Check for updates

Received: 13 September 2022 | Revised: 29 February 2024 | Accepted: 1 March 2024

DOI: 10.1112/jlms.12893

RESEARCH ARTICLE

Journal of the London Mathematical Society

# Foundations of the wald space for phylogenetic trees

**Jonas Lueg**[1] | **Maryam K. Garba**[2] | **Tom M. W. Nye**[3] | **Stephan F. Huckemann**[1]

[1]Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, Georg-August-Universität, Göttingen, Germany

[2]Department of Mathematics, Physics and Electrical Engineering, Northumbria University, Newcastle upon Tyne, UK

[3]School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, UK

**Correspondence**
Tom M. W. Nye, School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, UK.
Email: tom.nye@ncl.ac.uk

**Funding information**
DFG, Grant/Award Numbers: RTG 2088, HU 1575-7; Volkswagen Foundation

## Abstract

Evolutionary relationships between species are represented by phylogenetic trees, but these relationships are subject to uncertainty due to the random nature of evolution. A geometry for the space of phylogenetic trees is necessary in order to properly quantify this uncertainty during the statistical analysis of collections of possible evolutionary trees inferred from biological data. Recently, the wald space has been introduced: a length space for trees which is a certain subset of the manifold of symmetric positive definite matrices. In this work, the wald space is introduced formally and its topology and structure is studied in detail. In particular, we show that wald space has the topology of a disjoint union of open cubes, it is contractible, and by careful characterisation of cube boundaries, we demonstrate that wald space is a Whitney stratified space of type (A). Imposing the metric induced by the affine invariant metric on symmetric positive definite matrices, we prove that wald space is a geodesic Riemann stratified space. A new numerical method is proposed and investigated for construction of geodesics, computation of Fréchet means and calculation of curvature in wald space. This work

© 2024 The Authors. *Journal of the London Mathematical Society* is copyright © London Mathematical Society. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

is intended to serve as a mathematical foundation for further geometric and statistical research on this space.

**MSC 2020**
30L05, 57N80, 53A35 (primary)

# 1 | INTRODUCTION

## 1.1 | Background

Over billions of years, evolution has been driven by unobserved random processes. Inferences about evolutionary history, which by necessity are largely based on observations of present-day species, are therefore always subject to some level of uncertainty. Phylogenetic trees are used to represent possible evolutionary histories relating a set of species, or taxa, which form the leaves of each tree. Internal vertices on phylogenetic trees usually represent speciation events, and edge lengths represent the degree of evolutionary divergence over any given edge. Trees are typically inferred from genetic sequence data from extant species, and a variety of well-established statistical methods exist for phylogenetic inference [14]. These generally output a sample of trees (a collection of possible evolutionary histories compatible with the data). Moreover, evolutionary relationships can vary stochastically from one gene to another, giving a further source of random variation in samples of trees [27]. It is then natural to pose statistical questions about such samples: for example identifying a sample mean, identifying principal modes of variation in the sample, or testing differences between samples. This, in turn, calls for the design of suitable metric spaces in which each element is a phylogenetic tree on some fixed set of taxa, and which are ideally both biologically substantive and computationally tractable.

The design of these *tree spaces* is aggravated by the continuous and combinatorial nature of phylogenetic trees and furthermore, a metric space that is also a geodesic space (so that distance corresponds to the length of shortest paths, also called geodesics) is to be preferred, as it facilitates computation of statistics like the Fréchet mean significantly. The first geodesic space of phylogenetic trees was introduced by [7] and is called the BHV space, where BHV is an acronym of the authors Billera, Holmes and Vogtmann. For a fixed set of species $L = \{1, \dots, N\}$, also called taxa or labels, with $3 \leqslant N \in \mathbb{N}$, BHV space is constructed via embedding all phylogenetic trees into a Euclidean space $\mathbb{R}^M$, where $M \in \mathbb{N}$ is exponentially growing in $N$, and then taking the infinitesimally induced intrinsic distance on this embedded subset, giving a metric space. As a result, BHV space features a very rich and computationally tractable geometry as it is a CAT(0) space, that is, globally of non-positive curvature, and thus has unique geodesics and Fréchet means. Following the development of a polynomial time algorithm for computing geodesics that overcame the combinatorial difficulties [35], various algorithms have been derived for computing statistics like sample means [6, 29] and variance [9], confidence regions for the population mean [43] and principal component analysis [32–34], Feragen et al. 2013). The BHV paper has had considerable influence more widely on research in phylogenetics (see [42], for example), non-Euclidean statistics [28], algebraic geometry [1], probability theory [13] and other areas of mathematics [2]. In addition to the BHV tree space, a variety of alternative tree spaces have been proposed, both for discrete and continuous underlying

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

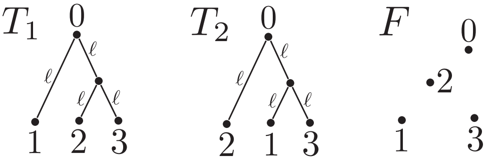


**FIGURE 1** Two trees $T_1$ and $T_2$ with positive edge length $\ell \in (0, \infty)$. Letting $\ell \to \infty$, the intuitive limit element for both trees is the forest $F$, as species are considered not related if their evolutionary distance approaches infinity. In wald space, the distance between $T_1$ and $T_2$ goes to zero accordingly as $\ell \to \infty$ and their limit is the forest $F$ that is also contained in wald space. In BHV space, however, their distance goes to infinity as $\ell \to \infty$ and $F$ is not an element of the space.

point sets of trees. For example, in the *tropical tree space* [30, 40] edge weights are times, not evolutionary divergences, thus allowing for a distance metric between two trees involving tropical algebra.

The geometries of the BHV and tropical tree spaces are unrelated to the methods used to infer phylogenies from sequence data. In contrast, there are substantially different tree spaces that originate via the evolutionary genetic substitution models used by molecular phylogenetic methods for tree inference (see [44] for details of these). Evolutionary substitution models are essentially Markov processes on a phylogenetic tree with state space $\Omega$. For DNA sequence data, the state space is $\Omega = \{\mathrm{A}, \mathrm{C}, \mathrm{T}, \mathrm{G}\}$. Under an appropriate set of assumptions on the substitution model, each tree determines a probability distribution on the set of possible letter patterns at the labelled vertices $L$, (i.e., a probability mass function $p : \Omega^N \to [0, 1]$), $N = |L|$, and this can be used to compute the likelihood of any tree. At about the same that BHV space was introduced, [23] provided a geometrical interpretation of tree estimation methods, where, given the substitution model, an embedding of phylogenetic trees into an $|\Omega|^N$-dimensional simplex using the likelihoods was discussed informally. The concept was then picked up by [31], introducing the topological space known as the *edge-product space*, taking not only into account phylogenetic trees but also forests, characterising each forest via a vector containing correlations between all pairs of labels in $L$ under the induced distribution $p$. This representation is then an embedding of all phylogenetic forests into a $N(N-1)/2$-dimensional space. Using the same characterisation of phylogenetic trees via distributions on $\Omega^N$ obtained from a fixed substitution model, [17] considered probabilistic distances to obtain metrics on tree space, but these metrics do not yield length spaces. Therefore, in [16], the fact that all phylogenetic trees with a fixed fully resolved tree topology are a manifold was used to apply the Fisher information geometry for statistical manifolds on each such piece of the space to eventually obtain a metric space that is a length space. Additionally, instead of using substitution models with finite state space $\Omega$, [16] considered a Gaussian model with state space $\Omega = \mathbb{R}$ in order to deal with the problem of computational tractability. The distributions characterising phylogenetic trees are then zero-mean multivariate Gaussians, and sums over $\Omega^N$ for discrete $\Omega$ are replaced with integrals over $\mathbb{R}^N$. The characterisation with this Gaussian model together with the choice of the Fisher information geometry and the extension to phylogenetic forests ultimately leads to the *wald space*, which is essentially an embedding of the phylogenetic forests into the real symmetric $N \times N$-dimensional strictly positive definite matrices $\mathcal{P}$ [16]. The elements of wald space are called *wälder* ('Wald' is a German word meaning 'forest').

The geometry of wald space is fundamentally different from BHV space [16, 26], as illustrated in Figure 1, which also underlines the biological reasonability of the wald space. Loosely speaking, wald space can be viewed topologically as being obtained by compactifying the

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

boundaries at the 'infinities' of BHV space, which comes with the price of fundamentally changing the geometry that is not locally Euclidean anymore. We avoid, though, the compactification at the 'zeroes' of the edge-product space proposed by [31] which suggests itself by mathematical elegance. It is biologically questionable, however, as it would allow different taxa to agree with one another. In [16], apart from defining the wald space, certain properties of the space were established, such as showing the distance between any two points to be finite, and algorithms for approximating geodesics were proposed. In [26], a compact definition of wald space as well as more refined algorithms for approximating geodesics were introduced.

## 1.2 | Contribution of this paper

Previous work on wald space established the space as a length space, and this paper was originally motivated by the aim of proving the existence of a minimising geodesic between every two points, that is, establishing wald space as a geodesic metric space, as the existence of geodesics is crucial for performing statistical analysis within the space. This aim is achieved in Theorem 4.2.1. The proof involves three essential characterisations of the elements of wald space (as graph-theoretic forests; as split systems; and as certain symmetric positive definite matrices). In turn, these enable a rigourous analysis of the topology of wald space, such as Theorem 3.3.5 about its stratified structure, in addition to providing a foundation for further research on this space.

The remainder of the paper is structured as follows. In Section 2, we define the wald space $\mathcal{W}$ for a fixed set of labels $\{1, \dots, N\}$ as equivalence classes of partially labelled graph-theoretic forests. The topology on $\mathcal{W}$ is obtained by defining a map $\psi$ from $\mathcal{W}$ into the set of $N \times N$ symmetric positive definite matrices and requiring $\psi$ to be a homeomorphism onto its image. We then provide an equivalent, but more tractable, definition in terms of *splits* or bipartitions of labels, and an equivalent map $\phi$ from split-representations of wälder to symmetric positive definite matrices. In particular, we show that wald space can be identified topologically with a disjoint union of open unit cubes. Each open unit cube is called a *grove*. In Section 3, we describe the structure or stratification of the wald space by investigating on how the groves are glued together along their respective boundaries. This is achieved by first providing in Subsection 3.1 a detailed characterisation of the matrices in the image $\phi(\mathcal{W})$ in terms of a set of algebraic constraints on the matrix elements. Using this characterisation, for example, we show that wald space is contractible. Then in Subsection 3.2 we use a partial ordering of forest topologies, first introduced by [31] to establish results about the boundaries of groves and the stratification of wald space. This culminates in Subsection 3.3 in which we prove wald space satisfies certain axioms at grove boundaries, collectively known as *Whitney condition (A)* [36], which ensure that tangent spaces behave well as the boundaries of strata are approached. We then go on to consider the induced affine invariant or information geometry on wald space in Section 4. We show the topology induced by the metric is the same as the previous topology defined using $\phi$, and hence show that $\mathcal{W}$ is a geodesic metric space (i.e., every two points are connected by a minimising geodesic). Finally in Section 5, we use a new algorithm for computing approximate geodesics to explore the geometry on wald space, specifically computing sectional curvatures within groves and Alexandrov curvatures for fundamental examples. We also investigate the behaviour of the sample Fréchet mean, in particular with reference to the issue of *stickiness* observed in in BHV space (see, e.g., [19, 22] for a description). In Section 6, we discuss the contributions of the

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

paper and some of the many open questions and unsolved problems about the geometry of wald space.

## 1.3 | Notation

Throughout the paper, we use the following notation and concepts, where points 4–6 below can be found in standard textbooks of differential geometry, for example, [24, chapter XII].

(1) $2 \leqslant N \in \mathbb{N}$ is a fixed integer defining the *set of labels* $L = \{1, \dots, N\}$.
(2) $\sqcup_{i=1}^{n} A_i$ denotes the union if the $A_i$ are pairwise disjoint ($i = 1, \dots, n$).
(3) When we speak of *partitions*, no empty sets are allowed.
(4) For a set $E$, its cardinality is denoted by $|E|$.
(5) $\mathcal{S}$ is the Euclidean space of real symmetric $N \times N$ matrices.
(6) $\mathcal{P}$ is the space of real symmetric and positive definite $N \times N$ matrices. It is an open cone in $\mathcal{S}$ and carries the topology and smooth manifold structure inherited from $\mathcal{S}$. In particular, every tangent space $T_P\mathcal{P}$ at $P \in \mathcal{P}$ is isomorphic to $\mathcal{S}$.
(7) We equip $\mathcal{P}$ with the *affine invariant Riemannian metric*, also called information geometry, yielding a Cartan–Hadamard manifold. Its metric tensor is given by

$$\langle X, Y \rangle_P = \operatorname{trace}(P^{-1} X P^{-1} Y)$$

for $X, Y \in \mathcal{S} \cong T_P\mathcal{P}$ and the unique geodesic $\gamma$ through $P = \gamma(0), Q = \gamma(1) \in \mathcal{P}$ is given by

$$(-\infty, \infty) \to \mathcal{P},\ t \mapsto \gamma(t) = \sqrt{P} \exp\left(t \log\left(\sqrt{P}^{-1} Q \sqrt{P}^{-1}\right)\right)\sqrt{P}$$

with the usual matrix exponential and logarithm, respectively. Here, $\sqrt{P}$ denotes the unique positive definite root of $P$.
(8) The Riemannian metric induces a metric on $\mathcal{P}$ denoted by $d_{\mathcal{P}}$ and for a rectifiable curve $\gamma : [a, b] \to \mathcal{P}$ let $L_{\mathcal{P}}(\gamma)$ be its length.

In a word of caution, we note that the term *topology* appears in two contexts: (i) as a system of open sets defining a topological space and (ii) as a branching structure of a graph-theoretic forest. The latter is standard in the phylogenetic literature, despite the potential for confusion.

# 2 | DEFINITION OF WALD SPACE VIA GRAPHS AND SPLITS

## 2.1 | From a Graph viewpoint

This section recalls definitions and results from [16] and [26].

**Definition 2.1.1.** A *forest* is a triple $(\mathfrak{V}, \mathfrak{E}, \ell)$, where

(PF1) $(\mathfrak{V}, \mathfrak{E})$ is a graph-theoretical undirected forest with *vertex* set $\mathfrak{V}$ such that $L \subseteq \mathfrak{V}$ and that $v \in \mathfrak{V} \setminus L$ implies $\deg(v) \geqslant 3$, where $\deg(v)$ is the degree of a vertex $v$, and *edge* set $\mathfrak{E} \subseteq \big\{\{u, v\} : u, v \in \mathfrak{V},\ u \neq v\big\}$, (PF2) and $\ell = (\ell_e)_{e\in\mathfrak{E}} \in (0,\infty)^{\mathfrak{E}}$.

**Definition 2.1.2.** Two forests $(\mathfrak{V},\mathfrak{E},\ell),(\mathfrak{V}',\mathfrak{E}',\ell')$ are *topologically* equivalent, if there is a bijection $f : \mathfrak{V} \to \mathfrak{V}'$ such that

(i) $f(u) = u$ for all $u \in L$,
(ii) $\{u,v\} \in \mathfrak{E} \Leftrightarrow \{f(u),f(v)\} \in \mathfrak{E}'$.

They are *phylogenetically* equivalent if additionally

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

Moreover,

(1) Every phylogenetic equivalence class is called a *phylogenetic forest* and denoted by $\mathfrak{F} = [\mathfrak{V},\mathfrak{E},\ell]$.
(2) $\mathcal{W}$ is the set of all phylogenetic forests.
(3) Every topological equivalence class is called a *forest topology* and denoted by $[\mathfrak{F}] = [\mathfrak{V},\mathfrak{E}]$.

**Definition 2.1.3.** Let $(\mathfrak{V},\mathfrak{E},\ell)$ be a forest. For two leaves $u,v \in L$ let $\mathfrak{E}(u,v)$ be the set of edges in $\mathfrak{E}$ of the unique path between $u$ and $v$, if $u$ and $v$ are connected, else set $\mathfrak{E}(u,v) = \emptyset$. Further define a mapping of forests via

$$\psi : (\mathfrak{V},\mathfrak{E},\ell) \mapsto (\rho_{uv})_{u,v=1}^{N} \in \mathcal{S}$$

where

$$\rho_{uv} = \begin{cases} \exp\left(-\sum_{e\in\mathfrak{E}(u,v)} \ell_e\right) & \text{if } u \neq v \text{ and } \mathfrak{E}(u,v) \neq \emptyset \\ 0 & \text{if } u \neq v \text{ and } \mathfrak{E}(u,v) = \emptyset \\ 1 & \text{if } u = v \end{cases} \tag{2.1}$$

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

By definition, the above matrix is the same for two forests representing the same phylogenetic forest. It is even positive definite and characterises phylogenetic forests uniquely as the following theorem shows.

**Theorem 2.1.4** ([16], Theorem 4.1). *For every forest* $(\mathfrak{V},\mathfrak{E},\ell)$*, we have*

$$\psi(\mathfrak{V},\mathfrak{E},\ell) \in \mathcal{P}$$

*and for any two forests* $(\mathfrak{V},\mathfrak{E},\ell)$ *and* $(\mathfrak{V}',\mathfrak{E}',\ell')$ *we have*

$$\psi(\mathfrak{V},\mathfrak{E},\ell) = \psi(\mathfrak{V}',\mathfrak{E}',\ell')$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

*if and only if*

$$[\mathfrak{V}, \mathfrak{E}, \ell] = [\mathfrak{V}', \mathfrak{E}', \ell'] .$$

In consequence of Theorem 2.1.4, $\psi$ induces a well-defined injection from $\mathcal{W}$ into $\mathcal{P}$. In slight abuse of notation we denote this mapping also by $\psi$, that is

$$\psi : \mathcal{W} \to \mathcal{P}, \quad \mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell] \mapsto \psi(\mathfrak{F}) := \psi(\mathfrak{V}, \mathfrak{E}, \ell). \tag{2.2}$$

**Definition 2.1.5.** The *wald space* is the topological space $\mathcal{W}$ equipped with the unique topology under which the map $\psi : \mathcal{W} \to \mathcal{P}$ from Equation (2.2) is a homeomorphism onto its image.

## 2.2 | From a split viewpoint

If $(\mathfrak{V}, \mathfrak{E}, \ell)$ is a representative of a phylogenetic forest $\mathfrak{F}$, there is $K \in \mathbb{N}$ such that the graph-theoretic forest $(\mathfrak{V}, \mathfrak{E})$ decomposes into $K$ disjoint non-empty graph-theoretic trees

$$(\mathfrak{V}_1, \mathfrak{E}_1), \dots, (\mathfrak{V}_K, \mathfrak{E}_K) .$$

In particular, this decomposition induces a partition $L_1, \dots, L_K$ of the leaf set $L$ with $L_\alpha \subseteq \mathfrak{V}_\alpha$, $1 \leqslant \alpha \leqslant K$.

Furthermore for $1 \leqslant \alpha \leqslant K$, taking away an edge $e \in \mathfrak{E}_\alpha$ decomposes $(\mathfrak{V}_\alpha, \mathfrak{E}_\alpha)$ into two disjoint graph-theoretic trees that *split* the leaf set $L_\alpha$ into two disjoint subsets $A$ and $B$.

The representation of phylogenetic trees via splits is more abstract than as graphs but more tractable. We first introduce the weighted split representation and then show equivalence of the concepts.

