SIAM J. DISCRETE MATH.
Vol. 28, No. 2, pp. 736–755

A SEMIALGEBRAIC DESCRIPTION OF THE GENERAL MARKOV MODEL ON PHYLOGENETIC TREES*

ELIZABETH S. ALLMAN†, JOHN A. RHODES†, AND AMELIA TAYLOR‡

Abstract. Many of the stochastic models used in inference of phylogenetic trees from biological sequence data have polynomial parameterization maps. The image of such a map—the collection of joint distributions for all parameter choices—forms the model space. Since the parameterization is polynomial, the Zariski closure of the model space is an algebraic variety which is typically much larger than the model space but amenable to study with algebraic methods. Of practical interest, however, is not the full variety but the subset formed by the model space. Here we develop complete semialgebraic descriptions of the model space arising from the k-state general Markov model on a tree, with slightly restricted parameters. Our approach depends upon both recently formulated analogues of Cayley's hyperdeterminant and on the construction of certain quadratic forms from the joint distribution whose positive (semi)definiteness encodes information about parameter values.

Key words. phylogenetic tree, phylogenetic variety, semialgebraic set, general Markov model

AMS subject classifications. 60J20, 92D15, 92D20, 62P10, 14P10

DOI. 10.1137/120901568

1. Introduction.

Statistical inference of evolutionary relationships among organisms from DNA sequence data is routinely performed using probabilistic models of sequence evolution along a tree. A site in a sequence is viewed as a 4-state (A,C,G, T) random variable, which undergoes state changes as it descends along the tree from an ancestral organism to its modern descendants. Such models exhibit a rich mathematical structure, which reflects both the combinatorial features of the tree and the algebraic way in which stochastic matrices associated to edges of the tree are combined to produce a joint probability distribution describing sequences of the extant organisms.

One thread in the literature on such models has utilized the viewpoint of algebraic geometry to understand the probability distributions that may arise. This is natural, since the distributions are in the image of a polynomial map, and the image thus lies in an algebraic variety. The defining equations of this variety (which depend on the tree topology) are called *phylogenetic invariants*. That a probability distribution satisfies them can be taken as evidence that it arose from sequence evolution along the particular tree. Phylogenetic invariants and varieties have been extensively studied by many authors [11, 23, 15, 19, 18, 2, 29, 6, 10, 25, 9] (see [5] for more references) with goals ranging from biological (improving data analysis) to statistical (establishing the identifiability of model parameters) to purely mathematical.

However, it has long been understood that, in addition to the equalities of phylogenetic invariants, inequalities should play a role in characterizing those distributions of interest for statistical purposes. Much of a phylogenetic variety is typically composed of points not arising from stochastic parameters but rather from applying the

*Received by the editors December 6, 2012; accepted for publication (in revised form) February 21, 2014; published electronically May 15, 2014.
http://www.siam.org/journals/sidma/28-2/90156.html
†Department of Mathematics and Statistics, University of Alaska, Fairbanks AK 99775 (e.allman@alaska.edu, j.rhodes@alaska.edu). The work of these authors was supported by a grant from the National Science Foundation, DMS 0954865.
‡Department of Mathematics, Colorado College, Colorado Springs, CO 80903 (amelia.taylor@coloradocollege.edu).

same polynomial parameterization to complex parameters. Thus the model space—the set of probability distributions arising as the image of stochastic parameters on a tree—can be considerably smaller than the set of all probability distributions on the variety. A recent computation [32] demonstrated that for the 2-state general Markov model on the 3-leaf tree, for example, the model space is only about 8% of the nonnegative real points on the variety. Inequalities can thus be crucial in determining if a probability distribution arises from a model.

In the pioneering 1987 paper of Cavender and Felsenstein [11], polynomial equalities and inequalities are given that can test which of the 3 possible unrooted leaf-labeled 4-leaf trees might have produced a given probability distribution, and thus in principle determine evolutionary relationships between 4 organisms. Despite many advances in understanding phylogenetic invariants in the intervening years, little has been accomplished in finding or understanding the necessary inequalities. The potential usefulness of such inequalities, meanwhile, has been demonstrated in [16], where an inequality that holds for the 2-state model on all tree topologies plays a key role in studying loss of biodiversity under species extinction. In [7] a small number of inequalities, dependent on the tree, were used to show that for certain mixture models trees were identifiable from probability distributions.

Recent independent works by Zwiernik and Smith [32] and by Klaere and Liebscher [20] provided the first substantial progress on the general problem of finding sufficient inequalities to describe the model space. Both groups successfully formulated inequalities for the 2-state general Markov model on trees, using different viewpoints. While the 2-state model has some applicability to DNA sequences, through a purine/pyrimidine encoding of nucleotides, it is not clear how to extend these methods to more general k-state models, or even to the $k = 4$ state model directly applicable to DNA sequences.

In this work we provide a novel approach to understanding the model space of the general Markov model on trees which has the advantage of extending from the 2-state to the k-state model with little modification. Our goal is a semialgebraic description (given by a boolean combination of polynomial equalities and inequalities) of the set of probability distributions that arise on a specific tree. Such a description exists by the Tarski–Seidenberg theorem [31, 26], since the stochastic parameter space for any k-state general Markov model is a semialgebraic set, so its image under the polynomial parameterization must be as well. However, we seek an *explicit* description, and the Tarski–Seidenberg theorem does not provide a useful means of obtaining it.

Our method for obtaining a semialgebraic model description applies equally easily for all k and all trees. We obtain inequalities using a recently formulated analogue of Cayley's $2 \times 2 \times 2$ hyperdeterminant from [1], and the construction of certain quadratic forms from the joint distribution whose positive (semi)definiteness encodes information about parameter values. Using Sylvester's classic theorem about the minors of real symmetric matrices defining such quadratic forms, we give the explicit semialgebraic description.

To prove our results for general k, we must impose some restrictions on the set of parameters under consideration and thus formulate a notion of *nonsingular parameters*. In the $k = 2$ case, this notion is particularly natural from a statistical point of view, as it expresses independence of certain subsets of variables. For any k, nonsingular parameters allow the use of natural $\mathrm{GL}(k, \mathbb{C})$ actions on probability distributions, which is fundamental to our techniques.

There is yet another method for obtaining a partial semialgebraic description of the general Markov model on trees, using Sturm sequences. From a probability distribution P, it is possible to construct matrices which—if P is in the image of the Markov parameterization on a tree T—are diagonalizable with eigenvalues equal to some of the numerical parameters. Using Sturm theory, one can then test that these parameters are actually probabilities, i.e., if they lie in the interval $(0, 1)$. Indeed, the Sturm approach can produce polynomial inequalities of *smaller* degree than the quadratic form approach introduced here. While this may have practical advantages (since the distributions arising on phylogenetic trees have very small entries with many sampling zeroes and numerical error is a real problem in polynomial evaluation at this magnitude), we do not develop the method here.

This paper is organized as follows: In section 2 we formally introduce the general Markov model on trees and set basic notations and terminology, including the notion of nonsingular parameters. In section 3, we give a semialgebraic description of the general Markov model on the 3-leaf tree using the work of [1] and Sylvester's theorem on quadratic forms. In section 4, we give the main result: a semialgebraic description of the k-state general Markov model on n-leaf trees for nonsingular parameters. For the 2-state model, we prove a stronger result, dropping the nonsingularity assumption.

2. Definitions and notations.

2.1. The general Markov model on trees. We review the k-state general Markov model on trees, GM(k), whose parameters consist of a combinatorial object, a tree, and a collection of numerical parameters that are associated to a rooted version of the tree. Let $T = (V, E)$ be a binary tree with leaves $L \subseteq V$, $|L| = n$, and $\{X_a\}_{a\in V}$ a collection of discrete random variables associated to the nodes, all with state space $[k] = \{1, 2, \dots, k\}$. Distinguish an internal node r of T to serve as its root, and direct all edges of T away from r. Though necessary for parameterizing the model, the choice of r will not matter in our final results, as will be shown in section 4.

For a tree T rooted at r, numerical parameters $\{\boldsymbol{\pi}, \{M_e\}_{e\in E}\}$ for the GM(k) model on T are:

(i) A *root distribution* row vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_k)$ with nonnegative entries summing to 1;

(ii) *Markov matrices* M_e, of size $k \times k$ with nonnegative entries and row sums equal to 1.

The vector $\boldsymbol{\pi}$ specifies the distribution of the random variable X_r, i.e., $\pi_i = \text{Prob}(X_r = i)$, and the Markov matrices M_e for $e = (a_e, b_e) \in E$ give transition probabilities $M_e(i, j) = \text{Prob}(X_{b_e} = j \mid X_{a_e} = i)$ of the various state changes in passing from the parent vertex a_e to the child vertex b_e. Letting $\mathbf{X} = (X_a)_{a\in V}$ and $\mathbf{j} \in [k]^{|V|}$, the joint probability distribution at all nodes of T is thus

$$\text{Prob}(\mathbf{X} = \mathbf{j}) = \pi_{j_r} \prod_{e\in E} M_e(j_{a_e}, j_{b_e}).$$

By marginalizing over all variables at internal nodes of T, we obtain the *joint distribution, P, of states at the leaves of T*; if $\mathbf{k} \in [k]^{|L|}$ is an assignment of states to leaf variables, then

$$P(\mathbf{k}) = \sum_{\mathbf{m}\in[k]^{|V\smallsetminus L|}} \text{Prob}(\mathbf{X} = (\mathbf{k}, \mathbf{m})),$$

where $(\mathbf{k}, \mathbf{m})$ is an assignment of states to all the vertices of T compatible with $\mathbf{k}$. It is natural to view P as an n-dimensional $k \times \cdots \times k$ array, or *tensor*, with one index for each leaf of the tree.

