THE SHAPE OF THE ONE-DIMENSIONAL PHYLOGENETIC LIKELIHOOD FUNCTION

VU DINH AND FREDERICK A. MATSEN IV

ABSTRACT. By fixing all parameters in a phylogenetic likelihood model except for one branch length, one obtains a one-dimensional likelihood function. In this work, we introduce a mathematical framework to characterize the shapes of such one-dimensional phylogenetic likelihood functions. This framework is based on analyses of algebraic structures on the space of all frequency patterns with respect to a polynomial representation of the likelihood functions. Using this framework, we provide conditions under which the one-dimensional phylogenetic likelihood functions are guaranteed to have at most one stationary point, and this point is the maximum likelihood branch length. These conditions cover an abundant class of evolution models that includes all binary models, the Jukes-Cantor model and the Felsenstein 1981 model.

We then prove that for the simplest model that does not satisfy our conditions, namely, the Kimura 2-parameter model, the one-dimensional likelihood functions may have multiple stationary points. As a proof of concept, we construct a non-degenerate example in which the phylogenetic likelihood function has two local maxima and a local minimum. To construct such examples, we derive a general method of constructing a tree and sequence data with a specified frequency pattern at the root. We then extend the result to prove that the space of all rescaled and translated one-dimensional phylogenetic likelihood functions under the Kimura 2-parameter model is dense in the space of all continuous functions on $[0, \infty)$ with finite limits. These results indicate that one-dimensional likelihood functions under advanced evolutionary models can be more complex than it is typically assumed by phylogenetic inference algorithms; however, these complexities can be effectively captured by the Kimura 2-parameter model.

1. INTRODUCTION

The likelihood of a phylogenetic model is a function of the parameters of continuous time Markov chains (CTMCs) used to model sequence evolution along each branch. It is common to assume a single rate matrix and stationary frequency for the CTMCs but allow the branch lengths to vary, representing a single evolutionary process but differing amounts of evolution along each branch. Commonly used maximum-likelihood phylogeny programs improve likelihood by modifying branch lengths iteratively and one at a time [1]. The general approach for numerical maximization of the one-dimensional likelihood function given by fixing every parameter except for one branch length is to iteratively sample the function at a number of points, use surrogate functions to fit simple curves to those points, and use those fits as approximations to locate the maximum branch length. For example, programs often employ Newton’s method, in which the intuitive idea is to use first and second derivatives to approximate the likelihood function (varying along that
branch) by a surrogate quadratic function. Since evaluations of the likelihoods (and their derivatives) are computationally expensive, many approaches have been tried to improve the efficiency of this optimization procedure [1].

Such approaches, however, rely on the assumptions that one-dimensional phylogenetic likelihood functions belong to some class of simple functions, and that the surrogate model can, at least, capture the shape of the functions. While there has been a considerable amount of work on finding multiple maxima of the multi-dimensional likelihood surfaces parameterized by all branch lengths for a tree [2–4], little has been done about the shapes of one-dimensional phylogenetic likelihood functions. Indeed, the answers to simple one-dimensional questions underlying practical methods are not known.

In this work, we introduce a mathematical framework to characterize the shapes of such one-dimensional phylogenetic likelihood functions. This framework is based on analyses of algebraic structures on the space of all frequency patterns with respect to a polynomial representation of the likelihood functions. Specifically, we introduce the new concept of logarithmic relative frequency patterns and analyze algebraic structures on the space of such patterns. These structures, along with the characteristic polynomial representations of one-dimensional phylogenetic likelihood functions, open a new way to explore the space of all possible likelihood functions. Moreover, by composing these structures, we are able to tackle the inverse problem of constructing a phylogenetic tree that has a given frequency pattern at the root. This enables us to construct phylogenetic trees that approximate any given likelihood function with arbitrary precision.

Using this framework, we provide conditions under which the one-dimensional phylogenetic likelihood functions are guaranteed to have at most one stationary point, and this point is the maximum likelihood branch length. These conditions cover an abundant class of evolution models that includes all binary models, the Jukes-Cantor model [5] and the Felsenstein 1981 model [6]. We then prove that for the simplest model that does not satisfy our conditions, namely, the Kimura 2-parameter model [7], the one-dimensional likelihood functions may have multiple stationary points. As a proof of concept, we construct a non-degenerate example in which the phylogenetic likelihood function has two local maxima and a local minimum.

We then extend the result to prove that the space of all rescaled and translated one-dimensional phylogenetic likelihood functions under the Kimura 2-parameter model is dense in the space of all continuous functions on \([0, \infty)\) with finite limits. These results indicates that one-dimensional likelihood functions under advanced evolutionary models can be more complex than it is typically assumed by phylogenetic inference algorithms; however, these complexities can be effectively captured by the Kimura 2-parameter model.

2. Background and Definitions

2.1. Notation. Our setting is the standard IID setting for likelihood-based phylogenetics with a finite number of sites; see [8] for more details. Let \(\Omega\) denote the set of states and let \(r = |\Omega|\). For an unrooted tree \(T\) with \(N\) taxa, we denote by \(E(T)\) and \(V(T)\) the set of edges and vertices of \(T\), respectively. We assume that the mutation events occur according to a continuous time Markov chain on states \(\Omega\). In this paper, an evolutionary model refers to a collection \(\mathcal{H}(\Omega)\) of \((Q, \pi)\) pairs,
where $\pi$ is a vector of stationary frequencies and $Q$ is a rate matrix on $\Omega$ that is reversible with respect to $\pi$. If at every edge of the tree $\tau$, the matrix-frequency pair $(Q_e, \pi)$ belongs to $\mathcal{H}$, we say that $\tau$ is a tree under evolutionary model $\mathcal{H}$. The rate matrix $Q_e$ and the branch length $t_e$ on each edge $e$ define the transition matrix $P_e = e^{Q_e t_e}$ on edge $e$.

Let $\psi = (\psi_1, \psi_2, ..., \psi_S) \in \Omega^{N \times S}$ be the observed sequences of length $S$ over $N$ taxa; our goal is to infer the maximum likelihood branch lengths for $\tau$ from $\psi$. We first orient the edges of $\tau$ away from an arbitrarily chosen root $\rho$ of the tree. (We can choose the root arbitrarily since each $P_e$ is reversible with respect to $\pi$). The likelihood of observing $\psi$ given the tree topology $\tau$ and the vector of branch lengths $t = (t_e)_{e \in E(T)}$ has the form

$$L(\psi|t) = \prod_{s=1}^{S} \sum_{a} \pi(a) \prod_{(u,v) \in E(T)} P^{uv}_{a_u a_v}(t_{uv})$$

where $a$ ranges over all extensions of $\psi$ to the internal nodes of $T$ and $a_u$ denotes the assigned state of node $u$ by $a$.

To investigate the one-dimensional likelihood function on one branch $e_0$, we fix all other branches, partition the set of all extensions of $\psi$ according to their labels at the end points of $e_0$, and split $E(T)$ into two sets of edges $E_{\text{left}}$ and $E_{\text{right}}$ corresponding to the location of the edges with respect to $e_0$. The likelihood function can be rewritten as a univariate function of $t$, the branch length of $e_0$:

$$L(\psi|t) = \prod_{s=1}^{S} \sum_{ij} \sum_{a \in A_{ij}} \pi(a) \left( \prod_{e \in E_{\text{left}}} P^{e}_{a_u a_v}(t_{uv}) \right) \times P^{e_0}_{ij}(t) \times \left( \prod_{e \in E_{\text{right}}} P^{e}_{a_u a_v}(t_{uv}) \right)$$

where $A_{ij}$ denotes the set of all extensions of $\psi$ for which the labels at the left end point and the right end point of $e_0$ are $i$ and $j$, respectively.

By grouping the products over $E_{\text{left}}$ and $E_{\text{right}}$ as well as the sum over $a$ in a single term $b^{e_0}_{ij}$, we can define the one-dimensional log-likelihood function as

$$\ell_{e_0}(t) = \log L(\psi|t) = \sum_{s=1}^{S} \log \left( \sum_{ij} b^{e_0}_{ij} P^{e_0}_{ij}(t) \right).$$

Such $\ell_{e_0}(t)$ are the object of study of this paper. For convenience, we will assume that $e_0$ has been chosen and will drop the index $e_0$ hereafter.