**Definition 2.2.1.** A tuple $F = (E, \lambda)$ with $E \neq \emptyset$ is a *split-based phylogenetic forest* if

(i) there is $1 \leqslant K \leqslant N$ and a partition $L_1, \dots, L_K$ of the leaf set $L$;
(ii) every element $e \in E$ is of the form $e = \{A, B\}$, called a *split*, where for some $1 \leqslant \alpha \leqslant K$, $A, B$ is a partition of $L_\alpha$; $E_\alpha$ denotes the elements in $E$ that are splits of $L_\alpha$; for notational ease we write interchangeably

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

whenever $A = \{a_1, \dots, a_r\}$, $B = \{b_1, \dots, b_s\}$;
(iii) all splits in $E_\alpha$ $(1 \leqslant \alpha \leqslant K)$ are pairwise *compatible* with one another, where two splits $A|B$ and $C|D$ of $L_\alpha$ are *compatible* with one another if one of the sets below is empty:

$$A \cap C, \quad A \cap D, \quad B \cap C, \quad B \cap D ;$$

(iv) for all distinct $u, v \in L_\alpha$, $1 \leqslant \alpha \leqslant K$, there exists a split $e = A|B \in E_\alpha$ such that $u \in A$ and $v \in B$;
(v) $\lambda := (\lambda_e)_{e \in E} \in (0, 1)^E$.

Moreover, $F_\infty$ with $E = \emptyset$ and void array $\lambda$ is the completely disconnected *split-based phylogenetic forest* with leaf partion $\{1\}, \dots, \{N\}$.

The partition $L_1, \dots, L_K$ is not mentioned explicitly in the definition of a split-based phylogenetic forest $F = (E, \lambda)$ because it can be derived from $E$ via $\{L_1, \dots, L_{\tilde{K}}\} := \{A \cup B : A|B \in E\}$, where $\tilde{K} \leqslant K$, and for all $u \in L \setminus \bigcup_{\alpha=1}^{\tilde{K}} L_\alpha$, the singleton $\{u\}$ is added to the collection to obtain $L_1, \dots, L_K$.

**Theorem 2.2.2.** *There is a one-to-one correspondence between split-based phylogenetic forests* $F = (E, \lambda)$ *from Definition 2.2.1 and phylogenetic forests* $\mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell]$ *from Definition 2.1.2 with* $\ell$ *and* $\lambda$ *related by*

$$\lambda_s := 1 - \exp(-\ell_e) \tag{2.3}$$

*with an arbitrary but fixed representative* $(\mathfrak{V}, \mathfrak{E}, \ell)$. *Furthermore, there is a one-to-one correspondence between compatible split sets E from Definition 2.2.1(i)–(iv), and phylogenetic forest topologies* $[\mathfrak{V}, \mathfrak{E}]$.

*Proof.* Case I. Suppose $K = 1$, that is, $\mathfrak{F}$ comprises only one tree: We take recourse to [39, Theorem 3.1.4] who establish a one-to-one correspondence of compatible split sets $E$ from Definition 2.2.1(i)–(iv), and phylogenetic forest topologies $[\mathfrak{V}, \mathfrak{E}]$, in case these are taken from graph-theoretic trees. Indeed, our phylogenetic forest topologies correspond to *isomorphic X-trees* there (our $L$ is $X$ there and the *labelling map* from [39, Definition 2.1.1] is the identity in our case) and for every representative $(\mathfrak{V}, \mathfrak{E}) \in [\mathfrak{V}, \mathfrak{E}]$ there is a unique compatible split set $E$ from Definition 2.2.1(i)–(iv) ((iv) is a consequence of $L \subseteq \mathfrak{V}$, $K = 1$ and injectivity of the labelling map). Vice versa, there is a bijection $e \mapsto s_e, \mathfrak{E} \to E$ that, removing the edge $e$ from $\mathfrak{E}$ produces two disconnected trees, yields a unique split $s = s_e = A|B$ of the leaf set $L = A \cup B$. This yields the second assertion, namely a one-to-one correspondence between compatible split sets $E$ from Definition 2.2.1 (i)–(iv) and phylogenetic forest topologies $[\mathfrak{V}, \mathfrak{E}]$ in case of underlying graph-theoretic trees. The first assertion follows from the correspondence in (2.3), which thus yields, due to phylogenetic equivalence in Definition 2.1.2(iii), a one-to-one correspondence between split based phylogenetic forests $F = (E, \lambda)$ and phylogenetic forests $[\mathfrak{V}, \mathfrak{E}, \ell]$, in case of underlying graph-theoretic trees.

Case II. Suppose $\mathfrak{F}$ comprises several $K > 1$ trees: Here, consider two phylogenetic forests representatives $(\mathfrak{V}, \mathfrak{E}, \ell), (\mathfrak{V}', \mathfrak{E}', \ell') \in [\mathfrak{V}, \mathfrak{E}, \ell]$. Due to Definition 2.1.2(i) and (ii), both $(\mathfrak{V}, \mathfrak{E}, \ell)$ and $(\mathfrak{V}', \mathfrak{E}', \ell')$ have the same number of connected components, each of which is a graph-theoretic tree and the bijection $f$ from Definition 2.1.2 restricts to bijections between the corresponding graph-theoretic trees. For each of these, Case I ($K = 1$) is applicable, thus yielding the assertion in the general case. □

In consequence of Theorem 2.2.2, we introduce the following additional notation.

**Definition 2.2.3.** From now on, we identify split-based phylogenetic forests $F = (E, \lambda)$ with phylogenetic forests $\mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell]$ and say that $F$ is a *wald*, in plural *wälder*, so that $F \in \mathcal{W}$, and use interchangeably the name split and edges for the elements of $E$ (as they are 'edges' in equivalence classes). In particular, the $\lambda_e, e \in E$, from Definition 2.2.1, are called *edge weights*. Furthermore,

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(1) $[F] := E$ also denotes the *topology* $[\mathfrak{V}, \mathfrak{E}]$ of $F$ and

$$\mathcal{E} := \{E : \exists \lambda \in (0,1)^E \text{such that} (E, \lambda) \text{ is a split-based phylogenetic forest}\} \cup \{\emptyset\}$$

denotes the set of all possible topologies;

(2) wälder of the same topology $E$ form a *grove*

$$\mathcal{G}_E = \left\{F = (E', \lambda') \in \mathcal{W} : E = E'\right\},$$

(3) for any two $u, v \in L$ with leaf partition $L_1, \dots, L_K$, define

$$E(u, v) := \{A|B : \exists\, 1 \leqslant \alpha \leqslant K \text{ and } e \in \mathfrak{E}(u, v) \text{ that splits } L_\alpha \text{ into } A \text{ and } B\},$$

which also denotes set of edges between $u$ and $v$, that may be empty;

(4) the edge length based matrix representation $\psi$ from Equation (2.2) translates to the *edge weight* based matrix representation $\phi$ defined by

$$\phi : \mathcal{W} \to \mathcal{P}, \quad F = (E, \lambda) \mapsto (\rho_{uv})_{u,v=1}^N := \left(\prod_{e \in E(u,v)} (1 - \lambda_e)\right)_{u,v=1}^N, \tag{2.4}$$

with the agreement that in case of empty $E(u, v)$

$$\left.\begin{aligned} \rho_{uv} &:= 1 \text{ whenever } u = v \text{ and} \\ \rho_{uv} &:= 0 \text{ whenever } u \in L_\alpha \text{ and } v \in L_\beta, \alpha \neq \beta, \alpha, \beta \in \{1, \dots, K\} \end{aligned}\right\}; \tag{2.5}$$

here $\lambda$ is computed from $\ell$ as defined in Equation (2.3).

*Remark* 2.2.4. In light of Definition 2.1.5, the *wald space* is the topological space $\mathcal{W}$ equipped with the unique topology such that the map $\phi : \mathcal{W} \to \mathcal{P}$ is a homeomorphism onto its image. Thus, groves can be identified topologically with open unit cubes

$$\mathcal{G}_E \cong (0, 1)^E \tag{2.6}$$

and the *wald space* thus with the disjoint union

$$\mathcal{W} = \bigsqcup_{E \in \mathcal{E}} \mathcal{G}_E \cong \bigsqcup_{E \in \mathcal{E}} (0, 1)^E, \tag{2.7}$$

where we note that $|E|$ runs from 0 (corresponding to $F_\infty$) to $2N - 3$ (for fully resolved trees), as is easily seen upon induction on $N$.

Furthermore, observe that

(1) Equation (2.3) links strictly monotonous edge weights with edge lengths so that the limits $\lambda_e \to 0, 1$ correspond to the limits $\ell_e \to 0, +\infty$, respectively;

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

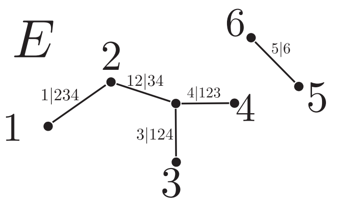


FIGURE 2 The topology $E$ as defined in Example 2.2.5 with the corresponding splits annotated to the edges.

(2) for any partition $A, B$ of $L_\alpha$ ($1 \leqslant \alpha \leqslant K$, as above), we have that

$$e = A|B \iff e \in E(u, v) \text{ for all } u \in A, v \in B\,, \tag{2.8}$$

where the implication to the right is a consequence of $E_\alpha$ being a tree topology and the reverse implication is a consequence of $A$ and $B$ being a partition of $L_\alpha$.

**Example 2.2.5.** Let $N = 6$. Consider

$$E = \{1|234, 3|124, 4|123, 12|34, 5|6\},$$

that is, the partition of labels is $L_1 = \{1, 2, 3, 4\}$, $L_2 = \{5, 6\}$, the corresponding graph is depicted in Figure 2. One can easily check that all edges that are splits of $L_1$ are compatible, likewise for all splits of $L_2$ (there is only one split, 5|6, in this case).

Moreover, observe that the unique path from 1 to 4 contains the edges $E(1, 4) = \{1|234, 4|123, 12|34\}$, that is, all splits separating 1 and 4.

Indeed, for every connected pair of leaves, there is a split separating this pair, for instance, for all $u, v \in L_1$ there is a split $e = A|B \in E$ such that $u \in A$ and $v \in B$. Removing the edge 1|234 from the subtree comprising the leaf set $L_1$ violates this condition: If there is no split separating 1 and 2, which remain connected, then one vertex is labelled twice with 1 and 2. Note that [39, e.g., section 3.1] allow such trees, we, however, exclude them.

# 3 | TOPOLOGY AND STRATIFICATION OF WALD SPACE

## 3.1 | Embedding

Recall from Theorem 2.1.4 that $\psi : \mathcal{W} \to \mathcal{P}$ from Equation (2.1) is injective and so is the equivalent $\phi : \mathcal{W} \to \mathcal{P}$ from Equation (2.4). Its image is characterised by algebraic equalities and inequalities, as shown by the following theorem. Further exploration will yield that the topology of wald space is that of a stratified union of disjoint open unit cubes, each corresponding to a grove from Definition 2.2.3.

**Theorem 3.1.1.** *A matrix $P = (\rho_{uv})_{u,v=1}^N \in \mathcal{P}$ is the $\phi$-image of a wald $F \in \mathcal{W}$ if and only if all of the following conditions are satisfied for arbitrary $u, v, s, t \in L$:*

(R1) $\rho_{uu} = 1$,

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(R2) *two of the following three are equal and smaller than (or equal to) the third*

$$\rho_{uv}\rho_{st}, \quad \rho_{us}\rho_{vt}, \quad \rho_{ut}\rho_{vs},$$

(R3) $\rho_{uv} \geqslant 0$.

*Furthermore, the wald $F \in \mathcal{W}$ is then uniquely determined.*

Before proving Theorem 3.1.1, we elaborate on the above algebraic conditions.

*Remark* 3.1.2.

(1) Condition (R2) above is called the *four-point condition*. In its *non-strict* version, all three products are equal and this indicates some degeneracy, namely that some internal vertices have degree four or higher. The four-point condition is equivalent to (e.g., [11] or [39, p. 147])

$$\rho_{uv}\rho_{st} \geqslant \min\left\{\rho_{us}\rho_{vt}, \ \rho_{ut}\rho_{vs}\right\} \tag{3.1}$$

and implies (e.g., setting $s = t$ in (R2) and exploiting (R1))
(R4) $\rho_{uv} \geqslant \rho_{us}\rho_{sv}$ for all $u, v, s \in L$.
Notably (R1) and (R2) imply, in conjunction with $P \in \mathcal{P}$ that
(R5) $\rho_{uv} < 1$ for all $u \neq v$,
for otherwise, if $\rho_{uv} = 1$ for some $u \neq v$, Condition (R4) implied for any $s \in L$ that

$$\rho_{us} \geqslant \rho_{uv}\rho_{vs} = \rho_{vs} \quad \text{and} \quad \rho_{vs} \geqslant \rho_{uv}\rho_{us} = \rho_{us},$$

so $\rho_{us} = \rho_{vs}$ and hence, $P$ would be singular, a contradiction to $P \in \mathcal{P}$.

(2) Observe that $\phi(F) = (\exp(-d_{uv}))_{u,v=1}^{N}$, where the $d_{uv}$ are the finite or infinite distances

$$d_{uv} := \sum_{e \in E(u,v)} \ell_e = -\log \rho_{uv},$$

between leaves $u, v \in L$, and, with Definition 2.2.3 4, this translates to $d_{uu} = 0$ and $d_{uv} = \infty$ whenever $u$ and $v$ are in different components. In the literature, $(d_{uv})_{u,v=1}^{N}$ is also called *tree metric* (e.g., [39, chapter 7]) or distance matrix (e.g., [14, chapter 11]). Indeed, it conveys a metric on $L$ as Condition (R4) encodes the *triangle inequality* (for any $u, v, s \in L$)

$$d_{uv} \leqslant d_{us} + d_{vs}.$$

(3) In particular, the unit $N \times N$ matrix $I = (\delta_{uv})_{u,v \in L} \in \mathcal{P}$ is the $\phi$-image of the complete disconnected wald $F_\infty \in \mathcal{W}$ with topology $E_\infty = \emptyset$ in which each leaf comprises one of the $K = N$ single element trees.

(4) For a given $P \in \mathcal{P}$ satisfying conditions (R1), (R2) and (R3) there are neighbour joining algorithms in [39, Section 7.3], determining its split $E \in \mathcal{E}$.

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

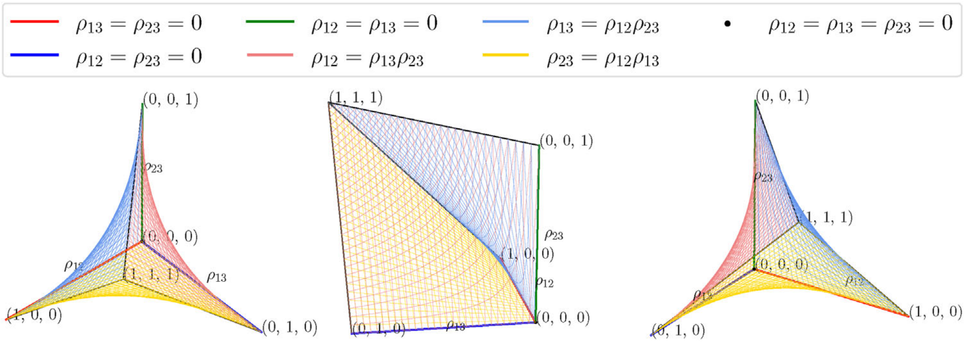


**FIGURE 3** $\phi(\mathcal{W})$ for $N = 3$ embedded in $\mathcal{P}$, where only the off-diagonal entries on the boundary are depicted. Note that the geometry of the wald space is not Euclidean and thus this depiction may be deceiving (as it is a non-isometric embedding into $\mathbb{R}^3$), for example, the regions where one coordinate equals 1 are infinitely far away.

*Proof of Theorem* 3.1.1. '$\Longrightarrow$'. Let $F \in \mathcal{W}$ and $(\rho_{uv})_{u,v=1}^N = \phi(F)$. (R1) and (R3) hold by definition. Further, applying Semple and Steel [39, Theorem 7.2.6] to each connected component asserts (R2) for all $u, v, s, t \in L_\alpha$ for all $\alpha = 1, \dots, K$. Furthermore, as $\rho_{uv} = 0$ whenever $u \neq v$ are in different components, for any $s \in L \setminus \{u, v\}$, $\rho_{us} = 0$ or $\rho_{vs} = 0$, so (R2) holds true in general.

'$\Longleftarrow$'. Let $P = (\rho_{uv})_{u,v=1}^N \in \mathcal{P}$ satisfy (R1), (R2) and (R3) (and thus by Remark 3.1.2 also (R4) and (R5)). The equivalence relation on $L$, defined by $u \sim v \iff \rho_{uv} \neq 0$ partitions $L$ into $L_1, \dots, L_K$ for some $K \in \{1, \dots, N\}$. For each $\alpha = 1, \dots, K$, apply Semple and Steel [39, Theorem 7.2.6] to each tree metric $(d_{uv})_{u,v \in L_\alpha}$ (defined in Remark 3.1.2 2.) to obtain a unique corresponding tree, say $[\mathfrak{V}_\alpha, \mathfrak{E}_\alpha, \ell^{(\alpha)}]$, where, in contrast to our definition, [39] allow leaves on top of each other, in their language, vertices *labelled* more than once. The union of trees gives a forest $\mathfrak{F} = [\mathfrak{V}, \mathfrak{E}, \ell]$ with label set $L$ with $\mathfrak{V} = \bigcup_\alpha \mathfrak{V}_\alpha$, $\mathfrak{E} = \bigcup_\alpha \mathfrak{E}_\alpha$, satisfying $\psi(\mathfrak{F}) = P$. Suppose now a vertex was labelled more than once, say, with distinct leaf labels $u, v \in L_\alpha$, that is, $u \neq v$, for some $\alpha = 1, \dots, K$. Then, $u$ and $v$ have zero distance $d_{uv}$, hence $\rho_{uv} = 1$, yielding a contradiction to (R5) (i.e., $P \notin \mathcal{P}$ as argued in Remark 3.1.2 1.) Thus, $\mathfrak{F} \in \mathcal{W}$ and with Definition 2.2.3, we obtain $F \in \mathcal{W}$ with $\phi(F) = \psi(\mathfrak{F}) = P$. ☐

As $\phi(\mathcal{W})$ is defined by algebraic equalities and non-strict inequalities, we have the following corollary at once.

**Corollary 3.1.3.** *$\phi(\mathcal{W}) \subseteq \mathcal{P}$ is a closed subset of $\mathcal{P}$.*

**Example 3.1.4** ($\mathcal{W}$ for $N = 3$). For $N = 3$, all matrices $P = \phi(F)$ with $F \in \mathcal{W}$ are given by (using Theorem 3.1.1)

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} \quad \text{satisfying the triangle inequalities} \quad \begin{cases} \rho_{12} \geqslant \rho_{13}\rho_{23}, \\ \rho_{13} \geqslant \rho_{12}\rho_{23}, \\ \rho_{23} \geqslant \rho_{12}\rho_{13}, \end{cases}$$

and $0 \leqslant \rho_{12}, \rho_{13}, \rho_{23} < 1$. This set in coordinates $\rho_{12}, \rho_{13}, \rho_{23}$ is depicted in Figure 3, where the 2-dimensional surfaces correspond to the non-linear boundaries resulting from the triangle

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

inequalities. Note that the regions, where at least one coordinate is one, are not included in $\phi(\mathcal{W})$, as the corresponding matrix is no longer strictly positive definite.

