GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN IN PUBLICA COMMODA
SEIT 1737

Phylogenetic Forest Spaces

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

“Doctor rerum naturalium”

der Georg-August-Universität Göttingen

im Promotionsprogramm “Mathematical Sciences”

der Georg-August University School of Science (GAUSS)

vorgelegt von

Jonas Lueg

aus Hameln

Göttingen, 2022

Betreuungsausschuss:

Prof. Dr. Stephan F. Huckemann
Felix-Bernstein-Institut für Mathematische Statistik in den Biowissenschaften,
Georg-August-Universität Göttingen

Prof. Dr. D. Russell Luke
Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Prof. Dr. Max Wardetzky
Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Mitglieder der Prüfungskommission:

Referent:
Prof. Dr. Stephan F. Huckemann
Felix-Bernstein-Institut für Mathematische Statistik in den Biowissenschaften,
Georg-August-Universität Göttingen

Korreferent:
Senior Lecturer Dr. Tom M. W. Nye
School of Mathematics, Statistics and Physics, Newcastle University

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. D. Russell Luke
Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Prof. Dr. Thomas Schick
Mathematisches Institut, Georg-August-Universität Göttingen

Jun.-Prof. Dr. Anne Wald
Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Prof. Dr. Max Wardetzky
Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 04. Oktober 2022

Acknowledgements

First of all, I would like to express utmost my gratitude to Stephan Huckemann, who has accompanied and supervised me over the course of the last four years, including my work on my master's thesis. I also thank him for working with me, for his hospitality and for sharing his many experiences of life with me.

Second, I would like to thank Tom Nye for inviting me to Newcastle for a one month visit, for his hospitality and for sharing his views with me. The experiences I made during this visit shaped me and influenced my perspective on many things.

Furthermore, I would like to thank Russell Luke for supervising me and having me as a frequent member of his Oberseminar.

Finally, I would like to thank all the colleagues and friends that made up a big portion of my everyday life with all its struggles and joyful moments in the last three and a half years. Every second at the Institut für Mathematische Stochastik I have felt at home and this was due to the wonderful, kind and fascinating people that I have met and befriended there – among those friendships many that will last for a long time. Last but not least, I would like to thank my long-standing friends and my closest, that is my family: my parents Gerda and Dirk, as well as my siblings Marie, Hannes and Paul.

Preface

Given a fixed set of operational taxonomic units (OTUs), when studying their evolutionary relationships among each other, a common way of depicting those is to construct a phylogenetic tree, a graph-theoretic tree with a root, that is the common ancestor (which one assumes to exist). Herein each leaf represents a unique OTU and edges have real positive lengths representing the evolutionary distance of some kind, which can be time, but also other metrics.

Nowadays, those trees are generated from multiple measured sequence data such as DNA, RNA or protein sequences, and having various measurements from the same species, one usually obtains many different proposals for phylogenetic trees representing the evolutionary relationships. As one believes that there exists one true phylogenetic tree, one wishes to combine the different proposals into a single phylogenetic tree. As a statistician, the natural thing to do is to average the proposals in some sense in order to obtain a mean, that is again a phylogenetic tree, and which allows for further statistical methodology, such as confidence regions or testing. However, this requires some appropriate mathematical structure on the set of phylogenetic trees, preferable with geometry, i.e. at least a metric space.

The Billera-Holmes-Vogtmann tree space (BHV space) is an example for such a structure, it is a metric space that is CAT(0) and furthermore a Riemann stratified space of type B, cf. Billera et al. (2001). This space has favorable properties like completeness and global non-positive curvature, which implies that between two points, there exists a unique geodesic connecting them. Furthermore, there is a polynomial time algorithm to compute exact geodesics in this space, cf. Owen & Provan (2011). Nonetheless, the space is artificially constructed from embedding it in a high dimensional Euclidean space and as a consequence does not behave as biological understanding would expect a metric space of phylogenetic trees to behave, as is discussed for example in Garba et al. (2018). They consider the distance between two trees with different structures in the case that the edge lengths go to infinity. The same example motivates one to actually include disconnected forests into the space.

A promising construction that covers the example from above and behaves as biological intuition suggests is our recently introduced Wald Space, cf. Garba et al. (2021a). It is based on the characterization of phylogenetic trees as covariance matrices, which is a byproduct of a generalization of the popular biological substitution models that are used to calculate likelihoods for trees given genetic sequence data, so those substitution models are the backbone of phylogenetic tree estimation. In other words, the Wald Space is a space that is consistent with the tree estimation methods that are currently used, up to the generalizations that have been made. More details can be found in Garba et al. (2021a).

In this work, we concentrate on the Wald Space purely from the perspective that it is a mathematical structure and thus we try to enable for a better understanding of the Wald Space. To this end, we introduce the mathematical structures required: metric spaces, Riemannian manifolds, Riemann stratified spaces, as well as, for our construction essential, various geometries on the manifold of strictly positive definite symmetric real matrices. Furthermore, we introduce various possible ways to represent the phylogenetic trees and forests that we consider. Having finished the introduction of the more general and known concepts, we briefly introduce the BHV Space. Then we define and describe the Wald Space, which is a topological stratified space, and we investigate its topological features. This part is the core of the thesis. Finally, we equip the Wald Space with a geometry that can be chosen to some degree and find that these spaces are then Riemann stratified spaces of type (A). Last but not least, we propose some numerical algorithms to calculate geodesics and distances in the Wald Space equipped with a geometry. Those are not tested in this work, but to some extent in Lueg et al. (2021).

Contents

Chapter 1

Introduction and Motivation

Among others, the subject of evolutionary biology is concerned with the origin of life on Earth and with the underlying principles that cause the diversification or evolution of the species over time. One of the fundamental assumptions or beliefs is the one of *common descent*, that is that all living things stem from one universal common ancestor. This assumption was already formulated by Charles Darwin in 1859 in his famous and at that time controversial publication *On the Origin of Species by Means of Natural Selection, Or The Preservation of Favoured Races in the Struggle for Life* (Darwin, 1859). The hypothesis of the common descent was also the subject of interest in recent publications such as Theobald (2010) and Weiss et al. (2016).

The hypothesis of the universal common ancestor is one of the reasons for representing the biological evolution and diversification of species as well as their evolutionary relationships via a *phylogenetic tree*. Usually, a phylogenetic tree consists of a *root vertex* and from it originating branches or *edges* that themselves branch again several times (the points at which they branch are called *interior vertices*), until they terminate in a vertex (called *leaf*). The leaves represent present-day species and the interior vertices represent common ancestors, we think of them as extinct or unobserved. The branching process represents a speciation event, i.e. new species evolve from a previous one. Finally, each edge has a length, that is a positive real number, describing the time or evolutionary distance between the incident vertices.

Up until the beginning of the second half of the 20th century, phylogenetic trees were mainly inferred from the morphological traits of the considered species. In fact, Ernst Haeckel predicted in 1866 that "building phylogenetic trees will be the most important and most interesting task of future morphology", cf. Haeckel (1866, p. xx, translated). However, in the second half of the 20th century, biologists developed computational methods to infer phylogenetic trees from measured biological sequence data of the respective species. The biological sequence data can be DNA (Deoxyribonucleic acid) or RNA (Ribonucleic acid), but also other sequences of large molecules such as protein sequences.

These methods, among others, are substitution models that consist of Markov models on trees that describe evolutionary change over time, and based on these methods, phylogenetic trees can be estimated based on the sequence data, e.g. Felsenstein (1981). However, due to these differing computational methods, often hundreds of trees are obtained as proposals for the true phylogenetic tree for the same set of species. This suggests to construct a sample space that is at least a metric space, such that one can then infer statistics on the set of trees, such as averaging (e.g. calculating the Fréchet mean) or confidence regions.

The Billera-Holmes-Vogtmann tree space (BHV space), cf. Billera et al. (2001), is the first metric space that serves as a sample space for phylogenetic trees where statistics can be performed, and is under ongoing investigation to this day (Owen & Provan (2011); Nye (2011); Barden et al. (2013); Nye (2014); Nye et al. (2016); Barden et al. (2016); Anaya et al. (2020), to name a few).

Furthermore, there exists tropical tree space which is based on representing each tree uniquely via its distance matrix and equipping these matrices with a very special vector space structure, cf. Lin et al. (2017, 2018); Yoshida et al. (2019).

However, both these spaces do not behave as one would expect when edges tend to infinite length, and this is due to the choice of representation and the way those spaces are constructed. In Kim (2000); Moulton & Steel (2004); Shiers et al. (2016), the idea of constructing a tree space based on representing the trees via correlation matrices or vectors was developed and further investigated, and this representation of trees is consistent with the substitution models used to estimate phylogenetic trees in the first place. This work then lead to the idea of our contribution called the *Wald Space*, first introduced in Garba et al. (2021a), and we made some numerical experiments in Lueg et al. (2021). In Garba et al. (2021a), there is a detailed introduction that motivates the Wald Space, and the construction of the Wald Space is given. Therefore, in this thesis, we do not focus on the motivation or justification of the construction of the Wald Space, but rather on the construction and the properties of the Wald Space itself.

Importantly, in this thesis, contrary to Garba et al. (2021a), we think of the Wald Space solely as the underlying stratified topological space. This modular formulation allows us then to flexibly equip the Wald Space with a metric which is induced by our choice of a Riemannian metric on the manifold of strictly positive definite symmetric real $N \times N$-dimensional matrices, which we abbreviate with $\mathcal{P}$ throughout the thesis. Choosing the Fisher-information metric on the statistical manifold of zero-mean multivariate Gaussians with covariance matrices in $\mathcal{P}$ then leads to the classical Wald Space that is discussed in Garba et al. (2021a).

To summarize, we start introducing in Chapter 2 the mathematical structures that we use throughout the thesis, e.g. metric spaces, manifolds and Riemann stratified spaces, and in Chapter 3 various Riemannian metrics for the manifold of strictly positive definite symmetric real $N \times N$-dimensional matrices $\mathcal{P}$, both of those chapters can be viewed as prerequisites that we need throughout the thesis. In Chapter 4, we then focus on the various representations of phylogenetic forests which are the elements of the Wald Space, and we refer to the elements of the Wald Space as *wälder* (singular *wald*, which is German for *forest*, and the plural is *wälder* and means *forests*). Notably, the word wälder is read *velder*, like the English word *elder*, but with a "v" in front. Each representation of the wälder has its advantages and disadvantages when it comes to proving or formulating statements of various kinds about the Wald Space, as each of these representations allows for viewing wälder from a different perspective. In Chapter 5, we introduce the BHV Space and in Chapter 6, the actual work of this thesis begins, that is the topological investigation of the Wald Space, where we benefit from the carefully and precisely designed notation and concepts that we developed in Chapter 4. Chapter 6 contains almost all the important theoretical results about Wald Space, and in Chapter 7, we show that in general, one can choose any Riemannian metric on $\mathcal{P}$ and obtains a well-defined induced metric on the Wald Space. We introduce various geometries on the Wald Space and refer to the Wald Space equipped with the geometry in Garba et al. (2021a) as the *Schwarzwald Space*, as the Schwarzwald in Germany was the place where the idea of the Wald Space first came up. Finally, in Chapter 8, we introduce some algorithms, which are already implemented. In Chapter 9 we summarize the contents of this work, as well as the contributions and the implementations that are partly already published, and pose some open questions that might be interesting to pursue in future work.

1.1 Notation

- $\mathbb{N} = \{1, 2, \dots\}$ are the natural numbers.
- $\mathbb{R}$ are the real numbers.
- For a finite set A, $\mathbb{R}^A$ contains all maps $\lambda\colon A \to \mathbb{R}$, sometimes we write elements $\lambda \in \mathbb{R}^A$ as vectors $\lambda = (\lambda_a)_{a \in A}$.

- We also allow $A = \emptyset$ and define $\mathbb{R}^\emptyset$ to contain exactly one element, which we do not describe explicitly, but in vector notation we would sometimes write $() \in \mathbb{R}^\emptyset$.
- For two sets A, B, the notation $A \subset B$ means A is a subset of B, which includes the possibility that A equals B. Sometimes, $A \subseteq B$ is used, where the author thinks it is important to highlight that A might be equal to B.
- For a matrix $X \in \mathbb{R}^{n \times n}$, $n \in \mathbb{N}$, the expression $\operatorname{diag}(X) \in \mathbb{R}^n$ is the diagonal of X as a vector.
- However, the expression $\operatorname{Diag}(X) \in \mathbb{R}^{n \times n}$ is the matrix that equals X on its diagonal and is zero elsewhere.

Chapter 2

Metric Spaces, Riemannian Manifolds and Stratified Spaces

We introduce the notion of metric spaces, Riemannian manifolds, Riemannian submersions and Riemann stratified spaces. We will also have a closer look at the topologies of the respective spaces.

2.1 Metric Spaces

Definition 2.1.1 (Metric space)**.** A *metric space* is a tuple $(\mathcal{M}, d)$, where $\mathcal{M}$ be a set and $d\colon \mathcal{M} \times \mathcal{M} \to [0, \infty]$ is a map such that for all $p, q, r \in \mathcal{M}$ we have

- $d(p, q) = 0 \iff p = q$, (identity of indiscernibles)
- $d(p, q) = d(q, p)$ and (symmetry)
- $d(p, q) \leq d(p, r) + d(r, q)$. (triangle inequality)

Furthermore, a metric space is also a topological space, where the topology is generated by all open balls around $p \in M$ with radius $\varepsilon > 0$:

$$B_{\mathcal{M},d,\varepsilon}(p) \coloneqq \{q \in \mathcal{M}\colon d(p, q) < \varepsilon\},$$

and note that d is continuous in this topology.

Definition 2.1.2 (Isometry)**.** Let $(\mathcal{M}, d)$ and $(\mathcal{M}', d')$ be two metric spaces. A map $f\colon \mathcal{M} \to \mathcal{M}'$ is an *isometry*, if it satisfies $d(p, q) = d'(f(p), f(q))$ for all $p, q \in \mathcal{M}$. If f is furthermore bijective, it is called *isometric isomorphism*, and in this case, $(\mathcal{M}, d)$ and $(\mathcal{M}', d')$ are called *isometric*.

Lemma 2.1.3 (Induced intrinsic metric). *Let $(\mathcal{M}, d)$ be a metric space. The* induced intrinsic metric *of $(\mathcal{M}, d)$ is the map $d^*: \mathcal{M} \times \mathcal{M} \to [0, \infty]$, given for all $p, q \in \mathcal{M}$ by*

$$d^*(p, q) = \inf_{\substack{\gamma\colon [0,1]\to\mathcal{M} \\ \gamma(0)=p,\gamma(1)=q \\ \gamma \text{ continuous}}} L_d(\gamma),$$

where $L_d(\gamma)$ is the length of γ measured using d, i.e.

$$L_d(\gamma) = \sup_{\substack{0=t_0\le t_1\le\ldots\le t_n=1 \\ n\in\mathbb{N}}} \sum_{i=0}^{n-1} d\big(\gamma(t_i), \gamma(t_{i+1})\big).$$

If no such path connecting p and q exists, we set $d^*(p, q) := \infty$.

Note that for some continuous path $\gamma\colon [0, 1] \to \mathcal{M}$ we might have $L_d(\gamma) = \infty$, or there need not exist a continuous path from p to q, so $d^*(p, q) = \infty$ is generally possible.

Proof. We show that d^* satisfies all three metric properties:

1. Using the triangle inequality of d, we find $d^*(p, q) \ge d(p, q)$ for all $p, q \in \mathcal{M}$, and thus $d^*(p, q) = 0 \implies d(p, q) = 0 \implies p = q$. If $p = q$, plugging in the constant path $\gamma\colon [0, 1] \to \mathcal{M}$ with $\gamma(t) = p$ for $t \in [0, 1]$, we find $0 \le d^*(p, p) \le L_d(\gamma) = 0$. Thus $d^*(p, q) = 0 \iff p = q$.
2. Symmetry follows directly from the symmetry of d.
3. For the triangle inequality, let $p, q, r \in \mathcal{M}$ let $(\gamma_k)_{k\in\mathbb{N}}$ with $\gamma_k\colon [0, 1] \to \mathcal{M}$ and $(\gamma'_k)_{k\in\mathbb{N}}$ with $\gamma'_k\colon [0, 1] \to \mathcal{M}$ be sequences of paths from p to r and from r to q, respectively, such that $\lim_{k\to\infty} L_d(\gamma_k) = d^*(p, r)$ and $\lim_{k\to\infty} L_d(\gamma'_k) = d^*(r, q)$ holds true. We concatenate γ_k and γ'_k to obtain a new path from p to q:

$$\gamma''_k(t) = \begin{cases} \gamma_k(2t), & t \in [0, \frac{1}{2}), \\ \gamma'_k(2t-1), & t \in [\frac{1}{2}, 1]. \end{cases}$$

   Thus

$$d^*(p, q) \le L_d(\gamma''_k) = L_d(\gamma_k) + L_d(\gamma'_k) \longrightarrow d^*(p, r) + d^*(r, q).$$

■

Note that the topology on $\mathcal{M}$ induced by d^* is finer or equal than the one induced by d since $d^* \ge d$ implies that every d-open ball is also d^*-open.

Definition 2.1.4 (Length spaces)**.** A metric space $(\mathcal{M}, d)$ is called *length space*, if $d = d^*$.

Definition 2.1.5 (Geodesics in metric spaces)**.** Let $(\mathcal{M}, d)$ be a metric space. We call a curve (continuous map) $\gamma\colon I \to \mathcal{M}$ from an interval $I \subset \mathbb{R}$ into $\mathcal{M}$ a *geodesic* if there exists a constant $c > 0$ such that locally for any $t, t' \in I$ we have

$$d\big(\gamma(t), \gamma(t')\big) = c|t - t'|.$$

If the above property holds for all $t, t' \in I$ (i.e. globally), then we say that γ is a *minimizing geodesic* or *shortest path*. We say that γ has *natural parametrization* if $c = 1$.

Let $a, b \in \mathbb{R}$ with $a < b$. Note that for a shortest path $\gamma\colon [a, b] \to \mathcal{M}$ with natural parametrization between $p = \gamma(a)$ and $q = \gamma(b)$ we have

$$b - a = d(\gamma(a), \gamma(b)) = d(p, q) \leq d^*(p, q) \leq L_d(\gamma) = b - a,$$

and thus the term minimizing geodesic or shortest path is justified. As this property can also be shown locally, geodesics in a metric space are locally minimizing curves.

The following definition is from Bridson & Haefliger (1999, Definition 1.3).

Definition 2.1.6 (Geodesic metric space)**.** Let $(\mathcal{M}, d)$ be a metric space. Then we say that it is

1. *geodesic metric space* or *geodesic space* if every two points in $\mathcal{M}$ are joined by a geodesic;
2. *uniquely geodesic* if there is exactly one geodesic joining p to q, for all $p, q \in \mathcal{M}$;
3. r*-geodesic* if for every pair of points $p, q \in \mathcal{M}$ with $d(p, q) < r$ there is a geodesic joining p to q.

It is immediate that a geodesic metric space is a length space as the length of a geodesic equals the distance between its end points. We introduce some notions that are needed in the next theorems.

Definition 2.1.7. Let $(\mathcal{M}, d)$ be a metric space.

1. A topological space is called *locally compact*, if each point has a locally compact neighborhood.
2. A sequence $(x_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ is called *Cauchy sequence in* $(\mathcal{M}, d)$, if for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n, m \geq N$ we have $d(x_n, x_m) < \varepsilon$.

3. The metric space $(\mathcal{M}, d)$ is called complete, if all Cauchy sequences in $\mathcal{M}$ converge within $\mathcal{M}$.

The following theorem shows under which assumptions a metric space is a geodesic space. It is taken from Bridson & Haefliger (1999, Proposition 3.7).

Theorem 2.1.8 (Hopf-Rinow for metric spaces)**.** *Let $(\mathcal{M}, d)$ be a length space. If $(\mathcal{M}, d)$ is complete and $\mathcal{M}$ locally compact as a topological space, then $(\mathcal{M}, d)$ is a geodesic space.*

An induced intrinsic metric inherits completeness under certain circumstances. This result can be found in Hu & Kirk (1978, p.123).

Theorem 2.1.9. *Let $(\mathcal{M}, d)$ be a complete metric space for which each two points are connected by a path that has finite length. Then the space $(\mathcal{M}, d^*)$ is also complete.*

Curvature in Metric Spaces

We introduce a notion of curvature in metric spaces. For simplicity and since we do not need other notions, we will introduce CAT(κ) spaces only for the special case $\kappa = 0$. The following is a consequence of Bridson & Haefliger (1999, Lemma 2.14).

Lemma 2.1.10. *Let p, q, r be three points in a metric space $(\mathcal{M}, d)$. Then there exist points $\overline{p}, \overline{q}, \overline{r} \in \mathbb{R}^2$ such that $d(p, q) = d_E(\overline{p}, \overline{q})$, $d(q, r) = d_E(\overline{q}, \overline{r})$ and $d(r, p) = d_E(\overline{r}, \overline{p})$, where $d_E(x, y) = \|x - y\|_2$ is the Euclidean distance on $\mathbb{R}^2$. The triangle $\overline{\Delta}(\overline{p}, \overline{q}, \overline{r}) \subset \mathbb{R}^2$ (the set of points on the triangle with vertices $\overline{p}, \overline{q}, \overline{r}$ and sides $[\overline{p}, \overline{q}]$, $[\overline{q}, \overline{r}]$ and $[\overline{r}, \overline{p}]$) is unique up to an isometry on $\mathbb{R}^2$.*

We call such a triangle $\overline{\Delta}$ a *comparison triangle* for the triple (p, q, r).

Definition 2.1.11 (Geodesic triangle)**.** Let $(\mathcal{M}, d)$ be a metric space.

1. A *geodesic triangle* $\Delta \subset \mathcal{M}$ consists of three points $p, q, r \in \mathcal{M}$, its *vertices*, and a choice of three geodesic segments $[p, q]$, $[q, r]$, $[r, p]$ joining them, its *sides* (where $[p, q]$ corresponds to the points on a geodesic between p and q).
2. For a geodesic triangle Δ with vertices $p, q, r \in \mathcal{M}$ and a comparison triangle $\overline{\Delta}$ (with vertices $\overline{p}$, $\overline{q}$ and $\overline{r}$) for (p, q, r), a point $\overline{x} \in [\overline{p}, \overline{q}]$ is called a *comparison point* for $x \in [p, q]$ if $d(p, x) = d_E(\overline{p}, \overline{x})$, and analogously for points in $[q, r]$ and $[r, p]$.

3. Let $\Delta \subset \mathcal{M}$ be a geodesic triangle with vertices $p, q, r \in \mathcal{M}$ and let $\overline{\Delta}$ be a comparison triangle for (p, q, r). Then, Δ is said to satisfy the CAT(0) *inequality* if for all $x, y \in \Delta$ and all comparison points $\overline{x}, \overline{y} \in \overline{\Delta}$,

$$d(x, y) \leq d_E(\overline{x}, \overline{y}).$$

Definition 2.1.12. A metric space $(\mathcal{M}, d)$ is called CAT(0) *space*, if $(\mathcal{M}, d)$ is a geodesic space and all its geodesic triangles satisfy the CAT(0) inequality. If $\mathcal{M}$ is additionally complete, we call $(\mathcal{M}, d)$ *Hadamard space.*

Theorem 2.1.13. *In a CAT(0) space $(\mathcal{M}, d)$ there is a unique geodesic segment joining each pair of points $x, y \in \mathcal{M}$ and $\mathcal{M}$ is contractible and simply connected.*

Proof. Cf. Bridson & Haefliger (1999, Chapter II.1, Proposition 1.4 and Corollary 1.5). ■

2.2 Smooth Manifolds

In this section, the basic definitions and notions of the theory of manifolds are introduced. A *manifold of dimension* $m \in \mathbb{N}$ is essentially a topological space that is locally homeomorphic to an open subset of $\mathbb{R}^m$. Most spaces that we deal with are manifolds, for example vector spaces, spheres, the torus, open subsets of $\mathbb{R}^m$, and many more. Manifolds of varying dimension are also the building blocks of *stratified spaces* and *Whitney stratified spaces.* This section is based on Lee (2003).

In Lee (2003), a manifold is a topological space that is required to be

- *second countable*: a topological space $\mathcal{T}$ is *second countable*, if it has a countable base, that is if there exists a collection $\{U_i\}_{i\in\mathbb{N}}$ of open subsets of $\mathcal{T}$ such that any open subset of $\mathcal{T}$ can be written as a union of elements of some subset of $\{U_i\}_{i\in\mathbb{N}}$;
- *Hausdorff*: a topological space $\mathcal{T}$ is *Hausdorff*, if for every two points $p, q \in \mathcal{T}$, there exist disjoint open subsets $U, V \subset \mathcal{T}$ such that $p \in U$ and $q \in V$.

As stated in Lee (2003, p.3), the Hausdorff property ensures for example that one-point sets are closed, as well as that limits of convergent sequences are unique, which we use extensively throughout later chapters.

Definition 2.2.1 (Manifold)**.** Let $\mathcal{M}$ be a topological space. $\mathcal{M}$ is a *topological manifold* of dimension m, if $\mathcal{M}$ is a Hausdorff space, second countable and locally Euclidean of dimension m, that is, every point $p \in \mathcal{M}$ has a neighborhood $p \in U \subset \mathcal{M}$ that is homeomorphic to an open subset of $\mathbb{R}^m$.

These homeomorphisms, denoted by $\varphi\colon U \to \varphi(U) \subset \mathbb{R}^m$, are called *charts* (more precisely, the pair (U, φ) is called chart). Loosely speaking, the idea of a chart is that whenever you "stand" somewhere in the topological space $\mathcal{M}$, there is no concept of direction, orientation or distance, but when you use a chart and translate your location $p \in \mathcal{M}$ into your coordinates, that is $\varphi(p) \in \mathbb{R}^m$, you can then move (at least locally) in a direction, or even measure a distance, and then translate back into the abstract topological space $\mathcal{M}$. Thus, a chart gives you a local "map" (local coordinates) that you can use to move in $\mathcal{M}$.

The following terminology is taken from Lee (2003, p.7ff). Two charts (U, φ), (V, ψ) are called *smoothly compatible*, if either $U \cap V = \emptyset$ or the *transition map* $\psi \circ \varphi^{-1}\colon \varphi(U \cap V) \to \psi(U \cap V)$ is a diffeomorphism, that is bijective, smooth (continuous partial derivatives of all orders exist) and smooth inverse. An *atlas* for $\mathcal{M}$ is a collection of charts whose domains cover $\mathcal{M}$, and an atlas $\mathcal{A}$ is called a *smooth atlas*, if any two charts in $\mathcal{A}$ are smoothly compatible. Such a smooth atlas is called *maximal* or *complete*, if it is not contained in a strictly larger smooth atlas. A *smooth structure* on a topological manifold $\mathcal{M}$ is a maximal smooth atlas, and any chart $(U, \varphi) \in \mathcal{A}$ will be called *smooth chart*.

Definition 2.2.2 (Smooth manifold)**.** A *smooth manifold* is a pair $(\mathcal{M}, \mathcal{A})$, where $\mathcal{M}$ is a topological manifold of dimension m and $\mathcal{A}$ is a smooth structure on $\mathcal{M}$.

If the smooth structure is clear, it is usually omitted and we say that $\mathcal{M}$ is a smooth manifold. Note that this definition implies that the topology of a smooth manifold $(\mathcal{M}, \mathcal{A})$ is characterized via

$$B \subset \mathcal{M} \text{ open} \iff \varphi(B \cap U) \subset \mathbb{R}^m \text{ open for all } (U, \varphi) \in \mathcal{A}.$$

If $\mathcal{M}$ is a smooth manifold, a function $f\colon \mathcal{M} \to \mathbb{R}^k$ is said to be *smooth* if, for every smooth chart (U, φ) on $\mathcal{M}$, the composite function $f \circ \varphi^{-1}$ is smooth on $\varphi(U) \subset \mathbb{R}^n$. Note that the technical definition of a smooth structure, that is a maximal atlas, is to avoid having several possible smooth structures that induce the same functions $f\colon \mathcal{M} \to \mathbb{R}^k$ to be smooth. This could be achieved in a similar way through a quotient, but via a maximal atlas, we can avoid this extra construction. The practicability is guaranteed by Lee (2003, Lemma 1.4) which states that every smooth atlas for $\mathcal{M}$ is contained in a unique maximal smooth atlas, and by Lee (2003, Lemma 2.1), which states that, in order to show that a function $f\colon \mathcal{M} \to \mathbb{R}^k$ is smooth, it suffices to show that $f \circ \varphi^{-1}$ is smooth on $\varphi(U) \subset \mathbb{R}^n$ for all charts (U, φ) of *some* smooth atlas for $\mathcal{M}$ (this means that the smooth atlas could consist of just one global chart).

We extend the notion of smoothness to maps between manifolds. The following definition is from Lee (2003, p.24)

Definition 2.2.3 (Smooth map between manifolds)**.** For two smooth manifolds $\mathcal{M}_1, \mathcal{M}_2$ of dimension m_1, m_2, respectively, a map $F\colon \mathcal{M}_1 \to \mathcal{M}_2$ is called *smooth map* if, for any smooth charts (U, φ) for $\mathcal{M}_1$ and (V, ψ) for $\mathcal{M}_2$, the composite map

$$\psi \circ F \circ \varphi^{-1}\colon \varphi(U \cap F^{-1}(V)) \to \psi(V)$$

is smooth.

In the same way it suffices to show smoothness of a function $f\colon \mathcal{M} \to \mathbb{R}^k$ with respect to some smooth atlas (which need not necessarily be a smooth structure), it suffices to do so for maps between manifolds, cf. Lee (2003, Lemma 2.2).

In the following, we will introduce the notion of tangent vectors on a manifold in the same way as in Lee (2003, Chapter 3) and thereby make the above mentioned concept of "directions" mathematically rigorous. We start with the general notion of derivations that are linear maps that satisfy the product rule, which we then push forward locally via charts into Euclidean spaces, where we can link the abstract idea of a tangent space with the very geometric imagination of directions in $\mathbb{R}^m$. Denote with $C^\infty(\mathcal{M})$ the set of all smooth functions $f\colon \mathcal{M} \to \mathbb{R}$ on a smooth manifold $\mathcal{M}$.

Definition 2.2.4. Let $\mathcal{M}$ be a smooth manifold and let $p \in \mathcal{M}$. A linear map $X\colon C^\infty(\mathcal{M}) \to \mathbb{R}$ is called a *derivation at* p if it satisfies the *Leibniz rule* for all $f, g \in C^\infty(\mathcal{M})$

$$X(fg) = f(p)Xg + g(p)Xf.$$

The set of all derivations of $C^\infty(\mathcal{M})$ at p is a vector space called the *tangent space* to $\mathcal{M}$ at p, and is denoted by $T_p\mathcal{M}$. An element of $T_p\mathcal{M}$ is called a *tangent vector* at p.

Note that for any constant function $f \in C^\infty(\mathcal{M})$, we have by linearity and the Leibniz rule of X that $f(p)Xf = X(ff) = f(p)Xf + f(p)Xf = 2f(p)Xf$, which is equivalent to $f(p)Xf = 0$, such that $Xf = 0$. In general, for a smooth map $F\colon \mathcal{M}_1 \to \mathcal{M}_2$ between two smooth manifolds $\mathcal{M}_1, \mathcal{M}_2$, the *push-forward* associated with F is the map (where $p \in \mathcal{M}_1$)

$$F_*\colon T_p\mathcal{M}_1 \to T_{F(p)}\mathcal{M}_2, \qquad (F_*X)(f) := X(f \circ F), \tag{2.2.1}$$

and it is easy to show that this is well-defined, as well as that F_*X is a derivation at $F(p)$. We also write $(\partial F)_p(X) := F_*X$.

Definition 2.2.5 (Submersions, immersions and embeddings)**.** Let $F\colon \mathcal{M}_1 \to \mathcal{M}_2$ be a smooth map between smooth manifolds. We call F

(a) *immersion*, if F_* is injective at each point $p \in \mathcal{M}_1$;

(b) *submersion*, if F_* is surjective at each point $p \in \mathcal{M}_1$;

(c) *smooth embedding*, if F is an injective immersion that is a homeomorphism onto its image $F(\mathcal{M}_1) \subset \mathcal{M}_2$.

With the concept of a push-forward F_*, we can, loosely speaking, translate directions in $\mathcal{M}_1$ into directions in $\mathcal{M}_2$. Thus, if $\mathcal{M}_2 = \mathbb{R}^{m_2}$, we can translate "abstract" directions in $\mathcal{M}_1$ into directions in $\mathbb{R}^{m_2}$, which we are familiar with, and those maps are given naturally by the charts of $\mathcal{M}_1$, although possibly only locally.

Let (U, φ) be a smooth coordinate chart on $\mathcal{M}$. Then, by Lee (2003, p.47), for any $p \in U$, $\varphi_*\colon T_p\mathcal{M} \to T_{\varphi(p)}\mathbb{R}^m$ is an isomorphism (between vector spaces). However, since $T_{\varphi(p)}\mathbb{R}^m$ has a basis consisting of the derivations $\partial/\partial x^i\big|_{\varphi(p)}$, $i = 1, \dots, m$, we can compute a basis of $T_p\mathcal{M}$ by applying $(\varphi^{-1})_*$:

$$\frac{\partial}{\partial x^i}\bigg|_p = (\varphi^{-1})_* \frac{\partial}{\partial x^i}\bigg|_{\varphi(p)},$$

which is a derivation that acts on a smooth function $f\colon \mathcal{M} \to \mathbb{R}$ by

$$\frac{\partial}{\partial x^i}\bigg|_p f = \frac{\partial}{\partial x^i}\bigg|_{\varphi(p)} (f \circ \varphi^{-1}),$$

which is just the ith partial derivative of the coordinate representation of f at the coordinate representation of p. Therefore, every tangent vector $X \in T_p\mathcal{M}$ can be written uniquely as a linear combination

$$X = \sum_{i=1}^m X^i \frac{\partial}{\partial x^i}\bigg|_p,$$

where $X^1, \dots, X^m \in \mathbb{R}$.

So far, we defined the tangent space $T_p\mathcal{M}$ to $\mathcal{M}$ at single points $p \in \mathcal{M}$, but there is no concept yet on how these tangent spaces at several points are connected or related to each other. This is realized via the tangent bundle.

Definition 2.2.6 (Tangent bundle)**.** Let $\mathcal{M}$ be a smooth manifold. The *tangent bundle* of $\mathcal{M}$, denoted by $T\mathcal{M}$, is the disjoint union of the tangent spaces at all points of $\mathcal{M}$,

$$T\mathcal{M} = \bigsqcup_{p \in \mathcal{M}} T_p\mathcal{M},$$

such that an element of $T\mathcal{M}$ is a pair (p, X), with $p \in \mathcal{M}$ and $X \in T_p\mathcal{M}$.

With this definition alone it is not clear how those tangent spaces are connected. However, by Lee (2003, Lemma 3.12), the tangent bundle $T\mathcal{M}$ is a smooth $2m$-dimensional manifold with a natural topology and smooth structure, such that the projection map $\pi\colon T\mathcal{M} \to \mathcal{M}, (p, X) \mapsto p$ is a smooth map. The charts of $T\mathcal{M}$ are constructed in the following way: for each chart (U, φ) of $\mathcal{M}$, define the chart $(\pi^{-1}(U), \tilde{\varphi})$, where $\tilde{\varphi}$ maps a point $(p, X) \in \pi^{-1}(U)$ onto

$$\tilde{\varphi}(p, X) = \big(\varphi(p)_1, \dots, \varphi(p)_m, X^1, \dots, X^m\big),$$

with $(X^1, \dots, X^m)$ such that $X = \sum_{i=1}^m X^i \frac{\partial}{\partial x^i}\big|_p$ holds, where smooth compatibility of two charts of $T\mathcal{M}$ is shown via smooth compatibility of the original charts of $\mathcal{M}$ and via a smooth coordinate change.

Definition 2.2.7 (Vector fields)**.** A *vector field* is a *smooth section* of the tangent bundle, that is a map $X\colon \mathcal{M} \to T\mathcal{M}$ such that $\pi \circ X = \mathrm{id}_{\mathcal{M}}$ and X is smooth as a map between manifolds. The space of all vector fields is denoted by $\mathcal{T}(\mathcal{M})$.

Note that we use the symbol X now for a vector field instead of just a vector in the tangent space.

In order to be able to describe *geodesics*, that are locally shortest paths (or that are locally straight lines or that have locally zero acceleration (second derivative)) and in that sense the generalization of the Euclidean straight line, we need to introduce *linear connections*. Loosely speaking, a linear connection is an operator that differentiates along vector fields (elements of $\mathcal{T}(\mathcal{M})$), and this concept will be linked to differentiation of vector fields along curves. The following is the definition of a *linear connection* as in Lee (2018, p.51).

Definition 2.2.8 (Linear connections)**.** Let $\mathcal{M}$ be a smooth manifold. A *linear connection* on $\mathcal{M}$ is a map

$$\nabla\colon \mathcal{T}(\mathcal{M}) \times \mathcal{T}(\mathcal{M}) \to \mathcal{T}(\mathcal{M}), \qquad (X, Y) \mapsto \nabla_X Y,$$

that satisfies the following properties:

(a) $\nabla_X Y$ is linear over $C^\infty(\mathcal{M})$ in X: for all $f, g \in C^\infty(\mathcal{M})$, $X_1, X_2, Y \in \mathfrak{T}(\mathcal{M})$,

$$\nabla_{fX_1+gX_2} Y = f\nabla_{X_1} Y + g\nabla_{X_2} Y,$$

(b) $\nabla_X Y$ is linear over $\mathbb{R}$ in Y: for all $a, b \in \mathbb{R}$, $X, Y_1, Y_2 \in \mathfrak{T}(\mathcal{M})$,

$$\nabla_X(aY_1 + bY_2) = a\nabla_X Y_1 + b\nabla_X Y_2,$$

(c) ∇ satisfies the following product rule: for all $f \in C^\infty(\mathcal{M})$, $X, Y \in \mathfrak{T}(\mathcal{M})$,

$$\nabla_X(fY) = f\nabla_X Y + (Xf)Y.$$

Note that due to linearity, one can show that $\nabla_X Y|_p$ depends on Y only in a local neighborhood of $p \in \mathcal{M}$, as well as it depends only on X_p, see e.g. Lee (2018, Lemma 4.1 and 4.2). This motivates to think of $\nabla_X Y|_p$ as the directional derivative of Y at p in the direction of the vector X_p and to also write $\nabla_{X_p} Y$. Using the basis $\partial_i \in \mathfrak{T}(\mathcal{M})$, $i = 1, \ldots, m$, with

$$\partial_i = \left(\frac{\partial}{\partial x^i}\bigg|_p \right)_{p\in\mathcal{M}}, \tag{2.2.2}$$

we can express the connection X, Y and ∇ in local coordinates: write $X = \sum_{i=1}^m X^i \partial_i$, $Y = \sum_{i=1}^m Y^i \partial_i$ (where $X^i, Y^i \in C^\infty(\mathcal{M})$ for $i = 1, \ldots, m$) as well as $\nabla_{\partial_i}\partial_j = \sum_{k=1}^m \Gamma_{ij}^k \partial_k$, where analogously $\Gamma_{ij}^k \in C^\infty(\mathcal{M})$ for all $i, j, k = 1, \ldots, m$. From now on, we use the Einstein summation convention, that is, whenever indices appear in a formula as once as an upper index and once as a lower index, we implicitly sum over this index. Exploiting linearity and the product rule of the linear connection ∇, we find

$$\nabla_X Y = \nabla_{X^i \partial_i} Y^j \partial_j = (XY^k + X^i Y^j \Gamma_{ij}^k)\partial_k. \tag{2.2.3}$$

Note that here, the vector field $X \in \mathfrak{T}(\mathcal{M})$ takes $Y^k \in C^\infty(\mathcal{M})$ as its argument, which means that at each point, the derivation $X_p \in T_p\mathcal{M}$ is applied to Y^k, giving a number, so $XY^k \in C^\infty(\mathcal{M})$. The term $X^i Y^j \Gamma_{ij}^k$ is just a product of maps in $C^\infty(\mathcal{M})$. By Lee (2018, Lemma 4.4), there is a one-to-one correspondence between the m^3 functions Γ_{ij}^k and the linear connections ∇ via (2.2.3).

We follow the section *Vector Fields Along Curves* from Lee (2018, p.55f). In this context, a *curve* is a smooth map $\gamma\colon I \to \mathcal{M}$, where $I \subset \mathbb{R}$ is some interval (where, if the interval has an endpoint, smooth means that we can extend the curve slightly more to an open interval such that it is a smooth map between manifolds. Recall the push-forward $\gamma_* \colon T_t I \to T_{\gamma(t)}\mathcal{M}$, where $T_t I$ is the tangent space of I at t with one-dimensional basis d/dt. For any $t \in I$, the *velocity* $\dot{\gamma}(t)$ defined as the push-forward $\dot{\gamma}(t) := \gamma_*(d/dt) \in T_{\gamma(t)}\mathcal{M}$, which acts on functions by

$$\dot{\gamma}(t)f = \frac{d}{dt}\big(f \circ \gamma\big)(t).$$

Definition 2.2.9 (Vector fields along curves)**.** A *vector field along a curve* $\gamma \colon I \to \mathcal{M}$ is a smooth map $V \colon I \to T\mathcal{M}$ such that $V(t) \in T_{\gamma(t)}\mathcal{M}$ for every $t \in I$. Denote with $\mathfrak{T}(\gamma)$ the space of vector fields along γ.

We call $V \in \mathfrak{T}(\gamma)$ *extendible*, if there exists a vector field $X \in \mathfrak{T}(\mathcal{M})$ such that for each $t \in I$, $V(t) = X_{\gamma(t)}$. The following lemma (which is exactly Lemma 4.9 from Lee (2018)) establishes the link between linear connections and the concept of directional derivatives of vector fields along curves.

Lemma 2.2.10. *Let ∇ be a linear connection on $\mathcal{M}$. For each curve $\gamma \colon I \to \mathcal{M}$, ∇ determines a unique operator*

$$D_t \colon \mathfrak{T}(\gamma) \to \mathfrak{T}(\gamma)$$

satisfying the following properties:

(a) Linearity over $\mathbb{R}$: for $a, b \in \mathbb{R}$, $V, W \in \mathfrak{T}(\gamma)$,

$$D_t(aV + bW) = aD_tV + bD_tW.$$

(b) Product rule: for $f \in C^\infty(I)$, $V \in \mathfrak{T}(\gamma)$,

$$D_t(fV) = \dot{f}V + fD_tV.$$

(c) If $V \in \mathfrak{T}(\gamma)$ is extendible, then for any extension $X \in \mathfrak{T}(\mathcal{M})$ of V (i.e. $V(t) = X_{\gamma(t)}$ for all $t \in I$),

$$D_tV(t) = \nabla_{\dot{\gamma}(t)}X.$$

For any $V \in \mathfrak{T}(\gamma)$, D_tV is called the *covariant derivative of V along γ*. Note that the expression $\nabla_{\dot{\gamma}(t)}$ is meant in the same way one writes $\nabla_{X_p}Y$ (as discussed below Definition 2.2.8). In coordinate notation, one can express the covariant derivative at $t_0 \in I$, where $V(t) = V^j(t)\partial_j$, so $V^j \in C^\infty(I)$, via

$$D_tV(t_0) = \big(\dot{V}^k(t_0) + V^j(t_0)\dot{\gamma}^i(t_0)\Gamma^k_{ij}(\gamma(t_0))\big)\partial_k, \tag{2.2.4}$$

where γ is plugged into the functions $\Gamma_{ij}^k \in C^\infty(\mathcal{M})$, as well as $\dot{V}^k(t_0)$ is the ordinary derivative of $V^k \colon I \to \mathbb{R}$ at t_0. The *acceleration* of a curve γ is the vector field $D_t\dot{\gamma}$ along γ.

Definition 2.2.11 (Geodesics). A curve γ is called a *geodesic* with respect to ∇ if its acceleration is zero: $D_t\dot{\gamma} \equiv 0$.

This definition together with (2.2.4) gives rise to a second-order system of ordinary differential equations: for coordinates (x^i) on some set $U \subset \mathcal{M}$, a curve $\gamma \colon I \to U$ is a geodesic if and only if its coordinate representation $\gamma(t) = (x^1(t), \dots, x^m(t))$ satisfies the *geodesic equation*:

$$\ddot{x}^k(t) + \dot{x}^i(t)\dot{x}^j(t)\Gamma_{ij}^k(x(t)) = 0.$$

If we fix some point $\gamma(t_0) = p \in U$ and initial velocity $\dot{\gamma}(t_0) = V \in T_p\mathcal{M}$ (which are initial conditions to the system), we get existence and uniqueness of geodesics, cf. Lee (2018, Theorem 4.10).

