Linear Algebra and its Applications 473 (2015) 37–53

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Tensors of nonnegative rank two ☆

Elizabeth S. Allman [a], John A. Rhodes [a,*], Bernd Sturmfels [b], Piotr Zwiernik [b]

[a] University of Alaska at Fairbanks, PO Box 756660, Fairbanks, AK 99709, United States
[b] University of California at Berkeley, United States

ARTICLE INFO

Article history:
Received 2 May 2013
Accepted 28 October 2013
Available online 19 November 2013
Submitted by L.-H. Lim

MSC:
15A69
62H17
14P10

Keywords:
Nonnegative tensor rank
Latent class model
Binary tree model

ABSTRACT

A nonnegative tensor has nonnegative rank at most 2 if and only if it is supermodular and has flattening rank at most 2. We prove this result, then explore the semialgebraic geometry of the general Markov model on phylogenetic trees with binary states, and comment on possible extensions to tensors of higher rank.

1. Introduction

This article offers a journey into *semialgebraic statistics*. By this we mean the systematic study of statistical models as semialgebraic sets. We shall give a semialgebraic description of binary latent class models in terms of binomials expressing supermodularity, and we determine the algebraic boundary of this and related models. Our discussion is phrased in the language of nonnegative tensor factorization [5,9].

☆ BS was supported by NSF (DMS-0968882) and DARPA (HR0011-12-1-0011), and PZ by the European Union 7th Framework Programme (PIOF-GA-2011-300975). We thank Luke Oeding and Giorgio Ottaviani for helpful conversations.
* Corresponding author.
E-mail addresses: e.allman@alaska.edu (E.S. Allman), j.rhodes@alaska.edu (J.A. Rhodes), bernd@math.berkeley.edu (B. Sturmfels), pzwiernik@berkeley.edu (P. Zwiernik).

0024-3795/$ – see front matter
http://dx.doi.org/10.1016/j.laa.2013.10.046

We consider real tensors $P = [p_{i_1 i_2 \cdots i_n}]$ of format $d_1 \times d_2 \times \cdots \times d_n$. Throughout this paper we shall assume that $n \geqslant 3$ and $d_1, d_2, \ldots, d_n \geqslant 2$. Such a tensor has *nonnegative rank at most* 2 if it can be written as

$$P = a_1 \otimes a_2 \otimes \cdots \otimes a_n + b_1 \otimes b_2 \otimes \cdots \otimes b_n, \tag{1}$$

where the vectors $a_i, b_i \in \mathbb{R}^{d_i}$ are nonnegative for $i = 1, 2, \ldots, n$. The set of such tensors is a closed semialgebraic subset of dimension $2(d_1 + d_2 + \cdots + d_n) - 2(n-1)$ in the tensor space $\mathbb{R}^{d_1 \times d_2 \times \cdots \times d_n}$; see [14, §5.5]. We present the following characterization of this semialgebraic set.

Theorem 1.1. *A nonnegative tensor P has nonnegative rank at most 2 if and only if P is supermodular and has flattening rank at most 2.*

Here, *flattening* means picking any subset A of $[n] = \{1, 2, \ldots, n\}$ with $1 \leqslant |A| \leqslant n-1$ and writing the tensor P as an ordinary matrix with $\prod_{i \in A} d_i$ rows and $\prod_{j \notin A} d_j$ columns. The *flattening rank* of P (also called the *multilinear rank* in the literature) is the maximal rank of any of these matrices. Landsberg and Manivel [15] proved that flattening rank $\leqslant 2$ is equivalent to border rank $\leqslant 2$, which for nonnegative tensors is equivalent to rank $\leqslant 2$ by [16, Proposition 6.2].

To define supermodularity, we first fix a tuple $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$ where π_i is a permutation of $\{1, 2, \ldots, d_i\}$. Then P is π*-supermodular* if

$$p_{i_1 i_2 \cdots i_n} \cdot p_{j_1 j_2 \cdots j_n} \leqslant p_{k_1 k_2 \cdots k_n} \cdot p_{l_1 l_2 \cdots l_n} \tag{2}$$

whenever $\{i_r, j_r\} = \{k_r, l_r\}$ and $\pi_r(k_r) \leqslant \pi_r(l_r)$ holds for $r = 1, 2, \ldots, n$. We call a tensor P *supermodular* if it is π-supermodular for some π.

Theorem 1.1 says that every tensor of the form (1) is π-supermodular. Here, the tuple of permutations π can be read off from the signs of the 2×2-minors of matrices A_i with two rows given by a_i and b_i. In particular, for $e = (\mathrm{id}, \ldots, \mathrm{id})$ the tensor P is e-supermodular if and only if these minors are all nonnegative (see Lemma 3.3), or all nonpositive. While conditions such as (2) have appeared before in the statistics literature, e.g. [12], the results in this paper are both fundamental and new.

Note that we are using multiplicative notation instead of the additive notation more commonly used for supermodularity. To be specific, if $d_1 = d_2 = \cdots = d_n = 2$, $\pi = (\mathrm{id}, \mathrm{id}, \ldots, \mathrm{id})$, and P is strictly positive, then P being π-supermodular means that $\log(P)$ lies in the convex polyhedral cone [18, §4] of supermodular functions $2^{\{1,2,\ldots,n\}} \to \mathbb{R}$.

The set of π-supermodular nonnegative tensors P of flattening rank $\leqslant 2$ is denoted $\mathcal{M}_\pi$ and called a *toric cell*. The number of toric cells is $d_1! d_2! \cdots d_n!/2$. Theorem 1.1 states that these cells stratify our model:

$$\mathcal{M} = \bigcup_{\pi} \mathcal{M}_\pi. \tag{3}$$

The term *model* refers to the fact that intersection of (3) with the probability simplex, where all coordinates of P sum to one, is a widely used statistical model. It is the mixture model for pairs of independent distributions on n discrete random variables.

The *Zariski closure* $\overline{\mathcal{S}}$ of a semialgebraic subset $\mathcal{S}$ of $\mathbb{R}^N$ is the complex zero set in $\mathbb{C}^N$ of all polynomials that vanish on $\mathcal{S}$. The *boundary* $\partial\mathcal{S}$ is the topological boundary of $\mathcal{S}$ inside $\overline{\mathcal{S}}$. We define the *algebraic boundary* of $\mathcal{S}$ to be the Zariski closure $\overline{\partial\mathcal{S}}$ of its topological boundary.

Our second theorem concerns the algebraic boundaries of the model $\mathcal{M}$ and of toric cells $\mathcal{M}_\pi$. We regard these boundaries as hypersurfaces inside the complex variety of tensors of border rank $\leqslant 2$. A *slice* of our tensor P is a subtensor of some format $d_1 \times \cdots \times d_{s-1} \times 1 \times d_{s+1} \times \cdots \times d_n$. Subtensors of format $d_1 \times \cdots \times d_{s-1} \times 2 \times d_{s+1} \times \cdots \times d_n$ are *double slices*.

Theorem 1.2. *The algebraic boundary of $\mathcal{M}$ has $\sum_{i=1}^n d_i$ irreducible components, given by slices having rank $\leqslant 1$. The algebraic boundary of any toric cell $\mathcal{M}_\pi$ has the same irreducible components plus $\sum_{i=1}^n \binom{d_i}{2}$ additional components given by linearly dependent double slices.*

A double slice is *linearly dependent* if its two slices are identical up to a multiplicative scalar. In the second component count of Theorem 1.2 we exclude the special case $2 \times 2 \times 2$ because the "further components" fail to be hypersurfaces. If $n = 2$ then the rank 1 constraint on slices is void, and the algebraic boundary consists of the $d_1 d_2$ coordinate hyperplanes in $\mathbb{R}^{d_1 \times d_2}$. This is consistent with the fact [7, Example 4.1.2] that all nonnegative matrices of rank 2 have nonnegative rank 2.

This paper is organized as follows. In Section 2 we derive our two theorems for tensors of format $2 \times 2 \times 2$. This extends results in [3,4,13,19,23] on this widely studied latent class model. Here, our semialgebraic set $\mathcal{M}$ is full-dimensional in $\mathbb{R}^{2\times2\times2}$, and it consists of four toric cells that are glued together. Any two cells intersect along the locus where one of the flattenings has rank one. The common intersection of all cells is the *independence model* (tensors of rank 1). In Section 3 we prove Theorems 1.1 and 1.2 for arbitrary $d_1, d_2, \dots, d_n$.

The set $\mathcal{M}$ above appears in phylogenetics as the *general Markov model on a star tree* with binary states. Section 4 develops the extension of our results to phylogenetic trees other than star trees. These models have another type of component in their algebraic boundaries, characterized by the constraint that the ranks of certain matrix flattenings of P inconsistent with the tree topology drop from 4 to 3. Supermodularity in this context was pioneered by Steel and Faller [22]. Our results refine earlier work on the general Markov model in [2,3,13,23].

Section 5 concerns the challenges to be encountered when trying to extend our results to tensors of higher rank. We present case studies of algebraic boundaries for one identifiable model ($3 \times 3 \times 2$-tensors of rank 3) and one non-identifiable model ($2 \times 2 \times 2 \times 2$-tensors of rank 3).