**Corollary 3.1.5.** *Conveyed by the homeomorphism $\phi$, $\mathcal{W}$ is star shaped as a subset $\mathbb{R}^{N\times N}$ with respect to $F_\infty$ and hence contractible.*

*Proof.* Let $F \in \mathcal{W}$ with $\phi(F) = P = (\rho_{uv})_{u,v\in L}$ satisfying (R1) - (R3) by Theorem 3.1.1. Recalling from Remark 3.1.2, 3., that $\phi(F_\infty) = I$, consider

$$\left(\rho_{uv}^{(x)}\right)_{u,v\in L} = P^{(x)} = xI + (1-x)P,$$

and observe that for all $x \in [0,1]$, $P^{(x)} \in \mathcal{P}$, $\rho_{uu}^{(x)} = 1 = \rho_{uu}$ for all $u \in L$ and $\rho_{uv}^{(x)} = (1-x)\rho_{uv} \geqslant 0$ for all $u, v \in L$ with $u \neq v$, that is, $P^{(x)}$ satisfies (R1) and (R3) for all $x \in [0,1]$. Moreover, to see that $P^{(x)}$ satisfies Equation (3.1) for all $x \in (0,1)$ for all $u, v, s, t \in L$, assume without loss of generality that

$$\rho_{uv}\rho_{st} = \rho_{us}\rho_{vt} \leqslant \rho_{ut}\rho_{vs}\,. \tag{3.2}$$

If all four $u, v, s, t$ are pairwise distinct, then

$$\rho_{uv}^{(x)}\rho_{st}^{(x)} = \rho_{us}^{(x)}\rho_{vt}^{(x)} \leqslant \rho_{ut}^{(x)}\rho_{vs}^{(x)}\,, \tag{3.3}$$

as well. If only one pair is equal, there are two typical cases. If $u = v$, say, we obtain a different but valid four-point condition

$$\rho_{uv}^{(x)}\rho_{st}^{(x)} \;\geqslant\; \rho_{us}^{(x)}\rho_{vt}^{(x)} \;=\; \rho_{ut}^{(x)}\rho_{vs}^{(x)}\,,$$

where the inequality is strict in case of $\rho_{st} > 0$ due to $1 - x > (1-x)^2$. If $u = t$, say, then we obtain Equation (3.3) where the inequality is strict if $\rho_{vs} > 0$. If exactly two pairs are the same, then, with the above setup only $u = t$ and $v = s$ is possible and both Equation (3.2) and Equation (3.3) are strict. In case of three equal indices, one different, or the same, Equation (3.3) holds again. Therefore, $P^{(x)}$ satisfies (R2) for all $x \in [0,1]$, and by Theorem 3.1.1 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. □

Showing contractibility of the edge-product space, Moulton and Steel contract to the same forest (cf. [31, Proposition 5.1]), employing a different proof, however.

*Remark* 3.1.6. We make the following observations about the proof of Corollary 3.1.5.

(1) All of the wälder $\phi^{-1}(P^{(x)})$, for $0 \leqslant x < 1$ in the proof share the same partition of leaves into connected tree components, due to $\rho_{uv} \neq 0 \iff (1-x)\rho_{uv} \neq 0$ for all $x \in [0,1)$ for all $u, v \in L$.
(2) For $0 < x < 1$, $P^{(x)}$ satisfies unchanged, strict or non-strict four-point conditions (R2), that may be different, though, from those of $P^{(0)} = \phi(F)$.
(3) All triangle inequalities (R4) involving initial non-zero $\rho_{uv}$ 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 labelled and the middle is

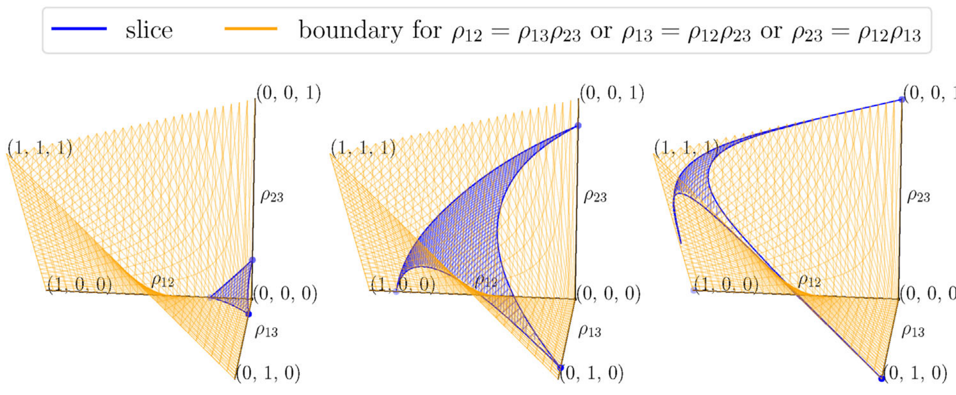


**FIGURE 4** Depicting the off-diagonal matrix entries of $\phi(\mathcal{W})$ embedded in $\mathcal{P}$ for $N = 3$ (orange boundary) in a 3-dimensional coordinate system (cf. Example 3.1.4) and the 2-dimensional images $\phi(B_a)$ (purple) of the slices $B_a$ for $a = 0.2, 0.87, 0.997$ (from left to right).

of degree two), it is immediately transformed into a fully resolved tree (and stays one for all $x \in (0, 1)$).

(4) The point $F_\infty$ can be viewed as a vantage point of $\mathcal{W}$ which is then a bounded part of a cone where every

$$B_a = \left\{ F \in \mathcal{W} \;\middle|\; \phi(F) = (\rho_{uv})_{u,v=1}^N,\ a = 1 - \prod_{\substack{u,v=1 \\ u<v}}^{N} (1 - \rho_{uv}) \right\}.$$

is a *slice* of *level* $a \in [0, 1)$. Then for every $F \in B_a$, there is $r_F > 1$ such that

$$F = \phi^{-1}((1 - x)\phi(F_\infty) + x\phi(F)) \in \mathcal{W}$$

for all $0 \leqslant x < r_F$ and $\phi(F)$ is singular for $x = r_F$. For $N = 3$, the set $B_a$ for several $a \in (0, 1]$ embedded into $\mathcal{P}$ is depicted in Figure 4.

We next consider the restriction of the map $\phi$ to each grove $\mathcal{G}_E$ explicitly in terms of edge weights.

**Definition 3.1.7.** With the agreement (2.5) in case of empty $E(u, v)$, we denote the restriction of $\phi : \mathcal{W} \to \mathcal{P}$ from Definition 2.2.3 to a grove $\mathcal{G}_E$ by

$$\phi_E : (0, 1)^E \to \mathcal{P}, \qquad \lambda \mapsto (\rho_{uv})_{u,v \in L} = \left( \prod_{e \in E(u,v)} (1 - \lambda_e) \right)_{u,v=1}^{N}; \tag{3.4}$$

its continuation onto all of $\mathbb{R}^E$ is denoted by

$$\bar{\phi}_E : \mathbb{R}^E \to \mathcal{S}, \qquad \lambda \mapsto (\rho_{uv})_{u,v \in L} = \left( \prod_{e \in E(u,v)} (1 - \lambda_e) \right)_{u,v=1}^{N}, \tag{3.5}$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

*Remark* 3.1.8. The continuation $\bar{\phi}_E$ is multivariate real analytic on all of $\mathbb{R}^E$.

The following theorem characterises each grove.

**Theorem 3.1.9.**

(1) *For* $F = (E, \lambda) \in \mathcal{W}$ *with* $\phi(F) = (\rho_{uv})_{u,v\in L}$ *we have*

$$\lambda_e = 1 - \max_{\substack{u,v\in A\\ s,t\in B}} \sqrt{\frac{\rho_{ut}\rho_{vs}}{\rho_{uv}\rho_{st}}}\,, \text{ for all } e = A|B \in E\,.$$

(2) *The derivative of* $\phi_E$ *has full rank* $|E|$ *throughout* $(0,1)^E$.
(3) *The map* $\phi_E : (0,1)^E \cong \mathcal{G}_E \to \mathcal{P}$ *is a smooth embedding.*

*Proof.* For the first assertion consider $e = A|B$, where $A \cup B = L_\alpha$, for some $1 \leqslant \alpha \leqslant K$ and where $L_1, \dots, L_K$ is the leaf partition induced by $E$. Then the matrix entries $d_{uv} := -\log \rho_{uv}$ $(u, v \in L_\alpha)$ define a metric on $L_\alpha$, as noted in Remark 3.1.2. For such a metric, [10, Lemma 8] asserts that one can assign a tree $(\mathfrak{V}_\alpha, \mathfrak{E}_\alpha, \ell^\alpha)$ where

$$\ell_e^\alpha = \min_{\substack{u,v\in A\\ s,t\in B}} \frac{1}{2}(d_{ut} + d_{vs} - d_{uv} - d_{st}), \tag{3.6}$$

which is uniquely determined by [10, Theorem 2]. Due to our uniqueness results from Theorem 2.2.2 and Theorem 3.1.1, due to Equation (2.3), $\lambda_e = 1 - \exp(-\ell_e^\alpha)$ and hence, using $\rho_{uv} = \exp(-d_{uv})$, the asserted equation follows at once from Equation (3.6).

For the second assertion, let $e \in E$ and suppose that $F = (E, \lambda)$ decomposes into $K$ subtrees inducing the leaf partition $L_1, \dots, L_K$. Using Equation (3.4), if either $u, v \in L$ are in different subtrees or $u = v$, then

$$\left(\frac{\partial \phi_E}{\partial \lambda_e}(\lambda)\right)_{uv} = 0\,.$$

Else, if $u, v \in L_\alpha$ for some $1 \leqslant \alpha \leqslant K$, then $\rho_{uv} > 0$ and with the Kronecker delta $\delta$,

$$\left(\frac{\partial \phi_E}{\partial \lambda_e}(\lambda)\right)_{uv} = -\delta_{e\in E(u,v)} \prod_{\substack{\tilde{e}\in E(u,v)\\ \tilde{e}\neq e}} (1 - \lambda_{\tilde{e}}) = -\frac{\rho_{uv}}{1-\lambda_e}\delta_{e\in E(u,v)}\,. \tag{3.7}$$

Thus, for every $x \in \mathbb{R}^E$, we have

$$((\mathrm{d}\phi_E)_\lambda(x))_{uv} = -\rho_{uv} \sum_{e\in E} \frac{x_e}{1-\lambda_e}\delta_{e\in E(u,v)}\,,$$

so that $((\mathrm{d}\phi_E)_\lambda(x))_{uv} = 0$ implies

$$0 = \sum_{e\in E(u,v)} \frac{x_e}{1-\lambda_e} =: h_{uv}. \tag{3.8}$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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 a representative $(\mathfrak{V}, \mathfrak{E})$ of $E$ with leaf set partition $L_1, \dots, L_K$, for every $e \in E$ there are $v_1, v_2 \in \mathfrak{V}_\alpha$ with suitable $1 \leqslant \alpha \leqslant K$ such that $e$ corresponds to $\{v_1, v_2\} \in \mathfrak{E}_\alpha$. In particular, as $(\mathfrak{V}_\alpha, \mathfrak{E}_\alpha)$ is a tree, there are $u, v, s, t \in L_\alpha$ (not necessarily all of them distinct), such that

$$\ell'_e = \frac{1}{2}(h_{uv} + h_{st} - h_{ut} - h_{vs}).$$

If the right-hand side is zero due to Equation (3.8), then $x_e = 0$, yielding that $(\mathrm{d}\phi_E)_\lambda$ has full rank, as asserted.

The third assertion follows directly from assertions 1 and 2 in the statement of the theorem, that is, $\phi_E$ is bijectively smooth onto its image and its differential is injective. □

In the following, we are concerned with $\bar{\phi}_E(\lambda)$ if $\lambda \in (0,1)^E$ approaches the boundary. The next result characterises exactly under which conditions $\bar{\phi}_E(\lambda)$ stays in the image $\phi(\mathcal{W})$ of wald space under $\phi$.

**Lemma 3.1.10.** *Let $F \in \mathcal{W}$ with topology $[F] = E$ and let $\lambda^* \in \partial([0,1]^E)$ with $\bar{\phi}_E(\lambda^*) = (\rho^*_{uv})_{u,v=1}^N$. 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.* The first equivalence follows from that Equation (3.9) is well-defined. We prove the second equivalence.

'⇒': Follows from Remark 3.1.2, Condition (R5).

'⇐': Analogously to the proof of Theorem 3.1.1, '⇐', we find a phylogenetic forest in the sense of [39, chapter 2.8], whose tree metric coincides with the one obtained from $\bar{\phi}_E(\lambda^*)$, but there might be multiply labelled vertices. However, this is impossible due to $\rho^*_{uv} < 1$ for any $u \neq v$, which is equivalent to a distance greater than zero between $u$ and $v$. Therefore, there exists a phylogenetic forest $F' \in \mathcal{W}$ with $\phi(F') = \bar{\phi}_E(\lambda^*)$, and thus by Theorem 3.1.1, $\bar{\phi}_E(\lambda^*) \in \mathcal{P}$. □

The previous result immediately shows which matrices in $\mathcal{P}$ form the boundary of a grove.

**Corollary 3.1.11.** *Let $E$ be a wald topology. Then the boundary of the grove $\mathcal{G}_E$ in $\mathcal{W}$ is given by*

$$\partial\mathcal{G}_E = \left\{ \phi^{-1}\big(\bar{\phi}_E(\lambda^*)\big) : \lambda^* \in \partial([0,1]^E),\ \bar{\phi}_E(\lambda^*) \in \mathcal{P} \right\}. \tag{3.9}$$

The following result gives a first glimpse on how different groves are connected through the convergence of wälder.

**Theorem 3.1.12.** *Let $\mathcal{W} \ni (E_n, \lambda^{(n)}) = F_n \to F' = (E', \lambda') \in \mathcal{W}$. 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 $E \neq E'$.*

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

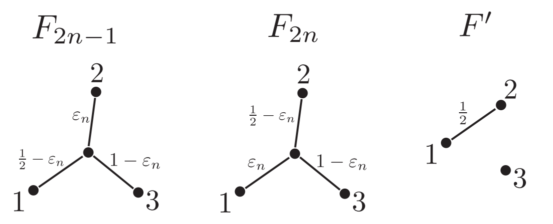


**FIGURE 5** A sequence of wälder (left and middle) converging (right) but having different $\lambda$ cluster points as detailed in Example 3.1.13.

*Proof.* For the first assertion, noting that there are only finitely many wald topologies, there needs to exist a subsequence $F_{n_k}$ of $F_n$ with $E_{n_k} = E$ for some topology $E$ for all $k \in \mathbb{N}$, and thus, as $F_{n_k} \in \mathcal{G}_E \cong (0,1)^E$, 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 assertion 1, by Bolzano–Weierstraß, there needs to exist a cluster point $\lambda^* \in [0,1]^E$ of $\lambda^{(n_k)}$.

For assertion 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 assertion 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. □

The following example teaches that when $F_n \to F$, $\lambda^{(n)}$ can have distinct cluster points.

**Example 3.1.13.** Let $N = 3$, set $e_i := u|(L \setminus \{u\})$, $u \in L = \{1,2,3\}$ and $E = \{e_1, e_2, e_3\}$. Define the sequence of wälder $F_n := (E, \lambda^{(n)})$, $n \in \mathbb{N}$, using a sequence $\varepsilon_n \in (0, \frac{1}{4})$ with $\varepsilon_n \to 0$ as $n \to \infty$, via

$$\lambda_{e_1}^{(2n-1)} := \frac{1}{2} - \varepsilon_n, \qquad \lambda_{e_2}^{(2n-1)} := \varepsilon_n, \qquad \lambda_{e_3}^{(2n-1)} := 1 - \varepsilon_n,$$

$$\lambda_{e_1}^{(2n)} := \varepsilon_n, \qquad \lambda_{e_2}^{(2n)} := \frac{1}{2} - \varepsilon_n, \qquad \lambda_{e_3}^{(2n)} := 1 - \varepsilon_n.$$

The corresponding forests are depicted in Figure 5. Clearly, the sequence $\lambda^{(n)}$ ($n \in \mathbb{N}$) has two distinct cluster points $(1/2, 0, 1), (0, 1/2, 1) \in (0,1)^3$. We observe, however, that

$$\phi(F_{2n-1}) = \begin{pmatrix} 1 & (1-\varepsilon_n)(\frac{1}{2}+\varepsilon_n) & (\frac{1}{2}+\varepsilon_n)\varepsilon_n \\ (1-\varepsilon_n)(\frac{1}{2}+\varepsilon_n) & 1 & (1-\varepsilon_n)\varepsilon_n \\ (\frac{1}{2}+\varepsilon_n)\varepsilon_n & (1-\varepsilon_n)\varepsilon_n & 1 \end{pmatrix} \overset{n\to\infty}{\longrightarrow} \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and similarly, $\phi(F_{2n})$ converges to the same matrix as $n \to \infty$. Letting $e' = 1|2$ and defining $F' = (E', \lambda') = (\{e'\}, \lambda')$ with $\lambda'_{e'} = \frac{1}{2}$ (i.e., label partitions $L'_1 = \{3\}$, $L'_2 = \{1,2\}$; cf. Figure 5), we have that $\phi(F_n) \to \phi(F')$, so $F_n \to F'$.

Theorem 3.1.12 shows that whenever a sequence of wälder $F_n \in \mathcal{G}_E$ converges to a wald $F' \in \mathcal{W}$ with topology $E'$ and $F' \notin \mathcal{G}_E$, then $F' \in \partial\mathcal{G}_E$. In the following section, we make this relationship between $E'$ and $E$ more precise and unravel the boundary correspondences via a partial ordering on the wald topologies.

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

## 3.2 | At Grove's end

In light of Theorem 3.1.12, we investigate how two wald topologies $E = [F]$ and $E' = [F']$ are related to each other.

**Definition 3.2.1.** Let $F \in \mathcal{W}$ be a wald with topology $E = [F]$. For an edge $e = A|B \in E$, we define the edge *restricted* to some subset $L' \subseteq L$ by

$$e|_{L'} := (A \cap L')|(B \cap L')$$

if both of the sets above are non-void, else, we say that the *restriction does not exist*. In case of existence, we also say that $e|_{L'}$ is a *valid split*.

The following definition is from [31] and translated into the language of wälder and their topologies.

**Definition 3.2.2.** For two wälder $F, F' \in \mathcal{W}$ with topologies $E = [F], E' = [F']$, respectively, we define a relation $\leq$ by

$$E' = [F'] \leqslant [F] = E \tag{3.10}$$

if all of the following three properties hold:

**Refinement**: with the partitions $L_1, \dots, L_K$ and $L'_1, \dots, L'_{K'}$ of $L$ induced by $E'$ and $E$, respectively, for every $1 \leqslant \alpha' \leqslant K'$ there is $1 \leqslant \alpha \leqslant K$ with $L'_{\alpha'} \subseteq L_\alpha$;
**Restriction**: for every $1 \leqslant \alpha' \leqslant K'$,

$$E'_{\alpha'} \subseteq E|_{L'_{\alpha'}} := \left\{\tilde{e} : \exists e \in E \text{such that} \tilde{e} := e|_{L'_{\alpha'}} \text{ is a valid split}\right\}$$

where the right-hand side is the set of splits $E$ *restricted to* $L'_{\alpha'}$;
**Cut**: for every $1 \leqslant \alpha'_1 \neq \alpha'_2 \leqslant K'$ and $1 \leqslant \alpha \leqslant K$ with $L'_{\alpha'_1}, L'_{\alpha'_2} \subset L_\alpha$, there is some

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

Further, we say $E' < E$ if $E \neq E' \leqslant E$. We also write $F' < F$ if $E' < E$.

The restriction condition above corresponds to the definition of a tree *displaying* another tree in [31]. From [31, Lemma 3.1], it follows at once that the relation $\leq$ as defined in Equation (3.10) is a partial ordering.

**Example 3.2.3.** Let $N = 5$, so $L = \{1, \dots, 5\}$. Define three wald topologies

$$E = \left\{1|2345, 12|345, 3|1245, 123|45, 1234|5, 1235|4\right\},$$
$$E'_1 = \{2|3, 4|5\},$$
$$E'_2 = \{2|5, 3|4\}.$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

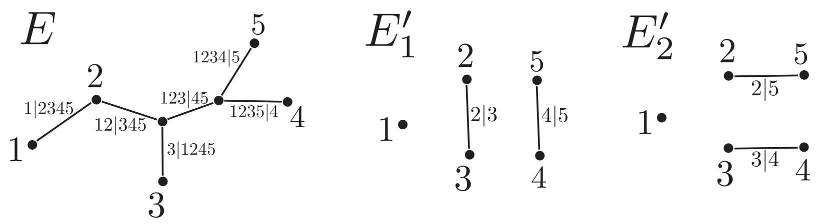


**FIGURE 6** Wald topologies $E$, $E_1'$ and $E_2'$ from Example 3.2.3, where $E_1' < E$ but $E_2' \nleq E$.

They are depicted in Figure 6. Then $E_1' < E$, as the refinement property holds, the restriction property, due to $E|_{\{2,3\}} = \{2|3\}, E|_{\{4,5\}} = \{4|5\}$ and the cut property, as the edge $1|2345$ separates $\{1\}$ from $\{2,3\}$ and $\{4,5\}$, and $123|45$ separates $\{2,3\}$ from $\{4,5\}$. Separating edges like $1|2345$ cannot be restricted to any of the leaf sets $\{1\}$, $\{2,3\}$ and $\{4,5\}$, and if edges are restricted, they can only be restricted to one leaf set, for example, $12|345$ can be restricted only to $\{2,3\}$ and not to any of the others.

