Abstract
The standard models of sequence evolution on a tree determine probabilities for every character or site pattern. A flattening is an arrangement of these probabilities into a matrix, with rows corresponding to all possible site patterns for one set $A$ of taxa and columns corresponding to all site patterns for another set $B$ of taxa. Flattenings have been used to prove difficult results relating to phylogenetic invariants and consistency and also form the basis of several methods of phylogenetic inference. We prove that the rank of the flattening equals $r v_T(A|B)$, where $r$ is the number of states and $v_T(A|B)$ is the minimum size of a vertex cut separating $A$ from $B$. When $T$ is binary the rank of the flattening equals $r \ell_T(A|B)$ where $\ell_T(A|B)$ equals the parsimony length of the binary character separating $A$ and $B$. We provide a direct proof that requires little more than undergraduate algebra, but note that the formula could also be derived from work by Casanellas and Fernández-Sánchez (2011) on phylogenetic invariants.

Keywords Phylogeny · Tensor flattening · Rank formula · Markov models on trees · Phylogenetic invariants

Mathematics Subject Classification 92D15 · 60J10 · 15A69

1 Introduction

Behind any statistical inference in phylogenetics is a model describing the evolution of the states (alleles/nucleotides/amino acids) observed at each site in the alignment. Under the standard models, the distribution of states at a site is determined by three types of parameters: the phylogeny itself, the distribution of the state at the root, and the transition probabilities along each edge. Together these generate the joint distribution
Fig. 1 An example of patterns or sites on a four taxa tree. The 16 possible patterns are listed in the table on the right - any one of these could be the pattern for a particular site for the state at each leaf, which in turn corresponds to a column of the alignment (reviewed in Bryant et al. (2005), Felsenstein (2004)).

An assignment of states to the taxa is often called a pattern. If there are \( r \) states (say \( r = 2 \) for binary, \( r = 4 \) for nucleotides, \( r = 20 \) for amino acids) and \( n \) taxa then there are \( r^n \) possible patterns. In the example in Fig. 1 we have \( r = 2 \) and \( n = 4 \), so \( r^n = 16 \) possible pattern for the joint distribution. The two states here code for purines (coded \( R \), standing for nucleotides \( A \) or \( G \)) and pyrimidines (coded \( Y \), standing for \( C \) or \( T \)). Let \( p_{i_1i_2i_3i_4} \) denote the joint probability of pattern \( i_1i_2i_3i_4 \), that is the probability under the model of observing state \( i_1 \) at leaf 1, \( i_2 \) at leaf 2 and so on. We can think of these as elements of an \( r^n \) dimensional vector

\[
\begin{bmatrix}
p_{RRRR} \\
p_{RRRY} \\
p_{RRYR} \\
\vdots \\
p_{YYYY}
\end{bmatrix}
\tag{1}
\]

or as a \( 2 \times 2 \times 2 \times 2 \) multidimensional array or tensor. Alternatively we can reshape the vector into a matrix, such as

\[
\begin{bmatrix}
p_{RRRR} & p_{RRRY} & p_{RRYR} & p_{RRYY} \\
p_{PRRR} & p_{PRRY} & p_{PRYR} & p_{PYYY} \\
p_{PYRR} & p_{PYRY} & p_{PYYR} & p_{PYYY} \\
p_{YYRR} & p_{YYRY} & p_{YYYR} & p_{YYYY}
\end{bmatrix}
\tag{2}
\]

In this matrix, the rows correspond to the \( r^2 \) possible ways of assigning states to taxa 1 and 2 while the columns correspond to the \( r^2 \) possible ways of assigning states to taxa 3 and 4. A matrix of this form is called a flattening. In tensor terminology, the flattening is an example of an unfolding of the tensor of pattern probabilities. The idea was introduced into phylogenetics by Pachter and Sturmfels (2004) and developed extensively by Allman and Rhodes to solve a wide range of mathematical problems in phylogenetics (Rhodes and Sullivant 2012; Allman et al. 2013; Allman and Rhodes 2003, 2005, 2006, 2007, 2008, 2009; Allman et al. 2014).
A split is a partition of the set of taxa into two non-empty parts. We can construct a flattening for any split. The rows of the flattening are indexed by all $r^{|A|}$ ways of assigning a state to the taxa in $A$ and the columns are indexed by all $r^{|B|}$ ways of assigning a state to the taxa in $B$. Each entry equals a term $p_{i_1 i_2 \cdots i_n}$ with each state $i_k$ determined by the row index if $k \in A$ and by the column index if $k \in B$. We denote this matrix $\text{flat}_{A|B}$. The matrix in (2) is $\text{flat}_{\{1, 2\}|\{3, 4\}}$ and corresponds to the split $\{1, 2\}|\{3, 4\}$. The flattening for split $\{1, 3\}|\{2, 4\}$ is
\[
\text{flat}_{\{1, 3\}|\{2, 4\}} = \begin{bmatrix}
p_{RRRR} & p_{RRRY} & p_{RYRR} & p_{RYRY} 
p_{PRYY} & p_{PRYR} & p_{PRYY} & p_{PRYY} 
p_{PYRR} & p_{PYRY} & p_{PYRR} & p_{PYRY} 
p_{PYYR} & p_{PYYR} & p_{PYYR} & p_{PYYY}
\end{bmatrix}
\]
and the flattening for $\{1\}|\{2, 3, 4\}$ is
\[
\text{flat}_{\{1\}|\{2, 3, 4\}} = \begin{bmatrix}
p_{RRRR} & p_{RRRY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} & p_{RRYY} \end{bmatrix}.
\]