In the rest of the paper, we will consider a number of different evolutionary models. These DNA substitution models differ in terms of the parameters used to describe the rates at which one state replaces another during evolution and the stationary frequencies:

- Jukes-Cantor model [5]: this model assumes equal stationary frequencies ($\pi_A = \pi_C = \pi_T = \pi_G = 1/4$) and equal mutation rates.
- Felsenstein 1981 model [6]: this is an extension of the Jukes-Cantor model in which stationary frequencies are allowed to vary.
- Kimura 2-parameter model [7]: this model assumes equal stationary frequencies, but distinguishes between the rates of transitions ($A \leftrightarrow G$, i.e.
from purine to purine, or $C \leftrightarrow T$, i.e. from pyrimidine to pyrimidine) and transversions (from purine to pyrimidine or vice versa). Following common usage, we use $\kappa$ to denote the transition/transversion rate ratio.

While the focus here is on DNA models, we emphasize that our theoretical framework is capable of analyzing any time-reversible evolutionary model on any state space. In fact, we do not assume a uniform molecular clock, or even a single evolutionary model along the edges of the tree.

2.2. Characteristic polynomials of one-dimensional phylogenetic likelihood functions. We will frequently use the following assumption:

Assumption 2.1. The eigenvalues of the rate matrix $Q$ are equal to

$$0 = d_0 \gamma \geq -d_1 \gamma \geq -d_2 \gamma \geq \ldots \geq -d_p \gamma$$

for some positive number $\gamma$ and non-negative integers $d_1, \ldots, d_p$.

The following remark, whose proof is provided in the Appendix, guarantees that Assumption 2.1 does not affect the generality of our analyses up to an arbitrarily small approximation error:

Remark 2.1. The set of rate matrices $Q$ for a given evolutionary model that satisfy Assumption 2.1 is dense in the set of rate matrices under the same evolutionary model.

Under Assumption 2.1, if we denote the entries of the diagonalizing matrix $M$ and $N$ of $Q$ by $m_{ij}$ and $n_{ij}$, respectively, then the transition probabilities can be computed as

$$P_{ij}(t) = \sum_k m_{ik}e^{-d_k \gamma t}n_{kj}.$$ 

By reparametrizing with $x := e^{-\gamma t}$, we can represent these transition probabilities as polynomial functions

$$P_{ij}(x) = \sum_k m_{ik}x^{d_i}n_{kj}.$$ 

Similarly, the log-likelihood function can be rewritten as

$$\ell(x) = \sum_{s=1}^S \log(\lambda_s(x))$$

where

$$\lambda_s(x) = \sum_{ij} b_{ij}^s P_{ij}(x).$$

Hereafter, we will refer to $P_{ij}(x)$ and $\lambda_s(x)$ as the transition polynomials of the evolutionary model and the characteristic polynomials of the one-dimensional phylogenetic likelihood function, respectively.

As we will see in later sections, this polynomial representations enable us to exploit many algebraic and analytic properties of the likelihood functions. The most noticeable feature is that one can use the Fundamental Theorem of Algebra to factorize $\lambda_s(x)$ as products of linear and quadratic polynomials. As a result, the log-likelihood function can be written in the form

$$\ell(x) = \sum_{s=1}^S \sum_{i=1}^{i_{s,1}} \log(\alpha_{s,i} + \beta_{s,i}x)$$

$$+ \sum_{s=1}^S \sum_{i=1}^{i_{s,2}} \log(\mu_{s,i} + \nu_{s,i}x + \omega_{s,i}x^2).$$
SHAPE OF ONE-DIMENSIONAL PHYLOGENETIC LIKELIHOOD FUNCTION

where \( \mu_{s,i}, \nu_{s,i}, \omega_{s,i} \) are the (real) coefficients of the quadratic polynomials in the decomposition of \( \lambda_s \), while \( \alpha_{s,i}, \beta_{s,i} \) are coefficients of the linear terms in the decomposition.

This enables us to decompose a complicated evolutionary model into smaller modules, each of which can be approximated either by a “linear” model (like the binary symmetric model) or by a “quadratic” model (like the Kimura 2-parameter model). In Section 3, we use this formulation prove that if the phylogenetic log-likelihood function is essentially linear (that is, there are no quadratic terms in the expression), its shape resembles those generated by the binary symmetric model, with a unique stationary point that is also the maximum point. In Section 5, we illustrate that this property does not hold for quadratic models by constructing a counter-example with the Kimura 2-parameter model. Finally, in Section 6, we use this formulation once again to prove that the space of all rescaled and translated one-dimensional phylogenetic likelihood functions under the Kimura 2-parameter model is dense in the space of all continuous functions on \([0, \infty)\) with finite limits.

3. Uniqueness of the stationary point

In this section, we discuss a condition under which the uniqueness of the stationary branch length is guaranteed. The analyses in this section stem from two observations:

1. If for every site index \( s \), the characteristic polynomial \( \lambda_s \) has no non-real root, then the likelihood function can be decomposed into smaller modules, each of which resemble the the binary symmetric model.
2. The likelihood functions of binary symmetric models and summations of such models are incave.

**Definition 3.1** (Hanson [9]). A vector-valued function \( f \) is said to be incave in \( \mathbb{R}^n \) if there exists a vector-valued function \( \phi(t,u) \) such that

\[
 f(t) - f(u) \leq \phi(t,u) \cdot \nabla f(u), \quad \forall t, u \in \mathbb{R}^n
\]

where \( \nabla f \) denotes the gradient of \( f \).

Incave functions were introduced in the optimization literature as a generalization of concave functions [9]. It can be proved that function is incave if and only if every stationary point is a global maximum [10]. Moreover, for real-valued functions, the following result also holds:

**Lemma 3.1.** If \( f \) is a real-valued incave function with a finite number of stationary points, then \( f \) has at most one stationary point. Moreover, if such a point exists, it is also a global maximum.

**Proof.** Denote \( A = \{ t \in [0, \infty) : f'(t) = 0 \} \) and assume that \( A \) has more than one element. Since \( A \) is finite, we can choose two elements \( t_1 \) and \( t_2 \) in \( A \) such that the interval \((t_1, t_2) \subset \mathbb{R} - A\). Since \( f \) is incave, every stationary point of \( f \) is a global maximum. We deduce that \( t_1 \) and \( t_2 \) are both global maxima of \( f \) and \( f(t_1) = f(t_2) \). Using the mean value theorem, there exists \( t \in (t_1, t_2) \) such that \( f'(t) = 0 \). This is a contradiction. \( \square \)

This enables us to prove the following theorem.
Theorem 3.1. If for every site index $s$, the polynomial $\lambda_s$ has only real roots, then $\ell$ has at most one stationary point. Moreover, if such point exists, it is also a global maximum.

Proof. Since $\lambda_s$ has only real roots, it can be written as product of linear functions

$$\lambda_s(x) = \prod_{i=1}^{d_p} (\alpha_{s,i} + \beta_{s,i}x)$$

where $d_p$, defined in Assumption 2.1, is the degree of the polynomial $\lambda_s$.

The log-likelihood function $\ell$ can be computed as

$$\ell(t) = \sum_{s=1}^{S} \log(\lambda_s(e^{-\gamma t}))$$

$$= \sum_{s=1}^{S} \log \left( \prod_{i=1}^{d_p} (\alpha_{s,i} + \beta_{s,i}e^{-\gamma t}) \right)$$

$$= \sum_{s=1}^{S} \sum_{i=1}^{d_p} \log(\alpha_{s,i} + \beta_{s,i}e^{-\gamma t}).$$

For any $t, u > 0$, we have

$$\ell(t) - \ell(u) = \sum_{s=1}^{S} \sum_{i=1}^{d_p} \log \left( \frac{\alpha_{s,i} + \beta_{s,i}e^{-\gamma t}}{\alpha_{s,i} + \beta_{s,i}e^{-\gamma u}} \right)$$

$$\leq \sum_{s=1}^{S} \sum_{i=1}^{d_p} \left( \frac{\alpha_{s,i} + \beta_{s,i}e^{-\gamma t}}{\alpha_{s,i} + \beta_{s,i}e^{-\gamma u}} - 1 \right)$$

$$= \sum_{s=1}^{m} \sum_{i=1}^{d_p} \left( \frac{\beta_{s,i}(e^{-\gamma t} - e^{-\gamma u})}{\alpha_{s,i} + \beta_{s,i}e^{-\gamma u}} \right)$$

$$= \frac{1}{\gamma} \left( 1 - e^{-\gamma(t-u)} \right) \sum_{s=1}^{S} \sum_{i=1}^{d_p} \left( \frac{-\beta_{s,i} \gamma e^{-\gamma u}}{\alpha_{s,i} + \beta_{s,i}e^{-\gamma u}} \right)$$

$$= \frac{1}{\gamma} \left( 1 - e^{-\gamma(t-u)} \right) \ell'(u).$$

Hence, $\ell$ is an incave function.

