The Annals of Statistics
2012, Vol. 40, No. 1, 238–261
DOI: 10.1214/11-AOS957

GEOMETRY OF MAXIMUM LIKELIHOOD ESTIMATION IN GAUSSIAN GRAPHICAL MODELS

BY CAROLINE UHLER

Institute of Science and Technology Austria

We study maximum likelihood estimation in Gaussian graphical models from a geometric point of view. An algebraic elimination criterion allows us to find exact lower bounds on the number of observations needed to ensure that the maximum likelihood estimator (MLE) exists with probability one. This is applied to bipartite graphs, grids and colored graphs. We also study the ML degree, and we present the first instance of a graph for which the MLE exists with probability one, even when the number of observations equals the treewidth.

1. Introduction. In current statistical applications, we are often faced with problems involving a large number of random variables, but only a small number of observations (e.g., [15], Chapter 18). This problem arises, for example, when studying genetic networks: We seek a model potentially involving a vast number of genes, while we are only given gene expression data of a few individuals. Gaussian graphical models have frequently been used to study gene association networks. The maximum likelihood estimator (MLE) of the covariance matrix is computed to describe the interaction between different genes (e.g., [19, 22]). So the following question is of great interest from an applied as well as a theoretical point of view: What is the minimum number of observations needed to guarantee the existence of the MLE in a Gaussian graphical model? It is well known that the MLE exists with probability one if the number of observations is at least as large as the number of variables. In this paper we examine the case of fewer observations.

Gaussian graphical models have been introduced by Dempster [8] under the name of covariance selection models. Subsequently, the graphical representation of these models increased in importance. Lauritzen [17] and Whittaker [21] give introductions to graphical models in general and discuss the connection between graph and probability distribution for Gaussian graphical models.

Gaussian graphical models are regular exponential families. The statistical theory of exponential families, as presented, for example, by Brown [5] or Barndorff-Nielsen [2], is a strong tool to establish existence and uniqueness of the MLE. The MLE exists and is unique if and only if the sufficient statistic lies in the interior of

Received April 2011; revised November 2011.
MSC2010 subject classifications. 62H12, 14Q10, 90C25.
Key words and phrases. Gaussian graphical model, maximum likelihood estimation, matrix completion problems, duality, algebraic statistics, algebraic variety, number of observations, sufficient statistics, treewidth, elimination ideal, ML degree, bipartite graphs.

its convex support. We will give a geometric description of the convex support of the sufficient statistics and discuss the connection to the number of samples.

This paper is organized as follows. In Section 2, we explain the connection between maximum likelihood estimation in Gaussian graphical models and positive definite matrix completion problems. In Section 3, we give a geometric description of the problem, and we develop an exact algebraic algorithm to determine lower bounds on the number of observations needed to ensure existence of the MLE with probability one. In Section 4, we discuss the existence of the MLE for bipartite graphs. Section 5 deals with small graphs. The 3×3 grid motivated this paper and is the original problem posed by Steffen Lauritzen during his lecture on the existence of the MLE in Gaussian graphical models at the "Durham Symposium on Mathematical Aspects of Graphical Models" on July 8, 2008. The 3×3 grid is also the first example of a graph for which the MLE exists with probability one even when the number of observations equals the treewidth of the underlying graph. We conclude this paper with a characterization of Gaussian models on colored 4-cycles in Section 6.

2. Positive definite matrix completion. Let $G = ([m], E)$ be an undirected graph on the vertex set $[m] = \{1, \ldots, m\}$ with edge set E. To simplify notation, we assume that E contains all self-loops, that is, $(i, i) \in E$ for all $i \in [m]$. Let q denote the maximal clique size of G. A graph G is *chordal* if it contains no chordless cycle of length greater than 3. For a nonchordal graph $G = ([m], E)$ one can define a *chordal cover* $G^+ = ([m], E^+)$, which is a chordal graph satisfying $E \subset E^+$. We denote its maximal clique size by q^+. It is useful to introduce the notion of a *minimal chordal cover* $G^* = ([m], E^*)$, where minimality refers to the maximal clique size in the chordal cover, that is, $q^* = \min(q^+)$. The *treewidth* of a graph $\tau(G)$ is defined as

$$\tau(G) = q^* - 1.$$

A random vector X taking values in $\mathbb{R}^m$ is said to satisfy the Gaussian graphical model with graph G if X follows a multivariate normal distribution obeying the undirected pairwise Markov property (e.g., [17, 21]). Assuming the mean to be zero, this property is as follows:

$$(1) \quad X \sim \mathcal{N}(0, \Sigma), \qquad \Sigma \text{ positive definite with } (\Sigma^{-1})_{ij} = 0 \qquad \forall (i, j) \notin E.$$

The results in this paper are based on the assumption that the mean is a *known* vector. In particular, we study the case where the mean is zero. The case where the mean is unknown or partially known is more complex, since mean and covariance matrix can generally not be estimated independently. Gehrmann and Lauritzen [11] describe symmetry relations on the underlying graph which ensure estimability of the mean vector independently from the true covariance matrix Σ.

We denote by $\mathbb{S}^m$ the set of symmetric $m \times m$ matrices and by $\mathbb{S}^m_{\succ 0}$ the open convex cone of positive definite matrices. For a matrix $M \in \mathbb{S}^m$ let M_G denote the *G-partial matrix* consisting of all entries of M corresponding to edges in the graph G, that is,

$$M_G = \big(M_{ij} \mid (i,j) \in E\big).$$

In particular, all diagonal entries of the partial matrix M_G are specified, because we assume that the edge set E contains all self-loops. Equivalently, M_G is the projection of M onto the (coordinates indexed by the) edge set of the graph G:

$$\pi_G : \mathbb{S}^m \to \mathbb{R}^E, \qquad M \mapsto M_G.$$

Let $X_1, \dots, X_n$ denote n independent draws from the distribution $\mathcal{N}(0, \Sigma)$. Then the *sample covariance matrix* is given by

$$S = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T.$$

The G-partial sample covariance matrix S_G plays an important role when studying the existence of the MLE, as seen in the following theorem first proven by Dempster [8].

