The Annals of Statistics
2015, Vol. 43, No. 1, 422–461
DOI: 10.1214/14-AOS1282

FIXED POINTS OF THE EM ALGORITHM AND NONNEGATIVE RANK BOUNDARIES

BY KAIE KUBJAS, ELINA ROBEVA[1] AND BERND STURMFELS[2]

Aalto University, University of California, Berkeley and University of California, Berkeley

Mixtures of r independent distributions for two discrete random variables can be represented by matrices of nonnegative rank r. Likelihood inference for the model of such joint distributions leads to problems in real algebraic geometry that are addressed here for the first time. We characterize the set of fixed points of the Expectation–Maximization algorithm, and we study the boundary of the space of matrices with nonnegative rank at most 3. Both of these sets correspond to algebraic varieties with many irreducible components.

1. Introduction. The rth mixture model $\mathcal{M}$ of two discrete random variables X and Y expresses the conditional independence statement $X \perp\!\!\!\perp Y|Z$, where Z is a hidden (or latent) variable with r states. Assuming that X and Y have m and n states, respectively, their joint distribution is written as an $m \times n$-matrix of nonnegative rank $\leq r$ whose entries sum to 1. This mixture model is also known as the *naive Bayes model*. Its graphical representation is shown in Figure 1.

A collection of i.i.d. samples from a joint distribution is recorded in a nonnegative matrix

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mn} \end{bmatrix}.$$

Here, u_{ij} is the number of observations in the sample with $X = i$ and $Y = j$. The sample size is $u_{++} = \sum_{i,j} u_{ij}$. It is standard practice to fit the model to the data U using the Expectation–Maximization (EM) algorithm. However, it has been pointed out in the literature that EM has several issues (see the next paragraph for details) and one has to be careful when using it. Our goal is to better understand this algorithm by studying its mathematical properties in some detail.

One of the main issues of Expectation–Maximization is that it does not provide a certificate for having found the global optimum. The geometry of the algorithm has

Received December 2013; revised October 2014.

[1]Supported by a UC Berkeley Graduate Fellowship.

[2]Supported by NSF Grant DMS-09-68882.

MSC2010 subject classifications. 62F10, 13P25.

Key words and phrases. Maximum likelihood, EM algorithm, mixture model, nonnegative rank.

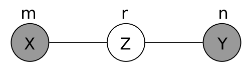


FIG. 1. *Graphical model on two observed variables and one hidden variable.*

been a topic for debate among statisticians since the seminal paper of Dempster, Laird and Rubin [13]. Murray [30] responded with a warning for practitioners to be aware of the existence of multiple stationary points. Beale [6] also brought this up, and Fienberg [18] referred to the possibility that the MLE lies on the boundary of the parameter space. A recent discussion of this issue was presented by Zwiernik and Smith [36], Section 3, in their analysis of inferential problems arising from the semialgebraic geometry of a latent class model. The fact that our model fails to be identifiable was highlighted by Fienberg et al. in [19], Section 4.2.3. This poses additional difficulties, and it forces us to distinguish between the boundary of the parameter space and the boundary of the model. The image of the former contains the latter.

The EM algorithm aims to maximize the log-likelihood function of the model $\mathcal{M}$. In doing so, it approximates the data matrix U with a product of nonnegative matrices $A \cdot B$ where A has r columns and B has r rows. In Section 3, we review the EM algorithm in our context. Here, it is essentially equivalent to the widely used method of Lee and Seung [26] for *nonnegative matrix factorization*. The nonnegative rank of matrices has been studied from a broad range of perspectives, including computational geometry [1, 10], topology [29], contingency tables [7, 19], complexity theory [28, 33] and convex optimization [17]. We here present the approach from algebraic statistics [14, 31].

Maximum likelihood estimation for the model $\mathcal{M}$ is a nonconvex optimization problem. Any algorithm that promises to compute the MLE $\widehat{P}$ will face the following fundamental dichotomy. The optimal matrix $\widehat{P}$ either lies in the relative interior of $\mathcal{M}$ or it lies in the model boundary $\partial\mathcal{M}$.

If $\widehat{P}$ lies in the relative interior of $\mathcal{M}$, then the situation is nice. In this case, $\widehat{P}$ is a critical point for the likelihood function on the manifold of rank r matrices. There are methods by Hauenstein et al. [24] for finding the MLE with certificate. The ML degree, which they compute, bounds the number of critical points, and hence all candidates for the global maximizer $\widehat{P}$. However, things are more difficult when $\widehat{P}$ lies in the boundary $\partial\mathcal{M}$. In that case, $\widehat{P}$ is generally not a critical point for the likelihood function in the manifold of rank r matrices, and none of the results on ML degrees in [14, 19, 23–25] are applicable. The present paper is the first to address the question of how $\widehat{P}$ varies when it occurs in the boundary $\partial\mathcal{M}$. Table 1 underscores the significance of our approach. As the matrix size grows, the boundary case is much more likely to happen for randomly chosen input U. The details for choosing U and the simulation study that generated Table 1 will be described in Example 3.4.

TABLE 1
Percentage of data matrices whose maximum likelihood estimate $\widehat{P}$ lies in the boundary $\partial\mathcal{M}$

	Size				
Rank	**4 × 4**	**5 × 5**	**6 × 6**	**7 × 7**	**8 × 8**
3	4.4%	23%	49%	62%	85%
4		7%	37%	71%	95%
5			10%	55%	96%
6				20%	75%
7					24%

We now summarize the contents of this article. Section 2 furnishes an introduction to the geometry of the mixture model $\mathcal{M}$ from Figure 1. We define the *topological boundary* of $\mathcal{M}$ and the *algebraic boundary* of $\mathcal{M}$, and we explain how these two notions of boundary differ. Concrete numerical examples for 4×4-matrices of rank 3 demonstrate how $\widehat{P}$ behaves as the data U vary.

In Section 3, we review the EM algorithm for the model $\mathcal{M}$, and we identify its fixed points in the parameter space. The main result is the characterization of the set of fixed points in Theorem 3.5.

In Section 4, we identify $\mathcal{M}$ with the set of matrices of nonnegative rank at most 3. Theorem 4.1 gives a quantifier-free formula for this semialgebraic set. The importance of finding such a formula was already stressed in the articles [3, 4]. The resulting membership test for $\mathcal{M}$ is very fast and can be applied to matrices that contain parameters. The proof of Theorem 4.1 is based on the familiar characterization of nonnegative rank in terms of nested polytopes [1, 10, 33], and, in particular, on work of Mond et al. [29] on the structure of critical configurations in the plane (shown in Figure 5).

In Section 5, we return to Expectation–Maximization, and we study the system of equations that characterize the EM fixed points. Proposition 5.1 characterizes its solutions in the interior of $\mathcal{M}$. Even in the smallest interesting case, $m = n = 4$ and $r = 3$, the variety of all EM fixed points has a huge number of irreducible components, to be determined and interpreted in Theorem 5.5.

The most interesting among these are the 288 components that delineate the topological boundary $\partial\mathcal{M}$ inside the simplex Δ_{15}. These are discussed in Examples 5.7 and 6.2. Explicit matrices that lie on these components are featured in (6.5) and in Examples 2.1, 2.2 and 3.2. In Proposition 6.3, we resolve a problem left open in [24, 25] concerning the ML degree arising from $\partial\mathcal{M}$. The main result in Section 6 is Theorem 6.1 which characterizes the algebraic boundary of $m \times n$-matrices of nonnegative rank 3. The commutative algebra of the irreducible components in that boundary is the content of Theorem 6.4. Corollary 6.6 furnishes a quantifier-free semialgebraic formula for $\partial\mathcal{M}$.

The proofs of all lemmas, propositions and corollaries appear in Appendix A. A review of basic concepts in algebraic geometry is given in Appendix B. This will help the reader understand the technicalities of our main results. Supplementary materials and software are posted at the website http://math.berkeley.edu/~bernd/EM/boundaries.html. Our readers will find code in `R`, `Macaulay2` and `Magma` for various sampling experiments, prime decompositions, semialgebraic formulas and likelihood equations discussed in this paper.

The methods presented here are not limited to the matrix model $\mathcal{M}$, but are applicable to a wide range of statistical models for discrete data, especially those used in computational biology [31]. Such models include phylogenetic models [2, 4] and hidden Markov models [12]. The most immediate generalization is to the rth mixture model of several random variables. It consists of all distributions corresponding to tensors of nonnegative rank at most r. In other words, we replace $m \times n$-matrices by tensors of arbitrary format. The geometry of the case $r = 2$ was studied in depth by Allman et al. [3]. For each of these models, there is a natural EM algorithm, with an enormous number of stationary points. The model itself is a complicated semialgebraic set, and the MLE typically occurs on the boundary of that set. For binary tree models, this was shown in [36], Section 3.

This article introduces tools needed to gain a complete understanding of these EM fixed points and model boundaries. We here study them for the graphical model in Figure 1. Already in this very simple case, we discovered patterns that are surprisingly rich. Thus, the present work serves as a blueprint for future research in real algebraic geometry that underlies statistical inference.

2. Model geometry. We begin with a geometric introduction of the likelihood inference problem to be studied. Let Δ_{mn-1} denote the probability simplex of nonnegative $m \times n$-matrices $P = [p_{ij}]$ with $p_{++} = 1$. Our model $\mathcal{M}$ is the subset of Δ_{mn-1} consisting of all matrices of the form

$$P = A \cdot \Lambda \cdot B, \tag{2.1}$$

where A is a nonnegative $m \times r$-matrix whose columns sum to 1, Λ is a nonnegative $r \times r$ diagonal matrix whose entries sum to 1, and B is a nonnegative $r \times n$-matrix whose rows sum to 1. The triple of parameters (A, Λ, B) represents conditional probabilities for the graphical model in Figure 1. In particular, the kth column of A is the conditional probability distribution of X given that $Z = k$, the kth row of B is the conditional probability distribution given that $Z = k$, and the diagonal of Λ is the probability distribution of Z. The parameter space in which A, Λ, B lie is the convex polytope $\Theta = (\Delta_{m-1})^r \times \Delta_{r-1} \times (\Delta_{n-1})^r$. Our model $\mathcal{M}$ is the image of the trilinear map

$$\phi : \Theta \to \Delta_{mn-1}, \qquad (A, \Lambda, B) \mapsto P. \tag{2.2}$$

We seek to learn the model parameters (A, Λ, B) by maximizing the likelihood function

$$\binom{u_{++}}{u} \cdot \prod_{i=1}^{m} \prod_{j=1}^{n} p_{ij}^{u_{ij}} \tag{2.3}$$

over $\mathcal{M}$. This is equivalent to maximizing the log-likelihood function

$$\ell_U = \sum_{i=1}^{m} \sum_{j=1}^{n} u_{ij} \cdot \log\left(\sum_{k=1}^{r} a_{ik}\lambda_k b_{kj}\right) \tag{2.4}$$

over $\mathcal{M}$. One issue that comes up immediately is that the model parameters are not identifiable:

$$\dim(\Theta) = r(m+n) - r - 1 \quad \text{but} \quad \dim(\mathcal{M}) = r(m+n) - r^2 - 1. \tag{2.5}$$

The first expression is the sum of the dimensions of the simplices in the product that defines the parameter space Θ. The second one counts the degrees of freedom in a rank r matrix of format $m \times n$. The typical fiber, that is, the preimage of a point in the image of (2.2), is a semialgebraic set of dimension $r^2 - r$. This is the *space of explanations* whose topology was studied by Mond et al. in [29]. Likelihood inference cannot distinguish among points in each fiber, so it is preferable to regard MLE not as an unconstrained optimization problem in Θ but as a constrained optimization problem in $\mathcal{M}$. The aim of this paper is to determine its constraints.

Let $\mathcal{V}$ denote the set of real $m \times n$-matrices P of rank $\leq r$ satisfying $p_{++} = 1$. This set is a variety because it is given by the vanishing of a set of polynomials, namely, the $(r+1) \times (r+1)$ minors of the matrix P plus the linear constraint $p_{++} = 1$. A point $P \in \mathcal{M}$ is an *interior point* of $\mathcal{M}$ if there is an open ball $U \subset \Delta_{mn-1}$ that contains P and satisfies $U \cap \mathcal{V} = U \cap \mathcal{M}$. We call $P \in \mathcal{M}$ a *boundary point* of $\mathcal{M}$ if it is not an interior point. The set of all such points is denoted by $\partial\mathcal{M}$ and called the *topological boundary* of $\mathcal{M}$. In other words, $\partial\mathcal{M}$ is the boundary of $\mathcal{M}$ inside $\mathcal{V}$. The variety $\mathcal{V}$ is the Zariski closure of the set $\mathcal{M}$; see Appendix B. In other words, the set of polynomials that vanish on $\mathcal{M}$ is exactly the same as the set of polynomials that vanish on $\mathcal{V}$. Our model $\mathcal{M}$ is a full-dimensional subset of the variety $\mathcal{V}$ and is given by a set of polynomial inequalities inside $\mathcal{V}$.

Fix U, r and $P \in \mathcal{M}$ as above. A matrix P is a nonsingular point on $\mathcal{V}$ if and only if the rank of P is exactly r. In this case, its tangent space $\mathrm{T}_P(\mathcal{V})$ has dimension $r(m+n) - r^2 - 1$, which, as expected, equals $\dim(\mathcal{M})$. We call P a *critical point* of the log-likelihood function ℓ_U if $P \in \mathcal{M}$, P is a nonsingular point for $\mathcal{V}$, that is, $\mathrm{rank}(P) = r$, and the gradient of ℓ_U is orthogonal to the tangent space $\mathrm{T}_P(\mathcal{V})$. Thus, the critical points are the nonnegative real solutions of the various *likelihood equations* derived in [14, 24, 31, 35] to address the MLE problem for $\mathcal{M}$. In other words, the critical points are the solutions obtained by using the Lagrange multipliers method for maximizing the likelihood function over the set $\mathcal{V}$. In the language of algebraic statistics, the critical points are those points in $\mathcal{M}$ that are accounted for by the *ML degree* of the variety $\mathcal{V}$.

Table 1 shows that the global maximum $\widehat{P}$ of ℓ_U is often a noncritical point. This means that the MLE lies on the topological boundary $\partial\mathcal{M}$. The ML degree of the variety $\mathcal{V}$ is irrelevant for assessing the algebraic complexity of such $\widehat{P}$. Instead, we need the ML degree of the boundary, as given in Proposition 6.3, as well as the ML degrees for the lower-dimensional boundary strata.

The following example illustrates the concepts we have introduced so far and what they mean.

EXAMPLE 2.1. Fix $m = n = 4$ and $r = 3$. For any integers $a \geq b \geq 0$, consider the data matrix

$$U_{a,b} = \begin{bmatrix} a & a & b & b \\ a & b & a & b \\ b & a & b & a \\ b & b & a & a \end{bmatrix}. \tag{2.6}$$

Note that $\operatorname{rank}(U_{a,b}) \leq 3$. For $a = 1$ and $b = 0$, this is the standard example [10] of a nonnegative matrix whose nonnegative rank exceeds its rank. Thus, $\frac{1}{8}U_{1,0}$ is a probability distribution in $\mathcal{V} \setminus \mathcal{M}$. Within the 2-parameter family (2.6), the topological boundary $\partial\mathcal{M}$ is given by the linear equation $b = (\sqrt{2} - 1)a$. This follows from the computations in [7], Section 5, and [29], Section 5. We conclude that

$$\frac{1}{8(a+b)}U_{a,b} \qquad \text{lies in } \mathcal{V} \setminus \mathcal{M} \quad \text{if and only if} \quad b < (\sqrt{2} - 1)a. \tag{2.7}$$

For integers $a > b \geq 0$ satisfying (2.7), the likelihood function (2.3) for $U_{a,b}$ has precisely eight global maxima on our model $\mathcal{M}$. These are the following matrices, each divided by $8(a + b)$:

$$\begin{bmatrix} a & a & b & b \\ v & w & t & u \\ w & v & u & t \\ s & s & r & r \end{bmatrix}, \quad \begin{bmatrix} v & t & w & u \\ a & b & a & b \\ s & r & s & r \\ w & u & v & t \end{bmatrix}, \quad \begin{bmatrix} t & v & u & w \\ r & s & r & s \\ b & a & b & a \\ u & w & t & v \end{bmatrix},$$

$$\begin{bmatrix} r & r & s & s \\ t & u & v & w \\ u & t & w & v \\ b & b & a & a \end{bmatrix}, \quad \begin{bmatrix} a & v & w & s \\ a & w & v & s \\ b & t & u & r \\ b & u & t & r \end{bmatrix}, \quad \begin{bmatrix} v & a & s & w \\ t & b & r & u \\ w & a & s & v \\ u & b & r & t \end{bmatrix},$$

$$\begin{bmatrix} t & r & b & u \\ v & s & a & w \\ u & r & b & t \\ w & s & a & v \end{bmatrix}, \quad \begin{bmatrix} r & t & u & b \\ r & u & t & b \\ s & v & w & a \\ s & w & v & a \end{bmatrix}.$$

This claim can be verified by exact symbolic computation, or by validated numerics as in the proof of [24], Theorem 4.4. Here, t is the unique simple real root of the cubic equation

$$\begin{aligned}&(6a^3+16a^2b+14ab^2+4b^3)t^3-(20a^4+44a^3b+8ab^3+32a^2b^2)t^2\\&\quad+(22a^5+43a^4b+30a^3b^2+7a^2b^3)t-(8a^6+16a^5b+10a^4b^2+2a^3b^3)\\&\quad\quad=0.\end{aligned}$$

To fill in the other entries of these nonnegative rank 3 matrices, we use the rational formulas

$$\begin{aligned}s&=\frac{(a+b)t-a^2}{a},\qquad u=\frac{tb}{a},\\w&=-\frac{t(3a^2+5ab+2b^2)t-4a^3-5a^2b-2ab^2}{2a^3+a^2b},\\r&=\frac{2a^2+ab-(a+b)t}{a},\\v&=\frac{(3a^2+5ab+2b^2)t^2-(6a^3+8a^2b+3ab^2)t+6a^3b+2a^2b^2+4a^4}{2a^3+a^2b}.\end{aligned}$$

These formulas represent an exact algebraic solution to the MLE problem in this case. They describe the multivalued map $(a,b)\mapsto \widehat{P}_{a,b}$ from the data to the eight maximum likelihood estimates. This allows us to understand exactly how these solutions behave as the matrix entries a and b vary.