Theorem 2.2.12 (Existence and uniqueness of geodesics). *Let $\mathcal{M}$ be a smooth manifold with a linear connection ∇. For any $p \in \mathcal{M}$, any $V \in T_p\mathcal{M}$, and any $t_0 \in \mathbb{R}$, there exist an open interval $I \subset \mathbb{R}$ containing t_0 and a geodesic $\gamma \colon I \to \mathcal{M}$ satisfying $\gamma(t_0) = p$, $\dot{\gamma}(t_0) = V$. Any two such geodesics agree on their common domain.*

Furthermore, from this theorem it follows that for any $p \in \mathcal{M}$, any $V \in T_p\mathcal{M}$, there is a unique *maximal geodesic* (one that cannot be extended to any larger interval) $\gamma \colon I \to \mathcal{M}$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = V$. This geodesic is denoted by γ_V. Consider the definition of the *exponential map* from Lee (2018, p.72)

Definition 2.2.13 (Exponential map). Let $\mathcal{M}$ be a smooth manifold. Define the set $\mathcal{E} \subset T\mathcal{M}$, that is the *domain of the exponential map*, by

$$\mathcal{E} := \left\{V \in T\mathcal{M} \colon \gamma_V \text{ is defined on an interval containing } [0,1]\right\},$$

and define the *exponential map* $\operatorname{Exp} \colon \mathcal{E} \to \mathcal{M}$ by

$$\operatorname{Exp}(V) = \gamma_V(1).$$

Furthermore, for $p \in \mathcal{M}$ the *restricted exponential map* Exp_p is the restriction of Exp to the set $\mathcal{E}_p := \mathcal{E} \cap T_p\mathcal{M}$.

Note that by Theorem 2.2.12 that $\mathcal{E}_p$ cannot be empty. From Lee (2018, Proposition 5.7), it follows that each set $\mathcal{E}_p \subset T_p\mathcal{M}$ is star-shaped with respect to $0 \in T_p\mathcal{M}$, that $\gamma_V(t) =$

$\mathrm{Exp}(tV)$ for all $V \in T\mathcal{M}$ and $t \in \mathbb{R}$ such that either side is defined, as well as that Exp is smooth.

2.2.1 Example: Symmetric Positive Definite Matrices

To this end, we introduce the smooth manifold of symmetric positive definite matrices and determine the tangent spaces. We denote the set of all strictly positive definite real-valued symmetric $n \times n$ matrices by $\mathcal{P}$, in particular,

$$\mathcal{P} = \Big\{P \in \mathbb{R}^{n\times n} \colon P = P^T, x^T P x > 0 \text{ for all } x \in \mathbb{R}^n, x \neq 0\Big\}.$$

Here, $x \neq 0$ means that not all entries are zero. It is clear from the definition that $\mathcal{P}$ is an open subset of the real-value symmetric $n \times n$ matrices, denoted by $\mathcal{S}$, i.e.

$$\mathcal{S} = \Big\{S \in \mathbb{R}^{n\times n} \colon S = S^T\Big\}.$$

As $\mathcal{S}$ is a Euclidean space of dimension $n(n+1)/2$, one finds the trivial smooth global chart

$$\mathrm{id} \colon \mathcal{P} \to \mathcal{S}, \quad P \mapsto P,$$

which determines a smooth atlas and thus a smooth structure, so $\mathcal{P}$ is a smooth manifold. Note that by construction of the smooth structure a function $f \colon \mathcal{P} \to \mathbb{R}^k$ is smooth if and only if $f = f \circ \mathrm{id}^{-1} \colon \mathcal{P} \to \mathbb{R}^k$ is smooth in the usual sense, i.e. all partial derivatives of all orders are continuous and exist.

Since $\mathcal{P} \subset \mathcal{S}$ is an open subset, the tangent space of $\mathcal{P}$ at P is $T_P\mathcal{P} \cong T_P\mathcal{S} \cong \mathcal{S}$ (use Lee (2003, Proposition 3.7)). We avoid the vectorization of the symmetric matrices to $\mathbb{R}^{n(n+1)/2}$ and thus simply keep the matrix shape for the tangent vectors. To be precise,

$$T_P\mathcal{P} = \left\{ \left.\frac{\partial}{\partial x^{ij}}\right|_P : i, j = 1, \dots, n,\ i \leq j \right\},$$

where, for smooth functions $f \colon \mathcal{P} \to \mathbb{R}$ mapping $(x^{ij})_{i,j=1; i\leq j}^n$ to a real number,

$$\left.\frac{\partial}{\partial x^{ij}}\right|_P f$$

is in this case just the partial derivative of f in x^{ij}. A derivation $X \in T_P\mathcal{P}$ can thus be represented as (where $X^{ij} \in \mathbb{R}$)

$$X = \sum_{\substack{i,j=1 \\ i \leq j}}^{n} X^{ij} \frac{\partial}{\partial x^{ij}}\bigg|_P,$$

and with slight abuse of notation we can express X as the symmetric matrix $X = (X^{ij})_{i,j=1}^n$ with $X^{ij} = X^{ji}$ and continue calculations with symmetric matrices $X \in \mathcal{S} \cong T_P\mathcal{P}$.

2.3 Riemannian Manifolds

Following Lee (2018, Chapter 2), with $k, l = 0, 1, \ldots$, we denote the *bundle of mixed* $\binom{k}{l}$-*tensors* on $\mathcal{M}$ (also called k-*covariant*, l-*contravariant tensor*) by

$$T^k_l\mathcal{M} := \bigsqcup_{p \in \mathcal{M}} T^k_l(T_p\mathcal{M}),$$

where $T^k_l(T_p\mathcal{M})$ is the space of all multilinear maps

$$F\colon \underbrace{T^*_p\mathcal{M} \times \cdots \times T^*_p\mathcal{M}}_{l \text{ copies}} \times \underbrace{T_p\mathcal{M} \times \cdots \times T_p\mathcal{M}}_{k \text{ copies}} \to \mathbb{R}.$$

We use the convention $T^k_0\mathcal{M} = T^k\mathcal{M}$ and $T^0_l\mathcal{M} = T_l\mathcal{M}$. As in the definition of the tangent bundle, for each such bundle we have a natural projection $\pi\colon T^k_l\mathcal{M} \to \mathcal{M}$. A $\binom{k}{l}$-*tensor field* is a *smooth section* of a tensor bundle $T^k_l\mathcal{M}$, that is a map $F\colon \mathcal{M} \to T^k_l\mathcal{M}$ such that $\pi \circ F = \mathrm{id}_{\mathcal{M}}$ and F is smooth as a map between manifolds. Then, the space of all $\binom{k}{l}$-tensor fields is denoted by $\mathcal{T}^k_l(\mathcal{M})$.

Definition 2.3.1 (Riemannian manifold)**.** Let $\mathcal{M}$ be a smooth manifold. A *Riemannian metric* on $\mathcal{M}$ is a 2-tensor field $g \in \mathcal{T}^2(\mathcal{M})$ that is symmetric, i.e. $g(X, Y) = g(Y, X)$ and positive definite, i.e. $g(X, X) > 0$ for $X \neq 0$, where $X, Y \in T\mathcal{M}$. The pair $(\mathcal{M}, g)$ is called a *Riemannian manifold.* Define $\langle X, Y\rangle := g(X, Y)$.

The Riemannian metric essentially determines an inner product on each tangent space $T_p\mathcal{M}$, and we define the *length* or *norm of a vector* $X \in T_p\mathcal{M}$ to be $|X| = \langle X, X\rangle_p^{1/2}$.

A linear connection ∇ is *compatible with* g if it satisfies the following product rule for all $X, Y, Z \in \mathcal{T}(\mathcal{M})$:

$$\nabla_X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle.$$

Furthermore, a linear connection is *symmetric*, if for all $X, Y \in \mathcal{T}(\mathcal{M})$

$$\nabla_X Y - \nabla_Y X \equiv [X, Y],$$

where $[X, Y]$ is the *Lie bracket*, that is $[X, Y] = XY - YX \in \mathcal{T}(\mathcal{M})$, i.e. for all $f \in C^\infty(\mathcal{M})$ and $p \in \mathcal{M}$, we have $[X, Y]_p f = X_p(Yf) - Y_p(Xf)$ (where $Xf \in C^\infty(\mathcal{M})$). We cite Theorem 5.4 from Lee (2018).

Theorem 2.3.2 (Fundamental lemma of Riemannian geometry). *Let $(\mathcal{M}, g)$ be a Riemannian manifold. There exists a unique linear connection ∇ on $\mathcal{M}$ that is compatible with g and symmetric.*

This connection is called the *Riemannian connection* or *Levi-Civita connection* of g. Geodesics on $(\mathcal{M}, g)$ with respect to the Levi-Civita connection are called *Riemannian geodesics*.

Using symmetry and compatibility, we can express the symbols Γ_{ij}^k from $\nabla_{\partial_i} \partial_j = \Gamma_{ij}^m \partial_m$ via

$$\Gamma_{ij}^k = \frac{1}{2} g^{kl} \big(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij} \big), \tag{2.3.1}$$

where $g_{ij} = \langle \partial_i, \partial_j \rangle$ and $(g^{lk})_{l,k=1}^m$ is the inverse matrix of the matrix $(g_{ij})_{i,j=1}^m$.

Definition 2.3.3 (Isometries). Let $(\mathcal{M}, g)$ and $(\widetilde{\mathcal{M}}, \tilde{g})$ be Riemannian manifolds. A smooth map $\varphi\colon \mathcal{M} \to \widetilde{\mathcal{M}}$ is an *isometry*, if for all $p \in \mathcal{M}$ and $X, Y \in T_p\mathcal{M}$ (recall the push-forward $\varphi_*\colon T_p\mathcal{M} \to T_{\varphi(p)}\widetilde{\mathcal{M}}$)

$$\langle X, Y \rangle_p = \langle \varphi_* X, \varphi_* Y \rangle_{\varphi(p)}.$$

By Lee (2018, Proposition 5.9), if $\varphi\colon \mathcal{M} \to \widetilde{\mathcal{M}}$ is an isometry, then for any $p \in \mathcal{M}$, the following diagram commutes (with Exp_p from Definition 2.2.13):

$$\begin{array}{ccc} T_p\mathcal{M} & \xrightarrow{\varphi_*} & T_{\varphi(p)}\widetilde{\mathcal{M}} \\ {\scriptstyle \mathrm{Exp}_p}\downarrow & & \downarrow{\scriptstyle \widetilde{\mathrm{Exp}}_{\varphi(p)}} \\ \mathcal{M} & \xrightarrow{\varphi} & \widetilde{\mathcal{M}} \end{array}$$

By Lee (2018, Lemma 5.10), for any $p \in \mathcal{M}$, there is a neighborhood $\mathcal{V}$ of $0 \in T_p\mathcal{M}$ and a neighborhood $\mathcal{U}$ of p in $\mathcal{M}$ such that $\mathrm{Exp}_p\colon \mathcal{V} \to \mathcal{U}$ is a diffeomorphism, i.e. bijective and smooth with smooth inverse.

Definition 2.3.4 (Logarithm map)**.** Consider the setup as above. We denote the inverse by $\mathrm{Log}_p\colon \mathcal{U} \to \mathcal{V}$, mapping elements from $\mathcal{U} \subset \mathcal{M}$ to $\mathcal{V} \subset T_p\mathcal{M}$. Log_p is called the *logarithm map*.

We continue to define the notion of lengths of curves as well as distance on a Riemannian manifold, eventually ending up in the realm of metric spaces. The easiest "class" of curves in a manifold whose length one wants to determine are smooth curve segments (analogously to Lee (2018, p.92)): if $\gamma\colon [a,b] \to \mathcal{M}$ is a curve segment in a Riemannian manifold $(\mathcal{M}, g)$, we define the *length* of γ to be

$$L_g(\gamma) \coloneqq \int_a^b |\dot{\gamma}(t)| \mathrm{d}t.$$

We introduce more classes of curves. A *regular curve* is a smooth curve $\gamma\colon I \to \mathcal{M}$ such that $\dot{\gamma}(t) \neq 0$ for all $t \in I$. A continuous map $\gamma\colon [a,b] \to \mathcal{M}$ is called a *piece-wise regular curve segment* if there exists a finite subdivision $a = a_0 < a_1 < \ldots < a_k = b$ such that $\gamma\big|_{[a_{i-1},a_i]}$ is a regular curve for $i = 1, \ldots, k$. Distances on Riemannian manifolds are measured along such curve segments, and we call such curves *admissible curves*. The *length* of an admissible curve is defined as the sum of the regular curve segments, i.e.

$$L_g(\gamma) \coloneqq \sum_{i=1}^{k} L_g\big(\gamma\big|_{[a_{i-1},a_i]}\big).$$

Definition 2.3.5 (Riemannian distance)**.** Let $(\mathcal{M}, g)$ be a connected Riemannian manifold. Define the *Riemannian distance* $d_g(p,q)$ between two points $p, q \in \mathcal{M}$ as the infimum of the lengths of all admissible curves from p to q, that is

$$d_g(p,q) = \inf_{\substack{\gamma\colon [0,1]\to\mathcal{M} \\ \gamma \text{ admissible} \\ \gamma(0)=p;\ \gamma(1)=q}} L_g(\gamma).$$

From Lee (2018, Lemma 6.2) or Bridson & Haefliger (1999, Proposition 3.18) it follows that $(\mathcal{M}, d_g)$ is a metric space and that the induced topology from the metric d_g coincides with the topology of the manifold $\mathcal{M}$. Furthermore, every minimizing curve $\gamma\colon [a,b] \to \mathcal{M}$, i.e. $L_g(\gamma) = d_g(\gamma(a), \gamma(b))$, is a geodesic when it is given a unit speed parametrization (i.e. $|\dot{\gamma}(t)| = 1$ for all $t \in [a,b]$), by Lee (2018, Theorem 6.6). Furthermore, by Lee (2018, Theorem 6.12), every Riemannian geodesic is locally minimizing, i.e. for $\gamma\colon I \to \mathcal{M}$, for any $t_0 \in I$ there exists a neighborhood $J \subset I$ such that $\gamma|_J$ is minimizing between each pair of its points. The following result from Bridson & Haefliger (1999, Proposition 3.18) relates the Riemannian manifold to length spaces (as in Definition 2.1.4).

Proposition 2.3.6. *Let $(\mathcal{M}, g)$ be a connected Riemannian manifold. The metric space $(\mathcal{M}, d_g)$ is a length space.*

Note that this means that $d_g = d_g^*$ and therefore, the length of an admissible curve in the metric space sense, $L_{d_g}(\gamma)$, coincides with the length of γ in the Riemannian manifold sense $L_g(\gamma)$. We cite Bridson & Haefliger (1999, Proposition 3.23):

Proposition 2.3.7. *Let $(\mathcal{M}, g)$ be a Riemannian manifold and $\widetilde{\mathcal{M}} \subset \mathcal{M}$ be a smoothly embedded submanifold. Then the restriction $\tilde{g}_p = g_p\big|_{T_p\widetilde{\mathcal{M}} \times T_p\widetilde{\mathcal{M}}}$ gives a Riemannian metric for $\widetilde{\mathcal{M}}$. It holds that $(d_g\big|_{\widetilde{\mathcal{M}}})^* = d_{\tilde{g}}$, that is, the induced intrinsic metric of $(\widetilde{\mathcal{M}}, d_g\big|_{\widetilde{\mathcal{M}}})$ coincides with the Riemannian distance on $\widetilde{\mathcal{M}}$ induced by $\tilde{g}$.*

Finally, we will cite the Theorem of Hopf-Rinow for Riemannian manifolds, cf. Lee (2018, Theorem 6.13) or Lang (1999, p.224-226). Note that a Riemannian manifold is geodesically complete if every maximal geodesic is defined for all $t \in \mathbb{R}$. Recall that a metric space is called *complete*, if all Cauchy sequences converge in that space.

Theorem 2.3.8 (Hopf-Rinow for Riemannian manifolds)**.** *A connected Riemannian manifold is geodesically complete if and only if it is complete as a metric space.*

2.4 Curvature

We introduce the curvature tensor on Riemannian manifolds and the concept of sectional curvatures. We follow Lee (2018, Chapter 7).

Definition 2.4.1 (Riemann curvature tensor)**.** Let $(\mathcal{M}, g)$ be a Riemannian manifold. The *Riemann curvature endomorphism* is the map $R\colon \mathfrak{T}(\mathcal{M}) \times \mathfrak{T}(\mathcal{M}) \times \mathfrak{T}(\mathcal{M}) \to \mathfrak{T}(\mathcal{M})$ defined by

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z. \tag{2.4.1}$$

This is by Lee (2018, Proposition 7.1) a $\binom{3}{1}$-tensor field and by Lee (2018, Lemma 7.2) locally invariant under isometries, and thus Riemannian manifolds that are locally isometric and in this sense considered to be "equal", have locally the same "curvature". The local coordinate representation of R is (where ∂_i as in Equation (2.2.2))

$$R(\partial_i, \partial_j)\partial_k = \sum_{s=1}^{m} R_{ijk}^s \partial_s,$$

and from Do Carmo (1992, p.93) (where, one needs to take the negative of their formula as their definition of the curvature endomorphism is $-R(X,Y)Z$), we have

$$R_{ijk}^s = \sum_{h=1}^m \Gamma_{jk}^h \Gamma_{ih}^s - \sum_{h=1}^m \Gamma_{ik}^h \Gamma_{jh}^s + \partial_i \Gamma_{jk}^s - \partial_j \Gamma_{ik}^s. \tag{2.4.2}$$

From the Riemann curvature endomorphism, we can derive its corresponding $\binom{4}{0}$-tensor, called the *Riemann curvature tensor*, defined by

$$Rm(X,Y,Z,W) = \langle R(X,Y)Z, W \rangle. \tag{2.4.3}$$

Its coordinate representation is consequently

$$R_{ijks} = \sum_{h=1}^m R_{ijk}^h g_{hs}. \tag{2.4.4}$$

Using the Riemann curvature tensor, we can define sectional curvatures (cf. Lee (2018, Proposition 8.8)).

Definition 2.4.2. Let $(\mathcal{M}, g)$ be a Riemannian manifold. The *sectional curvature* of $\mathcal{M}$ associated with the 2-plane $\Pi \subset T_p\mathcal{M}$ spanned by any basis $X, Y \in T_p\mathcal{M}$ with $p \in \mathcal{M}$ to be

$$K(X,Y) = \frac{Rm(X,Y,Y,X)}{|X|^2|Y|^2 - \langle X,Y \rangle^2}.$$

Remark 2.4.3.

1. In Lee (2018), the sectional curvature associated with the 2-plane Π is defined as the *Gaussian curvature* (cf. Lee (2018, p.142)) of the 2-dimensional submanifold locally spanned by the tangent vectors in Π and then the above formula is proven in Lee (2018, Proposition 8.8) using Gauss's *Theorema Egregium*, e.g. Lee (2018, Theorem 8.6).
2. There are more concepts of curvature directly derived from the Riemann curvature tensor, such as the *scalar curvature* and the *Ricci curvature* (e.g. Lee (2018, p.124)).

We establish the correspondence between the sectional curvatures and CAT(0) spaces from Definition 2.1.12. We cite Bridson & Haefliger (1999, Part II, Theorem 1A.6).

Theorem 2.4.4. *Let $(\mathcal{M}, g)$ be a smooth Riemannian manifold, and let $(\mathcal{M}, d_g)$ be the induced metric space. Then $(\mathcal{M}, d_g)$ is a CAT(0) space if and only if for all $p \in \mathcal{M}$ and for all choices of bases $X, Y \in T_p\mathcal{M}$ it holds that $K(X,Y) \leq 0$.*

2.5 Riemannian Submersions

We follow the notation from Lee (1997, Exercise 3-8). Let $\mathcal{M}$ and $\widetilde{\mathcal{M}}$ be smooth manifolds and suppose the map $\pi\colon \widetilde{\mathcal{M}} \to \mathcal{M}$ is a surjective submersion. For each $q \in \mathcal{M}$, the *fiber over* p, denoted by $\widetilde{\mathcal{M}}_q$, is the inverse image $\pi^{-1}(q) \subset \widetilde{\mathcal{M}}$ and by the implicit function theorem a closed, embedded submanifold (cf. Lee (1997, Exercise 3-8); and $\widetilde{\mathcal{M}}_q$ is non-empty since π is surjective). Suppose furthermore that $(\widetilde{\mathcal{M}}, \widetilde{g})$ is a Riemannian manifold. Recall the push-forward of π, that is the linear map and additionally surjective as π is a submersion,

$$(\partial\pi)_p\colon T_p\widetilde{\mathcal{M}} \to T_{\pi(p)}\mathcal{M}.$$

We consider the following decomposition of $T_p\widetilde{\mathcal{M}}$ into an orthogonal direct sum,

$$T_p\widetilde{\mathcal{M}} = H_p\widetilde{\mathcal{M}} \oplus V_p\widetilde{\mathcal{M}},$$

where $V_p\widetilde{\mathcal{M}}$ is the *vertical space* and $H_p\widetilde{\mathcal{M}}$ is the *horizontal space*, given by

$$V_p\widetilde{\mathcal{M}} \coloneqq \ker\big((\partial\pi)_p\big) = T_p\widetilde{\mathcal{M}}_{\pi(p)}, \qquad H_p\widetilde{\mathcal{M}} \coloneqq (V_p\widetilde{\mathcal{M}})^{\perp},$$

note that the Riemannian metric $\widetilde{g}$ is used here to determine orthogonality and that the map $(\partial\pi)_p\big|_{H_p\widetilde{\mathcal{M}}}\colon H_p\widetilde{\mathcal{M}} \to T_{\pi(p)}\mathcal{M}$ is an isomorphism.

Definition 2.5.1 (Riemannian submersion)**.** Let $(\widetilde{\mathcal{M}}, \widetilde{g})$ and $(\mathcal{M}, g)$ be smooth Riemannian manifolds. A smooth map $\pi\colon \widetilde{\mathcal{M}} \to \mathcal{M}$ is a *Riemannian submersion* if

1. π is a surjective submersion,
2. π_* is an isometry on vectors in the horizontal space.

The second condition can be rephrased in the sense that the map

$$(\partial\pi)_p\big|_{H_p\widetilde{\mathcal{M}}}\colon H_p\widetilde{\mathcal{M}} \to T_{\pi(p)}\mathcal{M}$$

is an isometry for all $p \in \widetilde{\mathcal{M}}$, i.e.

$$\widetilde{g}_p(\widetilde{X}, \widetilde{Y}) = g_{\pi(p)}\Big((\partial\pi)_p(\widetilde{X}), (\partial\pi)_p(\widetilde{Y})\Big)$$

for all $\widetilde{X}, \widetilde{Y} \in H_p\widetilde{\mathcal{M}}$.

With the setup of Definition 2.5.1, any vector field $\widetilde{X} \in \mathfrak{T}(\mathcal{M})$ can be written uniquely as $\widetilde{X} = \widetilde{X}^H + \widetilde{X}^V$ where X^H is horizontal and $\widetilde{X}^V$ is vertical, and both are smooth (cf. Lee

(1997, Exercise 3-8, (a))). Furthermore, if $X \in \mathfrak{T}(\mathcal{M})$ is a vector field on $\mathcal{M}$, then there is a unique smooth horizontal vector field $X^{\#}$ on $\widetilde{\mathcal{M}}$, called the *horizontal lift* of X, such that $(\partial\pi)_p(X_p^{\#}) = X_{\pi(p)}$ for each $p \in \mathcal{M}$ (cf. Lee (1997, Exercise 3-8, (b))). In the following, we will use $\cdot^{\#}$ as the horizontal lift operator. We state some well-known properties of the horizontal lift from Lee (1997, Exercise 5-9).

Lemma 2.5.2. *Let $\pi\colon (\widetilde{\mathcal{M}}, \widetilde{g}) \to (\mathcal{M}, g)$ be a Riemannian submersion and denote by $\widetilde{\nabla}$ and ∇ their respective Riemannian connections. Then, for any vector fields $X, Y \in \mathfrak{T}(\mathcal{M})$, it holds that for all $p \in \widetilde{\mathcal{M}}$ with $\pi(p) = q \in \mathcal{M}$ that*

1. $\tilde{g}_p(X_p^{\#}, Y_p^{\#}) = g_q(X_q, Y_q)$,
2. $[X^{\#}, Y^{\#}]^H = [X, Y]^{\#}$,
3. $\widetilde{\nabla}_{X^{\#}} Y^{\#} = \big(\nabla_X Y\big)^{\#} + \frac{1}{2}\big[X^{\#}, Y^{\#}\big]^V$.

Proof. A proof for 1. and 2. can be found in O'Neill (1983, Lemma 7.45), and 3. follows immediately from using O'Neill (1966, Lemma 3, 4.) and plugging in O'Neill (1966, Lemma 1, Lemma 2). ∎

We state *O'Neills formula*, that is the fundamental relation between the sectional curvatures of $\widetilde{\mathcal{M}}$ and $\mathcal{M}$ (cf. O'Neill (1966, Corollary 1), O'Neill (1983, Theorem 7.47) or Lee (1997, Exercise 8-11)).

Theorem 2.5.3 (O'Neills formula)**.** *Let $\pi\colon (\widetilde{\mathcal{M}}, \widetilde{g}) \to (\mathcal{M}, g)$ be a Riemannian submersion and let $X, Y \in \mathfrak{T}(\mathcal{M})$ be two orthonormal vector fields. Then the sectional curvatures of $\widetilde{\mathcal{M}}$ and $\mathcal{M}$ are related via*

$$K(X,Y) = \widetilde{K}\big(X^{\#}, Y^{\#}\big) + \frac{3}{4}\Big|\big[X^{\#}, Y^{\#}\big]^V\Big|^2,$$

This means that the sectional curvature on $\mathcal{M}$ cannot decrease with respect to the sectional curvature on $\widetilde{\mathcal{M}}$.

Next, we will consider a special case of a Riemannian submersion: the quotient obtained from a Lie group acting on a Riemannian manifold by isometries. For a start, we cite the definition of a *Lie group* from Lee (2003, page 30) and *Lie group actions* from Lee (2003, Chapter 7).

Definition 2.5.4 (Lie group, Lie group action)**.**

1. A *Lie group* is a smooth manifold G that is also a *group* in the algebraic sense, such that the the maps (where the first denotes the group operation)

$$G \times G \to G, \quad (h_1, h_2) \mapsto h_1 h_2,$$
$$G \to G, \quad h \mapsto h^{-1},$$

are smooth. Denote the identity element of G by e.

2. Let G be a Lie group and $\mathcal{M}$ be a smooth manifold. A *left action* of G on $\mathcal{M}$ is a map $\theta\colon G \times \mathcal{M} \to \mathcal{M}$, written as $\theta_h(p) = h \cdot p$, that satisfies (where $h_1, h_2 \in G$, $p \in \mathcal{M}$)

$$h_1 \cdot (h_2 \cdot p) = (h_1 h_2) \cdot p,$$
$$e \cdot p = p.$$

We further cite the definition of several properties of Lie group actions from Lee (2003, Chapter 7), as well as Huckemann et al. (2010).

Definition 2.5.5. Let $\theta\colon G \times \mathcal{M} \to \mathcal{M}$ be a left action of a Lie group G on a smooth manifold $\mathcal{M}$. Then

- the action is said to be *smooth*, if $\theta_h(p)$ depends smoothly on (h, p);
- for any $p \in \mathcal{M}$, the *fiber* or *orbit* of p under the action is the set $[p] = G \cdot p = \{h \cdot p\colon h \in G\}$;
- the action is *transitive* if for any two points $p, q \in \mathcal{M}$, there exists $h \in G$ with $h \cdot p = q$;
- the action is said to be acting *freely* on $\mathcal{M}$ if $h_1 \cdot p = h_2 \cdot p$ implies $h_1 = h_2$ for any $h_1, h_2 \in G$ and $p \in \mathcal{M}$.
- the action is said to be acting *properly* on $\mathcal{M}$ if for all $p_n, p, p' \in \widetilde{\mathcal{M}}$, $h_n \in G$ such that $h_n \cdot p_n \to p'$, $p_n \to p$, it follows that h_n has a point of accumulation $h \in G$ with $h \cdot p = p'$.

Note that whenever θ acts freely on $\mathcal{M}$, the fibers $[p]$ are isomorphic to G for all $p \in \mathcal{M}$ (cf. Huckemann et al. (2010)). Having introduced the necessary notation, we proceed to prove the following result (cf. Lee (1997, Exercise 3-8)).

Theorem 2.5.6. *Let $\pi\colon \widetilde{\mathcal{M}} \to \mathcal{M}$ be a surjective submersion from a smooth Riemannian manifold $(\widetilde{\mathcal{M}}, \widetilde{g})$ to a smooth manifold $\mathcal{M}$. Let furthermore G be a Lie group acting smoothly on $\widetilde{\mathcal{M}}$ by $\theta\colon G \times \widetilde{\mathcal{M}} \to \widetilde{\mathcal{M}}$, and suppose that*

1. *G acts on $\widetilde{\mathcal{M}}$ by isometries of $\widetilde{g}$, i.e. for any two vector fields $\widetilde{X}, \widetilde{Y} \in \mathfrak{T}(\widetilde{\mathcal{M}})$ and for any $p \in \widetilde{\mathcal{M}}$, $h \in G$, it holds that*

$$\widetilde{g}_p(\widetilde{X}_p, \widetilde{Y}_p) = \widetilde{g}_{\theta_h(p)}\Big((\partial\theta_h)_p(\widetilde{X}_p), (\partial\theta_h)_p(\widetilde{Y}_p)\Big),$$

2. *$\pi \circ \theta_h = \pi$ for every $h \in G$, and*
3. *G acts transitively on each fiber $\widetilde{\mathcal{M}}_q$, $q \in \mathcal{M}$, i.e. for any two points $p_1, p_2 \in \widetilde{\mathcal{M}}_q$, there exists $h \in G$ with $\theta_h(p_1) = p_2$.*

Then there is a unique Riemannian metric g on $\mathcal{M}$ such that π is a Riemannian submersion.

Proof. Recall that $V_p\widetilde{\mathcal{M}} = \ker\big((\partial\pi)_p\big)$, and since by 2., $\pi \circ \theta_h = \pi$, we have $(\partial\pi)_p = (\partial\pi)_{\theta_h(p)} \circ (\partial\theta_h)_p$ and thus

$$(\partial\theta_h)_p\big(V_p\widetilde{\mathcal{M}}\big) = V_{\theta_h(p)}\widetilde{\mathcal{M}} \qquad \text{and furthermore} \qquad (\partial\theta_h)_p\big(H_p\widetilde{\mathcal{M}}\big) = H_{\theta_h(p)}\widetilde{\mathcal{M}}.$$

Observe that for any $q \in \mathcal{M}$, from 2. and 3. it follows $\widetilde{\mathcal{M}}_q = \pi^{-1}(q) = [p]$, with arbitrary $p \in \widetilde{\mathcal{M}}_q$. Let $q \in \mathcal{M}$ and $X, Y \in \mathfrak{T}(\mathcal{M})$. Let $p \in \widetilde{\mathcal{M}}_q = [p]$ be arbitrary. Define

$$g_q(X_q, Y_q) := \widetilde{g}_p\big(X_p^{\#}, Y_p^{\#}\big) \tag{2.5.1}$$

To show that this is well-defined, let $p' \in \widetilde{\mathcal{M}}_q$ with $p' \neq p$. By 3., there exists an element $h \in G$ with $p' = \theta_h(p)$, and we have, since θ_h is an isometry,

$$\widetilde{g}_p\big(X_p^{\#}, Y_p^{\#}\big) = \widetilde{g}_{p'}\big((\partial\theta_h)_p(\widetilde{X}_p^{\#}), (\partial\theta_h)_p(\widetilde{Y}_p^{\#})\big),$$

and due to $(\partial\theta_h)_p\big(H_p\widetilde{\mathcal{M}}\big) = H_{\theta_h(p)}\widetilde{\mathcal{M}}$ the vector $(\partial\theta_h)_p(\widetilde{X}_p^{\#})$ is horizontal and thus

$$(\partial\theta_h)_p(\widetilde{X}_p^{\#}) = \widetilde{X}_{\theta_h(p)}^{\#} = \widetilde{X}_{p'}^{\#},$$

analogously for $Y_p^{\#}$, and therefore

$$\widetilde{g}_p\big(X_p^{\#}, Y_p^{\#}\big) = \widetilde{g}_{p'}\big(X_{p'}^{\#}, Y_{p'}^{\#}\big),$$

so $g_q(X_q, Y_q)$ is well-defined for any $q \in \mathcal{M}$ and by construction smooth. Now, g makes π a Riemannian submersion, since for any $\widetilde{X}, \widetilde{Y} \in H_p\widetilde{\mathcal{M}}$, $p \in \widetilde{\mathcal{M}}$, we have

$$\widetilde{g}_p(\widetilde{X}_p, \widetilde{Y}_p) = \widetilde{g}_p\Big(\big((\partial\pi)_p(\widetilde{X}_p)\big)^{\#}, \big((\partial\pi)_p(\widetilde{Y}_p)\big)^{\#}\Big) = g_{\pi(p)}\big((\partial\pi)_p(\widetilde{X}_p), (\partial\pi)_p(\widetilde{Y}_p)\big).$$

Uniqueness follows from π_* being an isometric diffeomorphism. ■

Remark 2.5.7. Note that in Theorem 2.5.6, we assume the existence of a manifold $\mathcal{M}$ and a submersion $\pi\colon \widetilde{\mathcal{M}} \to \mathcal{M}$. In contrast to these assumptions that are made in Lee (1997, Exercise 3-8), the authors in Huckemann et al. (2010) and Abraham & Marsden (2008, p.266) work with the setting that $\pi\colon \widetilde{\mathcal{M}} \to \widetilde{\mathcal{M}}/G$, $p \mapsto [p]$, and make the additional assumption that the group action is

- proper, such that the fibers $[p]$ for $p \in \widetilde{\mathcal{M}}$ are closed and thus $\widetilde{\mathcal{M}}/G$ is Hausdorff;
- free, which implies that all the fibers $[p]$ have the same dimension.

Then one can show that $\widetilde{\mathcal{M}}/G$ carries a unique smooth manifold structure compatible with its quotient topology (Abraham & Marsden (2008, p.266)) such that π is a surjective submersion.

From Huckemann et al. (2010), in the setting of Theorem 2.5.6, we find that the distance $d_{\mathcal{M}}$ induced from the Riemannian metric on $\mathcal{M}$ as defined in Theorem 2.5.6 is given by (where $d_{\widetilde{\mathcal{M}}}$ is the induced distance on $\widetilde{\mathcal{M}}$ and let $p_1, p_2 \in \widetilde{\mathcal{M}}$, $q_1 = \pi(p_1)$, $q_2 = \pi(p_2)$)

$$d_{\mathcal{M}}(q_1, q_2) = \inf_{h_1, h_2 \in G} d_{\widetilde{\mathcal{M}}}(h_1 \cdot p_1, h_2 \cdot p_2),$$

and since the action is isometric, this can be simplified to

$$d_{\mathcal{M}}(q_1, q_2) = \inf_{h \in G} d_{\widetilde{\mathcal{M}}}(h \cdot p_1, p_2). \tag{2.5.2}$$

2.6 Riemann Stratified Spaces

First, we introduce Stiefel manifolds and Grassmannian manifolds, for the latter see e.g. Lee (2003, Chapter 7). Let $k, m \in \mathbb{N}$. Any k-dimensional linear subspace $\mathcal{V}$ of $\mathbb{R}^m$ is the span of the linearly independent columns $x_1, \ldots, x_k$ of a matrix $X = (x_1, \ldots, x_k) \in \mathbb{R}^{m \times k}$, and the space of all such matrices is the *Stiefel manifold*

$$\mathrm{St}(m, k) = \{X \in \mathbb{R}^{m \times k} : \mathrm{rk}(X) = k\},$$

and we write

$$\mathcal{V} = \mathrm{span}\{x_1, \ldots, x_k\} =: \mathrm{col}(X).$$

The *Grassmannian manifold* is the quotient

$$\mathrm{Gr}(m,k) := \mathrm{St}(m,k)/\mathrm{St}(k,k),$$

which can be identified with the space

$$\{\mathcal{V} \subset \mathbb{R}^m \colon \mathcal{V} \text{ linear subspace, } \dim(\mathcal{V}) = k\},$$

which is due to $\mathrm{col}(X) = \mathrm{col}(XY)$ for all $Y \in \mathrm{St}(k,k)$ and $X \in \mathrm{St}(m,k)$. Hence, the Grassmannian can be viewed as the manifold of all k-dimensional linear subspaces in $\mathbb{R}^m$, and consequently, a sequence of k-dimensional linear subspaces $\mathcal{V}_n$, $n \in \mathbb{N}$, of $\mathbb{R}^m$, $1 \le k \le m$, converges in the Grassmannian $\mathrm{Gr}(m,k)$ to a k-dimensional linear subspace $\mathcal{V}$ if there are $X_n, X \in \mathrm{St}(m,k)$ and $Y_n \in \mathrm{St}(k,k)$ such that $\mathrm{col}(X_n) = \mathcal{V}_n$ for all $n \in \mathbb{N}$, $\mathrm{col}(X) = \mathcal{V}$ and $\|X_n Y_n - X\| \to 0$ as $n \to \infty$. The proof of the next lemma is from our unpublished work and due to Stephan Huckemann.

Lemma 2.6.1. *Let* $X_n \in \mathrm{St}(m,k)$, $n \in \mathbb{N}$ *such that* $\lim_{n\to\infty} X_n \in \mathbb{R}^{m\times k}$ *exists and such that* $\lim_{n\to\infty} \mathrm{col}(X_n)$ *exists. Then*

$$\mathrm{col}\Big(\lim_{n\to\infty} X_n\Big) \subseteq \lim_{n\to\infty} \mathrm{col}(X_n).$$

Proof. Let $x \in \mathbb{R}^m$ with $x \perp \lim_{n\to\infty} \mathrm{col}(X_n)$. Once we show $x \perp X := \lim_{n\to\infty} X_n$, the assertion follows. By hypothesis, for every $\varepsilon > 0$ there are $N \in \mathbb{N}$ and $Y_n \in \mathrm{St}(k,k)$ such that

$$|x^T X_n Y_n| < \varepsilon \;\forall n > N.$$

In the case that there is a subsequence n_m with $\|Y_{n_m}\| > 1$, define $Z_{n_m} = Y_{n_m}/\|Y_{n_m}\| \in \mathrm{St}(k,k)$ to find

$$|x^T X_{n_m} Z_{n_m}| < \frac{\varepsilon}{\|Y_{n_m}\|} < \varepsilon \;\forall n_m > N,$$

and as Z_{n_m} is bounded, there is a cluster point $Z \in \mathrm{St}(k,k)$ with $|x^T XZ| \le \varepsilon$, and as $\varepsilon > 0$ was arbitrary, we have $x^T XZ = 0$, so $x^T X = 0$. In the case that there is no such subsequence, Y_{n_m} has a cluster point $Y \in \mathrm{St}(k,k)$ such that $|x^T XY| \le \varepsilon$ and thus again $x^T X = 0$. Therefore, $x \perp X$ as asserted. ■

Definition 2.6.2 (Stratified space, Whitney stratified space of type (A) and (B)). A *stratified space* $\mathcal{S}$ of dimension m embedded in a Euclidean space (possibly of higher dimension $M \geq m$) is a direct sum

$$\mathcal{S} = \bigsqcup_{i=1}^{k} S_i$$

such that $0 \leq d_1 < \ldots < d_k = m$, each S_i is a d_i-dimensional manifold and $S_i \cap S_j = \emptyset$ for $i \neq j$ and if $S_i \cap \overline{S_j} \neq \emptyset$ then $S_i \subset \overline{S_j}$.

A stratified space $\mathcal{S}$ is *Whitney stratified of type (A)*,

(A) if for a sequence $q_1, q_2, \cdots \in S_j$ that converges to some point $p \in S_i$, such that the sequence of tangent spaces $T_{q_n}S_j$ converges in $\mathrm{Gr}(M, d_j)$ to some d_j-dimensional linear space T as $n \to \infty$, then $T_pS_i \subset T$, where all the linear spaces are seen as subspaces of $\mathbb{R}^M$.

Moreover, a stratified space $\mathcal{S}$ is a *Whitney stratified space of type (B)*,

(B) if for sequences $p_1, p_2, \cdots \in S_i$ and $q_1, q_2, \cdots \in S_j$ which converge to the same point $p \in S_i$ such that the sequence of secant lines c_n between p_n and q_n converges to a line c as $n \to \infty$ (in $\mathrm{Gr}(M, 1)$), and such that the sequence of tangent planes $T_{q_n}S_j$ converges to a d_j-dimensional plane T as $n \to \infty$ (in $\mathrm{Gr}(M, d_j)$), then $c \subset T$.

For the definition of Riemann stratified spaces, we use the definition from Huckemann & Eltzner (2020, Section 10.6).

Definition 2.6.3 (Riemann stratified space of type (A) and (B)). A *Riemann stratified space* of type (A) (type (B)) is a Whitney stratified space $\mathcal{S}$ of type (A) (type (B)) such for each $i = 0, \ldots, k$, S_i is a d_i-dimensional Riemannian manifold with Riemannian metric $g^{(i)}$, and if a sequence $q_1, q_2, \cdots \in S_j$ converges to a point $p \in S_i$ such that as above $T_{q_n}S_j$ converges to some T as $n \to \infty$, then the Riemannian metric $g^{(j)}_{q_n}$ converges to some 2-tensor $g^*_p \colon T \otimes T \to \mathbb{R}$ as $n \to \infty$, then $g^{(i)}_p \equiv g^*_q\big|_{T^2_pS_i}$.

Chapter 3

Geometries for Strictly Positive Definite Matrices

Let $N \in \mathbb{N}$. Recall from Section 2.2.1 the symmetric matrices and the symmetric positive definite matrices, i.e.

$$\mathcal{S} = \left\{ S \in \mathbb{R}^{N \times N} \colon S = S^T \right\}.$$

and

$$\mathcal{P} = \left\{ P \in \mathcal{S} \colon x^T P x > 0 \text{ for all } x \in \mathbb{R}^N, x \neq 0 \right\}.$$

Furthermore, we introduce the set of matrices $P \in \mathcal{P}$ that have ones on the diagonal, these are the correlation matrices:

$$\mathcal{C} = \left\{ P \in \mathcal{P} \colon \operatorname{diag}(P) = 1 \right\}.$$

Finally, denote all positive diagonal matrices by

$$\mathcal{D} = \left\{ P = (P_{ij})_{i,j=1}^N \in \mathcal{P} \colon P_{ij} = 0, P_{ii} > 0, i \neq j \right\}.$$

Furthermore, define the *matrix exponential function* $\exp \colon \mathcal{S} \to \mathcal{P}$ (cf. Lang (1999, Chapter XII,§1)) to be

$$\exp(X) = \sum_{k=0}^{\infty} \frac{X^k}{k!}, \tag{3.0.1}$$

which, given the spectral decomposition $U \Lambda U^T = X$, where Λ is a diagonal matrix containing the eigenvalues λ_i, $i = 1, \ldots, N$ of X (e.g. Hoffman & Kunze (1971, Chapter 8, Theorem 18) for the spectral decomposition), can be calculated via $\exp(X) = U \exp(\Lambda) U^T$, where $\exp(\Lambda)$ is the diagonal matrix with entries $\exp(\lambda_i)$, $i = 1, \ldots, N$. Since this is a one-to-one mapping from $\mathcal{S}$ to $\mathcal{P}$, we can define its inverse to be the *matrix logarithm* $\log\colon \mathcal{P} \to \mathcal{S}$ (e.g. Pennec et al. (2006)). Furthermore, the *matrix square root* of a positive definite matrix $P \in \mathcal{P}$ is denoted by $\sqrt{P}$ or $P^{1/2}$, which is the unique matrix in $\mathcal{P}$ that satisfies $\sqrt{P}^2 = P$.

