distance from the root. Gavryushkin and Drummond [28] considered two different geometries on this space, one of which consists of a CAT(0) cubical complex. Finally, Devadoss and Petti [16] have described a space of certain phylogenetic networks which are generalizations of trees. This is a cubical complex, but it is not CAT(0).

1.3.2 STATISTICAL METHODOLOGY IN BHV TREE SPACE

Throughout this section we assume that $D = \{x_1, \ldots, x_n\}$ is a sample of trees in either $\mathcal{T}_N$ or $\mathcal{U}_N$.

1.3.2.1 Fréchet mean and variance

Although biologists have defined and computed the mean of a sample D in a variety of ways, it is natural to consider the Fréchet mean $\bar{x}$ of D, defined in equation (1.8). If d is the largest distance between two points in D, and r is the distance between the origin and the furthest point, then any x which minimizes the Fréchet variance lies in the ball centred at the origin and with radius $d + r$. As tree space is CAT(0), the Fréchet function $\sum_i d(x, x_i)^2$ is convex in x and so attains a unique minimum within this ball. It follows that the Fréchet sample mean of D exists and is unique.

An algorithm originally due to Sturm [54], later extended and modified by other authors [9, 42], has been used for computing the Fréchet mean and variance in BHV tree space. The algorithms work in an iterative way, maintaining some estimate μ of $\hat{x}$. At each iteration, a data point x_i is selected either deterministically or by sampling from D; the geodesic from the current estimate μ to x_i is constructed; and μ is replaced with a point a certain proportion along this geodesic. While the algorithms are guaranteed to converge to $\bar{x}$, convergence can be slow in practice. Bačák's algorithms [9] are able to incorporate a weight for each data point in D, and can also be used to compute sample medians. Methods for minimizing the Fréchet function within a fixed orthant by making use of the local differentiable structure, have also been developed [53]. Owen and Brown have carried out a simulation study [13] to investigate behaviour of the Fréchet mean for samples from particular distributions of interest to biologists.

Asymptotic results have also been established for the Fréchet mean in tree space, under the limit of increasing sample size [6, 5]. These reflect the "stickiness" of the estimator, as described in Section 1.2. The asymptotic distribution of the sample mean consists of various Gaussian distributions on orthants, and in some situations non-maximal orthants can have strictly positive mass, corresponding to stickiness.

An example of a mean tree computed with the Sturm algorithm [54, 42] is shown in Figure 1.11. Here, the tree is formed by the centerlines passing through the tubular airway tree in the lung, segmented from chest CT scans [40], as originally presented in [21]. In order to model the airway trees using BHV tree-space, they are labelled using the automatic airway labelling algorithm of [22, 26], and the tree is cut off below the segment branches. Each branch of the tree is represented by 5 equidistant 3D landmark points, leading to 15-dimensional edge attributes (which is an easy extension of BHV tree-space, as remarked on page 20). The mean was computed from 8016 airway trees from the Danish Lung Cancer Screening Trial [49].

Note that the mean tree does not have any nodes with degree higher than 3. This indicates that the mean tree sits in the top dimensional orthant of BHV tree-space, and that it does not display sticky behavior. This lack of stickiness indicates that the population of trees is topologically relatively homogeneous, but does *not* mean that the airway trees all have the *same* topology; indeed, the dataset contains 1385 distinct topologies. However, about a third of the airway trees are contained in the 10 most frequent topologies, and more than 800 topologies only contain a single tree. Moreover, given the relatively fixed strucure of airway trees, we expect different

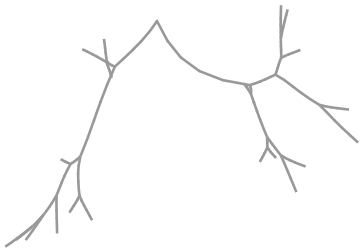

Figure 1.11: The mean airway centerline tree computed from a population of 8016 labelled airway centerline trees.

topologies to be quite similar to each other. These facts together explain why the mean airway tree is *not* sticky.

1.3.2.2 Principal component analysis

Principal component analysis (PCA) is a widely-used method for exploratory analysis and dimension reduction of high-dimensional data sets. It operates by identifying the main directions or modes of variation in a sample of vectors by eigendecomposition of the associated sample covariance matrix. As such, it inherently relies on the linear structure of the sample space, but the analysis can be re-expressed in a number of different, though equivalent, ways. In particular, PCA is equivalent to fitting affine subspaces to the data, in such a way as to minimize the sum of the squared distances of the data points from their projections onto each subspace. In Euclidean space, this amounts to finding the Fréchet mean (the zero-th order component), then a line of best fit to the data (the first principal component), then a plane of best fit, and so on. In Euclidean space the affine subspaces are necessarily nested, so, for example, the first principal geodesic passes through the Fréchet mean.

A best-fit geodesic is a natural analog in tree space to the Euclidean first principal component. Nye [44] first considered PCA in tree space, presenting an algorithm for constructing geodesics of best fit, constrained to pass through some choice of mean tree. The algorithm works by firing geodesics forward from the mean, with a greedy search to identify the optimal direction in which to fire. Golden ratio search is used to project data points onto candidate geodesics. The class of geodesics explored by this approach as candidates is limited by the constraint of passing through a given fixed mean, and as shown in Section 1.2.2, in tree space the principal geodesic does not necessarily contain the Fréchet mean. Furthermore, in tree space it is more natural to consider finite geodesic segments rather than infinitely long geodesic lines as principal components, since many infinite lines can share the same best-fit geodesic determined by the data. As an extreme example, consider the situation when all the data points are tightly clustered within the interior of a maximal orthant. Conventional

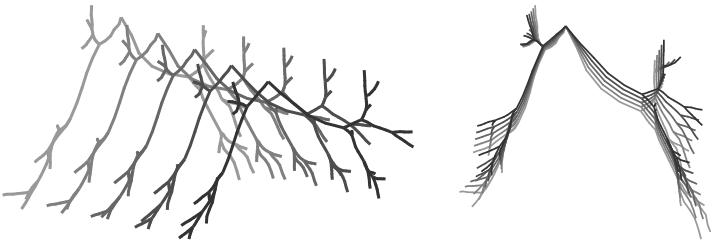

Figure 1.12: Airway trees sampled at 5 equidistant locations along the first principal component of the 8016 airway centerline trees, as computed in [21].

PCA could be used to construct a principal line within the orthant, but there would be many ways to extend this beyond the orthant into tree space. In view of this, Feragen et al [21] fitted finite geodesics to data by searching over geodesic segments whose end points were taken from the data set D. Examples can be constructed where such geodesics fit poorly in comparison to an unconstrained geodesic [62]. The constraint on the end points was subsequently dropped [46] by employing a stochastic search algorithm to vary the end points in tree space and search for the geodesic segment of best fit.

Construction of higher-dimensional objects in tree space to act as analogs of Euclidean principal components proved challenging. For example, the convex hull of three points in tree space – a natural candidate for a second order principal component – can have dimension strictly greater than 2. Examples of convex hulls in tree space with the "wrong" dimension in comparison to Euclidean space were first constructed by Sean Skwerer (personal communication). Details of a similar example based on his construction can be found in [45]. As an alternative to the convex hull, Nye et al [45] considered the locus of the weighted Fréchet mean of three given points, as the weights vary over the standard simplex. Also known as *barycentric subspaces* [50], these objects have the correct dimension, although they are not necessarily convex. It is possible that other analogs of PCA could be developed in tree space in the future, for example with different objects playing the role of higher dimensional components, potentially a nested version of PCA, or via some probabilistic model.

Figure 1.12 shows 5 sampled trees along the first principal component of the same 8016 airway centerline trees as used in Figure 1.11. The principal component was computed using the algorithm from [21]. We see the airways from two different views to emphasize the development throughout the principal component, which appears to capture breathing motion. Note that some of the trees along the principal component *do* contain a single node of degree 4, which indicates that the principal component partly runs along a codimension 1 stratum in BHV tree-space.

1.3.2.3 Other approaches

The preceeding methods all rely on least squares estimation but a few methods have been developed in BHV tree space which take a different approach. First, Weyenberg et al [59, 58] used Gaussian kernels in tree space to identify outliers in data sets. Gaussian kernels have density functions of the form $f(x) \propto \exp(d(x, x_0)^2/\sigma^2)$ where $d(\cdot, \cdot)$ is the BHV metric, x_0 is a point in tree space representing the mode of the distribution, and σ a dispersion parameter. The normalizing constant for these kernels is challenging to compute, and various computational approximations were employed. Secondly, Willis [60] developed analysis via projection onto a tangent space based at the Fréchet mean, by an analog of the Riemannian log-map. Here, the log-map is used to represent trees as points in a vector space in which existing Euclidean statistical methods can be applied. In [52], a set of brain artery trees were mapped to points in BHV tree space, and the data set was analysed by a variety of methods, including construction of the Fréchet mean, multidimensional scaling and minimal spanning trees. Finally, Chakerian and Holmes [15] presented a method for evaluating how close a data set comes to lying on a tree within tree space (a "tree of trees") in addition to various methods based on multidimensional scaling.