In contrast, $E_2' \nleq E$, although the refinement and restriction properties are satisfied, the cut property is not, as there is no edge $A|B = e \in E$ with $\{2,5\} \subseteq A$ and $\{3,4\} \subseteq B$.

**Definition 3.2.4.** Let $E, E'$ be wald topologies with $E' \leqslant E$.

(1) For each edge $e' \in E'_{\alpha'}$, $1 \leqslant \alpha' \leqslant K'$, denote the set of all *corresponding* splits in $E$ by

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

(2) Furthermore, denote the set of all *disappearing* splits in $E$ with

$$R_{\text{dis}} := \left\{e \in E : \exists \alpha' \text{ such that } e|_{L'_{\alpha'}} \text{ is a valid split of } L'_{\alpha'}, \text{ but } e|_{L'_{\alpha'}} \notin E'\right\}.$$

(3) Denote the set of all *cut* splits with

$$R_{\text{cut}} := \left\{e \in E : \nexists \alpha' \text{ such that } e|_{L'_{\alpha'}} \text{ is a valid split of } L'_{\alpha'}\right\}.$$

**Example 3.2.5.**

(1) We revisit Example 3.2.3, cf. also Figure 6. Note that with respect to $E_1' < E$, we have, for instance, with $e' = 2|3$ that $R_{e'} = \{12|345, 3|1245\}$, $R_{\text{dis}} = \emptyset$ and $R_{\text{cut}} = \{1|2345, 123|45\}$. By definition, none of the cut edges can be restricted.

(2) Let $N = 4$, so $L = \{1,2,3,4\}$. Define two wald topologies with

$$E = \left\{1|234, 2|134, 3|124, 123|4, 12|34\right\},$$
$$E' = \left\{1|234, 2|134, 3|124, 123|4\right\},$$

where $E$ is a fully resolved tree with interior edge $12|34$ and $E'$ is a *star tree*, that is, four leaves attached to one interior vertex, cf. Figure 7. Then $E' < E$ because $E' \subset E$, and the split $12|34$

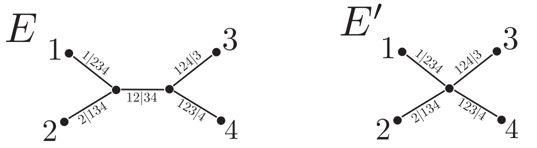


**FIGURE 7** Wald topologies $E$ and $E'$ from Example 3.2.5 (2), where $E' < E$.

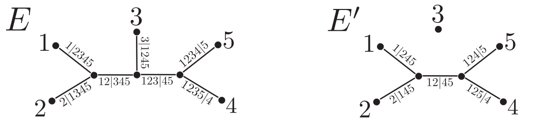


**FIGURE 8** Wald topologies $E$ and $E'$ from Example 3.2.5 (3), where $E' < E$.

disappears, that is, $R_{\text{dis}} = \{12|34\}$. Furthermore, $R_{\text{cut}} = \emptyset$ and $R_{e'} = \{e'\}$ for all $e' \in E'$.

(3) Let $N = 5$, so $L = \{1, 2, 3, 4, 5\}$. Define two wald topologies with

$$E = \{1|2345, 2|1345, 3|1245, 4|1235, 5|1234, 12|345, 123|45\},$$
$$E' = \{1|245, 2|145, 4|125, 5|124, 12|45\},$$

where $E$ is a fully resolved tree with two *cherries* containing 1,2 and 4,5, respectively, and 3 attached as a leaf to an interior vertex, cf. Figure 8. Furthermore, $E'$ has two connected components, a fully resolved tree with labels 1,2,4,5 and isolated label 3, cf. Figure 8. Then $E' < E$ and $R_{12|45} = \{12|345, 123|45\}$, $R_{\text{cut}} = \{3|1245\}$ and $R_{\text{dis}} = \emptyset$.

**Lemma 3.2.6.** *Let $E' \leqslant E$ with label partitions $L_1, \dots, L_K$ and $L'_1, \dots, L'_{K'}$, respectively, and $u, v \in L$. Then the following hold.*

- (i) *If $K = K'$, then without loss of generality $L'_\alpha = L_\alpha$ and $E'_\alpha \subseteq E_\alpha$ for all $\alpha = 1, \dots, K$ and $R_{e'} = \{e'\}$ for all $e' \in E'$.*
- (ii) *$K < K' \iff R_{\text{cut}} \neq \emptyset$.*
- (iii) *If $K = K'$, then $E' < E \iff R_{\text{dis}} \neq \emptyset$.*
- (iv) *$R_{e'} \neq \emptyset$ for all $e' \in E'$ and if $\exists e' \in E'_{\alpha'}$ with $|R_{e'}| > 1$ and $L'_{\alpha'} \subseteq L_\alpha$, then $L'_{\alpha'} \subsetneq L_\alpha$.*
- (v) *$E = E' \iff (R_{\text{dis}} = \emptyset$ and $R_{\text{cut}} = \emptyset)$.*
- (vi) *$R_{e'} \cap R_{e''} = \emptyset$ for all $E' \ni e' \neq e'' \in E'$.*
- (vii) *The splits in $E|_{L'_{\alpha'}}$ are pairwise compatible.*
- (viii) *$e' \in E'(u, v) \iff R_{e'} \cap E(u, v) \neq \emptyset \iff R_{e'} \subseteq E(u, v)$.*
- (ix) *$R_{\text{dis}}, R_{\text{cut}}$ in conjunction with the $R_{e'}$ over all $e' \in E'$ give a pairwise disjoint union of $E$, where $R_{\text{dis}}$ and $R_{\text{cut}}$ might be empty.*
- (x) *Let $u, v \in L'_{\alpha'}$ for some $1 \leqslant \alpha' \leqslant K'$. Then $R_{\text{dis}} \cap E(u, v)$ in conjunction with the $R_{e'}$ over all $e' \in E'(u, v)$ give a pairwise disjoint union of $E(u, v)$, where $R_{\text{dis}} \cap E(u, v)$ might be empty.*
- (xi) *For any $L'_{\alpha'}, L'_{\alpha''} \subset L_\alpha$ with $\alpha' \neq \alpha''$, there exists a split $A|B = e \in E$ with $L'_{\alpha'} \subseteq A$, $L'_{\alpha''} \subseteq B$ and $e \in R_{\text{cut}}$.*

*Let $F, F' \in \mathcal{W}$ with $\rho = \phi(F)$, $\rho' = \phi(F')$ and topologies $E$ and $E'$, respectively, with label partitions $L_1, \dots, L_K$ and $L'_1, \dots, L'_{K'}$, respectively. Then the following hold*

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(xii) *If for all $u, v \in L$: $\rho_{uv} = 0 \Rightarrow \rho'_{uv} = 0$, then $L'_1, \dots, L'_{K'}$ is a refinement of $L_1, \dots, L_K$.*

*Finally, we have the general result.*

(xiii) *For every wald topology $E'$ with $|E'| < 2N - 3$, there is a wald topology $E$ with $|E| = |E'| + 1$ and $E' < E$.*

*Proof.* As Assertions (ii) and (v) follow from Assertion (xi), we proceed in the following logical order.

(i) $K = K'$ implies without loss of generality $L_\alpha = L'_\alpha$ for all $\alpha = 1, \dots, K$ and therefore $e|_{L'_\alpha} = e|_{L_\alpha} = e$ are valid splits for all $e \in E_\alpha$ for all $\alpha = 1, \dots, K$, so $E'_\alpha \subseteq E_\alpha$ as well as $R_{e'} = \{e'\}$ for all $e' \in E'$.

(iii) From (i) without loss of generality $L_\alpha = L'_\alpha$ and $E'_\alpha \subseteq E_\alpha$, $\alpha = 1, \dots, K$. Thus, $R_{\text{dis}} = \emptyset \iff$ (for all $\alpha = 1, \dots, K$, $E'_\alpha = E_\alpha$) $\iff E' = E$.

(iv) By the restriction property of $E' \leqslant E$, 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 $e_1 = e_1|_{L'_{\alpha'}} = e_2|_{L'_{\alpha'}} = e_2$, a contradiction.

(vi) 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 $e' = e|_{L'_{\alpha'}} = e''$, a contradiction to $e' \neq e''$, so $\alpha' \neq \alpha''$. As $e$ 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 $E' \leqslant E$, by the cut property there exists $C|D = \tilde{e} \in E_\alpha$ separating $L'_{\alpha'}$ and $L'_{\alpha''}$, that is, $L'_{\alpha'} \subseteq C$ and $L'_{\alpha''} \subseteq D$. But then $\tilde{e}, e \in E_\alpha$ cannot be compatible, a contradiction.

(vii) Let $L'_{\alpha'} \subset L_\alpha$ and $e'_1, e'_2 \in E|_{L'_{\alpha'}}$ such that $e'_i = e_i|_{L'_{\alpha'}}$ for some $e_1, e_2$ with $e_i = A_i|B_i \in E$, $i = 1, 2$. Then $e_1, e_2 \in E_\alpha$ for otherwise their restrictions to $L'_{\alpha'}$ would not be valid splits. As $e_1$ and $e_2$ are compatible, without loss of generality $A_1 \cap A_2 = \emptyset$. Consequently, $e'_i = (A_i \cap L'_{\alpha'})|(B_i \cap L'_{\alpha'})$ for $i = 1, 2$ and so $e'_1$ and $e'_2$ are compatible as $(A_1 \cap L'_{\alpha'}) \cap (A_2 \cap L'_{\alpha'}) = \emptyset$.

(viii) We show $e' \in E'(u, v) \Rightarrow R_{e'} \subseteq E(u, v) \Rightarrow R_{e'} \cap E(u, v) \neq \emptyset \Rightarrow e' \in E'(u, v)$.

If $e' \in E'(u, v)$, then due to (iv), $R_{e'} \neq \emptyset$. Hence, $e' := e|_{L'_{\alpha'}} = (A \cap L'_{\alpha'})|(B \cap L'_{\alpha'})$ for some $e = A|B \in R_{e'}$, and thus $u \in A$, $v \in B$, or vice versa, that is, $e \in E(u, v)$. As the choice $e \in R_{e'}$ was arbitrary, $R_{e'} \subseteq E(u, v)$. If $e \in R_{e'} \cap E(u, v)$, $u, v \in L'_{\alpha'}$, then $e' = e|_{L'_{\alpha'}}$ and $e' \in E'(u, v)$ due to Equation (2.8).

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

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

(xi) Without loss of generality, let $K = 1 < K'$ and suppose that $\alpha' = 1$, $\alpha'' = 2$.

In the first step, note that it suffices to find a split $e = A|B$ that separates $L'_1$ from $L'_{\alpha'}$ for all $2 \leqslant \alpha' \leqslant K'$ for then, without loss of generality $L'_1 \subseteq A$, $L'_2, \dots L'_{K'} \subset B$, which implies $L'_1 = A$, $L'_2 \cup \dots \cup L'_{K'} = B$, so that none of the $e|_{L'_1}, \dots e|_{L'_{K'}}$ is a valid split and in consequence $e \in R_{\text{cut}}$ as desired.

In the second step, we show the existence of such a $e$. In fact, to this end, it suffices to establish the following claim for all $3 \leqslant J \leqslant K'$, invoke induction and separately show the assertion for $K' = 2$.

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Claim: If $\exists$ split $f = C|D$ separating $L'_1$ from all of $L'_1, \dots, L'_{J-1}$, that is, without loss of generality $L'_1 \subseteq C,\ L'_2, \dots, L'_{J-1} \subset D$, that has the property $C \cap L'_J \neq \emptyset \neq D \cap L'_J$ then $\forall$ compatible splits $e = A|B$ separating $L'_1$ from $L'_J$ where, without loss of generality $L'_1 \subseteq A, L'_J \subseteq B$ we have that $e$ separates $L'_1$ from all of $L'_1, \dots, L'_J$, that is, equivalently

$$L'_{\alpha'} \subset B\ \forall 2 \leqslant \alpha' \leqslant J\,.$$

Indeed, if $K' = 2$ and $e = A|B$ separates $L'_1$ from $L'_2$ then, without loss of generality, $A = L'_1$ and $B = L'_2$.

In the third step, we show the claim. To this end let $K' \geqslant 3$, $3 \leqslant J \leqslant K'$, $f = C|D$ as in the claim's hypothesis and suppose that $e = A|B$ is an arbitrary compatible split with $L'_1 \subseteq A, L'_J \subseteq B$. Then

$$C \cap A \supseteq L'_1 \neq \emptyset,\ C \cap B \supseteq C \cap L'_J \neq \emptyset,\ D \cap B \supseteq D \cap L'_J \neq \emptyset\,.$$

By compatibility of splits we have thus $\emptyset = D \cap A \supseteq L'_{\alpha'} \cap A$ for all $2 \leqslant \alpha' \leqslant J$ by hypothesis, yielding

$$L'_{\alpha'} \subset B\ \forall 2 \leqslant \alpha' \leqslant J\,,$$

thus establishing the claim.

(ii) We show equivalently $K = K' \Leftrightarrow R_{\text{cut}} = \emptyset$. '$\Rightarrow$': If $K = K'$, then by (i) without loss of generality $L'_\alpha = L_\alpha$ and in particular $e|_{L'_\alpha} = e|_{L_\alpha} = e$ are valid splits for all $e \in E_\alpha$, $\alpha = 1, \dots, K$, so that $R_{\text{cut}} = \emptyset$. '$\Leftarrow$' follows at once from (xi).

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

(xii) Let $1 \leqslant \alpha' \leqslant K'$ and $u \in L'_{\alpha'}$. Then, there is $1 \leqslant \alpha \leqslant K$ such that $u \in L_\alpha$. For any other $v \in L'_{\alpha'}$, $\rho'_{uv} > 0$, so by assumption $\rho_{uv} > 0$, thus $v \in L_\alpha$, yielding $L'_{\alpha'} \subseteq L_\alpha$.

(xiii) Suppose that $F'$ is a wald with leaf partition $L'_1, \dots, L'_{K'}$ and $|E'| < 2N - 3$.

In case of $K' = 1$ there is a vertex of degree $k \geqslant 4$, that is, there is a partition $A_1, \dots, A_k$ of $L = L'_1$ with splits

$$A_i | L \setminus A_i \in E',\ 1 \leqslant i \leqslant k$$

and all other splits in $E'$ are of form

$$A'_i | L \setminus A'_i \in E',\ 1 \leqslant i \leqslant k,$$

where $A'_i$ is a suitable subset of $A_i$. Then one verifies at once that the new split $e := A_1 \cup A_2 | L \setminus (A_1 \cup A_2)$ is compatible with all splits in $E'$ so that $E := E' \cup \{e\}$ is a wald topology with the desired properties $|E| = |E'| + 1$ and $E' < E$. For the latter note that $R_{e'} = \{e'\}$ for all $e' \in E'$, $R_{\text{cut}} = \emptyset$ and $R_{\text{dis}} = \{e\}$.

In case of $K' \geqslant 2$ introduce the new split $f := L'_1 | L'_2$ and for every $e'_1 = A|B \in E'_1$ let $e(e'_1) := A|B \cup L'_2$, so that $e(e'_1)|_{L'_1} = e'_1$. Similarly, for every $e'_2 = C|D \in E'_2$ let $e(e'_2) := C|D \cup L'_1$, so that $e(e'_2)|_{L'_2} = e'_2$. Setting

$$E := \{e(e') : e' \in E'_1 \cup E'_2\} \cup \{f\} \cup E'_3 \dots \cup E'_{K'}$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

one verifies that all splits in $E$ are pairwise compatible. Hence, $E$ is a wald topology with $|E| = |E'| + 1$ and $E' < E$. Indeed, for the latter note that $R_{e'} = \{e(e')\}$ for all $e' \in E'_1 \cup E'_2$, $R_{e'} = \{e'\}$ for all $e' \in E'_3 \cup \ldots \cup E'_{K'}$, $R_{\mathrm{cut}} = \{f\}$ and $R_{\mathrm{dis}} = \emptyset$. □

In the following theorem, we characterise the boundaries of groves via the partial ordering on wald topologies.

**Theorem 3.2.7.** *For wald topologies $E$ and $E'$, the following three statements are equivalent (with $\partial\mathcal{G}_E$ as in Equation 3.9):*

(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 $L_1, \ldots, L_K$.

'$(i) \Rightarrow (ii)$'. Assume that $F' = (E', \lambda') \in \mathcal{G}_{E'}$ with partition $L'_1, \ldots, L'_{K'}$. Using Lemma 3.2.6(ix), set

$$\lambda^*_e := \begin{cases} 0 & e \in R_{\mathrm{dis}} \\ 1 & e \in R_{\mathrm{cut}} \\ 1 - (1 - \lambda'_{e'})^{1/|R_{e'}|} & e \in R_{e'}, e' \in E' \end{cases}$$

to obtain $\lambda^* \in \partial([0,1]^E)$ because $R_{\mathrm{cut}} \cup R_{\mathrm{dis}} \neq \emptyset$ due to $E' < E$ by Lemma 3.2.6(v). By injectivity of $\phi$, it suffices to show $(*)$:

$$\bar{\phi}_E(\lambda^*) =: (\rho^*_{uv})_{u,v=1}^N \overset{(*)}{=} (\rho'_{uv})_{u,v=1}^N := \phi(F') .$$

First, observe by Agreement (2.5) that for all $u \in L$,

$$\rho^*_{uu} = 1 = \rho'_{uu} .$$

Next, again from Agreement (2.5), 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 $\alpha'_1, \alpha'_2 \in \{1, \ldots, K'\}$, 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 3.2.6(xi), there exists an edge $A|B = e \in R_{\mathrm{cut}}$ 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 3.2.6(x),

$$\begin{aligned} \rho^*_{uv} &= \prod_{e \in E(u,v)} (1 - \lambda^*_e) \\ &= \left( \prod_{e \in R_{\mathrm{dis}} \cap E(u,v)} (1 - \lambda^*_e) \right) \left( \prod_{e' \in E'(u,v)} \prod_{e \in R_{e'}} (1 - \lambda'_{e'})^{1/|R_{e'}|} \right) \end{aligned}$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$= \prod_{e' \in E'(u,v)} (1 - \lambda'_{e'}) = \rho'_{uv}.$$

Thus, we have shown $\phi(F') = \bar{\phi}_E(\lambda^*)$. As $F' = (E', \lambda')$ was arbitrary, we have shown $\mathcal{G}_{E'} \subset \partial\mathcal{G}_E$ where equality cannot be due to $\lambda^* \in \partial([0,1]^E)$.

'$(ii) \Rightarrow (iii)$' is trivial.

'$(iii) \Rightarrow (i)$'. Let $F' = (E', \lambda') \in \mathcal{G}_{E'} \cap \partial\mathcal{G}_E$, that is, there exists $\lambda^* \in \partial([0,1]^E)$ with $\bar{\phi}_E(\lambda^*) = \phi(F') \in \mathcal{P}$. In the following, we will construct $F^\circ = (E^\circ, \lambda^\circ)$ with $\lambda^\circ \in (0,1)^{E^\circ}$ and show that

**Claim I**: $E^\circ < E$, and
**Claim II**: $\phi(F^\circ) = \phi(F')$.

As Claim II implies $F^\circ = F'$ and $E^\circ = E'$, in conjunction with Claim I we then obtain the assertion $E' < E$.

