STATISTICAL METHODS FOR PHYLOGENETIC TREES WITH NON-IDENTICAL LEAF SETS

MARIA ALEJANDRA VALDEZ CABRERA

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

University of Washington
2024

Reading Committee:
Amy Willis, Chair
Ali Shojaie
Eardi Lila

Program Authorized to Offer Degree:
Department of Biostatistics

UNIVERSITY OF WASHINGTON

Abstract

Statistical methods for Phylogenetic Trees with Non-identical leaf sets
Maria Alejandra Valdez Cabrera

Chair of the Supervisory Committee:
Amy Willis
Department of Biostatistics

Phylogenetic trees, which describe shared ancestry between organisms through their branching structure (topology) and branch lengths, are a fundamental tool for analyzing evolutionary relationships. However, limited statistical tools exist for analyzing collections of phylogenetic trees with non-identical leaf sets. Here we present the first algorithm for computing distances between any pair of phylogenetic trees via Billera-Holmes-Vogtmann (BHV) extension spaces. Extension spaces represent trees with fewer leaves within the metric space of a larger leaf set, thus enabling comparisons in a common metric space. Motivated by limitations of extension spaces, we introduce the "towering" tree space, a metric space defined by transitions between nested BHV spaces. We describe an algorithm for computing distances within towering tree space, and propose Fréchet means to summarize collections of trees. We illustrate our proposed methods on gene trees spanning multiple domains of life.

ACKNOWLEDGMENTS

This project began as an intriguing idea suggested by my advisor, Amy Willis, during a chance encounter on a bus commute, and I will be forever thankful for that auspicious moment. Throughout this process, Amy has kindly embraced all my crazy ideas, my "intuition tells me this should go this way" moments, and all the redrafts and do-overs, guiding me with patience and expertise. I am grateful for her encouragement and the freedom she gave me in the development of this project, as well as her emotional support during the trying moments.

I would like to express my gratitude to Megan Owen, who generously offered advice during the conception of these projects and patiently answered my questions about her previous work. Our conversations significantly increased my understanding and shaped what became Chapter 3. She also provided invaluable support with coding.

I acknowledge all the wonderful mentors and professors who guided me during my six years in the Department of Biostatistics. Special thanks to Ruth Etzioni and Elizabeth Brown, who were early mentors in my career and helped shape me into the researcher I am today. I am grateful to Ali Shojaie and Eardi Lila for serving as my reading committee members and providing invaluable feedback during my general exam and on the final version of this document. I also thank Ali Shojaie for his brilliant course on scientific writing, which was one of my favorites.

To all the amazing (past and present) members of the StatDivLab, who have made collaboration fun and meaningful since I joined. To David Clausen and Pauline Trinh, for being a source of inspiration through their hard work, passion for their respective projects, and being very unique individuals in their own way. I also acknowledge Pauline for being the best dog sitter in Seattle. To my close colleague, Sarah Teichman, who provided a comfortable and productive environment through coffee-shop writing sessions and ice-cream hangouts. I want to thank the current members of the lab, Shirley Mathur, Ameer Dharamshi

and Grant Hopkins, for their company at coffee time and listening to many practice talks.

I want to recognize all of my cohort members: Avi Kenny, Taylor Okonek, Charlie Wolock, Pearl Liu, and Si Cheng. We braved all the core courses together, and I would surely have gone insane without our study sessions before the qualifying exams. I will always remember with joy our cohort outings to different breweries and for mushroom hunting. You guys always reminded me to enjoy the outdoors that Seattle has to offer. To other students in the Department, especially upperclassmen Subodh Selukar and Kendrick Li, for their career and life advice during my early years.

For more personal acknowledgments, I would like to recognize all the amazing people I have met since moving to Seattle. To Elliot Lee, thank you for inviting me to join your DnD group and for all the laughter on Thursday nights ever since. To Jacob Frick, thank you for understanding my community references and being an amazing friend to chill on the couch with. To Charles King, thank you for leaving me random messages on the phone that make me laugh, and crushing me in all the video games. To Fernanda Moreno, thank you for all the dinner and coffee dates where I could always vent about life.

To my life-long friends, who have stayed with me through the distance. To Ornella Tonda, Zul Giron, and Samantha Guerrero, thank you for being the first people I think about whenever something happens in my life. Having you three on the other side of the phone since high school has been one of the biggest joys in my life and has definitely held me together during my breaking points. To Cecilia Hernandez, for being the best roommate I could have asked for in college, and even better, for still being part of my life despite our paths drifting apart. To German Puga and Moises Pelayo, for all the afternoons playing board games on the computer that kept me sane during lock-downs.

To my beloved family, I want to thank you from the deepest part of my heart for all the support and endearment. To my mother, Karen Cabrera, for always having my back and being a force to be reckoned with. I know my being away is difficult sometimes, but your

support is what has brought me this far. To my father, thank you for all the love and advice, and for not getting too mad when I do not call as often as I should. To my sister, thank you for always being interested in what I am working on, both professionally and in my random hobbies. Also thank you for keeping me updated on what is happening back home. I know I can count on you for everything. And to my brother, for making me laugh with random memes and always showing me love and affection. And to my cherished Dog, Shippo, for all the tail waggles he throws my way when I am feeling low.

Finally, I want to recognize my partner, Edward Arnold. We have been working side by side, each of us on our own PhDs. The road has been long, but you have always kept my morale high and the good times coming. Thank you for keeping me strong when I wanted to give up, and for keeping my belly full when I didn't have time to cook. Wherever we take our journey next, I am sure it will be a bright and fun adventure.

To my mother, Karen Cabrera,

who exemplifies the person I aspire to become.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER 1

INTRODUCTION

Phylogenetic trees are fundamental tools in modern biology for understanding evolutionary relationships. In addition to organizing life using genetic similarity, modern phylogenetics enables the prediction of characteristics of uncultivated organisms (Parks et al., 2020; Rinke et al., 2013), lineage tracing of pathogens and drug-resistance genes (Chen et al., 2016; Gardy & Loman, 2018), and forensic investigations (Scaduto et al., 2010), amongst other applications. Broadly, this work constructs new mathematical, algorithmic, and statistical tools to analyze collections of phylogenetic trees with non-identical leaf sets.

Phylogenetic reconstruction methods (Drummond et al., 2005; Guindon et al., 2003; Stamatakis, 2014) leverage genomic data encompassing one or more genes shared between organisms to estimate phylogenetic trees. However, due to natural biological processes like deep coalescence or horizontal gene transfer, phylogenetic trees that describe different genes frequently exhibit differing topologies or significantly different branch lengths (Maddison, 1997), potentially leading to incorrect species tree estimations (Edwards, 2009; Kubatko et al., 2007). This variability has led to a shift towards analyzing entire collections of trees, necessitating the development of advanced mathematical methodologies to handle their complex structure.

Billera et al. (2001) introduced the BHV tree space $(\mathcal{T}^{\mathcal{N}}, d)$, where $\mathcal{T}^{\mathcal{N}}$ denotes the space of all trees with leaf set $\mathcal{N}$, and $d : \mathcal{T}^{\mathcal{N}} \times \mathcal{T}^{\mathcal{N}} \mapsto \mathbb{R}_{\geq 0}$ is the distance function. This metric space is Hadamard (Bridson & Haefliger, 1999, Definition 1.1), enabling the simultaneous comparison of topologies and branch lengths through continuous, smooth geodesic paths that represent gradual changes in these features. This has led to multiple tools for the analysis of tree collections, such as means and variances (Barden et al., 2013; Benner et al., 2014; Brown & Owen, 2020; Willis & Bell, 2018), confidence sets (Willis, 2019), density estimation

(Weyenberg et al., 2014), and clustering (Gori et al., 2016).

However, the BHV distance is only defined between trees with the same leaf sets. As a result, BHV-based tools are restricted to trees on the same set of organisms, limiting their practical applicability. For example, recent advances in genomic sequencing have led to a substantial increase in the number of near-complete archaeal genomes (Baker et al., 2020), with the potential for generating new insights into the origins of complex multicellular life (Imachi et al., 2020; Zhu et al., 2019). However, few genes are shared across archaeal, bacterial, and eukaryotic organisms, complicating the study of ancient evolution. To address the limitation that BHV-based tools can only analyze phylogenetic trees with identical leaf sets, Grindstaff and Owen (2019) introduced *extension spaces*. Briefly, for a tree T with leaves $\mathcal{L} \subseteq \mathcal{N}$, the elements of the extension space $E_T^{\mathcal{N}} \subset \mathcal{T}^{\mathcal{N}}$ are the trees created by attaching the leaves in $\mathcal{N} \setminus \mathcal{L}$ to T in every possible way. Extension spaces provide an avenue for analyzing trees with non-identical leaf sets inside the same BHV space, because $\mathcal{N}$ can be constructed as the union of all trees' leaf sets. I consider the shortest BHV path between extension spaces as a universally-applicable compatibility measure between trees, including those with differing leaves, and propose the first algorithm to compute distances between extension spaces.

An overview of the BHV tree space is presented in Chapter 2, along with a discussion on recent work regarding trees with non-identical leaf sets. In Chapter 3, I present the first algorithm to compute distances between extension spaces, an analysis of its computational performance, and an application comparing prokaryotic gene trees. Chapter 4 introduces a novel family of metric spaces for phylogenetic trees with non-identical leaf sets, and an exploration of the geometrical properties of this space. This metric space, called the Towering Tree Space, is the first of its kind, inheriting BHV-like behavior but enabling the analysis of collections of trees with non-identical leaf sets. Chapter 5 contains a discussion of limitations and future work, including algorithms for averaging trees in the Towering Space.

CHAPTER 2

PHYLOGENETIC TREES AND THEIR ANALYSIS

A phylogenetic tree is a graphical representation of the shared evolutionary history among a set of organisms, and is a key object in many biological studies. Analyzing phylogenetic trees with standard statistical tools is challenging due to their complex structure, and specialized tools have been built for their analysis. Significant progress has been made in analyzing trees with identical leaf sets using the BHV tree space (Billera et al., 2001). In this chapter, I review key results on the BHV metric space and topics related to extending these results to trees with non-identical leaf sets.

2.1 Phylogenetic tree structure

A phylogenetic tree encodes evolutionary divergence events through its topological shape, and encodes the distances between these events (e.g., time or number of genetic mutations) by the lengths of its edges. Formally, phylogenetic trees connect a set of organisms $\mathcal{L}$, which label leaves on the tree. Sometimes, one leaf is labeled as the "root," which represents a common ancestor among the organisms of interest. Nevertheless, the tools for rooted and un-rooted trees in the BHV spaces are equivalent, and so proceed by focusing on unrooted trees. Thus, I use the following definition for phylogenetic trees on $\mathcal{L}$.

Definition 2.1. A **phylogenetic tree** T on a set of organisms $\mathcal{L}$ is a connected weighted acyclic graph whose **leaves** (nodes of degree 1) are labelled by the elements of $\mathcal{L}$, and all interior nodes are at least of degree 3. All edge-weights are non-negative values referred to as **lengths**. The branching pattern of the graph gives the **topology** of the tree.

Removing an edge from a tree T divides it into two connected graphs, each with a subset of the leaves as part of its nodes. Each edge in T can be uniquely identified by the partition

$\mathcal{L} = \mathcal{G} \sqcup \{\mathcal{L} \setminus \mathcal{G}\}$ induced when removed. Throughout I denote this partition by $[\mathcal{G} \ddagger \mathcal{L} \setminus \mathcal{G}]$ and call it a **split** on $\mathcal{L}$.

Edges that induce the same split on topologically distinct trees are considered to be the same edge. Because of this, I use the terms edge and split interchangeably. It is frequently useful to know when two splits can be present in the same tree, for which there is a clear criterion in the following definition.

Definition 2.2. Two splits $s = [\mathcal{G} \ddagger \mathcal{L} \setminus \mathcal{G}]$ and $s' = [\mathcal{K} \ddagger \mathcal{L} \setminus \mathcal{K}]$ are considered **compatible** (Billera et al., 2001, Section 3.2) if they can be part of the edges of the same tree. Equivalently, s and s' are compatible if:

(i) At least one of the intersections $\mathcal{G} \cap \mathcal{K}$, $\{\mathcal{L} \setminus \mathcal{G}\} \cap \mathcal{K}$, $\mathcal{G} \cap \{\mathcal{L} \setminus \mathcal{K}\}$ or $\{\mathcal{L} \setminus \mathcal{G}\} \cap \{\mathcal{L} \setminus \mathcal{K}\}$ is empty.

(ii) One of the subsets in s is contained in one of the subsets of s', meaning $\mathcal{G} \subseteq \mathcal{K}$, $\mathcal{G} \subseteq \{\mathcal{L} \setminus \mathcal{K}\}$, $\{\mathcal{L} \setminus \mathcal{G}\} \subseteq \mathcal{K}$ or $\{\mathcal{L} \setminus \mathcal{G}\} \subseteq \{\mathcal{L} \setminus \mathcal{K}\}$.

Edges connecting to leaves are called the **external** edges, which correspond to splits $|\mathcal{G}| = 1$ or $|\mathcal{L} \setminus \mathcal{G}| = 1$, while all others, with $2 \leq |\mathcal{G}|, |\mathcal{L} \setminus \mathcal{G}|$, are the **internal** edges.

The topology of a tree is fully defined by all its internal splits. I use $\mathcal{S}(T)$ to refer to the set of internal splits in T (equivalently, the topology of T). For a tree with l leaves, the cardinality of this set falls in $\{0, 1, \ldots, l-3\}$, and the tree is binary when it has $l-3$ internal edges, which is also referred to as fully resolved. Additionally, T has l external edges, one per leaf. $\mathcal{H}(T)$ refers to this set of external edges and $\mathcal{P}(T) = \mathcal{S}(T) \cup \mathcal{H}(T)$ is the set of all edges in T. Later in this chapter, I discuss how the size of each edge indexed by $\mathcal{P}(T)$ affects the distance to T and how this influence varies slightly if the edge belongs to $\mathcal{S}(T)$ or $\mathcal{H}(T)$.

2.2 The Billera-Holmes-Vogtmann tree space

The BHV tree space proposed by Billera et al. (2001) is a metric space for all phylogenetic trees sharing a common leaf set. The BHV distance simultaneously compares topology and edge lengths between trees. The geometrical properties of the space makes it a Hadamard space (Bridson & Haefliger, 1999, I.1, Definition 1.1), which guarantees continuous paths connecting any pair of trees and a unique path achieving the minimal path length. This section overviews the definition and properties of the BHV space $\left(\mathcal{T}^{\mathcal{N}}, d_{\text{BHV}}\right)$ for a set of leaves $\mathcal{N}$ of size $n = |\mathcal{N}|$.

Consider all possible topologies for trees with leaf set $\mathcal{N}$ of size $n = |\mathcal{N}|$, meaning all possible subsets of internal splits that are mutually compatible. For one such set S, assign an order to its elements ($S = \{s_1, s_2, \ldots, s_m\}$) in any consistent way; for example, an order can be assigned to the leaves and the edges may inherit the lexicographic order. A tree with the topology given by S may be represented by a $(n+m)-$dimensional non-negative vector, where the first n coordinates represent the lengths of the external edges and the last m coordinates represent the lengths of the internal branches in the given order (Figure 2.1). All trees with topology S can be represented by a $(n+m)-$dimensional non-negative vector, which together form an orthant in $\mathbb{R}^{n+m}$, called the **topology orthant** and denoted by $\mathcal{O}(S)$ (Owen & Provan, 2011, Section 2.1).

Topology orthants are connected to each other along appropriate equivalence classes. A tree T with an internal edge $e \in \mathcal{S}(T)$ of size zero looks identical to a tree without e among its internal edges, provided all other edges remain unchanged (Figure 2.2). These two trees can be considered to be equal. Given a set of internal splits S with m elements, any topology orthant $\mathcal{O}(S')$ corresponding to a proper subset $S' \subset S$ of these internal edges can be viewed as a **face** of the larger topology orthant $\mathcal{O}(S)$ (Billera et al., 2001, Section 2), corresponding to the trees with length zero in all edges in the difference $S \setminus S'$. I use the notation $O' \subset O$

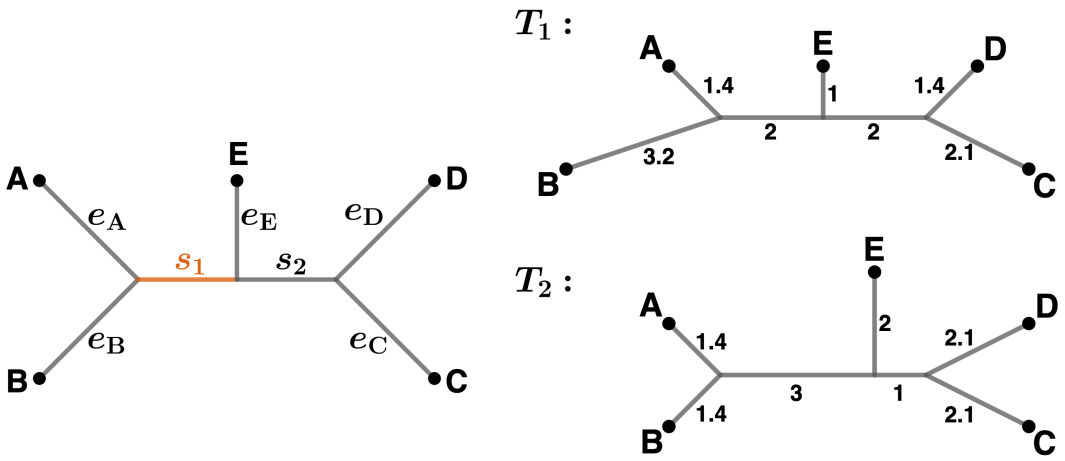


Figure 2.1. (a) A phylogenetic tree with leaf set $\mathcal{N} = \{A, B, C, D, E\}$ and internal split $s_1 = [\{A, B\} \ddagger \{C, D, E\}]$ highlighted in orange. This topology can be summarized by 7-dimensional vectors $(e_A, e_B, e_C, e_D, e_E, s_1, s_2)$. (b) Two different trees with the same topology. In its topology orthant, T_1 corresponds to the vector $(1.4, 3.2, 2.1, 1.4, 1, 2, 2)$ while T_2 corresponds to the vector $(1.4, 1.4, 2.1, 2.1, 2, 3, 1)$.

to refer to O' as a face of O. Two topology orthants $\mathcal{O}(S_1)$ and $\mathcal{O}(S_2)$ such that $S := S_1 \cap S_2$ are glued along the face corresponding to the topology orthant $\mathcal{O}(S)$ (Figure 2.2(b)). The only topology orthants that are not faces of another topology orthant are those of maximum dimension $2n - 3$, corresponding to the topologies of binary trees.

Remark 2.3. For any equivalence class, there is a unique representative of the class with all internal edges of positive length. This corresponds to the tree obtained from dropping any internal edge of length zero from the set of internal edges. From now on I adopt the convention that when referring to a tree T, $S(T)$ contains strictly positive edges (unless indicated otherwise), and employ the notation $\mathcal{O}(T)$ to refer to the lowest-dimensional orthant containing T; that is, $\mathcal{O}(T) = \mathcal{O}(S(T))$. For a topology orthant O, $\mathcal{S}(O)$ denotes the set of internal splits represented by the topology of O.

A path can be traced between any two trees, where a path is a piecewise connected curve. Each piece is fully contained within a single orthant, and contiguous pieces connect through

a common face of the orthants containing them. The BHV space is a complete geodesic space with non-positive curvature (Billera et al., 2001, Lemma 4.1), and thus the shortest path between any two trees (the **geodesic**) is unique. The **distance** between two trees $d_{\text{BHV}}(T_1, T_2)$ is the length of the geodesic connecting them.

Remark 2.4. In the BHV space the geodesic between two trees in the same orthant is the line segment between them contained in the orthant. Thus, the BHV metric coincides with the Euclidean metric when restricted to a single orthant. An orthant can be viewed as a Euclidean subspace of $\mathcal{T}^{\mathcal{N}}$.

The set $\mathcal{T}^{\mathcal{N}}$ that I consider differs from the original BHV space provided in Billera et al. (2001) by including the lengths of the external edges as part of the coordinates in the topology orthants. The subspace of $\mathcal{T}^{\mathcal{N}}$ comprising all trees with external edges of length zero, denoted as $\text{BHV}_{\mathcal{N}}$, is isometric to the original definition. Henceforth, I consider $\mathcal{T}^{\mathcal{N}}$ as $\mathbb{R}^n_{\geq 0} \times \text{BHV}_{\mathcal{N}}$. This definition aligns with that of Grindstaff and Owen (2019). Keeping track of external edge lengths becomes necessary when working with different leaf sets. Nevertheless, focusing on the space $\text{BHV}_{\mathcal{N}}$ (i.e., considering internal edges only) is beneficial when discussing the geometric and combinatorial properties of the BHV space, as I explore briefly in the next section.

I conclude this section by noting that $\mathcal{T}^{\mathcal{N}}$ being the Cartesian product of the space of external edge lengths $\mathbb{R}^n_{\geq 0}$ and the internal-edges-only space $\text{BHV}_{\mathcal{N}}$ implies the distance between two trees T_1 and T_2 in $\mathcal{T}^{\mathcal{N}}$ can be expressed as the L^2-norm combination of the distances in these two spaces. In other words, given T_1' and T_2' the trees with internal edges identical to T_1 and T_2 but external edge lengths set to zero, the the distance in $\mathcal{T}^{\mathcal{N}}$ is the same as:

$$d_{\text{BHV}}(T_1, T_2) = \sqrt{\sum_{e \in \mathcal{H}(T_1)} (|e|_{T_1} - |e|_{T_2})^2 + d^2_{\text{BHV}}(T_1', T_2')},$$

where $|e|_T$ is the length the edge e has in tree T.

2.2.1 Combinatorial representation of the BHV space: link of the origin

The BHV tree space smoothly incorporates differences in both topologies and edge lengths when comparing two trees. Trees with identical topologies reside within the same topology orthant, and their distance is determined by the Euclidean distance between their edge lengths. Conversely, trees in different topology orthants have distances influenced by the lengths of the uncommon internal edges, which must be dropped (reduced to zero-length) when crossing orthant faces. In the next section, I discuss the polynomial-time algorithm of Owen and Provan (2011) to construct geodesics and find the distance between any two trees. First, however, I provide an alternative combinatorial description of the BHV space, which aids in understanding the dependence of the BHV distance on both topologies and edge lengths.

As mentioned above, each possible topology for a tree on $\mathcal{N}$ is uniquely defined by a set S of pairwise compatible internal edges. Any internal edge corresponds to a partition on the leaves $\mathcal{N}$ of the form $[\mathcal{G} \ddagger \mathcal{L} \setminus \mathcal{G}]$ with $|\mathcal{G}|, |\mathcal{L} \setminus \mathcal{G}| \geq 2$; there are a total of $2^{n-1} - (n+1)$ such edges, but S can contain at most $n-3$ of these at a time. Moreover, there are a total of $(2n-3)!!$ different topologies for binary trees on $\mathcal{N}$.

Considering all possible internal splits on $\mathcal{N}$ as vertices (0-simplices), it is possible to construct an abstract simplicial complex (Jonsson, 2008, Section 2.3.1), named the link of the origin and denoted by $L_{\mathcal{N}}$, where every k-simplex is a set S of $k+1$ pairwise compatible internal edges; i.e, each simplex is a valid topology for a tree on $\mathcal{N}$ (Billera et al., 2001, Section 3.2).

This link of the origin can be extended to a spherical complex, by describing each k-simplex as a right-angled spherical simplex where the distance between each vertex pair in the simplex (i.e. the length of the edges) is $\frac{\pi}{2}$, endowed with the spherical intrinsic metric (Bridson & Haefliger, 1999, I.7, Definition 7.4). These simplices naturally arise by

considering all the trees in $\text{BHV}_{\mathcal{N}}$ that intersect the unit sphere centered at the origin of each topology orthant. Then, $\text{BHV}_{\mathcal{N}}$ is a 0-cone on $L_{\mathcal{N}}$ (Bridson & Haefliger, 1999, I.5, Definition 5.6), where each tree $T \in \text{BHV}_{\mathcal{N}}$ can be represented by the point (d, t) in the cone, with t the projection of T onto $L_{\mathcal{N}}$ (the intersection of the ray from the origin towards T and $L_{\mathcal{N}}$) and d is the distance from the origin tree to T. So, the distance between any two trees $T_1 = (d_1, t_1)$ and $T_2 = (d_2, t_2)$ in the internal-edges-only BHV space can be expressed by,

$$d^2_{\text{BHV}}(T_1, T_2) = d_1^2 + d_2^2 - 2d_1 d_2 \cos\left(\min\left\{\pi, \measuredangle(t_1, t_2)\right\}\right),$$

where $\measuredangle(t_1, t_2)$ is the spherical distance on the link of the origin. This spherical distance heavily reflects the differences in the topologies of the two trees. Billera et al. (2001, Section 4.2) utilized this to demonstrate that the orthants traversed by the geodesic between any two trees always include the common internal edges in their topologies. Additionally, all other splits in the topologies of these traversed orthants are part of the topology of one of the two trees in the endpoints of the geodesic. In the next section I show how this fact was leveraged into an algorithm to compute distances in the BHV space.

2.2.2 Geodesics in BHV space

Here I summarize the algorithm of Owen and Provan (2011) to find geodesics in the BHV space. This algorithm also produces a closed-form expression for the length of these geodesics. This is useful to describe how small changes to endpoint trees in a geodesic influences its length, which I utilize in Chapter 3.

Definition 2.5. (Owen, 2011, Definition 3.3) For trees T_1 and T_2, consider the sets of internal splits $S_1 = \mathcal{S}(T_1)$, $S_2 = \mathcal{S}(T_2)$ and the set of common internal splits $C = S_1 \cap S_2$. For a sequence of subsets $S_1 = G_0 \supset G_1 \supset \ldots \supset G_{k-1} \supset G_k = C$ and $C = F_0 \subset F_1 \subset \ldots \subset F_{k-1} \subset F_k = S_2$ such that $G_i \cup F_i$ is a set of pairwise compatible splits, denote by

$O_i = \mathcal{O}(G_i \cup F_i)$ for all $i \in \{0, \ldots, k\}$. The sequence of orthants,

$$\mathcal{O}(T_1) = O_0 \to O_1 \to \ldots \to O_k = \mathcal{O}(T_2),$$

is a **path space** from T_1 to T_2.

Remark 2.6. In the previous definition, since T_1 and T_2 are not necessarily binary, it may happen that an edge p belongs to S_1 and is absent from S_2, but it is pairwise compatible with every edge in S_2; meaning it could be added to S_2 with length zero. In this case, p is considered part of the common edges C. The same rule applies to all p in $S_2 \setminus S_1$ that are pairwise compatible with S_1.

Definition 2.7. (Owen & Provan, 2011, Section 2.3) Consider a path space $O_0 \to \ldots \to O_k$. When transitioning from O_{i-1} to O_i, denote by $A_i = G_{i-1} \setminus G_i$ the set of dropped edges and by $B_i = F_i \setminus F_{i-1}$ the set of added edges. For ordered sets $\mathcal{A} = \{A_1, \ldots, A_k\}$ and $\mathcal{B} = \{B_1, \ldots, B_k\}$, the pair $(\mathcal{A}, \mathcal{B})$ gives the **support** of the path space. Each (A_i, B_i) is also called a **support pair**.

A path space describes a way to efficiently move from the topology orthant of T_1 to the topology orthant of T_2 through connected orthants. Definition 2.5 differs slightly from that of Owen (2011, Definition 3.3) in that it allows T_1 and T_2 to have common edges. This change is congruent with Billera et al. (2001, Proposition 4.1), which ensures common edges are part of the topologies for all orthants the geodesic traverses, while uncommon edges in S_1 (corresponding to the A_i's) are gradually replaced by uncommon edges in S_2 (corresponding to the B_i's).

Given a support for the path space containing the geodesic, the length of the geodesic depends on the lengths of the uncommon edges through the L^2-norm of each A_i and B_i in the support; and on the difference in length of the common edges, including the external edges $H = \mathcal{H}(T_1) = \mathcal{H}(T_2)$. Denoting by $|p|_T$ the length of the edge p in tree T and by

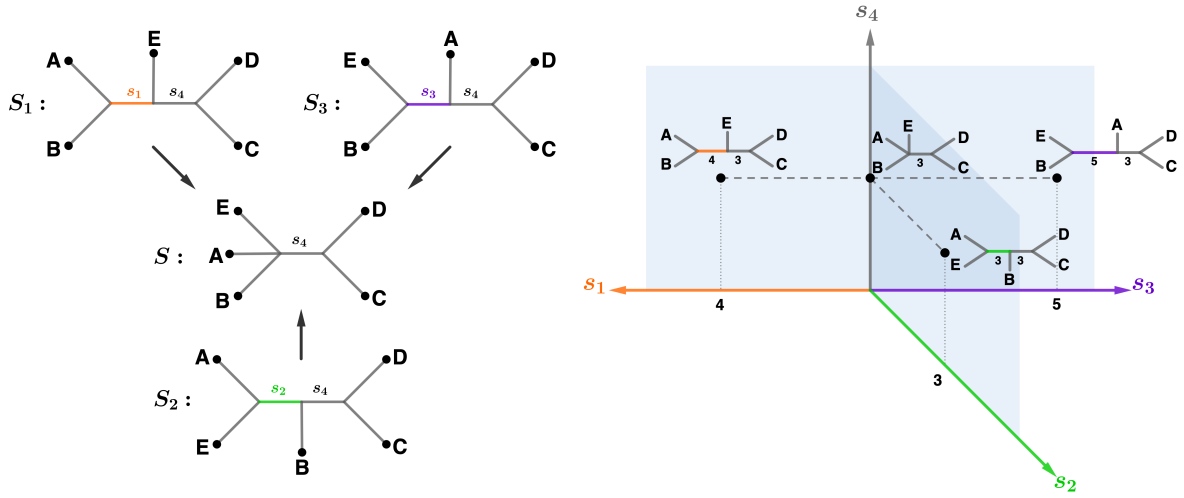


Figure 2.2. (a) Topologies S_1, S_2 and S_3 share topology S as a face. The topology S_i is composed of internal edges $S_i = \{s_i, s_4\}$, while the topology S only has $S = \{s_4\}$; thus $\mathcal{O}(S)$ is a face of $\mathcal{O}(S_i)$ for $i = 1, 2, 3$. (b) The topology orthants $\mathcal{O}(S_1)$, $\mathcal{O}(S_2)$ and $\mathcal{O}(S_3)$ are "glued" at $\mathcal{O}(S)$. The dimensions arising from external edges are not shown.

$||P||_T = \sqrt{\sum_{p \in P} |p|_T^2}$ the L^2-norm of the lengths in T of all edges in the set of splits P, the length of the geodesic can be written in these terms by

$$d_{\text{BHV}}(T_1, T_2) = \left|\left|\left(||A_1||_{T_1} + ||B_1||_{T_2}, \ldots, ||A_k||_{T_1} + ||B_k||_{T_2}, (|s|_{T_1} - |s|_{T_2})_{s \in K}\right)\right|\right|, \qquad (2.1)$$

where $(a_i)_{i \in I}$ denotes the vector containing objects in I as entries, and $K = C \cup H$ (Miller et al., 2015, Theorem 1.2) (see also Owen and Provan (2011, Section 4)).

There are specific properties that can be verified to ensure the path space associated with the support contains the geodesic between the two trees. These properties were first detailed in Section 2.3 of Owen and Provan (2011), and used in Section 4 to develop the *Geodesic Treepath Problem* (GTP) algorithm, which finds the appropriate support in polynomial time. These properties are:

(P1) For each $i > j$, the edges $A_i \cup B_j$ are pairwise compatible (Billera et al., 2001, Proposition 4.1).

(P2) $\frac{||A_1||_{T_1}}{||B_1||_{T_2}} \leq \frac{||A_2||_{T_1}}{||B_2||_{T_2}} \leq \ldots \leq \frac{||A_k||_{T_1}}{||B_k||_{T_2}}$ (Owen & Provan, 2011, Theorem 2.3).

(P3) For each support pair (A_i, B_i), there is no nontrivial partitions $A_i = C_1 \sqcup C_2$, $B_i = D_1 \sqcup D_2$, such that $D_1 \cup C_2$ are pairwise compatible and $\frac{||C_1||_{T_1}}{||D_1||_{T_2}} \leq \frac{||C_2||_{T_1}}{||D_2||_{T_2}}$ (Owen & Provan, 2011, Theorem 2.5).

Of note, Theorem 2.4 in Owen and Provan (2011) ensures that finding a support satisfying (P1) and (P2) provides a path from T_1 to T_2 with length given by (2.1), but this path is only the geodesic when (P3) is also true.

When finding the appropriate support through the GTP algorithm, the first step separates each tree into r subtrees $(T_1^1, T_2^1), \ldots, (T_1^r, T_2^r)$ by bisecting the common edges in the middle and adding a new label at the end of each half as a new leaf, in a way that the trees in each pair (T_1^j, T_2^j) still have the same leaves, but no common edges. For each tree pair, a proper support containing the geodesic is found through results based on trees without common edges (Owen & Provan, 2011, Section 3). When combining these separate supports in (2.1), I require that for any support pair (A_i, B_i) there is a $j \in \{1, \ldots, r\}$ such that all edges in A_i correspond to edges in T_1^j and all edges in B_i correspond to edges in T_2^j. This requirement does not impact the results presented here and it is useful in Chapter 4.

2.3 Analyzing trees with non-identical leaf sets

The BHV space was introduced as a common space to perform analyzes on tree samples, which are effectively collections of trees $\{T_1, \ldots, T_n\}$, each of them a full phylogenetic tree on a leaf set $\mathcal{N}$. I now discuss on the setting where not all trees in this collection are phylogenetic trees on the same leaf set, but rather the sample contains trees with varying leaves $(T_1 \in \mathcal{T}^{\mathcal{N}_1}, \ldots, T_n \in \mathcal{T}^{\mathcal{N}_n})$. Foundational work by Ren et al. (2017) introduced a combinatorial approach to relate different BHV spaces, which was later expanded by

Grindstaff and Owen (2019) to introduce *Extension Spaces*. The tools I propose in chapters 3 and 4 are closely related to these contributions.

2.3.1 The tree dimensionality reduction map

A way to reduce the number of leaves in a tree was first introduced by Zairis et al. (2016) through a tool to visualize trees with large leaf sets ($|\mathcal{N}| >> 1$). This approach was used to connect different BHV spaces by Grindstaff and Owen (2019). Although originally proposed as a mapping between phylogenetic trees relating to both their topology and branch lengths, it is also useful to consider the counterpart of this map as a function that maps splits on one set of leaves to splits on a subset of these leaves, making it a combinatorial tool to map topologies into topologies. I provide both definitions here.

A tree T with leaf set $\mathcal{L}$ gives rise to a distance between any two members of $\mathcal{L}$, defined as the shortest path between them if traversing the graph T. For each tree, I consider the discrete metric space on $\mathcal{L}$ with metric $d_T(\ell_1, \ell_2)$ defined to be the distance between $\ell_1, \ell_2 \in \mathcal{L}$ in T (Figure 2.3a).

Definition 2.8. For $\mathcal{L} \subseteq \mathcal{N}$, the **tree dimensionality reduction** (TDR) map is the projection function $\Psi_{\mathcal{L}} : \mathcal{T}^{\mathcal{N}} \mapsto \mathcal{T}^{\mathcal{L}}$ that maps each tree $T' \in \mathcal{T}^{\mathcal{N}}$ to the **unique** tree $T \in \mathcal{T}^{\mathcal{L}}$ such that $(\mathcal{T}^{\mathcal{L}}, d_T)$ is the metric subspace of $(\mathcal{T}^{\mathcal{N}}, d_{T'})$ restricted to $\mathcal{L}$.

The previous definition differs from that of Zairis et al. (2016, Definition 4.1), which defined the TDR map as the product of removing any edges not belonging to a path connecting leaves in $\mathcal{L}$ from T, then deleting every induced degree-2 vertex v by merging their incident edges, replacing the two edges (u_1, v) and (v, u_2) (with lengths w_1 and w_2) with edge (u_1, u_2) (with length $w_1 + w_2$). I use the definition of Grindstaff and Owen (2019, Proposition 2.6), who proved the equivalence between these two definitions. They also suggested the use of $\Psi_{\mathcal{L}}$ as a map between split- and orthant-valued arguments.

Definition 2.9. (Grindstaff & Owen, 2019, Definition 2.4) Given the split $s = [\mathcal{G} \ddagger \mathcal{N} \setminus \mathcal{G}]$ on $\mathcal{N}$, take $\Psi_{\mathcal{L}}(s)$ to be the split on $\mathcal{L}$ that separates all leaves of $\mathcal{L}$ in $\mathcal{G}$ from those not in $\mathcal{G}$; that is, $\Psi_{\mathcal{L}}(s) = [\mathcal{L} \cap \mathcal{G} \ddagger \mathcal{L} \setminus \mathcal{G}]$.

Remark 2.10. It might be the case that either $\mathcal{G} \cap \mathcal{L} = \emptyset$ or $\mathcal{L} \setminus \mathcal{G} = \emptyset$, which means the projection $\Psi_{\mathcal{L}}(s)$ for $s = [\mathcal{G} \ddagger \mathcal{N} \setminus \mathcal{G}]$ is not a valid edge. This happens when this split is not involved in the shortest path between any pair of leaves contained in $\mathcal{L}$, which means this edge should be removed from the tree when applying $\Psi_{\mathcal{L}}$. In these cases I write $\Psi_{\mathcal{L}}(s) = \emptyset$.

2.3.2 Extension Spaces

Extension spaces are a way to represent trees with a smaller leaf set $\mathcal{L} \subset \mathcal{N}$ in the BHV space $\mathcal{T}^{\mathcal{N}}$, thereby providing a framework for analyzing trees with non-identical leaf sets as subsets of $\mathcal{T}^{\mathcal{N}}$. Their definition, provided below, coincides with the pre-image of T under the tree dimensionality reduction map $\Psi_{\mathcal{L}}$.

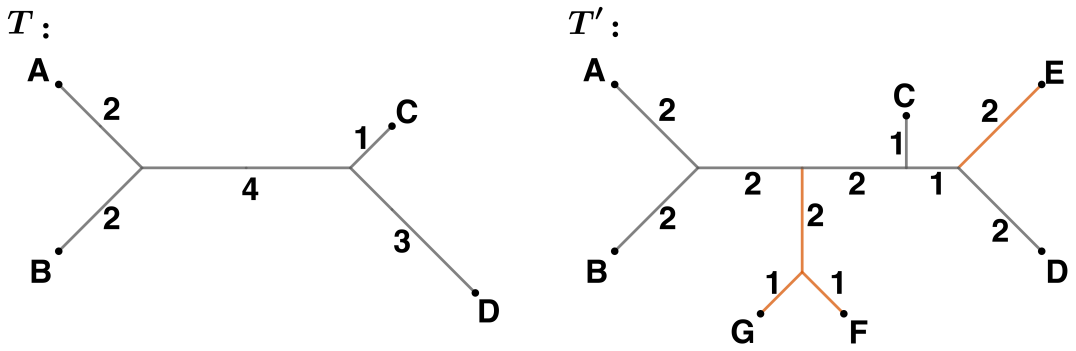
Definition 2.11. (Grindstaff & Owen, 2019, Definition 3.9) Given a tree T with leaf set $\mathcal{L} \subseteq \mathcal{N}$, the **extension space** of T in the BHV space $\mathcal{T}^{\mathcal{N}}$ is

$$E_T^{\mathcal{N}} = \left\{ T' \in \mathcal{T}^{\mathcal{N}} | d_{T'}(\ell_1, \ell_2) = d_T(\ell_1, \ell_2) \text{ for all } \ell_1, \ell_2 \in \mathcal{L} \right\}. \tag{2.2}$$

While I describe the structure of these extension spaces in detail in the next chapter, I conclude this chapter by introducing two key concepts that aid in this description.

Definition 2.12. (Ren et al., 2017, Section 1.3) For $\mathcal{L} \subseteq \mathcal{N}$ and a binary tree $T \in \mathcal{T}^{\mathcal{L}}$, the **connection cluster** $C_T^{\mathcal{N}}$ of T in $\mathcal{T}^{\mathcal{N}}$ is the set of maximum-dimensional orthants in $\Psi_{\mathcal{L}}^{-1}(\mathcal{O}(T))$.

Definition 2.13. (Grindstaff & Owen, 2019, Definition 3.9) Given a binary tree $T \in \mathcal{T}^{\mathcal{L}}$, with $\mathcal{L} \subseteq \mathcal{N}$, and an orthant $O \in C_T^{\mathcal{N}}$, let the **orthant-specific extension space**, E_T^O, be all trees in O that are part of the extension space; that is, $E_T^O = E_T^{\mathcal{N}} \cap O$.