2. The base case

Let $P = [p_{ijk}]$ be a real $2 \times 2 \times 2$ tensor. Then P has nonnegative rank at most 2 if there exist three nonnegative 2×2-matrices

$$A_1 = \begin{bmatrix} a_{11} & a_{12} \\ b_{11} & b_{12} \end{bmatrix}, \qquad A_2 = \begin{bmatrix} a_{21} & a_{22} \\ b_{21} & b_{22} \end{bmatrix} \quad \text{and} \quad A_3 = \begin{bmatrix} a_{31} & a_{32} \\ b_{31} & b_{32} \end{bmatrix}$$

such that

$$p_{ijk} = a_{1i} a_{2j} a_{3k} + b_{1i} b_{2j} b_{3k} \quad \text{for } i, j, k \in \{1, 2\}. \tag{4}$$

For $\pi = (\mathrm{id}, \mathrm{id}, \mathrm{id})$, the binomial inequalities for supermodularity are

$$\begin{array}{lll} p_{111}p_{222} \geqslant p_{112}p_{221}, & p_{111}p_{222} \geqslant p_{121}p_{212}, & p_{111}p_{222} \geqslant p_{211}p_{122}, \\ p_{112}p_{222} \geqslant p_{122}p_{212}, & p_{121}p_{222} \geqslant p_{122}p_{221}, & p_{211}p_{222} \geqslant p_{212}p_{221}, \\ p_{111}p_{122} \geqslant p_{112}p_{121}, & p_{111}p_{212} \geqslant p_{112}p_{211}, & p_{111}p_{221} \geqslant p_{121}p_{211}. \end{array} \tag{5}$$

Nonnegative $2 \times 2 \times 2$ tensors P that satisfy these nine inequalities lie in the toric cell $\mathcal{M}_{\mathrm{id},\mathrm{id},\mathrm{id}} = \mathcal{M}_{(12),(12),(12)}$. By label swapping $1 \leftrightarrow 2$, we obtain three other toric cells $\mathcal{M}_{\mathrm{id},\mathrm{id},(12)} = \mathcal{M}_{(12),(12),\mathrm{id}}$, $\mathcal{M}_{\mathrm{id},(12),\mathrm{id}} = \mathcal{M}_{(12),\mathrm{id},(12)}$, and $\mathcal{M}_{(12),\mathrm{id},\mathrm{id}} = \mathcal{M}_{\mathrm{id},(12),(12)}$. Thus, by definition, the semialgebraic set of all supermodular $2 \times 2 \times 2$-tensors is the union

$$\mathcal{M} = \mathcal{M}_{\mathrm{id},\mathrm{id},\mathrm{id}} \cup \mathcal{M}_{\mathrm{id},\mathrm{id},(12)} \cup \mathcal{M}_{\mathrm{id},(12),\mathrm{id}} \cup \mathcal{M}_{(12),\mathrm{id},\mathrm{id}}. \tag{6}$$

Theorem 1.1 states that $P \in \mathbb{R}^{2\times2\times2}$ has nonnegative rank at most 2 if and only if P lies in $\mathcal{M}$. We begin by proving the only-if direction.

Lemma 2.1. *If $P \in \mathbb{R}^{2\times2\times2}$ has nonnegative rank at most 2, then P is supermodular. More precisely, define $\pi = (\pi_1, \pi_2, \pi_3)$ by $\pi_i = \mathrm{id}$ if $\det(A_i) \geqslant 0$ and $\pi_i = (12)$ if $\det(A_i) < 0$. Then $P \in \mathcal{M}_\pi$.*

Proof. Let P be as in (4). The last six constraints in (5) specialize to

$$\begin{aligned}
p_{112}p_{222} - p_{122}p_{212} &= a_{32}b_{32}\det(A_1)\det(A_2),\\
p_{121}p_{222} - p_{122}p_{221} &= a_{22}b_{22}\det(A_1)\det(A_3),\\
p_{211}p_{222} - p_{212}p_{221} &= a_{12}b_{12}\det(A_2)\det(A_3),\\
p_{111}p_{122} - p_{112}p_{121} &= a_{11}b_{11}\det(A_2)\det(A_3),\\
p_{111}p_{212} - p_{112}p_{211} &= a_{21}b_{21}\det(A_1)\det(A_3),\\
p_{111}p_{221} - p_{121}p_{211} &= a_{31}b_{31}\det(A_1)\det(A_2).
\end{aligned} \tag{7}$$

First suppose that all 12 parameters a_{ij} and b_{ij} and the three determinants $\det(A_k)$ are positive. Then the six expressions in (7) are positive. The first three constraints in (5) are also satisfied, as seen from

$$p_{111}p_{222} - p_{121}p_{212} = \big(p_{111}(p_{112}p_{222} - p_{122}p_{212}) + p_{212}(p_{111}p_{122} - p_{112}p_{121})\big)/p_{112}. \tag{8}$$

Second, consider all tensors P where the parameters a_{ij}, b_{ij} and determinants $\det(A_k)$ are nonnegative. These lie in the closure of the previous case, so the nine binomials will be nonnegative.

If P has nonnegative rank $\leqslant 2$ then $\pi P = (p_{\pi_1(i)\pi_2(j)\pi_3(k)})$ also has nonnegative rank $\leqslant 2$, with parameterization given by swapping the columns of A_i whenever $\pi_i = (12)$. This changes the sign of $\det A_i$. Hence, for some π we can assume that $\pi P \in \mathcal{M}_{\mathrm{id,id,id}}$. However, $\pi P \in \mathcal{M}_{\mathrm{id,id,id}}$ if and only if $P \in \mathcal{M}_\pi$. $\square$

We now prove Theorem 1.1 for $2 \times 2 \times 2$ tensors. In this special case, the flattening rank is automatically $\leqslant 2$, so there are no equational constraints, and our model $\mathcal{M}$ is a full-dimensional subset of $\mathbb{R}^{2\times 2\times 2}$.

Proposition 2.2. *Let P be a nonnegative $2 \times 2 \times 2$-tensor. Then P has nonnegative rank $\leqslant 2$ if and only if P is supermodular.*

Proof. If P has nonnegative rank $\leqslant 2$, then P is supermodular by Lemma 2.1. For the converse, suppose that P is supermodular. Define

$$\begin{aligned}
U_{12} &= p_{11+}p_{22+} - p_{12+}p_{21+},\\
U_{13} &= p_{1+1}p_{2+2} - p_{1+2}p_{2+1},\\
U_{23} &= p_{+11}p_{+22} - p_{+12}p_{+21},
\end{aligned} \tag{9}$$

where a subscript $+$ refers to summing over all values of the given index. For example, $p_{22+} = p_{221} + p_{222}$. Similarly, for $i = 1, 2$, define

$$\begin{aligned}
U_{12}^i &= p_{11i}p_{22i} - p_{12i}p_{21i},\\
U_{13}^i &= p_{1i1}p_{2i2} - p_{1i2}p_{2i1},\\
U_{23}^i &= p_{i11}p_{i22} - p_{i12}p_{i21}.
\end{aligned} \tag{10}$$

Our strategy is to first show that the following hold for P:

(i) $U_{12}U_{13}U_{23} \geqslant 0$,
(ii) U_{ij}^1 and U_{ij}^2 have the same sign as U_{ij} for every $i < j$, and
(iii) if $U_{ij} = 0$, then $U_{ij}^1 = U_{ij}^2 = 0$.

Subsequently, in the second step, we will show that (i), (ii), (iii) imply that P has the form (4) with A_1, A_2, A_3 nonnegative. That second step will follow proofs of closely related results in [3,4,13,19,23].

Let $e = (\mathrm{id}, \mathrm{id}, \mathrm{id})$. Since $\pi\mathcal{M}_e = \mathcal{M}_\pi$, and the conditions (i), (ii), (iii) are invariant under label swapping, it suffices to consider $P \in \mathcal{M}_e$. By definition of e-supermodularity, $U^1_{ij}, U^2_{ij} \geqslant 0$. We need to show that $U_{ij} \geqslant 0$ also. By symmetry it suffices to show that $U_{12} \geqslant 0$. We have

$$U_{12} = U^1_{12} + U^2_{12} + (p_{111}p_{222} + p_{221}p_{112} - p_{211}p_{122} - p_{121}p_{212}). \tag{11}$$

We show that the expression in parentheses is nonnegative for $P \in \mathcal{M}_e$. We write this expression as $R = f_{11} + f_{22} - f_{12} - f_{21}$, where

$$f_{11} = p_{111}p_{222}, \qquad f_{12} = p_{121}p_{212}, \qquad f_{21} = p_{211}p_{122}, \qquad f_{22} = p_{221}p_{112}.$$

Note that, by (5), we have $f_{11} \geqslant \max\{f_{12}, f_{21}\}$. This implies $R \geqslant 0$ if either $p_{ijk} = 0$ for some i, j, k, or if $f_{22} \geqslant \min\{f_{12}, f_{21}\}$. Thus, we assume that $f_{ij} > 0$ and $f_{22} < \min\{f_{12}, f_{21}\}$. The supermodular inequalities $p_{121}p_{211} \leqslant p_{111}p_{221}$ and $p_{212}p_{122} \leqslant p_{222}p_{112}$ imply

$$f_{12}f_{21} = p_{121}p_{212}p_{211}p_{122} \leqslant p_{111}p_{222}p_{221}p_{112} = f_{11}f_{22}.$$

Hence $[f_{ij}]$ is supermodular itself. As a consequence, we have

$$\frac{f_{21}}{f_{11}} - 1 \leqslant \frac{f_{22}}{f_{12}} - 1 \leqslant \left(\frac{f_{22}}{f_{12}} - 1\right)\frac{f_{12}}{f_{11}},$$

where the second inequality holds since $f_{22} < f_{12} \leqslant f_{11}$.

After multiplying both sides by f_{11} we obtain

$$f_{21} - f_{11} \leqslant f_{22} - f_{12}$$

or equivalently $R \geqslant 0$. It follows that $U_{12} \geqslant 0$ and, by symmetry, that $U_{ij} \geqslant 0$ for all $i < j$; therefore (i) and (ii) hold. The identity (11) and the inequality $R \geqslant 0$ together imply that (iii) holds as well.

We take up separately the cases where the product $U_{12}U_{13}U_{23}$ is positive or is zero. Suppose first that $U_{12}U_{13}U_{23} > 0$. A special role in our argument will be played by the *hyperdeterminant*,

$$\begin{aligned}\mathrm{Det}(P) = {} & 4p_{111}p_{122}p_{212}p_{221} + 4p_{112}p_{121}p_{211}p_{222} + p_{111}^2p_{222}^2 + p_{122}^2p_{211}^2 + p_{112}^2p_{221}^2 \\ & - 2p_{111}p_{112}p_{221}p_{222} - 2p_{111}p_{121}p_{212}p_{222} - 2p_{111}p_{122}p_{211}p_{222} + p_{121}^2p_{212}^2 \\ & - 2p_{112}p_{121}p_{212}p_{221} - 2p_{112}p_{122}p_{211}p_{221} - 2p_{121}p_{122}p_{211}p_{212}.\end{aligned}$$

One can verify the identity

$$p_{+++}^2\,\mathrm{Det}(P) = \mu^2 + 4U_{12}U_{13}U_{23}, \tag{12}$$

where

$$\mu = p_{+++}^2p_{222} - p_{+++}(p_{2++}p_{+22} + p_{+2+}p_{2+2} + p_{++2}p_{22+}) + 2p_{2++}p_{+2+}p_{++2}.$$

Then (12) implies $\mathrm{Det}(P) > 0$.