Furthermore, since $\lambda_s$ are polynomial and $e^{-\gamma t}$ is a bijective map from $[0, \infty)$ to $(0,1]$, we deduce that $\ell(t)$ only has a finite number of stationary points. Using Lemma 3.1, we conclude that $\ell$ has at most one stationary point; moreover, if such a point exists, it is also a global maximum.

We note that Theorem 3.1 imposes a condition on the characteristic polynomials rather than the evolutionary model, and can be applied to assess the uniqueness of the stationary point of any time-reversible evolutionary model satisfying Assumption 2.1. In fact, Theorem 3.1 does not assume a uniform molecular clock, or even a single evolutionary model along the edges of the tree. However, it is worth noting that for the class of models on which the rate matrices have only one non-zero eigenvalue, the result automatically holds:
Corollary 3.1. For binary, Jukes-Cantor and Felsenstein 1981 models, the one-dimensional likelihood function has at most one stationary point; if such point exists, it is the global maximum.

We also note that the results in previous studies about the number of maxima of likelihood surfaces [2–4] are derived for binary models. Theorem 3.1 complements those results in the sense that while the likelihood surfaces considered in those work may have multiple (or even a continuum of) local maxima, the stationary points of one-dimensional likelihood functions are still unique.

This result also provides a full characterization of likelihood functions of binary models (and those considered by Corollary 3.1). Indeed, since log-likelihood functions are smooth, this result implies that:

1. If there is no stationary point, then \( \ell(t) \) is a monotonic function (either strictly decreasing or strictly increasing).
2. If the stationary point \( t_0 \) exists and is unique, then the function is increasing in the interval \((0, t_0)\) and is decreasing in \((t_0, \infty)\).

This simplicity of the shapes of phylogenetic likelihood functions provides a strong theoretical foundation for the use of simple optimization methods to locate the maximum likelihood branch length.

4. Algebraic structures on the space of all logarithmic relative frequency patterns under the Kimura 2-parameter model

While Section 3 provides a uniqueness result for the maximum likelihood branch lengths under three simple models, the result does not extend to more general models. In fact, as we will illustrate in the next section, the shapes of likelihood functions under the Kimura 2-parameter model [7] can be quite complicated, for example with multiple local and global maxima.

In order to enable theoretical analyses of phylogenetic likelihood functions under more complex evolutionary models, here we introduce the concept of conditional logarithmic frequency patterns and study the algebraic structures on the space of such patterns.

Definition 4.1. Given a rooted tree \( \tau_\rho \) with root \( \rho \) and \( N \) taxa, some labelings \( \psi = (\psi_1, \ldots, \psi_S) \in \Omega^{N \times S} \) of its taxa and a vector of real constants \((c_1, \ldots, c_S)\), we define the logarithmic relative frequency pattern \( \phi(\tau_\rho, \psi, c) \) as the \( 4 \times S \) matrix with entries

\[
\phi_{i,s} = c_s + \log \sum_{a \in \mathcal{L}_{i,s}} \pi(i) \prod_{(u,v) \in E(\tau_\rho)} P_{a_{uv}}^{uv}(t_{uv})
\]

for \( i \in \Omega, s = 1, \ldots, S \) and \( \mathcal{L}_{i,s} \) being the set of all extensions \( a \) of \( \psi_s \) to all the nodes of \( \tau \) such that \( a(\rho) = i \). For convenience, we will use the shorter term frequency pattern to refer to a logarithmic relative frequency pattern.

In probabilistic terms, for a fixed site index \( s \), the \((i, s)\)-entry of a logarithmic relative frequency pattern \( \phi(\tau_\rho, \psi, c) \) is (up to a constant \( c_s \)) the logarithm of the likelihood of observing state \( i \) at the root of the tree, given leaf states \( \psi_s \). This definition is directly related to the formulation of the characteristic polynomials \( \lambda_s \), whose coefficients \( b_{ij}^s \) are the product of the probabilities of observing state \( i \) and \( j \) at the two end points of an edge, given that the labeling \( \psi_s \) is observed at the taxa. Hence, to characterize the space of all phylogenetic characteristic polynomials under
a given evolutionary model, we just need to characterize the space of all possible logarithmic relative frequency patterns under that model.

**Definition 4.2.** We denote the space of all possible logarithmic relative frequency patterns under the Kimura 2-parameter model by

\[ G = \{ \phi(\tau, \psi, c) : \tau \in \mathcal{T}, \psi \in \Psi_\tau^S, c \in \mathbb{R}^S \} \]

where \( \mathcal{T} \) denotes the set of all rooted trees and \( \Psi_\tau^S \) denotes the set of all tuples \((\psi_1, \ldots, \psi_S)\) of \( S \) labelings of the taxa of \( \tau \).

The goal of this section is to establish that for any sequence of \( S \) column vectors \( v_1, v_2, \ldots, v_S \) in \( \mathbb{R}^4 \), there exists a tree \( \tau \) under the Kimura 2-parameter model, labelings \( \psi = (\psi_1, \psi_2, \ldots, \psi_S) \) of its taxa and a vector of real constants \( c \) such that

\[ \phi(\tau, \psi, c) = [v_1 \ v_2 \ \ldots \ v_S]. \]

The existence of such a tree is guaranteed indirectly by proving that under the Kimura 2-parameter model:

1. \( G \) is an algebraic subgroup of \((\mathbb{R}^{4 \times S}, +)\).
2. \( G \) is path-connected.
3. \( G \) is a linear subspace of \( \mathbb{R}^{4 \times S} \).

The first two steps are confirmed by the following theorem.

**Theorem 4.1.** If the stationary frequency of the evolutionary model is the same for every state, then the following properties hold:

1. \((G, +)\) is a subgroup of \((\mathbb{R}^{4 \times S}, +)\).
2. \( G \) is path-connected.
3. \( G \) is a linear subspace of \( \mathbb{R}^{4 \times S} \).

A detailed proof of this Theorem is provided in the Appendix, but the main arguments can be simply illustrated. The fact that \( G \) is closed under addition follows because we can add two frequency patterns just by gluing the roots of the two corresponding trees, labeling the taxa of \( \tau \) correspondingly and taking the pattern at the new root. Similarly, we can create the inverse of a pattern by gluing all permuted versions of its corresponding tree (with an appropriate vector of real constants).

To prove that \( G \) is path-connected, given two arbitrary trees with roots \( \rho_1, \rho_2 \), we create a new tree by adding a new root \( \rho \), joining \( \rho_1, \rho_2 \) with \( \rho \) by two new edges of length \( t \) and \( 1/t \), respectively, and making \( \rho \) the root of \( \tau \). By varying \( t \) continuously from zero to infinity, we can make a continuous path in \( G \) that connects the two frequency patterns. Since any path-connected subgroup of \( \mathbb{R}^n \) is a linear subspace \([11]\), so is \( G \).

**Remark 4.1.** Since Theorem 4.1 only requires that the stationary frequency of the evolutionary model is the same for every state, this result also extends to models with more parameters.

We then establish that \( G = \mathbb{R}^{4 \times S} \) through proving by induction that \( G \) contains \( 4 \times S \) independent frequency patterns (also proven in the Appendix):

**Theorem 4.2.** The set of all possible logarithmic conditional frequency patterns with \( S \) sites under the Kimura 2-parameter model with \( \kappa = 3 \) is equal to \( \mathbb{R}^{4 \times S} \).