3.1 The Fisher-Information Geometry

Recall from Section 2.2.1 the smooth manifold of symmetric strictly positive definite matrices $\mathcal{P}$, where the tangent space of $\mathcal{P}$ at $P \in \mathcal{P}$ is just $\mathcal{S} \cong T_P\mathcal{P}$. We introduce a Riemannian metric (cf. Definition 2.3.1) for $\mathcal{P}$, which is referred to in this thesis as the *Fisher-information metric*, given at $P \in \mathcal{P}$ for all $X, Y \in \mathcal{S}$ by

$$g_P(X,Y) = \operatorname{Tr}\left[P^{-1}XP^{-1}Y\right],$$

cf. Lang (1999, Chapter XII), so we obtain a Riemannian manifold $(\mathcal{P}, g)$. One can show that the induced distance is ($P, Q \in \mathcal{P}$; cf. Definition 2.3.5)

$$d(P,Q)^2 = \operatorname{Tr}\left[\log\left(\sqrt{P}^{-1}Q\sqrt{P}^{-1}\right)^2\right] = \sum_{i=1}^{N} \log(\mu_i)^2, \tag{3.1.1}$$

where $\mu_1, \dots, \mu_N$ are the eigenvalues of PQ^{-1} and $\log\colon \mathcal{P} \to \mathcal{S}$ is the matrix logarithm. The Riemann exponential and logarithm map are given by (e.g. Pennec et al. (2006))

$$\begin{aligned}\operatorname{Exp}_P(X) &= \sqrt{P}\exp\left(\sqrt{P}^{-1}X\sqrt{P}^{-1}\right)\sqrt{P},\\ \operatorname{Log}_P(Q) &= \sqrt{P}\log\left(\sqrt{P}^{-1}Q\sqrt{P}^{-1}\right)\sqrt{P}.\end{aligned}$$

With $t \in \mathbb{R}$, $\gamma_{P,Q}(0) = P$ and $\gamma_{P,Q}(1) = Q$, the geodesic from P to Q is given explicitly by (e.g. Moakher (2005, p.5) or Moakher & Zerai (2011, Theorem 3))

$$\gamma_{P,Q}(t) = \operatorname{Exp}_P\left(t \operatorname{Log}_P(Q)\right) = \sqrt{P}\exp\left(t\,\log\left(\sqrt{P}^{-1}Q\sqrt{P}^{-1}\right)\right)\sqrt{P}.$$

Parallel transport of $X \in T_P\mathcal{P}$ from P to Q is given by

$$\begin{aligned}\Pi_{P,Q}(X) &= \sqrt{P}\sqrt{\sqrt{P}^{-1}Q\sqrt{P}^{-1}}\sqrt{P}^{-1}X\sqrt{P}^{-1}\sqrt{\sqrt{P}^{-1}Q\sqrt{P}^{-1}}\sqrt{P}\\ &= \sqrt{QP^{-1}}X\sqrt{P^{-1}Q},\end{aligned}$$

where the first equality is derived using the *exact pole ladder scheme for symmetric spaces* from Pennec (2018), and the second equality is from Yair et al. (2019), which can be derived using

$$\sqrt{QP^{-1}} = \sqrt{P}\sqrt{\sqrt{P}^{-1}Q\sqrt{P}^{-1}}\sqrt{P}^{-1}.$$

The Riemannian manifold $(\mathcal{P}, g)$ has global non-positive sectional curvature and is therefore CAT(0) by Theorem 2.4.4. Consequently, between any two points $P, Q \in \mathcal{P}$, there exists a unique geodesic connecting them (cf. Theorem 2.1.13), and furthermore, its geodesics are globally defined (e.g. Lang (1999, Chapter XII)).

Remark 3.1.1. In Lang (1999), this metric is referred to as the *trace metric*. This geometry is also called *affine-invariant geometry*, as the metric is invariant under transformations $P \mapsto GPG^T$ for invertible matrices G, which correspond to affine transformations Gx of the corresponding zero-mean multivariate Gaussian $x \sim \mathcal{N}(0, P)$ (cf. Pennec et al. (2006)). Furthermore, this metric is also referred to as the *scaled Frobenius metric*, as it is a scaled version of the Frobenius metric, which is given by $\tilde{g}_P(X, Y) = \mathrm{Tr}[XY]$ (cf. Schwartzman (2006, Def. 2.2.2)). We refer to this metric as the *Fisher-information metric*, because it is the metric (times a constant) inherited from using the general Fisher-information metric for statistical manifolds onto the zero-mean multivariate Gaussians parameterized by covariance matrices $P \in \mathcal{P}$.

We to verify that the geodesics in $(\mathcal{P}, g)$ are also geodesics with respect to the induced Riemannian distance d.

With $P \in \mathcal{P}$, $X \in \mathcal{S}$ and $t \in \mathbb{R}$ denote the geodesic starting from P and emanating into the direction X by

$$P(t) := \sqrt{P}\exp\left(t\sqrt{P}^{-1}X\sqrt{P}^{-1}\right)\sqrt{P},$$

For two points on the geodesic, say $P(t_1), P(t_2)$ for some $t_1, t_2 \in \mathbb{R}$, we find that

$$P(t_1)P(t_2)^{-1} = \sqrt{P}\exp\left((t_1 - t_2)\sqrt{P}^{-1}X\sqrt{P}^{-1}\right)\sqrt{P}^{-1},$$

which has the same eigenvalues as $\exp\left((t_1 - t_2)\sqrt{P}^{-1}X\sqrt{P}^{-1}\right)$, so its eigenvalues are

$$\exp((t_1 - t_2)\mu_i), \quad i = 1, \ldots, N,$$

where $\mu_i, i = 1, \dots, N$ are the eigenvalues of $\sqrt{P}^{-1} X \sqrt{P}^{-1}$, or equivalently of $P^{-1}X$ or XP^{-1}, and they do not depend on t_1 nor t_2. Writing $c = \sqrt{\sum_{i=1}^{N} \mu_i^2}$, we conclude

$$d\big(P(t_1), P(t_2)\big) = \sqrt{\sum_{i=1}^{N} \log\big(\exp\big((t_1 - t_2)\mu_i\big)\big)^2} = c|t_1 - t_2|,$$

so $P(t)$ is indeed globally a geodesic in the sense of the metric space $(\mathcal{P}, d)$.

3.2 The Euclidean Geometry

This section provides a summary of the properties of the Riemannian manifold $(\mathcal{P}, g)$ equipped with the well-known *Euclidean metric*. Identify again $T_P\mathcal{P} \cong \mathcal{S}$ for each $P \in \mathcal{P}$. The *Euclidean metric* is a Riemannian metric (cf. Definition 2.3.1) defined on the manifold $\mathcal{P}$, given by (where $X, Y \in \mathcal{S}$)

$$g_P(X, Y) = \mathrm{Tr}[XY].$$

for each $P \in \mathcal{P}$, $X, Y \in \mathcal{S} \cong T_P\mathcal{P}$, i.e. it is independent of the point $P \in \mathcal{P}$, where the Riemannian metric is almost the scalar product of the flat Euclidean space $\mathbb{R}^{n(n+1)/2}$, except for that the off-diagonal entries are weighted twice. The induced distance is thus ($P, Q \in \mathcal{P}$; cf. Definition 2.3.5)

$$d(P, Q)^2 = \|P - Q\|_F^2,$$

where $\|\cdot\|_F$ is the Frobenius norm. The Riemann exponential and logarithm map are given by (for $g_P(X, X) < \delta$ for some $\delta > 0$)

$$\mathrm{Exp}_P(X) = P + X,$$
$$\mathrm{Log}_P(Q) = Q - P.$$

Note that we have the restriction by δ since geodesics might leave the open set $\mathcal{P} \subset \mathcal{S}$, which also implies that $(\mathcal{P}, g)$ is not geodesically complete (cf. also Schwartzman (2006, Section 2.2.5)). The geodesic from P to Q is given explicitly by

$$\gamma_{P,Q}(t) = \mathrm{Exp}_P(t\,\mathrm{Log}_P(Q)) = P + t(Q - P) = (1 - t)P + tQ.$$

Parallel transport of $X \in T_P\mathcal{P}$ from P to Q is given by

$$\Pi_{P,Q}(X) = X,$$

use Lee (1997, Equation (4.13), page 61) and the fact that the Christoffel symbols of the corresponding Levi-Civita connection are all zero (Lee (1997, page 52)).

3.3 The Bures-Wasserstein Geometry

This section provides a summary of the properties of the Riemannian manifold $(\mathcal{P}, g)$ equipped with the *Bures-Wasserstein metric*. Identify again $T_P\mathcal{P} \cong \mathcal{S}$ for each $P \in \mathcal{P}$. Most of those results are from Bhatia et al. (2017), although the metric is defined for complex-valued matrices, whereas we restrict ourselves to the real-valued case. It is a Riemannian metric defined on the manifold $\mathcal{P}$, given by

$$g_P(X, Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_j \frac{\tilde{X}_{ij}\tilde{Y}_{ij}}{(\lambda_i + \lambda_j)^2} = \mathbb{1}'(\tilde{X} \circ M \circ \tilde{Y})\lambda,$$

where $P = U\,\mathrm{Diag}(\lambda)\,U^T$ is the unique eigenvalue decomposition with eigenvalues $\lambda = (\lambda_i)_{i=1}^n > 0$ and U orthonormal, and where

$$M = \left(\frac{1}{(\lambda_i + \lambda_j)^2}\right)_{i,j=1}^{n}, \qquad \tilde{X} = (\tilde{X}_{ij})_{i,j=1}^n = U^T X U, \qquad \tilde{Y} = U^T Y U = (\tilde{Y}_{ij})_{i,j=1}^n.$$

In a more implicit manner, the metric can also be expressed as

$$g_P(X, Y) = \mathrm{Tr}\left[KAH\right],$$

where K and H are the unique solutions to the equations $KP + PK = Y$ and $HP + PH = X$, respectively (cf. Bhatia et al. (2017)). This yields the distance

$$\begin{aligned} d(P, Q)^2 &= \mathrm{Tr}\left[P\right] + \mathrm{Tr}\left[Q\right] - 2\,\mathrm{Tr}\left[\left(\sqrt{P}Q\sqrt{P}\right)^{1/2}\right] \\ &= \mathrm{Tr}\left[P\right] + \mathrm{Tr}\left[Q\right] - 2\,\mathrm{Tr}\left[\sqrt{PQ}\right]. \end{aligned}$$

The geodesic from P to Q is given explicitly by

$$\gamma_{P,Q}(t) = (1-t)^2 P + t^2 Q + t(1-t)\left[\sqrt{PQ} + \sqrt{QP}\right].$$

An expression for the Riemannian exponential can be derived by setting (for the third equality, see eg. Yair et al. (2019))

$$\begin{aligned} X = \gamma'_{P,Q}(0) &= -2P + \sqrt{PQ} + \sqrt{QP} \\ &= -2P + \sqrt{P}^{-1}\big(\sqrt{P}Q\sqrt{P}\big)^{1/2}\sqrt{P} + \sqrt{P}\big(\sqrt{P}Q\sqrt{P}\big)^{1/2}\sqrt{P}^{-1} \end{aligned}$$

which one can then rewrite with $H^2 = \sqrt{P}Q\sqrt{P}$ to

$$\sqrt{P}X\sqrt{P} + 2P^2 = HP + PH,$$

which has a unique solution $H = H(P, X)$ (solving a continuous Lyapunov equation), such that we find $Q = Q(P, X) = \sqrt{P}^{-1}H^2\sqrt{P}^{-1}$. The Riemann exponential and logarithm map are given by

$$\begin{aligned} \mathrm{Exp}_P(X) &= \gamma_{P,Q(P,X)}(1) = Q(P, X) \\ \mathrm{Log}_P(Q) &= -2P + \sqrt{PQ} + \sqrt{QP}. \end{aligned}$$

Remark 3.3.1. Notably, Bhatia et al. (2017) use the framework of a Riemannian submersion (cf. Definition 2.5.1) to derive the formulas (in their case complex matrices, we use real matrices). Denote the general linear group of invertible $n \times n$ matrices by $GL(n)$, the orthonormal matrices by $O(n)$. Then the map $\pi\colon GL(n) \to \mathcal{P}$, $\pi(A) = P$, where $A = UP$ is the polar decomposition (cf. Hall (2015, Proposition 2.19)) of a matrix $A \in GL(n)$, where $U \in O(n)$ and $P \in \mathcal{P}$, is a submersion. The Lie group $O(n)$ acting on $GL(n)$ via $\theta_V(A) = VA$ satisfies then all the requirements in Theorem 2.5.6, such that π is a Riemannian submersion.

3.4 The Log-Euclidean Distance

Identify again $T_P\mathcal{P} \cong \mathcal{S}$ for each $P \in \mathcal{P}$. The idea of this metric is to pull back points from $\mathcal{P}$ to $\mathcal{S}$ via the matrix logarithm $\log\colon \mathcal{P} \to \mathcal{S}$ and impose the Euclidean geometry of $\mathcal{S}$ (cf. Arsigny et al. (2005, 2006a,b)). The distance between two points $P, Q \in \mathcal{P}$ is then simply

$$d(P, Q) = \|\log(P) - \log(Q)\|_2.$$

To derive the Riemannian metric, one needs to compute the push-forwards $\log_*$ and $\exp_*$ (cf. Equation (2.2.1)), where we use the notation (making the points at which the derivative is taken explicit)

$$(\partial \log)_P \coloneqq \log_* \colon \mathcal{S} \cong T_P\mathcal{P} \to T_{\log(P)}\mathcal{S} \cong \mathcal{S},$$
$$(\partial \exp)_S \coloneqq \exp_* \colon \mathcal{S} \cong T_S\mathcal{S} \to T_{\exp(S)}\mathcal{P} \cong \mathcal{S}.$$

A formula for $(\partial \exp)_S$ is provided in Arsigny et al. (2006a, Equation (2.1)) through differentiation of Equation (3.0.1), as well as a formula for $(\partial \log)_P$ via the inverse of $(\partial \exp)_S$, where $X \in \mathcal{S}$:

$$(\partial \exp)_S(X) = \sum_{k=1}^{\infty} \frac{1}{k!}\Bigg(\sum_{l=0}^{k-1} S^{k-l-1} X S^l\Bigg),$$
$$(\partial \log)_P(X) = \big(\partial \exp\big)^{-1}_{\log(P)}(X).$$

Remarkably, Thanwerdas & Pennec (2021b) exploit an idea, originally developed in Bhatia (1997, page 124), to calculate the differential of univariant functions, that are extensions of maps (we assume they are analytic) $f \colon \mathbb{R} \to \mathbb{R}$ to $\mathcal{S} \to \mathcal{S}$ via the definition

$$f(S) = f\big(U \operatorname{Diag}(\lambda_1, \dots, \lambda_n)\, U^T\big) \coloneqq U \operatorname{Diag}(f(\lambda_1), \dots, f(\lambda_n))\, U^T,$$

where $U\operatorname{Diag}(\lambda_1, \dots, \lambda_n)U^T$ is the spectral decomposition of the symmetric matrix $S \in \mathcal{S}$ (e.g. Hoffman & Kunze (1971, Chapter 8, Theorem 18)). Then by Bhatia (1997, Corollary V.3.2) the differential of f is (where $X \in \mathcal{S}$ and with spectral decomposition $S = U\Lambda U^T$)

$$(\partial f)_S(X) = U\Big(f^{[1]}(\Lambda) \circ \big(U^T X U\big)\Big)U^T,$$

where $\circ$ is the Hadamard product between matrices and $f^{[1]}(\Lambda)$ is the *first divided difference* of f at $\Lambda = \operatorname{Diag}(\lambda_1, \dots, \lambda_n)$, defined as

$$\big(f^{[1]}(\Lambda)\big)_{ij} \coloneqq \begin{cases} \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j, \\ f'(\lambda_i) & \text{else,} \end{cases}$$

where f' is the first derivative of $f \colon \mathbb{R} \to \mathbb{R}$ (which exists as f is analytic). We can apply this idea to calculate $(\partial \exp)_S(X)$ and $(\partial \log)_P(X)$ directly (where $S = U\Lambda U^T$ is again the spectral decomposition, as well as (with abuse of notation) $P = U\Lambda U^T$, where, due to $P \in \mathcal{P}$, $\lambda_i > 0$ for $i = 1, \dots, n$)

$$\begin{aligned}(\partial \exp)_S(X) &= U\Big(\exp^{[1]}(\Lambda) \circ \big(U^T X U\big)\Big) U^T,\\ (\partial \log)_P(X) &= U\Big(\log^{[1]}(\Lambda) \circ \big(U^T X U\big)\Big) U^T.\end{aligned}$$

We proceed to verify that with these formulas we have indeed $(\partial \log)_P(X) \overset{!}{=} \big(\partial \exp\big)^{-1}_{\log(P)}(X)$ (as stated in Arsigny et al. (2006a)). Note that $\log(P) = U \log(\Lambda) U^T$. Observe that (where $Y \in \mathcal{S}$)

$$\begin{aligned}X = \big(\partial \exp\big)_{\log(P)}(Y) &= U\Big(\exp^{[1]}(\log(\Lambda)) \circ \big(U^T Y U\big)\Big) U^T\\ U^T X U &= \exp^{[1]}(\log(\Lambda)) \circ \big(U^T Y U\big),\end{aligned}$$

where

$$\Big(\exp^{[1]}(\log(\Lambda))\Big)_{ij} = \begin{cases} \frac{\lambda_i - \lambda_j}{\log \lambda_i - \log \lambda_j} & \text{if } \lambda_i \neq \lambda_j,\\ \lambda_i & \text{else,}\end{cases}$$

and since the derivative of $g(x) = \log(x)$ is $f'(x) = 1/x$, the element-wise inverse of this matrix is just $\log^{[1]}(\Lambda)$, which yields

$$\begin{aligned}U^T X U &= \exp^{[1]}(\log(\Lambda)) \circ \big(U^T Y U\big)\\ \log^{[1]}(\Lambda) \circ \big(U^T X U\big) &= U^T Y U\\ (\partial \log)_P(X) = U\Big(\log^{[1]}(\Lambda) \circ \big(U^T X U\big)\Big) U^T &= Y = \big(\partial \exp\big)^{-1}_{\log(P)}(X),\end{aligned}$$

which yields the assertion.

Let $P, Q \in \mathcal{P}$, $X, Y \in \mathcal{S}$. The Riemannian metric is given by (cf. Arsigny et al. (2006a))

$$g_P(X, Y) = \mathrm{Tr}\Big[(\partial \log)_P(X)\, (\partial \log)_P(Y)\Big].$$

The Riemann exponential and logarithm map are given by (cf. Arsigny et al. (2006a))

$$\begin{aligned}\mathrm{Exp}_P(X) &= \exp\Big(\log(P) + (\partial \log)_P(X)\Big),\\ \mathrm{Log}_P(Q) &= (\partial \exp)_{\log(P)}\big(\log(Q) - \log(P)\big).\end{aligned}$$

The geodesic from P to Q, $t \in \mathbb{R}$, is given explicitly by (cf. Arsigny et al. (2006a))

$$\gamma_{P,Q}(t) = \exp\Big((1-t)\log(P) + t\log(Q)\Big).$$

3.5 Quotient Geometry for Correlation Matrices

This Riemannian metric on the manifold of correlation matrices $\mathcal{C}$ was originally studied by Paul David in his PhD thesis (David, 2019) and subsequent publication (David & Gu, 2019). Thanwerdas & Pennec (2021a) provide calculation recipes for many quantities.

For a vector $x \in \mathbb{R}^n$, define $\mathrm{Diag}(x) \in \mathbb{R}^{n\times n}$ to be the $n \times n$ matrix with $x_1, \dots, x_n$ on its diagonal entries, zeros elsewhere. For a matrix $X \in \mathbb{R}^{n\times n}$, define $\mathrm{Diag}(X) \in \mathbb{R}^{n\times n}$ to be the $n\times n$ matrix with $X_{11}, \dots, X_{nn}$ on its diagonal entries, zeros elsewhere. Finally, denote by $\mathrm{diag}(X)$ the vector $(X_{11}, \dots, X_{nn}) \in \mathbb{R}^n$. Denote by $X \circ Y$ the *Hadamard product*, that is the element-wise multiplication of the respective entries.

For a matrix $P \in \mathcal{P}$, define $\Lambda_P := \sqrt{\mathrm{Diag}(P)}^{-1}$, where the square root is the unique matrix square root corresponding to taking the square root each element for diagonal matrices, and note that the inverse of a diagonal matrix corresponds to inverting the diagonal entries. Thus we can write $\Lambda_P = \mathrm{Diag}(\sqrt{P_{11}}^{-1}, \dots, \sqrt{P_{nn}}^{-1})$. Define the map

$$\pi\colon \mathcal{P} \to \mathcal{C}, \quad P \mapsto \Lambda_P P \Lambda_P, \tag{3.5.1}$$

which is clearly well-defined (i.e. $\pi(P) \in \mathcal{C}$), since $(\pi(P))_{ii} = \sqrt{P_{ii}}^{-1} P_{ii} \sqrt{P_{ii}}^{-1} = 1$ for $i = 1, \dots, n$, and since Λ_P is invertible and $P \in \mathcal{P}$, $\pi(P) \in \mathcal{P}$. Note that $\mathcal{C}$ is an $n(n-1)/2$-dimensional manifold with the Euclidean topology, and that the tangent space of $\mathcal{C}$ at some $C \in \mathcal{C}$ can be represented via the *hollow matrices*, i.e.

$$T_C\mathcal{C} \cong \mathcal{H} = \{H \in \mathcal{S} \colon \mathrm{diag}(H) = 0\}.$$

Theorem 3.5.1. *Let $(\mathcal{P}, g^{(\mathcal{P})})$ be the Riemannian manifold equipped with the Fisher-information geometry from Section 3.1.*

1. *The map π as defined in Equation* (3.5.1) *is a surjective submersion.*
2. *The map $\theta\colon \mathcal{D} \times \mathcal{P} \to \mathcal{P}$, $\theta_D(P) = DPD$, is a Lie group action on $\mathcal{P}$ that satisfies the requirements from Theorem 2.5.6.*
3. *There exists a unique Riemannian metric $g^{(\mathcal{C})}$ on $\mathcal{C}$ such that π is a Riemannian submersion (and $(\mathcal{C}, g^{(\mathcal{C})})$ is a smooth Riemannian manifold).*

Proof. 1. Clearly π is surjective since $\pi(C) = C$ for all $C \in \mathcal{C} \subset \mathcal{P}$. Further, π is smooth as it is smooth in its matrix entries. To see that π is a submersion, compute its differential (cf. Thanwerdas & Pennec (2021a, Appendix A)). Let $P(t)$ be a curve in $\mathcal{P}$ with $P(0) = P$ and $P'(0) = V \in T_P\mathcal{P} \cong \mathcal{S}$, where $t \in (-\varepsilon, \varepsilon)$ for some small $\varepsilon > 0$. Then

$$\begin{aligned}(\partial\pi)_P(V) &= \frac{\partial}{\partial t}\Big(P_{ii}(t)^{-1/2}\, P_{ij}(t)\, P_{jj}(t)^{-1/2}\Big)_{i,j=1}^{n}\Big|_{t=0} \\ &= \Big(-\frac{1}{2}V_{ii}P_{ii}^{-3/2}P_{ij}P_{jj}^{-1/2} + P_{ii}^{-1/2}V_{ij}P_{jj}^{-1/2} - \frac{1}{2}P_{ii}^{-1/2}P_{ij}V_{jj}P_{jj}^{-3/2}\Big)_{i,j=1}^{n} \qquad (3.5.2)\\ &= \Lambda_P^{-1}\Big(V - \frac{1}{2}\Big(\Lambda_P^{-2}\,\mathrm{Diag}(V)\,P + P\,\mathrm{Diag}(V)\,\Lambda_P^{-2}\Big)\Big)\Lambda_P^{-1}, \qquad (3.5.3)\end{aligned}$$

where from Equation (3.5.2) it is immediate that the diagonal is zero, so $(\partial\pi)_P(V) \in \mathcal{H}$. To see that $(\partial\pi)_P$ is surjective, let $H \in \mathcal{H}$ and let $\Lambda_P H \Lambda_P \in \mathcal{S} \cong T_P\mathcal{P}$, where $\mathrm{Diag}(\Lambda_P H \Lambda_P) = 0$ and thus $(\partial\pi)_P(\Lambda_P H \Lambda_P) = H$.
2. First, to see that $\mathcal{D}$ acts on $\mathcal{P}$ by isometries of $g^{(\mathcal{P})}$, observe that $(\partial\theta_D)_P(X) = DXD$ for any $X \in \mathcal{S} \cong T_P\mathcal{P}$, and thus for all $X, Y \in \mathcal{S} \cong T_P\mathcal{P}$,

$$\begin{aligned}g^{(\mathcal{P})}_{\theta_D(P)}\Big((\partial\theta_D)_P(X), (\partial\theta_D)_P(Y)\Big) &= \mathrm{Tr}\left[D^{-1}P^{-1}D^{-1}DXDD^{-1}P^{-1}D^{-1}DYD\right] \\ &= \mathrm{Tr}\left[P^{-1}XP^{-1}Y\right] \\ &= g^{(\mathcal{P})}_P(X, Y).\end{aligned}$$

Secondly, we show $\pi \circ \theta_D = \pi$. Let $P \in \mathcal{P}$, $D \in \mathcal{D}$ and observe that

$$\Lambda_{\theta_D(P)} = \mathrm{Diag}(D_{11}^{-1}\sqrt{P_{11}}^{-1}, \ldots, D_{nn}^{-1}\sqrt{P_{nn}}^{-1})$$

and consequently

$$\pi(\theta_D(P)) = \Lambda_{\theta_D(P)}\theta_D(P)\Lambda_{\theta_D(P)} = \Big(D_{ii}^{-1}\sqrt{P_{ii}}^{-1}D_{ii}P_{ij}D_{jj}\sqrt{P_{jj}}^{-1}D_{jj}^{-1}\Big)_{i,j=1}^{n} = \Lambda_P P \Lambda_P = \pi(P).$$

Thirdly, to see that $\mathcal{D}$ acts transitively on each fiber $\mathcal{P}_C = \pi^{-1}(C)$ for any $C \in \mathcal{C}$, let $P, Q \in \mathcal{P}$ such that $\pi(P) = \pi(Q) = C$. Then, for all $i, j = 1, \ldots, n$,

$$\sqrt{P_{ii}}^{-1}P_{ij}\sqrt{P_{jj}}^{-1} = \sqrt{Q_{ii}}^{-1}Q_{ij}\sqrt{Q_{jj}}^{-1} \iff P_{ij} = \frac{\sqrt{P_{ii}}}{\sqrt{Q_{ii}}}Q_{ij}\frac{\sqrt{P_{jj}}}{\sqrt{Q_{jj}}},$$

and thus we have $P = \theta_D(Q)$ for $D = \mathrm{Diag}(\sqrt{P_{11}}/\sqrt{Q_{11}}, \dots, \sqrt{P_{nn}}/\sqrt{Q_{nn}}) \in \mathcal{D}$ yields the assertion. Therefore θ fulfills the requirements of Theorem 2.5.6.
3. Follows from Theorem 2.5.6. ■

Having established the foundation that we are in the setting of a Riemannian submersion (cf. Section 2.5), we continue to compute the quantities of the Riemannian manifold $(\mathcal{C}, g^{(\mathcal{C})})$, and thereby follow Thanwerdas & Pennec (2021a) very closely. For $P \in \mathcal{P}$, the tangent space $T_P\mathcal{P}$ decomposes into the orthogonal sum of the vertical space $V_P\mathcal{P}$ and horizontal space $H_P\mathcal{P}$, i.e. $T_P\mathcal{P} = V_P\mathcal{P} \oplus H_P\mathcal{P}$. We find for $X \in \mathcal{S} \cong T_P\mathcal{P}$ (where $\mathbf{1}$ denotes the n-dimensional vector of ones; cf. also Thanwerdas & Pennec (2021a, Appendix A))

$$\begin{aligned} X \in V_P\mathcal{P} &\iff (\partial\pi)_P(X) = 0 \\ &\overset{(3.5.2)}{\iff} X_{ij} = \frac{1}{2}P_{ij}\left(\frac{X_{ii}}{P_{ii}} + \frac{X_{jj}}{P_{jj}}\right) \quad \text{for all } i,j = 1,\dots,n \\ &\iff X \in \left\{P \circ \left(x\mathbf{1}^T + \mathbf{1}x^T\right) : x \in \mathbb{R}^n\right\} \\ &\iff X \in \left\{DP + PD : D \in \mathcal{D}\right\}. \end{aligned}$$

Note that from this calculation it is evident that $V_P\mathcal{P}$ is an n-dimensional vector space, which makes sense since $\dim(\mathcal{P}) = n(n+1)/2$, $\dim(\mathcal{C}) = n(n-1)/2$, and $H_P\mathcal{P}$ is isomorphic to $T_{\pi(P)}\mathcal{C}$, and $\dim(\mathcal{P}) - \dim(\mathcal{C}) = n$. For the horizontal space $H_P\mathcal{P} = (V_P\mathcal{P})^\perp$, we find (cf. also Thanwerdas & Pennec (2021a, Appendix A))

$$\begin{aligned} X \in H_P\mathcal{P} &\iff \mathrm{Tr}\left[P^{-1}XP^{-1}Y\right] = 0 \quad \text{for all } Y \in V_P\mathcal{P} \\ &\iff \mathrm{Tr}\left[P^{-1}\big(DP + PD\big)P^{-1}X\right] = 0 \quad \text{for all } D \in \mathcal{D} \\ &\iff \mathrm{Tr}\left[D\big(P^{-1}X + XP^{-1}\big)\right] = 0 \quad \text{for all } D \in \mathcal{D} \\ &\iff P^{-1}X + XP^{-1} \in \mathcal{H} \iff P^{-1}X \in \mathcal{H} \iff XP^{-1} \in \mathcal{H} \\ &\iff P^{-1}X + XP^{-1} = H \text{ for some } H \in \mathcal{H}, \end{aligned}$$

where the last expression means that X is the solution to Sylvester's equation $P^{-1}X + XP^{-1} = H$ for some $H \in \mathcal{H}$.

Remark 3.5.2. Notably, vectors $X \in H_P\mathcal{P}$ need not necessarily satisfy $\mathrm{diag}(X) = 0$, as opposed to what one might think in the first place. This is only true for all vectors in the tangent spaces of the correlation matrices $\mathcal{C}$.

We continue to compute the orthogonal projections from $T_P\mathcal{P}$ onto $V_P\mathcal{P}$ and $H_P\mathcal{P}$, respectively (cf. Thanwerdas & Pennec (2021a, Appendix A)). We start with the vertical projection

$\mathfrak{v}_P\colon T_P\mathcal{P} \to V_P\mathcal{P}$ by solving $X = DP + PD + Y$ for $D \in \mathcal{D}$, where $X \in \mathcal{S} \cong T_P\mathcal{P}$ and $Y \in H_P\mathcal{P}$. Observe that

$$\begin{aligned} X = DP + PD + Y &\iff P^{-1}X = P^{-1}DP + D + P^{-1}Y \\ &\overset{P^{-1}Y \in \mathcal{H}}{\Longrightarrow} \operatorname{diag}(P^{-1}X) = (P^{-1} \circ P + I)\operatorname{diag}(D) \\ &\iff D = \operatorname{Diag}\Big((I + P^{-1} \circ P)^{-1}\operatorname{diag}(P^{-1}X)\Big). \end{aligned} \tag{3.5.4}$$

Thus, the vertical projection $\mathfrak{v}_P\colon T_P\mathcal{P} \to V_P\mathcal{P}$ is

$$\mathfrak{v}_P(X) = DP + PD, \qquad D := \operatorname{Diag}\Big((I + P^{-1} \circ P)^{-1}\operatorname{diag}(P^{-1}X)\Big). \tag{3.5.5}$$

Consequently, the horizontal projection $\mathfrak{h}_P\colon T_P\mathcal{P} \to H_P\mathcal{P}$ is given by

$$\mathfrak{h}_P(X) = X - \mathfrak{v}_P(X). \tag{3.5.6}$$

We are now ready to compute the horizontal lift of a vector $H \in \mathcal{H} \cong T_C\mathcal{C}$, $C \in \mathcal{C}$ (cf. Thanwerdas & Pennec (2021a, Theorem 2)).

Lemma 3.5.3. *The horizontal lift of $H \in \mathcal{H} \cong T_C\mathcal{C}$, $C \in \mathcal{C}$, is calculated by*

$$H^{\#} = \mathfrak{h}_P(\Lambda_P H \Lambda_P) \in H_P\mathcal{P}.$$

Proof. Let $H \in \mathcal{H} \cong T_C\mathcal{C}$. Since $(\partial\pi)_P$ is an isomorphism from $H_P\mathcal{P}$ to $T_{\pi(P)}\mathcal{C}$, it is enough to find any vector $X \in T_P\mathcal{P}$ that satisfies $(\partial\pi)_P(X) = H$ and then $H^{\#} = \mathfrak{h}_P(X) \in H_P\mathcal{P}$. Considering Equation (3.5.3), the obvious candidate is $X = \Lambda_P H \Lambda_P$ as from $\operatorname{Diag}(H) = 0$ we have that $\operatorname{Diag}(X) = 0$, and thus $(\partial\pi)_P(X) = 0$. Consequently, setting $H^{\#} = \mathfrak{h}_P(X)$ we have

$$(\partial\pi)_P(H^{\#}) = (\partial\pi)_P(\mathfrak{h}_P(\Lambda_P H \Lambda_P)) = (\partial\pi)_P(\Lambda_P H \Lambda_P) \overset{(3.5.3)}{=} H.$$

■

Using the horizontal lift, we can determine the Riemannian metric on $\mathcal{C}$ via Equation (2.5.1) (cf. Thanwerdas & Pennec (2021a, Theorem 3)).

Theorem 3.5.4. *The Riemannian metric $g^{(\mathcal{C})}$ on $\mathcal{C}$ induced by the Riemannian submersion from Equation* (3.5.1) *in the sense of Theorem 2.5.6 is given by (where $C \in \mathcal{C}$, $X, Y \in \mathcal{H} \cong T_C\mathcal{C}$, and let $P \in \pi^{-1}(C)$ be any point in the fiber of C)*

$$g_C^{(\mathcal{C})}(X,Y) = g_C^{(\mathcal{P})}\big(\Lambda_P^2 X \Lambda_P^2, \Lambda_P^2 Y \Lambda_P^2\big) - 2\,\mathrm{diag}(C^{-1}\Lambda_P^2 X \Lambda_P^2)^T (I + C \circ C^{-1})^{-1} \mathrm{diag}(C^{-1}\Lambda_P^2 Y \Lambda_P^2)$$

In particular, for $P = C \in \pi^{-1}(C)$, the formula simplifies to

$$g_C^{(\mathcal{C})}(X,Y) = g_C^{(\mathcal{P})}\big(X,Y\big) - 2\,\mathrm{diag}(C^{-1}X)^T (I + C \circ C^{-1})^{-1} \mathrm{diag}(C^{-1}Y).$$

Proof. Note that $C = \pi(P) = \Lambda_P P \Lambda_P$ implies that $P^{-1} = \Lambda_P C^{-1} \Lambda_P$. We compute the Riemannian metric using Equation (2.5.1). Let $X \in \mathcal{H} \cong T_C\mathcal{C}$ and with Equation (3.5.5), recall that (where $D \in \mathcal{D}$)

$$\mathfrak{v}_P(\Lambda_P X \Lambda_P) = DP + PD, \quad \text{where} \quad \mathrm{diag}(D) = (I + P^{-1} \circ P)^{-1} \mathrm{diag}(P^{-1}\Lambda_P X \Lambda_P). \tag{3.5.7}$$

Then

$$\begin{aligned}
g_C^{(\mathcal{C})}(X,X) &:= g_P^{(\mathcal{P})}(X^\#, X^\#) \\
&= g_P^{(\mathcal{P})}\big(\mathfrak{h}_P(\Lambda_P X \Lambda_P), \mathfrak{h}_P(\Lambda_P X \Lambda_P)\big) \\
&\overset{(3.5.6)}{=} g_P^{(\mathcal{P})}(\Lambda_P X \Lambda_P, \Lambda_P X \Lambda_P) - g_P^{(\mathcal{P})}(\mathfrak{v}_P(\Lambda_P X \Lambda_P), \mathfrak{v}_P(\Lambda_P X \Lambda_P)) \\
&\overset{(3.5.7)}{=} \mathrm{Tr}\left[P^{-1}\Lambda_P X \Lambda_P P^{-1} \Lambda_P X \Lambda_P\right] - \mathrm{Tr}\left[P^{-1}(DP + PD)P^{-1}(DP + PD)\right] \\
&= g_C^{(\mathcal{P})}\big(\Lambda_P^2 X \Lambda_P^2, \Lambda_P^2 X \Lambda_P^2\big) - \mathrm{Tr}\left[P^{-1}D^2P + DP^{-1}DP + P^{-1}DPD + D^2\right] \\
&= g_C^{(\mathcal{P})}\big(\Lambda_P^2 X \Lambda_P^2, \Lambda_P^2 X \Lambda_P^2\big) - 2\ \mathrm{Tr}\left[P^{-1}DPD + D^2\right] \\
&= g_C^{(\mathcal{P})}\big(\Lambda_P^2 X \Lambda_P^2, \Lambda_P^2 X \Lambda_P^2\big) - 2\,\mathrm{diag}(D)^T (I + P \circ P^{-1}) \mathrm{diag}(D) \\
&\overset{(3.5.7)}{=} g_C^{(\mathcal{P})}\big(\Lambda_P^2 X \Lambda_P^2, \Lambda_P^2 X \Lambda_P^2\big) - 2\,\mathrm{diag}(P^{-1}\Lambda_P X \Lambda_P)^T (I + P \circ P^{-1})^{-1} \mathrm{diag}(P^{-1}\Lambda_P X \Lambda_P) \\
&= g_C^{(\mathcal{P})}\big(\Lambda_P^2 X \Lambda_P^2, \Lambda_P^2 X \Lambda_P^2\big) - 2\,\mathrm{diag}(C^{-1}\Lambda_P^2 X \Lambda_P^2)^T (I + C \circ C^{-1})^{-1} \mathrm{diag}(C^{-1}\Lambda_P^2 X \Lambda_P^2).
\end{aligned}$$

Plugging this into $g_C^{(\mathcal{C})}(X,Y) = \frac{1}{2} g_C^{(\mathcal{C})}(X-Y, X-Y) - g_C^{(\mathcal{C})}(X,X) - g_C^{(\mathcal{C})}(Y,Y)$ yields

$$g_C^{(\mathcal{C})}(X,Y) = g_C^{(\mathcal{P})}\big(\Lambda_P^2 X \Lambda_P^2, \Lambda_P^2 Y \Lambda_P^2\big) - 2\,\mathrm{diag}(C^{-1}\Lambda_P^2 X \Lambda_P^2)^T (I + C \circ C^{-1})^{-1} \mathrm{diag}(C^{-1}\Lambda_P^2 Y \Lambda_P^2)$$

and if we choose $P = C \in \pi^{-1}(C)$, then $\Lambda_P = I$ and thus

$$g_C^{(\mathcal{C})}(X,Y) = g_C^{(\mathcal{P})}\big(X,Y\big) - 2\,\mathrm{diag}(C^{-1}X)^T (I + C \circ C^{-1})^{-1} \mathrm{diag}(C^{-1}Y).$$

■

To compute the distance, plugging the distance $d_{\mathcal{P}}$ of $\mathcal{P}$ from Equation (3.1.1) into Equation (2.5.2) yields

$$d^2_{\mathcal{C}}(C_1, C_2) = \inf_{D\in\mathcal{D}} d^2_{\mathcal{P}}(C_1, DC_2D). \tag{3.5.8}$$

This is a minimization problem over n variables on the set $\mathcal{D} \cong (0,\infty)^n$. Let $C \in \mathcal{C}$ and $X \in \mathcal{H} \cong T_C\mathcal{C}$. In Thanwerdas & Pennec (2021a, Theorem 4), the Riemann exponential map is derived for $P = C \in \pi^{-1}(C)$, but we shall also provide the formula for general $P \in \pi^{-1}(C)$, where $X^\# \in H_P\mathcal{P}$ is the horizontal lift of X at P:

$$\mathrm{Exp}^{(\mathcal{C})}_C(X) = \pi\big(\mathrm{Exp}^{(\mathcal{P})}_P(X^\#)\big). \tag{3.5.9}$$

The Riemannian logarithm between $C_1, C_2 \in \mathcal{C}$ is computed by searching for a horizontal direction $X \in H_{C_1}\mathcal{P}$ with $\mathrm{Exp}^{(\mathcal{P})}_{C_1}(X) = C_2$, which is achieved by finding a minimizer

$$D^* = \arg\min_{D\in\mathcal{D}} d^2_{\mathcal{P}}(C_1, DC_2D), \tag{3.5.10}$$

and then setting $X := \mathrm{Log}^{(\mathcal{P})}_{C_1}(D^*C_2D^*) \in H_{C_1}\mathcal{P}$. Hence the Riemannian logarithm in $\mathcal{C}$ is (cf. Thanwerdas & Pennec (2021a, p.5); and note that again, another choice than $C_1 \in \pi^{-1}(C_1)$ and $C_2 \in \pi^{-1}(C_2)$ can be made)

$$\mathrm{Log}^{(\mathcal{C})}_{C_1}(C_2) = (\partial\pi)_{C_1}\Big(\mathrm{Log}^{(\mathcal{P})}_{C_1}(D^*C_2D^*)\Big). \tag{3.5.11}$$

Consequently, the geodesic from C_1 to $C_2 \in \mathcal{C}$ is

$$\gamma^{(\mathcal{C})}_{C_1,C_2}(t) = \mathrm{Exp}^{(\mathcal{C})}_{C_1}\Big(t\,\mathrm{Log}^{(\mathcal{C})}_{C_1}(C_2)\Big) = \pi\Big(\gamma^{(\mathcal{P})}_{C_1,D^*C_2D^*}(t)\Big). \tag{3.5.12}$$

Note that $\gamma^{(\mathcal{C})}_{C_1,C_2}(0) = \pi(C_2) = C_2$ and $\gamma^{(\mathcal{C})}_{C_1,C_2}(1) = \pi(D^*C_2D^*) = C_2$.

Remark 3.5.5. For the minimization problem stated in Equation (3.5.10), which is also necessary to solve in order to compute the distance on $\mathcal{C}$ (cf. Equation (3.5.8)), in their first version, the authors of Thanwerdas & Pennec (2021a) state that the uniqueness and even existence of the minimizer D^* is not proven yet. Prior to the International Conference on Geometric Science of Information (GSI) in July 2021 in Paris, where Thanwerdas & Pennec (2021a) was published, I have made several attempts to find a solution to Equation (3.5.10), but apparently with no success. However, when I met Yann Thanwerdas at GSI 2021, we discussed this problem and could prove coercivity of the function to be minimized on the spot, which implies existence of a solution to Equation (3.5.10). This result in addition to the computation of the gradient of the function to be minimized, which is an outcome of my attempts of solving the minimization problem, is stated in the next lemma.