(a) *(b)*

Figure 2.3. For the complete leaf set $\mathcal{N} = \{A, B, C, D, E, F, G\}$ and subset $\mathcal{L} = \{A, B, C, D\}$. (a) A tree $T \in \mathcal{T}^{\mathcal{L}}$. For the discrete metric space over $\mathcal{L}$ given by T, $d_T(A, D) = 9$. (b) A tree $T' \in \mathcal{T}^{\mathcal{N}}$ in the extension space of T. The inconsequential edges for $E_T^{\mathcal{N}}$ are highlighted in orange.

Armed with the above definitions, I move on to defining and exploring distances between extension spaces.

CHAPTER 3

DISTANCES BETWEEN EXTENSION SPACES OF PHYLOGENETIC TREES

Extension spaces, as discussed in Section 2.3.2, were proposed as a natural way to represent in a BHV space for trees with leaves $\mathcal{N}$ trees with fewer leaves. These can also be used to represent two trees with non-identical leaf sets in a common BHV space and compare them based on these spaces. For this, Grindstaff and Owen (2019, Section 5) introduces two measures of "compatibility" between trees with non-identical leaf sets. These measures are based on constructing neighborhoods around the trees' extension spaces and determining the smallest radius for these neighborhoods to intersect. Trees in these intersections can be seen as "supertrees" that combine the two phylogenies. However, the neighborhoods do not intersect if the extension spaces do not share at least one common topology. Therefore, these measures are only defined for trees that are topologically compatible.

To address the need for a compatibility measure that is defined for *any* pair of trees, I consider the shortest BHV path between their extension spaces (Grindstaff & Owen, 2019, Section 3.4). That is:

$$d\left(E_{T_1}^{\mathcal{N}}, E_{T_2}^{\mathcal{N}}\right) := \inf_{(t_1,t_2)\in E_{T_1}^{\mathcal{N}}\times E_{T_2}^{\mathcal{N}}} d_{\mathrm{BHV}}(t_1,t_2). \tag{3.1}$$

In this chapter I propose the first algorithm to compute this distance, leveraging the polynomial-time BHV distance algorithm developed by Owen and Provan (2011). Although this distance does not define a metric between trees with non-identical leaf sets (Grindstaff & Owen, 2019, Remark 3.16), it is still an objective BHV based comparison that can be performed for any pair of trees with non-identical leaves and it could be employed to build super-trees based on tree samples.

The algorithm is introduced at the end of Section 3.3. It is based on the structure of the

orthant-specific extension spaces, discussed in Section 3.2 and employs the theory developed around non-linear optimization of convex functions through gradient of descent methods, which I review in the next section. The chapter ends with a discussion on the computational performance of the algorithm and a practical application on prokaryotic gene trees.

3.1 Reduced gradient methods

Gradient descent methods are a class of iterative algorithms that find stationary points of a function. At each iteration of a gradient descent algorithm, the local changes of the function around the current point are evaluated, and a "step" in best local direction is taken.

In their simplest form, these methods are defined on a continuous differentiable function $f : D \mapsto \mathbb{R}$ on a domain $D \subset \mathbb{R}^n$. They rely on the idea that the gradient ∇f describes local behavior of f and at each $x \in D$, and the vector $-\nabla f(x)$ provides a direction towards which the value of f is locally decreasing the most. By starting at x_0 and repeatedly taking small steps in a decreasing direction, the algorithm approaches a local minimum. While these algorithms are not guaranteed to find the global minimum of the function (or even to converge), certain conditions on f and D ensure good behavior. In this overview, I focus on convex functions over convex and compact domains. In general, for convex differentiable functions it is guaranteed the existence of a global minimum, and that this minimum is reachable by gradient descent methods (Ruszczyński, 2006, Theorem 3.1, Theorem 5.4).

The reduced gradient method is applicable when there are constraints on the domain of f. That is, the goal is to minimize $f : \mathbb{R}^n \mapsto \mathbb{R}$ subject to two constraints:

- **Linear constraints:** $A\mathbf{x} = \mathbf{b}$ for some matrix $A \in \mathbb{R}^{m \times n}$ of rank m and a fixed $\mathbf{b} \in \mathbb{R}^m$, and
- **Inequality constraints:** $\mathbf{x} \geq 0$.

The set of values $\mathcal{X} \subset \mathbb{R}^n$ holding these two constraints is called the **feasible set**. Within this feasible set, the algorithm searches by moving along different facets, where a **facet** is a subset of $\mathcal{X}$ with certain entries fixed at zero. Specifically, a facet consists of all vectors $x \in \mathcal{X}$ such that $x_i = 0$ for each i in some pre-specified subset of indices.

At each step of the reduced gradient method, the direction of change is selected by considering the gradient of f while limiting the options based on the constraints. Since A is a matrix of rank m, the system of equations $A\mathbf{x} = \mathbf{b}$ has $n-m$ degrees of freedom; that is, m entries of $\mathbf{x}$ can be seen as depending on the other $n-m$ as free variables. Then the gradient of f can be considered as a function on the free variables, determining the change on the dependent variables afterwards. This guaranteed the next point in the process remains in the feasible set.

In what follows, I use notation similar to Ruszczyński (2006, Section 6.1.2). At each step of the reduced gradient method, I classify the variables in $\mathbf{x}$ into three groups:

- Null variables $\mathbf{N}$: Variables fixed at zero. $x_i = 0$ when $i \in \mathbf{N}$.
- Free variables $\mathbf{F}$: Independent variables. x_i can take any non-negative value when $i \in \mathbf{F}$.
- Dependent variables $\mathbf{D}$: Knowing the values of all other variables allows us to find x_i, $i \in \mathbf{D}$, to satisfy the linear constraints.

I use the notation $\mathbf{x_D}$ to refer to the vector containing only the entries of $\mathbf{x}$ corresponding to the dependent variables, $\mathbf{x_F}$ the vector containing all the free variables, $\mathbf{x_N}$ the vector containing all the null variables, and write $f(\mathbf{x}) = f(\mathbf{x_D}, \mathbf{x_F}, \mathbf{x_N})$ to make explicit the dependence of f on the variables as the index sets vary. I also use $M_\mathbf{I}$ to refer to the matrix obtained by subsetting the columns of M indexed by $\mathbf{I}$.

The reduced gradient method begins by initializing the index sets, selecting m variables to be the dependent variables in $\mathbf{D}^0 \subset \{1, \ldots, n\}$ so that $A_{\mathbf{D}^0}$ is a square matrix of rank m, then taking $\mathbf{F}^0 = \{1, \ldots, n\} \setminus \mathbf{D}^0$ and $\mathbf{N}^0 = \emptyset$. An initial point $\mathbf{x}^0$ such that $A\mathbf{x}^0 = \mathbf{b}$ and $\mathbf{x}^0 > 0$ is also initialized. In the k-th iteration:

1. Define the function φ that takes the values of f but depends only on the free variables. Null variables are fixed at zero and the dependent variables are those that guarantee the linear constraints, $\mathbf{x}_{\mathbf{D}^k} = A_{\mathbf{D}^k}^{-1}[\mathbf{b} - A_{\mathbf{F}^k}\mathbf{x}_{\mathbf{F}^k}]$. Therefore, $\varphi(\mathbf{x}_{\mathbf{F}^k}) = f\left(A_{\mathbf{D}^k}^{-1}\left[\mathbf{b} - A_{\mathbf{F}^k}\mathbf{x}_{\mathbf{F}^k}\right], \mathbf{x}_{\mathbf{F}^k}, 0\right)$.

2. Compute the gradient of this function by applying the chain rule:

$$\nabla\varphi(\mathbf{x}_{\mathbf{F}^k}) = \nabla_{\mathbf{F}^k} f(\mathbf{x}) - A_{\mathbf{F}^k}^{\top}\left[A_{\mathbf{D}^k}^{-1}\right]^{\top}\nabla_{\mathbf{D}^k} f(\mathbf{x}). \tag{3.2}$$

3. Find an optimal direction of change $\mathbf{d}^k = (\mathbf{d}_{\mathbf{D}}^k, \mathbf{d}_{\mathbf{F}}^k, \mathbf{d}_{\mathbf{N}}^k)$ in the feasible set by:

   i. Employing an unconstrained optimization method to find $\mathbf{d}_{\mathbf{F}}^k$ based on $\nabla\varphi(\mathbf{x}_{\mathbf{F}^k})$ (e.g., via the conjugate gradient method (Ruszczyński, 2006, Section 5.5)).

   ii. Setting $\mathbf{d}_{\mathbf{D}}^k = -A_{\mathbf{D}}^{-1}A_{\mathbf{F}}\mathbf{d}_{\mathbf{F}}^k$ to ensure linear constraints.

   iii. Keeping the null variables fixed at zero, i.e., $\mathbf{d}_{\mathbf{N}}^k = 0$.

4. Find the value τ^* that minimizes $f(\mathbf{x}^k + \tau\mathbf{d}^k)$, maintaining the inequality constraint, $\mathbf{x}^k + \tau^*\mathbf{d}^k \geq 0$ (e.g., using a line search method (Ruszczyński, 2006, Section 5.2)).

5. Select the next point by taking $\mathbf{x}^{k+1} = \mathbf{x}^k + \tau^*\mathbf{d}^k$.

6. Verify if variables need to be reclassified. If $\langle\nabla f(\mathbf{x}^{k+1}), d^k\rangle < 0$, then τ^* was selected for being the maximum value of τ holding the inequality constraint. Some non-null variable in $\mathbf{x}^{k+1}$ has value zero and needs to be reclassified:

i. If $x_i^{k+1} = 0$ for some $i \in \mathbf{F}^k$, then reclassify x_i as a null variable (if there is multiple options, select one), setting $\mathbf{F}^{k+1} = \mathbf{F}^k \setminus \{i\}$ and $\mathbf{N}^{k+1} = \mathbf{N}^k \cup \{i\}$.

ii. Otherwise, $x_i^{k+1} = 0$ for some $i \in \mathbf{D}^k$ (if there is multiple, select one). In this case, x_i^{k+1} is reclassified as a null variable, but a new dependent variable must be found to satisfy the linear constraint. Find $j \in \mathbf{F}^k$ such that setting $\mathbf{D}^{k+1} = \{j\} \cup \mathbf{D}^k \setminus \{i\}$ makes $A_{\mathbf{D}^{k+1}}$ still of rank m (Ruszczyński, 2006, Lemma 6.2). Then, $\mathbf{F}^{k+1} = \mathbf{F}^k \setminus \{j\}$ and $\mathbf{N}^{k+1} = \mathbf{N}^k \cup \{i\}$.

7. Repeat 1-6 until either $\nabla\varphi(\mathbf{x}^k_{\mathbf{F}^k}) = 0$ or $\mathbf{F}^k = \emptyset$.

8. Check if the current point $\mathbf{x}^k$ is the global solution by computing $\bar{\mu}^k = \nabla_{\mathbf{N}^k} f\left(\mathbf{x}^k\right) - A^\top_{\mathbf{N}^k}\left\{A^\top_{\mathbf{D}^k}\right\}^{-1} \nabla_{\mathbf{D}^k} f\left(\mathbf{x}^k\right)$ Ruszczyński (2006, Lemma 6.3). If $\bar{\mu}^k_j \geq 0$ for every entry j, end the algorithm and return $\mathbf{x}^k$. If $\bar{\mu}^k_j < 0$ for any entry j, increase the search region by taking $\mathbf{N}_0 \subset \mathbf{N}^k$ so that $\bar{\mu}^k_i < 0$ for all $i \in \mathbf{N}_0$ and reclassifying $\mathbf{F}^{k+1} = \mathbf{F}^k \cup \mathbf{N}_0$ and $\mathbf{N}^{k+1} = \mathbf{N}^k \setminus \mathbf{N}_0$, and return to 1-6.

If f is convex and the feasible set $\mathcal{X} \subset \mathbb{R}^n$ is compact, the number of stationary points that are be reached in Step 7 is finite, and therefore the global minimum is eventually attained (Ruszczyński, 2006, Theorem 6.4).

3.2 Structure of the Extension Spaces

The extension space of a tree $T \in \mathcal{T}^{\mathcal{L}}$ in the BHV space $\mathcal{T}^N$ can be divided into a finite number of subspaces by considering all the orthant-specific extension spaces indexed by the connection cluster $C_T^{\mathcal{N}}$; i.e. the extension space can be described through the finite union $E_T^{\mathcal{N}} = \bigcup_{O \in C_T^{\mathcal{N}}} E_T^O$. Moreover, each E_T^O can be characterized through a set of linear equations. Namely, given a binary tree $T \in \mathcal{T}^{\mathcal{L}}$ with l leaves and an orthant $O \in C_T^{\mathcal{N}}$, consider the list of $2l-3$ splits $q_1, \ldots, q_{2l-3}$ in $\mathcal{S}(T)$ and all $2n-3$ splits $p_1, \ldots, p_{2n-3}$ in $\mathcal{S}(O)$. Note

that for every split p_j, either $\Psi_{\mathcal{L}}(p_j) = q_i$ for some unique value $i \in \{1, \ldots, 2l-3\}$, or $\Psi_{\mathcal{L}}(p_j) = \emptyset$. The TDR map acts linearly on the lengths of the splits because the removal of a leaf bisecting a branch results in a branch length that is the sum of the lengths of the adjacent branches. Formally,

$$|q_i|_{\Psi_{\mathcal{L}}(T')} = \sum_{p_j : \Psi_{\mathcal{L}}(p_j) = q_i} |p_j|_{T'},$$

for all $T' \in O$. Thus, the TDR map for the orthant O can be described by a linear map given by a $(2l-3) \times (2n-3)$-dimensional matrix,

$$M_{\mathcal{L}}^{O}[i,j] = \begin{cases} 1 & \text{if } \Psi_{\mathcal{L}}(p_j) = q_i, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, E_T^O can be described by the system of equations

$$\begin{aligned} M_{\mathcal{L}}^{O}\mathbf{x} &= \mathbf{v}_T, \\ \mathbf{x} &\geq 0, \end{aligned} \tag{3.3}$$

where $\mathbf{x}$ is a $(2n-3)-$dimensional vector in orthant O that gives the branch lengths of trees in E_T^O and $\mathbf{v}_T$ is the vector representation of T in $\mathcal{O}(T)$; that is, it is the $(2l-3)-$dimensional vector with the edge lengths of T.

Definition 3.1. The j-th column of the matrix $M_{\mathcal{L}}^{O}$ corresponds to the edge p_j of the trees in O, which by definition is a zero column if there is no edge q_i to which p_j maps into; i.e. $\Psi_{\mathcal{L}}(p_j) = \emptyset$. Thus, these edges as **inconsequential** for this extension space, since their length is unimportant for $\mathbf{x}$ to be a solution to (3.3). If an edge is not inconsequential, it is **consequential**.

Each column in $M_{\mathcal{L}}^{O}$ corresponding to a consequential edge has exactly one entry equal to 1 and the rest equal to 0, implying the rows are linearly independent. Moreover, every

edge q_i in T has a non-empty pre-image under $\Psi_{\mathcal{L}}$, which means the matrix has no zero rows. Thus, $M_{\mathcal{L}}^{O}$ is of rank $2l-3$.

Example 3.2. Consider again the trees in Figure 2.3, and take O to be the orthant of T'. The internal splits of trees in O are: $s_1 = [\{A,B\} \ddagger \{C,D,E,F,G\}]$, $s_2 = [\{C,D,E\} \ddagger \{A,B,F,G\}]$, $s_3 = [\{D,E\} \ddagger \{A,B,C,F,G\}]$ and $s_4 = [\{F,G\} \ddagger \{A,B,C,D,E\}]$. Any tree in E_T^O solves (3.3) for the following values of $M_{\mathcal{L}}^{O}$ and $\mathbf{v}_T$. All zero columns in the projection matrix are highlighted. These coincide with with the inconsequential edges shown in Figure 2.3b.

$$\mathbf{x}^\top = \begin{pmatrix} e_A & e_B & e_C & e_D & e_E & e_F & e_G & s_1 & s_2 & s_3 & s_4 \end{pmatrix},$$

$$M_{\mathcal{L}}^{O} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{v}_T = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 3 \\ 4 \end{pmatrix}$$

Extension spaces can be described through the union of a finite number of subsets, each contained in a single maximum-dimensional orthant. These maximum-dimensional orthants are Euclidean spaces (Remark 2.4) and the extension spaces restricted to maximum-dimensional orthants are affine subspaces. These observations lay the groundwork for the development of an algorithm to compute distances between orthant-specific extension spaces based on gradient of descent methods.

3.3 Distances between extension spaces

Given two trees T_1 and T_2 with leaf sets $\mathcal{L}_1, \mathcal{L}_2 \subseteq \mathcal{N}$, in this section I address the process of computing the distance between their extension spaces $E^{\mathcal{N}}_{T_1}$ and $E^{\mathcal{N}}_{T_2}$ (see (3.1)). A pair of trees $(t_1^*, t_2^*) \in E^{\mathcal{N}}_{T_1} \times E^{\mathcal{N}}_{T_2}$ is called an **optimal pair** if $d_{\mathrm{BHV}}(t_1^*, t_2^*) = d\left(E^{\mathcal{N}}_{T_1}, E^{\mathcal{N}}_{T_2}\right)$. I discuss below the convexity properties of orthant-specific extension spaces that guarantees the existence of an optimal pair. However, this may not be unique.

3.3.1 Search region

The algorithm for finding the distance between extension spaces begins by determining the minimal distance between all possible pairings of the orthant-specific extension spaces (Definition 2.13) that comprise the extension spaces. To achieve this, I consider every orthant in each connection cluster (Definition 2.12), and form all orthant pairs $(O_1, O_2) \in C^{\mathcal{N}}_{T_1} \times C^{\mathcal{N}}_{T_2}$. This process constructs a finite number of candidates for the optimal pair.

For every orthant pair, I construct a convex and compact search region by excluding poor candidates, as described in the following definition.

Definition 3.3. Given trees $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$, where $\mathcal{L}_1, \mathcal{L}_2 \subseteq \mathcal{N}$, and an orthant pair $(O_1, O_2) \in C^{\mathcal{N}}_{T_1} \times C^{\mathcal{N}}_{T_2}$, define the **orthant-specific mutually restricted extension space**, $\lceil E^{O_1}_{T_1} \times E^{O_2}_{T_2} \rceil$, as the subset of all tree pairs $(T_1', T_2') \in E^{O_1}_{T_1} \times E^{O_2}_{T_2}$ for which common edges that are inconsequential in either T_1 or T_2 are the same length in both trees, common edges that are inconsequential in both trees are length zero, and inconsequential uncommon edges are of length zero. That is, $(T_1', T_2') \in E^{O_1}_{T_1} \times E^{O_2}_{T_2}$ is in $\lceil E^{O_1}_{T_1} \times E^{O_2}_{T_2} \rceil$ when the **conditions for mutual restriction** hold:

- $|p|_{T_1'} = |p|_{T_2'}$ for all $p \in \mathcal{P}(O_1) \cap \mathcal{P}(O_2)$ with $\Psi_{\mathcal{L}_1}(p) = \emptyset$ or $\Psi_{\mathcal{L}_2}(p) = \emptyset$,
- $|p|_{T_1'} = |p|_{T_2'} = 0$ for all $p \in \mathcal{P}(O_1) \cap \mathcal{P}(O_2)$ with $\Psi_{\mathcal{L}_1}(p) = \emptyset$ and $\Psi_{\mathcal{L}_2}(p) = \emptyset$,

- $|p|_{T_1'} = 0$ for every $p \in \mathcal{P}(O_1) \setminus \mathcal{P}(O_2)$ with $\Psi_{\mathcal{L}_1}(p) = \emptyset$,

- $|p|_{T_2'} = 0$ for every $p \in \mathcal{P}(O_2) \setminus \mathcal{P}(O_1)$ with $\Psi_{\mathcal{L}_2}(p) = \emptyset$.

This definition is motivated by the length of any inconsequential edge p in $T_1' \in O_1$ not affecting whether T_1' is a part of the extension of T_1. However, the length of $|p|_{T_1'}$ influences the distance from this tree to trees in the orthant-specific extension space $E_{T_2}^{O_2}$. When building a tree T_1' as a candidate for the optimal pair, once the lengths of consequential edges are chosen to ensure $T_1' \in E_{T_1}^{O_1}$, the length of p can freely be selected to minimize the distance.

Lemmas 3.4 and 3.5 formalize this idea, showing that for any pair $(T_1', T_2') \in E_{T_1}^{O_1} \times E_{T_2}^{O_2}$, it is possible to construct a new tree pair $(T_1^*, T_2^*) \in \lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ such that $d_{\text{BHV}}(T_1^*, T_2^*) \leq d_{\text{BHV}}(T_1', T_2')$. The first result concerns uncommon inconsequential edges.

Lemma 3.4. *Consider two trees T_1 and T_2 in $\mathcal{T}^{\mathcal{N}}$ such that there is at least one edge p that is present in the edge set $\mathcal{P}(T_1)$ of edges of T_1 but not in the edge set $\mathcal{P}(T_2)$ of edges of T_2. Define $T_1^{\perp p}$ to be the projection of T_1 onto the face of $\mathcal{O}(T_1)$ defined by the length of p being equal to zero; that is, $T_1^{\perp p}$ is such that $T_1^{\perp p} \in \mathcal{O}(T_1)$ with $|p|_{T_1^{\perp p}} = 0$ and $|p'|_{T_1^{\perp p}} = |p'|_{T_1}$ for all $p' \neq p$. Then, $d_{BHV}(T_1^{\perp p}, T_2) \leq d_{BHV}(T_1, T_2)$.*

Proof. Consider the support $(\mathcal{A}, \mathcal{B})$ of the path space of the geodesic from T_1 to T_2. For $i = 1, \ldots, k$, denote by $t_i \in \mathcal{T}^{\mathcal{N}}$ the first tree on the geodesic from T_1 to T_2 that belongs to the orthant $O_i = \mathcal{O}(C \cup B_1 \cup \ldots \cup B_i \cup A_{i+1} \cup \ldots A_k)$ in the path space of the geodesic. This means t_i is a tree in the face shared by O_{i-1} and O_i. Since p is an edge of T_1 and not of T_2, then there is a value $r \in 1, \ldots, k$ such that $p \in A_r$. A shorter path from $T_1^{\perp p}$ to t_r than the geodesic from T_1 to t_r is constructed by projecting every t_i for $i \leq r$ towards the face of the respective orthant where the length of the edge p is zero.

$p \in S(O_i)$ for every $i < r$. Similarly as before, for each $i \leq r$, define $t_i^{\perp p}$ to be the tree in O_i such that $|p|_{t_i^{\perp p}} = 0$ and $|p'|_{t_i^{\perp p}} = |p'|_{t_i}$ for every other edge $p' \in \mathcal{P}(O_i)$ such that $p' \neq p$. In particular, since t_r is a tree in the face shared by O_{r-1} and O_r, and $p \in A_r$ implies $p \notin \mathcal{P}(O_r)$, then $|p|_{t_r} = 0$ and $t_r^{\perp p} = t_r$.

Denote $t_0 = T_1$. For every $i = 1, \ldots, r$, both t_{i-1} and t_i are in O_{i-1}, so $d_{\text{BHV}}(t_{i-1}, t_i) = \sqrt{\sum_{p' \in \mathcal{P}(O_{i-1})} \left(|p'|_{t_{i-1}} - |p'|_{t_i}\right)^2}$ and $d_{\text{BHV}}(t_{i-1}^{\perp p}, t_i^{\perp p}) = \sqrt{\sum_{p' \in \mathcal{P}(O_{i-1})} \left(|p'|_{t_{i-1}^{\perp p}} - |p'|_{t_i^{\perp p}}\right)^2}$. In the latter expression the difference of lengths for edge p are zero and all other differences are as in the former expression, therefore $d_{\text{BHV}}(t_{i-1}^{\perp p}, t_i^{\perp p}) \leq d_{\text{BHV}}(t_{i-1}, t_i)$. Thus,

$$\begin{aligned} d_{\text{BHV}}(T_1, T_2) \geq & d_{\text{BHV}}(T_0^{\perp p}, t_1^{\perp p}) + d_{\text{BHV}}(t_1^{\perp p}, t_2^{\perp p}) + \ldots \\ & + d_{\text{BHV}}(t_{r-1}^{\perp p}, t_r) + d_{\text{BHV}}(t_r, T_2) \geq d_{\text{BHV}}(T_1^{\perp p}, T_2) \end{aligned}$$

□

I now give a stronger result which demonstrates minimal distances between extension spaces can be found by searching orthant-specific mutually restricted extension spaces.

Lemma 3.5. *Given two trees T_1 and T_2 with leaf sets $\mathcal{L}_1, \mathcal{L}_2 \subseteq \mathcal{N}$ and an orthant pair $(O_1, O_2) \in C_{T_1}^{\mathcal{N}} \times C_{T_2}^{\mathcal{N}}$, the distance between the orthant-specific extension spaces equals the distance between their orthant-specific mutually restricted extension space, that is,*

$$\inf \left\{ d_{BHV}(T_1', T_2') \,\middle|\, (T_1', T_2') \in E_{T_1}^{O_1} \times E_{T_2}^{O_2} \right\} = \inf \left\{ d_{BHV}(T_1', T_2') \,\middle|\, (T_1', T_2') \in \lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil \right\}.$$

Proof. Since $\lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil \subseteq E_{T_1}^{O_1} \times E_{T_2}^{O_2}$, it is only necessary to show that for any pair $(T_1', T_2') \in E_{T_1}^{O_1} \times E_{T_2}^{O_2}$ there is a pair $(T_1^*, T_2^*) \in \lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ such that $d_{\text{BHV}}(T_1^*, T_2^*) \leq d_{\text{BHV}}(T_1', T_2')$. Such a pair is constructed by first setting all inconsequential common edges to a proper length and all uncommon inconsequential edges to length zero.

Consider (T_1^0, T_2^0) such that $\mathcal{O}(T_1^0) \subseteq O_1$ and $\mathcal{O}(T_2^0) \subseteq O_2$, $|s|_{T_1^0} = |s|_{T_1'}$ for all $s \in \mathcal{P}(O_1) \setminus \mathcal{P}(O_2)$, $|s|_{T_2^0} = |s|_{T_2'}$ for all $s \in \mathcal{P}(O_2) \setminus \mathcal{P}(O_1)$, and the conditions for mutual

restriction hold for common edges $p \in \mathcal{P}(O_1) \cap \mathcal{P}(O_2)$:

$$|p|_{T_1^0} = |p|_{T_1'}, |p|_{T_2^0} = |p|_{T_2'} \text{ when } \Psi_{\mathcal{L}_1}(p) \neq \emptyset \text{ and } \Psi_{\mathcal{L}_2}(p) \neq \emptyset$$
$$|p|_{T_1^0} = |p|_{T_2^0} = |p|_{T_2'} \text{ when } \Psi_{\mathcal{L}_1}(p) = \emptyset \text{ and } \Psi_{\mathcal{L}_2}(p) \neq \emptyset$$
$$|p|_{T_2^0} = |p|_{T_1^0} = |p|_{T_1'} \text{ when } \Psi_{\mathcal{L}_1}(p) \neq \emptyset \text{ and } \Psi_{\mathcal{L}_2}(p) = \emptyset$$
$$|p|_{T_1^0} = |p|_{T_2^0} = 0 \text{ when } \Psi_{\mathcal{L}_1}(p) = \emptyset \text{ and } \Psi_{\mathcal{L}_2}(p) = \emptyset$$

Since the lengths of the uncommon internal splits are unchanged from (T_1', T_2') to (T_1^0, T_2^0), the support of the path space is the same for both pairs (common edges lengths do not influence the support for the geodesic (Owen & Provan, 2011, Section 4)). For those edges that are inconsequential in one or both trees, the difference in length drops to zero; it remains the same for edges that are consequential for both (for such an edge p, $|p|_{T_1'} - |p|_{T_2'} = |p|_{T_1^0} - |p|_{T_2^0}$). Denote the common edges (including external edges) that are consequential for both extension spaces by K^0 (i.e. $s \in K^0 \subseteq K = C \cup H$ when $\Psi_{\mathcal{L}_1}(s) \neq \emptyset$ and $\Psi_{\mathcal{L}_2}(s) \neq \emptyset$)

$$\begin{aligned}
d_{\text{BHV}}(T_1^0, T_2^0) &= \left|\left|\left(||A_1||_{T_1^0} + ||B_1||_{T_2^0}, \ldots, ||A_k||_{T_1^0} + ||B_k||_{T_2^0}, \left(|s|_{T_1^0} - |s|_{T_2^0}\right)_{s \in K}\right)\right|\right| \\
&= \left|\left|\left(||A_1||_{T_1^0} + ||B_1||_{T_2^0}, \ldots, ||A_k||_{T_1^0} + ||B_k||_{T_2^0}, \left(|s|_{T_1^0} - |s|_{T_2^0}\right)_{s \in K^0}, 0\right)\right|\right| \\
&= \left|\left|\left(||A_1||_{T_1'} + ||B_1||_{T_2'}, \ldots, ||A_k||_{T_1'} + ||B_k||_{T_2'}, \left(|s|_{T_1'} - |s|_{T_2'}\right)_{s \in K^0}, 0\right)\right|\right| \\
&\leq d_{\text{BHV}}(T_1', T_2').
\end{aligned}$$

Although all common edges in the tree pair (T_1^0, T_2^0) hold the conditions for mutual restrictions, some of the uncommon edges between both trees may not. Define the inconsequential edges in T_1^0 that are not common with T_2^0 by $(p_1^1, \ldots, p_{r_1}^1)$, and the inconsequential edges in T_2^0 not common with T_1^0 by $(p_1^2, \ldots, p_{r_2}^2)$. For each $i = 1, 2$, define $T_i^j = T_i^{(j-1)\perp p_j^i}$ the projection of T_i^{j-1} towards the face of O_i defined by the length of p_j^i being equal to zero. Applying Lemma 3.4 repeatedly results in,

$$d_{\text{BHV}}(T_1^0, T_2^0) \geq d_{\text{BHV}}(T_1^1, T_2^0) \geq \ldots \geq d_{\text{BHV}}(T_1^{r_1}, T_2^0) \geq d_{\text{BHV}}(T_1^{r_1}, T_2^1) \geq \ldots \geq d_{\text{BHV}}(T_1^{r_1}, T_2^{r_2}).$$

Thus, $d_{\text{BHV}}(T_1^{r_1}, T_2^{r_2}) \leq d_{\text{BHV}}(T_1', T_2')$, and by construction, $(T_1^{r_1}, T_2^{r_2}) \in \lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$. $\square$

Having identified a subspace for each orthant pair that contains a minimum-distance tree pair, I can now guarantee algorithmic convergence. The goal is to minimize $d : \mathcal{T}^{\mathcal{N}} \times \mathcal{T}^{\mathcal{N}} \mapsto \mathbb{R}_{\geq 0}$. Since $\mathcal{T}^N$ is of non-positive curvature, the function d is doubly convex (Sturm, 2003, Definition 1.9, Corollary 2.5); i.e., d is a convex function on $\mathcal{T}^{\mathcal{N}} \times \mathcal{T}^{\mathcal{N}}$, which is in itself also a geodesic space (Bridson & Haefliger, 1999, Proposition 5.3).

The following result shows the search region $\lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ is convex and compact, which ensures the convergence of my proposed gradient descent method (Ruszczyński, 2006, Theorem 6.4).

Lemma 3.6. *For any two trees T_1 and T_2 with leaf sets $\mathcal{L}_1, \mathcal{L}_2 \subseteq \mathcal{N}$ and an orthant pair $(O_1, O_2) \in C_{T_1}^{\mathcal{N}} \times C_{T_2}^{\mathcal{N}}$, the orthant-specific mutually restricted extension space $\lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ is a convex, closed and bounded subspace of $\mathcal{T}^{\mathcal{N}} \times \mathcal{T}^{\mathcal{N}}$.*

Proof. $O_1 \times O_2$ is a Euclidean subspace of the geodesic space $\mathcal{T}^{\mathcal{N}} \times \mathcal{T}^{\mathcal{N}}$, since each of these orthants is a Euclidean subspaces (Remark 2.4) and this extends to the product. Additionally, $E_{T_1}^{O_1}$ and $E_{T_2}^{O_2}$ can both be described through a set of linear equations (3.3), which implies $E_{T_1}^{O_1} \times E_{T_2}^{O_2}$ is closed and convex. For each $p \in \mathcal{P}(O_1) \cup \mathcal{P}(O_2)$, define $\lceil O_1 \times O_2 \rceil^p$ to be the subset of trees that hold the condition for mutual restriction that applies to p. Each such subset can be expressed by a system of linear equations, which means each is a closed and convex subset. Thus, the mutually restricted extension space is the intersection of a finite number of closed and convex subspaces,

$$\lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil = E_{T_1}^{O_1} \times E_{T_2}^{O_2} \cap \left\{ \bigcap_{\substack{p \in \mathcal{P}(O_1) \cup \mathcal{P}(O_2) \\ \Psi_{\mathcal{L}_1}(p) = \emptyset \text{ or } \Psi_{\mathcal{L}_2}(p) = \emptyset}} \lceil O_1 \times O_2 \rceil^p \right\},$$

which is therefore a convex and closed subspace.

Finally, consider a tree pair $(T_1', T_2') \in \lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ and consider $|p|_{T_1'}$ for any edge $p \in \mathcal{P}(O_1)$. If p is consequential, there is an edge $q \in \mathcal{P}(T_1)$ such that $q = \Psi_{\mathcal{L}_1}(p)$ and $|q|_{T_1} = \sum_{p'|q=\Psi_{\mathcal{L}_1}(p)} |p'|_{T_1'}$, so that $0 \leq |p|_{T_1'} \leq |q|_{T_1}$. If p is inconsequential, then either $|p|_{T_1'} = 0$, or $|p|_{T_1'} = |p|_{T_2'}$ where p is consequential for O_2 (implying $0 \leq |p|_{T_1'} = |p|_{T_2'} \leq |q|_{T_2}$ for $q \in \mathcal{P}(T_2)$ such that $q = \Psi_{\mathcal{L}_2}(p)$). Thus, all edges in T_1' are bounded, and likewise for T_2'. Therefore, $\lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ is bounded. □

3.3.2 Distances as a reduced gradient problem

Having established properties of the search region, I now formulate the search for an optimal pair as a reduced gradient problem. Given T_1 and T_2 with leaf sets $\mathcal{L}_1, \mathcal{L}_2 \subseteq \mathcal{N}$ and orthants $O_1 \in C_{T_1}^{\mathcal{N}}$, $O_2 \in C_{T_2}^{\mathcal{N}}$ the objective is to find $(T_1^*, T_2^*) \in \lceil E_{T_1}^{O_1} \times E_{T_1}^{O_2} \rceil$ such that $d_{\text{BHV}}(T_1^*, T_2^*) \leq d_{\text{BHV}}(T_1', T_2')$ for all $(T_1', T_2') \in E_{T_1}^{O_1} \times E_{T_2}^{O_2}$. Consider the projection matrices $M_{\mathcal{L}_1}^{O_1}$ and $M_{\mathcal{L}_2}^{O_2}$ and fixed vectors $\mathbf{v}_{T_1}$ and $\mathbf{v}_{T_2}$ that describe the orthant-specific extension spaces $E_{T_1}^{O_1}$ and $E_{T_2}^{O_2}$. That is, the edge-lengths vectors $\mathbf{x}_{T_1'}$ and $\mathbf{x}_{T_2'}$ corresponding to any pair $(T_1', T_2') \in \lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ are such that $M_{\mathcal{L}_1}^{O_1} \mathbf{x}_{T_1'} = \mathbf{v}_{T_1}$ and $M_{\mathcal{L}_2}^{O_2} \mathbf{x}_{T_2'} = \mathbf{v}_{T_2}$. By Lemma 3.5, some of the elements of $\mathbf{x}_{T_i'}$ $(i = 1, 2)$ equal each other or equal zero. These values correspond to all the inconsequential edges, i.e., the zero columns in the projection matrices $M_{\mathcal{L}_1}^{O_1}$ and $M_{\mathcal{L}_2}^{O_2}$. Define the reduced matrices $\ddot{M}^i = \left[M_{\mathcal{L}_i}^{O_i}\right]_{\mathbf{R_i}}$, where $\mathbf{R_i}$ gives the indices of non-zero columns in $M_{\mathcal{L}_i}^{O_i}$. Similarly, $\ddot{\mathbf{x}}_{T_i'} := [\mathbf{x}_{T_i'}]_{\mathbf{R_i}}$ by subsetting to consequential edges in $\mathbf{x}_{T_i'}$. Note that since the only columns eliminated from the projection matrices are zero vectors, the system of linear equations defined by $M_{T_i}^{O_i} \mathbf{x}_{T_i'} = \mathbf{v}_i$ is equivalent to $\ddot{M}^i \ddot{\mathbf{x}}_{T_i'} = \mathbf{v}_i$. Therefore, defining

$$\dot{\mathbf{M}} = \begin{pmatrix} \ddot{M}^1 & 0 \\ 0 & \ddot{M}^2 \end{pmatrix}, \dot{\mathbf{x}} = \begin{pmatrix} \ddot{\mathbf{x}}_{T_1'} \\ \ddot{\mathbf{x}}_{T_2'} \end{pmatrix} \text{ and } \dot{\mathbf{v}} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix}, \tag{3.4}$$

it is possible to describe $\lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ through the system of linear equations

$$\dot{\mathbf{M}}\dot{\mathbf{x}} = \dot{\mathbf{v}} \text{ with } \dot{\mathbf{x}} \geq 0. \tag{3.5}$$

Given a solution to (3.5), trees $(T_1', T_2') \in \lceil E_{T_1}^{O_1} \times E_{T_1}^{O_2} \rceil$ can be constructed by assigning the values in $\dot{\mathbf{x}}$ corresponding to edges in O_i to the lengths of the edges in T_i', letting $|p|_{T_1} = |p|_{T_2}$ for all common edges p that are inconsequential for one of the trees, $|p|_{T_1} = |p|_{T_2} = 0$ for common edges inconsequential in both trees and $|p|_{T_i} = 0$ for uncommon edges that are inconsequential in the respective tree. Note that in this way, all values of edges in T_1' and T_2' are unambiguously defined. Let $(T_1'(\dot{\mathbf{x}}), T_2'(\dot{\mathbf{x}}))$ refer to the unique pair of trees constructed in this manner from a solution vector $\dot{\mathbf{x}}$. I summarize the above in the following result.

Lemma 3.7. *Given* $\dot{\mathbf{M}}$ *and* $\dot{\mathbf{v}}$ *for trees* $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ *and* $T_2 \in \mathcal{T}^{\mathcal{L}_2}$*, consider*

$$\dot{\mathbf{x}}^* \in \underset{\dot{\mathbf{M}}\dot{\mathbf{x}}=\dot{\mathbf{v}}, \dot{\mathbf{x}}\geq 0}{\arg\min}\, d_{BHV}\left(T_1'(\dot{\mathbf{x}}), T_2'(\dot{\mathbf{x}})\right). \tag{3.6}$$

Then $d_{BHV}(T_1'(\dot{\mathbf{x}}^*), T_2'(\dot{\mathbf{x}}^*)) \leq d_{BHV}(T_1', T_2')$ *for any* $(T_1', T_2') \in O_1 \times O_2$.

Proof. Take $\mathbf{U} = \left\{\dot{\mathbf{x}} \geq 0 | \dot{\mathbf{M}}\dot{\mathbf{x}} = \dot{\mathbf{v}}\right\}$. The function $\chi : \mathbf{U} \mapsto \lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ that maps $\chi(\dot{\mathbf{x}}) = (T_1'(\dot{\mathbf{x}}^*), T_2'(\dot{\mathbf{x}}^*))$ is bijective:

1. If $\dot{\mathbf{x}}_1 \neq \dot{\mathbf{x}}_2$ are two different solutions to (3.3), then at least one consequential edge in $T_1'(\dot{\mathbf{x}}_1)$ or $T_2'(\dot{\mathbf{x}}_1)$ has a different length than the same consequential edge in $T_1'(\dot{\mathbf{x}}_2)$ or $T_2'(\dot{\mathbf{x}}_2)$. Thus, χ is injective.

2. Given a pair of trees $(T_1', T_2') \in \lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$, construct the vector $\dot{\mathbf{x}}$ by subsetting $\mathbf{x}_{T_1'}$ and $\mathbf{x}_{T_2'}$ to only consequential edges. Since these trees are in the extension spaces, $\dot{\mathbf{x}}$ would be a solution to (3.5), so χ is surjective.