The key point is that the eight global maxima lie in the model boundary $\partial\mathcal{M}$. They are not critical points of ℓ_U on the rank 3 variety $\mathcal{V}$. They will not be found by the methods in [24, 31, 35]. Instead, we used results about the algebraic boundary in Section 5 to derive the eight solutions.

We note that this example can be seen as an extension of [24], Theorem 4.4, which offers a similar parametric analysis for the data set of the "100 Swiss Francs Problem" studied in [19, 35].

We now introduce the concept of algebraic boundary. Recall that the topological boundary $\partial\mathcal{M}$ of the model $\mathcal{M}$ is a semialgebraic subset inside the probability simplex Δ_{mn-1}. Its dimension is

$$\dim(\partial\mathcal{M})=\dim(\mathcal{M})-1=rm+rn-r^2-2.$$

Any quantifier-free semialgebraic description of $\partial\mathcal{M}$ will be a complicated Boolean combination of polynomial equations and polynomial inequalities. This can be seen for $r=3$ in Corollary 6.6.

To simplify the situation, it is advantageous to relax the inequalities and keep only the equations. This replaces the topological boundary of $\mathcal{M}$ by a much simpler object, namely the algebraic boundary of $\mathcal{M}$. To be precise, we define the *algebraic boundary* to be the Zariski closure $\overline{\partial\mathcal{M}}$ of the topological boundary $\partial\mathcal{M}$. Thus, $\overline{\partial\mathcal{M}}$ is a subvariety of codimension 1 inside the variety $\mathcal{V} \subset \mathbb{P}^{mn-1}$. Theorem 6.1 will show us that $\overline{\partial\mathcal{M}}$ can have many irreducible components.

The following two-dimensional family of matrices illustrates the results to be achieved in this paper. These enable us to discriminate between the topological boundary $\partial\mathcal{M}$ and the algebraic boundary $\overline{\partial\mathcal{M}}$, and to understand how these boundaries sit inside the variety $\mathcal{V}$.

EXAMPLE 2.2. Consider the following 2-parameter family of 4×4-matrices:

$$P(x,y) = \begin{bmatrix} 51 & 9 & 64 & 9 \\ 27 & 63 & 8 & 8 \\ 3 & 34 & 40 & 31 \\ 30 & 25 & 80 & 35 \end{bmatrix} + x \cdot \begin{bmatrix} 1 & 1 & 3 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} + y \cdot \begin{bmatrix} 5 & 4 & 1 & 1 \\ 5 & 1 & 5 & 1 \\ 1 & 5 & 1 & 5 \\ 1 & 1 & 5 & 5 \end{bmatrix}.$$

This was chosen so that $P(0,0)$ lies in a unique component of the topological boundary $\partial\mathcal{M}$. The equation $\det(P(x,y)) = 0$ defines a plane curve $\mathcal{C}$ of degree 4. This is the thin black curve shown in Figure 2. In our family, this quartic curve $\mathcal{C}$ represents the Zariski closure $\mathcal{V}$ of the model $\mathcal{M}$.

The algebraic boundary $\overline{\partial\mathcal{M}}$ is the variety described in Example 6.2. The quartic curve $\mathcal{C}$ meets $\overline{\partial\mathcal{M}}$ in 1618 real points (x, y). Of these 1618 points, precisely 188 satisfy the constraint $P(x, y) \geq 0$. These 188 points are the landmarks for our analysis. They are shown in blue on the right in Figure 2. In addition, we mark the unique point where the curve $\mathcal{C}$ intersects the boundary polygon defined by $P(x, y) \geq 0$. This is the leftmost point, defined by $\{\det(P(x, y)) = x + 5y + 8 = 0\}$. It equals

$$(-3.161429, -0.967714). \tag{2.8}$$

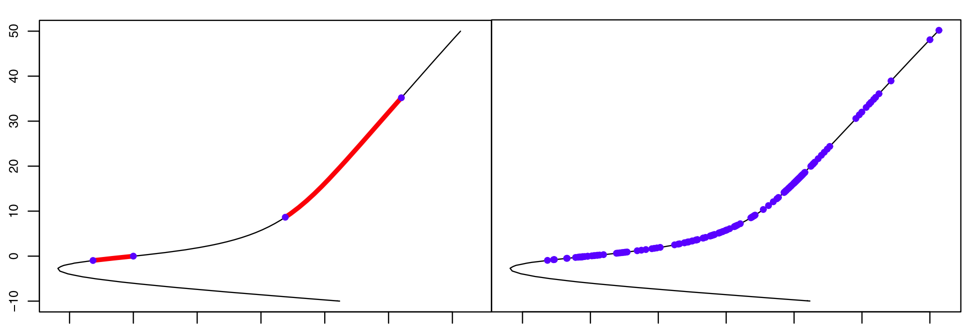


FIG. 2. *In a two-dimensional family of* 4×4-*matrices, the matrices of rank* 3 *form a quartic curve. The mixture model, shown in red, has two connected components. Its topological boundary consists of four points* (*on the left*). *The algebraic boundary includes many more points* (*on the right*). *Currently, there is no known way to obtain the four points on the topological boundary* (*in the left picture*) *without first considering all points on the algebraic boundary* (*in the right picture*).

We examined the 187 arcs on $\mathcal{C}$ between consecutive points of $\overline{\partial\mathcal{M}}$ as well as the two arcs at the ends. For each arc we checked whether it lies in $\mathcal{M}$. This was done by a combination of the EM algorithm in Section 3 and Theorem 4.1. Precisely 96 of the 189 arcs were found to lie in $\mathcal{M}$. These form two connected components on the curve $\mathcal{C}$, namely 19 arcs between (2.8) and (0, 0), and

$$\begin{aligned}&\text{76 arcs between}\\ &\qquad (11.905773, 8.642630) \quad \text{and} \quad (21.001324, 35.202110).\end{aligned} \tag{2.9}$$

These four points represent the topological boundary $\partial\mathcal{M}$. We conclude that, in the 2-dimensional family $P(x, y)$, the model $\mathcal{M}$ is the union of the two red arcs shown on the left in Figure 2.

Our theory of EM fixed points distinguishes between the (relatively open) red arcs and their blue boundary points. For the MLE problem, the red points are critical while the blue points are not critical. By Table 1, the MLE is more likely to be blue than red, for larger values of m and n.

This example demonstrates that the algebraic methods of Sections 4, 5 and 6 are indispensable when one desires a reliable analysis of model geometries, such as that illustrated in Figure 2. To apply a method for finding the critical points of a function, for example, Lagrange multipliers, the domain of the function needs to be given by equality constraints only. But using only these constraints, one cannot detect the maxima lying on the topological boundary. For finding the critical points of the likelihood function on the topological boundary by using the same methods, one needs to relax the inequality constraints and consider only the equations defining the topological boundary. Therefore, one needs to find the critical points on the algebraic boundary $\overline{\partial\mathcal{M}}$ of the model.

3. Fixed points of Expectation–Maximization. The EM algorithm is an iterative method for finding local maxima of the likelihood function (2.3). It can be viewed as a discrete dynamical system on the polytope $\Theta = (\Delta_{m-1})^r \times \Delta_{r-1} \times (\Delta_{n-1})^r$. Algorithm 1 presents the version in [31], Section 1.3.

The alternating sequence of E-steps and M-steps defines trajectories in the parameter polytope Θ. The log-likelihood function (2.4) is nondecreasing along each trajectory (cf. [31], Theorem 1.15). In fact, the value can stay the same only at a fixed point of the EM algorithm. See Dempster et al. [13] for the general version of EM and its increasing behavior and convergence.

DEFINITION 3.1. An *EM fixed point* for a given table U is any point (A, Λ, B) in the polytope $\Theta = (\Delta_{m-1})^r \times \Delta_{r-1} \times (\Delta_{n-1})^r$ to which the EM algorithm can converge if it is applied to (U, r).

Every global maximum $\widehat{P}$ of ℓ_U is among the EM fixed points. One hopes that $\widehat{P}$ has a large basin of attraction, and that the initial parameter choice (A, Λ, B)

Algorithm 1 Function EM(U, r)

Select random $a_1, a_2, \ldots, a_r \in \Delta_{m-1}$, random $\lambda \in \Delta_{r-1}$, and random $b_1, b_2, \ldots, b_r \in \Delta_{n-1}$.

Run the following steps until the entries of the $m \times n$-matrix P converge.

E-step: *Estimate the $m \times r \times n$-table that represents this expected hidden data*:

  Set $v_{ikj} := \frac{a_{ik}\lambda_k b_{kj}}{\sum_{l=1}^{r} a_{il}\lambda_l b_{lj}} u_{ij}$ for $i = 1, \ldots, m$, $k = 1, \ldots, r$ and $j = 1, \ldots, n$.

M-step: *Maximize the likelihood function of the model* ●—●—● for the hidden data:

  Set $\lambda_k := \sum_{i=1}^{m} \sum_{j=1}^{n} v_{ikj}/u_{++}$ for $k = 1, \ldots, r$.

  Set $a_{ik} := (\sum_{j=1}^{n} v_{ikj})/(u_{++}\lambda_k)$ for $k = 1, \ldots, r$ and $i = 1, \ldots, m$.

  Set $b_{kj} := (\sum_{i=1}^{m} v_{ikj})/(u_{++}\lambda_k)$ for $k = 1, \ldots, r$ and $j = 1, \ldots, n$.

Update *the estimate of the joint distribution for our mixture model* ●—○—●:

  Set $p_{ij} := \sum_{k=1}^{r} a_{ik}\lambda_k b_{kj}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

Return P.

---

gives a trajectory that converges to $\widehat{P}$. However, this need not be the case, since the EM dynamics on Θ has many fixed points other than $\widehat{P}$. Our aim is to understand all of these.

EXAMPLE 3.2. The following data matrix is obtained by setting $a = 1, b = 0$ in Example 2.1:

$$U = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Among the EM fixed points for this choice of U with $r = 3$ we find the probability distributions

$$P_1 = \frac{1}{24}\begin{bmatrix} 3 & 3 & 0 & 0 \\ 2 & 0 & 4 & 0 \\ 0 & 2 & 0 & 4 \\ 1 & 1 & 2 & 2 \end{bmatrix}, \qquad P_2 = \frac{1}{16}\begin{bmatrix} 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \end{bmatrix} \quad \text{and}$$

$$P_3 = \frac{1}{48}\begin{bmatrix} 4 & 8 & 0 & 0 \\ 3 & 0 & 4 & 5 \\ 5 & 4 & 0 & 3 \\ 0 & 0 & 8 & 4 \end{bmatrix},$$

and their orbits under the symmetry group of U. For instance, the orbit of P_1 is obtained by setting $s = \frac{1}{3}, r = \frac{2}{3}, v = \frac{2}{3}, t = \frac{4}{3}, w = u = 0$ in the eight matrices in Example 2.1. Over 98% of our runs with random starting points in Θ converged to one of these eight global maximizers of ℓ_U. Matrices in the orbits of P_2, respectively, P_3 were approached only rarely (less than 2%) by the EM algorithm.

LEMMA 3.3. *The following are equivalent for a point (A, Λ, B) in the parameter polytope Θ*:

(1) *The point (A, Λ, B) is an EM fixed point.*

(2) *If we start EM with (A, Λ, B) instead of a random point, then EM converges to (A, Λ, B).*

(3) *The point (A, Λ, B) remains fixed after one completion of the E-step and the M-step.*

It is often believed (and actually stated in [31], Theorem 1.5) that every EM fixed point is a critical point of the log-likelihood function ℓ_U. This statement is not true for the definition of "critical" given in Section 2. In fact, for many instances U, the global maximum $\widehat{P}$ is not critical.

To underscore this important point and its statistical relevance, we tested the EM algorithm on random data matrices U for a range of models with $m = n$. The following example explains Table 1.

EXAMPLE 3.4. In our first simulation, we generated random matrices U from the uniform distribution on Δ_{mn-1} by using R and then scaling to get integer entries. For each matrix U, we ran the EM algorithm 2000 times to ensure convergence with high probability to the global maximum $\widehat{P}$ on $\mathcal{M}$. Each run had 2000 steps. We then checked whether $\widehat{P}$ is a critical point of ℓ_U using the rank criterion in [24], equation (2.3). Our results are reported in Table 1. The main finding is that, with high probability as the matrix size increases, the MLE $\widehat{P}$ lands on the topological boundary $\partial\mathcal{M}$, and it fails to be critical.

In a second simulation, we started with matrices $A \in \mathbb{N}^{m\times r}$ and $B \in \mathbb{N}^{r\times n}$ whose entries were sampled uniformly from $\{0, 1, \dots, 100\}$. We then fixed $P \in \mathcal{M}$ to be the $m \times n$ probability matrix given by AB divided by the sum of its entries. We finally took Tmn samples from the distribution P and recorded the results in an $m \times n$ data matrix U. Thereafter, we applied EM to U. We observed the following. If $T \geq 20$ then the fraction of times the MLE lies in $\partial\mathcal{M}$ is very close to 0. When $T \leq 10$ though, this fraction was higher than the results reported in Table 1. For $T = 10$ and $m = n = 4$, $r = 3$, this fraction was 13%, for $m = n = 5$, $r = 3$, it was 23%, and for $m = n = 5$, $r = 4$, it was 17%. Therefore, based on these experiments, in order to have the MLE be a critical point in $\mathcal{M}$, one should have at least 20 times more samples than entries of the matrix.

This brings our attention to the problem of identifying the fixed points of EM. If we could compute all EM fixed points, then this would reveal the global maximizer of ℓ_U. Since a point is EM fixed if and only if it stays fixed after an E-step and an M-step, we can write rational function equations for the EM fixed points in Θ:

$$\lambda_k = \frac{1}{u_{++}} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{a_{ik}\lambda_k b_{kj}}{\sum_{l=1}^{r} a_{il}\lambda_l b_{lj}} u_{ij} \qquad \text{for all } k,$$

$$a_{ik} = \frac{1}{\lambda_k u_{++}} \sum_{j=1}^{n} \frac{a_{ik}\lambda_k b_{kj}}{\sum_{l=1}^{r} a_{il}\lambda_l b_{lj}} u_{ij} \qquad \text{for all } i, k,$$

$$b_{kj} = \frac{1}{\lambda_k u_{++}} \sum_{i=1}^{m} \frac{a_{ik}\lambda_k b_{kj}}{\sum_{l=1}^{r} a_{il}\lambda_l b_{lj}} u_{ij} \qquad \text{for all } k, j.$$

Our goal is to understand the solutions to these equations for a fixed positive matrix U. We seek to find the variety they define in the polytope Θ and the image of that variety in $\mathcal{M}$.