An important property of flattenings from phylogenies is their rank. Suppose that $e$ is an edge in a phylogeny. Removing $e$ partitions the tree, and hence the set of leaves, into two parts, inducing a split $A|B$ of the set of taxa. We say that $A|B$ is the split of the tree corresponding to edge $e$. Allman and Rhodes (2007) (Proposition 11) proved that, under minor assumptions, if $A|B$ is a split of the tree then the rank of $\text{flat}_{A|B}$ is at most $r$, while if $A|B$ is not a split of the tree then the rank of $\text{flat}_{A|B}$ is at least $r^2$.

Because of this property, flattenings and their ranks have played a prominent role in the mathematics of phylogenetics, particularly with respect to the development and construction of phylogenetic invariants (Allman and Rhodes 2008, 2005; Pachter and Sturmfels 2004). Roughly speaking, a phylogenetic invariant for a tree is a function on vectors of pattern probabilities such as (1) which is zero when the probability distribution comes from that tree and non-zero otherwise. Invariants provide an elegant and distinctive means to infer single phylogenies or classes of phylogenies, although the approach does not appear to have reached its potential.

Several other phylogenetic methods are based on flattenings. In an original and influential chapter, Eriksson (2005) outlined a polynomial-time method for inferring phylogenies with few assumptions about the evolutionary process. The SVDQuartets method (Chifman and Kubatko 2014) uses flattenings to infer trees for subsets of four taxa, subsequently assembling these four-taxon trees into one for the complete set of taxa. The method is statistically consistent even in the presence of incomplete lineage sorting. Quartet-based approaches based on flattening have also been developed by Fernández-Sánchez and Casanellas (2016), and Casanellas et al. (2021). An excellent review of many of these developments is found in Allman et al. (2017).

The calculation of the split scores described in Allman et al. (2017) is justified by rank bounds of flattenings. Given a split (and $r$-state sequence data) the split score is defined using the sum of all except the largest $r$ singular values, divided by the sum of all singular values. If the data exactly reflects the pattern probabilities of a tree containing that split then the fact that the flattening has rank $r$ gives a split score of
0. For other splits, the score will be higher, the intuition being that the worse a split fits the tree, the higher its split score will be. We provide a theoretical foundation for this intuition by showing that a split has low rank if and only if the split belongs to a nearby tree.

We note that our expression for the rank of a flattening $\text{flat}_{A|B}$ in a binary tree can be derived as a direct consequence of Proposition 3.1 in Casanellas and Fernández-Sánchez (2011). Casanellas and Fernández-Sánchez use group representation theory to prove a more general result, revealing structure within the flattenings due to symmetries in the transition matrices.

In this paper we extend and complete the theorem of Allman and Rhodes and derive an exact formula for the rank of arbitrary flattenings of a tree for both binary and non-binary trees. Our result corrects a formula appearing in Eriksson (2005), see (Casanellas and Fernández-Sánchez 2011), remark 3.3. We show that the rank of $\text{flat}_{A|B}$ is given by $r^{\nu_T(A|B)}$, where $\nu_T(A|B)$ is the size of the smallest vertex cut separating $A$ and $B$ (defined below). If $T$ is binary, then $\nu_T(A|B)$ equals the parsimony length $\ell_T(A|B)$ of the binary character separating $A$ and $B$, a classical measure of fit between a character and a tree (Felsenstein 2004). In contrast to Casanellas and Fernández-Sánchez (2011) we identify exactly the (modest) constraints on parameters required for the rank result to hold (conditions (C1)–(C3) below).

