**Algebraic Structures in Phylogenetics:**
**Insights from Tropical Convexity and Likelihood Geometry**

by

Shelby Cox

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mathematics)
in the University of Michigan
2024

Doctoral Committee:

        Professor David Speyer, Chair
        Professor Marisa Eisenberg
        Professor Thomas Lam
        Assistant Professor Aida Maraj

Shelby Cox

spcox@umich.edu

ORCID iD: 0000-0003-4378-5103

For my family, especially my husband, Mathew.

# ACKNOWLEDGEMENTS

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF ACRONYMS

**BHV** Billera-Holmes-Vogtman

**LCA** Least common ancestor

**NNI** Nearest neighbor interchange

**FCA** Four clade rearrangement

**MLD** Maximum likelihood degree

**MLE** Maximum likelihood estimator

# ABSTRACT

This thesis explores two areas of algebraic statistics: tropical phylogenetics and likelihood geometry. Part I concerns the space of equidistant phylogenetic trees, which is a convex tropical variety. Specifically, I classify the tree topology changes that can occur along a tropical line segment between two tree data points and compute the tropical geometric median (Fermat-Weber points) of weighted tree data. Part II is concerned with the maximum likelihood degree of Brownian motion tree models. The main result is that the maximum likelihood degree of a Brownian motion tree model on the star tree with $n$ leaves is $2^{n+1}-2n-3$. Another important result in this part is a generalization of the Cayley-Prüfer Theorem to complete graphs with a weighting determined by the paths in a tree.

# CHAPTER 1

# Introduction

This thesis spans three papers in mathematical phylogenetics. Phylogenetics is a field of computational biology concerned with studying the evolutionary relationships among a group of biological units (genes, individuals, species, taxa, etc.). Recently, it has become evident that tools from tropical geometry and algebraic geometry can be useful in studying problems that arise in phylogenetics, and in computational biology and statistics more broadly. One central insight is that the geometric spaces that arise from studying phylogenetics can be interpreted as tropical or algebraic varieties. In this thesis, I focus on two geometric objects arising in phylogenetics: the (tropical) space of phylogenetic trees and an (algebraic) model for Brownian motion along a phylogenetic tree.

Part I concerns averages in the space of phylogenetic trees. The space of phylogenetic trees, $\text{Tree}_n$, is a geometric space where each phylogenetic tree on $n$ leaves is represented by a point. Specifically, each tree corresponds to the vector of pairwise distances among its leaves. $\text{Tree}_n$ is not a Euclidean space and it is not classically convex, which means that standard approaches to computing averages, variances, distributions, random samples, and confidence intervals are not available. For example, simply averaging the pairwise leaf distances of two trees will usually yield a vector that is not a vector of pairwise leaf distances for any tree. However, under a certain embedding, the space of phylogenetic trees is tropically convex, meaning that every point on the tropical line segment between two phylogenetic trees itself represents a phylogenetic tree. This fact opens the door to several new methods for computing averages of phylogenetic trees using polyhedral complexes. Averages can be used as a tool to create consensus trees, which is important for phylogenetic reconstruction, a central problem in phylogenetics.

One way to compute an average of phylogenetic trees is to move along the geodesic (i.e., shortest path) between them in the space of phylogenetic trees. In Chapter 2, I classify the tree topologies (i.e., tree structures) that occur along the tropical line segment between two trees, which is one of many geodesics in tropical tree space [32]. Thus, the results of

Chapter 2 inform our understanding of a tropical average of two phylogenetic trees. To average more than two phylogenetic trees, Chapter 3 studies weighted Fermat-Weber points.

Part II is concerned with likelihood inference for a certain evolutionary model, called a Brownian motion tree model. A Brownian motion tree model is a family of Gaussian probability distributions that describe the evolution of a continuous, real-valued trait over a phylogenetic tree. Brownian motion tree models are canonical models of genetic drift, which refers to the accumulation of random (non-selective) genetic mutations over time, and they also have a nice algebraic interpretations. Likelihood geometry uses tools from algebraic geometry to infer the parameters of an algebraic model that maximize the likelihood of observing some fixed data. In the context of Brownian motion tree models, likelihood geometry can be used to infer a phylogenetic tree, or to infer edge lengths of a known phylogenetic tree, from biological data. Unlike purely combinatorial methods for inferring a phylogenetic tree, with this method it is relatively easy to compute standard errors and confidence in the estimate [17]. In Chapter 4, I use results from intersection theory to compute the complexity of inferring the parameters (edge lengths in the tree) for certain Brownian motion tree models.

The following sections introduce the background material for the following three chapters. Section 1.1 describes phylogenetic trees and tree spaces from a mathematical point of view. Section 1.2 and Section 1.5 contain brief introductions to the concepts in convexity and computational algebraic geometry that are used in this thesis. Section 1.3 introduces tropical geometry, tropical convexity, and the space of phylogenetic trees as a tropical variety. Section 1.4 defines statistical models, maximum likelihood, and Brownian motion tree models, the topics of Chapter 4. Finally, Section 1.6 includes a summary of results and contributions.

## 1.1 Phylogenetics

Trees are the central information structures in phylogenetics. Below, I briefly define the the data that accompanies a phylogenetic tree. In the following subsections, I introduce notation associated to phylogenetic trees and rigorously define tree spaces, including the Nearest Neighbor Interchange graph and the Billera-Holmes-Vogtman space of trees. Parts of this section are from [12], which was first published by MSP in *Algebraic Statistics* in volume 14(1), pages 71-90.

**Definition 1.1.1.** A *phylogenetic tree* is a metric leaf-labeled tree with no degree two ver-

tices. Internal edge[1] lengths must be real and non-negative; edges adjacent to leaves can have any real length. A *rooted phylogenetic tree* is a phylogenetic tree where one leaf is distinguished as the root of the tree. The edges of a rooted phylogenetic tree are implicitly directed away from the root.

Phylogenetic trees are used for a wide variety of models, and so the vertices of the tree may represent a range of data, which may be discrete or continuous. Genetic data and discrete traits are examples of discrete data. Genetic data may refer to individual bases from the set $\{A, C, G, T\}$, or it may refer to an *alignment*, which is a sequence of genes. Discrete traits refer to the presence, absence, or (whole) number of a certain discrete observable feature. For example, the presence, absence, or number of legs an individual millipede has is a discrete trait. Continuous biological data includes gene expression data and continuous traits like mass, length, etc. Gene expression data is a continuous measure of the quantity of genetic products (proteins or RNA) at the cellular level; it is the first level at which the phenotype can be measured. Depending on the model and experiment, the source of the data also varies. Thus, vertices of the tree may correspond to data from individual cells, data from individuals within the same species, or an average across a larger group of individuals – a species, family, genus, etc.

When I wish to be agnostic about the data on the vertices of the tree and at what level it is measured, I will refer to it as *biological data* measured on *biological units*. Edge lengths in a phylogenetic tree, called the *speciation times*, represent the amount of mutation that has occurred along the branch in the observed biological data between the biological units on either end of the branch.

### 1.1.1   Notation

I will briefly introduce some notation for graphs and trees that will be used throughout the thesis. For unrooted trees, I will often identify the leaf labels with the set $[n] = \{1, 2, \ldots, n\}$; for rooted trees, I will usually use the leaf labels $\{0\} \cup [n]$, where $0$ is the root, unless stated otherwise. The non-root leaves are called *species leaves*. The vertex directly below the root leaf is denoted by $\rho$.

Let $G$ be any graph, and let $v$ be a vertex of $G$. The vertices and edges of $G$ are denoted $\mathrm{Ver}(G)$ and $E(G)$ respectively. The degree of $v$ in $G$ is $\deg_G(v)$. For a phylogenetic tree $T$, the internal vertices and (non-root) leaves of $T$ are $\mathrm{Int}(T)$ and $\mathrm{Lv}(T)$ respectively. Let $T_0$ is the tree in Figure 1.1. Then $\mathrm{Int}(T_0) = \{5, 6\}$ and $\mathrm{Lv}(T_0) = \{1, 2, 3, 4\}$. The notation $i \leftrightsquigarrow_T j$, or simply $i \leftrightsquigarrow j$ when $T$ is clear from context, refers to the unique path in $T$

---

[1]An internal edge is an edge that is not connected to a leaf.
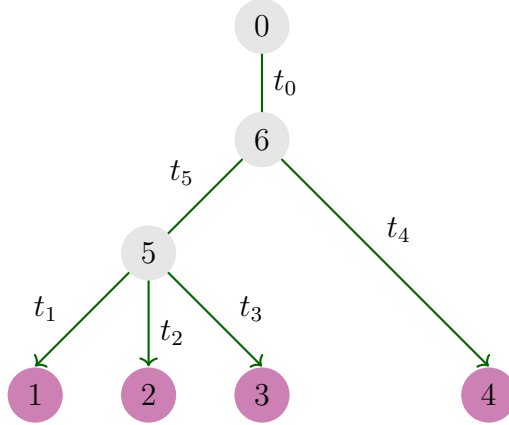
Figure 1.1: A rooted phylogenetic tree with 4 species leaves, $T_0$.

between $i, j \in \operatorname{Ver}(T)$. The distance between $u, v \in \operatorname{Ver}(T)$, denoted $d_T(u, v)$, is the sum of the edge lengths over edges in $u \leftrightsquigarrow v$. For example, $d_{T_0}(1, 4) = t_1 + t_5 + t_4$.

The *descendants* of a vertex $v \in \operatorname{Ver}(T)$ are all the vertices $u \in \operatorname{Ver}(T)$ that lie below $v$ in $T$, meaning that the path $v \leftrightsquigarrow u$ does not intersect the path $v \leftrightsquigarrow \rho$. The *ancestors* of $v$ are all vertices on the path $v \leftrightsquigarrow \rho$. The adjacent descendants are called the *children of $v$*, and the adjacent ancestor is called the *parent of $v$*. For example, the descendants of vertex 5 in $T_0$ (which are also the children of 5) are 1, 2 and 3. The parent and only ancestor of 5 in $T_0$ is $6 = \rho$.

Note that the internal vertices of $T$, while unlabeled by definition, are uniquely identified by their set of descendants leaves, which is called a *clade*. Conversely, given a pair of leaves $i, j$ of $T$, there is a unique closest vertex that is an ancestor of both $i$ and $j$, which is called the *least common ancestor of $i$ and $j$*. For example, $\operatorname{lca}_{T_0}(1, 2) = 5$, and $\operatorname{lca}(T_0)(1, 4) = 6$.

**Definition 1.1.2.** Given two leaves $i, j$ of $T$, the *least common ancestor* of $i$ and $j$, denoted $\operatorname{lca}_T(i, j)$, is the internal vertex in the intersection of the paths $i \leftrightsquigarrow j$, $\rho \leftrightsquigarrow i$, and $\rho \leftrightsquigarrow j$.

A tree is *fully resolved* or *binary* if every interior vertex has degree exactly three. A tree with exactly one internal node is a *star tree*. The tree in Figure 1.1 is only partially resolved (it is not a binary tree nor is it a star tree).

## 1.1.2 Tree Topology

Informally, the *topology* of a phylogenetic tree encodes its combinatorial (and sometimes, metric) structure. *Edge splits*, defined below, are a key tool in determining the topology of a tree.

4

**Definition 1.1.3.** Given a phylogenetic tree $T$ and an edge $e \in E(T)$, the *edge split for $T$ at $e$* is the bipartition that the removal of $e$ induces on the leaves of $T$.

I will use several versions of tree topology, which may or may not depend on the the edge lengths. Without using the edge lengths, there are two types of (unrooted) tree topology. First, the *(leaf-labeled) tree topology of $T$* is the data of the edge splits of $T$. This is the most common tree structure in this thesis, so I will refer to it as the *tree topology of $T$* from now on. Second, the *unlabeled tree topology of $T$* is the data of the edge splits of $T$ up to a permutation of the leaf labels. Note that if $T$ and $T'$ have the same tree topology, then they also have the same unlabeled tree topology, but the converse is not necessarily true. Thus, we say that the tree topology is a finer tree structure than the unlabeled tree topology.

Taking the edge lengths into account allows for even finer distinctions between tree structures. Specifically, the internal vertices in each tree are ranked according to their distance from the root. Recall that the internal vertices are canonically labeled by the set of their descendants, so it is possible to compare the internal vertex rankings between two trees with the same leaf-labeled tree topology. The *ranked tree topology of $T$* is the data of the (leaf-labeled) topology and the ranking on the internal vertices of $T$ by distance from the root. Definition 1.1.4 contains a summary of the different tree structures.

**Definition 1.1.4.** Let $T$ and $T'$ be phylogenetic trees. Three versions of tree topology are summarized below, from coarsest to finest.

1. If $T$ and $T'$ have the same set of edge splits after a permutation of the leaf labels, then they have the same *unlabeled tree topology*;

2. If $T$ and $T'$ have exactly the same set of edge splits, then they have the same *tree topology*;

3. If $T$ and $T'$ have exactly the same set of edge splits, and the same ranking on internal vertices, then they have the same *ranked tree topology*.

We can also use the tree metric to determine the tree topology. If $T$ is a tree and $S \subset \binom{[n]}{2}$, define $\mathrm{argmax}_T^1 = \mathrm{argmax}_T$ to be the subset of index pairs $\{i,j\} \in S$ which maximize $\mathbf{d}_T(i,j)$. Then define $\mathrm{argmax}_T^m\{S\}$ to be the subset of indices of $S$ which achieve the $m$th largest value of $\mathbf{d}_T(i,j)$ among indices in $S$. Tree topology can be tested using the argmax function.

**Lemma 1.1.5.** *Two phylogenetic trees $T$, $T'$ have the same topology if*

$$argmax_T\{ij, ik, jk\} = argmax_{T'}\{ij, ik, jk\}$$

*for all $i < j < k \in [n]$.*
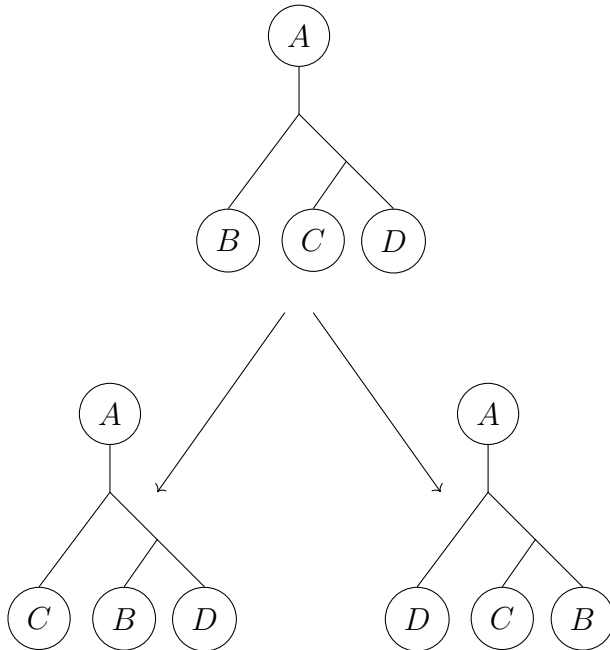
Figure 1.2: Possible NNI moves.

### 1.1.3 Rearrangement Graphs

The collection of phylogenetic trees on $n$ leaves can be realized discretely or continuously. The discrete realizations forget the edge lengths of the tree and rely only on the tree topology. A rearrangement move, R, is a rule that changes the connectivity of subtrees within a binary branching tree $T$. A rearrangement graph, $\mathcal{R}_n$, has one vertex for each binary branching tree topology on $n$ leaves. Vertices in $\mathcal{R}_n$ are connected by an edge if and only if the corresponding tree topologies differ by exactly one rearrangement by $R$. The possible topologies of trees along the tropical line segment between two trees are closely related to a rearrangement move called *Nearest Neighbor Interchange (NNI)*.

**Definition 1.1.6.** Let $T$ be a binary phylogenetic tree. Let $A$, $B$, $C$, and $D$ be the four subtrees connected to an internal edge $e \in T$. A *Nearest Neighbor Interchange move (NNI)*, illustrated in Figure 1.2, is a rearrangement of $A$, $B$, $C$, and $D$ around $e$.

NNI was introduced independently in [33] and [37], and the NNI graph is a well-studied graph. In 1997, DasGupta et al showed that computing shortest paths in the NNI graph is NP-complete, and [28] provide a polynomial time $O(n^2)$ algorithm that approximates the length of the shortest NNI path up to a factor of $4\log(n) + 4$. The diameter of the NNI graph is known to be $O(n \log n)$, and the expected distance between two randomly chosen
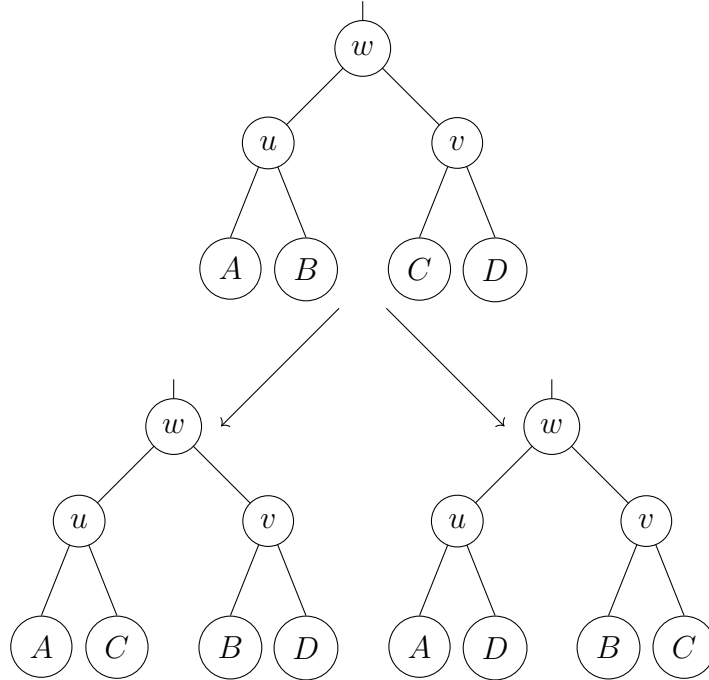
Figure 1.3: Possible four clade rearrangements.

vertices is also $O(n \log n)$. In Chapter 2, I will compare the number of NNI along a tropical line segment between two trees to the polynomial time approximation to the NNI distance given in [28].

Another relevant rearrangement move is a *four clade rearrangement*.

**Definition 1.1.7.** Let $T$ be a rooted binary branching tree with $w \in \text{Int}(T)$, and let $u, v$ be the children of $w$. A *four clade rearrangement (FCR)*, illustrated in Figure 1.3, is any rearrangement of the four subtrees below $u$ and $v$. More precisely, if $A, B$ are the subtrees descending from the children of $u$, and $C, D$ are the two subtrees descending from the children of $v$, then an FCR swaps either $A$ or $B$ with either $C$ or $D$.

Note that an FCR is equivalent to a sequence of three NNI moves.

### 1.1.4 The Space of Phylogenetic Trees

In this subsection, I begin by defining the space of unrooted phylogenetic trees using pairwise distances on the leaves. This first space is the Billera-Holmes-Vogtman (BHV) space of trees [3], which has been extensively studied by mathematicians [31, 36], statisticians [22, 35, 50], and biologists [2, 52]. Next, I define the space of equidistant (rooted) phylogenetic trees,

7

which is homeomorphic to a BHV space and is tropically convex. Tropical convexity of equidistant phylogenetic tree space will be addressed fully in Section 1.3.

An unrooted phylogenetic tree, $T$, is completely determined by the pairwise distances between its leaves. We denote this vector of distances by $\mathbf{d}_T$

$$\mathbf{d}_T := (d_T(i,j))_{\substack{i,j \in \mathrm{Lv}(T) \\ i \neq j}}.$$

The constraints that a tree imposes on its distance vector are summarized in the *four point condition*, which is defined below.

**Theorem 1.1.8** (Four Point Condition [29]). *A vector* $\mathbf{u} \in \mathbb{R}^{\binom{n}{2}}$ *is equal to* $\mathbf{d}_T$ *for some phylogenetic tree if and only if*

$$\max\{u_{ij} + u_{kl}, u_{ik} + u_{jl}, u_{il} + u_{jk}\}$$

*is achieved at least twice for all distinct* $i, j, k, l \in [n]$.

**Definition 1.1.9.** The space of phylogenetic trees on $n$ leaves is the collection of vectors $\mathbf{u} \in \mathbb{R}^{\binom{n}{2}}$ that satisfy the 4-point condition. This is the space of phylogenetic trees introduced by Billera, Holmes, and Vogtman in [3], so we denote it by $\mathrm{BHV}_n$.

$$\mathrm{BHV}_n := \left\{ \mathbf{u} \in \mathbb{R}^{\binom{n}{2}} : \mathbf{u} = \mathbf{d}_T, \text{ for some phylogenetic tree } T \text{ on } n \text{ leaves} \right\}$$
$$= \left\{ \mathbf{u} \in \mathbb{R}^{\binom{n}{2}} : \mathbf{u} \text{ satisfies the 4-point condition} \right\}.$$

Geometrically, $\mathrm{BHV}_n$ is a fan in $\mathbb{R}^{\binom{n}{2}}$, with the subspace metric inherited from Euclidean space [3]. Each cone, $C_\tau$, corresponds to a tree topology, $\tau$, and the points in the interior of $C_\tau$ parameterize all unrooted phylogenetic trees with topology $\tau$ [3]. The dimension of $C_\tau$ is the number of edges $\tau$ has, so more resolved topologies correspond to higher dimensional cones. For example, $\mathrm{BHV}_5$, depicted in Figure 1.4, is $\mathbb{R}^5$ times the cone over the Petersen graph. Each edge of the Petersen graph represents a 7-dimensional cone (a two-dimensional orthant, $\mathbb{R}^2_{\geq 0}$, times $\mathbb{R}^5$), which corresponds to a fully resolved tree topology with two internal edges and 5 edges adjacent to leaves.

Next, I define the space of equidistant trees (rooted at 0). Recall that a rooted phylogenetic tree on $[n]$ species leaves is a tree with leaf set $\{0\} \cup [n]$, where the leaf labeled 0 is distinguished as the root of the tree. The vertex directly below 0 is denoted $\rho$, and I will assume for simplicity that the length of the edge from 0 to $\rho$ is 0.

**Definition 1.1.10.** A rooted phylogenetic tree, $T$, is *equidistant* if $\mathbf{d}_T(0, \bullet)$ is constant on $\mathrm{Lv}(T)$. That is, if the distance from the root to any leaf is the same.
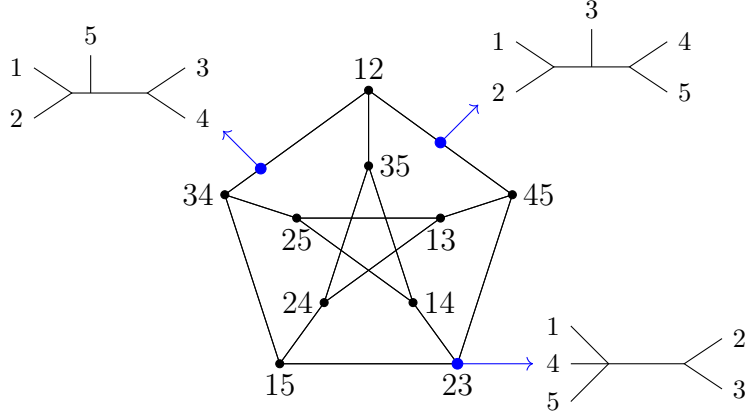
Figure 1.4: $BHV_5$ is the cone over the Petersen graph.

For a rooted equidistant tree, the height of an internal vertex (from the leaves) is well-defined and it will be useful to have a notation for it.

**Definition 1.1.11.** Let $x$ be an internal vertex of $T$. For any $i, j \in [n]$ such that $\text{lca}_T(i, j) = x$,

$$h_T(x) = \frac{1}{2}\mathbf{d}_T(i, j).$$

Note that $\mathbf{d}_T \in \mathbb{R}^{\binom{n+1}{2}}$, but the distances $\mathbf{d}_T(0, i)$, between 0 and each other leaf are redundant. This is because $\mathbf{d}_T(0, i)$ is the height of $\rho$, which is the maximum of the heights of internal vertices. I will denote the vector with the distances $\mathbf{d}_T(0, \bullet)$ omitted by $\mathbf{u}_T \in \mathbb{R}^{\binom{n}{2}}$. Vectors that arise in this way are also called *ultrametrics*, and they are characterized by the 3-point condition, which is defined below.

**Theorem 1.1.12** (Three Point Condition [29]). *A vector $\mathbf{u} \in \mathbb{R}^{\binom{n}{2}}$ is equal to $\mathbf{u}_T$ for some equidistant tree $T$ on $n$ species leaves (i.e., is an ultrametric) if and only if $\max\{u_{ij}, u_{ik}, u_{jk}\}$ is achieved at least twice for all distinct leaves $i, j, k \in [n]$.*

**Definition 1.1.13.** The *space of ultrametrics* is the is the collection of all ultrametrics $\mathbf{u} \in \mathbb{R}^{\binom{n}{2}}$; equivalently, it is the collection of equidistant trees on $n$ species leaves

$$\mathcal{U}_n^{\text{tr}} :=\{\mathbf{u} \in \mathbb{R}^{\binom{n}{2}} \mid \mathbf{u} \text{ is an ultrametric}\}$$
$$=\{\mathbf{u} \in \mathbb{R}^{\binom{n}{2}} \mid \mathbf{u} = \mathbf{u}_T \text{ for some equidistant tree } T \text{ rooted at } 0\}.$$

Biologically, equidistant trees follow a *molecular clock* hypothesis, which assumes that genetic mutation occurs at a roughly constant rate across species and over time [27, 25, 30].

9

This assumption is especially useful when studying viral and bacterial evolution over short periods of time, and it was used early in the COVID-19 pandemic to identify the origins of various strains of COVID-19 [53].

## 1.2 Convexity

In this subsection briefly introduce objects from convex geometry that play key roles in Chapters 2 and 3. The majority of the material in this section comes from the preprint "The tropical polytope is the set of all weighted tropical Fermat-Weber points" [13], which is joint work with Mark Curiel.

A set $S \in \mathbb{R}^n$ is (classically) *convex* if for every pair of points $x, y$ in $S$, the line segment between $x$ and $y$ also lies in $S$. We are particularly interested in convex sets with linear boundaries.

A *polyhedron* $P \subset \mathbb{R}^n$ is an intersection of finitely many half spaces. We call $P$ a *polytope* if this intersection is bounded. In this case, we write $P = \text{conv}(A)$ to mean $P$ is the convex hull of some finite set $A \subset \mathbb{R}^n$. For instance, the standard simplex is the convex hull of the $n$ standard basis vectors in $\mathbb{R}^n$ and we denote it by $\Delta^{n-1}$. A *face* of a polytope $P$ is the collection of points in $P$ that minimizes the dot product with a fixed vector $u$, specifically

$$\text{face}_{\mathbf{u}}(P) := \{\mathbf{x} \in P \mid \mathbf{u} \cdot \mathbf{x} \leq \mathbf{u} \cdot \mathbf{y}, \forall \mathbf{y} \in P\}. \tag{1.1}$$

A *polyhedral complex* is a collection of polyhedra $S = \{C_i\}$ with the following properties: (1) $S$ is closed under taking faces, and (2) $C_i \cap C_j$ is a face of both $C_i$ and $C_j$, or is empty. The *normal cone to $F$ in $P$*, denoted $\sigma_F$, is the set of all vectors $u \in (\mathbb{R}^m)^*$ such that $\text{face}_{\mathbf{u}}(P) \supseteq F$. Alternatively, $\sigma_F$ is the closure of $\{u \in (\mathbb{R}^m)^* \mid \text{face}_{\mathbf{u}}(P) = F\}$. The *normal fan* of a polytope $P$ is the collection of cones $\{\sigma_F \mid F \text{ is a face of } P\}$.

For the remainder of this section we are concerned with subdivisions of polytopes. Consider $m$ polytopes $P_1, \ldots, P_m \subset \mathbb{R}^n$ where $P_i = \text{conv}(A_i)$ for some finite sets $A_i \subset \mathbb{R}^n$. A *subdivision* of a polytope $P$ is a polyhedral complex $S = \{C_i\}$ such that $P = \bigcup_i C_i$.

**Definition 1.2.1.** Given a polytope $P = \text{conv}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$, and weight $w_i \in \mathbb{R}$ on $\mathbf{v}_i$, the *lift* of $P$ with respect to $\mathbf{w}$ is

$$\tilde{P} := \text{conv}\{(\mathbf{v}_i, w_i) \mid i = 1, \ldots, k\} \subset \mathbb{R}^{n+1} \tag{1.2}$$

A face $F = \text{face}_{\mathbf{u}}(\tilde{P})$ of $\tilde{P}$ is an *upper face* if $u_{n+1} < 0$. The *regular subdivision of $P$ with respect to $\mathbf{w}$* is the projection of the upper faces of $\tilde{P}$ onto $P$ (forgetting the last coordinate).

An example is given in Example 1.3.11.

**Definition 1.2.2.** The *normal complex* of a regular subdivision is the projection of the cones of the normal fan of $\tilde{P}$ that are normal to upper faces of $\tilde{P}$.

**Notation:** A subdivision of $P$ will be denoted $\underline{P}$. Regular subdivisions induced by a piecewise-linear convex function $\lambda$ will be denoted $\underline{P}_\lambda$.

## 1.3 Tropical Geometry

Tropical geometry is the study of the polyhedral skeletons of algebraic varieties. While algebraic geometry is the study of solutions to polynomial equations, tropical geometry is the study of regions of non-linearity of piece-wise linear functions, which are defined over a tropical semi-ring. This section uses material from [12] and [13].

### 1.3.1 Tropical Arithmetic

There are two tropical semi-rings which I will use in this thesis. In the *tropical max-plus semi-ring* $\mathbb{R}_+ := (\mathbb{R} \cup \{-\infty\}, \oplus, \odot)$, tropical addition $\oplus$ and tropical multiplication $\odot$ are defined to be max and $+$ respectively

$$a \oplus b = \max\{a, b\}, \quad a \odot b = a + b \quad \text{where } a, b \in \mathbb{R} \cup \{-\infty\}.$$

The max-plus multiplicative identity is 0, and the additive identity is $-\infty$. The *tropical min-plus semi-ring* is $\mathbb{R}_- := (\mathbb{R} \cup \{+\infty\}, \oplus, \odot)$, where $\oplus = \min$ and $\odot = +$. The min-plus multiplicative identity is also 0, and the min-plus additive identity is $+\infty$. The max- and min-plus semi-rings are equivalent up to a sign, but for certain applications it is more convenient to use one or the other. For example, when taking the tropical average of two trees, the max-plus semi-ring is the most natural; when tropicalizing an algebraic variety, it is sometimes easier use the min-plus semi-ring. The max-plus semi-ring will be the default when max/min is unspecified.

These tropical operations can be extended component-wise to the tropical projective torus, $\mathbb{R}^n/\mathbb{R}\mathbb{1}$. For vectors $\mathbf{v}, \mathbf{u} \in \mathbb{R}^n/\mathbb{R}\mathbb{1}$, the notation $\mathbf{u} \oplus \mathbf{v}$ and $\mathbf{u} \odot \mathbf{v}$ denotes component-wise max and addition, respectively. Tropical scalar multiplication of a vector amounts to adding a (classical) multiple of the all ones vector $\mathbb{1}_n$, namely $\lambda \odot \mathbf{v} = \lambda \mathbb{1} + \mathbf{v} = (\lambda + v_1, \ldots, \lambda + v_n)$ for any $\lambda \in \mathbb{R}$.