In the EM algorithm, we usually start with parameters $a_{ik}, \lambda_k, b_{kj}$ that are strictly positive. The a_{ik} or b_{kl} may become zero in the limit, but the parameters λ_k always remain positive when the u_{ij} are positive since the entries of each column of A and each row of B sum to 1. This justifies that we cancel out the factors λ_k in our equations. After this, the first equation is implied by the other two. Therefore, the set of all EM fixed points is a variety, and it is characterized by

$$a_{ik} = \frac{1}{u_{++}} \sum_{j=1}^{n} \frac{a_{ik} b_{kj}}{\sum_{l=1}^{r} a_{il}\lambda_l b_{lj}} u_{ij} \qquad \text{for all } i, k,$$

$$b_{kj} = \frac{1}{u_{++}} \sum_{i=1}^{m} \frac{a_{ik} b_{kj}}{\sum_{l=1}^{r} a_{il}\lambda_l b_{lj}} u_{ij} \qquad \text{for all } k, j.$$

Suppose that a denominator $\sum_l a_{il}\lambda_l b_{lj}$ is zero at a point in Θ. Then $a_{ik}b_{kj} = 0$ for all k, and the expression $\frac{a_{ik}b_{kj}}{\sum_{l=1}^{r} a_{il}\lambda_l b_{lj}}$ would be considered 0. Using the identity $p_{ij} = \sum_{l=1}^{r} a_{il}\lambda_l b_{lj}$, we can rewrite our two fixed point equations in the form

$$(3.1)\qquad \begin{aligned} a_{ik}\left(\sum_{j=1}^{n}\left(u_{++} - \frac{u_{ij}}{p_{ij}}\right) b_{kj}\right) &= 0 \qquad \text{for all } k, i \quad \text{and} \\ b_{kj}\left(\sum_{i=1}^{m}\left(u_{++} - \frac{u_{ij}}{p_{ij}}\right) a_{ik}\right) &= 0 \qquad \text{for all } k, j. \end{aligned}$$

Let R denote the $m \times n$ matrix with entries $r_{ij} = u_{++} - \frac{u_{ij}}{p_{ij}}$. The matrix R is the gradient of the log-likelihood function $\ell_U(P)$, as seen in [24], equation (3.1). With this, our fixed point equations are

$$(3.2)\qquad \begin{aligned} a_{ik}\left(\sum_{j=1}^{n} r_{ij} b_{kj}\right) &= 0 \qquad \text{for all } k, i \quad \text{and} \\ b_{kj}\left(\sum_{i=1}^{m} r_{ij} a_{ik}\right) &= 0 \qquad \text{for all } k, j. \end{aligned}$$

We summarize our discussion in the following theorem, with (3.2) rewritten in matrix form.

THEOREM 3.5. *The variety of EM fixed points in the polytope Θ is defined by the equations*

$$A \star (R \cdot B^T) = 0, \qquad B \star (A^T \cdot R) = 0, \tag{3.3}$$

where R is the gradient matrix of the log-likelihood function and $\star$ denotes the Hadamard product. The subset of EM fixed points that are critical points is defined by $R \cdot B^T = 0$ and $A^T \cdot R = 0$.

PROOF. Since (3.3) is equivalent to (3.2), the first sentence is proved by the derivation above. For the second sentence, we consider the normal space of the variety $\mathcal{V}$ at a rank r matrix $P = A\Lambda B$. This is the orthogonal complement of the tangent space $\mathrm{T}_P(\mathcal{V})$. The normal space can be expressed as the kernel of the linear map $Q \mapsto (Q \cdot B^T, A^T \cdot Q)$. Hence, $R = \mathrm{grad}_P(\ell_U)$ is perpendicular to $\mathrm{T}_P(\mathcal{V})$ if and only if $R \cdot B^T = 0$ and $A^T \cdot R = 0$. Therefore, the polynomial equations (3.3) define the Zariski closure of the set of parameters for which P is critical. $\square$

The variety defined by (3.3) is reducible. In Section 5, we shall present a detailed study of its irreducible components, along with a discussion of their statistical interpretation. As a preview, we here decompose the variety of EM fixed points in the simplest possible case.

EXAMPLE 3.6. Let $m = n = 2$, $r = 1$, and consider the ideal generated by the cubics in (3.3):

$$\begin{aligned}\mathcal{F} = \langle & a_{11}(r_{11}b_{11} + r_{12}b_{12}), a_{21}(r_{21}b_{11} + r_{22}b_{12}), \\ & b_{11}(a_{11}r_{11} + a_{21}r_{21}), b_{12}(a_{11}r_{12} + a_{21}r_{22})\rangle.\end{aligned}$$

The software `Macaulay2` [22] computes a primary decomposition into 12 components:

$$\begin{aligned}\mathcal{F} = {} & \langle r_{11}r_{22} - r_{12}r_{21}, a_{11}r_{11} + a_{21}r_{21}, a_{11}r_{12} + a_{21}r_{22}, b_{11}r_{11} \\ & \qquad + b_{12}r_{12}, b_{11}r_{21} + b_{12}r_{22}\rangle \\ & \cap \langle a_{11}, r_{21}, r_{22}\rangle \cap \langle a_{21}, r_{11}, r_{12}\rangle \cap \langle r_{12}, r_{22}, b_{11}\rangle \cap \langle r_{11}, r_{21}, b_{12}\rangle \\ & \cap \langle a_{11}, r_{22}, b_{11}\rangle \cap \langle a_{11}, r_{21}, b_{12}\rangle \cap \langle a_{21}, r_{12}, b_{11}\rangle \cap \langle a_{21}, r_{11}, b_{12}\rangle \\ & \cap \langle a_{11}, a_{21}\rangle \cap \langle b_{11}, b_{12}\rangle \cap (\langle a_{11}, a_{21}\rangle^2 + \langle b_{11}, b_{12}\rangle^2 + \mathcal{F}).\end{aligned} \tag{3.4}$$

The last primary ideal is embedded. Thus, $\mathcal{F}$ is not a radical ideal. Its radical requires an extra generator of degree 5. The first 11 ideals in (3.4) are the minimal primes of $\mathcal{F}$. These give the irreducible components of the variety $V(\mathcal{F})$. The first ideal represents the critical points in $\mathcal{M}$.

4. Matrices of nonnegative rank three. While the EM algorithm operates in the polytope Θ of model parameters (A, Λ, B), the mixture model $\mathcal{M}$ lives in the simplex $\Delta_{mn-1} \subset \mathbb{R}^{m\times n}$ of all joint distributions. The parametrization ϕ is not identifiable. The topology of its fibers was studied by Mond et al. [29], with focus on the first nontrivial case, when the rank r is three. We build on their work to derive a semialgebraic characterization of $\mathcal{M}$. This section is self-contained. It can be read independently from our earlier discussion of the EM algorithm. It is aimed at all readers interested in nonnegative matrix factorization, regardless of its statistical relevance.

We now fix $r = 3$. Let A be a real $m \times 3$-matrix with rows $a_1, \dots, a_m$, and B a real $3 \times n$-matrix with columns $b_1, \dots, b_n$. The vectors $b_j \in \mathbb{R}^3$ represent points in the projective plane $\mathbb{P}^2$. We view the a_i as elements in the dual space $(\mathbb{R}^3)^*$. These represent lines in $\mathbb{P}^2$. Geometric algebra (a.k.a. Grassmann–Cayley algebra [34]) furnishes two bilinear operations,

$$\vee : \mathbb{R}^3 \times \mathbb{R}^3 \to (\mathbb{R}^3)^* \quad \text{and} \quad \wedge : (\mathbb{R}^3)^* \times (\mathbb{R}^3)^* \to \mathbb{R}^3.$$

These correspond to the classical cross product in 3-space. Geometrically, $a_i \wedge a_j$ is the intersection point of the lines a_i and a_j in $\mathbb{P}^2$, and $b_i \vee b_j$ is the line spanned by the points b_i and b_j in $\mathbb{P}^2$. The pairing $(\mathbb{R}^3)^* \times \mathbb{R}^3 \to \mathbb{R}$ can be denoted by either $\vee$ or $\wedge$. With these conventions, the operations $\vee$ and $\wedge$ are alternating, associative and distributive. For instance, the minor

$$a_i \wedge a_j \wedge a_k = \det(a_i, a_j, a_k) \tag{4.1}$$

vanishes if and only if the lines a_i, a_j and a_k are concurrent. Likewise, the polynomial

$$\begin{aligned}(a_i \wedge a_j) &\vee b_{i'} \vee b_{k'} \\ &= a_{i1}a_{j2}b_{1i'}b_{2k'} - a_{i1}a_{j2}b_{1k'}b_{2i'} + a_{i1}a_{j3}b_{1i'}b_{3k'} - a_{i1}a_{j3}b_{1k'}b_{3i'} \\ &\quad - a_{i2}a_{j1}b_{1i'}b_{2k'} + a_{i2}a_{j1}b_{1k'}b_{2i'} + a_{i2}a_{j3}b_{2i'}b_{3k'} - a_{i2}a_{j3}b_{2k'}b_{3i'} \\ &\quad - a_{i3}a_{j1}b_{1i'}b_{3k'} + a_{i3}a_{j1}b_{1k'}b_{3i'} - a_{i3}a_{j2}b_{2i'}b_{3k'} + a_{i3}a_{j2}b_{2k'}b_{3i'}\end{aligned} \tag{4.2}$$

expresses the condition that the lines a_i and a_j intersect in a point on the line given by $b_{i'}$ and $b_{k'}$. Of special interest is the following formula involving four rows of A and three columns of B:

$$\big(((a_i \wedge a_j) \vee b_{i'}) \wedge a_k\big) \vee \big(((a_i \wedge a_j) \vee b_{j'}) \wedge a_l\big) \vee b_{k'}. \tag{4.3}$$

Its expansion is a bihomogeneous polynomial of degree $(6, 3)$ with 330 terms in (A, B).

A matrix $P \in \mathbb{R}^{m\times n}$ has nonnegative rank ≤ 3 if it admits a factorization $P = AB$ with A and B nonnegative. The set of such matrices P with $p_{++} = 1$ is precisely the mixture model $\mathcal{M}$ discussed in the earlier sections. Comparing with (2.1), we here subsume the diagonal matrix Λ into either A or B. In what follows, we consider the set $\mathcal{N}$ of pairs (A, B) whose product AB has nonnegative rank ≤ 3. Thus, $\mathcal{N}$ is a semialgebraic subset of $\mathbb{R}^{m\times 3} \oplus \mathbb{R}^{3\times n}$. We shall prove:

THEOREM 4.1. *A pair (A, B) is in $\mathcal{N}$ if and only if $AB \geq 0$ and the following condition holds: either* $\operatorname{rank}(AB) < 3$, *or* $\operatorname{rank}(AB) = 3$ *and there exist indices* $i, j \in [m]$, $i', j' \in [n]$ *such that*:

sign(4.1) *is the same or zero for all* $k \in [m] \setminus \{i, j\}$,
and sign(4.2) *is the same or zero for all* $k' \in [n] \setminus \{i'\}$,
and sign((4.2)$[i' \to j']$) *is the same or zero for all* $k' \in [n] \setminus \{j'\}$,
and (4.3) $\cdot$ (4.3) $[k \leftrightarrow l] \geq 0$ *for all* $\{k, l\} \subseteq [m] \setminus \{i, j\}$ *and* $k' \in [n] \setminus \{i', j'\}$,
or there exist $i, j \in [n]$, $i', j' \in [m]$ *such that these conditions hold after swapping* A *with* B^T.

Here, $[m] = \{1, 2, \ldots, m\}$, and the notation $[i' \to j']$ means that the index i' is replaced by the index j' in the preceding expression, and $[k \leftrightarrow l]$ means that k and l are switched.

Theorem 4.1 is our main result in Section 4. It gives a finite disjunction of conjunctions of polynomial inequalities in A and B, and thus a quantifier-free first order formula for $\mathcal{N}$. This represents our mixture model as follows: to test whether P lies in $\mathcal{M}$, check whether $\operatorname{rank}(P) \leq 3$; if yes, compute any rank 3 factorization $P = AB$ and check whether (A, B) lies in $\mathcal{N}$. Code for performing these computations in `Macaulay2` is posted on our website.

Theorem 4.1 is an algebraic translation of a geometric algorithm. For an illustration, see Figure 3. In the rest of the section, we will study the geometric description of nonnegative rank that leads to the algorithm. Let P be a nonnegative $m \times n$ matrix of rank r. We write $\operatorname{span}(P)$ and $\operatorname{cone}(P)$ for the linear space and the cone spanned by the columns of P, and we define

$$\mathcal{A} = \operatorname{span}(P) \cap \Delta_{m-1} \quad \text{and} \quad \mathcal{B} = \operatorname{cone}(P) \cap \Delta_{m-1}. \tag{4.4}$$

The matrix P has a size r nonnegative factorization if and only if there exists a polytope Δ with r vertices such that $\mathcal{B} \subseteq \Delta \subseteq \mathcal{A}$; see [29], Lemma 2.2. Without loss of generality, we will assume in the rest of this section that the vertices of Δ lie on the boundary of $\mathcal{A}$. We write $\mathcal{M}_r$ for the set of $m \times n$-matrices of nonnegative rank $\leq r$. Here is an illustration that is simpler than Example 2.2:

EXAMPLE 4.2. In [17], Section 2.7.2, the following family of matrices of rank ≤ 3 is considered:

$$P(a, b) = \begin{bmatrix} 1-a & 1+a & 1+a & 1-a \\ 1-b & 1-b & 1+b & 1+b \\ 1+a & 1-a & 1-a & 1+a \\ 1+b & 1+b & 1-b & 1-b \end{bmatrix}. \tag{4.5}$$

Here, $\mathcal{B}$ is a rectangle and $\mathcal{A} = \{x \in \Delta_3 : x_1 - x_2 + x_3 - x_4 = 0\}$ is a square, see Figure 4. Using Theorem 4.1, we can check that $P(a, b)$ lies in $\mathcal{M}_3$ if and only if $ab + a + b \leq 1$.

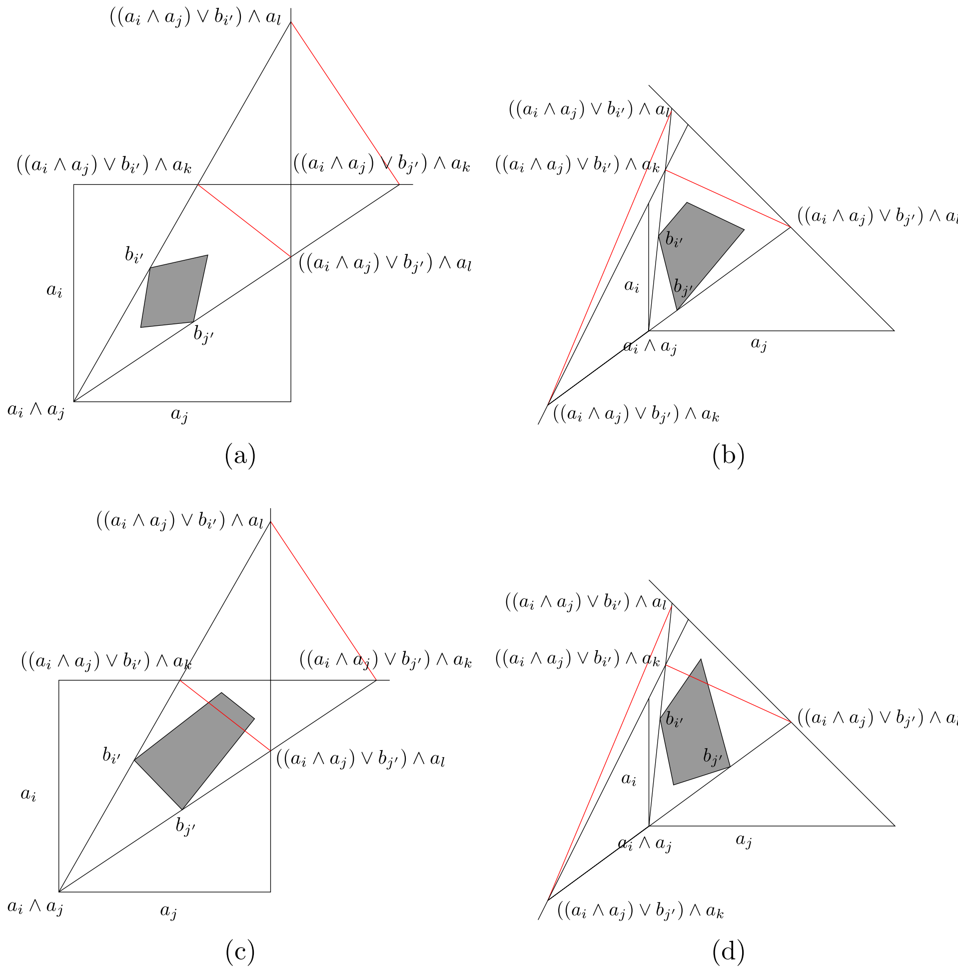


FIG. 3. *In the diagrams* (a) *and* (b), *the conditions of Theorem* 4.1 *are satisfied for the chosen* i, j, i', j'. *In the diagrams* (c) *and* (d), *the conditions of Theorem* 4.1 *fail for the chosen* i, j, i', j'.