Lemma 3.5.6. *Let $P, Q \in \mathcal{P}$ be arbitrary. Define the map $F_{P,Q} \colon (0,\infty)^n \to [0,\infty)$ (where $D_x := \operatorname{Diag}(x)$ for $x \in (0,\infty)^n$) with*

$$F_{P,Q}(x) = d^2_{\mathcal{P}}(P, D_x Q D_x) = \operatorname{Tr}\left[\log\left(\sqrt{P}^{-1} D_x Q D_x \sqrt{P}^{-1}\right)^2\right].$$

The gradient of $F_{P,Q}$ at $x \in (0,\infty)^n$ is

$$\vec{\nabla} F_{P,Q}(x) = 4\operatorname{diag}\Big(\sqrt{D_x}^{-1}\sqrt{Q}^{-1}\log\left(\sqrt{Q} D_x P^{-1} D_x \sqrt{Q}\right)\sqrt{Q}\sqrt{D_x}^{-1}\Big) \in \mathbb{R}^n.$$

Furthermore, $F_{P,Q}$ is coercive (i.e. $F_{P,Q}(x) \to \infty$ whenever x tends to some boundary of $(0,\infty)^n$, including going to ∞) and there is a solution to the minimization problem

$$x^* \in \operatorname*{arg\,min}_{x \in (0,\infty)^n} F_{P,Q}(x).$$

Proof. Coercivity and existence of a solution will be in a future version of Thanwerdas & Pennec (2021a). We continue to compute the gradient of $F_{P,Q}$. Let $x = (x_1, \dots, x_n) \in (0,\infty)^n$ and let $k = 1, \dots, n$. The equalities that are given numbers are explained below.

$$\begin{aligned}
\frac{\partial F_{P,Q}}{\partial x_k}(x) &= \frac{\partial}{\partial x_k}\operatorname{Tr}\left[\log\left(\sqrt{P}^{-1} D_x Q D_x \sqrt{P}^{-1}\right)^2\right] \\
&\overset{(1)}{=} 2\operatorname{Tr}\left[\log\left(\sqrt{P}^{-1} D_x Q D_x \sqrt{P}^{-1}\right)\sqrt{P} D_x^{-1} Q^{-1} D_x^{-1}\sqrt{P}\frac{\partial}{\partial x_k}\left(\sqrt{P}^{-1} D_x Q D_x \sqrt{P}^{-1}\right)\right] \\
&\overset{(2)}{=} 2\operatorname{Tr}\left[\left(Q^{-1/2}\log\left(Q^{1/2} D_x P^{-1} D_x Q^{1/2}\right) Q^{-1/2}\right)\left(D_x^{-1}\frac{\partial}{\partial x_k}\left(D_x Q D_x\right) D_x^{-1}\right)\right],
\end{aligned} \tag{3.5.13}$$

where (1) follows from Moakher (2005, Proposition 2.1), where a general formula for differentiating the trace of a squared matrix logarithm is given; and (2) follows from the invariance of the matrix logarithm under the congruence action, i.e. $\log(A^{-1}BA) = A^{-1}\log(B)A$ for, say, $A, B \in \mathcal{P}$ (e.g. Moakher (2005, Equation (2.1))), as well as the cyclic property of the trace: $\mathrm{Tr}[AB] = \mathrm{Tr}[BA]$. In the following, let denote by δ_{ij} the Kronecker delta, i.e. $\delta_{ij} = 0$ whenever $i \neq j$, $\delta_{ij} = 1$ if $i = j$. Then

$$
\begin{aligned}
D_x^{-1} \frac{\partial}{\partial x_k} \big(D_x Q D_x\big) D_x^{-1} &= D_x^{-1} \Big(\delta_{ik} Q_{ij} x_j + \delta_{jk} Q_{ij} x_i\Big)_{i,j=1}^{n} D_x^{-1} \\
&= \left(\frac{\delta_{ik} Q_{ij}}{x_i} + \frac{\delta_{jk} Q_{ij}}{x_j}\right)_{i,j=1}^{n} = \frac{1}{x_k} \Big(\delta_{ik} Q_{ij} + \delta_{jk} Q_{ij}\Big)_{i,j=1}^{n}.
\end{aligned}
$$

For any matrix $A = (A_{ij})_{i,j=1}^{n}$, we compute

$$
\begin{aligned}
2\,\mathrm{Tr}\left[A\, D_x^{-1} \frac{\partial}{\partial x_k} \big(D_x Q D_x\big) D_x^{-1}\right] &= \frac{2}{x_k} \mathrm{Tr}\left[\Big(\sum_{l=1}^{n} A_{il} \big(\delta_{lk} Q_{lj} + \delta_{jk} Q_{lj}\big)\Big)_{i,j=1}^{n}\right] \\
&= \frac{2}{x_k} \sum_{i=1}^{n} \Big(\sum_{l=1}^{n} A_{il} \big(\delta_{lk} Q_{li} + \delta_{ik} Q_{li}\big)\Big) \\
&= \frac{2}{x_k} \sum_{i=1}^{n} \Big(\sum_{l=1}^{n} A_{il} \big(\delta_{lk} Q_{li} + \delta_{ik} Q_{li}\big)\Big) \\
&= \frac{2}{x_k} \Big(\sum_{i=1}^{n} A_{ik} Q_{ki} + \sum_{l=1}^{n} A_{kl} Q_{lk}\Big) \\
&= \frac{4}{x_k} (AQ)_{kk}.
\end{aligned}
$$

Setting $A := \sqrt{Q}^{-1} \log\big(\sqrt{Q} D_x P^{-1} D_x \sqrt{Q}\big) \sqrt{Q}^{-1}$ and using the above result, we obtain for Equation (3.5.13)

$$
\vec{\nabla} F_{P,Q}(x) = 4\,\mathrm{diag}\Big(\sqrt{D_x}^{-1} \sqrt{Q}^{-1} \log\big(\sqrt{Q} D_x P^{-1} D_x \sqrt{Q}\big) \sqrt{Q} \sqrt{D_x}^{-1}\Big).
$$

■

Remark 3.5.7. 1. The gradient computed in Lemma 3.5.6 can be implemented in order to use gradient descent methods to solve the minimization problem. Observe that by setting the gradient equal to zero, i.e. $\vec{\nabla} F_{P,Q}(x) = 0$, we can reduce the equation to a seemingly simple equation:

$$
\begin{aligned}
& 0 = \vec{\nabla} F_{P,Q}(x) \\
\iff\ & 0 = 4\,\mathrm{diag}\Big(\sqrt{D_x}^{-1} \sqrt{Q}^{-1} \log\big(\sqrt{Q} D_x P^{-1} D_x \sqrt{Q}\big) \sqrt{Q} \sqrt{D_x}^{-1}\Big) \\
\iff\ & 0 = \mathrm{diag}\Big(\log\big(D_x P^{-1} D_x Q\big)\Big).
\end{aligned}
$$

Moreover, observe that whenever $P = D_x Q D_x$, then $\log\left(D_x P^{-1} D_x Q\right) = \log(I) = 0$, where not only the diagonal entries are zero, but the whole matrix.

2. Note that the formula for the gradient given in Lemma 3.5.6 is not the shortest version, as the $\sqrt{Q}$ terms can be pulled into the matrix logarithm (as has been done in 1. of this remark). In practice however, taking the matrix logarithm of a symmetric matrix is much more stable, as the singular value decomposition is more stable.
3. Furthermore, one could try to compute the Hessian. The difficulty here is that differentiating $\log(X(t))$ leads to $X'(t)X^{-1}(t)$ only if $X'(t)$ and $X^{-1}(t)$ commute. This is certainly not true in general for $QD_xP^{-1}D_x$ and $\frac{\partial}{\partial x_k} QD_xP^{-1}D_x$, so we cannot use this rule. There is a more general and also more involved expression using an integral over a product of matrices, but I did not manage to make something out of it (e.g. Moakher (2005, Proof of Proposition 2.1)).

Chapter 4

Phylogenetic Forests

To this end, we introduce phylogenetic forests starting with the intuitive representation via graphs. We continue to introduce the concept of splits and derive a representation of phylogenetic forests via splits, before we define distance matrices of phylogenetic forests and from that the correlation matrix representation. All of those representations are particularly relevant for defining and working on the Wald Space that we construct in Chapter 6. In the following, $L = \{1, \dots, N\}$ is the set of labels.

4.1 Representation via Graphs

4.1.1 Some Theory on Graphs

We introduce the basic notation for graphs (e.g. Semple & Steel (2003, Section 1.1 and 1.2)).

Definition 4.1.1. 1. A *graph* is a tuple $(\mathfrak{V}, \mathfrak{E})$ with a finite non-empty set of *vertices* $\mathfrak{V}$ and *edges* $\mathfrak{E} \subset \big\{\{u, v\} \colon u, v \in \mathfrak{V}, u \neq v\big\}$.

Moreover, if $(\mathfrak{V}, \mathfrak{E})$ is a graph, then we say that

2. two vertices $u, v \in \mathfrak{V}$ are *adjacent*, if $\{u, v\} \in \mathfrak{E}$,
3. two edges $e, e' \in \mathfrak{E}$ with $e' \neq e$ are *incident*, if $|e \cap e'| = 1$,
4. an edge $e \in \mathfrak{E}$ is *incident with* a vertex $v \in \mathfrak{V}$, if $v \in e$.
5. For $v \in \mathfrak{V}$, the *degree* of v is the number of incident edges to v, i.e.

$$\deg(v) = |\{e \in \mathfrak{E} \colon v \in e\}|.$$

6. A vertex $v \in \mathfrak{V}$ is a *leaf* or *pendant vertex* if $\deg(v) = 1$ and an *interior vertex* if $\deg(v) \geq 2$.
7. A vertex $v \in \mathfrak{V}$ is an *isolated vertex*, if $\deg(v) = 0$.
8. An edge $e \in \mathfrak{E}$ is a *pendant edge*, if it is incident with a leaf $v \in \mathfrak{V}$.
9. An edge that is not a pendant edge is called *interior edge*.
10. A *path* between vertices $u, u' \in \mathfrak{V}$ is a set of edges $\mathfrak{P} \subset \mathfrak{E}$ with

$$\mathfrak{P} = \big\{\{v_1, v_2\}, \{v_2, v_3\}, \ldots, \{v_{m-1}, v_m\}\big\}$$

for a sequence of pairwise distinct vertices $v_1, v_2, \ldots, v_m$, except $v_1 = v_m$ is allowed, with $2 \leq m \in \mathbb{N}$ such that $u = v_1$ and $u' = v_m$. We say that $\mathfrak{P}$ *connects* its *endpoints* u and u'.

11. We say that two vertices $u, u' \in \mathfrak{V}$ are *connected* if there exists a path $\mathfrak{P}$ that connects u and u'.
12. We say that $(\mathfrak{V}, \mathfrak{E})$ is *connected*, if all vertices $u, v \in \mathfrak{V}$ with $u \neq v$ are connected.
13. A path is called a cycle, if it is a path from v to v for some vertex $v \in \mathfrak{V}$.

Definition 4.1.2. A graph $(\mathfrak{V}, \mathfrak{E})$ is

1. a (graph-theoretical) *forest* if whenever two vertices $u, v \in \mathfrak{V}$ with $u \neq v$ are connected, then the path that connects them is unique. In this case, denote the unique path between u and v by $\mathfrak{E}(u, v)$.
2. a (graph-theoretical) *tree*, if it is connected and a forest.

Only to this end we will refer to graph-theoretical trees and forests as trees and forests, respectively. We state part of Theorem 9.1 from Wilson (1996, p.44).

Proposition 4.1.3. *Let $(\mathfrak{V}, \mathfrak{E})$ be a graph. Then the following statements are equivalent:*

(i) $(\mathfrak{V}, \mathfrak{E})$ *is a tree;*

(ii) $(\mathfrak{V}, \mathfrak{E})$ *contains no cycles and* $|\mathfrak{E}| = |\mathfrak{V}| - 1$*;*

(iii) $(\mathfrak{V}, \mathfrak{E})$ *is connected and* $|\mathfrak{E}| = |\mathfrak{V}| - 1$*;*

(iv) $(\mathfrak{V}, \mathfrak{E})$ *is connected and for any* $e \in \mathfrak{E}$ *the graph* $(\mathfrak{V}, \mathfrak{E} \setminus \{e\})$ *is not connected.*

Definition 4.1.4. We say that a tree $(\mathfrak{V}, \mathfrak{E})$ is *binary* or *fully resolved*, if $\deg(v) = 3$ for every interior vertex $v \in \mathfrak{V}$.

The next proposition is from Semple & Steel (2003, Proposition 1.2.3).

Proposition 4.1.5. *Let* $(\mathfrak{V}, \mathfrak{E})$ *be a graph. Then,*

1. *if* $(\mathfrak{V}, \mathfrak{E})$ *is a tree, then* $|\mathfrak{E}| \geq |\mathfrak{V}| - 1$*;*
2. *if* $(\mathfrak{V}, \mathfrak{E})$ *is a binary tree, then* $|\mathfrak{E}| = 2l - 3$*, where* $l \in \mathbb{N}$ *is the number of leaves of* $(\mathfrak{V}, \mathfrak{E})$*.*

4.1.2 Phylogenetic Forests via Graphs

Definition 4.1.6. 1. A *graph-based forest representative* is a triple $(\mathfrak{V}, \mathfrak{E}, \ell)$, where

(G1) $(\mathfrak{V}, \mathfrak{E})$ is a graph-theoretical forest, called the *underlying graph* and $L \subset \mathfrak{V}$, such that $v \in \mathfrak{V} \setminus L$ implies $\deg(v) \geq 3$,

(G2) and $\ell = (\ell_e)_{e \in \mathfrak{E}} \in (0, \infty)^{\mathfrak{E}}$.

We also refer to the underlying graph $(\mathfrak{V}, \mathfrak{E})$ as the *topology* of $(\mathfrak{V}, \mathfrak{E}, \ell)$. If the underlying graph is a tree, we say *graph-based tree representative*. If the underlying graph is a fully resolved tree and $\deg(u) = 1$ for all $u \in L$, we call the graph-based tree representative and its topology *fully resolved*.

Remark 4.1.7. 1. The definition of graph-based tree representatives without the edge lengths ℓ is exactly the definition of *phylogenetic X-trees* as in Semple & Steel (2003, Definition 2.1.1 and 2.1.2) with $X = L$, where, instead of using a labeling function $\phi\colon L \to \mathfrak{V}$ that is injective, we identify $\phi(L)$ with L and have $L \subset \mathfrak{V}$.

2. The condition that $v \in \mathfrak{V} \setminus L$ implies $\deg(v) \geq 3$ means that any unlabeled vertex must have degree three or higher. This is equivalent to saying that whenever a vertex has degree two or less, it must be labeled, i.e. for any $v \in \mathfrak{V}$, $\deg(v) \leq 2 \implies v \in L$.
3. Consequently, a fully resolved tree representative has N labeled leaves, and all other vertices have degree three.
4. For any graph-based forest representative $(\mathfrak{V}, \mathfrak{E}, \ell)$ and any pair of labels $u, v \in L$ with $u \neq v$ that are in the same connected component, it follows that $\mathfrak{E}(u, v) \neq \emptyset$.

Proposition 4.1.8. *Let* $(\mathfrak{V}, \mathfrak{E}, \ell)$ *be a graph-based forest representative. If* $N = 1$*, then* $|\mathfrak{V}| = 1$ *and* $|\mathfrak{E}| = 0$*. Let* $N \geq 2$*. Then*

1. $N \leq |\mathfrak{V}| \leq 2N - 2$ *and* $0 \leq |\mathfrak{E}| \leq 2N - 3$*;*
2. $N \leq |\mathfrak{V}| \leq 2N - 2$ *and* $N - 1 \leq |\mathfrak{E}| \leq 2N - 3$ *if the underlying graph is a tree;*

3. $|\mathfrak{V}| = 2N-2$ *and* $|\mathfrak{E}| = 2N-3$ *if and only if* $(\mathfrak{V}, \mathfrak{E}, \ell)$ *is a fully resolved graph-based tree representative.*

Proof. If $N = 1$, then $|\mathfrak{V}| = 1$, else there would exist unlabeled vertices with degree two or less. It follows that $|\mathfrak{E}| = 0$.

3. The equivalence that $(\mathfrak{V}, \mathfrak{E}, \ell)$ is a fully resolved graph-based tree representative if and only if $|\mathfrak{E}| = 2N - 3$ follows from Moulton & Steel (2004, Theorem 4.2, (v)). Then by Proposition 4.1.3, $|\mathfrak{V}| = |\mathfrak{E}| + 1 = 2N - 2$.

1. Any forest representative can be extended by adding edges and vertices until it is a fully resolved tree representative, so the upper bounds are $|\mathfrak{E}| \leq 2N - 3$ and $|\mathfrak{V}| \leq 2N - 2$ (cf. Moulton & Steel (2004, Theorem 4.2, (i) and (v))). For the lower bounds, the tree with N vertices $\mathfrak{V} = L$ and no edges is a graph-based forest representative and yields $|\mathfrak{V}| = N$ and $|\mathfrak{E}| = 0$.

2. Analogously to 1., we obtain the same upper bounds. For the lower bounds, we have at least N vertices, each labeled once. The minimum number of edges we can have that connect the N vertices form a chain of $N-1$ edges, yielding $|\mathfrak{V}| \geq N$ and $|\mathfrak{E}| \geq N-1$. ∎

Definition 4.1.9. Two forest representatives $(\mathfrak{V}, \mathfrak{E}, \ell)$, $(\mathfrak{V}', \mathfrak{E}', \ell')$ are *topologically equivalent*, if there is a bijection $f \colon \mathfrak{V} \to \mathfrak{V}'$ such that

(i) $\{u, v\} \in \mathfrak{E} \iff \{f(u), f(v)\} \in \mathfrak{E}'$,

(ii) $f(u) = u$ for all $u \in L$.

They are *isomorphic* if additionally

(iii) $\ell(\{u, v\}) = \ell'\big(\{f(u), f(v)\}\big)$ for all edges $\{u, v\} \in \mathfrak{E}$.

Moreover,

1. every isomorphy class of a graph-based forest representative is called a *graph-based forest* and denoted by $\mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell]$, analogously for graph-based tree representatives and fully resolved graph-based tree representatives;
2. every topological equivalence class of a graph-based forest representative is called its *topology* $[\mathfrak{F}] = [\mathfrak{V}, \mathfrak{E}]$.

Remark 4.1.10. 1. Note that the above definition is well-defined: if one representative of $[\mathfrak{V}, \mathfrak{E}, \ell]$ is a tree, so will be every representative; analogously for fully resolved graph-based tree representatives.

2. Topological equivalence between graph-based tree representatives corresponds to the notion of *isomorphisms between phylogenetic X-trees* as defined in Semple & Steel (2003, p.17) with $X = L$, cf. also Remark 4.1.7.

4.2 Representation via Splits

4.2.1 Some Theory on Splits

To this end, let X be a finite set.

Definition 4.2.1. 1. A *split of X* or *X-split* is a two-set partition of X into two non-empty sets, i.e. a split of X is a set $\{A, B\}$ of sets $A, B \subset X$, $A, B \neq \emptyset$, $A \cup B = X$ and $A \cap B = \emptyset$. We write interchangeably

$$\{A, B\} = A|B = a_1 \dots a_r|b_1 \dots b_s = a_1 \dots a_r|B = A|b_1 \dots b_s\,,$$

whenever $A = \{a_1, \dots, a_r\}$, $B = \{b_1, \dots, b_s\}$.

2. Two splits are called *compatible* with each other, if at least one of the four expressions is empty:

$$A \cap C, \qquad A \cap D, \qquad B \cap C, \qquad B \cap D. \tag{4.2.1}$$

3. A set of splits is called *compatible*, if all its splits are pair-wise compatible.

Lemma 4.2.2. *Consider two X-splits $A|B$ and $C|D$ with $A|B \neq C|D$. The following statements are equivalent:*

(i) *$A|B$ and $C|D$ are compatible,*

(ii) *exactly one of the four expressions in* (4.2.1) *is empty:*

$$A \cap C, \qquad A \cap D, \qquad B \cap C, \qquad B \cap D.$$

(iii) *exactly one of the following statements is true:*

$$A \subset D, \quad A \subset C, \quad B \subset D, \quad B \subset C, \tag{4.2.2}$$

(iv) *exactly one of the following statements is true:*

$$C \subset B, \quad D \subset B, \quad C \subset A, \quad D \subset A. \tag{4.2.3}$$

Proof. Since $A = X \setminus B$ and $C = X \setminus D$, we find that

$$A \cap C = \emptyset \iff C \subset B \iff A \subset D,$$

yielding (ii) $\iff$ (iii) $\iff$ (iv). By definition of compatibility, (ii) $\implies$ (i). Again, from $A \cap C = \emptyset \iff C \subset B \iff A \subset D$, we find $A \cap C = \emptyset \implies B \cap C \neq \emptyset$ and $A \cap C = \emptyset \implies A \cap D \neq \emptyset$, and finally, if $B \cap D = \emptyset \implies B \subset C$, contradicting $C \subset B$ implied by $A \cap C = \emptyset$. Thus, $A \cap C = \emptyset \implies B \cap D \neq \emptyset$, and thus (i) $\implies$ (ii), which yields the assertion. ■

Note that (i) $\iff$ (ii) is mentioned but not proven in Semple & Steel (2003, p.44). Furthermore, Buneman (1971, p.388) mentions that if two intersections in (4.2.1) are empty, then $A|B = C|D$.

Definition 4.2.3. Let $\emptyset \neq Y \subset X$ be another finite set and let $s = A|B$ be a split of X. The *restriction of s to Y* is

$$s\big|_Y := (A \cap Y)|(B \cap Y). \tag{4.2.4}$$

If $s\big|_Y$ is a split of Y we say that $s\big|_Y$ is *valid*. If $s\big|_Y$ is not a split of Y we say that $s\big|_Y$ has *vanished*.

4.2.2 Phylogenetic Forests via Splits

Definition 4.2.4. A tuple (E, λ) is a *split-based forest* if

1. $\mathcal{L} = \{L_1, \dots, L_K\}$ with $1 \leq K \leq N$ such that the non-empty sets $L_1, \dots, L_K$ form a partition of the label set L;
2. each $e \in E$ is a split of L_α for some $1 \leq \alpha \leq K$;
3. E_α denotes the elements in E that are splits of L_α;
4. each E_α is compatible;
5. for all $u, v \in L_\alpha$ with $u \neq v$ there exists $e = A|B \in E$ such that $u \in A$ and $v \in B$;
6. the *edge weights* are $\lambda = (\lambda_e)_{e \in E} \in (0, 1)^E$.

We say that (E, λ) is a *split-based tree*, if $K = 1$. We call the set E the *topology* of (E, λ).

Remark 4.2.5. The partition of the label set $\mathcal{L}$ can be omitted from the tuple (E, λ) as it can be reconstructed from E: set without loss of generality

$$\{L_1, \dots, L_{\tilde{K}}\} := \{A \cup B \colon A|B \in E\},$$

where $\tilde{K} \leq K$, and for all $u \in L \setminus \bigcup_{\alpha=1}^{\tilde{K}} L_\alpha$, the singleton $\{u\}$ is added to the collection and we obtain $\mathcal{L} = \{L_1, \ldots, L_K\}$.

We will elaborate on how to obtain a split-based tree from a graph-based tree representative $(\mathfrak{V}, \mathfrak{E}, \ell)$ following Semple & Steel (2003, Section 3.1). By Proposition 4.1.3, we know that cutting an edge $e \in \mathfrak{E}$ (i.e. considering the graph $(\mathfrak{V}, \mathfrak{E} \setminus \{e\})$) splits the underlying graph into two connected components. Furthermore, it partitions the labels $L \subset \mathfrak{V}$ into two non-empty sets A_e and B_e (if one of them was empty, we would have had a vertex $v \in \mathfrak{V} \setminus \{L\}$ with $\deg(v) \leq 2$, which is by definition not possible). Note that no two edges give the same partition (cf. Semple & Steel (2003, Section 3.1)). Thus, we have a map from $\mathfrak{E}$ to a set of splits,

$$e \mapsto s(e) = A_e|B_e, \quad E := \{s_e \colon e \in \mathfrak{E}\}. \tag{4.2.5}$$

We say that $(\mathfrak{V}, \mathfrak{E}, \ell)$ *induces* the splits E. From Semple & Steel (2003, Theorem 3.1.4), we can derive the following result.

Lemma 4.2.6. *Let E be a set of splits of L. Then there exists a graph-based tree representative $(\mathfrak{V}, \mathfrak{E}, \ell)$ that induces E if and only if E is compatible and for any $u, v \in L$ there exists a split $A|B \in E$ such that $u \in A$ and $v \in B$. In this case, $(\mathfrak{V}, \mathfrak{E}, \ell)$ is unique up to topological equivalence.*

Proof. Semple & Steel (2003, Theorem 3.1.4) yields this result for X-trees as defined in Semple & Steel (2003, Definition 2.1.1) with $X = L$, where multiple labeled vertices are allowed and E are compatible split sets, the condition that for any $u, v \in L$ there exists a split $A|B \in E$ with $u \in A$ and $v \in B$ is dropped. But this is exactly the condition that characterizes that labels are "separated" at least by any split and thus there cannot be multiply labeled vertices, and vice versa. As pointed out in Remark 4.1.10, 2., topological equivalence between graph-based tree representatives corresponds to isomorphisms between phylogenetic X-trees (here $X = L$), and thus we get the uniqueness result from Semple & Steel (2003, Theorem 3.1.4). In particular, for two topological equivalent graph-based tree representatives $(\mathfrak{V}, \mathfrak{E}, \ell)$ and $(\mathfrak{V}', \mathfrak{E}', \ell')$ with $f \colon \mathfrak{V} \to \mathfrak{V}'$ satisfying (i) and (ii) from Definition 4.1.9, it holds that

$$s(\{u, v\}) = s'(\{f(u), f(v)\}) \quad \text{for all } \{u, v\} \in \mathfrak{E}, \tag{4.2.6}$$

where s' is the map defined in Equation (4.2.5) with respect to $(\mathfrak{V}', \mathfrak{E}', \ell')$. ∎

For a graph-based tree representative $(\mathfrak{V}, \mathfrak{E}, \ell)$, define the *induced edge weights* $(\lambda_s)_{s\in E} \in (0,1)^E$ with respect to the induced splits E by (where $s = s(e)$ for $e \in \mathfrak{E}$)

$$\lambda_s := 1 - \exp(-\ell_e). \tag{4.2.7}$$

Note that this is a strictly monotonically increasing correspondence and thus $\ell_e \to 0 \iff \lambda_s \to 0$ as well as $\ell_e \to \infty \iff \lambda_s \to 1$. From Equation (4.2.6) and from (iii), Definition 4.1.9, it follows immediately, that the definition of λ is unique up to isomorphy of graph-based tree representatives. To be precise, for two topological equivalent graph-based tree representatives $(\mathfrak{V}, \mathfrak{E}, \ell)$ and $(\mathfrak{V}', \mathfrak{E}', \ell')$ with $f\colon \mathfrak{V} \to \mathfrak{V}'$ satisfying (i), (ii) and (iii) from Definition 4.1.9, since they are topologically equivalent, they induce the same set of splits E, and let $\lambda \in (0,1)^E$ and $\lambda' \in (0,1)^E$ be the edge weights induced by $(\mathfrak{V}, \mathfrak{E}, \ell)$ and $(\mathfrak{V}', \mathfrak{E}', \ell')$, respectively. Then, for any split $s \in E$ with $s = s(e) = s'(e')$ with $e \in \mathfrak{E}$ and $e' \in \mathfrak{E}'$, we know from the uniqueness of induced splits and Equation (4.2.6) that if $e = \{u, v\}$, then $e' = \{f(u), f(v)\}$. Therefore,

$$\lambda_s = 1 - \exp(-\ell_e) \overset{\text{isom.}}{=} 1 - \exp(-\ell'_{e'}) = \lambda'_{s'(e')} = \lambda'_s.$$

We summarize these observations in the following lemma.

Lemma 4.2.7. *Let E be a set of splits of L and let $\lambda \in (0,1)^E$ be some edge weights. Then there exists a graph-based tree representative $(\mathfrak{V}, \mathfrak{E}, \ell)$ that induces E and λ if and only if E is compatible, for all $u, v \in L$ there exists a split $A|B \in E$ such that $u \in A$ and $v \in B$ and λ satisfies Equation* (4.2.7)*. In this case, (E, λ) is a split-based tree and $(\mathfrak{V}, \mathfrak{E}, \ell)$ is unique up to isomorphy.*

We can extend the previous results to graph-based and split-based forests.

Theorem 4.2.8. *There is a one-to-one correspondence between split-based forests as in Definition 4.2.4 and graph-based forests as in Definition 4.1.9.*

Proof. Each graph-based forest $[\mathfrak{V}, \mathfrak{E}, \ell]$ corresponds one-to-one to a collection of graph-based trees, say $[\mathfrak{V}_1, \mathfrak{E}_1, \ell^{(1)}], \dots, [\mathfrak{V}_K, \mathfrak{E}_K, \ell^{(K)}]$ for some $K \in \mathbb{N}$ with non-empty label sets $L_1, \dots, L_K$, respectively, with $L_1, \dots, L_K$ being a partition of L.

To see this, note that the topology of any representative $(\mathfrak{V}, \mathfrak{E}, \ell)$ yields connected components, say, $(\mathfrak{V}_1, \mathfrak{E}_1), \dots, (\mathfrak{V}_K, \mathfrak{E}_K)$ with $L_\alpha = \mathfrak{V}_\alpha \cap L$ and $\ell_e^{(\alpha)} = \ell_e$ for all $e \in \mathfrak{E}_\alpha$, $\alpha = 1, \dots, K$. For any other representative $(\mathfrak{V}', \mathfrak{E}', \ell')$ that is isomorphic to $(\mathfrak{V}, \mathfrak{E}, \ell)$, its topology yields the same number of connected components, say, $(\mathfrak{V}'_\alpha, \mathfrak{E}'_\alpha, \ell^{(\alpha)\prime})$, $\alpha = 1, \dots, K$, such that w.l.o.g. $\mathfrak{V}'_\alpha \cap L = L_\alpha$ and hence by restricting an isomorphism $f\colon \mathfrak{V} \to \mathfrak{V}'$ to $f\big|_{\mathfrak{V}_\alpha}$ (i.e. f as in Definition 4.1.9), it holds that $(\mathfrak{V}'_\alpha, \mathfrak{E}'_\alpha, \ell^{(\alpha)\prime})$ is isomorphic to $\mathfrak{V}_\alpha, \mathfrak{E}_\alpha, \ell^{(\alpha)}$, $\alpha = 1, \dots, K$.

Then, from Lemma 4.2.7 we have a one-to-one correspondence between collections of graph-based trees $[\mathfrak{V}_1, \mathfrak{E}_1, \ell^{(1)}], \dots, [\mathfrak{V}_K, \mathfrak{E}_K, \ell^{(K)}]$ with label sets $L_1, \dots, L_K$ that are a partition of L, and collections of split-based trees $(E_1, \lambda^{(1)}), \dots, (E_K, \lambda^{(K)})$ for label sets $L_1, \dots, L_K$, respectively (to be precise, each $(E_\alpha, \lambda^{(\alpha)})$ has a partition $\mathcal{L}_\alpha = \{L_\alpha\}$ of the label set L_α, for $\alpha = 1, \dots, K$).

Finally, there is an obvious one-to-one correspondence between such collections of split-based trees $(E_1, \lambda^{(1)}), \dots, (E_K, \lambda^{(K)})$ for label sets $L_1, \dots, L_K$, respectively, and split-based forests (E, λ), given by $E = \cup_\alpha E_\alpha$ and $\lambda\big|_{E_\alpha} = \lambda^{(\alpha)}$. ■

Due to this one-to-one correspondence, we also use the symbol $e \in E$ for splits and will refer to them as *edges*. Furthermore, we say that two labels $u, v \in L$ are *connected* in a split-based forest (E, λ), if $u, v \in L_\alpha$ for some $\alpha = 1, \dots, K$, where $\mathcal{L} = \{L_1, \dots, L_K\}$ is its corresponding label partition. Let $(\mathfrak{V}, \mathfrak{E}, \ell)$ be a graph-based forest representative. For $u, v \in \mathfrak{V}$ that are connected, denote by $\mathfrak{E}(u, v)$ the set of edges in the unique path connecting u and v (recall the notation introduced in Definition 4.1.1). The following result states that an edge is in the path between u and v whenever its corresponding split "separates" u from v.

Lemma 4.2.9. *Let $(\mathfrak{V}, \mathfrak{E}, \ell)$ be a graph-based forest representative and (E, λ) be its corresponding split-based forest, and let s be defined as in Equation* (4.2.5). *Let $u, v \in L$ be connected and let $e \in \mathfrak{E}$ with corresponding split $s = s(e) \in E$. Then*

$$e \in \mathfrak{E}(u, v) \iff s = A|B \text{ and } u \in A, v \in B.$$

Proof. Cutting an edge e from the underlying graph of $(\mathfrak{V}, \mathfrak{E}, \ell)$ divides the connected component that e is contained in into two parts, and accordingly the labels, which yield the corresponding split $s(e)$. Thus e is on the unique path between u and v if and only if its corresponding split is $A|B$ with $u \in A$ and $v \in B$. ■

The previous lemma motivates the following definition.

Definition 4.2.10. Let (E, λ) be a split-based forest with partition $\mathcal{L} = \{L_1, \dots, L_K\}$. Then, for labels $u, v \in L_\alpha$, $\alpha = 1, \dots, K$, define

$$E(u, v) = \{e \in E\colon e = A|B, u \in A, v \in B\}.$$

Remark 4.2.11. In light of the previous definition, it follows from Lemma 4.2.9 and Remark 4.1.7 that $E(u, v) \neq \emptyset$ for all $u, v \in L_\alpha$, $1 \leq \alpha \leq K$, which is equivalent condition 5. in Definition 4.2.4.

Partial Ordering on Split-Based Forest Topologies

We introduce a partial ordering on split-based forest topologies from Moulton & Steel (2004, Section 3), although they define it for a broader class of forests. Whenever two split-based forest topologies E and E' satisfy the relationship $E' \leq E$, one can intuitively think of it as saying that "E' can be derived from E by contracting and cutting edges", where split-based forest topologies corresponding to fully resolved graph-based trees (using the one-to-one correspondence from Theorem 4.2.8) are the greatest upper bounds with respect to this partial ordering, and the split-based forest topology $E = \emptyset$ is the unique smallest element.

Definition 4.2.12. For two topologies E,E', of two split-based forests, respectively, and with label partitions $\mathcal{L} = \{L_1, \ldots, L_K\}$, $\mathcal{L}' = \{L'_1, \ldots, L'_{K'}\}$, respectively, we say that

$$E' \leq E \tag{4.2.8}$$

if the following three properties hold:

Refinement: $\mathcal{L}'$ is a refinement of $\mathcal{L}$, that is for every $1 \leq \alpha' \leq K'$ there is $1 \leq \alpha \leq K$ with $L'_{\alpha'} \subset L_\alpha$;

Restriction: for every such α' and α above with $L'_{\alpha'} \subset L_\alpha$,

$$E'_{\alpha'} \subset \left\{\tilde{e} \colon \tilde{e} := e\big|_{L'_{\alpha'}} \text{ is a valid split}, e \in E_\alpha\right\}$$

Cut: for every $1 \leq \alpha'_1 \neq \alpha'_2 \leq K'$ and $1 \leq \alpha \leq K$, if $L'_{\alpha'_1}, L'_{\alpha'_2} \subset L_\alpha$, then there is some

$$A|B \in E \text{ with } L'_{\alpha'_1} \subset A,\ L'_{\alpha'_2} \subset B\,.$$

Further, we say $E' < E$ if $E \neq E' \leq E$.

The following result is due to Moulton & Steel (2004, Lemma 3), and we prove it for convenience.

Proposition 4.2.13. *The relation $E' \leq E$ in split-based forest topologies as defined in Definition 4.2.12 is a partial ordering.*

Proof. We show reflexivity ($E \leq E$), antisymmetry ($E' \leq E$ and $E \leq E'$ then $E = E'$) and transitivity ($E'' \leq E'$ and $E' \leq E$ then $E'' \leq E$).

Reflexivity. Clearly, $\mathcal{L}$ is a refinement of itself and for the restriction property, since $e\big|_{L_\alpha} = e$ for any $e \in E_\alpha$, it boils down to an equality $E_\alpha = \{e\big|_{L_\alpha} : e \in E_\alpha\}$. The cut property does not come into play at all, and thus $E \leq E$.

Antisymmetry. Since $\mathcal{L}'$ is a refinement of $\mathcal{L}$ and vice versa, $\mathcal{L} = \mathcal{L}'$. Then analogously to reflexivity, the restriction property becomes trivial and the cut property does not come into play.

Transitivity. Let $E'' \leq E' \leq E$ with label partitions $\mathcal{L}'', \mathcal{L}'$ and $\mathcal{L}$, respectively. Since $\mathcal{L}''$ is a refinement of $\mathcal{L}'$, for any $1 \leq \alpha'' \leq K''$, there exists $1 \leq \alpha' \leq K'$ such that $L''_{\alpha''} \subset L'_{\alpha'}$, and since $\mathcal{L}'$ is a refinement of $\mathcal{L}$, there exists $1 \leq \alpha \leq K$ such that $L'_{\alpha'} \subset L_\alpha$, and thus $L''_{\alpha''} \subset L_\alpha$. Therefore, $\mathcal{L}''$ is a refinement of $\mathcal{L}$, so the refinement property with respect to E'' and E holds.

For the restriction property, let α'', α' and α be as above such that $L''_{\alpha''} \subset L'_{\alpha'} \subset L_\alpha$. Let $e'' \in E''_{\alpha''}$. Then from the restriction property of $E'' \leq E'$ there exists $e' \in E'_{\alpha'}$ with $e'' = e'\big|_{L''_{\alpha''}}$, and by the restriction property of $E' \leq E$ there exists $e \in E_\alpha$ with $e' = e\big|_{L'_{\alpha'}}$, but due to $L''_{\alpha''} \subset L'_{\alpha'}$, we have $e'' = \left(e\big|_{L'_{\alpha'}}\right)\big|_{L''_{\alpha''}} = e\big|_{L''_{\alpha''}}$ and thus the restriction property for E'' and E holds.

For the cut property, let $1 \leq \alpha''_1, \alpha''_2 \leq K''$ and $1 \leq \alpha \leq K$ such that $L''_{\alpha''_1}, L''_{\alpha''_2} \subset L_\alpha$. Then by the refinement property with respect to $E'' \leq E'$ and $E' \leq E$ there exist $1 \leq \alpha'_1, \alpha'_2 \leq K'$ such that $L''_{\alpha''_1} \subset L'_{\alpha'_1} \subset L_\alpha$ and $L''_{\alpha''_2} \subset L'_{\alpha'_2} \subset L_\alpha$. We distinguish two cases:

1. If $\alpha'_1 = \alpha'_2$, then by the cut property of $E'' \leq E'$ there exists $e' = A'|B'$ such that $L''_{\alpha''_1} \subset A'$ and $L''_{\alpha''_2} \subset B'$, and by the restriction property of $E' \leq E$ there exists $e = A|B$ with $e' = e\big|_{L'_{\alpha'_1}}$, say $A \cap L'_{\alpha'_1} = A'$ and $B \cap L'_{\alpha'_1} = B'$, so $L''_{\alpha''_1} \subset A$ and $L''_{\alpha''_2} \subset B$ and the cut property holds.
2. If $\alpha'_1 \neq \alpha'_2$ then by the cut property of $E' \leq E$ there exists an edge $e = A|B \in E_\alpha$ with $L''_{\alpha''_1} \subset L'_{\alpha'_1} \subset A$ and $L''_{\alpha''_2} \subset L'_{\alpha'_2} \subset B$, so the cut property holds true as well.

We conclude $E'' \leq E$, yielding the assertion. ■

The partial ordering simplifies significantly if E and E' are topologies of split-based trees.

Proposition 4.2.14. *Let E, E' be split-based tree topologies. Then*

$$E' \leq E \iff E' \subset E.$$

Proof. The label partitions are just $\mathcal{L} = \mathcal{L}' = \{L\}$, and thus the refinement property is always satisfied and the cut property never comes into play. For the refinement property, observe that in this case

$$E' \subset \left\{\tilde{e} \colon \tilde{e} := e\big|_L \text{ is a valid split}, e \in E\right\} = E,$$

yielding the assertion. ■

Back to the general case, let E, E' be two split-based forest topologies. By definition of the restriction property, some edges $e \in E$ yield edges $e' \in E'$ when being restricted to the respective label set. The following definition categorizes edges $e \in E$ with respect to the partial order $E' \leq E$ into three categories: those that vanish due to them being "cut", those that vanish due to them being "contracted" and finally those that yield valid splits that one can rediscover in E'.

Definition 4.2.15. Let E, E' be split-based forest topologies with $E' \leq E$.

1. Let $e' \in E'_{\alpha'}$, $1 \leq \alpha' \leq K'$. Define the set of all edges in E that yield e' to be

$$R_{e'} := \left\{e \in E \colon e' = e\big|_{L'_{\alpha'}}\right\}.$$

2. Denote the set of all *disappearing* splits in E by

$$R_{\text{dis}} := \left\{e \in E \colon \exists \alpha' \text{ s.t. } e\big|_{L'_{\alpha'}} \text{ is valid, but } e\big|_{L'_{\alpha'}} \notin E'\right\}.$$

3. Denote the set of all *cut* splits in E by

$$R_{\text{cut}} := \left\{e \in E \colon \nexists \alpha' \text{ s.t. } e\big|_{L'_{\alpha'}} \text{ is valid}\right\}.$$

The following lemma gives intuition for the behavior of these sets.

Lemma 4.2.16. *Let $E' \leq E$ with label partitions $\mathcal{L} = \{L_1, \ldots, L_K\}$ and $\mathcal{L}' = \{L'_1, \ldots, L'_{K'}\}$, respectively, and $u, v \in L$. Then the following hold*

(i) If $K = K'$ then, say, $L'_\alpha = L_\alpha$, and $E'_\alpha \subset E_\alpha$ for all $\alpha = 1, \ldots, K$ and $R_{e'} = \{e'\}$ for all $e' \in E'$. Furthermore, in this case,

$$E' < E \iff \exists\, \alpha \text{ with } E'_\alpha \subsetneq E_\alpha \iff R_{dis} \neq \emptyset.$$

(ii) $K < K' \iff R_{cut} \neq \emptyset$.

(iii) $R_{e'} \neq \emptyset$ *for all* $e' \in E'$ *and if* $\exists e' \in E'_{\alpha'}$ *with* $|R_{e'}| > 1$ *then* $L'_{\alpha'} \subsetneq L_\alpha$.

(iv) $E = E' \iff (R_{dis} = \emptyset$ *and* $R_{cut} = \emptyset)$.

(v) $R_{e'} \cap R_{e''} = \emptyset$ *for all* $e', e'' \in E'$ *with* $e' \neq e''$.

(vi) The set of splits from the restriction property in Definition 4.2.12, for any $1 \leq \alpha' \leq K'$, *given by*

$$\left\{\tilde{e}\colon \tilde{e} := e\big|_{L'_{\alpha'}} \text{ is valid}, e \in E\right\},$$

is compatible.

(vii) $e' \in E'(u,v) \iff R_{e'} \cap E(u,v) \neq \emptyset \iff R_{e'} \subset E(u,v)$.

(viii) R_{dis}, R_{cut} *in conjunction with the* $R_{e'}$ *over all* $e' \in E'$ *form a partition of* E, *where* R_{dis} *and* R_{cut} *might be empty.*

(ix) Let $u, v \in L'_{\alpha'}$ *for some* $1 \leq \alpha' \leq K'$. *Then* $R_{dis} \cap E(u,v)$ *in conjunction with the* $R_{e'}$ *over all* $e' \in E'(u,v)$ *form a partition of* $E(u,v)$, *where* $R_{dis} \cap E(u,v)$ *might be empty.*

(x) For any $L'_{\alpha'}, L'_{\alpha''} \subset L_\alpha$ *with* $\alpha' \neq \alpha''$, *there exists an edge* $A|B = e \in E$ *with* $L'_{\alpha'} \subseteq A$, $L'_{\alpha''} \subseteq B$ *and* $e \in R_{cut}$.

Proof. (i) $K = K'$ implies that $L_\alpha = L'_{\sigma(\alpha)}$ for some permutation σ on $\{1, \ldots, K\}$, so without loss of generality we assume in this case $L_\alpha = L'_\alpha$. Then $e\big|_{L'_\alpha} = e\big|_{L_\alpha} = e$ are valid splits for all $e \in E_\alpha$ for all $\alpha = 1, \ldots, K$, so the restriction property of $E' \leq E$ reads $E'_\alpha \subset E_\alpha$, and $R_{e'} = \{e'\}$ for all $e' \in E'$.

The equivalences are immediate from

$$R_{\text{dis}} = \emptyset \iff (\text{for all } \alpha = 1, \ldots, K,\ E'_\alpha = E_\alpha) \iff E' = E.$$

(ii) "$\Rightarrow$": Follows from the stronger statement (x). "$\Leftarrow$": If $K = K'$, then by (i) w.l.o.g. $L'_\alpha = L_\alpha$ and in particular $e\big|_{L'_\alpha} = e\big|_{L_\alpha} = e$ are valid splits for all $e \in E_\alpha$, $\alpha = 1, \ldots, K$, so $R_{cut} = \emptyset$, a contradiction. This yields $K < K'$.

(iii) By the restriction property of $E' \leq E$, each $e' \in E'_{\alpha'}$ is the restriction of some $e \in E_\alpha$, thus $e \in R_{e'} \neq \emptyset$. Assume that there exist $e_1, e_2 \in R_{e'}$ with $e_1 \neq e_2$. If $L'_{\alpha'} = L_\alpha$ was true, then $e_1 = e_1\big|_{L'_{\alpha'}} = e_2\big|_{L'_{\alpha'}} = e_2$, a contradiction.

(iv) "$\Rightarrow$": Trivial. "$\Leftarrow$": $R_{\text{cut}} = \emptyset \implies K = K'$ due to (ii) and thus $R_{\text{dis}} = \emptyset \implies E = E'$ due to (iv).

(v) Assume the contrary: let $A|B = e \in R_{e'} \cap R_{e''}$, where $e' \in L'_{\alpha'} \subset L_\alpha$ and $e'' \in L'_{\alpha''} \subset L_\alpha$.