Our most important contribution is the simplicity of our proof, which uses little more machinery than undergraduate linear algebra. Nevertheless, it demonstrates a fundamental link between combinatorial and algebraic approaches to phylogenetics, domains which have been mainly progressing in parallel. As a direct application we prove an exact formula connecting flattening rank and standard distances between trees (Theorem 9).

In the next section we present the notation and basic results needed to establish this result. We note that we generalise from characters to maps on arbitrary subsets of vertices, and define our concepts in this more general context. The main result is proved in a series of lemmas and propositions in the final section.

## 2 Background and definitions

In this section we give a fairly minimal set of definitions sufficient to describe the main result. Excellent introductions to this material can be found in Felsenstein (2004), Semple and Steel (2003), Steel (2016).

### 2.1 Trees and characters

**Definition 1** An *unrooted phylogeny* is an acyclic, connected graph $T = (V, E)$ with leaf set (taxon set) $L(T)$. We say that $T$ is binary (full resolved) if every non-leaf vertex has degree three. In a *rooted phylogeny* $T = (V, E, \rho)$ one vertex is selected as the root $\rho$ and edges are directed away from $\rho$. The rooted phylogeny $T$ is binary if every non-leaf vertex has out-degree 2.
**Definition 2** A character is a function $f$ from $L(T)$ to a set of states $[r] = \{1, 2, \ldots, r\}$. We will consider a more general situation where $f$ has a domain $\text{dom}(f)$ equal to any non-empty subset of $V$, not just $L(T)$.

One example of a character is the map from $L(T)$ to the set of nucleotides given by a column of the aligned sequences. In this case, $r = 4$ and the four states correspond to $A$, $C$, $G$ or $T$. In some situations, it can be advantageous to recode the four states as two: one for purines ($A$ or $G$, coded as $R$) and pyrimidines ($C$ or $T$, coded as $Y$), giving an example where $r = 2$. With protein-coding sequence, it is common to define characters using the amino acids determined by triples of nucleotides (codons), giving characters with $r = 20$. In fact $r$ can become arbitrarily large in covarion or profile models since the set of states for these models is the product of the set of nucleotides (or proteins) and a set of rate or profile classes (Yourdkhani et al. 2021; Tuffley and Steel 1998; Galtier 2001).

**Definition 3** The length of a function $F : V \to [r]$ is defined as

$$\ell_T(F) = |\{\{u, v\} \in E : F(u) \neq F(v)\}|$$

and the (parsimony) length of a function $f$ with domain $\text{dom}(f) \subseteq V$ is the length of a minimum extension

$$\ell_T(f) = \min\{\ell_T(F) : F|_{\text{dom}(f)} = f\}.$$

The length of a function or character can be expressed equivalently using vertex and edge cuts (Semple and Steel 2003). For $E' \subseteq E$ or $V' \subseteq V$ we let $T \setminus E'$ and $T \setminus V'$ denote the graphs resulting from deleting $E'$ or $V'$ respectively. The length $\ell_T(f)$ equals the minimum cardinality of an edge cut $E' \subseteq E$ such that $f(u) = f(v)$ whenever $u$ and $v$ are in the same component of $T \setminus E'$. In a similar way, we let $\nu_T(f)$ denote the minimum cardinality of a vertex cut $V' \subseteq V$ such that $f(u) = f(v)$ whenever $u$ and $v$ are in the same component of $T \setminus V'$. Note that $V'$ can intersect $\text{dom}(f)$, and that for all $f$ we have $\nu_T(f) \leq \ell_T(f)$.

If $A$ and $B$ are disjoint subsets of $V$ then we let $\ell_T(A|B)$ and $\nu_T(A|B)$ denote $\ell_T(f)$ and $\nu_T(f)$, where $f$ equals the indicator function

$$f(x) = \begin{cases} 0 & \text{if } x \in A \\ 1 & \text{if } x \in B. \end{cases}$$

Steel (1993); Semple and Steel (2003) observed that the parsimony length of a character with two states can be expressed in terms of the size of disjoint path sets by Menger’s theorem.

**Proposition 4** Let $T = (V, E)$ be an unrooted phylogeny and let $A$ and $B$ be disjoint subsets of $V$. Then $\ell_T(A|B)$ equals the maximum cardinality of a set of edge-disjoint paths connecting vertices in $A$ with vertices in $B$, while $\nu_T(A|B)$ equals the maximum cardinality of a set of vertex-disjoint paths connecting vertices in $A$ with vertices in $B$. 
Fig. 2 An example of a minimal edge cut (dotted edge) and a minimal vertex cut (square nodes) for a split \( A \mid B \) represented here using filled and unfilled circles.