With those results, we finally can establish the main theorems of the section.
Theorem 4.3. For any sequence of column vectors \(v_1, v_2, \ldots, v_S\) in \(\mathbb{R}^4\), there exists a rooted tree \(\tau\) under the Kimura 2-parameter model with \(\kappa = 3\), \(S\) labelings \(\psi_1, \psi_2, \ldots, \psi_S\) of its taxa, and a vector of real constants \(c\) such that
\[
\phi(\tau, \psi, c) = [v_1 \ v_2 \ldots \ v_S].
\]

While Theorem 4.3 provides a theoretical guarantee about the existence of a tree under Kimura model with a given frequency patterns, the proof is not constructive. This raises some concerns about the practicality of the approach. For example, one can not derive an estimation of the number of edges required to produce a given frequency pattern. Those concerns are addressed by the following theorem.

Theorem 4.4. A tree as in Theorem 4.3 can be constructed with at most \(64S\) edges.

Not only does the theorem provide an upper bound on the number of edges required to construct a tree with a given frequency pattern, its proof also provides a simple algorithm to construct such a tree.

Proof of Theorem 4.4. The main steps of the proof are as follows:

Step 1. As shown in the Appendix, any frequency pattern of the form \([x, 0, 0, 0]^t\) can be produced (up to a real constant \(c_1\)) by a tree \(\tau\) with 4 edges and some labeling \(\psi\) of its taxa.

Step 2. Using \(\tau\) from Step 1, we create a tree \(\tau'\) of 16 edges by gluing the roots of 4 different versions \(\tau_1, \tau_2, \tau_3, \tau_4\) of \(\tau\) together and define \(S\) labelings of \(\tau'\) as follows.

- For \(s = 1\), we copy the labeling of \(\tau\) onto \(\tau'\).
  \[
  \psi_1(a) = \psi(a)
  \]
  for each taxon \(a\) of \(\tau_1, \tau_2, \tau_3, \tau_4\).

- For all \(s \geq 2\), the labelings are defined as follows:
  \[
  \psi_s(a) = \sigma^s(\psi(a)) \quad \text{if } a \text{ is a taxon of } \tau_j
  \]
  where \(\sigma\) is the permutation \((A\ G\ T\ C)\) in cycle notation.

The construction of \(\tau'\) is similar to the construction of the inverse of elements in the group \(G\) in the proof of Theorem 4.1. Because of symmetry, for \(s \geq 2\), the frequency pattern corresponding to site \(s\) at the root of the newly created tree will be the same for every state while for \(s = 1\), the frequency pattern of \(\tau'\) is obtained by multiplying the frequency pattern of \(\tau\) by a factor of 4.

We deduce that the pattern created by \((\tau', \{\psi_1\})\) is:
\[
\begin{pmatrix}
4x & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
\end{pmatrix}
+ \begin{pmatrix}
c_1 & c_2 & \ldots & c_S \\
c_1 & c_2 & \ldots & c_S \\
c_1 & c_2 & \ldots & c_S \\
c_1 & c_2 & \ldots & c_S \\
\end{pmatrix}
\]
for some real constants \(c_1, c_2, \ldots, c_S\).

Step 3. By similar arguments, for any \(i = 1, 2, 3, 4\) and \(s = 1, 2, \ldots, S\), we can construct a tree of 16 edges for any patterns with \(S\) sites whose only non-zero entry is at the \((i, s)\)-position. Hence, it takes \(16 \times 4S = 64S\) edges to construct a tree with an arbitrary given frequency pattern.
5. Non-uniqueness of stationary points: Kimura 2-parameter model

In this section, we provide an example for which there are multiple stationary points of the likelihood function. To construct such an example, we find two polynomials \( p_1(x) \) and \( p_2(x) \) with coefficients \( b_1, b_2 \) such that the product \( p_1 p_2 \) has 2 local maxima in [0, 1], and \( p_1 \) and \( p_2 \) can be expressed as positive linear combination of the basis polynomial functions \( P_i \) derived from an evolutionary model.

Consider the Kimura 2-parameter model with \( \kappa = 3 \) which has the rate matrix

\[
Q = \begin{pmatrix}
-5/8 & 3/8 & 1/8 & 1/8 \\
3/8 & -5/8 & 1/8 & 1/8 \\
1/8 & 1/8 & -5/8 & 3/8 \\
1/8 & 1/8 & 3/8 & -5/8 \\
\end{pmatrix}.
\]

This matrix has eigenvalues \( 0 > -\gamma > -2\gamma \) where \( \gamma = 0.5 \). The transition probabilities under this evolutionary model can be computed explicitly by

\[
P_1(t) = 0.25 + 0.25 \exp(-0.5t) + 0.5 \exp(-t)
\]
\[
P_2(t) = 0.25 + 0.25 \exp(-0.5t) - 0.5 \exp(-t)
\]
\[
P_3(t) = P_4(t) = 0.25 - 0.25 \exp(-0.5t)
\]

where \( P_1(t), P_2(t), P_3(t), P_4(t) \) are the probabilities of transitioning from state \( A \) to state \( A, T, G, C \), respectively. This simple model is “universal” in an appropriate sense as shown in the end of the paper.

This leads to a representation of the likelihood as the product of two different linear combinations of the transition polynomials

\[
P_1(x) = 0.25 + 0.25x + 0.5x^2
\]
\[
P_2(x) = 0.25 + 0.25x - 0.5x^2
\]
\[
P_3(x) = P_4(x) = 0.25 - 0.25x
\]

where \( x = \exp(-0.5t) \).

We assume that the likelihood is computed by observing two sites \( s_1 \) and \( s_2 \), and that the edge of interest \( e \) is a pendant edge with the observed values at that tip being \( A \) for both sites. Assume further that the state observation probabilities at the inner node of the edge \( e \) are provided by

\[
b_1 = [0.24977275, 0.34067358, 0.2051904, 0.20436327]
\]
and

\[
b_2 = [0.25, 0.16087344, 0.29328435, 0.29584221].
\]

As discussed earlier, the log-likelihood function can be computed as

\[
\ell(t) = \log(\lambda_1(t)) + \log(\lambda_2(t))
\]

where

\[
\lambda_i(t) = \sum_{i \in \Omega} b^*(i) P_i(t).
\]

The graph of the log-likelihood function \( \ell \) and its perturbations (by varying the coefficients slightly) in terms of \( x \) and \( t \) are provided in Figure 3 and Figure 4 respectively. The figures show that \( \ell \) has three stationary points (two local maxima at \( t_1 < t_2 \) and one local minimum), all in the interval \([0, 1]\). The fact that \( \ell(t_1) > \ell(t_2) \) for some cases and \( \ell(t_1) < \ell(t_2) \) for some others indicates that there exist some...
values of $b_i^s$ such that $\ell(t_1) = \ell(t_2)$, i.e. the smoothly varying likelihood function can even have two global maxima.

We note that these examples can be achieved under the assumption that given any positive coefficients $b_i^s$ of the inner node, we can find some trees under Kimura 2-parameter model with these precise coefficients. This assumption is confirmed by the following result, proven in the Appendix.

**Theorem 5.1.** For every set of positive coefficients $\eta_i^s$, there exist a phylogenetic tree $\tau$ and $S$ labelings $\psi_1, \psi_2, \ldots, \psi_S$ of the taxa such that for some edge $e$ in $\tau$, the one-dimensional likelihood function on $e$ under the Kimura 2-parameter model with $\kappa = 3$ satisfies

$$\ell(\tau, t) = C_0 + \sum_{s=1}^{S} \log \left( \sum_{i} \eta_i^s P_i(t) \right)$$

where $P_i(t)$ is the probability of transition from state $A$ to state $i$ and $C_0$ is a constant. Moreover, such a tree $\tau$ can be constructed with at most $64S + 1$ edges.

In our examples, the upper bound on the number of edges to produce the given frequency pattern is $64 \times 2 + 1 = 129$ edges.

**Remark 5.1.** While the algorithm to construct a tree given a frequency pattern given by Theorem 4.4 always outputs a star-tree (a tree without internal edges), we note that

1. We can approximate any star tree by resolved trees with arbitrary precision.
2. The maximum number of stationary points of a polynomial of degree four is 3, hence small perturbations on the coefficient of a polynomial of degree

![Likelihood functions with multiple stationary points](image)

**Figure 1.** The log-likelihood (5.3) as a function of $x = \exp(-0.5t)$ for various values of the coefficients of the characteristic polynomial.
Figure 2. The log-likelihood of (5.3) as a function of branch length $t$ for various values of the coefficients of the characteristic polynomial.

four with three stationary points do not change the number of stationary points.