If $\alpha' = \alpha''$, then $e' = e\big|_{L'_{\alpha'}} = e''$, a contradiction to $e' \neq e''$, so $\alpha' \neq \alpha''$. Since e is in both $R_{e'}$ and $R_{e''}$, both restrictions to $L'_{\alpha'}$ and $L'_{\alpha''}$ exist and therefore

$$A \cap L'_{\alpha'} \neq \emptyset, \quad B \cap L'_{\alpha'} \neq \emptyset, \quad A \cap L'_{\alpha''} \neq \emptyset, \quad B \cap L'_{\alpha''} \neq \emptyset.$$

Due to $E' \leq E$, by the cut property there exists $C|D = \tilde{e} \in E_\alpha$ separating $L'_{\alpha'}$ and $L'_{\alpha''}$, i.e. $L'_{\alpha'} \subseteq C$ and $L'_{\alpha''} \subseteq D$. But then $\tilde{e}, e \in E_\alpha$ cannot be compatible, a contradiction.

(vi) Let $L'_{\alpha'} \subset L_\alpha$. Let $e', e'' \in E\big|_{L'_{\alpha'}}$ such that $e' = e\big|_{L'_{\alpha'}}$ and $e'' = e^\circ\big|_{L'_{\alpha'}}$ with $e = A|B \in E$ and $e^\circ = A^\circ|B^\circ \in E$. Then $e, e^\circ \in E_\alpha$ for otherwise their restriction with $L'_{\alpha'}$ would be empty. Since e and e° are compatible, w.l.o.g. $A \cap A^\circ = \emptyset$. Consequently, $e' = (A \cap L'_{\alpha'})|(B \cap L'_{\alpha'})$ and $e'' = (A^\circ \cap L'_{\alpha'})|(B^\circ \cap L'_{\alpha'})$ are compatible as $(A \cap L'_{\alpha'}) \cap (A^\circ \cap L'_{\alpha'}) = \emptyset$.

(vii) We proceed to show $e' \in E'(u,v) \implies R_{e'} \subset E(u,v) \implies R_{e'} \cap E(u,v) \neq \emptyset \implies e' \in E'(u,v)$.
If $e' \in E'(u,v)$ and due to (iii), $R_{e'} \neq \emptyset$, so $e' := e|_{L'_{\alpha'}} = (A \cap L'_{\alpha'})|(B \cap L'_{\alpha'})$ for some $e = A|B \in R_{e'}$, hence $u \in A$, $v \in B$, or vice versa, i.e. $e \in E(u,v)$. Since the choice $e \in R_{e'}$ was arbitrary, $R_{e'} \subset E(u,v)$.

If $e \in R_{e'} \cap E(u,v)$, $u, v \in L'_{\alpha'}$, such that $e' = e|_{L'_{\alpha'}}$ then $e' \in E'(u,v)$.

(viii) By definition of R_{dis} and R_{cut}, they are disjoint and furthermore have empty intersection with each $R_{e'}$, $e' \in E'$ and the latter are pair-wise disjoint due to (v).

(ix) By definition, $R_{\text{cut}} \cap E(u,v) = \emptyset$ for all $u, v \in L'_{\alpha'}$ (else R_{cut} would contain valid splits). Then (v) in conjunction with (viii) yields the assertion.

(x) Without loss of generality, let $K = 1$. We prove by induction over K'.

Base case: Let $K' = 2$. Then $L'_1 \cup L'_2 = L$, then by the cut property of $E' < E$, there exists an edge $A|B = e \in E$ with $L'_1 \subset A$, $L'_2 \subset B$ so neither $e\big|_{L'_1}$ nor $e\big|_{L'_2}$ yield a valid split, so $e \in R_{\text{cut}}$.

Induction step: Suppose w.l.o.g. the assumption holds true for $K' - 1 \geq 2$ and let $L = \bigcup_{\alpha=1}^{K'} L'_\alpha$. By the cut property of $E' \leq E$, there exists an edge $A|B = e \in E$ with $L'_1 \subseteq A$ and $L'_{K'} \subseteq B$. If $e \in R_{\text{cut}}$, we are done, so assume $e \notin R_{\text{cut}}$. By definition of R_{cut}, w.l.o.g. $e\big|_{L'_2}$ is a valid split, so $A \cap L'_2 \neq \emptyset$ and $B \cap L'_2 \neq \emptyset$.

In the following, we construct two wald topologies $\overline{E}$ and $\overline{E}'$, with respect to the labels $\overline{L} := L \setminus L'_{K'}$, via "deletion" of $L'_{K'}$ from E and E', respectively:

$$\overline{E} := \{e\big|_{\overline{L}} \colon e \in E, e\big|_{\overline{L}} \text{ is a valid split}\},$$
$$\overline{E}' := E' \setminus E'_{K'}.$$

Both are wald topologies, $\overline{E}$ due to (vi) and $\overline{E}'$ as we are deleting one connected component.

Then, $\overline{E}' \leq \overline{E}$ is an immediate consequence of $E' \leq E$.

Thus, by induction hypothesis, there exists a split $\overline{e} = \overline{C}|\overline{D} \in \overline{R}_{cut}$ of $\overline{L}$ (where $\overline{R}_{cut}$ from Definition 4.2.15 with respect to $\overline{E}' \leq \overline{E}$) with $L'_1 \subseteq \overline{C}$ and $L'_2 \subseteq \overline{D}$ such that for each $\alpha = 1, \dots, K' - 1$ the restriction $\overline{e}\big|_{L'_\alpha}$ does not yield a valid split. Furthermore, by construction of $\overline{E}$, there exists $e^\circ = C|D \in E$ with $e^\circ\big|_{\overline{L}} = \overline{e}$, i.e. $\overline{C} = C \setminus L'_{K'}$ and $\overline{D} = D \setminus L'_{K'}$, and such that e° and e are compatible; and note that due to $\overline{e} \in \overline{R}_{cut}$, for each $\alpha = 1, \dots, K' - 1$ that $e^\circ\big|_{L'_\alpha}$ does not yield a valid split. Recalling from above, we have that $L'_1 \subseteq A$, $L'_2 \cap A \neq \emptyset$, $L'_2 \cap B \neq \emptyset$, $L'_1 \subseteq C$ and $L'_2 \subseteq D$, therefore

$$A \cap C \neq \emptyset, \quad B \cap D \neq \emptyset, \quad A \cap C \neq \emptyset,$$

so by compatibility of e_0 and e it must be that $B \cap C = \emptyset$, so from $L'_{K'} \subseteq B$ we find $L'_{K'} \subseteq D$, i.e. $e_0 = C|D = \overline{C}|(\overline{D} \cup L'_{K'})$, and thus $e_0\big|_{L'_{K'}}$ does not yield a valid split, so $e_0 \in R_{\text{cut}}$, as well as $L'_1 \subseteq C$ and $L'_{K'} \subseteq D$, which yields the assertion.

■

4.3 Representation via Distance Matrices

Within a graph-based forest, we will set the distance between two vertices that are not connected to be infinity. Therefore, we define a calculus on $[0, \infty]$, where the element ∞ has been added to the interval $[0, \infty)$. We define the following rules:

(i) $x + \infty = \infty + x = \infty$ for all $x \in [0, \infty]$,

(ii) $x < \infty$ for any $x \in [0, \infty)$.

With this definition, we introduce forest distance matrices.

Definition 4.3.1. Let $D = (d_{uv})_{u,v=1}^N \in [0, \infty]^{N \times N}$ be a matrix such that for all $u, v \in L$

(D1) $d_{uv} = 0 \iff u = v$ and

(D2) $d_{uv} = d_{vu}$.

Then we say that D is a *forest distance matrix* if it satisfies the *four-point condition*, that is if

(D3) for every four (not necessarily distinct) labels $u, v, s, t \in L$, two of the following three terms are equal and greater or equal than the third:

$$d_{uv} + d_{st}, \qquad d_{us} + d_{vt}, \qquad d_{ut} + d_{vs}. \tag{4.3.1}$$

Remark 4.3.2. In the literature, D is also called *tree metric* (e.g. Semple & Steel (2003, Chapter 7)) or distance matrix (e.g. Felsenstein (2003, Chapter 11)).

The four-point condition can be expressed in a more useful manner.

Proposition 4.3.3. *The four-point condition is equivalent to the statement that for every four (not necessarily distinct) labels $u, v, s, t \in L$, it holds that*

$$d_{uv} + d_{st} \leq \max\{d_{us} + d_{vt}, d_{ut} + d_{vs}\} \tag{4.3.2}$$

Proof. "$\Rightarrow$". If one of the terms in Equation (4.3.1) equals ∞, as two of them must be equal and greater or equal than the third, at least two terms equal ∞, so (D3) holds.

Suppose that all terms in Equation (4.3.1) are finite. Then, independent of which two terms are equal and greater or equal than the third, Equation (4.3.2) holds.

"$\Leftarrow$". From permuting the indices we find that

$$\begin{aligned} (*) \quad & d_{uv} + d_{st} \leq \max\{d_{us} + d_{vt}, d_{ut} + d_{vs}\}, \\ (**) \quad & d_{us} + d_{vt} \leq \max\{d_{uv} + d_{st}, d_{ut} + d_{vs}\}, \\ (***) \quad & d_{ut} + d_{vs} \leq \max\{d_{us} + d_{vt}, d_{uv} + d_{st}\}, \end{aligned}$$

and thus, if one of the three terms in Equation (4.3.1) equals ∞, at least one other term equals ∞, so (D3) holds.

Suppose that all terms in Equation (4.3.1) are finite. Without loss of generality, let $d_{uv} + d_{st}$ be greater or equal than the other two terms, i.e. $d_{uv}+d_{st} \geq \max\{d_{us}+d_{vt}, d_{ut}+d_{vs}\}$. From $(*)$, we have equality (whenever another term is the largest, use $(**)$ or $(***)$ accordingly),

$$d_{us} + d_{vt} = \max\{d_{uv} + d_{st}, d_{ut} + d_{vs}\},$$

which implies (D3). ■

Corollary 4.3.4. *Let $D = (d_{uv})_{u,v=1}^N \in [0,\infty]^{N\times N}$ be a forest distance matrix. Then D satisfies the triangle inequalities (cf. Definition 2.1.1), that is*

(D4) $d_{uv} \leq d_{us} + d_{sv}$ for all $u, v, s \in L$.

Furthermore, D is a metric on L in the sense of Definition 2.1.1.

Proof. Plug $u, v, s, t \in L$ with $s = t$ into Equation (4.3.2) to obtain $d_{uv} = d_{uv} + d_{st} \leq \max\{d_{us} + d_{sv}, d_{us} + d_{sv}\} = d_{us} + d_{sv}$. Then, D satisfies all properties of a metric from Definition 2.1.1 (to be precise, the metric is $d(u,v) := d_{uv}$). ■

The following lemma teaches that the matrix entries of a forest distance matrix D that are equal to infinity can be "cleanly separated" from the entries that are finite.

Lemma 4.3.5. *Let $D = (d_{uv})_{u,v=1}^N \in [0,\infty]^{N\times N}$ be a forest distance matrix. The relation on L defined by $u \sim v \iff d_{uv} < \infty$ is an equivalence relation.*

Proof. Trivially, $u \sim u$ for all $u \in L$ and $u \sim v \iff v \sim u$ for all $u, v \in L$. For transitivity, let $u, v, s \in L$ with $u \sim s$ and $s \sim v$, i.e. $d_{us} < \infty$ and $d_{sv} < \infty$. Then, the triangle inequality (D4) yields $d_{uv} \leq d_{us} + d_{sv} < \infty$ and thus $u \sim v$. ■

Definition 4.3.6. Let $(\mathfrak{V}, \mathfrak{E}, \ell)$ be a graph-based forest representative as defined in Definition 4.1.6. Define

$$\mu(\mathfrak{V}, \mathfrak{E}, \ell) = (d_{uv})_{u,v=1}^N$$

where

$$d_{uv} = \begin{cases} \sum\limits_{e\in\mathfrak{E}(u,v)} \ell_e, & \text{if } u \neq v \text{ connected,} \\ 0, & \text{if } u = v, \\ \infty, & \text{else.} \end{cases} \tag{4.3.3}$$

for $1 \leq u, v \leq N$.

Theorem 4.3.7.

1. *Let $(\mathfrak{V}, \mathfrak{E}, \ell)$ be a graph-based forest representative. Then $\mu(\mathfrak{V}, \mathfrak{E}, \ell)$ is a forest distance matrix.*
2. *For any forest distance matrix D, there exists a graph-based forest representative $(\mathfrak{V}, \mathfrak{E}, \ell)$ with $D = \mu(\mathfrak{V}, \mathfrak{E}, \ell)$, and it is unique up to isomorphy.*

Proof. 1. By definition of μ, $D = (d_{uv})_{u,v=1}^N := \mu(\mathfrak{V}, \mathfrak{E}, \ell)$ is symmetric and $d_{uu} = 0$. For any $u, v \in L$ with $u \neq v$, either $d_{uv} = \infty \neq 0$ or $d_{uv} = \sum_{e \in \mathfrak{E}(u,v)} \ell_e > 0$, since by Remark 4.1.7, $\mathfrak{E}(u, v) \neq \emptyset$.

Hence D satisfies (D1) and (D2), and for each $1 \leq \alpha \leq K$, this makes $(d_{uv})_{u,v \in L_\alpha}$ a *tree metric* in the sense of Semple & Steel (2003, Definition 7.1.2) such that we can apply Semple & Steel (2003, Lemma 7.1.7), stating that the four-point condition holds for all choices (not necessarily distinct) $u, v, s, t \in L_\alpha$.

For all other choices $u, v, s, t \in L$, at least two of the four labels are not connected, say without loss of generality $d_{uv} = \infty$. Both s and t can thus only be connected to either u or v, respectively. Thus in all four cases, at least two of the three terms of Equation (4.3.1) equal ∞, which implies that the four-point condition holds for all choices of $u, v, s, t \in L$, yielding the assertion.

2. Let $D = (d_{uv})_{u,v=1}^N \in [0, \infty]^{N \times N}$ be a forest distance matrix. From Lemma 4.3.5, we have a partition of L into equivalence classes $L_1, \ldots, L_K$ for some $1 \leq K \leq N$ with $d_{uv} < \infty$ if and only if $u, v \in L_\alpha$ for some $1 \leq \alpha \leq K$. By assumption, each sub-matrix $(d_{uv})_{u,v \in L_\alpha}$ satisfies the four-point condition and thus by Semple & Steel (2003, Theorem 7.2.6) each sub-matrix $(d_{uv})_{u,v \in L_\alpha}$ is a *tree metric* on L_α in the sense of Semple & Steel (2003, Definition 7.1.2). This means that there exists an X-tree with $X = L_\alpha$ in the sense of Semple & Steel (2003, Definition 2.1.1), and positive edge lengths, that induces $(d_{uv})_{u,v \in L_\alpha}$, and $d_{uv} > 0$ for all $u, v \in L_\alpha$ with $u \neq v$ is equivalent to that it is a phylogenetic L_α-tree, which corresponds by Remark 4.1.7 (and with the edge lengths) to a graph-based tree representative $(\mathfrak{V}_\alpha, \mathfrak{E}_\alpha, \ell^{(\alpha)})$ with labels L_α, such that $\mu(\mathfrak{V}_\alpha, \mathfrak{E}_\alpha, \ell^{(\alpha)}) = (d_{uv})_{u,v \in L_\alpha}$. By Semple & Steel (2003, Theorem 7.1.8), $(\mathfrak{V}, \mathfrak{E}, \ell)$ is unique up to isomorphism.

This collection of graph-based tree representatives corresponds uniquely a graph-based forest representative $(\mathfrak{V}, \mathfrak{E}, \ell)$ by $\mathfrak{V} = \cup_\alpha \mathfrak{V}_\alpha$, $\mathfrak{E} = \cup_\alpha \mathfrak{E}_\alpha$, $\ell_e = \ell_e^{(\alpha)}$ for $e \in \mathfrak{E}_\alpha$, $1 \leq \alpha \leq K$, and note that $\mu(\mathfrak{V}, \mathfrak{E}, \ell) = D$, and that $(\mathfrak{V}, \mathfrak{E}, \ell)$ is unique up to isomorphism as every component is unique up to isomorphism. ■

The previous result motivates the following definition.

Definition 4.3.8. Let $\mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell]$ be a graph-based forest. Define

$$\mu(\mathfrak{F}) := \mu(\mathfrak{V}, \mathfrak{E}, \ell).$$

As the graph-based forest representative with forest distance matrix is unique up to isomorphy, we have at once the following result.

Corollary 4.3.9. *For any forest distance matrix D there exists a unique graph-based forest $\mathfrak{F}$ with $\mu(\mathfrak{F}) = D$.*

Let $\mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell]$ be a graph-based forest with corresponding forest distance matrix $D = (d_{uv})_{u,v=1}^{N}$ and corresponding split-based forest (E, λ) with label partition $\mathcal{L} = \{L_1, \ldots, L_K\}$ (cf. Theorem 4.2.8). Then, by definition of $\mu(\mathfrak{V}, \mathfrak{E}, \ell) = D$, cf. Equation (4.3.3), and using Lemma 4.2.9 as well as Equation (4.2.7), it follows for all $u, v \in L_\alpha$ with $u \neq v$, $1 \leq \alpha \leq K$,

$$d_{uv} = \sum_{e \in \mathfrak{E}(u,v)} \ell_e = \sum_{e \in E(u,v)} \ell_e = -\sum_{e \in E(u,v)} \log(1-\lambda_e) = -\log\left(\prod_{e \in E(u,v)} (1-\lambda_e)\right). \quad (4.3.4)$$

This motivates the following definition.

Definition 4.3.10. Let (E, λ) be a split-based forest. Define

$$\nu(E, \lambda) = (d_{uv})_{u,v=1}^{N}$$

where

$$d_{uv} = \begin{cases} -\log\left(\prod_{e \in E(u,v)} (1-\lambda_e)\right), & \text{if } u \neq v \text{ connected,} \\ 0, & \text{if } u = v, \\ \infty, & \text{else.} \end{cases} \quad (4.3.5)$$

for $1 \leq u, v \leq N$.

The calculations above yield the following result.

Theorem 4.3.11. *For any forest distance matrix D there exists a unique split-based forest (E, λ) with corresponding graph-based forest $\mathfrak{F}$ (in the sense of Theorem 4.2.8), such that*

$$\nu(E, \lambda) = \mu(\mathfrak{F}) = D.$$

We conclude this section with a characterization of the forest topology by the forest distance matrix.

Proposition 4.3.12. *Two graph-based forests $\mathfrak{F}, \mathfrak{F}'$ are topologically equivalent if and only if their corresponding forest distance matrices satisfy the same equalities and strict inequalities in (D3).*

Proof. This follows directly from Semple & Steel (2003, Example 7.1.6). ■

4.4 Representation via Correlation Matrices

Definition 4.4.1. A matrix $P = (\rho_{uv})_{u,v=1}^{N} \in \mathbb{R}^{N\times N}$ is a *forest correlation matrix*, if

$$\rho_{uv} = \exp(-d_{uv})$$

for all $u, v \in L$, where $D = (d_{uv})_{u,v=1}^{N}$ is a forest distance matrix and $\exp(-\infty) := 0$.

Translating the conditions for D being a forest distance matrix to forest correlation matrices, we obtain at once the following result.

Proposition 4.4.2. *A matrix $P = (\rho_{uv})_{u,v=1}^{N} \in \mathbb{R}^{N\times N}$ is a forest correlation matrix if and only if $P \in [0,1]^{N\times N}$ such that for all $u, v \in L$*

(C1) $\rho_{uv} = 1 \iff u = v$ *and*

(C2) $\rho_{uv} = \rho_{vu}$*,*

(C3) for every four (not necessarily distinct) labels $u, v, s, t \in L$, two of the following three terms are equal and less or equal than the third:

$$\rho_{uv}\rho_{st}, \qquad \rho_{us}\rho_{vt}, \qquad \rho_{ut}\rho_{vs}. \tag{4.4.1}$$

Translating Corollary 4.3.4 and Equation (4.3.2) to forest correlation matrices, we obtain at once the following corollary.

Corollary 4.4.3. *Let $P = (\rho_{uv})_{u,v=1}^{N}$ be a forest correlation matrix. Then*

(C4) $\rho_{uv} \geq \rho_{us}\rho_{sv}$ *for all* $u, v, s \in L$.

Furthermore, (C3) is equivalent to the statement that for every four (not necessarily distinct) labels $u, v, s, t \in L$, it holds that

$$\rho_{uv}\rho_{st} \geq \min\{\rho_{us}\rho_{vt}, \rho_{ut}\rho_{vs}\}. \tag{4.4.2}$$

Just as the maps μ and ν map graph-based forests and split-based forests to their corresponding forest distance matrices, respectively, we define the maps ψ and ϕ that map graph-based forests and split-based forests to their corresponding forest correlation matrices, respectively.

Definition 4.4.4. Let $\mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell]$ be a graph-based forest with corresponding split-based forest (E, λ), with distance matrix $\mu(\mathfrak{F}) = \nu(E, \lambda) = (d_{uv})_{u,v=1}^N$. Define ψ and ϕ by

$$\psi(\mathfrak{F}) = \psi(\mathfrak{V}, \mathfrak{E}, \ell) = \big(\exp(-d_{uv})\big)_{u,v=1}^N,$$
$$\phi(E, \lambda) = \big(\exp(-d_{uv})\big)_{u,v=1}^N.$$

The next result is an immediate consequence of Equation (4.3.3).

Corollary 4.4.5. *Let $\mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell]$ be a graph-based forest with corresponding forest correlation matrix $\psi(\mathfrak{F}) = (\rho_{uv})_{u,v=1}^N$. Then*

$$\rho_{uv} = \begin{cases} \prod_{e \in \mathfrak{E}(u,v)} \exp(-\ell_e), & \text{if } u \neq v \text{ connected}, \\ 1, & \text{if } u = v, \\ 0, & \text{else}. \end{cases} \tag{4.4.3}$$

for $1 \leq u, v \leq N$.

The next result is an immediate consequence of Equation (4.3.5).

Corollary 4.4.6. *Let (E, λ) be a split-based forest with corresponding forest correlation matrix $\phi(E, \lambda) = (\rho_{uv})_{u,v=1}^N$. Then*

$$\rho_{uv} = \begin{cases} \prod_{e \in E(u,v)} (1 - \lambda_e), & \text{if } u \neq v \text{ connected}, \\ 1, & \text{if } u = v, \\ 0, & \text{else}. \end{cases} \tag{4.4.4}$$

for $1 \leq u, v \leq N$.

The next result is the fundamental motivation for the definition of Wald Space in Chapter 6 and can be found in Garba et al. (2021a, Theorem 4.1).

Theorem 4.4.7. *Let (E, λ) be a split-based forest. Then its forest correlation matrix $\phi(E, \lambda)$ is strictly positive definite and thus $\phi(E, \lambda) \in \mathcal{P}$.*

Proof. This is proven in Garba et al. (2021a, Appendix E), but we shall state a more detailed version here. By definition, if we can show that for any graph-based forest representative

$(\mathfrak{V}, \mathfrak{E}, \ell)$, the matrix $\psi(\mathfrak{V}, \mathfrak{E}, \ell)$ is strictly positive definite, we are done. We prove the assertion by induction on N.

Base case. For $N = 2$ we have two possible graph-based forest representative topologies. First, $\mathfrak{V} = \{1, 2\}$, $\mathfrak{E} = \emptyset$ (so ℓ is an "empty vector") and thus $\psi(\mathfrak{V}, \mathfrak{E}, \ell) = I \in \mathcal{P}$, that is the 2×2 unit matrix. Secondly, $\mathfrak{V} = \{1, 2\}$, $\mathfrak{E} = \{1|2\}$ and $\ell_{1|2} \in (0, \infty)$, therefore

$$\psi(\mathfrak{V}, \mathfrak{E}, \ell) = \begin{pmatrix} 1 & \exp(-\ell_{1|2}) \\ \exp(-\ell_{1|2}) & 1 \end{pmatrix},$$

which is strictly positive definite since $\exp(-\ell_{1|2}) \in (0, 1)$.

Induction hypothesis. Now let $N \geq 3$ and assume that for any graph-based forest representative with respect to a label set of size $N - 1$ or less, its corresponding forest correlation matrix is strictly positive definite.

Induction step. Let $(\mathfrak{V}, \mathfrak{E}, \ell)$ be a graph-based forest representative with forest correlation matrix $\psi(\mathfrak{V}, \mathfrak{E}, \ell) = P = (\rho_{uv})_{u,v=1}^N$.

If it has more than one connected component, then the label set L is divided into a partition $L_1, \ldots, L_K$, such that $\rho_{uv} > 0$ for all $u, v \in L_\alpha$, $\alpha = 1, \ldots, K$, and $\rho_{uv} = 0$ otherwise. For any $\alpha = 1, \ldots, K$, the matrix $P_\alpha = (\rho_{uv})_{u,v \in L_\alpha}$ is strictly positive definite by induction hypothesis as it is the forest correlation matrix of a graph-based tree representative with label set L_α and $|L_\alpha| < |L| = N$. Thus we find for any vector $x \in \mathbb{R}^N$ with $x \neq 0$, where $x^{(\alpha)} = (x_u)_{u \in L_\alpha}$,

$$x^T P x = \sum_{\alpha=1}^{K} (x^{(\alpha)})^T P_\alpha \, x^{(\alpha)} > 0,$$

since $x^{(\alpha)} \neq 0$ for at least one $\alpha = 1, \ldots, K$.

Let $(\mathfrak{V}, \mathfrak{E}, \ell)$ be a graph-based tree representative with forest correlation matrix $\psi(\mathfrak{V}, \mathfrak{E}, \ell) = P = (\rho_{uv})_{u,v=1}^N$.

Sylvester's criterion tells us that a matrix is strictly positive definite if and only if all principal minors have strictly positive determinant. Since the matrix $(\rho_{uv})_{u,v=1}^{N-1}$ is also a forest correlation matrix, it must be corresponding to some graph-based forest, and thus by induction hypothesis, it is strictly positive definite. Therefore, it is enough to show that $\det(P) > 0$.

Within the tree $(\mathfrak{V}, \mathfrak{E})$, there must exist labels $u, v \in L$ with $u \neq v$, such that u is a leaf that is

1. either adjacent to v,

2. or adjacent to an unlabeled vertex $w \in \mathfrak{V} \setminus L$ such that v is also a leaf adjacent to w.

If there was not such labels u and v, then for any label $u \in L$ there must exists an unlabeled vertex to which no other label is adjacent, so $|\mathfrak{V}| \geq 2N$, a contradiction to Proposition 4.1.8. Without loss of generality, we can assume that $u = N$ and $v = N - 1$, as otherwise we permute the labels accordingly with an invertible permutation matrix R, and the forest correlation matrix of the thus obtained graph-based tree representative is $R^T PR$, which is strictly positive definite if and only if P is.

In both cases 1. and 2., denote the edge incident by N by e_N, and define $x := \exp(-\ell_{e_N})$, and since $e_N \in \mathfrak{E}(u, N)$ for all $u = 1, \ldots, N-1$, with Equation (4.4.3) we can write $\rho_{uN} = x\, c_u$ for some positive constant $c_u > 0$, $u = 1, \ldots, N-1$.

With $\mathfrak{S}$ being the set of permutations on L, the Leibniz formula for determinants gives (where $\operatorname{sgn}(\sigma) = (-1)^m$, where m is the number of transpositions of any decomposition into transpositions of σ)

$$\begin{aligned}
\det(P) &= \sum_{\sigma \in \mathfrak{S}} \operatorname{sgn}(\sigma) \prod_{u=1}^{N} \rho_{u\sigma(u)} \\
&= \Bigg(\sum_{\substack{\sigma \in \mathfrak{S} \\ \sigma(N)=N}} \operatorname{sgn}(\sigma) \prod_{u=1}^{N} \rho_{u\sigma(u)} \Bigg) + \Bigg(\sum_{\substack{\sigma \in \mathfrak{S} \\ \sigma(N)\neq N}} \operatorname{sgn}(\sigma) \prod_{u=1}^{N} \rho_{u\sigma(u)} \Bigg) \\
&= \det\Big((\rho_{uv})_{u,v=1}^{N-1} \Big) + x^2 \Bigg(\sum_{\substack{\sigma \in \mathfrak{S} \\ \sigma(N)\neq N}} \operatorname{sgn}(\sigma) c_{\sigma(N)} c_{\sigma^{-1}(N)} \prod_{\substack{u=1 \\ u \neq \sigma^{-1}(N)}}^{N-1} \rho_{u\sigma(u)} \Bigg),
\end{aligned}$$

so the $\det(P)$ as a function in x^2 is a straight line (denote this function by $f(x^2)$).

Case 1. In this case, as discussed above, the leaf N is adjacent to $N - 1$ via the edge e_N. Then $\rho_{N(N-1)} = x$ and $\rho_{uN} = x\rho_{u(N-1)}$, and $c_u = \rho_{u(N-1)}$ for $u = 1, \ldots, N-2$, and $c_{N-1} = 1$. Now, if we plug in $x = 0$, we obtain

$$f(0) = \det(P) = \det\Big((\rho_{uv})_{u,v=1}^{N-1} \Big) > 0,$$

by induction hypothesis. If we plug in $x = 1$, we find that $\rho_{uN} = \rho_{u(N-1)}$ for $u \in L$ and thus $\det(P) = 0$, giving $f(1) = 0$. Since $f(x^2)$ is a straight line and $f(0) > f(1) = 0$, we conclude $\det(P) = f(x^2) > 0$ for all $x \in (0, 1)$, which is the case by definition of $x = \exp(-\ell_{e_N})$.

Case 2. Let e_{N-1} be the edge that is incident with the leaf $N - 1$ and adjacent to the same unlabeled vertex $w \in \mathfrak{V} \setminus L$ that N is adjacent to via e_N. Define $y := \exp(-\ell_{e_{N-1}})$, and with the same argument using the Leibniz formula for determinants we know that $\det(P)$ is a straight line with parameter y^2 (as $N-1$ is also a leaf as N is). We make the dependency visible via defining $\det(P) := g(x^2, y^2)$, where g is an affine function in (x^2, y^2).

As N and $N-1$ are both adjacent to the same unlabeled vertex, we have that $\mathfrak{E}(u, N) \setminus \{e_N\} = \mathfrak{E}(u, N-1) \setminus \{e_{N-1}\}$ for all $u = 1, \ldots, N-2$, and $\mathfrak{E}(N-1, N) = \{e_{N-1}, e_N\}$, and for $u = 1, \ldots, N-2$, by Equation (4.4.3),

$$\rho_{uN} = x\, c_u, \quad \rho_{u(N-1)} = y\, c_u, \quad \rho_{N(N-1)} = xy,$$

where $c_u = \prod_{e \in \mathfrak{E}(w,u)} \exp(-\ell_e)$. Thus if $x = y = 1$ then $\rho_{u(N-1)} = \rho_{uN}$ for all $u \in L$, so $\det(P) = g(1,1) = 0$. If $x = 0$ and $y \neq 0$, then $\det(P) = \det((\rho_{uv})_{u,v=1}^{N-1}) > 0$ which is positive by induction hypothesis, if $y = 0$ and $x \neq 0$ then $\det(P) = \det((\rho_{uv})_{u,v \in L \setminus \{N-1\}}) > 0$ which is positive by induction hypothesis as well. If $x = 0$ and $y = 0$, then $\det(P) = \det((\rho_{u,v})_{u,v=1}^{N-2}) > 0$ which is again positive by induction hypothesis. Thus, since g is an affine function in (x^2, y^2) we have that $g(x^2, y^2) > 0$ for all $(x^2, y^2) \in (0,1)^2$, which is the case by definition of $x = \exp(-\ell_{e_N})$ and $y = \exp(-\ell_{e_{N-1}})$.

We conclude that $P = \psi(\mathfrak{V}, \mathfrak{E}, \ell)$ is strictly positive definite. ■

We summarize the previous results in the following corollary.

Corollary 4.4.8.

1. *Each split-based forest $F = (E, \lambda)$ corresponds uniquely to a forest correlation matrix given by $\phi(F) \in [0,1]^{N \times N}$, characterized by the properties (C1), (C2) and (C3).*
2. *Furthermore, ϕ maps injectively from the split-based forests into $\mathcal{P}$.*
3. *Finally, two split-based forests F, F' have the same topologies if and only if their corresponding forest correlation matrices satisfy the same equalities and strict inequalities in (C3).*

Chapter 5

BHV Tree Space

In 2001, Louis Billera, Susan Holmes and Karen Vogtmann introduced the BHV Space (the acronym stands for the authors names; cf. Billera et al. (2001)). Let $N \in \mathbb{N}$ with $N \geq 3$. The BHV space is a metric space where the elements are rooted phylogenetic trees with labels $\{1, \ldots, N\}$ on the leaves. We will label the root 0 and add it to the set of labels, so $L_0 = \{0, 1, \ldots, N\}$. Originally, the distance on BHV space was defined taking into account the interior edges only, and one would then add the pendant edges of the tree via a Cartesian product. We will take also the pendant edges into account and interpret the root 0 as another label. We will rigorously define the BHV Space using the notation from Chapter 4.

5.1 BHV Space without Pendant Edges

Recall properties and notation for *splits* from Section 4.2. Let $L_0 = \{0, 1, \ldots, N\}$. Denote the set of all possible *interior edges* with respect to the label set L_0 by

$$\mathcal{E} = \big\{A|B \colon L_0 = A \sqcup B, |A|, |B| \geq 2\big\}.$$

Note that the number of possible splits in $\mathcal{E}$ is the cardinality of the power set of L_0 divided by 2, minus the number of subsets $A \subset L_0$ with $|A| \leq 1$, that is $N + 2$, we conclude

$$|\mathcal{E}| = 2^N - N - 2.$$

Definition 5.1.1. 1. The set of all *tree trunks* is defined by

$$\mathcal{B}^\circ = \big\{(E, \ell) \colon E \subset \mathcal{E} \text{ compatible}, \ell \in (0, \infty)^E\big\}.$$

2. Define the map $\chi\colon \mathcal{B}^\circ \to [0,\infty)^{\mathcal{E}}$ by

$$T = (E,\ell) \mapsto \chi(T) := x,$$

where for each $e \in \mathcal{E}$,

$$x_e := \begin{cases} \ell_e, & \text{if } e \in E, \\ 0, & \text{else.} \end{cases}$$

3. Define the metric $d_{\mathcal{B}^\circ}\colon \mathcal{B}^\circ \times \mathcal{B}^\circ \to [0,\infty)$ on $\mathcal{B}^\circ$ by (where $T, T' \in \mathcal{B}^\circ$)

$$d_{\mathcal{B}^\circ}(T,T') := \inf_{\substack{\gamma\colon [0,1]\to\mathcal{B}^\circ \\ \gamma(0)=T,\gamma(1)=T' \\ \chi\circ\gamma \text{ continuous}}} L(\chi\circ\gamma),$$

where $L(\chi\circ\gamma)$ is the length of the path $\chi\circ\gamma$ with respect to the Euclidean distance on $[0,\infty)^{\mathcal{E}}$, i.e.

$$L(\chi\circ\gamma) = \sup_{\substack{0=t_0<t_1<\ldots<t_n=1 \\ n\in\mathbb{N}}} \sum_{i=0}^{n-1} \big\|(\chi\circ\gamma)(t_i) - (\chi\circ\gamma)(t_{i+1})\big\|.$$

Proposition 5.1.2. *The tuple $(\mathcal{B}^\circ, d_{\mathcal{B}^\circ})$ is a metric space. For any $T, T' \in \mathcal{B}^\circ$ we have that $d_{\mathcal{B}^\circ}(T,T') < \infty$.*

Proof. The metric space property follows directly from Lemma 2.1.3. As $0 \in \chi(\mathcal{B}^\circ)$ and the straight line segments from $\chi(T)$, $\chi(T')$ to 0, respectively, are contained in $\chi(\mathcal{B}^\circ)$, we find $d_{\mathcal{B}^\circ}(T,T') \le \|\chi(T)\|_2 + \|\chi(T')\|_2 < \infty$. ■

Remark 5.1.3. The metric space $(\mathcal{B}^\circ, d_{\mathcal{B}^\circ})$ introduced in Definition 5.1.1 is the original BHV Space, as is immediate from Billera et al. (2001, Section 3.2). This is essentially a metric space on split-based trees, but taking only the interior edges (splits) into account, ignoring all pendant edges (i.e. splits where one part contains only one label). However, as mentioned in Billera et al. (2001, p.743), one can include the pendant edges via a product of an $(N+1)$-dimensional Euclidean space, and that allows us to compare BHV Space with other tree or forest spaces.

The following result is from Billera et al. (2001, Lemma 4.1).

Theorem 5.1.4. *The metric space $(\mathcal{B}^\circ, d_{\mathcal{B}^\circ})$ is CAT(0).*

This property has strong implications: geodesics exist and are unique, cf. Theorem 2.1.13. We investigate the structure of the BHV Space which is a composition of *orthants*.

Definition 5.1.5. Let $E \subset \mathcal{E}$ be a compatible set of interior splits. The *orthant of* E is the set

$$\mathcal{O}_E = \Big\{ x \in [0,\infty)^{\mathcal{E}} \colon x_e \in (0,\infty) \text{ for } e \in E,\ x_e = 0 \text{ else} \Big\}.$$

All of the following statements are trivial by construction.

Lemma 5.1.6. *(i)* $\mathcal{O}_\emptyset = \{0\} \subset \mathbb{R}^{\mathcal{E}}$.

(ii) Let $E, E', E'' \subset \mathcal{E}$ *be compatible, respectively. Then*

$$\begin{aligned} E' \neq E &\iff \mathcal{O}_{E'} \cap \mathcal{O}_E = \emptyset, \\ E' \subseteq E &\iff \mathcal{O}_{E'} \subseteq \overline{\mathcal{O}}_E, \\ E \cap E' = E'' &\iff \overline{\mathcal{O}}_E \cap \overline{\mathcal{O}}_{E'} = \overline{\mathcal{O}}_{E''} \end{aligned}$$

and furthermore,

$$\overline{\mathcal{O}_E} = \bigsqcup_{E' \subseteq E} \mathcal{O}_{E'}.$$

(iii) For any compatible set of interior splits $E \subset \mathcal{E}$*, it holds that*

$$\mathcal{O}_E = \chi\Big(\big\{ (E,\ell) \colon \ell \in (0,\infty)^E \big\} \Big).$$

(iv) $\mathcal{B}^\circ$ *is a decomposition into its orthants, i.e.*

$$\chi(\mathcal{B}^\circ) = \bigsqcup_{\substack{E \subset \mathcal{E} \\ E \text{ compatible}}} \mathcal{O}_E.$$

The notation of orthants simplifies the expression of $d_{\mathcal{B}^\circ}$. For two tree trunks $T = (E,\ell), T' = (E',\ell') \in \mathcal{B}^\circ$ with their image being in the same orthant, i.e. $E = E'$, their distance is simply

$$d_{\mathcal{B}^\circ}(T,T') = \|\chi(T) - \chi(T')\|_2 = \|\ell - \ell'\|_2.$$

Furthermore, every continuous path $\gamma \colon [0,1] \to \mathcal{B}^\circ$ consists of segments such that their image under χ traverse an orthant, and we can always shorten γ by replacing the segment through an orthant by the straight line segment with the same end points. This motivates the following definition.

Definition 5.1.7. Let $T = T_0, T_1, \dots, T_n = T'$ be a finite sequence of points in $\mathcal{B}^\circ$ with $\chi(T_k) = x^{(k)} \in \mathcal{O}_{E_k}$ for some $E_k \subset \mathcal{E}$ compatible for each $k = 0, \dots, n$. The sequence

$T_0, T_1, \ldots, T_n$ is *admissible*, if for all $k = 1, \ldots, n$, the sets $E_{k-1} \cup E_k \subset \mathcal{E}$ are compatible, respectively.

Note that if $E_{k-1} \cup E_k$ is compatible then $x^{(k-1)}, x^{(k)} \in \overline{\mathcal{O}}_{E_{k-1} \cup E_k}$, and therefore the straight line segment through $\overline{\mathcal{O}}_{E_{k-1} \cup E_k}$ will be the shortest path between $x^{(k-1)}$ and $x^{(k)}$. Thus we can rewrite

$$d_{\mathcal{B}^\circ}(T, T') = \inf_{\substack{T=T_0,T_1,\ldots,T_n=T' \\ \text{admissible} \\ n\in\mathbb{N}}} \sum_{k=1}^{n} \big\|\chi(T_{k-1}) - \chi(T_k)\big\|.$$

By Theorem 5.1.4, there exists a unique shortest path between any two points in $\mathcal{B}^\circ$. In Owen & Provan (2011), M. Owen and J. S. Provan present an algorithm for computing geodesics in BHV Space in polynomial time, where one starts with the trivial path over the cone point and augments it step by step.

We now consider two examples, where we examine the BHV Space for $N = 3$ and $N = 4$.

Example 5.1.8 (BHV Space for $N = 3$)**.** We have $L_0 = \{0, 1, 2, 3\}$ and thus

$$\mathcal{E} = \big\{01|23, 02|13, 03|12\big\}.$$

Thus there are four possible choices for compatible subsets of $\mathcal{E}$, namely

$$E_0 = \emptyset, \quad E_1 = \{01|23\}, \quad E_2 = \{02|13\}, \quad E_3 = \{03|12\}.$$

The respective orthants are, say,

$$\begin{aligned}
\mathcal{O}_{E_0} &= \{(0,0,0) \in \mathbb{R}^3\}, \\
\mathcal{O}_{E_1} &= \{(a,0,0) \in \mathbb{R}^3 \colon a > 0\}, \\
\mathcal{O}_{E_2} &= \{(0,b,0) \in \mathbb{R}^3 \colon b > 0\}, \\
\mathcal{O}_{E_3} &= \{(0,0,c) \in \mathbb{R}^3 \colon c > 0\}.
\end{aligned}$$

They are depicted in Figure 5.1.

Example 5.1.9 (BHV Space for $N = 4$)**.** We have $L_0 = \{0, 1, 2, 3, 4\}$. Then $|\mathcal{E}| = 2^N - N - 2 = 10$ and according to Billera et al. (2001, p.743) there are $(2N-3)!! := 5 \cdot 3 \cdot 1 = 15$ different possible choices of compatible subsets $E \subset \mathcal{E}$ with $|E| = N - 2 = 2$. For simplicity and presentability, we consider only the splits

$$e_1 = 12|034, \;\; e_2 = 012|34, \;\; e_3 = 123|04, \;\; e_4 = 124|03,$$

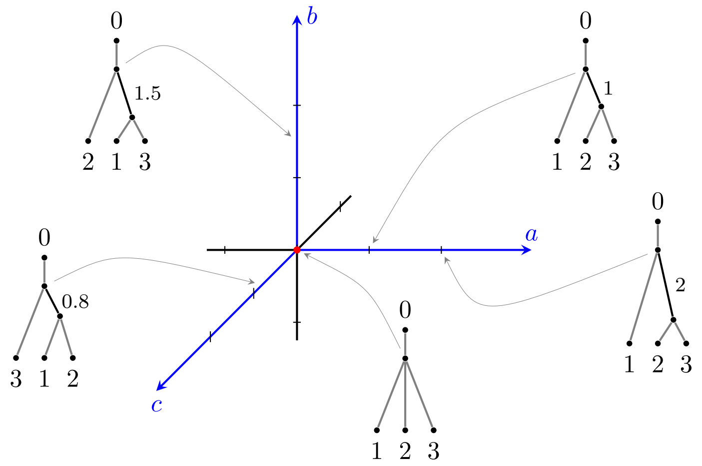

Figure 5.1: *The BHV Space embedded in $\mathbb{R}^3$ via χ for $N = 3$, cf. Example 5.1.8. The blue axes are the orthants $\mathcal{O}_{E_k}$ for $k = 1, 2, 3$. The red point corresponds to the orthant $\mathcal{O}_{E_0} = \{0\}$. The pendant edges of the trees are in gray to indicate that they are not taken into account.*

and the following compatible split sets,

$$E_1 = \{e_1, e_2\}, \quad E_2 = \{e_1, e_3\}, \quad E_3 = \{e_1, e_4\}.$$

With, say, $\mathcal{E} = \{e_1, e_2, e_3, e_4, e_5, \dots, e_{10}\}$, we have the following orthants,

$$\begin{aligned} \mathcal{O}_{E_1} &= \{(a, b, 0, 0, 0, \dots, 0) \in \mathbb{R}^{10} \colon a, b > 0\}, \\ \mathcal{O}_{E_2} &= \{(a, 0, c, 0, 0, \dots, 0) \in \mathbb{R}^{10} \colon a, c > 0\}, \\ \mathcal{O}_{E_3} &= \{(a, 0, 0, d, 0, \dots, 0) \in \mathbb{R}^{10} \colon a, d > 0\}. \end{aligned}$$

They are depicted in Figure 5.2.