1.4 THE SPACE OF UNLABELLED TREES

While BHV tree space has the advantage of CAT(0) geometry and polynomial time algorithms for computing geodesics, it comes with the assumption that all trees have the same labeled set of leaves. In many applications, including most anatomical trees, this assumption does not hold. In this section, we review the space of unlabelled trees (tree-like shapes) as defined in [25]. Versions of this space have been used to study airway trees from human lungs [25], blood arteries on the surface of the heart [31] and neuronal trees [19].

As we shall see below, the geometry of the space of unlabelled trees is more complicated than the geometry of BHV tree space. Among other things, geodesics are not generally unique, and curvature is not generally bounded. As a result, most research has so far gone into algorithms or heuristics for computing geodesics [25, 31, 19], as well as analysis that only requires geodesics or their lengths, such as labelling [31], clustering and classification [19]. Heuristic "means" have been proposed in place of the Fréchet mean [20], but these are not exact. In order to study more complex statistics, a better understanding of tree space geometry is needed. This section contributes to that by giving a thorough definition of the space of unlabelled trees, and linking its geodesics to geodesics in BHV tree space. We use this link to prove that for two points sampled from a natural class of probability distributions, their connecting geodesic is almost surely unique. Such uniqueness of geodesics is important for geometric statistical methods to be well-defined.

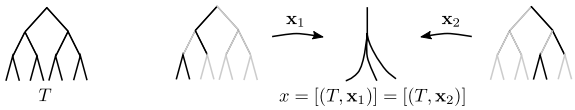

Figure 1.13: The supertree T must be large enough to span all trees of interest. It can represent smaller trees using 0-valued attributes to represent contracted branches. This allows representation of higher order vertices, but also results in multiple equivalent representations of unlabelled trees, as shown on the right.

1.4.1 WHAT IS AN UNLABELLED TREE?

Unlabelled trees are represented as pairs $(T, \mathbf{x})$ consisting of a *combinatorial tree* T and a *branch attribute map* $\mathbf{x}$, where T plays a role similar to the tree topology in Section 1.3. A combinatoral tree is a triple $T = (V, E, r)$ where V is a vertex set, E is an edge set so that the resulting graph is connected and does not have cycles, and $r \in V$ is a designated root vertex. Edges are undirected, so that the vertex pairs (u, v) and (v, u) define the same edge in E. Given an edge $e \in E$, any other edge $e' \in E$ on the path from e to r is said to be *above* e. If e' is above e, then we say that e is *below* e'. Parent, child and sibling relationships between edges can be similarly defined using the root.

A branch attribute map is a mapping $\mathbf{x}\colon E \to A$ associating to each edge $e \in E$ an edge attribute $\mathbf{x}(e) \in A$, where A is called the edge attribute space. In all our applications, A will contain a 0 element, which represents a *contracted* branch. Through contracted branches, we can represent many different unlabelled trees using the same combinatorial tree, and we can also represent higher order vertices using a binary combinatorial tree, as in Figure 1.13. This leads us to define *minimal* representations of unlabelled trees: A representation $(T, \mathbf{x})$ with combinatorial tree $T = (V, E, r)$ is *minimal* if $\mathbf{x}(e) \neq 0$ for all $e \in E$. Given an unlabelled tree representation $(T, \mathbf{x})$, we denote by $(\hat{T}, \hat{\mathbf{x}})$ its minimal representation with $\hat{T} = (\hat{V}, \hat{E}, \hat{r})$ and $\hat{\mathbf{x}} = \mathbf{x}|\hat{E}$, where $\hat{T}$ is obtained from T by contracting all edges that have 0 attribute.

An unlabelled tree is *spanned* by the combinatorial tree T if it can be represented as a pair $(T, \mathbf{x})$. Two unlabelled trees $(T_0, \mathbf{x}_0)$ and $(T_1, \mathbf{x}_1)$ are *equivalent*, denoted $(T_0, \mathbf{x}_0) \sim (T_1, \mathbf{x}_1)$, if, for their minimal representations $(\hat{T}_0, \hat{\mathbf{x}}_0)$ and $(\hat{T}_1, \hat{\mathbf{x}}_1)$, there exists a tree isomorphism $\phi\colon \hat{T}_0 \to \hat{T}_1$ such that, if $\phi_{\hat{E}}\colon \hat{E}_0 \to \hat{E}_1$ is the restriction of ϕ to edges, then $\hat{\mathbf{x}}_1 \circ \phi_{\hat{E}} = \hat{\mathbf{x}}_0$. Finally, we define an *unlabelled tree* as an equivalence class $x = [(T, \mathbf{x})]$.

In some applications, such as retinal vessels [41] or coronary arteries [31], the tree might actually reside on a surface and therefore have a natural planar order. This can be encoded by requiring the tree isomorphism ϕ to be an isomorphism of ordered trees, resulting in ordered unlabelled trees.

Example 1.14 (Edge attributes)**.** *The edge attribute space A can be designed to encode application-dependent branch properties. Branch length is encoded using*

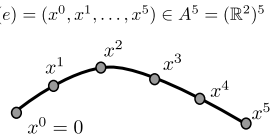

Figure 1.14: A simple model for branch geometry is obtained by representing the edge e by a set of n equidistant landmark points $x(e) \in (\mathbb{R}^N)^n$.

$A = \mathbb{R}_{\geq 0}$, and to encode branch geometry via landmarks (Figure 1.14), set $A = (\mathbb{R}^N)^n$. Here, a branch is described by n landmark points $x(e) = (x^0, x^1, x^2, \ldots, x^n) \in \{0\} \times (\mathbb{R}^N)^n$. The first landmark is typically translated to the origin ($x^0 = 0$) and left out of the analysis. Branch geometry can also be encoded via curves, giving $A = C^r(I, \mathbb{R}^N)$, the family of C^r curves. A version of this is used in [19].

Most results in this chapter assume that A is a finite dimensional vector space (or its positive orthant), whose metric is given by the Euclidean norm.

Definition 1.15 (The space of unlabelled trees)**.** *Fix a (possibly infinite) binary combinatorial tree T which is sufficiently large to span all the unlabelled trees of interest; we will henceforth refer to T as the* supertree*, see Figure 1.13. The space*

$$X = \prod_{e \in E} A = \{\mathbf{x} \colon E \to A\} \tag{1.16}$$

of all branch attribute maps $\mathbf{x}$ on E, contains all possible representations $(T, \mathbf{x})$ of unlabelled trees spanned by T. As shown in Figure 1.13, some unlabelled trees have multiple representations (T, x), and we construct the space of unlabelled trees spanned by T as the quotient of X with respect to the equivalence $\sim$ defined above:

$$\mathfrak{X} = X/\sim .$$

This definition covers both a space of ordered *unlabelled trees, and a space of* unordered *unlabelled trees, as accounted for in the definition of the equivalence $\sim$.*

The identifications made by the equivalence induce singularities in the tree space $\mathfrak{X}$. The metric on X induces a quotient metric [12] on $\mathfrak{X}$, called the *QED metric* (short for quotient Euclidean distance, as the original metric used on X was Euclidean).

As opposed to BHV tree space, not much is known about the geometry of $\mathfrak{X}$. The following theorem summarizes what we *do* know:

Theorem 1.17 (Geometry and topology of $\mathfrak{X}$ [25])**.**

i) The tree space $\mathfrak{X}$ with the QED metric is a contractible, complete, proper geodesic space.

ii) At generic points $x \in \mathfrak{X}$, the tree space $\mathfrak{X}$ is locally CAT(0).

iii) There exist $x_0, x_1 \in \mathfrak{X}$ with more than one geodesic connecting them. □

Due to the far more complex geometry of $\mathfrak{X}$, the computational tools and statistics in $\mathfrak{X}$ are far less developed than in BHV tree space. Computing geodesics is NP complete [23], but some heuristics have appeared [31].