For fixed T and choice of r, we use ψ_T to denote the *parameterization map*

$$\psi_T : \{\boldsymbol{\pi}, \{M_e\}_{e\in E}\} \mapsto P.$$

That the coordinate functions of ψ_T are polynomial is apparent, and essential to our work here. Note that we may naturally extend the domain of the polynomial map to larger sets, by dropping the nonnegativity assumptions in (i) and (ii) but retaining the condition that rows must sum to 1. We will consider *real parameters* and a real parameterization map, as well as *complex parameters* and a complex parameterization map. In contrast, we refer to the original probabilistic model as having *stochastic parameters*. Since the parameterization maps are all given by the same formula, we use ψ_T to denote them all, but will always indicate the current domain of interest.

The image of complex, real, or stochastic parameters under ψ_T is an n-dimensional $k\times\cdots\times k$ tensor, whose k^n entries sum to 1. When parameters are not stochastic, this tensor generally does not specify a probability distribution, as there can be negative or complex entries. We refer to any tensor whose entries sum to 1, regardless of whether the entries are complex, real, or nonnegative, as a *distribution*, but reserve the term *probability distribution* for a nonnegative distribution. With this language, the image of complex parameters under ψ_T is a distribution, but may or may not be a probability distribution. Similarly, while the matrix parameters M_e have rows summing to one even for complex parameters, we reserve the term *Markov* matrix exclusively for the stochastic setting.

2.2. Algebraic and semialgebraic model descriptions. Most previous algebraic analysis of the GM(k) model has focused on the *algebraic variety* associated to it for each choice of tree T. With this viewpoint one is essentially passing from the parameterization of the model, as given above, to an implicit description of the image of the parameterization as a zero set of certain polynomial functions, traditionally called *phylogenetic invariants* [11, 23, 5].

Whether one considers stochastic, real, or complex parameters, the collection of phylogenetic invariants for GM(k) on a tree T are the same. Thus they cannot distinguish probability distributions that arise from stochastic parameters from those arising from nonstochastic real or complex ones. To complicate matters further, there exist distributions that satisfy all phylogenetic invariants for the model on a given tree but are not even in the image of complex parameters. Though the algebraic issues behind this are well understood, they prevent classical algebraic geometry from being a sufficient tool to focus exclusively on the distributions of statistical interest.

To gain a more detailed understanding, we seek to refine the algebraic description of the model given by phylogenetic invariants into a *semialgebraic* description: to do this, we must supplement the polynomial equations holding on a superset of the image with polynomial inequalities sufficient to distinguish the stochastic image precisely.

Recall that a subset of $\mathbb{R}^n$ is called a *semialgebraic set* if it is a boolean combination (finite intersections and unions) of sets each of which is defined by a single polynomial equality or inequality. The Tarski–Seidenberg theorem [31, 26] implies that the image of a semialgebraic set under a polynomial map is also semialgebraic. Since for all T the stochastic parameter space of ψ_T is clearly semialgebraic, this implies that semialgebraic descriptions exist for the images of the ψ_T. Determining such descriptions explicitly is our goal.

2.3. Nonsingular parameters, positivity, and independence. Some of our results will be stated with additional mild conditions placed on the allowed parameters for the GM(k) model. We state these conditions here and explore their meaning.

Definition 2.1. *A choice* $\{\boldsymbol{\pi}, \{M_e\}_{e\in E}\}$ *of stochastic, real, or complex parameters for* GM(k) *on a tree* T *with root* r *is said to be* nonsingular *provided*

(i) *at every (hidden or observed) node* a*, the marginal distribution* $\mathbf{v}_a$ *of* X_a *has no zero entry, and*

(ii) *for every edge* e*, the matrix* M_e *is nonsingular.*

Parameters which are not nonsingular are said to be singular*.*

For stochastic parameters, the first condition in this definition can be replaced with a simpler one:

(i′) *the root distribution* $\boldsymbol{\pi}$ *has no zero entry.*

Statement (i) follows from (i′) and (ii) inductively, since if all entries of $\mathbf{v}_a$ are positive and $M_{(a,b)}$ is a nonsingular Markov matrix, then the distribution $\mathbf{v}_b = \mathbf{v}_a M_{(a,b)}$ at a child b of a has positive entries. However, for complex or real parameters requirement (i) is not implied by (i′) and (ii), as a simple example shows:

$$\mathbf{v}_a = (1/2, 1/2) \text{ and } M_{(a,b)} = \begin{pmatrix} s & 1-s \\ 2-s & s-1 \end{pmatrix} \tag{2.1}$$

are singular parameters since $\mathbf{v}_b = (1, 0)$, even though $\mathbf{v}_a$ has no zero entries and $M_{(a,b)}$ is a nonsingular for $s \neq 1$.

It is also natural to require that all numerical parameters of GM(k) on a tree T be strictly positive. This means that all states may occur at the root, and every state change is possible in passing along any edge of the tree. This assumption is plausible from a modeling point of view and can be desirable for technical statistical issues as well. Note that positivity of parameters does not ensure nonsingularity, since a Markov matrix may be singular despite all its entries being greater than zero. Similarly, nonsingularity of parameters does not ensure positivity since a nonsingular Markov matrix may have zero entries.

Given a joint probability distribution of random variables, two subsets of variables are *independent* when the marginal distribution for the union of the sets is the product of the marginal distributions for the two sets individually. We also use this term, in a nonstandard way, to apply to complex or real distributions when the same factorization holds.

To illustrate this usage, consider a tree T with two nodes, r, a, and one edge (r, a). For complex parameters $\boldsymbol{\pi}$ and $M_{(r,a)}$, the joint distribution of X_r and X_a is given by the matrix

$$P = \operatorname{diag}(\boldsymbol{\pi}) M_{(r,a)}.$$

Then the variables are independent exactly when P is a rank 1 matrix: $P = \boldsymbol{\pi}^T \mathbf{v}_a$. For $k = 2$ this occurs precisely when the parameters are singular. For $k > 2$, however, independence implies that the parameters are singular, but not vice versa. In general, singular parameters ensure that P has rank strictly less than k, but not that P has rank 1. These comments easily extend to larger trees to give the following.

Proposition 2.2. *Suppose* $P = \psi_T(\boldsymbol{\pi}, \{M_e\})$ *for a choice of complex GM(k) parameters on an* n*-leaf tree* T*. If the parameters are nonsingular, then there is no proper partition of the indices of* P *into independent sets. For* $k = 2$*, the converse also holds.*

That the converse is false for $k > 2$ is a complicating factor for the generalization of our results from the $k = 2$ case. Indeed, this is the reason we ultimately restrict to nonsingular parameters, avoiding a detailed analysis for all intermediate ranks $1 < \operatorname{rank}(P) < k$.

In closing this section, we note that for any $P \in \mathrm{Im}(\psi_T)$, there is an inherent and well-understood source of nonuniqueness of parameters giving rise to P, sometimes called "label-swapping." Since internal nodes of T are unobservable variables, the distribution P is computed by summing over all assignments of states to such variables. As a result, if the state names were permuted for such a variable, and corresponding changes made in numerical parameters, P would be left unchanged. At best, parameters leading to P can be determined only up to such permutations.

In the case of nonsingular parameters, label-swapping is the only source of nonuniqueness of parameters leading to P [13]. (See also [22].) This uniqueness result is used repeatedly in subsequent results. In contrast, for singular parameters there are additional sources of nonuniqueness beyond label-swapping.

2.4. Marginalizations, slices, group actions, and flattenings. Viewing probability distributions on n variables as n-dimensional tensors gives natural associations between statistical notions and tensor operations. For example, summing tensor entries over an index, or a collection of indices, corresponds to marginalizing over a variable, or collection of variables. Considering only those entries with a fixed value of an index, or collection of indices, corresponds (after renormalization) to conditioning on an observed variable, or collection of variables. Rearranging array entries into a new array, with fewer dimensions but larger size, corresponds to agglomerating several variables into a composite one with a larger state space. Here we introduce the necessary notation to formalize these tensor operations.

Definition 2.3. *For an n-dimensional $k \times \cdots \times k$ tensor P, integer $i \in [n]$, and vector $\mathbf{v} = (v_1, \cdots, v_k)$, define the $(n-1)$-dimensional tensor $P *_i \mathbf{v}$ by*

$$(P *_i \mathbf{v})(j_1, \ldots, \hat{j}_i, \ldots, j_n) = \sum_{j_i=1}^{k} P(j_1, \ldots, j_i, \ldots, j_n) v_{j_i},$$

where $\hat{}$ denotes omission. Similarly for a $k \times k$ matrix M, define the n-dimensional tensor $P *_i M$ by

$$(P *_i M)(j_1, \ldots, j_n) = \sum_{\ell=1}^{k} P(j_1, \ldots, j_{i-1}, \ell, j_{i+1}, \ldots, j_n) M(\ell, j_i).$$

The ℓth slice of P in the ith index *is defined by $P_{\cdots \ell \cdots} = P *_i \mathbf{e}_\ell$, where $\mathbf{e}_\ell$ is the ℓth standard basis vector, and the ith* marginalization of P *is defined by $P_{\cdots + \cdots} = P *_i \mathbf{1}$, where $\mathbf{1}$ is the vector of all* 1*s.*

When the above operations on a tensor are performed in different indices, they commute. This allows the use of n-tuple notation for the operation of matrices in all indices of a tensor, such as the following:

$$P \cdot (M_1, M_2, \ldots, M_n) = (\cdots((P *_1 M_1) *_2 M_2) \cdots) *_n M_n.$$