Thus, finding the minimum distance pair in $\lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$ is equivalent to solving (3.6). The result follows from Lemma 3.5. □

3.3.3 Objective function and gradient

While the goal is to minimizing $d_{\text{BHV}}(\cdot,\cdot)$, in practice I solve the equivalent problem of minimizing d^2_{BHV},

$$d^2_{\text{BHV}}\left(T_1', T_2'\right) = \sum_{i=1}^{k} \left(||A_i||_{T_1'} + ||B_i||_{T_2'}\right)^2 + \sum_{s\in K} \left(|s|_{T_1'} - |s|_{T_2'}\right)^2. \tag{3.7}$$

Define $\delta\left(\dot{\mathbf{x}}\right) := d^2_{\text{BHV}}\left(T_1'\left(\dot{\mathbf{x}}\right), T_2'\left(\dot{\mathbf{x}}\right)\right)$. Let $\dot{x}^j_p$ be the entry in $\dot{\mathbf{x}}$ corresponding to the edge $p \in \bigcup_{i=1}^{k}\{A_i \cup B_i\} \cup K$ in T_j'. By construction of $\dot{\mathbf{x}}$, $\dot{x}^j_p$ is only well-defined if p is a consequential edge for T_j. For any subset of edges $S \subseteq \mathcal{P}(T_j')$, denote by $\dot{S}$ the edges in S that are consequential for T_j, and define $||\dot{S}|| = \sqrt{\sum_{p\in\dot{S}}\left(\dot{x}^j_p\right)^2}$. Given that any inconsequential uncommon edge is of length zero for trees in $\lceil E^{O_1}_{T_1} \times E^{O_2}_{T_2}\rceil$, then $||A_i||_{T_1'(\dot{\mathbf{x}})} = ||\dot{A}_i||$ and $||B_i||_{T_2'(\dot{\mathbf{x}})} = ||\dot{B}_i||$ for every $i = 1,\ldots,k$. I also use $\dot{K}$ to refer to the set of common splits that are consequential for both trees. Since any squared term in the last sum in (3.7) involving an inconsequential edge is zero as well, the function δ is

$$\delta\left(\dot{\mathbf{x}}\right) = \sum_{i=1}^{k} \left(||\dot{A}_i|| + ||\dot{B}_i||\right)^2 + \sum_{s\in\dot{K}} \left(\dot{x}^1_s - \dot{x}^2_s\right)^2 \tag{3.8}$$

Note the support $(\mathcal{A},\mathcal{B})$ depends on $\dot{\mathbf{x}}$ indirectly through the trees $T_1'(\dot{\mathbf{x}})$ and $T_2'(\dot{\mathbf{x}})$, but the following lemma ensures that this dependence does not affect the continuity and convexity of the function. Additionally, the gradient method requires $\delta(\dot{\mathbf{x}})$ to be continuously differentiable, which I also address.

Lemma 3.8. *The function $\delta : \mathbb{R}^{\mathbf{r_1}+\mathbf{r_2}}_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$, where $\mathbf{r}_j$ is the number of non-zero columns in $M^{O_j}_{\mathcal{L}_j}$, is a continuous and convex function. Moreover, δ is continuously differentiable in the interior of the domain $\mathbb{R}^{\mathbf{r_1}+\mathbf{r_2}}_{>0}$.*

Proof. Consider the map $\chi : \mathbb{R}^{\mathbf{r_1}+\mathbf{r_2}}_{\geq 0} \mapsto O_1 \times O_2$ given by $\chi(\dot{\mathbf{x}}) = (T_1'(\dot{\mathbf{x}}), T_2'(\dot{\mathbf{x}}))$. As discussed previously, $O_1 \times O_2$ is a $(4n-6)-$dimensional Euclidean space, and by definition, each of

the $4n-6$ coordinates in the image of $\chi(\dot{\mathbf{x}})$ is either one of the values of $\dot{\mathbf{x}}$ (the value corresponding to the consequential edge) or fixed to zero. Thus, χ is a linear mapping between Euclidean spaces. And since the function $d^2_{\text{BHV}} : O_1 \times O_2 \mapsto \mathbb{R}_{\geq 0}$ is continuous and convex, then $\delta = d^2_{\text{BHV}} \circ \chi$ is also continuous and convex (Boyd & Vandenberghe, 2004, Page 79).

Note each variable $\dot{x}_p^j$ in (3.8) contributes to exactly one quadratic term. Thus, the gradient of $\delta : \mathbb{R}_{\geq 0}^{\mathbf{r_1}+\mathbf{r_2}} \mapsto \mathbb{R}_{\geq 0}$ has entries

$$\frac{\partial \delta(\dot{\mathbf{x}})}{\partial \dot{x}_p^j} = \begin{cases} 2\dot{x}_p^j \left(1 + \frac{||\dot{B}_i||}{||\dot{A}_i||}\right) & j = 1, p \in \dot{A}_i \\ 2\left(\dot{x}_p^1 - \dot{x}_p^2\right) & j = 1, p \in \dot{K} \\ 2\dot{x}_p^j \left(1 + \frac{||\dot{A}_i||}{||\dot{B}_i||}\right) & j = 2, p \in \dot{B}_i \\ 2\left(\dot{x}_p^2 - \dot{x}_p^1\right) & j = 2, p \in \dot{K}. \end{cases} \tag{3.9}$$

Since I use the unique minimal support of the geodesic between trees $T_1'(\dot{\mathbf{x}})$ and $T_2'(\dot{\mathbf{x}})$ in (3.8) and (3.9), and these trees are uniquely and well-defined by $\dot{\mathbf{x}}$, the partial derivatives given by (3.9) are well-defined as long as $||\dot{A}_i||, ||\dot{B}_i|| \neq 0$, which is the case within the domain's interior. Other support $(\mathcal{A}', \mathcal{B}')$ for the geodesic from $T_1'(\dot{\mathbf{x}})$ to $T_2'(\dot{\mathbf{x}})$ will hold the property

$$\frac{||\dot{B}_l'||}{||\dot{A}_l'||} = \frac{||B_l'||_{T_2'(\dot{\mathbf{x}})}}{||A_l'||_{T_1'(\dot{\mathbf{x}})}} = \frac{||B_i||_{T_2'(\dot{\mathbf{x}})}}{||A_i||_{T_1'(\dot{\mathbf{x}})}} = \frac{||\dot{B}_i||}{||\dot{A}_i||} \text{ when } p \in A_i \text{ and } p \in A_l',$$

$$\frac{||\dot{A}_l'||}{||\dot{B}_l'||} = \frac{||A_l'||_{T_1'(\dot{\mathbf{x}})}}{||B_l'||_{T_2'(\dot{\mathbf{x}})}} = \frac{||A_i||_{T_1'(\dot{\mathbf{x}})}}{||B_i||_{T_2'(\dot{\mathbf{x}})}} = \frac{||\dot{A}_i||}{||\dot{B}_i||} \text{ when } p \in B_i \text{ and } p \in B_l',$$

which means the partial derivative with respect to any $\dot{x}_p^i$ would be equal under equivalent supports, which implies it is unambiguously defined. Since the map $\dot{\mathbf{x}} \mapsto \frac{||A||_{T_1'(\dot{\mathbf{x}})}}{||B||_{T_2'(\dot{\mathbf{x}})}}$ for any nonempty subsets $A \subseteq \mathcal{S}(T_1')$ and $B \subseteq \mathcal{S}(T_2')$ is continuous in the interior of the domain, the partial derivatives are continuous as well. □

The continuous differentiability of δ extends beyond the interior of the domain. The gradient remains continuous at boundary points of the domain where (3.9) is well-defined,

i.e., where $||\dot{A}_i|| \neq 0$ and $||\dot{B}_i|| \neq 0$. However, the gradient does not exist at points where $\dot{x}_p^1 = 0$ for all $p \in \dot{A}_i$ or $\dot{x}_p^2 = 0$ for all $p \in \dot{B}_i$. In these cases a subgradient can be used in place of the non-existent gradient (Ruszczyński, 2006, Definition 2.72) without impacting the convergence of the algorithm. Specifically, in place of $\nabla\varphi(\dot{\mathbf{x}}^k_{\mathbf{F}^k})$ in the pausing condition $\nabla\varphi(\dot{\mathbf{x}}^k_{\mathbf{F}^k}) = 0$ (Step 7, Section 3.1), I use a subgradient, replacing every undefined gradient entry $\frac{\partial\delta(\dot{\mathbf{x}})}{\partial x_p^j}$ with zero. Note that the new (sub)gradient vector will equal zero when zero belongs to the set of subgradients $\partial\delta(\dot{\mathbf{x}}^k)$, and that a sufficient condition for attaining a minimum of a convex function is for zero to belong to this set (Ruszczyński, 2006, Theorem 3.5). The optimality condition for the global minimum (Step 8, Section 3.1) can be replaced by an equivalent condition in which I require the existence of $\eta \in \partial\delta(\mathbf{x}^k)$ such that $\eta_{\mathbf{N}^k} - \dot{\mathbf{M}}^\top_{\mathbf{N}^k}\left\{\dot{\mathbf{M}}^\top_{\mathbf{D}^k}\right\}^{-1}\eta_{\mathbf{D}^k} \geq 0$ (Ruszczyński, 2006, Theorem 3.34).

3.3.4 Algorithm for distances between extension spaces

I am now able to describe my algorithm to solve (3.1) and find an optimal pair. It starts with Algorithm 1, a reduced gradient method adapted to identify the closest trees within *orthant-specific* extension spaces.

Theorem 3.9. *Algorithm 1 converges to trees* $(T_1^*, T_2^*) \in (E^{O_1}_{T_1}, E^{O_2}_{T_2})$ *such that*

$$d_{BHV}(T_1^*, T_2^*) = \inf_{(t_1,t_2)\in(E^{O_1}_{T_1},E^{O_2}_{T_2})} d_{BHV}(t_1, t_2).$$

Proof. Algorithm 1 is a reduced gradient method to minimize $\delta : \mathbb{R}^{\mathbf{r_1}+\mathbf{r_2}}_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$ subject to constraints $\dot{\mathbf{M}}\dot{\mathbf{x}} = \dot{\mathbf{v}}$ and $\dot{\mathbf{x}} \geq 0$. This function is continuous and convex (Lemma 3.8). Consider the feasible set $\mathbf{U} = \left\{\dot{\mathbf{x}} \geq 0 | \dot{\mathbf{M}}\dot{\mathbf{x}} = \dot{\mathbf{v}}\right\}$. Given that the function $\chi : \mathbf{U} \mapsto \lceil E^{O_1}_{T_1} \times E^{O_2}_{T_2} \rceil$ that maps $\chi(\dot{\mathbf{x}}) = (T_1'(\dot{\mathbf{x}}), T_2'(\dot{\mathbf{x}}))$ is a continuous bijective linear map, and $\lceil E^{O_1}_{T_1} \times E^{O_2}_{T_2} \rceil$ is a convex, closed and bounded set, then $\mathbf{U}$ is as well. It follows that the algorithm converges to a point minimizing δ inside the feasible set (Ruszczyński, 2006, Theorem 6.4). Because

of the bijection between $\mathbf{U}$ and $\lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$, the algorithm's solution $\dot{\mathbf{x}}^*$ minimizes d_{BHV}^2 (and by extension d_{BHV}) on $\lceil E_{T_1}^{O_1} \times E_{T_2}^{O_2} \rceil$. The result follows from Lemma 3.5. □

Theorem 3.10. *For each orthant pair $(O_1, O_2) \in C_{T_1}^{\mathcal{N}} \times C_{T_2}^{\mathcal{N}}$, apply Algorithm 1 to construct a candidate pair (T_1^*, T_2^*) for the optimal pair. Among all these candidate pairs, the one with the minimum distance will be a solution to* (3.1).

Proof. Since the number of orthants in each connection cluster is finite, all orthants in it can be listed, $C_{T_i}^{\mathcal{N}} = \{O_i^1, \ldots, O_i^{c_i}\}$. Denote by $(T_{1,j_1}^*, T_{2,j_2}^*)$ the tree pair obtained by applying Algorithm 1 to $(O_1^{j_1}, O_2^{j_2})$, $j_1 = 1, \ldots, c_1$, $j_2 = 1, \ldots, c_2$. $\{d_{\text{BHV}}(T_{1,j_1}^*, T_{2,j_2}^*)\}$ is a finite set, and I define (T_1^*, T_2^*) to be the pair achieving its minimum. For any $(T_1', T_2') \in E_{T_1}^{\mathcal{N}} \times E_{T_2}^{\mathcal{N}}$, $\mathcal{O}(T_1) \subseteq O_1^{j_1}$ and $\mathcal{O}(T_2) \subseteq O_1^{j_2}$ for some j_1 and j_2. Therefore, $d_{\text{BHV}}(T_1^*, T_2^*) \leq d_{\text{BHV}}(T_{1,j_1}^*, T_{2,j_2}^*) \leq d_{\text{BHV}}(T_1', T_2')$ □

3.4 Algorithmic complexity and runtime

Having described my method in Theorem 3.10, I now turn my attention to studying its performance as the shared and total number of leaves varies. The algorithm is implemented in Java (version 20.0.2) and it is available as part of the `BHVExtMinDistance` library, which can be accessed freely from the `ExtnSpaces` repository at https://github.com/statdivlab/ExtnSpaces.git. My implementation depends on the `distanceAlg1` and `polyAlg` libraries, available at the `BHVtreespace` github repository: https://github.com/megan-owen/BHVtreespace.git. Code and instructions to reproduce the following two sections' analysis are available at https://github.com/statdivlab/ExtnSpaces_supplementary.git. To my knowledge, no other algorithms exist to find distances between extension spaces, and therefore, there are no methods to benchmark against.

Algorithm 1: A reduced gradient method to find BHV distances between orthant-specific extension spaces (see additional details in Appendix A)

Set initial values: $\dot{\mathbf{x}}_j^0 = \sum_i \mathbf{1}_{\{\dot{\mathbf{M}}[i,j]>0\}} \times \dot{\mathbf{v}}_i / \#\{j : \dot{\mathbf{M}}[i,j] = 1\}$.

Define initial index sets: For each i, add the index j of the first column such that $\dot{\mathbf{M}}[i,j] = 1$ to $\mathbf{D}$. Add all $j' > j$ such that $\dot{\mathbf{M}}[i,j'] = 1$ to $\mathbf{F}$. Set $\mathbf{N} = \emptyset$.

Initialize $c_{\text{conj}} = 1$. Set tolerance thresholds `Tol1` and `Tol2`.

while global minimum not reached **do**

  Compute gradient $\nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^t) = \nabla_{\mathbf{F}}\delta(\dot{\mathbf{x}}^t) - \dot{\mathbf{M}}_{\mathbf{F}}^\top \dot{\mathbf{M}}_{\mathbf{D}}^{-\top} \nabla_{\mathbf{D}}\delta(\dot{\mathbf{x}}^t)$

  if $||\nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^t)||_\infty <$ `Tol1` **then**

    Compute $\overline{g}_{\mathbf{N}} = \nabla_{\mathbf{N}}\delta(\dot{\mathbf{x}}^t) - \dot{\mathbf{M}}_{\mathbf{N}}^\top \dot{\mathbf{M}}_{\mathbf{D}}^{-\top} \nabla_{\mathbf{D}}\delta(\dot{\mathbf{x}}^t)$.

    if $\overline{g}_{\mathbf{N}} \geq 0$ **then**

      stop while: global minimum has been reached.

    else

      Define $\mathbf{N}_p = \{j \in \mathbf{N} \mid \overline{g}_{\mathbf{N}}[j] < 0\}$

      Update $\mathbf{F} = \mathbf{F} \cup \mathbf{N}_p$ and $\mathbf{N} = \mathbf{N} \setminus \mathbf{N}_p$

    end if

  end if

  Compute $\mathbf{d}_{\mathbf{F}}^t = -\nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^t) + \mathbf{1}_{\{c_{\text{conj}}=1\}} \frac{\langle \nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^t), \nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^t) - \nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^{t-1}) \rangle}{||\nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^{t-1})||^2} \mathbf{d}_{\mathbf{F}}^{t-1}$

  Set $\mathbf{d}_{\mathbf{D}}^t = -\dot{\mathbf{M}}_{\mathbf{D}}^{-1} \dot{\mathbf{M}}_{\mathbf{F}} \mathbf{d}_{\mathbf{F}}^t$ and $\mathbf{d}_{\mathbf{N}}^t = 0$. Increase $c_{\text{conj}} \leftarrow c_{\text{conj}} + 1$

  Find $\tau_{\max} = \max\{\tau \geq 0 \mid \dot{\mathbf{x}}^t + \tau \mathbf{d}^t \geq 0\}$ and $h(\tau_{\max}) = \langle \nabla\delta(\dot{\mathbf{x}}^t + \tau_{\max}\mathbf{d}^t), \mathbf{d}^t \rangle$

  if $h(\tau_{\max}) \leq 0$ **then**

    Set $\tau_0 = \tau_{\max}$

  else

    Set $\tau_{\text{left}} = 0$, $\tau_{\text{right}} = \tau_{\max}$ and $\tau_0 = \frac{\tau_{\text{left}} + \tau_{\text{right}}}{2}$

    while $|h(\tau_0) = \langle \nabla\delta(\dot{\mathbf{x}}^t + \tau_0 \mathbf{d}^t), \mathbf{d}^t \rangle| >$ `Tol2` **do**

      if $h(\tau^*) > 0$ **then** set $\tau_{\text{right}} = \tau^*$ and $\tau^* = \frac{\tau_{\text{left}} + \tau_{\text{right}}}{2}$

      else set $\tau_{\text{left}} = \tau^*$ and $\tau^* = \frac{\tau_{\text{left}} + \tau_{\text{right}}}{2}$.

    end while

  end if

  if $\dot{\mathbf{x}}_j^t + \tau_0 d_j^t = 0$ for some $j \in \mathbf{D} \cup \mathbf{F}$ **then**

    Select $j \in \mathbf{D} \cup \mathbf{F}$ such that $\dot{\mathbf{x}}_j^t + \tau_0 d_j^t = 0$

    if $j \in \mathbf{F}$ **then**

      Set $\mathbf{F} = \mathbf{F} \setminus \{j\}$ and $\mathbf{N} = \mathbf{N} \cup \{j\}$

    else

      Select $j' \in \mathbf{F}$ such that $\dot{\mathbf{M}}[i,j] = \dot{\mathbf{M}}[i,j'] = 1$ for some index i

      Set $\mathbf{D} = \mathbf{D} \setminus (\{j\} \cup \{j'\})$, $\mathbf{F} = \mathbf{F} \setminus \{j'\}$ and $\mathbf{N} = \mathbf{N} \cup \{j\}$

    end if

  end if

  Update $\dot{\mathbf{x}}^{t+1} = \dot{\mathbf{x}}^t + \tau^* \mathbf{d}_t$

  if $c_{\text{conj}} + 1 > 15$ **then** $c_{\text{conj}} = 1$ **else** $c_{\text{conj}} \leftarrow c_{\text{conj}} + 1$

end while

return $(T_1'(\dot{\mathbf{x}}^t), T_2'(\dot{\mathbf{x}}^t))$ and $\sqrt{\delta(T_1'(\dot{\mathbf{x}}^t), T_2'(\dot{\mathbf{x}}^t))}$

Since my algorithm performs an optimization routine for each orthant pair, the total number of orthant pairs is a major driver of the complexity of my algorithm. The number of orthants in the extension space in $\mathcal{T}^{\mathcal{N}}$ ($|\mathcal{N}| = n$) of a tree with $|\mathcal{L}_i| = l_i$ leaves is $(2n-5)!!/(2l_i-5)!!$ (Ren et al., 2017, Theorem 2.1), and therefore, the number of orthant pairs considered is $\Omega := \{(2n-5)!!\}^2/\{(2l_1-5)!! \times (2l_2-5)!!\}$. This value has the potential to be considerably large, since the growth rate of the value $(2n-5)!!$ is super-exponential. Using Stirling's approximation (for large l_1, l_2 and n)

$$\Omega \sim \left(2e^{-1}\right)^{2n-l_1-l_2} (n-2)^{2n-4}(l_1-2)^{2-l_1}(l_2-2)^{2-l_2}.$$

and therefore, $\Omega = O(n^{2n-(l_1+l_2)})$ when $n-l_1$ and $n-l_2$ are constant.

Table 3.1. The runtime of Algorithm 1 in practice. For each setting $\mathcal{S}$, I report the number of pairs of orthants to search over (Ω); the total runtime for computing distances between $E_{T_1}^{\mathcal{N}}$ and $E_{T_2}^{\mathcal{N}}$ (min:sec); the number of iterations for each reduced gradient method to converge (mean [median, 90% quantile and maximum]); the distance between $E_{T_1}^{\mathcal{N}}$ and $E_{T_2}^{\mathcal{N}}$; and the number of optimal pairs ("# pairs"). I observe that the number of orthant pairs is the largest factor contributing to runtime.

$\mathcal{S}$	$\lvert\mathcal{L}_1 \cup \mathcal{L}_2\rvert$	$\lvert\mathcal{L}_1\rvert$	$\lvert\mathcal{L}_2\rvert$	Ω	Runtime	Iterations	$d(E_{T_1}^{\mathcal{N}}, E_{T_2}^{\mathcal{N}})$	# pairs
Unimodal distribution for Edge Lengths								
a	7	6	4	2835	00:05	5.50 [5, 7, 174]	6.675	1
b	7	6	4	2835	00:03	4.94 [5, 7, 22]	4.378	1
c	7	5	4	19845	00:34	6.41 [6, 10, 64]	0.268	1
d	10	9	8	2925	00:10	3.96 [4, 6, 17]	18.497	1
e	10	8	8	38025	02:27	4.98 [5, 8, 46]	15.710	1
f	10	8	7	418275	50:17	7.33 [6, 12, 68]	9.459	1
Bimodal distribution for Edge Lengths								
a	7	6	4	2835	00:04	4.56 [4, 7, 28]	108.284	6
b	7	6	4	2835	00:02	4.60 [5, 6, 18]	108.667	1
c	7	5	4	19845	00:29	6.00 [6, 9, 36]	24.880	1
d	10	9	8	2925	00:08	4.03 [4, 6, 23]	132.690	1
e	10	8	8	38025	02:27	4.97 [5, 8, 40]	104.722	2
f	10	8	7	418275	47:45	7.53 [6, 12, 91]	123.323	2

I study the in-practice scalability and performance of the algorithm using simulated

phylogenetic trees. I selected 6 pairs of topologies each with a different combination of $\mathcal{L}_1 \cup \mathcal{L}_2$, $\mathcal{L}_1$ and $\mathcal{L}_2$, and considered $\mathcal{N} = \mathcal{L}_1 \cup \mathcal{L}_2$. I consider two distributions for the edge lengths, resulting in 12 total simulation settings. The first edge length distribution is a lognormal distribution (mean = 5, variance = 1), reflecting a low-variance edge length scenario. The second distribution is a mixture of two lognormal distributions. The first component has a mean of 5 and a variance of 1 (sampled with 75% probability), and the second component has a mean of 60 and a variance of 10 (sampled with 25% probability). This second, high-variance distribution reflects long-branch scenarios that commonly arise in practice.

The results of the exploration can be found in Table 3.1. Running clock-times are based on an 8-core Apple M1 processor with 16GB of RAM. These times reflect the process of computing the distances for each orthant pair sequentially, and applying a merge sort algorithm. My library supports multi-threading, allowing distances in different orthant pairs to be computed concurrently to reduce run-times, but I report single-thread times here for transparency. For this simulation, I selected a tolerance for $\nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}})$ of 10^{-8}.

As my algorithm performs an optimization process per each orthant pair, I expected the runtime to be approximately proportional to the number of orthant pairs for a fixed search space dimension, which I generally find to be the case. For example, when the number of orthant pairs increased by a factor of 7 (from 2835 pairs in (a) and (b) to 19845 in (c)), the increase in runtime was approximately 7-fold, from 2-5 seconds to around 29-34 seconds. Similarly, when the number of orthant pairs increased by 11-fold (from (e) to (f)), runtimes increased by $\sim$ 11-fold, and when the number of orthant pairs increased by 13-fold (from (d) to (e)), runtimes increased by $\sim$ 16-fold.

The number of orthant pairs appears to affect the runtime through two avenues: directly (via the number of reduced gradient runs to be performed), and indirectly (due to an increase in the average number of iterations required for convergence). The number of iterations to

convergence is directly influenced by the number of leaves being attached to each tree to create their extension space, as each reduced gradient problem involves the linear constraints $\dot{\mathbf{M}}\dot{\mathbf{x}} = \dot{\mathbf{v}}$ with $(r_1 + r_2) - \{2(l_1 + l_2) - 6\}$ (the difference between the number of consequential edges and the number of original edges in T_1 and T_2) degrees of freedom, which is upper-bounded by the number of leaves being added to the trees. Consequentially, the number of free variables in each iteration of my optimization mechanism is at most $2n - l_1 - l_2$. Convergence will take longer when this number is higher.

Unsurprisingly, I find that distances between extension spaces tend to be higher when edge lengths are heavy-tailed (bimodal distribution). Interestingly, while in all unimodal distribution cases, only one optimal pair was found, it was common to find more than one optimum under the bimodal distribution. Both of these observations can be explained by how BHV distances depend on edge lengths. As described in Section 2.2.2, when going from a tree t_1 to t_2, the common edges are present in the topologies of all trees on the geodesic, while uncommon edges present in t_1 are gradually swapped for uncommon edges in t_2. Intuitively, the size of the uncommon edges in t_1 and t_2 gradually change between zero and their original size. Similarly, the lengths of common edges gradually change from their lengths in t_1 to their lengths in t_2. Thus, BHV distances are increased by longer uncommon edges, and by common edges with significantly different lengths. The bimodal distribution allows for longer uncommon edges, and introduces more variability in the sizes of the common edges, which explains why the distances are higher.

Another effect of edge lengths on BHV distances is that, in practice, if an edge in a tree is decidedly longer than the others, topology orthants in the connection cluster that involve attaching new edges to this edge tend to produce shorter geodesics. If this particularly long edge is such that edges mapping into it (under the tree dimensionality reduction map) are uncommon edges, then the large value of the length must reduce to zero at some point along the geodesic. Thus, dividing this long edge into several smaller edges through attaching

edges into it will reduce the size of the geodesic. This also explains why more than one optimal pair was found in some of the cases where edge lengths were assigned through the bimodal distribution. If one of the edges in one of the trees is long, then the best candidates for the optimal pair arise from those attaching leaves along that edge, and the same geodesic length can be achieved by attaching the edges in the same places along the long edge, but in a different order. For example, T_1 (case (a), bimodal) has a long external edge incident to the leaf **L06**, and all 6 optimal pairs are such that T_2^* has edges **L03**, **L04** and **L05** attached to that edge. Thus, I obtain 6 optimal pairs because there 6 different ways these three external edges can be ordered across the long edge.

3.5 Application to prokaryotic gene trees

Here, I illustrate my method on gene trees spanning phylogenetically diverse prokaryotic lineages. Prokaryotes (bacteria and archaea) display a high degree of discordance in the genes they carry, with fewer than $\sim$1% of a given organism's genes distributed "universally" across all bacteria (Dagan & Martin, 2006). Thus, the comparison of two prokaryotic gene phylogenies will almost always require tools that can handle non-identical leaf sets, motivating the development of my method.

I analyze gene trees from Zhu et al. (2019), focusing on two genes involved in essential tasks: cell division and repair. Specifically, I consider T_1 to be the gene tree for *ftsA* (coding for a protein involved in cell division) and T_2 to be the gene tree for *dinB* (coding for a DNA polymerase protein involved in translesion repair). I restrict my analysis to 10 phylogenetically diverse organisms spanning 2 domains of life; the complete leaf set $\mathcal{N}$ is given in Table 3.2. These organisms are found in diverse habitats, including the human gut, oral cavity, and tumors; as well as groundwater, treated water, and deep-sea hydrothermal vents. Out of ten total organisms, only five organisms have both genes, with 3 and 2 unshared

genes carried by *ftsA* and *dinB*, respectively. Note that these genes could be truly absent, or they could be unobserved due to imperfections in genome reconstruction from metagenomes (Duarte et al., 2020; Royalty & Steen, 2019; Zaheer et al., 2018).

Table 3.2. The complete leaf set $\mathcal{N}$ for the *ftsA* and *dinB* gene trees.

Species	Domain	*ftsA* tree	*dinB* tree
Actinomyces odontolyticus	Bacteria	No	Yes
Fusobacterium nucleatum	Bacteria	Yes	Yes
Pseudomonas pelagia	Bacteria	Yes	Yes
Bacteroides fragilis	Bacteria	Yes	Yes
Candidatus Saccharibacteria TM7x	Bacteria	Yes	No
Sphingomonas hengshuiensis	Bacteria	Yes	Yes
Parcubacteria SG8-24	Bacteria	Yes	No
Vibrio scophthalmi	Bacteria	Yes	Yes
Candidatus Lokiarchaeota CR4	Archea	No	Yes
Candidatus Odinarchaeota LCB4	Archea	No	Yes

T_1 and T_2 are shown in Figure 3.1(a). While $E^{\mathcal{N}}_{T_1}$ spans 2145 orthants, and $E^{\mathcal{N}}_{T_2}$ spans 195 orthants, none of these orthants are shared between the two extension spaces. As a result, the compatibility measures of Grindstaff and Owen (2019) are not defined for these two trees. In contrast, my distance $d(E^{\mathcal{N}}_{T_1}, E^{\mathcal{N}}_{T_2})$ is always defined. I applied Algorithm 1 to every orthant pair in $C^{\mathcal{N}}_{T_1} \times C^{\mathcal{N}}_{T_2}$ and found that the distance between the extension spaces is 4.234, and that this value was attained as the distance between the trees in $\mathcal{T}^{\mathcal{N}}$ shown in Figure 3.1(b). To search through 418275 pairs of orthants in $\mathcal{T}^{\mathcal{N}}$ took 21 minutes on a 8-core Apple M1 processor with 16GB of RAM in multi-threaded setting with a thread pool of size 8.

In addition, my approach to computing distances between trees with non-identical leaf sets naturally lends itself to a "supertree" method for combining information from two trees. Specifically, because the minimum distance between extension spaces is the size of geodesic paths in $\mathcal{T}^{\mathcal{N}}$ the tree in $\mathcal{T}^{\mathcal{N}}$ that is the midpoint on such a geodesic averages the information

across the true trees (Brown & Owen, 2020; Miller et al., 2015). However, because minimum distance paths between extension spaces are not necessarily unique, this supertree measure may not be unique. This is in contrast with Fréchet means of trees that are all contained within the same BHV space, whose Fréchet means are uniquely defined (Sturm, 2003).

Two paths were minimum distance between the extension spaces of the *ftsA* and *dinB* trees. Similar to the cases discussed in the previous section, the two trees (T_A and $T_{A'}$ in Figure 3.1(b)) in the extension space for *ftsA* are both produced by attaching new edges (corresponding to external edges to *Ca. Lokiarchaeota CR4* and *Ca. Odinarchaeota LCB4*) to a particularly long edge (the external edge to *S. hengshuiensis*) in the same locations but in a different order — a phenomenon discussed at the end of Section 3.4. The internal edges resulting from attaching these edges (with lengths 0.72 and 1.30) are in both cases uncommon to the tree $T_B \in E^{\mathcal{N}}_{T_2}$, and thus these edges reduce to zero in length and are then dropped. Because these edges have the same lengths in T_A and $T_{A'}$ and all other edges are equal length, the tree along the geodesics where the last of these two edges (the edge separating *Ca. Lokiarchaeota CR4*, *Ca. Odinarchaeota LCB4* and *S. hengshuiensis* from all other organisms) is dropped is the same, and then the geodesic follows the same path to T_B. Although theoretically the existence of two optimal pairs could admit two supertrees, in this case the length of the geodesic section from the starting tree (either T_A or T'_A) to the tree where both geodesics coincide is less than half the length of the total geodesic. Because of this, the mid-point is unchanged between T_A and T'_A. This unique supertree is shown in Figure 3.1(c). If the length of the section where both geodesics do not coincide were longer than half the length of the geodesics, then the midpoints would differ from each other. Nevertheless they would still share many edges in common and with the same length, and uncommon edges would have a counterpart with the same length. Thus, I expect non-unique supertrees to be similar in general.

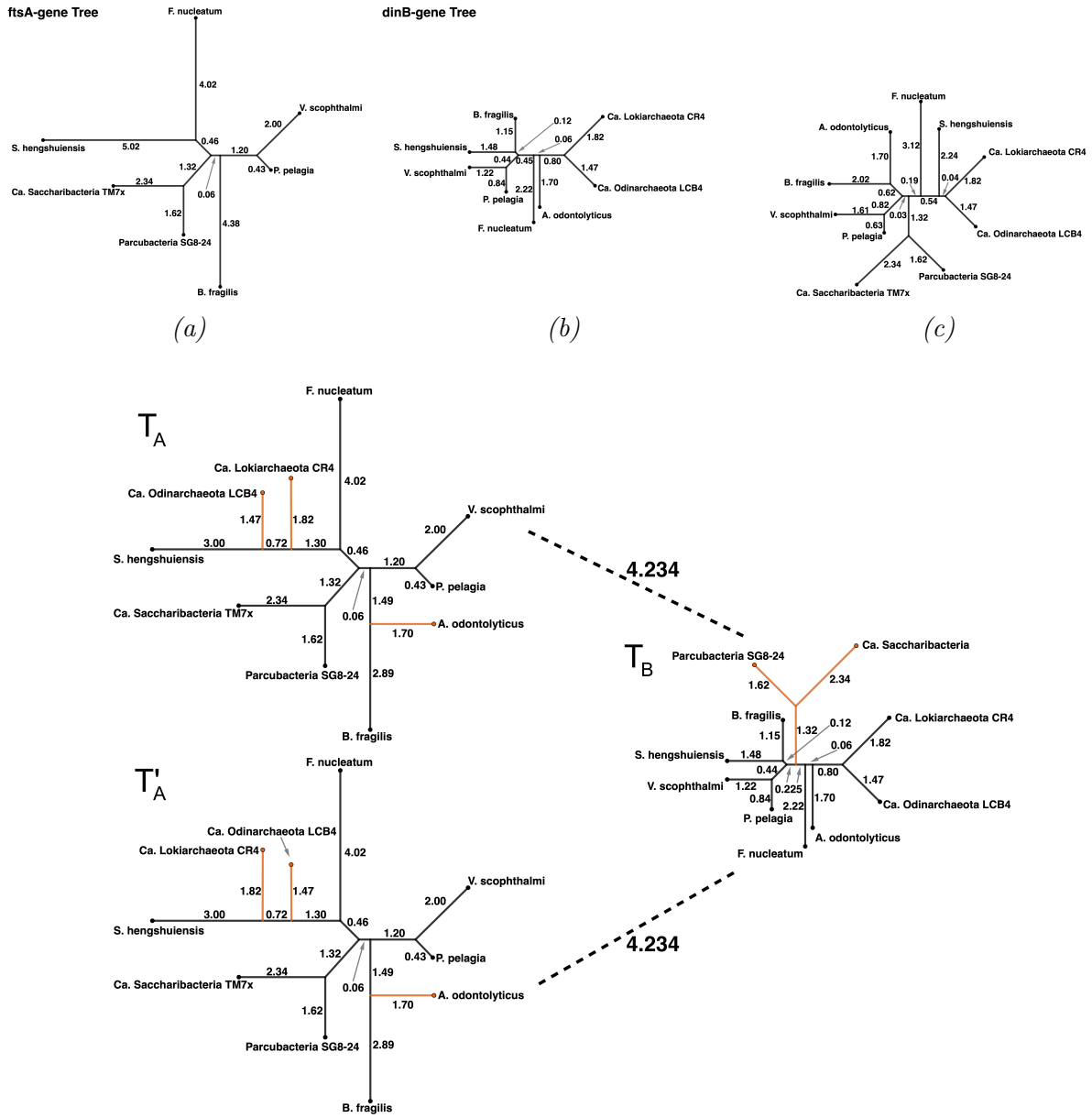


Figure 3.1. The estimated evolutionary history of (a) the *ftsA* gene and (b) the *dinB* gene for 10 organisms. (c) The midpoint of the geodesics between (T_A, T_B) and (T'_A, T_B). This midpoint is the same for both geodesics. (d) The minimal distance between the extension spaces of these trees is 4.234, which can be obtained via two geodesic paths. The two tree pairs (T_A, T_B) and (T'_A, T_B) that achieve the minimal distance are shown.

3.6 Discussion

Extension spaces provide an intuitive approach to contextualizing phylogenetic trees with reduced leaf sets in higher-dimensional BHV spaces. In this chapter, I proposed a method for finding the minimum distance between extension spaces by implementing a reduced gradient algorithm, as a comparison method between trees with non-identical leaf sets. A major advantage of this approach is that it gives a measure of dissimilarity applicable to any two trees. It therefore addresses some of the limitations in Grindstaff and Owen (2019), such as that the trees under comparison must share common orthants in their extension spaces, and must have all internal branches of strictly positive length. This methodology has the potential to be applied in future studies to evaluate whether two trees share a compatible evolutionary history or to assess the performance of tree reconstruction methods by comparing original versus reconstructed trees in simulation studies.

An additional advantage of my approach is that it enables construction of a "supertree" that summarizes a pair of trees with respect to topological and edge length differences, even when those trees have non-identical leaf sets. This suggests a measure of compatibility among a collection of trees as the Fréchet means of their extension spaces $E^{\mathcal{N}}_{T_1}, E^{\mathcal{N}}_{T_2}, \ldots, E^{\mathcal{N}}_{T_r}$, which for $r = 2$ reduces to a midpoint along a geodesic. One of the primary benefits of utilizing the BHV space is its ability to provide a measure of variability and formalize uncertainty by examining the distances between trees in the sample, as opposed to only obtaining a single consensus tree.

Note that as minimal distance paths between extension spaces are not necessarily unique, Fréchet means of extension spaces are also not necessarily unique. Interestingly, in my applied data example, while there were two minimal distance paths between the trees, both paths had the same midpoint. I conjecture that in applied data examples, Fréchet means of extension spaces may often be unique. I leave the construction and study of Fréchet means

to future work.

This algorithm enables the comparison of phylogenetic trees without the limitations of considering only subtrees with common leaves. Existing tools that focus solely on common leaves require discarding valuable information collected during data acquisition. While this approach may suffice for datasets with extensive taxonomic diversity, it becomes a concern when available data is disregarded due to tool limitations rather than developing tools to accommodate real data structure. Additionally, in analyses involving multiple trees, the common leaves often represent only a small fraction of the total taxonomic dataset. Thus, the algorithm I present allows biologists to compare trees with non-identical leaf sets without pre-pruning, providing a more accurate representation of biological complexity.

While finding the minimal BHV distance between extension spaces is highly intuitive, my proposed distance is not formally a metric between trees (Grindstaff & Owen, 2019, Section 3.4). Specifically, distinct trees can have intersecting extension spaces (Grindstaff & Owen, 2019, Example 4.1), and therefore a zero distance, violating positivity. Furthermore, it is possible to find trees T_1, T_2 and T_3 for which $d(E^{\mathcal{N}}_{T_1}, E^{\mathcal{N}}_{T_3}) = 0$ and $d(E^{\mathcal{N}}_{T_2}, E^{\mathcal{N}}_{T_3}) = 0$, but for which $d(E^{\mathcal{N}}_{T_1}, E^{\mathcal{N}}_{T_2}) > 0$, thus violating the triangle inequality (Grindstaff & Owen, 2019, Remark 3.6). Despite this, my distance still provides a useful measure of similarity between phylogenies with non-identical leaf sets, and when $\mathcal{L}_1 = \mathcal{L}_2 = \mathcal{N}$, my distance reduces to the classical BHV distance. In addition, the algorithm I developed is broadly applicable to the minimization of any convex function over a subset of a BHV space defined by linear constraints (see also Miller et al. (2015)), which could be broadly useful in other mathematical or computational phylogenetics problems.

In practice, my algorithm runs within an hour on a modern laptop for up to 10 total leaves without multithreading. Alas, computation time grows quickly in the total number of leaves. For example, if the largest example in Table 3.1 had one more leaf not included in the second tree (increasing $|\mathcal{L}_1|$ to 9 and $|\mathcal{L}_1 \cup \mathcal{L}_2|$ to 11), the number of orthant pairs would grow

by a factor of ~ 22, from 418,275 to 9,298,575, for which I estimate a single-thread runtime of $\sim$ 24 hours. That said, the method can be trivially parallelized across orthant pairs, making it well-suited to distributed computing environments. While I report single-thread runtimes for transparency, my open-source software package implements multi-threading, conveniently accelerating the method for typical (non-distributed) computing environments.

Because extension spaces can be characterized within a given orthant as a linear system of equations, my algorithm employs a reduced gradient method. Reduced gradient methods are iterative procedures, and therefore their computational complexity is challenging to characterize. That said, in practice,I find that the number of iterations per orthant pair is generally low, with 50% converging with a gradient of $< 10^{-8}$ within 4-6 iterations and 90% converging within 6-12 iterations. Each iteration, however, involves the computation of multiple BHV geodesics. As computing a geodesic is $O(n^4)$-time (Owen & Provan, 2011, Theorem 3.5), each iteration of Algorithm 1 is $O(n^4 d)$, where d is the number of iterations required for the line search. The number of orthant pairs grows at $O(|\mathcal{N}|^c)$, where c is the number of leaves to be added to the trees, further contributing to the runtime of the algorithm. Unsurprisingly, in practice, I find that the number of orthant pairs, rather than the geodesic computations, is the major limiting factor in calculating my distance. As a result, future work to accelerate computation could consider excluding suboptimal orthant pairs from consideration, such as by excluding highly dissimilar topologies while prioritizing the attachment of inconsequential edges to long edges. That said, as previously mentioned, orthant pair comparisons can be parallelized across distributed computing architecture, reducing the in-practice computation time by a factor equal to the number of machines available.