Fig. 3 The tree on the left is an example where the minimum vertex cut (the central square node) is strictly smaller than a minimum edge cut. In the example on the right, the sets \( A \) and \( B \) are arbitrary disjoint subsets of vertices. Furthermore the minimum vertex cut (the square node and interior circle node) includes a node in \( A \cup B \).

Suppose that \( T \) is binary and that \( p_1 \) and \( p_2 \) are any two paths beginning and ending at leaves of \( T \). We then have that \( p_1 \) and \( p_2 \) are edge disjoint if and only if they are vertex disjoint, and so obtain the following corollary of Proposition 4.

**Corollary 5** Let \( T = (V, E) \) be a binary phylogeny and suppose \( A \) and \( B \) are disjoint subsets of \( L(T) \). Then \( \ell_T(A \mid B) = \nu_T(A \mid B) \).

Figure 2 illustrates \( \ell_T \) and \( \nu_T \). Leaves in \( A \) and \( B \) are indicated using filled and unfilled leaf nodes. The dotted edges form a minimal edge cut while the square vertices form a minimal vertex cut.

Corollary 5 states that if \( A \) and \( B \) are sets of leaves in a binary tree \( T \) then \( \ell_T(A \mid B) = \nu_T(A \mid B) \). Equality does not hold when \( T \) is not binary nor when \( A \) and \( B \) are not restricted to leaves (Fig. 3).

### 2.2 Distances between trees

Bryant (2004) showed that the parsimony score of a binary character (or split) can be expressed in terms of either distances based on rearrangements of phylogenetic trees (Fig. 4). In a subtree-prune-and-regraft (SPR) rearrangement an edge in the tree is removed and one endpoint of the edge is connected to an arbitrary location in the opposite component. In a tree-bisection-and-regraft (TBR) an edge is removed and a new edge is inserted between an arbitrary location in one component and an arbitrary location in the other component. By definition, any SPR rearrangement is also a TBR rearrangement, but there are TBR rearrangements which are not SPR rearrangements.
Fig. 4 Two standard rearrangements for phylogenetic trees. In the first, subtree-prune-and-regraft (SPR) an edge in the tree (left) is removed and one of its endpoints is connected to a point in the opposite component (top middle) after which degree two vertices are surpressed. In the second, tree-bisection-and-regraft (TBR) an edge in the tree (left) is removed. A new edge which connects any two points in each component, and degree two vertices are surpressed. Every SPR operation is a TBR operation but not vice versa.

Both SPR and TBR rearrangements are extensively used in optimization and sampling algorithms for phylogenetic inference, and have been for some time (Swofford et al. 1996). The operations induce a natural edit distance on phylogenetic trees. Let \( T_1 \) and \( T_2 \) be two fully-resolved trees. The SPR-distance \( d_{SPR}(T_1, T_2) \) is the smallest number of SPR rearrangements required to transform \( T_1 \) into \( T_2 \). The TBR-distance \( d_{TBR}(T_1, T_2) \) is the smallest number of TBR rearrangements required to transform \( T_1 \) into \( T_2 \) (Allen and Steel 2001).

2.3 The general Markov model of sequence evolution

We assume the general Markov model of sequence evolution, which we introduce here. The model was first proposed by Barry and Hartigan (1987). It is general enough that essentially all standard models for sequence evolution on a phylogeny with iid sites can be considered as special cases (Jayaswal et al. 2005).

Let \( T = (V, E_\rho) \) be a rooted tree with root \( \rho \). Let \( \pi_\rho \) denote the root distribution and for each directed edge \((u, v)\) we associate an \( r \times r \) transition probability matrix \( P_{uv} \). Let \( X_v \) denote the random state associated with vertex \( v \) and, for \( A \subseteq V \), let \( X_A \) denote the joint random variable \( X_a : a \in A \).

The joint probability that \( X_v = F(v) \) for all \( v \in V \) is defined

\[
\pi(X_V = F) = \pi_\rho(F(\rho)) \prod_{(u,v) \in E_\rho} P_{uv}(F(u), F(v)).
\]
The marginal probabilities for maps \( f : A \rightarrow [r] \) on some subset \( A \subseteq V \) are then given by

\[
\pi(X_A = f) = \sum_{F : F|A = f} \pi(F).
\]

With this notation, the probability for a character \( f : L(T) \rightarrow [r] \) is \( \pi(X_{L(T)} = f) \).

In addition to the general Markov model we will be making the following, modest, assumptions about transition matrices:

(C1) The transition matrices \( P_{uv} \) are non-singular.

(C2) \( \pi_{\rho}(i) > 0 \) for all \( i \in [r] \).

(C3) \( P_{uv}(i, j) > 0 \) for all internal edges \( (u, v) \) and \( i, j \in [r] \).