To see Claim I, let $\bar{\phi}_E(\lambda^*) = (\rho^*_{uv})^N_{u,v=1}$. Denote the connectivity classes of $L$, where $u, v \in L$ are connected if and only if $\rho^*_{uv} > 0$, by

$$L^\circ_1, \dots, L^\circ_{K^\circ},$$

with $1 \leqslant K^\circ \leqslant N$. As $\rho^*_{uv} > 0$ implies $\rho_{uv} > 0$, we have that $L^\circ_1, \dots, L^\circ_{K^\circ}$ is a *refinement* of $L_1, \dots, L_K$ by Lemma 3.2.6(xii).

Define $E^\circ$ by setting for each $1 \leqslant \alpha^\circ \leqslant K^\circ$ (where, say, $L^\circ_{\alpha^\circ} \subset L_\alpha$ for some $1 \leqslant \alpha \leqslant K$)

$$E^\circ_{\alpha^\circ} := \left\{ e|_{L^\circ_{\alpha^\circ}} : \exists e \in E \text{such that } e|_{L^\circ_{\alpha^\circ}} \text{ is a valid split and } \lambda^*_e \neq 0 \right\}, \tag{3.11}$$

and $E^\circ := \bigcup_{\alpha^\circ} E^\circ_{\alpha^\circ}$. By Lemma 3.2.6(vii), each $E^\circ_{\alpha^\circ}$ comprises compatible splits only so that $E^\circ$ satisfies the *restriction property* from Definition 3.2.2.

Verifying the cut property, suppose there exist $1 \leqslant \alpha^\circ_1 \neq \alpha^\circ_2 \leqslant K^\circ$ and $1 \leqslant \alpha \leqslant K$ such that $L^\circ_{\alpha^\circ_1}, L^\circ_{\alpha^\circ_2} \subset L_\alpha$. Hence, by construction

$$\rho^*_{us} = 0,\ \rho^*_{uv} > 0 \text{ and } \rho^*_{st} > 0 \text{ for all } u, v \in L^\circ_{\alpha_1} \text{ and } s, t \in L^\circ_{\alpha_2}. \tag{3.12}$$

Let now $u \in L^\circ_{\alpha^\circ_1}$ and $s \in L^\circ_{\alpha^\circ_2}$, then by definition of $\bar{\phi}_E$, $\rho^*_{us} = \prod_{e \in E(u,s)} (1 - \lambda^*_e) = 0$, so there must exist $e = A|B \in E(u,s)$ with $\lambda^*_e = 1$. This implies $L^\circ_{\alpha^\circ_1} \subseteq A$ and $L^\circ_{\alpha^\circ_2} \subseteq B$, for otherwise, if $A \not\ni v \in L^\circ_{\alpha^\circ_1}$, say, then $v \in B$ and hence $e \in E(u,v)$ due to Equation (2.8) and hence $\rho^*_{uv} = 0$, due to $\lambda^*_e = 1$, a contradiction to Equation (3.12). Thus, the *cut property* holds.

Having verified all of the properties from Definition 3.2.2, we have shown $E^\circ \leqslant E$, and we can use the notation introduced in Definition 3.2.4 and Lemma 3.2.6 is applicable for $E^\circ \leqslant E$. As $\lambda^*$ 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 3.2.6, (v), yielding $E^\circ < E$, which was Claim I.

To see Claim II, we define suitable edge weights $\lambda^\circ$. Let $1 \leqslant \alpha^\circ \leqslant K^\circ$ be arbitrary and let $1 \leqslant \alpha \leqslant K$ be such that $L^\circ_{\alpha^\circ} \subseteq L_\alpha$. For each $e^\circ \in E^\circ_{\alpha^\circ}$, define

$$\lambda^\circ_{e^\circ} := 1 - \prod_{e \in R_{e^\circ}} (1 - \lambda^*_e). \tag{3.13}$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Indeed, $\lambda^\circ_{e^\circ} \in (0,1)$, as by Lemma 3.2.6(ix), none of the $e \in R_{e^\circ}$ lie in $R_{\text{cut}}$, we have $\lambda^*_e < 1$, and, as at least for one $e \in R_{e^\circ}$, we have $\lambda^*_e > 0$ by Equation (3.11). 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 $\phi(F') = \bar{\phi}_E(\lambda^*) = (\rho^*_{uv})^N_{u,v=1}$ and let $\phi(F^\circ) = (\rho^\circ_{uv})^N_{u,v=1}$. By Agreement (2.5), for all $u \in L$ we have $\rho^*_{uu} = 1 = \rho^\circ_{uu}$ and by definition of the connectivity classes $L^\circ_1, \dots, L^\circ_{K^\circ}$ 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 $1 \leqslant \alpha^\circ \leqslant K^\circ$ and $1 \leqslant \alpha \leqslant K$. By Lemma 3.2.6(viii) and (ix), the sets $R_{\text{dis}} \cap E(u,v)$ in conjunction with $R_{e^\circ}$ for all $e^\circ \in E^\circ(u,v)$ form a partition of $E(u,v)$. For the first set we have

$$e \in R_{\text{dis}} \cap E(u,v) \Rightarrow \lambda^*_e = 0\,. \tag{3.14}$$

Indeed, if $e \in R_{\text{dis}} \cap E(u,v)$ then the restriction $e^\circ := e|_{L^\circ_{\alpha^\circ}}$ 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 $E^\circ$ which, taking into account Equation (3.11), 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 $R_{e^\circ}$, $e^\circ \in E^\circ(u,v)$ partition $E(u,v)$ and the third uses for the first factor (3.14) and (3.13) for the second factor)

$$\begin{aligned}
\rho^*_{uv} &= \prod_{e \in E(u,v)} (1 - \lambda^*_e) \\
&= \left( \underbrace{\prod_{e \in R_{\text{dis}} \cap E(u,v)} (1 - \lambda^*_e)}_{=1} \right) \left( \prod_{e^\circ \in E^\circ(u,v)} \underbrace{\prod_{e \in R_{e^\circ}} (1 - \lambda^*_e)}_{=1-\lambda^\circ_{e^\circ}} \right) \\
&= \prod_{e^\circ \in E^\circ(u,v)} (1 - \lambda^\circ_{e^\circ}) \\
&= \rho^\circ_{uv},
\end{aligned}$$

completing the proof. □

From the above theorem and its proof, we collect at once the following key relationships.

**Corollary 3.2.8.** *Let $F \in \mathcal{W}$ with topology $E$. Then*

$$\partial \mathcal{G}_E = \bigsqcup_{E' < E} \mathcal{G}_{E'}.$$

*Further for $F' \in \mathcal{W}$ with topology $E' < E$, $\phi_{E'}(\lambda') = \phi(F') = \bar{\phi}_E(\lambda^*)$ for $\lambda' \in (0,1)^{E'}$ and $\lambda^* \in \partial([0,1]^E)$, the following hold:*

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

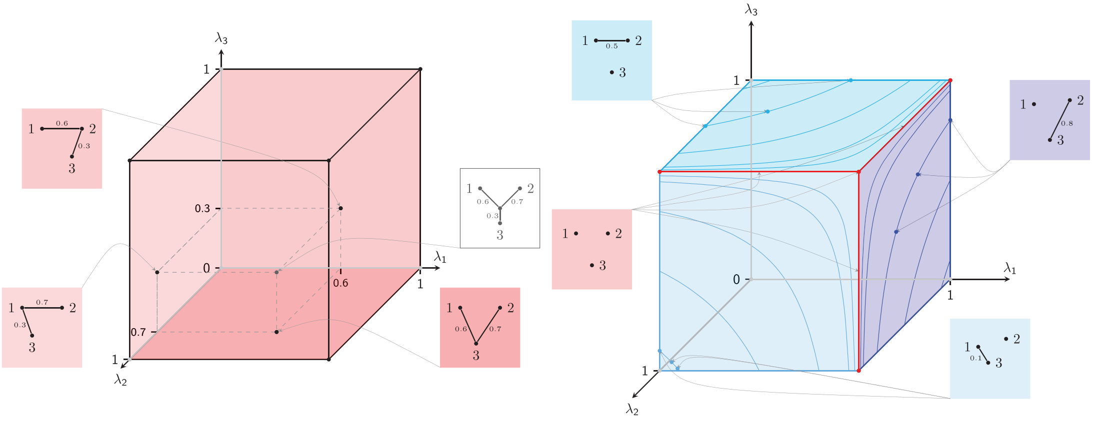


**FIGURE 9** Depicting the grove $\mathcal{G}_E \cong (0,1)^3$ of a fully resolved tree with $N = 3$ leaves, and its boundary $\partial\mathcal{G}_E$, as discussed in Example 3.2.9. Left: $\mathcal{G}_E$ and its 2-dimensional 'boundary at zero' (coordinate axes are excluded). Right: the 'boundary at one' comprising the 1-dimensional component (points on same blue curves represent a single wald) and 0-dimensional component (points on the red spider) represent $F_\infty$.

(1) *for each* $1 \leqslant \alpha' \leqslant K'$, *we have that*

$$E'_{\alpha'} = \left\{ e|_{L'_{\alpha'}} : e \in E \textit{such that} e|_{L'_{\alpha'}} \textit{ is a valid split and } \lambda^*_e \neq 0 \right\},$$

(2) *for any* $e' \in E'$,

$$\lambda'_{e'} = 1 - \bigcup_{e \in R_{e'}} (1 - \lambda^*_e).$$

**Example 3.2.9.** Let $N = 3$, so $L = \{1, 2, 3\}$ and let

$$E = \{e_1 = 1|23,\ e_2 = 2|13,\ e_3 = 3|12\}.$$

Abbreviating $\lambda_{e_i} = \lambda_i$ for $i = 1, 2, 3$ we have $\mathcal{G}_E \cong (0,1)^3$, cf. Equation (2.6), and the map $\bar{\phi}_E : [0,1]^3 \to \mathcal{S}$ from Equation (3.5) has the form

$$\bar{\phi}_E(\lambda) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} = \begin{pmatrix} 1 & (1-\lambda_1)(1-\lambda_2) & (1-\lambda_1)(1-\lambda_3) \\ (1-\lambda_1)(1-\lambda_2) & 1 & (1-\lambda_2)(1-\lambda_3) \\ (1-\lambda_1)(1-\lambda_3) & (1-\lambda_2)(1-\lambda_3) & 1 \end{pmatrix}.$$

One can easily see from Lemma 3.1.10 that $\bar{\phi}_E(\lambda) \in \mathcal{P}$ if and only if at most one coordinate of $\lambda$ is zero, otherwise there would exist $u, v \in L$ with $u \neq v$ such that $\rho_{uv} = 1$. This means that the origin and the intersections of the coordinate axes with the cube $[0,1]^3$ are not mapped into $\phi(\mathcal{W})$ and thus not part of the boundary of $\mathcal{G}_E$. The cube and the corresponding wälder $F' \in \mathcal{W}$ of the grove's boundaries (i.e., there exists $\lambda \in \partial([0,1]^E)$ with $\bar{\phi}_E(\lambda) = \phi(F')$) are depicted in Figure 9.

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Note that for the boundaries where at least one $\lambda$ coordinate is one, infinitely many coordinates give the same wald: let $F' = (\{2|3\}, \lambda')$ with $\lambda'_{2|3} = 0.8$, then all coordinates $\lambda^* = (1, \lambda_2^*, \lambda_3^*) \in \partial([0,1]^E)$ that satisfy $1 - (1 - \lambda_2^*)(1 - \lambda_3^*) = 0.8$ give $\bar{\phi}_E(\lambda^*) = \phi(F')$. This is also illustrated in Figure 9 (right panel), where several arrows point to the coordinates on curves that correspond the same wald. This means that a 2-dimensional boundary of the cube collapses into a 1-dimensional grove.

If at least two coordinates of $\lambda^*$ are equal to 1, then the corresponding phylogenetic forest will be the forest consisting of three isolated vertices, and in this case, four points as well as the three segments where two coordinates are 1 and one is strictly between zero and one on the boundary of the cube collapse to only one point in $\mathcal{W}$, marked red in Figure 9 (right panel).

**Corollary 3.2.10.** *Let $F, F' \in \mathcal{W}$ with topologies $E, E'$, respectively, and let $(F_n)_{n\in\mathbb{N}} \subset \mathcal{G}_E \subset \mathcal{W}$ be a sequence. If $F_n \to F'$, then $E' \leqslant E$.*

*Proof.* Let $\lambda^{(n)} \in (0,1)^E \cong \mathcal{G}_E$ such that $\phi_E(\lambda^{(n)}) = \phi(F_n)$ for all $n \in \mathbb{N}$. With the same argument as in the proof of Theorem 3.1.12, there exists at least one subsequence $(\lambda^{(n_k)})_{k\in\mathbb{N}}$ such that $\lambda^{(n_k)} \to \lambda^* \in [0,1]^E$ with $\bar{\phi}_E(\lambda^*) = \phi(F') \in \mathcal{P}$, so either $F' \in \mathcal{G}_E$, then $E' = E$, or $F' \in \partial\mathcal{G}_E$ (by definition of $\partial\mathcal{G}_E$ from Equation 3.9), then by Theorem 3.2.7 it follows that $E' < E$, so in general $E' \leqslant E$. □

## 3.3 | Whitney stratification of wald space

Recall from Subsection 1.3 the differentiable manifold of strictly positive definite matrices $\mathcal{P}$, and that the tangent space $T_P\mathcal{P}$ at $P \in \mathcal{P}$ is isomorphic to the vector space of symmetric matrices $\mathcal{S}$. To study convergence of linear subspaces of $\mathcal{S}$, we recall the *Grassmannian manifold* of $k$-dimensional linear subspaces in $\mathbb{R}^m$, $0 \leqslant k \leqslant m$, see, for example, [25, chapter 7].

Every $k$-dimensional linear subspace $\mathcal{V}$ of $\mathbb{R}^m$ is the span of the columns $v_1, \dots, v_k$ of a matrix $(v_1, \dots, v_k) = V \in S(m,k)$, the *column space*,

$$\mathcal{V} = \mathrm{span}\{v_1, \dots, v_k\} = \mathrm{col}(V),$$

where

$$S(m,k) = \{V \in \mathbb{R}^{m\times k} \,:\, \mathrm{rank}(V) = k\}$$

is the *Stiefel manifold* of maximal rank $(m \times k)$-matrices equipped with the smooth manifold structure inherited from embedding in the Euclidean $\mathbb{R}^{m\times k}$. As $\mathrm{col}(V) = \mathrm{col}(VG)$ for every $G \in S(k,k)$ and $V \in S(k,m)$, the space

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

can be identified with the *Grassmannian*

$$G(m,k) := S(m,k)/S(k,k)\,.$$

As every orbit $\{VG : G \in S(k,k)\}$ of $V \in S(m,k)$ is closed in $S(m,k)$ and as for every $V \in S(m,k)$ its *isotropy group* $\{G \in S(k,k) : VG = V\}$ contains the unit matrix only, the quotient carries a canonical smooth manifold structure.

**Definition 3.3.1.** With the above notation, a sequence of $k$-dimensional linear subspaces $\mathcal{V}_n$, $n \in \mathbb{N}$, of $\mathbb{R}^m$, $1 \leqslant k < m$, *converges in the Grassmannian* $G(m,k)$ to a $k$-dimensional linear subspace $\mathcal{V}$ if there are $V_n, V \in S(m,k)$ and $G_n \in S(k,k)$ such that

$$\operatorname{col}(V_n) = \mathcal{V}_n \text{ for all } n \in \mathbb{N},\ \operatorname{col}(V) = \mathcal{V} \text{ and } \|V_n G_n - V\| \to 0 \text{ as } n \to \infty\,.$$

*Remark* 3.3.2.

(1) Note that none of the cluster points of $G_n$ or $G_n/\|G_n\|$ can be singular, hence they are all in $S(k,k)$
(2) There may be, however, a sequence $V_n \in S(m,k)$ and $V \in S(m,k)$, $W \in \mathbb{R}^{m\times k} \setminus S(m,k)$ with

$$\operatorname{col}(V_n) = \mathcal{V}_n \to \mathcal{V} = \operatorname{col}(V)$$

in the Grassmanian $G(m,k)$ but

$$\|V_n - W\| \to 0$$

in $\mathbb{R}^{m\times k}$. Nevertheless, we have the following relationship.

**Lemma 3.3.3.** *Let* $V_n \in S(m,k)$ *and assume that the two limits below exist. Then*

$$\operatorname{col}\Big(\lim_{n\to\infty} V_n\Big) \subseteq \lim_{n\to\infty} \operatorname{col}(V_n)\,.$$

*Proof.* Let $v \in \mathbb{R}^m$ with $v \perp \lim_{n\to\infty} \operatorname{col}(V_n)$. Then the assertion follows, once we show $v \perp W$ with $W = \lim_{n\to\infty} V_n$.

By hypothesis, for every $\epsilon > 0$ there are $N \in \mathbb{N}$ and $G_n \in S(k,k)$ such that

$$|v^T V_n G_n| < \epsilon\ \forall n > N\,.$$

Let us first assume that there is a subsequence $n_k$ with $\|G_{n_k}\| > 1$. Then

$$|v^T V_{n_k} R_{n_k}| < \frac{\epsilon}{\|G_{n_k}\|} < \epsilon\ \forall n_k > N\,,$$

where $R_{n_k} = \frac{G_{n_k}}{\|G_{n_k}\|} \in S(k,k)$ is of unit norm, hence it has a cluster point $R$ satisfying $|v^T W R| \leqslant \epsilon$. As $\epsilon > 0$ was arbitrary, we have $v^T W R = 0$. As $R \in S(k,k)$ by Remark 3.3.2 we have thus $v^T W = 0$ as asserted.

If there is no such subsequence, without loss of generality we may assume $\|G_n\| \leqslant 1$ for all $n \geqslant N$. Again, $G_n$ has a cluster point $R$ and thus $|v^T W R| \leqslant \epsilon$ which implies, as above, $v^T W R = 0$. As $R \in S(k,k)$ by Remark 3.3.2 we have $v^T W = 0$ as asserted. □

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In the following, recall the definition of a Whitney stratified space of type (A) and (B), respectively, taken from the wording of Huckemann and Eltzner [21, section 10.6].

**Definition 3.3.4.** A *stratified space* $\mathcal{S}$ of dimension $m$ embedded in a Euclidean space (possibly of higher dimension $M \geqslant m$) is a direct sum

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

such that $0 \leqslant d_1 < \ldots < d_k = m$, 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 $q_1, q_2, \dots \in S_j$ that converges to some point $p \in S_i$, such that the sequence of tangent spaces $T_{q_n} S_j$ converges in the Grassmannian $G(M, d_j)$ to a $d_j$-dimensional linear space $T$ as $n \to \infty$, then $T_p S_i \subseteq T$, 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 $p_1, p_2, \dots \in S_i$ and $q_1, q_2, \dots \in S_j$ 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 the Grassmannian $G(M, 1)$), 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 the Grassmannian $G(M, d_j)$), then $c \subset T$.