By [5, Proposition 5.9] we can write P in terms of *real* vectors a_i, b_i as in (4). We obtain the identities

$$\begin{aligned}U_{12} &= \det A_1 \det A_2(a_{31} + a_{32})(b_{31} + b_{32}),\\ U_{13} &= \det A_1 \det A_3(a_{21} + a_{22})(b_{21} + b_{22}),\\ U_{23} &= \det A_2 \det A_3(a_{11} + a_{12})(b_{11} + b_{12}).\end{aligned}$$

Since $U_{12}U_{13}U_{23}$ is strictly positive, the coordinate sum of each vector a_i, b_i is nonzero. Hence our model can be equivalently parametrized by

$$P = sa_1 \otimes a_2 \otimes a_3 + tb_1 \otimes b_2 \otimes b_3, \tag{13}$$

where $s, t \in \mathbb{R}$ and the coordinates of a_i, b_i sum to 1. We now show that (i)–(iii) ensures these parameters to be nonnegative. Note that

$$\begin{aligned} U_{12} &= \det A_1 \det A_2 st, \\ U_{12}^1 &= \det A_1 \det A_2 st a_{31} b_{31}, \\ U_{12}^2 &= \det A_1 \det A_2 st a_{32} b_{32}, \end{aligned} \tag{14}$$

and similar formulas hold for $U_{13}, U_{13}^1, U_{13}^2$ and $U_{23}, U_{23}^1, U_{23}^2$.

Under the specialization (13), the hyperdeterminant factors as

$$\mathrm{Det}(P) = (st \det A_1 \det A_2 \det A_3)^2.$$

This gives

$$st = \frac{U_{12} U_{13} U_{23}}{\mathrm{Det}(P)} > 0,$$

and thus either $s, t > 0$ or $s, t < 0$. By (ii), $U_{12}, U_{12}^1, U_{12}^2$ have the same signs. Hence $a_{31}b_{31} \geqslant 0$ and $a_{32}b_{32} \geqslant 0$ by (14). This, together with the fact that $[p_{++i}] = sa_3 + tb_3$ is a nonnegative vector, implies that $a_3, b_3 \in \mathbb{R}^2_{\geqslant 0}$ if $s, t > 0$ and $a_3, b_3 \in \mathbb{R}^2_{\leqslant 0}$ if $s, t < 0$. The same argument shows that $a_1, b_1, a_2, b_2 \in \mathbb{R}^2_{\geqslant 0}$ if $s, t > 0$ and $a_1, b_1, a_2, b_2 \in \mathbb{R}^2_{\leqslant 0}$ if $s, t < 0$. Hence, we obtain a nonnegative decomposition in (13).

Suppose now that $U_{12}U_{13}U_{23} = 0$. Without loss of generality, assume $U_{12} = 0$. Hypothesis (iii) implies $U_{12}^1 = U_{12}^2 = 0$. Regard the expressions in (9) and (10) as elements in the polynomial ring $\mathbb{Q}[p_{111}, p_{112}, \ldots, p_{222}]$. A computation reveals the prime decomposition

$$\langle U_{12}, U_{12}^1, U_{12}^2 \rangle = \langle 2 \times 2\text{-minors of } \mathrm{Flat}_{1|23}(p) \rangle \cap \langle 2 \times 2\text{-minors of } \mathrm{Flat}_{2|13}(p) \rangle,$$

where $\mathrm{Flat}_{1|23}(p) = \begin{pmatrix} p_{111} & p_{112} & p_{121} & p_{122} \\ p_{211} & p_{212} & p_{221} & p_{222} \end{pmatrix}$, and similarly for $\mathrm{Flat}_{2|13}(p)$. Hence one of these two flattenings of the tensor $P \in \mathcal{M}_e$ has rank 1. Suppose it is the first. We can find $v \in \mathbb{R}^2_{\geqslant 0}$ such that $p_{ijk} = v_i \cdot p_{+jk}$ for every $i, j, k \in \{1, 2\}$. Since the 2×2-matrix (p_{+jk}) can be written as $(p_{+jk}) = a_2 \otimes a_3 + b_2 \otimes b_3$ for some $a_2, b_2, a_3, b_3 \in \mathbb{R}^2_{\geqslant 0}$, we obtain the desired nonnegative representation (13) by setting $a_{1i} = b_{1i} = v_i$. □

Theorem 1.2 tells us that the algebraic boundary of $\mathcal{M}$ equals

$$\begin{aligned} &\{p_{112}p_{222} = p_{122}p_{212}\} \cup \{p_{121}p_{222} = p_{122}p_{221}\} \cup \{p_{211}p_{222} = p_{212}p_{221}\} \\ &\qquad \cup \{p_{111}p_{122} = p_{112}p_{121}\} \cup \{p_{111}p_{212} = p_{112}p_{211}\} \\ &\qquad \cup \{p_{111}p_{221} = p_{121}p_{211}\}. \end{aligned}$$

Each toric cell $\mathcal{M}_\pi$ has exactly the same algebraic boundary because the linear dependence constraint on double slices is void in the $2 \times 2 \times 2$-case. The coordinate planes $\{p_{ijk} = 0\}$ are *not* part of the algebraic boundary of $\mathcal{M}$ or $\mathcal{M}_\pi$. Indeed, the inverse image of $\{p_{ijk} = 0\}$ under the parametrization lies in the boundary. But, if any of the parameters a_{ij} or b_{ij} is zero then the tensor P has a rank 1 slice. Hence, the set $\{p_{ijk} = 0\} \cap \mathcal{M}$ lies in the union above. Similarly, the hyperdeterminant $\{\mathrm{Det}(P) = 0\}$ is *not* a component in the algebraic boundary of $\mathcal{M}$.

Example 2.3. It is instructive to look at a 3-dimensional picture of our 7-dimensional model $\mathcal{M}$. We consider the *Jukes–Cantor slice* given by

$$\begin{bmatrix} p_{111} & p_{112} \\ p_{121} & p_{122} \end{bmatrix} = \begin{bmatrix} x & y \\ z & w \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} p_{211} & p_{212} \\ p_{221} & p_{222} \end{bmatrix} = \begin{bmatrix} w & z \\ y & x \end{bmatrix}.$$

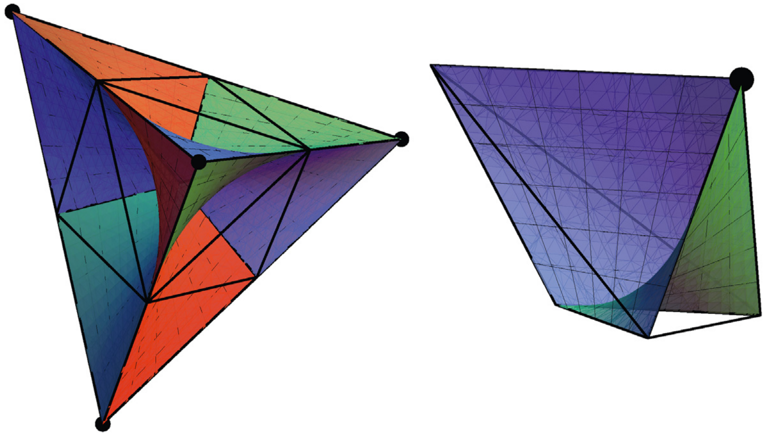

Fig. 1. Jukes–Cantor slice showing $2 \times 2 \times 2$ tensors of nonnegative rank $\leqslant 2$. Each toric cell is bounded by three quadrics and contains a vertex of the tetrahedron.

Under this specialization, the hyperdeterminant factors as

$$\mathrm{Det}(P) = (x+y+z+w)(x+y-z-w)(x-y+z-w)(x-y-z+w). \tag{15}$$

Consider the tetrahedron $\{(x, y, z, w) \in \mathbb{R}^4_{\geqslant 0}\colon\ x+y+z+w = 1/2\}$. Fixing the signs of the last three factors in (15) divides the tetrahedron into four bipyramids and four smaller tetrahedra. Inside our slice, the four toric cells of (6) occupy the bipyramids. Each toric cell is precisely the object in [8, Fig. 1]. Redrawn on the right in Fig. 1, its convex hull is the bipyramid, and it contains six of the nine edges. Any two of the toric cells meet in a line segment such as $\{x+y-z-w = x-y+z-w = 0,\ x-y-z+w \geqslant 0\}$. The algebraic boundary of each toric cell consists of the same three quadrics $\{xy = zw\}$, $\{xz = yw\}$ and $\{xw = yz\}$. Neither the three planes in (15) nor the four facet planes of the tetrahedron are in the algebraic boundary. $\square$

3. The general case

Before embarking on the general proofs of Theorems 1.1 and 1.2, let us briefly go over an example that exhibits the general behavior.

Example 3.1. Consider the semialgebraic set $\mathcal{M}$ of $3 \times 3 \times 3$-tensors of nonnegative rank $\leqslant 2$. The Zariski closure $\overline{\mathcal{M}}$ of $\mathcal{M}$ in $\mathbb{R}^{3\times3\times3}_{\geqslant 0}$ has dimension 14 and is defined by 222 cubic equations [10, Table 3], namely 3×3-minors of the 3×9-matrices $\mathrm{Flat}_{1|23}(P)$, $\mathrm{Flat}_{2|13}(P)$ and $\mathrm{Flat}_{3|13}(P)$. The model $\mathcal{M}$ decomposes into 108 toric cells $\mathcal{M}_\pi$, each defined in $\overline{\mathcal{M}}$ by 162 quadratic binomial inequalities of the form (2).

A quick way to generate these inequalities, for $\pi = (\mathrm{id}, \mathrm{id}, \mathrm{id})$, is to run the following code in the computer algebra system `Macaulay2` [11]:

```
R = QQ[p111,p112,p113,p121,p122,p123,p131,p132,p133,
       p211,p212,p213,p221,p222,p223,p231,p232,p233,
       p311,p312,p313,p321,p322,p323,p331,p332,p333];
S = QQ[a1,a2,a3,b1,b2,b3,c1,c2,c3];
f=map(S,R,{a1*b1*c1,a1*b1*c2,a1*b1*c3,a1*b2*c1,a1*b2*c2,a1*b2*c3,
a1*b3*c1,a1*b3*c2,a1*b3*c3,a2*b1*c1,a2*b1*c2,a2*b1*c3,a2*b2*c1,
a2*b2*c2,a2*b2*c3,a2*b3*c1,a2*b3*c2,a2*b3*c3,a3*b1*c1,a3*b1*c2,
a3*b1*c3,a3*b2*c1,a3*b2*c2,a3*b2*c3,a3*b3*c1,a3*b3*c2,a3*b3*c3});
gens gb kernel f
```

Being π-supermodular means that each of the binomials in the resulting Gröbner basis, such as `p223*p312-p212*p323`, must be nonpositive.