Although the M_i need not be invertible, restricting to that case gives the natural (right) group action of $GL(k, \mathbb{C})^n$ on $k \times \cdots \times k$ tensors. This generalizes the familiar operation on two-dimensional tensors P, i.e., on matrices, where

$$P \cdot (M_1, M_2) = (P *_1 M_1) *_2 M_2 = M_1^T P M_2.$$

If $\mathbf{v} \in \mathbb{C}^k$, then $\mathrm{Diag}(\mathbf{v})$ denotes the three-dimensional $k \times k \times k$ diagonal tensor whose only nonzero entries are the v_i in the (i, i, i) positions. That this notion is useful for the GM(k) model is made clear by the observation that for a 3-leaf star tree T, rooted at the central node,

$$\psi_T(\boldsymbol{\pi}, \{M_1, M_2, M_3\}) = \mathrm{Diag}(\boldsymbol{\pi}) \cdot (M_1, M_2, M_3). \tag{2.2}$$

If P is an n-dimensional $k \times \cdots \times k$ tensor and $[n] = A \sqcup B$ is a disjoint union of nonempty sets, then the *flattening of P with respect to this bipartition*, $\mathrm{Flat}_{A|B}(P)$, is the $k^{|A|} \times k^{|B|}$ matrix with rows indexed by $\mathbf{i} \in [k]^{|A|}$ and columns indexed by $\mathbf{j} \in [k]^{|B|}$, with

$$\mathrm{Flat}_{A|B}(P)(\mathbf{i}, \mathbf{j}) = P(\mathbf{k}),$$

where $\mathbf{k} \in [k]^n$ has entries matching those of $\mathbf{i}$ and $\mathbf{j}$, appropriately ordered. Thus the entries of P are simply rearranged into a matrix, in a manner consistent with the original tensor structure. When P specifies a joint distribution for n random variables, this flattening corresponds to treating the variables in A and B as two agglomerate variables, with state spaces the product of the state spaces of the individual variables. We assume throughout that all flattenings use some fixed orderings on the agglomerate state spaces but do not usually specify it explicitly since it plays no substantial role.

Notation such as $\mathrm{Flat}_{1|23}(P)$, for example, will be used to denote the matrix flattening obtained from a three-dimensional tensor using the partition of indices $A = \{1\}$, $B = \{2, 3\}$. If e is an edge in an n-leaf tree, then e naturally induces a bipartition of the leaves, by removing the edge and grouping leaves according to the resulting connected components. A flattening for such a bipartition is denoted by $\mathrm{Flat}_e(P)$.

Finally, we note that flattenings naturally occur in the notion of independence: If $[n] = A \sqcup B$, then the sets are independent precisely when $\mathrm{Flat}_{A|B}(P)$ is a rank 1 matrix.

3. GM(k) on 3-leaf trees.

In this section we derive a semialgebraic description of GM(k) on the 3-leaf tree, the smallest example of interest. Results for the 3-leaf tree also serve as a building block for the study of the model on larger trees in section 4. For this section, T is fixed, with leaves 1, 2, 3 and root r at the central node.

When $k = 2$, Cayley's hyperdeterminant plays a critical role, as was highlighted in [33]. Though our formulation will be different, we take the hyperdeterminant [12, 17, 14] as our starting point. For a $2 \times 2 \times 2$ tensor $A = (a_{ijk})$, the hyperdeterminant $\Delta(A)$ is

$$\begin{aligned}\Delta(A) = \big(&a_{111}^2 a_{222}^2 + a_{112}^2 a_{221}^2 + a_{121}^2 a_{212}^2 + a_{122}^2 a_{211}^2\big)\\ &- 2\big(a_{111}a_{112}a_{221}a_{222} + a_{111}a_{121}a_{212}a_{222} + a_{111}a_{122}a_{211}a_{222}\\ &\quad + a_{112}a_{121}a_{212}a_{221} + a_{112}a_{122}a_{221}a_{211} + a_{121}a_{122}a_{212}a_{211}\big)\\ &+ 4\big(a_{111}a_{122}a_{212}a_{221} + a_{112}a_{121}a_{211}a_{222}\big).\end{aligned}$$

The function Δ has the invariance property

$$\Delta(P \cdot (g_1, g_2, g_3)) = \det(g_1)^2 \det(g_2)^2 \det(g_3)^2 \Delta(P) \tag{3.1}$$

for $(g_1, g_2, g_3) \in GL(2, \mathbb{C})^3$. This fact, combined with a study of canonical forms for $GL(2, \mathbb{C})^3$-orbit representatives, leads to the following theorem.

THEOREM 3.1 (see [14, Theorem 7.1]). *A complex $2 \times 2 \times 2$ tensor P is in the $\mathrm{GL}(2, \mathbb{C})^3$-orbit of $D = \mathrm{Diag}(1, 1)$ if, and only if, $\Delta(P) \neq 0$. A real tensor is in the $\mathrm{GL}(2, \mathbb{R})^3$-orbit of D if, and only if, $\Delta(P) > 0$.*

Suppose that $k = 2$ and $P = \psi_T(\boldsymbol{\pi}, \{M_1, M_2, M_3\})$ arises from real nonsingular parameters on T. Then, (2.2) states $P = \mathrm{Diag}(\boldsymbol{\pi}) \cdot (M_1, M_2, M_3)$, but letting $M_1' = \mathrm{diag}(\boldsymbol{\pi})M_1$ we also have

$$P = D \cdot (M_1', M_2, M_3).$$

Thus P is in the $\mathrm{GL}(2, \mathbb{R})^3$-orbit of D, and by Theorem 3.1, $\Delta(P) > 0$. However, if P is a probability distribution satisfying $\Delta(P) > 0$, we cannot conclude that P arises from stochastic, or even real, nonsingular parameters; additional inequalities are needed for a semialgebraic model description.

Nonetheless, motivated by the role the hyperdeterminant plays in the semialgebraic description of the GM(2) model, in a separate work Allman et al. [1] construct generalizations of Δ for $k \geq 2$. These functions are defined by

$$f_i(P; \mathbf{x}) = \det(H_{\mathbf{x}}(\det(P *_i \mathbf{x}))),$$

where $\mathbf{x}$ is a vector of auxiliary variables, $H_{\mathbf{x}}$ denotes the Hessian operator, and $i \in \{1, 2, 3\}$. The next theorem establishes that the nonvanishing of these polynomials, in conjunction with the vanishing of some others, identifies the orbit of $\mathrm{Diag}(\mathbf{1})$, yielding an analogue of Theorem 3.1 for larger k.

THEOREM 3.2 (see [1]). *A complex $k \times k \times k$ tensor P lies in the $GL(k, \mathbb{C})^3$-orbit of* $\mathrm{Diag}(\mathbf{1})$ *if, and only if, for some $i \in \{1, 2, 3\}$, the following hold:*

(i) $(P *_i \mathbf{e}_j) \,\mathrm{adj}(P *_i \mathbf{x})(P *_i \mathbf{e}_\ell) - (P *_i \mathbf{e}_\ell) \,\mathrm{adj}(P *_i \mathbf{x})(P *_i \mathbf{e}_j) = 0$ *for all* $j, \ell \in [k]$. *Here* adj *denotes the classical adjoint, and equality means as a matrix of polynomials in* $\mathbf{x}$.

(ii) $f_i(P; \mathbf{x})$ *is not identically zero as a polynomial in* $\mathbf{x}$.

Moreover, if the enumerated conditions hold for one i, then they hold for all.

When $k > 2$, the $GL(k, \mathbb{C})^3$-orbit of $\mathrm{Diag}(\mathbf{1})$ is not dense among all $k \times k \times k$ tensors; rather its closure is a lower dimensional subvariety. This explains the necessity of the equalities in item (i). In the case $k = 2$, the equalities of (i) hold for all tensors and, in addition, $f_i(P; \mathbf{x}) = \Delta(P)$. Thus Theorem 3.2 includes the first statement of Theorem 3.1.

We emphasize that for $k > 2$, the functions f_i are *not* the ones usually referred to as hyperdeterminants [17], but rather a different generalization of Δ. Moreover, sign properties of $f_i(P; \mathbf{x})$ similar to that given in Theorem 3.1 for $\Delta(P)$ do not exist in general. See [1] for details.

With semialgebraic conditions ensuring that a tensor is in the $GL(k, \mathbb{C})^3$ orbit of $\mathrm{Diag}(\mathbf{1})$ in hand, we seek further conditions to ensure that it arises from nonsingular stochastic parameters. We proceed in two steps: first, we give requirements that a tensor is the image of complex parameters under ψ_T, and then that these parameters are nonnegative.

PROPOSITION 3.3. *Let P be a complex $k \times k \times k$ distribution. Then P is in the image of nonsingular complex parameters for GM(k) on the* 3*-leaf tree if, and only if, P is in the $GL(k, \mathbb{C})^3$-orbit of* $\mathrm{Diag}(\mathbf{1})$ *and* $\det(P *_i \mathbf{1}) \neq 0$ *for $i = 1, 2, 3$. Moreover, the parameters are unique up to label-swapping.*

Proof. To establish the reverse implication, suppose $P = \mathrm{Diag}(\mathbf{1}) \cdot (g_1, g_2, g_3)$ for some $g_i \in \mathrm{GL}(k, \mathbb{C})$, and let $\mathbf{r}^i = g_i \mathbf{1}$ denote the vector of row sums of g_i. A computation shows that

$$P *_3 \mathbf{1} = g_1^T \,\mathrm{diag}(\mathbf{r}^3) g_2.$$

Thus $\det(P *_3 \mathbf{1}) \neq 0$ is equivalent to the row sums of g_3 being nonzero, and similarly for the other g_i.