5.2 BHV Space with Pendant Edges

As already discussed in Remark 5.1.3, pendant edges can be added to $\mathcal{B}^\circ$ by a Cartesian product. Define the *set of all pendant edges* by

$$E_{\text{pen}} \coloneqq \big\{\{u\}|(L \setminus \{u\}) \colon u \in L_0\big\}.$$

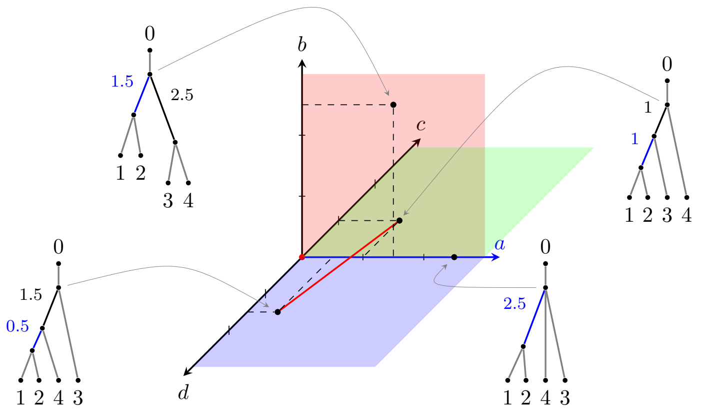

Figure 5.2: *Part of the BHV Space for $N = 4$ embedded in $\mathbb{R}^{10}$ via χ, where only a three dimensional slice is shown, cf. Example 5.1.9. The planes are the three orthants $\mathcal{O}_{E_1}$ (red), $\mathcal{O}_{E_2}$ (green) and $\mathcal{O}_{E_3}$ (blue) The blue axis is the orthant $\mathcal{O}_{\{e_1\}}$ and the red dot marks the star tree $0 \in \mathbb{R}^{10}$. Furthermore, four trees are depicted with the topology of the respective orthant they belong to. The red line marks the shortest path between the respective endpoints, and note that only due to the visualization the path appears to be a straight line.*

For any split-based tree (E, λ) (cf. Definition 4.2.4), we reparametrize the edge weights $\lambda \in (0,1)^E$ back to edge lengths $\ell \in (0,\infty)^E$ in a strictly monotonically increasing fashion (in particular, one-to-one), for all $e \in E$, by

$$\ell_e = -\log(1 - \lambda_e), \tag{5.2.1}$$

and write (E, ℓ) instead of (E, λ). Then, for any reparametrized split-based tree (E, ℓ) with $E_{\text{pen}} \subset E$, the edges are divided into $E_{\text{int}} = E \setminus E_{\text{pen}} \subset \mathcal{E}$ and E_{pen}, and the edge lengths $\ell \in (0,\infty)^E$ are separated into $\ell_{\text{pen}} \in (0,\infty)^{E_{\text{pen}}}$ and $\ell_{\text{int}} \in (0,\infty)^{E_{\text{int}}}$ accordingly.

Definition 5.2.1. Denote the set of all reparametrized split-based trees (E, ℓ) with $E_{\text{pen}} \subset E$ by $\mathcal{B}$. Furthermore, define the metric $d_{\mathcal{B}}$ on $\mathcal{B}$ by (where $T = (E, \ell) \in \mathcal{B}$, $T' = (E', \ell') \in \mathcal{B}$)

$$d^2_{\mathcal{B}}(T, T') = d^2_{\mathcal{B}^\circ}\big((E_{\text{int}}, \ell_{\text{int}}), (E'_{\text{int}}, \ell'_{\text{int}})\big) + \|\ell_{\text{pen}} - \ell'_{\text{pen}}\|_2^2.$$

The BHV Space $(\mathcal{B}, d_{\mathcal{B}})$ is the set $\mathcal{B}$ of all split-based trees

Note that $\mathcal{B}$ is the Cartesian product $\mathcal{B}^\circ \times (0,\infty)^{E_{\text{pen}}}$ and its metric is a product of metrics of the form as in Bridson & Haefliger (1999, Part I, Definition 5.1) and thus from Bridson & Haefliger (1999, Part II, Example 1.15, (3)) it follows that $(\mathcal{B}, d_{\mathcal{B}})$ is also CAT(0). Finally, geodesics in $(\mathcal{B}, d_{\mathcal{B}})$ are just geodesics in $(\mathcal{B}^\circ, d_{\mathcal{B}^\circ})$ coupled with straight line segments in $(0,\infty)^{E_{\text{pen}}}$.

Chapter 6

Wald Space of Phylogenetic Forests

Let $N \in \mathbb{N}$ be the number of labels with $N \geq 2$, and define $L = \{1, \dots, N\}$. Recall from Chapter 4 the definitions of graph-based forests $\mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell]$ (cf. Definition 4.1.9) and split-based forests $F = (E, \lambda)$ (cf. Definition 4.2.4), as well as their one-to-one correspondence we established in Theorem 4.2.8. Furthermore, recall the maps ψ and ϕ (cf. Equation (4.4.3) and Equation (4.4.4), respectively), that map graph-based forests and split-based forests one-to-one to strictly positive definite forest correlation matrices, respectively (positive definiteness from Theorem 4.4.7). Recall further the manifold of strictly positive definite matrices $\mathcal{P}$ from Section 2.2.1 and Chapter 3 with the topology inherited from the Euclidean space, $\mathcal{P} \subset \mathbb{R}^{N \times N}$. In the following section, we define the Wald Space.

6.1 Definition and Topology

Definition 6.1.1. The *Wald Space* is a topological space consisting of the set

$$\mathcal{W}$$

of all split-based forests $F = (E, \lambda)$ with labels L, equipped with the unique topology such that the map $\phi \colon \mathcal{W} \to \mathcal{P}$ is a homeomorphism onto its image. According to Theorem 4.4.7, we use split-based forests F and graph-based forests $\mathfrak{F}$ interchangeably and call an element of $\mathcal{W}$ a *wald*, in plural *wälder*. The topology E of a wald $(E, \lambda) \in \mathcal{W}$ is called *wald topology*. Further, wälder that have the same topology E form a *grove*

$$\mathcal{G}_E = \big\{ F' = (E', \lambda') \in \mathcal{W} \colon E = E' \big\}.$$

Remark 6.1.2. With $\mathcal{E}$, the set of all wald topologies $E = E_1 \cup \ldots \cup E_K$, i.e. each E_α is a set of compatible splits of L_α, $\alpha = 1, \ldots, K$, satisfying Definition 4.2.4, 1.-5., where $L_1, \ldots, L_K$ runs over all partitions of L and $1 \leq K \leq N$, the Wald Space can be bijectively identified with the disjoint union

$$\mathcal{W} \cong \bigsqcup_{E \in \mathcal{E}} (0, 1)^E.$$

Note that this is not a topological statement.

In order to explore the topology of $\mathcal{W}$, by definition we need to understand the topology of the subset $\phi(\mathcal{W}) \subset \mathcal{P}$. From Theorem 4.3.7 and by definition Definition 4.4.1 of forest correlation matrices, the image $\phi(\mathcal{W})$ is exactly the set of all forest correlation matrices. Thus by Proposition 4.4.2, $\phi(\mathcal{W})$ is characterized by algebraic equalities and inequalities (namely (C1)-(C3) in Proposition 4.4.2), and we have the following corollary.

Proposition 6.1.3. *$\phi(\mathcal{W}) \subset \mathcal{P}$ is a closed subset of $\mathcal{P}$.*

Proof. For any sequence $P^{(k)} \in \phi(\mathcal{W})$ with limit $P = (\rho_{uv})_{u,v=1}^N \in \mathcal{P}$, trivially, (C2) and (C3), as well as (C4) are satisfied. For (C1), trivially $\rho_{uu} = 1$ for $u \in L$, and for $u, v \in L$ with $u \neq v$, assume the contrary of (C1), namely suppose that $\rho_{uv} = 1$. Then by the triangle inequalities (C4),

$$\Big(\rho_{us} \geq \rho_{uv}\rho_{vs} = \rho_{vs} \quad \wedge \quad \rho_{vs} \geq \rho_{uv}\rho_{us} = \rho_{us}\Big) \implies \rho_{us} = \rho_{vs},$$

for all $s \in L$, so the two rows coincide and $\det(P) = 0$, a contradiction to $P \in \mathcal{P}$. Thus $\rho_{uv} < 1$ for all $u, v \in L$ with $u \neq v$, and (C1) holds true as well. As P satisfies (C1)-(C3) from Proposition 4.4.2, we deduce that $P \in \phi(\mathcal{W})$. ■

Proposition 6.1.4. *Let E be the wald topology of a wald $F = (E, \lambda)$ that is fully resolved tree. Then the grove $\mathcal{G}_E$ is open in $\mathcal{W}$.*

Proof. All matrices $P \in \phi(\mathcal{G}_E)$ are characterized by the fact that the matrix entries satisfy the same equalities and the same strict inequalities (cf. Corollary 4.4.8). Fully resolved trees are the highest dimensional forests (maximal number of edges) where each label is a leaf, and thus $\phi(\mathcal{G}_E)$ is open in $\phi(\mathcal{W})$, i.e. $\mathcal{G}_E$ is open in $\mathcal{W}$. ■

Define the completely disconnected wald F_∞ to be $F_\infty = (E, \lambda)$ with $E = \emptyset$ and $\lambda = ()$, with partition $\mathcal{L} = \{\{u\} \colon u \in L\}$ accordingly. This is the wald with $K = N$ connected components, where each component is a single vertex u with $u \in L$. In particular, the unit $N \times N$ matrix $I = (\delta_{uv})_{u,v \in L} \in \mathcal{P}$ is the ϕ-image of $F_\infty \in \mathcal{W}$, i.e. $\phi(F_\infty) = I$. We can directly show another simple result that is due to the characterization of forest correlation matrices.

Proposition 6.1.5. *$\mathcal{W}$ is star shaped in the Euclidean sense of $\mathcal{P} \subset \mathbb{R}^{N \times N}$ with respect to F_∞ and hence contractible.*

Proof. Let $F \in \mathcal{W}$ with $\phi(F) = P = (\rho_{uv})_{u,v=1}^N \in \mathcal{P}$ satisfying (C1)-(C3) from Proposition 4.4.2. Recall that $\phi(F_\infty) = I$ and consider

$$P^{(x)} = (\rho_{uv}^{(x)})_{u,v=1}^N = x\, I + (1-x)P,$$

and observe for all $x \in [0,1]$ that $\rho_{uv}^{(x)} \in [0,1]$, further $\rho_{uv}^{(x)} = 1 \iff u = v$ for all $u, v \in L$, and that $P^{(x)}$ is symmetric. Thus for all $x \in [0,1]$, $P^{(x)}$ satisfies (C1) and (C2) from Proposition 4.4.2. Moreover, to see that $P^{(x)}$ satisfies (C3), we show that $P^{(x)}$ satisfies Equation (4.4.2) for all choices of $u, v, s, t \in L$ (and note that P does satisfy Equation (4.4.2), and we will use this fact everywhere). For pair-wise distinct $u, v, s, t \in L$, observe that

$$\rho_{uv}^{(x)} \rho_{st}^{(x)} = (1-x)^2 \rho_{uv}\rho_{st} \geq (1-x)^2 \min\{\rho_{us}\rho_{vt}, \rho_{ut}\rho_{vs}\} = \min\{\rho_{us}^{(x)}\rho_{vt}^{(x)}, \rho_{ut}^{(x)}\rho_{vs}^{(x)}\}.$$

In the case that two of the pair-wise distinct $u, v, s, t \in L$ are equal, it is enough to verify the cases, say, $s = t$ and $u = s$. For $s = t$, so $\rho_{st}^{(x)} = 1 = \rho_{st}$, it holds that

$$\begin{aligned}\rho_{uv}^{(x)} \rho_{st}^{(x)} = (1-x)\rho_{uv}\rho_{st} &\geq (1-x) \min\{\rho_{us}\rho_{vt}, \rho_{ut}\rho_{vs}\} \\ &\geq (1-x)^2 \min\{\rho_{us}\rho_{vt}, \rho_{ut}\rho_{vs}\} = \min\{\rho_{us}^{(x)}\rho_{vt}^{(x)}, \rho_{ut}^{(x)}\rho_{vs}^{(x)}\};\end{aligned}$$

and for $u = s$, so $\rho_{us}^{(x)} = 1 = \rho_{us}$ we trivially find

$$\rho_{uv}^{(x)} \rho_{st}^{(x)} = \rho_{vs}^{(x)} \rho_{st}^{(x)} \geq \min\{\rho_{vt}^{(x)}, \rho_{vs}^{(x)}\rho_{st}^{(x)}\} = \min\{\rho_{us}^{(x)}\rho_{vt}^{(x)}, \rho_{ut}^{(x)}\rho_{vs}^{(x)}\}.$$

If two or more pairs of the pair-wise distinct $u, v, s, t \in L$ are equal, then Equation (4.4.2) is trivially satisfied.

We conclude that $P^{(x)}$ satisfies (C1)-(C3) from Proposition 4.4.2, so $P^{(x)}$ is a forest correlation matrix for all $x \in [0,1]$, and by Theorem 4.3.7 the entire continuous path $x \mapsto P^{(x)}$, $[0,1] \to \mathcal{P}$ corresponds to a path $F^{(x)} := \phi^{-1}(P^{(x)}) \in \mathcal{W}$, connecting $F = F^{(0)}$ with $F_\infty = F^{(1)}$ as asserted. ■

Remark 6.1.6. 1. For showing contractibility of the edge-product space the authors of Moulton & Steel (2004, Proposition 5.1) contract to the same point.

2. All of the wälder $F^{(x)}$, for $0 \leq x < 1$ constructed in the proof of Proposition 6.1.5 have the same partition of labels into connected tree components, respectively, due to $\rho_{uv} \neq 0 \iff (1-x)\rho_{uv} \neq 0$ for all $x \in [0,1)$ for all $u, v \in L$.

3. For $0 < x < 1$, $P^{(x)}$ satisfies unchanged, strict or non-strict four-point conditions (C3), that may be different, though, from those of $P^{(0)} = \phi(F)$. Thus, $F^{(x)}$ are in the same grove for all $x \in (0,1)$.

4. All triangle inequalities (C4) from Corollary 4.4.3 involving initial nonzero ρ_{uv} are strict, however, for $0 < x < 1$, so that for $\phi^{-1}(P^{(x)})$ none of the leaves have degree 2. For example, starting with the wald consisting of a chain of three vertices with $N = 3$ (so each vertex is labeled and the middle is of degree two), it is immediately transformed into a fully resolved tree (and stays one for all $x \in (0,1)$).

Definition 6.1.7. Let E be a wald topology.

1. We identify a grove $\mathcal{G}_E$ of a wald $F = (E, \lambda)$ with topology E with the open cube

$$\mathcal{G}_E = \{F = (E,\lambda) \in \mathcal{W} \colon \lambda \in (0,1)^E\} \cong (0,1)^E . \tag{6.1.1}$$

2. We denote the restriction of $\phi \colon \mathcal{W} \to \mathcal{P}$ from Equation (4.4.3) to a grove $\mathcal{G}_E$ by $\phi_E \colon \mathcal{G}_E \to \mathcal{P}$ such that

$$\phi_E \colon (0,1)^E \to \mathcal{P}, \qquad \lambda \mapsto \phi_E(\lambda) \coloneqq \phi(E,\lambda) = (\rho_{uv})_{u,v=1}^N, \tag{6.1.2}$$

such that

$$\rho_{uv} = \begin{cases} \prod\limits_{e \in E(u,v)} (1-\lambda_e), & \text{if } u \neq v \text{ connected,} \\ 1, & \text{if } u = v, \\ 0, & \text{else.} \end{cases} \tag{6.1.3}$$

3. The continuation of ϕ_E from Equation (6.1.2) onto all of $\mathbb{R}^E$ is denoted by

$$\bar{\phi}_E \colon \mathbb{R}^E \to \mathcal{S}, \qquad \lambda \mapsto (\rho_{uv})_{u,v=1}^N, \tag{6.1.4}$$

where

$$\rho_{uv} = \begin{cases} \prod_{e \in E(u,v)} (1-\lambda_e), & \text{if } u \neq v \text{ connected,} \\ 1, & \text{if } u = v, \\ 0, & \text{else.} \end{cases} \tag{6.1.5}$$

Note that although the formulas are the same, for sake of completeness we have repeated them here in detail. Furthermore, note that the continuation $\bar{\phi}_E$ is multivariate real analytic. The properties of ϕ_E are summarized in the following theorem, which thereby characterize each grove.

Theorem 6.1.8. *Let E be a wald topology.*

1. *The inverse of $\phi_E \colon (0,1)^E \to \mathcal{P}$ for a forest correlation matrix $P = (\rho_{uv})_{u,v=1}^N \in \phi_E(\mathcal{G}_E)$ is given by*

$$\phi_E^{-1}(P) = \lambda,$$

with

$$\lambda_e = 1 - \max_{\substack{u,v \in A \\ s,t \in B}} \sqrt{\frac{\rho_{ut}\rho_{vs}}{\rho_{uv}\rho_{st}}}\,, \text{ for all } e = A|B \in E\,.$$

2. *The derivative of ϕ_E has full rank $|E|$ throughout $(0,1)^E$,*
3. *The map $\phi_E \colon (0,1)^E \to \mathcal{P}$ is a smooth embedding.*

Proof. For the first assertion consider $e = A|B$, where $A \cup B = L_\alpha$, for some $1 \leq \alpha \leq K$ and where $L_1, \ldots, L_K$ is the leaf partition induced by E. Let $D = (d_{uv})_{u,v=1}^N$ be the forest distance matrix corresponding to P (cf. Definition 4.3.1 and Definition 4.4.1). Then, by Corollary 4.3.4, the matrix entries $d_{uv} = -\log \rho_{uv}$ $(u, v \in L_\alpha)$ define a metric on L_α. For such a metric, Buneman (1971, Lemma 8) asserts that one can assign a graph-based tree representative $(\mathfrak{V}_\alpha, \mathfrak{E}_\alpha, \ell^\alpha)$ where

$$\ell_e^\alpha = \min_{\substack{u,v \in A \\ s,t \in B}} \frac{1}{2}\big(d_{ut} + d_{vs} - d_{uv} - d_{st}\big), \tag{6.1.6}$$

which is uniquely determined by Buneman (1971, Theorem 2). Due to our uniqueness results from Theorem 4.2.8 and Theorem 4.3.7, due to Equation (4.2.7), $\lambda_e = 1 - \exp(-\ell_e^\alpha)$ and thus from Equation (6.1.6) and by $\rho_{uv} = \exp(-d_{uv})$,

$$\lambda_e = 1 - \exp\left(- \min_{\substack{u,v \in A \\ s,t \in B}} \frac{1}{2}\big(d_{ut} + d_{vs} - d_{uv} - d_{st}\big)\right) = 1 - \max_{\substack{u,v \in A \\ s,t \in B}} \sqrt{\frac{\rho_{ut}\rho_{vs}}{\rho_{uv}\rho_{st}}}.$$

For the second assertion, let $e \in E$ and suppose that $F = (E, \lambda)$ with label partition $\mathcal{L} = \{L_1, \ldots, L_K\}$, $1 \leq K \leq N$. Using Equation (6.1.3), for any $u, v \in L$, if either $u = v$ or $u \in L_\alpha$, $v \in L_{\alpha'}$ with $\alpha \neq \alpha'$, then

$$\left(\frac{\partial \phi_E}{\partial \lambda_e}(\lambda)\right)_{uv} = 0\,.$$

Else, if $u, v \in L_\alpha$ for some $1 \leq \alpha \leq K$, then $\rho_{uv} > 0$ and with the Kronecker delta δ,

$$\left(\frac{\partial \phi_E}{\partial \lambda_e}(\lambda)\right)_{uv} = -\delta_{e \in E(u,v)} \prod_{\substack{\tilde{e} \in E(u,v)\\ \tilde{e} \neq e}} \left(1 - \lambda_{\tilde{e}}\right) = -\frac{\rho_{uv}}{1 - \lambda_e}\delta_{e \in E(u,v)}\,, \tag{6.1.7}$$

Thus, for every $x \in \mathbb{R}^E$, we have for the differential of ϕ_E,

$$\Big((\mathrm{d}\phi_E)_\lambda(x)\Big)_{uv} = -\rho_{uv} \sum_{e \in E} \frac{x_e}{1 - \lambda_e}\delta_{e \in E(u,v)}\,,$$

so that $\big((\mathrm{d}\phi_E)_\lambda(x)\big)_{uv} = 0$ implies for all $u, v \in L_\alpha$, for each $\alpha = 1, \ldots, K$,

$$0 = \sum_{e \in E(u,v)} \frac{x_e}{1 - \lambda_e} =: d'_{uv}. \tag{6.1.8}$$

We now view each of the $\ell'_e := \frac{x_e}{1-\lambda_e}$, $e \in E$ as a real valued "length" of e. With the corresponding graph-based forest topology $[\mathfrak{V}, \mathfrak{E}]$ corresponding to E, for every $e \in E$ there are $v_1, v_2 \in \mathfrak{V}_\alpha$ with suitable $1 \leq \alpha \leq K$ such that e corresponds to $\{v_1, v_2\} \in \mathfrak{E}_\alpha$. In particular, since $(\mathfrak{V}_\alpha, \mathfrak{E}_\alpha)$ is a tree, there are $u, v, s, t \in L_\alpha$ (not necessarily all of them distinct), such that

$$\ell'_e = \frac{1}{2}\big(d'_{uv} + d'_{st} - d'_{ut} - d'_{vs}\big).$$

As the r.h.s. is zero due to Equation (6.1.8), it follows that $x_e = 0$, yielding that $(\mathrm{d}\phi_E)_\lambda$ has full rank, as asserted.

The third assertion follows directly from 1. and 2., i.e. ϕ_E is bijectively smooth onto its image and its differential is injective. ■

Having characterized each grove, we are interested in what happens with $\bar{\phi}_E$ if $\lambda \in (0,1)^E$ converges to the boundary of the cube. The next result characterizes exactly under which conditions $\bar{\phi}_E(\lambda)$ stays in the image of Wald Space under ϕ, i.e. $\phi(\mathcal{W})$.

Lemma 6.1.9. *Let E be a wald topology and $\lambda^* \in \partial([0,1]^E)$ with $\bar{\phi}_E(\lambda^*) = (\rho^*_{uv})_{u,v=1}^N \in \mathcal{S}$. Then*

$$\bar{\phi}_E(\lambda^*) \in \phi(\mathcal{W}) \iff \bar{\phi}_E(\lambda^*) \in \mathcal{P} \iff \rho^*_{uv} < 1 \text{ for all } u, v \in L \text{ with } u \neq v.$$

Proof. For the first equivalence, by Proposition 6.1.3 $\phi(\mathcal{W})$ is closed in $\mathcal{P}$ and by continuity of $\bar{\phi}_E$, $\bar{\phi}_E(\lambda^*)$ is a limit of matrices $\bar{\phi}_E(\lambda^{(n)}) = \phi_E(\lambda^{(n)}) \in \phi(\mathcal{W})$, so $\bar{\phi}_E(\lambda^*) \in \phi(\mathcal{W}) \iff \bar{\phi}_E(\lambda^*) \in \mathcal{P}$.

For the second equivalence, if $\bar{\phi}_E(\lambda^*) \in \phi(\mathcal{W})$ then by (C1) of Proposition 4.4.2, for all $u, v \in L$, $\rho^*_{uv} = 1 \iff u = v$.

For the other direction, suppose that $\rho^*_{uv} < 1$ for all $u, v \in L$ with $u \neq v$. Again, by continuity of $\bar{\phi}_E$, $\bar{\phi}_E(\lambda^*)$ is a limit of matrices $\bar{\phi}_E(\lambda^{(n)}) = \phi_E(\lambda^{(n)})$ and thus $\bar{\phi}_E(\lambda^*)$ satisfies (C2), (C3) of Proposition 4.4.2, as well as $\rho^*_{uu} = 1$ for $u \in L$. Now, since $\rho^*_{uv} < 1$ for all $u, v \in L$ with $u \neq v$, we have $\rho^*_{uv} = 1 \iff u = v$, so (C1) is also satisfied and thus $\bar{\phi}_E(\lambda^*) \in \phi(\mathcal{W})$. ■

So, as long as all off-diagonal entries are strictly less than one, or equivalently, as long as the matrix is strictly positive definite, we stay in the image of the Wald Space under ϕ. Accordingly, we define the boundary of groves which is by Lemma 6.1.9 well-defined.

Definition 6.1.10. Let E be a wald topology. The *boundary of the grove $\mathcal{G}_E$* is defined by

$$\partial\mathcal{G}_E \coloneqq \Big\{F = \phi^{-1}\big(\bar{\phi}_E(\lambda^*)\big) : \lambda^* \in \partial([0,1]^E),\ \bar{\phi}_E(\lambda^*) \in \mathcal{P}\Big\} \subset \mathcal{W}. \tag{6.1.9}$$

Accordingly, define $\overline{\mathcal{G}_E} \coloneqq \mathcal{G}_E \cup \partial\mathcal{G}_E$.

The following result gives a first glimpse on how different groves are connected through the convergence of wälder.

Theorem 6.1.11. *Consider a sequence of wälder $(F_n)_{n\in\mathbb{N}} \subset \mathcal{W}$, such that $F_n \to F' \in \mathcal{W}$, where $F_n = (E_n, \lambda^{(n)})$, $n \in \mathbb{N}$, $F' = (E', \lambda')$. Then there is a subsequence n_k, $k \in \mathbb{N}$, and a common topology E such that $E_{n_k} = E$ for all $k \in \mathbb{N}$. Furthermore*

1. *$\lambda^{(n_k)}$ has a cluster point $\lambda^* \in [0,1]^E$,*
2. *and $\phi(F') = \bar{\phi}_E(\lambda^*)$ for every of such cluster point $\lambda^* \in [0,1]^E$,*
3. *and $F' \in \partial\mathcal{G}_E$ whenever $E \neq E'$.*

Proof. For the first assertion, as there are only finitely many wald topologies, there needs to exist a subsequence $(F_{n_k})_{k\in\mathbb{N}} \subset (F_n)_{n\in\mathbb{N}}$ with $E_{n_k} = E$ for some topology E for all $k \in \mathbb{N}$, and thus, since $F_{n_k} \in \mathcal{G}_E \cong (0,1)^E$, there exists $\lambda^{(n_k)} \in (0,1)^E$ with $\phi_E(\lambda^{(n_k)}) = \phi(F_{n_k})$ for all $k \in \mathbb{N}$.

For 1., by Bolzano-Weierstraß, there needs to exist a cluster point $\lambda^* \in [0,1]^E$ of $(\lambda^{(n_k)})_{k\in\mathbb{N}}$.

For 2., for any cluster point $\lambda^* \in [0,1]^E$, from the continuity of $\bar{\phi}_E$, $\bar{\phi}_E(\lambda^*)$ is a cluster point of $(\phi(F_n))_{n\in\mathbb{N}}$ and by $F_n \to F'$ we find $\phi(F_n) \to \phi(F')$ and thus $\bar{\phi}_E(\lambda^*) = \phi(F')$.

For 3., let $\lambda^* \in [0,1]^E$ be a cluster point. If $\lambda^* \in (0,1)^E$ then $F' \in \mathcal{G}_E$ and $E = E'$, a contradiction. Thus $\lambda^* \in \partial([0,1]^E)$, and due to $\bar{\phi}_E(\lambda^*) = \phi(F') \in \mathcal{P}$, the assertion follows. ■

Theorem 6.1.11 proves that whenever a sequence of wälder $F_n \in \mathcal{G}_E$ converges $F_n \to F' \in \mathcal{W}$ topology E' and $F' \notin \mathcal{G}_E$, then $F' \in \partial\mathcal{G}_E$. In this sense we have found a relationship between E' and E. In the following section we make this relationship precise and unravel the boundary correspondences via a partial ordering on wald topologies, which was introduced in Definition 4.2.12.

6.2 At Grove's End

In the following theorem, we characterize the boundaries of groves via the partial ordering on wald topologies (cf. Definition 4.2.12).

Theorem 6.2.1. *For wald topologies E and E', the following three statements are equivalent, where the boundary of groves $\partial\mathcal{G}_E$ is as defined in Equation* (6.1.9)*:*

(i) $E' < E$,

(ii) $\mathcal{G}_{E'} \subset \partial\mathcal{G}_E$,

(iii) $\mathcal{G}_{E'} \cap \partial\mathcal{G}_E \neq \emptyset$.

Proof. Let E have label partition $\mathcal{L} = \{L_1, \dots, L_K\}$.

"$(i) \implies (ii)$". Assume that $F' = (E', \lambda') \in \mathcal{G}_{E'}$ with partition $\mathcal{L}' = \{L'_1, \dots, L'_{K'}\}$. By assumption $E' < E$, and we can use the sets from Definition 4.2.15 as well as Lemma 4.2.16. Using Lemma 4.2.16, (ix), set

$$\lambda^*_e := \begin{cases} 0 & e \in R_{\text{dis}} \\ 1 & e \in R_{\text{cut}} \\ 1 - \left(1 - \lambda'_{e'}\right)^{1/|R_{e'}|} & e \in R_{e'},\ e' \in E' \end{cases}$$

to obtain $\lambda^* \in [0,1]^E$, and note that $\lambda^* \in \partial([0,1]^E)$ since $R_{\text{cut}} \cup R_{\text{dis}} \neq \emptyset$ due to $E' < E$ by Lemma 4.2.16, (v). By injectivity of ϕ, it suffices to show $(*)$ (in this case, $F' \in \partial\mathcal{G}_E$):

$$\left(\rho^*_{uv}\right)_{u,v=1}^N := \bar{\phi}_E(\lambda^*) \overset{(*)}{=} \phi(F') =: \left(\rho'_{uv}\right)_{u,v=1}^N \in \mathcal{P}.$$

First, observe that by Equation (6.1.5) for all $u \in L$,

$$\rho^*_{uu} = 1 = \rho'_{uu}\,.$$

Next, again from Equation (6.1.5), for all $u, v \in L$ with $u \neq v$ that are not connected in F', say $u \in L'_{\alpha'_1}$, $v \in L'_{\alpha'_2}$ for some $\alpha'_1, \alpha'_2 \in \{1, \dots, K'\}$, we have $\rho'_{uv} = 0$. If u and v are also not connected in E, then $\rho^*_{uv} = 0 = \rho'_{uv}$. Assume now that u and v are connected in E. Then, by Lemma 4.2.16, (xi), there exists an edge $A|B = e \in R_{\text{cut}}$ with $u \in A$ and $v \in B$, and due to $\lambda^*_e = 1$ by construction, $\rho^*_{uv} = 0 = \rho'_{uv}$.

Finally, for all $u, v \in L$ that are connected in F', we have, due to construction and Lemma 4.2.16, (x),

$$\begin{aligned}
\rho^*_{uv} &= \prod_{e \in E(u,v)} \left(1 - \lambda^*_e\right) \\
&= \left(\prod_{e \in R_{\text{dis}} \cap E(u,v)} \left(1 - \lambda^*_e\right)\right)\left(\prod_{e' \in E'(u,v)} \prod_{e \in R_{e'}} \left(1 - \lambda'_{e'}\right)^{1/|R_{e'}|}\right) \\
&= \prod_{e' \in E'(u,v)} \left(1 - \lambda'_{e'}\right) \;=\; \rho'_{uv}\,.
\end{aligned}$$

Thus, we have shown $\phi(F') = \bar{\phi}_E(\lambda^*)$.

"$(ii) \implies (iii)$" is trivial.

"$(iii) \implies (i)$". Let $F' = (E', \lambda') \in \mathcal{G}_{E'} \cap \partial\mathcal{G}_E$, i.e. there exists $\lambda^* \in \partial([0,1]^E)$ with $\bar{\phi}_E(\lambda^*) = \phi(F') \in \mathcal{P}$. In the following, we will construct a wald $F^\circ = (E^\circ, \lambda^\circ)$ and show that

Claim I: $E^\circ < E$, and

Claim II: $\phi(F^\circ) = \phi(F')$, implying $F^\circ = F'$ and $E^\circ = E'$,

which, in conjunction, yield $E' < E$.

Let $\bar{\phi}_E(\lambda^*) = (\rho^*_{uv})_{u,v=1}^N$. Denote the connectivity classes of L, where $u, v \in L$ are connected if and only if $\rho^*_{uv} > 0$, by

$$\mathcal{L}^\circ = \{L_1^\circ, \dots, L_{K^\circ}^\circ\},$$

with $1 \leq K^\circ \leq N$. By Equation (6.1.5), if $\rho^*_{uv} > 0$ for $u \neq v$, it follows that $u, v \in L_\alpha$ for some $1 \leq \alpha \leq K$ and therefore also $\rho_{uv} > 0$ by Equation (6.1.3). So $\rho^*_{uv} > 0$ implies $\rho_{uv} > 0$, and we have that $L_1^\circ, \dots, L_{K^\circ}^\circ$ is a *refinement* of $L_1, \dots, L_K$ by Lemma 4.2.16, (xii).

Define E° by setting for each $1 \leq \alpha^\circ \leq K^\circ$ (where, say, $L^\circ_{\alpha^\circ} \subset L_\alpha$ for some $1 \leq \alpha \leq K$)

$$E^\circ_{\alpha^\circ} := \left\{ e\big|_{L^\circ_{\alpha^\circ}} : e \in E, \text{ if } e\big|_{L^\circ_{\alpha^\circ}} \text{ exists and } \lambda^*_e \neq 0 \right\}, \tag{6.2.1}$$

and consequently $E^\circ := \bigcup_{\alpha^\circ} E^\circ_{\alpha^\circ}$. To verify that E° is a wald topology, by Lemma 4.2.16 (vii), each $E^\circ_{\alpha^\circ}$ comprises of compatible splits. We need to check Definition 4.2.4, 5. For any labels $u, v \in L^\circ_{\alpha^\circ} \subset L_\alpha$, since E is a wald topology, there exists an edge $e = A|B \in E_\alpha$ with $u \in A$ and $v \in B$, and any such edge yields a valid split $e^\circ = A^\circ|B^\circ = e\big|_{L^\circ_{\alpha^\circ}}$ with $u \in A^\circ|B^\circ$. Suppose for any split $e \in E$ such that $e^\circ = e\big|_{L^\circ_{\alpha^\circ}}$ it holds that $\lambda^*_e = 0$. Then, by Equation (6.1.5), we had that $\rho^*_{uv} = 1$, so by Lemma 6.1.9 $\bar{\phi}_E(\lambda^*) \notin \mathcal{P}$, a contradiction. Thus there exists at least one edge $e^\circ \in E^\circ$ separating u and v. We conclude that E° is a wald topology. Furthermore, the *restriction property* is satisfied trivially.

Verifying the cut property, suppose there exist $1 \leq \alpha_1^\circ \neq \alpha_2^\circ \leq K^\circ$ and $1 \leq \alpha \leq K$ such that $L^\circ_{\alpha_1^\circ}, L^\circ_{\alpha_2^\circ} \subset L_\alpha$. Hence by construction

$$\rho^*_{us} = 0,\ \rho^*_{uv} > 0 \text{ and } \rho^*_{st} > 0 \text{ for all } u, v \in L^\circ_{\alpha_1} \text{ and } s, t \in L^\circ_{\alpha_2}. \tag{6.2.2}$$

Let now $u \in L^\circ_{\alpha_1^\circ}$ and $s \in L^\circ_{\alpha_2^\circ}$, then by Equation (6.1.5), $\rho^*_{us} = \prod_{e \in E(u,s)} (1 - \lambda^*_e) = 0$, so there must exist $e = A|B \in E(u, s)$ with $\lambda^*_e = 1$. This implies $L^\circ_{\alpha_1^\circ} \subseteq A$ and $L^\circ_{\alpha_2^\circ} \subseteq B$, for otherwise, if $A \not\ni v \in L^\circ_{\alpha_1^\circ}$, say, then $v \in B$ and hence $e \in E(u, v)$ by Definition 4.2.10, so $\rho^*_{uv} = 0$, due to $\lambda^*_e = 1$, a contradiction to Equation (6.2.2). Thus the *cut property* holds.

Having verified all of the properties from Definition 4.2.12, we have shown $E^\circ \leq E$, and we can use the notation introduced in Definition 4.2.15 and Lemma 4.2.16 is applicable for $E^\circ \leq E$. Since λ^* is on the boundary, there must be some $e \in E$ with either $\lambda^*_e = 1 > \lambda_e > 0$ or all $\lambda^*_e < 1$ and there is $\lambda^*_e = 0 < \lambda_e$. In the first case, $e \in R_{\text{cut}}$, in the second case $e \in R_{\text{dis}}$, so that in both cases $E^\circ \neq E$ by Lemma 4.2.16, (v), yielding $E^\circ < E$, which was Claim I.

In order to see Claim II we define suitable edge weights λ°. Let $1 \le \alpha^\circ \le K^\circ$ be arbitrary and let $1 \le \alpha \le K$ be such that $L^\circ_{\alpha^\circ} \subseteq L_\alpha$. For each $e^\circ \in E^\circ_{\alpha^\circ}$, define

$$\lambda^\circ_{e^\circ} := 1 - \prod_{e \in R_{e^\circ}} \left(1 - \lambda^*_e\right). \tag{6.2.3}$$

Indeed, $\lambda^\circ_{e^\circ} \in (0,1)$, since by Lemma 4.2.16 (ix), none of the $e \in R_{e^\circ}$ lie in R_{cut}, we have $\lambda^*_e < 1$, and by Equation (6.2.1) and Lemma 4.2.16 (iii) there exists $e \in R_{e^\circ}$ with $\lambda^*_e > 0$. Thus $F^\circ := (E^\circ, \lambda^\circ)$ is a well defined wald.

We now show the final part of Claim II, namely that $\phi(F') = \phi(F^\circ)$. Recall that $\phi(F^\circ) = \bar{\phi}_E(\lambda^*) = (\rho^*_{uv})^N_{u,v=1}$ and let $\phi(F^\circ) = (\rho^\circ_{uv})^N_{u,v=1}$. By Equation (6.1.5), for all $u \in L$ we have $\rho^*_{uu} = 1 = \rho^\circ_{uu}$ and by definition of the connectivity classes $L^\circ_1, \ldots, L^\circ_{K^\circ}$ we have $\rho^*_{uv} = 0$ if and only if $\rho^\circ_{uv} = 0$ for all $u, v \in L$.

For all other $u, v \in L$, we may assume that $u, v \in L^\circ_{\alpha^\circ}$ with $L^\circ_{\alpha^\circ} \subseteq L_\alpha$ for some $1 \le \alpha^\circ \le K^\circ$ and $1 \le \alpha \le K$. By Lemma 4.2.16 (vii), the sets $R_{\text{dis}} \cap E(u,v)$ in conjunction with R_{e° for all $e^\circ \in E^\circ(u,v)$ form a partition of $E(u,v)$. For the first set we have

$$e \in R_{\text{dis}} \cap E(u,v) \implies \lambda^*_e = 0\,. \tag{6.2.4}$$

Indeed, if $e \in R_{\text{dis}} \cap E(u,v)$ then the restriction $e^\circ := e\big|_{L^\circ_{\alpha^\circ}}$ is a valid split as it splits $L^\circ_{\alpha^\circ}$ into two non empty sets. But as $e \in R_{\text{dis}}$ this split does not exist in E° which, taking into account Equation (6.2.1), is only possible for $\lambda^*_e = 0$.

In consequence, we have (the first and the last equality are the definitions, respectively, the second uses that $R_{\text{dis}} \cap E(u,v)$ and R_{e°, $e^\circ \in E^\circ(u,v)$ partition $E(u,v)$ and the third uses for the first factor (6.2.4) and (6.2.3) for the second factor)

$$\begin{aligned}
\rho^*_{uv} &= \prod_{e \in E(u,v)} \left(1 - \lambda^*_e\right) \\
&= \Bigg(\underbrace{\prod_{e \in R_{\text{dis}} \cap E(u,v)} \left(1 - \lambda^*_e\right)}_{=1}\Bigg)\Bigg(\prod_{e^\circ \in E^\circ(u,v)} \underbrace{\prod_{e \in R_{e^\circ}} \left(1 - \lambda^*_e\right)}_{=1-\lambda^\circ_{e^\circ}}\Bigg) \\
&= \prod_{e^\circ \in E^\circ(u,v)} \left(1 - \lambda^\circ_{e^\circ}\right) \\
&= \rho^\circ_{uv},
\end{aligned}$$

completing the proof. ■

Corollary 6.2.2. *Let E be a wald topology. Then*

$$\partial\mathcal{G}_E = \bigcup_{E'<E} \mathcal{G}_{E'}. \tag{6.2.5}$$

Corollary 6.2.3. *Let E be a wald topology and let $(F_n)_{n\in\mathbb{N}} \subset \mathcal{G}_E \subset \mathcal{W}$ be a sequence of wälder such that $F_n \to F' \in \mathcal{W}$ for some wald $F' \in \mathcal{W}$ with wald topology E'. Then $E' \leq E$.*

Proof. Either $E' = E$ or by Theorem 6.1.11, 3., $F' \in \partial\mathcal{G}_E$, which, in turn, implies by Theorem 6.2.1 that $\partial\mathcal{G}_E \cap \mathcal{G}_{E'} \neq \emptyset$, so $E' < E$. In both cases, we find $E' \leq E$. ■

6.3 Stratification

Theorem 6.3.1. *Wald Space with the smooth structure on every grove $\mathcal{G}_E$ conveyed by ϕ_E, is a Whitney stratified space of type (A).*

Proof. First, we show that $\mathcal{W}$ is a stratified space. Let E be a wald topology with $|E| = i$ for some $i = 0, \ldots, 2N-4$. In the case that the topology of the graph-based forest representative corresponding to E has more than one connected component, one can connect two of the components by inserting an edge connecting two labeled vertices to obtain a new graph-based forest representative and thus wald topology E', and as E is *displayed* by E' in the sense of Moulton & Steel (2004), we have $E < E'$ and $|E'| = i+1$. In the case that E has only one component, and since $i < 2N-3$, one can add another compatible split to obtain a new wald topology E' with $|E'| = i+1$.

We conclude that for any E with $|E| = i$ there exists E' with $|E'| = i+1$ and $E < E'$, and by Theorem 6.2.1, $\mathcal{G}_E \subset \overline{\mathcal{G}_{E'}}$. As E was arbitrary with $|E| = i$, we have that $S_i \subset \overline{S_{i+1}}$ and thus for any $i < j$ that $S_i \subset \overline{S_j}$ as required.

For Whitney condition (A), let $F_1, F_2, \dots \in S_j$ be a sequence of wälder that converges to some wald $F' \in S_i$, so $i < j$, and such that $T_{F_n}S_j \subset \mathcal{S}$ converges to a j-dimensional plane $T \subset \mathcal{S}$ as $n \to \infty$ (in the Grassmannian $G(\dim(\mathcal{S}), j)$).