In the final section we show that these conditions are required for the main rank result to hold.

We will make use of the following consequences of (C1)–(C3), all of which can be found in the literature (e.g. Allman and Rhodes 2003; Steel et al. 1994; Steel 2016) or are proved directly from positivity and invertibility requirements.

**Proposition 6** Suppose that conditions (C1), (C2) and (C3) are satisfied. Then

1. \( \pi(X_v = i) > 0 \) for all \( v \in V \) and \( i \in [r] \).
2. For all \( u, v \in V \) (not necessarily adjacent) the \( r \times r \) matrix with entries \( P_{uv} = \pi(X_v = j | X_u = i) \) is non-singular.
3. For all \( u, v \in V \) the \( r \times r \) matrix with entries \( M_{ij} = \pi(X_u = i, X_v = j) \) is non-singular.
4. We can move the root to any internal vertex and update root distribution and transition matrices so that conditions (C1)–(C3) are still satisfied and all character probabilities are unchanged.

Conditional independence for the variables \( X_v \) is, as with any graphical model, determined by cuts in the graph. More specifically, if \( A_1, A_2, \ldots, A_k \) are the components of \( T \setminus \{v\} \) then \( X_{A_1}, X_{A_2}, \ldots, X_{A_k} \) are conditionally independent, given \( X_v \), see (Steel 2016, Lemma 7.1). The following Proposition is a direct consequence of this result.

**Proposition 7** If \( V' \subseteq V \) and \( A_1, A_2, \ldots, A_k \) are the components of \( T \setminus \{V'\} \) then \( X_{A_1}, X_{A_2}, \ldots, X_{A_k} \) are conditionally independent, given \( X_{V'} \).

**Proof** Apply Lemma 7.1 of Steel (2016) with respect to some \( v \in V' \) and then apply the result recursively on the components. \( \square \)

### 2.4 The main theorems

Let \( T = (V, E) \) be a phylogeny and suppose that \( A \) and \( B \) are disjoint subsets of \( V \). We let \( \text{flat}_{A \mid B} \) denote the matrix with rows indexed by maps \( f_A : A \rightarrow [r] \), columns indexed by maps \( f_B : B \rightarrow [r] \) and \( (f_A, f_B) \) entry given by

\[
\text{flat}_{A \mid B}(f_A, f_B) = \pi(X_A = f_A, X_B = f_B).
\]
Assumptions (C1)–(C3) imply that flat_{A|B} is an $r^{|A|} \times r^{|B|}$ matrix with entries which are strictly positive and sum to 1.

**Theorem 8** Let $T = (V, E)$ be a phylogeny with root distribution and transition matrices satisfying (C1)–(C3). If $A$ and $B$ are disjoint subsets of $V$ then

$$\text{rank}(\text{flat}_{A|B}) = r^{\nu_T(A|B)}.$$  

If $T$ is binary and $A$ and $B$ subsets of $L(T)$ then

$$\text{rank}(\text{flat}_{A|B}) = r^{\ell_T(A|B)}.$$  

As a corollary of Theorem 8 we obtain the result of Allman and Rhodes (2007) that if $A|B$ is a split in the tree then the rank of flat_{A|B} is $r$, while if $A|B$ is not a split in the tree then the rank of flat_{A|B} is at least $r^2$.

Combining Theorem 8 with results of Bryant (2004) we connect the rank of flattenings with neighborhoods of trees.

**Theorem 9** Let $T = (V, E)$ be a binary phylogeny with root distribution and transition matrices satisfying (C1)–(C3). Let $A\mathbin{|}B$ be a split and let $T_1$ and $T_2$ be the closest trees to $T$ which contain the split $A\mathbin{|}B$, with respect to SPR and TBR distances respectively. Then

$$\text{rank}(\text{flat}_{A\mathbin{|}B}) = r^{d_{\text{SPR}}(T, T_1)+1} = r^{d_{\text{TBR}}(T, T_2)+1}$$  

3 Proof of theorem 8

To prove the main theorem we first show that $r^{\nu_T(A|B)}$ provides an upper bound for the rank of flat_{A|B} and then show that this upper bound is actually achieved.