We deduce that there are resolved trees for which the one-dimensional likelihood function on certain edges have multiple maxima.

Since a resolved tree with $n$ taxa has $2n - 3$ edges, the upper bound on the number of edges of a resolved tree for which the one-dimensional likelihood function on certain edges has multiple maxima is $2 \times 129 - 3 = 255$ edges.

6. Universality and complexity of the Kimura 2-parameter model

As we discussed earlier in the paper, the main idea behind the results in Section 3 and Section 4 is that by using the Fundamental Theorem of Algebra, we can decompose a complicated evolutionary model into smaller modules, each of which can be approximated either by a “linear” model or by a “quadratic” model. This paradigm focuses on the branch lengths of the tree and is independent of the state space $\Omega$ of the evolutionary model, which provides a way to represent advanced evolutionary models (amino-acid models, codon models) by simple ones (nucleotide models).

This motivates the problem of constructing a complete characterization of one-dimensional likelihood functions. The main question is: does there exist an evolutionary model that can represent all one-dimensional likelihood functions of any time-reversible evolutionary model?

Such a model $M$, if it exists, and which we will refer to as a universal model, needs to satisfy the following two conditions:
1. All one-dimensional likelihood functions under any reversible evolutionary model can be written as a product of polynomials, each of which is a positive linear combination of the transition polynomials of $\mathcal{M}$.

2. For every set of positive coefficients $b_{ij}$, there exists a phylogenetic tree $\tau$ and $S$ labelings $\psi_1, \psi_2, \ldots, \psi_S$ of the taxa such that for some edge $e$ in $\tau$, the one-dimensional likelihood function on $e$ under the $\mathcal{M}$ satisfies

$$\ell(\tau, t) = C_1 + \sum_{s=1}^{S} \log \left( \sum_{ij} b_{ij}^s P_{ij}(t) \right)$$

for some constant $C_1$.

In this section, we will prove that the Kimura 2-parameter model with $\kappa = 3$ is, in fact, a universal model. The key components of the proof are Theorem 5.1, the Fundamental Theorem of Algebra and the fact that the transition polynomials of the Kimura 2-parameter model effectively span a large class of linear and quadratic polynomials.

### 6.1. Universality of the Kimura 2-parameter model.

We first make the following observation, proven in the Appendix.

**Lemma 6.1.** If $f$ is a real-coefficient polynomial that satisfies

1. $f$ is positive on $[0, 1]$,
2. $\deg f = 1$ or $f$ is a quadratic polynomial with no real root,

then $f$ can be written as positive linear combination of the transition polynomials of the Kimura 2-parameter model if and only if $f$ has no root inside the set

$$(6.1) \quad B = \{ z \in \mathbb{C} : |z + 1| \leq 1 \text{ or } |z - 1| \leq \sqrt{2} \}.$$

This enables us to establish the universality of the Kimura 2-parameter model.

**Theorem 6.1** (Universality). If $L$ is a one-dimensional phylogenetic likelihood function of a tree under an arbitrary time-reversible model that satisfies Assumption 2.1, then up to translation and rescaling, $L$ is equal to a one-dimensional likelihood under the Kimura 2-parameter model.

That is, there exist $c_1, c_2, c_3 > 0$ such that

$$L(t) = c_2 L_{K2P}(\tau, \psi, c_3 t) - c_1, \quad \forall t \in [0, \infty),$$

where $L_{K2P}(\tau, \psi, \cdot)$ is the one-dimensional likelihood function under the Kimura 2-parameter model on some edge of a tree $\tau$ with labeling $\psi$.

**Proof.** Assumption 2.1 implies that the function

$$L(x) := L \left( -\frac{1}{\gamma} \log x \right)$$

is a polynomial in $x$ for some $\gamma > 0$. Since $L$ is continuous and the set $B$ defined by (6.1) is compact, if we define

$$c_1 = 1 + \sup_{z \in B} |L(z)|,$$

then by the triangle inequality, the polynomial $L(x) + c_1$ has no root in $B$. 
By the Fundamental Theorem of Algebra, the polynomial \( L(x) + c_1 \) can be written as
\[
L(x) + c_1 = \prod_{s=1}^{S} g_s(x),
\]
where each \( g_s \) is either a quadratic polynomial with no real root, or a polynomial of degree 1. Moreover, each \( g_s \) is positive on \([0, 1]\) and has no root in \( B \). Lemma 6.1 implies that each \( g_s \) can be written as a positive linear combination of the transition polynomials of the Kimura 2-parameter model
\[
g_s(x) = \sum_{ij} b_{ij}^s P_{ij}(x).
\]
We deduce that
\[
\log(L(x) + c_1) = \sum_{s=1}^{S} \log \left( \sum_{ij} b_{ij}^s P_{ij}(x) \right).
\]
We recall that the Kimura 2-parameter model has symmetries such that any transition probability \( P_{ij}(t) \) is in fact equal to \( P_{ik}(t) = P_k(t) \) for some \( k \). Therefore, by grouping
\[
\eta_t^s = \sum_{P_{ij} = P_{il}} b_{ij}^s,
\]
we have
\[
\log(L(x) + c_1) = \sum_{s=1}^{S} \log \left( \sum_l \eta_t^s P_l(x) \right).
\]
Also, the characteristic polynomial for the Kimura 2-parameter model \( 5.1 \) with \( \kappa = 3 \) is parameterized by \( x = \exp(-0.5t) \) such that the one-dimensional likelihood \( L_{K2P}(\tau, \psi, t) \) satisfies
\[
L_{K2P}(\tau, \psi, t) = L_{K2P}(\tau, \psi, \exp(-0.5t)).
\]
Now, Theorem 5.1 guarantees that there exists a tuple \( (\tau, \psi) \) under the Kimura 2-parameter model on an edge of the tree such that
\[
\log L_{K2P}(\tau, \psi, x) = -\log c_2 + \sum_{s=1}^{S} \log \left( \sum_l \eta_t^s P_l(x) \right)
\]
for some positive constant \( c_2 \).
In other words, we have
\[
L(x) = c_2 L_{K2P}(\tau, \psi, x) - c_1, \quad \forall x \in (0, 1].
\]
Hence,
\[
L \left( -\frac{1}{\gamma} \log x \right) = c_2 L_{K2P}(\tau, \psi, -2 \log x) - c_1, \quad \forall x \in (0, 1],
\]
or
\[
L(t) = c_2 L_{K2P}(\tau, \psi, c_3 t) - c_1, \quad c_3 = \gamma/2, \quad \forall t \in [0, \infty).
\]
That is, up to translation and rescaling, \( L \) is equal to a one-dimensional phylogenetic likelihood function under the Kimura 2-parameter model. \( \square \)
Since the set of rate matrices for a given evolutionary model that satisfy Assumption 2.1 is dense in the set of all possible rate matrices under the same evolutionary model (Remark 2.1), we also have the following corollary.

**Corollary 6.1.** Any one-dimensional phylogenetic likelihood function under an arbitrary time-reversible evolutionary model can be uniformly approximated with arbitrary precision by (rescaled and translated) one-dimensional phylogenetic likelihood functions under the Kimura 2-parameter model.

We also note that the rescaling and translation constants in the statements of Theorem 6.1 can not be removed: Lemma 6.1 indicates that any polynomial function with roots in $B$ can not be represented exactly as a Kimura 2-parameter likelihood function. For example, one of the transition polynomials of the Jukes-Cantor model is

$$J(x) = 0.25 + 0.75x$$

which has a root in $B$. For this reason, some likelihood functions under the Jukes-Cantor model may not be represented exactly by the Kimura 2-parameter model without adding an additive constant.

### 6.2. Complexity of the Kimura 2-parameter model.

The universality results in the previous section can be adapted easily to analyze the set of all one-dimensional phylogenetic likelihood functions under the Kimura 2-parameter model. The following complexity results imply that one-dimensional likelihood functions under advanced evolutionary models can be more complex than it is typically assumed by phylogenetic inference algorithms.

First, it is straightforward to check that Theorem 6.1 still holds (without changing the proof) if we replace the one-dimensional phylogenetic likelihood function $L$ with an arbitrary polynomial $P$ in $x = \exp(-\gamma t)$ for some $\gamma > 0$ and relax Assumption 2.1. Moreover, if $P$ is of degree $n$, then by Theorem 5.1, it can be represented by a one-dimensional likelihood function of a tree with at most $(64n + 1)$ edges with respect to some $n$-site labeling of its taxa.