**Theorem 3.3.5.** *Wald space with the smooth structure on every grove $\mathcal{G}_E$ conveyed by $\phi_E$ from (3.4), is a Whitney stratified space of type (A).*

*Proof.* First, we show that $\mathcal{W}$ is a stratified space. In conjunction with Remark 2.2.4, the manifolds $S_i$ of dimension $d_i = i$ are the unions over disjoint groves of $\mathcal{W}$ of equal dimension $i = 0, \dots, 2N - 3 = m$, counting the number of edges, each diffeomorphic to an $i$-dimensional open unit cube,

$$S_i = \bigsqcup_{|E|=i} \mathcal{G}_E .$$

If $S_i \cap \overline{S_j} \neq \emptyset$ for some $0 \leqslant i \neq j \leqslant m$ then there are wald topologies $E, E'$ with $j = |E|, i = |E'|$ and $\mathcal{G}_{E'} \cap \in \overline{\mathcal{G}_E} \neq \emptyset$, implying $\mathcal{G}_{E'} \subset \overline{\mathcal{G}_E}$ by Theorem 3.2.7. In particular, then $i < j$. Further, if $\widetilde{E}'$ with $i = |E'|$ is any other wald topology, induction on Lemma 3.2.6(xiii) shows that it can be extended to a wald topology $\widetilde{E}$ with $j = |E|$ such that $\widetilde{E}' < \widetilde{E}$ and hence $\mathcal{G}_{\widetilde{E}'} \subset \overline{\mathcal{G}_{\widetilde{E}'}}$ by Theorem 3.2.7. Thus, we have shown that $S_i \subset \overline{S_j}$, as required.

To show Whitney condition (A), it suffices to assume $i \neq j$. Let $F_1, F_2, \dots \in S_j$ be a sequence of wälder that converges to some wald $F' = (E', \lambda') \in S_i$, so $i < j$. As $S_j$ is a disjoint union of finitely many groves, without loss of generality we may assume that $F_1, F_2, \dots \in \mathcal{G}_E$ for some wald topology $E > E'$ with $|E| = j$. Hence, under the hypothesis that $T_{F_n} \mathcal{G}_E \cong T_{\Phi(F_n)} \Phi_E(\mathcal{G}_E) \subset \mathcal{S}$ converges in the Grassmannian $G(\dim(\mathcal{S}), j)$, to a $j$-dimensional linear space $T \subset \mathcal{S}$ as $n \to \infty$, we need to

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

show that

$$T_{F'}\mathcal{G}_{E'} \cong T_{\Phi(F')}\Phi_{E'}(\mathcal{G}_{E'}) \subseteq T\,. \tag{3.15}$$

With the analytic continuation $\bar{\phi}_E$ of $\phi_E$, see Remark 3.1.8, a cluster point $\lambda^* \in [0,1]^E$ of $\lambda^{(n)} = \phi_E^{-1}(F_n) \in (0,1)^E$, $F' = \phi^{-1}\circ\bar{\phi}_E(\lambda^*)$, see Theorem 3.1.12, and the unit standard basis $\partial/\partial\lambda_e$, $e \in E$ of $\mathcal{G}_E \cong (0,1)^E$ we have thus

$$T_{\Phi(F_n)}\Phi_E(\mathcal{G}_E) = \operatorname{span}\left\{\frac{\partial\bar{\phi}_E}{\partial\lambda_e}\left(\lambda^{(n)}\right):\, e\in E\right\}$$

and, due to Lemma 3.3.3,

$$\begin{aligned}\operatorname{span}\left\{\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*):\, e\in E\right\} &= \operatorname{span}\left\{\lim_{n\to\infty}\frac{\partial\bar{\phi}_E}{\partial\lambda_e}\left(\lambda^{(n)}\right):\, e\in E\right\}\\ &\subseteq \lim_{n\to\infty}\operatorname{span}\left\{\frac{\partial\bar{\phi}_E}{\partial\lambda_e}\left(\lambda^{(n)}\right):\, e\in E\right\} = T.\end{aligned}$$

As likewise

$$T_{\phi(F')}\phi(\mathcal{G}_{E'}) = \operatorname{span}\left\{\frac{\partial\phi_{E'}}{\partial\lambda'_{e'}}(\lambda'):\, e'\in E'\right\}$$

showing assertion (3.15) is equivalent to showing

$$\operatorname{span}\left\{\frac{\partial\phi_{E'}}{\partial\lambda'_{e'}}(\lambda'):\, e'\in E'\right\} \subseteq \operatorname{span}\left\{\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*):\, e\in E\right\}.$$

To see this, it suffices to show that for each $e' \in E'$, there exists a constant $c > 0$ and an edge $e \in E$ such that

$$\frac{\partial\phi_{E'}}{\partial\lambda'_{e'}}(\lambda') = c\,\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*). \tag{3.16}$$

In the following, we show (3.16).

Recalling for $u, v \in L$

$$\left(\bar{\phi}_E(\lambda^*)\right)_{uv} = \prod_{e\in E(u,v)}\left(1-\lambda^*_e\right),\quad \left(\phi_{E'}(\lambda')\right)_{uv} = \prod_{e'\in E'(u,v)}\left(1-\lambda'_{e'}\right)$$

from Definition 3.1.7, obtain their derivatives

$$\left(\frac{\partial\bar{\phi}_E}{\partial\lambda_e}(\lambda^*)\right)_{uv} = -\mathbf{1}_{e\in E(u,v)}\prod_{\substack{\tilde{e}\in E(u,v)\\ \tilde{e}\neq e}}\left(1-\lambda^*_{\tilde{e}}\right), \tag{3.17}$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv} = -\mathbf{1}_{e'\in E'(u,v)} \prod_{\substack{\tilde{e}'\in E'(u,v)\\ \tilde{e}'\neq e'}} \left(1-\lambda'_{\tilde{e}'}\right). \tag{3.18}$$

Recall from Corollary 3.2.8 the two relationships between $F'$ and $\bar{\phi}_E(\lambda^*)$:

$$E'_{\alpha'} = \left\{e' : e' = e|_{L'_{\alpha'}} \text{ is a valid split of } L'_{\alpha'}, e \in E \text{ and } \lambda^*_e \neq 0\right\}$$

as well as for each $e' \in E'$,

$$\lambda'_{e'} = 1 - \prod_{e\in R_{e'}} (1-\lambda^*_e) \neq 0. \tag{3.19}$$

Consequently, for any $e' \in E'_{\alpha'}$ there exists $e \in R_{e'}$ with $\lambda^*_e \neq 0$.

Now, let $u, v \in L$ be arbitrary and for every $e' \in E'$, we consider $e$ as above.

(1) Case $e \notin E(u,v)$. Then $(\bar{\phi}_E(\lambda^*))_{uv}$ is constant as $\lambda_e$ varies and as $R_{e'} \ni e \notin E(u,v)$, that is, $R_{e'} \nsubseteq E(u,v)$, we have $e' \notin E'(u,v)$ by Lemma 3.2.6 (viii) so that likewise $(\phi_{E'}(\lambda'))_{uv}$ is constant as $\lambda'_{e'}$ varies, yielding

$$\left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv} = 0 = \left(\frac{\partial \bar{\phi}_E}{\partial \lambda_e}(\lambda^*)\right)_{uv}.$$

Thus, for $c$ in (3.16) any positive constant can be chosen.

(2) Case $e = A|B \in E(u,v)$. Without loss of generality, assume that $u \in A$ and $v \in B$. Then there are two subcases:

(a) $e' \notin E'(u,v)$. On the one hand, as above this implies $\left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv} = 0$, on the other hand, as $e' = A\cap L'_{\alpha'}|B \cap L'_{\alpha'} \notin E'(u,v)$, either $u \notin L'_{\alpha'}$ or $v \notin L'_{\alpha'}$, implying

$$0 = \left(\phi_{E'}(\lambda')\right)_{uv} = \left(\bar{\phi}_E(\lambda^*)\right)_{uv} = \prod_{\tilde{e}\in E(u,v)} \left(1-\lambda^*_{\tilde{e}}\right).$$

Thus, $\lambda^*_{\tilde{e}} = 1$ for some $\tilde{e} \in E(u,v)$ with $\tilde{e} \neq e$ (recall that $\lambda^*_e < 1$ for otherwise $\lambda'_{e'} = 1$ by (3.19)), which implies in conjunction with (3.17) that

$$\left(\frac{\partial \bar{\phi}_E}{\partial \lambda_e}(\lambda^*)\right)_{uv} = 0 = \left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv}.$$

Again, for $c$ in (3.16)) any positive constant can be chosen.

(b) $e' \in E'(u,v)$. Then Lemma 3.2.6(viii) yields $R_{e'} \subseteq E(u,v)$ and we have, invoking (3.18) as well as (3.19), that

$$\left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv} = -\prod_{\substack{\tilde{e}'\in E'(u,v)\\ \tilde{e}'\neq e'}} \left(1-\lambda'_{\tilde{e}'}\right) = -\prod_{\substack{\tilde{e}'\in E'(u,v)\\ \tilde{e}'\neq e'}} \prod_{\tilde{e}\in R_{\tilde{e}'}} \left(1-\lambda^*_{\tilde{e}}\right)$$

$$= -\prod_{\substack{\tilde{e}\in E(u,v)\\ \tilde{e}\notin R_{e'}}} \left(1-\lambda^*_{\tilde{e}}\right), \tag{3.20}$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where the last equality follows from observing that $\tilde{e} \notin R_{e'} \Leftrightarrow \exists \tilde{e}' \in E', \tilde{e}' \neq e'$ such that $\tilde{e} \in R_{\tilde{e}'}$, due to Lemma 3.2.6(vi). Furthermore, again by (3.17), recalling from above that $R_{e'} \subseteq E(u, v)$ and (3.20),

$$\left(\frac{\partial \bar{\phi}_E}{\partial \lambda_e}(\lambda^*)\right)_{uv} = - \prod_{\substack{\tilde{e}\in E(u,v)\\ \tilde{e}\neq e}} \left(1-\lambda^*_{\tilde{e}}\right)$$

$$= -\left(\prod_{\substack{\tilde{e}\in R_{e'}\\ \tilde{e}\neq e}} \left(1-\lambda^*_{\tilde{e}}\right)\right)\left(\prod_{\substack{\tilde{e}\in E(u,v)\\ \tilde{e}\notin R_{e'}}} \left(1-\lambda^*_{\tilde{e}}\right)\right)$$

$$= \left(\prod_{\substack{\tilde{e}\in R_{e'}\\ \tilde{e}\neq e}} \left(1-\lambda^*_{\tilde{e}}\right)\right)\left(\frac{\partial \phi_{E'}}{\partial \lambda'_{e'}}(\lambda')\right)_{uv}.$$

Thus,

$$c = \prod_{\substack{\tilde{e}\in R_{e'}\\ \tilde{e}\neq e}} \left(1-\lambda^*_{\tilde{e}}\right)$$

satisfies (3.16) as it does not depend on $u$ and $v$ and is non-zero by Equation (3.19).

Having thus shown (3.16), as detailed above we have established (3.15) thus verifying Whitney condition (A). □

Whitney condition (B) is a conjecture.

## 4 | INFORMATION GEOMETRY FOR WALD SPACE

In [16], we equipped the space of phylogenetic forests with a metric induced from the metric of the *Fisher-information* Riemannian metric $g$ on $\mathcal{P}$ (see Subsection 1.3), where the latter induces the metric $d_{\mathcal{P}}$ on $\mathcal{P}$. In this section we show, first that this induced metric is compatible with the stratification structure of $\mathcal{W}$, and second that this turns $\mathcal{W}$ into a geodesic Riemann stratified space.

### 4.1 | Induced intrinsic metric

In [16], we introduced a metric on $\mathcal{W}$ induced from the geodesic distance metric $d_{\mathcal{P}}$ of $\mathcal{P}$ introduced in Subsection 1.3. Recalling also the definition of path length $L_{\mathcal{P}}$ from Subsection 1.3, for two wälder $F, F' \in \mathcal{W}$, set

$$d_{\mathcal{W}}(F, F') := \inf_{\substack{\gamma:\,[0,1]\to\mathcal{W}\\ \phi\circ\gamma \text{ continuous in } \mathcal{P},\\ \gamma(0)=F,\gamma(1)=F'}} L_{\mathcal{P}}(\phi\circ\gamma).$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This metric defines the *induced intrinsic metric* topology on $\mathcal{W}$. Although in general this topology may be finer than the one conveyed by making an embedding a homeomorphism, as the following example teaches, this is not the case for wald space.

**Example 4.1.1.** Consider

$$\mathcal{M} := (\{1\} \times [0,1]) \quad \cup \bigcup_{y \in \{1/n : n \in \mathbb{N}\} \cup \{0\}} [-1,1) \times \{y\},$$

an infinite union of half open intervals in $\mathbb{R}^2$ connected vertically on the right. In the trace topology where the canonical embedding $\iota : \mathcal{M} \hookrightarrow \mathbb{R}^2$ is a homeomorphism, the sequence $q_n = (0, 1/n)$ converges to $q = (0,0)$. For the induced intrinsic metric

$$d_{\mathcal{M}}(x,y) = \inf_{\substack{\gamma:\ [0,1] \to \mathcal{M} \\ \gamma \text{ continuous in } \mathbb{R}^2, \\ \gamma(0)=x, \gamma(1)=y}} L_{\mathbb{R}^2}(\gamma),$$

with the Euclidean length $L_{\mathbb{R}^2}$, we have, however, $d_W(q_n, q) \geqslant 2$ for all $n \in \mathbb{N}$.

**Theorem 4.1.2.** *The topology of $\mathcal{W}$ obtained from making $\phi$ a homeomorphism agrees with the topology induced from the induced intrinsic metric $d_{\mathcal{W}}$. In particular, $d_{\mathcal{W}}$ turns $\mathcal{W}$ into a metric space.*

*Proof.* By definition, we have that $d_{\mathcal{W}} \geqslant d_{\mathcal{P}}$, which implies that sequences that converge with respect to $d_{\mathcal{W}}$ also converge with respect to $d_{\mathcal{P}}$.

For the converse, assume that $\mathcal{W} \ni F_n \to F' \in \mathcal{W}$ with respect to $d_{\mathcal{P}}$, as $n \to \infty$. As there are only finitely many groves in $\mathcal{W}$ it suffices to show that $d_{\mathcal{W}}(F_n, F) \to 0$ for $F_n \in \mathcal{G}_E$ and $F \in \overline{\mathcal{G}_E}$ with a common grove $\mathcal{G}_E$. Hence, we assume that $\bar{\phi}_E^{-1} \circ \phi(F_n) = \lambda_n \in (0,1)^E$ and $\bar{\phi}_E^{-1} \circ \phi(F') = \lambda' \in [0,1]^E$ with $\lambda_n \to \lambda'$, due to Theorem 3.1.12. Then, with $\delta(t) = t\lambda' + (1-t)\lambda_n$,

$$\gamma : [0,1] \to \mathcal{P}, t \mapsto \phi^{-1} \circ \bar{\phi}_E \circ \delta(t)) =: \gamma(t)$$

is a path in $\mathcal{W}$ connecting $\gamma(0) = F$ with $\gamma(1) = F_n$. For $k \in \mathbb{N}$ and $j = 1, \dots, k$ we note that

$$\bar{\phi}_E \circ \delta_n\left(\frac{j}{k}\right) = \bar{\phi}_E \circ \delta_n\left(\frac{j-1}{k}\right) + (D\bar{\phi}_E) \circ \delta_n\left(\frac{j-1}{k}\right) \cdot \frac{\lambda' - \lambda_n}{k} + o\left(\frac{\|\lambda' - \lambda_n\|}{k}\right)$$

where both terms

$$\bar{\phi}_E \circ \delta_n\left(\frac{j-1}{k}\right), (D\bar{\phi}_E) \circ \delta_n\left(\frac{j-1}{k}\right)$$

are bounded, also uniformly $n \in \mathbb{N}$, due to Remark 3.1.8. In consequence, in conjunction with Subsection 1.3,

$$\begin{aligned} &d_{\mathcal{W}}(F_n, F) \\ &\leqslant \lim_{k \to \infty} \sum_{j=1}^{k} d_{\mathcal{P}}\left(\bar{\phi}_E \circ \delta\left(\frac{j-1}{k}\right), \bar{\phi}_E \circ \delta\left(\frac{j}{k}\right)\right) \end{aligned}$$

$$= \lim_{k\to\infty} \sum_{j=1}^{k} \left\| \log\left( \sqrt{\bar{\phi}_E \circ \delta\left(\frac{j-1}{k}\right)}^{-1} \bar{\phi}_E \circ \delta_n\left(\frac{j}{k}\right) \sqrt{\bar{\phi}_E \circ \delta_n\left(\frac{j-1}{k}\right)}^{-1} \right) \right\|$$

$$\leqslant C\|\lambda' - \lambda_n\|$$

with a constant $C > 0$ independent of $n$. Letting $n \to \infty$ thus yields the assertion. □

## 4.2 | Geodesic space and Riemann stratification

Having established the equivalence between the stratification topology and that of the Fisher information metric, we longer distinguish between them.

**Theorem 4.2.1.** *The wald space equipped with the information geometry is a geodesic metric space, that is, every two points in $(\mathcal{W}, d_{\mathcal{W}})$ are connected by a minimising geodesic.*

*Proof.* By [24, p. 325], $(\mathcal{P}, g)$ is geodesically complete as a Riemannian manifold and thus by the Hopf–Rinow theorem for Riemannian manifolds (among others, [24, p. 224]), it follows that $(\mathcal{P}, d_{\mathcal{P}})$ is complete and locally compact. By Corollary 3.1.3, $\phi(\mathcal{W})$ is a closed subset of the complete and locally compact metric space $\mathcal{P}$ and so $(\phi(\mathcal{W}), d_{\mathcal{P}})$ itself is, and so is $(\mathcal{W}, d_{\mathcal{P}})$. By [16, Theorem 5.1], any two wälder in are connected by a continuous path of finite length in $(\mathcal{W}, d_{\mathcal{P}})$, which is complete, and thus applying [20, Corollary, p. 123] yields that $(\mathcal{W}, d_{\mathcal{W}})$ is complete. Applying the Hopf–Rinow theorem for metric spaces [8, p. 35] to $(\mathcal{W}, d_{\mathcal{W}})$, the assertion holds. □

Following Huckemann and Eltzner [21, section 10.6], extend the notion of a Whitney stratified space in Definition 3.3.4 to the notion of a Riemann stratified space.

**Definition 4.2.2.** A *Riemann stratified space* is a Whitney stratified space $\mathcal{S}$ of type (A) such that each stratum $S_i$ is a $d_i$-dimensional Riemannian manifold with Riemannian metric $g^i$, respectively, if whenever a sequence $q_1, q_2, \dots \in S_j$ which converges to a point $p \in S_i$ (where, assume again that the sequence of tangent planes $T_{q_n} S_j$ converges to some $d_j$-dimensional plane $T$ as $n \to \infty$), then the Riemannian metric $g^j_{q_n}$ converges to some two form $g^*_p : T \otimes T \to \mathbb{R}$ with $g^i_p \equiv g^*_p|_{T_p S_i \otimes T_p S_i}$.

**Theorem 4.2.3.** *The wald space $\mathcal{W}$ equipped with the information geometry is a Riemann stratified space.*

*Proof.* As we impose the Riemannian metric $g$ from $\mathcal{P}$ onto all of $\phi(\mathcal{W}) \subset \mathcal{P}$, the assertion follows immediately. □

**Example 4.2.4** (Geometry of wald space for $N = 2$). For $N = 2$, $L = \{1, 2\}$, there is one edge $e = 1|2$, and two different topologies, namely

$$E = \{1|2\} \qquad \text{and} \qquad E' = \emptyset.$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

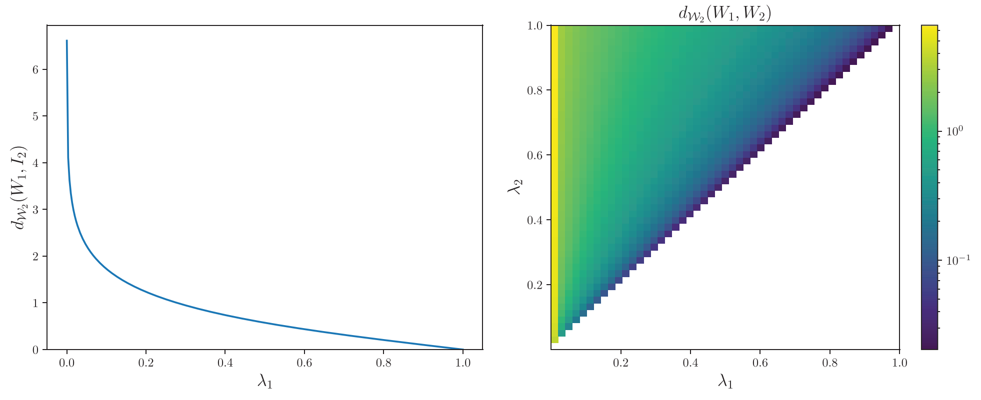


**FIGURE 10** Depicting the distance $d_{\mathcal{W}}$ of an arbitrary wald, to the disconnected forest $F_\infty$ as a function of $\lambda_1$ (left) and between any two other wälder in $\mathcal{W}$, as a function of $(\lambda_1, \lambda_2)$ (right) as detailed in Example 4.2.4.