**Example 1.3.1.** If $\mathbf{v}_1 = (1, 2, -3)$ and $\mathbf{v}_2 = (-5, 3, 2)$ are points in $\mathbb{R}^3/\mathbb{R}\mathbb{1}$, then

$$
\begin{aligned}
\mathbf{v}_1 \oplus \mathbf{v}_2 &= (\max\{1, -5\}, \max\{2, 3\}, \max\{-3, 2\}) \\
&= (1, 3, 2) = (-1, 1, 0) \\
\mathbf{v}_1 \odot \mathbf{v}_2 &= (1 - 5, 2 + 3, -3 + 2) \\
&= (-4, 5, -1) \\
3 \odot v_2 &= (3 - 5, 3 + 3, 3 + 2) \\
&= (-2, 6, 5)
\end{aligned}
$$

When I draw pictures in $\mathbb{R}^n/\mathbb{R}\mathbb{1}$, I will tropically scale points to have last coordinate zero, then project away the last coordinate and draw the point in $\mathbb{R}^{n-1}$. For example, $(3, 1, 2) \equiv (1, -1, 0)$ will be drawn in the plane at the location $(1, -1)$.

## 1.3.2 Tropical Polynomials

Let $\mathbf{x}, \mathbf{a} \in \mathbb{R}^n$, and define the tropical monomial: $\mathbf{x}^{\mathbf{a}} := \bigoplus_{i=1}^n a_i \odot x_i$. Note that when the entries of $\mathbf{a}$ are non-negative integers, we have

$$
x_1^{a_1} \odot \cdots \odot x_n^{a_n} = \underbrace{x_1 \odot \cdots \odot x_1}_{a_1 \text{ times}} \odot \cdots \odot \underbrace{x_n \odot \cdots \odot x_n}_{a_n \text{ times}},
$$

which explains the notation. Note that for $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{x}^{\mathbf{a}}$ is still a well-defined tropical function, but not a tropical polynomial.

**Definition 1.3.2.** Let $\mathbf{x} \in \mathbb{R}^n/\mathbb{R}\mathbb{1}$. A *tropical signomial in* $\mathbf{x}$ is a finite linear combination of tropical monomials, i.e.

$$
f(\mathbf{x}) = \bigoplus_{\mathbf{a} \in A} \lambda_{\mathbf{a}} \odot \mathbf{x}^{\mathbf{a}}
$$

where $A \subset \mathbb{R}_{\geq 0}^n$ is finite and $\lambda_{\mathbf{a}} \in \mathbb{R}$ If $A \subset \mathbb{Z}_{\geq 0}^n$, then $f(\mathbf{x})$ is a *tropical polynomial*.

**Example 1.3.3.** Let $f(x) = 1 \oplus 3 \odot x \oplus -1 \odot x^{\sqrt{2}}$. In terms of classical arithmetic operations:

$$
f(x) = \max\{1, 3 + x, -1 + \sqrt{2}x\}.
$$
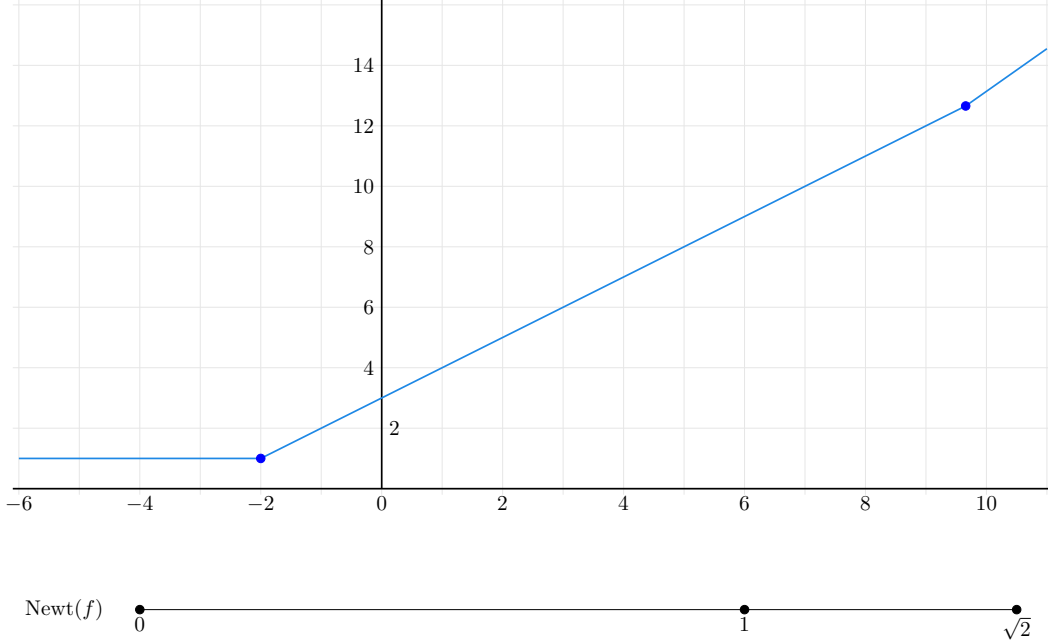
Figure 1.5: The graph of $f(x) = 1 \oplus 3 \odot x \oplus -1 \odot x^{\sqrt{2}}$. The connection to the Newton polytope of $f$ is explained in section 1.3.3.

The graph of $f(x)$ is depicted in fig. 1.5; it has three linear pieces:

$$f(x) = \begin{cases} 1 & \text{if } x \leq -2, \\ 3 + x & \text{if } -2 \leq x \leq \frac{4}{\sqrt{2}-1}, \\ -1 + \sqrt{2}x & \text{if } \frac{4}{\sqrt{2}-1} \leq x. \end{cases}$$

### 1.3.3 Tropical Varieties

I begin by defining the tropical hypersurface of a tropical polynomial, and briefly describe tropical varieties more generally. Then, I explain how a tropical hypersurface is encoded by a subdivision of the Newton polytope of its defining polynomial. Later, in Section 1.3.5, I introduce two relevant examples of tropical varieties: the tropical Grassmannian and the space of ultrametrics.

**Definition 1.3.4.** The *tropical vanishing set* or *tropical hypersurface* of a tropical signomial $f = \bigoplus_i c_i \odot x^{\alpha_i}$, denoted $\text{tropV}(f)$, is the set of $x \in \mathbb{R}^n$ for which the max in $f(x)$ is achieved at least twice,

$$\text{tropV}(f) = \{x \in \mathbb{R}^n \mid \max \text{ in } f(x) \text{ is achieved at least twice}\}. \tag{1.3}$$

13

A *tropical variety* is an intersection of finitely many tropical hypersurfaces.

*Remark* 1.3.5. *Tropicalization* is a process that transforms polynomials and algebraic varieties (i.e., the solutions to a finite set of polynomial equations) into their tropical counterparts. Under tropicalization, polynomials become piece-wise linear functions (tropical polynomials) and algebraic varieties become polyhedral complexes (tropical varieties). [In this thesis I will only describe the tropical hypersurfaces and varieties using tropical polynomials]. For details on the process of tropicalization see [29, §3].

When $f$ is a product of tropical polynomials $f_i$, tropV($f$) is the union of the tropical hypersurfaces tropV($f_i$). A more general result is the following.

**Lemma 1.3.6.** *The tropical vanishing set of a product of real positive powers of polynomials, $f_i^{w_i}$, is (as a set) the union of tropical vanishing sets of the $f_i$. That is,*

$$trop\,V\left(\bigodot_{i=1}^{m} f_i^{w_i}\right) = \bigcup_{i=1}^{m} trop\,V(f_i), \ \ for \ w_i > 0.$$

*Proof.* Let $f_i = \bigoplus_{\alpha \in A_i} c_\alpha \odot \mathbf{x}^\alpha$, $A_i \subset \mathbb{R}_{>0}^n$. If the maximum in $f_j$ is achieved twice by $\mathbf{x}$, then $f_j^{w_j}(\mathbf{x}) = w_j(c_{\alpha_1} + \alpha_1 \cdot \mathbf{x}) = w_j(c_{\alpha_2} + \alpha_2 \cdot \mathbf{x})$ for some $\alpha_1 \neq \alpha_2 \in A_j$. It follows that the maximum in $\bigodot_{i=1}^{m} f_i^{w_i}$ is also achieved at least twice:

$$\bigodot_{i=1}^{m} f_i^{w_i} = w_j(c_{\alpha_1} + \alpha_1 \cdot \mathbf{x}) + \sum_{i \neq j} f_i^{w_i}(\mathbf{x}) = w_j(c_{\alpha_2} + \alpha_2 \cdot \mathbf{x}) + \sum_{i \neq j} f_i^{w_i}(\mathbf{x}).$$

On the other hand, if the maximum is achieved twice in $\bigodot_{i=1}^{m} f_i^{w_i}$ at $\mathbf{x}$, then we must be able to write the maximum as two distinct sums: $\sum_{i=1}^{m} w_i(c_{\alpha_i^1} + \alpha_i^1 \cdot \mathbf{x}) = \sum_{i=1}^{m} w_i(c_{\alpha_i^2} + \alpha_i^2 \cdot \mathbf{x})$, where $\alpha_i^1, \alpha_i^2 \in A_i$. It follows that for some $j \in [m]$, $\alpha_j^1 \neq \alpha_j^2$, so the maximum in $f_j^{w_j}$ is achieved at least twice. $\qquad \square$

A tropical hypersurface is a polyhedral complex, and it can be understood combinatorially in terms of the Newton polytope of $f$, defined below.

**Definition 1.3.7** (Newton polytope). Suppose $f(\mathbf{x}) = \bigoplus_{\mathbf{a} \in A} c_\mathbf{a} \odot \mathbf{x}^\mathbf{a}$ is a multivariate polynomial for some finite $A \subset \mathbb{R}^n$ and $c_\mathbf{a} \in \mathbb{R}$. The *support* of $f(\mathbf{x})$ is the set, denoted supp($f$), containing all $\mathbf{a} \in A$ such that $c_\mathbf{a} \neq -\infty$. The *Newton polytope* Newt($f$) of the polynomial $f(\mathbf{x})$ is the convex hull of its support, i.e. Newt($f$) = conv(supp($f$)). If $f, g$ are polynomials, then Newt($fg$) = Newt($f$) + Newt($g$).

**Proposition 1.3.8** (Proposition 3.1.6, [29]). *Given a tropical signomial $f = \sum_i c_i x^{\alpha_i}$, let $\underline{N}_f$ denote the regular subdivision of Newt($f$) induced by the weighting $w(a_i) = c_i$. Then trop V($f$) is the codimension-1 skeleton of the normal complex of $\underline{N}_f$.*

14
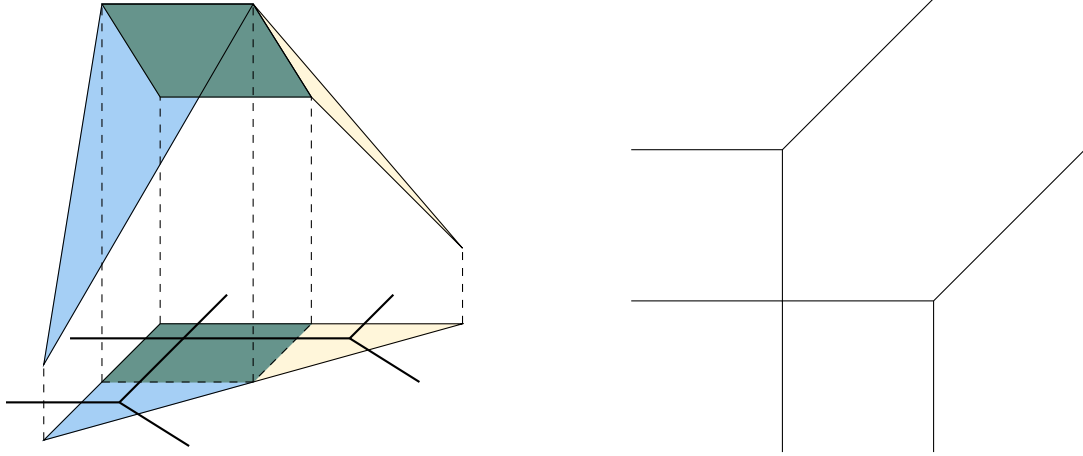
Figure 1.6: Left: A lift of $2\Delta^2 = \text{Newt}(f)$ with weights given by the coefficients of $f$, overlaid with $\text{tropV}(f)$ in black; right: $\text{tropV}(f)$.

*Remark* 1.3.9. Although Proposition 3.1.6 in [29] is originally stated for tropical polynomials, the arguments clearly hold for tropical signomials as well.

**Definition 1.3.10.** Given a tropical polynomial $f$, let $\underline{N}$ be the regular subdivision of $\text{Newt}(f)$ induced by the coefficients of $f$. The *normal complex of $f$* is normal complex of $\underline{N}$; it is a subdivision of $\mathbb{R}^n$.

**Example 1.3.11.** The following is an example of Proposition 1.3.8. Let

$$f = x^2 \oplus 4 \odot xy \oplus 3 \odot xz \oplus y^2 \oplus 4 \odot yz \oplus 3 \odot z^2. \tag{1.4}$$

Figure 1.6 depicts $\text{tropV}(f)$, the subdivision of the Newton polytope dual to it, and the lift of the Newton polytope that induces that subdivision.

## 1.3.4 Tropical Convexity

**Definition 1.3.12.** The *min-tropical convex hull* (or *max-tropical convex hull*) of a set of points $A \subset \mathbb{R}^n/\mathbb{R}\mathbb{1}$, denoted $\text{tconv}^{\min}(A)$ (respectively, $\text{tconv}^{\max}(A)$), is the set of all tropical linear combinations of points in $A$, that is,

$$\text{tconv}(A) := \{\lambda_1 \odot \mathbf{a}_1 \oplus \cdots \oplus \lambda_k \odot \mathbf{a}_k \mid \lambda_i \in \mathbb{R}, \mathbf{a}_i \in A, k \in \mathbb{N}\}. \tag{1.5}$$

If $A$ is a finite set, then $\text{tconv}(A)$ is called a *tropical polytope.*

Note that the tropical convex hull is independent of the representatives in $\mathbb{R}^n/\mathbb{R}\mathbb{1}$ we

choose for the points in $A$. That is, if $\mathbf{v}'_i = \alpha_i \odot \mathbf{v}_i$, $\lambda_i \in \mathbb{R}$, then with $\lambda'_i = \lambda_i - \alpha_i$

$$\lambda_1 \odot v_1 \oplus \cdots \oplus \lambda_m \odot v_m = (\lambda_1 - \alpha_1) \odot v'_1 \oplus \cdots \oplus (\lambda_m - \alpha_m) \odot v_m = \lambda'_1 \odot v'_1 \oplus \cdots \oplus \lambda'_m \odot v'_m.$$

**Example 1.3.13.** Let $v_1 = (0, 0, 0)$, and $v_2 = (1, -1, 0)$. The tropical line segment $\text{tconv}(v_1, v_2)$ consists of points of the form $\lambda_1 \odot v_1 \oplus \lambda_2 \odot v_2$, and $\text{tconv}^{\min}(v_1, v_2)$ is illustrated in Figure 1.7. The tropical polytope $\text{tconv}^{\min}(v_1, v_2)$ consists of three points, connected by two classical line segments.

$$
\begin{array}{c}
\lambda_1 - \lambda_2 \geq 1 \rightarrow \\
1 > \lambda_1 - \lambda_2 > 0 \rightarrow \\
\lambda_1 - \lambda_2 = 0 \rightarrow
\end{array}
$$
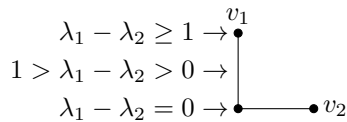
Figure 1.7: The tropical line segment between $v_1 = (0, 0, 0)$ and $v_2 = (1, -1, 0)$.

A subset $S$ of the tropical projective torus is *tropically convex* if the tropical line segment between any two points of $S$ is fully contained in $S$. Tropical line segments are discussed in greater detail in Chapter 2.

## 1.3.5   Tropical Phylogenetic Tree Space

In Section 1.1.4 I defined BHV space and the space of equidistant trees. These spaces, when embedded in the tropical projective torus, are also tropical varieties. The tropical space of equidistant trees is particularly interesting because it is tropically convex.

The (rank 2) tropical Grassmannian, $\text{tropGr}^\circ(2, n)$, encodes unrooted trees on $n$ leaves. Recall that a point $x$ in $\text{Gr}(2, n)$ corresponds to a rank 2 linear subspace of $\mathbb{P}^n$, which may be represented by a $2 \times n$ matrix $M_x$. The $2 \times 2$ minors of $M_x$, $p_{ij}$, are called the *Plücker coordinates* of $x$, and they give an embedding of the Grassmannian into a large projective space, $\mathbb{P}^{\binom{n}{2}}$. The subset $\text{Gr}^\circ(2, n)$ consists of all points $x \in \text{Gr}(2, n)$ all of whose Plücker coordinates are non-vanishing.

The *Plücker relations* are polynomial equations in the $p_{ij}$ that cut out this embedding of the Grassmannian. The *tropical Plücker relations* are

$$(p_{ij} \odot p_{kl}) \oplus (p_{ik} \odot p_{jl}) \oplus (p_{il} \odot p_{jk}), \text{ for all distinct } i, j, k, l \in [n].$$

This is exactly the 4-point condition (see Theorem 1.1.8). Thus, the tropical Grassmannian (which is tropically cut out by the tropical Plücker relations) is the tropical embedding of

16

the space of unrooted phylogenetic trees [42]. The following example shows that the tropical Grassmannian is not tropically convex.

**Example 1.3.14.** The two vectors below satisfy the 4-point condition. However, their tropical sum does not.

$$(u_{12}, u_{34}, u_{13}, u_{24}, u_{14}, u_{23})$$
$$A = (3, 9, 8, 10, 9, 9)$$
$$B = (10, 8, 3, 9, 9, 9)$$
$$A \oplus B = (10, 9, 8, 10, 9, 9).$$

The tropical embedding of the space of equidistant trees arises as the tropicalization of the complete graph, $\mathcal{K}_n$. The tropicalization of a graph $\mathcal{G}$ lies in $\mathbb{R}^{|E(\mathcal{G})|}/\mathbb{R}\mathbb{1}$ and is cut out by the tropical polynomials of the form $\bigoplus_{e \in C} x_e$, where $C$ is a cycle of $\mathcal{G}$. These tropical polynomials are the *tropical circuit relations* of $\mathcal{G}$. For more details on this construction, see [29, §4]. In a complete graph, every triangle is a circuit. Therefore, the tropical circuit relations of $\mathcal{K}_n$ are

$$x_{ij} \oplus x_{jk} \oplus x_{ik}, \text{ for all distinct } i, j, k \in [n].$$

The above relations exactly describe the 3-point condition, so $\mathrm{trop}(\mathcal{K}_n)$ is the space of ultrametrics (i.e., equidistant phylogenetic trees) [1]. I will denote this space by $\mathcal{U}_n$.

Tropical varieties that arise from graphs in the way described above are *linear* tropical varieties, and therefore, they are always tropically convex [29, Theorem 5.2.8]. For the space of ultrametrics, it is also possible to give an explicit argument for tropical convexity using the tropical circuit relations of $\mathcal{K}_n$.

**Theorem 1.3.15** ([29]). *The space of ultrametrics is (max)-tropically convex.*

*Proof.* A space is tropically convex if it contains the tropical line segment between any two points in the space. I will show that for any ultrametrics, $u, v \in \mathcal{U}_n$, the space $\mathcal{U}_n$ contains $u \oplus v$; this implies that $\mathcal{U}_n$ contains the entire tropical line segment $\mathrm{tconv}(u, v)$. By definition,

$$u \oplus v = (\max\{u_{ij}, v_{ij}\})_{ij \in \binom{[n]}{2}}.$$

We want to prove that this vector satisfies the 3-point condition, i.e. we want to show that for distinct $i, j, k \in [n]$, $\max\{\max\{u_{ij}, v_{ij}\}, \max\{u_{ik}, v_{ik}\}, \max\{u_{jk}, v_{jk}\}\}$ is achieved at least twice. Since $u$ itself satisfies the 3-point condition, suppose without loss of generality that $u_{ij} = u_{ik} \geq u_{jk}$. If $\max\{v_{ij}, v_{ik}, v_{jk}\} \leq u_{ij}$, then $\max\{u_{ij}, v_{ij}\} = \max\{u_{ik}, v_{ik}\} = u_{ij} = $

$u_{ik} \geq \max\{u_{jk}, v_{jk}\}$ and we are done. On the other hand, if $\max\{v_{ij}, v_{ik}, v_{jk}\} \geq u_{ij} = \max\{u_{ij}, u_{ik}, u_{jk}\}$, then by a symmetric argument we are also done. $\qquad\square$

*Remark* 1.3.16. In terms of trees, tropical shifting by $\lambda\mathbb{1}$ means adding $\lambda/2$ to the length of each edge that is adjacent to a leaf. Thus, lengths of edges adjacent to leaves are no longer well-defined.

*Remark* 1.3.17. Another fan structure on the tropical Grassmannian is described by Ardila and Klivans in [1]. It is finer than the fan structure we described above; two trees $u$, $v$ lie in the same cone of the finer structure and have the same *ranked tree topology* if they are combinatorially isomorphic, and the chronological order of all $n-1$ internal vertices is the same in both trees, or equivalently if

$$\mathrm{argmax}_u^k \binom{[n]}{2} = \mathrm{argmax}_v^k \binom{[n]}{2}, \text{ for all } k \in \mathbb{N},$$

where $\mathrm{argmax}_u^k(S)$ denotes the elements of $S$ where $u$ achieves its $k$th largest value over elements in $S$.

Note that in Figure 2.2, the second tree, $w = u \oplus -1 \odot v$, lies in a codimension 1 cone in the fine fan structure (since the tree is binary branching, but two internal vertices are at the same height). More precisely, the line segment between $w$ and $u \oplus v$ (the second and third trees in Figure 2.2) lies in the intersection of two cells representing distinct **ranked** tree topologies: $w_{12} \leq w_{34} \leq w_{13} = w_{14} = w_{23} = w_{24}$ and $w_{34} \leq w_{12} \leq w_{13} = w_{14} = w_{23} = w_{24}$.

The number of turning points is close to the NNI distance, because each turning point sits in a coarse cone of codimension at most two, hence uses at most three NNI moves. In the finer fan structure, the tropical line segment can pass through cones of codimension up to $n/2$, which indicates that many *ranked* NNI moves may be required at a single turning point. Additionally, in the coarse fan structure, the tropical line segment only intersects non-maximal cones at points, but in the finer fan structure, the tropical line segment may intersect non-maximal cones in a whole classical line segment.

## 1.4   Statistical Models

In this section, I begin by defining models and then introduce two relevant examples: *Gaussian models* and *Brownian motion tree models*, the latter of which are the topic of Chapter 4. Afterwards, I define the *maximum likelihood estimator*, which is a method for fitting a probability distribution to observed data. The primary reference for the general material on models is *Algebraic Statistics* by Seth Sullivant [47]; for Brownian motion tree models a

reference is [21].

A *statistical model*, $\mathcal{M}$, is a collection of probability distributions or density functions which all share some underlying assumptions. Models may be explicitly parameterized, or they may be defined implicitly, for example by specifying independence constraints on a collection of random variables. Brownian motion tree models, and linear Gaussian models more generally, are parameterized, so this introduction will be limited to parameterized models, which are defined formally below.

**Definition 1.4.1** (Definition 5.1.1, [47]). A *(parametric) statistical model* is a mapping, $\theta \mapsto p_\theta$, from a finite dimensional parameter space $\Theta \subseteq \mathbb{R}^d$ to a space of probability distributions or density functions.

One example of a model is the *binomial random variable model*, defined below; this will be the running example throughout this section.

**Example 1.4.2** (Binomial variable with $r$ trials. Example 5.1.2, [47]). Let $X$ be a random variable that records the number of successes in $r$ trials of an experiment with two possible outcomes – "success" and "failure". The *binomial random variable model* consists of all probability distributions of the form

$$P_\theta(X = i) = \binom{r}{i} \theta^i (1 - \theta)^{r-i}.$$

The model has a single parameter, $\theta$, which is the probability of success; the parameter space of the model is $\Theta = [0, 1]$.

## 1.4.1 Gaussian Models

A Gaussian probability distribution on $m$ variables has two parameters: the mean, $\mu \in \mathbb{R}^m$, and the covariance matrix, $\Sigma \in \mathrm{PD}_m(\mathbb{R})$, where $\mathrm{PD}_m(\mathbb{R})$ denotes the set of all real positive definite symmetric matrices. A vector of random variables, $\mathbf{x}$, which is distributed according to a normal distribution with mean $\mu$ and covariance $\Sigma$, is denoted $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$. The probability density function of for a vector of jointly distributed Gaussian variables is then given by the following function

$$Q_{\Sigma,\mu}(x_1, \ldots, x_m) := \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp\left( -\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu) \right). \qquad (1.6)$$

A *Gaussian model* is a collection of Gaussian probability distributions. For the Gaussian models in this thesis, every probability distribution in the model is mean zero (i.e. $\mu = 0$).

Such Gaussian models are frequently defined by specifying polynomial constraints on their covariance matrices or concentration matrix (i.e., inverse covariance matrix). In this case, the model is referred to as a *covariance model* (resp. *concentration model*). Since a mean zero Gaussian model is completely determined by the admissible covariance (or, concentration) matrices, I will identify it with the collection of admissible covariance (concentration) matrices. An important class of Gaussian models are *linear covariance models*.

**Definition 1.4.3.** Given a linear space $\mathcal{L} \subseteq \mathbb{R}^{m \times m}$, the associated *linear covariance model*, $\mathcal{M}_\mathcal{L}$, is the collection of mean zero Gaussian density functions whose covariance matrices lie in $\mathcal{L} \cap \mathrm{PD}_m(\mathbb{R})$,
$$\mathcal{M}_\mathcal{L} := \{Q_{\Sigma,0} : \Sigma \in \mathcal{L} \cap \mathrm{PD}_m(\mathbb{R})\}.$$

I may also simply write $\mathcal{M}_\mathcal{L} = \mathcal{L} \cap \mathrm{PD}_m(\mathbb{R})$, identifying the model with its set of covariance matrices.

Brownian motion tree models, defined in the next section, are an example of a linear covariance model.

## 1.4.2  Brownian Motion Tree Models

Brownian motion is a stochastic process that models the movement of an object subject to a large number of very small forces. While these models were motivated by understanding the movement of a small particle in air or liquid, they have also found applications in a variety of fields. In phylogenetics, Brownian motion can be used to model the evolution of real-valued continuous traits through *genetic drift*[2], which refers to the accumulation of purely random genetic mutations over time. Brownian motion tree models have been used to determine how quickly certain traits evolved, and whether a continuous trait evolved because of selective pressure or genetic drift [21, 40].

For a single variable, a Brownian motion is a family of normal distributions, $B_t$, whose variance is proportional to (time) $t$

$$B_t \sim \mathcal{N}(\mu, \sigma^2 t).$$

For a group of biological units, Brownian motion can be modeled continuously, discretely, or algebraically along a phylogenetic tree $T$. These models, called *Brownian motion tree models*, were introduced in [17]; recently, these models have been studied as algebraic statistical models [46, 4, 45].

---

[2]Brownian motion tree models can also be used to model evolution due to selective pressure, with some additional assumptions [19].

$$Y_x = Y_v + B_{\mathrm{dist}(x,v)}$$
$$Y_z = Y_v + B_{\mathrm{dist}(z,v)}$$
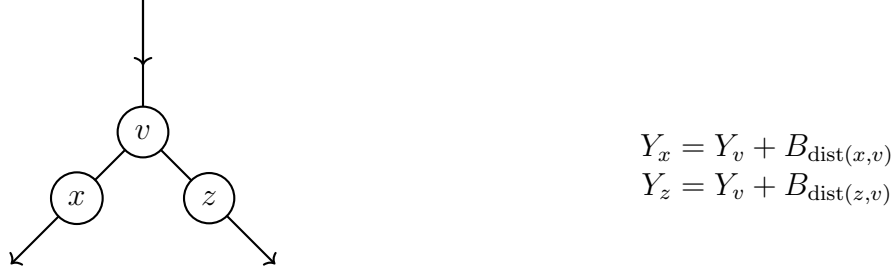
Figure 1.8: Brownian motion along a phylogenetic tree at a branch point.

In the continuous version of a Brownian motion tree model, which is illustrated in Figure 1.8, each point $x$ along the tree (including vertices and all the points on edges between them) corresponds to a random variable $Y_x$, which is normally distributed. Starting at the top, the distribution assigned to the root is 0. Moving downwards towards the leaves, the distribution evolves as a mean zero Brownian motion along each branch with variance $\sigma_T^2$ and splits into two independent mean zero Brownian motions at each internal vertex. The normal distribution assigned to an arbitrary point $z$ on the tree is normal distribution of its parent vertex, plus the Brownian motion of the current branch up to $z$. Thus, if $x, z$ are points on different edges directly below an internal vertex $v$, we have $Y_x = Y_v + B_{t_1}$, and $Y_z = Y_v + B_{t_2}$, where $t_1, t_2$ are the distances between the vertices $v$ and $x$ and between $v$ and $z$ when $T$ is considered as a metric tree. Note that $\sigma_T$ is constant over the whole tree; thus, we may only estimate the $t_i$ up to a factor of $\sigma_T^2$.

In the discrete version of a Brownian motion tree model, only the distributions at vertices (both internal and leaves) are recorded. The distribution at each vertex is a sum of mean-zero normal distributions, hence the discrete Brownian motion tree model is a vector of normal distributions, $(Y_v)_{v \in V(T)}$. The algebraic version of a Brownian motion tree model is a joint normal distribution on the leaf variables only. The following proposition computes the covariances among leaf variables.

**Proposition 1.4.4** (Proposition 2.1, [46])**.** *The random vector $(Y_1, \ldots, Y_n)$ is normally distributed with mean zero, and the entries $\sigma_{ij} = \mathrm{cov}(Y_i, Y_j)$ of its covariance matrix $\Sigma_\theta$ are*

$$\sigma_{ij} = \sum_{v \leq \mathrm{lca}(i,j)} \theta_v, \qquad \text{for } i, j = 1, \ldots, n. \tag{1.7}$$

Thus, the (algebraic) Brownian motion tree model is a linear covariance model where the covariance matrix is constrained by least common ancestry in a fixed evolutionary tree $T$.

A more precise definition is given below.

**Definition 1.4.5.** Given a rooted tree $T$ with $n + 1$ leaves (where one leaf is the root), the associated linear space $\mathcal{L}_T \subseteq \mathrm{PD}_n$ consists of $\Sigma = (\sigma_{ij})_{1 \leq i,j \leq n}$ satisfying $\sigma_{ij} = \sigma_{kl}$ if $\mathrm{lca}_T(i,j) = \mathrm{lca}_T(k,l)$.

In the example below, I provide the Brownian motion tree model for the tree in Figure 1.1.

**Example 1.4.6.** For the tree in Figure 1.1, the corresponding linear space is

$$
\mathcal{L}_T(\mathbb{R}) = \left\{ \begin{bmatrix} t_1 & t_5 & t_5 & t_6 \\ t_5 & t_2 & t_5 & t_6 \\ t_5 & t_5 & t_3 & t_6 \\ t_6 & t_6 & t_6 & t_4 \end{bmatrix} : t_1, \ldots, t_6 \in \mathbb{R} \right\}.
$$

The Brownian motion tree model for $T$ is $\mathcal{M}_T = \mathcal{L}_T(\mathbb{R}) \cap \mathrm{PD}_4$.