**Corollary 6.2.** Given an arbitrary polynomial $P$ of degree $n$ and $\gamma > 0$, then up to translation and rescaling, $P(\exp(-\gamma t))$ is equal to a one-dimensional likelihood under the Kimura 2-parameter model on a phylogeny with at most $64n + 1$ edges.

This corollary indicates that by increasing the number of sites and the size of the tree, we can obtain likelihood functions shaped like an arbitrary polynomial in the interval $[0, 1]$. For example, given an arbitrary finite sequence $t_1, t_2, \ldots, t_k \in (0, \infty)$, we can construct a polynomial $P_k$ that peaks precisely at $x_k = \exp(-0.5t_k)$ and use Corollary 6.2 to obtain the following result.

**Corollary 6.3.** Given an arbitrary finite sequence $t_1, t_2, \ldots, t_k \in (0, \infty)$, there exists a phylogenetic tree $\tau$ and some labeling of its taxa such that for some edge of the tree, the one-dimensional likelihood function under the Kimura 2-parameter model peaks precisely at $t_1, t_2, \ldots, t_k$.

Furthermore, since rescaling and translation do not change the relative order of the likelihood values at the stationary points, we can make any of the $t_i$’s (or all of them) the function’s global maxima.

Finally, we can replace the phylogenetic likelihood functions in Corollary 6.1 by an arbitrary continuous function $f$ with finite limit to obtain the following density result.
Corollary 6.4. The space of all rescaled and translated one-dimensional phylogenetic likelihood functions under the Kimura 2-parameter model is dense in the space of all continuous functions on $[0, \infty)$ with finite limits.

Proof. Let $f$ be a continuous function on $[0, \infty)$ with finite limit. Define

$$g(x) = f(-\log(x)) \quad \forall x \in (0, 1],$$

then $g(x)$ can be extended continuously to $[0, 1]$. By Weierstrass’s theorem, there exists a sequence of polynomials $\{P_n\}$ such that

$$\sup_{x \in [0, 1]} |P_n(x) - g(x)| \to 0.$$ 

This implies that

$$\sup_{t \in [0, \infty)} |P_n(exp(-t)) - f(t)| \to 0.$$ 

On the other hand, we deduce from Corollary 6.1 that $P_n(exp(-t))$ is, up to rescaling and translation, a one-dimensional likelihood under the Kimura 2-parameter model. This completes the proof.

\[\square\]

7. Conclusions and discussion

In this work, we investigate the problem of characterizing the shape of one-dimensional phylogenetic likelihood functions. Our results classify all evolutionary models into two categories:

1. For binary, Jukes-Cantor and Felsenstein 1981 models: the one-dimensional likelihood function has at most one stationary point.

2. For Kimura 2-parameter model and more advanced evolutionary models: the shape of the one-dimensional likelihood function can be much more complex. In fact, the space of all rescaled and translated one-dimensional phylogenetic likelihood functions under such a model is dense in the space of all continuous functions on $[0, \infty)$ with finite limits.

Despite the complexity of the one-dimensional likelihood functions under advanced evolutionary models, we prove that all one-dimensional phylogenetic likelihood function are essentially Kimura 2-parameter likelihood functions. This result establishes a strong foundation for the use of the Kimura 2-parameter as the building block of all evolutionary models.

Our results are based on two novel techniques. First, we introduce and use characteristic polynomial representations of one-dimensional phylogenetic likelihood functions and the Fundamental Theorem of Algebra to decompose any evolutionary models into smaller modules, each of which resembles the Kimura 2-parameter model. Second, we introduce the new concept of logarithmic relative frequency patterns and analyze algebraic structures on the space of such patterns. These structures open a new way to explore the space of all possible likelihood functions. Moreover, by analyzing these structures, we are able to tackle the inverse problem of constructing a phylogenetic tree that has a given frequency pattern at the root. This enables us to construct phylogenetic trees that approximate any given likelihood function with arbitrary precision.

There are several avenues for improvements. Firstly, while we know that the shape of one-dimensional likelihood function can be very complex, it is not clear how frequently multimodality might be encountered in practice and to which degree it affects the accuracy of phylogenetic algorithms. Since the space of high
degree polynomials are dominated by multimodal functions, one might expect that as the number of sites and the size of the tree increase, multimodality becomes more likely. However, since the space of phylogenies is known to possess considerable hidden structure which sometimes lead to counter-intuitive properties, careful analysis of the space of all rescaled and translated one-dimensional phylogenetic likelihood functions under the Kimura 2-parameter model are required to evaluate this hypothesis. Secondly, although the focus of this work is on one-dimensional phylogenetic likelihood functions, it is possible to utilize the framework we propose to study full phylogenetic likelihood functions. This will be a subject for future work.

Acknowledgments. This research was supported by National Science Foundation award DMS-1223057.

REFERENCES

[1] D. Bryant, N. Galtier, and M.-A. Poursat, “Likelihood calculation in molecular phylogenetics,” Mathematics of Evolution and Phylogeny, pp. 33–62, 2005.
[2] M. Steel, “The maximum likelihood point for a phylogenetic tree is not unique,” Systematic Biology, pp. 560–564, 1994.
[3] B. Chor, M. D. Hendy, B. R. Holland, and D. Penny, “Multiple maxima of likelihood in phylogenetic trees: an analytic approach,” Molecular Biology and Evolution, vol. 17, no. 10, pp. 1529–1541, 2000.
[4] J. S. Rogers and D. L. Swofford, “Multiple local maxima for likelihoods of phylogenetic trees: a simulation study,” Molecular biology and evolution, vol. 16, no. 8, pp. 1079–1085, 1999.
[5] T. H. Jukes and C. R. Cantor, “Evolution of protein molecules,” Mammalian protein metabolism, vol. 3, pp. 21–132, 1969.
[6] J. Felsenstein, “Evolutionary trees from DNA sequences: a maximum likelihood approach,” Journal of molecular evolution, vol. 17, no. 6, pp. 368–376, 1981.
[7] M. Kimura, “A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences,” Journal of molecular evolution, vol. 16, no. 2, pp. 111–120, 1980.
[8] J. Felsenstein, “Inferring phylogenies,” 2004.
[9] M. A. Hanson, “On sufficiency of the Kuhn-Tucker conditions,” Journal of Mathematical Analysis and Applications, vol. 80, no. 2, pp. 545–550, 1981.
[10] A. Ben-Israel and B. Mond, “What is invexity?,” The Journal of the Australian Mathematical Society, Series B. Applied Mathematics, vol. 28, no. 01, pp. 1–9, 1986.
[11] T. Hayashida, “Arc-wise connected subgroup of a vector group,” in Kodai Mathematical Seminar Reports, vol. 1, pp. 16–16, 1949.

8. Appendix

Proof of Remark 2.1 If we denote the entries of the diagonalizing matrix $M$ and $N$ of $Q$ by $m_{ij}$ and $n_{ij}$, respectively, then

$$Q_{ij} = \sum_k m_{ik} e^{-r_k} n_{kj}.$$ 

where $-r_k$ are the eigenvalues of $Q$. (The eigenvalues are known to be non-positive, so $r_k$ are non-negative.)

Since the set of rational numbers $\mathbb{Q}$ is dense in $\mathbb{R}^+$, we can find $r(k, l) \in \mathbb{Q}^+$ such that for all $k$, $r(k, l) \to r_k$ as $l$ approaches infinity. If we define

$$Q^{l}_{ij} = \sum_k m_{ik} e^{-r(k,l)} n_{kj},$$

Appendix

Proof of Remark 2.1 If we denote the entries of the diagonalizing matrix $M$ and $N$ of $Q$ by $m_{ij}$ and $n_{ij}$, respectively, then

$$Q_{ij} = \sum_k m_{ik} e^{-r_k} n_{kj}.$$ 

where $-r_k$ are the eigenvalues of $Q$. (The eigenvalues are known to be non-positive, so $r_k$ are non-negative.)