First of all, we can assume without loss of generality that $F_1, F_2, \dots \in \mathcal{G}_E$ for some wald topology E with $|E| = j$ since S_j is a disjoint union of finitely many groves, and by Corollary 6.2.3 we have that $F_n \to F' \in \mathcal{G}_{E'}$ implies $F' \in \overline{\mathcal{G}_{E'}}$, and furthermore, without loss of generality that $F_n = \phi_E(\lambda^{(n)})$, where $\lambda^{(n)} \in (0,1)^E$, $n \in \mathbb{N}$, with $\lambda^{(n)} \to \lambda^* \in [0,1]^E$ satisfying $\phi(F') = \bar{\phi}_E(\lambda^*)$ (by taking subsequences accordingly). Then, using

$$T_{F_n}\mathcal{G}_E = \operatorname{span}\left\{\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^{(n)})\colon e \in E\right\} \subset \mathcal{S},$$

and using Lemma 2.6.1, we find

$$\begin{aligned}\operatorname{span}\left\{\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*)\colon e \in E\right\} &= \operatorname{span}\left\{\lim_{n\to\infty}\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^{(n)})\colon e \in E\right\}\\ &\subset \lim_{n\to\infty}\operatorname{span}\left\{\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^{(n)})\colon e \in E\right\} = T.\end{aligned}$$

Therefore, showing that $(*)$ holds in

$$T_{F'}\mathcal{G}_{E'} := \operatorname{span}\left\{\frac{\partial\phi_{E'}}{\partial\lambda'_{e'}}(\lambda')\colon e' \in E'\right\} \overset{(*)}{\subset} \operatorname{span}\left\{\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*)\colon e \in E\right\} \subset T$$

would yield the assertion. We will show $(*)$ by showing that for each $e' \in E'$, there exists a constant $c > 0$ and an edge $e \in E$ such that

$$\frac{\partial\phi_{E'}}{\partial\lambda'_{e'}}(\lambda') = c\,\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*).$$

Recall from Equation (6.1.3) and Equation (6.1.5), respectively, the uv-th matrix entry of $\bar{\phi}_E$ and $\phi_{E'}$ for connected u, v,

$$\begin{aligned}\left(\bar{\phi}_E(\lambda^*)\right)_{uv} &= \prod_{e\in E(u,v)}(1-\lambda^*_e),\\ \left(\phi_{E'}(\lambda')\right)_{uv} &= \prod_{e'\in E'(u,v)}(1-\lambda'_{e'}),\end{aligned}$$

and we calculate their partial derivatives, where the symbol 1 is the indicator function:

$$\begin{aligned}\left(\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*)\right)_{uv} &= -1_{e\in E(u,v)}\prod_{\substack{\tilde{e}\in E(u,v)\\ \tilde{e}\neq e}}(1-\lambda^*_{\tilde{e}}),\\ \left(\frac{\partial\phi_{E'}}{\partial\lambda'_{e'}}(\lambda')\right)_{uv} &= -1_{e'\in E'(u,v)}\prod_{\substack{\tilde{e}'\in E'(u,v)\\ \tilde{e}'\neq e'}}(1-\lambda'_{\tilde{e}'}).\end{aligned}$$

The relationship between F' and $\bar{\phi}_E(\lambda^*)$ as derived in the "$(iii) \Rightarrow (i)$" part in the proof of Theorem 6.2.1 holds true here as well, in particular Equation (6.2.1)

$$E'_{\alpha'} = \left\{e' \colon e \in E, e' := e\big|_{L'_{\alpha'}} \text{ is a valid split of } L'_{\alpha'} \text{ and } \lambda^*_e \neq 0\right\}$$

and for each $e' \in E'$ from Equation (6.2.3),

$$\lambda'_{e'} = 1 - \prod_{e \in R_{e'}} (1 - \lambda^*_e) \neq 0. \tag{6.3.1}$$

Consequently, by Lemma 4.2.16 (iii), for any $e' \in E'_{\alpha'}$ there exists $e \in R_{e'}$ with $\lambda^*_e \neq 0$. Let $u, v \in L$ be arbitrary.

1. Assume $e \notin E(u,v)$. By Lemma 4.2.16 (vii), $e' = e\big|_{L'_{\alpha'}} \notin E'(u,v)$ and thus

$$\left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv} = 0 = \left(\frac{\partial \bar{\phi}_E}{\partial \lambda_e}(\lambda^*)\right)_{uv}.$$

2. Assume $e \in E(u,v)$. Then there are two cases:

   a) $e' \notin E'(u,v)$. On the one hand, this implies $\left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv} = 0$, on the other hand either $u \notin L'_{\alpha'}$ or $v \notin L'_{\alpha'}$, implying

   $$0 = \big(\phi_{E'}(\lambda')\big)_{uv} = \big(\bar{\phi}_E(\lambda^*)\big)_{uv} \stackrel{\text{def}}{=} \prod_{\tilde{e} \in E(u,v)} \big(1 - \lambda^*_{\tilde{e}}\big),$$

   so $\lambda^*_{\tilde{e}} = 1$ for some $\tilde{e} \in E(u,v)$ with $\tilde{e} \neq e$, which implies

   $$\left(\frac{\partial \bar{\phi}_E}{\partial \lambda_e}(\lambda^*)\right)_{uv} = 0 = \left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv}.$$

   b) $e' \in E'(u,v)$. In this case by Lemma 4.2.16 (vii), $R_{e'} \subset E(u,v)$ and we have

   $$\left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv} = - \prod_{\substack{\tilde{e}' \in E'(u,v) \\ \tilde{e}' \neq e'}} \big(1 - \lambda'_{\tilde{e}'}\big) = - \prod_{\substack{\tilde{e} \in E(u,v) \\ \tilde{e} \notin R_{e'}}} \big(1 - \lambda^*_{\tilde{e}}\big)$$

and furthermore,

$$\begin{aligned}\left(\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*)\right)_{uv} &= -\prod_{\substack{\tilde{e}\in E(u,v)\\ \tilde{e}\neq e}}(1-\lambda^*_{\tilde{e}})\\ &= -\Bigg(\prod_{\substack{\tilde{e}\in R_{e'}\\ \tilde{e}\neq e}}(1-\lambda^*_{\tilde{e}})\Bigg)\Bigg(\prod_{\substack{\tilde{e}\in E(u,v)\\ \tilde{e}\notin R_{e'}}}(1-\lambda^*_{\tilde{e}})\Bigg)\\ &= \Bigg(\prod_{\substack{\tilde{e}\in R_{e'}\\ \tilde{e}\neq e}}(1-\lambda^*_{\tilde{e}})\Bigg)\left(\frac{\partial\phi_{E'}}{\partial\lambda'_{e'}}(\lambda')\right)_{uv},\end{aligned}$$

where the first factor (call it c) does not depend on u and v and must be non-zero by Equation (6.3.1).

From all these cases, we find

$$\frac{\partial\phi_{E'}}{\partial\lambda'_{e'}}(\lambda') = c\,\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*),$$

concluding the proof. ■

Having Whitney condition (A), this later allows for interpreting Wald Space as a Riemann stratified space.

Chapter 7

Geometries for Wald Space

In Chapter 6, the Wald Space is introduced, that is a topological space $\mathcal{W}$ with topology inherited from embedding it into the manifold of strictly positive definite $N \times N$ symmetric matrices $\mathcal{P}$ via $\phi\colon \mathcal{W} \to \mathcal{P}$. In this chapter, we define a general framework on how to obtain a metric on $\mathcal{W}$ via choosing a Riemannian metric on $\mathcal{P}$ and state some general results. Recall the definitions of Riemannian manifolds, Riemannian metrics in Definition 2.3.1 and the Riemannian distance in Definition 2.3.5. Furthermore, recall the concept of an induced intrinsic metric and lengths of curves in metric spaces (Lemma 2.1.3).

7.1 General Results for Geometries on Wald Space

If g is a Riemannian metric on $\mathcal{P}$, we construct a length space $(\mathcal{W}, d_g^*)$ by first taking the induced intrinsic metric d_g^* with respect to the metric space $(\phi(\mathcal{W}), d_g)$, where d_g is the Riemannian distance on $(\mathcal{P}, g)$, and then pull-back this construction to $\mathcal{W}$ via ϕ. But one has to be cautious with the topologies: the topology induced by the metric space $(\phi(\mathcal{W}), d_g^*)$ might not necessarily be the same (it can be finer) as the topology induced by $(\phi(\mathcal{W}), d_g)$, and thus the metric space $(\mathcal{W}, d_g^*)$ can carry a topology that is finer than the topology of the Wald Space. However, we show that this is not the case.

Lemma 7.1.1. *Let g be a Riemannian metric on $\mathcal{P}$ and let E be a wald topology. For any point $P \in \phi(\overline{\mathcal{G}_E})$ with $P = \bar{\phi}_E(\lambda)$ for some $\lambda \in [0,1]^E$ there exists a compact neighborhood $U \subset \mathbb{R}^E$ of λ with $\bar{\phi}_E(U) \subset \mathcal{P}$ and a constant $c > 0$ such that for all $\lambda' \in U$ it holds that*

$$d_g^*\big(P, \bar{\phi}_E(\lambda')\big) \leq c\|\lambda - \lambda'\|_2.$$

Proof. First of all, since $\bar{\phi}_E$ is continuous, $\mathcal{P} \subset \mathcal{S}$ is open and $\bar{\phi}_E(\lambda) \in \mathcal{P}$, there must exist an open neighborhood $U' \subset \mathbb{R}^E$ around λ such that $\bar{\phi}_E(\lambda') \in \mathcal{P}$ for all $\lambda' \in U'$. Let $U \subset U'$

be a compact and convex set and $\varepsilon > 0$ with $B_\varepsilon = \{x \in \mathbb{R}^E : \|x\|_2 \leq \varepsilon\}$ such that for any $\lambda' \in U$, we have $\lambda' + B_\varepsilon \subset U'$. Then, for any $\lambda' \in U$ and $x \in B_\varepsilon$ the function

$$f_{\lambda'}(x) = d_g^2\big(\bar{\phi}_E(\lambda'), \bar{\phi}_E(\lambda' + x)\big)$$

is well-defined and smooth. Note that $f_{\lambda'}(0) = 0$ and furthermore

$$f'_{\lambda'}(0) = 2d_g\big(\bar{\phi}_E(\lambda), \bar{\phi}_E(\lambda + x)\big) \text{ grad}\, d_g\big(\bar{\phi}_E(\lambda), \bar{\phi}_E(\lambda + \cdot)\big)\big|_{x=0} = 0,$$

and thus using the Taylor expansion of f at $x = 0$ yields $f_{\lambda'}(x) = O(\|x\|_2^2)$ in particular $f_{\lambda'}(x) \leq c_{\lambda'}^2 \|x\|_2^2$ for some constant $c_{\lambda'} > 0$ depending on λ', for all $x \in B_\varepsilon$.

Let $\lambda' \in U$ be arbitrary and set $Q = \bar{\phi}_E(\lambda')$. Furthermore, denote the straight line segment from λ to λ' by

$$\gamma\colon [0,1] \to U, \qquad \gamma(t) = \lambda + t(\lambda' - \lambda),$$

where l is well-defined as U is convex, and observe that $\|\gamma(t') - \gamma(t)\|_2 = |t' - t| \|\lambda' - \lambda\|_2$ for all $t, t' \in [0,1]$. Finally, note that $c := \sup_{\lambda' \in U} c_{\lambda'} < \infty$ as U is compact. By definition of the induced intrinsic metric (cf. Lemma 2.1.3), we can bound d_g^* from above by plugging in the curve $\bar{\phi}_E \circ \gamma$ instead of taking the infimum over all continuous paths connecting P and Q. This yields, where for $t_0 < t_1 < \ldots < t_n$ in the supremum, we can assume without loss of generality that $n \in \mathbb{N}$ is large enough and $\|\gamma(t_{i+1}) - \gamma(t_i)\|_2 = (t_{i+1} - t_i)\|\lambda' - \lambda\|_2 \leq \varepsilon$ for all $i = 0, \ldots, n-1$,

$$\begin{aligned}
d_g^*(P, Q) &= d_g^*\big(\bar{\phi}_E(\lambda), \bar{\phi}_E(\lambda')\big) \\
&\leq \sup_{\substack{t_0 < t_1 < \ldots < t_n \\ t_0 = 0, t_n = 1 \\ n \in \mathbb{N}}} \sum_{i=0}^{n-1} d_g\Big(\bar{\phi}_E\big(\gamma(t_i)\big), \bar{\phi}_E\big(\gamma(t_{i+1})\big)\Big) \\
&= \sup_{\substack{t_0 < t_1 < \ldots < t_n \\ t_0 = 0, t_n = 1 \\ n \in \mathbb{N}}} \sum_{i=0}^{n-1} \Big(f_{\gamma(t_i)}\big(\gamma(t_{i+1}) - \gamma(t_i)\big)\Big)^{1/2} \\
&\leq \sup_{\substack{t_0 < t_1 < \ldots < t_n \\ t_0 = 0, t_n = 1 \\ n \in \mathbb{N}}} \sum_{i=0}^{n-1} c_{\gamma(t_i)} \,(t_{i+1} - t_i)\, \|\lambda' - \lambda\|_2 \\
&\leq c \|\lambda' - \lambda\|_2.
\end{aligned}$$

We conclude the assertion. ■

Proposition 7.1.2. *If g is a Riemannian metric on $\mathcal{P}$, the topologies of the metric spaces $(\phi(\mathcal{W}), d_g)$ and $(\phi(\mathcal{W}), d_g^*)$ agree.*

Proof. By definition we have that $d_g^* \geq d_g$, which implies that sequences that converge with respect to d_g^* also converge with respect to d_g. For the other direction, let $(P_n)_{n\in\mathbb{N}} \subset \phi(\mathcal{W})$ be a sequence that converges to $P \in \phi(\mathcal{W})$ with respect to d_g, i.e. $d_g(P, P_n) \to 0$ as $n \to \infty$. Using that the topology of the Wald Space is by definition the one induced by $(\phi(\mathcal{W}), d_g)$, we can apply Theorem 6.1.11 to find a subsequence n_k, $k \in \mathbb{N}$, and a wald topology E, such that there are wälder $F_{n_k} = (E, \lambda^{(k)}) \in \mathcal{W}$ with $\bar{\phi}_E(\lambda^{(k)}) = P_{n_k}$ for all $k \in \mathbb{N}$ and such that $\lambda^{(k)} \to \lambda^*$ for some $\lambda^* \in [0,1]^E$ and $\bar{\phi}_E(\lambda^*) = P$. By Lemma 7.1.1,

$$d_g^*(P, P^{(k)}) \leq c\|\lambda^* - \lambda^{(k)}\|_2 \to 0, \quad \text{as } k \to \infty.$$

■

The following result is a direct consequence of Proposition 7.1.2 and Burago et al. (2001, Proposition 2.3.12).

Proposition 7.1.3. *Let g be a Riemannian metric on $\mathcal{P}$. Then any continuous curve $\gamma\colon [0,1] \to \phi(\mathcal{W})$ in $(\phi(\mathcal{W}), d_g)$ is continuous in $(\phi(\mathcal{W}), d_g^*)$ and vice versa, and $L_{d_g}(\gamma) = L_{d_g^*}(\gamma)$.*

Another practical result is the following consequence of Lemma 7.1.1.

Lemma 7.1.4. *Let g be a Riemannian metric on $\mathcal{P}$ and let E be a wald topology. Let $\gamma\colon [0,1] \to [0,1]^E$ be a continuous curve and assume that $\bar{\phi}_E \circ \gamma\colon [0,1] \to \bar{\phi}_E(\overline{\mathcal{G}_E}) \subset \mathcal{P}$ is well-defined. Then, if γ has finite length measured with respect to the Euclidean distance on $[0,1]^E$, the continuous curve $\bar{\phi}_E \circ \gamma$ has finite length measured in $(\phi(\mathcal{W}), d_g^*)$ and $(\phi(\mathcal{W}), d_g)$.*

Proof. Denote the length of γ with respect to the Euclidean distance on $[0,1]^E$ by L. Use Lemma 7.1.1 and the fact that without loss of generality for the supremum we can assume that $\|\gamma(t_i) - \gamma(t_{i+1})\| < \varepsilon$ for all $i = 0, \dots, n-1$, for any $\varepsilon > 0$, and we obtain

$$\begin{aligned} L_{d_g}(\bar{\phi}_E \circ \gamma) = L_{d_g^*}(\bar{\phi}_E \circ \gamma) &\leq \sup_{\substack{t_0 < t_1 < \ldots < t_n \\ t_0 = 0, t_n = 1 \\ n \in \mathbb{N}}} \sum_{i=0}^{n-1} d_g\Big(\bar{\phi}_E\big(\gamma(t_i)\big), \bar{\phi}_E\big(\gamma(t_{i+1})\big)\Big) \\ &\leq \sup_{\substack{t_0 < t_1 < \ldots < t_n \\ t_0 = 0, t_n = 1 \\ n \in \mathbb{N}}} \sum_{i=0}^{n-1} c_{\gamma(t_i)} \big\|\gamma(t_i) - \gamma(t_{i+1})\big\|_2 \\ &\leq cL, \end{aligned}$$

where $c = \sup_{\gamma(t)\in[0,1]} c_{\gamma(t)} < \infty$ as the supremum is taken over a compact set. So in particular, $L_{d_g}(\bar{\phi}_E \circ \gamma) \leq cL < \infty$ by assumption, which yields the assertion. ■

In particular, we have the following useful result.

Proposition 7.1.5. *Let g be a Riemannian metric on $\mathcal{P}$. For any two points $P, Q \in \phi(\mathcal{W})$ there exists a continuous path connecting them with finite length with respect to d_g. In particular, $d_g^*(P, Q) < \infty$ for all $P, Q \in \phi(\mathcal{W})$.*

Proof. Any point $P \in \phi(\mathcal{W})$ corresponds to a wald $F \in \mathcal{G}_E$ for some wald topology E with $P = \phi_E(\lambda)$ for some $\lambda \in (0,1)^E$. The straight line segment in $[0,1]^E$ from λ to the vector $\mathbf{1} \in [0,1]^E$ that has ones in each coordinate, where $\bar{\phi}_E(\mathbf{1}) = I \in \mathcal{P}$, the unit matrix, yields a continuous curve $\gamma\colon [0,1] \to [0,1]^E$ that has finite Euclidean length. Therefore, using Lemma 7.1.4, the continuous curve $\bar{\phi}_E \circ \gamma$ connects P and I and has finite length with respect to d_g and d_g^*. Thus, for arbitrary points $P, Q \in \phi(\mathcal{W})$, take the paths connecting them with I, respectively, and concatenate them, yielding a path from P to Q with finite length.

For the second assertion, denote this path by $\tilde{\gamma}$. Then, $d_g^*(P, Q) \leq L_{d_g}(\tilde{\gamma}) < \infty$ yields the assertion. ■

To conclude, Proposition 7.1.2 justifies the following definition, where we use that ϕ is injective (cf. Corollary 4.4.8).

Definition 7.1.6. Let g be a Riemannian metric on $\mathcal{P}$ inducing the Riemannian distance d_g on $\mathcal{P}$. The *Wald Space equipped with metric induced by g* is the metric space $(\mathcal{W}, \tilde{d}_g)$ defined by

$$\tilde{d}_g(F, F') := d_g^*(\phi(F), \phi(F'))$$

for all $F, F' \in \mathcal{W}$, where d_g^* is the induced intrinsic metric from $(\phi(\mathcal{W}), d_g)$.

Remark 7.1.7. First of all, note that the Wald Space from Definition 6.1.1 is a topological space and Proposition 7.1.2 ensures that the topology induced by $\tilde{d}_g$ is the same as the topology of the Wald Space.

Second, we could have started with just a metric d on $\mathcal{P}$ instead of the Riemannian manifold structure and assumed that the topologies of $(\phi(\mathcal{W}), d)$ and $(\phi(\mathcal{W}), d^*)$ agree, and we would still have obtained a well-defined Wald Space equipped with a metric. However, the Riemann exponential and logarithm will prove to be very useful for designing numerical algorithms that compute approximated geodesics on $(\mathcal{W}, d_g^*)$. That is why from the very beginning, we use a Riemannian manifold structure on $\mathcal{P}$. And as we have seen in Chapter 3, there exist several of Riemannian metrics on $\mathcal{P}$ to choose from.

For the next theorem, recall completeness and local compactness from Definition 2.1.7.

Theorem 7.1.8. *Let g be a Riemannian metric on $\mathcal{P}$ with Riemannian distance d_g and suppose that $(\mathcal{P}, d_g)$ is complete. Then the Wald Space equipped with metric induced by g, $(\mathcal{W}, \tilde{d}_g)$, is a geodesic space.*

Proof. Proposition 6.1.3 yields that $\phi(\mathcal{W})$ is closed in $\mathcal{P}$. Consequently, as any manifold is locally compact, $\phi(\mathcal{W})$ is locally compact, and since $(\mathcal{P}, d_g)$ is complete by assumption, $(\phi(\mathcal{W}), d_g)$ is complete.

Furthermore, since by Proposition 7.1.5, any two points are connected by a continuous path of finite length in $(\phi(\mathcal{W}), d_g)$, which is complete, applying Theorem 2.1.9 yields that $(\phi(\mathcal{W}), d_g^*)$ is complete.

Finally, as $(\phi(\mathcal{W}), d_g^*)$ is by definition a length space that is complete and locally compact (the latter is the case since $(\phi(\mathcal{W}), d_g)$ and $(\phi(\mathcal{W}), d_g^*)$ carry the same topology due to Proposition 7.1.2), we can apply the Hopf-Rinow Theorem for metric spaces (cf. Theorem 2.1.8), which concludes the theorem. ■

Recall from Definition 2.6.3 the definition of a Riemann stratified space.

Theorem 7.1.9. *Let g be a Riemannian metric on $\mathcal{P}$. The Wald Space equipped with metric induced by g, $(\mathcal{W}, \tilde{d}_g)$ is a Riemann stratified space of type (A).*

Proof. As we equip all of $\phi(\mathcal{W}) \subset \mathcal{P}$ (i.e. all strata) with g, the assertion follows immediately. ■

Geometry on Maximal Groves

Given a wald topology E, on can construct an intrinsically induced distance on $\phi(\mathcal{G}_E)$ via $(\phi(\mathcal{G}_E), d_g^*)$, where the $*$ operator is with respect to $\phi(\mathcal{G}_E)$. Pulling back onto $\mathcal{G}_E$ then yields another metric space $(\mathcal{G}_E, \tilde{d}_g^E)$, and it is not clear whether $\tilde{d}_g$ restricted to $\mathcal{G}_E$ equals $\tilde{d}_g^E$, in general it does not need to be true.

However, Proposition 6.1.4 teaches that for a fully resolved tree topology E, $\mathcal{G}_E$ is open in $\mathcal{W}$, and thus, as $\mathcal{G}_E$ is an embedded submanifold of $\mathcal{P}$ by Theorem 6.1.8, the geometries agree locally. Therefore, geodesics computed in the thus obtained Riemannian manifold

$\mathcal{G}_E$ (by restricting the metric g on $\mathcal{P}$ to $\phi(\mathcal{G}_E)$ are geodesics in the metric space $(\mathcal{W}, \tilde{d}_g)$, in a small enough neighborhood. This is the motivation to view groves of fully resolved trees as Riemannian manifolds.

In the following, we identify $\mathcal{G}_E \cong (0,1)^E$ as in Definition 6.1.7 and recall the embedding $\phi_E\colon (0,1)^E \cong \mathcal{G}_E \to \mathcal{P}$ from Definition 6.1.7. Further, recall Theorem 6.1.8 stating that ϕ_E is a smooth embedding between manifolds.

The tangent space at $\lambda \in \mathcal{G}_E$ is then given by $T_\lambda \mathcal{G}_E \cong \mathbb{R}^E$, and denote the differential of ϕ_E by

$$(\mathrm{d}\phi_E)_\lambda(x) = \sum_{i\in E} x_i \frac{\partial \phi_E}{\partial \lambda_i}(\lambda).$$

The Riemannian metric on $\mathcal{G}_E$ is computed by requiring ϕ_E to be an isometric embedding in the manifold sense, i.e. for any $x, y \in \mathbb{R}^E$, setting $P = \phi_E(\lambda) \in \mathcal{P}$, we have

$$g^E_\lambda(x,y) = g_P\Big((\mathrm{d}\phi_E)_\lambda(x),\ (\mathrm{d}\phi_E)_\lambda(y)\Big) = \sum_{i\in E}\sum_{j\in E} x_i y_j g^E_{ij},$$

where for all $i, j \in E$,

$$g^E_{ij} = g^E_{ij}(\lambda) = g_P\left(\frac{\partial \phi_E}{\partial \lambda_i}(\lambda),\ \frac{\partial \phi_E}{\partial \lambda_j}(\lambda)\right), \tag{7.1.1}$$

The Gram matrix of g^E is then $(g^E_{ij})_{i,j\in E}$ with inverse denoted by $(g_E^{ij})_{i,j\in E}$. By definition of the Christoffel symbols in Equation (2.3.1) we have with $i, j, k \in E$,

$$\Gamma^k_{ij} = \frac{1}{2}\sum_{l\in E}\left(\frac{\partial g^E_{jl}}{\partial \lambda_i} + \frac{\partial g^E_{li}}{\partial \lambda_j} - \frac{\partial g^E_{ij}}{\partial \lambda_l}\right) g_E^{lk} \tag{7.1.2}$$

Consequently, recall from Equation (2.4.4) that the symbols of the curvature tensor are computed via

$$R_{ijkl} = \sum_{s\in E} R^s_{ijk} g^E_{sl} = \sum_{s\in E}\left(\sum_{h\in E}\Gamma^h_{ik}\Gamma^s_{jh} - \sum_{h\in E}\Gamma^h_{jk}\Gamma^s_{ih} + \frac{\partial}{\partial \lambda_j}\Gamma^s_{ik} - \frac{\partial}{\partial \lambda_i}\Gamma^s_{jk}\right) g^E_{sl}. \tag{7.1.3}$$

We continue to introduce several different geometries for Wald Space.

7.2 Schwarzwald Space: Fisher-Information Geometry

Recall the Fisher-information metric for $\mathcal{P}$ from Section 3.1. Recall the corresponding Riemannian metric, given by

$$g_P(X, Y) = \mathrm{Tr}\left[P^{-1}XP^{-1}Y\right]$$

for $P \in \mathcal{P}$, $X, Y \in \mathcal{S} \cong T_P\mathcal{P}$. Thus, with Definition 7.1.6, we can construct a metric for the Wald Space.

Definition 7.2.1. We call the metric space that is the Wald Space equipped with the metric induced by the Fisher-information metric g the *Schwarzwald Space*, and denote it by $(\mathcal{W}, d_{\mathrm{S}})$.

We call this space the Schwarzwald Space, as it was first proposed at the Oberwolfach workshop 1804 (2018) in the black forest which is the Schwarzwald in German (the reason for the name was already mentioned in Lueg et al. (2021)). A direct consequence of Theorem 7.1.8 and Theorem 7.1.9 is the following theorem.

Theorem 7.2.2. *The Schwarzwald Space is a geodesic space. Furthermore, it is a Riemann stratified space of type (A).*

Proof. By Lang (1999, p.325) and as mentioned in Section 3.1 $(\mathcal{P}, g)$ is geodesically complete, and by the Hopf-Rinow theorem for Riemannian manifolds (cf. Theorem 2.3.8) it follows that $(\mathcal{P}, d_g)$ is complete. Thus by Theorem 7.1.8, the Schwarzwald Space is a geodesic metric space. The second assertion follows immediately from Theorem 7.1.9. ■

7.2.1 Schwarzwald Space for $N = 2$

Let $N = 2$, and consequently $L = \{1, 2\}$. Then, there are two possible wald topologies:

$$E = \{1|2\} \qquad \text{and} \qquad E_\infty = \emptyset,$$

where $\mathcal{G}_{E_\infty} = \partial\mathcal{G}_E$ and $\mathcal{W} = \mathcal{G}_E \cup \mathcal{G}_{E_\infty}$. Identifying the grove $\mathcal{G}_E \cong (0,1)^E \cong (0,1)$ yields $\bar{\phi}_E\colon [0,1]^E \to \mathcal{S}$ with

$$\bar{\phi}_E(\lambda) = \begin{pmatrix} 1 & 1-\lambda \\ 1-\lambda & 1 \end{pmatrix},$$

where due to Lemma 6.1.9 we have that $\bar{\phi}_E(\lambda) \in \mathcal{P}$ whenever $\lambda \in (0, 1]$.

Since $\overline{\mathcal{G}_E} = \mathcal{W}$ is one-dimensional, we can compute the distance between two wälder $F_1, F_2 \in \mathcal{W}$ with $\phi(F_1) = \bar{\phi}_E(\lambda_1)$ and $\phi(F_2) = \bar{\phi}_E(\lambda_2)$ by choosing any non self-intersecting differentiable curve γ from λ_1 to λ_2 and compute the length of $\bar{\phi}_E \circ \gamma$ to obtain their distance $d_S(F_1, F_2)$. To be precise, we need to compute

$$d_S(F_1, F_2) = \int_0^1 g_{\bar{\phi}_E(\gamma(t))}\Big((\bar{\phi}_E \circ \gamma)'(t), (\bar{\phi}_E \circ \gamma)'(t) \Big)^{1/2} \mathrm{d}t.$$

We will compute the different terms involved in this equation step by step. The first quantity is the matrix inverse of $\bar{\phi}_E(\lambda)$, $\lambda \in (0, 1]$,

$$\bar{\phi}_E(\lambda)^{-1} = \frac{1}{\lambda(2-\lambda)} \begin{pmatrix} 1 & \lambda - 1 \\ \lambda - 1 & 1 \end{pmatrix}.$$

Choosing the curve $\gamma(t) = \lambda_1 + t(\lambda_2 - \lambda_1)$ then yields for any $t \in [0, 1]$,

$$(\bar{\phi}_E \circ \gamma)'(t) = \frac{\partial \bar{\phi}_E}{\partial \lambda}(\gamma(t)) \cdot \gamma'(t) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} (\lambda_2 - \lambda_1),$$

and we conclude

$$\begin{aligned} g_{\bar{\phi}_E(\gamma(t))}\Big((\bar{\phi}_E \circ \gamma)'(t), (\bar{\phi}_E \circ \gamma)'(t) \Big) &= \mathrm{Tr}\left[\Big(\bar{\phi}_E(\gamma(t))^{-1} (\bar{\phi}_E \circ \gamma)'(t) \Big)^2 \right] \\ &= (\lambda_2 - \lambda_1)^2 \frac{1}{\gamma(t)^2 \, (2 - \gamma(t))^2} \mathrm{Tr}\left[\begin{pmatrix} 1 - \gamma(t) & -1 \\ -1 & 1 - \gamma(t) \end{pmatrix}^2 \right] \\ &= 2(\lambda_2 - \lambda_1)^2 \frac{1 + (1 - \gamma(t))^2}{\gamma(t)^2 \, (2 - \gamma(t))^2}, \end{aligned}$$

and therefore (where in the step $\overset{(*)}{=}$ we substitute $\gamma(t) = \lambda$ with $\mathrm{d}\lambda = \gamma'(t)\mathrm{d}t = (\lambda_2 - \lambda_1)\mathrm{d}t$),

$$\begin{aligned} d_S(F_1, F_2) &= 2\,|\lambda_2 - \lambda_1| \int_0^1 \frac{\sqrt{1 + (1 - \gamma(t))^2}}{\gamma(t)\,(2 - \gamma(t))} \mathrm{d}t \\ &\overset{(*)}{=} \sqrt{2} \left| \int_{\lambda_1}^{\lambda_2} \frac{\sqrt{1 + (1 - \lambda)^2}}{\lambda\,(2 - \lambda)} \mathrm{d}\lambda \right|. \end{aligned}$$

Then with

$$f(\lambda) := \sqrt{2}\sqrt{1 + (1 - \lambda)^2}$$

we find

$$\int \frac{\sqrt{1+(1-\lambda)^2}}{\lambda\,(2-\lambda)}\,\mathrm{d}\lambda = \sinh^{-1}(1-\lambda) + \frac{1}{\sqrt{2}}\tanh^{-1}\left(\frac{\lambda}{f(\lambda)}\right) - \frac{1}{\sqrt{2}}\tanh^{-1}\left(\frac{2-\lambda}{f(\lambda)}\right),$$

so using $\tanh^{-1}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$ for $|x| < 1$, we find

$$\begin{aligned}\int \frac{\sqrt{1+(1-\lambda)^2}}{\lambda\,(2-\lambda)}\,\mathrm{d}\lambda &= \sinh^{-1}(1-\lambda) + \frac{1}{2\sqrt{2}}\ln\left(\frac{f(\lambda)+\lambda}{f(\lambda)-\lambda}\right) - \frac{1}{2\sqrt{2}}\ln\left(\frac{f(\lambda)+2-\lambda}{f(\lambda)-2+\lambda}\right)\\ &= \sinh^{-1}(1-\lambda) + \frac{1}{2\sqrt{2}}\ln\left(\frac{\big(f(\lambda)-(1-\lambda)\big)^2-1}{\big(f(\lambda)+(1-\lambda)\big)^2-1}\right).\end{aligned}$$

We conclude with $\sinh^{-1}(x) = \ln(x+\sqrt{x^2+1})$ that

$$\begin{aligned}d_{\mathrm{S}}(F_1,F_2) = \Bigg|&\sqrt{2}\ln\left(\frac{1-\lambda_2+\frac{1}{\sqrt{2}}f(\lambda_2)}{1-\lambda_1+\frac{1}{\sqrt{2}}f(\lambda_1)}\right)\\ &+\frac{1}{2}\ln\left(\frac{\big(f(\lambda_2)-(1-\lambda_2)\big)^2-1}{\big(f(\lambda_1)-(1-\lambda_1)\big)^2-1}\cdot\frac{\big(f(\lambda_1)+(1-\lambda_1)\big)^2-1}{\big(f(\lambda_2)+(1-\lambda_2)\big)^2-1}\right)\Bigg|.\end{aligned}$$

Plugging in $\lambda_2 = 1$, we find

$$d_{\mathrm{S}}(F_1,F_\infty) = \sqrt{2}\left|\frac{1}{2\sqrt{2}}\ln\left(\frac{(f(\lambda_1)+(1-\lambda_1))^2-1}{(f(\lambda_1)-(1-\lambda_1))^2-1}\right) - \ln\left(1-\lambda_1+\frac{1}{\sqrt{2}}f(\lambda_1)\right)\right|$$

which is finite for all $\lambda_1 \in (0,1]$ and *strictly monotonically decreasing* as $\lambda_1 \nearrow 1$, see also Figure 7.1, where the distance from a wald $F_\lambda = (E,\lambda)$ to F_∞ for any value of $\lambda \in (0,1)$ is depicted. Note that the distance behaves almost linear when λ is close to one, which is consistent with the fact that at $\phi(F_\infty) = I$, the Riemannian metric is $g_I(X,Y) = \mathrm{Tr}[XY]$, which is the same as the Euclidean inner product, and thus locally at I the space $\mathcal{P}$ should "behave Euclidean".

7.2.2 Sectional Curvature in Groves

Using the Riemannian metric introduced on groves and the consequential formulas for the Christoffel symbols and Riemann curvature tensor, defined in Equation (7.1.1), Equation (7.1.2) and Equation (7.1.3), respectively, we compute the sectional curvature on a grove $\mathcal{G}_E$ for a wald topology E. For the sectional curvature from Definition 2.4.2, using the formula for the Riemann curvature tensor symbols R_{ijkl} computed in Equation (A.1.11)

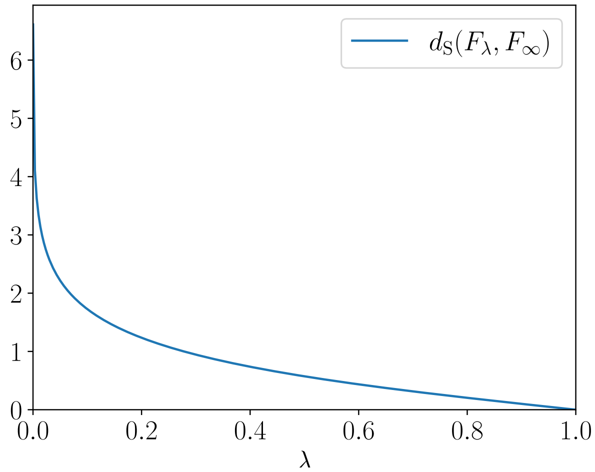

Figure 7.1: *The distance* $d_S(F_\lambda, F_\infty)$ *for* $\lambda \in (0,1)$*, where* $F_\lambda = (E, \lambda)$*, with* $E = \{1|2\}$*.*

and consequently the sectional curvature symbols R_{ijji} from Equation (A.1.12) from Appendix A.1, we conclude that the sectional curvature of the spanned tangent plane of two linearly independent vectors $x, y \in \mathbb{R}^E \cong T_\lambda \mathcal{G}_E$ at some coordinate $\lambda \in (0,1)^E$ is computed by

$$K(x,y) = \frac{\sum_{i,j \in E} R_{ijij}\, x_i\, y_i}{|x|^2|y|^2 - g_\lambda^E(x,y)^2},$$

where with $Q_i := P^{-1}\frac{\partial \phi_E}{\partial \lambda_i}(\lambda)$ and $Q_{ij} := P^{-1}\frac{\partial^2 \phi_E}{\partial \lambda_i \partial \lambda_j}(\lambda)$, where $P = \phi_E(\lambda)$,

$$\begin{aligned} R_{ijji} = \;& \frac{1}{4} \sum_{a,h \in E} g_E^{ah} \operatorname{Tr}\Big[Q_a(2Q_{ij} - Q_iQ_j - Q_jQ_i)\Big] \operatorname{Tr}\Big[Q_h(2Q_{ij} - Q_iQ_j - Q_jQ_i)\Big] \\ & - \sum_{a,h \in E} g_E^{ah} \operatorname{Tr}\Big[Q_a Q_j^2\Big] \operatorname{Tr}\Big[Q_h Q_i^2\Big]. \\ & - \operatorname{Tr}\Big[Q_{ij}(Q_{ij} - Q_iQ_j - Q_jQ_i)\Big]. \end{aligned}$$

7.3 Euclidean Induced Geometry

Recall the Euclidean geometry on $\mathcal{P}$ from Section 3.2, where $g_P(X,Y) = \operatorname{Tr}[XY]$ for all $P \in \mathcal{P}$, $X, Y \in \mathcal{S}$. We use Definition 7.1.6 for the following definition.

Definition 7.3.1. We call the metric space that is the Wald Space equipped with the metric induced by the Euclidean metric g the *Wald Space with induced Euclidean geometry*, and denote it by $(\mathcal{W}, d_{\mathrm{E}})$.

From Theorem 7.1.9, we have immediately the following result.

Theorem 7.3.2. *The Wald Space with induced Euclidean geometry is a Riemann stratified space of type (A).*

7.4 Bures-Wasserstein Induced Geometry

Recall the Bures-Wasserstein metric g on $\mathcal{P}$ from Section 3.3. We use Definition 7.1.6 for the following definition.

Definition 7.4.1. We call the metric space that is the Wald Space equipped with the metric induced by the Bures-Wasserstein metric g the *Wald Space with induced Wasserstein geometry*, and denote it by $(\mathcal{W}, d_{\mathrm{W}})$.

From Theorem 7.1.9, we have immediately the following result.

Theorem 7.4.2. *The Wald Space with induced Wasserstein geometry is a Riemann stratified space of type (A).*

7.5 Log-Euclidean Induced Geometry

Recall the Log-Euclidean metric g on $\mathcal{P}$ from Section 3.4. We use Definition 7.1.6 for the following definition.

Definition 7.5.1. We call the metric space that is the Wald Space equipped with the metric induced by the Log-Euclidean metric g the *Wald Space with induced Log-Euclidean geometry*, and denote it by $(\mathcal{W}, d_{\mathrm{LE}})$.

From Theorem 7.1.9, we have immediately the following result.

Theorem 7.5.2. *The Wald Space with induced Log-Euclidean geometry is a Riemann stratified space of type (A).*

7.6 Correlation Quotient Geometry

Note that the map $\phi\colon \mathcal{W} \to \mathcal{P}$ is also well-defined as $\phi\colon \mathcal{W} \to \mathcal{C}$. Recall from Section 3.5, in particular Theorem 3.5.4 that $(\mathcal{C}, g^{(\mathcal{C})})$ is a Riemannian manifold, where the topology is the trace topology from being a subset of the Euclidean space $\mathcal{S}$, i.e. $\mathcal{C} \subset \mathcal{P} \subset \mathcal{S}$, due to the design of the surjective submersion $\pi\colon \mathcal{P} \to \mathcal{C}$ from Equation (3.5.1). Therefore, we can analogously apply Lemma 7.1.1 and the succeeding results in Section 7.1. So, again, we use Definition 7.1.6 for the following definition.

Definition 7.6.1. We call the metric space that is the Wald Space equipped with the metric induced by the *Wald Space with induced correlation quotient geometry* $g^{(\mathcal{C})}$, and denote it by $(\mathcal{W}, d_{\mathrm{C}})$.

From Theorem 7.1.9, we have the following result.

Theorem 7.6.2. *The Wald Space with induced correlation quotient geometry is a Riemann stratified space of type (A).*

Chapter 8

Algorithms for Geodesics in Wald Space

We introduce a general framework with algorithms that are supposed to approximate geodesics in Wald Space equipped with some geometry as in Definition 7.1.6.

Let E be a fully resolved wald topology. Recall once again from Chapter 6 that $\phi_E\colon (0,1)^E \cong \mathcal{G}_E \to \mathcal{P}$ is an embedding (analogously for $\phi_E\colon (0,1)^E \to \mathcal{C}$) and let g be some Riemannian metric on $\mathcal{P}$, such that $(\mathcal{W}, \tilde{d}_g)$ is the Wald Space with geometry induced by g as defined in Definition 7.1.6.

Recall the notation from Definition 2.3.4, i.e. the open neighborhoods $\mathcal{U}_P \subset \mathcal{P}$ of $P \in \mathcal{P}$ and $\mathcal{V}_P \subset \mathcal{S} \cong T_P\mathcal{P}$ of $0 \in T_P\mathcal{P}$, such that the Riemann exponential and Riemann logarithm, as well as the geodesic between two points $P, Q \in \mathcal{P}$, with respect to the Riemannian manifold $(\mathcal{P}, g)$, are given by

$$\mathrm{Exp}_P^g\colon \mathcal{V}_P \to \mathcal{P}, \quad \mathrm{Log}_P^g\colon \mathcal{U}_P \to T_P\mathcal{P} \quad \text{and} \quad \gamma_{P,Q}^g\colon [0,1] \to \mathcal{P},$$

respectively. For any $\lambda \in (0,1)^E \cong \mathcal{G}_E$, the tangent space $\mathbb{R}^E \cong T_\lambda\mathcal{G}_E$ can be embedded as a linear subspace into $T_P\mathcal{P}$ via $(\mathrm{d}\phi_E)_\lambda$, and we write $T_P\mathcal{G}_E := (\mathrm{d}\phi_E)_\lambda\big(T_\lambda\mathcal{G}_E\big)$, such that the orthogonal projection from $T_P\mathcal{P}$ onto $T_P\mathcal{G}_E$ is

$$\pi_P\colon T_P\mathcal{P} \to T_P\mathcal{G}_E,$$

this projection is calculated using an orthonormal basis of $T_P\mathcal{G}_E$ obtained from applying Gram-Schmidt to the following basis of $T_P\mathcal{G}_E$:

$$\frac{\partial\phi_E}{\partial\lambda_e}(\lambda), \quad e \in E.$$

Finally, we will need a projection from $\mathcal{P}$ onto $\mathcal{W} \cong \phi(\mathcal{W}) \subset \mathcal{P}$, which we define by

$$\pi\colon \mathcal{P} \to \mathcal{W}, \qquad P \mapsto \pi(P) \in \operatorname*{arg\,min}_{F\in\mathcal{W}} d_g^2\big(P, \phi(F)\big),$$

where π is only well-defined for $P \in \mathcal{P}$ close enough to the embedded Riemann stratification $\phi(\mathcal{W}) \subset \mathcal{P}$.

The projection π is computed by starting to search in some grove $\mathcal{G}_E$ for some fully resolved wald topology E, and if we have found a local minimum $F = (E, \lambda^*)$ with $\lambda^* \in (0,1)^E$, set $\pi(P) := F$. However, the search might end up on the boundary of the grove, i.e. $\lambda^* \in \partial([0,1]^E)$, and we will terminate the search with an error, if any coordinate $\lambda_e^* = 1$ for some $e \in E$, or if two or more coordinates are zero. However, if exactly one coordinate is zero, so $\lambda_e^* = 0$ for some $A|B = e \in E$ and $0 < \lambda_{e'}^* < 1$ for all $e' \in E \setminus \{e\}$, there are exactly two splits e_1, e_2 that are compatible with $E \setminus \{e\}$, i.e. two neighboring fully resolved wald topologies $E_1 = E \setminus \{e\} \cup \{e_1\}$ and $E_2 = E \setminus \{e\} \cup \{e_2\}$, where then $F := \phi^{-1}(\bar{\phi}_E(\lambda^*))$ satisfies $F \in \overline{\mathcal{G}_E} \cap \overline{\mathcal{G}_{E_1}} \cap \overline{\mathcal{G}_{E_2}}$ (that there are exactly two splits can be explained with the *nearest neighbor interchange (NNI)* operation that is described in Semple & Steel (2003, p.31)).

This motivates to continue the search in the groves of the wald topologies E_1 and E_2, respectively, and repeat the procedure, where the algorithm memorizes the groves which were already searched through, and excludes those from being searched twice.

Definition 8.0.1 (Projection). Let $P \in \mathcal{P}$. Let $F \in \mathcal{G}_E$ be a fully resolved wald and let $\lambda_0 \in (0,1)^E$ with $\phi_E(\lambda_0) = F$. Setup the set of possible wald topologies to be $\mathcal{E} = \{E\}$ and let $\mathcal{E}_{\text{old}} = \emptyset$. Repeat the following steps until the algorithm stops.