### 1.4.3 Maximum Likelihood Estimation

Given a model $\mathcal{M}$ and independent, identically distributed data (iid) data set $S$, we may ask:

**Question 1.4.7.** Which model in $\mathcal{M}$ maximizes the likelihood of $S$ being observed?

A distribution that answers this question is called a *maximum likelihood estimator (MLE)*. The assumption that $S$ is independent and identically distributed means that the probability, or likelihood, of observing $S$ is the product of probabilities of observing each data point in $S$. Assuming the probability density function for $\mathbf{x}$ is $p_\theta$, we denote this likelihood by $L(\theta|S)$. That is,

$$
L_{\mathcal{M}}(\theta|S) := \prod_{\mathbf{x} \in S} p_\theta(\mathbf{x}).
$$

It is often easier to work with the log-likelihood function, denoted $\ell(\theta|S)$ and defined below, which has extrema at the same locations as $L(\theta|S)$.

$$
\ell_{\mathcal{M}}(\theta|S) = \log L_{\mathcal{M}}(\theta|S) = \sum_{i=1}^{k} \log L_{\mathcal{M}}(\theta|t_i). \tag{1.8}
$$

Given a parameterized model $\mathcal{M}$, the problem of computing the maximum likelihood estimator is equivalent to finding the parameters $\theta$ that maximize $L_{\mathcal{M}}(\theta|S)$ (or $\ell_{\mathcal{M}}(\theta|S)$) when $S$ is fixed. This can be done by finding the points where all $\theta$ partial derivatives

of $\ell(\theta|S)$ are set to zero. These partial derivatives are called the *score equations*. The maximum likelihood estimates are critical points of the likelihood function, so the number of critical points is an algebraic measure of the complexity of the problem, called the *maximum likelihood degree*.

**Definition 1.4.8.** The *maximum likelihood degree* of a model, $\mathcal{M}$, is the number of complex critical points of $\ell(\theta|S)$ over $\mathcal{M}$.

*Remark* 1.4.9. Although linear Gaussian models usually require real, positive definite covariance matrices, the ML degree usually refers to the number of critical points to the likelihood function over the complex linear space model, intersected with invertible matrices $\mathcal{L}(\mathbb{C}) \cap \mathrm{GL}_n(\mathbb{C})$. This is because tools from algebraic geometry are naturally suited to compute all complex solutions, and the condition that a matrix is invertible is easier to enforce than positive definite. However, an MLE is always a real, positive definite matrix, if it exists. In Chapter 4, we compute this modified ML degree.

Chapter 3 of this thesis is concerned with computing the maximum likelihood degree for Brownian motion tree models. Below is an example where the maximum likelihood degree and estimate are computed for the running example.

**Example 1.4.10** (MLE of a binomial random variable. Example 5.3.6, [47]). Let $\mathcal{M}_r$ be the model of $r$ trials of a binomial random variable. Each probability distribution in $\mathcal{M}_r$ takes the form $p_\theta(i) = \binom{r}{i}\theta^i(1-\theta)^{r-i}$ for some $\theta \in (0,1)$. The domain $p_\theta$ is $\{0\} \cup [r]$, which represents the number of successes among the $r$ trials. The sample is a vector of counts $\mathbf{s} = (s_i)$, where $s_i$ is the number of times the experiment resulted in $i$ successes for $i = 0, \ldots, r$. The likelihood function is

$$L(\theta|\mathbf{s}) = \prod_{i=0}^{r} p_\theta(i)^{s_i} = \prod_{i=0}^{r} \left( \binom{r}{i}\theta^i(1-\theta)^{r-i} \right)^{s_i}.$$

And the log-likelihood function is

$$\ell(\theta|\mathbf{s}) = \sum_{i=0}^{r} \left( \log \binom{r}{i} + s_i \left( i \log \theta + (r-i) \log(1-\theta) \right) \right).$$

Thus, the (one) score equation for this model is

$$\frac{\partial \ell}{\partial \theta} = \sum_{i=0}^{r} \frac{i s_i}{\theta} - \sum_{i=0}^{r} \frac{(r-i)s_i}{1-\theta} = 0.$$

In this example, the only critical point is at $\widehat{\theta} = \frac{\sum_{i=0}^{r} i s_i}{rn}$, where $n$ is the sum of the entries of $\mathbf{s}$. Hence the ML degree is one, and the MLE is the distribution $p_{\widehat{\theta}}(\mathbf{x})$. Note the the optimal parameter $\widehat{\theta}$ is the sample probability of success, i.e. the total number of successes in the sample divided by the total number of trials.

## 1.5 Computational Algebraic Geometry

A classical problem in algebraic geometry is to compute the number of solutions to a system of $n$ homogeneous polynomials in $n+1$ variables. For a statistical model, $\mathcal{M}$, computing the ML degree means finding the number of solutions to the system of score equations over $\mathcal{M}$. In this section, I introduce the tools from algebraic geometry needed to carry out this computation for Brownian motion tree models, where the score equations are rational. First, I explain how to transform the rational system of score equations into a polynomial system of equations by clearing denominators and saturating. In the following subsections, I introduce Bézout's Theorem and standard bases, which we use in Chapter 4 to compute the solutions to this new polynomial system.

Given a system of rational functions, $\bar{F} = \{p_1(\mathbf{x})/q_1(\mathbf{x}), \ldots, p_k(\mathbf{x})/q_k(\mathbf{x})\}$, a *solution* to the system is a point $\mathbf{x}$ so that $p_i(\mathbf{x}) = 0$ and $q_i(\mathbf{x}) \neq 0$ for $i = 1, \ldots, k$. Let $F$ be denote the system after clearing denominators, i.e., $F = \{p_1(\mathbf{x}), \ldots, p_k(\mathbf{x})\}$. $F$ is a polynomial system. The solutions to $\bar{F}$ are contained in $V(F)$, but $V(F)$ is generally bigger. The excess solutions in $V(F)$ are points where $q_i$ vanishes for some $i$. Thus, it is possible to find the solutions to $\bar{F}$ by first finding solutions to $F$, and then removing solutions to the $q_i$. That is,

$$\text{solutions}(\bar{F}) = V(F) \setminus V(q_1, \ldots, q_k).$$

### 1.5.1 Bézout's Theorem

Consider the following statements about intersections of algebraic curves in the affine plane:

1. Any two distinct lines intersect in one point, unless they are parallel.

2. A parabola and a line intersect in 0, 1, or 2 points.

Three assumptions allow us to simplify the statements above. First, we compute intersections in a topologically compact space, like $\mathbb{P}^n$. This addresses the issue in the first statement – in $\mathbb{P}^n$, parallel lines intersect "at infinity", so any two lines intersect in exactly one point. Next, we work over an algebraically closed field, like $\mathbb{C}$, to avoid the issue that arises for the first picture in Figure 1.9. Finally, we count with *multiplicity*. Multiplicity is a rigorous
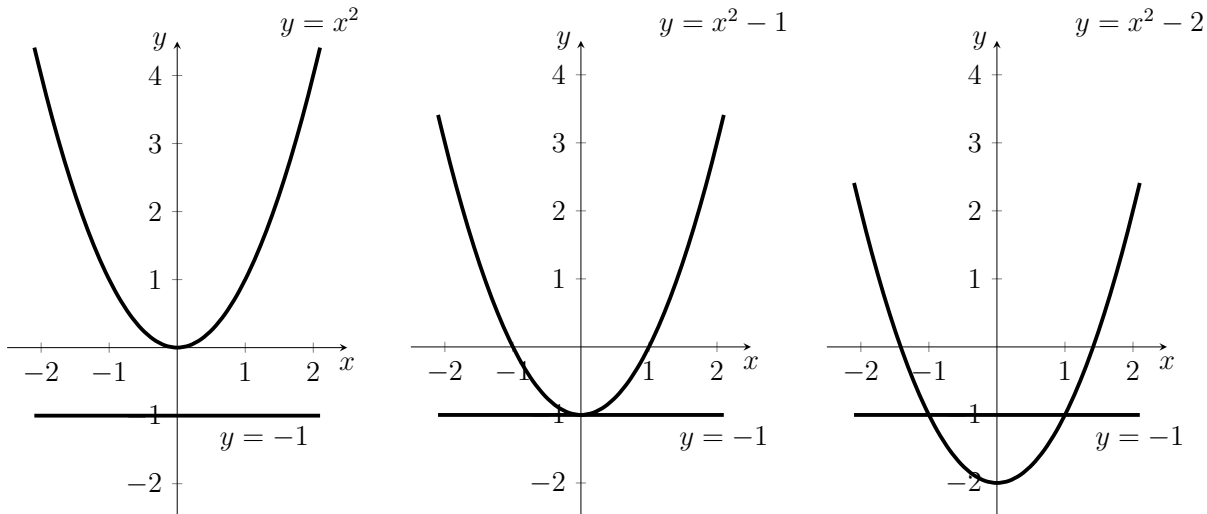
Figure 1.9: Parabolas intersecting the line $y = -1$ in 0, 1, or 2 points.

way of assigning an integer weight to each intersection point, which addresses the middle picture in Figure 1.9. In the parabola and line intersection, the multiplicity arises intuitively by studying a family of parabolas $f_t(x) = x^2 + t$. For $t < -1$, $f_t(x) = 0$ intersects the line $g(x) = -1$ in exactly two points, which we will call $a$ and $b$. As $t$ increases, the parabola moves upwards, and as $t$ approaches $-1$, the points $a$ and $b$ come together. Thus, the single intersection point of $f_{-1}(x) = x^2 - 1$ and $g(x) = -1$ should have 2, since the two points $a$ and $b$ collided.

With these additional assumptions, we can simplify the intersection statements as follows:

1. Any two distinct lines intersect in exactly one point.

2. A parabola and a line intersect in exactly two points.

These additional assumptions permit statements that talk about the number of intersection points without knowing the coefficient of the polynomials involved nor ever explicitly solving for the intersection points. For example, for a system of homogeneous polynomial equations over an algebraically closed field, the number of solutions in $\mathbb{P}^n$ can be found using Bézout's Theorem (Theorem 1.5.1).

**Theorem 1.5.1** (Bézout's Theorem [41, §II.2]). *Given $n$ projective hypersurfaces of degrees $d_1, \ldots, d_n$ in a projective space of dimension $n$ over an algebraically closed field, if the intersection of the hypersurfaces is zero dimensional, then the number of intersection points, counted with multiplicity, is equal to the product of the degrees $d_1 \cdots d_n$.*

25

We will now formally define multiplicity. Let $I \subseteq k[x_1, \ldots, x_n]$ be an ideal in a polynomial ring with finitely many variables, $x_1, \ldots, x_n$, over an algebraically closed field $k$. Then $I$ is *zero-dimensional* if the corresponding variety $V(I) \subseteq \mathbb{P}^n_k$ contains only finitely many points, or equivalently, if $k[x_1, \ldots, x_n]/I$ is finite dimensional. For a zero-dimensional ideal $I$, we can assign to each point $p \in V(I)$ a positive integer, $m_I(p)$, called the *multiplicity*. Multiplicities are formally defined using local rings.

**Definition 1.5.2** ((2.1), [11]). Let $I$ be a zero-dimensional ideal in $k[x_1, \ldots, x_n]$, and assume that $\mathbf{p} = (p_1, \ldots, p_n) \in V(I)$. Then the *multiplicity* of $\mathbf{p}$ as a point in $V(I)$ is

$$\dim_k k[x_1, \ldots, x_n]_{\langle x_1 - p_1, \ldots, x_n - p_n \rangle} / I k[x_1, \ldots, x_n]_{\langle x_1 - p_1, \ldots, x_n - p_n \rangle}.$$

Let $X(I)$ denote the vanishing of a homogeneous ideal $I$ in complex projective space, $\mathbb{P}^n_{\mathbb{C}}$. When counted with multiplicity, the number of points in $X(I)$ is always the expected count

$$\sum_{p \in V(I)} m_I(p) = \dim_k k[x_1, \ldots, x_n]/I.$$

This expected count is also referred to as the *degree* of $I$, denoted $\deg I$. If $f_1, \ldots, f_n$ are generators for $I$, Bézout's Theorem tells us the degree of $I$ is the product of the degrees of the $f_i$.

## 1.5.2 Standard Bases

In Chapter 4, we compute the multiplicity of specific points which are counted by Bézout's Theorem, but which do not lie in the model. It can be difficult in general to compute dimensions of quotient rings, but it is much simpler when $I$ is generated by monomials. In this case $I$ is called a *monomial ideal*, and the dimension of $R/I$ is the number of monomials not in $I$. Monomial orders and Gröbner bases provide the tools needed to convert any computation $\dim_k R/I$ in a polynomial ring into a computation with monomial ideals by taking leading terms. The leading term of a polynomial $f$ under the monomial order $>$ is the term of $f$ whose monomial is largest according to $>$. The leading term ideal of $I$ is generated by the leading terms of all elements of $I$; if $G = \{g_1, \ldots, g_k\}$ is a Gröbner basis of $I$, then the leading term ideal is generated by the leading terms of $G$. That is,

$$\mathrm{LT}_>(I) = \langle \mathrm{LT}_>(f) \mid f \in I \rangle = \langle \mathrm{LT}_>(g_1), \ldots, \mathrm{LT}(g_k) \rangle.$$

The operation of taking leading terms preserves the dimension but replaces $I$ with a monomial ideal,

$$\dim_k R/I = \dim_k R/\mathrm{LT}_>(I).$$

Local orders and standard bases, which I introduce below, are the analog for local rings and counting the multiplicity of a point. Let $R$ be a local ring. For example, let $R$ be a polynomial ring localized at a point, $k[x_1, \ldots, x_n]_{\langle x_1 - p_1, \ldots, x_n - p_n \rangle}$, or a ring of power series $k[\![x_1, \ldots, x_n]\!]$. Analogous to a monomial order for a polynomial ring, on a local ring can have a *local order*.

**Definition 1.5.3.** A *local order* is any total order $>$ such that

1. $>$ is compatible with multiplication, meaning $x^a > x^b$ implies $x^a x^c > x^b x^c$, and

2. $1 > x_i$ for $1 \le i \le n$.

In analogy to a Gröbner basis, a generating set $\{f_1, \ldots, f_k\}$ of an ideal $I$ is a *standard basis* for $I$ under local order $>$ if $\{\mathrm{LT}_>(f_1), \ldots, \mathrm{LT}_>(f_k)\}$ is a generating set for $\mathrm{LT}_>(I)$.

With a local order, an analogous dimension counting theorem holds:

**Theorem 1.5.4** (§4(4.3), [11])**.** *Let $\widehat{R} = \mathbb{C}[\![x_0, \ldots, x_n]\!]$. Let $\widehat{J} \subset \widehat{R}$ be an ideal, $>$ a local order, and $\langle \mathrm{LT}_>(\widehat{J}) \rangle$ be the leading term ideal for $\widehat{J}$ with respect to $>$. If $\widehat{R}/\widehat{J}$ contains finitely many standard monomials; that is, monomials $x^\alpha \in \widehat{R}$ such that $x^\alpha \notin \langle \mathrm{LT}_>(\widehat{J}) \rangle$, then*

$$\dim_{\mathbb{C}}(\widehat{R}/\widehat{J}) = \dim_{\mathbb{C}}(\widehat{R}/\mathrm{LT}_>(\widehat{J})).$$

The analogs of the polynomial division algorithm and Buchburger's criterion, which give algorithms to compute standard bases, are the following.

**Theorem 1.5.5** (Mora's Normal Form Algorithm §4 (3.10), [11])**.** *Given homogeneous polynomials $F, F_1, \ldots, F_s$ in $K[t, x_1, \ldots, x_n]$ and the monomial order $>'$ extending the semigroup order $>$ on monomials in the $x_i$, there is an algorithm for producing homogeneous polynomials $H, U, A_1, \ldots, A_s \in k[x_1, \ldots, x_n]$ satisfying*

$$U \cdot F = A_1 F_1 + \cdots + A_s F_s + H,$$

*where $\mathrm{LT}_{>'}(U) = t^a$ for some $a$, $a + \deg F + \deg A_i + \deg F_i = \deg H$ whenever $A_i, H \ne 0$, and no $\mathrm{LT}_{>'}(F_i)$ divides $t^b \mathrm{LT}_{>'}(H)$ for any $b \ge 0$.*

**Theorem 1.5.6** (Analog of Buchburger's Criterion [11, §4 (4.2)b]). *A subset* $S = \{g_1, \ldots, g_t\} \subset I$ *is a standard basis for* $I$ *if and only if applying the Mora normal form algorithm Theorem 1.5.5 to every S-polynomial formed from the elements of the set of homogenizations* $S^h = \{g_1^h, \ldots, g_t^h\}$ *yields a zero remainder.*

We will use the following shortcut when computing a standard basis in Chapter 4, which applies when $R$ is a a polynomial ring, power series ring, or a localization of a polynomial ring.

**Lemma 1.5.7.** *Let* $f, g \in R$, *and let* $>$ *be a local order (or monomial order) on* $R$. *If* $f, g$ *have relatively prime leading terms, then* $S(f, g) = 0$. *Thus, generators of an ideal whose leading terms are relatively prime form a standard basis (Gröbner basis) for the ideal.*

*Proof.* $S$-pairs of elements with relatively prime leading terms vanish. Therefore, Buchburger's algorithm immediately terminates, and we conclude that the generators form a standard basis. □

## 1.6 Results and Contributions

Part 1 of this thesis is concerned with tropical phylogenetics, i.e., problems in phylogenetics involving the tropical space of phylogenetic trees. In Chapter 2, I study the tree topology changes that occur along the tropical line segment between two phylogenetic trees. The main theorem is Theorem 2.3.1, which states that trees along the tropical line segment can only change by Nearest Neighbor Interchange and Four Clade Rearrangement. This disproves an earlier conjecture that the topology changes are only NNI illustrates that tropical tree space is very different from BHV space – the trees along geodesic in BHV space can be less resolved or even star trees [3, Corollary 4.1]. In addition, I show that the number of NNI moves occurring along the tropical line segment can be as large as $n^2$, but the average number of moves when the two endpoint trees are chosen at random is $O(n(\log n)^4)$. This is in contrast with $O(n \log n)$, the average number of NNI moves needed to transform one tree into another. This chapter was first published by MSP in *Algebraic Statistics* in volume 14(1), pages 71-90 [12], and builds off an earlier joint work with Ruriko Yoshida [51].

Chapter 3 concerns geometric medians of weighted tropical data points. Specifically, let $V = \{v_1, \ldots, v_m\}$ be points in $\mathbb{R}^n$, and let $w_1, \ldots, w_m$ be positive real weights. A weighted geometric median of $V$ with respect to $\mathbf{w}$ is a point $z$ which minimizes $\sum w_i d(v_i, z)$. Such a point is called a Fermat-Weber point. Fermat-Weber points in tropical space were first studied by Comăneci and Joswig [7], in the case where all the weights are equal, i.e., $w_1 = \cdots = w_m$. Comăneci and Joswig showed that the set of (unweighted) Fermat-Weber points

is the "central" cell of the tropical convex hull of $v_1, \ldots, v_m$. In Theorem 3.1.1, we show that for any fixed data points $v_1, \ldots, v_m$, as the weights $w_i$ vary, the set of all Fermat-Weber points is the entire tropical convex hull of the $v_i$. This chapter appears in "The tropical polytope is the set of all weighted tropical Fermat-Weber points" [13], which is joint work with Mark Curiel.

Part 2 is about likelihood inference for Brownian motion tree models. We study the complexity of inferring the maximum likelihood estimator for a Brownian motion tree model by computing its maximum likelihood degree. Our goal is to obtain the maximum likelihood degree solely from the unlabeled topology of $T$. Our main result is that the maximum likelihood degree of the Brownian motion tree model on a star tree with $n + 1$ leaves is $2^{n+1} - 2n - 3$. This result was previously conjectured by Améndola and Zwiernik. In addition, we find a combinatorial formula for the determinant of the concentration matrix of a Brownian motion tree model (for any tree $T$), which generalizes the Cayley-Prüfer theorem. We also prove that the maximum likelihood of a Brownian motion tree model is independent of the choice of root. This work will appears in the preprint "ML Degrees of Brownian Motion Tree Models: Star Trees and Root Invariance", which is joint work with Jane Ivy Coons, Aida Maraj, and Ikenna Nometa.

# Part I

# Tropical Phylogenetics

## CHAPTER 2

## Classifying Tree Topologies along Tropical Line Segments

## 2.1   Introduction

Tropical polytopes in tree space play a critical role in understand sampling methods, yet it is poorly understood which trees appear in a tropical polytope with tree vertices. Therefore, the first goal of this chapter is to answer the following question:

**Question 2.1.1.** How do tree topologies (i.e. tree structures) change along the tropical line segment?

It was conjectured previously that the trees along the tropical line segment change by Nearest Neighbor Interchange (NNI). This is not quite true. For example, in [51] it was shown that the tropical line between two trees on four leaves may pass through the star tree with probability greater than 0, and this is not an NNI move. However, the behavior does not get much worse.

The tropical line segment is a concatenation of classical line segments. The tree topology can change at the points where the classical line segments connect, but not on the interior of any of the classical line segments. We call these points *turning points* (whether or not the tree topology changes). In Section 2.3, we prove the following result:
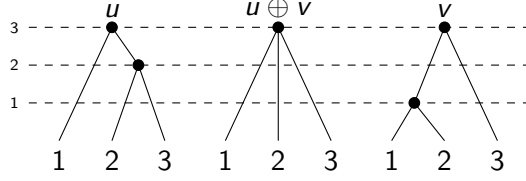
Figure 2.1: The turning points on a tropical line segment with a single NNI move.
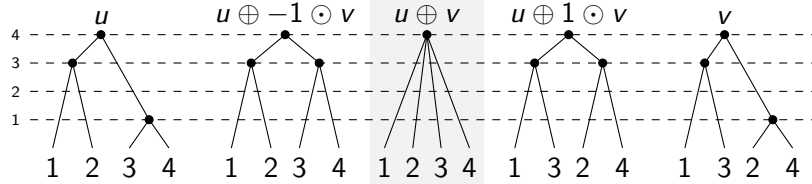


Figure 2.2: The turning points on a tropical line segment with a four clade rearrangement.

**Theorem 2.3.1.** *Tree topology changes only occur at turning points; at each turning point, one of three things can happen:*

1. *(Nearest Neighbor Interchange) One internal vertex has three children, see Figure 2.1.*

2. *(Four Clade Rearrangement) One internal vertex has four children, see Figure 2.2.*

3. *(No Topology Change) All internal vertices have two children, see Figure 2.3.*

The four clade rearrangement move can be achieved by three nearest neighbor interchange moves, so it makes sense to ask how many NNI moves occur along the tropical line segment between two general trees – each turning point can contribute 0, 1, or 3 NNI moves. We will refer to the number of NNI moves occurring along the tropical path as the *tropical NNI number* (counting a four-clade rearrangement as three NNI moves).
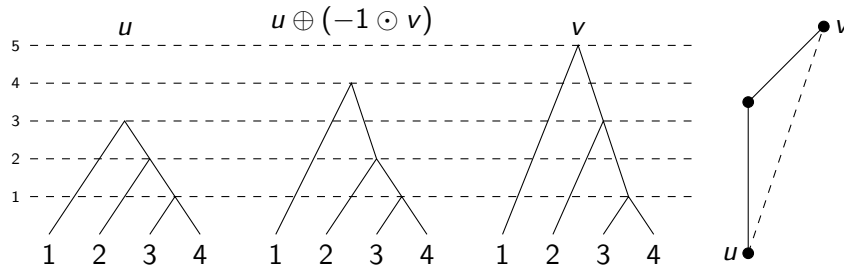


Figure 2.3: Left: the turning points on a tropical line segment with constant tree topology. Right: the tropical line segment (solid), and straight line segment (dashed) between $u$ and $v$.

**Question 2.1.2.** What is the tropical NNI number between a random pair of trees on $n$ leaves? How does it compare to the *NNI distance*, the minimal number of NNI moves required to transform one tree into the other?

By a pair of *randomly chosen phylogenetic trees* on $[n]$ leaves, we mean that the topologies of the trees are uniformly randomly chosen from the (finitely many) binary tree topologies on $[n]$ leaves, and whose corresponding tree metrics are sufficiently general, in a way that is made precise in Section 2.2. For a random pair of phylogenetic trees on $n$ leaves, it is known that the average NNI distance is $O(n \log n)$, and the minimal number of NNI moves is NP-hard to compute [14]. Theorem 2.3.1 implies that the number of turning points is an upper bound on triple the number of NNI moves on the tropical line segment. First we provide an example where the tropical NNI number can be much larger than the minimal number of NNI moves.

**Theorem 2.4.1.** *For each $n$, there exist generic pairs of trees, $T_1, T_2$, on $n$ leaves such that that the tropical NNI number from $T_1$ to $T_2$ is $\binom{n-1}{2}$.*

Then we prove an upper bound for the average case.

**Corollary 2.5.2.** *The expected number of NNI moves occurring along the tropical line segment between two randomly chosen trees on $n$ leaves is $O(n(\log(n))^4)$.*

This chapter is structured as follows. In Section 2.2, we derive a non-standard definition of turning points, which will be useful throughout the rest of the chapter. In Section 2.3 we classify tree topology changes occurring along the tropical line segment between two generic trees, proving Theorem 2.3.1. In Section 2.4, we describe a family of generic pairs of trees whose tropical NNI number far exceeds the expected tropical NNI number (proving Theorem 2.4.1). In Section 2.5, we bound the expected NNI distance along the tropical line segment between two randomly chosen trees on $n$ leaves (proving Theorem 2.5.1).

## 2.2 Redefining Turning Points for Trees

The tropical line segment between two points in the tropical projective torus can be found using the following algorithm.

**Definition 2.2.1** (Algorithm for the Tropical Line Segment)**.**

1. **input:** $u, v \in \mathbb{R}^n$.

2. $\Lambda(u, v) = [u_i - v_i \mid i \in [n]]$, sorted from least to greatest.

$u = u \oplus (-2 \odot v)$

$u \oplus (1 \odot v) = v \bullet \quad\quad \bullet w_2 = u \oplus (0 \odot v)$
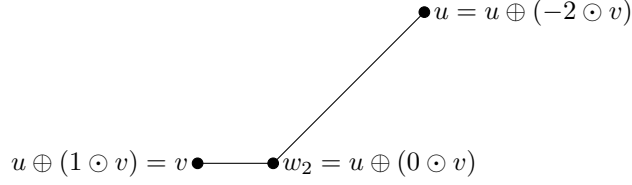
Figure 2.4: The tropical line segment from $u$ to $v$.

3. $w_i = u \oplus (\lambda_i \odot v)$, where $\lambda_i$ is the $i$th smallest element of $\Lambda$.

4. **output:** the concatenation of Euclidean line segments from $w_i$ to $w_{i+1}$.

**Definition 2.2.2.** In the algorithm above, $\Lambda(u, v) = \{u_i - v_i \mid i \in [n]\}$ is the *set of turning point scalars*, and $w_i$ is a *turning point* for the tropical line segment from $u$ to $v$.

**Example 2.2.3.** Let $u = (3, 3, 1)$ and $v = (3, 2, 3)$. Then $\Lambda = [-2, 0, 1]$, and the turning points on the tropical line are:

$$
\begin{aligned}
u \oplus (-2 \odot v) = (3, 3, 1) \oplus (1, 0, 1) = (3, 3, 1) && = u \\
u \oplus (0 \odot v) = (3, 3, 1) \oplus (3, 2, 3) = (3, 3, 3) && = (0, 0, 0) \\
u \oplus (1 \odot v) = (3, 3, 1) \oplus (4, 3, 4) = (4, 3, 4) && = v.
\end{aligned}
$$

This tropical line segment is depicted in Figure 2.4.

Now we will reinterpret the turning point scalars in the tropical line segment algorithm (Definition 2.2.1) in terms of internal vertices of the endpoint trees, which will allow us to describe turning points without referring to the tree metrics.

**Definition 2.2.4.** Given a pair of phylogenetic trees $T_1, T_2$ on $n$ leaves, define the *essential pairs* of $T_1, T_2$ to be

$$
\Pi(T_1, T_2) := \{(x_1, x_2) \mid x_k \in T_k, \exists i, j \in [n], i \neq j \text{ s.t. } x_k = \text{lca}_{T_k}(i, j)\}.
$$

In the sections that follow, we will find that the results are much nicer if we consider trees with sufficiently generic ultrametrics. We qualify "sufficiently generic ultrametric" with the definition of *generic* below.

**Definition 2.2.5.** An equidistant tree $T$ on $n$ leaves with distance vector $u$ is *generic* if any of the following equivalent conditions hold:

1. $u$ lies in the relative interior of a full-dimensional cone of $\text{Tree}_n^{\text{tr}}$,

33

2. $T$ has exactly $n - 1$ internal vertices (including the root), and

3. all internal vertices of $T$ have degree 3, except the root, which has degree 2.

Note that if $u$ is generic, then so is $\lambda \odot u$ for any $\lambda \in \mathbb{R}$, since tropical multiplication affects only pendant edge lengths. A tree that is not generic is unresolved.

**Definition 2.2.6.** A tree $T$ is unresolved (at internal vertex $v$) if $v$ has more than two children.

It is also possible to encounter highly unresolved trees along the tropical line segment between $T_1$ and $T_2$ if certain edge lengths in the trees coincide. Therefore, we also require that $T_1$ and $T_2$ are generic as a pair. We call such a pair a *generic pair*, and in Definition 2.2.7 we formally define such a pair.

**Definition 2.2.7.** A pair of equidistant trees $T_1, T_2$ on $n$ leaves, with distance vectors $u$ and $v$, respectively, is a *generic pair* if:

1. $u$ and $v$ are each generic, and

2. for all $x_i, y_i$ distinct internal vertices of $T_i$, $h_{T_1}(x_1) - h_{T_2}(x_2) \neq h_{T_1}(y_1) - h_{T_2}(y_2)$.

For a generic pair of trees, the number of turning points is independent of the specific choice of ultrametric.

**Lemma 2.2.8.** *Let $T_1, T_2$ be phylogenetic trees with leaf set $[n]$. Then for any sufficiently general choice of ultrametrics $u^k$ on $T_k$ (i.e. $u^1, u^2$ is a generic pair),*

$$\Lambda(u_1, u_2) = \{2(h_{u_1}(x_1) - h_{u_2}(x_2)) \mid (x_1, x_2) \in \Pi(T_1, T_2)\}.$$

*Proof.* According to Definition 2.2.1, the turning points of the tropical line segment between $T_1$ and $T_2$ occur at $u^1 \oplus \lambda_{ij} \odot u^2$, where $\lambda_{ij} = u_{ij}^1 - u_{ij}^2$.

We can rewrite the expression above in terms of heights on internal vertices.

$$\lambda_{ij} = 2(h_{T_1}(x_1) - h_{T_2}(x_2)), \text{ where } x_k = \text{lca}_{T_k}(i, j).$$

Conversely, given internal vertices $x_1 \in T_1$ and $x_2 \in T_2$, $\mu_{x_1 x_2} = 2(h_{T_1}(x_1) - h_{T_2}(x_2))$ is the scalar for a turning point if and only if $\mu_{x_1 x_2} = \lambda_{ij}$ for some $i, j \in [n]$, which happens if and only if there are $i, j \in [n]$ with $x_k = \text{lca}_{T_k}(i, j)$ for $k = 1, 2$. $\qquad\square$

**Corollary 2.2.9.** *Let $T_1, T_2$ be phylogenetic trees with leaf set $[n]$. Then for any sufficiently general choice of ultrametrics $u_k$ on $T_k$, $\#\Lambda(u_1, u_2) = \#\Pi(T_1, T_2)$.*

Therefore, the number of turning points between two combinatorial (non-metric) tree structures can be defined as:

**Definition 2.2.10.** The *tropical interchange number* for trees $T_1$, $T_2$, denoted by $tI(T_1, T_2)$, is the number of turning points between $T_1$ and $T_2$ for any choice of generic weights $u_1, u_2$ on $T_1, T_2$.