Now $M_i = \operatorname{diag}(\mathbf{r}^i)^{-1} g_i$ is a complex matrix with row sums equal to one. Letting $\boldsymbol{\pi} = (\prod_{i=1}^3 r_1^i, \dots, \prod_{i=1}^3 r_k^i)$ be the vector of entry-wise products of the $\mathbf{r}^i$, the entries of $\boldsymbol{\pi}$ are nonzero and

$$P = \operatorname{Diag}(\boldsymbol{\pi}) \cdot (M_1, M_2, M_3).$$

Since P is a distribution,

$$\begin{aligned} 1 &= ((P *_1 \mathbf{1}) *_2 \mathbf{1}) *_3 \mathbf{1} \\ &= (((\operatorname{Diag}(\boldsymbol{\pi}) \cdot (M_1, M_2, M_3)) *_1 \mathbf{1}) *_2 \mathbf{1}) *_3 \mathbf{1} \\ &= ((\operatorname{Diag}(\boldsymbol{\pi}) *_1 M_1\mathbf{1}) *_2 M_2\mathbf{1}) *_3 M_3\mathbf{1} \\ &= ((\operatorname{Diag}(\boldsymbol{\pi}) *_1 \mathbf{1}) *_2 \mathbf{1}) *_3 \mathbf{1} \\ &= \boldsymbol{\pi} \cdot \mathbf{1}, \end{aligned}$$

so $\boldsymbol{\pi}$ is a valid complex root distribution. Thus, P is in the image of ψ_T for complex, nonsingular parameters.

The forward implication in the theorem is straightforward. ∎

Combining this proposition with Theorems 3.1 and 3.2, we obtain the following.

COROLLARY 3.4. *A $k \times k \times k$ complex distribution P is the image of complex, nonsingular parameters for GM(k) on the* 3*-leaf tree if, and only if, it satisfies the semialgebraic conditions* (i) *and* (ii) *of Theorem* 3.2 *and*

(iii) *for $i = 1, 2, 3$, $\det(P *_i \mathbf{1}) \neq 0$.*

For $k = 2$, P is the image of real nonsingular parameters for GM(2) *on the* 3*-leaf tree if, and only if, it satisfies $\Delta(P) > 0$ and the semialgebraic conditions* (iii).

The key to ensuring that nonsingular parameters are stochastic is the construction of certain quadratic forms whose positive semidefiniteness (respectively, definiteness) encodes nonnegativity (respectively, positivity) of the numerical parameters. A classic result of Sylvester [30], in the form of sign conditions on minors, then gives a semialgebraic version of these conditions. Since our goal is an explicit semialgebraic description, we state the theorem for reference.

Recall that a *principal minor* of a matrix is the determinant of a submatrix chosen with the same row and column indices, and that a *leading* principal minor is one of these where the chosen indices are $\{1, 2, 3, \dots, k\}$ for any k.

THEOREM 3.5 (Sylvester's theorem). *Let A be an $n \times n$ real symmetric matrix and $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ the associated quadratic form on $\mathbb{R}^n$. Then*

1. *Q is positive semidefinite if, and only if, all principal minors of A are nonnegative, and*
2. *Q is positive definite if, and only if, all leading principal minors of A are strictly positive.*

In conjunction with Sylvester's theorem, the following gives a semialgebraic characterization of our model of interest.

THEOREM 3.6. *A $k \times k \times k$ probability distribution P is the image of nonsingular stochastic parameters for the GM(k) model on the* 3*-leaf tree if, and only if, conditions* (i), (ii), *and* (iii) *of Theorem* 3.2 *and Corollary* 3.4 *are satisfied, and*

(iv) *the matrix*

$$\det(P_{\cdot\cdot+}) P_{+\cdot\cdot}^T \operatorname{adj}(P_{\cdot\cdot+}) P_{\cdot+\cdot} \tag{3.2}$$

is positive definite and the matrices

$$\begin{aligned} &\det(P_{\cdot\cdot+})\, P_{i\cdot\cdot}^T \,\mathrm{adj}(P_{\cdot\cdot+})P_{\cdot+\cdot} \quad \text{for } i=1,\dots,k,\\ &\det(P_{\cdot\cdot+})P_{+\cdot\cdot}^T \mathrm{adj}(P_{\cdot\cdot+})P_{\cdot i\cdot} \quad \text{for } i=1,\dots,k,\\ &\det(P_{+\cdot\cdot})P_{\cdot+\cdot}\mathrm{adj}(P_{+\cdot\cdot})P_{\cdot\cdot i}^T \quad \text{for } i=1,\dots,k. \end{aligned} \tag{3.3}$$

are all positive semidefinite.

Moreover, the probability distribution P is the image of nonsingular positive parameters if, and only if, conditions (i), (ii), *and* (iii) *are satisfied and*

(iv′) *all of the matrices in* (3.2) *and* (3.3) *are positive definite.*

In both cases, the nonsingular parameters are unique up to label-swapping.

Proof. Let P be an arbitrary $k \times k \times k$ probability distribution. By Corollary 3.4, the first 3 conditions are equivalent to $P = \psi_T(\boldsymbol{\pi}, \{M_1, M_2, M_3\})$ for complex nonsingular parameters. We need to show the addition of assumption (iv) is equivalent to parameters being nonnegative.

Note that

$$\begin{aligned} P_{\cdot\cdot+} &= P *_3 \mathbf{1} = M_1^T \,\mathrm{diag}(\boldsymbol{\pi}) M_2,\\ P_{\cdot+\cdot} &= P *_2 \mathbf{1} = M_1^T \,\mathrm{diag}(\boldsymbol{\pi}) M_3,\\ P_{+\cdot\cdot} &= P *_1 \mathbf{1} = M_2^T \,\mathrm{diag}(\boldsymbol{\pi}) M_3. \end{aligned}$$

Since $P_{\cdot\cdot+}$ is nonsingular, the product

$$P_{+\cdot\cdot}^T P_{\cdot\cdot+}^{-1} P_{\cdot+\cdot} = M_3^T \,\mathrm{diag}(\boldsymbol{\pi}) M_3 \tag{3.4}$$

is symmetric and the matrix of a real quadratic form. Similarly, using slices, additional quadratic forms are defined by the symmetric matrices

$$\begin{aligned} P_{i\cdot\cdot}^T P_{\cdot\cdot+}^{-1} P_{\cdot+\cdot} &= M_3^T \,\mathrm{diag}(\boldsymbol{\pi})\Lambda_{1,i} M_3,\\ P_{+\cdot\cdot}^T P_{\cdot\cdot+}^{-1} P_{\cdot i\cdot} &= M_3^T \,\mathrm{diag}(\boldsymbol{\pi})\Lambda_{2,i} M_3,\\ P_{\cdot+\cdot} P_{+\cdot\cdot}^{-1} P_{\cdot\cdot i}^T &= M_1^T \,\mathrm{diag}(\boldsymbol{\pi})\Lambda_{3,i} M_1, \end{aligned} \tag{3.5}$$

where $\Lambda_{j,i} = \mathrm{diag}(M_j \mathbf{e}_i)$ is the diagonal matrix with entries from the ith column of M_j. Multiplying all these matrices by the square of an appropriate nonzero determinant clears denominators and preserves signs yielding the matrices in (3.2) and (3.3). It follows that condition (iv) is necessary if P arises from nonsingular stochastic parameters.

It remains to show that if the matrices in (3.2) and (3.3) are positive definite and positive semidefinite as in (iv), then all the M_i are real and stochastic, and $\boldsymbol{\pi}$ has positive entries. Letting $\mathbf{r}_{ij}$ denote the jth row of matrix M_i, by Kruskal's theorem [21, 22] the summands in

$$P = \sum_{j=1}^{k} \pi_j \,\mathbf{r}_{1j} \otimes \mathbf{r}_{2j} \otimes \mathbf{r}_{3j}$$

are uniquely determined, up to order. Since P is real, $P = \overline{P}$ and this uniqueness implies that any complex summands occur in conjugate pairs. It follows easily that we may assume only the first 2ℓ rows of all M_i are complex and that $\mathbf{r}_{i,2j-1} = \overline{\mathbf{r}_{i,2j}}$ and $\pi_{2j-1} = \overline{\pi_{2j}}$ for $j = 1, \dots, \ell$.

We argue next that the positive definiteness of the matrix in (3.2), or equivalently of that in (3.4), implies the M_i are real, and the entries of $\boldsymbol{\pi}$ are positive. To this end, suppose the first two rows of M_3 are complex (nonreal) conjugates, and consider any nonzero $\mathbf{v} \in \mathbb{R}^k$, orthogonal to the real and imaginary parts of the last $k-2$ rows of M_3. Then

$$Q(\mathbf{v}) = \mathbf{v}^T \left(M_3^T \operatorname{diag}(\boldsymbol{\pi}) M_3\right) \mathbf{v} = \pi_1 \left(\mathbf{r}_{31} \cdot \mathbf{v}\right)^2 + \overline{\pi}_1 \left(\overline{\mathbf{r}}_{31} \cdot \mathbf{v}\right)^2.$$

If $\mathbf{v}$ is chosen to also be orthogonal to $\Im(\mathbf{r}_{31})$, then $Q(\mathbf{v}) = 2\,\Re(\pi_1)\,(\Re(\mathbf{r}_{31}) \cdot v)^2$, and since Q is positive definite we find $\Re(\pi_1) > 0$. If instead $\mathbf{v}$ is chosen to be orthogonal to $\Re(\mathbf{r}_{31})$, then $Q(\mathbf{v}) = -2\,\Re(\pi_1)\,(\Im(\mathbf{r}_{31}) \cdot v)^2$, implying $\Re(\pi_1) < 0$. This contradiction establishes that M_3 is real. Since M_3 is nonsingular, it therefore has no conjugate rows. Thus all M_i and $\boldsymbol{\pi}$ are real. Moreover, since Q is positive definite and M_3 is real, the entries of $\boldsymbol{\pi}$ are positive.