Since the set of rational numbers $\mathbb{Q}$ is dense in $\mathbb{R}^+$, we can find $r(k, l) \in \mathbb{Q}^+$ such that for all $k$, $r(k, l) \to r_k$ as $l$ approaches infinity. If we define

$$Q^{l}_{ij} = \sum_k m_{ik} e^{-(r(k,l))} n_{kj},$$
then \( Q^l \to Q \) element-wise as \( l \) approaches infinity. Since \( r(k, l) \) are all rational the matrices are of fixed finite dimension, we can also find \( \gamma_l > 0 \) and \( d(k, l) \in \mathbb{N} \) such that \( r(k, l) = d(k, l) \gamma_l \).

**Proof of Theorem 4.1** We define the equivalence relation \( \sim \) on \( \mathbb{R}^{4 \times S} \) as follows: \( u \sim v \) if and only if there exists a vector of real constants \( c = (c_1, \ldots, c_S) \) such that for all \( i = 1, 2, 3, 4 \) and \( s = 1, \ldots, S \), we have

\[
 u_{i,s} = v_{i,s} + c_s.
\]

If we define

\[
 [h(\tau, \psi)]_{i,s} = \log \sum_{a \in L_{i,s}} \prod_{(u,v) \in E(\tau)} P_{a_{u}a_{v}}^{uv}(t_{uv})
\]

for \( i \in \Omega, s = 1, \ldots, S \) and \( L_{i,s} \) being the set of all extensions \( a \) of \( \psi_s \) to all the nodes of \( \tau \) such that \( a(\rho) = i \).

Then for all \( \tau, \psi, c \), we have \( \phi(\tau, \psi, c) \sim h(\tau, \psi) \).

1. (Addition): Consider any two elements \( x_1, x_2 \in G \). By the definition of \( G \) and since the stationary frequency of the evolutionary model is the same for every state, there exist trees \( \tau_1, \tau_2 \) with \( n_1, n_2 \) taxa and labelings \( \psi_1, \psi_2 \) such that

\[
x_i \sim \phi(\tau_i, \psi_i), \quad i = 1, 2.
\]

If we construct a new tree \( \tau \) from \( \tau_1 \) and \( \tau_2 \) by gluing the roots \( \rho_1, \rho_2 \) and label the taxa of \( \tau \) corresponding to \( \psi_1, \psi_2 \), then we have

\[
 [h(\tau, \psi)]_{i,s} = \log \sum_{a \in L^i_s} \prod_{(u,v) \in E(\tau_1)} P_{a_{u}a_{v}}^{uv}(t_{uv}) \prod_{(u,v) \in E(\tau_2)} P_{a_{u}a_{v}}^{uv}(t_{uv})
\]

where each term \( a \) in the sum corresponds uniquely to a pair of extensions \( (a^1, a^2) \) of \( \psi_1^i, \psi_2^i \) to the internal nodes of \( \tau_1, \tau_2 \), respectively, such that \( a^1(\rho) = a^2(\rho) = i \).

Therefore,

\[
 [h(\tau, \psi)]_{i,s} = \log \sum_{a^1} \prod_{(u,v) \in E(\tau_1)} P_{a_{u}a_{v}}^{uv}(t_{uv}) + \log \sum_{a^2} \prod_{(u,v) \in E(\tau_2)} P_{a_{u}a_{v}}^{uv}(t_{uv})
\]

for all \( i \in \Omega \) and \( s = 1, \ldots, S \).

Therefore

\[
 \phi(\tau, \psi) \sim \phi(\tau_1, \psi_1) + \phi(\tau_2, \psi_2)
\]

and

\[
 x_1 + x_2 \sim \phi(\tau, \psi) \in G
\]

which implies that \( G \) is closed under addition.

2. (Inverse): Consider any element \( x \in G \) and its corresponding representative tree \( \tau \) and labeling \( \psi \). For any permutation \( \sigma \) of the states, we define the labeling \( \psi_{\sigma} \) as

\[
 \psi_{\sigma}(\omega) = \sigma(\psi(\omega))
\]

for every taxon \( \omega \) of \( T \). For example, if \( \sigma \) is the permutation \( (A \ G \ T \ C) \) in cycle notation, then \( \psi_{\sigma} \) is obtained from \( \psi \) by replacing \( A \) by \( G \), \( G \) by \( T \), \( T \) by \( C \) and \( C \) by \( A \).

Now let \( \sigma_0 \) be a permutation of order \( r \) on the state space \( \Omega \), create \( r \) identical copies \( \tau_1, \tau_2, \ldots, \tau_r \) of the tree \( \tau \) with labelings \( \psi_{\sigma_0}, \psi_{\sigma_0}^2, \ldots, \psi_{\sigma_0}^r \).
\( \psi_{\gamma} \) and glue the root of all the trees together with taxon labeling \( \gamma \) corresponding to the labelings of \( \tau_1, \tau_2, \ldots, \tau_r \). Then because of symmetry, the frequency pattern \( f \) at the root of the newly created tree \( \mu \) will be the same for every state, i.e., \( f \sim 0 \). We deduce that \( 0 \in G \) and for every \( x \in G \), there exists \( y \in G \) such that \( x + y = 0 \).

This property and the fact that \( G \) is closed under addition prove that \((G, +)\) is a subgroup of \((\mathbb{R}^{n \times S}, +)\).

3. (Connectedness): Consider any two elements \( x_1, x_2 \in G \) and their corresponding trees \( \tau_1, \tau_2 \), labelings \( \psi_1, \psi_2 \) and vectors of real constants \( c_1, c_2 \). For any \( \alpha \in (0, 1) \), we create a new tree \( \tau(\alpha) \) by adding a new root \( \rho \), joining \( \rho_1, \rho_2 \) by new edges of length \( t_1 = \tan(\frac{\pi}{2} \alpha), t_2 = 1/t_1 \), respectively. We make \( \rho \) the root of \( \tau \) and label the taxa of \( \tau \) according to \( \psi_1, \psi_2 \).

Now we note that when \( \alpha \to 0 \), we have
\[
h(\tau(\alpha), \psi) \to h(\tau_1, \psi_1) + \log \frac{1}{r}
\]
since the contribution of \( \tau_2 \) becomes stationary (the stationary frequency is \( 1/r \) because of the model’s symmetry). Similarly, when \( \alpha \to 1 \), we have
\[
h(\tau(\alpha), \psi) \to h(\tau_2, \psi_2) + \log \frac{1}{r}.
\]
Therefore, the function \( g(\alpha) = \phi(\tau(\alpha), \psi) \) can be extended continuously to the closed interval \([0, 1]\). By changing \( c \) continuously from \( c_1 \) to \( \log(1/r) \), varying \( \alpha \) continuously from \( 0 \) to \( 1 \), then changing \( c \) continuously from \( \log(1/r) \) to \( c_2 \), we can make a path in \( G \) that connects \( x_1 \) and \( x_2 \).

4. Since any path-connected subgroup of \( \mathbb{R}^n \) is a linear subspace \([\Pi]\), so is \( G \).

\[\square\]

**Proof of Theorem 4.2**

Denote by \( H \) the set of all rooted trees with one edge (that is, any tree in \( H \) has only one root and one taxon) and
\[
H = \{ \phi(\tau, \psi, c) : \tau \in \mathcal{H}, \psi = (\psi_1, \psi_2, \ldots, \psi_S) \in \mathbb{R}^S, c \in \mathbb{R}^S \}.
\]
We have
\[
(8.1) \quad [h(\tau, \psi)]_{j,s} = \log P_{\psi,j}(x)
\]
where \( x = \exp(-0.5t) \), \( j = A, G, T, C \), and \( t \) is the length of the unique edge of \( \tau \).

Let \( x_1 = 1/4, x_2 = 1/2, x_3 = 3/4 \). We will prove, by induction on \( S \), that \( H \) contains \( 4 \times S \) independent frequency patterns.

For \( S = 1 \), by considering the 4 patterns \((A), (G), (T), (C)\) and the values of \( x \) described above, we can create a set of 12 different pairs \((\tau, \psi)\). A quick check by computer shows that the corresponding frequency patterns generated by those pairs span the whole vector space \( \mathbb{R}^{4 \times 1} \). We can achieve similar result for \( S = 2 \) with the patterns \((A, G), (G, T), (T, C), (C, A)\).

Now assume that for \( S = n \), \( H \) contains \( 4 \times n \) independent frequency patterns of the form \((8.1)\).