## 2.3 The Tropical Line and NNI

Now we come to the first main question of the paper: what kinds of tree topology changes can occur along the tropical line segment between two trees $T_1$ and $T_2$?

It turns out that even for a generic pair of trees, there may be a turning point which is not a single NNI. For example, the endpoint trees in Figure 2.2 are a generic pair, but the tropical line segment passes through a tree with an internal vertex with four descendants (and this is independent of the specific choice of metrics on the trees). In this section, we prove that when the endpoint trees are chosen to be sufficiently general, the only moves that can occur are: a single NNI, and a single four clade rearrangement.

### 2.3.1 Possible Turning Points

The main theorem for this section is the following:

**Theorem 2.3.1.** *At each turning point of the tropical line segment between a generic pair of trees, the intermediate tree takes one of the following three forms:*

1. *a tree that is trivalent except for one internal vertex with three children (the middle of an NNI, passing through a co-dimension 1 cell), see Figure 2.1 for an example;*

2. *a tree that is trivalent except for one internal vertex with four children (the middle of a four clade rearrangement, passing through a co-dimension 2 cell), see Figure 2.2 for an example;*

3. *a generic tree, (remaining in a top-dimensional cell), see Figure 2.3 for an example.*

**Corollary 2.3.2.** *On the tropical line between two generic trees $u, v$, there are no trees with an internal vertex with five or more children.*

The following definition will be useful in proving the theorem:

**Definition 2.3.3.** For $u, v$ a pair of trees on $n$ leaves, let $G(u \geq v)$ be the graph with vertices $[n]$ and edges $(i, j)$ such that $u_{ij} \geq v_{ij}$.

Note that $G(u \geq v) \cup G(u \leq v) = K_n$.

**Lemma 2.3.4.** *If $G$ and $H$ are subgraphs of $K_5$, with $G \cup H = K_5$, then at least one of $G$, $H$ contains an odd cycle.*

*Proof.* Assume $G$ and $H$ each contain no odd cycles. Then $G$, $H$ are bipartite graphs on the five vertices. Let $A, B, C$ be distinct vertices of $G$ that share the same color. Then at least one of the pairs $\{A, B\}$, $\{A, C\}$, and $\{B, C\}$ has constant color in $H$. So neither $G$ nor $H$ contains an edge between that pair. This contradicts the assumption that $G \cup H = K_5$. $\square$

*Proof of Theorem 2.3.1.* Suppose for a contradiction that there is some $\lambda \in \Lambda(u, v)$ such that the tree $w = u \oplus \lambda \odot v$ has an internal vertex with five or more children. Denote that internal vertex $w_0$. Pick one leaf descending from each child of $w_0$, and call them $1, 2, 3, 4, 5, \ldots$. Then $10 = \binom{5}{2}$ distances coincide:

$$\max\{u_{12}, \lambda + v_{12}\} = \max\{u_{13}, \lambda + v_{13}\} = \cdots = \max\{u_{45}, \lambda + v_{45}\}.$$

Lemma 2.3.4 shows that one of $G(u \geq \lambda + v)$ and $G(u \leq \lambda + v)$ must contain an odd cycle. This means that for $k = 1$ or $k = 2$:

$$u_{12} = u_{23} = \cdots = u_{2k+1,1}$$

(or symmetrically for $\lambda \odot v$). Without loss of generality, suppose that the equations involving $u$ hold. Then all but the last distance tells us that $u_0 = \mathrm{lca}_u(i, j)$ for all $i \in \{1, 3, \ldots, 2k+1\}$ and $j \in \{2, 4, \ldots, 2k\}$, meaning that the odd numbered and even numbered leaves descend from different children of $u_0$. But then the last distance, $u_{2k+1,1}$ says that $1, 2k+1$ also have least common ancestor $u_0$, so the internal vertex must have at least three children, which contradicts the assumption that $u$ is generic.

Therefore, one of $u$, $\lambda \odot v$ is not generic. Since $\lambda \odot v$ is generic if and only if $v$ is generic, this contradicts our assumption that $u, v$ are generic (i.e. trivalent) trees. Thus, no intermediate tree on the tropical line between a generic pair $u, v$ has an internal vertex with five or more children. $\square$

**Proposition 2.3.5.** *On the tropical line segment between a generic pair of trees, there are no intermediate trees with two internal vertices that have three or more children each.*

*Proof of Proposition 2.3.5.* Suppose $w = u \oplus (\lambda \odot v)$ has two internal vertices with at least three children each. Name three of the descendants of the first high-degree internal vertex 1, 2, and 3, and name three descendants of the second high-degree internal vertex 4, 5, and

36

6. Then
$$w_{12} = w_{13} = w_{23} \text{ and } w_{45} = w_{46} = w_{56}.$$

If we expand $w_{12} = w_{13} = w_{23}$ in terms of $u$ and $\lambda \odot v$, we see

$$\max\{u_{12}, \lambda + v_{12}\} = \max\{u_{13}, \lambda + v_{13}\} = \max\{u_{23}, \lambda + v_{23}\}. \tag{2.1}$$

We will show that these equations imply $u_{ij} = \lambda + v_{ij}$ for some $ij \subset \{1, 2, 3\}$ and $u_{k\ell} = \lambda + v_{k\ell}$ for some $k\ell \subset \{4, 5, 6\}$, and this implies that $u_{ij} - v_{ij} = \lambda = u_{k\ell} - v_{k\ell}$, which contradicts the assumption that $u, v$ is a generic pair.

The definition of ultrametric tells us that two of $u_{12}, u_{13}, u_{23}$ are equal and greater than the third; without loss of generality, assume $u_{12} = u_{13} \geq u_{23}$. The fact that $u$ is generic implies the inequality is strict: $u_{12} = u_{13} > u_{23}$. Using a similar argument for $\lambda \odot v$, there are three possibilities:

$$\lambda + v_{12} = \lambda + v_{13} > \lambda + v_{23} \tag{2.2}$$
$$\lambda + v_{12} = \lambda + v_{23} > \lambda + v_{13} \tag{2.3}$$
$$\lambda + v_{13} = \lambda + v_{23} > \lambda + v_{12} \tag{2.4}$$

However, if (2) holds, then $w_{13} = \max\{u_{13}, \lambda + v_{13}\} > \max\{u_{23}, \lambda + v_{23}\} = w_{23}$, and this contradicts $w_{13} = w_{23}$. Therefore, one of (3) or (4) must hold (and they are equivalent up to permuting the indices of $u$ and $v$ simultaneously), so without loss of generality assume $\lambda + v_{12} = \lambda + v_{23} > \lambda + v_{13}$.

Now we must have $w_{13} = u_{13}$. If not, then $w_{13} = \lambda + v_{13} < \lambda + v_{23} \leq w_{23}$, and this contradicts $w_{13} = w_{23}$. Similarly, if $w_{23} \neq \lambda + v_{23}$, then $w_{23} = u_{23} < u_{13} \leq w_{13}$, which again contradicts $w_{23} = w_{13}$. It follows that:

$$w_{12} = w_{23} = w_{13} = u_{13} = u_{12} > u_{23}, \tag{2.5}$$
$$w_{12} = w_{13} = w_{23} = \lambda + v_{23} = \lambda + v_{12} > \lambda + v_{13} \tag{2.6}$$
$$\implies \lambda + v_{12} = u_{12} = u_{13} = \lambda + v_{23} > u_{23} \text{ and } \lambda + v_{13}. \tag{2.7}$$

In particular, $\lambda + v_{12} = u_{12}$. Following the same argument after replacing 123 with 456:

$$\lambda + v_{45} = u_{45} = u_{46} = \lambda + v_{56} > u_{56} \text{ and } \lambda + v_{46}.$$

This implies $u_{12} - v_{12} = \lambda = u_{45} - v_{45}$, so $u_{12} - v_{12} = u_{45} - v_{45}$, and this contradicts the assumption that $u, v$ is a generic pair. $\qquad\square$

## 2.4  A Very Long Tropical Line

Although Theorem 2.5.1 limits the average tropical NNI number to $O(n(\log n)^4)$, specific pairs of trees can have a much larger tropical NNI number. Figure 2.5 depicts a pair of trees on $[n]$ leaves, which differ by $n - 2$ NNI moves, but whose tropical line segment has $\binom{n-1}{2} \sim n^2$ single NNI moves. Denote the trees in Figure 2.5 below by $u^n$ and $v^n$ respectively.



Figure 2.5: $u^n, v^n$ is a generic pair of trees on $[n]$ leaves with tropical NNI number $\binom{n-1}{2}$, with $u_{ij} = n(n - \min(i,j))$ and $v_{ij} = \max\{i,j\} - 1$. Only $n - 2$ NNI moves are needed to transform $u^n$ into $v^n$.

**Theorem 2.4.1.** *Along the tropical line segment from $v^n$ to $u^n$, there are $\binom{n-1}{2}$ single NNIs and no four clade rearrangements.*

*Proof.* First note that $(\mathrm{lca}_u(i,j), \mathrm{lca}_v(i,j)) = (u_{\min(i,j)}, v_{\max\{i,j\}})$. Thus, $\#\Pi(u,v) = \binom{n}{2}$. Furthermore, $u, v$ is a generic pair, so $\#\Lambda(u,v) = \#\Pi(u,v)$ and the only possible turning points are those outlined in Theorem 2.3.1.

Fix $i < j$. By definition, $u_{ij}$ depends only on $\min(i,j) = i$, and $v_{ij}$ depends only on $\max\{i,j\} = j$. We will show that at the turning point with scalar $\lambda = u_{ij} - v_{ij}$, there is a NNI move if $|i - j| > 1$ and a generic tree if $|i - j| = 1$. The intermediate tree $w$ is illustrated in Figure 2.6.

Now let $k \in [i + 1, j - 1] \cap [n]$. In tree $u$, we have

$$u_{ij} = u_{ik} > u_{kj},$$
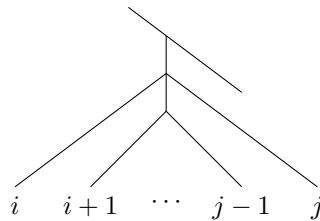


Figure 2.6: A part of the turning point tree $w = u \oplus (\lambda_{ij} \odot v)$. When $i + 1 = j$, there are no leaves in the subtree between $i$ and $j$, and $w$ is generic.

38

and in tree $v$, we have

$$v_{ij} = v_{kj} > v_{ik}.$$

Recall that we set $\lambda = u_{ij} - v_{ij}$. Let $w = u \oplus (\lambda \odot v)$. Then

$$u_{ik} = u_{ij} = v_{ij} + \lambda = v_{jk} + \lambda > v_{ik} + \lambda \text{ and } u_{jk}$$

which means $w_{ij} = w_{ik} = w_{jk}$, i.e., $\mathrm{lca}_w(i,j) = \mathrm{lca}_w(i,k) = \mathrm{lca}_w(j,k)$. This implies $w$ is non-generic when $|i - j| > 1$.

Observe that if $k \notin [i,j] \cap [n]$, then $\mathrm{lca}_w(k,i) \neq \mathrm{lca}_w(i,j)$. We can see this by comparing $w_{ki}$ or $w_{jk}$ to $w_{ij}$. If $k < i$, then $u_{ij} < u_{jk}$, and $v_{ij} = v_{jk}$, so

$$w_{jk} = \max\{u_{jk}, v_{jk} + \lambda\} = \max\{u_{jk}, v_{ij} + \lambda\} = \max\{u_{jk}, u_{ij}\} = u_{jk} > w_{ij}.$$

An analogous argument shows that $w_{jk} > w_{ij}$ for $k > j$. Therefore, $w$ is non-generic *only when* $|i - j| > 1$. We are just left to rule out a four-clade rearrangement.

If $k_1, k_2 \in [i+1, j-1] \cap [n]$, then

$$u_{ij} = u_{ik_2} > u_{k_1 k_2}, \text{ and } v_{ij} = v_{k_1 j} > v_{k_1 k_2}$$

and this implies $w_{k_1 k_2} < w_{ij}$, so $w$ has a NNI and not a four-clade rearrangement. □

**Warning:** The tropical NNI number is not a metric because it does not satisfy the triangle inequality. For example, there is a concatenation of tropical line segments from $v^n$ to $u^n$, over which $n - 2$ single NNI moves occur (the minimum possible number), and this is smaller than the number of NNI moves occurring on the tropical line segment from $v^n$ to $u^n$.

## 2.5 Expected Length of the Tropical Line Segment

Now that we know tree topologies along the tropical line segment change by single NNI or four clade rearrangement moves, we can ask how many NNI moves occur along the tropical line segment between two randomly chosen trees. The goal of this section is to prove Theorem 2.5.1 bounding the number of turning points. A pair of randomly chosen phylogenetic trees on $[n]$ leaves means that the topologies of the trees are uniformly randomly chosen from the (finitely many) binary tree topologies on $[n]$ leaves, and the corresponding ultrametrics form a generic pair.

**Theorem 2.5.1.** *The expected number of turning points of the tropical line segment between two randomly chosen phylogenetic trees on n leaves is $O\left(n\left(\log(n)\right)^4\right)$.*

**Corollary 2.5.2.** *The expected tropical NNI number between two randomly chosen phylogenetic trees on n leaves is $O\left(n\left(\log(n)\right)^4\right)$.*

## 2.5.1 Sample Spaces, Notation, and Equivalence of Probabilities

We want to bound the number of turning points of the tropical line segment between two phylogenetic trees on $n$ leaves. After a series of translations, we reduce this to a question on the space of unlabeled rooted binary plane trees with one marked internal vertex.

**Definition 2.5.3.** A rooted tree is a ***plane tree*** if the children of each vertex are assigned an ordering. For rooted binary plane trees, we will refer to the *left child* and *right child* of an internal vertex, and denote the descendants of the left child by $L(v)$ and the descendants of the right child by $R(v)$. Let $\ell(v) = |L(v)|$ and $r(v) = |R(v)|$.

We will consider the following sample spaces of trees:

$$\mathcal{T}_{[n]} := \{T : T \text{ is a rooted binary tree with } n \text{ labeled leaves}\}.$$
$$\mathcal{T}_{[n]}^{\text{plane}} := \{T : T \text{ is a rooted binary plane tree with } n \text{ labeled leaves}\}.$$
$$\mathcal{T}_{[n],v}^{\text{plane}} := \{(T,v) : T \in \mathcal{T}_{[n]}^{\text{plane}}, v \text{ is an internal vertex of } T\}.$$
$$\mathcal{T}_{n}^{\text{plane}} := \{T : T \text{ is a rooted binary plane tree with } n \text{ unlabeled leaves}\}.$$
$$\mathcal{T}_{n,v}^{\text{plane}} := \{(T,v) : T \in \mathcal{T}_{n}^{\text{plane}}, v \text{ is an internal vertex of } T\}.$$
$$\mathcal{T}_{n,v}^{\text{plane}}(a,b) := \{(T,v) \in \mathcal{T}_{n,v}^{\text{plane}} : \ell(v) = a, r(v) = b\}.$$

We now re-frame the number of turning points in terms of the last tree sample space above. This proposition makes it clear that the number of turning points is reduced from the naive bound of $\binom{n}{2}$ only because different leaf pairs $(i,j)$ can have the same common ancestor in $T$.

**Definition 2.5.4.** Given $A_i, B_i \subseteq [n]$ with $A_i \cap B_i = \emptyset$ and $|A_i| = a_i$, $|B_i| = b_i$, let $Q(a_1, b_1, a_2, b_2)$ be the probability that the following intersections are non-trivial:

$$A_1 \cap A_2 \neq \emptyset, B_1 \cap B_2 \neq \emptyset.$$

**Proposition 2.5.5.**

$$E\left(\#\Pi(T_1, T_2) \mid T_k \in \mathcal{T}_{[n]}\right)$$
$$\leq 2(n-1)^2 E\left(Q\left(|L(v_1)|, |R(v_1)|, |L(v_2)|, |R(v_2)|\right) \mid {}^{(T_k, v_k) \in \mathcal{T}_{n,v}^{\text{plane}}}_{k=1,2}\right).$$

*Proof.* By definition, the left-hand side is the probability that a pair of internal vertices $x_1 \in T_1$, $x_2 \in T_2$ are the least common ancestor of some leaves $i, j$ in their respective trees, times the number of pairs $(x_1, x_2)$. There are $n-1$ internal vertices of each tree $T_k$, so there are $(n-1)^2$ possible pairs. So we find that (LHS) is equal to:

$$(n-1)^2 P((x_1, x_2) \in \Pi(T_1, T_2) \mid (T_k, x_k) \in \mathcal{T}_{[n],v}^{\text{plane}}, k = 1, 2). \tag{2.8}$$

We can further rephrase (2.8) in terms of leaf-labeled plane trees:

$$(n-1)^2 P(L(x_1) \cap L(x_2) \neq \emptyset, R(x_1) \cap R(x_2) \neq \emptyset$$
$$\text{or } L(x_1) \cap R(x_2) \neq \emptyset, R(x_1) \cap L(x_2) \neq \emptyset \mid (T_k, x_k) \in \mathcal{T}_{[n],v}^{\text{plane}}). \tag{2.9}$$

We can simplify (2.9) by using only one of the conditions. Then, (2.9) is bounded by:

$$2(n-1)^2 P\left(L(x_1) \cap L(x_2) \neq \emptyset, R(x_1) \cap R(x_2) \neq \emptyset \mid (T_k, x_k) \in \mathcal{T}_{[n],v}^{\text{plane}}\right). \tag{2.10}$$

Finally, we forget about the leaf-labels and introduce the variable $Q$, so (2.10) equals the right-hand side of Proposition 2.5.5. $\square$

### 2.5.1.1 Counting Trees

In this section, we count trees to provide an explicit formula for the expected value bound that we proved in the last subsection.

**Lemma 2.5.6.** *[43, Example 5.3.12]*

$$|\mathcal{T}_n^{\text{plane}}| = \frac{1}{n}\binom{2n-2}{n-1}.$$

**Lemma 2.5.7.**

$$|\mathcal{T}_{n,v}^{\text{plane}}| = \frac{n-1}{n}\binom{2n-2}{n-1}.$$
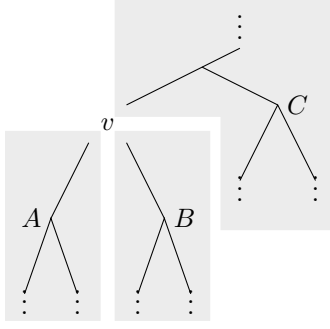
Figure 2.7: Splitting a binary tree into three binary trees at an internal vertex.

*Proof.* A rooted binary tree on $n$ leaves has $n-1$ internal vertices (including the root), so

$$|\mathcal{T}_{n,v}^{\text{plane}}| = (n-1) \cdot \frac{1}{n}\binom{2n-2}{n-1} = \frac{n-1}{n}\binom{2n-2}{n-1}. \qquad \square$$

**Lemma 2.5.8.**

$$|\mathcal{T}_{n,v}^{\text{plane}}(\ell(v) = a, r(v) = b)| = \frac{1}{ab}\binom{2a-2}{a-1}\binom{2b-2}{b-1}\binom{2(n-a-b)}{n-a-b}.$$

*Proof.* We want to count the number of binary plane trees on $n$ leaves with one marked internal vertex $v$, which has $a$ left descendants and $b$ right descendants. We can construct such a tree by gluing together three binary plane trees $A$, $B$, and $C$ on $a$, $b$, and $c = (n-a-b)+1$ leaves respectively, as in Figure 2.7. Specifically, we pick a leaf of tree $C$ to be the marked vertex $v$, then attach tree $A$ as the left descendants of $v$ and tree $B$ as the right descendants of $B$.

It follows that:

$$\begin{aligned}
|\mathcal{T}_{n,v}^{\text{plane}}(a,b)| &= |\mathcal{T}_a^{\text{plane}}||\mathcal{T}_b^{\text{plane}}|(n-a-b+1)|\mathcal{T}_{n-a-b+1}^{\text{plane}}| \\
&= \frac{1}{a}\binom{2a-2}{a-1}\frac{1}{b}\binom{2b-2}{b-1}\frac{(n-a-b+1)}{n-a-b+1}\binom{2(n-a-b)}{n-a-b} \\
&= \frac{1}{ab}\binom{2a-2}{a-1}\binom{2b-2}{b-1}\binom{2(n-a-b)}{n-a-b}. \qquad \square
\end{aligned}$$

## 2.5.2   Bounding the Sum

Let $P(a,b;n)$ be the probability of picking a tree-vertex pair $(T,v)$ from $\mathcal{T}_{n,v}^{\text{plane}}$ with $L(v) = a$, $R(v) = b$. In Proposition 2.5.5, we bound the number we want to compute (average tropical NNI number over pairs of trees on $[n]$ leaves) by the following expectation, which we have

written in terms of $P(a_i, b_i; n)$ and $Q(a_1, b_1, a_2, b_2; n)$.

$$S_n := \sum_{\substack{a_i+b_i+c_i=n \\ a_i, b_i \geq 1, c_i \geq 0}} Q(a_1, b_1, a_2, b_2; n) P(a_1, b_1; n) P(a_2, b_2; n) \qquad (2.11)$$

We will work to bound (2.11) by bounding $P(a, b; n)$ and $Q(a_1, b_1, a_2, b_2; n)$. It will be useful to have some bounds on the central binomial coefficients.

**Proposition 2.5.9** ([38]).

$$\frac{4^n}{\sqrt{\pi \left(n + \frac{1}{3}\right)}} \leq \binom{2n}{n} \leq \frac{4^n}{\sqrt{\pi \left(n + \frac{1}{4}\right)}}$$

**Lemma 2.5.10.** *For $n \geq 1$, we have the bounds*

$$\frac{4^{n-1}}{n\sqrt{\pi \left(n - \frac{2}{3}\right)}} \leq \left|\mathcal{T}_n^{\text{plane}}\right| \leq \frac{4^{n-1}}{n\sqrt{\pi \left(n - \frac{3}{4}\right)}},$$

$$\frac{4^{n-1}(n-1)}{n\sqrt{\pi \left(n - \frac{2}{3}\right)}} \leq \left|\mathcal{T}_{n,v}^{\text{plane}}\right| \leq \frac{4^{n-1}(n-1)}{n\sqrt{\pi \left(n - \frac{3}{4}\right)}}.$$

*Proof.* The bounds are obtained by plugging the bounds from Proposition 2.5.9 into the formula we found for $\left|\mathcal{T}_n^{\text{plane}}\right|$ in Lemma 2.5.6.

$$\left|\mathcal{T}_n^{\text{plane}}\right| = \frac{1}{n}\binom{2n-2}{n-1} \leq \frac{4^{n-1}}{n\sqrt{\pi \left(n - 1 + \frac{1}{4}\right)}} = \frac{4^{n-1}}{n\sqrt{\pi \left(n - \frac{3}{4}\right)}}$$

$$\left|\mathcal{T}_n^{\text{plane}}\right| = \frac{1}{n}\binom{2n-2}{n-1} \geq \frac{4^{n-1}}{n\sqrt{\pi \left(n - 1 + \frac{1}{3}\right)}} = \frac{4^{n-1}}{n\sqrt{\pi \left(n - \frac{2}{3}\right)}}$$

The bounds for $\left|\mathcal{T}_{n,v}^{\text{plane}}\right|$ are obtained through an analogous argument, observing that as we saw in Lemma 2.5.7 $\left|\mathcal{T}_{n,v}^{\text{plane}}\right| = (n-1)\left|\mathcal{T}_n^{\text{plane}}\right|$. $\qquad \square$

**Corollary 2.5.11.**

$$P(a, b; n) \leq \frac{1}{2\pi} \frac{\sqrt{n}}{a \left(a - \frac{3}{4}\right)^{\frac{1}{2}} b \left(b - \frac{3}{4}\right)^{\frac{1}{2}} \left(c + \frac{1}{4}\right)^{\frac{1}{2}}}$$

*Proof.* We can express $P(a, b; n)$ as a fraction of the number of tree-vertex pairs $(T, v) \in$

43

$\mathcal{T}_{n,v}^{\text{plane}}$ with $\ell(v) = a$, $r(v) = b$ over the number of all trees in $\mathcal{T}_{n,v}^{\text{plane}}$.

$$P(a, b; n) = \frac{\left|\mathcal{T}_{n,v}^{\text{plane}}(a, b)\right|}{\left|\mathcal{T}_{n,v}^{\text{plane}}\right|} = \frac{\left|\mathcal{T}_a^{\text{plane}}\right| \left|\mathcal{T}_b^{\text{plane}}\right| (c+1) \left|\mathcal{T}_{c+1}^{\text{plane}}\right|}{\left|\mathcal{T}_{n,v}^{\text{plane}}\right|}$$

Applying the bounds from Lemma 2.5.10 to the expression above yields the claimed inequality.

$$P(a, b; n) \leq \frac{1}{4\pi} \frac{(n-1)\sqrt{n}}{a \left(a - \frac{3}{4}\right)^{\frac{1}{2}} b \left(b - \frac{3}{4}\right)^{\frac{1}{2}} \left(c + \frac{1}{4}\right)^{\frac{1}{2}} n} \leq \frac{1}{2\pi} \frac{\sqrt{n}}{a \left(a - \frac{3}{4}\right)^{\frac{1}{2}} b \left(b - \frac{3}{4}\right)^{\frac{1}{2}} \left(c + \frac{1}{4}\right)^{\frac{1}{2}}}$$

$\square$

**Lemma 2.5.12.** *Let $A_1$ and $A_2$ be subsets of $[n]$, chosen uniformly at random from all subsets of size $a_1$ and $a_2$ respectively. Then,*

$$P(A_1 \cap A_2 \neq \emptyset) \leq \frac{a_1 a_2}{n}$$

*Proof.* The probability that $1 \in A_1 \cap A_2$ is $(a_1/n) \cdot (a_2/n)$. Thus, the probability that at least one element of $[n]$ lies in the intersection of $A_1, A_2$ is at most $n(a_1/n)(a_2/n) = a_1 a_2/n$. $\square$

**Corollary 2.5.13.**

$$Q(a_1, b_1, a_2, b_2) \leq \min\left(\frac{a_1 a_2}{n}, \frac{b_1 b_2}{n}\right) \leq \frac{(a_1 a_2 b_1 b_2)^{1/2}}{n}$$

**Definition 2.5.14.**

$$\tilde{Q}(a_1, b_1, a_2, b_2; n) := \frac{(a_1 a_2 b_1 b_2)^{1/2}}{n} \quad \text{and} \quad \tilde{Q}_0(a, b; n) := \left(\frac{ab}{n}\right)^{\frac{1}{2}}.$$

Corollary 2.5.13 says

$$Q(a_1, b_1, a_2, b_2; n) \leq \tilde{Q}(a_1, b_1, a_2, b_2; n) = \tilde{Q}_0(a_1, b_1; n)\tilde{Q}_0(a_2, b_2; n).$$

We are not yet ready to compute the whole sum in (2.11), since we do not have bounds for $P(a, b; n)$ in some edge cases, but we can bound a significant subsum.

**Proposition 2.5.15.**

$$S_n \leq \left( \sum_{\substack{a+b\leq n \\ a,b\geq 1}} \frac{1}{\left(a - \frac{1}{2}\right)\left(b - \frac{1}{2}\right)\left(c + \frac{1}{4}\right)^{\frac{1}{2}}} \right)^2 .$$

*Proof.* We plug in the bounds we computed in Corollary 2.5.11 and Corollary 2.5.13 and simplify. The key idea here is that the symmetry of the upper bound for $Q$ allows us to express the sum in six variables as the square of a sum in three variables.

$$S_n = \sum_{\substack{a_i+b_i+c_i=n \\ a_i,b_i\geq 1,c_i\geq 0}} Q(a_1, b_1, a_2, b_2)P(a_1, b_1; n)P(a_2, b_2; n)$$

$$\leq \sum_{\substack{a_i+b_i+c_i=n \\ a_i,b_i\geq 1,c_i\geq 0}} \tilde{Q}(a_1, b_1, a_2, b_2; n)P(a_1, b_1; n)P(a_2, b_2; n)$$

$$= \left( \sum_{\substack{a+b\leq n \\ a,b\geq 1}} \tilde{Q}_0(a, b; n)P(a, b; n) \right)^2$$

Now we expand and bound the terms of the sum in the square above using the bound on $P(a, b; n)$ from Corollary 2.5.11.

$$\tilde{Q}_0(a, b; n)P(a, b; n) \leq \frac{a^{\frac{1}{2}}b^{\frac{1}{2}}}{\sqrt{n}} \frac{\sqrt{n}}{a\left(a - \frac{3}{4}\right)^{\frac{1}{2}} b\left(a - \frac{3}{4}\right)^{\frac{1}{2}} \left(c + \frac{1}{4}\right)^{\frac{1}{2}}}$$

$$= \frac{1}{a^{\frac{1}{2}}\left(a - \frac{3}{4}\right)^{\frac{1}{2}} b^{\frac{1}{2}}\left(b - \frac{3}{4}\right)^{\frac{1}{2}} \left(c + \frac{1}{4}\right)^{\frac{1}{2}}}$$

We can simplify further by applying the arithmetic-geometric mean inequality:

$$\frac{1}{a^{\frac{1}{2}}\left(a - \frac{3}{4}\right)^{\frac{1}{2}}} \leq \frac{1}{a - \frac{1}{2}}$$

Thus we derive the following bound for $S_n$.

$$S_n \leq \left( \sum_{\substack{a+b\leq n \\ a,b\geq 1}} \frac{1}{\left(a - \frac{1}{2}\right)\left(b - \frac{1}{2}\right)\left(c + \frac{1}{4}\right)^{\frac{1}{2}}} \right)^2 . \qquad \square$$

The following lemma will be useful in completing the computation of the sum on the right hand side of Proposition 2.5.15.

**Lemma 2.5.16.**
$$\sum_{\substack{a+b=d \\ a,b\geq 1}} \frac{1}{\left(a-\frac{1}{2}\right)\left(b-\frac{1}{2}\right)} \leq \frac{2\left(2+\log\left(d-\frac{3}{2}\right)\right)}{d-1}.$$

*Proof.* First, when $a+b=d$ we have:
$$\frac{1}{\left(a-\frac{1}{2}\right)\left(b-\frac{1}{2}\right)} = \frac{1}{d-1}\left(\frac{1}{a-\frac{1}{2}} + \frac{1}{b-\frac{1}{2}}\right).$$

Also,
$$\sum_{a=1}^{d-1} \frac{1}{a-\frac{1}{2}} = 2 + \frac{1}{2-\frac{1}{2}} + \cdots + \frac{1}{d-\frac{3}{2}} \leq 2 + \log\left(d-\frac{3}{2}\right).$$

It follows that
$$\sum_{\substack{a+b=d \\ a,b\geq 1}} \frac{1}{\left(a-\frac{1}{2}\right)\left(b-\frac{1}{2}\right)} = \frac{1}{d-1}\sum_{\substack{a+b=d \\ a,b\geq 1}}\left(\frac{1}{a-\frac{1}{2}} + \frac{1}{b-\frac{1}{2}}\right)$$
$$= \frac{2}{d-1}\sum_{a=1}^{d-1}\frac{1}{a-\frac{1}{2}}$$
$$\leq \frac{2\left(2+\log\left(d-\frac{3}{2}\right)\right)}{d-1}. \qquad \square$$

Recall the following inequality.