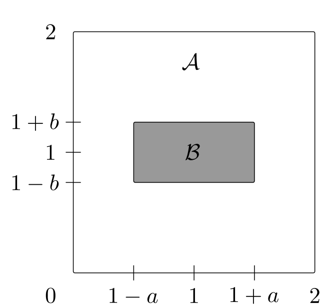
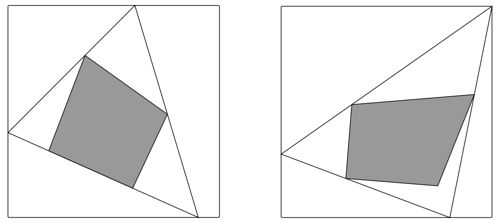


FIG. 4. *The matrix* $P(a, b)$ *defines a nested pair of rectangles.*



(a) (b)

FIG. 5. *Critical configurations.*

LEMMA 4.3. *A matrix $P \in \mathbb{R}_{\geq 0}^{m\times n}$ of rank r lies in the interior of $\mathcal{M}_r$ if and only if there exists an $(r-1)$-simplex $\Delta \subseteq \mathcal{A}$ such that $\mathcal{B}$ is contained in the interior of Δ. It lies on the boundary of $\mathcal{M}_r$ if and only if every $(r-1)$-simplex Δ with $\mathcal{B} \subseteq \Delta \subseteq \mathcal{A}$ contains a vertex of $\mathcal{B}$ on its boundary.*

For $r = 3$, Mond et al. [29] prove the following result. Suppose $\mathcal{B} \subseteq \Delta \subseteq \mathcal{A}$ and every edge of Δ contains a vertex of $\mathcal{B}$. Then $t\mathcal{B} \subseteq \Delta' \subseteq \mathcal{A}$ for some triangle Δ' and some $t > 1$, unless:

(a) an edge of Δ contains an edge of $\mathcal{B}$, or
(b) a vertex of Δ coincides with a vertex of $\mathcal{A}$.

Here, the dilate $t\mathcal{B}$ is taken with respect to a point in the interior of $\mathcal{B}$. By Lemma 4.3, this means that P lies in the interior of $\mathcal{M}_3^{m\times n}$ unless one of (a) and (b) holds. The conditions (a) and (b) are shown in Figure 5. For the proof of this result, we refer to [29], Lemmas 3.10 and 4.3.

COROLLARY 4.4. *A matrix $P \in \mathcal{M}_3$ lies on the boundary of $\mathcal{M}_3$ if and only if*:

- *P has a zero entry, or*
- $\operatorname{rank}(P) = 3$ *and if Δ is any triangle with $\mathcal{B} \subseteq \Delta \subseteq \mathcal{A}$ then every edge of Δ contains a vertex of $\mathcal{B}$, and* (a) *or* (b) *holds.*

COROLLARY 4.5. *A matrix $P \in \mathbb{R}_{\geq 0}^{m\times n}$ has nonnegative rank ≤ 3 if and only if*:

- $\operatorname{rank}(P) < 3$, *or*
- $\operatorname{rank}(P) = 3$ *and there exists a triangle Δ with $\mathcal{B} \subseteq \Delta \subseteq \mathcal{A}$ such that a vertex of Δ coincides with a vertex of $\mathcal{A}$, or*
- $\operatorname{rank}(P) = 3$ *and there exists a triangle Δ with $\mathcal{B} \subseteq \Delta \subseteq \mathcal{A}$ such that an edge of Δ contains an edge of $\mathcal{B}$.*

Corollary 4.5 provides a geometric algorithm similar to that of Aggarwal et al. [1] for checking whether a matrix has nonnegative rank 3. For the algorithm, we need to consider one condition for every vertex of $\mathcal{A}$ and one condition for every edge of $\mathcal{B}$. We now explain these conditions.

Let v be a vertex of $\mathcal{A}$. Let b_1, b_2 be the vertices of $\mathcal{B}$ such that $l_1 = \overline{vb_1}$ and $l_2 = \overline{vb_2}$ support $\mathcal{B}$. Let Δ be the convex hull of v and the other two intersection points of the lines l_1, l_2 with the boundary of $\mathcal{A}$. If $\mathcal{B} \subseteq \Delta$, then P has nonnegative rank 3.

Let l be the line spanned by an edge of $\mathcal{B}$. Let v_1, v_2 be the intersection points of l with $\partial\mathcal{A}$. Let b_1, b_2 be the vertices of $\mathcal{B}$ such that $l_1 = \overline{v_1b_1}$ and $l_2 = \overline{v_2b_2}$ support $\mathcal{B}$. Let v_3 be the intersection point of l_1 and l_2. If $\operatorname{conv}(v_1, v_2, v_3) \subseteq \mathcal{A}$, then P has nonnegative rank 3.

PROOF OF THEOREM 4.1. Let $\operatorname{rank}(P) = 3$ and consider any factorization $P = AB$ where $a_1, \dots, a_m \in (\mathbb{R}^3)^*$ are the row vectors of A and $b_1, \dots, b_n \in \mathbb{R}^3$ are the column vectors of B. The map $x \mapsto Ax$ identifies $\mathbb{R}^3$ with the common column space of A and P. Under this identification, and by passing from 3-dimensional cones to polygons in $\mathbb{R}^2$, we can assume that the edges of $\mathcal{A}$ are given by $a_1, \dots, a_m$ and the vertices of $\mathcal{B}$ are given by $b_1, \dots, b_n$.

To test whether P belongs to $\mathcal{M}_3$, we use the geometric conditions in Corollary 4.5. These still involve a quantifier over Δ. Our aim is to translate them into the given quantifier-free formula, referring only to the vertices b_i of $\mathcal{B}$ and the edges a_j of $\mathcal{A}$. First, we check with the sign condition on (4.1) that the intersection point $a_i \wedge a_j$ defines a vertex of $\mathcal{A}$. Next we verify that the lines $(a_i \wedge a_j) \vee b_{i'}$ and $(a_i \wedge a_j) \vee b_{j'}$ are supporting $\mathcal{B}$, that is, all vertices of $\mathcal{B}$ lie on the same side of the lines $(a_i \wedge a_j) \vee b_{i'}$ and $(a_i \wedge a_j) \vee b_{j'}$. For this, we use the sign conditions on (4.2) and (4.2) $[i' \to j']$.

Finally, we need to check whether all vertices of $\mathcal{B}$ belong to the convex hull of $a_i \wedge a_j$ and the other two intersection points of the lines $(a_i \wedge a_j) \vee b_{i'}$ and $(a_i \wedge a_j) \vee b_{j'}$ with the boundary of $\mathcal{A}$. Fix $\{k, l\} \subseteq [m] \setminus \{i, j\}$. If either the line $(a_i \wedge a_j) \vee b_{i'}$ intersects a_k or the line $(a_i \wedge a_j) \vee b_{j'}$ intersects a_l outside $\mathcal{A}$, then the polygon $\mathcal{B}$ lies completely on one side of the line $(((a_i \wedge a_j) \vee b_{i'}) \wedge a_k) \vee (((a_i \wedge a_j) \vee b_{j'}) \wedge a_l)$. Similarly, if either the line $(a_i \wedge a_j) \vee b_{i'}$ intersects a_l or the line $(a_i \wedge a_j) \vee b_{j'}$ intersects a_k outside $\mathcal{A}$, then the polygon $\mathcal{B}$ lies completely on one side of the line $(((a_i \wedge a_j) \vee b_{i'}) \wedge a_l) \vee (((a_i \wedge a_j) \vee b_{j'}) \wedge a_k)$. Then the condition (4.3) $\cdot$ (4.3) $[k \leftrightarrow l] \geq 0$ is automatically satisfied for all $k' \in [n] \setminus \{i', j'\}$. If the intersection points $((a_i \wedge a_j) \vee b_{i'}) \wedge a_k$ and $((a_i \wedge a_j) \vee b_{j'}) \wedge a_l$ are on the boundary of $\mathcal{A}$, then the polygon $\mathcal{B}$ is on one side of $(((a_i \wedge a_j) \vee b_{i'}) \wedge a_l) \vee (((a_i \wedge a_j) \vee b_{j'}) \wedge a_k)$. In this case, we use the conditions (4.3) $\cdot$ (4.3) $[k \leftrightarrow l] \geq 0$ to check whether $\mathcal{B}$ is also on one side of the line $(((a_i \wedge a_j) \vee b_{i'}) \wedge a_k) \vee (((a_i \wedge a_j) \vee b_{j'}) \wedge a_l)$. For an illustration, see Figure 3. $\square$

We wish to reiterate that the semialgebraic formula for our model in Theorem 4.1 is quantifier-free. It is a finite Boolean combination of polynomial inequalities with rational coefficients.

COROLLARY 4.6. *If a rational $m \times n$ matrix P has nonnegative rank ≤ 3, then there exists a nonnegative rank ≤ 3 factorization $P = AB$ where all entries of A and B are rational numbers.*

This answers a question of Cohen and Rothblum in [10] for matrices of nonnegative rank 3. It is not known whether this result holds in general. In Section 6, we apply Theorem 4.1 to derive the topological boundary and the algebraic boundary of $\mathcal{M}$. Also, using what follows in Section 5, we shall see how these boundaries are detected by the EM algorithm.

5. Decomposing the variety of EM fixed points. After this in-depth study of the geometry of our model, we now return to the fixed points of Expectation–Maximization on $\mathcal{M}$. We fix the polynomial ring $\mathbb{Q}[A, R, B]$ in $mr + mn + rn$ indeterminates a_{ik}, r_{ij} and b_{kj}. Let $\mathcal{F}$ denote the ideal generated by the entries of the matrices $A \star (R \cdot B^T)$ and $B \star (A^T \cdot R)$ in (3.3). Also, let $\mathcal{C}$ denote the ideal generated by the entries of $R \cdot B^T$ and $A^T \cdot R$. Thus, $\mathcal{F}$ is generated by $mr + rn$ cubics, $\mathcal{C}$ is generated by $mr + rn$ quadrics, and we have the inclusion $\mathcal{F} \subset \mathcal{C}$. By Theorem 3.5, the variety $V(\mathcal{C})$ consists of those parameters A, R, B that correspond to critical points for the log-likelihood function ℓ_U, while the variety $V(\mathcal{F})$ encompasses all the fixed points of the EM algorithm. We are interested in the irreducible components of the varieties $V(\mathcal{F})$ and $V(\mathcal{C})$. These are the zero sets of the minimal primes of $\mathcal{F}$ and $\mathcal{C}$, respectively. More precisely, if $\mathcal{F}$ has minimal primes $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_N$, then $V(\mathcal{F}_i)$ are the irreducible components of $V(\mathcal{F})$, and $V(\mathcal{F}) = \bigcup_i V(\mathcal{F}_i)$.

Recall that the matrix R represents the gradient of the log-likelihood function ℓ_U, that is,

$$r_{ij} = u_{++} - \frac{u_{ij}}{p_{ij}} = u_{++} - \frac{u_{ij}}{\sum_k a_{ik}\lambda_k b_{kj}}. \tag{5.1}$$

The set of EM-fixed points corresponding to a data matrix $U \in \mathbb{N}^{m \times n}$ is defined by the ideal $\mathcal{F}' \subset \mathbb{Q}[A, B, \Lambda]$ that is obtained from $\mathcal{F}$ by substituting (5.1), clearing denominators, and saturating. Note that $V(\mathcal{F}') = \bigcup_i V(\mathcal{F}'_i)$. So, studying the minimal primes $\mathcal{F}_i$ will help us study the fixed points of EM. A big advantage of considering $\mathcal{F}$ rather than $\mathcal{F}'$ is that $\mathcal{F}$ is much simpler. Also, it does not depend on the data U. This allows a lot of the work in exact MLE using algebraic methods (as in Example 2.1) to be done in a preprocessing stage.

There are two important points we wish to make in this section:

1. the minimal primes of $\mathcal{F}$ have interesting statistical interpretations, and
2. the nontrivial boundaries of the mixture model $\mathcal{M}$ can be detected from this.

We shall explain these points by working out two cases that are larger than Example 3.6.

Example 3.6 showed that $\mathcal{F}$ is not radical but has embedded components. Here, we focus on the minimal primes $\mathcal{F}_i$ of $\mathcal{F}$, as these correspond to geometric components of $V(\mathcal{F})$. If $\mathcal{F}_i$ is also a minimal prime of $\mathcal{C}$ then $\mathcal{F}_i$ is a *critical* prime of $\mathcal{F}$. Not every minimal prime of $\mathcal{C}$ is a minimal prime of $\mathcal{F}$. For instance, for $m = n = 2, r = 1$, the ideal $\mathcal{C}$ is the intersection of the first prime in Example 3.6 and $\langle a_{11}, a_{21}, b_{11}, b_{12}\rangle$. The latter is not minimal over $\mathcal{F}$. We now generalize this example:

PROPOSITION 5.1. *The ideal $\mathcal{C}$ has precisely $r + 1$ minimal primes, indexed by $k = 1, \ldots, r + 1$:*

$$\mathcal{C} + \langle k\text{-minors of } A\rangle + \big\langle (m-k+2)\text{-minors of } R\big\rangle + \big\langle (n-m+k)\text{-minors of } B\big\rangle \qquad \text{if } m \le n,$$

$$\mathcal{C} + \big\langle (m-n+k)\text{-minors of } A\big\rangle + \big\langle (n-k+2)\text{-minors of } R\big\rangle + \langle k\text{-minors of } B\rangle \qquad \text{if } m \ge n.$$

Moreover, the ideal $\mathcal{C}$ is radical, and hence, it equals the intersection of its minimal primes.

We refer to Example A.1 for an illustration of Proposition 5.1. The proof we give in Appendix A relies on methods from representation theory. The duality relation (A.2) plays an important role.

We now proceed to our case studies of the minimal primes of the EM fixed ideal $\mathcal{F}$.

EXAMPLE 5.2. Let $m = n = 3$ and $r = 2$. The ideal $\mathcal{F}$ has 37 minimal primes, in six classes. The first three are the minimal primes of the critical ideal $\mathcal{C}$, as seen in Proposition 5.1:

$$\begin{aligned}
I_1 = \langle & r_{23}r_{32} - r_{22}r_{33}, r_{13}r_{32} - r_{12}r_{33}, r_{23}r_{31} - r_{21}r_{33}, r_{22}r_{31} - r_{21}r_{32}, \\
& r_{13}r_{31} - r_{11}r_{33}, r_{12}r_{31} - r_{11}r_{32}, r_{13}r_{22} - r_{12}r_{23}, r_{13}r_{21} - r_{11}r_{23}, \\
& r_{12}r_{21} - r_{11}r_{22}, b_{21}r_{31} + b_{22}r_{32} + b_{23}r_{33}, b_{11}r_{31} + b_{12}r_{32} + b_{13}r_{33}, \\
& b_{21}r_{21} + b_{22}r_{22} + b_{23}r_{23}, b_{11}r_{21} + b_{12}r_{22} + b_{13}r_{23}, \\
& a_{12}r_{13} + a_{22}r_{23} + a_{32}r_{33}, a_{11}r_{13} + a_{21}r_{23} + a_{31}r_{33}, \\
& a_{12}r_{12} + a_{22}r_{22} + a_{32}r_{32}, a_{11}r_{12} + a_{21}r_{22} + a_{31}r_{32}, \\
& b_{21}r_{11} + b_{22}r_{12} + b_{23}r_{13}, b_{11}r_{11} + b_{12}r_{12} + b_{13}r_{13}, \\
& \qquad a_{12}r_{11} + a_{22}r_{21} + a_{32}r_{31}, a_{11}r_{11} + a_{21}r_{21} + a_{31}r_{31}\rangle, \\
I_2 = \langle & r_{13}r_{22}r_{31} - r_{12}r_{23}r_{31} - r_{13}r_{21}r_{32} + r_{11}r_{23}r_{32} + r_{12}r_{21}r_{33} - r_{11}r_{22}r_{33}, \\
& b_{21}r_{31} + b_{22}r_{32} + b_{23}r_{33}, b_{11}r_{31} + b_{12}r_{32} + b_{13}r_{33}, \\
& b_{21}r_{21} + b_{22}r_{22} + b_{23}r_{23}, b_{11}r_{21} + b_{12}r_{22} + b_{13}r_{23}, \\
& a_{12}r_{13} + a_{22}r_{23} + a_{32}r_{33}, a_{11}r_{13} + a_{21}r_{23} + a_{31}r_{33},
\end{aligned}$$

$$a_{12}r_{12}+a_{22}r_{22}+a_{32}r_{32}, a_{11}r_{12}+a_{21}r_{22}+a_{31}r_{32},$$
$$b_{21}r_{11}+b_{22}r_{12}+b_{23}r_{13}, b_{11}r_{11}+b_{12}r_{12}+b_{13}r_{13},$$
$$a_{12}r_{11}+a_{22}r_{21}+a_{32}r_{31}, a_{11}r_{11}+a_{21}r_{21}+a_{31}r_{31},$$
$$b_{13}b_{22}-b_{12}b_{23}, b_{13}b_{21}-b_{11}b_{23}, b_{12}b_{21}-b_{11}b_{22},$$
$$a_{31}a_{22}-a_{21}a_{32}, a_{31}a_{12}-a_{11}a_{32}, a_{21}a_{12}-a_{11}a_{22}\rangle,$$
$$I_3 = \langle a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}, b_{11}, b_{12}, b_{13}, b_{21}, b_{22}, b_{23}\rangle.$$