Finally, I acknowledge that there are metrics for phylogenetic trees that focus exclusively on topological comparisons, including adaptations like the generalized Robinson-Foulds distance (Briand et al., 2020) for trees with non-identical leaf labels. While these methods provide valuable insights into the structural relationships between trees, they overlook the

evolutionary information conveyed by branch lengths. Incorporating branch lengths into phylogenetic analysis offers several advantages, such as assessing evolutionary distances, identifying varying rates of evolution, and highlighting functional or phenotypic differences between organisms.

Adapting the methods presented in this dissertation for scenarios where only topological comparisons are of interest is beyond the scope of this project. Furthermore, evaluating the effectiveness of such adaptations compared to methods specifically designed for topological analysis would require additional research.

CHAPTER 4

THE TOWERING TREE SPACE

My primary goal for this chapter is the construction of a metric space for phylogenetic trees with non-identical leaf sets that, similar to the BHV space, takes into account both topological and edge length differences between trees. I introduce the Towering Tree Spaces, a family of metric spaces, for this purpose. These spaces are constructed as unions of BHV spaces, allowing for transitions between them.

Throughout this chapter, I refer to the action of transitioning from one BHV space to a another on fewer leaves as "moving to a lower BHV level". Similarly, transitioning to a BHV space with a larger leaf set is referred to as "moving to a higher BHV level". Thus, the union of BHV spaces is referred to as a "tower" of tree spaces. I formalize this with the following definition.

Definition 4.1. Given two BHV spaces $\mathcal{T}^{\mathcal{L}}$ and $\mathcal{T}^{\mathcal{L}'}$, $\mathcal{T}^{\mathcal{L}}$ is a **higher level** than $\mathcal{T}^{\mathcal{L}'}$ whenever $\mathcal{L}' \subset \mathcal{L}$. In this case, $\mathcal{T}^{\mathcal{L}'}$ is a **lower level** than $\mathcal{T}^{\mathcal{L}}$.

I begin by introducing common structures and properties among the family of towering spaces. Members of the family will be indexed by an operation used to combine edge lengths when transitioning between BHV levels. There will be multiple options for this combining operation, but later I suggest the selection of one of these options based on its nice geometrical and computational properties.

4.1 Preliminary structures

Here I provide the first steps towards the construction of a towering tree space, and tools for analysis within these spaces. For a set of n leaves $\mathcal{N}$, consider all possible subsets $\mathcal{L} \subset \mathcal{N}$, $|\mathcal{L}| \geq 3$, and corresponding BHV spaces $\mathcal{T}^{\mathcal{L}}$. A towering tree space is defined on the union

of these BHV spaces $\mathcal{T}^{\mathbf{P}(\mathcal{N})} = \bigcup_{\mathcal{L}\in\mathbf{P}(\mathcal{N})} \mathcal{T}^{\mathcal{L}}$, where $\mathbf{P}(\mathcal{N}) = \{\mathcal{L} \subseteq \mathcal{N} \mid |\mathcal{L}| \geq 3\}$. To simplify working with several BHV spaces, I maintain the convention given in Remark 2.3 for trees throughout this chapter. Unless indicated otherwise, a tree T is represented by the tree in the lowest-dimensional orthant possible in its BHV space; that is, all internal edges of length zero are dropped and $\mathcal{S}(T)$, includes only splits with a positive length. In contrast, external edges may have length zero. These external edges are only "dropped" from a tree when transitioning from one BHV space to another.

4.1.1 Topological transitions between BHV spaces

Transitions between BHV spaces in a towering space are through trees in the same equivalence class. To define equivalence classes, I introduce pruning and re-grafting operations. As the name suggests, pruning leaves involve removing some of the edges in a tree containing these leaves, while regrafting involves attaching new edges with new leaves. The idea of removing and adding leaves is a long-standing technique in phylogenetics; it serves as the basis for the subtree prune-and-regraft distance between tree topologies (Hein et al., 1996; Whidden et al., 2014) and it also plays a role in the description of extension spaces, as seen in Grindstaff and Owen (2019, Definition 3.1). However, my operations differ from previous examples in *when* pruning can be performed and *how* edge lengths are merged.

Similar to dropping zero-length *internal* edges when transitioning between *topology orthants* in BHV space, zero-length external edges drop when transitioning between *BHV spaces*. The trees where leaf prunes may be performed are defined below.

Definition 4.2. Given a subset of leaves $\mathcal{M} \subset \mathcal{L}$ and $T \in \mathcal{T}^{\mathcal{L}}$, the leaves in $\mathcal{M}$ are **mutually prunable** from T if each external edge to a leaf in $\mathcal{M}$ is of length zero, and every internal edge of T maps to an internal split of the leaves $\mathcal{L}' = \mathcal{L} \setminus \mathcal{M}$ under the TDR map. This is, for every $s \in \mathcal{S}(T)$, $\Psi_{\mathcal{L}'}(s) = [\mathcal{L}_1 \ddagger \mathcal{L}' \setminus \mathcal{L}_1]$ such that $|\mathcal{L}_1|, |\mathcal{L} \setminus \mathcal{L}_1| \geq 2$.

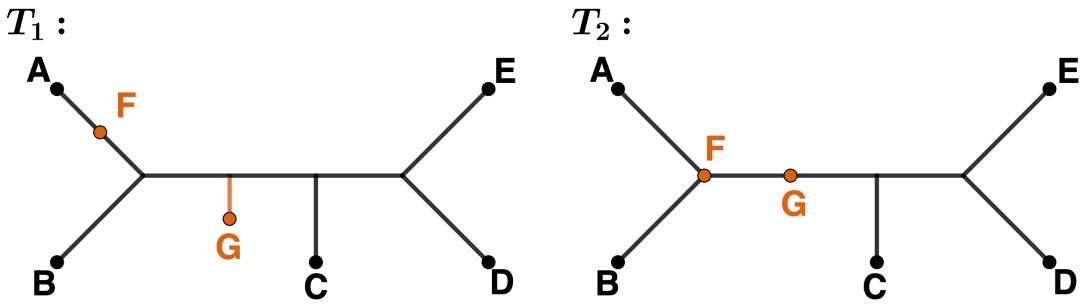


Figure 4.1. An example of mutually prunable leaves. The set $\mathcal{M} = \{F, G\}$ is not mutually prunable from T_1 but it is mutually prunable from T_2. In T_1, the external edge towards G is of positive length, and the internal edge $[\{A, F\} \ddagger \{B, C, D, E, G\}] \in \mathcal{S}(T_1)$ maps to an external edge under the TDR map.

If $\mathcal{M}$ is not mutually prunable from T, the **set of edges preventing $\mathcal{M}$ from being prunable from** T is a set of interest, which I denote by

$$P^{\downarrow\mathcal{M}}(T) = \{p \in \mathcal{P}(T) \mid \Psi_{\mathcal{L}'}(p) = \emptyset \text{ or } \Psi_{\mathcal{L}'}(p) = [\ell \ddagger \mathcal{L}' \setminus \{\ell\}] \text{ for some leaf } \ell\}\,.$$

Remark 4.3. A subset of leaves $\mathcal{M}$ with zero-length external edges on T is mutually prunable from that tree if and only if for every non-empty subset $\mathcal{M}' \subseteq \mathcal{M}$, the edges $[\mathcal{M}' \ddagger \mathcal{L} \setminus \mathcal{M}']$ and $[\mathcal{M}' \cup \{\ell\} \ddagger \mathcal{L} \setminus (\mathcal{M}' \cup \{\ell\})]$ for all $\ell \in \mathcal{L} \setminus \mathcal{M}$ do *not* belong to $\mathcal{S}(T)$. This implies $P^{\downarrow\mathcal{M}}(T)$ only contains external edges, and these are of length zero in T.

The definition of mutually prunable ensures that all internal edges map to another internal edge. Thus, when assessing the similarity of BHV topologies, each internal edge affects the topology of the tree after the transformation, since the topology of a tree is fully defined by its internal edges. Moreover, this restriction aids in maintaining a strictly positive distance between trees that are clearly distinct to each other: trees with strictly positive external edges will be at a positive distance from each other. As shown in the next remark, relaxing this condition could result in counter-intuitive results.

Remark 4.4. If prunes and regrafts could be performed from and onto external edges, then a regraft of new leaves could happen at the endpoint of the external edge corresponding to the

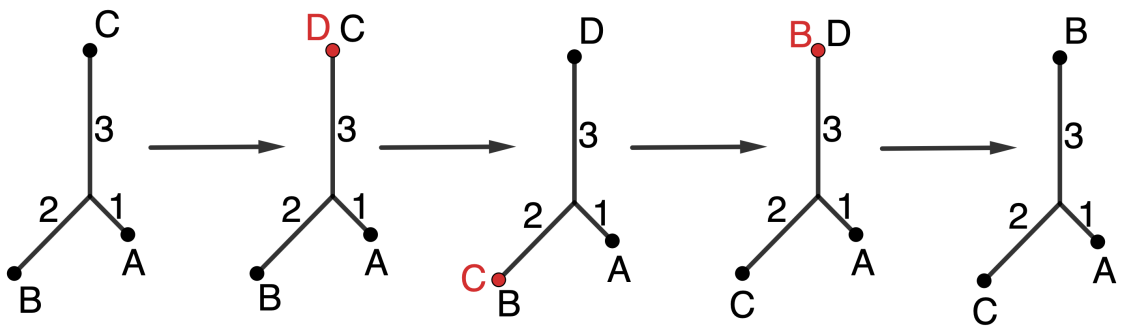


Figure 4.2. Example of "leaf swapping" via pruning and regrafting at external edges endpoints. By a series of prunes and regrafts, leaves B and C in the first tree get swapped. Between the first and second tree, the new leaf D is regrafted directly next to leaf C, Then C is pruned and regrafted next to B, which is later pruned and regrafted next to D. D is finally pruned, producing a new tree with the leaves B and C in opposite positions.

leaf node, allowing for "leaf swaps". As an example, consider the trees shown in Figure 4.2, in which through a series of prunes and regrafts, the positions of two leaves are swapped. Under these transformations, all trees shown in the this figure would be at distance zero, which is clearly undesirable.

As leaf prunings form the basis for moving between levels of towering space, trees with prunable leaves form an important subspace. Later in the chapter, I discuss finding the nearest tree where a prune can be performed, which is crucial for identifying short paths in towering space. For this, I introduce the following subspace, followed by some of its properties.

Definition 4.5. Consider the BHV space $\mathcal{T}^{\mathcal{L}}$ and a subset of leaves $\mathcal{M} \subset \mathcal{L}$. The **$\mathcal{M}$-trimmable subspace**, denoted by $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$, is the subset of trees in $\mathcal{T}^{\mathcal{L}}$ where $\mathcal{M}$ is mutually prunable.

Lemma 4.6. *For any BHV space $\mathcal{T}^{\mathcal{L}}$ and a subset of leaves $\mathcal{M} \subset \mathcal{L}$, the $\mathcal{M}$-trimmable subspace $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$ is convex and closed.*

Proof. Refer to Appendix B.1. □

From this lemma, the corollary below is directly obtained by Sturm (2003, Proposition 2.6).

Corollary 4.7. *Given the $\mathcal{M}$-trimmable subspace $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$, for every tree $T \in \mathcal{T}^{\mathcal{L}}$ there exists a unique tree $t^* \in \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$ such that $d_{BHV}(T, t^*) = \inf_{t \in \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}} d_{BHV}(T, t)$. Moreover, $d^2_{BHV}(T, t) \geq d^2_{BHV}(T, t^*) + d^2_{BHV}(t^*, t)$ for every $t \in \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$.*

Definition 4.8. Given the $\mathcal{M}$-trimmable subspace $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$, for every tree $T \in \mathcal{T}^{\mathcal{L}}$, **the projection of T onto $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$**, is defined as the unique tree $T^{\perp\mathcal{M}} \in \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$ such that

$$d_{\mathrm{BHV}}(T, T^{\perp\mathcal{M}}) = \inf_{t \in \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}} d_{\mathrm{BHV}}(T, t).$$

The projection of any tree $T \in \mathcal{T}^{\mathcal{L}}$ onto the $\mathcal{M}$-trimmable subspace can directly be found by focusing on the positive-length external edges to leaves in $\mathcal{M}$ and the internal edges that do not map into internal edges under the TDR map $\Psi_{\mathcal{L}\setminus\mathcal{M}}$; that is, the edges $P^{\downarrow\mathcal{M}}(T)$ preventing $\mathcal{M}$ from being mutually prunable from T. I show in the following result how to compute the distance between trees and their projections.

Lemma 4.9. *For $T \in \mathcal{T}^{\mathcal{L}}$ and a subset of leaves $\mathcal{M} \subset \mathcal{L}$, consider the set of edges $P^{\downarrow\mathcal{M}}(T)$. The projection of T onto $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$ is the tree $T^{\perp\mathcal{M}}$ with topology given by internal edges*

$$\mathcal{S}(T^{\perp\mathcal{M}}) = \left\{ s \in \mathcal{S}(T) \mid s \notin P^{\downarrow\mathcal{M}}(T) \right\},$$

and the lengths of these edges coinciding with the corresponding lengths in T. Similarly, all external edges to leaves not in $\mathcal{M}$ in $T^{\perp\mathcal{M}}$ are of the same length as the external edges in T, but all external edges to leaves in $\mathcal{M}$ equal to zero. Furthermore, $d_{BHV}(T, T^{\perp\mathcal{M}}) = ||P^{\downarrow\mathcal{M}}(T)||_T$.

Proof. By definition, all external edges to leaves in $\mathcal{M}$ in $T^{\perp\mathcal{M}}$ are of size zero, and every internal edge maps to an internal split under the TDR map $\Psi_{\mathcal{L}\setminus\mathcal{M}}$, implying $T^{\perp\mathcal{M}} \in \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$.

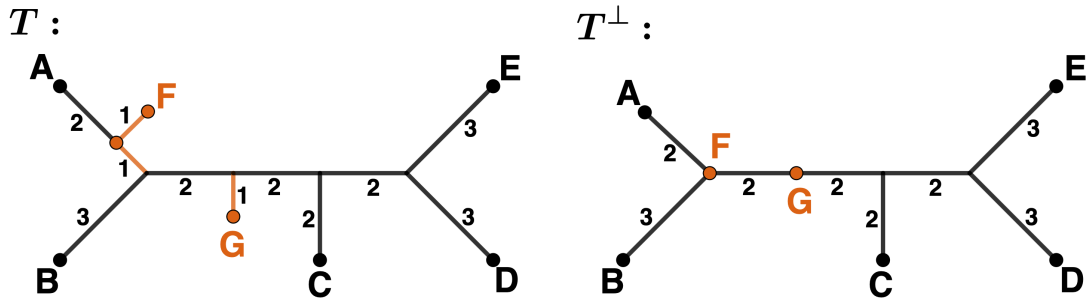


Figure 4.3. The projection of T onto the $\mathcal{M}$-trimmable space. (left) $\mathcal{M} = \{F, G\}$ is not mutually prunable from T, because the external edges to F and G are of positive lengths, and the internal edge $[\{A, F\} \ddagger \{B, C, D, E, G\}]$ would map to an external edge under the TDR map. The edges in $P^{\downarrow\mathcal{M}}(T)$ shown in orange. (right) The projection $T^{\perp\mathcal{M}}$ onto the $\mathcal{M}$-trimmable space. The distance between the two trees is $\sqrt{3}$.

It is straightforward to see $d_{\text{BHV}}(T, T^{\perp\mathcal{M}}) = ||P^{\downarrow\mathcal{M}}(T)||_T$, since all edges in $P^{\downarrow\mathcal{M}}$ are of size zero (or not present) in $T^{\perp\mathcal{M}}$ and those are the only edges where the two trees differ.

Finally, note that for any tree $t' \in \mathcal{T}^{\mathcal{L}}$ such that $d_{\text{BHV}}(T, t') < ||P^{\downarrow\mathcal{M}}(T)||_T$ at least one of the edges in $P^{\downarrow\mathcal{M}}(T)$ is of positive length, thus not belonging to the $\mathcal{M}$-trimmable subspace. □

For brevity, henceforth I use $||P^{\downarrow\mathcal{M}}(T)||$ (in place of $||P^{\downarrow\mathcal{M}}(T)||_T$) to refer to the distance from T to the $\mathcal{M}$-trimmable subspace.

The transition from a BHV space $\mathcal{T}^{\mathcal{L}}$ to a lower-level space $\mathcal{T}^{\mathcal{L}'}$ occurs through one of the trees in the $\mathcal{M}$-trimmable space, where $\mathcal{M} = \mathcal{L} \setminus \mathcal{L}'$. The transformation on the topology of T where $\mathcal{M}$ is mutually prunable into a topology in $\mathcal{T}^{\mathcal{L}'}$ is found by using the TDR map (Definition 2.9) on splits in $\mathcal{S}(T)$. The topology of the tree in the lower-level is given by $\Psi_{\mathcal{L}'}(\mathcal{S}(T)) = \{\Psi_{\mathcal{L}}(s) | \, s \in \mathcal{S}(T)\}$, which (by the properties of the $\mathcal{M}$-trimmable space) contains only internal splits. The final tree resulting from pruning $\mathcal{M}$ from T will be determined by combining the lengths of splits that map into the same split, which I detail in the next section.

I now present an additional property of the $\mathcal{M}$-trimmable spaces that will later be useful.

Lemma 4.10. *Given subsets of leaves $\mathcal{M}' \subseteq \mathcal{M} \subset \mathcal{L}$, the $\mathcal{M}$-trimmable subspace is contained in the $\mathcal{M}'$-trimmable subspace; i.e., $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}} \subseteq \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}'}$.*

Proof. Assume $\mathcal{M}$ is mutually trimmable from T. Note that any leaf in $\mathcal{M}'$ belongs to $\mathcal{M}$ as well, so all external edges to leaves in $\mathcal{M}'$ will be of length zero in T. Moreover, given an internal split $s = [\mathcal{L}_1 \ddagger \mathcal{L}_2] \in \mathcal{S}(T)$, $T \in \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$ implies $|\mathcal{L}_1 \setminus \mathcal{M}|, |\mathcal{L}_2 \setminus \mathcal{M}| \geq 2$, which in turn implies $|\mathcal{L}_1 \setminus \mathcal{M}'|, |\mathcal{L}_2 \setminus \mathcal{M}'| \geq 2$. Thus, if $\mathcal{M}$ is mutually prunable from T, so is $\mathcal{M}'$. □

4.1.2 Pruning and Regrafting Leaf Sets

In the process of removing leaves and using the TDR map to determine new internal splits, the lengths of these new splits should reflect the combined lengths of all the splits that map into them. In the original TDR definition (Zairis et al., 2016, Definition 4.1), this was achieved by adding the lengths of the edges that map into the same split. In the notation below, this amounts to choosing $\beta(x, y) = ||(x, y)||_1$. However, I argue that the method used to combine edge lengths after removing leaves directly impacts interpretability, as the combined value may need to be compared against the length of a single edge on a different tree when determining distances. Careful consideration should be given to this decision. Providing flexibility in how edge lengths are combined is both computationally and mathematically advantageous and can lead to different interpretations. This flexibility allows us to consider Towering Tree Spaces that emphasize various aspects of topological and branch length changes when determining distances between trees. I first introduce a general method for combining edge lengths and later contrast some of these methods in the chapter.

Definition 4.11. A binary operation on the non-negative reals, $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$ is a **merging operation** if it is continuous, commutative ($\beta(x, y) = \beta(y, x)$), associative

$(\beta(\beta(x,y),z) = \beta(x,\beta(y,z)))$, has 0 as the identity element (i.e $\beta(x,0) = x$ for all x), and is non-vanishing ($\beta(x,y) = 0$ only if $x = y = 0$).

Example 4.12. The L^p norms $\beta(x,y) = ||(x,y)||_p$ in $\mathbb{R}^2_{\geq 0}$, are merging operations: $||(x,y)||_1 = x + y$, $||(x,y)||_2 = \sqrt{x^2+y^2}$ and $|||(x,y)||_\infty = \max(x,y)$.

The properties provided in the definition of merging functions guarantee favorable behavior when assigning edge lengths. Firstly, commutativity and associativity guarantee that the final length of an internal split is independent of the order in which the contributing splits are combined. These properties also ensure that the length remains consistent under sequential leaf-prunings (see Lemma 4.15). Secondly, having 0 as the identity element ensures that equivalent trees in the BHV space (those with length-zero internal splits) map to the same tree, maintaining the integrity of the space. Thirdly, the non-vanishing property ensures that all internal splits resulting from the TDR map are preserved in the new tree. Finally, the continuity of the β function ensures that trees close to each other in the higher space remain close when mapped. I use the notation $\mathbf{B}_{i=1}^{m} a_i = \beta(\beta(\ldots\beta(\beta(a_1,a_2),a_3),\ldots),a_m)$ to refer to the process of applying the binary operation repeatedly through a sequence of values.

Example 4.13. Returning to Example 4.12, consider the case $\beta(x,y) = \max(x,y)$. With values $a_1 = 2$, $a_2 = 5$, $a_3 = 1$, $a_4 = 3$, the notation $\mathbf{B}$ summarizes:

$$\mathop{\mathbf{B}}_{i=1}^{4} a_i = \max(\max(\max(2,5),1),3) = \max(\max(5,1),3) = \max(5,3) = 5.$$

I now formally define leaf prunings and regraftings.

Definition 4.14. Consider $T \in \mathcal{T}^{\mathcal{L}}$ with a mutually prunable subset of leaves $\mathcal{M}$, and denote $\mathcal{L}' = \mathcal{L} \setminus \mathcal{M}$. Given a merging operation β, **the β-pruning of $\mathcal{M}$ from** T is the operation $\psi_\beta(T,\mathcal{M})$ that produces $T' \in \mathcal{T}^{\mathcal{L}'}$ with the following properties:

- The topology of T' is given by the TDR map applied to the internal edges in T; i.e.,

$$\mathcal{S}(T') = \Psi_{\mathcal{L}'}\left(\mathcal{S}(T)\right).$$

- Every external edge remaining in T' (those corresponding to leaves in $\mathcal{L}'$) has the length of the corresponding external edge in T.

- For every internal edge $p \in \mathcal{S}(T')$, the lengths of the edges in the pre-image of p under the TDR map restricted to the edges in T, $\Psi^{-1}_{\mathcal{L}'}(p)\big|_{\mathcal{S}(T)} = \{q \in \mathcal{S}(T) \mid \Psi_{\mathcal{L}'}(q) = p\}$, are combined through β and assigned to the length of p:

$$|p|_{T'} = \underset{q \in \Psi^{-1}_{\mathcal{L}'}(p)\big|_{\mathcal{S}(T)}}{\mathbf{B}} |q|_T.$$

As mentioned previously, the properties of the merging operations ensure β-prunings can be performed in different orders to obtain the same tree, which is essential for constructing equivalence classes. We see this in the following lemma.

Lemma 4.15. *Given $T \in \mathcal{T}^{\mathcal{L}}$ in the $\mathcal{M}$-trimmable subspace, and a partition of the leaves to be pruned $\mathcal{M} = \mathcal{M}_1 \sqcup \mathcal{M}_2$, pruning $\mathcal{M}_1$ from T produces a tree where $\mathcal{M}_2$ is mutually prunable. Additionally, pruning $\mathcal{M}_2$ after pruning $\mathcal{M}_1$ is equivalent to pruning $\mathcal{M}$; i.e., $\psi_\beta(\psi_\beta(T, \mathcal{M}_1), \mathcal{M}_2) = \psi_\beta(T, \mathcal{M})$.*

Proof. Refer to Appendix B.1. □

The action of pruning is one of the two actions I define to transition between BHV spaces, from a higher to a lower level. The opposite operation, regrafting new leaves onto a tree, is its counterpart, which allow transitions from lower to higher levels in the Towering Space.

Definition 4.16. Given $T \in \mathcal{T}^{\mathcal{L}}$ and a set of leaves $\mathcal{M}$ such that $\mathcal{M} \cap \mathcal{L} = \emptyset$, we define a **$\beta$-regraft of $\mathcal{M}$ onto** T to be the operation of selecting $T' \in \mathcal{T}^{\mathcal{L} \cup \mathcal{M}}$ such that $\mathcal{M}$ is mutually prunable from T' and $\psi_\beta(T', \mathcal{M}) = T$.

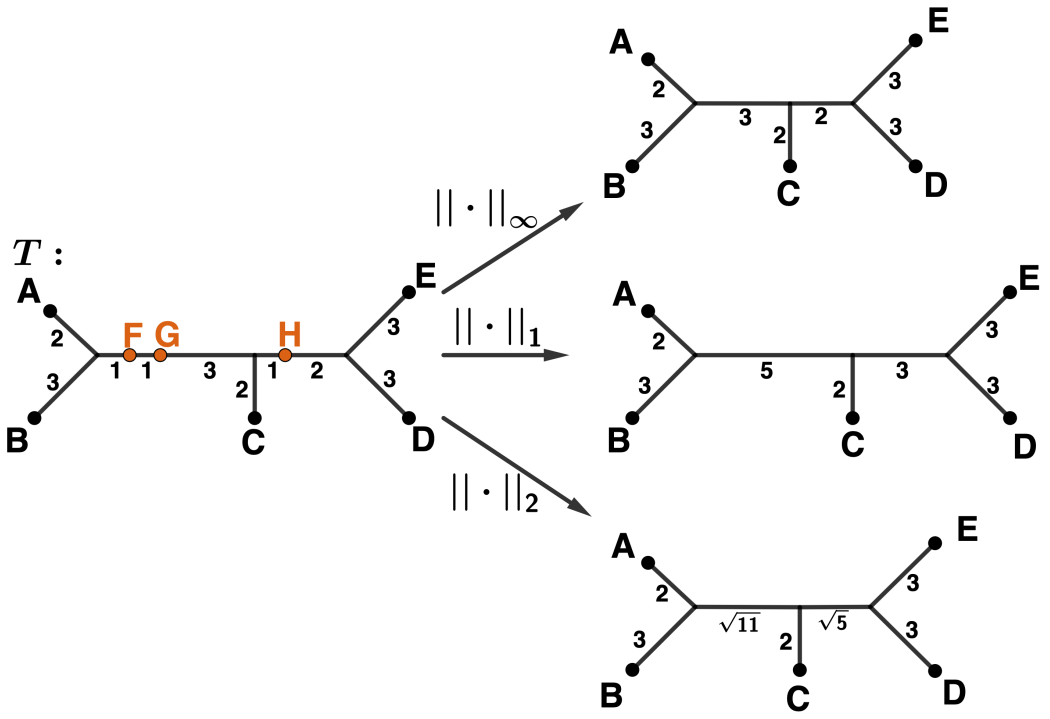


Figure 4.4. Example of different leaf prunings given by different choices of merging operators β. Given the tree T (left) where $\mathcal{M} = \{F, G, H\}$, the β-pruning of these leaves from T are shown, using as the merging operation the L^p norms: $||\cdot||_\infty$ (top-right), $||\cdot||_1$ (center-right) and $||\cdot||_2$ (bottom-right).

Note that while a β-pruning is a function that returns a unique, unambiguously defined tree, a β-regraft is not a function, as it could return several trees. Thus, we consider the following space.

Definition 4.17. Given $T \in \mathcal{T}^{\mathcal{L}'}$ and a superset of leaves $\mathcal{L} \supseteq \mathcal{L}'$, **the β-sprouting space** of T in $\mathcal{T}^{\mathcal{L}}$, denoted by $\Lambda_\beta^{\mathcal{L}}(T)$ is the set of all the trees in $\mathcal{T}^{\mathcal{L}}$ from which $\mathcal{M} = \mathcal{L} \setminus \mathcal{L}'$ are mutually prunable and that map to T under the β-pruning of $\mathcal{M}$. This is, $\Lambda_\beta^{\mathcal{L}}(T) = \{t \in \mathcal{Z}_{\mathcal{M}}^{\mathcal{L}} \mid \psi_\beta(t, \mathcal{M}) = T\}$.

The β-sprouting spaces are closely related to extension spaces (Section 3.2). In the same way as extension spaces are subsets of trees in the BHV space that represent a tree at a

lower level — where these trees map into the lower-level tree under the TDR map (Definition 2.6) — sprouting spaces are also subsets representing a single tree. In the case of sprouting spaces, the trees map into the base tree under a β-pruning operation. That said, there are key differences between extension spaces and sprouting spaces. Extension spaces include trees with positive-length external edges to leaves that are missing from the lower-level tree they represent. In contrast, sprouting spaces require the external edges of these leaves to have zero length, as only those leaves with zero-length edges can be pruned. Additionally, in extension spaces, there is no distinct separation between external and internal edges: leaves may be regrafted onto external edges of the lower-level tree. However, this is not permitted in the construction of sprouting spaces; regrafted leaves must be attached to internal edges to prevent the behavior shown in Figure 4.2. Through the merging operation, sprouting spaces provide greater options on what trees in a BHV space represent a tree with fewer leaves. In fact, when selecting the merging operation $\beta(x, y) = ||(x, y)||_1$, the sprouting space of a tree is a subspace of its extension space, $\Lambda^{\mathcal{L}}_{\beta}(T) \subset E^{\mathcal{L}}_{T}$. Here, the sprouting space consists of all trees in the extension space that are formed by attaching new leaves to interior edges, with external edges of length zero.

Before moving unto the formal definition of the Towering Metric Spaces' family, I give an additional lemma that provides conditions under which leaves remain mutually prunable after the pruning of other leaves, which simplifies contiguous leaf prunings later on this chapter.

Lemma 4.18. *Consider a set of leaves $\mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}$, and denote by $\mathcal{M}_2 = \mathcal{L}_1 \setminus \mathcal{L}_2$ and $\mathcal{M}_1 = \mathcal{L} \setminus \mathcal{L}_1$. Given $T \in \mathcal{T}^{\mathcal{L}_1}$, any tree $T^{\uparrow} \in \Lambda^{\mathcal{L}}(T)$ belongs to the $\mathcal{M}_1$-trimmable subspace $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}_1}$. Furthermore, the subset of leaves $\mathcal{M} = \mathcal{M}_1 \sqcup \mathcal{M}_2$ is mutually prunable from $T^{\uparrow}$ if and only if $\mathcal{M}_2$ is mutually prunable from T.*

Proof. All lengths of the external edges to leaves in $\mathcal{L}_1$ are of the same length in T as

the corresponding external edges in $T^{\uparrow}$. Thus, the external edges to leaves in $\mathcal{M}_2$ are of length zero in T if and only if they are of length zero in $T^{\uparrow}$. Since $T^{\uparrow} \in \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}_1}$, all internal edges in $T^{\uparrow}$ map into internal edges under the TDR map $\Psi_{\mathcal{L}_1}$. Moreover, for any $s = [\mathcal{L}'_1 \ddagger \mathcal{L}_1 \setminus \mathcal{L}'_1] \in \mathcal{S}(T)$, any edge $s^{\uparrow} \in \mathcal{S}(T^{\uparrow})$ such that $\Psi_{\mathcal{L}_1}(s^{\uparrow}) = s$ is of the form $s^{\uparrow} = [\mathcal{L}'_1 \cup \mathcal{M}'_1 \ddagger (\mathcal{L}_1 \setminus \mathcal{L}'_1) \cup \mathcal{M}''_1]$ where $\mathcal{M}_1 = \mathcal{M}'_1 \sqcup \mathcal{M}''_1$. Since $\mathcal{L}'_1 \setminus \mathcal{M}_2 = (\mathcal{L}'_1 \cup \mathcal{M}'_1) \setminus \mathcal{M}$ and $(\mathcal{L}_1 \setminus \mathcal{L}'_1) \setminus \mathcal{M}_2 = [(\mathcal{L}_1 \setminus \mathcal{L}'_1) \cup \mathcal{M}''_1] \setminus \mathcal{M}$, then $\Psi_{\mathcal{L}_2}(s) = \Psi_{\mathcal{L}_2}(s^{\uparrow})$, and thus s maps to an internal edge under $\Psi_{\mathcal{L}_2}$ if and only if $s^{\uparrow}$ also maps to an internal edge. □

4.2 The family of Towering Tree Spaces

So far, I have defined an operation that transforms the topology and branch lengths of a tree to morph it into another tree with fewer leaves. This serves as the foundation for defining transitions between different BHV levels. In this section, I briefly formalize the distance operation on the union of BHV spaces based on the given merging operation β, thereby completing the definition of the entire family of towering tree spaces. After this section, I will focus on one of these towering spaces by selecting a convenient β.

4.2.1 Equivalence classes

While β-prunings are functions mapping trees from a higher to a lower BHV space, β-regrafts are the inverse operation, returning the pre-image of a lower tree in a higher BHV level. I refer to β-prunings and β-regrafts as β-transformations. These transformations determine the equivalence classes that connect BHV spaces of different dimension, thus creating a towering space.

Definition 4.19. Given a merging operation β, two trees $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$ are **equivalent under β-transformations**, denoted $T_1 \simeq_{\beta} T_2$, if there is a finite number of β-transformations that can be applied sequentially, starting at T_1 and ending at T_2.

Lemma 4.20. *The relationship $\simeq_\beta$ is an equivalence relationship.*

Proof. The properties of reflexivity, symmetry and transitivity follow directly from the definition. □

4.2.2 Metric definition

Here I introduce the β-towering tree space, a metric tree space defined over the quotient space $\mathcal{T}^{\mathbf{P}(\mathcal{N})}/\simeq_\beta$ for a given merging operation β. To define distances in this space, I first consider a preliminary metric on $\mathcal{T}^{\mathbf{P}(\mathcal{N})}$ that is later refined to the (strictly finite-valued) quotient pseudometric. The preliminary metric uses the BHV distance for trees in the same BHV space and treats trees with different leaves as incomparably distant:

$$d^*(T_1, T_2) = \begin{cases} d_{\text{BHV}}(T_1, T_2) & \text{if } \mathcal{L}(T_1) = \mathcal{L}(T_2), \\ \infty & \text{otherwise.} \end{cases}$$

Definition 4.21. The **β-towering distance** between any two trees T_1 and T_2 is the quotient pseudometric of d^* on $\mathcal{T}^{\mathbf{P}(\mathcal{N})}/\simeq_\beta$ (Bridson & Haefliger, 1999, I.5, Definition 5.19); that is, the size of the smallest possible path between T_1 and T_2, allowing for a finite number of distance-zero β-transformations between trees within an equivalence class.

$$d_\beta(T_1, T_2) = \inf \left\{ \sum_{i=1}^{k} d_{\text{BHV}}(t_i, t_i') \middle| T_1 \simeq_\beta t_1, t_i' \simeq_\beta t_{i+1} \forall i = 1, \ldots, k-1, t_k' \simeq_\beta T_2 \right\}, \quad (4.1)$$

with the infimum taken over all finite sequences $\{t_i\}_{i=1}^k, \{t_i'\}_{i=1}^k$.

By definition, $d_\beta(\cdot, \cdot)$ is guaranteed to be a pseudometric, that is, symmetry and the triangle inequality hold. Moreover, it is also a complete metric on the quotient space, denoted by $\mathcal{T}_\beta^{\mathbf{P}(\mathcal{N})} = \mathcal{T}^{\mathbf{P}(\mathcal{N})}/\simeq_\beta$, that is, $d_\beta(T_1, T_2) = 0$ if and only if $T_1 \simeq_\beta T_2$.

4.2.3 Leaf-distance preserving Towering Tree Space

Given its connection to the TDR map defined in Zairis et al. (2016, Definition 4.1) and extension spaces, using $\beta_1(x, y) = x + y$ as the merging operation appears to be a natural choice. I refer to this as the leaf-distance preserving towering tree space. In this tree space, the tree produced from a leaf pruning maintains the same discrete metric on its leaves as the original tree did for that subset of leaves; specifically, $d_T(\ell_1, \ell_2) = d_{\psi_{\beta_1}(T,\mathcal{M})}(\ell_1, \ell_2)$ for $\ell_1, \ell_2 \notin \mathcal{M}$ (see Figure 2.3). While this interpretation provides a clear understanding of what it means for two trees to belong to the same equivalence class, it also results in counter-intuitive properties and computational challenges in practice. Here, I briefly discuss some of its advantages and disadvantages. In the next section, however, I introduce an alternative approach that offers more practical and intuitive properties.

A property of the β_1-towering distance is that it allows for the creation of "shortcuts" when constructing a path between two trees by regrafting new leaves that were not originally part of their leaf sets. For example, in Figure 4.5, two trees in the same BHV space are initially separated by a distance of $\sqrt{7} \approx 2.65$. However, by regrafting new leaves, this distance is reduced. Intuitively, one might expect shorter paths to be achieved by pruning common leaves between two trees and regrafting them in strategic locations, thus making the topologies more similar. Nevertheless, the ability to shorten paths by adding leaves that are not part of either tree is counter-intuitive, as these leaves do not provide additional information on the similarity of the trees. Theoretically, one could continue adding leaves indefinitely, reducing the length of the paths without adding any meaningful value. This introduces a paradox where the metric's flexibility undermines its practical utility in measuring tree similarity.

This phenomenon can be partially explained by the difference in how edge lengths are treated in these spaces. In the β_1-towering distance space, edges are combined using the

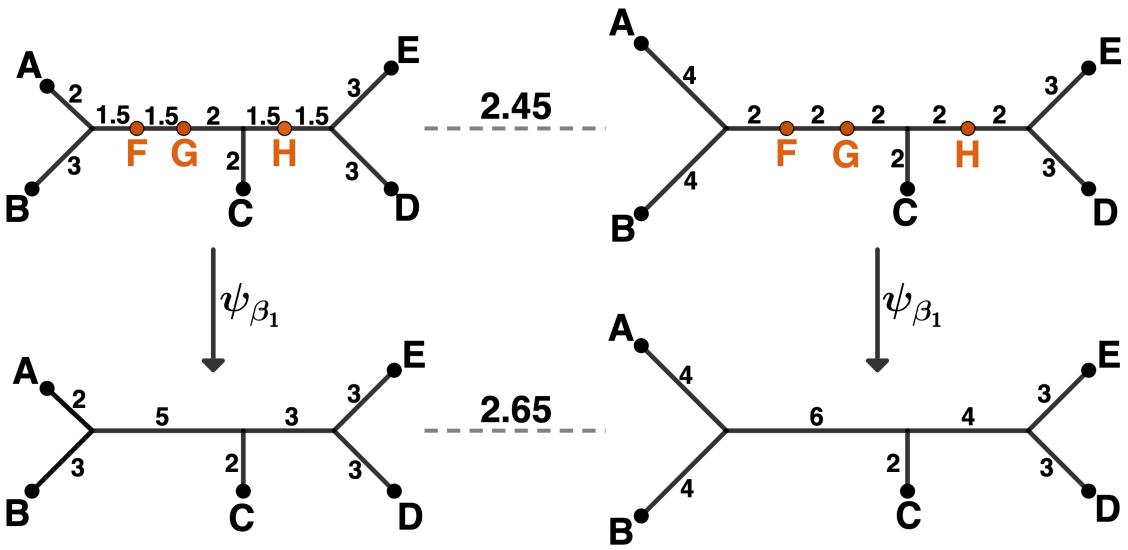


Figure 4.5. Example of a shorter path through β_1-regrafts. Given the two trees with leaves $\{A, B, C, D, E\}$ at BHV distance ≈ 2.65 (bottom), it is possible to regraft new leaves $\{F, G, H\}$ to find a shorter path of length ≈ 2.45 at a higher BHV space (top).

L^1-norm, whereas in each BHV space, edge lengths are compared using the L^2-norm of the differences. Dividing internal branches by regrafting new leaves allows the path to effectively move in L^2 distances instead of L^1 distances, thereby artificially reducing its length. This behavior introduced challenges in my preliminary explorations of this distance. Specifically, when constructing short paths between trees in this space, I found that the shortest paths were often achieved by regrafting leaves into particularly long edges, irrespective of topological or branch length similarities. This undermines the intended metric properties and complicates meaningful comparisons between trees.

This space maintains a close relationship with extension spaces — specifically, sprouting spaces are a subset of extension spaces ($\Lambda_\beta^{\mathcal{L}}(T) \subset E_T^{\mathcal{L}}$) —. This connection suggests that the gradient methods developed in Chapter 3 could be adapted to find distances within sprouting spaces, thereby aiding in determining distances in the β_1-towering space. However, this process remains computationally intensive. As will be discussed in the next section, utiliz-

ing the L^2-norm as the merging operation offers improved interpretability, more intuitive behavior for short paths, and the potential for greater computational efficiency.