**Lemma 2.5.17** (Chebyshev's Sum Inequality, [20]). *If $a_n$ is a decreasing sequence, and $b_n$ is an increasing sequence, then*
$$\frac{1}{n}\sum_{k=1}^{n} a_k b_k \leq \left(\frac{1}{n}\sum_{k=1}^{n} a_k\right)\left(\frac{1}{n}\sum_{k=1}^{n} b_k\right)$$

**Proposition 2.5.18.** $\sqrt{S_n}$ *is* $O\left((\log(n))^2/n^{\frac{1}{2}}\right)$.

*Proof.* We first bound the sum inside the square by applying Lemma 2.5.16.

$$\sum_{\substack{a+b\leq n \\ a,b\geq 1}} \frac{1}{\left(a-\frac{1}{2}\right)\left(b-\frac{1}{2}\right)\left(c+\frac{1}{4}\right)^{\frac{1}{2}}} = \sum_{c=0}^{n-2}\left(c+\frac{1}{4}\right)^{-\frac{1}{2}}\sum_{\substack{a+b=n-c \\ a,b\geq 1}}\frac{1}{\left(a-\frac{1}{2}\right)\left(b-\frac{1}{2}\right)}$$

$$\leq \sum_{c=0}^{n-2}\frac{2\left(2+\log\left(n-c-\frac{3}{2}\right)\right)}{\left(c+\frac{1}{4}\right)^{\frac{1}{2}}(n-c-1)}$$

$$\leq \sum_{c=0}^{n-2}\frac{2\left(2+\log\left(n-c-1\right)\right)}{\left(c+\frac{1}{4}\right)^{\frac{1}{2}}(n-c-1)} \qquad (*)$$

The sequence $\left(c+\frac{1}{4}\right)^{-\frac{1}{2}}$ is decreasing in $c$, and the sequence $(2+\log(n-c-1))/(n-c-1)$ is increasing in $c$ on the interval $\left[0, n-\frac{1}{e}-1\right]$. Therefore, we can apply Chebyshev's sum inequality in Lemma 2.5.17 to bound the sum in $(*)$.

$$\sum_{c=0}^{n-2}\frac{1}{\left(c+\frac{1}{4}\right)^{\frac{1}{2}}}\frac{2\left(2+\log\left(n-c-1\right)\right)}{(n-c-1)} \leq \frac{1}{n-1}\left(\sum_{c=0}^{n-2}\left(c+\frac{1}{4}\right)^{-\frac{1}{2}}\right)$$

$$\cdot\left(\sum_{c=0}^{n-2}\frac{2(2+\log(n-c-1))}{(n-c-1)}\right)$$

$$\sim \frac{1}{n-1}\left(\sum_{c=0}^{n-2}\left(c+\frac{1}{4}\right)^{-\frac{1}{2}}\right)$$

$$\cdot\left(\sum_{c=0}^{n-2}\frac{\log(n-c-1)}{(n-c-1)}\right)$$

$$\sim \frac{n^{\frac{1}{2}}(\log(n))^2}{n-1} \sim \frac{(\log(n))^2}{n^{\frac{1}{2}}}$$

In the third line, we use asymptotics derived by thinking of the sums are left-hand or right-hand Riemann sums of monotone functions.

$$\int_0^{n-1}\left(x+\frac{1}{4}\right)^{-\frac{1}{2}}dx \leq \sum_{c=0}^{n-2}\left(c+\frac{1}{4}\right)^{-\frac{1}{2}} \leq 2+\int_0^{n-2}\left(x+\frac{1}{4}\right)^{-\frac{1}{2}}dx \sim n^{\frac{1}{2}}$$

$$\int_{-1}^{n-2}\frac{\log(n-x-1)}{n-x-1}dx \leq \sum_{c=0}^{n-2}\frac{\log(n-c-1)}{(n-c-1)} \leq \int_0^{n-2}\frac{\log(n-x-1)}{n-x-1}dx \sim (\log(n))^2 \qquad \square$$

*Proof of Theorem 2.5.1.* The results in this section prove the following inequality:

$$E\left(\#\Pi(T_1,T_2)\mid T_1,T_2\in\mathcal{T}_{[n]}\right) \leq 2(n-1)^2 S_n.$$

Furthermore, $S_n$ is $O\left((\log(n))^4/n\right)$ by Proposition 2.5.18. It follows that $(n-1)^2 S_n$ is $O\left(n\,(\log(n))^4\right)$, as claimed. $\qquad\square$

# CHAPTER 3

# Weighted Tropical Fermat-Weber Points

## 3.1   Introduction

This chapter is joint work with Mark Curiel and appears in the preprint "The tropical polytope is the set of all weighted tropical Fermat-Weber points" [13].

The Fermat-Weber problem was first posed by Fermat before 1640 to compute the Fermat-Weber point when $P$ is a triangle in Euclidean space. It was solved geometrically by Evangelista Torricelli in 1645. We note that the sum $\sum_{i=1}^{m} d(x, v_i)$ assumes that the distances are weighted equally. Motivated by unequal attracting forces on particles, a generalization of the problem was introduced and solved by Thomas Simpson in 1750, later popularized by Alfred Weber in 1909, by considering weighted distances. In that spirit, this paper is concerned with generalizing the result by Comăneci and Joswig, namely, we are interested in locating the specific cells for which the Fermat-Weber points live in $P$ by minimizing the sum $\sum_{i=1}^{m} w_i d(x, v_i)$ for some choice of positive real weights $w_i$. Our main theorem states that the Fermat-Weber set is a cell of the tropical convex hull $P$ and that, by choosing weights appropriately, it can be any cell of $P$.

**Theorem 3.1.1.** *Given data points $v_1, \ldots, v_m$, the collection of asymmetric tropical weighted Fermat-Weber points over all possible positive real weights $w_i$ is $P = \text{tconv}\{v_1, \ldots, v_m\}$.*

This chapter is organized as follows. In Section 3.2, we begin by formulating the Fermat-Weber problem in the language of tropical convexity. Then we provide further background in tropical geometry and polyhedral geometry necessary to understand our approach. Of particular interest we recall the Cayley trick which gives a correspondence between mixed subdivisions of the Minkowski sum of polytopes and subdivisions of the corresponding Cayley polytope. In Section 3.3, we prove our main result Theorem 3.1.1 as a corollary to Theorem 3.3.2.

## 3.2 Background

### 3.2.1 Fermat-Weber Problems

A Fermat-Weber problem is a geometric problem seeking the median of a collection of data points $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \subset X$, where $X$ is a metric space with distance $d(\mathbf{x}, \mathbf{y})$. We are particularly interested in the Fermat-Weber points for a collection of data points in $\mathbb{R}^n/\mathbb{R}\mathbb{1}$, where the points could represent phylogenetic trees. The goal of this section is to introduce the Fermat-Weber problem, and reframe a tropical version as a problem on Newton polytopes.

In general, the median of a collection of points is not unique and hence we seek the set of all such medians, called the called *Fermat-Weber set*. The medians belonging to the Fermat-Weber set are called *Fermat-Weber points*. Formally, the Fermat-Weber points are the points $\mathbf{x} \in X$ minimizing the sum in (3.1).

**Definition 3.2.1.** The *Fermat-Weber* points on the data $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \subset X$ are the points $\mathbf{x} \in X$ minimizing the following sum

$$\text{FW}(V) := \frac{1}{m} \sum_{i=1}^{m} d(\mathbf{x}, \mathbf{v}_i). \tag{3.1}$$

In this paper, we are interested in a variant of the Fermat-Weber problem, called the *weighted Fermat-Weber problem*. This new problem seeks the points $\mathbf{x}$ minimizing the sum in Equation (3.2), where the weights $w_i$ are positive real numbers.

$$\text{FW}(V, \mathbf{w}) := \frac{1}{m} \sum_{i=1}^{m} w_i d(\mathbf{x}, \mathbf{v}_i). \tag{3.2}$$

We will use the asymmetric tropical distance first defined by Comăneci and Joswig in [7].

**Definition 3.2.2.** The asymmetric tropical distance, $d_\Delta(\mathbf{x}, \mathbf{y})$ is:

$$d_\Delta(\mathbf{x}, \mathbf{y}) := n \max_{i \in [n]} \{x_i - y_i\} + \sum_{i \in [n]} (y_i - x_i). \tag{3.3}$$

When the points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n/\mathbb{R}\mathbb{1}$ are given by their unique representative in $H_0$ (the subspace where the coordinates sum to zero), the metric $d_\Delta(\mathbf{x}, \mathbf{y})$ can be simplified to the following

$$d_\Delta(\mathbf{x}, \mathbf{y}) := n \max_{i \in [n]} \{x_i - y_i\}, \quad \mathbf{x}, \mathbf{y} \in H_0. \tag{3.4}$$

Note that $d_\Delta(x, y)$ is invariant under independent scalar multiplication of the input vectors, so it is well-defined on $\mathbb{R}^n/\mathbb{R}1$. From now on, we will assume that all points in $\mathbb{R}^n/\mathbb{R}1$ are given by their representative in $H_0$. With this assumption, the distance to a point $v_i$ can be reinterpreted as a power of a tropical linear equations, and the sum in equation 3.2 can be realized as a tropical product of tropical linear functions (possibly with real exponents).

The distance to a point $v_i$, denote by $f_{v_i}(\mathbf{x})$ is

$$f_{v_i}^n(\mathbf{x}) := d_\Delta(x, v_i) = n \max_j \{x_j - v_{ij}\} = \left( \bigoplus_{j=1}^{n} -v_{ij} \odot x_j \right)^n. \tag{3.5}$$

It follows that the sum in (eq. (3.2)) for $d = d_\Delta$ is

$$\frac{1}{m} \sum_{i=1}^{m} w_i d_\Delta(\mathbf{x}, \mathbf{v}_i) = \frac{1}{m} \sum_{i=1}^{m} w_i f_{v_i}^{nw_i}(\mathbf{x}) = \frac{n}{m} \bigodot_{i=1}^{m} f_{v_i}^{w_i}. \tag{3.6}$$

**Definition 3.2.3.** We define *the tropical signomial associated to data $V$ with weights* $\mathbf{w}$, $f_{V,\mathbf{w}}$, to be the following tropical function:

$$f_{V,\mathbf{w}}(\mathbf{x}) := \bigodot_{i=1}^{m} f_{v_i}^{w_i} = \bigodot_{i=1}^{m} \left( \bigoplus_{j=1}^{n} -v_{ij} \odot x_j \right)^{w_i}.$$

The tropical hypersurface $\text{trop}V(f_{v_i})$ is a tropical hyperplane centered at $v_i$; it is the codimension-1 skeleton of the normal fan of the standard simplex $\Delta^{n-1}$. By lemma 1.3.6, the hypersurface $\text{trop}V(f_{V,\mathbf{w}})$ is the union of tropical hyperplanes centered at the data points $v_i$. The Newton polytope of $f_{V,\mathbf{w}} = \bigodot_{i=1}^{m} f_{v_i}^{w_i}$ is $\sum_{i=1}^{m} w_i \cdot \Delta^{n-1}$.

**Example 3.2.4.** The polynomial $f_{V,\mathbf{w}}(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^3/\mathbb{R}1$ and $\mathbf{w} \in \mathbb{R}^2$ has nine terms for generic $V$ and $\mathbf{w}$.

$$f_{V,\mathbf{w}} = (-v_{11} \odot x_1 \oplus -v_{12} \odot x_2 \oplus -v_{13} \odot x_3)^{w_1}$$
$$\odot \left(-v_{21} \odot x_1 \oplus -v_{22} \odot x_2 \oplus -v_{23} \odot x_3\right)^{w_2}$$
$$= (-v_{11} \odot -v_{21}) \odot x_1^{w_1+w_2} \oplus (-v_{11} \odot -v_{22}) \odot x_1^{w_1} x_2^{w_2} \oplus (-v_{11} \odot -v_{23}) \odot x_1^{w_1} x_3^{w_2}$$
$$\oplus (-v_{12} \odot -v_{21}) \odot x_1^{w_2} x_2^{w_1} \oplus (-v_{12} \odot -v_{22}) \odot x_2^{w_1+w_2} \oplus (-v_{12} \odot -v_{23}) \odot x_2^{w_1} x_3^{w_2}$$
$$\oplus (-v_{13} \odot -v_{21}) \odot x_1^{w_2} x_3^{w_1} \oplus (-v_{13} \odot -v_{22}) \odot x_2^{w_2} x_3^{w_1} \oplus (-v_{13} \odot -v_{23}) \odot x_3^{w_1+w_2}.$$

Its Newton polytope is $(w_1 + w_2) \cdot \Delta^2$, and it is depicted in the lower right of Figure 3.3, with $w_2 > w_1$. The image on the lower left of Figure 3.3 is the Newton polytope in the special case where $w_1 = w_2 = 1$.

We now apply the results of the previous subsection to translate the problem of optimizing $f_{V,\mathbf{w}}$ into a problem on $\text{Newt}(f_{V,\mathbf{w}})$. Since $f_{V,\mathbf{w}}$ is a function from $\mathbb{1}^{\perp}$ rather than $\mathbb{R}^n$, we need the following result to apply the results of the previous subsection.

**Proposition 3.2.5.** *Given a max-plus tropical polynomial $f : \mathbb{1}^{\perp} \to \mathbb{R}$, let $\underline{N}$ be the subdivision of $\text{Newt}(f)$ induced by the coefficients of $f$. The minimum of $f$ is achieved on the cell dual to the cell of the Newton polytope containing $\lambda\mathbb{1}$ for any $\lambda \in \mathbb{R} \setminus \{0\}$.*

*Proof.* Consider the following isomorphism $\mathbb{R}^n/\mathbb{R}\mathbb{1} \cong \mathbb{1}^{\perp}$.

$$\psi(x_1, \ldots, x_{n-1}) = \left( x_1, \ldots, x_{n-1}, -\sum_{i=1}^{n-1} x_i \right) \tag{3.7}$$

The Newton polytope of $f$ lives in $\left(\mathbb{1}^{\perp}\right)^*$. The dual function of $\psi$ is below.

$$\psi^*(x_1, \ldots, x_n) = (x_1 - x_n, \ldots, x_{n-1} - x_n) \tag{3.8}$$

Then $\psi^*(\lambda\mathbb{1}) = \lambda(1 - 1, \ldots, 1 - 1) = \overline{\mathbf{0}}$. Lemma 3.2.7 says that for a tropical function $g : \mathbb{R}^n \to \mathbb{R}$, the linear piece of $g$ whose dual cell contains $\overline{\mathbf{0}}$ is the linear piece achieving the minimum. Combining this result with the map $\psi$, it follows that $f : \mathbb{1}^{\perp} \to \mathbb{R}$ achieves its minimum on the linear piece dual to the cell containing $\lambda\mathbb{1}_n$. $\qquad\square$

In this chapter, we will use the specific representative of $v \in \mathbb{R}^n/\mathbb{R}\mathbb{1}$ where the sum of the coordinates is zero. We denote by $H_0$ the hyperplane where these points are located:

$$H_0 := \left\{ \mathbf{z} \in \mathbb{R}^n \ \middle| \ \mathbf{z} \cdot \mathbb{1} = \sum_i z_i = 0 \right\} \subset \mathbb{R}^n.$$

Each point in $\mathbb{R}^n/\mathbb{R}\mathbb{1}$ has a unique representative in $H_0$.

## 3.2.2 Optimization

A tropical max-plus signomial is a piecewise-linear, continuous, convex function on $\mathbb{R}^n$. For a convex function, any local minimum is a global minimum. This minimum can be identified by locating the tangent plane with zero slope, which we formalize using subgradients.

**Definition 3.2.6** ([34, §3.1.5])**.** Given a convex function $f : \mathbb{R}^n \to \mathbb{R}$, the *subdifferential* of $f$ at $x$ is:

$$\partial_f(x) := \{u \in \mathbb{R}^n \mid \forall z \in \text{dom}(f), f(z) \geq f(x) + u^\top \cdot (z - x)\}. \tag{3.9}$$

A *subgradient* of $f$ at $x$ is any element of $\partial_f(x)$.

The subdifferential of any function is a closed convex set. If $f$ is convex and differentiable at $x$, then the subdifferential of $f$ at $x$ is a singleton. In particular, if $f$ is linear then the subdifferential contains only the slope of $f$. And if $f$ is piecewise-linear, then $\partial_f(x)$ is constant on the linear pieces of $f$.

**Lemma 3.2.7** ([34, Theorem 3.1.15]). *For any function $f$, the subdifferential at $x$ contains $\overline{\mathbf{0}}$ if and only if $x$ is a global minimizer for $f$.*

*Proof.* By definition, $\overline{\mathbf{0}} \in \partial_f(x)$ if and only if

$$\forall z \in \mathrm{dom}(f), f(z) \geq f(x) + \overline{\mathbf{0}}^\top \cdot (z - x) \iff \forall z \in \mathrm{dom}(f), f(z) \geq f(x),$$

which is if and only if $x$ is a global minimizer for $f$. $\qquad\square$

**Example 3.2.8.** Let $f(x) = 1 \oplus 3 \odot x \oplus -1 \odot x^{\sqrt{2}}$. The subdifferential of $f(x)$ on each linear piece is the slope of that piece. The subdifferential at $x = -2$ is $\partial_f(-2) = [0, 1]$, and the subdifferential at $x = 4(1 + \sqrt{2})$ is $[1, \sqrt{2}]$. Note that 0 is in the subdifferential of the constant (left-most) linear piece, and this is where the global minimum of $f(x)$ is achieved. See fig. 1.5.

For a tropical polynomial $f : \mathbb{R}^n \to \mathbb{R}$, subdifferentials of linear pieces of $f$ are encoded by a subdivision of $\mathrm{Newt}(f)$. It is this connection that will allow us to convert the problem of optimizing $f$ into a polyhedral geometry problem.

Moreover, the proof in Proposition 3.1.6 in [29] shows that the subdifferential of a linear piece of $f$ consists of the points in the corresponding cell of $\underline{N}$.

**Lemma 3.2.9.** *The minimum of a max-plus tropical polynomial is achieved on the cell dual to the cell of the Newton polytope containing $\overline{\mathbf{0}}$.*

*Proof.* Let $f = \sum_i c_i x^{\alpha_i}$ be a max-plus tropical polynomial (so $f$ is a piecewise-linear convex function). Let $\underline{N}$ be the regular subdivision of $\mathrm{Newt}(f)$ induced by the weighting $w(\alpha_i) = c_i$. Let $L$ be a linear piece of $f$, and let $N_L$ be the cell dual to it in $\underline{N}$. If $\overline{\mathbf{0}} \in N_L$, then by Proposition 1.3.8 $\overline{\mathbf{0}}$ is in the subdifferential of $f$ at any point in $L$. It then follows from Lemma 3.2.7 that the minimum of $f$ is achieved on $L$. $\qquad\square$

### 3.2.3 Cayley Polytopes

**Definition 3.2.10.** Let $P_1, \ldots, P_m$ be polytopes in $\mathbb{R}^n$. The Cayley polytope, denoted $\mathrm{Cayley}(P_1, \ldots, P_m)$, is the convex hull of $\bigcup_{i=1}^m e_i \times P_i$ in $\mathbb{R}^m \times \mathbb{R}^n$. In the special case where $P_1 = P_2 = \cdots = P_m$, the Cayley polytope is $\Delta^{r-1} \times P$. For an example see Figure 3.3.

**Definition 3.2.11.** The *Minkowski sum* of $P_1, \ldots, P_m$ is the set

$$P = P_1 + \cdots + P_m = \{\mathbf{p}_1 + \cdots \mathbf{p}_m \mid \mathbf{p}_i \in P_i\} \subset \mathbb{R}^n.$$

The Minkowski sum of polytopes is indeed a polytope since its vertices are necessarily sums of vertices of the summands, hence the Minkowski sum is a convex hull of a finite set.

**Example 3.2.12** (Minkowski Sum). Let $A_1 = \{a, b, c, d\}$ and $A_2 = \{e, f, g\}$ be the vertex sets in $\mathbb{R}^2$ of $P_1$ and $P_2$ respectively. The Minkowski sum $P_1 + P_2$ is the convex hull of $\{a+e, b+f, c+f, c+g, d+g\}$ and is shown in Figure 3.1.
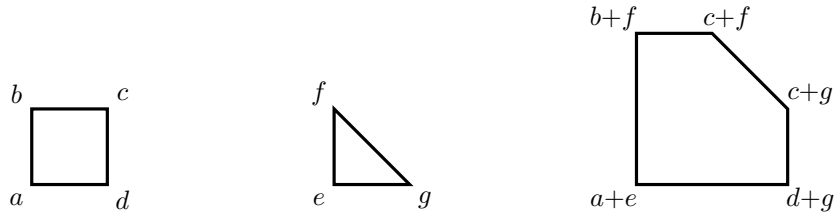


Figure 3.1: A square (left), a triangle (middle), and their Minkowski sum (right).

**Definition 3.2.13.** A *cell* of the minkowski sum $P = P_1 + \ldots + P_m$ is a tuple $C = (C_1, \ldots, C_m)$ where $C_i \subseteq A_i$ for all $i$.

*Remark* 3.2.14. Each cell $(C_1, C_2, \ldots, C_m)$ gives a polytope $\sum C_i \subseteq \sum P_i$, and we will often abuse notation by identifying $(C_1, C_2, \ldots, C_m)$ with this sum. If $C$ and $C'$ are two such cells, then $C \cap C'$ refers to the intersection $(\sum_i \mathrm{conv}(C_i)) \cap (\sum_i \mathrm{conv}(C_i'))$. Additionally, a cell of $P$ may be the Minkowski sum of two (or more) different ordered $m$-tuples, and we would like to consider these as different cells.

The Cayley trick is a correspondence between mixed subdivisions of the Minkowski sum $P_1 + \cdots + P_m$ and subdivisions of $\mathrm{Cayley}(P_1, \ldots, P_m)$, illustrated in Figure 3.3. The top right polytope in Figure 3.3 is $\mathrm{Cayley}(\Delta^2, \Delta^2) \cong \Delta^1 \times \Delta^2$. Explicitly, a subdivision of the Cayley polytope gives rise to a subdivision of the Minkowski sum after forgetting the first $m$ coordinates. For more details, see [44, Section 5] for coherent/regular subdivisions, and [23, Theorem 3.1] for all subdivisions. Although often stated as a theorem, we will use the Cayley trick to define *mixed subdivisions*.

**Definition 3.2.15.** A *mixed subdivision* of $P = (P_1, \ldots, P_m)$ is a collection of cells $C^1, \ldots, C^k$ so that $\{\mathrm{Cayley}(\mathrm{conv}(C_1^j), \ldots, \mathrm{conv}(C_m^j)) \mid j = 1, \ldots, k\}$ is a subdivision of $\mathrm{Cayley}(P_1, \ldots, P_m)$.
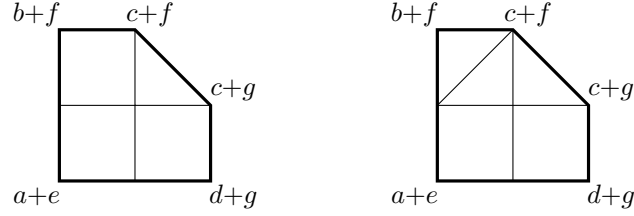
Figure 3.2: A mixed subdivision of $P_1 + P_2$ (left) and a subdivision of $P_1 + P_2$ that is not mixed (right).

**Example 3.2.16.** A subdivision of the polytope $P_1 + P_2$ of Example 3.2.12 consists of a collection of cells $\{C^1, C^2, C^3, C^4\}$ with

$$
\begin{aligned}
C^1 &= \operatorname{conv}\{a, b, c, d\} + \operatorname{conv}\{e\} & C^3 &= \operatorname{conv}\{b, c\} + \operatorname{conv}\{e, f\} \\
C^2 &= \operatorname{conv}\{c, d\} + \operatorname{conv}\{e, g\} & C^4 &= \operatorname{conv}\{c\} + \operatorname{conv}\{e, f, g\}
\end{aligned}
$$

It is mixed since the cells Cayley($\{a, b, c, d\}, \{e\}$), Cayley($\{b, c\}, \{e, f\}$), Cayley($\{c, d\}, \{e, g\}$), Cayley($\{c\}, \{e, f, g\}$) give a subdivision of Cayley($P_1, P_2$). Another subdivision of $P_1 + P_2$ can be achieved with the cells

$$
\begin{aligned}
C^1 &= \operatorname{conv}\{a, b, c, d\} + \operatorname{conv}\{e\} \\
C^2 &= \operatorname{conv}\{c, d\} + \operatorname{conv}\{e, g\} \\
C^3 &= \operatorname{conv}\{c\} + \operatorname{conv}\{e, f, g\} \\
C^4 &= \operatorname{conv}\{a, b, c\} + \operatorname{conv}\{f\} \\
C^5 &= \operatorname{conv}\{a, c, d\} + \operatorname{conv}\{f\}
\end{aligned}
$$

However, it is not mixed since for example the cells Cayley($\{a, b, c, d\}, \{e\}$) and Cayley($\{a, b, c\}, \{f\}$) intersect on their interior.

We will refer to Definition 3.2.15 above as the *combinatorial Cayley trick*. In addition to the combinatorial correspondence above, there is also an explicit geometric correspondence between a subdivision of the Cayley polytope and a mixed subdivision, sometimes called the *geometric Cayley trick*.

**Theorem 3.2.17** ([23, Theorem 3.1])**.** *Let $\underline{C}$ be a subdivision of* Cayley($P_1, \ldots, P_m$). *Then the corresponding mixed subdivision of $P = \sum_{i=1}^m P_i$ is $\underline{P} = n \cdot \underline{C} \cap \{\frac{1}{m}\mathbb{1}_m\} \times \mathbb{R}^n$.*

Proposition 3.2.18 is essentially stated in [39, §1.3]. It allows us to think about mixed subdivisions of a *weighted* Minkowski sum, $P^{\mathbf{w}} = \sum_{i=1}^m w_i P_i$, in terms of the Cayley polytope

55

without weights, $\text{Cayley}(P_1, \ldots, P_m)$, by slicing at $\{\mathbf{w}\} \times \mathbb{R}^n$. We provide a proof that explicitly states the map that induces the bijection, which we will use later in the paper.

**Proposition 3.2.18.** *Let $C = Cayley(P_1, \ldots, P_m)$, and let $D = cone(C)$ denote the cone over it. Let $C^{\mathbf{w}}$, $D^{\mathbf{w}}$, and $P^{\mathbf{w}}$ denote the corresponding weighted versions, for $\mathbf{w} \in \mathbb{R}^m$ with $|\mathbf{w}| = 1$. Let $\lambda$ be a piecewise-linear convex function on $Cayley(w_1 P_1, \ldots, w_m P_m)$, and let $g(\mathbf{x}, \mathbf{y}) = (w_1 x_1, \ldots, w_m x_m, \mathbf{y})$. If $\lambda' = \lambda(g^{-1}(x, y))$, then the following diagram commutes.*

$$
\begin{array}{ccc}
\underline{D}^{\mathbf{w}}_\lambda & \xrightarrow{\quad g \quad} & \underline{D}_{\lambda'} \\
\uparrow & & \uparrow {\scriptstyle \cap\{\sum_{i=1}^m x_i = 1\}} \\
\underline{C}^{\mathbf{w}}_\lambda & \dashrightarrow & \underline{C}_{\lambda'}
\end{array}
$$

$$n \cdot \left(\left(\{\tfrac{1}{m}\} \times \mathbb{R}^n\right) \cap \bullet\right) \searrow \quad \swarrow \left(\bullet \cap (\{\mathbf{w}\} \times \mathbb{R}^n)\right)$$

$$\underline{P}^{\mathbf{w}}_\lambda$$

*Proof.* The subdivision $\underline{C}^{\mathbf{w}}_\lambda$ induces a subdivsion $\underline{D}^{\mathbf{w}}_\lambda$. The function $g$ is an invertible linear function. In particular, this means $g$ preserves convexity, dimension, and the containment relations within a polyhedral subdivision. Thus, $\underline{D}^{\mathbf{w}}_\lambda$ induces a subdivision $\underline{D}_{\lambda'}$ via $\lambda'(z) = \lambda(g^{-1}(z))$. This in turn induces a subdivision $\underline{C}_{\lambda'}$ with the weightings $\lambda'(e_i, v_{ij}) = \lambda(e_i, w_i v_{ij})$, which is combinatorially equivalent to $\underline{C}^{\mathbf{w}}_{\lambda'}$. Moreover, $n \cdot \underline{C}^{\mathbf{w}}_{\lambda'} \cap \left(\{\tfrac{1}{m}\} \times \mathbb{R}^n\right)$, and $\underline{C}_{\lambda'} \cap (\{\mathbf{w}\} \times \mathbb{R}^n)$ give the same subdivision of $\underline{P}^{\mathbf{w}}_\lambda$. $\qquad \square$

### 3.2.4 Decomposing Tropical Hypersurfaces

Let $f = \bigodot_{i=1}^m f_i$ be a tropical signomial that factors into a product of tropical signomials. The following theorem tells us how to compute the tropical hypersurface $\text{tropV}(f)$ in terms of a subdivision of they Cayley polytope. Let $P_i = \text{Newt}(f_i)$, $P = \sum_i P_i$, and write $f_i = \sum c_{i,\alpha} \mathbf{x}^\alpha$.

**Theorem 3.2.19** (Corollary 4.9 in [24])**.** *Let $\underline{C}$ be the regular subdivision of $Cayley(P_1, \ldots, P_m)$ induced by the weights $w(e_i, \alpha) = c_{i,\alpha}$. Then the mixed subdivision of $P$ corresponding to $\underline{C}$ coincides with the regular subdivsion of $P$ induced by the coefficients of $f$.*

Recall that in the weighted tropical Fermat-Weber problem, $f_{V,\mathbf{w}}$ factors into linear pieces, and so theorem 3.2.19 applies.