1.4.2 GEODESICS BETWEEN UNLABELLED TREES

In this section we describe a previously unpublished relation between BHV geodesics and QED geodesics, which hints at a potential algorithm for computing or approximating geodesics in $\mathfrak{X}$. Additionally, we use this relation to prove almost sure uniqueness of geodesics between pairs of points in $\mathfrak{X}$. Throughout the section, an unlabelled tree $x \in \mathfrak{X}$ will be analyzed via its minimal representation $(\hat{T}, \hat{\mathbf{x}})$, with minimal combinatorial tree $\hat{T} = (\hat{V}, \hat{E}, \hat{r})$.

1.4.2.1 Mappings, geodesics and compatible edges

A tree space geodesic from x_0 to x_1 in $\mathfrak{X}$ carries with it an identification of subsets of the corresponding edge sets $\hat{E}_0$ and $\hat{E}_1$. A *mapping* [10] between $\hat{T}_0$ and $\hat{T}_1$ is defined as a subset $M \subset \hat{E}_0 \times \hat{E}_1$ such that for any two $(a, b), (c, d) \in M \subset \hat{E}_0 \times \hat{E}_1$ we have

i) $a = c$ if and only if $b = d$, and
ii) a is an ancestor of c if and only if b is an ancestor of d.

The mapping M identifies subsets of $\hat{E}_0$ and $\hat{E}_1$, in the sense that if the pair of edges $(a, b) \in \hat{E}_0 \times \hat{E}_1$ is in the subset $M \subset \hat{E}_0 \times \hat{E}_1$, then the edge a from $\hat{T}_0$ is identified with the edge b from $\hat{T}_1$. In view of this, the condition i) is a 1-1 identification condition on the edges; the edge a from $\hat{T}_0$ can only be identified with a single edge in $\hat{T}_1$ and vice versa. The condition ii) ensures that when all un-identified edges are contracted, the identification is a tree isomorphism between $\hat{T}_0$ and $\hat{T}_1$.

Given two trees $x_0, x_1 \in \mathfrak{X}$ and a mapping M between their minimal combinatorial trees $\hat{T}_0$ and $\hat{T}_1$, we say that a pair of unmapped edges $(e_0, e_1) \in \hat{E}_1 \times \hat{E}_2 \setminus M$ are *compatible* with M if $M \cup (e_0, e_1)$ is also a mapping between $\hat{T}_0$ and $\hat{T}_1$. A single unmapped edge $e_0 \in \hat{E}_0 \setminus \mathrm{pr}_{\hat{E}_0} M$ is *compatible* with the mapping M if $\hat{T}_1$ can be transformed into a tree $\hat{T}_1'$ by adding a zero-attributed edge e_1' so that $M \cup (e_0, e_1')$ is a mapping between $\hat{T}_0$ and $\hat{T}_1'$. Compatibility with M of a single unmapped edge $e_1 \in \hat{E}_1 \setminus \mathrm{pr}_{\hat{E}_1} M$ is defined analogously.

A path $\gamma\colon [0, 1] \to \mathfrak{X}$ from $x_0 = \gamma(0)$ to $x_1 = \gamma(1)$ naturally induces a mapping $M \subset \hat{E}_0 \to \hat{E}_1$: If the edge $a \in \hat{E}_0$ is identified with the edge $b \in \hat{E}_1$ by the path γ as illustrated in Figure 1.15, then $(a, b) \in M$, and vice versa. The *unmapped edges compatible with the mapping* are edges that do not "disturb" the shortest path associated with M, where length is measured with respect to the QED metric. In the shortest path from x_0 to x_1 associated with M, such edges will appear or disappear at one of the geodesic endpoints, shrinking to or growing from 0 at constant speed throughout the path. The *unmapped edges incompatible with the mapping* are edges from $\hat{T}_0$ that, in the shortest path associated with M, will have to disappear before

Figure 1.15: A geodesic γ from x_0 to x_1 induces a mapping between the minimal combinatorial trees $\hat{T}_0$ and $\hat{T}_1$ as indicated by colors. The black edges are unmapped, and incompatible with the mapping. In γ, the black edge from $\hat{T}_1$ cannot appear until the black edge from $\hat{T}_0$ has been contracted.

other edges from $\hat{T}_1$ can appear, or edges from $\hat{T}_1$ which cannot appear before other edges from $\hat{T}_0$ have disappeared. See Figure 1.15 for an example.

In the QED metric, the order and speed of edge deletions and additions in a tree space path affect the length of the path. In particular a path will, when possible, be shorter if it continuously performs two branch deformations simultaneously rather than first performing one, then the other. Thus, in order to find a QED geodesic, it is not enough to know which branches will be identified and which branches appear and disappear throughout the geodesic. That is, it is not enough to know the mapping. We also need to know at which point in the geodesic the branches will appear/disappear.

1.4.2.2 Link between QED geodesics and BHV geodesics

Consider two unlabelled trees x_0 and x_1, and let $\gamma\colon [0,1] \to \mathfrak{X}$ be a geodesic from x_0 to x_1 with mapping $M \subset \hat{E}_0 \times \hat{E}_1$ between the minimal representations $\hat{T}_0$ and $\hat{T}_1$.

Definition 1.18 (Subtrees spanned by a mapping)**.** *The subtrees $\tilde{x}_0$ and $\tilde{x}_1$ spanned by the mapping M are the subtrees of x_0 and x_1 obtained by removing all edges* below *the edges from $\hat{T}_0$ and $\hat{T}_1$ that appear in M. More precisely, $\tilde{x}_0 = (\tilde{T}_0, \tilde{\mathbf{x}}_0)$, where $\tilde{T}_0 = (\tilde{V}_0, \tilde{E}_0, \tilde{r}_0 = \hat{r}_0)$ is the combinatorial tree obtained by keeping all those vertices and edges from $\hat{T}_0$ that are found on the path from the root $\hat{r}_0$ to some edge in $pr_{\hat{E}_0}(M)$. The branch attribute mapping is defined by restriction: $\tilde{\mathbf{x}}_0 = \hat{\mathbf{x}}_0|\tilde{E}_0$. The subtree $\tilde{x}_1$ of x_1 is defined similarly. The remaining edges are collected in residual edge sets $R_i = \hat{E}_i \setminus \tilde{E}_i$, and attributed residual edge sets $r_i = (R_i, \mathbf{x}_i|R_i)$, $i = 0, 1$. See Figure 1.16.*

We now show that there is a leaf labeling of the subtrees $\tilde{x}_0$ and $\tilde{x}_1$ spanned by the

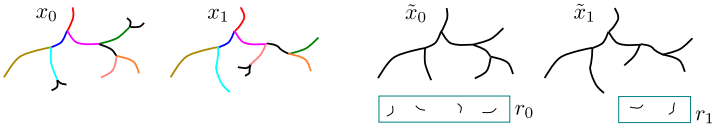

Figure 1.16: The mapping indicated on the left spans the subtrees $\tilde{x}_0$ and $\tilde{x}_1$ shown on on the right, leaving the residual edge sets r_0 and r_1.

mapping M, such that the geodesic γ decomposes as a product of a BHV geodesic between the leaf-labeled $\tilde{x}_0$ and $\tilde{x}_1$, and constant-speed interpolations between the origin and the attributed residual sets r_0 and r_1, respectively.

Consider M as a subset of $\tilde{E}_0 \times \tilde{E}_1$; then M is a mapping from $\tilde{T}_0$ to $\tilde{T}_1$. Each leaf in $\tilde{T}_0$ is mapped to a leaf in $\tilde{T}_1$, so by arbitrarily labeling the leaves in $\tilde{T}_0$, there is a labeling of the leaves in $\tilde{T}_1$ with the same labels, which is consistent with the mapping: If $(a, b) \in M$ where a and b are both leaves, then b is given the same label as a. The trees $\tilde{T}_0$ and $\tilde{T}_1$ might contain nodes of order 2, in which case a labeled "ghost" leaf with zero attribute is added at the node in order to raise its order. It may be necessary to add corresponding ghost edges and leaves in the other tree as well. Assume that the number of leaves in $\tilde{T}_0$ and $\tilde{T}_1$ is N; now the trees $\tilde{T}_0$ and $\tilde{T}_1$ can be considered as trees in BHV tree space $\mathcal{T}_N$ (with edge attribute set A, see page 20). The geodesic γ restricts to a map $\tilde{\gamma}\colon [0,1] \to \mathcal{T}_N$ which takes $\tilde{x}_0$ to $\tilde{x}_1$; this is a geodesic in $\mathcal{T}_N$:

Theorem 1.19. *A geodesic $\gamma\colon [0,1] \to \mathfrak{X}$ from x_0 to x_1 in $\mathfrak{X}$ decomposes as a BHV geodesic between $\tilde{x}_0$ and $\tilde{x}_1$, and Euclidean geodesics from r_0 to 0 and from 0 to r_1. In particular:*

i) All edges in R_0 and R_1 are compatible with the mapping M.
ii) All leaves in $\tilde{T}_0$ are mapped with leaves in $\tilde{T}_1$.
iii) The map $\tilde{\gamma} = pr_{\mathcal{T}_N} \circ \gamma$ is a geodesic in the BHV space $\mathcal{T}_N$.
iv) Denote by $\mathfrak{R}_0 = \prod_{R_0} A$ and $\mathfrak{R}_1 = \prod_{R_1} A$; now $pr_{\mathfrak{R}_i} \circ \gamma\colon [0,1] \to \mathfrak{R}_i$ for $i = 1, 2$ are constant-speed parametrizations of straight lines in $\mathfrak{R}_i$, the first from r_0 to 0 and the second from 0 to r_1.