4.3 Definition and Preliminary Results of the Towering Tree Space

In the previous sections, I defined a family of tree spaces for trees with non-identical leaves, indexed by the merging operations that define equivalence classes among trees in different BHV spaces. There are many options for the merging operation, each with its own advantages, disadvantages and interpretations. In the prior section, I briefly discussed some of these for the the merging operation $\beta_1(x, y) = x+y$, which is a natural choice based on the TDR map and its connection to extension spaces. Here, I argue the most suitable option for the merging operation is the function $\beta_2(x, y) = \sqrt{x^2 + y^2}$ (the L^2-norm), since it produces a towering space that is both interpretable and has favorable geometrical properties that make distance computation feasible.

For the rest of this chapter, I focus on this towering space and explore the process of computing distances inside it. Henceforth, $\mathbb{T}^{\mathcal{N}} = \mathcal{T}^{\mathbf{P}(\mathcal{N})}/\simeq_{\beta_2}$, $d = d_{\beta_2}$ and I refer to $(\mathcal{T}^{\mathbf{P}(\mathcal{N})}/\simeq_{\beta_2}, d_{\beta_2})$ as *the towering tree space*, which I denote by $(\mathbb{T}^{\mathcal{N}}, d)$. Where unambiguous, I drop β from notation (such as β-transformations and β-sprouting subspaces), with the understanding that I am now only considering β_2 as the merging operation.

4.3.1 Justification for Using L^2-Norm as Merging Operation

One motivation for $\beta_2(x, y) = \sqrt{x^2 + y^2}$ relates to the biological interpretation of the distances between tree and the origin tree in the BHV space, as well as the relationship between the link of the origin in the BHV space Billera et al. (2001, Section 4.2), geodesic distances, and the distances between topologies. The origin tree in BHV space is the tree with all branches of zero length, representing no structured evolutionary divergence among

the organisms labeling its leaves. The BHV distance of a tree to the origin represents the cumulative divergence represented by all the branches in the tree. Trees further from the origin indicate evolutionary processes with more substantial changes than those closer to the origin. Using β_2 to merge branch lengths after leaf prunings results in trees that remain equidistant to the origin (in their respective BHV space); i.e., $d_{BHV}(0,T) = d_{BHV}(0, \psi_{\beta_2}(T, \mathcal{M}))$ for any $\mathcal{M}$. Thus, a β-transformation produces a new tree with a similar topology and reflecting a comparable amount of overall evolutionary change among the organisms in the leaf set.

In addition, the BHV space can be interpreted as a 0-cone of the link of the origin (see Section 2.2.1). This implies that the BHV distance between $T_1, T_2 \in \mathcal{T}^{\mathcal{L}}$ is

$$d^2_{\text{BHV}}(T_1,T_2) = d^2_{\text{BHV}}(0,T_1) + d^2_{\text{BHV}}(0,T_2) - 2d_{\text{BHV}}(0,T_1)d_{\text{BHV}}(0,T_2)\cos\left[\min\{\pi, \measuredangle(T_1,T_2)\}\right],$$

where $\measuredangle(T_1,T_2)$ is the spherical distance (angle) between the projections of T_1 and T_2 onto the link of the origin. Since the angle between the two points in the link of the origin heavily depends on the topology of the trees, and $d_{\text{BHV}}(0,T_1) = d_{\text{BHV}}(0,T_1')$ for any $T_1' \simeq_{\beta_2} T_1$, the tree in the same equivalence class as T_1 closer to T_2 with respect to the BHV distance will depend primarily on which has the most similar topology. This contrasts with the case where $\beta_1(x,y) = x + y$, where the closest tree in a β_1-sprouting space to another tree in the same BHV space is sometimes produced by regrafting new leaves onto particularly long edges in the lower-level tree, artificially reducing the distance regardless of the properties of the second tree This concurs with the intuition that when comparing trees with different leaf sets, trees with similar topologies (indicating they potentially come from similar evolutionary history, but are missing leaves) should lie close to each other.

4.3.2 First geometrical results

For trees $t \in \mathbb{T}^{\mathcal{N}}$ in the sprouting space $\Lambda^{\mathcal{L}}(T)$ of $T \in \mathcal{T}^{\mathcal{L}'}$ with $\mathcal{L}' \subseteq \mathcal{L}$, all edges mapping to an internal edge $p \in \mathcal{S}(T)$ under the TDR map contribute to the length of p by

$$|p|_T = \sqrt{|q_1|_t^2 + \ldots + |q_k|_t^2},$$

where $\Psi_{\mathcal{L}'}^{-1}(p)\big|_{\mathcal{S}(t)} = \{q_1, \ldots, q_k\}$. As a result, I use the following inequality throughout this chapter.

Lemma 4.22. *For sequences of real numbers $\{a_i\}_{i=1}^k$ and $\{b_i\}_{i=1}^k$,*

$$\sum_{i=1}^{k}(a_i - b_i)^2 \geq \left(\sqrt{\sum_{i=1}^{k} a_i^2} - \sqrt{\sum_{i=1}^{k} b_i^2}\right)^2.$$

Proof. By the triangle inequality,

$$\sqrt{\sum_{i=1}^{k}(a_i - b_i)^2} + \sqrt{\sum_{i=1}^{k} b_i^2} \geq \sqrt{\sum_{i=1}^{k} a_i^2} \text{ and } \sqrt{\sum_{i=1}^{k}(a_i - b_i)^2} + \sqrt{\sum_{i=1}^{k} a_i^2} \geq \sqrt{\sum_{i=1}^{k} b_i^2},$$

and therefore $\sqrt{\sum_{i=1}^{k}(a_i - b_i)^2} \geq \left|\sqrt{\sum_{i=1}^{k} a_i^2} - \sqrt{\sum_{i=1}^{k} b_i^2}\right| \geq 0$. The result follows. □

This inequality ensures that the difference in length of a common edge between two distinct trees cannot be artificially diminished by attaching new edges through regrafting. Consequently, the distance between two trees within the same BHV space cannot be reduced by taking shortcuts at higher BHV levels with leaves not originally present in the trees, as demonstrated in the following lemma.

Theorem 4.23. *Given two trees $T_1, T_2 \in \mathcal{T}^{\mathcal{L}'}$ for a subset of leaves $\mathcal{L}' \subseteq \mathcal{L}$, then $d_{BHV}(T_1, T_2) \leq d_{BHV}(T_1^{\uparrow}, T_2^{\uparrow})$ for any trees $T_1^{\uparrow} \in \Lambda^{\mathcal{L}}(T_1)$ and $T_2^{\uparrow} \in \Lambda^{\mathcal{L}}(T_2)$.*

Proof. Proof in Appendix B.2 □

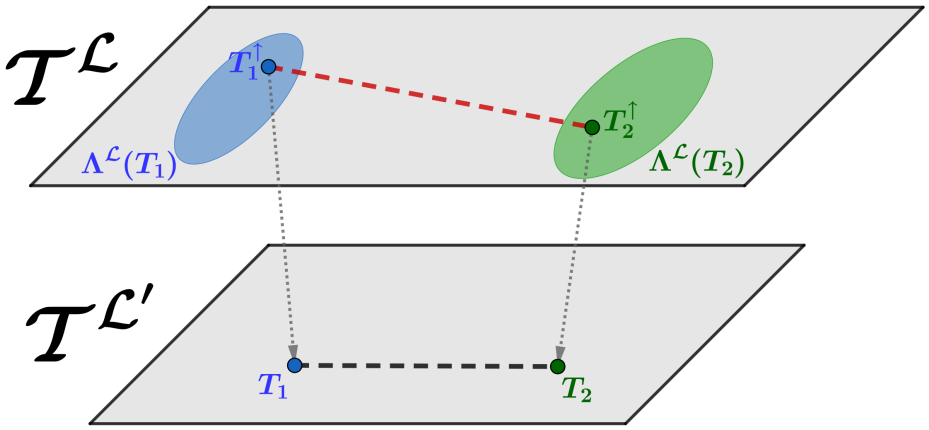


Figure 4.6. Representation of BHV geodesics between two trees $T_1, T_2 \in \mathcal{T}^{\mathcal{L}'}$ and two trees in their respective sprouting subspaces $T_1^{\uparrow} \in \Lambda^{\mathcal{L}}(T_1)$ and $T_2^{\uparrow} \in \Lambda^{\mathcal{L}}(T_2)$. The sprouting spaces for T_1 and T_2 are respectively represented by shaded areas in blue and green. By Theorem 4.23, the dashed red line (representing the geodesic between $T_1^{\uparrow}$ and $T_2^{\uparrow}$ in $\mathcal{T}^{\mathcal{L}}$) is always longer than the dashed black line (representing the geodesic between T_1 and T_2 in $\mathcal{T}^{\mathcal{L}'}$).

The previous lemma demonstrates that when two trees share the same leaf set, no shorter paths can be found by introducing new leaves that were not originally present in either of the trees being compared. This is an intuitive property, as the addition of extraneous leaves should not alter distances between trees. It stands in direct contrast to the counter-intuitive behavior I discussed in Section 4.2.3.

The result above extends to the conclusion that when a set of leaves are prunable from two trees, a shorter geodesic can be found in the lower BHV level resulting from this prune.

Corollary 4.24. *Given two trees in the same $\mathcal{M}$-trimmable space, $T_1, T_2 \in \mathcal{Z}_{\mathcal{M}}^{\mathcal{L}}$, the length of the geodesic between them is lower-bounded by the length of the geodesic between their respective prunings of $\mathcal{M}$; i. e. $d_{BHV}(T_1, T_2) \geq d_{BHV}(\psi(T_1, \mathcal{M}), \psi(T_2, \mathcal{M}))$.*

In general, I will focus on results that involve the pruning and regraft of leaves present in at least one of the trees for which the distance is being computed, as Theorem 4.23 suggests the regraft of leaves outside the two leaf sets will not be beneficial.

4.4 Short paths through lower BHV levels

We now explore the construction of towering space geodesics in more general settings. Shorter paths should intuitively be possible by pruning and regrafting leaves at optimal places, transforming the original trees to more topologically similar trees in the process. The following theorem and subsequent lemmas provide insights on how to efficiently construct short paths through these transformations.

Before presenting these results, I provide a structure that divides the edges of a tree T in $\mathcal{P}^{\downarrow\mathcal{M}}(T)$ into separate independent subtrees (see Figure 4.7). This division will play a role on the results and algorithms throughout the rest of this chapter.

Definition 4.25. Given a tree $T \in \mathcal{T}^{\mathcal{L}}$ and a subset of leaves $\mathcal{M} \subset \mathcal{L}$, consider the natural partition of $\mathcal{M}$ into subsets that group into a single node when T is projected onto the $\mathcal{M}$-trimmable space. This is, I partition the set into subsets $\mathcal{M} = \mathcal{M}_1 \sqcup \ldots \sqcup \mathcal{M}_r$, called **the independent maximal sets** of $\mathcal{M}$ in T, such that for each $\mathcal{M}_i$,

1. There is a leaf $\ell \in \mathcal{L} \setminus \mathcal{M}$ such that $s_i = [\mathcal{M}_i \cup \{\ell\} \ddagger \mathcal{L} \setminus (\mathcal{M}_i \cup \{\ell\})] \in \mathcal{S}(T)$ and there is no $\mathcal{M}' \subseteq \mathcal{M}$ such that $\mathcal{M}_i \subset \mathcal{M}'$ and $[\mathcal{M}' \cup \{\ell\} \ddagger \mathcal{L} \setminus (\mathcal{M}' \cup \{\ell\})] \in \mathcal{S}(T_1)$. In this case, $\mathcal{M}_i$ is **the independent maximal set neighboring** ℓ; or

2. $s_i = [\mathcal{M}_i \ddagger \mathcal{L} \setminus \mathcal{M}_i] \in \mathcal{P}(T_1)$ and there is no $\mathcal{M}' \subseteq \mathcal{M}$ such that $\mathcal{M}_i \subset \mathcal{M}'$ and either $[\mathcal{M}' \ddagger \mathcal{L} \setminus \mathcal{M}'] \in \mathcal{S}(T_1)$ or $[\mathcal{M}' \cup \{\ell\} \ddagger \mathcal{L} \setminus (\mathcal{M}' \cup \{\ell\})] \in \mathcal{S}(T_1)$ for some leaf $\ell \in \mathcal{L} \setminus \mathcal{M}$. I refer to $\mathcal{M}_i$ as **an independent maximal set attached to internal edges**, since the edge s_i must be adjacent to two internal edges in T_1 or incident on a node with degree higher than 3.

Given an independent set $\mathcal{M}_i$, consider the set

$$P_i^{\downarrow\mathcal{M}}(T) = \{s \in \mathcal{S}(T) \mid s = [\mathcal{M}' \ddagger \mathcal{L} \setminus \mathcal{M}'] \text{ or } [\mathcal{M}' \cup \{\ell\} \ddagger \mathcal{L} \setminus (\mathcal{M}' \cup \{\ell\})] \text{ for } \mathcal{M}' \subseteq \mathcal{M}_i\}$$

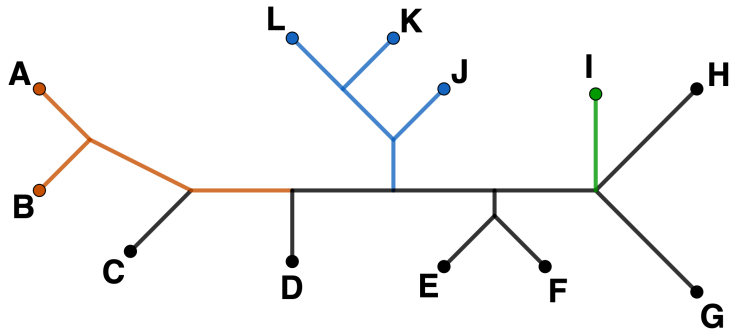


Figure 4.7. Example of independent maximal sets on a phylogenetic tree. In the tree shown above, consider the subset of leaves $\mathcal{M} = \{A, B, I, J, K, L\}$. The independent maximal sets are: $\mathcal{M}_1 = \{A, B\}$, neighboring C and $P_1^{\downarrow\mathcal{M}}(T)$ highlighted in orange; $\mathcal{M}_2 = \{I\}$ attached to internal edges, and $P_2^{\downarrow\mathcal{M}}(T)$ highlighted in green; and $\mathcal{M}_3 = \{J, K, L\}$ attached to internal edges, and $P_i^{\downarrow\mathcal{M}}(T)$ highlighted in blue.

which I call **the edges in T belonging to $\mathcal{M}_i$**. Note $P_i^{\downarrow\mathcal{M}}(T) \subset P^{\downarrow\mathcal{M}}(T)$.

Theorem 4.26. *Consider $T_1 \in \mathcal{T}^{\mathcal{L}}$ and $T_2 \in \mathcal{T}^{\mathcal{L}'}$ with $\mathcal{L} = \mathcal{L}' \sqcup \mathcal{M}$. Take $T_1' = \psi(T_1^{\perp\mathcal{M}}, \mathcal{M})$, the pruning of $\mathcal{M}$ from the projection of T onto $\mathcal{Z}_{\mathcal{M}}^{\mathcal{L}}$. Then:*

(i) The BHV distance between T_1 and trees in the sprouting space $\Lambda^{\mathcal{L}}(T_2)$ is lower-bounded by the L^2-norm of the distances between T_1 and the $\mathcal{M}$-trimmable subspace, and between T_2 and the pruning of $\mathcal{M}$ from the projection of T_1, that is,

$$d_{BHV}(T_1, t) \geq \sqrt{||P^{\downarrow\mathcal{M}}(T_1)||^2 + d_{BHV}^2(T_1', T_2)} \text{ for every } t \in \Lambda^{\mathcal{L}}(T_2).$$

(ii) There is a tree (not necessarily unique) in the sprouting subspace $\Lambda^{\mathcal{L}}(T_2)$ that achieves this lower bound; i.e., there exists $T^* \in \Lambda^{\mathcal{L}}(T_2)$ such that

$$d_{BHV}(T_1, T^*) = \sqrt{||P^{\downarrow\mathcal{M}}(T_1)||^2 + d_{BHV}^2(T_1', T_2)}. \tag{4.2}$$

Proof. While point (i) directly follows from Corollary 4.7 and Theorem 4.23, point (ii) involves constructing T^* by partitioning $\mathcal{M}$ into the independent maximal sets in T_1, and then regrafting each of these sets onto T_2 at locations that maximize the number of common edges

between T_1 and T^*. Additionally, the differences in lengths among these common edges are optimized to match the respective differences in T_1' and T_2. For details, refer to the proof in Appendix B.3. □

The previous theorem will be key to constructing shortest possible paths via leaf prunings. Construction of short paths are supported by Lemma 4.27 and Lemma 4.28 given below. They establish that when a portion of a path goes exclusively downwards in the towering space (i.e. the only transitions along the portion are prunings), the path can be shortened by performing all prunings simultaneously at the end, traversing only the lowest level.

Lemma 4.27. *Consider $T_1 \in \mathcal{T}^{\mathcal{L}}$, $T_2 \in \mathcal{T}^{\mathcal{L}'}$ such that $\mathcal{L}' \subset \mathcal{L}$. Consider all paths from T_1 to T_2 where the only transition between BHV spaces is the leaf pruning of $\mathcal{M} = \mathcal{L} \setminus \mathcal{L}'$. The lengths of these paths are lower bounded by the length of the direct path from T_1 to T_2 by pruning $\mathcal{M}$ at a tree in $\Lambda^{\mathcal{L}}(T_2)$, with length given by* (4.2).

Proof. Refer to Appendix B.3 □

Lemma 4.28. *Consider $T_1 \in \mathcal{T}^{\mathcal{L}}$ and $T_2 \in \mathcal{T}^{\mathcal{L}'}$ where $\mathcal{L}' \subset \mathcal{L}$. Partition the set of leaves $\mathcal{M} = \mathcal{L} \setminus \mathcal{L}'$ into $\mathcal{M} = \mathcal{M}_1 \sqcup \mathcal{M}_2$. Consider all paths from T_1 to T_2 where only two transitions are performed: a pruning of $\mathcal{M}_1$, followed by a pruning of $\mathcal{M}_2$ at some point later. All of these paths are at least as long as the direct path from T_1 to T_2, where $\mathcal{M}$ is pruned at a tree in $\Lambda^{\mathcal{L}}(T_2)$, with length given by* (4.2).

Proof. Consider any path from T_1 to T_2 where a pruning of $\mathcal{M}_1$ is performed in the sprouting space $\Lambda^{\mathcal{L}}(X)$ for some tree $X \in \mathcal{T}^{\mathcal{L}_1}$, where $\mathcal{L}_1 = \mathcal{L} \setminus \mathcal{M}_1$ and then the pruning of $\mathcal{M}_2$ is performed in some other tree $Y \in \mathcal{T}^{\mathcal{L}_1}$. Assume $Y \in \Lambda^{\mathcal{L}_1}(T_2)$, so that the path from X to T_2 is as short as possible (Lemma 4.27). Lemma 4.27 also implies that going from T_1 to Y by performing the pruning of $\mathcal{M}_1$ at a tree $Y^{\uparrow} \in \Lambda^{\mathcal{L}}(Y)$ produces a shorter path. From Lemma

4.18, $\mathcal{M}$ is mutually prunable from $Y^{\uparrow}$ and by the transitivity of leaf prunings (Lemma 4.15), $\psi_\beta(Y^{\uparrow}, \mathcal{M}) = T_2$. □

An important implication of the last two lemmas is that any path between two trees T_1 and T_2 in $\mathcal{T}^{\mathcal{N}}$ can be optimized (i.e., shortened) by refining certain portions of the path (see Figure 4.8). Specifically, sections where all transitions between BHV spaces involve downward movements (strictly leaf prunings) can be consolidated into a single, large pruning operation, allowing a direct jump to a tree at the lowest level of that section. Similarly, sections of the path that exclusively move upwards (strictly regrafting leaves) can be combined into a single regrafting operation, thereby reducing the overall path length. This culminates in the following theorem.

Theorem 4.29. *Consider two trees $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$, with $\mathcal{L}_1 = \mathcal{L}' \sqcup \mathcal{M}_1$ and $\mathcal{L}_2 = \mathcal{L}' \sqcup \mathcal{M}_2$ for some subset $\mathcal{L}' \subseteq \mathcal{L}_1 \cap \mathcal{L}_2$. Among the paths from T_1 to T_2 where the leaves in $\mathcal{M}_1$ are all pruned, followed by regrafts of leaves in $\mathcal{M}_2$, the shortest paths will be of length*

$$\sqrt{\left[||P^{\downarrow\mathcal{M}}(T_1)|| + ||P^{\downarrow\mathcal{M}}(T_2)||\right]^2 + d^2_{BHV}(T_1', T_2')}, \tag{4.3}$$

where T_i' is the pruning of $\mathcal{M}_i$ from $T_i^{\perp\mathcal{M}_i}$, for $i = 1, 2$.

Proof. To prove this theorem, I select a generic tree X in the lowest BHV level $\mathcal{T}^{\mathcal{L}'}$ and employ Theorem 4.26, along with the previous two lemmas to find the shortest paths from each T_1 and T_2 to X. I then prove the best option for X is a tree on the geodesic from T_1' to T_2'. See details in Appendix B.3. □

Notably, the expression for the shortest path through a common lower level provided in (4.3) relies entirely on computing the L^2-norm of the edges that prevent $\mathcal{M}$ from being pruned in each tree, along with the BHV distance between the two trees at the lower level. This expression closely resembles the formula for geodesic lengths in the BHV space. In the

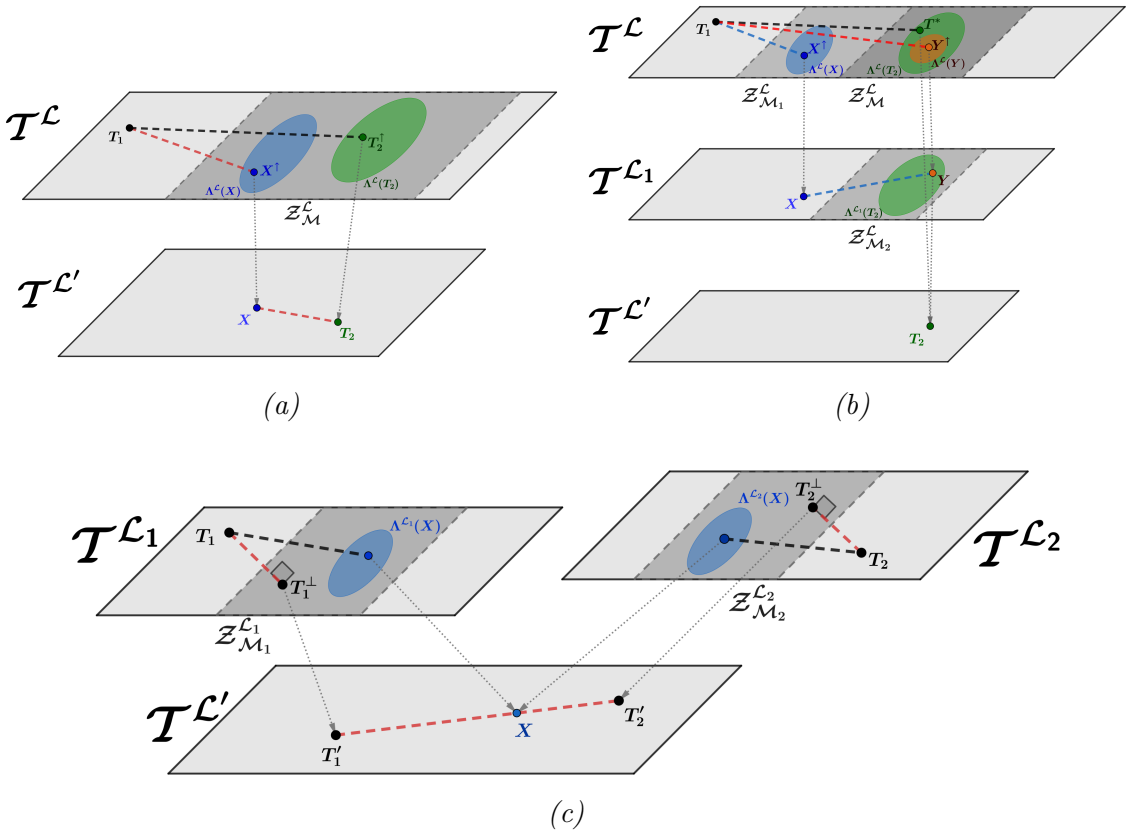


Figure 4.8. Refinements of paths between trees in the towering space. In all figures, gray shaded areas represent relevant trimmable subspaces, and the shortest possible path is given by black dashed lines (a) Refinement of a path by pruning at the end. The red dashed line indicates a path from T_1 to T_2 where leaves $\mathcal{M}$ are pruned in the sprouting space of some tree X (shaded blue area). A shorter path (black dashed line) is found by pruning $\mathcal{M}$ in the sprouting space of T_2 (shaded green area). (b) Path refinement with sequential pruning of leaves $\mathcal{M}_1$ and $\mathcal{M}_2$. The blue dashed line represents a general path from T_1 to T_2 at a lower level with initial pruning in the sprouting space of some tree X (shaded blue area), followed by pruning in some tree Y in the sprouting space of T_2 (shaded green area). A shorter path (red dashed) prunes all leaves in the sprouting space of Y (shaded orange area). Further refinement is achieved by pruning in the optimal tree in the sprouting space of T_2 in the higher BHV level (shaded green area). (c) Shortest path via a common lower level. $T_1^{\perp}$ and $T_2^{\perp}$ represent the projections of trees to the trimmable spaces on gray. Red dashed lines in the higher BHV levels show distances to these projections. In the lower BHV space, the red dashed line is the geodesic between the pruned trees T_1' and T_2'. Shaded blue areas represent the sprouting spaces for tree X on this geodesic. The shortest path (black dashed lines) traverses the top-level spaces to reach sprouting space $\Lambda^{L_1}(X)$, then a prune and immediate regraft to $\Lambda^{L_2}(X)$ is performed, before again traversing the top level space.

geodesic length calculation provided in Section 2.2.2 (see (2.1)), the contribution of common edges is determined by the L^2-norm combination of their differences. Uncommon edges are grouped through support pairs, and the contribution of each support pair (A, B) to the geodesic length is the squared sum of the L^2-norms of each set in the pair, $([||A|| + ||B||]^2$. The edges' lengths contribute similarly in Theorem 4.29, except the classification of "common" and "uncommon" edges is adjusted based on which leaves are predetermined to be pruned. I provide a concrete example of these parallels below.

Example 4.30. Parallels between BHV geodesics and short paths in the towering tree spaces.

Consider the trees T_1 and T_2 in Figure 4.9. While in the same BHV space, we can consider the set of common edges $C = \{p_1 = q_1, \ldots, p_8 = q_8, p_9 = q_9, p_{12} = q_{12}, p_{13} = q_{13}\}$, while the uncommon edges that must be swapped at some point along the geodesic are $\{p_{10}, p_{11}, p_{12}\}$ in T_1 and $\{q_{10}, q_{11}, q_{12}\}$ in T_2.

Nevertheless, if I construct the shortest path between T_1 and T_2, with the pre-determined decision of pruning and regrafting G and H, Theorem 4.29 indicates that its length is given by

$$\sqrt{\left[\sqrt{p_7^2 + p_8^2 + p_{13}^2} + \sqrt{q_7^2 + q_8^2 + q_{13}^2}\right]^2 + \sum_{i=1}^{6}(p_i - q_i)^2 + (p'_9 - q_9)^2 + (p_{11} - q_{11})^2 + (p_{12} - q'_{11})^2},$$

where $p'_9 = \sqrt{p_9^2 + p_{10}^2}$ and $q'_{11} = \sqrt{q_{11}^2 + q_{12}^2}$ is the length of the new edge after pruning G and H from T_1 and T_2 respectively. Comparing this against the expression for computing geodesic lengths in (2.1), the value of the length is analogous to taking common edges $\{p_1 = q_1, \ldots, p_6 = q_6, p'_9 = q_9, p_{11} = q_1 0, p_1 2 = q'_{11}\}$, and the uncommon edges are part of the same "support pair" (A, B) with $A = \{p_7, p_8, p_{13}\}$ and $B = \{q_7, q_8, q_{13}\}$.

One can interpret this as follows: since it has been pre-determined that G and H will be pruned and regrafted, the edges in orange in T_1 must reduce to zero and be dropped before

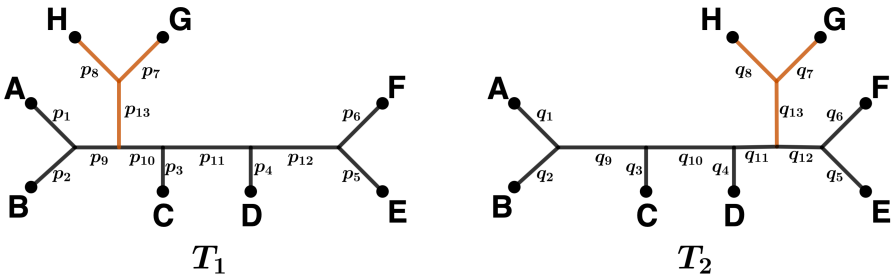


Figure 4.9. Example of two trees in the same BHV space where a shorter path through pruning and regrafting can be found. The trees T_1 and T_2 have similar topologies, except for the position of leaves G and H. The edges preventing $\mathcal{M} = \{G, H\}$ from being prunable are shown in orange. Employing Theorem 4.29, the shortest path through pruning and regrafting $\mathcal{M}$ is possible, with potential improvement versus the BHV distance.

the edges in orange in T_2 are added and increased to the size they have in T_2. Thus, the prune and regraft of G and H transform these effectively into uncommon edges in practice.

While the previous theorem gives the shortest path between $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$ through a common lower level, we must also consider paths between T_1 and T_2 through higher levels. Paths through higher levels have the potential to be shorter than those in the lower levels (Figure 4.10). While in paths through lower levels the edges in T_1 preventing uncommon leaves from being prunable must be reduced to zero-length before the uncommon leaves in T_2 are regrafted (producing the combined term $\left[d_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}_1}) + d_{\text{BHV}}(T_2^{\perp\mathcal{M}_2}, T_2)\right]^2$ in (4.3)), going through common higher BHV levels may allow for these edges to gradually change simultaneously, reducing the size of paths.

4.5 Short paths through higher BHV levels

In this section we explore how to build the short paths between trees $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$ traversing BHV levels that are higher than those containing the trees. The lowest BHV level that is higher than both $\mathcal{T}^{\mathcal{L}_1}$ and $\mathcal{T}^{\mathcal{L}_2}$ is $\mathcal{T}^{\mathcal{L}}$ for $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$. Based on Theorem

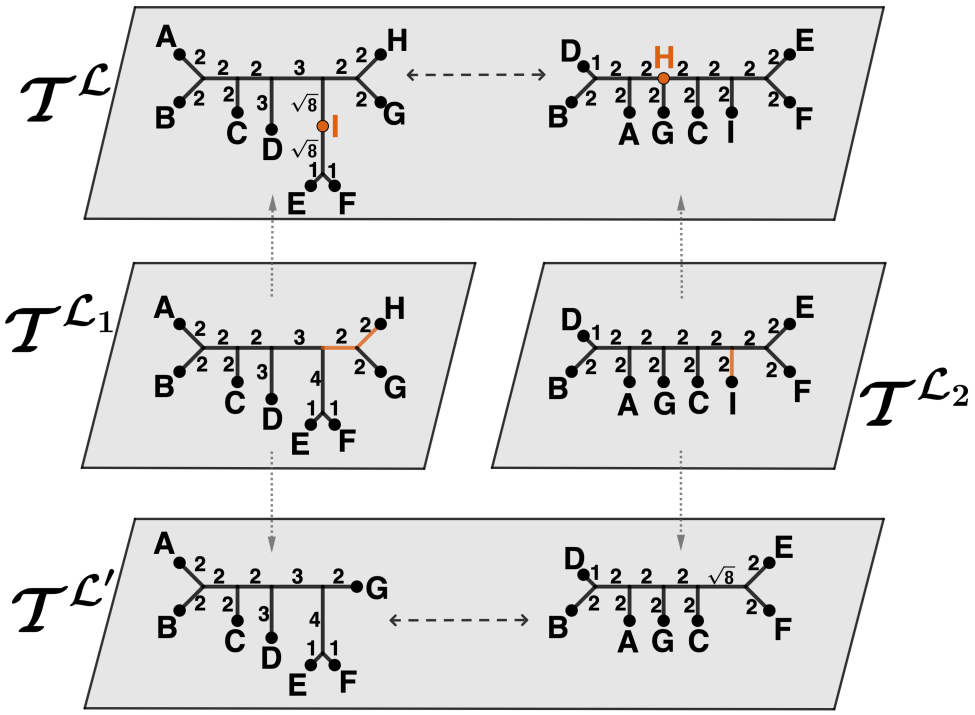


Figure 4.10. Paths between trees in different BHV spaces $\mathcal{T}^{\mathcal{L}_1}$ and $\mathcal{T}^{\mathcal{L}_2}$ via a common lower level and via a common higher level. In the lower path, node H is pruned from the tree on the center-left, and node I is pruned from the tree on the center-right, reducing the corresponding edges (highlighted in orange) to zero length. The shortest path through the lower BHV level, $\mathcal{T}^{\mathcal{L}'}$, has length ≈ 9.36. In the higher level path, node I is regrafted onto the tree on the center-left, and node H is regrafted onto the tree on the center-right. The path length is then the geodesic distance between the trees in the upper level, which is approximately ≈ 8.74.

4.23, I conjecture that no improvements on paths can be reached by traversing higher BHV levels.

Consider all paths from T_1 to T_2 through $\mathcal{T}^{\mathcal{L}}$ that involve regrafting all leaves $\mathcal{M} = \mathcal{L} \setminus \mathcal{L}_1$ missing from T_1 at different points before traversing $\mathcal{T}^{\mathcal{L}}$, and subsequently pruning leaves $\mathcal{K} = \mathcal{L} \setminus \mathcal{L}_2$ missing from T_2. By Lemmas 4.27 and 4.28, the shortest of such paths performs the regrafting of all leaves in $\mathcal{M}$ onto T_1, forming a tree $T_1^{\uparrow} \in \Lambda^{\mathcal{L}}(T_1)$, and then pruning all leaves in $\mathcal{K}$ from a tree $T_2^{\uparrow} \in \Lambda^{\mathcal{L}}(T_2)$. Thus, finding the shortest path from T_1 to T_2 through $\mathcal{T}^{\mathcal{L}}$ is equivalent to the problem of finding the minimum BHV distance between

sprouting spaces; i.e. finding $T_1^* \in \Lambda^{\mathcal{L}}(T_1)$ and $T_2^* \in \Lambda^{\mathcal{L}}(T_2)$ such that $d_{\text{BHV}}(T_1^*, T_2^*) = \min_{t_1 \in \Lambda^{\mathcal{L}}(T_1), t_2 \in \Lambda^{\mathcal{L}}(T_2)} d_{\text{BHV}}(t_1, t_2)$. A subproblem of this optimization is addressed in Theorem 4.26, where the distance from any tree to a sprouting space in the same BHV space is explicitly given in (4.2).

I now discuss how to extend Theorem 4.26 to find the distance between two sprouting spaces in the same BHV space. This extension is based on two observations. Firstly, the minimum distance given in (4.2) depends on two values: the length of the projection from T to the $\mathcal{M}$-trimmable space ($||P^{\downarrow\mathcal{M}}(T)||$) and the BHV distance between the lower level tree T' and the tree resulting from pruning $\mathcal{M}$ from $T^{\perp\mathcal{M}}$ ($d^2_{\text{BHV}}(T', \psi(T^{\perp\mathcal{M}}, \mathcal{M}))$). Secondly, the construction of the optimal tree in the sprouting space $\Lambda^{\mathcal{L}}(T')$ in the proof of Theorem 4.26, leaves in $\mathcal{M}$ are regrafted onto T' to maximize the number of common edges between the optimal tree and T_1. While this strategy of maximizing the number of common edges may reduce $d^2_{\text{BHV}}(T', \psi(T^{\perp\mathcal{M}}, \mathcal{M}))$, regrafting new leaves to edges in $P^{\downarrow\mathcal{M}}(T_1)$ may reduce $||P^{\downarrow\mathcal{M}}(T)||$, resulting in a net decrease in the length of the path.

Given the above observations, I propose the following process to find the shortest path from T_1 to T_2 through $\mathcal{T}^{\mathcal{L}}$: start with a tree $T_1^{\uparrow}$ in the sprouting space $\Lambda^{\mathcal{L}}(T_1)$, and apply Theorem 4.26 to find its distance to the second sprouting space $\Lambda^{\mathcal{L}}(T_2)$, which is given by $\sqrt{||P^{\downarrow\mathcal{K}}(T_1^{\uparrow})||^2 + d^2_{\text{BHV}}(\psi(T_1^{\uparrow\perp\mathcal{K}}, \mathcal{K}, T_2))}$. Then, optimize this expression with respect to the second term while keeping the first term fixed. I formalize this approach in Theorem 4.33 below. Subsequently, in Section 4.5.1, I discuss applying this result to find the optimal solution.

I begin by introducing a subset of sprouting spaces containing trees that are guaranteed to be at the same distance to a trimmable subspace. This facilitates the process described above by aiding in the step of fixing the first term $||P^{\downarrow\mathcal{K}}(T_1^{\uparrow})||^2$. I then provide the preliminary Lemma 4.32.

Definition 4.31. Consider a tree $T_1 \in \mathcal{T}^{\mathcal{L}_1}$, a subset $\mathcal{K} \subset \mathcal{L}_1$ of its leaves, and a subset of leaves $\mathcal{M}$ such that $\mathcal{L}_1 \cap \mathcal{M} = \emptyset$. Given a tree $T_1^{\uparrow} \in \Lambda^{\mathcal{L}}(T_1)$ resulting from regrafting $\mathcal{M}$ (i.e. $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{M}$) and the partition of $\mathcal{K} = \mathcal{K}_1 \sqcup \ldots \sqcup \mathcal{K}_r$ into independent maximal sets in T_1, I define a partition of $\mathcal{M} = \bar{\mathcal{M}}_0 \cup \bar{\mathcal{M}}_1 \cup \ldots \cup \bar{\mathcal{M}}_r$ with respect on how these leaves were regrafted onto T_1 to form $T_1^{\uparrow}$.

In this partition, for each $i = 1, \ldots, r_1$, all leaves in $\bar{\mathcal{M}}_i$ were regrafted on edges $P_i^{\downarrow\mathcal{K}}(T_1)$ belonging to $\mathcal{K}_i$, meaning $\mathcal{S}(T_1^{\uparrow})$ contains the edge $\left[\mathcal{K}_i \cup \bar{\mathcal{M}}_i \ddagger \mathcal{L} \setminus \left(\mathcal{K}_i \cup \bar{\mathcal{M}}_i\right)\right]$ if $\mathcal{K}_i$ is attached to internal edges, or contains $\left[\mathcal{K}_i \cup \bar{\mathcal{M}}_i \cup \{\ell\} \ddagger \mathcal{L} \setminus \left(\mathcal{K}_i \cup \bar{\mathcal{M}}_i \cup \{\ell\}\right)\right]$ if $\mathcal{K}_i$ neighbors ℓ. Let $\bar{\mathcal{M}}_0$ be all other leaves that were regrafted on edges in $\mathcal{S}(T_1) \setminus P^{\downarrow\mathcal{K}}(T_1)$. I denote by $\Lambda^{\mathcal{L}}_{\mathcal{K}}(T_1, T_1^{\uparrow})$ the set of all trees $t_1 \in \Lambda^{\mathcal{L}}_{\mathcal{K}}(T_1, T_1^{\uparrow}) \subseteq \Lambda^{\mathcal{L}}(T_1)$ where all leaves in $\mathcal{M} \setminus \bar{\mathcal{M}}_0$ are regrafted at the same position as in $T_1^{\uparrow}$ (equivalently, $t_1 \in \Lambda^{\mathcal{L}}_{\mathcal{K}}(T_1, T_1^{\uparrow})$ if every $s \in \mathcal{S}(T_1^{\uparrow})$ such that $\Psi_{\mathcal{L}_1}(s) \in P^{\downarrow\mathcal{K}}(T_1)$ also belongs to the edges in t_1 with the same length $|s|_{T_1^{\uparrow}} = |s|_{t_1}$).