**Lemma 10** Let $T = (V, E)$ be a phylogeny and suppose that $A$ and $B$ are disjoint subsets of $V$. Then

$$\text{rank}(\text{flat}_{A\mathbin{|}B}) \leq r^{\nu_T(A\mathbin{|}B)}.$$  

**Proof** Let $C$ be a minimum cardinality vertex cut of $T$ such that each component of $T \setminus C$ contains vertices from at most one of $A$ or $B$. By Proposition 7, $X_{A\setminus C}$ and $X_{B\setminus C}$ are conditionally independent given $X_C$. Hence for all maps $f_A : A \to [r]$ and $f_B : B \to [r]$ we can factor $\pi(X_A = f_A, X_B = f_B)$ as

$$\pi(X_A = f_A, X_B = f_B) = \sum_{f_C} \pi(X_A = f_A, X_B = f_B | X_C = f_C) \pi(X_C = f_C) = \sum_{f_C} \pi(X_A = f_A | X_C = f_C) \pi(X_B = f_B, X_C = f_C)$$
Define the \( r^{|A|} \times r^{|C|} \) matrix \( R \) by

\[
R(f_A, f_C) = \pi(X_A = f_A|X_C = f_C),
\]

and the \( r^{|C|} \times r^{|B|} \) matrix \( S \) by

\[
S(f_C, f_B) = \pi(X_B = f_B, X_C = f_C).
\]

Then

\[
\text{flat}_{A|B} = RS
\]

and

\[
\text{rank}(\text{flat}_{A|B}) \leq \text{rank}(R) \leq r^{|C|} = r^{v_T(A|B)}.
\]

To illustrate, consider the tree in Fig. 2. The sets \( A \) and \( B \) are indicated by filled and unfilled leaves respectively. Removing the three marked vertices separates all leaves in \( A \) from leaves in \( B \). Hence \( v_T(A|B) \leq 3 \) and \( \text{rank}(\text{flat}_{A|B}) \leq r^{v_T(A|B)} \). We note that the formula in Eriksson (2005) gives an incorrect generic rank of \( r^4 \) for this flattening, as observed by Casanellas and Fernández-Sánchez (2011).

We now work towards finding a matching lower bound for Lemma 10. We start by demonstrating that the rank of \( \text{flat}_{A|B} \) does not increase if we marginalise over a subset of the variables.

**Lemma 11** Let \( A \) and \( B \) be disjoint subsets of \( V \) and suppose \( A' \subseteq A \) and \( B' \subseteq B \). Then

\[
\text{rank}(\text{flat}_{A'|B'}) \leq \text{rank}(\text{flat}_{A|B}).
\]

**Proof** The idea behind this proof is to show that \( \text{flat}_{A'|B'} \) can be expressed as the product of \( \text{flat}_{A|B} \) with two other matrices, from which the rank inequality follows.

For any \( f_{A'} : A' \to [r] \) and \( f_{B'} : B' \to [r] \) we have

\[
\text{flat}_{A'|B'}(f_{A'}, f_{B'}) = \pi(X_{A'} = f_{A'}, X_{B'} = f_{B'})
\]

\[
= \sum_{g_A: A \to [r]} \sum_{g_B: B \to [r]} \pi(X_A = g_A, X_B = g_B)
\]

\[
= \sum_{g_A: A \to [r]} \sum_{g_B: B \to [r]} \text{flat}_{A|B}(g_A, g_B).
\]
Hence there is an $r^{|A'|} \times r^{|A|} 0-1$ (binary) matrix $U_A$ and a $r^{|B'|} \times r^{|B|} 0-1$ matrix $U_B$ such that

$$\text{flat}_{A'|B'} = U_A \text{flat}_{A|B} U_B^T$$

and rank($\text{flat}_{A'|B'}$) $\leq$ rank($\text{flat}_{A|B}$).

The next Lemma establishes Theorem 8 in the extremal case that $\nu_T(A\mid B) = |A| = |B|$. This is where the bulk of the work proving the main theorem is carried out.

**Lemma 12** Suppose that $T = (V, E)$, $A$ and $B$ satisfy the conditions of Theorem 8. If $|A| = |B| = m$ and there are $m$ disjoint paths connecting elements of $A$ to elements of $B$ then rank($\text{flat}_{A|B}$) = $r^m$.

**Proof** We prove the result by induction on $m$.

For the base case suppose that $m = 1$, $A = \{a\}$ and $B = \{b\}$. Then

$\text{flat}_{A|B}(i, j) = \pi(X_a = i, X_B = j)$

so $\text{flat}_{A|B}$ is full rank $r = r^m$ by Proposition 6.

Next, assume that the result holds whenever $|A| = |B| = m$. Suppose that there is a collection $P'$ of $m+1$ vertex disjoint paths connecting elements of $A$ and $B$, where $|A| = |B| = m+1$. There is a path $p$ which is separated from the rest of the paths in $P'$ by removal of a single edge in the tree. This can be seen by fixing an arbitrary vertex $u$ and then taking $p$ to be the path which is furthest from $u$. We let $v$ be the vertex on $p$ which is closest to $u$, and let $e = \{v, w\}$ denote the first edge on the path from $v$ to $u$ (Fig. 5).