The algebraic boundary of $\mathcal{M}$ has nine irreducible components, corresponding to the nine slices of P. It is instructive to see how our 162 hypersurfaces, like $\{p_{223}p_{312} = p_{212}p_{323}\} \cap \overline{\mathcal{M}}$, break into these components. Each individual toric cell $\mathcal{M}_\pi$ has 18 irreducible components in its algebraic boundary: now also the 9 double slices kick in. The intersection of all 108 toric cells is the *Segre variety* of rank 1 tensors, whose reverse lexicographic Gröbner basis we identified with (2). □

A *marginalization* of P is any tensor obtained from P by summing all slices for some fixed indices. For instance, the 2×2-matrix (p_{ij+}) is a marginalization of the $2 \times 2 \times 2$-tensor $P = (p_{ijk})$. The following lemma, whose proof is delayed, will be useful in proving Theorem 1.1.

Lemma 3.2. *All marginalizations of a supermodular tensor are supermodular, and ditto for e-supermodular with* $e = (\mathrm{id}, \mathrm{id}, \ldots, \mathrm{id})$. *In addition, all flattenings of a supermodular tensor are supermodular.*

In this lemma, and in the remainder of the paper, we use the term *flattening* to include all tensor flattenings, not just the matrix flattenings described in the introduction. We now prove our first main result.

Proof of Theorem 1.1. Suppose first that P has nonnegative rank $\leqslant 2$. Then P has the form (1) with $a_i, b_i \in \mathbb{R}^{d_i}_{\geqslant 0}$. As tensor rank cannot increase under flattening, we conclude that P has flattening rank $\leqslant 2$.

Consider the $d_i \times 2$-matrix with columns a_i, b_i. By swapping rows we can make all 2×2-subdeterminants of these n matrices (a_i, b_i) nonnegative. But swapping rows in these matrices corresponds to acting on P by π, where $\pi P := [p_{\pi^{-1}(\mathbf{i})}]$ for $\mathbf{i} = (i_1, \ldots, i_n)$ and $\pi^{-1}(\mathbf{i}) = (\pi_1^{-1}(i_1), \ldots, \pi_n^{-1}(i_n))$. Since $P \in \mathcal{M}_e$ if and only if $\pi P \in \mathcal{M}_\pi$, it suffices to prove the following result.

Lemma 3.3. *If P has the form (1) with $a_{ik}b_{il} \geqslant a_{il}b_{ik}$ for every i and all $k \leqslant l$ then $P \in \mathcal{M}_e$.*

To prove this we define an auxiliary $2 \times d_1 \times \cdots \times d_n$ tensor $\hat{P}$ by

$$\hat{p}_{1i_1\cdots i_n} = a_{1i_1}a_{2i_2}\cdots a_{ni_n} \quad \text{and} \quad \hat{p}_{2i_1\cdots i_n} = b_{1i_1}b_{2i_2}\cdots b_{ni_n}.$$

We claim that $\hat{P}$ is e-supermodular. For this, we need to check that

$$\hat{p}_{i_0i_1\cdots i_n}\hat{p}_{j_0j_1\cdots j_n} \leqslant \hat{p}_{k_0k_1\cdots k_n}\hat{p}_{l_0l_1\cdots l_n}, \tag{16}$$

for all i_r, j_r such that $k_r = \min\{i_r, j_r\}$, $l_r = \max\{i_r, j_r\}$. This holds with equality if $i_0 = j_0$. If $i_0 \neq j_0$ we have two cases to consider, and our claim (16) is equivalent to the inequality

$$\max\{a_{1j_1}b_{1i_1}\cdots a_{nj_n}b_{ni_n}, a_{1i_1}b_{1j_1}\cdots a_{ni_n}b_{nj_n}\} \leqslant a_{1k_1}b_{1l_1}\ldots a_{nk_n}b_{nl_n}.$$

Our assumption on the 2×2-subdeterminants of (a_r, b_r) ensures

$$\max\{a_{rj_r}b_{ri_r}, a_{ri_r}b_{rj_r}\} \leqslant a_{rk_r}b_{rl_r} \quad \text{for every } r \in [n].$$

This gives the desired inequality, and therefore $\hat{P}$ is e-supermodular. But P is a marginalization of $\hat{P}$ because $p_{i_1\cdots i_n} = \hat{p}_{1i_1\cdots i_n} + \hat{p}_{2i_1\cdots i_n}$, so Lemma 3.2 then implies that P is e-supermodular.

For the converse, consider any supermodular $d_1 \times \cdots \times d_n$ tensor P of flattening rank $\leqslant 2$. Let F_i be the flattening of P given by the partition $\{i\}, [n]\backslash\{i\}$. Suppose $\mathrm{rank}(F_i) < 2$ for some i, say $i = 1$. Then $P = v \otimes P'$ for some $v \in \mathbb{R}^{d_1}_{\geqslant 0}$ and $P' = [p_{+i_2\cdots i_n}]$. By Lemma 3.2, the marginalization P' is supermodular with flattening rank $\leqslant 2$. By repeated application of this argument, we may reduce to tensors P whose $d_i \times (d_1 \cdots d_{i-1}d_{i+1} \cdots d_n)$-flattenings F_i all have rank exactly 2.

We next reduce to tensors of format $2 \times \cdots \times 2$. Let P be a supermodular $d_1 \times \cdots \times d_n$ tensor all of whose flattenings are of rank 2, and $L_i \subseteq \mathbb{R}^{d_i}$ the span of the columns of a flattening F_i. Two suitable columns of F_i give a nonnegative basis $\{a, b\}$ of L_i. We modify this basis to $\{a', b'\}$ so that, after permuting entries, it is nonnegative and

$$a' = (1, 0, *, \dots, *), \qquad b' = (0, 1, *, \dots, *).$$

To obtain this nonnegative basis first set $a'' = a - tb$, using the maximal t for which a'' is nonnegative. Then set $b'' = b - sa''$ with the maximal s for which b'' is nonnegative. The vectors a'', b'' each have an entry of 0 in a position where the other does not. Rescaling so the nonzero entries in these positions become 1, and permuting entries to bring these positions to the first two, we obtain the desired a', b'.

Now every column of F_i is in the nonnegative span of a', b'. More concretely, we have $F_i = C_i^T \cdot F_i'$, where C_i has rows a', b', and F_i' is the first two rows of F_i. On tensors, this is expressed by

$$P = P' *_i C_i,$$

where P' is the double slice of P with ith index in $\{1, 2\}$ and $P' *_i C_i$ denotes the linear action of C_i on the ith index of P'. Applying this construction in each index we find (after suitable relabellings) that

$$P = P_0 * (C_1, \dots, C_n), \tag{17}$$

where P_0 is the $2 \times \cdots \times 2$ subtensor of P obtained by restricting all indices to $\{1, 2\}$, and the $2 \times d_i$-matrices C_i are real and nonnegative.

Our hypotheses ensure that P_0 is supermodular with all flattening ranks 2. Moreover, if P_0 has nonnegative rank 2, then it follows from Eq. (17) that P also has nonnegative rank 2. Explicitly, if $P_0 = a_1 \otimes \cdots \otimes a_n + b_1 \otimes \cdots \otimes b_n$ is a nonnegative decomposition, then $P = \tilde{a}_1 \otimes \cdots \otimes \tilde{a}_n + \tilde{b}_1 \otimes \cdots \otimes \tilde{b}_n$ with $\tilde{a}_i = a_i C_i$, $\tilde{b}_i = b_i C_i$ nonnegative.

It remains to show the result for $2 \times \cdots \times 2$ tensors. Let P' denote the $2 \times 2 \times 2^{n-2}$ flattening of P from the tripartition $\{1\}, \{2\}, [n]\backslash\{1, 2\}$. By Lemma 3.2, P' is supermodular. By Proposition 2.2, each $2 \times 2 \times 2$ subtensor of P' has nonnegative rank $\leqslant 2$. The argument of the last three paragraphs implies that P' itself has nonnegative rank $\leqslant 2$, so

$$P' = a_1 \otimes a_2 \otimes a_3 + b_1 \otimes b_2 \otimes b_3,$$

with $a_1, a_2, b_1, b_2 \in \mathbb{R}^2_{\geqslant 0}$, $a_3, b_3 \in \mathbb{R}^{2^{n-2}}_{\geqslant 0}$. The matrices $A = (a_1, b_1)^T$ and $B = (a_2, b_2)^T$ are invertible, by our assumptions on the $2 \times 2^{n-1}$ flattening ranks of P. Acting on the tensor P by their inverses, we get

$$\tilde{P} = e_1 \otimes e_1 \otimes N_1 + e_2 \otimes e_2 \otimes N_2,$$

where N_1, N_2 are nonnegative tensors whose vector flattenings are a_3, b_3.

Consider any bipartition A, B of $\{3, \dots, n\}$. The $2^{|A|+1} \times 2^{|B|+1}$ flattening of $\tilde{P}$ using the bipartition $\{1\} \cup A$, $\{2\} \cup B$ is block-diagonal, with blocks given by $A|B$ flattenings of N_1, N_2. This $2^{|A|+1} \times 2^{|B|+1}$-matrix has rank $\leqslant 2$, so either both flattenings of N_i have rank $\leqslant 1$, or one N_i is zero. But $N_i = 0$ is impossible since that would mean some $2 \times 2^{n-1}$ flattening of P has rank 1. Hence the $A|B$ flattenings of N_1, N_2 have rank 1. Since A, B were arbitrary, both N_i have (nonnegative) rank 1. Consequently, $\tilde{P}$ has nonnegative rank 2, and so does P. □

It remains to prove Lemma 3.2. We shall use the *Four Function Theorem* of Ahlswede and Daykin [1], here presented in a special case:

Proposition 3.4 *(Ahlswede–Daykin). Fix $n \geqslant 2$ and a nonnegative $d_1 \times \cdots \times d_n$-tensor $P = [p_{i_1 \dots i_n}]$. For any collection $\mathcal{C}$ of indices $\mathbf{i} = (i_1, \dots, i_n)$ in $[d_1] \times \cdots \times [d_n]$ define $p_{\mathcal{C}} = \sum_{\mathbf{i} \in \mathcal{C}} p_{\mathbf{i}}$. Suppose that*

$$p_{\mathbf{i}} \cdot p_{\mathbf{j}} \leqslant p_{\mathbf{i} \vee \mathbf{j}} \cdot p_{\mathbf{i} \wedge \mathbf{j}} \quad \text{for any two indices } \mathbf{i}, \mathbf{j}, \tag{18}$$

where $\vee$, $\wedge$ are join and meet operations that gives $[d_1] \times \cdots \times [d_n]$ a lattice structure. Then for any two collections $\mathcal{C}, \mathcal{C}'$, we have

$$p_{\mathcal{C}} \cdot p_{\mathcal{C}'} \leqslant p_{\mathcal{C}\vee\mathcal{C}'} \cdot p_{\mathcal{C}\wedge\mathcal{C}'},$$

where $\mathcal{C} \vee \mathcal{C}' = \{\mathbf{i} \vee \mathbf{j}\colon \mathbf{i} \in \mathcal{C}, \mathbf{j} \in \mathcal{C}'\}$ *and* $\mathcal{C} \wedge \mathcal{C}' = \{\mathbf{i} \wedge \mathbf{j}\colon \mathbf{i} \in \mathcal{C}, \mathbf{j} \in \mathcal{C}'\}$.