Lemma 4.32. *Consider two trees $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$, with $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$. Let $\mathcal{L}' = \mathcal{L}_1 \cap \mathcal{L}_2$, $\mathcal{K} = \mathcal{L}_1 \setminus \mathcal{L}'$ and $\mathcal{M} = \mathcal{L}_2 \setminus \mathcal{L}'$. Fix a tree $T_1^{\uparrow} \in \Lambda^{\mathcal{L}}(T_1)$ and take the partition $\mathcal{K} = \mathcal{K}_1 \sqcup \ldots \sqcup \mathcal{K}_r$ into independent maximal sets in T_1 and the partition of $\mathcal{M} = \bar{\mathcal{M}}_0 \cup \bar{\mathcal{M}}_1 \cup \ldots \cup \bar{\mathcal{M}}_r$, so that in $T_1^{\uparrow}$, all leaves in $\bar{\mathcal{M}}_i$ are regrafted on edges belonging to $\mathcal{K}_i$, as in Definition 4.31. Then $||P^{\downarrow\mathcal{K}}(T_1^{\uparrow})|| = ||P^{\downarrow\mathcal{K}}(t_1)||$ for all $t_1 \in \Lambda^{\mathcal{L}}_{\mathcal{K}}(T_1, T_1^{\uparrow})$, $\bar{\mathcal{M}}_0$ is mutually prunable from $t_1^{\perp\mathcal{K}}$, and $T_1^{\downarrow} = \psi(T_1^{\uparrow\perp\mathcal{K}}, \mathcal{K} \cup \bar{\mathcal{M}}_0) = \psi(t_1^{\perp\mathcal{K}}, \mathcal{K} \cup \bar{\mathcal{M}}_0)$ for all $t_1 \in \Lambda^{\mathcal{L}}_{\mathcal{K}}(T_1, T_1^{\uparrow})$.*

Proof. See Appendix B.4 □

In the previous lemma, I show that by pre-determining which leaves from $\mathcal{M}$ are regrafted onto the edges belonging to the independent sets of $\mathcal{K}$ in T_1 and specifying their positions, the length of the projection onto the $\mathcal{K}$-trimmable subspace is fixed. Furthermore, the trees resulting from pruning $\mathcal{K}$ are in the same equivalent class, since they map to the same tree after pruning $\bar{\mathcal{M}}_0$ whose regrafting positions were not pre-determined. I use this fact in the

following theorem, where I establish the shortest path from T_1 to T_2 through trees with these fixed leaf positions.

Theorem 4.33. *Given $T_1 \in \mathcal{T}^{\mathcal{L}_1}$, $T_2 \in \mathcal{T}^{\mathcal{L}_2}$, $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$, $\mathcal{L}' = \mathcal{L}_1 \cap \mathcal{L}_2$, $\mathcal{K} = \mathcal{L}_1 \setminus \mathcal{L}'$ and $\mathcal{M} = \mathcal{L}_2 \setminus \mathcal{L}'$. Consider a fixed tree $T_1^{\uparrow} \in \Lambda^{\mathcal{L}}(T_1)$. Then, the BHV distance between any $t_1 \in \Lambda^{\mathcal{L}}_{\mathcal{K}}(T_1, T_1^{\uparrow})$ and any $t_2 \in \Lambda^{\mathcal{L}}(T_2)$ is lower bounded by*

$$d_{BHV}(t_1, t_2) \geq \sqrt{||P^{\downarrow\mathcal{K}}(T_1^{\uparrow})||^2 + ||P^{\downarrow\bar{\mathcal{M}}_0}(T_2)||^2 + d^2_{BHV}(T_1^{\downarrow}, T_2^{\downarrow})},$$

where $T_1^{\downarrow} = \psi(T_1^{\uparrow\perp\mathcal{K}}, \mathcal{K} \cup \bar{\mathcal{M}}_0)$ and $T_2^{\downarrow} = \psi(T_2^{\perp\bar{\mathcal{M}}_0}, \bar{\mathcal{M}}_0)$. Additionally, there exists $T_1^* \in \Lambda^{\mathcal{L}}(T_1)$ and $T_2^* \in \Lambda^{\mathcal{L}}(T_2)$ with $||P^{\downarrow\mathcal{K}}(T_1^{\uparrow})|| = ||P^{\downarrow\mathcal{K}}(T_1^*)||$ that achieve this lower bound, i.e.,

$$d_{BHV}(T_1^*, T_2^*) = \sqrt{||P^{\downarrow\mathcal{K}}(T_1^{\uparrow})||^2 + ||P^{\downarrow\bar{\mathcal{M}}_0}(T_2)||^2 + d^2_{BHV}(T_1^{\downarrow}, T_2^{\downarrow})} \tag{4.4}$$

Proof. See Appendix B.4. □

The previous theorem shows how to determine the optimal locations for regrafting the remaining leaves in $\mathcal{M}$, given which leaves are regrafted onto T_1 within the edges of $P^{\downarrow\mathcal{K}}(T_1)$, as well as the positions where these are regrafted. Additionally, it identifies where to regraft leaves in $\mathcal{K}$ onto T_2. This results in the expression (4.4). In the next section I discuss the procedure to minimize this expression by selecting the best position for the leaves regrafted onto $P^{\downarrow\mathcal{K}}(T_1)$.

Theorem 4.33 suggests a method for finding short paths from $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ to $T_2 \in \mathcal{T}^{\mathcal{L}_2}$ through the common higher BHV level $\mathcal{T}^{\mathcal{L}_1 \cup \mathcal{L}_2}$, while Theorem 4.29 identifies the shortest paths through common lower BHV levels. Building on these results, I now present a more general result that offers a method to connect the two trees through paths traversing BHV spaces that are neither entirely above nor below, achieving the shortest possible connection (Theorem 4.34). In the subsequent corollary, I establish that, despite the multiple choices for pruning and regrafting uncommon leaves to form general paths as presented in Theorem

4.34, the optimal strategy involves first regrafting all leaves missing from T_1, then pruning and regrafting some of the common leaves, and finally pruning all leaves missing from T_2.

Theorem 4.34. *Given two trees $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$, with $\mathcal{L}' = \mathcal{L}_1 \cap \mathcal{L}_2$, take $\mathcal{K} = \mathcal{L}_1 \setminus \mathcal{L}'$ the leaves in T_1 missing from T_2 and $\mathcal{M} = \mathcal{L}_2 \setminus \mathcal{L}'$ the leaves in T_2 missing from T_1. Consider partitions of the uncommon and common leaves $\mathcal{K} = \mathcal{K}^U \sqcup \mathcal{K}^D$, $\mathcal{M} = \mathcal{M}^U \sqcup \mathcal{M}^D$, and $\mathcal{L} = \mathcal{L}^U \sqcup \mathcal{L}^D$. For these partitions, take $\mathcal{K}^* = \mathcal{K} \cup \mathcal{L}^D$ and $\mathcal{M}^* = \mathcal{M} \cup \mathcal{L}^D$, and their partitions into independent sets $\mathcal{K}^* = \mathcal{K}^*_1 \sqcup \ldots \sqcup \mathcal{K}^*_{r_1}$ and $\mathcal{M}^* = \mathcal{M}^*_1 \sqcup \ldots \sqcup \mathcal{M}^*_{r_2}$ in T_1 and T_2 respectively. Fix two trees $T_1^{\uparrow} = \Lambda^{\mathcal{L}' \cup \mathcal{K} \cup \mathcal{M}^U}(T_1)$ and $T_2^{\uparrow} = \Lambda^{\mathcal{L}' \cup \mathcal{M} \cup \mathcal{K}^U}(T_2)$, and consider the partition of the new leaves in them $\mathcal{M}^U = \bar{\mathcal{M}}^U_0 \sqcup \ldots \sqcup \bar{\mathcal{M}}^U_{r_1}$ and $\mathcal{K}^U = \bar{\mathcal{K}}^U_0 \sqcup \ldots \sqcup \bar{\mathcal{K}}^U_{r_2}$, so that in $T_1^{\uparrow}$ all leaves in $\bar{\mathcal{M}}^U_i$ were regrafted onto edges in $P_i^{\downarrow \mathcal{K}^*}(T_1)$ for $i = 1, \ldots, r_1$, and in $T_2^{\uparrow}$, leaves in $\bar{\mathcal{K}}^U_i$ were regrafted onto edges $P_i^{\downarrow \mathcal{M}^*}(T_2)$ for $i = 1, \ldots, r_2$ as in Definition 4.31*

The length of a path from any tree $t_1 \in \Lambda^{\mathcal{L}' \cup \mathcal{K} \cup \mathcal{M}^U}_{\mathcal{K}^*}(T_1, T_1^{\uparrow})$ to any tree $t_2 \in \Lambda^{\mathcal{L}' \cup \mathcal{M} \cup \mathcal{K}^U}_{\mathcal{M}^*}(T_2, T_2^{\uparrow})$ by pruning $\mathcal{K}^{D*} = \mathcal{K}^D \cup \mathcal{L}^D$ and regrafting $\mathcal{M}^{D*} = \mathcal{M}^D \cup \mathcal{L}^D$ is lower bounded by

$$\sqrt{\left[||P^{\downarrow \mathcal{K}^{D*}}(T_1^{\uparrow})|| + ||P^{\downarrow \mathcal{M}^{D*}}(T_2^{\uparrow})||\right]^2 + ||P^{\downarrow \bar{\mathcal{K}}^U_0}(T_1^{\downarrow})||^2 + ||P^{\downarrow \bar{\mathcal{M}}^U_0}(T_2^{\downarrow})||^2 + d^2_{BHV}(T_1^{\downarrow\downarrow}, T_2^{\downarrow\downarrow})}, \tag{4.5}$$

where $T_1^{\downarrow} = \psi(T_1^{\uparrow \perp \mathcal{K}^{D*}}, \mathcal{K}^{D*} \cup \bar{\mathcal{M}}^U_0)$, $T_2^{\downarrow} = \psi(T_2^{\uparrow \perp \mathcal{M}^{D*}}, \mathcal{M}^{D*} \cup \bar{\mathcal{K}}^U_0)$, $T_1^{\downarrow\downarrow} = \psi(T_1^{\downarrow \perp \bar{\mathcal{K}}^U_0}, \bar{\mathcal{K}}^U_0)$ and $T_2^{\downarrow\downarrow} = \psi(T_2^{\downarrow \perp \bar{\mathcal{M}}^U_0}, \bar{\mathcal{M}}^U_0)$. Moreover, it is possible to find optimal trees $T_1^* \in \Lambda^{\mathcal{L}' \cup \mathcal{K} \cup \mathcal{M}^U}(T_1)$ and $T_2^* \in \Lambda^{\mathcal{L}' \cup \mathcal{M} \cup \mathcal{K}^U}(T_2)$ whose shortest path through the common space $\mathcal{T}^{\mathcal{L}^U \cup \mathcal{K}^U \cup \mathcal{M}^U}$ achieves this lower bound.

Proof. Refer to Appendix B.4 □

Corollary 4.35. *Given two trees $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$, with $\mathcal{L}' = \mathcal{L}_1 \cap \mathcal{L}_2$, take $\mathcal{K} = \mathcal{L}_1 \setminus \mathcal{L}'$ the leaves in T_1 missing from T_2 and $\mathcal{M} = \mathcal{L}_2 \setminus \mathcal{L}'$ the leaves in T_2 missing from T_1. Consider partitions of the uncommon and common leaves $\mathcal{K} = \mathcal{K}^U \sqcup \mathcal{K}^D$, $\mathcal{M} = \mathcal{M}^U \sqcup \mathcal{M}^D$, and $\mathcal{L} = \mathcal{L}^U \sqcup \mathcal{L}^D$. Consider the set of paths from T_1 to T_2 where the order of the transitions*

between BHV levels is: regrafting $\mathcal{M}^U$, pruning $\mathcal{K}^D \cup \mathcal{L}^D$, regrafting $\mathcal{M}^D \cup \mathcal{L}^D$ and finally pruning $\mathcal{K}^U$. The shortest path among these set of paths is at least as long as the shortest path among those where: $\mathcal{M}$ is regrafted first, then $\mathcal{L}^D$ is pruned and regrafted, and $\mathcal{K}$ is pruned at the end.

Proof. For a path as the one described in Theorem 4.34, new trees $T_1^{\uparrow\uparrow} = \Lambda^{\mathcal{L}' \cup \mathcal{K} \cup \mathcal{M}}(T_1)$ and $T_2^{\uparrow\uparrow} = \Lambda^{\mathcal{L}' \cup \mathcal{M} \cup \mathcal{K}}(T_2)$ are build by regrafting $\mathcal{M}^D$ and $\mathcal{K}^D$ onto $T_1^{\uparrow}$ and $T_2^{\uparrow}$ respectively, in such a way that Theorem 4.34 applied to those new trees returns a shorter path. See details in Appendix B.4. □

In the last corollary, I have proven that the shortest path from $T_1 \in \mathcal{T}^{\mathcal{L}_1}$, $T_2 \in \mathcal{T}^{\mathcal{L}_2}$, where a certain $\mathcal{L}^* \subset \mathcal{L}_1 \cap \mathcal{L}_2$ is to be pruned and regrafted at some point along the path, starts at some tree in $\Lambda^{\mathcal{L}_1 \cup \mathcal{L}_2}(T_1)$, performing the pruning and regrafting of $\mathcal{L}^*$ at some point, and ending at a tree in $\Lambda^{\mathcal{L}_1 \cup \mathcal{L}_2}(T_2)$. The length of this shortest path depends on the positions where $\mathcal{M}$ and $\mathcal{K}$ are regrafted to create the starting and ending trees in $\Lambda^{\mathcal{L}_1 \cup \mathcal{L}_2}(T_1)$ and $\Lambda^{\mathcal{L}_1 \cup \mathcal{L}_2}(T_2)$ respectively. This process is discussed in the following section.

4.5.1 Optimization of paths through high BHV levels

Consider the setting of Theorem 4.33, and the partition $\mathcal{K} = \mathcal{K}_1 \sqcup \ldots \sqcup \mathcal{K}_r$ into independent maximal sets. Let $\mathcal{M} = \bar{\mathcal{M}}_0 \sqcup \bar{\mathcal{M}}_1 \sqcup \ldots \sqcup \bar{\mathcal{M}}_r$ be a pre-fixed partition of $\mathcal{M}$, with $\bar{\mathcal{M}}_i$ corresponding to leaves to be regrafted onto edges belonging to $\mathcal{K}_i$. Observe that in (4.4), the position of each leaf in $\bar{\mathcal{M}}_i$ for $i = 1, \ldots, r$ does not affect the value of the term $||P^{\downarrow \bar{\mathcal{M}}_0}(T_2)||^2$, so it only affects the distance by changing the value of the first and last terms ($||P^{\downarrow \mathcal{K}}(T_1^{\uparrow})||^2$ and $d^2_{\mathrm{BHV}}(T_1^{\downarrow}, T_2^{\downarrow})$). The process of selecting the positions for leaves in each $\bar{\mathcal{M}}_i$ for $i = 1, \ldots, r$ is based on balancing these two values.

Knowing beforehand the partition $\mathcal{M} = \bar{\mathcal{M}}_0 \sqcup \bar{\mathcal{M}}_1 \sqcup \ldots \sqcup \bar{\mathcal{M}}_r$, it is possible to find the best positions for leaves in each $\bar{\mathcal{M}}_i$ independently from the other subsets. For simplicity, I

will explain the process for the independent set $\mathcal{K}_1$, assuming it is attached to internal edges, following Example 4.36. The process is analogous for every independent set.

When attaching leaves from $\bar{\mathcal{M}}_1$ onto edges in $P_1^{\downarrow\mathcal{K}}(T_1)$, new edges are introduced to the leaf. The squared length of some of these edges contribute to $||P^{\downarrow\mathcal{K}}(T_1^{\uparrow})||^2 = ||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow})||^2 + ||P^{\downarrow(\mathcal{K}\setminus\mathcal{K}_1)}(T_1^{\uparrow})||^2$ (shown in orange in Figure 4.11), while the squared lengths of others (shown in black in Figure 4.11) contribute to $d^2_{\text{BHV}}(T_1^{\downarrow}, T_2^{\downarrow})$.

When all leaves in $\bar{\mathcal{M}}_1$ are regrafted at the node connecting the subtree defined by $P_1^{\downarrow\mathcal{K}}(T_1)$ (as in $T_1^{\uparrow 0}$ in Figure 4.11) the length of the edges in this subtree all contribute to the size of the projection, meaning $||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow})||^2 = ||P^{\downarrow\mathcal{K}_1}(T_1)||^2$; and none contribute to the distance $d^2_{\text{BHV}}(T_1^{\downarrow}, T_2^{\downarrow})$.

Now imagine some of the leaves $\mathcal{M}' \subset \bar{\mathcal{M}}_1$ are regrafted within the first edge up in the independent set instead, (as in trees $T_1^{\uparrow 1}$ and $T_1^{\uparrow 2}$ in Figure 4.11). The value $||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow})||^2$ decreases by the squared length of the size of the (new) edge $a_1 = |s_1|_{T_1^{\uparrow 1}} = |[\mathcal{K}_1 \cup \mathcal{M}' \ddagger \mathcal{L} \setminus (\mathcal{K}_1 \cup \mathcal{M}')]|_{T_1^{\uparrow 1}}$, but this value affects the distance between $T_1^{\downarrow 1}$ and $T_2^{\downarrow}$, increasing or decreasing depending on $T_2^{\downarrow}$. In fact, if $s_1^{\downarrow} = \Psi_{\mathcal{L}_2\setminus\bar{\mathcal{M}}_0}(s_1)$ belongs to the edges of $T_2^{\downarrow}$ with length b, then the contribution of this edge to the overall value of $d^2_{\text{BHV}}(T_1^{\downarrow}, T_2^{\downarrow})$ is $(a_1 - b)^2$, which is smaller than the value a_1^2 that decreased in $||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow})||^2$, and this regraft is overall beneficial. If, on the contrary, $s_1^{\downarrow}$ is not part of the edges in $T_2^{\downarrow}$, then it maps to an uncommon edge between $T_1^{\downarrow 1}$ and $T_2^{\downarrow}$ and the contribution to length of the geodesic ends up being higher than a_1^2. This regraft would be overall disadvantageous. See Example 4.36 for details on this for trees $T_1^{\uparrow 1}$ and $T_1^{\uparrow 2}$ in Figure 4.11 in particular.

I conclude from this that, knowing which leaves are in $\bar{\mathcal{M}}_1$ and the set of edges $\mathcal{S}(T_2^{\downarrow})$, the regraft positions for leaves in $\bar{\mathcal{M}}_1$ reduce the length of the path given in Theorem 4.33 (equation (4.4)) when introducing edges already in $\mathcal{S}(T_2^{\downarrow})$ into the tree $T_1^{\downarrow}$. Moreover, it is better to introduce as many as these common edges as possible, and for them to decrease the lengths of the edges in $P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow})$ as much as possible (see trees $T_1^{\uparrow 3}$, $T_1^{\uparrow 4}$ and $T_1^{\uparrow 5}$ in Example

4.36).

Example 4.36. Selecting regraft positions for new leaves onto independent set: Consider the setting of Theorem 4.33, with tree $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ where the subset of leaves $\mathcal{K}_1 = \{A, B, C, D, E\}$ is an independent maximal set attached to internal edges and the set of leaves $\bar{\mathcal{M}}_1 = \{u, v, w, z\}$ being regrafted onto $P_1^{\downarrow\mathcal{K}}(T_1)$. Figure 4.11 shows 6 different ways to perform the regraft. Assume edges $p_1 = \big[\{u, v\} \ddagger (\mathcal{L}_2 \setminus \bar{\mathcal{M}}_0) \setminus \{u, v\}\big]$, $p_2 = \big[\{u, v, z\} \ddagger (\mathcal{L}_2 \setminus \bar{\mathcal{M}}_0) \setminus \{u, v, z\}\big]$ and $p_3 = \big[\{u, v, w, z\} \ddagger (\mathcal{L}_2 \setminus \bar{\mathcal{M}}_0) \setminus \{u, v, w, z\}\big]$ belongs to the internal edges of $T_2^{\downarrow}$ with lengths $|p_1|_{T_2^{\downarrow}} = 4$, $|p_2|_{T_2^{\downarrow}} = 3$ and $|p_3|_{T_2^{\downarrow}} = 3$.

In the first tree $T_1^{\uparrow 0}$ in Figure 4.11 all leaves in $\bar{\mathcal{M}}_1$ are regrafted at the endpoint of the edge $s = [\mathcal{K}_1 \ddagger \mathcal{L}_1 \setminus \mathcal{K}_1]$, so $\mathcal{S}(T_1^{\uparrow 0})$ contains the edge $s_1 = [\mathcal{K}_1 \ddagger \mathcal{L} \setminus \mathcal{K}_1]$ with the same length as s in T_1. In this case $||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 0})||^2 = ||P^{\downarrow\mathcal{K}_1}(T_1)||^2$ and none of the lengths of the edges belonging to $\mathcal{K}_1$ contribute to the distance $d^2_{\text{BHV}}(T_1^{\downarrow 0}, T_2^{\downarrow})$. Comparing this values against the corresponding values for all other 5 trees in the figure, I can determine which regraft positions reduces the value of (4.4) the most.

Comparing $T_1^{\uparrow 0}$ and $T_1^{\uparrow 1}$, $||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 1})||^2 = ||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 0})||^2 - 5$ and $d^2_{\text{BHV}}(T_1^{\downarrow 1}, T_2^{\downarrow}) = d^2_{\text{BHV}}(T_1^{\downarrow 0}, T_2^{\downarrow}) - 3^2 + (3 - \sqrt{5})^2$; because in the geodesic from $T_1^{\downarrow 0}$ to $T_2^{\downarrow}$ the edge p_1 goes from length zero to length 3, while in the geodesic from $T_1^{\downarrow 1}$ to $T_2^{\downarrow}$ the size of this edge changes from $\sqrt{5}$ to 3. So overall the size of the path decreased by $6\sqrt{5}$.

In contrast, when comparing $T_1^{\uparrow 0}$ against $T_1^{\uparrow 2}$, both p_1 and p_2 are the only non-compatible edges to $p_4 = \big[\{u, w\} \ddagger (\mathcal{L}_2 \setminus \bar{\mathcal{M}}_0) \setminus \{u, w\}\big]$ in $T_2^{\downarrow}$, so these two edges need to be swapped with p_4 at some point in the geodesic from $T_1^{\downarrow 2}$ and $T_2^{\downarrow}$. Thus, $||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 2})||^2 = ||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 0})||^2 - 5$ as with $T_1^{\uparrow 1}$, but $d^2_{\text{BHV}}(T_1^{\downarrow 1}, T_2^{\downarrow}) = d^2_{\text{BHV}}(T_1^{\downarrow 0}, T_2^{\downarrow}) - 3^2 - 4^2 + (\sqrt{3^2 + 4^2} + \sqrt{5})^2$, so overall the size of the path increases by $10\sqrt{5}$.

Following the same process for the rest of the trees in Figure 4.11, I obtain:

- $||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 3})||^2 = ||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 0})||^2 - 3^2$ and $d^2_{\text{BHV}}(T_1^{\downarrow 3}, T_2^{\downarrow}) = d^2_{\text{BHV}}(T_1^{\downarrow 0}, T_2^{\downarrow}) - 3^2 + (3 - 3)^2$,

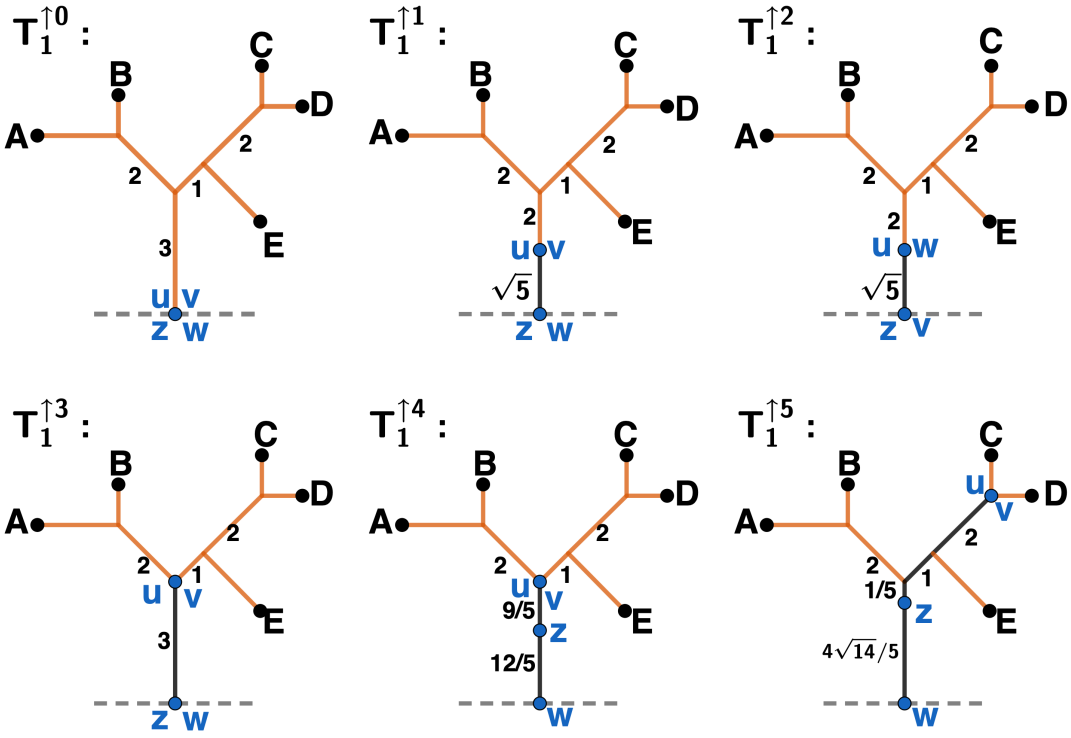


Figure 4.11. Example of finding optimal places to regraft new leaves onto edges belonging to an independent maximal set. The section of a tree T_1 containing independent maximal set $\mathcal{K}_1 = \{A, B, C, D, E\}$, onto which new leaves $\bar{\mathcal{M}}_1 = \{u, v, w, z\}$ are regrafted. In each figure, leaves in $\bar{\mathcal{M}} = \{u, v, w, z\}$ (blue) are regrafted at different positions to create trees $T_1^{\uparrow j}$; edges shown in orange are part of in $P^{\downarrow\mathcal{K}}(T_1^{\uparrow j})$, while edges in black contribute to the distance between $T_1^{\downarrow j}$ and $T_2^{\downarrow}$.

so the overall decrease is 18.

- $||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 4})||^2 = ||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 0})||^2 - 3^2$ and $d^2_{\text{BHV}}(T_1^{\downarrow 4}, T_2^{\downarrow}) = d^2_{\text{BHV}}(T_1^{\downarrow 0}, T_2^{\downarrow}) - 3^2 - 4^2 + (3 - \frac{9}{5})^2 + (4 - \frac{12}{5})^2$, so the overall decrease is 30.

- $||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 5})||^2 = ||P^{\downarrow\mathcal{K}_1}(T_1^{\uparrow 0})||^2 - 3^2 + 1^2 + 2^2$ and $d^2_{\text{BHV}}(T_1^{\downarrow 4}, T_2^{\downarrow}) = d^2_{\text{BHV}}(T_1^{\downarrow 0}, T_2^{\downarrow}) - 3^2 - 4^2 + \left(3 - \sqrt{\left(\frac{1}{5}\right)^2 + 1^2 + 2^2}\right)^2 + \left(4 - \frac{4\sqrt{14}}{5}\right)^2$, so the overall decrease is $-39 + \left(4 - \frac{4\sqrt{14}}{5}\right)^2 + \left(3 - \frac{3\sqrt{14}}{5}\right)^2 \approx 37.42$.

The regraft of $\bar{\mathcal{M}}_1$ in $T_1^{\uparrow 5}$ is actually the optimal in this case. In deciding where to regraft the leaves for this case, the leaves u and v were regrafted as far as possible from the node connecting the edges in orange to the rest of the tree, introducing the common edge p_1 and reducing the lengths of the orange edges still present in the tree as much as possible. Afterwards, the common p_2 can also be introduced, by regrafting z somewhere within the edges from that node to the node connecting to u and v. The position was selected so the squared lengths for p_1 and p_2 in $T_1^{\downarrow 5}$ ($\left(\frac{1}{5}\right)^2+1^2+2^2 = \frac{125}{25}$ and $\left(\frac{4\sqrt{14}}{5}\right)^2 = \frac{224}{25}$) are proportional to the squared lengths of the same edges in $T_2^{\downarrow}$ ($3^3 = 9$ and $4^2 = 16$).

In general, a similar process as the one described at the end of the example can be followed to find the optimal points for any pre-fixed set $\bar{\mathcal{M}}_1$ of leaves to be regrafted onto the edges belonging to $\mathcal{K}_1$, so that the length of the path described by (4.4) is minimized. These would be a process with incremental improvement, detailed in Algorithm 2.

Algorithm 2: Process to optimize regraft locations for leaves $\bar{\mathcal{M}}_1$.

1: Regraft all leaves at the node connecting $P_1^{\downarrow\mathcal{K}}(T_1)$ to the rest of the tree, creating $T_1^{\uparrow 0}$.
2: Identify all edges in $\mathcal{S}(T_2^{\downarrow})$ that could potentially be introduced as edges in $T_1^{\downarrow}$. This is,
$P^{\text{new}} = \left\{ \left[\mathcal{M}' \ddagger (\mathcal{L}_2 \setminus \bar{\mathcal{M}}_0) \setminus \mathcal{M}' \right] \in \mathcal{S}(T_2^{\downarrow}) \mid \mathcal{M}' \subseteq \bar{\mathcal{M}}_1 \right\}$
3: Select $\mathcal{M}' \subseteq \bar{\mathcal{M}}_1$ forming an edge in P^{new} such that $\mathcal{M}'$ is one of the sets with fewer elements forming an edge in P^{new}.
4: Change regraft of leaves in $\mathcal{M}'$ to a positions as far as possible (with respect to the L^2-norm of the edges between them) from the initial node.
5: Continue this process with the other leaves remaining at the initial node, introducing more common edges, and selecting their regrafting position such that is the same position (proportional wise) as in $T_2^{\downarrow}$.

Consider now the setting where one wishes to find the shortest path between $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$ where a pre-fixed set $\mathcal{L}^* \subset \mathcal{L}_1 \cap \mathcal{L}_2$ is to be pruned and regrafted along the path. Per Corollary 4.35, the shortest path where $\mathcal{L}^*$ is pruned and regrafted involves starting by regrafting all leaves in $\mathcal{M}$ onto T_1 and ending by pruning all leaves in $\mathcal{K}$ to end in $\mathcal{T}_2$, and the length of this path will be given by (4.5) for some regraft positions. The process

of finding the best regrafting positions to minimize the value of (4.5) in thist setting is more challenging. I explain briefly a greedy approach to solving this problem.

Following Theorem 4.34, fix $T_1^{\uparrow} \in \Lambda^{\mathcal{L}_1 \cap \mathcal{L}_2}(T_1)$, $T_2^{\uparrow} \in \Lambda^{\mathcal{L}_1 \cap \mathcal{L}_2}(T_2)$ and, given the partitions $\mathcal{K} \cup \mathcal{L}^* = \mathcal{K}^* = \mathcal{K}_1^* \sqcup \ldots \sqcup \mathcal{K}_{r_1}^*$ and $\mathcal{M} \cup \mathcal{L}^* = \mathcal{M}^* = \mathcal{M}_1^* \sqcup \ldots \sqcup \mathcal{M}_{r_2}^*$ into independent maximal sets, consider the partitions of $\mathcal{M} = \bar{\mathcal{M}}_0 \sqcup \ldots \sqcup \bar{\mathcal{M}}_{r_1}$ and $\mathcal{K} = \bar{\mathcal{K}}_0 \sqcup \ldots \sqcup \bar{\mathcal{K}}_{r_2}$, so that in $T_1^{\uparrow}$ all leaves in $\bar{\mathcal{M}}_i$ were regrafted onto edges belonging to $\mathcal{K}_i^*$ for $i = 1, \ldots, r_1$, and equivalently, in $T_2^{\uparrow}$, leaves in $\bar{\mathcal{K}}_i$ were regrafted onto edges belonging to $\mathcal{M}_i^*$ for $i = 1, \ldots, r_2$ as in Definition 4.31. The expression to be minimized in this case is

$$\sqrt{\left[||P^{\downarrow \mathcal{L}^*}(T_1^{\uparrow})|| + ||P^{\downarrow \mathcal{L}^*}(T_2^{\uparrow})||\right]^2 + ||P^{\downarrow \bar{\mathcal{K}}_0}(T_1^{\downarrow})||^2 + ||P^{\downarrow \bar{\mathcal{M}}_0}(T_2^{\downarrow})||^2 + d^2_{\mathrm{BHV}}(T_1^{\downarrow\downarrow}, T_2^{\downarrow\downarrow})}, \quad (4.6)$$

where $T_1^{\downarrow} = \psi(T_1^{\uparrow \perp \mathcal{L}^*}, \mathcal{L}^* \cup \bar{\mathcal{M}}_0)$, $T_2^{\downarrow} = \psi(T_2^{\uparrow \perp \mathcal{L}^*}, \mathcal{L}^* \cup \bar{\mathcal{K}}_0)$, $T_1^{\downarrow\downarrow} = \psi(T_1^{\downarrow \perp \bar{\mathcal{K}}_0}, \bar{\mathcal{K}}_0)$ and $T_2^{\downarrow\downarrow} = \psi(T_2^{\downarrow \perp \bar{\mathcal{M}}_0}, \bar{\mathcal{M}}_0)$.

In this case, the positions of the leaves in each $\bar{\mathcal{M}}_i$ for $i = 1, \ldots, r_1$ affect the size of the terms $||P^{\downarrow \mathcal{L}^*}(T_1^{\uparrow})||$, $||P^{\downarrow \bar{\mathcal{K}}_0^U}(T_1^{\downarrow})||$, and $d^2_{\mathrm{BHV}}(T_1^{\downarrow\downarrow}, T_2^{\downarrow\downarrow})$. Similarly, the positions of the leaves in each $\bar{\mathcal{K}}_i$ for $i = 1, \ldots, r_2$ affect the size of the terms $||P^{\downarrow \mathcal{L}^*}(T_2^{\uparrow})||$, $||P^{\downarrow \bar{\mathcal{M}}_0^U}(T_2^{\downarrow})||$, and $d^2_{\mathrm{BHV}}(T_1^{\downarrow\downarrow}, T_2^{\downarrow\downarrow})$. I propose an iterative process:

1. Fix the positions where leaves in $\mathcal{K} \setminus \bar{\mathcal{K}}_0$ are regrafted and optimize the positions of each $\bar{\mathcal{M}}_i$ for $i = 1, \ldots, r_1$ using the method discussed in Algorithm 2, with $T_2^{downarrow\downarrow}$ with the fixed positions of $\mathcal{K} \setminus \bar{\mathcal{K}}_0$ in place of $\mathcal{T}_2^{\downarrow}$

2. Then, with the new positions for leaves in $\mathcal{M} \setminus \bar{\mathcal{M}}_0$ fixed, optimize the positions of each $\bar{\mathcal{K}}_i$ for $i = 1, \ldots, r_2$ in the same way.

3. Repeat this process until no further improvement in the expression (4.6) is achieved.

I conjecture that this greedy algorithm will obtain the shortest possible for (4.6).

4.6 Distance computation algorithm

I now combine the above into an algorithm to find the distance between any two trees $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$ in the towering space $(\mathbb{T}, d)$.

In this algorithm, I loop through every possible subset of the common leaves $\mathcal{L}^* \subset \mathcal{L}_1 \cap \mathcal{L}_2$ (including the empty set) and construct short paths where these subsets are pruned and regrafted based on Theorem 4.34 and Corollary 4.35. I conjecture that the shortest path between T_1 and T_2 must be as described in the Corollary 4.35, since Theorem 4.23 suggests pruning and regrafting leaves outside $\mathcal{L}_1 \cup \mathcal{L}_2$ is not beneficial at any point of the path. Furthermore, Lemmas 4.27 and 4.28 indicate the prune and regraft of leaves in $\mathcal{L}^*$ should happen at the same time.

4.7 Discussion

In this chapter, I introduce a comprehensive method to construct metric spaces for phylogenetic trees with varying leaf sets through the family of *towering tree spaces*. These spaces are complete length spaces with continuous paths connecting any pair of trees, enabling smooth changes of edge lengths and topologies along these paths as they traverse the space. This approach borrows properties from the BHV tree space, allowing for the development of similar mathematical and statistical tools. Towering tree spaces are the first metric spaces for phylogenetic trees with non-identical leaf sets, enabling smooth comparisons of trees based on topological differences and edge lengths.

While previous works have attempted to utilize the BHV space to compare phylogenetic trees with differing leaf sets (Grindstaff & Owen, 2019; Ren et al., 2017), the tools presented in this chapter take a more expansive approach. Instead of restricting the analysis to a single BHV space, they allow for movement across multiple BHV spaces with as many transitions

Algorithm 3: A greedy algorithm to compute Towering Tree Space distance $(\mathbb{T}, d)$.

INPUT: $T_1 \in \mathcal{T}^{\mathcal{L}_1}$ and $T_2 \in \mathcal{T}^{\mathcal{L}_2}$
for Every $\mathcal{L}^* \subseteq \mathcal{L}_1 \cap \mathcal{L}_2$ **do**
  Take $\mathcal{K}^* = \mathcal{K} \cup \mathcal{L}^*$ and $\mathcal{M}^* = \mathcal{M} \cup \mathcal{L}^*$
  Find independent maximal set partitions in T_1 and T_2:
    $\mathcal{K}^* = \mathcal{K}_1^* \sqcup \ldots \sqcup \mathcal{K}_{r_1}^*$
    $\mathcal{M}^* = \mathcal{M}_1^* \sqcup \ldots \sqcup \mathcal{M}_{r_2}^*$
  for Every partition $\mathcal{M} = \bar{\mathcal{M}}_0 \sqcup \ldots \sqcup \bar{\mathcal{M}}_{r_1}$ and $\mathcal{K} = \bar{\mathcal{K}}_0 \sqcup \ldots \sqcup \bar{\mathcal{K}}_{r_2}$ **do**
    Regraft $\bar{\mathcal{M}}_i$ at node connecting $P^{\downarrow \mathcal{K}_i^*}$ to T_1, for all $i = 1, \ldots, r_1$. Creating $T_1^{\uparrow}$.
    Regraft $\bar{\mathcal{K}}_i$ at node connecting $P^{\downarrow \mathcal{M}_i^*}$ to T_2, for all $i = 1, \ldots, r_2$. Creating $T_2^{\uparrow}$.
    Compute initial trees involved in shortest path length in (4.6)
      $T_1^{\downarrow} = \psi(T_1^{\uparrow \perp \mathcal{L}^*}, \mathcal{L}^* \cup \bar{\mathcal{M}}_0)$ and $T_1^{\downarrow\downarrow} = \psi(T_1^{\downarrow \perp \bar{\mathcal{K}}_0}, \bar{\mathcal{K}}_0)$
      $T_2^{\downarrow} = \psi(T_2^{\uparrow \perp \mathcal{L}^*}, \mathcal{L}^* \cup \bar{\mathcal{K}}_0)$ and $T_2^{\downarrow\downarrow} = \psi(T_2^{\downarrow \perp \bar{\mathcal{M}}_0}, \bar{\mathcal{M}}_0)$
    Compute current shortest path length:
    $D = \sqrt{\left[||P^{\downarrow \mathcal{L}^*}(T_1^{\uparrow})|| + ||P^{\downarrow \mathcal{L}^*}(T_2^{\uparrow})||\right]^2 + ||P^{\downarrow \bar{\mathcal{K}}_0}(T_1^{\downarrow})||^2 + ||P^{\downarrow \bar{\mathcal{M}}_0}(T_2^{\downarrow})||^2 + d_{\mathrm{BHV}}^2(T_1^{\downarrow\downarrow}, T_2^{\downarrow\downarrow})}$
    while Value D is improving **do**
      for $i = 1, \ldots, r_1$ **do**
        Find optimal positions for regrafting $\bar{\mathcal{M}}_i$ by maximizing common edges between $T_1^{\downarrow\downarrow}$ and $T_2^{\downarrow\downarrow}$ while minimizing $||P^{\downarrow \mathcal{L}^*}(T_1^{\uparrow})||$ and $||P^{\downarrow \bar{\mathcal{K}}_0}(T_1^{\downarrow})||^2$
        Update trees $T_1^{\uparrow}$, $T_1^{\downarrow}$ and $T_1^{\downarrow\downarrow}$
        Update value D
      end for
      for $i = 1, \ldots, r_2$ **do**
        Find optimal positions for regrafting $\bar{\mathcal{K}}_i$ by maximizing common edges between $T_1^{\downarrow\downarrow}$ and $T_2^{\downarrow\downarrow}$ while minimizing $||P^{\downarrow \mathcal{L}^*}(T_2^{\uparrow})||$ and $||P^{\downarrow \bar{\mathcal{M}}_0}(T_2^{\downarrow})||^2$
        Update trees $T_2^{\uparrow}$, $T_2^{\downarrow}$ and $T_2^{\downarrow\downarrow}$
        Update value D
      end for
    end while
  end for
end for
return Return best value D found

as necessary for a fair comparison between tree, exploiting the properties of a quotient pseudometric. This results in the towering distance being a proper metric between trees, rather than just a measure of compatibility. This is a significant advancement, as a proper metric enables the use of formal statistical tools developed for metric spaces. This approach not only facilitates the potential to use Fréchet means for constructing super-trees, but also paves the way for developing inference tools such as confidence sets and hypothesis testing.