**Corollary 3.2.20.** *Let $\underline{N}$ be the subdivision of $Newt(f_{V,\mathbf{w}})$ induced by the coefficients of $f_{V,\mathbf{w}}$. Then $\underline{N} = \underline{C} \cap \{\tfrac{1}{m} \mathbb{1}_m\} \times \mathbb{R}^n$. In particular, the cell of $\underline{N}$ containing $\lambda \mathbb{1}_n$ corresponds to the cell of $\underline{C}$ containing $\left(\tfrac{1}{m} \mathbb{1}_m, \tfrac{|\mathbf{w}|}{mn} \mathbb{1}_n\right)$.*

*Proof.* The point $(\frac{1}{m}\mathbb{1}_m, \frac{|\mathbf{w}|}{mn}\mathbb{1}_n)$ is the barycenter of $\mathrm{Cayley}(P_1, \ldots, P_m)$, so in particular, it lies in $\mathrm{Cayley}(P_1, \ldots, P_m) \cap \{\frac{1}{m}\mathbb{1}_m\} \times \mathbb{R}^n$. Apply theorem 3.2.19 and then theorem 3.2.17. $\quad\square$

**Proposition 3.2.21.** *Let $P_i = Newt(f_{v_i})$, and let $\underline{C}'$ be the subdivision of the Cayley polytope, $Cayley(P_1, \ldots, P_m)$, induced by $w(\alpha_j, e_i) = -v_{ij}$. Then the subdivision of $P_{\mathbf{w}} = \sum_{i=1}^{m} w_i P_i$ induced by the coefficients of $f_{V,\mathbf{w}}$ is $\underline{C}' \cap \{\mathbf{w}\} \times \mathbb{R}^n$. In particular, the cell of $\underline{P}$ containing $\mathbb{1}_n$ corresponds to the cell of $\underline{C}'$ containing $(\mathbf{w}, \frac{1}{n}\mathbb{1}_n)$.*

*Proof.* Apply proposition 3.2.18 to corollary 3.2.20. $\quad\square$
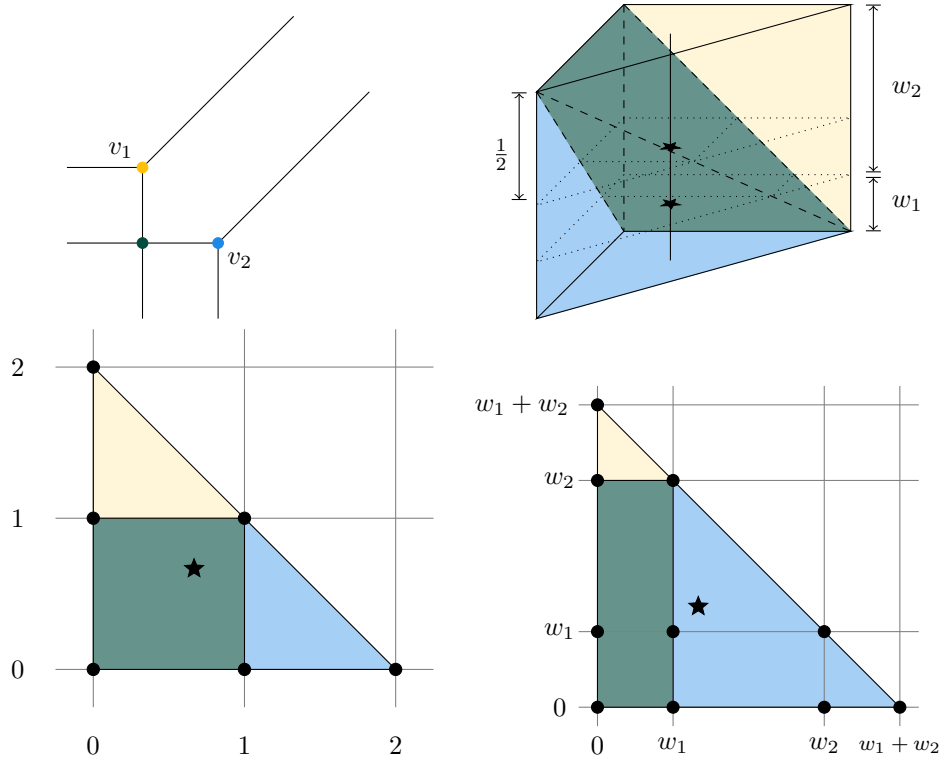


Figure 3.3: Subdivisions with weightings. Clockwise starting on top left: Two tropical hyperplanes in $\mathbb{TR}^2$, the corresponding regular subdivision of $\Delta^1 \times \Delta^2$, the corresponding mixed subdivision of $(w_1 + w_2)\Delta^2$ (weighted FW problem), and the corresponding mixed subdivision of $2\Delta^2$ (unweighted FW problem).

It turns out that the tropical convex hull of the data points coincides with the bounded part of the tropical hypersurface $\mathrm{tropV}(f_V)$.

**Theorem 3.2.22** (Theorem 5.2.11 in [29]). *The bounded part of the max-plus tropical hypersurface $f_V$ is $tconv^{\min}(v_1, \ldots, v_m)$.*

We now see that $f_V$ and $f_{V,\mathbf{w}}$ define the same tropical hypersurface. Thus, the bounded part $\mathrm{tropV}(f_{V,\mathbf{w}})$ is the tropical convex hull of the $v_i$'s.

**Lemma 3.2.23.** *For any* $\mathbf{w} \in \mathbb{R}^m$, *trop* $V(f_V) = trop\, V(f_{V,\mathbf{w}})$.

*Proof.* By lemma 1.3.6, $\mathrm{tropV}(f_v^w) = \mathrm{tropV}(f_v)$ for any $v \in \mathbb{1}^\perp$, and any $w > 0$. Applying *lemma* 1.3.6 to $f_V$ and $f_{V,\mathbf{w}}$, it follows that

$$\mathrm{tropV}(f_V) = \bigcup_{i=1}^{m} \mathrm{tropV}(f_{v_i}) = \mathrm{tropV}(f_{V,\mathbf{w}}). \qquad \square$$

**Corollary 3.2.24.** *The bounded part of the tropical hypersurface* $f_{V,\mathbf{w}}$ *is the min-tropical convex hull of the* $v_i$, *tconv*$^{\min}(v_1, \ldots, v_m)$.

# 3.3 Solving the Weighted Tropical Fermat-Weber Problem

In this section, we use combinatorics and tropical geometry to solve the weighted Fermat-Weber problem for $\mathbb{R}^n/\mathbb{R}\mathbb{1}$ equipped with the tropical asymmetric distance. We begin by discussing the extrema of tropical polynomials.

## 3.3.1 Containment

**Theorem 3.3.1.** *Given data points* $v_1, \ldots, v_m \in \mathbb{R}^n/\mathbb{R}\mathbb{1}$, *and weights* $w_1, \ldots, w_m > 0$, *the weighted Fermat-Weber points under the tropical asymmetric metric are a cell of* $tconv(v_1, \ldots, v_m)$.

*Proof.* According to corollary 3.2.24, $\mathrm{tconv}(v_1, \ldots, v_m)$ is the bounded part of the tropical hypersurface $\mathrm{tropV}(f_{V,\mathbf{w}})$. The bounded cells of $\mathrm{tropV}(f_{V,\mathbf{w}})$ are exactly those cells dual to interior cells of the Newton polytope, so by proposition 3.2.5, it suffices to show that $\lambda\mathbb{1}$ is in the interior of $\mathrm{Newt}(f_{V,\mathbf{w}})$. The vertices of $\mathrm{Newt}(f_{V,\mathbf{w}})$ are $|\mathbf{w}|e_i$, and their average, $\frac{|\mathbf{w}|}{n}\mathbb{1}$, is in the interior of the Newton polytope. This proves the Fermat-Weber points are achieved on a bounded cell of $\mathrm{tropV}(f_{V,\mathbf{w}})$, and therefore form a cell of $\mathrm{tconv}(v_1, \ldots, v_m)$. $\qquad \square$

## 3.3.2 Any Cell Can be the Weighted Fermat-Weber Cell

The following result shows that we can pick weights $w_i$ so that the weighted barycenter lies in any interior cell of the subdivision of the Cayley polytope. This finishes the proof of the main theorem.
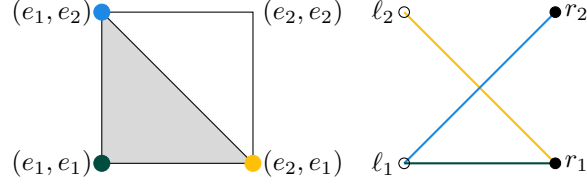
Figure 3.4: Vertices in the product of simplices (left) correspond to the color-coded edges of the bipartite graph (right).

**Theorem 3.3.2.** *Given some data points* $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n / \mathbb{R}\mathbb{1}$*, and any simplex $S$ in $\Delta^{r-1} \times \Delta^{n-1}$, which intersects the relative interior of $\Delta^{r-1} \times \Delta^{n-1}$, there is a choice of weights $w_1, \ldots, w_m \in [0, 1]$ with $\sum w_i = 1$, so that $S$ contains the point $(w_1, \ldots, w_m, \frac{1}{n}\mathbb{1}_n)$.*

The proof uses a well-known correspondence between subsets of the vertices of $\Delta^{m-1} \times \Delta^{n-1}$ and subgraphs of $K_{m,n}$ (the complete bipartite graph with $m$ left vertices, and $n$ right vertices), which we now briefly recall (see [15, §6.2.2] for more details). The vertex $(e_i, e_j)$ in a simplex $S \subseteq \Delta^{r-1} \times \Delta^{n-1}$ corresponds to the edge between left vertex $i$ and right vertex $j$ in the bipartite graph. Thus, a subset of vertices $A \subseteq \Delta^{r-1} \times \Delta^{n-1}$ corresponds to the subgraph of $K_{m,n}$.

**Example 3.3.3** (Simplex-Forest Correspondence for $m = n = 2$)**.** The product of two 2-simplices (i.e. line segments) is a square; the corresponding bipartite graph has two left vertices and two right vertices. Both are illustrated in fig. 3.4. The vertex $(e_i, e_j)$ in the simplex corresponds to the edge $(l_i, r_j)$ in the bipartite graph. For example, the top left vertex of the shaded gray simplex, $(e_1, e_2)$, corresponds to the edge $(l_1, r_2)$. The shaded simplex is full-dimensional, so it corresponds to a spanning tree of $K_{2,2}$ (see lemma 3.3.4).

**Lemma 3.3.4** (Lemma 6.2.8 in [15])**.** *Let $A$ be a subset of the vertices of $\Delta^{m-1} \times \Delta^{n-1}$. Then,*

*(a) $\mathrm{conv}(A)$ is a simplex if and only if the corresponding subgraph of $K_{m,n}$ is a forest.*

*(b) $\mathrm{conv}(A)$ is full dimensional if and only if the corresponding subgraph of $K_{m,n}$ is spanning and connected.*

*Proof of Theorem 3.3.2.* Let $S$ be a simplex in $\Delta^{r-1} \times \Delta^{n-1}$, and let $F$ be the corresponding forest in $K_{r,n}$. Assume that $S \cap \mathrm{int}\left(\Delta^{r-1} \times \Delta^{n-1}\right) \neq \emptyset$ (so $F$ is a spanning forest).

A point $(\mathbf{p}, \mathbf{q}) \in \mathbb{R}^m \times \mathbb{R}^n$ lies in $S$ if it can be written as a convex combination of the vertices of $S$. In terms of the forest $F$, $(\mathbf{p}, \mathbf{q})$ lies in $S$ if there exist $\lambda(e) > 0$ for each edge $e \in E(F)$ such that the sum of edge weights on any left vertex adds up to the corresponding

59

**p** coordinate, and the sum of edge weights on any right vertex adds up to the corresponding **q** coordinate. Let $r(e)$ be the node on the right side connected to $e$, and let $\ell(e)$ be the node on the left side connected to $e$. The choice of $\lambda$'s in (3.10) leads to a valid choice of weights $w_1, \ldots, w_m$ (given in (3.11)) so that $S$ contains the weighted barycenter.

$$\lambda(e) := \frac{1}{n \cdot \deg r(e)}. \tag{3.10}$$

$$w_i := \sum_{e \text{ s.t. } \ell(e)=i} \lambda(e). \tag{3.11}$$

The equations in (3.12) show that the weights on any right node sum to $\frac{1}{n}$ (since $F$ is spanning, every vertex has at least one edge); by definition, the weights on the $i$th left node sum to $w_i$. It follows that $\mathbf{b} = (w_1, \ldots, w_m, \frac{1}{n}\mathbb{1})$ lies in the relative interior of $S$.

$$\sum_{r(e)=j} \frac{1}{n \cdot \deg(j)} = \frac{1}{n} \sum_{r(e)=j} \frac{1}{\deg(j)} = \frac{1}{n} \deg(j) \frac{1}{\deg(j)} = \frac{1}{n}. \tag{3.12}$$

Moreover, $w_1, \ldots, w_m$ is a valid choice of weights for the Fermat-Weber problem. The weights $w_i$ are positive because $F$ is spanning, so the sum in (3.11) is never empty; The equations in (3.13) show that the $w_i$ sum to one.

$$\sum_{i=1}^{m} w_i = \sum_{e} \lambda(e) = \sum_{j=1}^{n} \sum_{r(e)=j} \frac{1}{n \cdot \deg(j)} = \sum_{j=1}^{n} \frac{1}{n} \sum_{r(e)=j} \frac{1}{\deg(j)} = \sum_{j=1}^{n} \frac{1}{n} = n\frac{1}{n} = 1. \tag{3.13}$$
$$\square$$

**Corollary 3.3.5.** *Given a cell $T$ in the tropical polytope* $\mathrm{tconv}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$, *there is a choice of weights* $w_1, \ldots, w_m$ *so that $T$ is the set of weighted tropical Fermat-Weber points for* $\mathbf{v}_1, \ldots, \mathbf{v}_m$ *with weights* $w_1, \ldots, w_m$.

# Part II

# Brownian Motion Tree Models

## CHAPTER 4

## ML Degree of Brownian Motion Tree Models

This chapter is joint work with Jane Ivy Coons, Aida Maraj and Ikenna Nometa. It appears in the preprint "Maximum Likelihood of Brownian Motion Tree Models" [8].

## 4.1   Introduction

A Brownian motion tree (BMT) model is a family of multivariate Gaussian distributions that describe the evolution of a continuous trait in a set of $n$ species along a phylogenetic tree. A phylogenetic tree is a rooted tree with no degree two vertices. The root is a distinguished leaf of the tree and each edge is implicitly directed away from it. Non-root leaves correspond to the extant species of interest, and the other nodes in the tree are the common ancestors of these species. We investigate the algebraic complexity of the maximum likelihood estimation problem for these models.

Introduced by Felsenstein [17] in 1973, BMT models enjoy many applications in phylogenetics. They have been applied to test for selective pressure [10, 18], often by serving as null model for evolution under genetic drift [40]. BMT models are commonly used to represent continuous molecular traits, such as gene expression profiles [5], and have even found use outside biology, such as in internet network tomography [16, 49]. Recent work in algebraic statistics [4, 45, 46, 48] uses algebraic geometry to study parameter inference problems for BMT models.

The BMT model on a tree has a simple description in terms of the covariance matrices of the densities in the model, stated in Definition 4.1.1 and illustrated in Figure 4.1. For simplicity, label the root of the phylogenetic tree $T$ by 0 and the rest of the leaves by $1, \ldots, n$.

Denote by lca$(i, j)$ the *least common ancestor* of non-root leaves $i$ and $j$; that is, lca$(i, j)$ is the first common node on the paths joining $i$ and $j$ to the root 0. Denote by Int$(T)$ the set of all internal nodes of $T$. Whenever $i \neq j$, lca$(i, j)$ belongs to Int$(T)$. Denote by $\mathbb{S}^n(\mathbb{R})$ the set of $n \times n$ symmetric matrices with real entries and let PD$^n(\mathbb{R})$ denote the cone of positive definite matrices within $\mathbb{S}^n(\mathbb{R})$. We can now define the BMT model.

**Definition 4.1.1.** Let $T$ be a phylogenetic tree on $n$ non-root leaves. Consider the linear space of symmetric matrices

$$\mathcal{L}_T(\mathbb{R}) := \{\Sigma \in \mathbb{S}^n(\mathbb{R}) \mid \sigma_{ij} = \sigma_{kl} \text{ if } \text{lca}(i, j) = \text{lca}(k, l)\}. \tag{4.1}$$

The *Brownian motion tree model*, $\mathcal{M}_T$, specified by $T$ is the set of all multivariate Gaussian distributions with mean $\mathbf{0}$ and whose covariance matrices lie in the set $\mathcal{L}_T(\mathbb{R}) \cap \text{PD}^n(\mathbb{R})$.



$$\mathcal{L}_T(\mathbb{R}) = \left\{ \Sigma = \begin{pmatrix} t_1 & t_5 & t_5 & t_6 \\ t_5 & t_2 & t_5 & t_6 \\ t_5 & t_5 & t_3 & t_6 \\ t_6 & t_6 & t_6 & t_4 \end{pmatrix} : t_1, \ldots, t_6 \in \mathbb{R} \right\}.$$
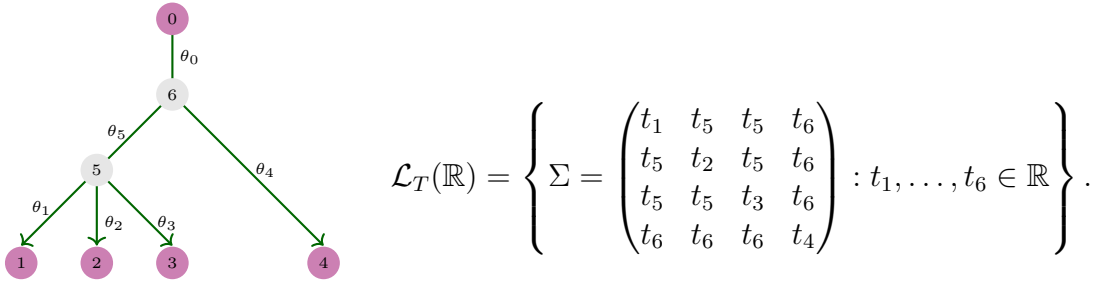
Figure 4.1: An evolutionary tree $T$ on species $1, 2, 3, 4$. A covariance matrix in the associated BMT model must be in the linear space $\mathcal{L}_T(\mathbb{R})$.

Finding the probability distribution in a fixed model that best fits observed data is a standard problem in statistics and data science. One popular method for inferring such a distribution is maximum likelihood estimation. The *maximum likelihood estimate* (MLE) for given data is the maximizer of the log-likelihood function (see Section 4.3.1) over the model. In this paper, we investigate the number of complex critical points of the log-likelihood function over a BMT model. This number, known to be invariant under a generic choice of data, is referred to as the maximum likelihood degree (ML degree) of a model. Since the MLE, if it exists, is one of these critical points, the ML degree of a model measures the algebraic complexity of this problem.

As BMT models are not exponential families and their likelihood functions are not typically convex, finding general formulae for their ML degree is challenging. Sturmfels, Timme, and Zwiernik in [45] use numerical algebraic geometry to compute ML degrees of BMT models for phylogenetic trees with up to 6 leaves. Améndola and Zwiernik conjectured a formula

for the ML degree of a star tree model[1]; that is, of a BMT model whose tree has exactly one internal node. The main result of this paper positively answers that conjecture, as stated below.

**Theorem 4.1.2.** *The ML degree of the BMT model on a star tree on $n + 1$ leaves is $2^{n+1} - 2(n + 1) - 1$.*

While computing the ML degree for trees with multiple internal nodes remains challenging, we are able to consolidate the problem to classes of trees with the same unlabeled, unrooted tree topology.

**Theorem 4.1.3.** *BMT models on phylogenetic trees with the same unlabeled, unrooted tree topology have the same ML degree.*

The proofs of these theorems rely on the toric geometry of the inverse linear space of $\mathcal{L}_T$ under the change of variables given by Sturmfels, Uhler, and Zwiernik [46] (see Theorem 4.2.2), and monomial parametrization provided previously by Boege et al. [4] (see Theorem 4.2.3). This parametrization, called the *path parametrization*, assigns to each edge $e$ in the tree a parameter $\theta_e$ and allows us to write each concentration matrix $K_T$ in $\mathcal{L}_T^{-1}$ as a function of the parameters $\theta_e$, which we call $K_T(\theta)$. Towards this end, we compute the degree of this parametrization in Theorem 4.2.5, which allows us compute the ML degree by counting the solutions in the new parameters. Though we focus on ML degrees for star trees in the present work, we envision that Theorem 4.2.5 will be useful for future work towards computing ML degrees for arbitrary trees.

In order to write the log-likelihood function in terms of the parameters $\theta_e$, we require an expression for the determinant of $K_T(\theta)$. Our result is a weighted analog of the Cayley-Prüfer formula, which is a factorization of the sum of the products of edge variables over spanning trees of an unweighted complete graph. This yields an explicit formula for $\det(K_T(\theta))$ for any tree $T$ (see Theorem 4.3.5), which is applicable to future work on the likelihood geometry of BMT models.

### 4.1.1 Structure of the Paper

In Section 4.2, we reframe Brownian motion tree models in terms of their concentration matrices. We recall the toric representation of this space of concentration matrices by the monomial path map and compute the degree of this map in Theorem 4.2.5. In Section 4.3, we introduce the maximum likelihood estimation problem for BMT models and define their

---

[1]shared at the program "Linear Spaces of Symmetric Matrices" held in Fall 2020 at the Max Planck Institute for Mathematics in the Sciences

ML degree via the score equations. In order to better understand these score equations, we prove Theorem 4.3.5, which writes the determinant of an arbitrary concentration matrix in the BMT model in terms of the parameters of the path map. Section 4.4 is devoted to the proof of Theorem 4.1.3, which states that the ML degree is invariant under rerooting. In Section 4.5, we prove the ML degree formula for the BMT model on a star tree. We end the paper with a discussion of our results and directions for future work.

## 4.2 Toric Geometry of Brownian Motion Tree Models

### 4.2.1 Monomial Parametrizations of Concentration Matrices

Recall that the covariance matrices in the Brownian motion tree model specified by a tree $T$ are exactly the positive definite matrices in $\mathcal{L}_T$, as described in Definition 4.1.1. The *concentration* (or precision) matrices for the BMT model are therefore $\mathcal{L}_T^{-1} \cap \mathrm{PD}^n(\mathbb{R})$, where $\mathcal{L}_T^{-1}$ is the Zariski closure of all matrices $K \in \mathbb{S}^n$ such that $K^{-1} \in \mathcal{L}_T$. The algebraic variety $\mathcal{L}_T^{-1}$ is the vanishing locus of the kernel of the rational map

$$\rho_T : \mathbb{R}[K] \to \mathbb{R}(\Sigma), \ k_{ij} \mapsto \frac{(-1)^{i+j} \det(\Sigma_{ij})}{\det(\Sigma)}, \tag{4.2}$$

where $\Sigma_{ij}$ is the submatrix of the symmetric matrix $\Sigma$ with its $i$th row and $j$th column deleted. The ideal $\ker(\rho_T)$ is referred to as the *vanishing ideal of the Brownian motion tree model* $\mathcal{M}_T$.

**Example 4.2.1.** Consider the phylogenetic tree $T$ in Figure 4.1. It is pictured along with a generic element of its associated linear space $\mathcal{L}_T$. The inverse linear space $\mathcal{L}_T^{-1}$ has vanishing ideal

$$\begin{aligned}
\ker(\rho_T) = \langle &k_{12}k_{14} + k_{14}k_{22} + k_{14}k_{23} - k_{11}k_{24} - k_{12}k_{24} - k_{13}k_{24}, \\
&k_{13}k_{14} + k_{14}k_{23} + k_{14}k_{33} - k_{11}k_{34} - k_{12}k_{34} - k_{13}k_{34}, \\
&k_{13}k_{24} + k_{23}k_{24} + k_{24}k_{33} - k_{12}k_{34} - k_{22}k_{34} - k_{23}k_{34}, \\
&k_{12}k_{23} + k_{12}k_{33} + k_{12}k_{34} - k_{13}k_{22} - k_{13}k_{23} - k_{13}k_{24}, \\
&k_{12}k_{13} + k_{12}k_{33} + k_{12}k_{34} - k_{11}k_{23} - k_{13}k_{23} - k_{14}k_{23}, \\
&k_{12}k_{34} - k_{13}k_{24}, \ k_{12}k_{34} - k_{14}k_{23} \rangle.
\end{aligned}$$

The variety $\mathcal{L}_T^{-1}$ is described in detail by Sturmfels, Uhler, and Zwiernik in [46] as a toric variety. To exhibit the toric structure of the space of concentration matrices, they show

that $\ker(\rho_T)$ is generated by binomials after a change of coordinates known as the *Farris transform*.

**Theorem 4.2.2** ([46], Theorem 1)**.** *The vanishing ideal for the Brownian motion tree model of the phylogenetic tree $T$ is toric in variables $p_{ij}$ with $i, j$ distinct elements of $\{0, \ldots, n\}$, where*

$$\begin{aligned} p_{ij} &= -k_{ij} && \text{for } i, j > 0, \text{ and} \\ p_{0i} &= \textstyle\sum_{j=1}^{n} k_{ij} && \text{for } 1 \le i \le n. \end{aligned} \tag{4.3}$$

*It is generated by the forms $p_{ik}p_{j\ell} - p_{i\ell}p_{jk}$, where $\{i, j\}$ and $\{k, \ell\}$ are cherries in the induced 4-leaf subtree on any quadruple $i, j, k, \ell \in \{0, \ldots, n\}$.*

We note that in Theorem 4.2.2, the subscripts on the variables $p_{ij}$ are unordered so that $p_{ij} = p_{ji}$. We therefore think of the concentration matrices in the BMT model in the coordinates $p_{ij}$ for $0 \le i < j \le n$ for the rest of the paper.

Denote by $I_T$ the ideal $\ker(\rho_T)$ in the coordinates $p_{ij}$. Since the ideal $I_T$ is toric, it is the kernel of a monomial map. Boege et al. [4] connect this ideal to the paths in the tree and use these paths to give the monomial parametrization $\varphi_T$ of this toric variety, called the *path parametrization*:

$$\varphi_T : \mathbb{R}[p_{ij} \mid 0 \le i < j \le n] \to \mathbb{R}[\theta_e \mid e \in E(T)], \; p_{ij} \mapsto \prod_{e \in i \leadsto j} \theta_e, \tag{4.4}$$

where $E(T)$ is the set of edges of $T$ and $i \leadsto j$ is the set of edges in the path between leaf $i$ and leaf $j$.

**Theorem 4.2.3** ([4], Proposition 3.1)**.** *The toric vanishing ideal of the Brownian motion tree model on the phylogenetic tree $T$ in the $p_{ij}$ coordinates defined in Theorem 4.2.2 is the kernel of the path map $\varphi_T$.*

**Example 4.2.4.** The Brownian motion tree model for $T$ in Figure 4.1 has toric vanishing ideal

$$\begin{aligned} I_T = \big\langle\, & p_{01}p_{24} - p_{02}p_{14}, \; p_{01}p_{34} - p_{03}p_{14}, \; p_{02}p_{34} - p_{03}p_{24}, \; p_{02}p_{13} - p_{03}p_{12}, \\ & p_{01}p_{23} - p_{03}p_{12}, \; p_{12}p_{34} - p_{13}p_{24}, \; p_{12}p_{34} - p_{23}p_{14} \,\big\rangle. \end{aligned}$$

Note that the seven generators in the $k_{ij}$ coordinates given in Example 4.2.1 can be obtained by applying the change of coordinates in Theorem 4.2.2 to each of these binomials. This

ideal is the kernel of the path map $\varphi_T$ whose exponent matrix is

$$
A_T = \begin{array}{c} \\ \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \end{array} \begin{array}{c} p_{01} \; p_{02} \; p_{03} \; p_{04} \; p_{12} \; p_{13} \; p_{14} \; p_{23} \; p_{24} \; p_{34} \\ \left(\begin{array}{cccccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right) \end{array}.
$$

The path parametrization proved to be essential in the computation of reciprocal (dual) maximum likelihood degree of BMT models [4]. It will continue to be instrumental in all of the proofs of the present work.

## 4.2.2 Degree of the Path Parametrization

Let $T$ be a tree on $n+1$ leaves with edge set $E(T)$. Let $\mathrm{Lv}(T) := \{0, \ldots, n\}$ denote the leaf set of $T$. Consider the pullback of the path parametrization:

$$
\varphi_T^* : \mathbb{C}^{\#E(T)} \to \mathbb{C}^{\binom{n+1}{2}}, \; (\theta_e)_{e \in E(T)} \mapsto \left( \prod_{e \in i \rightsquigarrow j} \theta_e \right)_{i \neq j \in \mathrm{Lv}(T)} .
$$

In the following proposition, we compute the degree of $\varphi_T^*$ by explicitly describing elements in its fibers. For each $S \subseteq \mathrm{Int}(T)$, let

$$
\epsilon^S \in \{-1, 1\}^{\#E(T)} \text{ such that } \epsilon_e^S = (-1)^{\#(S \cap e)} \text{ for each edge } e \in E(T).
$$

For any two vectors $u, v$ of the same length $k$, denote by $u * v$ their componentwise product; that is, $u * v = (u_i v_i)_{i=1}^k$.

**Theorem 4.2.5.** *Let $T$ be a rooted tree with no degree two nodes. Let $p \in \mathrm{im}(\varphi_T^*)$, with all coordinates non-zero. Let $\hat{\theta} \in \mathbb{C}^{\#E}$ such that $\varphi_T^*(\hat{\theta}) = p$. Then $(\varphi_T^*)^{-1}(p) = \{\epsilon^S * \hat{\theta} \mid S \subseteq \mathrm{Int}(T)\}$. In particular, the degree of $\varphi_T^*$ is $2^{\#Int(T)}$.*

*Proof.* For ease of readability, we split the proof into two parts. In the first part, we show that the coordinates of any point $\theta$ in the fiber $(\varphi_T^*)^{-1}(p)$ agree with the coordinates of $\hat{\theta}$ up to a sign. In the second part, we show that $\theta$ must be exactly of form $\epsilon^S * \hat{\theta}$ for some $S \subseteq \mathrm{Int}(T)$.

**Part 1.** Consider an edge $\{u, v\} \in E(T)$, and the associated coordinate $\theta_{uv}$ of $\theta \in (\varphi_T^*)^{-1}(p)$. At least one of the nodes, say $v$, is an internal node. Let $i$ be a leaf such that the path $v \leftrightsquigarrow i$ contains edge $\{u, v\}$. Let $j, k$ be two other leaves with property that $\{u, v\}$ is not in the paths $v \leftrightsquigarrow j$ or $v \leftrightsquigarrow k$. The existence of distinct leaves $i, j, k$ is guaranteed since $v$ has degree at least three. The condition $\theta \in (\varphi_T^*)^{-1}(p)$ implies that

$$p_{ij} = \prod_{e \in i \leftrightsquigarrow v} \theta_e \prod_{e \in v \leftrightsquigarrow j} \theta_e = \prod_{e \in i \leftrightsquigarrow v} \hat{\theta}_e \prod_{e \in v \leftrightsquigarrow j} \hat{\theta}_e \tag{4.5}$$

$$p_{ik} = \prod_{e \in i \leftrightsquigarrow v} \theta_e \prod_{e \in v \leftrightsquigarrow k} \theta_e = \prod_{e \in i \leftrightsquigarrow v} \hat{\theta}_e \prod_{e \in v \leftrightsquigarrow k} \hat{\theta}_e \tag{4.6}$$

$$p_{jk} = \prod_{e \in j \leftrightsquigarrow v} \theta_e \prod_{e \in v \leftrightsquigarrow k} \theta_e = \prod_{e \in j \leftrightsquigarrow v} \hat{\theta}_e \prod_{e \in v \leftrightsquigarrow k} \hat{\theta}_e. \tag{4.7}$$

Since each $p_{ij} \neq 0$, each of the products in the above equations is nonzero. Solving for $\prod_{e \in v \leftrightsquigarrow j} \theta_e$ in (4.5), for $\prod_{e \in v \leftrightsquigarrow k} \theta_e$ in (4.6), and substituting these expressions in (4.7), gives

$$\left( \prod_{e \in i \leftrightsquigarrow v} \theta_e \right)^2 = \left( \prod_{e \in i \leftrightsquigarrow v} \hat{\theta}_e \right)^2. \tag{4.8}$$

Therefore, the product of $\theta_e$'s over the path $v \leftrightsquigarrow i$ from an internal vertex $v$ to any leaf $i$ is equal to the product of $\hat{\theta}_e$'s over $v \leftrightsquigarrow i$ up to a sign. Factoring out the terms corresponding to the edge $\{u, v\}$ gives

$$\theta_{uv} \prod_{e \in i \leftrightsquigarrow u} \theta_e = \pm \hat{\theta}_{uv} \prod_{e \in i \leftrightsquigarrow u} \hat{\theta}_e. \tag{4.9}$$

If $u = i$ is a leaf, then we immediately obtain $\theta_{uv} = \epsilon_{uv} \hat{\theta}_{uv}$ for some $\epsilon_{uv} \in \{-1, 1\}$ and we are done. If $u$ is an internal node, then analogously to Equation (4.8) we have

$$\prod_{e \in i \leftrightsquigarrow u} \theta_e = \pm \prod_{e \in i \leftrightsquigarrow u} \hat{\theta}_e, \tag{4.10}$$

and Equation (4.9) and Equation (4.10) imply $\theta_{uv} = \epsilon_{uv} \hat{\theta}_{uv}$ for some $\epsilon_{uv} \in \{-1, 1\}$, as desired.