Proof of Lemma 3.2. Let P be a supermodular $d_1 \times \cdots \times d_n$-tensor. For the first assertion, it suffices to show that $P' = [p_{+i_2\cdots i_n}]$ is supermodular. The general statement for marginal tensors follows by induction.

If P is π-supermodular, define the lattice structure on $[d_1] \times \cdots \times [d_n]$ by taking $\mathbf{k} = \mathbf{i} \wedge \mathbf{j}$ if and only if $\pi(\mathbf{k})$ is the coordinatewise minimum of $\pi(\mathbf{i})$ and $\pi(\mathbf{j})$. Similarly, $\mathbf{l} = \mathbf{i} \vee \mathbf{j}$ if and only if $\pi(\mathbf{l})$ is the coordinatewise maximum of $\pi(\mathbf{i})$ and $\pi(\mathbf{j})$. Fix $\mathbf{i}', \mathbf{j}' \in [d_2] \times \cdots \times [d_n]$ and set

$$\mathcal{C} = \{(i_1, \mathbf{i}')\colon i_1 \in [d_1]\}, \qquad \mathcal{C}' = \{(i_1, \mathbf{j}')\colon i_1 \in [d_1]\}.$$

We have $p_{\mathcal{C}} = \sum_{\mathbf{i}\in\mathcal{C}} p_{\mathbf{i}} = p_{+\mathbf{i}'}$ and $p_{\mathcal{C}'} = p_{+\mathbf{j}'}$. The tensor $\pi P = (p_{\pi^{-1}(\mathbf{i})})$ is e-supermodular. Proposition 3.4 now gives

$$p_{+\mathbf{i}'} \cdot p_{+\mathbf{j}'} \leqslant p_{+(\mathbf{i}'\wedge\mathbf{j}')} \cdot p_{+(\mathbf{i}'\vee\mathbf{j}')}.$$

This means that P' is π'-supermodular, where $\pi' = (\pi_2, \ldots, \pi_n)$.

We now prove that every flattening of P is supermodular. Let $Q = [q_{\alpha_1\cdots\alpha_r}]$ be a flattening of P corresponding to the partition $A_1, \ldots, A_r$ of $\{1, \ldots, n\}$. Let $h_i = \prod_{j\in A_i} d_j$, then $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r) \in [h_1] \times \cdots \times [h_r]$. Without loss of generality we can assume that α_i indexes elements of $\prod_{j\in A_i}[d_j]$ ordered lexicographically. Every $q_{\boldsymbol{\alpha}}$ is equal to $p_{\mathbf{i}}$ for some $\mathbf{i}$, so that each $\boldsymbol{\alpha}$ corresponds to a unique $\mathbf{i}$. Since P is supermodular, there exists $\pi = (\pi_1, \ldots, \pi_n)$ such that for every $\mathbf{i}, \mathbf{j}$ we have $p_{\mathbf{i}} p_{\mathbf{j}} \leqslant p_{\mathbf{i}\wedge\mathbf{j}} p_{\mathbf{i}\vee\mathbf{j}}$, where $\mathbf{i} \wedge \mathbf{j}$ and $\mathbf{i} \vee \mathbf{j}$ is as defined in the previous paragraph.

Define now $\boldsymbol{\alpha} \wedge \boldsymbol{\beta}$ and $\boldsymbol{\alpha} \vee \boldsymbol{\beta}$ to be the r-tuples corresponding to $\mathbf{i} \wedge \mathbf{j}$ and $\mathbf{i} \vee \mathbf{j}$. The permutation π induces the corresponding r-tuple of permutations $\tilde{\pi} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_r)$ such that $\pi(\mathbf{i}) = \tilde{\pi}(\boldsymbol{\alpha})$. By construction, we have $p_{\boldsymbol{\alpha}} p_{\boldsymbol{\beta}} \leqslant p_{\boldsymbol{\alpha}\wedge\boldsymbol{\beta}} p_{\boldsymbol{\alpha}\vee\boldsymbol{\beta}}$, where $\tilde{\pi}(\boldsymbol{\alpha} \vee \boldsymbol{\beta}) \leqslant \tilde{\pi}(\boldsymbol{\alpha} \vee \boldsymbol{\beta})$. This implies that Q is $\tilde{\pi}$-supermodular. □

We now prove the second theorem stated in the introduction.

Proof of Theorem 1.2. The formula (1) defines a polynomial map

$$\phi : \mathbb{R}^{2(d_1+d_2+\cdots+d_n)} \to \mathbb{R}^{d_1\times d_2\times\cdots\times d_n}$$

such that $\mathcal{M} = \phi(\mathbb{R}^{2(d_1+\cdots d_n)}_{\geqslant 0})$ is the set of tensors of nonnegative rank $\leqslant 2$. We modify the domain by assuming the coordinate sums of all a_i and b_i are 1, while adding two mixture parameters s, t as in (13). This does not change the image, but makes the map generically 2-to-1. More specifically, ϕ is 2-to-1 on the open set where $st \neq 0$ and each pair a_i, b_i is linearly independent. Since this open set intersects the coordinate hyperplane $\{a_{ij} = 0\}$ (or $\{b_{ij} = 0\}$), the map ϕ is generically finite on that hyperplane. Hence the closure of the image $\phi(\{a_{ij} = 0\})$ is an irreducible subvariety of codimension 1 in $\overline{\mathcal{M}}$. Moreover, in any neighborhood of a point on $\{a_{ij} = 0\}$ there are points with $a_{ij} < 0$ that are not mapped into the interior of $\mathcal{M}$. Indeed, generically the fiber containing such a point only contains its image under label swapping, and thus all points in the fiber have a negative coordinate. Thus $\phi(\{a_{ij} = 0\})$ is a component of the algebraic boundary of $\mathcal{M}$.

By restricting to open subsets $\mathcal{U}_\pi$ where the signs of all 2×2-minors of the matrices (a_i, b_i) are fixed, we see that $\overline{\phi(\{a_{ij} = 0\})}$ is also a component in the algebraic boundary of $\mathcal{M}_\pi$. Additional pieces of the boundary of $\mathcal{U}_\pi$ are the quadrics $\{a_{ij}b_{ik} = a_{ik}b_{ij}\}$, on whose general points the map ϕ is also 2-to-1. Therefore the varieties $\overline{\phi(\{a_{ij}b_{ik} = a_{ik}b_{ij}\})}$ are irreducible of codimension 1 in $\overline{\mathcal{M}_\pi}$, and, by the same argument as above, they are also components of the algebraic boundary of $\mathcal{M}_\pi$.

We next argue that there are no components in the algebraic boundary of $\mathcal{M}$ or $\mathcal{M}_\pi$ other than the two types we just identified. This follows from Theorem 1.1. Let $P \in \partial\mathcal{M}_\pi$. Consider the binomials $p_{i_1i_2\cdots i_n} p_{j_1j_2\cdots j_n} - p_{k_1k_2\cdots k_n} p_{l_1l_2\cdots l_n}$ that correspond to facets of the polyhedral cone of supermodular functions. For such a facet binomial, the indices in the four appearing unknowns $p_\bullet$ agree in all but two of the positions. All other binomials (2) admit representations such as (8). The expansion of a facet binomial into parameters a_{ij}, b_{ij} factors into coordinates and 2×2-determinants as in (7). Hence, at the two points in $\phi^{-1}(P)$, one of these factors must vanish, and this implies that P lies on one of the hypersurfaces we already identified above.

We finally identify $\overline{\phi(\{a_{ij}=0\})}$ and $\overline{\phi(\{a_{ij}b_{ik}=a_{ik}b_{ij}\})}$ with the rank loci described in the statement of Theorem 1.2. If the coordinate a_{ij} vanishes then the jth slice of P in the ith dimension drops its rank from $\leqslant 2$ to $\leqslant 1$. Likewise, if $a_{ij}b_{ik}=a_{ik}b_{ij}$, then the jth and kth slices of P in dimension i becomes linearly dependent. Hence the irreducible components of the algebraic boundaries of $\mathcal{M}$ and $\mathcal{M}_\pi$ are uniquely characterized by lying in the following two types of rank loci:

(a) the variety of tensors P of border rank $\leqslant 2$ such that a particular slice has border rank $\leqslant 1$;
(b) the variety of tensors P of border rank $\leqslant 2$ such that a particular double slice is linearly dependent.

This completes the proof of Theorem 1.2. □

We believe that the rank loci in (a) and (b) are irreducible varieties, and that their prime ideals are generated by the relevant subdeterminants of format 2×2 and 3×3. At present we do not know how to prove this. A similar issue for tree models appears in Conjecture 4.2. For the case of Example 3.1, we proved irreducibility by computation:

Example 3.5. The variety $\overline{\mathcal{M}}$ of $3 \times 3 \times 3$ tensors of border rank $\leqslant 2$ has dimension 14 and degree 783. Using `Macaulay2` [11], we verified that both (a) and (b) define irreducible subvarieties of dimension 13. The variety (a) has degree 882, and its prime ideal is minimally generated by 9 quadrics and 187 cubics. The variety (b) has degree 342, and its prime ideal is minimally generated by 36 quadrics and 90 cubics. All ideal generators can be chosen from the relevant subdeterminants. □

One may ask how efficiently the model membership can be tested. The number of facets of the submodular cone is a polynomial in the size of the tensor, and each facet inequality involves precisely four of the unknowns. Hence supermodularity for positive tensors can be tested in polynomial time. For instance, a $2 \times 2 \times \cdots \times 2$-tensor has $N = 2^n$ cell entries, and the facets correspond to the 2-faces of the n-cube (see the proof of Theorem 1.2), of which there are only $n(n-1)2^{n-3} = O(N^{1+\epsilon})$.

4. Binary tree models

In this section we study the extension of our results to the *general Markov model* $\mathcal{M}_T$ on a phylogenetic tree T with binary states [2,3,6,13,23]. The special case when T is a *star tree*, with only one internal node, corresponds to $2 \times 2 \times \cdots \times 2$-tensors of nonnegative rank $\leqslant 2$. For arbitrary trees T, Steel and Faller [22] showed that distributions in $\mathcal{M}_T$ are supermodular, by a marginalization argument as in Lemma 3.2.