At the beginning of the chapter I introduced flexibility in the comparison for edge lengths when computing distances through the merging operation, which lead to a *family* of metric spaces. This allowed me to consider one towering space closely related to Extension Spaces (discussed in Chapter 3), by employing the L^1-norm as the merging operation. While this metric space has potential for phylogenetic analysis, and is supported by previous works in the field, I argue that another member of the family is a stronger candidate for future progress in phylogenetics due to its geometrical properties and interpretability. I consider the towering metric space (with the L^2-norm as the merging operation) to be a major contribution of this work, and likely to be highly useful for applied phylogenetics. As such, I briefly discuss its advantages.

4.7.1 On the Towering Space Interpretability

The BHV space can be viewed as a 0-cone of the link of the origin, where angle distances between trees on the link directly reflect differences in tree topologies. In the towering tree space, using β_2 as the merging operation ensures that trees in the same equivalence class remain equidistant from the origin tree. This emphasizes topological similarities as a key factor when computing distances, ensuring that trees with similar topologies, and potentially similar evolutionary histories, are positioned closer to each other in the metric space.

Furthermore, maintaining the equidistance of equivalent trees to the origin in their respec-

tive BHV spaces preserves the overall representation of evolutionary change. Consequently, the towering tree distance not only captures differences in topologies but also accurately reflects key variations in mutation rates between evolutionary events. This dual focus on topology and evolutionary rates provides a clear and intuitive framework for understanding phylogenetic relationships, thereby enhancing the utility of towering tree spaces in phylogenetic analysis.

4.7.2 On the Towering Space Geometric properties

The geometrical properties of the towering tree space closely resemble those of the BHV space. This similarity likely stems from the use of the L^2-norm as the merging operation. Since edge comparisons within each BHV space are conducted using the Euclidean distance, which is fundamentally based on the L^2-norm, this approach ensures consistency in the comparison of edge lengths. Even when the length of a single edge is compared to the result of combining several edges, the scale remains uniform. Consequently, adopting the L^2-norm for merging operations not only preserves meaningful biological interpretations but also facilitates the development of statistical tools for analyzing phylogenetic trees with differing leaf sets.

Throughout this chapter, I have demonstrated that the lengths of short paths can be expressed by a combination of L^2-norms of some edge lengths and BHV distances. This formulation results in paths that closely resemble geodesics in the BHV space. Specifically, edges that are common (under TDR mapping) between the two trees at the endpoints of the path gradually change in size from their initial length in one tree to their final length in the other. Meanwhile, uncommon edges — whether initially uncommon or made so by pre-specified prunings and regraftings — gradually reduce to zero and are replaced by other uncommon edges. A concrete example of this behavior is provided in Example 4.30,

illustrating that geometrically, the towering tree space shares significant parallels with the BHV space.

4.7.3 Future work: Distance algorithm improvement

In this work, I have presented a greedy algorithm for computing distances in the Towering Tree space. While this algorithm shows promise, there are several areas where it could be improved. Firstly, in Section 4.5.1, I outlined a general approach for determining optimal positions for regrafting uncommon leaves onto edges belonging to independent maximal sets. This approach is conjectured to yield the shortest possible path in the setting described in Theorem 4.34, when fixing which leaves are regrafted onto the edges belonging to each independent maximal set. However, I have not yet provided a proof that this algorithm indeed produces the optimal solution. Additionally, the current algorithm cycles through all possible ways to predefine which leaves are regrafted onto edges belonging to independent maximal sets. Given the results discussed in Section 4.5.1, which suggest that these predefined sets are only beneficial when potentially more common edges can be introduced between the trees, I believe there is a more systematic way to discard some of these sets beforehand. Finally, future work should focus on proving that the shortest path in the towering tree space between two trees is of the form described in Corollary 4.35. This would validate that Algorithm 3 effectively computes the distance.

Another promising direction for future research involves a deeper exploration of the parallels between constructing short paths in the towering tree space and computing geodesic lengths in the BHV space. In certain circumstances, such as those described in Theorem 4.29, it is theoretically possible to reclassify edges as "common" or "uncommon" after a predetermined selection of which leaves are to be pruned and regrafted to connect two trees (see Example 4.30). By developing a more precise mechanism to identify not only com-

mon and uncommon edges but also pairwise incompatible edges under this new pruning and regrafting-dependent classification, we could potentially create an algorithm similar to that presented by Owen and Provan (2011). In their work, an incompatibility graph is constructed where vertices represent uncommon edges between trees, and vertices representing incompatible edges are connected. A Min-Max algorithm is then used to determine which edges are swapped first based on their lengths, identifying the support pairs for the geodesic (see Section 3 of Owen and Provan (2011) for details). Exploring the implementation of an analogous method in the towering tree space could significantly enhance the accuracy and efficiency of distance computations in this space.

CHAPTER 5

DISCUSSION

5.1 Summary of contributions

The BHV tree space of Billera et al. (2001), with its intrinsic geodesic distance and interpretable geometric and topological properties, is a fundamental tool for analyzing collections of phylogenetic trees. Many methods for data analysis in this space exist, including fast algorithms for distance computation (Owen & Provan, 2011) and tree averaging (Benner et al., 2014; Brown & Owen, 2020; Miller et al., 2015); results on probabilistic uncertainty (Willis, 2019; Willis & Bell, 2018) and asymptotic behaviors (Barden et al., 2016; Nye, 2015); and advanced visualization tools (Nye, 2011; Teichman et al., 2023). Nevertheless, the application of BHV-based tools in the biological sciences has been limited. This can be partly attributed to BHV-based analyses requiring all trees to have identical leaf sets. As a result, researchers must choose between excluding trees from their analysis (those missing leaves), or excluding organisms from their analysis (those missing from one or more trees), neither of which is desirable. As noted by Ren et al. (2017, page 15), "supertree methods must take inputs with varying numbers of taxa to be useful in a biological context."

In this thesis, I address this key limitation of BHV space via two complementary approaches. My first contribution is to present an algorithm for measuring compatibility between any two trees with non-identical leaf sets within a common BHV space (Chapter 3). This work builds naturally on the contributions of Ren et al. (2017) and Grindstaff and Owen (2019). I then present a new family of BHV-like metric spaces for trees with non-identical leaf sets, focusing on one in particular: the Towering Tree Space (Chapter 4). Due to the similar geometric behavior of short paths in the BHV space and the Towering Space, I regard Towering Tree Space as a natural extension of BHV space, designed for the analysis of

trees with non-identical leaf sets. In this way, the Towering Space enables the development of practical tools analogous to those previously mentioned, but relevant to more general data-analytic settings.

Both extension spaces and towering tree spaces are built on the same principle as BHV space: distances between trees should be continuous, and based on both topological and branch length differences. In BHV space, the comparison of topologies relies on topological transformations, via nearest neighbor interchanges, while extension spaces and towering spaces incorporate additional operations such as pruning and regrafting. These additional topological operations introduce complications, as the length of an edge in one tree may need to be compared against the lengths of several edges in another tree because of leaf prunings. Extension spaces address this by summing the values of edges that merge after pruning. In contrast, towering spaces introduce flexibility via different methods to combine edge lengths after merging. Consequently, one of these towering spaces is closely related to extension spaces. However, I found that a different towering space, where edges are merged through L2-norms, results in a metric space with more favorable geometrical properties, as well as intuitive Euclidean-like distances.

5.2 Limitations

5.2.1 Scalability

A key consideration in any phylogenetic analysis is scalability, in part because the number of phylogenetic tree topologies grows exponentially in the number of leaves. As detailed in Section 3.4, my proposed algorithm for computing distances between extension spaces scales with the number of pairs of orthants in the BHV space, which grows super-exponentially. In practice, computing a distance between extension spaces in $\mathcal{T}^{\mathcal{L}}$ for $|\mathcal{L}| = 10$ can take up to an

hour on a modern laptop, though the runtime of a given analysis will depend on the overlap between the trees' leaf sets as well as the proximity of their extension spaces. While there are significant challenges in computing distances between extension spaces, my algorithm is the first method to compare trees with non-identical leaf sets using BHV distances, and thus addresses an important open problem in mathematical phylogenetics.

Although I did not explore the scalability of distances in the towering space, I anticipate that, in practice, its running time will be feasible. The algorithm involves iterating through various subsets of common leaves between the two trees, with each iteration's distance calculation relying on a combination of BHV distances (which is polynomial time) and L^2-norms (which is linear-time). Since introducing leaves not present in either of the two trees does not result in shorter paths (see Theorem 4.23), the paths connecting two trees are not expected to reach the top-dimensional BHV space. Consequently, the BHV distances computed in practice should be less computationally intensive.

I believe the algorithm for computing the towering tree distance between two trees will be more computationally efficient than the algorithm I proposed for computing the distance between extension spaces for the same two trees. Although both algorithms are exhaustive in nature, the algorithm for extension spaces must loop through all orthant pairs, whose number grows super-exponentially. In contrast, the towering tree distance algorithm only loops through all subsets of the common leaves, which grows exponentially. Additionally, each iteration in the extension spaces algorithm requires a gradient descent method, an iterative process that involves computing several BHV distances. Conversely, each iteration for different leaf subsets in the towering tree space requires significantly fewer BHV distance computations. As discussed in Section 4.7.3, there are numerous potential improvements for the towering tree space algorithm, which further underscores its computational advantages.

5.3 Future work

5.3.1 Improving scalability

As noted above, my algorithm for computing distances between extension spaces scales poorly with the number of leaves on each tree due to its need to compare all pairs of orthants. However, it is likely that the algorithm could be accelerated by developing criteria to exclude certain orthant pairs before applying the orthant-pair specific algorithm. Although I attempted to discard orthant pairs based on topological dissimilarities, the results varied depending on the dissimilarity of edge lengths. I observed that orthant pairs where new edges could be attached to particularly long edges tended to reduce the distance between the orthant-specific extension spaces, but no consistent pattern emerged. This would be a promising starting point for future investigations. I leave the development of these criteria to future work.

As discussed in Section 4.7.3, there is potential to improve the algorithm for computing distances in the Towering Tree space through two avenues. Firstly, refining the approach for regrafting uncommon leaves and developing a systematic way to discard unnecessary predefined sets of leaves to be regrafted onto edges belonging to independent maximal sets could enhance its efficiency. Secondly, exploring parallels with the BHV space and incorporating methods similar to those in Owen and Provan (2011, Section 3) could improve both the accuracy and efficiency of the algorithm.

5.3.2 Fréchet means

A practical problem in phylogenetics is the construction of a "consensus" (summary) tree based on the collection of gene trees. Fréchet means provide an interpretable summary of a collection of elements in a metric space, as they generalize the concept of a mean from

Euclidean spaces to metric spaces. The Fréchet mean is the minimizer of the expected squared distance under a probability distribution on a metric space, which could be the empirical distribution of trees (in the case of sample Fréchet means). Fréchet means have been extensively studied in the context of BHV tree space. Algorithms for computing sample Fréchet means have been developed (Benner et al., 2014; Miller et al., 2015), and results about their asymptotic properties have been established (Barden et al., 2016; Barden et al., 2013; Hotz et al., 2013). These results often rely on the geometric properties of spaces with non-positive curvature, such as the convexity of the distance function and the uniqueness of geodesics, which in turn ensure the uniqueness of Fréchet means. For a comprehensive overview, see Sturm (2003).

The Towering metric space is not a space with non-positive curvature, and shortest paths between trees are not necessarily unique. Thus, there is not guaranteed to be a single sample Fréchet mean, but rather a collection of sample Fréchet means. However, recent developments in asymptotic theory for Fréchet mean sets, as opposed to single means, provide a more general framework suitable for my metric space (Evans & Jaffe, 2024; Schötz, 2022). Inspired by this, I propose the use of sample Fréchet mean sets to summarize collections of trees in my metric space $(\mathbb{T}^{\mathcal{N}}, d)$. Since the towering metric space is a complete length space, continuous paths achieving the distance length exist, albeit without guarantees on their uniqueness. Using continuous paths and asymptotic results for Fréchet mean sets, I anticipate that it will be possible to develop algorithms to calculate Fréchet mean sets, similar to the approach provided in Bacák (2014, Section 3). I believe this is an is important open question to be explored alongside improvements to the algorithm discussed at the end of Chapter 4. Further work on these two topics will, in my opinion, enable the rigorous analysis of phylogenetic trees with non-identical leaf sets in general settings.

APPENDIX A

DISTANCES BETWEEN EXTENSION SPACES: ALGORITHM DETAILS

Here I provide additional details pertaining to Algorithm 1.

A.1 Reduced gradient directions

The direction of change will be based on the gradient of δ at the current point $\dot{\mathbf{x}}^t$ when all partial derivatives in (3.9) are well-defined, and a subgradient otherwise. To compute this (sub)gradient,I find the support for the geodesic from $T_1'(\dot{\mathbf{x}}^t)$ to $T_2'(\dot{\mathbf{x}}^t)$, and use its support to determine the entries of the (sub)gradient. I then focus on the reduced gradient within the current facet, which is $\nabla\varphi(\mathbf{x_F}) = \nabla_{\mathbf{F}} f(\mathbf{x}) - A_{\mathbf{F}}^{\top}\left[A_{\mathbf{D}}^{-1}\right]^{\top}\nabla_{\mathbf{D}} f(\mathbf{x})$ for the general case. Due to the structure of $\dot{\mathbf{M}}$, the partial derivative corresponding to a free variable $\dot{\mathbf{x}}_j^t$, $j \in \mathbf{F}$, is $\frac{\partial\varphi(\dot{\mathbf{x}}_{\mathbf{F}})}{\varphi\dot{\mathbf{x}}_j^t} = \nabla_j\delta(\dot{\mathbf{x}}) - \nabla_{j'}\delta(\dot{\mathbf{x}})$, where $j' \in \mathbf{D}$ is the only index in the dependent variable set such that $\dot{\mathbf{M}}[i,j] = \dot{\mathbf{M}}[i,j'] = 1$.

There are multiple ways to determine a good direction of change for the free variables at $\dot{\mathbf{x}}_{\mathbf{F}}^t$. I employed the conjugate gradient method Ruszczyński, 2006, Section 5.5 because of its simplicity and the potential gain in efficiency. In this method, the main driver for the direction of change is the (sub)gradient at the current point, but after the first iteration a correction is added to increase efficiency. The correction loses its advantages and new directions become inefficient after many iterations Ruszczyński, 2006, Section 5.5.2, thus I re-initialize the correction regularly. The direction of change $\mathbf{d}_{\mathbf{F}}^t$ for the free variables is computed by

$$\mathbf{d}_{\mathbf{F}}^t = \begin{cases} -\nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^t) & \text{if } c_{\text{conj}} = 1 \\ -\nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^t) + \frac{\langle\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^t),\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^t)-\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^{t-1})\rangle}{\left|\left|\varphi(\dot{\mathbf{x}}_{\mathbf{F}}^{t-1})\right|\right|^2}\mathbf{d}_{\mathbf{F}}^{t-1} & \text{if } c_{\text{conj}} > 1, \end{cases} \tag{A.1}$$

for a counter c_{conj}. In my method, c_{conj} is re-initialized (reset to 1) every time a new facet is reached (i.e. $\mathbf{F}$ is re-defined and $\mathbf{d}_{\mathbf{F}}^{t-1}$ is no longer of the same dimension) or when $c_{\text{conj}} > 15$. The threshold of 15 was recommended by Ruszczyński, 2006, Page 248, and I found it to work well in practice. Given the direction of change for the free variables $\mathbf{d}_{\mathbf{F}}^t$, I can find the direction of change for the dependent variables as $\mathbf{d}_{\mathbf{D}}^t = -\dot{\mathbf{M}}_{\mathbf{D}}^{-1}\dot{\mathbf{M}}_{\mathbf{F}}\mathbf{d}_{\mathbf{F}}^t$ and for the null variables as $\mathbf{d}_{\mathbf{N}}^t = 0$.

A.2 Selecting step sizes

Given a non-zero direction of change $\mathbf{d}^t$, I now discuss finding the best next point on the line segment $\dot{\mathbf{x}}^t + \tau\mathbf{d}^t$ such that both $\tau \geq 0$ and $\dot{\mathbf{x}}^t + \tau\mathbf{d}^t \geq 0$. Let $\tau_{\max}$ be the maximum value of τ such that $\dot{\mathbf{x}}^t + \tau\mathbf{d}^t \geq 0$, which is finite because $\dot{\mathbf{M}}\left(\dot{\mathbf{x}}^t + \tau\mathbf{d}^t\right) = \dot{\mathbf{v}}$ implies $\dot{\mathbf{M}}\mathbf{d}^t = 0$, and thus some entry of $\mathbf{d}^t$ is negative. Thus, I am looking for $\tau \in [0, \tau_{\max}]$ that minimizes $\delta(\dot{\mathbf{x}}^t + \tau\mathbf{d}^t)$.

Using the chain-rule, the derivative of $\delta(\dot{\mathbf{x}}^t + \tau\mathbf{d}^t)$ with respect to τ is

$$h(\tau) = \frac{\partial\delta(\dot{\mathbf{x}}^t + \tau\mathbf{d}^t)}{\partial\tau} = \langle\nabla\delta(\dot{\mathbf{x}}^t + \tau\mathbf{d}^t), \mathbf{d}^t\rangle. \tag{A.2}$$

The function $\tau \mapsto \delta(\dot{\mathbf{x}}^t + \tau\mathbf{d}^t)$ is convex and by construction of $\mathbf{d}^t$, $h(0) < 0$. From this, I employ a derivative-based bisection method to reach the minimum of this function. I start by checking the value of $h(\tau_{\max})$. If $h(\tau_{\max}) \leq 0$, then the minimum has been reached at $\tau_0 = \tau_{\max}$. Otherwise, I search for the value $\tau_0 \in [0, \tau_{\max}]$ such that $h(\tau_0) = 0$ as follows

1. Initialize $\tau_{\text{left}} = 0$ and $\tau_{\text{right}} = \tau_{\max}$.
2. Take $\tau^* = \frac{\tau_{\text{left}} + \tau_{\max}}{2}$.
3. Evaluate $h(\tau^*)$:
   - If $h(\tau^*) = 0$, return τ^*.

- If $h(\tau^*) < 0$, update $\tau_{\text{left}} = \tau^*$ and return to step 2.
- If $h(\tau^*) > 0$, update $\tau_{\text{right}} = \tau^*$ and return to step 2.

In practice, I do not require $h(\tau^*) = 0$ exactly; instead, I require $|h(\tau^*)|$ to be below a threshold. I find that $|h(\tau^*)| < 10^{-16}$ works well.

This threshold can be altered in my software via the flag `-Tol2`. After finding τ_0, I select the next point as $\dot{\mathbf{x}}^{t+1} = \dot{\mathbf{x}}^t + \tau_0 \mathbf{d}^t$.

A.3 Thresholds for convergence

My iterative algorithm is guaranteed to converge to stationary points with zero (sub)gradients. However, in practice, I employ a tolerance threshold to find a solution that is sufficiently close to a stationary point in a finite number of iterations. Specifically, before computing a new direction of change, I test if the global minimum has been reached by checking if every entry in $\nabla\varphi(\dot{\mathbf{x}}_{\mathbf{F}})$ is less than a given threshold `Tol1`. In practice, I find that choosing this threshold to be 10^{-8} produces good performance. The user can alter this value using the flag `-Tol1`.

APPENDIX B

TOWERING TREE SPACES: PROOFS FOR MAIN RESULTS

This appendix contains proofs for lemmas and theorems presented in Chapter 4.

B.1 Proofs for Section 4.1: Preliminary Structures

Proof of Lemma 4.6. (Closure) Consider $T \in \mathcal{T}^{\mathcal{L}}$ such that $T \notin \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$. Since $\mathcal{M}$ is not mutually prunable from T then either some external edge e to a leaf in $\mathcal{M}$ has a positive length, or there is an internal split s such that $\Psi_{\mathcal{L}\setminus\mathcal{M}}(s)$ is not an internal split on the leaves $\mathcal{L}\setminus\mathcal{M}$. In the first case, given a value $\epsilon < |e|_T$, is enough to guarantee $|e|_{T'} > 0$ for any tree T' such that $d_{\mathrm{BHV}}(T,T') < \epsilon$, which implies $T' \notin \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$. In the second case, if $\epsilon < |s|_T$ then similarly $|s|_{T'} > 0$ (so that $s \in \mathcal{S}(T')$) for any tree T' such that $d_{\mathrm{BHV}}(T,T') < \epsilon$, implying $T' \notin \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$. Thus, there is always a neighborhood of T not intersecting $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$. Therefore the subspace is closed.

(Convexity) Consider two trees $T_1, T_2 \in \mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$ and T on the geodesic from T_1 to T_2. Along the geodesic, the lengths of external edges change gradually from the lengths they have in T_1 to the lengths they have in T_2. Since all external edges to leaves in $\mathcal{M}$ are of length zero in both trees, then these external edges are of length zero as well in T. Given any internal split s of the leaves $\mathcal{L}$ such that $\Psi_{\mathcal{L}\setminus\mathcal{M}}(s)$ does not map to an internal split, s is not a part of the internal edges of T_1 nor the internal edges of T_2, so it is not part of $\mathcal{S}(T)$ as well. Thus, $\mathcal{M}$ is also mutually prunable from T, and $\mathcal{Z}^{\mathcal{L}}_{\mathcal{M}}$ is convex. □

Proof of Lemma 4.15. Since $\mathcal{M}$ is mutually prunable from T, by Lemma 4.10, $\mathcal{M}_1$ is as well, and thus $\psi_\beta(T,\mathcal{M}_1)$ is well-defined. For this proof take $\mathcal{L}' = \mathcal{L}\setminus\mathcal{M}_1$, $\mathcal{L}'' = \mathcal{L}\setminus\mathcal{M} = \mathcal{L}'\setminus\mathcal{M}_2$ and call $T_1 = \psi_\beta(T,\mathcal{M}_1)$, $T_2 = \psi_\beta(\psi_\beta(T,\mathcal{M}_1),\mathcal{M}_2)$ and $T' = \psi_\beta(T,\mathcal{M})$. Given that $|e|_{T_1} = |e|_T$ for any external edge e corresponding to a leaf in $\mathcal{L}'$, and $|e|_T = 0$ when e is

the external edge to a leaf in $\mathcal{M}$, it follows that $\mathcal{M}_2 \subset \mathcal{M}$ implies $|e|_{T_1} = 0$ for all external edges to leaves in $\mathcal{M}_2$. For internal splits, note that for any partition $\mathcal{L} = \mathcal{L}_1 \sqcup \mathcal{L}_2$ it is true that

$$(\mathcal{L}_i \setminus \mathcal{M}_1) \setminus \mathcal{M}_2 = \mathcal{L}_i \setminus (\mathcal{M}_1 \cup \mathcal{M}_2) = \mathcal{L}_i \setminus \mathcal{M},$$

for $i = 1, 2$. So an internal edge $q \in \mathcal{S}(T)$ maps to $p' \in \mathcal{S}(T')$ under the TDR map ($\Psi_{\mathcal{L}''}(q) = p'$) if and only if there is a split $p \in \mathcal{S}(T_1)$ such that q maps to p and then p maps to p' ($\Psi_{\mathcal{L}'}(q) = p$ and $\Psi_{\mathcal{L}''}(p) = p'$). For any internal edge $p \in \mathcal{S}(T_1)$, it follows that $\Psi_{\mathcal{L}''(p)} = \Psi_{\mathcal{L}''}(q)$ for any edge $q \in \mathcal{S}(T)$ where $\Psi_{\mathcal{L}'}(q) = p$ (by the definition of β-prunings, there must be at least one such q for each p). Since $\mathcal{M}$ is mutually prunable from T, $\Psi_{\mathcal{L}''(p)}$ is an internal split, implying $\mathcal{M}_2$ is mutually prunable from T_1. Moreover, this implies $\Psi_{\mathcal{L}''}(\mathcal{S}(T)) = \Psi_{\mathcal{L}''}(\Psi_{\mathcal{L}'}(\mathcal{S}(T)))$, so the topology of T coincides with the topology of T_2. Finally, we show that not only the topologies but also the edge lengths coincide. For any external edge e to a leaf in $\mathcal{L}''$, $|e|_{T_2} = |e|_{T_1}$ and $|e|_{T_1} = |e|_T$, so $|e|_{T_2} = |e|_{T'}$. And for an internal split $s \in \Psi_{\mathcal{L}''}(\mathcal{S}(T))$,

$$|s|_{T_2} = \underset{p \in \Psi^{-1}_{\mathcal{L}''}(s)\big|_{\mathcal{S}(T_1)}}{\mathbf{B}} |p|_{T_1} = \underset{p \in \Psi^{-1}_{\mathcal{L}''}(s)\big|_{\mathcal{S}(T_1)}}{\mathbf{B}} \left[\underset{q \in \Psi^{-1}_{\mathcal{L}'}(p)\big|_{\mathcal{S}(T)}}{\mathbf{B}} |q|_T \right] = \underset{p \in \Psi^{-1}_{\mathcal{L}''}(s)\big|_{\mathcal{S}(T)}}{\mathbf{B}} |p|_T = |s|_{T'}$$

□

B.2 Proofs for Section 4.3: Definition and Preliminary Results of the Towering Tree Space

Proof of Theorem 4.23. Consider first the case where the trees in the higher level are part of the same orthant; i.e., $\mathcal{O}(T_1^{\uparrow}), \mathcal{O}(T_2^{\uparrow}) \subseteq O$. Their geodesic in this case is a segment completely contained in O and its length can be expressed in terms of the edges in $\mathcal{P}(O)$ by $d_{\mathrm{BHV}}(T_1^{\uparrow}, T_2^{\uparrow}) = \sqrt{\sum_{q \in \mathcal{P}(O)} \left(|q|_{T_1^{\uparrow}} - |q|_{T_2^{\uparrow}} \right)^2}$. Since $\mathcal{O}(T_1), \mathcal{O}(T_2) \subseteq O' = \Psi_{\mathcal{L}'}(O)$, for each

edge $p \in \mathcal{P}(O')$, consider the set of edges $\Psi^{-1}_{\mathcal{L}'}(q)\big|_O$ in O that map to p under the TDR map. By Lemma 4.22,

$$\sum_{q \in \Psi^{-1}_{\mathcal{L}'}(p)|_O} \left(|q|_{T_1^\uparrow} - |q|_{T_2^\uparrow}\right)^2 \geq (|p|_{T_1} - |p|_{T_2})^2.$$

and therefore

$$d_{\text{BHV}}(T_1^\uparrow, T_2^\uparrow) = \sqrt{\sum_{p \in \mathcal{P}(O^\downarrow)} \sum_{q \in \Psi^{-1}(p)|_O} \left(|q|_{T_1^\uparrow} - |q|_{T_2^\uparrow}\right)^2} \geq \sqrt{\sum_{p \in \mathcal{P}(O^\downarrow)} (|p|_{T_1} - |p|_{T_2})^2} = d_{\text{BHV}}(T_1, T_2).$$

For the more general case where $T_1^\uparrow$ and $T_2^\uparrow$ are not part of the same topology orthant, consider the geodesic $\gamma : [0,1] \mapsto \mathcal{T}^{\mathcal{L}}$ from $T_1^\uparrow$ to $T_2^\uparrow$. Since $\mathcal{M} = \mathcal{L} \setminus \mathcal{L}'$ is mutually prunable from $T_1^\uparrow$ and $T_2^\uparrow$, then it is mutually prunable for any tree $T_\lambda^\uparrow = \gamma(\lambda)$ on the geodesic (Lemma 4.6), and it belongs to $\Lambda^{\mathcal{L}}(T_\lambda)$ for some tree $T_\lambda \in \mathcal{T}^{\mathcal{L}'}$.

Take the path space for the geodesic $O_0 \to O_1 \to \ldots \to O_k$, and consider $t_i^\uparrow$ the transition tree from O_{i-1} to O_i along γ. For simplicity, consider $t_0^\uparrow = T_1^\uparrow$ and $t_{k+1}^\uparrow = T_2^\uparrow$. For each $t_i^\uparrow$ there is a $t_i \in \mathcal{T}^{\mathcal{L}'}$ such that $t_i^\uparrow \in \Lambda^{\mathcal{L}}(t_i)$. Moreover, from construction $t_{i-1}^\uparrow$ and $t_i^\uparrow$ are part of the same orthant, and so are t_{i-1} and t_i. Thus,

$$d_{\text{BHV}}(T_1^\uparrow, T_2^\uparrow) = \sum_{i=1}^{k+1} d_{\text{BHV}}(t_{i-1}^\uparrow, t_i^\uparrow) \geq \sum_{i=1}^{k+1} d_{\text{BHV}}(t_{i-1}, t_i) \geq d_{\text{BHV}}(T_1, T_2).$$

□

B.3 Proofs for Section 4.4: Short paths through lower BHV spaces

Proof of Theorem 4.26. (i) By Corollary 4.7 $d^2_{\text{BHV}}(T_1, t) \geq d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}}) + d^2_{\text{BHV}}(T_1^{\perp\mathcal{M}}, t)$ for any tree $t \in \Lambda^{\mathcal{L}}(T_2)$. Since $T_1^{\perp\mathcal{M}} \in \Lambda^{\mathcal{L}}(T_1')$ as well, $d^2_{\text{BHV}}(T_1^{\perp\mathcal{M}}, t) \geq d^2_{\text{BHV}}(T_1', T_2)$ by Theorem 4.23. Therefore,

$$d^2_{\text{BHV}}(T_1, t) \geq d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}}) + d^2_{\text{BHV}}(T_1', T_2).$$

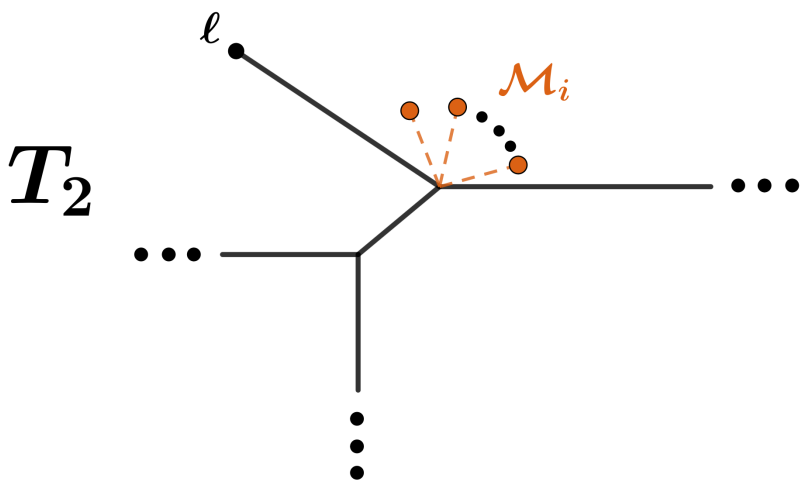


Figure B.1. Regrafting of an independent maximal set neighboring a leaf $\ell \in \mathcal{L}'$. A section of tree T_2 in $\mathcal{T}^{\mathcal{L}'}$ is showed in black. The new leaves $\mathcal{M}_i$ (in orange) form an independent maximal set neighboring ℓ in another tree T_1, so they are regrafted to be incident to the same node as the external edge to ℓ. The external edges of the new leaves are represented with dashed lines, indicating they are of length zero.

(ii) To construct the optimal tree $T^* \in \Lambda_\beta^{\mathcal{L}}(T_2)$, the positions where leaves in $\mathcal{M}$ are regrafted on T_2 are based on the partition of these leaves into independent maximal sets $\mathcal{M} = \mathcal{M}_1 \cup \ldots \cup \mathcal{M}_r$ in T_1 and the geodesic from T_1' to T_2. The process is as follows:

1. For every leaf $\ell \in \mathcal{L}'$ for which there is a neighboring independent maximal set $\mathcal{M}_i$, regraft these at the same node in T_2 to which the external edge to ℓ is incident (see Figure B.1).

2. For a common edge $p = [\mathcal{L}_1' \ddagger \mathcal{L}_2'] \in \mathcal{S}(T_1') \cap \mathcal{S}(T_2)$, consider all internal edges of $T_1^\perp$ that map to p under the TDR map $\Psi_{\mathcal{L}'}$. Consider a new partition of $\mathcal{M}$ given by $\mathcal{M} = \widehat{\mathcal{M}}_1^p \sqcup \mathcal{M}_1^p \sqcup \ldots \sqcup \mathcal{M}_{r_p}^p \sqcup \widehat{\mathcal{M}}_2^p$, where $\mathcal{M}_1^p, \ldots, \mathcal{M}_{r_p}^p$ are all the independent maximal sets that could be considered attached to p in T_1 (including attached to the endpoint nodes of the edge) and the two sets $\widehat{\mathcal{M}}_i^p$ are the rest of the leaves in $\mathcal{M}$ that are on the same side of p as $\mathcal{L}_i'$ in T_1; i.e. every edge in $T_1^\perp$ that maps to p under the TDR map

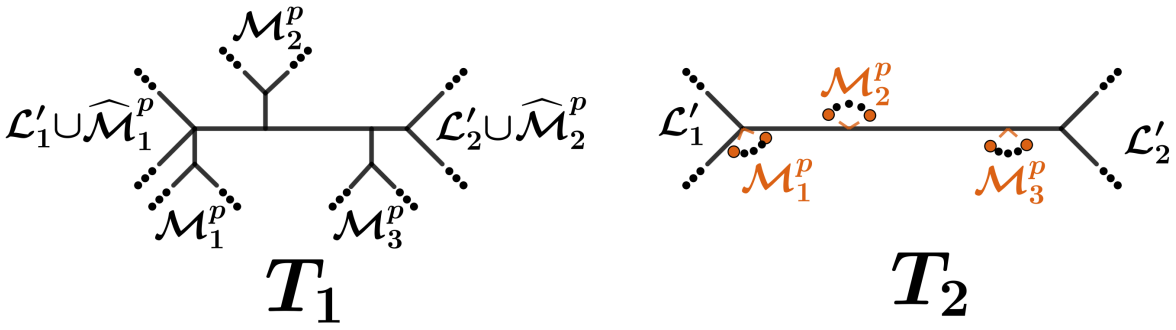


Figure B.2. Regrafting of independent maximal sets attached on a common edge $p \in \mathcal{S}(T_1') \cap \mathcal{S}(T_2)$. The section of tree T_1 containing the edges mapping to p under the TDR map is shown in the left, featuring three independent maximal sets attached to these edges. The common edge p in T_2 is depicted on the right, with the new leaves in the three independent maximal sets regrafted (orange) in the correct position (same position proportional wise as their counterparts in T_1), with dashed lines indicating the external edges are of length zero.

is of the form $q_j^p = \left[\mathcal{L}_1' \cup \widehat{\mathcal{M}}_1^p \cup \mathcal{M}_1^p \cup \ldots, \mathcal{M}_j^p \ddagger \mathcal{M}_{j+1}^p \cup \ldots, \mathcal{M}_{r_p}^p \cup \widehat{\mathcal{M}}_2^p \cup \mathcal{L}_2'\right]$. When regrafting $\mathcal{M}$ onto T_2 to create T^*, leaves in $\widehat{\mathcal{M}}_i^p$ are regrafted on the same side of the edge p, and each $\mathcal{M}_j^p$ is regrafted at the same position (proportional wise) inside the edge p in T_2. Thus, every q_j^p as given above will be part of the interior edges $\mathcal{S}(T^*)$ of T^*, with the lengths given by $\left|q_j^p\right|_{T^*} = \frac{\left|q_j^p\right|_{T_1^\perp}}{|p|_{T_1'}} |p|_{T_2}$ (see Figure B.2).

3. For any independent maximal set $\mathcal{M}_i$ of leaves that do not fall in one of the two previous scenarios, the split $[\mathcal{M}_i \ddagger \mathcal{L} \setminus \mathcal{M}_i]$ is adjacent to at least one edge $q_i \in \mathcal{S}(T_1)$ that under $\Psi_{\mathcal{L}'}$ maps to an internal edge $p_i \in \mathcal{S}(T_1')$ that is uncommon with edges in $\mathcal{S}(T_2)$. Given a support for the path space of the geodesic from T_1' to T_2, there must be a support pair (A, B) for which $p_i \in A$ and exists $p_i' \in B$ that is the product of a nearest neighbor interchange from the node resulting from collapsing p_i to zero. Regraft the external edges to leaves in $\mathcal{M}_i$ on the center of p_i' (see Figure B.3).

The new tree T^* created through these regrafts contains all leaves in $\mathcal{M}$ and is part of the sprouting space $\Lambda^{\mathcal{L}}(T_2)$. Moreover, the BHV distance from T_1 to T^* will precisely

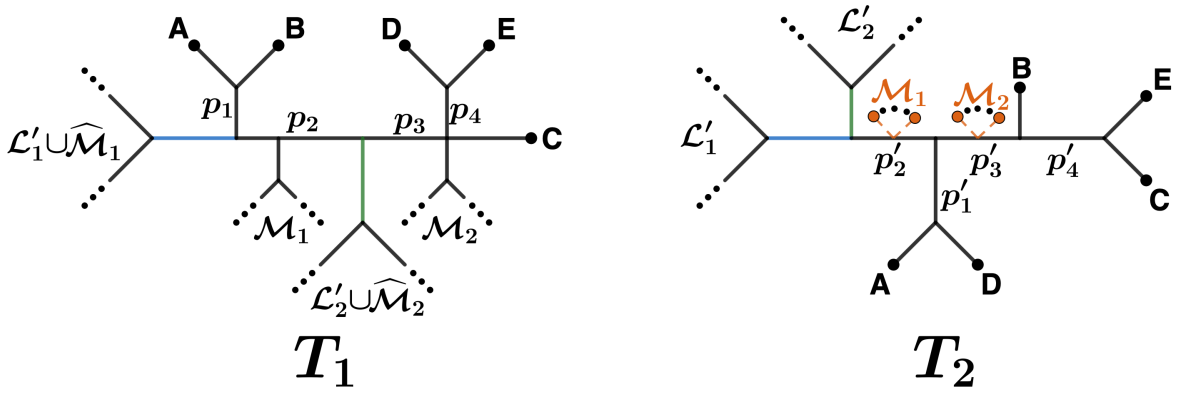


Figure B.3. Example of regrafting independent maximal sets attached to uncommon edges. Sections in trees $T_1 \in \mathcal{T}^{\mathcal{L}}$ (left) and $T_2 \in \mathcal{T}^{\mathcal{L}'}$ (right) where leaves $A, B, C, D, E \in \mathcal{L}'$ are located. After pruning $\mathcal{M}$ from $T_1^{\perp}$, the edges $[\{A, B, C, D, E\} \cup \mathcal{L}'_2 \ddagger \mathcal{L}'_1]$ (in blue) and $[\{A, B, C, D, E\} \cup \mathcal{L}'_1 \ddagger \mathcal{L}'_2]$ (in green) are common, while the edges $p_1, p_2, p_3, p_4 \in \mathcal{S}(T'_1)$ and $p'_1, p'_2, p'_3, p'_4 \in \mathcal{S}(T_2)$ are uncommon. In T_1, the independent maximal set $\mathcal{M}_1$ is adjacent to two edges mapping to p_2 under the TDR map, and $\mathcal{M}_2$ is adjacent to edges that map to p_3 and p_4. Assuming the support for the geodesic from T'_1 to T_2 includes the support pairs $(\{p_2\}, \{p'_2\})$, $(\{p_4\}, \{p'_4\})$, and $(\{p_1, p_3\}, \{p'_1, p'_3\})$, p'_2 can be selected to regraft $\mathcal{M}_1$ and p'_3 to regraft $\mathcal{M}_2$. These leaves are represented in orange in the correct position for the regraft, with dashed lines indicating the external edges are of length zero.

be $\sqrt{d^2_{\text{BHV}}(T_1, T_1^{\perp \mathcal{M}}) + d^2_{\text{BHV}}(T'_1, T_2)}$. To see this, consider again the edges $P^{\downarrow \mathcal{M}}(T_1)$. If an edge is of the form $[\mathcal{M}' \ddagger \mathcal{L} \setminus \mathcal{M}']$, by definition $\mathcal{M}' \subseteq \mathcal{M}_i$ for some $i = 1, \dots r$, and since all leaves in the maximal independent sets are regrafted together when constructing T^*, then $[\mathcal{M}' \ddagger \mathcal{L} \setminus \mathcal{M}']$ is pairwise compatible with $\mathcal{S}(T^*)$. Similarly, for an edge $[\mathcal{M}' \cup \{\ell\} \ddagger \mathcal{L} \setminus (\mathcal{M} \cup \{\ell\})] \in P^{\downarrow \mathcal{M}}(T_1)$, by definition $\mathcal{M}' \subseteq \mathcal{M}_i$ for the independent maximal set $\mathcal{M}_i$ neighboring ℓ, and since in $\mathcal{T}^*$ all external edges to leaves in $\mathcal{M}_i$ are adjacent to external edge to ℓ, then $[\mathcal{M}' \cup \{\ell\} \ddagger \mathcal{L} \setminus (\mathcal{M}' \cup \{\ell\})]$ is compatible with every edge in T^*. Since $P^{\downarrow \mathcal{M}}(T_1)$ is pairwise compatible with $\mathcal{S}(T^*)$, their size will gradually change from their size in T_1 to zero along the geodesic. Thus, the contribution of these edges to the geodesic length is $||P^{\downarrow \mathcal{M}}(T_1)|| = d_{\text{BHV}}(T_1, T_1^{\perp \mathcal{M}})$.