For the proof, we need the following well-known lemma on product geodesics:

Lemma 1.20. [12, Chapter I, Proposition 5.3] *Let A and B be geodesic metric spaces, and let $A \times B$ have the metric $d^2((a_1, b_1), (a_2, b_2)) = d_A^2(a_1, a_2) + d_B^2(b_1, b_2)$. Now a path $\gamma\colon [0,1] \to A \times B$ is a geodesic if and only if it is a product of geodesics $\gamma_A\colon [0,1] \to A$, $\gamma_B\colon [0,1] \to B$, that is $\gamma = (\gamma_A, \gamma_B) : [0,1] \to A \times B$.* □

We are now ready to prove Theorem 1.19.

Proof. i) This holds because R_i consists of subtrees rooted at leaves in $\tilde{T}_i$, and such subtrees can be mapped onto ghost subtrees in the other tree. ii) This follows from the definition of a mapping: A leaf e_0 in $\tilde{T}_0$ cannot be mapped to a non-leaf edge e_1 in $\tilde{T}_1$ while some other edge e_0' in $\tilde{T}_0$ is mapped to the child of e_1 in $\tilde{T}_1$. iii)-iv) The geodesic γ must necessarily correspond to some path μ in $\mathcal{T}_N \times \mathfrak{R}_0 \times \mathfrak{R}_1$, and the length of γ in $\mathfrak{X}$ is the same as the length of μ in $\mathcal{T}_N \times \mathfrak{R}_0 \times \mathfrak{R}_1$. Reversely, for any other path $\tilde{\mu}$ in $\mathcal{T}_N \times \mathfrak{R}_0 \times \mathfrak{R}_1$ there is a corresponding path γ' in $\mathfrak{X}$ of the same length. Thus, the geodesic γ must correspond to a geodesic μ in $\mathcal{T}_N \times \mathfrak{R}_0 \times \mathfrak{R}_1$. But a geodesic in $\mathcal{T}_N \times \mathfrak{R}_0 \times \mathfrak{R}_1$ consists precisely of a BHV geodesic in $\mathcal{T}_N$ and straight Euclidean lines in $\mathfrak{R}_0$ and $\mathfrak{R}_1$ by Lemma 1.20. □

The significance of Theorem 1.19 is that it hints at an algorithm for computing QED geodesics by searching over all possible leaf labelings of the two unlabelled trees $x_0, x_1 \in \mathfrak{X}$. For each leaf mapping, we can compute the corresponding BHV geodesic between the corresponding $\tilde{x}_0, \tilde{x}_1$, and the interpolation between the corresponding attributed residual edge sets. Combining these, we can form the corresponding path in $\mathfrak{X}$, where the shortest possible such path is indeed the geodesic. While such an algorithm is still NP complete due to the search over all possible leaf labelings, one might be able to utilize heuristics for tree matching to reduce the search space in practice.

1.4.3 UNIQUENESS OF QED GEODESICS

While we have just seen that geodesics in $\mathfrak{X}$ decompose into products of BHV geodesics and Euclidean interpolations in $\mathfrak{R}_1$ and $\mathfrak{R}_2$, this does *not* indicate that $\mathfrak{X}$ is a product of $\mathcal{T}_N$ and a Euclidean space. Both the leaf-number N for $\mathcal{T}_N$, the assignment of leaf labels, and the residual spaces $\mathfrak{R}_1$ and $\mathfrak{R}_2$ depend on the two unlabelled trees x_0 and x_1. Nevertheless, we can use the previous result to prove uniqueness results for geodesics in $\mathfrak{X}$. First, note that such geodesics in $\mathfrak{X}$ are not *generally* unique:

Example 1.21. *Consider the simple case of the tree space spanned by the combinatorial tree T with two edges rooted at the root vertex, representing two geometric trees x_0 and x_1 with one edge each, and with branch attributes $\mathbf{x}_0(e_1) = (0, 1) \in \mathbb{R}^2$ and $\mathbf{x}_1(e_2) = (1, 0)$. There are now two geodesics from x_0 to x_1 in $\mathfrak{X}$: one which maps the two edges onto each other, and one which does not.*

However, we show that for a natural family of probability distributions, two independently sampled trees will almost surely have a unique connecting geodesic.

Any measure μ on X can be pushed forward to a measure $\mu_\sharp(Y) = \mu(\phi^{-1}(Y))$ on $\mathfrak{X}$ through the quotient map $\pi\colon X \to \mathfrak{X}$. Thus, in the case where the edge attribute space A is a Euclidean space (or orthant), we can endow $\mathfrak{X}$ with the push-forward of the Lebesgue measure on X. We now state the main theorem of this section:

Theorem 1.22 (Main theorem)**.** *Assume that the edge attribute space A is Euclidean or a Euclidean orthant, and let f be any probability density distribution on $\mathfrak{X}$ with respect to the push-forward of the Lebesgue measure on X. If x_0 and x_1 are independently sampled from f, then with probability* 1*, there is a unique geodesic connecting x_0 and x_1.*

Note that in Theorem 1.22, the probability density function f will exist whenever the corresponding probability measure is absolutely continuous with respect to the push-forward of the Lebesgue measure, ensuring that positive probability mass does not concentrate on the cut locus where pairs of points can have multiple geodesics.

In order to prove Theorem 1.22, we need to link unlabelled tree-space geodesics to BHV geodesics by assigning artificial "leaf labels" to select subsets of edges that will play the role of leaves.

Definition 1.23 (Leaf mapping)**.** *Given a mapping M between combinatorial trees T_0 and T_1, and subtrees $\tilde{T}_0$ and $\tilde{T}_1$ spanned by the mapping M, define*

$$M_L = \{(e_0, e_1) \in M | e_0 \text{ is a leaf in } \tilde{T}_0 \text{ and } e_1 \text{ is a leaf in } \tilde{T}_1\}.$$

We call M_L the leaf mapping *associated with the mapping M (and, when relevant, with the geodesic γ whose mapping is M).*

Our proof relies on the following observations:

Lemma 1.24. *Assume that x_0 and x_1 are sampled independently from f, where f is any probability density distribution on $\mathfrak{X}$ with respect to the push-forward of the Lebesgue measure on X.*

i) Note that while every geodesic in $\mathfrak{X}$ induces a mapping M, and thus a leaf mapping M_L, there may be several mappings $M_1, \ldots, M_k$, not all associated with geodesics, that give the same leaf mapping M_L.

ii) For any leaf mapping M_L, there is almost surely a unique shortest path from x_0 to x_1 associated to the leaf mapping, as follows: The leaf mapping defines leaf-labeled subtrees $\tilde{x}_0$ and $\tilde{x}_1$ of x_0 and x_1, respectively, as above. Associated to the leaf-labeled subtrees $\tilde{x}_0$ and $\tilde{x}_1$ there is a BHV geodesic $\tilde{\gamma}$ and residual spaces $\mathfrak{R}_0$ and $\mathfrak{R}_1$, which give rise to a path γ from x_0 to x_1; this is the shortest possible path from x_0 to x_1 with the given leaf mapping M_L associated to it.

iii) There are finitely many possible leaf mappings $(M_L)_i$, $i = 1, \ldots, N$, between the trees x_0 and x_1, which almost surely give rise to N shortest possible paths γ_i from x_0 to x_1 in $\mathfrak{X}$ with that given leaf mapping.

iv) Associated with the $(M_L)_i$ and their associated shortest paths γ_i, there are finitely many possible distances $d_1, \ldots, d_N$ between x_0 and x_1. It is possible that $d_i = d_j$ for different i, j, for instance if the geodesic is not unique.