The corresponding groves are then $\phi(\mathcal{G}_{E'}) = \{I\}$, where $I$ is the $2 \times 2$ unit matrix, and $\mathcal{G}_E \cong (0,1)$ such that $\mathcal{W} = \mathcal{G}_E \sqcup \mathcal{G}_{E'} \cong (0,1]$. Using $\lambda \in (0,1)$ for the only edge $e = 1|2$, we have

$$\phi(\mathcal{G}_E) = \left\{ \phi_E(\lambda) = \begin{pmatrix} 1 & 1-\lambda \\ 1-\lambda & 1 \end{pmatrix} \,:\, \lambda \in (0,1) \right\},$$

Thus, with the definition of $d_{\mathcal{P}}$ in Subsection 1.3, the distance between two phylogenetic forests $F_1 = \bar{\phi}_E(\lambda_1)$, $F_2 = \bar{\phi}_E(\lambda_2)$ with $\lambda_1, \lambda_2 \in (0,1]$ can be calculated as

$$d_{\mathcal{W}}(F_1, F_2) = \left| \ln\left( \frac{1-\lambda_2 + \frac{1}{\sqrt{2}} p(\lambda_2)}{1-\lambda_1 + \frac{1}{\sqrt{2}} p(\lambda_1)} \right) + \frac{1}{2\sqrt{2}} \ln\left( \frac{(p(\lambda_1) + (1-\lambda_1))^2 - 1}{(p(\lambda_1) - (1-\lambda_1))^2 - 1} \cdot \frac{(p(\lambda_2) - (1-\lambda_2))^2 - 1}{(p(\lambda_2) + (1-\lambda_2))^2 - 1} \right) \right|,$$

where $p(x) = \sqrt{2}\sqrt{(1-x)^2 + 1}$ for $x \in [0,1]$. Figure 10 (right) depicts the distance as a function of $\lambda_1, \lambda_2$. We obtain the distance to the disconnected forest $F_\infty = \phi^{-1}\bar{\phi}_E(1)$:

$$d_{\mathcal{W}}(F_1, F_\infty) = \left| \frac{1}{2\sqrt{2}} \ln\left( \frac{(p(\lambda_1) + (1-\lambda_1))^2 - 1}{(p(\lambda_1) - (1-\lambda_1))^2 - 1} \right) - \ln\left( 1 - \lambda_1 + \frac{1}{\sqrt{2}} p(\lambda_1) \right) \right|.$$

This distance is depicted (as a function in $\lambda_1$) in Figure 10 (left).

## 5 | NUMERICAL EXPLORATION OF WALD SPACE

In this section, we propose a new algorithm to approximate geodesics between two fully resolved trees $F_1$ and $F_2$, that is a mixture of the *successive projection* algorithm and the *extrinsic path straightening* algorithm from [26]. Using this algorithm allows to explore curvature and so-called *stickiness* of Fréchet means.

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

## 5.1 | Approximating geodesics in wald space

From the ambient geometry of $\mathcal{P}$, recalling the notation from Subsection 1.3, we employ the globally defined Riemannian exponential Exp and logarithm Log at $P \in \mathcal{P}$, with $Q \in \mathcal{P}, X \in T_P\mathcal{P}$, as well as points on the unique (if $P \neq Q$) geodesic $\gamma_{P,Q}$ in $\mathcal{P}$ comprising $P$ and $Q$. Further, exp and log denote the matrix exponential and logarithm, respectively:

$$\begin{aligned}
\mathrm{Exp}_P &: T_P\mathcal{P} \to \mathcal{P}, && X \mapsto \sqrt{P}\exp\left(\sqrt{P}^{-1}X\sqrt{P}^{-1}\right)\sqrt{P},\\
\mathrm{Log}_P &: \mathcal{P} \to T_P\mathcal{P}, && Q \mapsto \sqrt{P}\log\left(\sqrt{P}^{-1}Q\sqrt{P}^{-1}\right)\sqrt{P},\\
\gamma^{(\mathcal{P})}_{P,Q} &: [0,1] \to \mathcal{P}, && t \mapsto \mathrm{Exp}_P\big(t\,\mathrm{Log}_P(Q)\big),
\end{aligned}$$

Furthermore, for a forest $F \in \mathcal{W}$ with topology $E$, denote the orthogonal projection from the tangent space $T_P\mathcal{P}$ at $P = \phi(F)$ onto the tangent space of the sub-manifold $T_P\phi_E(\mathcal{G}_E)$, as a subspace of $T_P\mathcal{P}$, with

$$\pi_P : T_P\mathcal{P} \to T_P\phi_E(\mathcal{G}_E).$$

This projection is computed using an orthonormal basis of $T_P\phi_E(\mathcal{G}_E)$ obtained from applying Gram–Schmidt to the basis

$$\left\{\frac{\partial\phi_E}{\partial\lambda_e}(\lambda) : e \in E\right\}$$

of $T_P\phi_E(\mathcal{G}_E)$. Finally, we make use of the projection

$$\pi : \mathcal{P} \to \mathcal{W}, \qquad P \mapsto \operatorname*{argmin}_{F\in\mathcal{W}} d_{\mathcal{P}}(P, \phi(F)),$$

where $\pi$ is well-defined for $P \in \mathcal{P}$ close enough to $\phi(\mathcal{W})$. The following algorithm is similar to the *extrinsic path straightening* algorithm from [26], which has been inspired by [38]. It starts with generating a discrete curve using the *successive projection* algorithm from [26] and then iteratively straightening it and adding more points in between the points of the discrete curve. To keep notation simple, we omit $\phi$ and identify a forest $F \in \mathcal{W}$ with its matrix representation $\phi(F)$.

**Definition 5.1.1** (Geodesic Approximation Algorithm). Let $5 \leqslant n_0 \in \mathbb{N}$ be the odd number of points in the *initial path*, $I \in \mathbb{N}$ the number of *extensions iterations* and $J \in \mathbb{N}$ the number of *straightening iterations* of the path.

**Input**: $F, F' \in \mathcal{W}$
**Initial path**: Set $F_1 := F, F_{n_0} := F'$, then, for $i = 2, \dots, (n_0 - 3)/2$, compute

$$\begin{aligned}
F_i &= \pi\left(\gamma_{F_{i-1},F_{n_0-i+2}}\left(\frac{1}{n_0 - 2i + 3}\right)\right),\\
F_{n_0-i+1} &= \pi\left(\gamma_{F_{n_0-i+2},F_{i-1}}\left(\frac{1}{n_0 - 2i + 3}\right)\right),
\end{aligned}$$

and, with $F_{(n_0-1)/2} := \pi(\gamma_{F_{(n_0-3)/2},F_{(n_0+1)/2}}(0.5))$, set the *current discrete path* to

$$\Gamma := \left(F_1, \dots, F_{(n_0-3)/2}, F_{(n_0-1)/2}, F_{(n_0+1)/2}, \dots, F_{n_0}\right).$$

**Iteratively extend and straighten**: Do $I$ times:

**Extend**: With the current discrete path $\Gamma = (F_0, \dots, F_n)$, for $i = 0, \dots, n-1$ compute $G_i := \pi\left(\gamma_{F_i,F_{i+1}}(0.5)\right)$ and define the new current discrete path

$$\Gamma := (F_0, G_0, F_1, G_1, \dots, F_{n-1}, G_{n-1}, F_n).$$

Do $J$ times:

**Straighten**: With the current discrete path $\Gamma = (F_0, \dots, F_n)$, for $i = 2, \dots, n-1$, compute $X_i = \frac{1}{2}(\mathrm{Log}_{F_i}(F_{i-1}) + \left(\mathrm{Log}_{F_i}(F_{i+1})\right)$, update $F_i := \pi\left(\mathrm{Exp}^{(\mathcal{P})}_{F_i}(X_i)\right)$, and define the new current discrete path

$$\Gamma := (F_0, F_1, \dots, F_n).$$

**Return**: The current discrete path $\Gamma$, which is a discrete approximation of the geodesic between $F$ and $F'$ with $2^I(n_0 - 1) + 1$ points.

Although Theorem 4.2.1 guarantees the existence of a shortest path between any $F, F' \in \mathcal{W}$, it may not be unique, and it is not certain whether the path found by the algorithm is near a shortest path or represents just a local approximation.

To better assess the quality of the approximation $\Gamma = (F_0, \dots, F_n)$ found by the algorithm, [37] propose considering its energy,

$$E(\Gamma) = \frac{1}{2}\sum_{i=0}^{n-1} d_{\mathcal{P}}(F_i, F_{i+1})^2,$$

yielding a means of comparison for discrete paths with equal number of points.

**Example 5.1.2** (Geodesics in wald space for $N = 3$). Revisiting $\mathcal{W}$ from Example 3.2.9 with unique top-dimensional grove $\mathcal{G}_E \cong (0,1)^3$, we approximate a shortest path between the two phylogenetic forests $F_1, F_2 \in \mathcal{W}$ with $\phi(F_1) = \phi_E(\lambda^{(1)})$ and $\phi(F_2) = \phi_E(\lambda^{(2)})$ using the algorithm from Definition 5.1.1, where

$$\lambda^{(1)} = (0.1, 0.9, 0.07) \qquad \text{and} \qquad \lambda^{(2)} = (0.3, 0.1, 0.9).$$

This path is depicted in Figure 11, as well as the BHV space geodesic (which is a straight line with respect to the $\ell$-parametrisation from Definition 2.1.2), first in the coordinates $\lambda \in (0,1)^3$ and second embedded into $\mathcal{P}$ viewed as $\mathbb{R}^3$, cf. Figure 3. In contrast to the BHV geometry, the shortest path in the wald space geometry sojourns on the 2-dimensional boundary, where the coordinate $\lambda_1$ is zero for some time. The end points $\lambda^{(1)}, \lambda^{(2)}$, are trees that show a high level of disagreement over the location of taxon 1, but a similar divergence between taxon 2 and taxon 3. The section of the approximate geodesic with $\lambda_1 = 0$ represents trees on which the overall divergence between taxon 1 and the other two taxa is reduced. In this way, the conflicting information in the end points

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

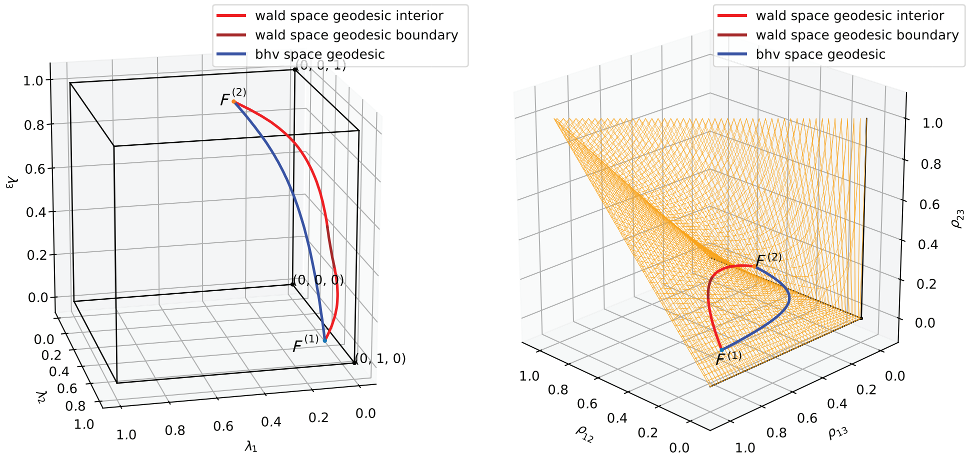


**FIGURE 11** The wald space geodesic (red) between fully resolved phylogenetic forests $F_1, F_2 \in \mathcal{W}$ $(N = 3)$ sojourns on the boundary (brown). The image of the BHV space geodesic (blue) remains in the grove as discussed in Example 5.1.2. In $\lambda$-representation (left) and embedded in $\mathcal{P}$ viewed as $\mathbb{R}^3$ (right, cf. Figure 3).

is resolved by reducing the divergence (and hence increasing the correlation) between taxon 1 and the other two taxa, in comparison to the BHV geodesic which has $\lambda_1 > 0$ along its length.

## 5.2 | Exploring curvature of wald space

As curvature computations involving higher order tensors are heavy on indices, we keep notation as simple as possible in the following by indexing splits in $E$ by

$$h, i, j, k, m, s, t \in E\,.$$

The concepts of transformation of metric tensors, Christoffel symbols and curvature employed in the following can be found in any standard text book on differential geometry, for example, [24, 25].

Recall that the Riemannian structure of wald space is inherited on each grove $\mathcal{G}_E \cong (0,1)^E$ from the information geometric Riemann structure of $\mathcal{P}$ pulled back from $\phi_E : (0,1)^E \to \mathcal{P}$. In consequence, the Riemannian metric tensor $g_\lambda^{(\mathcal{G}_E)}$ of $\mathcal{G}_E$, evaluated at $\lambda \in (0,1)^E$, is given by the Riemannian metric tensor $g_\lambda^{\mathcal{P}}$ at $\phi_E(\lambda) = P$, where base vectors transform under the derivative of $\phi_E$:

$$g_\lambda^{(\mathcal{G}_E)}(x, y) = \sum_{i\in E}\sum_{j\in E} x_i y_j\; g_P^{(\mathcal{P})}\left(\frac{\partial \phi_E}{\partial \lambda_i}(\lambda),\, \frac{\partial \phi_E}{\partial \lambda_j}(\lambda)\right)$$

for $(x, y) \in T_\lambda \mathcal{G}_E \times T_\lambda \mathcal{G}_E \cong \mathbb{R}^E \times \mathbb{R}^E$ and

$$(\mathrm{d}\phi_E)_\lambda(T_\lambda \mathcal{G}_E) = \mathrm{span}\left\{\frac{\partial \phi_E}{\partial \lambda_i}(\lambda) : i \in E\right\} \subseteq T_P \mathcal{P}.$$

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

As usual $(g_{ij})_{i,j\in E}$ denotes the matrix of $g_\lambda^{(\mathcal{G}_E)}$ in standard coordinates and $(g^{ij})_{i,j\in E}$ its inverse. This yields the *Christoffel symbols* for $i, j, m \in E$,

$$\Gamma_{ij}^m = \frac{1}{2}\sum_{k\in E}\left(\frac{\partial g_{jk}}{\partial\lambda_i} + \frac{\partial g_{ki}}{\partial\lambda_j} - \frac{\partial g_{ij}}{\partial\lambda_k}\right)g^{km},$$

which give the representation of the *curvature tensor*

$$R_{ijks} = \sum_{t\in E}\left(\sum_{h\in E}\Gamma_{ik}^h\Gamma_{jh}^t - \sum_{h\in E}\Gamma_{jk}^h\Gamma_{ih}^t + \frac{\partial}{\partial\lambda_j}\Gamma_{ik}^t - \frac{\partial}{\partial\lambda_i}\Gamma_{jk}^t\right)g_{ts}$$

in the coordinates $i, j, k, s \in E$.

Introducing the notation $(P = \phi_E(\lambda))$

$$Q_i = P^{-1}\frac{\partial\phi_E}{\partial\lambda_i}(\lambda) \qquad \text{and} \qquad Q_{ij} = P^{-1}\frac{\partial^2\phi_E}{\partial\lambda_i\partial\lambda_j}(\lambda)$$

and performing a longer calculation in coordinates $i, j \in E$, gives

$$\begin{aligned}R_{ijij} = \frac{1}{4}\sum_{a,h\in E} & g^{ah}\,\mathrm{Tr}\left[(2Q_{ij} - Q_jQ_i - Q_iQ_j)Q_a\right]\mathrm{Tr}\left[(2Q_{ij} - Q_jQ_i - Q_iQ_j)Q_h\right]\\ & - \sum_{a,h\in E} g^{ah}\,\mathrm{Tr}\left[Q_i^2Q_a\right]\mathrm{Tr}\left[Q_j^2Q_h\right]\\ & - \mathrm{Tr}\left[(2Q_{ij} - Q_jQ_i - Q_iQ_j)Q_{ij}\right].\end{aligned}$$

Evaluating the sectional curvature tensor at a pair of tangent vectors $x, y \in T_\lambda\mathcal{G}_E \cong \mathbb{R}^E$ at $\lambda$ gives the *sectional curvature* $K(x, y)$ at $\lambda$ of the local 2-dimensional subspace spanned by geodesics with initial directions generated by linear combinations of $x$ and $y$. Abbreviating $|x|_\lambda^{(\mathcal{G}_E)} := (g_\lambda^{(\mathcal{G}_E)}(x, x))^{1/2}$ we have

$$K(x, y) = \frac{\sum_{i\in E}\sum_{j\in E} x_iy_iR_{ijij}}{|x|_\lambda^{(\mathcal{G}_E)}|y|_\lambda^{(\mathcal{G}_E)} - g_\lambda^{(\mathcal{G}_E)}(x, y)}.$$

**Example 5.2.1.** Again revisiting $\mathcal{W}$ from Example 3.2.9 with $N = 3$, we first consider wälder in the unique top-dimensional grove $\mathcal{G}_E \cong (0, 1)^3$, and then on its boundary.

(1) We compute minimum and maximum sectional curvatures at the wälder $F$ with $\lambda = (a, a, a)$, for $a \in (0, 1)$, as displayed in Figure 12. Traversing along $0 < a < 1$ we find both positive and negative sectional curvatures and their extremes escape to positive and negative infinity as the vantage point, the isolated forest $F_\infty$, is approached, where all dimensions collapse.

(2) To assess *Alexandrov curvature* that measures 'fatness/slimless' of geodesic triangles (hence it does not require a Riemannian structure, a geodesic space suffices, see [41]) we compute several geodesic triangles and their respective angle sums within $\mathcal{W}$. The corners of the triangles are wälder $F_1, F_2, F_3 \in \mathcal{W}$, where $\{i, j, k\} = \{1, 2, 3\}$, $E_i := \{j|k\}$ and $\lambda_{j|k} \in (0, 1]$. Figure 13 depicts the geodesic triangles (left panel) non-isometrically embedded in $\mathbb{R}^3$ representing the off-diagonals in $\mathcal{P}$, as well as their respective angle sums (right panel). When the two

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

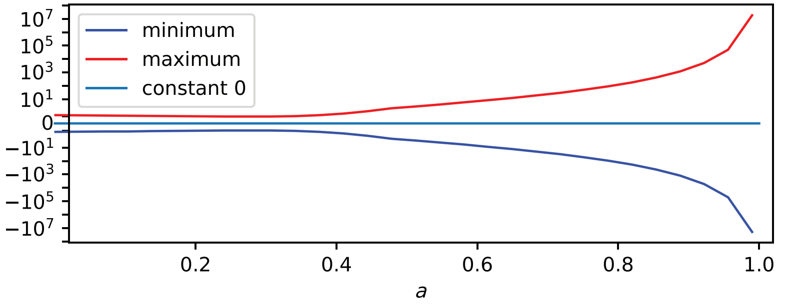


**FIGURE 12** Minimum and maximum sectional curvatures along $0 < a < 1$ of wald space ($N = 3$) at $F \in \mathcal{W}$ with $\lambda = (a, a, a)$, as described in Example 5.2.1.