We assume that T has $n \geqslant 3$ leaves, E is the set of edges of T, and one of the $|E|-n+1$ internal nodes is the root of T. We specify each probability distribution P in the model $\mathcal{M}_T$ by a nonnegative root distribution $\pi \in \mathbb{R}^2_{\geqslant 0}$, together with a 2×2 Markov matrix M_e for each edge e, directed away from the root. The entries of π and of each row of each M_e sum to 1. These choices determine a point $\theta = (\pi, (M_e)_{e\in E})$ in the cube $\Theta = [0,1]^{2|E|+1}$. That cube serves as the domain for the model parametrization $\phi: \Theta \twoheadrightarrow \mathcal{M}_T \subset \mathbb{R}^{2\times 2\times\cdots\times 2}_{\geqslant 0}$. This can be found in explicit form in [2, Eq. (1)]. The map ϕ is locally identifiable. To be precise, each general fiber consists of $2^{|E|-n+1}$ points, corresponding to label swapping on the internal nodes. Hence our binary tree model $\mathcal{M}_T = \phi(\Theta)$ is a compact semialgebraic set of dimension $2|E|+1$ inside the probability simplex of dimension $2^n - 1$. It is known that $\mathcal{M}_T$ is independent of the choice of the root node.

The prime ideal that defines the Zariski closure $\overline{\mathcal{M}_T}$ is known. It is generated by the 3×3-minors of all flattenings of P that are compatible with T. Here, a split (A, A^c) of $[n]$ is *compatible* with T if the intersection of any path between two leaves in A with any path between two leaves in A^c is either empty or just one internal node. This was first shown set-theoretically for trivalent trees by Allman and Rhodes [2]. The ideal-theoretic statement for arbitrary trees T is seen by combining the result of Draisma and Kuttler in [6] with the result of Raicu in [20].

Our main result of this section concerns the algebraic boundary of the general Markov model $\mathcal{M}_T$ inside the phylogenetic variety $\overline{\mathcal{M}_T}$.

Theorem 4.1. *The algebraic boundary of the binary tree model $\mathcal{M}_T$ has $n+|E|$ irreducible components, two for each of the n pendant edges, and one for each of the $|E|-n$ internal edges. The components are closures of images of facets of the cube Θ, as described below.*

The components of the algebraic boundary of $\mathcal{M}_T$ are as follows:

(1) For each pendant edge e with leaf ℓ, fix one row of the 2×2-matrix M_e. (The other row gives the same two components.) Nonnegativity of either entry determines a facet F of the cube Θ. Then $\overline{\phi(F)}$ is a component. It has the following *equational description* inside $\mathcal{M}_T$. If the internal node on e is r-valent, it gives a partition $(L_1=\{\ell\}, L_2, \dots, L_r)$ of $[n]$. Flatten P accordingly to a $2\times 2^{|L_2|}\times\cdots\times 2^{|L_r|}$ tensor. The rank of the $1\times 2^{|L_2|}\times\cdots\times 2^{|L_r|}$ slice selected by F drops to $\leqslant 1$ on $\overline{\phi(F)}$.

(2) For each internal edge e, fix any one entry of the 2×2-matrix M_e. (The other three entries give the same component.) Nonnegativity of that entry determines a facet F of the cube Θ. Then $\overline{\phi(F)}$ is a component. It has the following *equational description* inside $\mathcal{M}_T$. Let $T[e]$ be the tree obtained from T by contracting e. For either matrix flattening of P that is compatible with $T[e]$ but not with T, the rank drops to $\leqslant 3$ on $\overline{\phi(F)}$.

At present we do not know whether the equational descriptions above (in terms of tensor rank) are enough to cut out the codimension 1 subvarieties $\overline{\phi(F)}$ of $\mathcal{M}_T$. For this, it would suffice to prove the following:

Conjecture 4.2. *The rank varieties in* (1) *and* (2) *are irreducible.*

We have a computational proof of Conjecture 4.2 in the smallest non-trivial case, the trivalent tree on 4 taxa, which we discuss next.

Example 4.3. Let $n=4$ and T the trivalent tree with split 12|34. The phylogenetic variety lives in $\mathbb{P}^{15}$ and it has dimension 11:

$$\overline{\mathcal{M}_T} = \left\{ P\in\mathbb{P}^{15}\colon \operatorname{rank}\begin{bmatrix} p_{1111} & p_{1112} & p_{1121} & p_{1122}\\ p_{1211} & p_{1212} & p_{1221} & p_{1222}\\ p_{2111} & p_{2112} & p_{2121} & p_{2122}\\ p_{2211} & p_{2212} & p_{2221} & p_{2222} \end{bmatrix} \leqslant 2 \right\}. \tag{19}$$

The model $\mathcal{M}_T$ is composed of eight 11-dimensional cells $\mathcal{M}_\pi$, where $\pi=(\pi_1,\pi_2,\pi_3,\pi_4)$. As before, $\mathcal{M}_{\mathrm{id,id,id,id}} = \mathcal{M}_{(12),(12),(12),(12)}$. These cells are glued together along lower-dimensional models corresponding to forests obtained by deleting edges of the tree. For instance $\mathcal{M}_{\mathrm{id,id,id,id}}$ is glued to $\mathcal{M}_{\mathrm{id,id,(12),(12)}}$ along the model of two independent 2-leaf trees. It is also glued to $\mathcal{M}_{\mathrm{id,id,id,(12)}}$ along a model of a 3-leaf tree and an independent leaf, and similarly to 3 other cells. Finally, it is glued to the remaining cells $\mathcal{M}_{(12),\mathrm{id},(12),\mathrm{id}}$ and $\mathcal{M}_{(12),\mathrm{id,id},(12)}$ along even more degenerate models, of a forest with one 2-leaf tree and two singleton leaves. All eight cells intersect in the model of four independent leaves. The various strata correspond to $\mathbb{P}^3\times\mathbb{P}^3$, $\mathbb{P}^7\times\mathbb{P}^1$, $\mathbb{P}^3\times(\mathbb{P}^1)^2$ and $(\mathbb{P}^1)^4$.

The algebraic boundary of $\mathcal{M}_T$ has eight irreducible components of type (1), such as

$$\left\{ P\in\overline{\mathcal{M}_T}\colon \operatorname{rank}\begin{bmatrix} p_{1111} & p_{1112} & p_{1121} & p_{1122}\\ p_{1211} & p_{1212} & p_{1221} & p_{1222} \end{bmatrix} \leqslant 1 \right\}. \tag{20}$$

The 2×2-minors of (20) and 3×3-minors of (19) generate a prime ideal.

The ninth component of $\partial\overline{\mathcal{M}_T}$ comes from the internal edge and is of type (2). It is defined by the 4×4-determinant of either of the two flattenings other than (19). These two determinants are equal and irreducible on $\overline{\mathcal{M}_T}$, so they give the prime ideal of that component. $\square$

To prove Theorem 4.1, we consider the singular locus Θ_{sing} of the parametrization ϕ. By definition, Θ_{sing} is the closed subset of the cube Θ where the rank of the Jacobian matrix of ϕ drops to $2|E|$ or below.

Lemma 4.4. *Θ_{sing} is the subset of points in Θ where either the root distribution π has a zero entry, or some Markov matrix M_e is singular.*

Proof. A tree T_n with n leaves is obtained by attaching a cherry to a leaf ℓ of an $(n-1)$-leaf tree T_{n-1}. Assuming the matrices M_e on T_{n-1} are non-singular and π has nonzero entries, then the distribution for T_{n-1} flattens on the edge incident to ℓ to a $2 \times 2^{n-2}$-matrix A of rank 2. Let a_1, b_1 be the rows of A, and a_2, b_2 and a_3, b_3 the rows of the matrix parameters on the edges of the cherry. Then the distribution for T_n, appropriately flattened, is $a_1 \otimes a_2 \otimes a_3 + b_1 \otimes b_2 \otimes b_3$.

We next show that the map

$$\psi : (a_1, b_1, a_2, b_2, a_3, b_3) \mapsto a_1 \otimes a_2 \otimes a_3 + b_1 \otimes b_2 \otimes b_3$$

where the entries of a_2, b_2, a_3, b_3 sum to 1, is non-singular precisely at points where all pairs a_i, b_i are linearly independent. That ψ is singular at points where some pair a_i, b_i is dependent is straightforward. To show the rest of this claim, we allow arbitrary real entries in the vectors, to take advantage of a group action.

Let G be the subgroup of $GL(2^{n-2}) \times GL(2) \times GL(2)$ consisting of matrix triples (g_1, g_2, g_3) where the rows of g_2 and g_3 sum to 1. The group G acts on both the domain and range of ψ, and intertwines as

$$\psi(zg) = \psi(z)g, \quad g \in G.$$

Hence the Jacobian matrix of ψ has constant rank on each orbit. But the orbit of any point with all pairs a_i, b_i linearly independent is dense in the domain. Thus if ψ were singular at such a point, it would be singular everywhere. Since ψ is generically 2-to-1, that is impossible.

Note that the statement of the lemma for the 3-leaf tree follows from the previous paragraphs. Building the tree T_n inductively from T_{n-1} writes the Jacobian of ψ as a product of block matrices of smaller Jacobians. From this we see that Θ_{sing} consists of points where either π has a zero entry, or some M_e is singular. $\square$

Lemma 4.5. *If $\theta \in \Theta_{\text{sing}}$, then the fiber of θ intersects the boundary of Θ, i.e. there exists $\theta' \in \partial\Theta$ with $\phi(\theta') = \phi(\theta)$. Moreover, θ' can be found in a facet of Θ where some entry of a Markov matrix is zero.*

Proof. For a 3-leaf tree, rooted at the internal node, consider the parameters $\theta = (\pi, M_1, M_2, M_3) \in \Theta_{\text{sing}}$. If $\pi_i = 0$, then we may replace row i of any or all M_j with $(1, 0)$ to obtain θ'. Otherwise, suppose M_3 is singular yet there are no zeros in the parameters. Define θ' by $\pi' = \pi M_1$, $M_1' =$ the identity matrix, $M_2' = \operatorname{diag}(\pi')^{-1} M_1^T \operatorname{diag}(\pi) M_2$, and $M_3' = M_3$. One checks that $\phi(\theta') = \phi(\theta)$, and M_1' has a zero entry.

The result is derived inductively for larger trees, by viewing them as built from 3-leaf trees by attaching cherries. $\square$

Proof of Theorem 4.1. Points in $\text{Int}(\Theta) \backslash \Theta_{\text{sing}}$ must map to points in the relative interior of the model $\mathcal{M}_T$. Thus the boundary of $\mathcal{M}_T$ is a subset of $\phi(\partial\Theta) \cup \phi(\Theta_{\text{sing}})$. By Lemma 4.5, this is contained in $\phi(\partial\Theta)$.