In addition to these three, $\mathcal{F}$ has 12 noncritical components like

$$J_1 = \langle a_{11}, a_{21}, r_{31}, r_{32}, r_{33}, r_{13}r_{22}-r_{12}r_{23}, r_{13}r_{21}-r_{11}r_{23},$$
$$r_{12}r_{21}-r_{11}r_{22}, b_{21}r_{21}+b_{22}r_{22}+b_{23}r_{23}, b_{21}r_{11}+b_{22}r_{12}+b_{23}r_{13},$$
$$a_{12}r_{13}+a_{22}r_{23}, a_{12}r_{12}+a_{22}r_{22}, a_{12}r_{11}+a_{22}r_{21}\rangle,$$

four noncritical components like

$$J_2 = \langle a_{11}, a_{21}, a_{31}, r_{13}r_{22}r_{31}-r_{12}r_{23}r_{31}-r_{13}r_{21}r_{32}+r_{11}r_{23}r_{32}$$
$$+r_{12}r_{21}r_{33}-r_{11}r_{22}r_{33}, b_{21}r_{21}+b_{22}r_{22}+b_{23}r_{23}, b_{21}r_{11}$$
$$+b_{22}r_{12}+b_{23}r_{13}, b_{21}r_{31}+b_{22}r_{32}+b_{23}r_{33}, a_{12}r_{13}+a_{22}r_{23}$$
$$+a_{32}r_{33}, a_{12}r_{12}+a_{22}r_{22}+a_{32}r_{32}, a_{12}r_{11}+a_{22}r_{21}+a_{32}r_{31}\rangle$$

and 18 noncritical components like

$$J_3 = \langle a_{11}, a_{21}, b_{11}, b_{12}, r_{33}, r_{13}r_{22}r_{31}-r_{12}r_{23}r_{31}-r_{13}r_{21}r_{32}+r_{11}r_{23}r_{32},$$
$$b_{21}r_{31}+b_{22}r_{32}, b_{21}r_{21}+b_{22}r_{22}+b_{23}r_{23}, b_{21}r_{11}+b_{22}r_{12}+b_{23}r_{13},$$
$$a_{12}r_{13}+a_{22}r_{23}, a_{12}r_{12}+a_{22}r_{22}+a_{32}r_{32}, a_{12}r_{11}+a_{22}r_{21}+a_{32}r_{31}\rangle.$$

Each of the 34 primes J_1, J_2, J_3 specifies a face of the polytope Θ, as it contains two, three or four of the parameters a_{ik}, b_{kj}, and expresses rank constraints on the matrix $R = [r_{ij}]$.

REMARK 5.3. Assuming the sample size u_{++} to be known, we can recover the data matrix U from the gradient R using the formula $U = R \star P + u_{++}P$. In coordinates, this says

$$u_{ij} = (r_{ij} + u_{++}) \cdot p_{ij} \qquad \text{for } i \in [m], j \in [n].$$

This formula is obtained by rewriting (5.1). Hence, $r_{ij} = 0$ holds if and only if $p_{ij} = u_{ij}/u_{++}$. This can be rephrased as follows. If a minimal prime of $\mathcal{F}$ contains the unknown r_{ij}, then the corresponding fixed points of the EM algorithm maintain the cell entry u_{ij} from the data.

With this, we can now understand the meaning of the various components in Example 5.2. The prime I_1 parametrizes critical points P of rank 2. This represents the behavior of the EM algorithm when run with random starting parameters in the interior of Θ. For special data U, the MLE will be a rank 1 matrix, and such cases are captured by the critical component I_2. The components I_3 and J_2 can be disregarded because each of them contains a column of A. This would force the entries of that column to sum to 0, which is impossible in Θ.

The components J_1 and J_3 describe interesting scenarios that are realized by starting the EM algorithm with parameters on the boundary of the polytope Θ. On the components J_1, the EM algorithm produces an estimate that maintains one of the rows or columns from the data U, and it replaces the remaining table of format 2×3 or 3×2 by its MLE of rank 1. This process amounts to fitting a context specific independence (CSI) model to the data. Following Georgi and Schliep [21], CSI means that independence holds only for some values of the involved variables. Namely, J_1 expresses the constraint that X is independent of Y given that Y is either 1 or 2. Finally, on the components J_3, we have $\operatorname{rank}(A) = \operatorname{rank}(B) = 2$ and $r_{ij} = 0$ for one cell entry (i, j).

DEFINITION 5.4. Let $\mathcal{F} = \langle A \star (R \cdot B^T), B \star (A^T \cdot R)\rangle$ be the ideal of EM fixed points. A minimal prime of $\mathcal{F}$ is called *relevant* if it contains none of the mn polynomials $p_{ij} = \sum_{k=1}^{r} a_{ik} b_{kj}$.

In Example 3.6, only the first minimal prime is relevant. In Example 5.2, all minimal primes besides I_3 are relevant. Restricting to the relevant minimal primes is justified because the EM algorithm never outputs a matrix containing zeros for positive starting data. Note also that the p_{ij} appear in the denominators in the expressions (3.1) that were used in our derivation of $\mathcal{F}$.

Our main result in this section is the computation in Theorem 5.5. We provide a census of EM fixed points for 4×4-matrices of rank $r = 3$. This is the smallest case where rank can differ from nonnegative rank, and the boundary hypersurfaces (4.3) appear.

THEOREM 5.5. *Let $m = n = 4$ and $r = 3$. The radical of the EM fixed point ideal $\mathcal{F}$ has* 49,000 *relevant primes. These come in* 108 *symmetry classes, listed in Table* 2.

PROOF. We used an approach that mirrors the primary decomposition of binomial ideals [16]. Recall that the EM fixed point ideal equals

$$\begin{aligned}\mathcal{F} &= \langle A \star (R \cdot B^T), B \star (A^T \cdot R)\rangle \\ &= \left\langle a_{ik}\left(\sum_{l=1}^{n} r_{il} b_{kl}\right), b_{kj}\left(\sum_{l=1}^{m} r_{lj} a_{lk}\right) : k \in [r], i \in [m], j \in [n]\right\rangle.\end{aligned}$$

Any prime ideal containing $\mathcal{F}$ contains either a_{ik} or $\sum_{l=1}^{n} r_{il} b_{kl}$ for any $k \in [r]$, $i \in [m]$, and either b_{kj} or $\sum_{l=1}^{m} r_{lj} a_{lk}$ for any $k \in [r]$, $j \in [n]$. We enumerated all primes containing $\mathcal{F}$ according to the set S of unknowns a_{ik}, b_{kj} they contain.

TABLE 2
Minimal primes of the EM fixed ideal $\mathcal{F}$ for 4×4-matrices of rank 3

Set S	$\|S\|$	a's	b's	deg	codim	rA	rB	rR	rP	|orbit|
$\varnothing$	0	0	0	1	24	0	0	4	0	1
	0	0	0	1630	19	1	1	3	1	1
	0	0	0	3491	16	2	2	2	2	1
	0	0	0	245	15	3	3	1	3	1
$\{a_{11}\}$	1	1	0	245	16	3	3	1	3	24
	1	1	0	3491	17	2	2	2	2	24
$\{a_{11}, a_{21}\}$	2	2	0	20	17	3	3	1	3	36
	2	2	0	245	17	3	3	1	3	36
	2	2	0	1460	17	2	3	2	2	36
$\{a_{11}, a_{21}, a_{31}\}$	3	3	0	53	17	3	3	1	3	24
	3	3	0	188	17	2	3	2	2	24
$*\{a_{11}, a_{21}, b_{11}, b_{12}\}*$	4	2	2	245	19	3	3	1	3	108
	4	2	2	20	19	3	3	1	3	108×2
	4	2	2	1460	19	2	3	2	2	108×2
	4	2	2	2370	20	2	2	3	2	108
	4	2	2	240	19	3	3	2	3	108
$\{a_{11}, a_{21}, b_{21}, b_{22}\}$	4	2	2	825	18	3	3	2	3	216
$\{a_{11}, a_{21}, a_{31}, a_{41}\}$	4	4	0	689	16	2	3	2	2	6
	4	4	0	474	17	1	2	3	1	6
$\{a_{11}, a_{21}, a_{12}, a_{22}\}$	4	4	0	592	17	2	3	2	2	36
	4	4	0	9	17	3	3	1	3	36
$\{a_{11}, a_{21}, a_{32}, a_{42}\}$	4	4	0	20	19	3	3	1	3	36×2
	4	4	0	245	19	3	3	1	3	36
	4	4	0	400	18	2	3	2	2	36
$\{a_{11}, a_{21}, a_{31}, b_{11}, b_{12}\}$	5	3	2	474	20	2	2	3	2	144
	5	3	2	188	19	2	3	2	2	144
	5	3	2	448	19	3	3	2	3	144
	5	3	2	53	19	3	3	1	3	144
$\{a_{11}, a_{21}, a_{31}, b_{21}, b_{22}\}$	5	3	2	125	18	3	3	2	3	288
$\{a_{11}, a_{21}, a_{32}, a_{42}, b_{31}\}$	5	4	1	723	19	3	3	2	3	144
$\{a_{11}, a_{21}, a_{31}, b_{11}, b_{12}, b_{13}\}$	6	3	3	689	19	3	3	2	3	48
	6	3	3	474	20	2	2	3	2	48
$\{a_{11}, a_{21}, a_{31}, b_{21}, b_{22}, b_{23}\}$	6	3	3	21	18	3	3	2	3	96
$\{a_{11}, a_{21}, a_{32}, b_{11}, b_{12}, b_{33}\}$	6	3	3	2785	20	3	3	3	3	864
$*\{a_{11}, a_{22}, a_{33}, b_{11}, b_{22}, b_{33}\}*$	6	3	3	9016	21	3	3	4	3	576
	6	3	3	245	21	3	3	1	3	576
$\{a_{11}, a_{21}, a_{31}, a_{41}, b_{21}, b_{22}\}$	6	4	2	265	17	2	3	2	2	72
$\{a_{11}, a_{21}, a_{12}, a_{22}, b_{11}, b_{12}\}$	6	4	2	592	19	2	3	2	2	432
	6	4	2	9	19	3	3	1	3	432
	6	4	2	104	19	3	3	2	3	432
$\{a_{11}, a_{21}, a_{32}, a_{42}, b_{11}, b_{12}\}$	6	4	2	825	20	3	3	2	3	432
	6	4	2	100	20	3	3	2	3	432
	6	4	2	400	20	2	3	2	2	432
$\{a_{11}, a_{21}, a_{32}, a_{42}, b_{31}, b_{32}\}$	6	4	2	301	19	3	3	2	3	216

TABLE 2
(*Continued*)

Set S	$\lvert S\rvert$	a's	b's	deg	codim	rA	rB	rR	rP	|orbit|
$\{a_{11}, a_{21}, a_{31}, a_{41}, a_{12}, a_{22}\}$	6	6	0	265	17	2	3	2	2	72
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}\}$	6	6	0	35	16	2	3	2	2	24
$\{a_{11}, a_{21}, a_{12}, a_{22}, a_{33}, a_{43}\}$	6	6	0	180	18	2	3	2	2	36
	6	6	0	9	19	3	3	1	3	36
$\{a_{11}, a_{21}, a_{31}, a_{41}, b_{21}, b_{22}, b_{23}\}$	7	4	3	35	17	2	3	2	2	48
$\{a_{11}, a_{21}, a_{31}, a_{42}, b_{11}, b_{12}, b_{33}\}$	7	4	3	557	20	3	3	3	3	576
$\{a_{11}, a_{21}, a_{12}, a_{22}, b_{11}, b_{12}, b_{13}\}$	7	4	3	191	19	3	3	2	3	288
$\{a_{11}, a_{21}, a_{32}, a_{42}, b_{11}, b_{12}, b_{13}\}$	7	4	3	140	20	3	3	2	3	288
	7	4	3	125	20	3	3	2	3	288
$\{a_{11}, a_{21}, a_{32}, a_{42}, b_{11}, b_{12}, b_{33}\}$	7	4	3	835	20	3	3	3	3	864
$\{a_{11}, a_{21}, a_{32}, a_{42}, b_{31}, b_{32}, b_{33}\}$	7	4	3	49	19	3	3	2	3	144
$*\{a_{11}, a_{21}, a_{32}, a_{43}, b_{11}, b_{22}, b_{33}\}*$	7	4	3	3087	21	3	3	4	3	1728
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, b_{21}, b_{22}\}$	7	5	2	31	19	3	3	2	3	864
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{42}, b_{11}, b_{12}\}$	7	5	2	225	20	3	3	2	3	864
$\{a_{11}, a_{21}, a_{12}, a_{32}, a_{43}, b_{11}, b_{22}\}$	7	5	2	1193	21	3	3	3	3	1728
$\{a_{11}, a_{21}, a_{31}, a_{41}, b_{21}, b_{22}, b_{23}, b_{24}\}$	8	4	4	85	15	2	2	3	1	6
$\{a_{11}, a_{21}, a_{31}, a_{41}, b_{21}, b_{22}, b_{33}, b_{34}\}$	8	4	4	81	18	2	3	2	2	36
$\{a_{11}, a_{21}, a_{31}, a_{42}, b_{11}, b_{12}, b_{13}, b_{34}\}$	8	4	4	557	20	3	3	3	3	96
$\{a_{11}, a_{21}, a_{31}, a_{42}, b_{11}, b_{12}, b_{33}, b_{34}\}$	8	4	4	167	20	3	3	3	3	288
$\{a_{11}, a_{21}, a_{12}, a_{22}, b_{11}, b_{12}, b_{21}, b_{22}\}$	8	4	4	850	20	2	2	3	2	108
	8	4	4	45	19	3	3	2	3	108
$\{a_{11}, a_{21}, a_{12}, a_{22}, b_{11}, b_{12}, b_{23}, b_{24}\}$	8	4	4	9	21	3	3	1	3	216
	8	4	4	1024	21	3	2	3	2	216
	8	4	4	104	21	3	3	2	3	216×2
	8	4	4	592	21	2	3	2	2	216
$\{a_{11}, a_{21}, a_{12}, a_{32}, b_{11}, b_{12}, b_{21}, b_{23}\}$	8	4	4	2121	21	3	3	3	3	1728
$\{a_{11}, a_{21}, a_{12}, a_{32}, b_{11}, b_{12}, b_{23}, b_{24}\}$	8	4	4	2125	21	3	3	3	3	864
$\{a_{11}, a_{21}, a_{32}, a_{42}, b_{11}, b_{12}, b_{23}, b_{24}\}$	8	4	4	2125	21	3	3	3	3	108
$\{a_{11}, a_{21}, a_{32}, a_{42}, b_{11}, b_{12}, b_{33}, b_{34}\}$	8	4	4	265	20	3	3	3	3	216
$\{a_{11}, a_{21}, a_{32}, a_{43}, b_{11}, b_{12}, b_{23}, b_{34}\}$	8	4	4	2205	21	3	3	4	3	432
$\{a_{11}, a_{21}, a_{32}, a_{43}, b_{11}, b_{22}, b_{23}, b_{34}\}$	8	4	4	1029	21	3	3	4	3	864
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, b_{21}, b_{22}, b_{23}\}$	8	5	3	35	19	3	3	2	3	576
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{42}, b_{11}, b_{12}, b_{13}\}$	8	5	3	265	20	3	3	2	3	576
$\{a_{11}, a_{21}, a_{12}, a_{32}, a_{43}, b_{11}, b_{12}, b_{23}\}$	8	5	3	1185	21	3	3	3	3	3456
$\{a_{11}, a_{21}, a_{31}, a_{41}, a_{12}, a_{22}, b_{21}, b_{22}\}$	8	6	2	425	18	2	3	3	2	432
$\{a_{11}, a_{21}, a_{12}, a_{22}, a_{33}, a_{43}, b_{11}, b_{12}\}$	8	6	2	180	20	2	3	2	2	432
	8	6	2	45	20	3	3	2	3	432
$\{a_{11}, a_{21}, a_{31}, a_{41}, a_{12}, a_{22}, a_{32}, a_{42}\}$	8	8	0	85	15	1	3	3	1	6
$\{a_{11}, a_{21}, a_{31}, a_{41}, a_{12}, a_{22}, a_{33}, a_{43}\}$	8	8	0	81	18	2	3	2	2	36
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, b_{11}, b_{12}, b_{23}, b_{24}\}$	9	5	4	296	21	3	3	3	3	864
	9	5	4	31	21	3	3	2	3	864
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{42}, b_{11}, b_{12}, b_{21}, b_{23}\}$	9	5	4	425	21	3	3	3	3	3456
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{42}, b_{11}, b_{12}, b_{23}, b_{24}\}$	9	5	4	425	21	3	3	3	3	864
$\{a_{11}, a_{21}, a_{12}, a_{22}, a_{33}, b_{11}, b_{12}, b_{23}, b_{24}\}$	9	5	4	839	21	3	3	3	3	432