For \( l = A, T, G, C \) and \( x \in [0, 1] \), we define the building blocks
\[
R_l(x) := \begin{bmatrix} \log P_{A}(x) & \log P_{T}(x) & \log P_{G}(x) & \log P_{C}(x) \end{bmatrix}
\]
\[
W_l(x) := \begin{pmatrix} R_l(x_1) \\ R_l(x_2) \\ R_l(x_3) \end{pmatrix}
\]
The induction hypothesis implies that there exist $4 \times n$ independent frequency patterns of the form $[8.1]$. This means that for some labelings $\psi_1, \psi_2, \ldots, \psi_{4n}$, the block matrix

$$J = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_{4n} \end{pmatrix}$$

has maximal rank $4n$, where

$$B_s := \begin{pmatrix} R_{\psi_1}(x_1) & \cdots & R_{\psi_n}(x_1) \\ R_{\psi_1}(x_2) & \cdots & R_{\psi_n}(x_2) \\ R_{\psi_1}(x_3) & \cdots & R_{\psi_n}(x_3) \end{pmatrix}.$$  

For $s = 1, \ldots, 4n$, we consider all the labelings obtained by appending $\psi_s$ with one of the four nucleotides $A, T, G, C$. By doing so, we create a set of $48n$ different frequency patterns. We want to prove that the block matrix

$$C = \begin{pmatrix} B_1 & W_A(x) \\ B_2 & W_A(x) \\ \vdots \\ B_{4n} & W_A(x) \\ B_1 & W_G(x) \\ B_2 & W_G(x) \\ \vdots \\ B_{4n} & W_G(x) \\ B_1 & W_T(x) \\ B_2 & W_T(x) \\ \vdots \\ B_{4n} & W_T(x) \\ B_1 & W_C(x) \\ B_2 & W_C(x) \\ \vdots \\ B_{4n} & W_C(x) \end{pmatrix}$$

has maximal rank $4n + 4$.

Note that this matrix is row-equivalent to

$$\begin{pmatrix} J \\ 0 \end{pmatrix} U \begin{pmatrix} V \end{pmatrix}$$

where each row of $V$ is of the form $R_i(x_k) - R_A(x_k)$ for $i = G, T, C$. (This is done by subtracting the blocks $(B_s, R_i(x))$ by the block $(B_s, R_A(x))$ then rearranging the row to obtain the sub-matrix $J$ at the top-left corner.)

On the other hand, from the case $S = 1$, we have

$$\text{rank} \begin{pmatrix} W_A(x) \\ W_T(x) \\ W_G(x) \\ W_C(x) \end{pmatrix} = 4,$$

which implies that $\text{rank}(V) = 4$. Hence, $\text{rank}(C) = \text{rank}(J) + \text{rank}(V) = 4n + 4$.

We deduce that for every $S$, the set $G$ of all possible logarithmic conditional frequency patterns with $S$ sites under the Kimura 2-parameter model is a linear
subspace of $\mathbb{R}^{4 \times S}$ (Theorem 4.1) that contains 4 linearly independent vectors. This implies that $G = \mathbb{R}^{4 \times S}$.

\begin{proof}[Proof of Theorem 4.4 (Step 1)] (Any pattern of the form $v = [x \ 0 \ 0 \ 0]$ can be produced by a tree $\tau$ with four edges.)

Denote
\begin{align*}
x_1(t) &= P_{AA}(t) & x_2(t) &= P_{AG}(t) \\
x_3(t) &= P_{AT}(t) & x_4(t) &= P_{AC}(t)
\end{align*}

we note that in the Kimura 2-parameter model, $x_3(t) = x_4(t)$.

Now consider two trees $\tau_1$ and $\tau_2$, each with one edge, whose branch lengths are $t$ and $s$, respectively. We label the only nodes of $\tau_1$ and $\tau_2$ by the patterns $\psi_1 = (A)$ and $\psi_2 = (T)$, and obtain the frequency patterns $f_1(t)$ and $f_2(s)$ respectively. By gluing the roots of $\tau_1$ (1 edge) and the “inverse” of the tree $\tau_2$ (3 edges), we obtain a tree $T(t, s)$ with 4 edges whose frequency pattern is equivalent to
\[
f_1(t) - f_2(s) \sim \left[ \log \frac{x_1(t)x_4(s)}{x_4(t)x_2(s)} \log \frac{x_2(t)x_4(s)}{x_4(t)x_1(s)}, 0, 0 \right] .
\]

On the other hand, we note that for the Kimura 2-parameter model (5.1),
\[
\frac{x_2(t)}{x_4(t)} = 1 + 2\exp(-0.5t)
\]
only admits values in the interval $[1, 3]$, while $x_1(s)/x_4(s)$ is a continuous decreasing function in $s$ that admits all values in the interval $[1, \infty)$. Hence, for every $t > 0$, there exists a unique $k(t) > 0$ such that
\[
\frac{x_2(t)x_4(k(t))}{x_4(t)x_1(k(t))} = 1.
\]
Moreover, $k(t)$ is a continuous function in $t$ and
\[
\lim_{t \to \infty} k(t) = \infty \quad \lim_{t \to 0} k(t) = k_0
\]
where $k_0$ satisfies $x_1(k_0)/x_4(k_0) = 3$.

Now if we denote
\[
g(t) = \frac{x_1(t)x_4(k(t))}{x_4(t)x_2(k(t))}
\]
then $g(t)$ is a continuous function that satisfies
\[
\lim_{t \to \infty} g(t) = 1 \quad \lim_{t \to 0} g(t) = \infty.
\]
We deduce that for a range of $t$,
\[
f_1(t) - f_2(k(t)) \sim [\log g(t), 0, 0, 0]
\]
which admits every patterns of the form $[x, 0, 0, 0]$ with $x > 0$. Similarly
\[
f_2(k(t)) - f_1(t) \sim [-\log g(t), 0, 0, 0]
\]
admits every patterns of the form $[x, 0, 0, 0]$ with $x < 0$. This completes the proof. \qed
\end{proof}
Proof of Theorem 5.1. From Theorem 4.1, there exists a rooted tree $\tau$, a labeling $\psi$ and a vector of real constants $c = (c_1, \ldots, c_S)$ such that

$$c_s + \log \sum_{a \in \mathcal{L}_i, s} \pi(i) \prod_{u \bar{v} \in E(\tau)} P_{a_ua_v}(t_{uv}) = \log(\eta^*_s).$$

For any $t > 0$, we create a new tree $\tau(t)$ by adding an edge $e$ of length $t$ to the root $\rho$ and labeling the additional taxon by the constant vector $(A, A, \ldots, A)$. The log-likelihood function on $e$ of $\tau(t)$ given this taxon labeling is

$$\ell(t) = \sum_{s=1}^{S} \log \left( \sum_{i} \sum_{a \in \mathcal{L}_i, s} \pi(i) \prod_{u \bar{v} \in E(\tau)} P_{a_ua_v}(t_{uv})P_i(t) \right) = -\sum_{s=1}^{S} c_s + \sum_{s=1}^{S} \log \left( \sum_{i} \eta^*_s P_i(t) \right).$$

Theorem 4.4 implies that the tree $\tau$ can be constructed with at most $64S$ edges. Hence, $\tau(t)$ has at most $(64S + 1)$ edges. $\square$

Proof of Lemma 6.1. We first consider the case of linear functions. Assume that $f(x) = ax + b$ such that $f$ is positive in $[0, 1]$. We deduce that $f(-1) = b - a > 0$. Hence $f$ can be written as

$$f(x) = 2(b - a)P_3(x) + 2(b + a)P_2(x) + 2(b + a)P_1(x),$$

using the transition polynomials $P_i$ from (5.3).

The coefficients are positive if and only if $f(-1) = b - a > 0$, or in other words, $f$ has no root in $B$. If $f(x)$ is a monic polynomial of degree 2 with no real roots, then $f$ can be written as

$$f(x) = x^2 - 2ax + a^2 + b^2$$

$$= [(a - 1)^2 + b^2 - 1]P_1(x)$$

$$+ [(a - 1)^2 + b^2 - 2]P_2(x)$$

$$+ 2[(a + 1)^2 + b^2 - 1]P_3(x).$$

The coefficients are positive if and only if $a \pm bi$ do not belong to $B$. $\square$