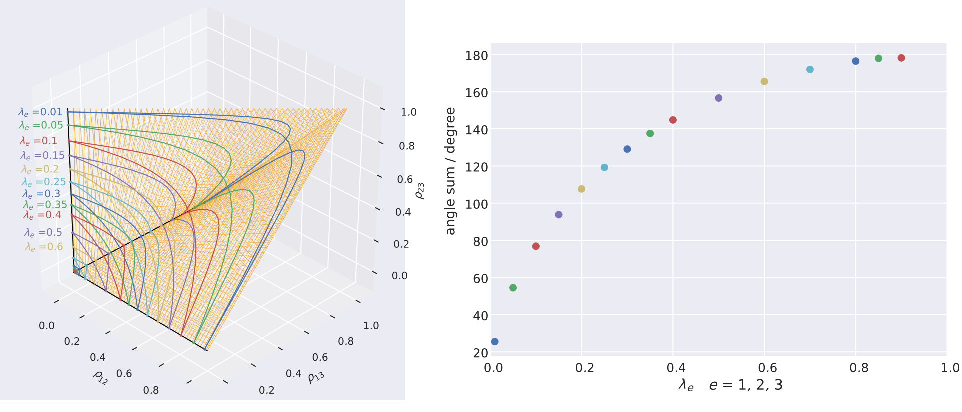


**FIGURE 13** Displaying sums of angles in degrees (right) of geodesic triangles spanned by three wälder for $N = 3$ with one disconnected leaf and edge weight $0 < \lambda_e < 1$ between the other two leaves as discussed in Example 5.2.1. Embedding $\mathcal{W}$ in $\mathcal{P}$ viewed (non-isometrically) as $\mathbb{R}^3$, the geodesic triangles are visualised on the left, where the origin corresponds to $\lambda_e = 1$.

connected leaves approach one another ($\lambda_e \approx 0$) triangles become infinitely thin, but near $F_\infty$ ($\lambda_e \approx 1$) the triangles become Euclidean.

**Conjecture 5.2.2.** *This example hints towards a general situation.*

(i) *Wald space groves feature positive and negative sectional curvatures alike, both of which become unbounded when approaching the vantage point $F_\infty$.*
(ii) *When approaching the infinitely far away boundary of $\mathcal{P}$ from within $\mathcal{W}$, some Alexandrov curvatures tend to negative infinity.*

## 5.3 | Exploring stickiness in wald space

Statistical applications in tree space often require the concept of a mean or average tree. As the expectation of a random variable taking values in a non-Euclidean metric space $(M, d)$ is not well-defined, [15] proposed to resort instead to a minimiser of expected squared distance to a random

1469750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

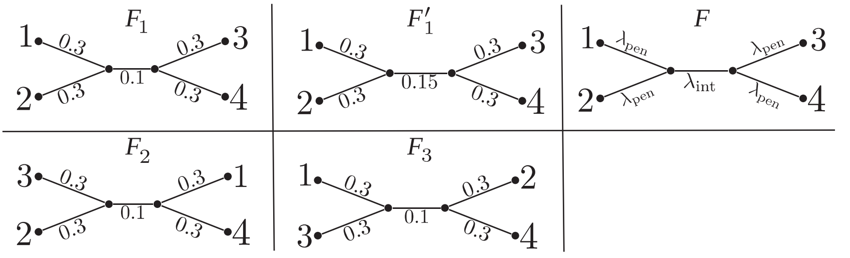


**FIGURE 14** Two samples of wälder: $F_1, F_2, F_3$ and $F_1', F_2, F_3$ where $F_1$ and $F_1'$ only differ by weights of their interior edges. By symmetry, $F$ is a candidate for each Fréchet mean, see Example 5.3.1.

element $X$ in $M$,

$$p^* \in \underset{p \in M}{\operatorname{argmin}} \, \mathbb{E}[d(p,X)^2]$$

called a *barycenter* or *Fréchet mean*. In a Euclidean geometry, if existent, the Fréchet mean is unique and identical to the expected value of $X$. Given a sample $X_1, \dots, X_n \overset{i.i.d.}{\sim} X$, measurable selections from the set

$$\underset{p \in M}{\operatorname{argmin}} \, \mathcal{F}(p), \quad \mathcal{F}(p) := \frac{1}{n} \sum_{j=1}^{n} d(p, X_j)^2$$

are called *empirical Fréchet means* and their asymptotic fluctuations allow for non-parametric statistics. Usually, $\mathcal{F}$ is called the *empirical Fréchet function*.

Recently, it has been discovered by [12, 18] that positive curvatures may increase asymptotic fluctuation by orders of magnitude, and by [19, 22] that infinite negative Alexandrov curvature may completely cancel asymptotic fluctuation, putting a dead end to this approach of non-Euclidean statistics. In particular, this can be the case for BHV spaces, cf. [3–5].

**Example 5.3.1** (Stickiness in wald space). Consider two samples $F_1, F_2, F_3 \in \mathcal{W}$ and $F_1', F_2, F_3 \in \mathcal{W}$ with $N = 4$, depicted in Figure 14, where $F_1$ and $F_1'$ only differ by weights of their interior edges. By symmetry, their Fréchet means are of form $F$ having equal but unknown pendent edge weights $0 < \lambda_{\text{pen}} < 1$ and unknown interior edge weights $0 \leqslant \lambda_{\text{int}} < 1$, as in Figure 14. It turns out that the Fréchet means of both samples agree in BHV with $\lambda_{\text{int}} = 0$, that is, the empirical mean *sticks* to the lower dimensional *star tree stratum* (featuring only pendant edges).

In contrast, the two empirical Fréchet functions in wald space

$$\mathcal{F}(F) = \frac{1}{3}\left(d_{\mathcal{W}}(F, F_1)^2 + d_{\mathcal{W}}(F, F_2)^2 + d_{\mathcal{W}}(F, F_3)^2\right)$$

$$\mathcal{F}'(F) = \frac{1}{3}\left(d_{\mathcal{W}}(F, F_1')^2 + d_{\mathcal{W}}(F, F_2)^2 + d_{\mathcal{W}}(F, F_3)^2\right)$$

have different minimisers, and, in particular the minimiser for $\mathcal{F}'$ does not stick to the star stratum but has $\lambda_{\text{int}} > 0$. Figure 15 illustrates the values of $\mathcal{F}$ and $\mathcal{F}'$ for different values of the parameters $\lambda_{\text{pen}}, \lambda_{\text{int}}$ of $F$ near the respective minima.

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

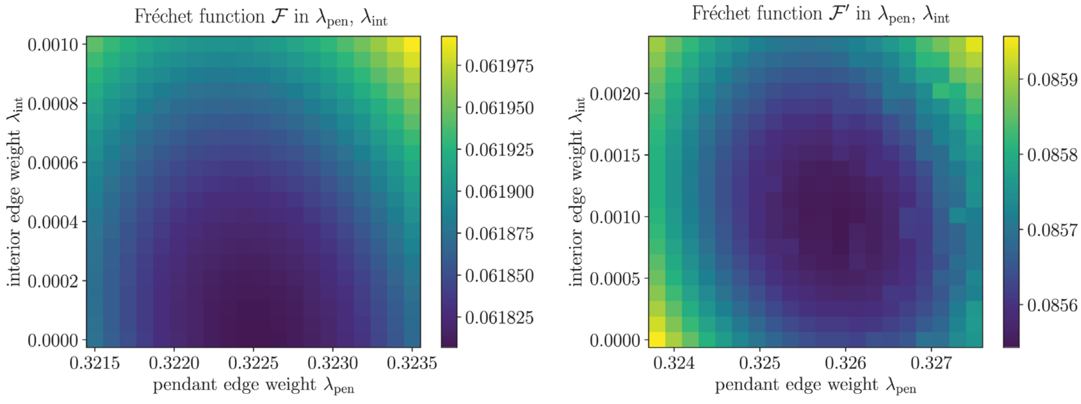


**FIGURE 15** Heat map for the values of the Fréchet functions $\mathcal{F}$ (left) and $\mathcal{F}'$ (right) of two samples as functions of $\lambda_{pen}, \lambda_{int}$ determining candidate minimisers $F$ as detailed in Example 5.3.1 and illustrated in Figure 14.

*Remark* 5.3.2. This preliminary research indicates that effects of stickiness, which are still expected where 'too many' lower dimensional strata hit higher dimensional strata, are less severe in wald space than in BHV space, thus making wald space more attractive for asymptotic statistics based on Fréchet means.

## 6 | DISCUSSION

In previous work [16], the wald space was introduced as a space for statistical analysis of phylogenetic trees, based on assumptions with a stronger biological motivation than existing spaces. In that work, the focus was primarily on geometry, whereas here we have provided a rigourous characterisation of the topology of wald space. Specifically, wald space $\mathcal{W}$ is a disjoint union of open cubes with the Euclidean toplogy, and as topological subspaces we have

$$\mathcal{BHV}_{N-1} \subset \mathcal{W} \subset \mathcal{E}_N$$

with the BHV space $\mathcal{BHV}_{N-1}$ from [7] and the edge-product space $\mathcal{E}_N$ from [31]. We have shown that this topology is the same as that induced by the information metric $d_{\mathcal{W}}$ defined in [16]. Furthermore, we have shown $\mathcal{W}$ is contractible, and so does not contain holes or handles of any kind. Examples suggest that $\mathcal{W}$ is a truncated cone in some sense (see Figure 4), but its precise formulation remains an open problem. As established in Theorem 3.3.5, boundaries between strata in wald space satisfy Whitney condition (A); whether Whitney condition (B) holds is an open problem, although we expect it to hold on the boundaries of any grove $(0,1)^E \cong \mathcal{G}_E$ corresponding to the limit as one or more coordinates $\lambda_e \to 0$ (i.e., the boundaries between strata in $\mathcal{BHV}_{N-1}$). Our key geometrical result is that with the metric $d_{\mathcal{W}}$, wald space is a geodesic metric space, Theorem 4.2.1. The existence of geodesics greatly enhances the potential of wald space as a home for statistical analysis.

The approximate geodesics computed via the algorithm in Definition 5.1.1 provide insight into the geometry and a source of conjectures. For example, unlike geodesics in BHV tree space, it appears that geodesics in wald space can run for a proportion of their length along grove

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

boundaries, even when the end points are within the interior of the same grove (see Example 5.1.2). If wald space is uniquely geodesic (so that there is a unique geodesic between any given pair of points), its potential as a home for statistical analysis would be improved further. However, the presence of positive and negative sectional curvatures for different pairs of tangent vectors at the same point, and an apparent lack of global bounds on these, suggests geodesics may be non-unique, or at least makes proving uniqueness more challenging. Finally, Example 5.3.1 which involves approximate calculation of Fréchet means, suggests that wald space is less 'sticky' than BHV tree space and hence more attractive for studying asymptotic statistics.

A variety of open problems remain, and we make the following conjectures.

(1) All points on any geodesic between two trees are also trees.
(2) Geodesics between trees in the same grove do not leave the closure of that grove.
(3) The disconnected forest $F_\infty$ is repulsive, in the sense that the only geodesics passing through the disconnect forest have an end point there.

Other open problems include the following, all mentioned elsewhere in the paper.

(4) Is wald space a truncated topological cone?
(5) Does Whitney condition (B) hold at grove boundaries?
(6) Most importantly for statistical applications, is wald space uniquely geodesic or can examples of exact non-unique geodesics be constructed? What is then the structure of cut loci?

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge helpful discussions with Fernando Galaz-García and support from the DFG RTG 2088 (Jonas Lueg), and from the DFG HU 1575-7 and the Niedersachsen Vorab of the Volkswagen Foundation (Stephan F. Huckemann).

Open access funding enabled and organized by Projekt DEAL.

## JOURNAL INFORMATION

The *Journal of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

## ORCID

*Tom M. W. Nye* https://orcid.org/0000-0002-8737-777X

## REFERENCES

1. F. Ardila and C. J. Klivans, *The Bergman complex of a matroid and phylogenetic trees*, J. Comb. Theory B **96** (2006), no. 1, 38–49.
2. J. C. Baez and N. Otter, *Operads and phylogenetic trees*, Theory and Applications of Categories, **32** (2017), No. 40, 1397–1453.
3. D. Barden and H. Le, *The logarithm map, its limits and Fréchet means in orthant spaces*, P. Lond. Math. Soc. **117** (2018), no. 4, 751–789.
4. D. Barden, H. Le, and M. Owen, *Central limit theorems for Fréchet means in the space of phylogenetic trees*, Electron. J. Probab. **18** (2013), 1–25.

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5. D. Barden, H. Le, and M. Owen, *Limiting behaviour of Fréchet means in the space of phylogenetic trees*, Ann. I. Stat. Math. **70** (2018), no. 1, 99–129.
6. M. Bačák, *Computing medians and means in Hadamard spaces*, SIAM J. Optimiz. **24** (2014), no. 3, 1542–1566.
7. L. J. Billera, S. P. Holmes, and K. Vogtmann, *Geometry of the space of phylogenetic trees*, Adv. Appl. Math. **27** (2001), no. 4, 733–767.
8. M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der mathematischen Wissenschaften, vol. 319, Springer, Berlin, Heidelberg, 1999.
9. D. G. Brown and M. Owen, *Mean and variance of phylogenetic trees*, Syst. Biol. **69** (2020), no. 1, 139–154.
10. P. Buneman, *The recovery of trees from measures of dissimilarity*, Mathematics in the Archeological and Historical Sciences: Proceedings of the Anglo-Romanian Conference, Mamaia, 1971, pp. 387–395.
11. P. Buneman, *A note on the metric properties of trees*, J. Comb. Theory B **17** (1974), no. 1, 48–50.
12. B. Eltzner and S. F. Huckemann, *A smeary central limit theorem for manifolds with application to high-dimensional spheres*, Ann. Stat. **47** (2019), no. 6, 3360–3381.
13. S. N. Evans, J. Pitman, and A. Winter, *Rayleigh processes, real trees, and root growth with re-grafting*, Probab. Theory Rel. **134** (2006), no. 1, 81–126.
14. J. Felsenstein, *Inferring phylogenies*, Oxford University Press, Oxford, 2003.
15. M. Fréchet, *Les éléments aléatoires de nature quelconque dans un espace distancié*, Ann. I. H. Poincaré **10** (1948), no. 4, 215–310.
16. M. K. Garba, T. M. W. Nye, J. Lueg, and S. F. Huckemann, *Information geometry for phylogenetic trees*, J. Math. Biol. **82** (2021), no. 3, 19.
17. M. K. Garba, T. M. W. Nye, and R. J. Boys, *Probabilistic distances between trees*, Syst. Biol. **67** (2018), no. 2, 320–327.
18. T. Hotz and S. Huckemann, *Intrinsic means on the circle: uniqueness, locus and asymptotics*, Ann. I. Stat. Math. **67** (2015), no. 1, 177–193.
19. T. Hotz, S. Skwerer, S. Huckemann, H. Le, J. S. Marron, J. C. Mattingly, E. Miller, J. Nolen, M. Owen, and V. Patrangenaru, *Sticky central limit theorems on open books*, Ann. Appl. Probab. **23** (2013), no. 6, 2238–2258.
20. T. Hu and W. A. Kirk, *Local contractions in metric spaces*, P. Am. Math. Soc. **68** (1978), no. 1, 121–124.
21. S. Huckemann and B. Eltzner, *Statistical methods generalizing principal component analysis to non-Euclidean spaces*, Handbook of variational methods for non-linear geometric data, P. Grohs, M. Holler, and A. Weinmann (eds.), Springer International Publishing, Berlin, 2020, pp. 317–338.
22. S. F. Huckemann, J. Mattingly, E. Miller, and J. Nolen, *Sticky central limit theorems at isolated hyperbolic planar singularities*, Electron. J. Probab. **20** (2015), 1–34.
23. J. Kim, *Slicing hyperdimensional oranges: the geometry of phylogenetic estimation*, Mol. Phylogenet. Evol. **17** (2000), no. 1, 58–75.
24. S. Lang, *Fundamentals of differential geometry*, Graduate texts in mathematics, Springer, Berlin, pp. 901–922 1999.
25. J. M. Lee, *Introduction to Riemannian manifolds*, Graduate texts in mathematics, 2nd ed., Springer, Berlin, 2018.
26. J. Lueg, M. K. Garba, T. M. W. Nye, and S. F. Huckemann, *Wald space for phylogenetic trees*, Geometric science of information, F. Nielsen and F. Barbaresco (eds.), Lecture Notes in Computer Science, Springer International Publishing, Berlin, 2021, pp. 710–717.
27. W. P. Maddison, *Gene trees in species trees*, Syst. Biol. **46** (1997), no. 3, 523–536.
28. J. S. Marron and A. M. Alonso, *Overview of object oriented data analysis*, Biometrical J. **56** (2014), no. 5, 732–753.
29. E. Miller, M. Owen, and J. S. Provan, *Polyhedral computational geometry for averaging metric phylogenetic trees*, Adv. Appl. Math. **68** (2015), 51–91.
30. A. Monod, B. Lin, R. Yoshida, and Q. Kang, *Tropical geometry of phylogenetic tree space: a statistical perspective*, arXiv:1805.12400, 2022.
31. V. Moulton and M. Steel, *Peeling phylogenetic 'oranges'*, Adv. App. Math. **33** (2004), no. 4, 710–727.
32. T. M. W. Nye, *Principal components analysis in the space of phylogenetic trees*, Ann. Stat. **39** (2011), no. 5, 2716–2739.
33. T. M. W. Nye, *An algorithm for constructing principal geodesics in phylogenetic treespace*, IEEE-ACM T. Comput. Bi. **11** (2014), no. 2, 304–315.


14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

34. T. M. W. Nye, X. Tang, G. Weyenberg, and R. Yoshida, *Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees*, Biometrika **104** (2016), no. 4.
35. M. Owen and J. S. Provan, *A fast algorithm for computing geodesic distances in tree space*, IEEE-ACM T. Comput. Bi. **8** (2011), no. 1, 2–13.
36. M. J. Pflaum, *Analytic and geometric study of stratified spaces*, Lecture Notes in Mathematics, vol. 1768, Springer, Berlin, 2001.
37. M. Rumpf and B. Wirth, *Variational time discretization of geodesic calculus*, IMA J. Numer. Anal. **35** (2015), no. 3, 1011–1046.
38. F. Schmidt, M. Clausen, and D. Cremers, *Shape matching by variational computation of geodesics on a manifold*, Pattern recognition, K. Franke, K.-R. Müller, B. Nickolay, and R. Schäfer (eds.), Lecture Notes in Computer Science, vol. 4174, DAGM (German Association for Pattern Recognition), Springer, Berlin, 2006, pp. 142–151.
39. C. Semple and M. Steel, *Phylogenetics*, Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford, 2003.
40. D. Speyer and B. Sturmfels, *The tropical Grassmannian*, Adv. Geom. **4** (2004), no. 3, 389–411.
41. K.-T. Sturm, *Probability measures on metric spaces of nonpositive curvature*, Heat kernels and analysis on manifolds, graphs, and metric spaces: lecture notes from a quarter program on heat kernels, random walks, and analysis on manifolds and graphs, P. Auscher, T. Coulhon, and A. Grigor'yan (eds.), Emile Borel Centre of the Henri Poincaré Institute, Paris, France, 2002, Contemporary Mathematics, vol. 338, American Mathematical Society, Providence, RI, pp. 357–390, 2003.
42. M. A. Suchard, *Stochastic models for horizontal gene transfer: taking a random walk through tree space*, Genetics **170** (2005), no. 1, 419–431.
43. A. Willis, *Confidence sets for phylogenetic trees*, J. Am. Stat. Assoc. **114** (2016), 235–244.
44. Z. Yang, *Computational molecular evolution*, Oxford University Press, Oxford, 2006.

14697750, 2024, 5, Downloaded from https://londmathsoc.onlinelibrary.wiley.com/doi/10.1112/jlms.12893 by Newcastle University, Wiley Online Library on [30/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License