TABLE 2
(*Continued*)

Set S	**$\vert S\vert$**	**a's**	**b's**	**deg**	**codim**	**rA**	**rB**	**rR**	**rP**	**$\vert$orbit$\vert$**
$\{a_{11}, a_{21}, a_{12}, a_{32}, a_{43}, b_{11}, b_{12}, b_{13}, b_{24}\}$	9	5	4	237	21	3	3	3	3	1152
$\{a_{11}, a_{21}, a_{12}, a_{32}, a_{43}, b_{11}, b_{12}, b_{23}, b_{24}\}$	9	5	4	875	21	3	3	3	3	864
$\{a_{11}, a_{21}, a_{31}, a_{41}, a_{12}, a_{22}, b_{21}, b_{22}, b_{23}\}$	9	6	3	85	18	2	3	3	2	288
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{43}, b_{11}, b_{12}, b_{23}\}$	9	6	3	163	21	3	3	3	3	1728
$\{a_{11}, a_{21}, a_{12}, a_{22}, a_{33}, a_{43}, b_{11}, b_{12}, b_{13}\}$	9	6	3	63	20	3	3	2	3	288
$\{a_{11}, a_{21}, a_{31}, a_{41}, a_{12}, a_{22}, a_{32}, b_{21}, b_{22}\}$	9	7	2	85	18	2	3	3	2	288
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, b_{11}, b_{12}, b_{13}, b_{21}, b_{24}\}$	10	5	5	425	21	3	3	3	3	1728
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, b_{11}, b_{12}, b_{21}, b_{22}, b_{23}\}$	10	5	5	85	20	3	3	3	3	864
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{42}, b_{11}, b_{12}, b_{13}, b_{21}, b_{24}\}$	10	5	5	425	21	3	3	3	3	864
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{42}, b_{11}, b_{12}, b_{21}, b_{23}, b_{24}\}$	10	5	5	85	21	3	3	3	3	864
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}, b_{11}, b_{12}, b_{21}, b_{22}\}$	10	6	4	85	19	2	3	3	2	144
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{42}, b_{11}, b_{12}, b_{21}, b_{23}\}$	10	6	4	85	21	3	3	3	3	1728
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{42}, b_{11}, b_{12}, b_{23}, b_{24}\}$	10	6	4	85	21	3	3	3	3	432
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{43}, b_{11}, b_{12}, b_{13}, b_{24}\}$	10	6	4	237	21	3	3	3	3	576
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{43}, b_{11}, b_{12}, b_{23}, b_{24}\}$	10	6	4	175	21	3	3	3	3	864
$\{a_{11}, a_{21}, a_{12}, a_{22}, a_{33}, a_{43}, b_{11}, b_{12}, b_{23}, b_{24}\}$	10	6	4	225	21	3	3	3	3	216
$\{a_{11}, a_{21}, a_{31}, a_{41}, a_{12}, a_{22}, a_{32}, b_{21}, b_{22}, b_{23}\}$	10	7	3	85	18	2	3	3	2	192
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{42}, b_{11}, b_{12}, b_{13}, b_{21}, b_{24}\}$	11	6	5	85	21	3	3	3	3	1728
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}, b_{11}, b_{12}, b_{13}, b_{21}, b_{22}, b_{23}\}$	12	6	6	85	20	2	2	3	2	48
$\{a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{42}, b_{11}, b_{12}, b_{13}, b_{21}, b_{22}, b_{24}\}$	12	6	6	85	21	3	3	3	3	432

There are 2^{24} subsets and the symmetry group acts on this power set by replacing A with B^T, permuting the rows of A, the columns of B, and the columns of A and the rows of B simultaneously. We picked one representative S from each orbit that is relevant, meaning that we excluded those orbits for which some $p_{ij} = \sum_{k=1}^{r} a_{ik} b_{kj}$ lies in the ideal $\langle S \rangle$. For each relevant representative S, we computed the *cellular component* $\mathcal{F}_S = ((\mathcal{F} + \langle S \rangle) : (\prod S^c)^\infty)$, where $S^c = \{a_{11}, \ldots, b_{34}\} \setminus S$. Note that $\mathcal{F}_\varnothing = \mathcal{C}$ is the critical ideal. We next minimalized our cellular decomposition by removing all representatives S such that $\mathcal{F}_T \subset \mathcal{F}_S$ for some representative T in another orbit. This led to a list of 76 orbits, comprising 42,706 ideals $\mathcal{F}_S$ in total. For the representative $\mathcal{F}_S$, we computed the set $\mathrm{Ass}(\mathcal{F}_S)$ of associated primes P. By construction, the sets $\mathrm{Ass}(\mathcal{F}_S)$ partition the set of relevant primes of $\mathcal{F}$. The block sizes $|\mathrm{Ass}(\mathcal{F}_S)|$ range from 1 to 7. Up to symmetry, each prime is uniquely determined by its attributes in Table 2. These are its set S, its degree and codimension, and the ranks $\mathrm{rA} = \mathrm{rank}(A)$, $\mathrm{rB} = \mathrm{rank}(B)$, $\mathrm{rR} = \mathrm{rank}(R)$, $\mathrm{rP} = \mathrm{rank}(P)$ at a generic point. Our list starts with the four primes from coming from $S = \varnothing$. See Example A.1. In each case, the primality of the ideal was verified using a linear elimination sequence as in [20], Proposition 23(b). Proofs in `Macaulay2` code are posted on our website. $\square$

Below is the complete list of all 108 classes of prime ideals in Theorem 5.5. Three components are marked with stars. After the table, we discuss these components in Examples 5.6, 5.7 and 5.8.

We illustrate our census of relevant primes for three sets S that are especially interesting.

EXAMPLE 5.6. Let $S = \{a_{11}, a_{21}, b_{11}, b_{12}\}$. The cellular component $\mathcal{F}_S$ is the ideal generated by $S, \det(R_{34}^{34}), \det(R)$, and the entries of the matrices $B^{23}R^T, B^1(R^T)_{34}, R^T A_{23}, (R^T)^{34}A_1$. In specifying submatrices, upper indices refer to rows and lower indices refer to columns. The ideal $\mathcal{F}_S$ is radical with 7 associated primes, to be discussed in order of their appearance in Table 2. For instance, the prime (1) below has degree 245. The phrase "Generated by" is meant modulo $\mathcal{F}_S$:

(1) Generated by entries of BR^T, $A^T R$, and 2×2-minors of R. This gives 60 quadrics.

(2) Generated by entries of $A^T R$, R^{34}, and 2×2-minors of R, A_{23}^{12}. This gives 19 quadrics.

(2′) Mirror image of (2) under swapping A and B^T.

(3) Generated by entries of $A^T R$, 2×2-minors of A_{23}^{12}, R^{34}, and 3×3-minors of A, R^{123}, R^{124}. This gives 29 quadrics and 10 cubics.

(3′) Mirror image of (3) under swapping A and B^T.

(4) Generated by 2×2-minors of A_{23} and B^{23}. This gives 33 quadrics and one quartic.

(5) Generated by entries of R_{34}^{34}, 2×2-minors of R_{34}^{12}, R_{12}^{34}, A_{23}^{12}, B_{12}^{23}, and 3×3-minors of R. This gives 20 quadrics and 4 cubics.

These primes have the following meaning for the EM algorithm:

(1) The fixed points $P = \phi(A, R, B)$ given by this prime ideal are those critical points for the likelihood function ℓ_U for which the parameters $a_{11}, a_{21}, b_{11}, b_{21}$ happen to be 0.

(2) The fixed points $P = \phi(A, R, B)$ given by this prime ideal have the last two rows of P fixed and equal to the last two rows of the data matrix U (divided by the sample size u_{++}). Therefore, the points coming from this ideal are the maximum likelihood estimates with these eight entries fixed and which factor so that $a_{11}, a_{21}, b_{11}, b_{21}$ are 0.

(3) Since the 3×3 minors of A lie in this ideal, we have $\text{rank}(P) \leq 2$. Therefore, these fixed points give an MLE of rank 2. This component is the restriction to $V(\mathcal{F}_S)$ of the generic behavior on the singular locus of $\mathcal{V}$.

(4) On this component, the duality relation in (A.2) fails since $\text{rank}(P) = 2$ but $\text{rank}(R) = 3$.

(5) The fixed points $P = \phi(A, R, B)$ given by this ideal have the four entries in the last 2 rows and last 2 columns of P fixed and equal to the corresponding entries

in U (divided by u_{++}). Therefore, the points coming from this ideal are maximum likelihood estimates with those four entries fixed, and parameters $a_{11}, a_{21}, b_{11}, b_{21}$ being 0.

EXAMPLE 5.7. Let $S = \{a_{11}, a_{21}, a_{32}, a_{43}, b_{11}, b_{22}, b_{33}\}$. The ideal $\mathcal{F}_S$ has codimension 21, degree 3087, and is generated modulo $\langle \mathcal{S} \rangle$ by 20 quadrics and two cubics. To show that $\mathcal{F}_S$ is prime, we use the elimination method of [20], Proposition 23(b), with the variable x_1 taken successively to be r_{44}, r_{43}, r_{34}, a_{13}, $r_{21}, r_{12}, r_{14}, r_{33}, b_{21}, a_{31}, r_{41}, a_{21}, a_{32}$.

The last elimination ideal is generated by an irreducible polynomial of degree 9, thus proving primality of $\mathcal{F}_S$.

If we add the relation $P = AB$ to $\mathcal{F}_S$ and thereafter eliminate $\{A, B, R\}$, we obtain a prime ideal in $\mathbb{Q}[P]$. That prime ideal has height one over the determinantal ideal $\langle \det(P) \rangle$. Any such prime gives a candidate for a component in the boundary of our model $\mathcal{M}$. By matching the set S with the combinatorial analysis in Section 4, we see that Figure 5(b) corresponds to $V(S)$. Hence, by Corollary 4.4, this component does in fact contribute to the boundary $\partial\mathcal{M}$. This is a special case of Theorem 6.1 below; see equation (6.2) in Example 6.2.

This component is the most important one for EM. It represents the typical behavior when the output of the EM algorithm is not critical. In particular, the duality relation (A.2) fails in the most dramatic form because $\operatorname{rank}(R) = 4$. As seen in Table 1, this failure is still rare (4.4%) for $m = n = 4$. For larger matrix sizes, however, the noncritical behavior occurs with overwhelming probability.

EXAMPLE 5.8. Let $S = \{a_{11}, a_{22}, a_{33}, b_{11}, b_{22}, b_{33}\}$. The computation for the ideal $\mathcal{F}_S$ was the hardest among all cellular components. It was found to be radical, with two associated primes of codimension 21. The first prime has the largest degree, namely 9016, among all entries in Table 2. In contrast to Example 5.7, the set S cannot contribute to $\partial\mathcal{M}$. Indeed, for both primes, the elimination ideal in $\mathbb{Q}[P]$ is $\langle \det(P) \rangle$. The degree 9016 ideal is the only prime in Table 2 that has $\operatorname{rank}(R) = 4$ but does not map to the boundary of the model $\mathcal{M}$. Starting the EM algorithm with zero parameters in S generally leads to the correct MLE.

6. Algebraic boundaries. In Section 4, we studied the real algebraic geometry of the mixture model $\mathcal{M}$ for rank three. In this section, we also fix $r = 3$ and focus on the algebraic boundary of our model. Our main result in this section is the characterization of its irreducible components.

THEOREM 6.1. *The algebraic boundary $\overline{\partial\mathcal{M}}$ is a pure-dimensional reducible variety in $\mathbb{P}^{mn-1}$. All irreducible components have dimension $3m + 3n - 11$ and their number equals*

$$mn + \frac{m(m-1)(m-2)(m+n-6)n(n-1)(n-2)}{4}.$$

Besides the mn components $\{p_{ij} = 0\}$ that come from $\partial\Delta_{mn-1}$ there are:

(a) $36\binom{m}{3}\binom{n}{4}$ *components parametrized by* $P = AB$, *where* A *has three zeros in distinct rows and columns*, *and* B *has four zeros in three rows and distinct columns*.

(b) $36\binom{m}{4}\binom{n}{3}$ *components parametrized by* $P = AB$, *where* A *has four zeros in three columns and distinct rows*, *and* B *has three zeros in distinct rows and columns*.

This result takes the following specific form in the first nontrivial case:

EXAMPLE 6.2. For $m = n = 4$, the algebraic boundary of our model $\mathcal{M}$ has 16 irreducible components $\{p_{ij} = 0\}$, 144 irreducible components corresponding to factorizations like

$$\begin{aligned}&\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \\ &\quad = \begin{bmatrix} 0 & a_{12} & a_{13} \\ a_{21} & 0 & a_{23} \\ a_{31} & a_{32} & 0 \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & b_{13} & b_{14} \\ b_{21} & b_{22} & 0 & b_{24} \\ b_{31} & b_{32} & b_{33} & 0 \end{bmatrix},\end{aligned} \tag{6.1}$$

and 144 irreducible components that are transpose to those in (6.1), that is,

$$\begin{aligned}&\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \\ &\quad = \begin{bmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & 0 & a_{33} \\ a_{41} & a_{42} & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & b_{12} & b_{13} & b_{14} \\ b_{21} & 0 & b_{23} & b_{24} \\ b_{31} & b_{32} & 0 & b_{34} \end{bmatrix}.\end{aligned} \tag{6.2}$$

The prime ideal of each component is generated by the determinant and four polynomials of degree six. These are the maximal minors of a 4×5-matrix. For the component (6.2), this can be chosen as

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & 0 \\ p_{21} & p_{22} & p_{23} & p_{24} & 0 \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{33}(p_{11}p_{22} - p_{12}p_{21}) \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{41}(p_{12}p_{23} - p_{13}p_{22}) + p_{43}(p_{11}p_{22} - p_{12}p_{21}) \end{bmatrix}. \tag{6.3}$$

This matrix representation was suggested to us by Aldo Conca and Matteo Varbaro.

We begin by resolving a problem that was stated in [24], Section 5, and [25], Example 2.13.

PROPOSITION 6.3. *The ML degree of each variety* (6.1) *in the algebraic boundary* $\overline{\partial\mathcal{M}}$ *is* 633.

Proposition 6.3 is a first step towards deriving an exact representation of the MLE function $U \mapsto \widehat{P}$ for our model $\mathcal{M} =$ ●–○–●. As highlighted in Table 1, the MLE $\widehat{P}$ typically lies on the boundary $\partial\mathcal{M}$. We now know that this boundary has $304 = 16 + 144 + 144$ strata $X_1, X_2, \dots, X_{304}$. If $\widehat{P}$ lies on exactly one of the strata (6.1) or (6.2), then we can expect the coordinates of $\widehat{P}$ to be algebraic numbers of degree 633 over the rationals $\mathbb{Q}$. This is the content of Proposition 6.3. By [24], Theorem 1.1, the degree of $\widehat{P}$ over $\mathbb{Q}$ is only 191 if $\widehat{P}$ happens to lie in the interior of $\mathcal{M}$.

In order to complete the exact analysis of MLE for the 4×4-model, we also need to determine which intersections $X_{i_1} \cap \cdots \cap X_{i_s}$ are nonempty on $\partial\mathcal{M}$. For each such nonempty stratum, we would then need to compute its ML degree. This is a challenge left for a future project.

PROOF OF THEOREM 6.1. By Corollary 4.4, an $m \times n$ matrix P of rank 3 without zero entries lies on $\partial\mathcal{M}_3^{m\times n}$ if and only if all triangles Δ with $\mathcal{B} \subseteq \Delta \subseteq \mathcal{A}$ contain an edge of $\mathcal{B}$ on one of its edges and a vertex of $\mathcal{B}$ on all other edges, or one of its vertices coincides with a vertex of $\mathcal{A}$ and all other edges contain a vertex of $\mathcal{B}$. We will write down these conditions algebraically.

The columns of A correspond to the vertices of Δ, and the columns of B correspond to the convex combinations of the vertices of Δ that give the columns of $P = AB$. If a vertex of Δ and a vertex of $\mathcal{A}$ coincide, then the corresponding column of A has two 0's. Otherwise the corresponding column of A has one 0. If a vertex of $\mathcal{B}$ lies on an edge of Δ, then one entry of B is zero.