Finally, consider the matrices from (3.3). Since $\boldsymbol{\pi}$ has positive entries and M_i are real, the associated real quadratic forms are all positive semidefinite only if the entries of the M_i are all nonnegative. Since the parameters are nonsingular, the only source of nonuniqueness is label-swapping.

The second statement of the theorem, about positive parameters, requires only a minor modification to this argument. ∎

Remark. The $j \times j$ minors of the matrices in (3.2) and (3.3) are polynomials in the entries of P of degree $j(2k+1)$ with $j = 1, \dots, k$. However, as the leading determinant in the products defining those matrices is real and nonzero, one can remove an even power of it without affecting the sign of the minors. Thus a polynomial inequality of degree $j(2k+1)$ can be replaced by one of lower degree, $j(k+1) + e_j k$, where $e_j = 0$ or 1 is the parity of j.

We illustrate Theorem 3.6 by considering the probability distribution with exact rational entries given by

$$P = \left[\begin{array}{ccc|ccc|ccc} 0.1500 & 0.0130 & 0.105\overline{3} & 0.0130 & 0.0050 & 0.015\overline{3} & 0.105\overline{3} & 0.015\overline{3} & 0.077\overline{6} \\ 0.0130 & 0.0050 & 0.015\overline{3} & 0.0050 & 0.0090 & 0.009\overline{3} & 0.015\overline{3} & 0.009\overline{3} & 0.018\overline{6} \\ 0.105\overline{3} & 0.015\overline{3} & 0.077\overline{6} & 0.015\overline{3} & 0.009\overline{3} & 0.018\overline{6} & 0.077\overline{6} & 0.018\overline{6} & 0.0620 \end{array}\right],$$

which by checking conditions (i-iii) above arises from nonsingular parameters. Considering the symmetric nonnegative matrix $Q = \det(P_{\cdot\cdot+})\, P_{1\cdot\cdot}^T \operatorname{adj}(P_{\cdot\cdot+}) P_{\cdot+\cdot}$, a computation shows that $\det(Q) < 0$ so that by Sylvester's theorem, Q does not define a positive semidefinite quadratic form, and P does not arise from stochastic parameters. Indeed, P was constructed from a stochastic vector $\boldsymbol{\pi}$, but complex matrices M_i.

In the case of the 2-state model, the above result can be made more complete by also explicitly describing the image of singular parameters. Such results are not new—see [27, 8, 32, 20] and for related work [24]—but much of the success of these previous analyses has been derived by careful consideration of statistical interpretations of particular quantities computed from a probability distribution P (e.g., covariances, conditional covariances, moments, tree cumulants). Such statistical interpretations, however, are specific to the binary model and do not generalize to larger state spaces. We include a novel proof here, based on our viewpoint.

THEOREM 3.7. *A probability distribution P is in the image of the stochastic parameterization map ψ_T for the GM(2) model on the 3-leaf tree if, and only if, one of the following occur:*

1. $\Delta(P) > 0$, $\det(P *_i \mathbf{1}) \neq 0$ *for* $i = 1, 2, 3$, *the matrix of* (3.2) *is positive definite, and the six matrices of* (3.3) *are positive semidefinite. In this case, P is the image of unique (up to label-swapping) nonsingular parameters.*

2. $\Delta(P) = 0$, *and all* 2×2 *minors of at least one of the matrices* $\mathrm{Flat}_{1|23}(P)$, $\mathrm{Flat}_{2|13}(P)$, $\mathrm{Flat}_{3|12}(P)$ *are zero. In this case,* P *arises from singular parameters. If* P *has all positive entries, then it is the image of infinitely many singular stochastic parameter choices.*

Proof. Using Theorem 3.6 and the comments immediately following Theorem 3.2, case 1 is already established under the weaker condition that $\Delta(P) \neq 0$. However, since the parameters are nonsingular when $\Delta(P) \neq 0$ and real when the conditions of case 1 are satisfied, by Theorem 3.1 we may assume equivalently that $\Delta(P) > 0$.

To establish case 2, first assume $P = \psi_T(\boldsymbol{\pi}, \{M_1, M_2, M_3\})$ is the image of singular stochastic parameters. Then certainly P has nonnegative entries summing to 1, and by (2.2) and (3.1), $\Delta(P) = 0$. Now

$$\mathrm{Flat}_{1|23}(P) = M_1^T \operatorname{diag}(\boldsymbol{\pi}) M,$$

where M is the 2×4 matrix with rows obtained by taking the tensor product of corresponding rows of M_2 and M_3. (That is, $M(i,(j,k)) = M_2(i,j)M_3(i,k)$.) Thus this flattening has rank 1 if $\boldsymbol{\pi}$ has a zero entry or M_1 has rank 1. Similar products for the other flattenings show that singular parameters imply that at least one of the flattenings $\mathrm{Flat}_{1|23}(P), \mathrm{Flat}_{2|13}(P), \mathrm{Flat}_{3|12}(P)$ has rank 1, and hence its 2×2 minors vanish.

Conversely, suppose $\Delta(P) = 0$ and at least one of the flattenings has vanishing 2×2 minors, and hence rank 1. Then by the classification of orbits given in [14, Table 7.1], P is in the $GL(2,\mathbb{R})^3$-orbit of one of the following four tensors: the tensor $\mathrm{Diag}(1,0)$ (in which case all three flattenings have rank 1) or one of the 3 tensors with parallel slices I and the zero matrix (in which case exactly one of the flattenings has rank 1).

If $P = \mathrm{Diag}(1,0) \cdot (g_1, g_2, g_3)$, then $P(i,j,k) = g_1(1,i)g_2(1,j)g_3(1,k)$. Since the entries of P are nonnegative and sum to 1, one sees the top rows of each g_i can be chosen to be nonnegative, summing to 1. The bottom row of each g_i can also be replaced with any nonnegative row summing to 1 that is independent of the top row. Taking $\boldsymbol{\pi} = (1,0)$, this gives us infinitely many choices of singular stochastic parameters giving rise to P. Alternatively, one could choose each Markov matrix to have two identical rows, and any $\boldsymbol{\pi}$ with nonzero entries to obtain other singular stochastic parameters leading to P.

For the remaining cases assume, without loss of generality, that $P = E \cdot (g_1, g_2, g_3)$, where $E_{\cdot\cdot 1} = I$, and $E_{\cdot\cdot 2}$ is the zero matrix. Then $P_{\cdot\cdot 1} = g_3(1,1)(g_1^T g_2)$ and $P_{\cdot\cdot 2} = g_3(1,2)(g_1^T g_2)$. Since the entries of P are nonnegative and add to 1, we may assume that the top row of g_3 is also nonnegative and adds to 1. Choose M_3 to have two identical rows matching the top row of g_3. Now $P_{\cdot\cdot +} = g_1^T g_2$ is a rank-2 nonnegative matrix with entries adding to 1. Such a matrix can be written in the form $P_{\cdot\cdot +} = M_1^T \operatorname{diag}(\boldsymbol{\pi}) M_2$ with, for instance, $M_1 = I$, $\boldsymbol{\pi} = P_{\cdot ++}$, $M_2 = \operatorname{diag}(\boldsymbol{\pi})^{-1} P_{\cdot\cdot +}$. Then one has $P = \psi_T(\boldsymbol{\pi}, \{M_1, M_2, M_3\})$. If P has positive entries, one may also choose M_1 sufficiently close to I so that M_1, $\boldsymbol{\pi} = (M_1^T)^{-1} P_{\cdot ++}$, $M_2 = \operatorname{diag}(\boldsymbol{\pi})^{-1} \left(M_1^T\right)^{-1} P_{\cdot\cdot +}$ all have nonnegative entries, thus obtaining infinitely many singular parameter choices leading to P. (The example of $P = (1/2)E$ shows that with only nonnegative entries there may be only finitely many singular parameter choices leading to P.) ∎

Remark. The analysis of the singular parameter case in this proof, by appealing without explanation to [14, Table 7.1], has not made explicit the importance of the notion of *tensor rank*. Indeed, that concept is central to both [14] and [1] and thus plays a crucial behind-the-scenes role in this work as well. The first singular case, a tensor in the orbit of $\mathrm{Diag}(1,0)$, is of tensor rank 1, while the second, a tensor in the orbit of E, is of tensor rank 2 yet multilinear rank $(2,2,1)$. The nonsingular case is those of tensor rank 2 and multilinear rank $(2,2,2)$.

Remark. The equality $\Delta(P)=0$ appearing in case 2 of Theorem 3.7 is redundant in that it is implied by the vanishing of the minors of any one of the flattenings. Moreover, these quadratics simply express that at least one of the leaf variables is independent of the others.

Minor modifications to the above argument extend the result to positive parameters.

THEOREM 3.8. *A probability distribution P is in the image of the positive stochastic parameterization map for the GM(2) model on the 3-leaf tree if, and only if, the conditions of Theorem 3.7 are met with the following modification to case 1: all of the matrices are positive definite.*

Proof. This is straightforward replacing "nonnegative" with "positive" in the argument of Theorem 3.7, though in case 2, if P is in the orbit of $\mathrm{Diag}(1,0)$, then one must use the second construction of singular parameters. ∎

4. GM(k) on n-leaf trees. We now extend the results of the previous sections to n-leaf trees for $n>3$. To vary the choice of the root node of the tree in our arguments, we need the following. Similar lemmas are given in [28, Theorem 2] and [2, Proposition 1].