v) Among the possible paths γ_i, $i = 1, \ldots, N$ enumerated in iii), the shortest one(s) will be the geodesic(s) between x_0 and x_1. Among the possible distances d_i in iv), the smallest distance $\min\{d_1, \ldots, d_N\}$ *is the QED distance between the unlabelled trees x_0 and x_1.*

Proof. i) If the subtrees spanned by the leaf mapping M_L do not have degree 2 vertices, then there is only one (maximal) mapping M with leaf mapping M_L. But if one of the subtrees, say $\tilde{x_0}$, spanned by the mapping *has* a degree 2 vertex, then there may be more than one way to add ghost vertices and edges to obtain a tree whose internal vertices have degree ≥ 3. ii) If the subtrees spanned by the leaf mapping M_L do not have degree 2 vertices, then there is a unique corresponding geodesic and mapping. If one of the subtrees *has* a degree 2 vertex, then there will only be more than one shortest path (and corresponding mapping) if there are different, equal-cost ways of matching the edges adjacent to the degree 2 mapping and the added ghost subtree, to corresponding edges and subtrees in the other tree. This will only happen if permutations of matched edges give the same total difference, which can only happen on a subset of measure 0. iii) Follows from ii). iv) Trivial. v) Follows from

Theorem 1.19. □

In a similar way as Theorem 1.19 we have:

Lemma 1.25. *Let γ be a geodesic from x_0 to x_1 in $\mathfrak{X}$ with a corresponding mapping M, and assume that $(e_0, e_1) \in M$, i.e. the edge e_0 in $\hat{T}_0$ is matched to the edge e_1 in $\hat{T}_1$ by the geodesic. Let $x_c(e_0)$ denote the child subtree of x_0 rooted at the end of e_0, and let $x_p(e_0)$ denote the remaining subtree of x_0 after removing e_0 and its child subtree $x_c(e_0)$. Similarly for x_1 and e_1. Then γ can also be represented as a product*

$$\gamma = (\gamma_1, \gamma_2, \gamma_3) \colon I \to \mathfrak{X} \times A \times \mathfrak{X},$$

where γ_1 is the shortest path from $x_c(e_0)$ to $x_c(e_1)$ respecting the restriction of M; γ_2 is the straight line from $\mathbf{x}(e_0)$ to $\mathbf{x}(e_1)$ in A; and γ_3 is the shortest path from $x_p(e_0)$ to $x_p(e_1)$ respecting the restriction of M. In particular, the length of γ in $\mathfrak{X}$ is the same as the length of $(\gamma_1, \gamma_2, \gamma_3)$ in $\mathfrak{X} \times A \times \mathfrak{X}$.

We are now ready to start the proof of Theorem 1.22. Suppose that x_0 is any tree in $\mathfrak{X}$ where no two nonzero edges have the same attribute. Denote by W_{x_0} the set of trees x_1 in $\mathfrak{X}$ such that there are at least two distinct geodesics γ_a and γ_b in $\mathfrak{X}$ connecting x_0 to x_1. If we can show that the set $\mathfrak{X} \setminus W_{x_0}$ is open and dense in $\mathfrak{X}$, then we have proven Theorem 1.22, since the measure on $\mathfrak{X}$ is the push-forward of the Lebesuge measure.

Lemma 1.26. *The complement $\mathfrak{X} \setminus W_{x_0}$ is open.*

Proof. Let $x_1 \in \mathfrak{X} \setminus W_{x_0}$, that is, x_1 is an unlabelled tree in $\mathfrak{X}$ with only one geodesic γ from x_0 to x_1, of length l_1. We prove the lemma by finding an $\varepsilon > 0$ such that $B(x_1, \varepsilon) \subset \mathfrak{X} \setminus W_{x_0}$.

There is a unique leaf mapping corresponding to γ; we denote it $(M_L)_1$. Associated with the finite number of other possible leaf mappings $(M_L)_2, \ldots, (M_L)_N$ there is a finite number of shortest possible path lengths $l_2 \leq \ldots \leq l_N$, where $l_1 < l_2$ since the geodesic from x_0 to x_1 is unique.

Set $\varepsilon = \frac{l_2 - l_1}{2}$, and assume that there exists some $x_1' \in B(x_1, \varepsilon) \cap W_{x_0}$, that is, such that $d(x_1, x_1') < \varepsilon$ and there are two geodesics γ_a and γ_b from x_0 to x_1'. Now, if γ_c is a geodesic from x_1' to x_1, then the concatenations of paths $\gamma_c * \gamma_a$ and $\gamma_c * \gamma_b$ give two distinct paths from x_0 to x_1, each of length

$$d(x_0, x_1') + d(x_1', x_1) \leq d(x_0, x_1) + 2d(x_1, x_1') < l_1 + 2\varepsilon = l_1 + 2\frac{l_2 - l_1}{2} = l_2.$$

But this is not possible since the shortest two paths of equal length from x_0 to x_1 have length at least l_2. Hence, there cannot be two geodesics from x_0 to x_1', i.e. $x_1' \notin W_{x_0}$ and $B(x_1, \varepsilon) \subset \mathfrak{X} \setminus W_{x_0}$. This completes the proof that $\mathfrak{X} \setminus W_{x_0}$ is open in $\mathfrak{X}$. □

First, note that for the set Δ of trees x_1 where at least two branches have identical attributes, its complement $\mathfrak{X} \setminus \Delta$ is open and dense in $\mathfrak{X}$. It is thus enough to show that

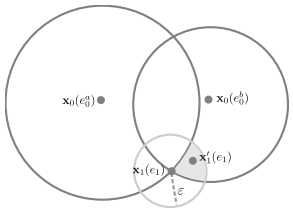

Figure 1.17: Since $\mathbf{x}_0(e_0^a) \neq \mathbf{x}_0(e_0^b)$, it is possible to find some $\mathbf{x}_1'$ such that the inequalities in equations (1.28) and (1.29) are satisfied.

$\mathfrak{X} \setminus (W_{x_0} \setminus \Delta)$ is open and dense in $\mathfrak{X}$, since the intersection of two open and dense sets is dense, and $\mathfrak{X} \setminus W_{x_0} \supset (\mathfrak{X} \setminus (W_{x_0} \setminus \Delta)) \cap (\mathfrak{X} \setminus \Delta)$.

Next, assume that $x_1 \in W_{x_0} \setminus \Delta$. We are going to show that for any $\varepsilon > 0$ the ball $B(x_1, \varepsilon)$ intersects $\mathfrak{X} \setminus W_{x_0}$. Since $x_1 \in W_{x_0}$, there are two distinct geodesics γ_a and γ_b from x_0 to x_1 in $\mathfrak{X}$. Let M_a and M_b be the corresponding mappings between $\hat{E}_0$ and $\hat{E}_1$, and denote by $\hat{E}_1^a = \mathrm{pr}_{\hat{E}_1}(M_a) \subset \hat{E}_1$ and $\hat{E}_1^b = \mathrm{pr}_{\hat{E}_1}(M_b) \subset \hat{E}_1$ the sets of edges in x_1 identified with some edge in x_0 by γ_a and γ_b, respectively. We divide the proof into two cases:

Case I: $\hat{E}_1^a = \hat{E}_1^b$, and **Case II:** $\hat{E}_1^a \neq \hat{E}_1^b$.

Proof of Case I: In this case, since $\gamma_a \neq \gamma_b$, there must be some edge e_1 in $\hat{E}_1^a = \hat{E}_1^b$ onto which two *different* edges e_0^a and e_0^b in $\hat{E}_0$ are mapped by $(M_L)_a$ and $(M_L)_b$, respectively. The two source edges have different edge attributes $\mathbf{x}_0(e_0^a)$ and $\mathbf{x}_0(e_0^b)$ by the assumption of the theorem, and e_1 has edge attribute $\mathbf{x}_1(e_1)$.