We can freely permute the columns of the left $m \times 3$ matrix A of a factorization—this corresponds to permuting the rows of the corresponding right $3 \times n$ matrix B. Thus we can assume that the first column contains two 0's and/or the rest of the 0's appear in the increasing order.

In the first case, there are $\binom{m}{3}$ possibilities for choosing the three rows of A containing 0's, there are 3 choices for the row of B with two 0's, $\binom{n}{2}$ possibilities for choosing the positions for the two 0's, and $(n-2)(n-3)$ possibilities for choosing the positions of the 0's in the other two rows of B. In the second case, there are $\binom{m}{2}$ possibilities for choosing the 0's in the first column of A and $\binom{m-2}{2}$ choices for the positions of the 0's in other columns. There are $\binom{n}{3}$ choices for the columns of B containing 0's and 3! choices for the positions of the 0's in these columns. $\square$

The prime ideal in (6.3) can be found and verified by direct computation, for example, by using the software `Macaulay2` [22]. For general values of m and n, the prime ideal of an irreducible boundary component is generated by quartics and sextics that generalize those in Example 6.2. The following theorem was stated as a conjecture in the original December 2013 version of this paper. That conjecture was proved in April 2014 by Eggermont, Horobeţ and Kubjas [15].

THEOREM 6.4 (Eggermont, Horobeţ and Kubjas). *Let $m \geq 4, n \geq 3$ and consider the irreducible component of $\overline{\partial\mathcal{M}}$ in Theorem 6.1(b). The prime ideal of this component is minimally generated by $\binom{m}{4}\binom{n}{4}$ quartics, namely the 4×4-minors of P, and by $\binom{n}{3}$ sextics that are indexed by subsets $\{i, j, k\}$ of $\{1, 2, \dots, n\}$. These form a Gröbner basis with respect to graded reverse lexicographic order. The sextic indexed by $\{i, j, k\}$ is homogeneous of degree $e_1 + e_2 + e_3 + e_i + e_j + e_k$ in the column grading by $\mathbb{Z}^n$ and homogeneous of degree $2e_1 + 2e_2 + e_3 + e_4$ in the row grading by $\mathbb{Z}^m$.*

The row and column gradings of the polynomial ring $\mathbb{Q}[P]$ are given by $\deg(p_{ij}) = e_i$ and $\deg(p_{ij}) = e_j$ where e_i and e_j are unit vectors in $\mathbb{Z}^m$ and $\mathbb{Z}^n$, respectively.

EXAMPLE 6.5. If $m = 5$ and $n = 6$, then our component is given by the parametrization

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} & p_{36} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} & p_{46} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} & p_{56} \end{bmatrix}$$
$$= \begin{bmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & 0 & a_{33} \\ a_{41} & a_{42} & 0 \\ a_{51} & a_{52} & a_{53} \end{bmatrix} \cdot \begin{bmatrix} 0 & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ b_{21} & 0 & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{31} & b_{32} & 0 & b_{34} & b_{35} & b_{36} \end{bmatrix}.$$

This parametrized variety has codimension 7 and degree 735 in $\mathbb{P}^{29}$. Its prime ideal is generated by 75 quartics and 20 sextics of the desired row and column degrees.

The base case for Theorem 6.4 is the case of 4×3-matrices, even though $\partial\mathcal{M} = \mathcal{M} \cap \Delta_{11}$ is trivial in this case.

The corresponding ideal is principal, and it is generated by the determinant of the 4×4-matrix that is obtained by deleting the fourth column of (6.3).

The sextics in Theorem 6.4 can be constructed as follows. Start with the polynomial

$$\big(\big((a_1 \wedge a_2) \vee b_1\big) \wedge a_3\big) \vee \big(\big((a_1 \wedge a_2) \vee b_2\big) \wedge a_4\big) \vee b_3$$

that is given in (4.3). Now multiply this with the 3×3-minor $b_i \vee b_j \vee b_k$ of B. The result has bidegree $(6, 6)$ in the parameters (A, B) and can be written as a sextic in $P = AB$. By construction, it vanishes on our component of $\overline{\partial\mathcal{M}}$, and it has the asserted degrees in the row and column gradings on $\mathbb{Q}[P]$. This is the generator of the prime ideal referred to in Theorem 6.4.

Theorem 6.1 characterizes the probability distributions in the algebraic boundary of our model, but not those in the topological boundary, since the following inclusion is strict:

$$\partial\mathcal{M} \subset \overline{\partial\mathcal{M}} \cap \Delta_{mn-1}. \tag{6.4}$$

In fact, the left-hand side is much smaller than the right-hand side.

To quantify the discrepancy between the two semialgebraic sets in (6.4), we conducted the following experiment in the smallest interesting case $m = n = 4$. We sampled from the component (6.1) of $\overline{\partial\mathcal{M}} \cap \Delta_{15}$ by generating random rational numbers for the nine parameters a_{ij} and the eight parameters b_{ij}. This was done using the built-in `Macaulay2` function `random(QQ)`. The resulting matrix in $\overline{\partial\mathcal{M}} \cap \Delta_{15}$ was obtained by dividing by the sum of the entries. For each matrix, we tested whether it lies in $\partial\mathcal{M}$. This was done using the criterion in Corollary 6.6. The answer was affirmative only in 257 cases out of 5000 samples. This suggests that $\partial\mathcal{M}$ occupies only a tiny part of the set $\overline{\partial\mathcal{M}} \cap \Delta_{15}$. One of those rare points in the topological boundary is the matrix

$$\begin{bmatrix} 6 & 13 & 3 & 1 \\ 4 & 16 & 6 & 2 \\ 12 & 4 & 8 & 12 \\ 5 & 9 & 10 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 4 \\ 4 & 4 & 0 \\ 4 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 2 & 2 \\ 3 & 1 & 0 & 1 \\ 1 & 4 & 1 & 0 \end{bmatrix}. \tag{6.5}$$

To construct this particular example, the parameters a_{ij} and b_{ij} were selected uniformly at random among the integers between 1 and 4. Only 1 out of 1000 samples gave a matrix lying in $\partial\mathcal{M}$. In fact, this matrix lies on precisely one of the 304 strata in the topological boundary $\partial\mathcal{M}$.

We close this paper with a quantifier-free semialgebraic formula for the topological boundary.

COROLLARY 6.6. *An $m \times n$-matrix P lies on the topological boundary $\partial\mathcal{M}$ if and only if*:

- *the conditions of Theorem 4.1 are satisfied*, *and*
- *P contains a zero*, *or* $\operatorname{rank}(P) = 3$ *and for each i, j, i', j' for which the conditions of Theorem 4.1 are satisfied there exist k, l such that* (4.3) $\cdot$ (4.3) $[k \leftrightarrow l] = 0$.

This corollary will be derived (in Appendix A) from our results in Section 4.

APPENDIX A: PROOFS

This appendix furnishes the proofs for all lemmas, propositions and corollaries in this paper.

PROOF OF LEMMA 3.3. (3) $\Rightarrow$ (2): If (A, Λ, B) remains fixed after one completion of the E-step and the M-step, then it will remain fixed after any number of rounds of the E-step and the M-step.

(2) $\Rightarrow$ (3): By the proof of [31], Theorem 1.15, the log-likelihood function ℓ_U grows strictly after the completion of an E-step and an M-step unless the parameters (A, Λ, B) stay fixed, in which case ℓ_U also stays fixed. Thus, the only way to start with (A, Λ, B) and to end with it is for (A, Λ, B) to stay fixed after every completion of an E-step and an M-step.

(2) $\Rightarrow$ (1): If (A, Λ, B) is the limit point of EM when we start with it, then it is in the set of all limit points. This argument is reversible, and so we also get (1) $\Rightarrow$ (2), (3). $\square$

PROOF OF LEMMA 4.3. The if-direction of the first sentence follows from the following two observations: 1. The function that takes $P \in \mathbb{R}^{m\times n}_{\geq 0}$ to the vertices of $\mathcal{B}$ is continuous on all $m \times n$ nonnegative matrices without zero columns, since the vertices of $\mathcal{B}$ are of the form P^j/P_{+j}, where P_{+j} denotes the jth column sum of P. 2. The function that takes $P \in \mathbb{R}^{m\times n}_{\geq 0}$ to the vertices of $\mathcal{A}$ is continuous on all $m \times n$ nonnegative matrices of rank r, since the vertices of $\mathcal{A}$ are solutions to a system of linear equations in the entries of P.

For the only-if-direction of the first sentence assume that P lies in the interior of $\mathcal{M}_r$. Each P' of rank r in a small neighborhood of P has nonnegative rank r. We can choose P' in this neighborhood such that the columns of P' are in $\operatorname{span}(P)$ and $\operatorname{cone}(P') = t \cdot \operatorname{cone}(P)$ for some $t > 1$. Since P' has nonnegative rank r, there exists an $(r-1)$-simplex Δ such that $\mathcal{B}' \subseteq \Delta' \subseteq \mathcal{A}$. Hence, $\mathcal{B}$ is contained in the interior of Δ'. Finally, the second sentence is the contrapositive of the first sentence. $\square$

PROOF OF COROLLARY 4.4. The if-direction follows from the second sentence of Lemma 4.3. For the only-if-direction, assume that $P \in \partial\mathcal{M}_3$ and it contains no zeros. We first consider the case $\operatorname{rank}(P) = 3$. By Lemma 4.3, every triangle Δ with $\mathcal{B} \subseteq \Delta \subseteq \mathcal{A}$ contains a vertex of $\mathcal{B}$ on its boundary. Moreover, by the discussion above, every edge of Δ contains a vertex of $\mathcal{B}$, and (a) or (b) must hold. It remains to be seen that $\operatorname{rank}(P) \leq 2$ is impossible on the strictly positive part of the boundary of $\mathcal{M}_3$. Indeed, for every rank 3 matrix P' in a neighborhood of P, the polygons $\mathcal{A}', \mathcal{B}'$ have the property that $\mathcal{B}'$ is very close to a line segment strictly contained in the interior of $\mathcal{A}'$. Hence, $t\mathcal{B}' \subseteq \Delta \subseteq \mathcal{A}'$ for some triangle Δ. Thus, $P' \notin \partial\mathcal{M}_3$ and, therefore, $P \notin \partial\mathcal{M}_3$. $\square$

PROOF OF COROLLARY 4.5. The if-direction is immediate. For the only-if direction, consider any $P \in \mathcal{M}_3$. If $P \in \partial\mathcal{M}_3$, then the only-if-direction follows from Corollary 4.4. If P lies in the interior of $\mathcal{M}_3$, then let t be maximal such that $t\mathcal{B} \subseteq \Delta' \subseteq \mathcal{A}$ for some triangle Δ'. Then either a vertex of Δ' coincides with a vertex of $\mathcal{A}$ or an edge of Δ' contains an edge of $t\mathcal{B}$. In the first case, we take $\Delta = \Delta'$. In the second case, we take $\Delta = \frac{1}{t}\Delta'$. In the first case, a vertex of Δ coincides with a vertex of $\mathcal{A}$, and in the second case, an edge of Δ contains an edge of $\mathcal{B}$. $\square$

PROOF OF COROLLARY 4.6. If P has a nonnegative factorization of size 3, then it has one that corresponds to a geometric condition in Corollary 4.5. The left matrix in the factorization can be taken to be equal to the vertices of the nested triangle, which can be expressed as rational functions in the entries of P. Finally, the right matrix is obtained from solving a system of linear equations with rational coefficients, hence its entries are again rational functions in the entries of P. $\square$

PROOF OF PROPOSITION 5.1. Consider the sequence of linear maps

$$\mathbb{R}^r \xrightarrow{B^T} \mathbb{R}^n \xrightarrow{R} \mathbb{R}^m \xrightarrow{A^T} \mathbb{R}^r. \tag{A.1}$$

The ideal $\mathcal{C}$ says that the two compositions are zero. It defines a *variety of complexes* [27], Example 17.8. The irreducible components of that variety correspond to *irreducible rank arrays* [27], Section 17.1, that fit inside the format (A.1) and are maximal with this property. By [27], Theorem 17.23, the quiver loci for these rank arrays are irreducible and their prime ideals are the ones we listed. These can also be described by *lacing diagrams* [27], Proposition 17.9.

The proof that $\mathcal{C}$ is radical was suggested to us by Allen Knutson. Consider the Zelevinski map [27], Section 17.2, that sends the triple (A^T, R, B^T) to the $(r+m+n+r) \times (r+m+n+r)$ matrix

$$\begin{bmatrix} 0 & 0 & B^T & 1 \\ 0 & R & 1 & 0 \\ A^T & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Next, apply the map that takes this matrix to the big cell (the open Borel orbit) in the flag variety $GL(2r+m+n)/\text{parabolic}(r, m, n, r)$ corresponding to the given block structure.

Our scheme is identified with the intersection of two Borel invariant Schubert varieties. The first Schubert variety encodes the fact that there are 0's in the North West block, and the $(r+n+m) \times (r+m)$ North West rectangle has rank $\leq m$. The second Schubert variety corresponds to the $(r+n) \times (r+m+n)$ North West rectangle having rank $\leq n$. The intersection of Schubert varieties is reduced by [9], Section 2.3.3, page 74. Hence, the original scheme is reduced, and we conclude that $\mathcal{C}$ is the radical ideal defining the variety of complexes (A.1). $\square$

The following relations hold for $P = AB$ and R on the variety of critical points $V(\mathcal{C})$:

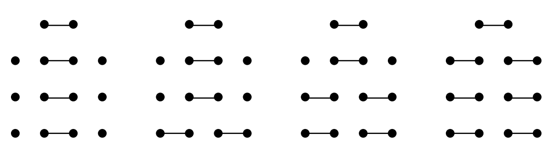
$$P^T \cdot R = 0 \quad \text{and} \quad R \cdot P^T = 0. \tag{A.2}$$

These bilinear equations characterize the *conormal variety* associated to a pair of determinantal varieties. Suppose P is fixed and has rank r. Then P is a nonsingular point in $\mathcal{V}$, and (A.2) is the system of linear equations that characterizes normal vectors R to $\mathcal{V}$ at P.

EXAMPLE A.1. Let $m = n = 4$ and $r = 3$. Then $\mathcal{C}$ has four minimal primes, corresponding to the four columns in the table below. These are the ranks for generic points on that prime:

$$\begin{array}{llll}
\operatorname{rank}(A) = 0, & \operatorname{rank}(A) = 1, & \operatorname{rank}(A) = 2, & \operatorname{rank}(A) = 3, \\
\operatorname{rank}(R) = 4, & \operatorname{rank}(R) = 3, & \operatorname{rank}(R) = 2, & \operatorname{rank}(R) = 1, \\
\operatorname{rank}(B) = 0, & \operatorname{rank}(B) = 1, & \operatorname{rank}(B) = 2 & \operatorname{rank}(B) = 3.
\end{array}$$

The lacing diagrams that describe these four irreducible components are as follows:



For instance, the second minimal prime is $\mathcal{C} + \langle 2 \times 2\text{-minors of } A \text{ and } B \rangle + \langle \det(R) \rangle$.

Note that the ranks of $P = AB$ and R are complementary on each irreducible component. They add up to 4. The last component gives the behavior of EM for random data: the MLE P has rank 3, it is a nonsingular point on the determinantal hypersurface $\mathcal{V}$, and the normal space at P is spanned by the rank 1 matrix R. This is the duality (A.2). The third component expresses the behavior on the singular locus of $\mathcal{V}$. Here, the typical rank of both P and R is 2.