To see that each of the components listed is a boundary component, we must show they have codimension 1 in the model, and a Zariski dense subset of points in them are limits of points outside the model. Since the complement of Θ_{sing} intersects these facets of Θ in non-empty open sets, the codimension is as needed. Since all elements of a fiber of the parameterization ϕ which contains non-singular points are related by label swapping, even when the map is extended outside Θ, one sees that non-singular points outside Θ cannot be mapped into the model, yet they are mapped arbitrarily close to the claimed component.

We have discussed all but two of the $4|E|+2$ facets of Θ. The remaining two facets, where an entry of the root distribution π is 0, contain only elements of Θ_{sing}, by Lemma 4.4. By Lemma 4.5, they lie in fibers with points where some entry of a Markov matrix is zero. Thus they are mapped into a component of the boundary already identified.

It remains to be shown that the equational descriptions given in (1) and (2) are valid on the respective components of the boundary of $\mathcal{M}_T$.

For a pendant edge e as in (1), we can assume that the root of the tree is located at the non-leaf end of e. The sets L_j span subtrees that intersect only at the root, and for each j there is a $2 \times 2^{|L_j|}$-matrix A_j, dependent only on the Markov matrices on edges of the subtrees, which expresses the joint probabilities of states at the leaves in L_j, conditioned on the root. In particular, $A_1 = M_e$. Denoting the rows of A_j by a_j, b_j, the r-dimensional flattening of our distribution is

$$\pi_1 \cdot a_1 \otimes a_2 \otimes \cdots \otimes a_r + \pi_2 \cdot b_1 \otimes b_2 \otimes \cdots \otimes b_r.$$

If $a_{1i} = 0$ (or $b_{1i} = 0$) then the ith slice in the first index has rank $\leqslant 1$.

For an internal edge e as in (2), assume the root of the tree is located at one end, and the Markov matrix on the edge is M_e, with rows a_e, b_e. Let L_1, L_2 and L_3, L_4 be the leaves of the subtrees attached to the respective ends of e. Then the $2^{|L_1 \cup L_3|} \times 2^{|L_2 \cup L_4|}$-matrix flattening incompatible with T can be expressed as

$$A^T \operatorname{diag}(\pi_1 a_e, \pi_2 b_e) B, \tag{21}$$

where A, B are $4 \times 2^{|L_1 \cup L_3|}$ and $4 \times 2^{|L_2 \cup L_4|}$-matrices, respectively. The entries of A depend on the parameters on the subtrees on L_1 and L_3, while those of B depend on the parameters on the subtrees on L_2 and L_4. Thus if M_e has a zero entry, then the 4×4-matrix $\operatorname{diag}(\pi_1 a_e, \pi_2 b_e)$ is singular, and hence the flattening (21) has rank at most 3. □

Several recent works found semialgebraic descriptions of the 2-state general Markov model on trees that is considered here. In [23] a different coordinate system is used, but [3] follows the same framework as this paper. Although some of the inequalities given in [3] hint at the form of the algebraic boundary determined in Theorem 4.1, those inequalities are considerably more complicated than our description here. While the inequalities provide tests for model membership, the relative simplicity of the algebraic boundary is expected to be advantageous for other purposes, such as understanding the geometry of log-likelihood functions over $\mathcal{M}_T$, and studying the limit behavior of iterative methods for parameter estimation such as Expectation Maximization (EM).

5. Towards higher rank

There are formidable obstacles to extending our results to tensors of rank $r > 2$. First of all, we do not know how to generalize the supermodular constraints. Second, we run into problems of non-identifiability, even in the case of matrices ($n = 2$). Recall also (e.g. from [7, Example 4.1.2]) that a nonnegative matrix of rank 3 need not have nonnegative rank 3. The topological analysis given by Mond et al. [17] illustrates well the difficulties involved in obtaining a characterization of the semialgebraic set of $d_1 \times d_2$-matrices of nonnegative rank $\leqslant 3$.

On the other hand, for tensors of dimension $n \geqslant 3$, rank decompositions are often identifiable when r is small relative to $d_1, d_2, \ldots, d_n$. In such situations, when the model is identifiable, one might hope for results similar to Theorems 1.2 and 4.1. However, a third obstacle arises: in order to characterize algebraic boundaries, one needs a version of Lemma 4.4 for the singular locus Θ_{sing} of the model parameterization ϕ.

In what follows, we illustrate these issues for two rank 3 examples.

Example 5.1. Consider the set $\mathcal{M}$ of $3 \times 3 \times 2$ tensors of nonnegative rank $\leqslant 3$. This is a smallest format for which rank 3 decompositions are generically unique, up to label swapping. Normalizing the tensor entries to sum to 1, we obtain $\mathcal{M}$ as the image of the map

$$\begin{aligned} &\phi : \Theta \to \Delta_{17}, \\ &(\pi; a_1, a_2, a_3; b_1, b_2, b_3; c_1, c_2, c_3) \\ &\quad \mapsto \pi_1 \cdot a_1 \otimes a_2 \otimes a_3 + \pi_2 \cdot b_1 \otimes b_2 \otimes b_3 + \pi_3 \cdot c_1 \otimes c_2 \otimes c_3, \end{aligned}$$

where $\pi \in \mathbb{R}^3_{\geqslant 0}$, $a_i, b_i, c_i \in \mathbb{R}^3_{\geqslant 0}$ for $i = 1, 2$, and $a_3, b_3, c_3 \in \mathbb{R}^2_{\geqslant 0}$ all have coordinate sum 1. The domain is the polytope $\Theta = \Delta_2 \times (\Delta_2 \times \Delta_2 \times \Delta_1)^3$. The facets of Θ are given by parameters being 0. This map is generically 6-to-1, so $\phi(\Theta) = \mathcal{M}$ is full-dimensional in the simplex Δ_{17}. The Zariski closure $\overline{\mathcal{M}}$ is the entire projective space $\mathbb{P}^{17}$ of $3 \times 3 \times 2$ tensors.

The algebraic boundary $\overline{\partial \mathcal{M}}$ has eight irreducible components:

(a) Two components $\overline{\phi(\{a_{3k} = 0\})} = \overline{\phi(\{b_{3k} = 0\})} = \overline{\phi(\{c_{3k} = 0\})}$, for $k = 1, 2$, given by the 3×3-slice $P_k = [p_{**k}]$ having rank $\leqslant 2$.
(b) Three components given, for $i = 1, 2, 3$, by the 3×3-matrix $P_1 \cdot (P_2)^{-1}$ having an eigenvector with zero ith coordinate.
(c) Three components given, for $j = 1, 2, 3$, by the 3×3-matrix $P_1^T \cdot (P_2)^{-T}$ having an eigenvector with zero jth coordinate.

The two components (a) are the cubic hypersurfaces given by the determinants of P_1 and P_2. The six components (b) and (c) are hypersurfaces of degree 6. For instance, the polynomial K that defines the (b) component $\overline{\phi(\{a_{13} = 0\})} = \overline{\phi(\{b_{13} = 0\})} = \overline{\phi(\{c_{13} = 0\})}$ equals

$$\begin{aligned} K = {} & p_{111}p_{212}p_{321}^2p_{332}^2 - 2p_{111}p_{212}p_{321}p_{322}p_{331}p_{332} + p_{111}p_{212}p_{322}^2p_{331}^2 \\ & - p_{111}p_{222}p_{311}p_{321}p_{332}^2 + p_{111}p_{222}p_{311}p_{322}p_{331}p_{332} - p_{111}p_{222}p_{312}p_{322}p_{331}^2 \\ & + p_{111}p_{222}p_{312}p_{321}p_{331}p_{332} + p_{111}p_{232}p_{311}p_{321}p_{322}p_{332} - p_{112}p_{211}p_{321}^2p_{332}^2 \\ & - p_{111}p_{232}p_{311}p_{322}^2p_{331} + p_{111}p_{232}p_{312}p_{321}p_{322}p_{331} - p_{111}p_{232}p_{312}p_{321}^2p_{332} \\ & + 2p_{112}p_{211}p_{321}p_{322}p_{331}p_{332} - p_{112}p_{211}p_{322}^2p_{331}^2 + p_{112}p_{221}p_{311}p_{321}p_{332}^2 \\ & - p_{112}p_{221}p_{311}p_{322}p_{331}p_{332} - p_{112}p_{221}p_{312}p_{321}p_{331}p_{332} + p_{112}p_{221}p_{312}p_{322}p_{331}^2 \\ & - p_{112}p_{231}p_{311}p_{321}p_{322}p_{332} + p_{112}p_{231}p_{311}p_{322}^2p_{331} + p_{112}p_{231}p_{312}p_{321}^2p_{332} \\ & - p_{112}p_{231}p_{312}p_{321}p_{322}p_{331} - p_{121}p_{212}p_{311}p_{321}p_{332}^2 + p_{121}p_{212}p_{311}p_{322}p_{331}p_{332} \\ & + p_{121}p_{212}p_{312}p_{321}p_{331}p_{332} - p_{121}p_{212}p_{312}p_{322}p_{331}^2 + p_{121}p_{222}p_{311}^2p_{332}^2 \\ & - 2p_{121}p_{222}p_{311}p_{312}p_{331}p_{332} + p_{121}p_{222}p_{312}^2p_{331}^2 - p_{121}p_{232}p_{311}^2p_{322}p_{332} \\ & + p_{121}p_{232}p_{311}p_{312}p_{321}p_{332} + p_{121}p_{232}p_{311}p_{312}p_{322}p_{331} - p_{121}p_{232}p_{312}^2p_{321}p_{331} \\ & + p_{122}p_{211}p_{311}p_{321}p_{332}^2 - p_{122}p_{211}p_{311}p_{322}p_{331}p_{332} - p_{122}p_{211}p_{312}p_{321}p_{331}p_{332} \\ & + p_{122}p_{211}p_{312}p_{322}p_{331}^2 - p_{122}p_{221}p_{311}^2p_{332}^2 + 2p_{122}p_{221}p_{311}p_{312}p_{331}p_{332} \\ & - p_{122}p_{221}p_{312}^2p_{331}^2 + p_{122}p_{231}p_{311}^2p_{322}p_{332} - p_{122}p_{231}p_{311}p_{312}p_{321}p_{332} \\ & - p_{122}p_{231}p_{311}p_{312}p_{322}p_{331} + p_{122}p_{231}p_{312}^2p_{321}p_{331} + p_{131}p_{212}p_{311}p_{321}p_{322}p_{332} \\ & - p_{131}p_{212}p_{311}p_{322}^2p_{331} - p_{131}p_{212}p_{312}p_{321}^2p_{332} + p_{131}p_{212}p_{312}p_{321}p_{322}p_{331} \\ & - p_{131}p_{222}p_{311}^2p_{322}p_{332} + p_{131}p_{222}p_{311}p_{312}p_{321}p_{332} + p_{131}p_{222}p_{311}p_{312}p_{322}p_{331} \end{aligned}$$

$$\begin{aligned}
&- p_{131}p_{222}p_{312}^2p_{321}p_{331} + p_{131}p_{232}p_{311}^2p_{322}^2 - 2p_{131}p_{232}p_{311}p_{312}p_{321}p_{322}\\
&+ p_{131}p_{232}p_{312}^2p_{321}^2 - p_{132}p_{211}p_{311}p_{321}p_{322}p_{332} + p_{132}p_{211}p_{311}p_{322}^2p_{331}\\
&+ p_{132}p_{211}p_{312}p_{321}^2p_{332} - p_{132}p_{211}p_{312}p_{321}p_{322}p_{331} + p_{132}p_{221}p_{311}^2p_{322}p_{332}\\
&- p_{132}p_{221}p_{311}p_{312}p_{321}p_{332} - p_{132}p_{221}p_{311}p_{312}p_{322}p_{331} + p_{132}p_{221}p_{312}^2p_{321}p_{331}\\
&- p_{132}p_{231}p_{311}^2p_{322}^2 + 2p_{132}p_{231}p_{311}p_{312}p_{321}p_{322} - p_{132}p_{231}p_{312}^2p_{321}^2.
\end{aligned}$$