**Part 2.** We will use induction on the number of internal nodes to show that

$$(\varphi_T^*)^{-1}(p) = \{\epsilon^S * \hat{\theta} \mid S \subseteq \text{Int}(T)\}. \tag{4.11}$$

The following observation will be useful:

"if $\epsilon * \hat{\theta} \in (\varphi_T^*)^{-1}(p)$, then $\epsilon_{iu}$ is fixed for all leaves $i$ with same parent $u$".  (4.12)

Indeed, let $i, j$ be leaves with the same parent $u$. The path between $i$ and $j$ in $T$ consists of the edges $iu$ and $ju$. Since $\epsilon * \hat{\theta}$ and $\hat{\theta}$ belong to the same fiber of $\varphi_T^*$, by definition of the path map we have $(\epsilon_{iu}\hat{\theta}_{iu})(\epsilon_{ju}\hat{\theta}_{ju}) = \hat{\theta}_{iu}\hat{\theta}_{ju}$. Hence, $\epsilon_{iu} = \epsilon_{ju}$.

Now we prove Equation (4.11) for a tree $T$ whose internal nodes have a degree of at least three by inducting on the number of internal nodes. First, if $T$ has no internal nodes, then $(\varphi_T^*)^{-1}(p) = \{\epsilon^{\emptyset} * \hat{\theta}\}$. Let $T$ have one internal node, $u$. Let $\epsilon * \hat{\theta} \in (\varphi_T^*)^{-1}(p)$ be a point in the fiber. Suppose $\epsilon_{iu} = -1$ for some leaf $i$. Since all leaves have the same parent $u$, by Observation 4.12, $\epsilon_{ju} = -1$ for any other leaf $j$. So, $\epsilon = \epsilon^{\{u\}}$ and $(\varphi_T^*)^{-1}(p) = \{\epsilon^{\emptyset} * \hat{\theta}, \epsilon^{\{u\}} * \hat{\theta}\}$.

Now suppose that Equation (4.11) holds for any tree whose internal nodes have degree at least three and with less than $m > 1$ internal nodes, and let $T$ have $m$ internal nodes. Select an internal node, $u$, in $T$ whose children are all leaves which we label $1, 2, \ldots, k$ without loss of generality. Note that since $T$ is not a star tree, any such $u$ must have a parent $v$ that is an internal node.

Let $T'$ be the tree obtained by removing from $T$ the leaves $1, \ldots, k$ along with the edges $\{u, i\}$ and the edge $\{u, v\}$. The new tree has $m - 1$ internal nodes and $\mathrm{Lv}(T') \subset \mathrm{Lv}(T)$. Take $p'$ to be the projection of $p$ in $\mathbb{C}^{\binom{n+1-k}{2}}$. The projection $(\epsilon * \hat{\theta})|_{E(T')} \in \mathbb{C}^{\#E(T')}$ of any $\epsilon * \hat{\theta} \in (\varphi_T^*)^{-1}(p)$ is in $(\varphi_{T'}^*)^{-1}(p')$. Hence the projection of the fiber of $p$ must lie in the fiber of $p'$. Denote by $\hat{\theta}'$ the projection of $\hat{\theta}$ onto the coordinates corresponding to edges of $T'$.

We consider two cases dependent on if the degree of $v$ in $T'$ is 2 or greater.

*Case 1.* Suppose $v$ is of degree greater than two in $T'$. Then Equation (4.11) holds for $T'$, points $p'$ and $(\hat{\theta}_e)_{e \in E(T')}$. From Observation 4.12, any point $\epsilon * \hat{\theta}$ in the fiber of $p$ has fixed $\epsilon_{iu} = \epsilon_u$ for $i = 1, \ldots, k$. Now, any choice of $\epsilon_u$ and any point $\epsilon^{S'} * \hat{\theta}' \in (\varphi_{T'}^*)^{-1}(p')$ for $S' \subseteq \mathrm{Int}(T) \setminus \{u\}$ extend to the unique point $\theta$ in $\mathbb{C}^{\#E(T)}$:

$$
\theta_e = \begin{cases}
\epsilon_e^{S'} \cdot \hat{\theta}_e & \text{for } e \in E(T') \\
\epsilon_u \cdot \hat{\theta}_e & \text{for } e = \{i, u\}, \ i = 1, \ldots, k \\
\left( \epsilon_u \prod_{e \in 0 \rightsquigarrow v} \epsilon_e^{S'} \right) \hat{\theta}_e & \text{for } e = \{u, v\}.
\end{cases}
$$ (4.13)

Note that by construction, $\theta = \epsilon^S * \hat{\theta}$ where $S = S'$ when $\epsilon_u = 1$ and $S = S' \cup \{v\}$ when $\epsilon_u = -1$.

*Case 2.* Suppose $v$ is of degree two. Let $v_1, v_2$ be the other nodes adjacent to $v$ in $T'$. Consider the tree $T''$ obtained by merging edges $\{v_1, v\}$ and $\{v, v_2\}$ into edge $\{v_1, v_2\}$.

69

The new tree $T''$ has $m - 2$ internal nodes, all of degree at least three. The trees $T'$ and $T''$ have the same set of leaves. The point $p' = \varphi_{T''}^*(\hat{\theta}'')$, where $\hat{\theta}'' \in \mathbb{C}^{\#E(T'')}$ has entries $\hat{\theta}''_{v_1 v_2} = \hat{\theta}_{vv_1}\hat{\theta}_{vv_2}$, and otherwise $\hat{\theta}''_e = \hat{\theta}_e$. By the induction hypothesis, $(\varphi_{T''}^*)^{-1}(p') = \{\epsilon'' * \hat{\theta}'' \mid S'' \subset \mathrm{Int}(T) \setminus \{u, v\}\}$.

Since any point $\epsilon^{S''} * \hat{\theta}''$ in this fiber has

$$(-1)^{S'' \cap \{v_1, v_2\}}\theta''_{v_1 v_2} = \epsilon^{S''}\hat{\theta}_{vv_1}\hat{\theta}_{vv_2}$$

we construct two points in the fiber $(\varphi_{T'}^*)^{-1}(p')$. These are precisely, $\epsilon^{S''} * \hat{\theta}'$ and $\epsilon^{S'' \cup \{v\}} * \hat{\theta}'$, which differ only by the signs of the $\{v, v_1\}$ and $\{v, v_2\}$ coordinates. Since all points of the form $\epsilon * \hat{\theta}'$ in $(\varphi_{T'}^*)^{-1}(p')$ are of one of the two forms above, we conclude that solutions in $T'$ of the form $\epsilon * \hat{\theta}'$ arise from $\epsilon^{S'}$ for $S' \subset \mathrm{Int}(T) \setminus \{u\}$. Proceed as in Case 1 to conclude the fiber of $(\varphi_T^*)^{-1}(p)$. $\qquad\square$

Theorem 4.2.5 deduced the following results, ready to be used when solving polynomials systems on a Brownian motion tree model.

**Corollary 4.2.6.** *Let $F = \{f_1(K), \ldots, f_\ell(K)\}$ be a polynomial system in the variables $p_{ij}$. Let $F \circ \varphi_T = \{f_1(K(\theta)), \ldots, f_\ell(K(\theta))\}$ be these polynomials written in variables $\theta_e$. Then,*

*a.   the solutions $K \in \mathcal{L}_T^{-1}$ to system $F$ are precisely $K(\theta)$ for $\theta$ a solution of $F \circ \varphi_T$,*

*b.   $\deg\left(\langle F \rangle + I_T\right) = \dfrac{\deg\left(\langle F \circ \varphi_T \rangle\right)}{2^{\#\mathrm{Int}(T)}}.$*

## 4.3   Score Equations

### 4.3.1   Maximum Likelihood Estimation in Brownian Motion Tree Models

Maximum likelihood estimation is a method for inferring the distribution in a statistical model that best explains a data set. Let $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{R}^n$ be independent, identically distributed data which we assume are sampled from a distribution in the BMT model on a tree $T$. The observed data has sample covariance matrix

$$S := \frac{1}{m}\sum_{j=1}^m \mathbf{u}_i\mathbf{u}_i^T \in \mathrm{PSD}^n. \tag{4.14}$$

A *maximum likelihood estimate* (MLE) for this data in the BMT model $\mathcal{M}_T$ is a concentration matrix $\hat{K} \in \mathcal{L}_T^{-1} \cap \mathrm{PD}^n(\mathbb{R})$ that maximizes the value of the density function for the

normal distribution $\mathcal{N}(\mathbf{0}, K^{-1})$ on this data, if such a maximizer exists. Equivalently, $\hat{K}$ is a global maximizer of

$$\ell(K|S) := \log \det(K) - \text{trace}(SK). \tag{4.15}$$

We note that, as written, the expression $\ell(K|S)$ is not exactly the logarithm of the likelihood function. However, they only differ by constant addition and multiplication and hence have the same critical points. So we slightly abuse terminology and refer to $\ell(K|S)$ as the *log-likelihood function*. We refer the reader to [47, Chapter 7] for background on algebraic geometry and maximum likelihood estimation. Section 2 of [9] also thoroughly introduces maximum likelihood estimation, specifically in linear covariance models.

The MLE is a critical point of the log-likelihood function. Hence, we may compute it by finding the common zeros of the partial derivatives of $\ell(K|S)$ and computing the likelihood at each critical point. Thus, the number of critical points of $\ell(K|S)$, called the *maximum likelihood degree* (ML degree), measures the algebraic complexity of computing the MLE. We now define the ML degree more precisely.

**Definition 4.3.1.** The *maximum likelihood degree* of the BMT model, which is denoted $\text{mld}(\mathcal{M}_T)$, is the number of complex critical points $\ell(K|S)$ over $\mathcal{M}_T$, counted with multiplicity, for a generic sample covariance matrix $S$.

In order to compute the ML degree of a Gaussian model, we begin by writing the log-likelihood $\ell(K|S)$ in terms of the parameters $\theta_e$ of the path map. Since we are interested in the critical points of the log-likelihood, we take the partial derivatives of $\ell(K|S)$ with respect to each $\theta_e$ and set these equal to zero. These partial derivatives are called the *score equations*. In the case of a linear Gaussian covariance model, they are rational functions. In fact, they are of the form

$$\frac{\partial \ell}{\partial \theta} = \frac{1}{\det(K)} \frac{\partial}{\partial \theta}(\det(K)) - \frac{\partial}{\partial \theta}(\text{trace}(SK)).$$

We can compute the vanishing locus of the rational score equations by finding the variety of their numerators and removing the variety of the product of their denominators. In the Gaussian case, note that since $\det(K)$ and $\text{trace}(SK)$ are both polynomials in the $\theta_e$ parameters, the only denominator that appears in any score equation is $\det(K)$. Hence, removing the vanishing locus of the denominators simply corresponds to removing any solutions for which the resulting concentration matrix would be singular. Next, we count the critical points in this variety in the $\theta_e$ coordinates, including their multiplicities. Finally, Theorem 4.2.5 allows us to divide the number of solutions in the $\theta_e$ parameters by $2^{\#\text{Int}(T)}$ to obtain

the desired ML degree.

**Example 4.3.2.** Consider the tree $T$ in Figure 4.1. Via the Farris transform and the path map, the concentration matrices in the BMT model on $T$ are of the following form in the $\theta_e$ parameters:

$$\begin{pmatrix} \theta_1\left(\theta_5\left(\theta_0+\theta_4\right)+\theta_2+\theta_3\right) & -\theta_1\theta_2 & -\theta_1\theta_3 & -\theta_5\theta_1\theta_4 \\ -\theta_1\theta_2 & \theta_2\left(\theta_5\left(\theta_0+\theta_4\right)+\theta_1+\theta_3\right) & -\theta_2\theta_3 & -\theta_5\theta_2\theta_4 \\ -\theta_1\theta_3 & -\theta_2\theta_3 & \theta_3\left(\theta_5\left(\theta_0+\theta_4\right)+\theta_1+\theta_2\right) & -\theta_5\theta_3\theta_4 \\ -\theta_5\theta_1\theta_4 & -\theta_5\theta_2\theta_4 & -\theta_5\theta_3\theta_4 & \theta_4\left(\theta_0+\theta_5\left(\theta_1+\theta_2+\theta_3\right)\right) \end{pmatrix}.$$

Consider a generic sample covariance matrix $S = (s_{ij})_{1\leq i,j\leq 4}$. The log-likelihood function for $S$ in this BMT model is

$$\begin{aligned} \ell(K|S) &= \log\det(K) - \operatorname{trace}(SK) \\ &= \log(\theta_0\theta_1\theta_2\theta_3\theta_4\theta_5(\theta_1\theta_5+\theta_2\theta_5+\theta_0+\theta_4)(\theta_0\theta_5+\theta_4\theta_5+\theta_1+\theta_2+\theta_3)^2) \\ &\quad - s_{11}\theta_1\left(\theta_5\left(\theta_0+\theta_4\right)+\theta_2+\theta_3\right) - s_{22}\theta_2\left(\theta_5\left(\theta_0+\theta_4\right)+\theta_1+\theta_3\right) \\ &\quad - s_{33}\theta_3\left(\theta_5\left(\theta_0+\theta_4\right)+\theta_1+\theta_2\right) - s_{44}\theta_4\left(\theta_0+\theta_5\left(\theta_1+\theta_2+\theta_3\right)\right) \\ &\quad + 2s_{12}\theta_1\theta_2 + 2s_{13}\theta_1\theta_3 + 2s_{14}\theta_1\theta_4\theta_5 + 2s_{23}\theta_2\theta_3 + 2s_{24}\theta_2\theta_4\theta_5 + 2s_{34}\theta_3\theta_4\theta_5. \end{aligned}$$

To find the ML degree, we need to set the system of score equations equal to zero and solve. Using `HomotopyContinuation.jl` [6] in `Julia` with generic values for the entries of $S$, we see that the system has 44 solutions in the torus. By Theorem 4.2.5, the degree of the path map in this case is 4. Hence, dividing by 4 gives that $\operatorname{mld}(\mathcal{M}_T) = 11$.

## 4.3.2  A Generalization of the Cayley-Prüfer Theorem

The classical Cayley-Prüfer Theorem provides an enumeration of the spanning trees of a complete graph in factored form.

**Theorem 4.3.3** (Classical Cayley-Prüfer Theorem)**.** *Let $\mathcal{K}_n$ be the complete graph on $n$ vertices. Then,*

$$\sum_{\substack{\Gamma\subseteq\mathcal{K}_n \\ spanning \\ tree}} \prod_{v\in V(\mathcal{K}_n)} x_v^{\deg_\Gamma(v)} = x_1\cdots x_n\left(x_1+\cdots+x_n\right)^{n-2}, \tag{4.16}$$

*where $V(\mathcal{K}_n)$ is the vertex set of $\mathcal{K}_n$ and $\deg_\Gamma(v)$ is the number of edges adjacent to $v$ in tree $\Gamma$.*

The goal of this section is to prove Theorem 4.3.5, which factorizes $\det(K_T(\theta))$ and specializes to the classical Cayley-Prüfer Theorem when $T$ is a star tree. We begin by recalling Kirchoff's Matrix-Tree Theorem. Let $\mathcal{G}$ be a weighted graph with vertex set $[n] := \{1,\ldots,n\}$.

Let $w_{ij}$ be the weight for the edge $\{i, j\} \in E(\mathcal{G})$. One naturally extends $w$ to all pairs of vertices in $\mathcal{G}$ by setting $w_{ij} = 0$ when $\{i, j\}$ is not an edge of $\mathcal{G}$. The *Laplacian of $\mathcal{G}$*, denoted $L_G$, is an $n \times n$ matrix which encodes the weights of $\mathcal{G}$ as follows:

$$
(L_{\mathcal{G}})_{ij} = \begin{cases} \sum\limits_{\substack{k=1 \\ k \neq i}}^{n} w_{ik} & \text{if } i = j, \\[2em] -w_{ij} & \text{if } i \neq j. \end{cases}
$$

Kirchoff's Matrix-Tree Theorem [26], applied to a complete graph, states that the left hand-side sum in Equation (4.16) is the determinant of any principal submatrix of the Laplacian of $\mathcal{K}_n$. Our concentration matrix, $K_T(\theta)$, is a principal submatrix of the Laplacian of a weighted complete graph, $\mathcal{K}^T$ with weights determined by the paths in tree $T$.

**Definition 4.3.4.** Let $T$ be a phylogenetic tree on $n + 1$ leaves. Define $\mathcal{K}^T$ to be the weighted complete graph on $n + 1$ vertices, where the weight of an edge $\{i, j\}$ is equal to $\varphi_T(p_{ij}) = \prod\limits_{e \in i \rightsquigarrow j} \theta_e$.
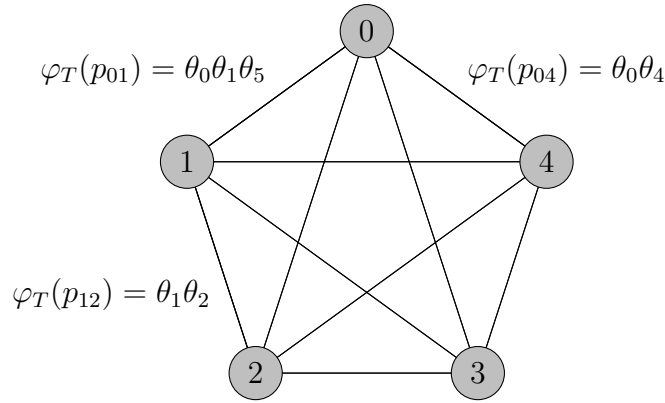


Figure 4.2: The weighted complete graph, $\mathcal{K}_5^T$, for the tree from Figure 4.1.

The Laplacian of $\mathcal{K}^T$ is the $n \times n$ matrix with entries

$$
(L_{\mathcal{K}^T})_{0 \leq i,j, \leq n} = \begin{cases} \sum\limits_{\substack{k=0 \\ k \neq i}}^{n} \prod\limits_{e \in i \rightsquigarrow k} \theta_e & \text{if } i = j, \\[2em] -\prod\limits_{e \in i \rightsquigarrow j} \theta_e & \text{if } i \neq j. \end{cases}
$$

Note that $K_T(\theta)$ is the submatrix of $L_{\mathcal{K}^T}$ with row and column corresponding to vertex 0 removed. By the Matrix-Tree Theorem, $\det(K_T(\theta))$ enumerates the weighted spanning trees of $\mathcal{K}^T$. The following theorem provides a factorization for this enumeration.

**Theorem 4.3.5.** *The determinant of a concentration matrix $K_T(\theta)$ in the BMT model on $T$ is*

$$\det(K_T(\theta)) = \left( \prod_{e \in E(T)} \theta_e \right) \prod_{v \in \mathrm{Int}(T)} \left( \sum_{\ell \in \mathrm{Lv}(T)} \prod_{e \in v \rightsquigarrow \ell} \theta_e \right)^{\deg_T(v)-2}. \tag{4.17}$$

*Proof.* Note that since $K_T$ is a principal submatrix of the Laplacian of a complete graph, its determinant is described by the matrix-tree theorem:

$$\det(K_T(\theta)) = \sum_{\substack{\Gamma \subset \mathcal{K}^T \\ \text{spanning} \\ \text{tree}}} \prod_{\{i,j\} \in E(\Gamma)} \varphi_T(p_{ij}).$$

Observe that $\theta_e$ divides $\det(K_T(\theta))$ for all edges $e \in E(T)$. Indeed, each spanning tree of $\mathcal{K}^T$ must contain at least one edge $\{i, j\}$ where the leaves $i$ and $j$ are in different connected components of $T \setminus \{e\}$, for every $e \in E(T)$. Thus, $\theta_e$ divides each term of the sum in $\det(K_T(\theta))$. Now define

$$D(T) := \prod_{v \in \mathrm{Int}(T)} \left( \sum_{\ell \in \mathrm{Lv}(T)} \prod_{e \in v \rightsquigarrow \ell} \theta_e \right)^{\deg_T(v)-2} \quad \text{and}$$

$$\Delta(T) := \frac{\det(K_T(\theta))}{\displaystyle\prod_{e \in E(T)} \theta_e} - D(T).$$

We want to show that $\Delta(T) = 0$ by induction on the number of leaves. Specifically, we show that $\Delta(T)$ is divisible by each of the $n + 1$ variables $\theta_e$ where $e$ contains a leaf; we call these variables *leaf variables* and denote by $\theta_i$ the leaf variable $i$. Since $\Delta(T)$ is homogeneous of degree $\sum_{v \in \mathrm{Int}(T)}(\deg_T(v) - 2) = n - 1$ in the leaf variables, this shows that $\Delta(T)$ is identically zero. If $T$ has two leaves, then it is easy to check that $\Delta(T) = 0$. Now let $T$ be a tree with $n + 1$ leaves (including the root leaf), and fix a leaf $i \in \mathrm{Lv}(T)$. We show that $\theta_i$ divides $\Delta(T)$, or equivalently, $\Delta(T)|_{\theta_i=0} = 0$. We claim that

$$\Delta(T)\Big|_{\theta_i=0} = \Delta(T \setminus \{i\}) \left( \sum_{j \in \mathrm{Lv}(T) \setminus \{i\}} \frac{\varphi_T(p_{ij})}{\theta_i} \right).$$

The tree $T \setminus \{i\}$ has fewer leaves than $T$, so by induction, $\Delta(T \setminus \{i\}) = 0$. Thus, if the above

74

equation holds, $\theta_i$ divides $\Delta(T)$. To prove the equation above, we will show the following:

$$\left.\frac{\det(K_T)}{\prod\limits_{e \in E(T)} \theta_e}\right|_{\theta_i = 0} = \frac{\det(K_{T \setminus \{i\}})}{\prod\limits_{e \in E(T) \setminus \{i\}} \theta_e} \cdot \left( \sum_{j \in \mathrm{Lv}(T) \setminus \{i\}} \frac{\varphi_T(p_{ij})}{\theta_i} \right) \text{ and} \qquad (4.18)$$

$$\left. D(T) \right|_{\theta_i = 0} = D(T \setminus \{i\}) \cdot \left( \sum_{j \in \mathrm{Lv}(T) \setminus \{i\}} \frac{\varphi_T(p_{ij})}{\theta_i} \right). \qquad (4.19)$$

On the left-hand side of Equation (4.18), the term corresponding to $\Gamma$ has degree $\deg_\Gamma(i) - 1$ in $\theta_i$. Therefore, the terms on the left-hand side that remain after setting $\theta_i = 0$ are those where $i$ is a leaf in $\Gamma$. The factorization on the right says that any such spanning tree is obtained by first finding a spanning tree on $\mathcal{K}^T \setminus \{i\}$, and then adding $i$ as a leaf. This proves Equation (4.18).

To prove Equation (4.19), let $v_i$ be the internal vertex of $T$ adjacent to $i$. Then

$$D(T) = \left( \theta_i + \sum_{\ell \in \mathrm{Lv}(T) \setminus \{i\}} \frac{\varphi_T(p_{i\ell})}{\theta_i} \right)^{\deg(v_i) - 2} \cdot \prod_{v \in \mathrm{Int}(T) \setminus \{v_i\}} \left( \sum_{\ell \in \mathrm{Lv}(T)} \prod_{e \in v \rightsquigarrow \ell} \theta_e \right)^{\deg(v) - 2}.$$

When $\theta_i = 0$, the leftmost factor is

$$\sum_{j \in \mathrm{Lv}(T) \setminus \{i\}} \frac{\varphi_T(p_{ij})}{\theta_i},$$

and the product of the rest of the factors is $D(T \setminus \{i\})$, which proves Equation (4.19). We conclude that $\theta_i$ divides $\Delta(T)$ for every $i \in \mathrm{Lv}(T)$. Now if $\Delta(T) \neq 0$, it contains at least one monomial term $m \cdot \theta_0^{a_0} \theta_1^{a_1} \cdots \theta_n^{a_n}$, where $\sum_{i=0}^n a_i = n - 1$, $a_i \geq 0$ for all $i$, and $m$ is a monomial in the non-leaf variables, We must have $a_i = 0$ for some $i$. But that contradicts that $\theta_i$ divides $\Delta(T)$, so we conclude that $\Delta(T)$ is identically zero. $\qquad \square$

We note that when $T$ is a star tree, Theorem 4.3.5 recovers the classical Cayley-Prüfer theorem.

**Corollary 4.3.6.** *Let $T$ be the star tree on $n + 1$ leaves, with each edge is weighted by $\theta_i$, where $i$ is the adjacent leaf. Then,*

$$\det(K_T(\theta)) = \prod_{i=0}^n \theta_i \left( \sum_{i=0}^n \theta_i \right)^{n-1}. \qquad (4.20)$$

## 4.4 Equivalence of Phylogenetic Trees up to Re-Rooting

In this section, we show that the ML degree of a Brownian motion tree model depends only on the (unlabeled) unrooted tree topology. In particular, we show that the ML degree does not depend on which of the $n+1$ leaves is chosen to be the root. Moreover, we show that if $T$ and $T'$ are two trees with the same unrooted tree topology, the MLE of $T'$ can be easily obtained from the MLE of $T$.

Let $T$ be a phylogenetic tree on leaves $\mathrm{Lv}(T) = \{0, \ldots, n\}$ with $0$ as its root. Let $r \in [n]$ and let $T'$ be the rooted phylogenetic tree with the same unrooted topology as $T$ obtained by rerooting at $r$. The non-root leaves of $T'$ are then $\mathrm{Lv}(T) \setminus \{r\}$. We consider them in the order $1, \ldots, r-1, 0, r+1, \ldots, n$. With this order, an arbitrary element $K'$ of $\mathcal{L}_{T'}^{-1}$ has entries

$$
k'_{ij} = \begin{cases}
\sum\limits_{t=1}^{n} p_{0t} & \text{for } i = j = r, \\
-p_{0j} & \text{for } i = r \text{ and } j \neq r, \\
\sum\limits_{\substack{t=0 \\ t \neq i}}^{n} p_{it} & \text{for } i = j \neq r, \\
-p_{ij} & \text{for } i \neq j \text{ and } i, j \neq r,
\end{cases}
\tag{4.21}
$$

where $p_{ij}$ are as in Equation (4.3). This gives an invertible linear transformation between $\mathcal{L}_T^{-1}$ and $\mathcal{L}_{T'}^{-1}$.

Given a symmetric matrix $S$, construct the symmetric matrix $S'$ by applying the following invertible linear transformation to the entries of $S$:

$$
\begin{aligned}
s'_{rr} &= s_{rr}, \\
s'_{rj} &= s_{rr} - s_{rj} \text{ for } j \neq r, \\
s'_{ii} &= s_{rr} + s_{ii} - 2s_{ri} \text{ for } i \neq r, \\
s'_{ij} &= s_{rr} - s_{ri} - s_{rj} + s_{ij} \text{ for } i \neq j \text{ and } i, j \neq r.
\end{aligned}
\tag{4.22}
$$

Note that this linear transformation is visibly invertible since each $s_{ij}$ can be written as a linear function of the entries of $S'$. Let $\mathrm{mle}(\mathcal{M}_T, S)$ denote the MLE for sample covariance matrix $S$ in $\mathcal{M}_T$. We consider $\mathrm{mle}(\mathcal{M}_T, S)$ to be written in the coordinates $(p_{ij})_{0 \leq i < j \leq n}$.

**Theorem 4.1.3.** *Let $T$ and $T'$ be phylogenetic trees with the same unlabeled, unrooted tree topology as $T$. Then*

*(a)* $\mathrm{mld}(\mathcal{M}_T) = \mathrm{mld}(\mathcal{M}_{T'})$ *and*

*(b)* $\mathrm{mle}(\mathcal{M}_T, S) = \mathrm{mle}(\mathcal{M}_{T'}, S')$ *for $S$ and $S'$ as in Equation (4.22), if both MLEs exist.*

*Proof.* Assume without loss of generality that $T$ has root $0$ and non-root leaf set $[n]$ and $T'$ has root $r \in \mathrm{Lv}(T)$ and the same unlabeled, unrooted topology as $T$. Let $K \in \mathcal{L}_T^{-1}$ in the $p_{ij}$ coordinates as in Equation (4.3). Let $K'$ be as Equation (4.21). Let $S$ be a sample covariance matrix for $\mathcal{M}_T$ and let $S'$ be as in Equation (4.22). The transformation Equation (4.22) was chosen so that $\mathrm{tr}(SK) = \mathrm{tr}(S'K')$. Indeed, we can compare the coefficients of each $s_{ij}$ in $\mathrm{tr}(SK)$ and $\mathrm{tr}(S'K')$, denoted $\mathrm{coeff}(s_{ij}, \mathrm{tr}(SK))$ and $\mathrm{coeff}(s_{ij}, \mathrm{tr}(S'K'))$ respectively. For the diagonal entries corresponding to leaves $i \in \mathrm{Lv}(T) \setminus \{r\}$, we have

$$\mathrm{coeff}(s_{ii}, \mathrm{tr}(S'K')) = \sum_{\substack{t=0 \\ t \neq i}}^{n} p_{it} = \mathrm{coeff}(s_{ii}, \mathrm{tr}(SK)).$$

For $s_{rr}$, we have

$$\mathrm{coeff}(s_{rr}, \mathrm{tr}(S'K')) = \sum_{t=1}^{n} p_{0t} + \sum_{\substack{i=1 \\ i \neq r}}^{n} \sum_{\substack{t=0 \\ t \neq i}}^{n} p_{it} - 2\sum_{j=1}^{n} p_{0j} - 2 \sum_{\substack{i<j \\ i,j \neq 0,r}} p_{ij}$$

$$= 2 \sum_{\substack{i=0 \\ i \neq r}}^{n} \sum_{\substack{j=0 \\ j \neq i,r}}^{n} p_{ij} + \sum_{\substack{j=0 \\ j \neq r}}^{n} p_{rj} - 2 \sum_{\substack{i=0 \\ i \neq r}}^{n} \sum_{\substack{j=0 \\ j \neq i,r}}^{n} p_{ij}$$

$$= \sum_{\substack{j=0 \\ j \neq r}}^{n} p_{rj} = \mathrm{coeff}(s_{rr}, \mathrm{tr}(S'K')).$$

Similar algebraic manipulations show that the coefficients of $s_{ij}$ for $i \neq j$ are also equal. Hence the traces are the same.

By Theorem 4.3.5, the term $\log \det(K)$ in the log-likelihood function depends only on the unrooted topology of the tree. So,

$$\ell_T(\theta|S) = \ell_{T'}(\theta|S'). \tag{4.23}$$

The map sending $S$ to $S'$ is invertible. Since the number of complex critical points of the log-likelihood function is fixed and equal to the ML degree for a generic choice of sample covariance matrix, (a) follows.

The map in Equation (4.21) sending $K$ to $K'$ and its inverse sending $K'$ to $K$ map the positive semidefinite cone to itself. Indeed, by the construction of the Farris transform, a matrix $K \in \mathcal{L}_T^{-1}$ is diagonally dominant with positive diagonal if and only if its image $K' \in \mathcal{L}_T^{-1}$ is diagonally dominant with positive diagonal. Equation (4.23) implies that the

maximizer of $\ell_T(\theta|S)$ is equal to the maximizer of $\ell_{T'}(\theta|S')$, and hence $(b)$. $\qquad\square$

## 4.5   ML Degrees of BMT Models on Star Trees

Let $T_n$ be the star tree on leaves $\{0, 1, \ldots, n\}$ with unique internal node $x$. We prove Theorem 4.1.2, which states that the maximum likelihood degree of its associated BMT model is $2^{n+1} - 2n - 3$. We will use Bézout's Theorem.