1. Choose $E \in \mathcal{E}$ and update $\mathcal{E} := \mathcal{E} \setminus \{E\}$ and $\mathcal{E}_{\text{old}} := \mathcal{E}_{\text{old}} \cup \{E\}$.
2. Define the functional $f_E\colon [0,1]^E \to [0,\infty)$ by

$$f_E(\lambda) = d_g^2\big(P, \bar{\phi}_E(\lambda)\big)$$

and compute $\lambda^* \in \arg\min_{\lambda\in[0,1]^E} f_E(\lambda)$. If $\lambda^* \in (0,1)^E$, stop and return $F^* = (E, \lambda^*)$.
3. If and only if $\lambda_e^* = 0$ for exactly one $e \in E$ and $\lambda_{e'}^* \in (0,1)$ for all other $e' \in E, e' \neq e$, determine the two fully resolved wald topologies E_1, E_2 that satisfy $\phi^{-1}(\bar{\phi}_E(\lambda^*)) \in \overline{\mathcal{G}_{E_1}} \cap \overline{\mathcal{G}_{E_2}}$ and such that E, E_1, E_2 are pair-wise distinct. Update $\mathcal{E} := \mathcal{E} \cup \{E_1, E_2\}$.
4. Else, stop and return an error message that the computation of the projection failed.

The minimization for the computation of λ^* is done using a sequential linear-quadratic programming algorithm[1], which is based on Kraft & Munchen (1994).

Remark 8.0.2. The calculation of this map is the bottleneck of the algorithms that are introduced later in this section. Note that in practice, one would expect the error message in 4. to be raised often, but to my surprise this usually does not happen in practice whenever one is close enough to a fully resolved wald.

We continue to define the algorithms for the approximation of geodesics between two points $F, F' \in \mathcal{W}$. The following is a very simple but naive algorithm. We have included it in our study of possible algorithms in Lueg et al. (2021).

Definition 8.0.3 (Naive Projection (NP))**.** Given $3 \leq n \in \mathbb{N}$ and two fully resolved wälder $F_1', F_2' \in \mathcal{W}$, for $i = 1, \ldots, n$ compute

(1) $F_i = \pi\big(\gamma^g_{F_1', F_2'}\big(\frac{i-1}{n-1}\big)\big)$.

Return $(F_1, \ldots, F_n)$.

Importantly, the weakness of this algorithm is two-fold. First, due to the projection, the discrete path between F_1', F_2' no longer needs to be equidistant. The other weakness is the design of the algorithm that points on the geodesic $\gamma_{F_1', F_2'}$ can be far away from $\phi(\mathcal{W}) \subset \mathcal{P}$, and thus the projection might be significantly inaccurate or fail.

The next algorithm makes small (approximately geodesic) steps and successively takes the geodesic from the newest point to the destination (note the F_{i-1} and G_{i-1} in the subscript in the update step). This algorithm was proposed in Garba et al. (2021a) and also included in the study of algorithms in Lueg et al. (2021).

Definition 8.0.4 (Symmetric Projection (SP))**.** Given odd $3 \leq n \in \mathbb{N}$ and two fully resolved wälder $F_1', F_2' \in \mathcal{W}$, set $F_1 \coloneqq F_1'$ and $G_1 \coloneqq F_2'$. For $i = 2, \ldots, \lfloor \frac{n}{2} \rfloor$, do

(1) $F_i \coloneqq \pi\big(\gamma^g_{F_{i-1}, G_{i-1}}\big(\frac{1}{n-2i+1}\big)\big)$ and

(2) $G_i \coloneqq \pi\big(\gamma^g_{G_{i-1}, F_{i-1}}\big(\frac{1}{n-2i+1}\big)\big)$.

Set $H \coloneqq \pi\big(\gamma^g_{F_{\lfloor \frac{n}{2} \rfloor}, G_{\lfloor \frac{n}{2} \rfloor}}(\frac{1}{2})\big)$ and return

$$(F_1, \ldots, F_{\lfloor \frac{n}{2} \rfloor}, H, G_{\lfloor \frac{n}{2} \rfloor}, \ldots, G_1).$$

Given a proposal $F_1, \ldots, F_n \in \mathcal{W}$ for a geodesic from $F_1' \in \mathcal{W}$ to $F_2' \in \mathcal{W}$ (i.e. $F_1 = F_1'$ and $F_n = F_2'$), the following algorithm iteratively improves the path by "straightening".

[1]https://docs.scipy.org/doc/scipy/reference/optimize.minimize-slsqp.html

It is originally inspired from Schmidt et al. (2006) and is the *extrinsic path straightening algorithm* from Lueg et al. (2021, Algorithm 3).

Definition 8.0.5 (Path Straightening (PS))**.** Let $F_1', F_2' \in \mathcal{W}$ and suppose that $(F_1, \ldots, F_n)$ is a discrete path from F_1' to F_2' with $3 \leq n \in \mathbb{N}$. Given $m \in \mathbb{N}$, for $j = 1, \ldots, m$, do

(1) for $i = 2, \ldots, n-1$ compute

$$X_i = \frac{1}{2}\Big(\operatorname{Log}^g_{\phi(F_i)} \big(\phi(F_{i-1})\big) + \operatorname{Log}^g_{\phi(F_i)} \big(\phi(F_{i+1})\big) \Big),$$

(2) and update $(F_2, \ldots, F_{n-1})$: for $i = 2, \ldots, n-1$ compute

$$F_i \coloneqq \pi\Big(\operatorname{Exp}^g_{\phi(F_i)}(X_i) \Big).$$

Return $(F_1, \ldots, F_n)$.

Note that this algorithm works properly only if the points on the path $F_1, \ldots, F_n$ are successively close enough such that the Riemann logarithm $\operatorname{Log}^g_{\phi(F_i)} \big(\phi(F_{i-1})\big)$ is well defined for all $i = 2, \ldots, n-1$, as well as if $\operatorname{Exp}^g_{\phi(F_i)}(X_i)$ and $\phi(F_i)$ are close enough. However, note that the closer the successive pair-wise points in $F_1, \ldots, F_n$ are (and n needs to be sufficiently big for that), the smaller $|X_i|$ will be, and the smaller the change of the iteration, which slows down the algorithm. Thus there is reason to have a sufficiently large n such that everything is well-defined and exists, but also such that n is not too big, as the computational cost will be unnecessarily large.

We measure the quality of a proposal $(F_1, \ldots, F_n)$, $3 \leq n \in \mathbb{N}$ by its energy,

$$E(F_1, \ldots, F_n) = \frac{1}{2} \sum_{i=1}^{n-1} d_g^2\big(\phi(F_i), \phi(F_{i+1})\big).$$

Note that in order two compare the quality of two discrete paths, they need to have the same number of points $n \in \mathbb{N}$.

Chapter 9

Discussion and Outlook

What are the main contributions of this thesis? We will answer this question chronologically with respect to the contents of the thesis. Chapter 2 concisely introduces the relevant background on Riemannian manifolds, in particular Riemannian submersions in Section 2.5. A summary about the most important geometries on $\mathcal{P}$ is of great interest for those who wish to implement them. Chapter 4 introduces and summarizes various representations of phylogenetic forests, serving as the notational foundation for possible future publications in this field. In Chapter 6, we have shown the most important results about Wald Space, that it is contractible, how it is embedded in $\mathcal{P}$ and finally we have understood the stratified nature of the space via the partial ordering on the wald topologies. Furthermore, in Chapter 7, we have shown that one can choose any Riemannian metric on $\mathcal{P}$ and obtains a well-defined geometry for the Wald Space, turning it into a metric space. Furthermore, we have computed the curvature symbols for the sectional curvatures in the groves of the Schwarzwald Space. In Chapter 8, we introduce a framework for algorithms to approximate geodesics. Although we do not make any simulations in this thesis, we have implemented those algorithms and made simulations in Garba et al. (2021a) and Lueg et al. (2021), and we make some more observations, including the sectional curvatures in Schwarzwald Space, in Lueg et al. (2022).

I have made contributions to Garba et al. (2021a,b); Lueg et al. (2021), and Lueg et al. (2022) includes many of the results presented in this thesis. Furthermore, over time I have developed an unpublished Python package, where I have also implemented the BHV Space that is defined in Chapter 5. Lately, I have begun to merge this code into the Python project *geomstats*, cf. Miolane et al. (2020), where many scientists from the field of non-Euclidean statistics contribute to this well-structured package, so far almost only on Riemannian manifolds. Together with Anna Calissano, we have started to extend the package to also deal with stratified spaces, and the BHV Space as well as a first version of the Wald Space is already online and part of the main repository. The aim of this is to enable anyone to be able to make simulations and computations within BHV Space and Wald Space despite their complicated nature.

There are still plenty of open questions regarding the Wald Space (equipped with some geometry), that include geodesics from trees to actual forests, curvature, stickiness, and in general the very non-intuitive behavior of stratified spaces. It will be exciting to see actual averages of data sets of phylogenetic trees. Finally, it will be interesting to compare the various geometries that one can equip the Wald Space with, and if the Schwarzwald Space proves to be the most reasonable space to use.

A Appendix

A.1 Schwarzwald: Sectional Curvature in Groves

Recall from Equation (7.1.2) and Equation (7.1.3) the curvature and Christoffel symbols, respectively. Let E be a wald topology and let $F \in \mathcal{G}_E$ with $P = \phi(F) = \phi_E(\lambda) \in \mathcal{P}$ for some $\lambda \in (0,1)^E$.

Throughout this chapter, parts of formulas will be highlighted in different colors to improve comprehensibility. For notational convenience, we introduce the following names, where $i, j, k \in E$.

$$Q_i = P^{-1}\frac{\partial \phi_E}{\partial \lambda_i}(\lambda), \qquad Q_{ij} = P^{-1}\frac{\partial^2 \phi_E}{\partial \lambda_i \partial \lambda_j}(\lambda), \quad \text{and} \quad Q_{ijk} = P^{-1}\frac{\partial^3 \phi_E}{\partial \lambda_i \partial \lambda_j \partial \lambda_k}(\lambda),$$

where note that $Q_{ij} = Q_{ji}$ and $Q_{ijk} = Q_{ikj} = Q_{kji} = \dots$ holds for all $i, j, k \in E$. For these symbols, we can use standard differentiation rules (the product rule and $(\partial/\partial\lambda_i)P^{-1} = -P^{-1}\big((\partial/\partial\lambda_i)P\big)P^{-1})$) such that we obtain the following useful rules:

$$\frac{\partial}{\partial \lambda_i}P^{-1} = -Q_iP^{-1}, \qquad \frac{\partial}{\partial \lambda_i}Q_j = Q_{ij} - Q_iQ_j, \quad \text{and} \quad \frac{\partial}{\partial \lambda_i}Q_{jk} = Q_{ijk} - Q_iQ_{jk}.$$

Thus, plugging the Fisher-information metric into the definition of the Gram matrix on groves from Equation (7.1.1) yields (using that the trace is a linear operator and its cyclic permutation property)

$$\frac{\partial g^E_{ij}}{\partial \lambda_l} = \frac{\partial}{\partial \lambda_l}\operatorname{Tr}\left[Q_iQ_j\right] = \operatorname{Tr}\left[Q_iQ_{jl} + Q_jQ_{il} - Q_iQ_jQ_l - Q_jQ_iQ_l\right]$$

Continuing with the Christoffel symbols from Equation (7.1.2),

$$\begin{aligned}\frac{\partial g^E_{jl}}{\partial \lambda_i} + \frac{\partial g^E_{li}}{\partial \lambda_j} - \frac{\partial g^E_{ij}}{\partial \lambda_l} &= \mathrm{Tr}\Big[Q_j Q_{li} + Q_l Q_{ji} - Q_j Q_l Q_i - Q_l Q_j Q_i \\ &\quad + Q_i Q_{jl} + Q_l Q_{ij} - Q_i Q_l Q_j - Q_l Q_i Q_j \\ &\quad - Q_i Q_{jl} - Q_j Q_{il} + Q_i Q_j Q_l + Q_j Q_i Q_l\Big] \\ &= \mathrm{Tr}\Big[2Q_{ij}Q_l - Q_i Q_j Q_l - Q_j Q_i Q_l\Big],\end{aligned}$$

consequently,

$$\Gamma^k_{ij} = \frac{1}{2}\sum_{l\in E}\left(\frac{\partial g^E_{jl}}{\partial \lambda_i} + \frac{\partial g^E_{li}}{\partial \lambda_j} - \frac{\partial g^E_{ij}}{\partial \lambda_l}\right) g_E^{lk} = \frac{1}{2}\sum_{l\in E}\mathrm{Tr}\Big[2Q_{ij}Q_l - Q_i Q_j Q_l - Q_j Q_i Q_l\Big] g_E^{lk}. \tag{A.1.1}$$

We continue to compute the curvature symbols as defined in Equation (7.1.3). For this, we need several results.

$$\begin{aligned}\frac{\partial}{\partial \lambda_k}\mathrm{Tr}\Big[2Q_{ij}Q_l - Q_i Q_j Q_l - Q_j Q_i Q_l\Big] &= \mathrm{Tr}\Big[2Q_{kij}Q_l - 2Q_k Q_{ij} Q_l + 2Q_{ij}Q_{kl} - 2Q_{ij}Q_k Q_l \\ &\quad - Q_{ki}Q_j Q_l + Q_k Q_i Q_j Q_l - Q_i Q_{kj} Q_l + Q_i Q_k Q_j Q_l \\ &\quad - Q_i Q_j Q_{kl} + Q_i Q_j Q_k Q_l \\ &\quad - Q_{kj} Q_i Q_l + Q_k Q_j Q_i Q_l - Q_j Q_{ki} Q_l + Q_j Q_k Q_i Q_l \\ &\quad - Q_j Q_i Q_{kl} + Q_j Q_i Q_k Q_l\Big].\end{aligned} \tag{A.1.2}$$

Using (where $g_E = (g^E)^{-1}$)

$$\frac{\partial g_E}{\partial \lambda_l} = -g_E \frac{\partial g^E}{\partial \lambda_l} g_E,$$

we find

$$\begin{aligned}\frac{\partial g^{ls}_E}{\partial \lambda_i} &= -\sum_{a\in E}\sum_{b\in E} g_E^{la}\frac{\partial g^E_{ab}}{\partial \lambda_i} g_E^{bs} \\ &= -\sum_{a\in E}\sum_{b\in E} g_E^{la} g_E^{bs}\,\mathrm{Tr}\Big[Q_a Q_{bi} + Q_b Q_{ai} - Q_a Q_b Q_i - Q_b Q_a Q_i\Big].\end{aligned} \tag{A.1.3}$$

We conclude

$$
\begin{aligned}
2\left(\frac{\partial\Gamma^s_{ik}}{\partial\lambda_j}-\frac{\partial\Gamma^s_{jk}}{\partial\lambda_i}\right) = &\quad \frac{\partial}{\partial\lambda_j}\sum_{l\in E}\operatorname{Tr}\left[2Q_{ik}Q_l - Q_iQ_kQ_l - Q_kQ_iQ_l\right]g_E^{ls}\\
&-\frac{\partial}{\partial\lambda_i}\sum_{l\in E}\operatorname{Tr}\left[2Q_{jk}Q_l - Q_jQ_kQ_l - Q_kQ_jQ_l\right]g_E^{ls}\\
= &\quad \sum_{l\in E}\left(\frac{\partial}{\partial\lambda_j}\operatorname{Tr}\left[2Q_{ik}Q_l - Q_iQ_kQ_l - Q_kQ_iQ_l\right]\right)g_E^{ls}\\
&+\sum_{l\in E}\operatorname{Tr}\left[2Q_{ik}Q_l - Q_iQ_kQ_l - Q_kQ_iQ_l\right]\frac{\partial g_E^{ls}}{\partial\lambda_j}\\
&-\sum_{l\in E}\left(\frac{\partial}{\partial\lambda_i}\operatorname{Tr}\left[2Q_{jk}Q_l - Q_jQ_kQ_l - Q_kQ_jQ_l\right]\right)g_E^{ls}\\
&-\sum_{l\in E}\operatorname{Tr}\left[2Q_{jk}Q_l - Q_jQ_kQ_l - Q_kQ_jQ_l\right]\frac{\partial g_E^{ls}}{\partial\lambda_i},
\end{aligned}
\tag{A.1.4}
$$

where using Equation (A.1.2) yields

$$
\begin{aligned}
&\frac{\partial}{\partial\lambda_j}\operatorname{Tr}\left[2Q_{ik}Q_l - Q_iQ_kQ_l - Q_kQ_iQ_l\right] - \frac{\partial}{\partial\lambda_i}\operatorname{Tr}\left[2Q_{jk}Q_l - Q_jQ_kQ_l - Q_kQ_jQ_l\right]\\
&\quad= \operatorname{Tr}\Big[2Q_{jik}Q_l - 2Q_jQ_{ik}Q_l + 2Q_{ik}Q_{jl} - 2Q_{ik}Q_jQ_l\\
&\qquad -Q_{ji}Q_kQ_l + Q_jQ_iQ_kQ_l - Q_iQ_{jk}Q_l + Q_iQ_jQ_kQ_l - Q_iQ_kQ_{jl} + Q_iQ_kQ_jQ_l\\
&\qquad -Q_{jk}Q_iQ_l + Q_jQ_kQ_iQ_l - Q_kQ_{ji}Q_l + Q_kQ_jQ_iQ_l - Q_kQ_iQ_{jl} + Q_kQ_iQ_jQ_l\Big]\\
&\quad\; - \operatorname{Tr}\Big[2Q_{ijk}Q_l - 2Q_iQ_{jk}Q_l + 2Q_{jk}Q_{il} - 2Q_{jk}Q_iQ_l\\
&\qquad -Q_{ij}Q_kQ_l + Q_iQ_jQ_kQ_l - Q_jQ_{ik}Q_l + Q_jQ_iQ_kQ_l - Q_jQ_kQ_{il} + Q_jQ_kQ_iQ_l\\
&\qquad -Q_{ik}Q_jQ_l + Q_iQ_kQ_jQ_l - Q_kQ_{ij}Q_l + Q_kQ_iQ_jQ_l - Q_kQ_jQ_{il} + Q_kQ_jQ_iQ_l\Big]\\
&\quad= \operatorname{Tr}\Big[2Q_{ik}Q_{jl} - 2Q_{jk}Q_{il} - Q_jQ_{ik}Q_l + Q_iQ_{jk}Q_l - Q_{ik}Q_jQ_l\\
&\qquad +Q_{jk}Q_iQ_l - Q_iQ_kQ_{jl} - Q_kQ_iQ_{jl} + Q_jQ_kQ_{il} + Q_kQ_jQ_{il}\Big],
\end{aligned}
$$

and thus two of the terms in Equation (A.1.4) simplify to

$$\begin{aligned}
&\sum_{l\in E}\Bigg(\frac{\partial}{\partial\lambda_j}\operatorname{Tr}\Big[2Q_{ik}Q_l - Q_iQ_kQ_l - Q_kQ_iQ_l\Big] \\
&\qquad - \frac{\partial}{\partial\lambda_i}\operatorname{Tr}\Big[2Q_{jk}Q_l - Q_jQ_kQ_l - Q_kQ_jQ_l\Big]\Bigg)g_E^{ls} \\
= \quad &\sum_{l\in E}\operatorname{Tr}\Big[2Q_{ik}Q_{jl} - 2Q_{jk}Q_{il} - Q_jQ_{ik}Q_l + Q_iQ_{jk}Q_l - Q_{ik}Q_jQ_l \\
&\qquad + Q_{jk}Q_iQ_l - Q_iQ_kQ_{jl} - Q_kQ_iQ_{jl} + Q_jQ_kQ_{il} + Q_kQ_jQ_{il}\Big]g_E^{ls}. \qquad \text{(A.1.5)}
\end{aligned}$$

Observe that only terms remain where no factor is differentiated more than twice and in each term, there exists at least one factor that is differentiated twice. For the remaining two terms in Equation (A.1.4), we use Equation (A.1.3) to find

$$\begin{aligned}
&\sum_{l\in E}\operatorname{Tr}\Big[2Q_{ik}Q_l - Q_iQ_kQ_l - Q_kQ_iQ_l\Big]\frac{\partial g_E^{ls}}{\partial\lambda_j} - \sum_{l\in E}\operatorname{Tr}\Big[2Q_{jk}Q_l - Q_jQ_kQ_l - Q_kQ_jQ_l\Big]\frac{\partial g_E^{ls}}{\partial\lambda_i} \\
&= \sum_{a,b,l\in E} g_E^{la}g_E^{bs}\operatorname{Tr}\Big[2Q_{jk}Q_l - Q_jQ_kQ_l - Q_kQ_jQ_l\Big]\operatorname{Tr}\Big[Q_aQ_{bi} + Q_bQ_{ai} - Q_aQ_bQ_i - Q_bQ_aQ_i\Big] \\
&- \sum_{a,b,l\in E} g_E^{la}g_E^{bs}\operatorname{Tr}\Big[2Q_{ik}Q_l - Q_iQ_kQ_l - Q_kQ_iQ_l\Big]\operatorname{Tr}\Big[Q_aQ_{bj} + Q_bQ_{aj} - Q_aQ_bQ_j - Q_bQ_aQ_j\Big].
\end{aligned} \qquad \text{(A.1.6)}$$

Furthermore, using Equation (A.1.1) we find

$$\sum_{h\in E}\Gamma^h_{ik}\Gamma^s_{jh} = \frac{1}{4}\sum_{a,b,h\in E} g_E^{ah}g_E^{bs}\operatorname{Tr}\Big[2Q_{ik}Q_a - Q_iQ_kQ_a - Q_kQ_iQ_a\Big]\operatorname{Tr}\Big[2Q_{jh}Q_b - Q_jQ_hQ_b - Q_hQ_jQ_b\Big] \qquad \text{(A.1.7)}$$

$$\sum_{h\in E}\Gamma^h_{jk}\Gamma^s_{ih} = \frac{1}{4}\sum_{a,b,h\in E} g_E^{ah}g_E^{bs}\operatorname{Tr}\Big[2Q_{jk}Q_a - Q_jQ_kQ_a - Q_kQ_jQ_a\Big]\operatorname{Tr}\Big[2Q_{ih}Q_b - Q_iQ_hQ_b - Q_hQ_iQ_b\Big] \qquad \text{(A.1.8)}$$

Recall that the curvature tensor is

$$\begin{aligned} R_{ijkl} &= \sum_{s\in E}\left(\sum_{h\in E}\Gamma^h_{jk}\Gamma^s_{ih} - \sum_{h\in E}\Gamma^h_{ik}\Gamma^s_{jh} + \frac{\partial\Gamma^s_{jk}}{\partial\lambda_i} - \frac{\partial\Gamma^s_{ik}}{\partial\lambda_j}\right)g^E_{sl} \\ &= \sum_{s,h\in E}\Gamma^h_{jk}\Gamma^s_{ih}g^E_{sl} - \sum_{s,h\in E}\Gamma^h_{ik}\Gamma^s_{jh}g^E_{sl} + \sum_{s\in E}\left(\frac{\partial\Gamma^s_{jk}}{\partial\lambda_i} - \frac{\partial\Gamma^s_{ik}}{\partial\lambda_j}\right)g^E_{sl}. \end{aligned} \tag{A.1.9}$$

Using that $\sum_{s\in E} g_E^{bs} g^E_{sl} = \delta_{bl}$ (since g_E is the inverse of g^E), where δ_{bl} is the Kronecker delta, we simplify Equation (A.1.7) and Equation (A.1.8) by

$$\sum_{s,h\in E}\Gamma^h_{ik}\Gamma^s_{jh}g^E_{sl} = \frac{1}{4}\sum_{a,h\in E} g_E^{ah}\operatorname{Tr}\Big[2Q_{ik}Q_a - Q_iQ_kQ_a - Q_kQ_iQ_a\Big]\operatorname{Tr}\Big[2Q_{jh}Q_l - Q_jQ_hQ_l - Q_hQ_jQ_l\Big],$$

$$\sum_{s,h\in E}\Gamma^h_{jk}\Gamma^s_{ih}g^E_{sl} = \frac{1}{4}\sum_{a,h\in E} g_E^{ah}\operatorname{Tr}\Big[2Q_{jk}Q_a - Q_jQ_kQ_a - Q_kQ_jQ_a\Big]\operatorname{Tr}\Big[2Q_{ih}Q_l - Q_iQ_hQ_l - Q_hQ_iQ_l\Big],$$

and using Equation (A.1.4), where we plug in Equation (A.1.5) and Equation (A.1.6) to obtain

$$\begin{aligned} &2\sum_{s\in E}\left(\frac{\partial\Gamma^s_{ik}}{\partial\lambda_j} - \frac{\partial\Gamma^s_{jk}}{\partial\lambda_i}\right)g^E_{sl} \\ = \quad &\operatorname{Tr}\Big[2Q_{ik}Q_{jl} - 2Q_{jk}Q_{il} - Q_jQ_{ik}Q_l + Q_iQ_{jk}Q_l - Q_{ik}Q_jQ_l \\ &\quad + Q_{jk}Q_iQ_l - Q_iQ_kQ_{jl} - Q_kQ_iQ_{jl} + Q_jQ_kQ_{il} + Q_kQ_jQ_{il}\Big] \\ &+ \sum_{a,h\in E} g_E^{ha}\operatorname{Tr}\Big[2Q_{jk}Q_h - Q_jQ_kQ_h - Q_kQ_jQ_h\Big]\operatorname{Tr}\Big[Q_aQ_{li} + Q_lQ_{ai} - Q_aQ_lQ_i - Q_lQ_aQ_i\Big] \\ &- \sum_{a,h\in E} g_E^{ha}\operatorname{Tr}\Big[2Q_{ik}Q_h - Q_iQ_kQ_h - Q_kQ_iQ_h\Big]\operatorname{Tr}\Big[Q_aQ_{lj} + Q_lQ_{aj} - Q_aQ_lQ_j - Q_lQ_aQ_j\Big]. \end{aligned}$$

Putting everything together, we obtain for Equation (A.1.9)

$$
\begin{aligned}
R_{ijkl} &= \sum_{s,h\in E} \Gamma^h_{jk}\Gamma^s_{ih} g^E_{sl} - \sum_{s,h\in E} \Gamma^h_{ik}\Gamma^s_{jh} g^E_{sl} + \sum_{s\in E}\left(\frac{\partial \Gamma^s_{jk}}{\partial \lambda_i} - \frac{\partial \Gamma^s_{ik}}{\partial \lambda_j}\right) g^E_{sl} \\
&= \frac{1}{4}\sum_{a,h\in E} g_E^{ah}\,\mathrm{Tr}\left[2Q_{jk}Q_a - Q_jQ_kQ_a - Q_kQ_jQ_a\right]\mathrm{Tr}\left[2Q_{ih}Q_l - Q_iQ_hQ_l - Q_hQ_iQ_l\right] \\
&- \frac{1}{4}\sum_{a,h\in E} g_E^{ah}\,\mathrm{Tr}\left[2Q_{ik}Q_a - Q_iQ_kQ_a - Q_kQ_iQ_a\right]\mathrm{Tr}\left[2Q_{jh}Q_l - Q_jQ_hQ_l - Q_hQ_jQ_l\right] \\
&+ \frac{1}{2}\sum_{a,h\in E} g_E^{ah}\,\mathrm{Tr}\left[2Q_{ik}Q_a - Q_iQ_kQ_a - Q_kQ_iQ_a\right]\mathrm{Tr}\left[Q_hQ_{lj} + Q_lQ_{hj} - Q_hQ_lQ_j - Q_lQ_hQ_j\right] \\
&- \frac{1}{2}\sum_{a,h\in E} g_E^{ah}\,\mathrm{Tr}\left[2Q_{jk}Q_a - Q_jQ_kQ_a - Q_kQ_jQ_a\right]\mathrm{Tr}\left[Q_hQ_{li} + Q_lQ_{hi} - Q_hQ_lQ_i - Q_lQ_hQ_i\right]. \\
&- \frac{1}{2}\mathrm{Tr}\Big[2Q_{ik}Q_{jl} - 2Q_{jk}Q_{il} - Q_jQ_{ik}Q_l + Q_iQ_{jk}Q_l - Q_{ik}Q_jQ_l \\
&\qquad + Q_{jk}Q_iQ_l - Q_iQ_kQ_{jl} - Q_kQ_iQ_{jl} + Q_jQ_kQ_{il} + Q_kQ_jQ_{il}\Big],
\end{aligned} \tag{A.1.10}
$$

and after merging the blue terms and the red terms, respectively, we have

$$
\begin{aligned}
R_{ijkl} = &\ \frac{1}{4}\sum_{a,h\in E} g_E^{ah}\,\mathrm{Tr}\left[2Q_{ik}Q_a - Q_iQ_kQ_a - Q_kQ_iQ_a\right]\mathrm{Tr}\left[2Q_hQ_{lj} - Q_hQ_lQ_j - Q_lQ_hQ_j\right] \\
&- \frac{1}{4}\sum_{a,h\in E} g_E^{ah}\,\mathrm{Tr}\left[2Q_{jk}Q_a - Q_jQ_kQ_a - Q_kQ_jQ_a\right]\mathrm{Tr}\left[2Q_hQ_{li} - Q_hQ_lQ_i - Q_lQ_hQ_i\right]. \\
&- \frac{1}{2}\,\mathrm{Tr}\Big[2Q_{ik}Q_{jl} - 2Q_{jk}Q_{il} - Q_jQ_{ik}Q_l + Q_iQ_{jk}Q_l - Q_{ik}Q_jQ_l \\
&\qquad + Q_{jk}Q_iQ_l - Q_iQ_kQ_{jl} - Q_kQ_iQ_{jl} + Q_jQ_kQ_{il} + Q_kQ_jQ_{il}\Big],
\end{aligned} \tag{A.1.11}
$$

Sectional Curvatures

The sectional curvature symbols are R_{ijji} (cf. Definition 2.4.2). Thus, from Equation (A.1.11), we have, where $Q_{ii} = Q_{jj} = 0$,

$$\begin{aligned} R_{ijji} = \quad & \frac{1}{4} \sum_{a,h \in E} g_E^{ah} \operatorname{Tr}\left[Q_a(2Q_{ij} - Q_iQ_j - Q_jQ_i)\right] \operatorname{Tr}\left[Q_h(2Q_{ij} - Q_iQ_j - Q_jQ_i)\right] \\ - \quad & \sum_{a,h \in E} g_E^{ah} \operatorname{Tr}\left[Q_aQ_j^2\right] \operatorname{Tr}\left[Q_hQ_i^2\right]. \\ - \quad & \operatorname{Tr}\left[Q_{ij}(Q_{ij} - Q_iQ_j - Q_jQ_i)\right], \end{aligned} \tag{A.1.12}$$

so in particular $R_{iiii} = 0$.

Bibliography

Abraham, R., & Marsden, J. (2008). *Foundations of Mechanics: Second Edition*, vol. 364 of *AMS Chelsea Publishing*. American Mathematical Society.
URL http://www.ams.org/chel/364

Anaya, M., Anipchenko-Ulaj, O., Ashfaq, A., Chiu, J., Kaiser, M., Ohsawa, M. S., Owen, M., Pavlechko, E., St. John, K., Suleria, S., Thompson, K., & Yap, C. (2020). Properties for the Fréchet mean in Billera-Holmes-Vogtmann treespace. *Advances in Applied Mathematics*, *120*, 102072.
URL https://www.sciencedirect.com/science/article/pii/S0196885820300750

Arsigny, V., Fillard, P., Pennec, X., & Ayache, N. (2005). Fast and Simple Calculus on Tensors in the Log-Euclidean Framework. In J. S. Duncan, & G. Gerig (Eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, Lecture Notes in Computer Science, (pp. 115–122). Berlin, Heidelberg: Springer.

Arsigny, V., Fillard, P., Pennec, X., & Ayache, N. (2006a). Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices. *SIAM J. Matrix Analysis Applications*, *29*, 328–347.

Arsigny, V., Fillard, P., Pennec, X., & Ayache, N. (2006b). Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, *56*(2), 411–421.

Barden, D., Le, H., & Owen, M. (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electronic Journal of Probability*, *18*, 1–25.
URL https://projecteuclid.org/journals/electronic-journal-of-probability/volume-18/issue-none/Central-limit-theorems-for-Fr%c3%a9chet-means-in-the-space-of/10.1214/EJP.v18-2201.full

Barden, D., Le, H., & Owen, M. (2016). Limiting behaviour of Fréchet means in the space of phylogenetic trees. *Annals of the Institute of Statistical Mathematics*, *70*(1).
URL https://nottingham-repository.worktribe.com/output/908832

Bhatia, R. (1997). Operator Monotone and Operator Convex Functions. In R. Bhatia (Ed.) *Matrix Analysis*, Graduate Texts in Mathematics, (pp. 112–151). New York, NY: Springer. URL `https://doi.org/10.1007/978-1-4612-0653-8_5`

Bhatia, R., Jain, T., & Lim, Y. (2017). On the Bures-Wasserstein distance between positive definite matrices. *arXiv:1712.01504 [math]*.
URL `http://arxiv.org/abs/1712.01504`

Billera, L. J., Holmes, S. P., & Vogtmann, K. (2001). Geometry of the Space of Phylogenetic Trees. *Advances in Applied Mathematics*, *27*(4), 733–767.
URL `http://www.sciencedirect.com/science/article/pii/S0196885801907596`

Bridson, M. R., & Haefliger, A. (1999). *Metric Spaces of Non-Positive Curvature*, vol. 319 of *Grundlehren der mathematischen Wissenschaften*. Berlin, Heidelberg: Springer Berlin Heidelberg.
URL `http://link.springer.com/10.1007/978-3-662-12494-9`

Buneman, P. (1971). The Recovery of Trees from Measures of Dissimilarity. *Mathematics the the Archeological and Historical Sciences: Proceedings of the Anglo-Romanian Conference, Mamaia, 1970*, (pp. 387–395).
URL `https://www.research.ed.ac.uk/portal/en/publications/the-recovery-of-trees-from-measures-of-dissimilarity(d8fe99ba-8e78-4026-b11c-40191daac38a)/export.html`

Burago, D., Burago, Y., & Ivanov, S. (2001). *A Course in Metric Geometry*, vol. 33 of *Graduate Studies in Mathematics*. Providence, Rhode Island: American Mathematical Society.
URL `http://www.ams.org/gsm/033`

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, Or The Preservation of Favoured Races in the Struggle for Life*. John Murray, Albemarle Street.

David, P. (2019). *A Riemannian Quotient Structure for Correlation Matrices with Applications to Data Science*. Ph.D., Ann Arbor, United States.
URL `https://search.proquest.com/docview/2234746919/abstract/99A16C5EA3F46C1PQ/1`

David, P., & Gu, W. (2019). A Riemannian structure for correlation matrices. *Operators and Matrices*, (3), 607–627.
URL `http://oam.ele-math.com/13-46`

Do Carmo, M. (1992). *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser Basel.
URL https://www.springer.com/de/book/9780817634902

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, *17*(6), 368–376.
URL https://doi.org/10.1007/BF01734359

Felsenstein, J. (2003). *Inferring Phylogenies*. Sunderland, Mass: OUP USA, 2003rd edition ed.

Garba, M. K., Nye, T. M. W., & Boys, R. J. (2018). Probabilistic Distances Between Trees. *Systematic Biology*, *67*(2), 320–327.
URL https://academic.oup.com/sysbio/article/67/2/320/4344841

Garba, M. K., Nye, T. M. W., Lueg, J., & Huckemann, S. F. (2021a). Information geometry for phylogenetic trees. *Journal of Mathematical Biology*, *82*(3), 19.
URL https://doi.org/10.1007/s00285-021-01553-x

Garba, M. K., Nye, T. M. W., Lueg, J., & Huckemann, S. F. (2021b). Information Metrics for Phylogenetic Trees via Distributions of Discrete and Continuous Characters. In F. Nielsen, & F. Barbaresco (Eds.) *Geometric Science of Information*, Lecture Notes in Computer Science, (pp. 701–709). Cham: Springer International Publishing.

Haeckel, E. (1866). *Generelle Morphologie der Organismen*. Berlin: Reimer, 1. auflage ed.

Hall, B. (2015). The Matrix Exponential. In B. C. Hall (Ed.) *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, Graduate Texts in Mathematics, (pp. 31–48). Cham: Springer International Publishing.
URL https://doi.org/10.1007/978-3-319-13467-3_2

Hoffman, K., & Kunze, R. A. (1971). *Linear Algebra 2Nd Ed.*. Prentice-Hall Of India Pvt. Limited.

Hu, T., & Kirk, W. A. (1978). Local contractions in metric spaces. *Proceedings of the American Mathematical Society*, *68*(1), 121–124.
URL https://www.ams.org/proc/1978-068-01/S0002-9939-1978-0464180-2/

Huckemann, S., & Eltzner, B. (2020). Statistical Methods Generalizing Principal Component Analysis to Non-Euclidean Spaces. In P. Grohs, M. Holler, & A. Weinmann (Eds.) *Handbook of Variational Methods for Nonlinear Geometric Data*, (pp. 317–338). Cham: Springer

International Publishing.
URL https://doi.org/10.1007/978-3-030-31351-7_10

Huckemann, S., Hotz, T., & Munk, A. (2010). Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric lie group actions. *Statistica Sinica*, *20*, 1–100.
URL http://goedoc.uni-goettingen.de:8081/handle/1/7238

Kim, J. (2000). Slicing Hyperdimensional Oranges: The Geometry of Phylogenetic Estimation. *Molecular Phylogenetics and Evolution*, *17*(1), 58–75.
URL http://www.sciencedirect.com/science/article/pii/S1055790300908169

Kraft, D., & Munchen, I. (1994). Algorithm 733: TOMP - Fortran modules for optimal control calculations. *ACM Trans. Math. Soft*, (pp. 262–281).

Lang, S. (1999). *Fundamentals of Differential Geometry*. Graduate Texts in Mathematics. New York: Springer-Verlag.
URL https://www.springer.com/de/book/9780387985930

Lee, J. (2018). *Introduction to Riemannian Manifolds*. Graduate Texts in Mathematics. Springer International Publishing, 2 ed.
URL https://www.springer.com/de/book/9783319917542

Lee, J. M. (1997). *Riemannian Manifolds: An Introduction to Curvature*. Graduate Texts in Mathematics. New York: Springer-Verlag.
URL https://www.springer.com/gp/book/9780387982717

Lee, J. M. (2003). *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. New York: Springer-Verlag.
URL https://www.springer.com/de/book/9780387217529

Lin, B., Monod, A., & Yoshida, R. (2018). Tropical Foundations for Probability & Statistics on Phylogenetic Tree Space. *arXiv:1805.12400 [math, q-bio, stat]*.
URL http://arxiv.org/abs/1805.12400

Lin, B., Sturmfels, B., Tang, X., & Yoshida, R. (2017). Convexity in Tree Spaces. *SIAM Journal on Discrete Mathematics*, *31*(3), 2015–2038.
URL http://arxiv.org/abs/1510.08797

Lueg, J., Garba, M. K., Nye, T. M. W., & Huckemann, S. F. (2021). Wald Space for Phylogenetic Trees. In F. Nielsen, & F. Barbaresco (Eds.) *Geometric Science of Information*, Lecture Notes in Computer Science, (pp. 710–717). Cham: Springer International Publishing.

Lueg, J., Garba, M. K., Nye, T. M. W., & Huckemann, S. F. (2022). Foundations of the Wald Space for Phylogenetic Trees. ArXiv:2209.05332 [math, stat].
URL `http://arxiv.org/abs/2209.05332`

Miolane, N., Guigui, N., Brigant, A. L., Mathe, J., Hou, B., Thanwerdas, Y., Heyder, S., Peltre, O., Koep, N., Zaatiti, H., Hajri, H., Cabanes, Y., Gerald, T., Chauchat, P., Shewmake, C., Brooks, D., Kainz, B., Donnat, C., Holmes, S., & Pennec, X. (2020). Geomstats: A python package for riemannian geometry in machine learning. *Journal of Machine Learning Research*, *21*(223), 1–9.
URL `http://jmlr.org/papers/v21/19-027.html`

Moakher, M. (2005). A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices. *SIAM J. Matrix Analysis Applications*, *26*, 735–747.

Moakher, M., & Zerai, M. (2011). The Riemannian Geometry of the Space of Positive-Definite Matrices and Its Application to the Regularization of Positive-Definite Matrix-Valued Data. *Journal of Mathematical Imaging and Vision*, *40*, 171–187.

Moulton, V., & Steel, M. (2004). Peeling phylogenetic 'oranges'. *Advances in Applied Mathematics*, *33*(4), 710–727.
URL `http://www.sciencedirect.com/science/article/pii/S0196885804000430`

Nye, T. (2014). An Algorithm for Constructing Principal Geodesics in Phylogenetic Treespace. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, *11*, 304–315.

Nye, T., Tang, X., Weyenberg, G., & Yoshida, R. (2016). Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees. *Biometrika*, *104*(4), 901–922.
URL `http://dx.doi.org/10.1093/biomet/asx047`

Nye, T. M. W. (2011). PRINCIPAL COMPONENTS ANALYSIS IN THE SPACE OF PHYLOGENETIC TREES. *The Annals of Statistics*, *39*(5), 2716–2739. Publisher: Institute of Mathematical Statistics.
URL `https://www.jstor.org/stable/41713594`

O'Neill, B. (1966). The fundamental equations of a submersion. *Michigan Mathematical Journal*, *13*(4), 459–469.
URL `https://projecteuclid.org/journals/michigan-mathematical-journal/volume-13/issue-4/The-fundamental-equations-of-a-submersion/10.1307/mmj/1028999604.full`

O'Neill, B. (1983). *Semi-Riemannian Geometry With Applications to Relativity*. Academic Press.

Owen, M., & Provan, J. S. (2011). A Fast Algorithm for Computing Geodesic Distances in Tree Space. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *8*(1), 2–13.

Pennec, X. (2018). Parallel Transport with Pole Ladder: a Third Order Scheme in Affine Connection Spaces which is Exact in Affine Symmetric Spaces. *arXiv:1805.11436 [cs, math]*.
URL `http://arxiv.org/abs/1805.11436`

Pennec, X., Fillard, P., & Ayache, N. (2006). A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, *66*(1), 41–66.
URL `https://hal.inria.fr/inria-00614990`

Schmidt, F., Clausen, M., & Cremers, D. (2006). Shape Matching by Variational Computation of Geodesics on a Manifold. vol. 4174, (pp. 142–151).

Schwartzman, A. (2006). *Random Ellipsoids and False Discovery Rates: Statistics for Diffusion Tensor Imaging Data*. Ph.D. thesis.

Semple, C., & Steel, M. (2003). *Phylogenetics*. Oxford ; New York: Oxford University Press, new edition ed.

Shiers, N., Zwiernik, P., Aston, J. A. D., & Smith, J. Q. (2016). The correlation space of Gaussian latent tree models and model selection without fitting. *arXiv:1508.00436 [stat]*.
URL `http://arxiv.org/abs/1508.00436`

Thanwerdas, Y., & Pennec, X. (2021a). Geodesic of the Quotient-Affine Metrics on Full-Rank Correlation Matrices. *arXiv:2103.04621 [math]*.
URL `http://arxiv.org/abs/2103.04621`

Thanwerdas, Y., & Pennec, X. (2021b). O(n)-invariant Riemannian metrics on SPD matrices.
URL `https://hal.archives-ouvertes.fr/hal-03338601`

Theobald, D. L. (2010). A formal test of the theory of universal common ancestry. *Nature*, *465*(7295), 219–222.
URL `https://www.nature.com/articles/nature09014`

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., & Martin, W. F. (2016). The physiology and habitat of the last universal common ancestor. *Nature Microbiology*, *1*(9), 1–8.
URL `https://www.nature.com/articles/nmicrobiol2016116`

Wilson, R. J. (1996). *Introduction to Graph Theory*. TBS.

Yair, O., Ben-Chen, M., & Talmon, R. (2019). Parallel Transport on the Cone Manifold of SPD Matrices for Domain Adaptation. *IEEE Transactions on Signal Processing*, *67*(7), 1797–1811. ArXiv: 1807.10479.
URL `http://arxiv.org/abs/1807.10479`

Yoshida, R., Zhang, L., & Zhang, X. (2019). Tropical Principal Component Analysis and Its Application to Phylogenetics. *Bulletin of Mathematical Biology*, *81*(2), 568–597.
URL `https://doi.org/10.1007/s11538-018-0493-4`