Lemma 1.27. *For any $\varepsilon > 0$, we can find an unlabelled tree x_1' with the same minimal combinatorial tree as x_1, such that $d(x_1, x_1') < \varepsilon$ and the number of geodesics from x_0 to x_1' is at most $p - 1$, where p is the number of geodesics from x_0 to x_1.*

Proof. By Lemma 1.24 there are (almost surely) finitely many leaf mappings between x_0 and x_1, denoted $(M_L)_1, \ldots, (M_L)_N$, with corresponding shortest possible paths $\gamma_1, \ldots, \gamma_N$ of lengths $l_1, \ldots, l_N$. We may assume $l_1 = l_2 = \ldots = l_p < l_{p+1} \leq \ldots \leq l_N$. The length of a geodesic from x_0 to x_1 is thus l_1. Without loss of generality, and possibly swapping a and b, we may assume that $\varepsilon < l_{p+1} - l_1$. Since $\varepsilon > 0$, we can find an attribute map $\mathbf{x}_1' \colon E_1 \to A$ such that $\mathbf{x}_1'|\hat{E}_1 \setminus \{e_1\} = \mathbf{x}_1|\hat{E}_1 \setminus \{e_1\}$, and

$$\|\mathbf{x}_1'(e_1) - \mathbf{x}_1(e_1)\| < \varepsilon, \tag{1.28}$$

while

$$\|\mathbf{x}_0(e_0^a) - \mathbf{x}_1(e_1)\| < \|\mathbf{x}_0(e_0^a) - \mathbf{x}_1'(e_1)\| \text{ and } \|\mathbf{x}_0(e_0^b) - \mathbf{x}_1(e_1)\| > \|\mathbf{x}_0(e_0^b) - \mathbf{x}_1'(e_1)\|. \tag{1.29}$$

Denote by x_1' the unlabelled tree whose edge attribute map is $\mathbf{x}'$. Note that the

leaf mappings $(M_L)_a$ and $(M_L)_b$ can be transferred to the pair (x_0, x_1') since x_1' has the same tree topology as x_1. This induces two paths γ_a' and γ_b' from x_0 to x_1', which are the shortest possible paths with the corresponding leaf mappings $(M_L)_a$ and $(M_L)_b$. By Lemma 1.25 and Eq. (1.29), the length $l(\gamma_b')$ of γ_b' satisfies

$$l(\gamma_b')^2 = l(\gamma_b)^2 + \|\mathbf{x}_0(e_0^b) - \mathbf{x}_1'(e_1)\|^2 - \|\mathbf{x}_0(e_0^b) - \mathbf{x}_1(e_1)\|^2 < l(\gamma_b)^2,$$

so $l(\gamma_b') < l(\gamma_b)$. Now, we see that the shortest path from x_0 to x_1' corresponds to a leaf mapping M_L which also gives a shortest path from x_0 to x_1.

To see this, let γ_c' be the geodesic from x_0 to x_1', with leaf mapping $(M_L)_c$; we then have $l(\gamma_c') \leq l(\gamma_b')$. The leaf mapping $(M_L)_c$ also generates path γ_c from x_0 to x_1 which is the shortest possible with leaf mapping $(M_L)_c$. We now have

$$\begin{aligned} l(\gamma_c) &\leq l(\gamma_c') + d(x_1', x_1) < l(\gamma_c') + \varepsilon < l(\gamma_c') + (l_{p+1} - l_1) \leq l(\gamma_b') + (l_{p+1} - l_1) \\ &< l(\gamma_b) + (l_{p+1} - l_1) = l_1 + (l_{p+1} - l_1) < l_1 + l_{p+1} - l_1 = l_{p+1}, \end{aligned}$$

so we must necessarily have $l(\gamma_c) = l_1$, i.e., γ_c is a shortest path from x_0 to x_1.

As a consequence, there are no new geodesic-generating leaf mappings between x_0 and x_1', which were not geodesic-generating between x_0 and x_1. Thus, the number of shortest paths from x_0 to x_1' is at most $p - 1$, where p is the number of shortest paths from x_0 to x_1. This concludes the proof of Lemma 1.27. □

By repeatedly using Lemma 1.27, we see that for any $\varepsilon > 0$ there is a tree x_1' with $d(x_1, x_1') < \varepsilon$ and a unique geodesic from x_0 to x_1', which proves Case I.

Proof of Case II: We must have $|\hat{E}_1^a| = |\hat{E}_1^b|$ by the definition of a mapping; therefore we may assume (by symmetry) that $e_1 \in \hat{E}_1^a \setminus \hat{E}_1^b$. That is, there exists some $e_0 \in \hat{E}_0$ which is identified with $e_1 \in \hat{E}_1^b$ by γ_b, whereas e_0 is not identified with any edge in $\hat{E}_1$ by γ_a. Let $t_0 \in [0, 1]$ be the time at which the zero attributed edge corresponding to e_1 appears in the geodesic γ_a. Let $x_{t_0}^b(e_0)$ be the attribute associated to the edge mapped from e_0 to e by γ_b at time t_0. Find an attribute map $x_1' : \hat{E}_1 \to \mathbb{R}^N$, such that $\mathbf{x}_1'|\hat{E}_1 \setminus \{e_1\} = \mathbf{x}_1|\hat{E}_1 \setminus \{e_1\}$, and $\|\mathbf{x}_1'(e_1) - \mathbf{x}_1(e_1)\| < \varepsilon$, while

$$\|(\mathbf{x}')_{t_0}^a(e_0) - \mathbf{x}_1(e_1)\| < \|\mathbf{x}_{t_0}^a(e_0) - \mathbf{x}_1'(e_1)\| \quad \text{and} \quad \|\mathbf{x}_1(e_1)\| > \|\mathbf{x}_1'(e_1)\|.$$

Now, γ_b' is shorter than γ_a', and in particular,

$$l(\gamma_b') \leq d(x_0, x_{b,t_0}) + d(x_{b,t_0}, x_1') < d(x_0, x_{b,t_0}) + d(x_{b,t_0}, x_1) = d(x_0, x_1) = l(\gamma_b).$$

The second inequality holds by Lemma 1.25. The proof wraps up as in Case I. □

1.5 BEYOND TREES

Tree space is so far the stratified data space that has seen the most attention, both in theoretical developments and in applications that perform statistical analysis in the stratified space (as opposed to reducing the data to Euclidean features). This is

most likely caused in part by the availability of efficient code for computing BHV geodesics [47], and in part the availability of tree-structured data [1].

However, a number of other applications generate data with combinatorial properties that are modelled well using stratified spaces. Below we discuss a few examples, some of which have seen some analysis – and some which are yet unexplored.

1.5.1 VARIABLE TOPOLOGY DATA

Stratified spaces are well suited for modeling data with variable topological structure, as we have already seen in the case of trees. This idea generalizes also to other examples.

Example 1 (Graphs)**.** Graph-structured data are often represented using adjacency matrices. An adjacency matrix representing a directed graph with n vertices is an $n \times n$ matrix M such that the entry M_{ij} contains a scalar or vector attribute which describes the directed edge from vertex i to vertex j, and the entry M_{ii} describes the vertex i. Undirected graphs are given by symmetric matrices, and graphs with different sizes can be represented as fixed-size $n \times n$ adjacency matrices by entering empty (ghost) vertices described by zero attributes, just like the unlabeled trees from Section 1.4. Such an approach is used by Jain and Obermayer [34], who build a space of attributed graphs as a quotient of the space of adjacency matrices on n vertices, where vertex permutations are factored out. In their graph-space, the zero attribute M_{ij} denotes a situation where there is no edge connecting the edges i and j, as in Fig. 1.18 (a), top. A similar space of attributed graphs was also studied recently by Kolaczyc et al [39], reproducing several results from [34].

In the space of attributed graphs from [34], the zero edge attribute corresponds to "no edge", as in Fig. 1.18 (a), top. Within such a model, Jain and Obermayer develop theory for statistics and machine learning such as means and medians, clustering and classification using Lipschitz optimization on the quotient, which is computationally efficient. However, this model does not accommodate continuous edge lengths well: In the case where the "shape" of a graph is simply described by the lengths of edges, we obtain converging sequences as illustrated in Figure 1.18 (b, top), where a sequence of cycles with one decreasing edge converges to a "line" graph with four vertices on it. In order to accommodate edge lengths, one might prefer a length 0 branch to be a contracted one, as shown in Figure 1.18 (a-c, bottom). This would imply identification of vertices, which has a drastic impact on the geometry of the space of graphs, which in the top case is a quotient with respect to the permutation group as introduced in [34] – but which in the bottom case is not.

Another example of variable topology data is given by *point sets*, which can be used to represent a wide variety of data objects.