Let $a \in A$ and $b \in B$ be the endpoints of the path $p$ and let $A' = A \setminus \{a\}$, $B' = B \setminus \{b\}$. Given any function $f : A \to [r]$ or $g : B \to [r]$ we use $f'$ and $g'$ to denote the restriction of $f$ to $A$ or $g$ to $B'$.

Define the matrices $F, G$ by

$$F(i, k) = \pi(X_a = i|X_v = k)$$

$$G(k, j) = \pi(X_v = k|X_b = j)$$

and for each $k \in [r]$ define the $r^m \times r^m$ matrix $H^{(k)}$ by

$$H^{(k)}(f', g') = \pi(X_{A'} = f', X_{B'} = g'|X_v = k).$$
We consider the tree to be rooted at \( v \). The matrices \( F, G, H^{(k)} \) provide a decomposition of \( \text{flat}_{A|B} \):

\[
\text{flat}_{A|B}(f, g) = \sum_{k=1}^{r} F(f(a), k) H^{(k)}(f', g') G(k, g(b))
\]

which we rewrite using the Kronecker product \((\otimes)\) as

\[
\text{flat}_{A|B} = (F \otimes I_{r-1}) \begin{bmatrix}
H^{(1)} \\
H^{(2)} \\
\vdots \\
H^{(r)}
\end{bmatrix} (G \otimes I_{r-1}). \tag{3}
\]

The matrices \( F \) and \( G \) are both non-singular, by Proposition 6. Furthermore, for each \( k \) the matrix \( H^{(k)} \) equals the flattening matrix \( \text{flat}_{A|B} \) but with respect to root \( w \) and root distribution

\[
\pi_w(i) = P_{vw}(k, i)
\]

which is strictly positive for all \( i \) by (C3). By the induction hypothesis, each matrix \( H^{(k)} \) is also non-singular. It follows from standard properties of the Kronecker product (Horn and Johnson 1991, Chapter 4) that all three matrices in the product (3) are non-singular, so that \( \text{flat}_{A|B} \) is also non-singular. \( \square \)

We can now prove the main theorem.

**Proof** (Theorem 8) Suppose that \( A \) and \( B \) are disjoint subsets of \( V \). From Lemma 10 we have

\[
\text{rank}(\text{flat}_{A|B}) \leq r^{\nu_T(A|B)}.
\]

By Proposition 4 there are \( \nu_T(A|B) \) vertex disjoint paths connecting vertices in \( A \) to vertices in \( B \). Let \( \hat{A} \) and \( \hat{B} \) be the endpoints of these paths, so \( |\hat{A}| = |\hat{B}| = \nu_T(\hat{A}|\hat{B}) = \nu_T(A|B) \). By Lemma 12 we have \( \text{rank}(\text{flat}_{\hat{A}|\hat{B}}) = r^{\nu_T(A|B)} \) and by Lemma 11 we have

\[
\text{rank}(\text{flat}_{A|B}) \geq \text{rank}(\text{flat}_{\hat{A}|\hat{B}}) = r^{\nu_T(A|B)}. \tag{5}
\]

The second part of the theorem follows from Corollary 5. \( \square \)

Theorem 9 follows as a corollary of Theorem 8 and Theorem 5.2 of Bryant (2004). We note that none of the conditions (C1)–(C3) on the root distribution and transition matrices can be eliminated. For example, consider the four taxa tree in Figure 1, a case studied in detail by Allman and Rhodes (2006). The joint probability distribution for \( X_1, X_2, X_3, X_4 \) is

\[
\pi(X_1 = a, X_2 = b, X_3 = c, X_4 = d)
\]
\[ = \sum_i \sum_j P_{u1}(i, a) P_{u2}(i, b) \pi_{\rho}(i) P_{uv}(i, j) P_{v3}(j, c) P_{v4}(j, d). \]

Gather terms, we obtain a decomposition

\[ \text{flat}_{\{1,3\}|\{2,4\}} = UDV \]

where \( U \) and \( V \) are \( r^2 \times r^2 \) matrices and \( D \) is a diagonal matrix with diagonal entries

\[ D_{ij;ij} = \pi_{\rho}(i) P_{uv}(i, j). \]

From here we see that if there is \( i \) such that \( \pi_{\rho}(i) = 0 \), or \( ij \) such that \( P_{uv}(i, j) = 0 \), then

\[ \text{rank}(\text{flat}_{\{1,3\}|\{2,4\}}) < r^2. \]

The example can be extended directly to any tree with more than four taxa.

**Acknowledgements**  We thank Elizabeth Allman and an anonymous reviewer for their many useful suggestions and comments.