This polynomial was found using the reduced Kalman matrix in [21, Eq. (1.5)]. Under the parametrization, this expression factors as

$$\begin{aligned}
K = {}&\pi_1^2\pi_2^2\pi_3^2 a_{13}b_{13}c_{13}(a_{31}b_{32} - a_{32}b_{31})(a_{31}c_{32} - a_{32}c_{31})(b_{31}c_{32} - b_{32}c_{31})\\
&\times \det[a_1, b_1, c_1]\det[a_2, b_2, c_2]^2
\end{aligned}$$

To prove that there is nothing else in $\overline{\partial\mathcal{M}}$, we proceed as in Lemma 4.5. We examine the Jacobian of ϕ, which has rank 17 at generic points of Θ. Using symbolic computation, we find that its singular locus Θ_{sing}, where the rank drops, decomposes into three types of components:

(1) points with π having a zero entry, at which the rank of the Jacobian is generically 12,
(2) points with two of a_3, b_3, c_3 equal, at which the rank of the Jacobian is generically 15,
(3) points with a_1, b_1, c_1 (or with a_2, b_2, c_2) linearly dependent, at which the rank of the Jacobian is generically 14.

We now show that every point of Θ_{sing} lies in a fiber of ϕ that intersects the boundary of the polytope Θ. For singular points of type (1), if say $\pi_1 = 0$, one may replace an a_i with any other vector to obtain another point in the fiber, so this is clear. For type (2), if say $b_3 = c_3$, then

$$\pi_2 \cdot b_1 \otimes b_2 \otimes b_3 + \pi_3 \cdot c_1 \otimes c_2 \otimes c_3 = A \otimes b_3$$

for a 3×3-matrix A of nonnegative rank 2. Since one can find a nonnegative rank 2 decomposition of A with zeros in some vector entry, we can construct the desired boundary point in the fiber.

For type (3) singular points we argue as follows. Suppose P is the image of parameters where a_2, b_2, c_2 are dependent. Let d be a nonzero, nonnegative vector in the span of a_1, b_1 and consider the line of tensors

$$B(t) = P - \pi_3 \cdot (c_1 + td) \otimes c_2 \otimes c_3 = a_1 \otimes a_2 \otimes a_3 + \pi_2 \cdot b_1 \otimes b_2 \otimes b_3 - t\pi_3 \cdot d \otimes c_2 \otimes c_3.$$

The sets $\{a_1, b_1, d\}$, $\{a_2, b_2, c_2\}$ and $\{a_3, b_3, c_3\}$ are dependent, so the three matrix flattenings of $B(t)$ have rank $\leqslant 2$. Thus for all t, the tensor $B(t)$ has border rank $\leqslant 2$. Also, $B(0)$ has nonnegative rank 2.

Since $B(t)$ fails to have nonnegative rank 2 for $t \gg 0$, there exists $t_0 \geqslant 0$ such that $B(t_0)$ lies on the boundary of the tensors of nonnegative rank 2. By Theorem 1.2, $B(t_0)$ has a nonnegative rank 2 decomposition with a zero coordinate in some parameter vector. Since

$$P = B(t_0) + \pi_3(c_1 + t_0 d) \otimes c_2 \otimes c_3,$$

the tensor P has a nonnegative rank 3 decomposition with a zero parameter. Hence P lies in one of the eight varieties seen in (a), (b), (c).

We note that a distribution P with invertible slices P_i lies in $\mathcal{M}$ if and only if all eigenvalues and eigenvectors of the 3×3-matrices $P_1 \cdot (P_2)^{-1}$ and $P_1^T \cdot (P_2)^{-T}$ are nonnegative. Here, one should be able to pass to the closure and infer a nice semialgebraic description of $\mathcal{M}$. □

Example 5.2. Let $\mathcal{M}$ be the set of $2 \times 2 \times 2 \times 2$ tensors of nonnegative rank $\leqslant 3$. As in the previous example, we normalize tensors to have entries summing to one. This model is not identifiable: the generic fiber of its stochastic parametrization ϕ is a curve. Facets of the parameter polytope $\Theta = \Delta_2 \times (\Delta_1)^{12}$ are mapped into subsets of the model $\mathcal{M}$ that are Zariski dense in $\overline{\mathcal{M}}$. We note that the

13-dimensional variety $\overline{\mathcal{M}}$ is a complete intersection of degree $16 = 4 \cdot 4$ in $\mathbb{P}^{15}$. It is defined by the determinants of any two of the three 4×4-flattenings of P.

Components of the algebraic boundary $\partial\overline{\mathcal{M}}$ might now be obtained from codimension 2 faces of the polytope Θ. For instance, write

$$P = \pi_1 a_1 \otimes a_2 \otimes a_3 \otimes a_4 + \pi_2 b_1 \otimes b_2 \otimes b_3 \otimes b_4 + \pi_3 c_1 \otimes c_2 \otimes c_3 \otimes c_4,$$

and consider the face $\{a_{11} = b_{22} = 0\}$ of Θ. Then $\overline{\phi(\{a_{11} = b_{22} = 0\})}$ is suspected to be a component in $\partial\overline{\mathcal{M}}$. This variety has dimension 12 and degree 56 in $\mathbb{P}^{15}$. It is defined, as a subscheme of $\overline{\mathcal{M}}$, by 55 polynomials of degree 8 in the 16 unknowns p_{ijkl}. The smallest of these degree 8 polynomials has 96 terms, and we shall resist the temptation to list them. The semialgebraic geometry of $\mathcal{M}$ deserves further analysis. □

References

[1] R. Ahlswede, D. Daykin, An inequality for the weights of two families of sets, their unions and intersections, Z. Wahrscheinlichkeitstheor. Verw. Geb. 43 (1978) 183–185.

[2] E. Allman, J. Rhodes, Phylogenetic ideals and varieties for the general Markov model, Adv. in Appl. Math. 40 (2008) 127–148.

[3] E.S. Allman, J.A. Rhodes, A. Taylor, A semialgebraic description of the general Markov model on phylogenetic trees, preprint, arXiv:1212.1200.

[4] D. Altmann, K. Altmann, Estimating vaccine coverage by computer algebra, IMA J. Math. Appl. Med. Biol. 17 (2000) 137–146.

[5] V. de Silva, L. Lim, Tensor rank and the ill-posedness of the best low-rank approximation problem, SIAM J. Matrix Anal. Appl. 30 (2008) 1084–1127.

[6] J. Draisma, J. Kuttler, On the ideals of equivariant tree models, Math. Ann. 344 (2009) 619–644.

[7] M. Drton, B. Sturmfels, S. Sullivant, Lectures on Algebraic Statistics, Oberwolfach Semin., vol. 39, Birkhäuser Verlag, Basel, 2009.

[8] A. Engström, P. Hersh, B. Sturmfels, Toric cubes, Circ. Mat. Palermo 62 (2013) 67–78.

[9] M. Friedlander, K. Hatz, Computing nonnegative tensor factorizations, Optim. Methods Softw. 23 (2008) 631–647.

[10] L. Garcia, M. Stillman, B. Sturmfels, Algebraic geometry of Bayesian networks, J. Symbolic Comput. 39 (2005) 331–355.

[11] D. Grayson, M. Stillman, Macaulay 2, a software system for research in algebraic geometry, available at http://www.math.uiuc.edu/Macaulay2/.

[12] P.W. Holland, P.R. Rosenbaum, Conditional association and unidimensionality in monotone latent variable models, Ann. Statist. 14 (1986) 1523–1543.

[13] S. Klaere, V. Liebscher, An algebraic analysis of the two state Markov model on tripod trees, Math. Biosci. 237 (2012) 38–48.

[14] J.M. Landsberg, Tensors: Geometry and Applications, Grad. Stud. Math., vol. 128, American Mathematical Society, Providence, RI, 2012.

[15] J.M. Landsberg, L. Manivel, On ideals of secant varieties of Segre varieties, Found. Comput. Math. 4 (2004) 397–422.

[16] L.-H. Lim, P. Comon, Nonnegative approximations of nonnegative tensors, J. Chemom. 23 (2009) 432–441.

[17] D. Mond, J.Q. Smith, D. van Straten, Stochastic factorizations, sandwiched simplices and the topology of the space of explanations, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 459 (2003) 2821–2845.

[18] J. Morton, L. Pachter, A. Shiu, B. Sturmfels, O. Wienand, Convex rank tests and semigraphoids, SIAM J. Discrete Math. 23 (2009) 1117–1134.

[19] J. Pearl, M. Tarsi, Structuring causal trees, in: Complexity of Approximately Solved Problems, Morningside Heights, 1985, J. Complexity 2 (1986) 60–77.

[20] C. Raicu, Secant varieties of Segre–Veronese varieties, Algebra Number Theory 6 (2012) 1817–1868.

[21] G. Ottaviani, B. Sturmfels, Matrices with eigenvectors in a given subspace, Proc. Amer. Math. Soc. 141 (2013) 1219–1232.

[22] M. Steel, B. Faller, Markovian log-supermodularity, and its applications in phylogenetics, Appl. Math. Lett. 22 (2009) 1141–1144.

[23] P. Zwiernik, J.Q. Smith, Implicit inequality constraints in a binary tree model, Electron. J. Stat. 5 (2011) 1276–1312.