Example 2. A *point set* is a finite set $\{x_1, \ldots, x_n\} \subset X$ of points in a geodesic metric

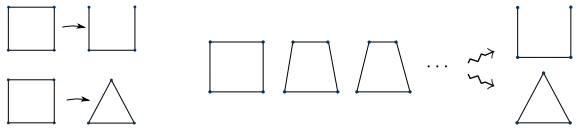

Figure 1.18: Having edge attribute 0 can be used to model either no branch (a, top), or a contracted edge, causing vertices to melt together (a, bottom). This choice drastically affects limits of sequences (b).

space X. We denote by $\mathcal{P} = \{\{x_1, \ldots, x_n\} \subset X | n \in \mathbb{N}\}$ the space of point sets of arbitrary cardinality. Geometric objects can be specified by point clouds or landmark points, where one traditionally requires exact correspondence between the point sets for different objects [38]. However, this does not always make sense, e.g. if there is occlusion or missing annotations; if the landmark points correspond to non-existent physical attributes; or if the landmark points represent geometric features such as high curvature, rather than specific physical attributes. Other examples of point set data include objects tracked over time or over different 2D slices of a 3D object. Persistence diagrams [56] are a special case of point sets, where the points lie above the diagonal, in the positive orthant of $\mathbb{R}^2$. In persistence diagrams, the diagonal itself represents an arbitrary number of "dummy" points to which points from another persistence diagram can be matched, in effect constituting a quotient space in which all points on the diagonal are identified. Returning to point sets in general, by imposing an order on the point sets, we obtain *sequences*, where examples include spike trains, time series and other discrete signals [3, 35].

There are different ways to interpolate between point sets with variable cardinality, and different modelling choices lead to different "point set spaces". In case of occlusion or missing labels, it is natural to introduce "dummy" points, which can be matched to points which exist in one object, but not the other. This is similar to the 0 attributed edges in the space of graphs. A different modelling choice could be to allow points to merge together. This would, for instance, make sense when tracking cells over time, which might divide.

In every example encountered in this chapter, a data space can be built as follows: Consider data objects belonging to a discrete set of topologies $\mathcal{T} = \{T_i : i \in I\}$. These topologies could be different different tree topology, different graph structures, different point set cardinality, or something else entirely. Restricting analysis to the set X_i of data objects that have the fixed topology T_i, we apply known techniques: Trees, graphs and point sets with a fixed topology are represented as fixed-length vectors or matrices, to which standard Euclidean or manifold statistics can be applied. Including all the different topologies in $\mathcal{T}$, we obtain a disjoint union $X = \dot{\bigcup} X_i$ of spaces

where different topologies are represented, but where we cannot yet interpolate between points in different subspaces X_i. Ultimately,we interpolate between different topologies by realizing that, as with the trees, the boundary of each fixed-topology stratum X_i consists of data whose topology is a degeneration of the topology found in X_i, and thus topologically different. We join the different X_i when their degenerated boundary topologies coincide, just like with the trees in Sections 1.3 and 1.4.

1.5.2 MORE GENERAL QUOTIENT SPACES

For all the examples above, different topologies are bridged by identifying different representations of the same degenerated topologies along stratum boundaries. As with trees, this can be thought of as creating a quotient $X/\sim$, where $x \sim x'$ whenever x and x' are two different representations, in two different strata, of the same point.

This quotient space approach extends beyond topological variation, for example to the case of symmetric, positive definite (SPD) matrices. SPD matrices are frequently encountered data objects, representing e.g. diffusion tensors [7, 4, 27, 8, 18] or covariance descriptors [57]. Any SPD matrix Σ can be interpreted as the covariance of a centered normal distribution, whose shape is characterized by its eigenvalues and whose orientation is characterized by its eigenvectors. Note that whenever all eigenvalues are distinct, the eigenvectors (orientations) are unique up to sign change, whereas when two or more eigenvalues coincide, the eigenvectors are no longer unique (corresponding to rotational symmetry of the normal distribution). This leads to a stratification of the set of eigenspace decompositions of $n \times n$ SPD matrices [29], where the stratification corresponds to the eigenvalue multiplicities.

Both the topologically variable data and the SPD matrices are quotient spaces with respect to equivalences. A particularly well-understood type of equivalence is defined by belonging to the same group action orbit. For instance, Kendall's shape space is a quotient of $(\mathbb{R}^d)^n$ with respect to the group of rotations, translations and rescaling, and the space of attributed graphs is a quotient of the space of adjacency matrices with respect to the node permutation group. Group quotients appear when we seek invariance with respect to a group action; invariant properties are properties which can be defined on orbits rather than on data points. However, unless the group action is particularly nice (and often it is not), group quotients are not generally smooth. For instance, Kendall's shape space [38] has singularities, and these singularities correspond to changes in the diffeomorphism class of the group orbit at the singularity. Other variants of shape space, such as projective shape space, have also been studied as stratified data spaces [37, 48]. When a Lie group acts properly on a smooth manifold, the quotient is a stratified space, stratified by orbit type – and the stratification can be helpful in understanding the geometric, computational and statistical properties of the quotient [29].

While imposing invariants means that one is actually working on a quotient space, geometric statistics is not always phrased as a problem on a quotient. Moreover, when it *has* been phrased as a statistical problem on a quotient, it has not typically been acknowledged that noise in the data might *not* live on the quotient. This is

the topic of recent work [43, 17], which shows that statistics on quotients can be biased by these modelling choices. It is still unknown whether this can be interpreted through stratified (singular) space geometry.

1.5.3 OPEN PROBLEMS

Stratified spaces have many attractive properties from the modelling point of view: They allow continuously bridging different topologies or structures, and they give a way to organize singularities in quotient spaces. However, a number of open problems are still left unanswered. As we have seen in Section 1.2, least squares statistics in stratified spaces exhibit unexpected and possibly unwanted properties due to the singularities in the stratified space. Can we design statistical methods that do not exhibit stickiness – or that, perhaps, are sufficiently rich to capture the variation ignored by stickiness? As we have seen in Sections 1.3 and 1.4, computing geodesics is another challenge, and thus also an open problem, at least in the case of unlabelled trees. However, the other data types above do not exhibit the hierarchical structure of trees, and might therefore prove less computationally challenging. A general open challenge is thus to utilize the modelling capabilities of stratified spaces in new applications.

[1] Treebase, a database of phylogenetic knowledge, `http://www.treebase.org`.

[2] A.D. Aleksandrov, V.N. Berestovskii, and I.G Nikolaev. Generalized Riemannian spaces. *Russian Mathematical Surveys*, 41(3):1, 1986.

[3] Dmitriy Aronov and Jonathan D. Victor. Non-Euclidean properties of spike train metric spaces. *Phys Rev E*, 69(6), 2004.

[4] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A Log-Euclidean framework for statistics on diffeomorphisms. In *MICCAI*, 2006.

[5] Dennis Barden and Huiling Le. The logarithm map, its limits and Fréchet means in orthant spaces. *arXiv preprint, arXiv:1703.07081*, 2017.

[6] Dennis Barden, Huiling Le, and Megan Owen. Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electron. J. Probab*, 18(25):1–25, 2013.

[7] Peter J Basser, James Mattiello, and Denis LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1):259–267, 1994.

[8] P. G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine*, 53(1):221–225, 2005.

[9] M. Bačák. Computing medians and means in Hadamard spaces. *SIAM J. Optimiz.*, 24(3):1542–1566, 2014.

[10] Philip Bille. A survey on tree edit distance and related problems. *Theor. Comput. Sci.*, 337(1-3):217–239, 2005.

[11] L. J. Billera, S. P. Holmes, and K. Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27(4):733–767, 2001.

[12] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*. Springer, 1999.

[13] D.G. Brown and M. Owen. Mean and variance of phylogenetic trees. *arXiv preprint, arXiv:1708.00294*, 2017.

[14] P. Buneman. The recovery of trees from measures of dissimilarity. In F.R. Hodson, D.G. Kendall, and P. Tautu, editors, *Mathematics in the Archaeological and Historical Sciences*, pages 387–395. Edinburgh Univeristy Press, Edinburgh, 1971.

[15] John Chakerian and Susan Holmes. Computational tools for evaluating phylogenetic and hierarchical clustering trees. *Journal of Computational and Graphical Statistics*, 21(3):581–599, 2012.

[16] Satyan L Devadoss and Samantha Petti. A space of phylogenetic networks. *SIAM Journal on Applied Algebra and Geometry*, 1(1):683–705, 2017.

[17] Loïc Devilliers, Stéphanie Allassonnière, Alain Trouvé, and Xavier Pennec. Inconsistency of template estimation by minimizing of the variance/pre-variance in the quotient space. *Entropy*, 19(6):288, 2017.

[18] Ian L. Dryden, Alexey Koloydenko, and Diwei Zhou. Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, 3(3):1102–1123, September 2009.

[19] Adam Duncan, Eric Klassen, and Anuj Srivastava. Statistical shape analysis of simplified neuronal trees. *Ann Appl Stat, to appear*, 2018.