**Theorem 4.5.1** (Bézout's Theorem [41, §II.2]). *Given $n$ hypersurfaces of degrees $d_1, \ldots, d_n$ in a projective space of dimension $n$ over an algebraically closed field, if the intersection of the hypersurfaces is zero-dimensional, then the number of intersection points, counted with multiplicity, is equal to the product of the degrees $d_1 \cdots d_n$.*

For ease of notation, denote by $\theta_i$ the parameter for edge $\{i, x\}$ in the path parametrization of $\mathcal{L}_T^{-1}$. Let $K_{T_n}(\theta)$ denote the image of $\theta$ under the pullback of the path parametrization. Let $S \in \mathrm{PSD}^n(\mathbb{R})$ be sample covariance matrix. We start by setting up the system of score equations of $\ell_{T_n}(\theta|S)$. Then we count, with multiplicity, the solutions $\theta \in \mathbb{C}^{n+1}$ that have $\det(K_{T_n}(\theta)) \neq 0$.

**Proposition 4.5.2.** *The score equations of $\ell_{T_n}(\theta|S)$ have the form*

$$\frac{\partial \ell_{T_n}(\theta|S)}{\partial \theta_i} = \frac{1}{\theta_i} + \frac{n-1}{\theta_0 + \theta_1 + \cdots + \theta_n} - \sum_{\substack{j=0 \\ j \neq i}}^{n} c_{ij}\theta_j, \ \text{for } i = 0, \ldots, n, \qquad (4.24)$$

*where $c_{0j} = s_{jj}$, and $c_{ij} = s_{ii} + s_{jj} - 2s_{ij}$ for $i > 0$.*

*Proof.* Expanding out the expression for the trace of $SK_{T_n}(\theta)$ and applying Theorem 4.3.5 to $T_n$ gives

$$\mathrm{tr}(SK_{T_n}(\theta)) = \sum_{i=0}^{n} \sum_{j>i} c_{ij}\theta_i\theta_j \ \text{and} \ \det(K_{T_n}(\theta)) = \prod_{i=0}^{n} \theta_i \left(\sum_{i=0}^{n} \theta_i\right)^{n-1}.$$

Substituting these expressions into $\ell_{T_n}(\theta|S)$ and taking its partial derivatives gives Equations in (4.24). $\qquad\square$

For a system $F$ of polynomials, we denote by $V(F)$ its complex affine variety. In the following steps, we will restate the problem of computing the ML degree as counting solutions to a polynomial system $F_n$ (Lemma 4.5.3) and then as counting solutions to a *homogeneous*

78

polynomial system $\tilde{F}_n$ in projective space (Lemma 4.5.6). In all the steps, we need to remove solutions for which $\det(K(\theta)) = 0$.

We introduce a new variable $\psi$ which plays the role of the inverse of $\theta_0 + \theta_1 + \cdots + \theta_n$ by adding the equation $1 - \psi \sum_{i=0}^{n} \theta_i = 0$ to the set of the score equations. For a fixed sample covariance matrix $S = (s_{ij})$ and values $c_{ij}$ as in (4.24), let $F_n = \{f_0, \ldots, f_{n+1}\}$ be the following system of $n + 2$ polynomials in $\mathbb{C}[\theta_0, \ldots, \theta_n, \psi]$:

$$F_n: \quad f_i = 1 + \theta_i \left( (n-1)\psi + \sum_{j \neq i} c_{ij}\theta_j \right), \quad \text{for } i = 0, \ldots, n$$

$$\text{and } f_{n+1} = 1 - \psi \left( \sum_{i=0}^{n} \theta_i \right). \tag{4.25}$$

Note that for $i = 0, \ldots, n$, the polynomial $f_i$ is $\dfrac{\partial \ell_{T_n}(\theta|S)}{\partial \theta_i}$ with its denominator cleared. The goal of Lemma 4.5.3 is to show that the degree of the affine variety of this system intersected with the algebraic torus is exactly the maximum likelihood degree of $\mathcal{M}_{T_n}$. Let $\mathbb{C}^*$ denote the complex numbers without zero.

**Lemma 4.5.3.** *The maximum likelihood degree of the BMT model on $T_n$ is the degree of the ideal generated by $F_n$; that is $\mathrm{mld}(\mathcal{M}_{T_n}) = \deg V(F_n)$.*

*Proof.* Let $L_n$ denote the ideal generated by the score equations in the ring

$$A_n = \mathbb{C}[\theta_0^{\pm}, \ldots, \theta_n^{\pm}, (\theta_0 + \cdots + \theta_n)^{-1}].$$

By definition, the ML degree of $\mathcal{M}_{T_n}$ is $\deg V(L_n)$. We use the variable $\psi$ to represent $(\sum_{i=0}^{n} \theta_i)^{-1}$. By the first isomorphism theorem,

$$A_n/L_n \cong \mathbb{C}[\theta_0^{\pm}, \ldots, \theta_n^{\pm}, \psi]/\left(L_n + \langle 1 - \psi(\theta_0 + \cdots + \theta_n)\rangle\right).$$

By clearing denominators in the score equations, we see that the vanishing locus of $L_n + \langle 1 - \psi(\sum_{i=0}^{n} \theta_i)\rangle$ is isomorphic to that of the saturated ideal, $\langle F_n \rangle : (\det(K_{T_n}))^{\infty}$ in $\mathbb{C}[\theta_0, \ldots, \theta_n, \psi]$. Using the factorization of $\det(K_{T_n})$ given in Corollary 4.3.6, we see that this is the same as $V(F_n) \cap (\mathbb{C}^*)^{n+2}$. But if $\theta_i = 0$, then $f_i = 1 \neq 0$, and if $\psi = 0$, then $f_{n+1} = 1 \neq 0$. It follows that $V(F_n) \cap (\mathbb{C}^*)^{n+2} = V(F_n)$. Thus, $\mathrm{mld}(\mathcal{M}_{T_n}) = \deg V(L_n) = \deg V(F_n)$. $\qquad\square$

In order to apply Bézout's Theorem, we consider the homogenization of the system $F_n$. Given a system $G$ of homogeneous polynomials in $m$ variables, let $X(G) \subset \mathbb{P}\mathbb{C}^{m-1}$ denote its

complex projective variety. The homogenization of the system $F_n$ is the system $\tilde{F}_n$ of $n + 2$ homogeneous polynomials in the $n + 3$ variables $\theta_0, \ldots, \theta_n, \psi, z$:

$$\tilde{F}_n: \quad \tilde{f}_i = z^2 + \theta_i \left( (n-1)\psi + \sum_{j \neq i} c_{ij}\theta_j \right), \quad \text{for } i = 0, \ldots, n,$$

$$\text{and } \tilde{f}_{n+1} = z^2 - \psi \sum_{i=0}^{n} \theta_i. \tag{4.26}$$

Its solution set $X(\tilde{F}_n)$ lives in $(n + 2)$-dimensional complex projective space. In the next lemma, we consider the standard points in $\mathbb{PC}^{n+2}$:

$$e_i = X(\langle z, \psi, \theta_j, \mid j \neq i \rangle) \text{ for } i = 0, \ldots, n, \quad \text{and } e_{n+1} = X(\langle z, \theta_i \mid i = 0, \ldots, n \rangle).$$

We show that the maximum likelihood degree of $\mathcal{M}_{T_n}$ is exactly the degree of the projective variety $X(\tilde{F}_n)$ with these standard points removed.

**Lemma 4.5.4.** *The number of affine solutions to $F_n$, counted with multiplicity, in the algebraic torus is equal to the number of projective solutions, counted with multiplicity, to $\tilde{F}_n$ that are not standard points; that is, $\deg \left( V(F_n) \cap (\mathbb{C}^*)^{n+2} \right) = \deg \left( X(\tilde{F}_n) \setminus \{e_0, \ldots, e_{n+1}\} \right).$*

*Proof.* We prove the lemma by showing that

$$X(\tilde{F}_n) \setminus (\mathbb{PC}^*)^{n+2} = X(\tilde{F}_n, z) = \{e_0, \ldots, e_{n+1}\}.$$

The bijection $(\mathbb{C}^*)^{n+2} \to (\mathbb{PC}^*)^{n+2}$ sending $(\theta_0, \ldots, \theta_n, \psi) \mapsto [\theta_0 : \cdots : \theta_n : \psi : 1]$ concludes the rest.

Take $P = [\theta_0 : \cdots : \theta_n : \psi : z] \in X(\tilde{F}_n)$. We first show that $z = 0$ if and only if $\psi = 0$ or $\theta_i = 0$ for some $i \in \{0, \ldots, n\}$. Points not in the torus $(\mathbb{PC}^*)^{n+2}$ have at least one coordinate equal to zero. If any of the $\theta_i$ are zero, then $\tilde{f}_i(P) = z^2$, so $z = 0$ as well. When $\psi = 0$, we have $\tilde{f}_{n+1}(P) = 0$, so again $z = 0$. Conversely, suppose $P$ has $z = 0$. Our system becomes

$$\tilde{f}_i(P) = \theta_i \left( (n-1)\psi + \sum_{j \neq i} c_{ij}\theta_j \right) \text{ for } i = 0, \ldots, n, \quad \text{and } \tilde{f}_{n+1}(P) = \psi \left( \sum_{i=0}^{n} \theta_i \right).$$

If none of the remaining coordinates are zero, then we have the linear conditions

$$(n-1)\psi + \sum_{j \neq i} c_{ij}\theta_j = \sum_{i=1}^{n} \theta_i = 0.$$

This is equivalent to the singularity of the matrix of coefficients,

$$
C = \begin{pmatrix}
0 & c_{01} & \cdots & c_{0n} & n-1 \\
c_{01} & 0 & \cdots & c_{1n} & n-1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
c_{0n} & c_{1n} & \cdots & 0 & n-1 \\
n-1 & n-1 & \cdots & n-1 & 0
\end{pmatrix}.
$$

The matrix $C$ being singular is a polynomial condition in the $c_{ij}$'s. Moreover, we may assume that any principal minor, except the diagonal entries themselves, has full rank by the same argument. Since the sample covariance $S$ is generic, so are the values $c_{ij}$. It follows all but one of $\psi$ and $\theta_j$ for $j = 0, \ldots, n$ must be zero. There are $n + 2$ such points, precisely $e_0, \ldots, e_{n+1}$, and they clearly all lie in $X(\tilde{F}_n)$. $\qquad \square$

To finish our proof, we must compute the multiplicities of points $e_0, \ldots, e_{n+1}$ in $X(\tilde{F}_n)$, and subtract them from $\deg X(\tilde{F}_n)$. Let $\mathbb{C}[\![x_0, \ldots, x_n]\!]$ denote the ring of formal power series in variables $x_0, \ldots, x_n$. In order to compute these multiplicities, we make use of Theorem 4.5.5. The *standard monomials* of an ideal $I$ with respect to a local order are the monomials $x^\alpha$ such that $x^\alpha$ does not belong to the leading term ideal $\mathrm{LT}(I)$ with respect to this order.

**Theorem 4.5.5** ([11], Theorem 4.3). *Let $\widehat{R} = \mathbb{C}[\![x_0, \ldots, x_n]\!]$. Let $\widehat{J} \subset \widehat{R}$ be an ideal, $>$ a local order, and $\mathrm{LT}(\widehat{J})$ the leading term ideal for $\widehat{J}$ with respect to $>$. If $\widehat{R}/\widehat{J}$ contains finitely many standard monomials, then $\dim_{\mathbb{C}}(\widehat{R}/\widehat{J})$ is the number of standard monomials.*

In the proof of the next lemma, we define a local order on a given power series ring. We set a variable $\theta_i$ or $\psi$ equal to 1 to localize at the prime ideal of the corresponding standard point $e_i$. Then we use the polynomials $\tilde{f}_0, \ldots, \tilde{f}_{n+1}$ to expand each $\theta_i$ and $\psi$ as a power series in $z$. This allows us to find the standard monomials and compute the multiplicity $e_i$.

**Lemma 4.5.6.** *1. The multiplicity of the standard point $e_i$ for $i = 0, \ldots, n$ in $X(\tilde{F}_n)$ is four.*

*2. The multiplicity of the standard point $e_{n+1}$ in $X(\tilde{F}_n)$ is two.*

*Proof of (1).* By symmetry, we only need to prove the lemma for $e_0$. Let $R = \mathbb{C}[\theta_1, \ldots, \theta_n, \psi, z]$ and $\widehat{R} = \mathbb{C}[\![\theta_1, \ldots, \theta_n, \psi, z]\!]$. Substituting $\theta_0 = 1$ to Equation (4.26), we

obtain

$$\bar{f}_0 = z^2 + (n-1)\psi + \sum_{j \neq 0} c_{0j}\theta_j,$$

$$\bar{f}_i = z^2 + \theta_i \left( (n-1)\psi + c_{0i} + \sum_{j \neq i,0} c_{ij}\theta_j \right) \quad \text{for } i = 1, \ldots, n,$$

$$\bar{f}_{n+1} = z^2 - \psi \left( 1 + \sum_{i=1}^{n} \theta_i \right).$$

Denote $J = \langle \bar{f}_0, \ldots, \bar{f}_n, \bar{f}_\psi \rangle \subseteq R$ and $p = \langle \theta_1, \ldots, \theta_n, z, \psi \rangle \subseteq R$. Let $(R/J)_p$ be the local ring at $p$. By definition, the multiplicity of $e_0$ is the length of $(R/J)_p$. So, to show that $\text{mult}(e_0) = 4$, we prove that $\text{length}\,(R/J)_p = 4$. Since $(R/J)_p$ is a Noetherian local ring, we have that $\text{length}\,(R/J)_p = \dim_{\mathbb{C}}(R/J)_p$. Since completion preserves dimension, we reduce to computing $\dim_{\mathbb{C}}(\widehat{R}/\widehat{J})$, where $\widehat{J}$ is ideal of the embedding of $J$ in $\widehat{R}$.

First, we use the functions $\bar{f}_1, \ldots, \bar{f}_{n+1}$ to write power series expansions for $\psi$ and $\theta_j$ for $j = 1, \ldots, n$ in terms of $z$. Since $\bar{f}_j = 0$ for $j = 1, \ldots, n$, we have

$$\theta_j = \frac{-\frac{1}{c_{0j}}z^2}{1 + (n-1)\psi + \sum_{i \neq 0,j} \frac{c_{ij}}{c_{0j}}\theta_i} = -\frac{1}{c_{0j}}z^2 \left( 1 - \frac{n-1}{c_{0j}}\psi - \sum_{i \neq 0,j} \frac{c_{ij}}{c_{0j}}\theta_i + \cdots \right).$$

Similarly, since $\bar{f}_{n+1} = 0$, we have

$$\psi = \frac{z^2}{1 + \sum_{i \neq 0} \theta_i} = z^2 \left( 1 - \sum_{i \neq 0} \theta_i + \cdots \right).$$

By substitution, we obtain power series expansions for $\theta_j$, $j \neq 0$, and $\psi$ up to degree 4 in $z$.

$$\theta_j = -\frac{1}{c_{0j}}z^2 + \left( -\frac{n-1}{c_{0j}} + \sum_{i \neq 0,j} \frac{c_{ij}}{c_{0j}^2} \right) z^4 + O(z^6), \tag{4.27}$$

$$\psi = z^2 + \left( \sum_{i \neq 0} \frac{1}{c_{0i}} \right) z^4 + O(z^6).$$

Denote by $\bar{g}_0$ the equation obtained by writing all the variables in $\bar{f}_0$ in terms of $z$ using the

equations in (4.27), so that

$$\bar{g}_0 = z^2 + (n-1)\left(z^2 + \left(\sum_{i \neq 0} \frac{1}{c_{0i}}\right)z^4 + O(z^6)\right)$$

$$+ \sum_{i \neq 0} c_{0i}\left(-\frac{1}{c_{0i}}z^2 + \left(-\frac{n-1}{c_{0i}} + \sum_{k \neq 0,i} \frac{c_{ik}}{c_{0i}^2}\right)z^4 + O(z^6)\right)$$

$$= \left(\sum_{i \neq 0} \frac{1}{c_{0i}}\left(1 + \sum_{k \neq 0,i} c_{ik}\right) - n(n-1)\right)z^4 + O(z^6).$$

By the genericity of the $c_{ij}$'s, the coefficient of $z^4$ in the power series $g_0$ is generically non-zero. Let $G = \{\bar{g}_0, \bar{f}_1, \ldots, \bar{f}_n, \bar{f}_{n+1}\}$. Note that $\widehat{J} = \langle G \rangle$. We further claim that $G$ is a standard basis for $\widehat{J}$. Indeed, take $<$ to be a negative graded monomial order, i.e. $1 > \theta_1, \ldots, \theta_j, \psi, z > \theta_i\theta_j, \ldots$, and so on. Since the $<$-leading terms of $\bar{g}_0, \bar{f}_j, \ldots, \bar{f}_{n+1}$ are relatively prime, the set $G$ is a standard basis for $\widehat{J}$. Thus, $\mathrm{LT}(\widehat{J}) = \langle \theta_1, \ldots, \theta_n, \psi, z^4 \rangle$. There are four standard monomials, $1, z, z^2, z^3$, not in $\mathrm{LT}(\widehat{J})$. By Theorem 4.5.5, $\dim_{\mathbb{C}} \widehat{R}/\widehat{J} = \mathrm{length}(R/J)_p = \mathrm{mult}(e_0) = 4$. $\qquad \square$

*Proof of (2).* After dehomogenizing by $\psi = 1$, we obtain the system

$$\bar{f}_i = z^2 + \theta_i\left((n-1) + \sum_{j \neq i} c_{ij}\theta_j\right) \quad \text{for } i = 1, \ldots, n,$$

$$\bar{f}_{n+1} = z^2 - \left(\sum_{i=0}^n \theta_i\right).$$

As in the previous case, let $R = \mathbb{C}[\theta_0, \ldots, \theta_n, z]$, $J = \langle \bar{f}_0, \ldots, \bar{f}_n, \bar{f}_{n+1} \rangle$ and $p = \langle \theta_0, \ldots, \theta_n, z \rangle$. Consider the localization $(R/J)_p$ and its completion $\widehat{R}/\widehat{J}$. We want to show that $\dim_{\mathbb{C}}(\widehat{R}/\widehat{J})_p = 2$. Since $\bar{f}_i = 0$ for $i = 0, \ldots, n$, we may again solve for $\theta_i$ in terms of $z$ and obtain

$$\theta_i = \frac{-z^2}{(n-1) + \sum_{j \neq i} c_{ij}\theta_j} = -\frac{z^2}{n-1}\left(1 - \sum_{j \neq i} c_{ij}\theta_j\right) = z^2\left(-\frac{1}{n-1} + O(z^4)\right).$$

Substituting the relations derived above into $\bar{f}_{n+1}$, the power series,

$$\bar{g}_\psi = z^2 + \frac{n+1}{n-1}z^2 + O(z^4) = \frac{2n}{n-1}z^2 + O(z^4).$$

Note that $\widehat{J} = \langle \bar{f}_0, \ldots, \bar{f}_n, \bar{g}_\psi \rangle$. For any negative graded monomial order $<$, the leading

terms of $\bar{f}_0, \ldots, \bar{f}_n, \bar{g}_\psi$ are relatively prime. Thus, $\bar{f}_0, \ldots, \bar{f}_n, \bar{g}_\psi$ form a standard basis of $\widehat{J}$. It follows that

$$\mathrm{LT}(\widehat{J}) = \langle \theta_0, \ldots, \theta_n, z^2 \rangle.$$

By Theorem 4.5.5, $\dim_{\mathbb{C}} \widehat{R}/\widehat{J} = \mathrm{mult}(e_{n+1}) = 2$. □

We are ready to prove the main result.

**Theorem 4.1.2.** *The maximum likelihood degree of the Brownian motion star tree model on $n + 1$ leaves is $2^{n+1} - 2n - 3$.*

*Proof.* Since $\tilde{F}_n$ is a homogeneous system of $n + 2$ quadratic polynomials satisfying the conditions of Bézout's Theorem, it has $2^{n+2}$ solutions, considering multiplicity. By Lemma 4.5.3 and Lemma 4.5.4, we have

$$\mathrm{mld}(\mathcal{M}_{T_n}) = \frac{1}{2} \deg \left( X(\tilde{F}_n) \setminus \{e_1, \ldots, e_{n+2}\} \right).$$

Applying Lemma 4.5.6 to remove these standard points with their multiplicities, we obtain

$$\mathrm{mld}(\mathcal{M}_{T_n}) = \frac{1}{2} \left( \deg X(\tilde{F}_n) - \sum_{i=1}^{n+2} \mathrm{mult}(e_i) \right)$$
$$= \frac{1}{2} \left( 2^{n+2} - 4(n+1) - 2 \right)$$
$$= 2^{n+1} - 2n - 3. \quad \square$$

## 4.6  Discussion

In this paper, we use algebraic techniques to give a formula for the ML degree of the BMT model on a star tree. Theorem 4.3.5 is a generalization of the Cayley-Prüfer Theorem, which gives a formula for the determinant of a matrix in $\mathcal{L}_T^{-1}$. We used this result to show that the ML degree of the BMT model is the same for all trees with the same unrooted topology.

Computational results show that our formula does not generalize to other BMT models. For example, the table in Figure 4.3 lists all possible non-star tree topologies in 7 leaves, the degree of the vanishing ideal (deg), reciprocal maximum likelihood degree (rmld), and the maximum likelihood degree (mld). We computed the ML degrees using the software `HomotopyContinuation.jl` [6]. Unlike in the case of the reciprocal ML degree (see [4]), there is not an obvious way to extend the formula for the ML degree of a star tree to a formula for trees of any topology. The methods that we use in Section 4.5 also do not directly extend to arbitrary trees. In many cases, the common vanishing locus of $\det(K_T(\theta))$

and the score equations with denominators cleared is positive dimensional, so one cannot directly apply Bézout's theorem. However, we are hopeful that our formulas for $\det(K_T(\theta))$ and the degree of the path map will be useful in future approaches to this problem.
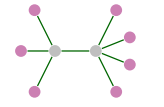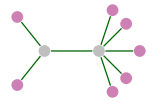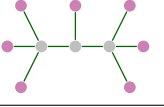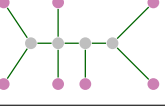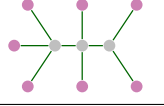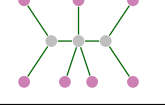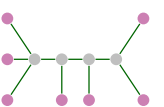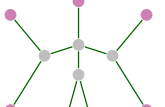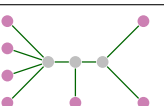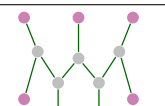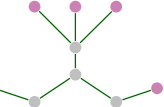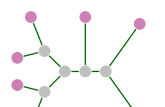
| tree topology | deg | rmldeg | mldeg | tree topology | deg | rmldeg | mldeg |
|---|---|---|---|---|---|---|---|
|  | 93 | 44 | 259 |  | 95 | 26 | 53 |
|  | 90 | 16 | 221 |  | 51 | 4 | 83 |
|  | 77 | 16 | 181 |  | 47 | 11 | 81 |
|  | 61 | 4 | 115 |  | 42 | 4 | 63 |
|  | 60 | 11 | 101 |  | 42 | 1 | 61 |
|  | 61 | 4 | 99 |  | 53 | 1 | 61 |

Figure 4.3: ML degrees, reciprocal ML degrees and algebraic degrees of BMT models on phylogenetic trees with 7 leaves.

# BIBLIOGRAPHY

[1] Federico Ardila and Caroline J. Klivans. The Bergman complex of a matroid and phylogenetic trees. *J. Combin. Theory Ser. B*, 96(1):38–49, 2006.

[2] Philipp Benner, Miroslav Bačák, and Pierre-Yves Bourguignon. Point estimates in phylogenetic reconstructions. *Bioinformatics*, 30(17):i534–i540, 2014.

[3] Louis J. Billera, Susan P. Holmes, and Karen Vogtmann. Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.*, 27(4):733–767, 2001.

[4] Tobias Boege, Jane Ivy Coons, Christopher Eur, Aida Maraj, and Frank Roettger. Reciprocal maximum likelihood degrees of Brownian motion tree models. *Le Matematiche*, 76(2):383–398, 2021.

[5] David Brawand, Magali Soumillon, Anamaria Necsulea, Philippe Julien, Gábor Csárdi, Patrick Harrigan, Manuela Weier, Angélica Liechti, Ayinuer Aximu-Petri, Martin Kircher, Frank W. Albert, Ulrich Zeller, Philipp Khaitovich, Frank Grützner, Sven Bergmann, Rasmus Nielsen, Svante Pääbo, and Henrik Kaessmann. The evolution of gene expression levels in mammalian organs. *Nature*, 478(7369):343–348, 2011.

[6] Paul Breiding and Sascha Timme. HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia. In *International Congress on Mathematical Software*, pages 458–465. Springer, 2018.

[7] Andrei Comăneci and Michael Joswig. Tropical medians by transportation. *Mathematical Programming*, pages 1–27, 2023.

[8] Jane Ivy Coons, Shelby Cox, Aida Maraj, and Ikenna Nometa. Maximum likelihood degrees of brownian motion tree models: Star trees and root invariance, 2024. preprint arXiv 2402.10322.

[9] Jane Ivy Coons, Orlando Marigliano, and Michael Ruddy. Maximum likelihood degree of the two-dimensional linear Gaussian covariance model. *Algebraic Statistics*, 11(2):107–123, 2020.

[10] Natalie Cooper and Andy Purvis. Body size evolution in mammals: complexity in tempo and mode. *The American Naturalist*, 175(6):727–738, 2010.

[11] David A. Cox, John Little, and Donal O'Shea. *Using algebraic geometry*, volume 185 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2005.

[12] Shelby Cox. Classifying tree topology changes along tropical line segments. *Algebraic Statistics*, 14(1):71–90, 2023.

[13] Shelby Cox and Mark Curiel. The tropical polytope is the set of all weighted tropical fermat-weber points, 2023. preprint arXiv 2310.07732.

[14] Bhaskar DasGupta, Xin He, Ming Li, John Tromp, and Louxin Zhang. *Nearest Neighbor Interchange and Related Distances*, pages 1402–1405. Springer New York, New York, NY, 2016.

[15] Jesús A. De Loera, Jörg Rambau, and Francisco Santos. *Triangulations*, volume 25 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, 2010. Structures for algorithms and applications.

[16] Brian Eriksson, Gautam Dasarathy, Paul Barford, and Robert Nowak. Toward the practical use of network tomography for internet topology discovery. In *2010 Proceedings IEEE INFOCOM*, pages 1–9. IEEE, 2010.

[17] Joseph Felsenstein. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of human genetics*, 25(5):471, 1973.

[18] Robert P Freckleton and Paul H Harvey. Detecting non-Brownian trait evolution in adaptive radiations. *PLoS biology*, 4(11):e373, 2006.

[19] Thomas F Hansen and Emília P Martins. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, 50(4):1404–1417, 1996.

[20] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. Reprint of the 1952 edition.

[21] Luke Harmon. *Phylogenetic comparative methods: learning from trees*. EcoEvoRxiv, 2019.

[22] Susan Holmes. Statistics for phylogenetic trees. *Theoretical population biology*, 63(1):17–32, 2003.

[23] Birkett Huber, Jörg Rambau, and Francisco Santos. The Cayley trick, lifting subdivisions and the Bohne-Dress theorem on zonotopal tilings. *J. Eur. Math. Soc. (JEMS)*, 2(2):179–198, 2000.

[24] Michael Joswig. *Essentials of tropical combinatorics*, volume 219 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, [2021] ©2021.

[25] Motoo Kimura et al. Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626, 1968.

[26] Gustav Kirchhoff. Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird. *Annalen der Physik*, 148(12):497–508, 1847.

[27] Sudhir Kumar. Molecular clocks: four decades of evolution. *Nature Reviews Genetics*, 6(8):654–662, 2005.

[28] Ming Li, John Tromp, and Louxin Zhang. Some notes on the nearest neighbour interchange distance. In *International Computing and Combinatorics Conference*, pages 343–351. Springer, 1996.

[29] Diane Maclagan and Bernd Sturmfels. *Introduction to tropical geometry*, volume 161 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015.

[30] Emanuel Margoliash. Primary structure and evolution of cytochrome c. *Proceedings of the National Academy of Sciences*, 50(4):672–679, 1963.

[31] Ezra Miller, Megan Owen, and J Scott Provan. Polyhedral computational geometry for averaging metric phylogenetic trees. *Advances in Applied Mathematics*, 68:51–91, 2015.

[32] Anthea Monod, Bo Lin, Ruriko Yoshida, and Qiwen Kang. Tropical geometry of phylogenetic tree space: A statistical perspective, 2022. arXiv preprint arXiv:1805.12400.

[33] G William Moore, M Goodman, and J Barnabas. An iterative approach from the standpoint of the additive hypothesis to the dendrogram problem posed by molecular data sets. *Journal of theoretical biology*, 38(3):423–457, 1973.

[34] Yurii Nesterov. *Introductory lectures on convex optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004. A basic course.

[35] Tom M. W. Nye. Principal components analysis in the space of phylogenetic trees. *Ann. Statist.*, 39(5):2716–2739, 2011.

[36] Megan Owen and J Scott Provan. A fast algorithm for computing geodesic distances in tree space. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(1):2–13, 2010.

[37] David F Robinson. Comparison of labeled trees with valency three. *Journal of Combinatorial Theory, Series B*, 11(2):105–119, 1971.

[38] robjohn (https://math.stackexchange.com/users/13854/robjohn). Elementary central binomial coefficient estimates. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/932509 (version: 2019-10-07).

[39] Francisco Santos. The Cayley trick and triangulations of products of simplices. In *Integer points in polyhedra—geometry, number theory, algebra, optimization*, volume 374 of *Contemp. Math.*, pages 151–177. Amer. Math. Soc., Providence, RI, 2005.

[40] Joshua G Schraiber, Yulia Mostovoy, Tiffany Y Hsu, and Rachel B Brem. Inferring evolutionary histories of pathway regulation from transcriptional profiling data. *PLoS computational biology*, 9(10):e1003255, 2013.

[41] I. R. Shafarevich. *Basic algebraic geometry.* Springer Study Edition. Springer-Verlag, Berlin-New York, 1977. Translated from the Russian by K. A. Hirsch, Revised printing of Grundlehren der mathematischen Wissenschaften, Vol. 213, 1974.

[42] David Speyer and Bernd Sturmfels. The tropical Grassmannian. *Adv. Geom.*, 4(3):389–411, 2004.

[43] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

[44] Bernd Sturmfels. On the Newton polytope of the resultant. *J. Algebraic Combin.*, 3(2):207–236, 1994.

[45] Bernd Sturmfels, Sascha Timme, and Piotr Zwiernik. Estimating linear covariance models with numerical nonlinear algebra. *Algebraic Statistics*, 11(1):31–52, 2020.

[46] Bernd Sturmfels, Caroline Uhler, and Piotr Zwiernik. Brownian motion tree models are toric. *Kybernetika (Prague)*, 56(6):1154–1175, 2020.

[47] Seth Sullivant. *Algebraic statistics*, volume 194 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.

[48] Michael Truell, Jan-Christian Hütter, Chandler Squires, Piotr Zwiernik, and Caroline Uhler. Maximum likelihood estimation for Brownian motion tree models based on one sample. *arXiv preprint arXiv:2112.00816*, 2021.

[49] Yolanda Tsang, Mehmet Yildiz, Paul Barford, and Robert Nowak. Network radar: tomography from round trip time measurements. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 175–180, 2004.

[50] Amy Willis. Confidence sets for phylogenetic trees. *J. Amer. Statist. Assoc.*, 114(525):235–244, 2019.

[51] Ruriko Yoshida and Shelby Cox. Tree topologies along a tropical line segment. *Vietnam Journal of Mathematics*, pages 1–25, 2022.

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

[53] Carl Zimmer. Most new york coronavirus cases came from europe, genomes show. *International New York Times*, pages NA–NA, 2020.