LEMMA 4.1. *Suppose stochastic parameters are given for the GM(k) model on a tree T with the root located at a specific node of T. Then there are stochastic parameters for T rooted at any other node of T, or at a node of valence 2 introduced along an edge of T, which lead to the same distribution. Moreover, if the original parameters were nonsingular and/or positive, then so are the new ones.*

Note that for nonstochastic parameters, Lemma 4.1 fails to hold as the examples in (2.1) show. The problem is simply that a column sum of $P=\mathrm{diag}(\boldsymbol{\pi})M$ can be zero though the column is not the zero vector, so one cannot solve for a factorization of $P=\widetilde{M}^T\,\mathrm{diag}(\widetilde{\boldsymbol{\pi}})$ as needed for moving the root from one vertex of an edge to the other. For real and complex parameters, moving the root is possible under the more stringent hypothesis that the parameters are nonsingular.

LEMMA 4.2. *Suppose real or complex nonsingular parameters are given for the GM(k) model on a tree T with the root located at a specific node of T. Then there are nonsingular parameters for T rooted at any other node of T, or at a node of valence 2 introduced along an edge of T, which lead to the same distribution.*

We now show that independent subsets of variables allow the question of determining if a distribution arises from parameters on a tree to be "decomposed" into the same question for the marginalizations to the subsets.

PROPOSITION 4.3. *Let P be a joint distribution of a set L of k-state variables such that for some partition $L_1|L_2|\cdots|L_s$ of L, the variable sets L_i and L_j are independent for all $i\neq j$. Suppose the marginal distribution of each L_i arises from nonsingular GM(k) parameters on a tree T_i. Then P arises from GM(k) parameters on any tree T which can be obtained by connecting the trees $T_1,T_2,\dots,T_s$ by the introduction of new edges between them, with endpoints possibly subdividing either edges of the T_i or previously introduced edges joining some of the T_i.*

If, in addition, the parameters for each of the marginal distributions L_i are stochastic, then P arises from stochastic parameters for the GM(k) model on an $|L|$-leaf tree. If the parameters for the $|L_i|$-leaf trees are positive, then so are those on the $|L|$-leaf tree.

Note that the converse of this statement—that if P arises from parameters for the GM(k) model on an $|L|$-leaf tree, then the marginal distributions of each L_i arise from parameters for the GM(k) model on an $|L_i|$-leaf tree—is well known and does not require the independence of the variable sets, nor nonsingularity of parameters.

Proof. It is enough to consider a partition of L into two independent subsets, $L_1|L_2$. Let T be any tree formed by connecting T_1 and T_2 by a single edge, possibly with endpoints introduced to subdivide edges of one or both of the T_i. If $e = (r_1, r_2)$ is the edge joining T_1 and T_2, with r_i in T_i, then under the nonsingularity assumption by Lemma 4.2 we may assume that parameters on T_1 and T_2 are given for roots r_1 and r_2. We root T at r_1 and then specify parameters on T as the root distribution $\boldsymbol{\pi}_1$ for T_1, all matrix parameters on the edges of T_1 and T_2, and for the edge e the matrix $M_e = \mathbf{1}\boldsymbol{\pi}_2$, where $\boldsymbol{\pi}_2$ is the root distribution on T_2.

Let $\tilde{P}$ denote the image of these parameters under ψ_T. The edge e of T induces the split $L_1|L_2$ of the leaf variables, and flattening with respect to e gives $\mathrm{Flat}_e(\tilde{P}) = A^T C B$ where A, B are $k \times k^{|L_1|}$ and $k \times k^{|L_2|}$ matrices depending only on the matrix parameters on the subtrees T_1 and T_2, and

$$C = \mathrm{diag}(\boldsymbol{\pi}_1) M_e = \mathrm{diag}(\boldsymbol{\pi}_1)\mathbf{1}\boldsymbol{\pi}_2 = \boldsymbol{\pi}_1^T \boldsymbol{\pi}_2.$$

Indeed, in the stochastic case, A gives probabilities of observations at the leaves L_1 conditioned on the state at r_1, B gives probabilities of observations at the leaves L_2 conditioned on the state at r_2, and C is a matrix giving the joint distribution of states at r_1 and r_2. Observing that $A^T C B = (\boldsymbol{\pi}_1 A)^T (\boldsymbol{\pi}_2 B)$, independence implies that $\tilde{P}$ is the product of the same marginal distributions on L_1 and L_2 as P, and hence $\tilde{P} = P$.

For stochastic parameters, the result follows by the same argument, but using Lemma 4.1 to move the root in T while preserving nonsingularity and/or positivity of parameters. ∎

By this proposition, the only sets we must understand to build a semialgebraic description for the full n-leaf stochastic model are the image of parameters for m-leaf trees, $m \leq n$, when no subsets of the m leaf variables are independent. In the case $k = 2$, by Proposition 2.2, this is precisely the images of nonsingular parameters. For $k > 2$, where this equivalence does not hold, we give a semialgebraic description for nonsingular parameters, avoiding the complications of intermediate ranks.

PROPOSITION 4.4. *Let P be an n-dimensional $k \times k \times \cdots \times k$ distribution with $n \geq 3$. Then P arises from nonsingular complex parameters on a binary tree T if, and only if,*

(i) *all marginalizations of P to 3 variables arise from nonsingular parameters on the induced 3-leaf, 3-edge trees, and*

(ii) *for all internal edges e of T, all $(k+1) \times (k+1)$ minors of the matrix flattening $Flat_e(P)$ are 0.*

Moreover, such nonsingular parameters are unique up to label-swapping at internal nodes of T.

Note that condition (i) can be stated in terms of explicit semialgebraic conditions, using Corollary 3.4. Also, the polynomial equalities of condition (ii) are usually called *edge invariants* [6].

Proof. For the forward implication, condition (i) follows since marginalizations arise from the model on the associated induced subtree, using Markov matrices that are products of the original ones. Item (ii) is from [6], where it is shown that all $P \in \mathrm{Im}(\psi_T)$ satisfy the edge invariants. (The nonsingularity of parameters is not required for either of these.)

For the reverse implication, we proceed by induction on the size n of the variable set L. The claim holds by assumption in the base case of $n = 3$. Assume the statement is true for fewer than $n \geq 4$ variables. We identify leaves of T with the variables associated to them. Choose some internal edge $e_0 = (a, b)$ of T, corresponding to the split $L_1|L_2$ of L, with $|L_1|, |L_2| \geq 2$, a in the subtree spanned by L_1, and b in the subtree spanned by L_2. Introducing a vertex c subdividing (a, b), let T_1 be the subtree with leaves $L_1 \cup \{c\}$ and T_2 the subtree with leaves $L_2 \cup \{c\}$. Thus (a, c) in T_1 and (b, c) in T_2 are the edges formed from dividing (a, b).

Since the edge invariants are satisfied by P, $\mathrm{Flat}_{e_0}(P)$ has rank at most k. Therefore, there exist $k^{|L_1|} \times k$ and $k \times k^{|L_2|}$ matrices A, B, both of rank at most k, with

$$\mathrm{Flat}_{e_0}(P) = AB.$$

Choose a single variable $\ell_2 \in L_2$ and let Q denote the marginalization of P to $L_1 \cup \{\ell_2\}$. Then there is a $k^{|L_2|} \times k$ matrix N such that

$$\mathrm{Flat}_{e_0}(P)N = ABN = \mathrm{Flat}_{e_1}(Q),$$

where this last flattening is along the edge $e_1 = (a, \ell_2)$ in the induced subtree on $L_1 \cup \{\ell_2\}$. Stated differently, multiplication by N marginalizes over all those leaves in L_2 except ℓ_2.

Since Q also satisfies conditions (i) and (ii), by the inductive hypothesis Q arises from nonsingular parameters. Moreover, we see that $\mathrm{Flat}_{e_1}(Q)$ has rank k, since marginalization over all but one variable in L_1 is seen to produce a rank k matrix from the nonsingular parameterization. It follows that the $k \times k$ matrix BN has rank k. Replacing A and B with AC and $C^{-1}B$ for some invertible $k \times k$ matrix C, we may further assume the rows of BN add to 1.

Now since Q arises from nonsingular parameters on a $(|L_1|+1)$-leaf tree isomorphic to T_1 rooted at a, we claim that $Q' = Q *_{\ell_2} (BN)^{-1}$ arises from nonsingular complex parameters on T_1 for some suitable choice of B. Indeed, Q' arises from the same parameters as Q, except that on the edge (a, c) we use the matrix parameter that is the product of the one on the edge leading to ℓ_2 and $(BN)^{-1}$. Since $(BN)^{-1}$ is a nonsingular matrix with rows summing to one, the only condition to check is that the marginalization of the resulting distribution to c has no zero entries. But this marginalization is $\mathbf{v}_c = \mathbf{v}_{\ell_2}(BN)^{-1}$ and has a zero entry only if $\mathbf{v}_{\ell_2}$ is in the left nullspace of one (or more) of the columns of $(BN)^{-1}$. However, replacing A and B with AC and $C^{-1}B$ for some appropriate nonsingular matrix C whose rows sum to one, we can ensure that $\mathbf{v}_c$ has no zero entries.

Since the parameters producing Q' are nonsingular, by Lemma 4.2 we may reroot T_1 at c with parameters the root distribution $\mathbf{v}_c$, matrices $\{M_e\}$ on all edges of T_1 corresponding to ones in T, and matrix $M_{(c,a)}$ on the edge (c, a).