[20] A. Feragen, S. Hauberg, M. Nielsen, and F. Lauze. Means in spaces of tree-like shapes. In *ICCV*, 2011.

[21] A. Feragen, M. Owen, J. Petersen, M.M.W. Wille, L.H. Thomsen, A. Dirksen, and M. de Bruijne. Tree-space statistics and approximations for large-scale analysis of anatomical trees. In *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, pages 74–85, 2013.

[22] A. Feragen, J. Petersen, M. Owen, P. Lo, L. H. Thomsen, M. M. W. Wille, A. Dirksen, and

M. de Bruijne. A hierarchical scheme for geodesic anatomical labeling of airway trees. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2012*, Lecture Notes in Computer Science, pages 147–155, 2012.

[23] Aasa Feragen. Complexity of computing distances between geometric trees. In *Structural, Syntactic, and Statistical Pattern Recognition - Joint IAPR International Workshop, SSPR&SPR 2012, Hiroshima, Japan, November 7-9, 2012. Proceedings*, pages 89–97, 2012.

[24] Aasa Feragen, Sean Cleary, Megan Owen, and Daniel Vargas. On tree-space PCA. In *Mini-Workshop: Asymptotic Statistics on Stratified Spaces*, volume 44 of *Mathematisches Forschungsinstitut Oberwolfach, Report No. 44/2014*, pages 2491–2495, 2014.

[25] Aasa Feragen, Pechin Lo, Marleen de Bruijne, Mads Nielsen, and François Lauze. Toward a theory of statistical tree-shape analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):2008–2021, 2013.

[26] Aasa Feragen, Jens Petersen, Megan Owen, Pechin Lo, Laura Hohwu Thomsen, Mathilde Marie Winkler Wille, Asger Dirksen, and Marleen de Bruijne. Geodesic atlas-based labeling of anatomical trees: Application and evaluation on airways extracted from CT. *IEEE Trans. Med. Imaging*, 34(6):1212–1226, 2015.

[27] P. Thomas Fletcher and Sarang Joshi. *Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors*, pages 87–98. 2004.

[28] Alex Gavryushkin and Alexei J Drummond. The space of ultrametric phylogenetic trees. *J. Theor. Bio.*, 403:197–208, 2016.

[29] David Groisser, Sungkyu Jung, and Armin Schwartzman. Geometric foundations for scaling-rotation statistics on symmetric positive definite matrices: Minimal smooth scaling-rotation curves in low dimensions. *Electron. J. Statist.*, 11(1):1092–1159, 2017.

[30] M. Gromov. Hyperbolic groups. In *Essays in group theory*, volume 8 of *Math. Sci. Res. Inst. Publ.*, pages 75–263. Springer, 1987.

[31] Mehmet A. Gülsün, Gareth Funka-Lea, Yefeng Zheng, and Matthias Eckert. CTA coronary labeling through efficient geodesics between trees using anatomy priors. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2014 - 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part II*, pages 521–528, 2014.

[32] Thomas Hotz, Stephan Huckemann, Huiling Le, J. S. Marron, Jonathan C. Mattingly, Ezra Miller, James Nolen, Megan Owen, Vic Patrangenaru, and Sean Skwerer. Sticky central limit theorems on open books. *Ann. Appl. Probab.*, 23(6):2238–2258, 12 2013.

[33] S. Huckemann, T. Hotz, and A. Munk. Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica*, 20(1):1–58, 2010.

[34] B. J. Jain and K. Obermayer. Structure spaces. *J. Mach. Learn. Res.*, 10:2667–2714, 2009.

[35] Brijnesh J. Jain. Generalized gradient learning on time series. *Machine Learning*, 100(2):587–608, 2015.

[36] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5):509–541, 1977.

[37] Florian Kelma. *Projective Shapes: Topology and Means*. PhD thesis, TU Ilmenau, 2017.

[38] D. G. Kendall. Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.*, 16(2):81–121, 1984.

[39] Eric Kolaczyk, Lizhen Lin, Steven Rosenberg, and Jackson Walters. Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *arXiv preprint, arXiv:1709.02793*, 2017.

[40] P. Lo, J. Sporring, H. Ashraf, J.J.H. Pedersen, and M. de Bruijne. Vessel-guided airway tree segmentation: A voxel classification approach. *Medical Image Analysis*, 14(4):527–538, 2010.

[41] M. E. Martinez-Perez, A. D. Highes, A. V. Stanton, S. A. Thorn, N. Chapman, A. A. Bharath, and K. H. Parker. Retinal vascular tree morphology: a semi-automatic quantification. *IEEE Transactions on Biomedical Engineering*, 49(8):912–917, 2002.

[42] E. Miller, M. Owen, and J. S. Provan. Polyhedral computational geometry for averaging metric phylogenetic trees. *Adv. Appl. Math.*, 68:51–91, 2015.

[43] Nina Miolane, Susan Holmes, and Xavier Pennec. Template shape estimation: Correcting an asymptotic bias. *SIAM J. Imaging Sciences*, 10(2):808–844, 2017.

[44] T.M.W. Nye. Principal components analysis in the space of phylogenetic trees. *Ann. Stat.*, 39:2716–2739, 2011.

[45] T.M.W. Nye, X. Tang, G. Weyenberg, and R. Yoshida. Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees. *Biometrika*, 104:901–922, 2017.

[46] Tom M. W. Nye. An algorithm for constructing principal geodesics in phylogenetic treespace. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 11(2):304–315, 2014.

[47] M. Owen and J. S. Provan. A fast algorithm for computing geodesic distances in tree space. *ACM/IEEE Transactions on Computational Biology and Bioinformatics*, 8(1):2–13, 2011.

[48] Victor Patrangenaru and Leif Ellingson. *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 2015.

[49] J. Pedersen, H. Ashraf, A. Dirksen, K. Bach, H. Hansen, P. Toennesen, H. Thorsen, J. Brodersen, B. Skov, M. Døssing, J. Mortensen, K. Richter, P. Clementsen, and N. Seersholm. The Danish randomized lung cancer CT screening trial - overall design and results of the prevalence round. *Journal of Thoracic Oncology*, 4(5):608–614, May 2009.

[50] Xavier Pennec. Barycentric subspace analysis on manifolds. *arXiv preprint, arXiv:1607.02833*, 2016.

[51] Markus J. Pflaum. *Analytic and geometric study of stratified spaces*, volume 1768 of *Lecture notes in mathematics*. Springer, 2001.

[52] Sean Skwerer, Elizabeth Bullitt, Stephan Huckemann, Ezra Miller, Ipek Oguz, Megan Owen, Vic Patrangenaru, Scott Provan, and J.S. Marron. Tree-oriented analysis of brain artery structure. *Journal of Mathematical Imaging and Vision*, 50:126–143, 2014.

[53] Sean Skwerer, Scott Provan, and J.S. Marron. Relative optimality conditions and algorithms for rreespace Fréchet means. *SIAM Journal on Optimization*, 28(2):959–988, 2018.

[54] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. volume 338 of *Contemp. Math.*, pages 357–390. 2003.

[55] Asuka Takatsu. Wasserstein geometry of gaussian measures. *Osaka J. Math.*, 48(4):1005–1026, 12 2011.

[56] K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet means for distributions of persistence diagrams. *Disc. Comp. Geom.*, 52(1):44–70, 2014.

[57] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. *European Conference on Computer Vision (ECCV)*, pages 589–600, 2006.

[58] G. Weyenberg, R. Yoshida, and D. Howe. Normalizing kernels in the billera-holmes-vogtmann treespace. *IEEE ACM T. Comput. Bi.*, page doi:10.1109/TCBB.2016.2565475, 2016.

[59] Grady Weyenberg, Peter M Huggins, Christopher L Schardl, Daniel K Howe, and Ruriko Yoshida. KDEtrees: non-parametric estimation of phylogenetic tree distributions. *Bioinformatics*, 30(16):2280–2287, 2014.

[60] A. Willis. Confidence sets for phylogenetic trees. *J. Am. Stat. Assoc.*, 2018.

[61] Sakellarios Zairis, Hossein Khiabanian, Andrew J. Blumberg, and Raul Rabadan. Genomic data analysis in tree spaces. *arXiv preprint, arXiv:1607.07503*, 2016.

[62] H. Zhai. *Principal component analysis in phylogenetic tree space*. PhD thesis, University of North Carolina at Chapel Hill, 2016.

ACKNOWLEDGMENTS

he authors wish to thank the editors and the anonymous reviewers for their insightful feedback, which helped greatly improve the quality of the paper. Aasa Feragen was supported by the Lundbeck Foundation, as well as by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.