We now focus on the remaining internal edges $(\mathcal{S}(T_1) \cup \mathcal{S}(T^*) \setminus P^{\downarrow \mathcal{M}}(T_1))$ and their respective contribution to the length of the geodesic. Of note is that all internal edges in

$\mathcal{S}(T_1)$ that are not in $P^{\downarrow\mathcal{M}}(T_1)$ belong to $\mathcal{S}(T_1^{\perp})$ as well, and with the same lengths. Step 2 of the above construction of T^* guarantees $q \in \mathcal{S}(T_1^{\perp}) \cap \mathcal{S}(T^*)$ if and only if $\Psi_{\mathcal{L}'}(q) \in \mathcal{S}(T_1') \cap \mathcal{S}(T_2)$. The contribution to the length of the geodesic of all common edges $q_1^p, ..., q_r^p$ in $\mathcal{S}(T_1^{\perp}) \cap \mathcal{S}(T^*)$ that map into a common edge p is given by

$$\begin{aligned}\sum_{k=1}^{r}(|q_k^p|_{T_1^{\perp}} - |q_k^p|_{T^*})^2 &= \sum_{k=1}^{r}\left(|q_k^p|_{T_1^{\perp}} - \frac{|q_k^p|_{T_1^{\perp}}}{|p|_{T_1'}}|p|_{T_2}\right)^2 \\ &= \left(1 - \frac{|p|_{T_2}}{|p|_{T_1^{\perp}}}\right)^2 \sum_{k=1}^{r}|q_k^p|_{T_1^{\perp}}^2 \\ &= \left(|p|_{T_1'} - |p|_{T_2}\right)^2.\end{aligned}$$

So the contribution of all these edges in the length of the geodesic from T_1 to T^* is the same as the contribution of p in the length of the geodesic from T_1' to T_2.

Finally, consider the contribution of uncommon edges between T_1 and T^*. If the support of the geodesic from T_1' to T_2 is $\mathcal{A} = \{A_1, \ldots, A_k\}$ and $\mathcal{B} = \{B_1, \ldots, B_k\}$. For each A_i, consider the set of edges in $T_1^{\perp}$ that map to an edge in A_i, $\Psi_{\mathcal{L}'}^{-1}(A_i)\big|_{\mathcal{S}(T_1^{\perp})}$, and similarly, the set of edges in T^* that map to edges in B_i, $\Psi_{\mathcal{L}'}^{-1}(B_i)\big|_{\mathcal{S}(T^*)}$. The positions for the independent maximal sets in Step 3 are selected to ensure that $\left(\Psi_{\mathcal{L}'}^{-1}(A_1)\big|_{\mathcal{S}(T_1^{\perp})}, \ldots, \Psi_{\mathcal{L}'}^{-1}(A_k)\big|_{\mathcal{S}(T_1^{\perp})}\right)$ and $\left(\Psi_{\mathcal{L}'}^{-1}(B_1)\big|_{\mathcal{S}(T^*)}, \ldots, \Psi_{\mathcal{L}'}^{-1}(B_k)\big|_{\mathcal{S}(T^*)}\right)$ form a valid support for a path space from T_1 to T^*, satisfying condition (P1) from Section 2.2.2. Since $||A_i||_{T_1'} = ||\ \Psi_{\mathcal{L}'}^{-1}(A_i)\big|_{\mathcal{S}(T_1^{\perp})}\ ||_{T_1^{\perp}}$ and $||B_i||_{T_2} = ||\ \Psi_{\mathcal{L}'}^{-1}(B_i)\big|_{\mathcal{S}(T^*)}\ ||_{T^*}$ the condition (P2) is also satisfied. Consequently, the contribution of the support pair $\left(\Psi_{\mathcal{L}'}^{-1}(A_i)\big|_{\mathcal{S}(T_1^{\perp})}, \Psi_{\mathcal{L}'}^{-1}(B_i)\big|_{\mathcal{S}(T^*)}\right)$ to the length of the geodesic from T_1 to T^* is the same as the contribution of the edges (A_i, B_i) to the length of the geodesic from T_1' to T_2. Thus, the internal edges in $\mathcal{S}(T_1) \cup \mathcal{S}(T^*) \setminus P^{\downarrow\mathcal{M}}(T_1)$ contribute to the length of the geodesic from T_1 to T^* an amount exactly equal to that the internal edges in $\mathcal{S}(T_1') \cup \mathcal{S}(T_2)$'s contribution to the geodesic from T_1' to T_2. The result follows.

□

Proof of Theorem 4.27. Consider a general path from T_1 to T_2 where the pruning of $\mathcal{M}$

occurs in the sprouting space $\Lambda^{\mathcal{L}}(X)$ of a tree $X \in \mathcal{T}^{\mathcal{L}'}$. By Theorem 4.26, the shortest the section of the path from T_1 to X can be is $\sqrt{d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}}) + d^2_{\text{BHV}}(T_1', X)}$, where $T_1' = \psi(T_1^{\perp\mathcal{M}}, \mathcal{M})$. And the shortest path from X to T_2 completely contained in $\mathcal{T}^{\mathcal{L}'}$ is simply the geodesic with length $d_{\text{BHV}}(X, T_2)$. On the other hand, again by Theorem 4.26, there is a path from T_1 to T_2 where the pruning of $\mathcal{M}$ is at a tree in $\Lambda^{\mathcal{L}}(T_2)$ and of length $\sqrt{d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}}) + d^2_{\text{BHV}}(T_1', T_2)}$.

By the triangle inequality, $d_{\text{BHV}}(T_1', X) + d_{\text{BHV}}(X, T_2) \geq d_{\text{BHV}}(T_1', T_2)$, and thus

$$d^2_{\text{BHV}}(T_1', X) + 2d_{\text{BHV}}(T_1', X)d_{\text{BHV}}(X, T_2) + d^2_{\text{BHV}}(X, T_2) \geq d^2_{\text{BHV}}(T_1', T_2).$$

It is clear that

$$\sqrt{d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}}) + d^2_{\text{BHV}}(T_1', X)} \geq d_{\text{BHV}}(T_1', X)$$

so:

$$d^2_{\text{BHV}}(T_1', X) + 2d_{\text{BHV}}(X, T_2)\sqrt{d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}}) + d^2_{\text{BHV}}(T_1', X)} + d^2_{\text{BHV}}(X, T_2) \geq d^2_{\text{BHV}}(T_1', T_2).$$

Adding $d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}})$ to both sides, we have that

$$\sqrt{d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}}) + d^2_{\text{BHV}}(T_1', X)} + d_{\text{BHV}}(X, T_2) \geq \sqrt{d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}}) + d^2_{\text{BHV}}(T_1', T_2)}.$$

□

Proof of Theorem 4.29. Any such path will pass through a tree $X \in \mathcal{T}^{\mathcal{L}'}$. Lemmas 4.27 and 4.28 imply (for each $i = 1, 2$) the shortest paths from T_i to X going through exclusively leaf prunings are given by performing the pruning of $\mathcal{M}_i$ at trees in the sprouting space $\Lambda^{\mathcal{L}_i}(X)$, and the length of these are $\sqrt{d^2_{\text{BHV}}(T_i, T_i^{\perp\mathcal{M}_i}) + d^2_{\text{BHV}}(T_i', X)}$. Thus, the shortest paths from T_1 to T_2, going strictly downwards, passing through X and then going strictly upwards to T_2 are of length

$$\sqrt{d^2_{\text{BHV}}(T_1, T_1^{\perp\mathcal{M}_1}) + d^2_{\text{BHV}}(T_1', X)} + \sqrt{d^2_{\text{BHV}}(X, T_2') + d^2_{\text{BHV}}(T_2^{\perp\mathcal{M}_2}, T_2)}.$$

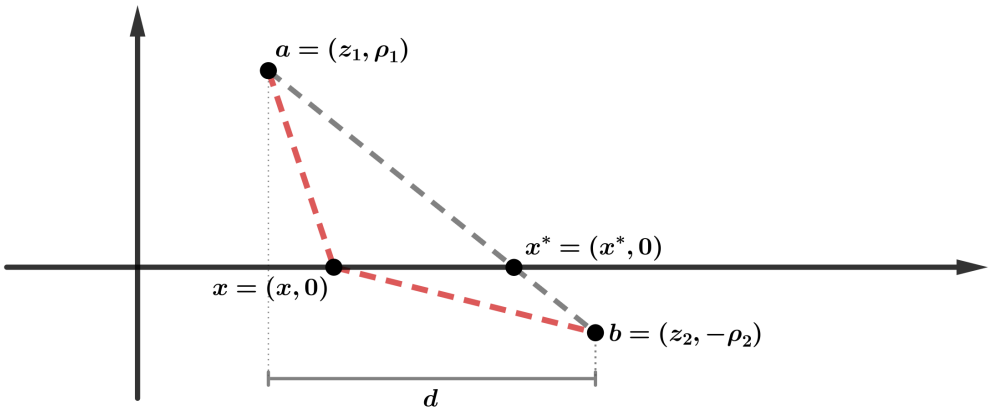


Figure B.4. Representation of minimizing function $f(x) = \sqrt{\rho_1^2 + (z_1 - x)^2} + \sqrt{\rho_2^2 + (z_2 - x)^2}$. The length of the path from $\mathbf{a} = (z_1, \rho_1)$ to $\mathbf{b} = (z_2, \rho_2)$ passing through $(0, x)$ (dashed red) coincides with $f(x)$, and any such path is longer than the direct segment (dashed gray).

The values $d^2_{\text{BHV}}(T_1, T_1^{\perp \mathcal{M}_1})$ and $d^2_{\text{BHV}}(T_2^{\perp \mathcal{M}_2}, T_2)$ are independent from the choice of $X \in \mathcal{T}^{\mathcal{L}'}$. For simplicity, denote $\rho_i = d_{\text{BHV}}(T_i, T_i^{\perp \mathcal{M}_i})$. Since T_1', X and T_2' are all in the same BHV space, $d_{\text{BHV}}(T_1', X) + d_{\text{BHV}}(X, T_2') \geq d_{\text{BHV}}(T_1', T_2')$. Let $y_1 = d_{\text{BHV}}(T_1', X)$, $y_2 = d_{\text{BHV}}(X, T_2')$ and $d = d_{\text{BHV}}(T_1', T_2')$. The problem of finding X that produces the shortest path between T_1 and T_2 is now equivalent to finding values y_1 and y_2 that minimize $\sqrt{\rho_1^2 + y_1^2} + \sqrt{\rho_2^2 + y_2^2}$ subject to $y_1 + y_2 \geq d$. Consider a geometrical approach by considering points $\mathbf{a} = (z_1, \rho_1)$ and $\mathbf{b} = (z_2, -\rho_2)$ in $\mathbb{R}^2$, where $|z_1 - z_2| = d$. Consider the piecewise linear path from $\mathbf{a}$ to $\mathbf{b}$ crossing the x-axis at a point $\mathbf{x} = (x, 0)$. The length of this path is given by $f(x) = \sqrt{\rho_1^2 + (z_1 - x)^2} + \sqrt{\rho_2^2 + (x - z_2)^2}$. But the shortest of these paths is the direct line from $\mathbf{a}$ to $\mathbf{b}$, that crosses the x-axis at $\left(\frac{\rho_1 z_2 + \rho_2 z_1}{\rho_1 + \rho_2}, 0\right)$ (see Figure B.4). This implies the minimum of $f(x)$ is reached by $x^* = \frac{\rho_1 z_2 + \rho_2 z_1}{\rho_1 + \rho_2}$, where $|z_i - x^*| = \frac{\rho_i}{\rho_1 + \rho_2}|z_1 - z_2|$ for $i = 1, 2$, and the minimum value is $f(x^*) = \sqrt{(\rho_1 + \rho_2)^2 + d^2}$. Equivalently, the value $\sqrt{\rho_1^2 + y_1^2} + \sqrt{\rho_2^2 + y_2^2}$ is lower-bounded by $\sqrt{(\rho_1 + \rho_2)^2 + d^2}$ and it is achieved when $y_1 = \frac{\rho_1}{\rho_1 + \rho_2} d$ and $y_2 = \frac{\rho_2}{\rho_1 + \rho_2} d$. This, in turn, is reached on the tree X^* on the geodesic from T_1' to T_2', at a distance $\frac{\rho_1}{\rho_1 + \rho_2} d_{\text{BHV}}(T_1', T_2')$ from T_1'.

□

B.4 Proofs for Section 4.5: Short paths through higher BHV spaces

Proof of Theorem 4.32. Take a generic $t_1 \in \Lambda^{\mathcal{L}}_{\mathcal{K}}(T_1, T_1^{\uparrow})$. Any $p \in P^{\downarrow\mathcal{K}}(T_1^{\uparrow})$ is such that $\Psi_{\mathcal{L}_2}(p) = \emptyset$ or $[\{\ell\} \ddagger \mathcal{L}_2 \setminus \{\ell\}]$ for some $\ell \in \mathcal{L}_2$, which implies also $\Psi_{\mathcal{L}'}[\Psi_{\mathcal{L}_1}(p)] = \emptyset$ or $[\{\ell\} \ddagger \mathcal{L}_2 \setminus \{\ell\}]$ if $\ell \in \mathcal{L}'$, where $\mathcal{L}' = \mathcal{L}_1 \cap \mathcal{L}_2$. Thus $|p|_{T_1^{\uparrow}} = |p|_{t_1}$ for every $p \in P^{\downarrow\mathcal{K}}(T_1^{\uparrow})$ and $||P^{\downarrow\mathcal{K}}(T_1^{\uparrow})|| = ||P^{\downarrow\mathcal{K}}(t_1)||$.

Since the leaves in $\bar{\mathcal{M}}_0$ are regrafted onto T_1 to form any tree $t_1 \in \Lambda^{\mathcal{L}}_{\mathcal{K}}(T_1, T_1^{\uparrow})$, then all external edges to leaves in $\mathcal{K}^+ = \mathcal{K} \cup \bar{\mathcal{M}}_0$ are zero-length in $t_1^{\perp\mathcal{K}}$.

This implies that for any $s = [\mathcal{G}_1 \ddagger \mathcal{G}_2]$ in $\mathcal{S}(t_1)$ such that $|\mathcal{G}_1 \cap \bar{\mathcal{M}}_0|, |\mathcal{G}_2 \cap \bar{\mathcal{M}}_0| \geq 0$ it must be true that $|\mathcal{G}_1 \cap \mathcal{L}'|, |\mathcal{G}_2 \cap \mathcal{L}'| \geq 2$. Since $\mathcal{K}^+ \subset \mathcal{L} \setminus \mathcal{L}'$, then s is still part of the internal edges of $t_1^{\perp\mathcal{K}}$ and $|\mathcal{G}_1 \cap [\mathcal{L} \setminus \mathcal{K}^+]|, |\mathcal{G}_1 \cap [\mathcal{L} \setminus \mathcal{K}^+]| \geq 2$ as well, so s maps to an internal edge when pruning $\mathcal{K}^+$ from $t_1^{\perp\mathcal{K}}$.

Any other $s' \in \mathcal{S}(t_1^{\perp\mathcal{K}})$ is of the form $\left[\mathcal{G}_1' \ddagger \mathcal{G}_2' \sqcup \bar{\mathcal{M}}_0\right]$, with $|\mathcal{G}_1' \setminus \mathcal{K}|, |\,[\mathcal{G}_2' \setminus \mathcal{K}] \cup \bar{\mathcal{M}}_0| \geq 2$. By the definition of $\bar{\mathcal{M}}_0$, $\mathcal{G}_2' \setminus \mathcal{K} \geq 2$ as well, so s' also maps to an internal edge under the TDR map $\Psi_{\mathcal{L}\setminus\mathcal{K}^+}$.

This proves $\mathcal{K} \cup \bar{\mathcal{M}}_0$ is mutually prunable from $t_1^{\perp\mathcal{K}}$. By Lemma 4.18, $\bar{\mathcal{M}}_0$ is mutually prunable from $\psi(t_1^{\perp\mathcal{K}}, \mathcal{K})$.

For the last part, consider an edge s in $\mathcal{S}(T_1) \setminus P^{\downarrow\mathcal{K}}(T_1)$; i.e. $s = [\mathcal{L}_1' \cup \mathcal{K}_1' \ddagger \mathcal{L}_2' \cup \mathcal{K}_2']$, with $|\mathcal{L}_1'|, |\mathcal{L}_2'| \geq 2$. By the definition of independent maximal sets, $\mathcal{K}_i \subseteq \mathcal{K}_1'$ or $\mathcal{K}_i \subseteq \mathcal{K}_2'$ for every $i = 1, \ldots, r$. For these, define $\bar{\mathcal{M}}_j' = \bigcup_{i|\mathcal{K}_i \subseteq \mathcal{K}_j'} \bar{\mathcal{M}}_i$ for$j = 1, 2$, and define $s' = \left[\mathcal{L}_1' \cup \bar{\mathcal{M}}_1' \ddagger \mathcal{L}_2' \cup \bar{\mathcal{M}}_2'\right]$. Note $s \in \mathcal{S}(T_1)$ if and only if $s' \in \mathcal{S}\left(\psi(t_1^{\perp\mathcal{K}}, \mathcal{K} \cup \bar{\mathcal{M}}_0)\right)$. Moreover, for any edge in $s^{\uparrow} \in \mathcal{S}(t_1)$, $\Psi_{\mathcal{L}_1}(s^{\uparrow}) = s$ if and only if $\Psi_{\mathcal{L}\setminus\mathcal{K}^+}(s^{\uparrow}) = s'$, which implies $|s|_{T_1} = |s'|_{\psi(t_1^{\perp\mathcal{K}}, \mathcal{K} \cup \bar{\mathcal{M}}_0)}$. Since any other internal edge p of t_1 is such that $\Psi_{\mathcal{L}_1}(p) \in P^{\downarrow\mathcal{K}}(T_1)$,

and these are identical between t_1 and $T_1^{\uparrow}$, then $\psi(T_1^{\uparrow \perp \mathcal{K}}, \mathcal{K} \cup \bar{\mathcal{M}}_0) = \psi(t_1^{\perp \mathcal{K}}, \mathcal{K} \cup \bar{\mathcal{M}}_0)$. $\square$

Proof of Theorem 4.33. By Theorem 4.26, $d^2_{\text{BHV}}(t_1, t_2) \geq ||P^{\downarrow \mathcal{K}}(t_1)||^2 + d^2_{\text{BHV}}(\psi(t_1^{\perp \mathcal{K}}, \mathcal{K}), T_2)$. Since $\psi(t_1^{\perp \mathcal{K}}, \mathcal{K}) \in \Lambda^{\mathcal{L}_2}(T_1^{\downarrow})$, then Theorem 4.26 applies again to show $d^2_{\text{BHV}}(\psi(t_1^{\perp \mathcal{K}}, \mathcal{K}), T_2) \geq ||P^{\downarrow \bar{\mathcal{M}}_0}(T_2)||^2 + d^2_{\text{BHV}}(T_1^{\downarrow}, T_2^{\downarrow})$. By Lemma 4.32:

$$d^2_{\text{BHV}}(t_1, t_2) \geq ||P^{\downarrow \mathcal{K}}(t_1)||^2 + ||P^{\downarrow \bar{\mathcal{M}}_0}(T_2)||^2 + d^2_{\text{BHV}}(T_1^{\downarrow}, T_2^{\downarrow})$$

By Theorem 4.26, there is a tree $t_1^* \in \Lambda^{\mathcal{L}_2}(T_1^{\downarrow})$ achieving the lower bound

$$d^2_{\text{BHV}}(t_1^*, T_2) = ||P^{\downarrow \bar{\mathcal{M}}_0}(T_2)||^2 + d^2_{\text{BHV}}(T_1^{\downarrow}, T_2^{\downarrow}).$$

It is possible to find $T_1^* \in \Lambda^{\mathcal{L}}(T_1)$ such that $||P^{\downarrow \mathcal{K}}(T_1^{\uparrow})|| = ||P^{\downarrow \mathcal{K}}(T_1^*)||$ and $t_1^* = \psi(T_1^{* \perp \mathcal{K}}, \mathcal{K})$, by regrafting onto T_1 the leaves in every $\bar{\mathcal{M}}_1, \ldots, \bar{\mathcal{M}}_r$ in the same positions as in $T_1^{\uparrow}$, and then regrafting $\bar{\mathcal{M}}_0$ at the positions indicated by t_1^*. The proof concludes by applying Theorem 4.26 to find the tree $T_2^* \in \Lambda^{\mathcal{L}}(T_2)$ achieving the lower boundary

$$d^2_{\text{BHV}}(T_1^*, T_2^*) = ||P^{\downarrow \mathcal{K}}(T_1^*)||^2 + d^2_{\text{BHV}}(t_1^*, T_2).$$

$\square$

Proof of Theorem 4.34. Given trees $t_1 \in \Lambda^{\mathcal{L}' \cup \mathcal{K} \cup \mathcal{M}^U}_{\mathcal{K}^*}(T_1, T_1^{\uparrow})$ and $t_2 \in \Lambda^{\mathcal{L}' \cup \mathcal{M} \cup \mathcal{K}^U}_{\mathcal{M}^*}(T_2, T_2^{\uparrow})$, theorem 4.29 indicates that the shortest path from t_1 to t_2 through pruning $\mathcal{K}^{D*}$ and later regrafting $\mathcal{M}^{D*}$ is of length

$$\begin{aligned} &\sqrt{\left[d_{\text{BHV}}(t_1, t_1^{\perp \mathcal{K}^{D*}}) + d_{\text{BHV}}(t_2, t_2^{\perp \mathcal{M}^{D*}})\right]^2 + d^2_{\text{BHV}}(t_1', t_2')} \\ &= \sqrt{\left[||P^{\downarrow \mathcal{K}^{D*}}(T_1^{\uparrow})|| + ||P^{\downarrow \mathcal{M}^{D*}}(T_2^{\uparrow})||\right]^2 + d^2_{\text{BHV}}(t_1', t_2')} \end{aligned}$$

where $t_1' = \psi(t_1^{\perp \mathcal{K}^{D*}}, \mathcal{K}^{D*})$ and $t_2' = \psi(t_2^{\perp \mathcal{M}^{D*}}, \mathcal{M}^{D*})$. By Lemmna 4.32, $\bar{\mathcal{M}}_0^U$ is prunable from t_1' and $T_1^{\downarrow} = \psi(t_1', \bar{\mathcal{M}}_0^U)$, and $\bar{\mathcal{K}}_0^U$ is mutually prunable from t_2' and $T_2^{\downarrow} = \psi(t_2', \bar{\mathcal{K}}_0^U)$, so $t_1' \in \Lambda^{\mathcal{L}^U \cup \mathcal{K}^U \cup \mathcal{M}^U}(T_1^{\downarrow})$ and $t_2' \in \Lambda^{\mathcal{L}^U \cup \mathcal{K}^U \cup \mathcal{M}^U}(T_2^{\downarrow})$. Moreover, the edges in $\bar{\mathcal{M}}_0^U$ are regrafted

onto $T_1^{\downarrow}$ to form t_1' in edges that are not in $P^{\downarrow\bar{\mathcal{K}}_0^U}(T_1^{\downarrow})$, since these are part of the image of edges in $P^{\downarrow\mathcal{K}}(t_1)$. Thus, Theorem 4.33 applies and there exists optimal trees $t_1^* \in \Lambda^{\mathcal{L}^U\cup\mathcal{K}^U\cup\mathcal{M}^U}(T_1^{\downarrow})$ and $t_2^* \in \Lambda^{\mathcal{L}^U\cup\mathcal{K}^U\cup\mathcal{M}^U}(T_2^{\downarrow})$ such that $d_{\text{BHV}}(t_1', t_2') \geq d_{\text{BHV}}(t_1^*, t_2^*)$, since their distance reaches the lower bound

$$d^2_{\text{BHV}}(t_1^*, t_2^*) = d^2_{\text{BHV}}(t_1', t_1'^{\perp\bar{\mathcal{K}}_0^U}) + d^2_{\text{BHV}}(T_2^{\downarrow}, T_2^{\downarrow\perp\bar{\mathcal{M}}_0^U}) + d^2_{\text{BHV}}(T_1^{\downarrow\downarrow}, T_2^{\downarrow\downarrow}).$$

Note that in the previous expression $d^2_{\text{BHV}}(t_1', t_1'^{\perp\bar{\mathcal{K}}_0^U})$ can be replaced by $d^2_{\text{BHV}}(T_1^{\downarrow}, T_1^{\downarrow\perp\bar{\mathcal{K}}_0^U})$. By regrafting the edges in $P^{\downarrow\mathcal{K}^{D*}}(t_1)$ onto t_1^* and $P^{\downarrow\mathcal{M}^{D*}}(t_2)$ onto t_2^*, we can find optimal trees $T_1^* \in \Lambda_0^{\mathcal{L}'\cup\mathcal{K}\cup\mathcal{M}^U}(T_1, T_1^{\uparrow})$ and $T_2^* \in \Lambda_0^{\mathcal{L}'\cup\mathcal{M}\cup\mathcal{K}^U}(T_2, T_2^{\uparrow})$ such that the path between them achieves the lower bound

$$\sqrt{\left[||P^{\downarrow\mathcal{K}^D}(T_1^{\uparrow})|| + ||P^{\downarrow\mathcal{M}^D}(T_2^{\uparrow})||\right]^2 + ||P^{\downarrow\bar{\mathcal{K}}_0^U}(T_1^{\downarrow})||^2 + ||P^{\downarrow\bar{\mathcal{M}}_0^U}(T_2^{\downarrow})||^2 + d^2_{\text{BHV}}(T_1^{\downarrow\downarrow}, T_2^{\downarrow\downarrow})}.$$

□

Proof of Corollary 4.35. Fix two trees $T_1^{\uparrow} = \Lambda^{\mathcal{L}'\cup\mathcal{K}\cup\mathcal{M}^U}(T_1)$ and $T_2^{\uparrow} = \Lambda^{\mathcal{L}'\cup\mathcal{M}\cup\mathcal{K}^U}(T_2)$ and build the partitions of $\mathcal{K}^*$, $\mathcal{M}^*$, $\mathcal{M}^U$ and $\mathcal{K}^U$ as in Theorem 4.34. Per the theorem, the shortest path between trees $t_1 \in \Lambda_{\mathcal{K}^*}^{\mathcal{L}'\cup\mathcal{K}\cup\mathcal{M}^U}(T_1, T_1^{\uparrow})$ to any tree $t_2 \in \Lambda_{\mathcal{M}^*}^{\mathcal{L}'\cup\mathcal{M}\cup\mathcal{K}^U}(T_2, T_2^{\uparrow})$ as described in the corollary is

$$\sqrt{\left[||P^{\downarrow\mathcal{K}^{D*}}(T_1^{\uparrow})|| + ||P^{\downarrow\mathcal{M}^{D*}}(T_2^{\uparrow})||\right]^2 + ||P^{\downarrow\bar{\mathcal{K}}_0^U}(T_1^{\downarrow})||^2 + ||P^{\downarrow\bar{\mathcal{M}}_0^U}(T_2^{\downarrow})||^2 + d^2_{\text{BHV}}(T_1^{\downarrow\downarrow}, T_2^{\downarrow\downarrow})},$$

where $T_1^{\downarrow} = \psi(T_1^{\uparrow\perp\mathcal{K}^{D*}}, \mathcal{K}^{D*}\cup\bar{\mathcal{M}}_0^U)$, $T_2^{\downarrow} = \psi(T_2^{\uparrow\perp\mathcal{M}^{D*}}, \mathcal{M}^{D*}\cup\bar{\mathcal{K}}_0^U)$, $T_1^{\downarrow\downarrow} = \psi(T_1^{\downarrow\perp\bar{\mathcal{K}}_0^U}, \bar{\mathcal{K}}_0^U)$ and $T_2^{\downarrow\downarrow} = \psi(T_2^{\downarrow\perp\bar{\mathcal{M}}_0^U}, \bar{\mathcal{M}}_0^U)$.

On the other hand, consider new fixed trees $T_1^{\uparrow\uparrow} = \Lambda^{\mathcal{L}'\cup\mathcal{K}\cup\mathcal{M}}(T_1)$ and $T_2^{\uparrow\uparrow} = \Lambda^{\mathcal{L}'\cup\mathcal{M}\cup\mathcal{K}}(T_2)$, build by regrafting $\mathcal{M}^U$ and $\mathcal{K}^U$ as in $T_1^{\uparrow}$ and $T_2^{\uparrow}$ respectively, and then regrafting all $\mathcal{M}^D$ and $\mathcal{K}^D$ in edges that do not belong to independent maximal sets. Thus, for these two trees the new partitions are $\mathcal{K}_2^U = \mathcal{K}$, $\mathcal{K}_2^D = \emptyset$, $\mathcal{M}_2^U = \mathcal{M}$, $\mathcal{M}_2^D = \emptyset$, and $\mathcal{K}_2^{D*} = \mathcal{M}_2^{D*} = \mathcal{L}^D$;

$\bar{\mathcal{K}}^U_{0,2} = \bar{\mathcal{K}}^U_0 \cup \mathcal{K}^D$ and $\bar{\mathcal{M}}^U_{0,2} = \bar{\mathcal{M}}^U_0 \cup \mathcal{M}^D$, while $\bar{\mathcal{M}}^U_{i,2} = \bar{\mathcal{M}}^U_i$ for all $i = 1, \ldots, r_1$ and $\bar{\mathcal{K}}^U_{i,2} = \bar{\mathcal{K}}^U_i$ for all $i = 1, \ldots, r_2$. Per Theorem 4.34 again, the shortest path is of length

$$\sqrt{\left[||P^{\downarrow\mathcal{L}^D}(T_1^{\uparrow\uparrow})|| + ||P^{\downarrow\mathcal{L}^D}(T_2^{\uparrow\uparrow})||\right]^2 + ||P^{\downarrow\bar{\mathcal{K}}^U_{0,2}}(T^{\downarrow}_{1,2})||^2 + ||P^{\downarrow\bar{\mathcal{M}}^U_{0,2}}(T^{\downarrow}_{2,2})||^2 + d^2_{\mathrm{BHV}}(T^{\downarrow\downarrow}_{1,2}, T^{\downarrow\downarrow}_{2,2})},$$

where $T^{\downarrow}_{1,2} = \psi(T_1^{\uparrow\uparrow\perp\mathcal{L}^D}, \mathcal{L}^D \cup \bar{\mathcal{M}}^U_{0,2})$, $T^{\downarrow}_{2,2} = \psi(T_2^{\uparrow\uparrow\perp\mathcal{L}^D}, \mathcal{L}^D \cup \bar{\mathcal{K}}^U_{0,2})$, $T^{\downarrow\downarrow}_{1,2} = \psi(T^{\downarrow\perp\bar{\mathcal{K}}^U_{0,2}}_{1,2}, \bar{\mathcal{K}}^U_{0,2})$ and $T^{\downarrow\downarrow}_{2,2} = \psi(T^{\downarrow\perp\bar{\mathcal{M}}^U_{0,2}}_{2,2}, \bar{\mathcal{M}}^U_{0,2})$.

But it turns out that $T^{\downarrow\downarrow}_1 = T^{\downarrow\downarrow}_{1,2}$, $T^{\downarrow\downarrow}_2 = T^{\downarrow\downarrow}_{2,2}$ and

$$\begin{aligned}
&\left[||P^{\downarrow\mathcal{K}^{D*}}(T_1^{\uparrow})|| + ||P^{\downarrow\mathcal{M}^{D*}}(T_2^{\uparrow})||\right]^2 + ||P^{\downarrow\bar{\mathcal{K}}^U_0}(T_1^{\downarrow})||^2 + ||P^{\downarrow\bar{\mathcal{M}}^U_0}(T_2^{\downarrow})||^2 \geq \\
&\geq \left[||P^{\downarrow\mathcal{L}^D}(T_1^{\uparrow\uparrow})|| + ||P^{\downarrow\mathcal{L}^D}(T_2^{\uparrow\uparrow})||\right]^2 + ||P^{\downarrow\bar{\mathcal{K}}^U_{0,2}}(T^{\downarrow}_{1,2})||^2 + ||P^{\downarrow\bar{\mathcal{M}}^U_{0,2}}(T^{\downarrow}_{2,2})||^2
\end{aligned}$$

□

REFERENCES CITED

Bacák, M., *Computing medians and means in Hadamard spaces*, SIAM journal on optimization **24** (2014), no. 3, 1542–1566.

Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., & Lloyd, K. G., *Diversity, ecology and evolution of Archaea* (eng), Nature microbiology **5** (2020), no. 7, 887–900.

Barden, D., Le, H., & Owen, M., *Limiting behaviour of Fréchet means in the space of phylogenetic trees*, Ann. Inst. Statist. Math. **24** (2016), 1542.

Barden, D., Le, H., & Owen, M., *Central limit theorems for Fréchet means in the space of phylogenetic trees*, Electron. J. Probab **18** (2013), no. 25, 1–25.

Benner, P., Bacák, M., & Bourguignon, P.-Y., *Point estimates in phylogenetic reconstructions*, Bioinformatics **30** (2014), no. 17, 534–540.

Billera, L. J., Holmes, S. P., & Vogtmann, K., *Geometry of the Space of Phylogenetic Trees*, Adv. in Appl. Math. **27** (2001), no. 4, 733–767, https://doi.org/10.1006/aama.2001.0759.

Boyd, S. P., & Vandenberghe, L., *Convex optimization*, Cambridge University Press, Cambridge, UK, 2004.

Briand, S., Dessimoz, C., El-Mabrouk, N., Lafond, M., & Lobinska, G., *A generalized Robinson-Foulds distance for labeled trees*, BMC genomics **21** (2020), 1–13.

Bridson, M. R., & Haefliger, A., *Metric Spaces of Non-Positive Curvature*, *Vol. 319*, Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

Brown, D. G., & Owen, M., *Mean and variance of phylogenetic trees*, Systematic biology **69** (2020), no. 1, 139–154.

Chen, Y., Burall, L. S., Luo, Y., Timme, R., Melka, D., Muruvanda, T., Payne, J., Wang, C., Kastanis, G., Maounounen-Laasri, A., et al., *Listeria monocytogenes in stone fruits linked to a multistate outbreak: enumeration of cells and whole-genome sequencing*, Applied and Environmental Microbiology **82** (2016), no. 24, 7030–7040.

Dagan, T., & Martin, W., *The tree of one percent*, Genome biology **7** (2006), 1–7.

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G., *Bayesian coalescent inference of past population dynamics from molecular sequences* (eng), Molecular biology and evolution **22** (2005), no. 5, 1185–1192.

Duarte, C. M., Ngugi, D. K., Alam, I., Pearman, J., Kamau, A., Eguiluz, V. M., Gojobori, T., Acinas, S. G., Gasol, J. M., Bajic, V., et al., *Sequencing effort dictates gene discovery in marine microbial metagenomes*, Environmental Microbiology **22** (2020), no. 11, 4589–4603.

Edwards, S. V., *Is a new and general theory of molecular systematics emerging?* (eng), Evolution **63** (2009), no. 1, 1–19.

Evans, S. N., & Jaffe, A. Q., *Limit theorems for Fréchet mean sets*, Bernoulli **30** (2024), no. 1, 419–447.

Gardy, J. L., & Loman, N. J., *Towards a genomics-informed, real-time, global pathogen surveillance system* (eng), Nature reviews. Genetics **19** (2018), no. 1, 9–20.

Gori, K., Suchan, T., Alvarez, N., Goldman, N., & Dessimoz, C., *Clustering genes of common evolutionary history*, Molecular biology and evolution **33** (2016), no. 6, 1590–1605.

Grindstaff, G., & Owen, M., *Representations of Partial Leaf Sets in Phylogenetic Tree Space* (eng), SIAM journal on applied algebra and geometry **3** (2019), no. 4, 691–720.

Guindon, S., Gascuel, O., & Rannala, B., *A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood* (eng), Systematic biology **52** (2003), no. 5, 696–704.

Hein, J., Jiang, T., Wang, L., & Zhang, K., *On the complexity of comparing evolutionary trees* (eng), DISCRETE APPLIED MATHEMATICS **71** (1996), no. 1, 153–169.

Hotz, T., Skwerer, S., Huckemann, S., Le, H., Marron, J., Mattingly, J. C., Miller, E., Nolen, J., Owen, M., & Patrangenaru, V., *Sticky central limit theorems on open books* (2013).

Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., et al., *Isolation of an archaeon at the prokaryote–eukaryote interface*, Nature **577** (2020), no. 7791, 519–525.

Jonsson, J., *Simplicial complexes of graphs*, Lecture notes in mathematics, 1928, Springer, Berlin ; 2008.

Kubatko, L. S., Degnan, J. H., & Collins, T., *Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence* (eng), Systematic biology **56** (2007), no. 1, 17–24.

Maddison, W. P., *Gene Trees in Species Trees*, Systematic Biology **46** (1997), no. 3, 523–536.

Miller, E., Owen, M., & Provan, J. S., *Polyhedral computational geometry for averaging metric phylogenetic trees* (eng), Advances in applied mathematics **68** (2015), 51–91.

Nye, T. M., *Principal components analysis in the space of phylogenetic trees*, Ann. Statist. **39** (2011), no. 5, 2716–2739.

———, *Convergence of random walks to Brownian motion on cubical complexes*, arXiv preprint arXiv:1508.02906 (2015).

Owen, M., *Computing geodesic distances in tree space*, SIAM J. Discrete Math. **25** (2011), no. 4, 1506–1529.

Owen, M., & Provan, J. S., *A Fast Algorithm for Computing Geodesic Distances in Tree Space* (eng), IEEE/ACM transactions on computational biology and bioinformatics **8** (2011), no. 1, 2–13.

Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., & Hugenholtz, P., *A complete domain-to-species taxonomy for Bacteria and Archaea* (eng), Nature biotechnology **38** (2020), no. 9, 1079–.

Ren, Y., Zha, S., Bi, J., Sanchez, J. A., Monical, C., Delcourt, M., Guzman, R. K., & Davidson, R., *A combinatorial method for connecting BHV spaces representing different numbers of taxa*, arXiv preprint arXiv:1708.02626v2 (2017).

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W.-T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., *Insights into the phylogeny and coding potential of microbial dark matter* (eng), Nature (London) **499** (2013), no. 7459, 431–437.

Royalty, T. M., & Steen, A. D., *Theoretical and Simulation-Based Investigation of the Relationship between Sequencing Effort, Microbial Community Richness, and Diversity in Binning Metagenome-Assembled Genomes*, Msystems **4** (2019), no. 5, 10–1128.

Ruszczyński, A. P., *Nonlinear optimization* (eng), Princeton University Press, Princeton, N.J, 2006.

Scaduto, D. I., Brown, J. M., Haaland, W. C., Zwickl, D. J., Hillis, D. M., & Metzker, M. L., *Source identification in two criminal cases using phylogenetic analysis of HIV-1 DNA sequences*, Proceedings of the National Academy of Sciences **107** (2010), no. 50, 21242–21247.

Schötz, C., *Strong laws of large numbers for generalizations of Fréchet mean sets*, Statistics **56** (2022), no. 1, 34–52.

Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies* (eng), BIOINFORMATICS **30** (2014), no. 9, 1312–1313.

Sturm, K.-T., *Probability measures on metric spaces of nonpositive curvature*, Contemporary mathematics **338** (2003), 357–390.

Teichman, S., Lee, M. D., & Willis, A. D., *Analyzing microbial evolution through gene and genome phylogenies*, bioRxiv (2023).

Weyenberg, G., Huggins, P. M., Schardl, C. L., Howe, D. K., & Yoshida, R., *KDETREES: non-parametric estimation of phylogenetic tree distributions*, Bioinformatics **30** (2014), no. 16, 2280–2287.

Whidden, C., Zeh, N., & Beiko, R. G., *Supertrees Based on the Subtree Prune-and-Regraft Distance* (eng), Systematic biology **63** (2014), no. 4, 566–581.

Willis, A., *Confidence sets for phylogenetic trees*, J. Amer. Statist. Assoc. **114** (2019), no. 525, 235–244.

Willis, A., & Bell, R., *Uncertainty in phylogenetic tree estimates*, J. Comput. Graph. Statist. **27** (2018), no. 3, 542–552.

Zaheer, R., Noyes, N., Ortega Polo, R., Cook, S. R., Marinier, E., Van Domselaar, G., Belk, K. E., Morley, P. S., & McAllister, T. A., *Impact of sequencing depth on the characterization of the microbiome and resistome*, Scientific reports **8** (2018), no. 1, 5890.

Zairis, S., Khiabanian, H., Blumberg, A. J., & Rabadan, R., *Genomic data analysis in tree spaces*, arXiv:1607.07503 (2016).

Zhu, Q., Mai, U., Pfeiffer, W., Janssen, S., Asnicar, F., Sanders, J. G., Belda-Ferre, P., Al-Ghalith, G. A., Kopylova, E., McDonald, D., et al., *Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea*, Nature Communications **10** (2019), no. 1, 1–14.