Now with K the matrix which marginalizes $\mathrm{Flat}_{e_0}(P)$ over all elements of L_1 but one, say ℓ_1, we see

$$K\,\mathrm{Flat}_{e_0}(P) = KAB = \mathrm{Flat}_{e_2}(U),$$

where U is the marginalization of P over the same elements of L_1 and the last flattening is on $e_2 = (b, \ell_1)$ in the induced subtree, which is isomorphic to T_2. But by induction U arises from nonsingular parameters on T_2 rooted at b. Let M be the product of the matrix parameters on the edges in the path from c to ℓ_1 in T_1. Then $U' = U *_{\ell_1} M^{-1}$ also arises from nonsingular parameters on T_2 (checking that its marginalization to c is $\mathbf{v}_{\ell_1} M^{-1} = \mathbf{v}_c$, which has no zeros by construction).

Note now that U' has flattening $(M^{-1})^T KAB$. But $(M^{-1})^T KA = \operatorname{diag}(\mathbf{v}_c)$ by construction. Thus $\operatorname{diag}(\mathbf{v}_c)B$ is the $c|L_2$ flattening of a tensor arising on T_2 from nonsingular complex parameters. With the root at c, let M_e be the Markov parameters for all edges of T_2 corresponding to ones in T, and $M_{(c,b)}$ the Markov matrix on (c,b). The root distribution $\mathbf{v}_c$ is the same as for T_1.

It remains to check that P is the image of the parameters on T with subdivided edges (c,a) and (c,b) rooted at c given by $\mathbf{v}_c$, $\{M_e\}_{e\neq(a,b)}$, and $M_{(c,a)}$ and $M_{(c,b)}$. But these parameters lead to the distributions Q' and U' on T_1 and T_2, respectively. Since $\operatorname{Flat}_{(a,c)}(Q') = A$ while $\operatorname{Flat}_{(c,b)}(U') = \operatorname{diag}(\mathbf{v}_c)B$, the equation $\operatorname{Flat}_e(P) = AB$ shows they produce P on T.

That the parameters are unique, up to label-swapping at the internal nodes of T, follows from the 3-leaf case. ∎

Note that in establishing the reverse implication in Proposition 4.4 we did not use condition (i) for every 3-variable marginalization. Informally, given a tree T one could choose a sequence of edges which can be successively cut (by the introduction of the node c in the inductive proof above) to produce a forest of 3-taxon trees. Then condition (i) is only needed for a subset of the 3-leaf marginalizations, determined by the sequence of edges chosen to cut and the choice of the variables denoted ℓ_1, ℓ_2 in the proof. Similarly, not all edge flattenings of condition (ii) are used: For the first edge to be cut, one uses the full edge flattening, but after that, only edge flattenings of marginalizations to subsets of variables are needed. Thus the full set of conditions given in this proposition is actually equivalent to a subset of them.

Supposing now that an n-dimensional distribution P arises from nonsingular complex parameters on a binary tree T, we wish to give semialgebraic conditions that are satisfied if, and only if, the parameters are stochastic. By considering only marginalizations to 3 variables and appealing to Theorem 3.6, we can give conditions that hold precisely when the root distribution and products of matrix parameters along any path leading from an interior vertex of T to leaves are stochastic. This immediately yields semialgebraic conditions that the root distribution and matrix parameters on terminal edges are stochastic. However, additional criteria are needed to ensure that matrix parameters on interior edges are stochastic. Adopting the convention that a 4-leaf tree is labeled by the partition of leaves induced by its single internal edge, in the next proposition we give conditions for stochastic parameters on a 4-leaf tree.

PROPOSITION 4.5. *Suppose a distribution P arises from nonsingular complex parameters for $GM(k)$ on the 4-leaf tree $12|34$. If the 3-marginalizations $P_{\cdots+}$ and $P_{+\cdots}$ arise from stochastic parameters and, in addition, the $k^2 \times k^2$ matrix*

$$\det(P_{+\cdot\cdot+})\det(P_{\cdot+\cdot+})\operatorname{Flat}_{13|24}\left(P *_2 (\operatorname{adj}(P^T_{+\cdot\cdot+})P^T_{\cdot+\cdot+})) *_3 (\operatorname{adj}(P_{\cdot+\cdot+})P_{\cdot++\cdot})\right) \tag{4.1}$$

is positive semidefinite, then P arises from stochastic parameters.

Note that the matrix in (4.1) could be replaced by ones where the roles of leaves 1 and 2 or of leaves 3 and 4 have been interchanged.

Proof. Root T at the interior node near leaves 1 and 2. Let M_i, $i = 1,2,3,4$, be the complex matrix parameter with row sums equal to one on the edge leading to leaf i, M_5 the matrix parameter on the internal edge, and $\boldsymbol{\pi}$ the root distribution. Define the matrices

$$N_{32} = P^T_{+\cdot\cdot+} = M_3^T M_5^T \operatorname{diag}(\boldsymbol{\pi}) M_2,$$
$$N_{31} = P^T_{\cdot+\cdot+} = M_3^T M_5^T \operatorname{diag}(\boldsymbol{\pi}) M_1.$$

Then

$$\overline{P} = P *_2 N_{32}^{-1} N_{31}$$

is a distribution arising from the same parameters as P except that M_2 has been replaced with M_1, so that the same matrix parameter is used on the edges leading to leaves 1 and 2.

Similarly with

$$N_{14} = P_{\cdot++\cdot} = M_1^T \operatorname{diag}(\boldsymbol{\pi}) M_5 M_4,$$
$$N_{13} = P_{\cdot+\cdot+} = M_1^T \operatorname{diag}(\boldsymbol{\pi}) M_5 M_3,$$

then

$$\widetilde{P} = \overline{P} *_3 N_{13}^{-1} N_{14} \tag{4.2}$$

is a distribution arising from the same parameters as $\overline{P}$ except that M_3 has been replaced with M_4.

Consider now the 13|24 flattening of $\widetilde{P}$, a flattening which is *not* consistent with the topology of the underlying tree. As shown in [4], this can be expressed as a product of $k \times k$ matrices

$$\operatorname{Flat}_{13|24}(\widetilde{P}) = A^T D A, \tag{4.3}$$

where D is the diagonal matrix with the k^2 entries of $\operatorname{diag}(\boldsymbol{\pi}) M_5$ on its diagonal, and $A = M_1 \otimes M_4$ is the Kronecker product. Because M_1, M_4 are nonsingular, so is A. Since conditions on 3-marginals ensure that $\boldsymbol{\pi}$ has positive entries, we can ensure M_5 has nonnegative entries by requiring that $\operatorname{Flat}_{13|24}(\widetilde{P})$ be positive semidefinite. Since the resulting inequalities would involve rational expressions, due to the inverses of matrices, we first multiply $\operatorname{Flat}_{13|24} \widetilde{P}$ by squares of nonzero determinants to remove denominators. ∎

Together with Theorems 3.6 and 3.7, the last two propositions yield the following theorem.

THEOREM 4.6. *Suppose P is an n-dimensional joint probability distribution for the k-state variables $Y_1, \dots, Y_n$. Then P arises from nonsingular stochastic parameters for GM(k) on an n-leaf binary tree T if, and only if,*

(i) *all marginalizations of P to* 3 *variables satisfy the conditions of Theorem* 3.6 *(or if $k = 2$ of Theorem* 3.7*) to arise from nonsingular stochastic parameters on a* 3*-leaf tree;*

(ii) *for all internal edges e of T, the edge invariants are satisfied by P, i.e., all $(k+1) \times (k+1)$ minors of the matrix flattening $Flat_e(P)$ are* 0*;*

(iii) *for each internal edge e of T, and some choice of* 4 *leaves inducing a quartet tree with internal edge e, the matrix flattening constructed in Proposition* 4.5 *for the four-dimensional marginalization is positive semidefinite.*

The full set of inequalities given in Theorem 4.6 also has additional redundancies. To illustrate, in the 4-leaf case checking that only two of the 3-marginals, say, $P_{+\cdots}$ and $P_{\cdots+}$ for the tree 12|34, satisfy the conditions of Theorem 3.6 is sufficient.

For a 4-variable distribution P, it is straightforward to obtain explicit semialgebraic conditions ensuring that P arises from strictly positive parameters: One need only require the more stringent condition (iv′) of Theorem 3.6. Then the argument of Theorem 4.6 extends to establish the following, whose detailed statement we leave to the reader.

THEOREM 4.7. *Semialgebraic conditions that a probability distribution P arises from nonsingular positive parameters for* GM(k) *on a tree T can be explicitly given. In particular, for nonsingular positive parameters, explicit conditions are obtained by modifying* (i), (ii), *and* (iii) *of Theorem* 4.6 *using Theorems* 3.6 *and* 3.8 *to express that the resulting quadratic forms are positive definite.*

Note that one can also handle nonbinary trees by the techniques of this section. To show that a distribution arises from nonsingular, or stochastic nonsingular, parameters on a nonbinary tree, one need only show that it arises from parameters on a binary resolution of the tree, and that the Markov matrix on each edge introduced to obtain the resolution is the identity. But semialgebraic conditions that the Markov matrix on an internal edge of a 4-leaf tree is I (or a permutation, since label-swapping prevents us from distinguishing these) amount to requiring that the matrix of (4.3) has rank k. Indeed, rank k implies that the Markov matrix on the internal edge has only k nonzero entries, and since other conditions we have derived imply nonsingularity, the matrix must be a permutation.

In closing, we give an example illustrating that the quadratic form approach of Proposition 4.5, and thus of Theorem 4.6, detects a probability distribution that is in the image of ψ_T for nonsingular real GM(2) parameters on the 4-taxon tree, where each matrix parameter on a terminal edge is stochastic but the one on the internal edge is not. By choosing parameters with some care, we can arrange that such a probability distribution P satisfies that all 3-marginalizations arise from stochastic parameters, yet P does not. Such examples are not new (see, for example, [3, 20, 32]), but we include one here to illustrate our methods.