**Funding**  Open Access funding enabled and organized by CAUL and its Member Institutions.

**Open Access**  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

**References**

Allen BL, Steel M (2001) Subtree transfer operations and their induced metrics on evolutionary trees. Ann Comb 5(1):1–15
Allman ES, Jarvis PD, Rhodes JA, Sumner JG (2013) Tensor rank, invariants, inequalities, and applications. SIAM J Matrix Anal Appl 34(3):1014–1045
Allman ES, Kubatko LS, Rhodes JA (2017) Split Scores: a tool to quantify phylogenetic signal in genome-scale data. Syst Biol 66(4):620–636
Allman ES, Rhodes (2005) Phylogenetic invariants. In: Gascuel Olivier (ed) Mathematics of Evolution and Phylogeny. Oxford University Press, Oxford, pp 108–146
Allman ES, Rhodes JA (2003) Phylogenetic invariants for the general markov model of sequence mutation. Math Biosci 186(2):113–144
Allman ES, Rhodes JA (2006) The identifiability of tree topology for phylogenetic models, including covarion and mixture models. J Comput Biol 13(5):1101–1113
Allman ES, Rhodes JA (2007) Identifying evolutionary trees and substitution parameters for the general markov model with invariable sites. Math Biosci 211(1):18–33
Allman ES, Rhodes JA (2008) Phylogenetic ideals and varieties for the general Markov model. Adv Appl Math 40(2):127–148
Allman ES, Rhodes JA (2009) The identifiability of covarion models in phylogenetics. IEEE/ACM Trans Comput Biol Bioinf 6(1):76–88
Allman ES, Rhodes JA, Taylor A (2014) A semialgebraic description of the general markov model on phylogenetic trees. SIAM J Discret Math 28(2):736–755
Barry D, Hartigan JA (1987) Statistical analysis of hominoid molecular evolution. Stat Sci 2(2):191–207
Bryant D (2004) The splits in the neighborhood of a tree. Ann Comb 8(1):1–11
Bryant D, Galtier N, Poursat M-A (2005) Likelihood calculation in molecular phylogenetics. In: Gascuel Olivier (ed) Mathematics of Evolution and Phylogeny. Oxford University Press, Oxford, UK, pp 33–62
Casanellas M, Fernández-Sánchez J (2011) Relevant phylogenetic invariants of evolutionary models. J Math Pures Appl 96(3):207–229
Casanellas M, Fernández-Sánchez J, Garrote-López M (2021) Saq: semi-algebraic quartet reconstruction. IEEE/ACM Trans Comput Biol Bioinform
Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. Bioinformatics 30(23):3317–3324
Eriksson N (2005) Tree construction using singular value decomposition. In: Pachter L, Sturmfels B. Editors (eds) Algebraic Statistics for Computational Biology. Cambridge University Press, pp 347–358
Felsenstein J (2004) Inferring Phylogenies. Sinauer Associates Sunderland, MA
Fernández-Sánchez J, Casanellas M (2016) Invariant versus classical quartet inference when evolution is heterogeneous across sites and lineages. Syst Biol 65(2):280–291
Galtier N (2001) Maximum-likelihood phylogenetic analysis under a covarion-like model. Mol Biol Evol 18(5):866–873
Horn RA, Johnson CR (1991) Topics in matrix analysis. Cambridge University Presss, Cambridge, pp 37–39
Jayaswal V, Jermiin LS, Robinson J (2005) Estimation of phylogeny using a general markov model. Evol Bioinforma 1:117693430500100000
Pachter L, Sturmfels B (2004) Tropical geometry of statistical models. Proc Natl Acad Sci 101(46):16132–16137
Rhodes JA, Sullivant S (2012) Identifiability of large phylogenetic mixture models. Bull Math Biol 74(1):212–231
Semple C, Steel M (2003) Phylogenetics, vol 24. Oxford University Press
Steel MA, Székely LA, Hendy MD (1994) Reconstructing trees when sequence sites evolve at variable rates. J Comput Biol 1(2):153–163
Steel MA (1993) Distributions on bicoloured binary trees arising from the principle of parsimony. Discret Appl Math 41(3):245–261
Steel M (2016) Phylogeny: discrete and random processes in evolution. SIAM
Swofford D, Olsen GI, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) Molecular Systematics, 2nd edn. Sinauer, pp 407–514
Tuffley C, Steel M (1998) Modeling the covarion hypothesis of nucleotide substitution. Math Biosci 147(1):63–91
Yourdkhani S, Allman ES, Rhodes JA (2021) Parameter identifiability for a profile mixture model of protein evolution. J Comput Biol 28(6):570–586

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.