PROOF OF PROPOSITION 6.3. Let f, g_1, g_2, g_3, g_4 denote the 4×4 minors of the matrix (6.3), where $\deg(f) = 4$ and $\deg(g_i) = 6$. Fix $i \in \{1, 2, 3, 4\}$, select $u_{11}, \dots, u_{44} \in \mathbb{N}$ randomly, and set

$$L = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{44} \\ p_{11} & p_{12} & \cdots & p_{44} \\ p_{11}\,\partial f/\partial p_{11} & p_{12}\,\partial f/\partial p_{12} & \cdots & p_{44}\,\partial f/\partial p_{44} \\ p_{11}\,\partial g_i/\partial p_{11} & p_{12}\,\partial g_i/\partial p_{12} & \cdots & p_{44}\,\partial g_i/\partial p_{44} \end{bmatrix}. \tag{A.3}$$

This is a 4×16 matrix. Let λ_1 and λ_2 be new unknowns and consider the row vector

$$[\,1 \quad -u_+ \quad \lambda_1 \quad \lambda_2\,] \cdot L. \tag{A.4}$$

Inside the polynomial ring $\mathbb{Q}[p_{ij}, \lambda_k]$ with 20 unknowns, let I denote the ideal generated by $\{f, g_1, g_2, g_3, g_4\}$, the 16 entries of (A.4), and the linear polynomial $p_{11} + p_{12} + \cdots + p_{44} - 1$. Thus, I is the ideal of *Lagrange likelihood equations* introduced in [23], Definition 2. Gross and Rodriguez [23], Proposition 3, showed that I is a 0-dimensional radical ideal, and its number of roots is the ML degree of the variety $V(f, g_1, g_2, g_3, g_4)$. We computed a Gröbner bases for I using the computer algebra software `Magma` [8]. This computation reveals that $V(I)$ consists of 633 points over $\mathbb{C}$. $\square$

PROOF OF COROLLARY 6.6. A matrix P has nonnegative rank 3 if and only if the conditions of Theorem 4.1 are satisfied. Assume $\operatorname{rank}(P) = 3$. By Corollary 4.4, a matrix $P \in \mathcal{M}$ lies on the boundary of $\mathcal{M}$ if and only if it contains a zero or for any triangle Δ with $\mathcal{B} \subseteq \Delta \subseteq \mathcal{A}$ every edge of Δ contains a vertex of $\mathcal{B}$ and (a) or (b) holds. By proof of Theorem 4.1, the latter implies that for each i, j, i', j' for which the conditions of Theorem 4.1 are satisfied there exist k, l such that (4.3) $\cdot$ (4.3) $[k \leftrightarrow l] = 0$. On the other hand, if P lies in the interior of $\mathcal{M}_3^{m\times n}$, then by the proof of Corollary 4.5, the following holds: there exists a triangle Δ with a vertex coinciding with a vertex of $\mathcal{A}$ or with an edge containing an edge of $\mathcal{B}$, and such that the inequality (4.3) $\cdot$ (4.3) $[k \leftrightarrow l] > 0$ holds for all k, l in the corresponding semialgebraic condition. $\square$

APPENDIX B: BASIC CONCEPTS IN ALGEBRAIC GEOMETRY

This appendix gives a synopsis of basic concepts from algebraic geometry that are used in this paper. It furnishes the language to speak about solutions to polynomial equations in many variables.

B.1. Ideals and varieties. Let $R = K[x_1, \ldots, x_n]$ be the ring of polynomials in n variables with coefficients in a subfield K of the real numbers $\mathbb{R}$, usually the rational numbers $K = \mathbb{Q}$. The concept of an ideal I in the ring R is similar to the concept of a normal subgroup in a group.

DEFINITION B.1. A subset $I \subseteq R$ is an *ideal* in R if I is an subgroup of R under addition, and for every $f \in I$ and every $g \in R$ we have $fg \in I$. Equivalently, an ideal I is closed under taking linear combinations with coefficients in the ring R.

Let T be any set of polynomials in R. Their set of zeros is called the *variety* of T. It is denoted

$$V(T) = \{P \in \mathbb{C}^n : f(P) = 0 \text{ for all } f \in T\}.$$

Here, we allow zeros with complex coordinates. This greatly simplifies the study of $V(T)$ because $\mathbb{C}$ is algebraically closed, *that is*, every nonconstant polynomial has a zero.

The *ideal generated by* T, denoted by $\langle T\rangle$, is the smallest ideal in R containing T. Note that

$$V(T) = V\big(\langle T\rangle\big).$$

In computational algebra, it is often desirable to replace the given set T by a *Gröbner basis* of $\langle T\rangle$. This allows us to test ideal membership and to determine geometric properties of the variety $V(T)$.

DEFINITION B.2. A subset $X \subseteq \mathbb{C}^n$ is a *variety* if $X = V(T)$ for some subset $T \subseteq R$.

Hilbert's basis theorem ensures that here T can always be chosen to be a finite set of polynomials. The concept of variety allows us to define a new topology on $\mathbb{C}^n$. It is coarser than the usual topology.

DEFINITION B.3. We define the *Zariski topology* on $\mathbb{C}^n$ by taking closed sets to be the varieties and open sets to be the complements of varieties. This topology depends on the choice of K.

If $K = \mathbb{Q}$, then $X = \{+\sqrt{2}, -\sqrt{2}\}$ is a variety (for $n = 1$) but $Y = \{+\sqrt{2}\}$ is not a variety. Indeed, $X = \overline{Y}$ is the *Zariski closure* of Y, *that is*, it is the smallest variety containing Y, because the minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$ is $x^2 - 2$. Likewise, the set of 1618 points in Example 2.2 is a variety in $\mathbb{C}^2$. It is the Zariski closure of the four points on the topological boundary on the left in Figure 2. The following proposition justifies the fact that the Zariski topology is a topology.

PROPOSITION B.4. *Varieties satisfy the following properties*:

1. *The empty set* $\varnothing = V(R)$ *and the whole space* $\mathbb{C}^n = V(\langle 0\rangle)$ *are varieties.*
2. *The union of two varieties is a variety*:
$$V(I) \cup V(J) = V(I \cdot J) = V(I \cap J).$$
3. *The intersection of any family of varieties is a variety*:
$$\bigcap_{i\in\mathcal{I}} V(I_i) = V\big(\langle I_i : i \in \mathcal{I}\rangle\big).$$

Given any subset $X \subseteq \mathbb{C}^n$ (not necessarily a variety), we define the *ideal* of X by

$$I(X) = \big\{f \in R : f(P) = 0 \text{ for all } P \in X\big\}.$$

Thus, $I(X)$ consists of all polynomials in R that vanish on X. The *Zariski closure* $\overline{X}$ of X equals

$$\overline{X} = V\big(I(X)\big).$$

B.2. Irreducible decomposition. A variety $X \subseteq \mathbb{C}^n$ is *irreducible* if we cannot write $X = X_1 \cup X_2$, where $X_1, X_2 \subsetneq X$ are strictly smaller varieties. An ideal $I \subseteq R$ is *prime* if $fg \in I$ implies $f \in I$ or $g \in I$. For instance, $I(\{\pm\sqrt{2}\}) = \langle x^2 - 2\rangle$ is a prime ideal in $\mathbb{Q}[x]$.

PROPOSITION B.5. *The variety X is irreducible if and only if $I(X)$ is a prime ideal.*

An ideal is *radical* if it is an intersection of prime ideals. The assignment $X \mapsto I(X)$ is a bijection between varieties in $\mathbb{C}^n$ and radical ideals in R. Indeed, every variety X satisfies $V(I(X)) = X$.

PROPOSITION B.6. *Every variety X can be written uniquely as $X = X_1 \cup X_2 \cup \cdots \cup X_m$, where $X_1, X_2, \ldots, X_m$ are irreducible and none of these m components contains any other. Moreover,*

$$I(X) = I(X_1) \cap I(X_2) \cap \cdots \cap I(X_m)$$

is the unique decomposition of the radical ideal $I(X)$ as an intersection of prime ideals.

For an explicit example, with $m = 11$, we consider the ideal (3.4) with the last intersectand removed. In that example, the EM fixed variety X is decomposed into 11 irreducible components.

All ideals I in R can be written as intersections of *primary ideals*. Primary ideals are more general than prime ideals, but they still define irreducible varieties. A *minimal prime* of an ideal I is a prime ideal J such that $V(J)$ is an irreducible component of $V(I)$. See [32], Chapter 5, for the basics on *primary decomposition*.

DEFINITION B.7. Let $I \subseteq R$ be an ideal and $f \in R$ a polynomial. The saturation of I with respect to f is the ideal

$$(I : f^\infty) = \langle g \in R : gf^k \in I \text{ for some } k > 0\rangle.$$

Saturating an ideal I by a polynomial f geometrically means that we obtain a new ideal $J = (I : f^\infty)$ whose variety $V(J)$ contains all components of the variety $V(I)$ except for the ones on which f vanishes. For the more on these concepts from algebraic geometry we recommend the text [11].

B.3. Semialgebraic sets. The discussion above also applies if we consider the varieties $V(T)$ as subsets of $\mathbb{R}^n$ instead of $\mathbb{C}^n$. This brings us to the world of *real algebraic geometry*. The field $\mathbb{R}$ of real numbers is not algebraically closed, it comes with a natural order, and it is fundamental for applications. These features explain why real algebraic geometry is a subject in its own right. In addition to the polynomial equations we discussed so far, we can now also introduce inequalities.

DEFINITION B.8. A *basic semialgebraic set* $X \subseteq \mathbb{R}^n$ is a subset of the form

$$X = \{P \in \mathbb{R}^n : f(P) = 0 \text{ for all } f \in T \text{ and } g(P) \geq 0 \text{ for all } g \in S\},$$

where S and T are finite subsets of R. A *semialgebraic set* is a subset $X \subseteq \mathbb{R}^n$ that is obtained by a finite sequence of unions, intersections and complements of basic semialgebraic sets.

In other words, semialgebraic sets are described by finite Boolean combinations of polynomial equalities and polynomial inequalities. For basic semialgebraic sets, only conjunctions are allowed. For example, the following two simple subsets of the plane are both semialgebraic:

$$X = \{(x, y) \in \mathbb{R}^2 : x \geq 0 \text{ and } y \geq 0\} \quad \text{and}$$
$$Y = \{(x, y) \in \mathbb{R}^2 : x \geq 0 \text{ or } y \geq 0\}.$$

The set X is basic semialgebraic, but Y is not. All convex polyhedra are semialgebraic. A fundamental theorem due to Tarski states that the image of a semialgebraic set under a polynomial map is semialgebraic. Applying this to the map (2.2), we see that the model $\mathcal{M}$ is semialgebraic. The boundary of any semialgebraic set is again semialgebraic. The formulas in Theorem 4.1 and Corollary 6.6 make this explicit. For more on semialgebraic sets and real algebraic geometry, see [5].

Acknowledgements. This work was carried out at the Max-Planck-Institut für Mathematik in Bonn, where all three authors were based during the Fall of 2013.

We thank Aldo Conca, Allen Knutson, Pierre-Jean Spaenlehauer and Matteo Varbaro for helping us with this project. Mathias Drton, Sonja Petrović, John Rhodes, Caroline Uhler and Piotr Zwiernik provided comments on various drafts of the paper. We thank Christopher Miller for pointing out an inaccuracy in Example 2.2.

REFERENCES

[1] AGGARWAL, A., BOOTH, H., O'ROURKE, J., SURI, S. and YAP, C.-K. (1989). Finding minimal convex nested polygons. *Inform. and Comput.* **83** 98–110. MR1019958

[2] ALLMAN, E. S., MATIAS, C. and RHODES, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* **37** 3099–3132. MR2549554

[3] ALLMAN, E. S., RHODES, J., STURMFELS, B. and ZWIERNIK, P. (2014). Tensors of nonnegative rank two. *Linear Algebra Appl. Special Issue on Statistics*. To appear. DOI:10.1016/j.laa.2013.10.046.

[4] ALLMAN, E. S., RHODES, J. A. and TAYLOR, A. (2014). A semialgebraic description of the general Markov model on phylogenetic trees. *SIAM J. Discrete Math.* **28** 736–755. MR3206983

[5] BASU, S., POLLACK, R. and ROY, M.-F. (2003). *Algorithms in Real Algebraic Geometry. Algorithms and Computation in Mathematics* **10**. Springer, Berlin. MR1998147

[6] BEALE, E. (1977). Discussion of "Maximum likelihood from incomplete data via the EM algorithm" by A. Dempster, N. Laird and D. Rubin. *J. Roy. Statist. Soc. Ser. B* **39** 22–23.

[7] BOCCI, C., CARLINI, E. and RAPALLO, F. (2011). Perturbation of matrices and nonnegative rank with a view toward statistical models. *SIAM J. Matrix Anal. Appl.* **32** 1500–1512. MR2869495

[8] BOSMA, W., CANNON, J. and PLAYOUST, C. (1997). The Magma algebra system. I. The user language. *J. Symbolic Comput.* **24** 235–265. MR1484478

[9] BRION, M. and KUMAR, S. (2005). *Frobenius Splitting Methods in Geometry and Representation Theory*. *Progress in Mathematics* **231**. Birkhäuser, Boston. MR2107324

[10] COHEN, J. E. and ROTHBLUM, U. G. (1993). Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra Appl.* **190** 149–168. MR1230356

[11] COX, D., LITTLE, J. and O'SHEA, D. (1992). *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, New York. MR1189133

[12] CRITCH, A. (2013). Binary hidden Markov models and varieties. *J. Algebr. Stat.* **4** 1–30. MR3084209

[13] DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–22. MR0501537

[14] DRTON, M., STURMFELS, B. and SULLIVANT, S. (2009). *Lectures on Algebraic Statistics*. *Oberwolfach Seminars* **39**. Birkhäuser, Basel. MR2723140

[15] EGGERMONT, R., HOROBEŢ, E. and KUBJAS, K. Algebraic boundary of matrices of nonnegative rank at most three. Available at arXiv:1412.1654.

[16] EISENBUD, D. and STURMFELS, B. (1996). Binomial ideals. *Duke Math. J.* **84** 1–45. MR1394747

[17] FAWZI, H. and PARRILO, P. Self-scaled bounds for atomic cone ranks: Applications to nonnegative rank and cp-rank. Available at arXiv:1404.3240.

[18] FIENBERG, S. (1977). Discussion of "Maximum likelihood from incomplete data via the EM algorithm" by A. Dempster, N. Laird and D. Rubin. *J. Roy. Statist. Soc. Ser. B* **39** 29–30.

[19] FIENBERG, S. E., HERSH, P., RINALDO, A. and ZHOU, Y. (2010). Maximum likelihood estimation in latent class models for contingency table data. In *Algebraic and Geometric Methods in Statistics* 27–62. Cambridge Univ. Press, Cambridge. MR2642657

[20] GARCIA, L. D., STILLMAN, M. and STURMFELS, B. (2005). Algebraic geometry of Bayesian networks. *J. Symbolic Comput.* **39** 331–355. MR2168286

[21] GEORGI, B. and SCHLIEP, A. (2006). Context-specific independence mixture modeling for positional weight matrices. *Bioinformatics* **22** 166–173.

[22] GRAYSON, D. and STILLMAN, M. Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.

[23] GROSS, E. and RODRIGUEZ, J. Maximum likelihood geometry in the presence of data zeros. In *Proceedings of the* 39*th International Symposium on Symbolic and Algebraic Computation* 232–239. ACM, New York.

[24] HAUENSTEIN, J., RODRIGUEZ, J. and STURMFELS, B. (2014). Maximum likelihood for matrices with rank constraints. *J. Algebr. Stat.* **5** 18–38.

[25] HUH, J. and STURMFELS, B. (2014). Likelihood geometry. In *Combinatorial Algebraic Geometry*. *Lecture Notes in Math.* **2108** 63–117. Springer, Cham.

[26] LEE, D. and SEUNG, H. (1999). Learning the parts of objects by nonnegative matrix factorization. *Nature* **401** 788–791.

[27] MILLER, E. and STURMFELS, B. (2005). *Combinatorial Commutative Algebra*. *Graduate Texts in Mathematics* **227**. Springer, New York. MR2110098

[28] MOITRA, A. (2012). An almost optimal algorithm for computing nonnegative rank. In *Proceedings of the Twenty-Fourth Annual ACM–SIAM Symposium on Discrete Algorithms* 1454–1464. SIAM, Philadelphia. MR3202990

[29] MOND, D., SMITH, J. and VAN STRATEN, D. (2003). Stochastic factorizations, sandwiched simplices and the topology of the space of explanations. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* **459** 2821–2845. MR2015992

[30] MURRAY, G. (1977). Discussion of "Maximum likelihood from incomplete data via the EM algorithm" by A. Dempster, N. Laird and D. Rubin. *J. Roy. Statist. Soc. Ser. B* **39** 27–28.

[31] PACHTER, L. and STURMFELS, B. (2005). *Algebraic Statistics for Computational Biology*. Cambridge Univ. Press, New York. MR2205865

[32] STURMFELS, B. (2002). *Solving Systems of Polynomial Equations*. *CBMS Regional Conference Series in Mathematics* **97**. AMS, Washington, DC. MR1925796

[33] VAVASIS, S. A. (2009). On the complexity of nonnegative matrix factorization. *SIAM J. Optim.* **20** 1364–1377. MR2587731

[34] WHITE, N. L. (1997). Geometric applications of the Grassmann–Cayley algebra. In *Handbook of Discrete and Computational Geometry* 881–892. CRC, Boca Raton. MR1730204

[35] ZHU, M., JIANG, G. and GAO, S. (2011). Solving the 100 Swiss francs problem. *Math. Comput. Sci.* **5** 195–207. MR2864064

[36] ZWIERNIK, P. and SMITH, J. Q. (2011). Implicit inequality constraints in a binary tree model. *Electron. J. Stat.* **5** 1276–1312. MR2842906

K. KUBJAS
AALTO SCIENCE INSTITUTE
AALTO UNIVERSITY
PO BOX 15500
FI-00076 AALTO
FINLAND
E-MAIL: kaie.kubjas@gmail.com

E. ROBEVA
B. STURMFELS
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: erobeva@gmail.com
bernd@berkeley.edu