To create such an example, set the Markov parameter on each terminal edge to have positive entries, using, for instance, the same M on each of these 4 edges. Then choose the matrix parameter N on the internal edge of the tree to have very small negative off-diagonal entries, so small that both MN and NM are Markov matrices. The root distribution may be taken to be any probability distribution with positive entries. An example of such an (exact) probability distribution is given by P with slices

$$P_{\cdot\cdot 11} = \begin{bmatrix} 0.4005062 & 0.0565718 \\ 0.0565718 & 0.0545702 \end{bmatrix}, \qquad P_{\cdot\cdot 12} = \begin{bmatrix} 0.0457358 & 0.0141662 \\ 0.0141662 & 0.0379118 \end{bmatrix},$$
$$P_{\cdot\cdot 21} = \begin{bmatrix} 0.0457358 & 0.0141662 \\ 0.0141662 & 0.0379118 \end{bmatrix}, \qquad P_{\cdot\cdot 22} = \begin{bmatrix} 0.0100222 & 0.0330958 \\ 0.0330958 & 0.1316062 \end{bmatrix}.$$

Here P satisfies all conditions of Theorem 4.6 except (iii). A computation shows that the leading principal minors of the matrix in (4.1) are, when rounded to eight decimal places, 0.00363408, 0.00001744, 0.00000060, and -0.00000005. The negativity of one of these shows that P does not arise from stochastic parameters.

We conclude with a complete semialgebraic description of the 2-state general Markov model on a 4-leaf tree without a restriction to nonsingular stochastic parameters. This is straightforward to give, since by Proposition 2.2 a distribution which arises from parameters either has independent leaf sets so we can decompose the tree using Proposition 4.3, or the parameters were nonsingular so Theorem 4.6 applies.

PROPOSITION 4.8. *For the* 4*-leaf tree* $12|34$*, the image of the stochastic parameter space under the general Markov model* $GM(2)$ *is the union of the following sets of nonnegative tensors whose entries add to* 1*:*

1. *Probability distributions of* 4 *independent variables: P such that all 2×2 minors of every edge flattening vanish (i.e., all edge flattenings have rank* 1*).*

2. *Probability distributions with partition into minimal independent sets of variables of size* 1, 3, *of which there are* 4 *cases: If the partition is $\{\{Y_1\}, \{Y_2, Y_3, Y_4\}\}$, then P such that all 2×2 minors of $\mathrm{Flat}_{1|234}(P)$ vanish, and $P_{+\cdots}$ satisfies the conditions of Theorem* 3.7, *case* 1*.*

3. *Probability distributions with partition into minimal independent sets of variables of size* 1, 1, 2, *of which there are* 6 *cases: If the partition is $\{Y_1\}|\{Y_2\}|\{Y_3, Y_4\}$, then P such that all 2×2 minors of $\mathrm{Flat}_{1|234}(P)$ and $\mathrm{Flat}_{2|134}(P)$ vanish, and $\det(P_{++\cdot\cdot})$ is nonzero.*

4. *Probability distributions with partition into minimal independent sets of variables $\{\{Y_1, Y_2\}, \{Y_3, Y_4\}\}$ of size* 2, 2*: P such that all 2×2 minors of $\mathrm{Flat}_{12|34}(P)$ vanish, yet $\det(P_{\cdot\cdot++})$ and $\det(P_{++\cdot\cdot})$ are nonzero.*

5. *Probability distributions with no independent sets of variables: P such that the edge invariants for* 12|34 *are satisfied, the three-dimensional marginalizations $P_{+\cdots}$ and $P_{\cdots+}$ satisfy the conditions of Theorem* 3.7, *case* 1, *and all principal minors of the matrix constructed in Proposition* 4.5 *are nonnegative.*

In case 1, the only edge flattenings that are needed are those associated to terminal edges. If these all have rank 1, then the flattening for the internal edge does as well. In cases 1, 2, 3, the distributions arise from stochastic parameters on all 3 of the binary topological trees with 4 leaves, as well as the star tree. Finally, note that all 15 possible partitions of variables do not appear, but only those 13 consistent with the tree topology.

The above proposition extends to trees of arbitrary size, as long as $k = 2$, but the number of possible partitions into independent sets of variables grows quickly.

Acknowledgments. We thank Piotr Zwiernik for his careful reading of an earlier version of this paper and his helpful comments. We are also grateful to the Mittag Leffler Institute for its hospitality in early 2011, where some of this work was undertaken.

REFERENCES

[1] E. S. Allman, P. Jarvis, J. A. Rhodes, and J. Sumner, *Tensor rank, invariants, inequalities, and applications*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 1014–1045.

[2] E. S. Allman and J. A. Rhodes, *Phylogenetic invariants for the general Markov model of sequence mutation*, Math. Biosci., 186 (2003), pp. 113–144.

[3] E. S. Allman and J. A. Rhodes, *Quartets and parameter recovery for the general Markov model of sequence mutation*, Appl. Math. Res. Express. ARMX, 2004, pp. 107–131.

[4] E. S. Allman and J. A. Rhodes, *The identifiability of tree topology for phylogenetic models, including covarion and mixture models*, J. Comput. Biol., 13 (2006), pp. 1101–1113.

[5] E. S. Allman and J. A. Rhodes, *Phylogenetic invariants*, in Reconstructing Evolution, Oxford University Press, Oxford, 2007, pp. 108–146.

[6] E. S. Allman and J. A. Rhodes, *Phylogenetic ideals and varieties for the general Markov model*, Adv. Appl. Math., 40 (2008), pp. 127–148.

[7] E. S. Allman, J. A. Rhodes, and S. Sullivant, *When do phylogenetic mixture models mimic other phylogenetic models?*, Syst. Biol., 26 (2012), pp. 472–481.

[8] V. Auvray, P. Geurts, and L. Wehenkel, *A semi-algebraic description of discrete naive Bayes models with two hidden classes*, in Proceedings of the Ninth International Symposium on Artificial Intelligence and Mathematics, 2006.

[9] W. Buczyńska, *Phylogenetic toric varieties on graphs*, J. Algebraic Combin., 35 (2012), pp. 421–460.

[10] M. Casanellas and J. Fernández-Sánchez, *Relevant phylogenetic invariants of evolutionary models*, J. Math. Pures Appl. (9), 96 (2011), pp. 207–229.

[11] J. A. Cavender and J. Felsenstein, *Invariants of phylogenies in a simple case with discrete states*, J. Classification, 4 (1987), pp. 57–71.
[12] A. Cayley, *On the theory of linear transformations*, Cambridge Math. J., 4 (1845), pp. 193–209.
[13] J. T. Chang, *Full reconstruction of Markov models on evolutionary trees: Identifiability and consistency*, Math. Biosci., 137 (1996), pp. 51–73.
[14] V. de Silva and L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.
[15] S. N. Evans and T. P. Speed, *Invariants of some probability models used in phylogenetic inference*, Ann. Statist., 21 (1993), pp. 355–377.
[16] B. Faller and M. Steel, *Trait-dependent extinction leads to greater expected biodiversity loss*, SIAM J. Discrete Math., 26 (2012), pp. 472–481.
[17] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, resultants and multidimensional determinants*, Modern Birkhäuser Classics, Birkhäuser Boston, Cambridge, MA, 2008.
[18] T. R. Hagedorn, *Determining the number and structure of phylogenetic invariants*, Adv. Appl. Math., 24 (2000), pp. 1–21.
[19] M. D. Hendy and D. Penny, *Complete families of linear invariants for some stochastic models of sequence evolution, with and without the molecular clock assumption*, J. Comput. Biol., 3 (1996), pp. 19–31.
[20] S. Klaere and V. Liebscher, *An algebraic analysis of the two state Markov model on tripod trees*, Math. Biosci., 237 (2012), pp. 38–48.
[21] J. B. Kruskal, *More factors than subjects, tests and treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling*, Psychometrika, 41 (1976), pp. 281–293.
[22] J. B. Kruskal, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.
[23] J. A. Lake, *A rate independent technique for analysis of nucleic acid sequences: Evolutionary parsimony*, Mol. Bio. Evol., 4 (1987), pp. 167–191.
[24] J. Pearl and M. Tarsi, *Structuring causal trees*, J. Complexity, 2 (1986), pp. 60–77.
[25] J. A. Rhodes and S. Sullivant, *Identifiability of large phylogenetic mixture models*, Bull. Math. Biol., 74 (2012), pp. 212–231.
[26] A. Seidenberg, *A new decision method for elementary algebra*, Ann. of Math., 60 (1954), pp. 365–374.
[27] R. Settimi and J. Q. Smith, *Geometry, moments and conditional independence trees with hidden variables*, Ann. Statist., 28 (2000), pp. 1179–1205.
[28] M. A. Steel, L. Székely, and M. D. Hendy, *Reconstructing trees from sequences whose sites evolve at variable rates*, J. Comput. Biol., 1 (1994), pp. 153–163.
[29] B. Sturmfels and S. Sullivant, *Toric ideals of phylogenetic invariants*, J. Comput. Biol., 12 (2005), pp. 204–228.
[30] J. J. Sylvester, *A demonstration of the theorem that every homogeneous quadratic polynomial is reducible by real orthogonal substitutions to the form of a sum of positive and negative squares*, Philos. Mag., IV (1852), pp. 138–142.
[31] A. Tarski, *A Decision Method for Elementary Algebra and Geometry*, 2nd ed., University of California Press, Berkeley, CA, 1951.
[32] P. Zwiernik and J. Smith, *Implicit inequality constraints in a binary tree model*, Electron. J. Stat., 5 (2011), pp. 1276–1312.
[33] P. Zwiernik and J. Smith, *Tree cumulants and the geometry of binary tree models*, Bernoulli, 18 (2012), pp. 290–321.