THEOREM 2.1. *In the Gaussian graphical model on G, the MLE of the covariance matrix Σ exists if and only if the G-partial sample covariance matrix S_G can be completed to a positive definite matrix. Then the MLE $\hat{\Sigma}$ is the unique completion satisfying $(\hat{\Sigma}^{-1})_{ij} = 0$ for all $(i,j) \notin E$.*

So checking existence of the MLE in a Gaussian graphical model is a special matrix completion problem with a rank constraint on the partial matrix given by the number of observations. Matrix completion problems have been extensively studied, and the following result from [14] is very useful in this context.

THEOREM 2.2. *For a graph G the following statements are equivalent*:

(i) *A G-partial matrix $M_G \in \mathbb{R}^E$ has a positive definite completion if and only if all submatrices corresponding to maximal cliques in M_G are positive definite.*

(ii) *G is chordal.*

By combining Theorems 2.1 and 2.2 we get the following result about the existence of the MLE in Gaussian graphical models (see also [6]).

COROLLARY 2.3. *If $n \geq q^*$, the MLE exists with probability* 1. *If $n < q$, the MLE does not exist.*

Note that chordal graphs have $q^* = q$. Therefore, existence of the MLE only depends on the number of observations. For nonchordal graphs, however, there is a gap $q \leq n < q^*$, in which existence of the MLE is not well understood. Cycles and wheels (cycles with one additional completely connected vertex) are the only nonchordal graphs, which have been studied [3, 4, 6]. We will extend the results on cycles and wheels to bipartite graphs $K_{2,m}$ and small grids.

3. Geometry of maximum likelihood estimation in Gaussian graphical models. Every concentration matrix (i.e., inverse of a covariance matrix) in a Gaussian graphical model satisfies the undirected pairwise Markov property (1). The set of all concentration matrices in the model is a convex cone

$$\mathcal{K}_G := \{K \in \mathbb{S}^m_{\succ 0} \mid K_{ij} = 0, \forall (i, j) \notin E\}.$$

Note again that the edge set contains all self-loops, that is, $(i, i) \in E$ for all $i \in [m]$. By taking the inverse of every matrix in $\mathcal{K}_G$, we get the set of all covariance matrices in the model denoted by $\mathcal{K}_G^{-1}$. This is an algebraic variety intersected with the positive definite cone $\mathbb{S}^m_{\succ 0}$ and shown in purple in Figure 1.

In a Gaussian graphical model, the G-partial matrix S_G is a *minimal sufficient statistic* of a sample covariance matrix S (e.g., [17, 21]). So Theorem 2.1 has the following geometric interpretation also explained in Figure 1:

COROLLARY 3.1. *The MLEs $\hat{\Sigma}$ and $\hat{K}$ exist for a given sample covariance matrix S if and only if*

$$\mathrm{fiber}_{\mathcal{G}}(S) := \{\Sigma \in \mathbb{S}^m_{\succ 0} \mid \Sigma_G = S_G\}$$

is nonempty, in which case $\mathrm{fiber}_{\mathcal{G}}(S)$ *intersects $\mathcal{K}_G^{-1}$ in exactly one point, namely the MLE $\hat{\Sigma}$.*

So the MLE $\hat{\Sigma}$ has an algebraic description in terms of the sufficient statistic S_G, that is, $\hat{\Sigma}$ can be represented as a solution to polynomial equations in the

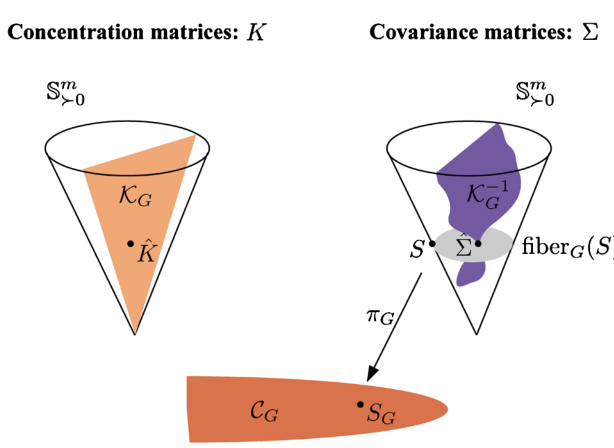


FIG. 1. *Geometry of maximum likelihood estimation in Gaussian graphical models. The cone $\mathcal{K}_G$ consists of all concentration matrices in the model, and $\mathcal{K}_G^{-1}$ is the corresponding set of covariance matrices. The cone of sufficient statistics $\mathcal{C}_G$ is defined as the projection of $\mathbb{S}^m_{\succ 0}$ onto the edge set of G. It is dual to $\mathcal{K}_G$. Given a sample covariance matrix S,* $\mathrm{fiber}_{\mathcal{G}}(S)$ *consists of all positive definite completions of the G-partial matrix S_G, and it intersects $\mathcal{K}_G^{-1}$ in at most one point, namely the MLE $\hat{\Sigma}$.*

sufficient statistic S_G. The maximal degree of these polynomials is called the ML degree. The ML degree describes the map taking a sample covariance matrix S to its maximum likelihood estimate $\hat{\Sigma}$ and is studied in more detail in Section 4.

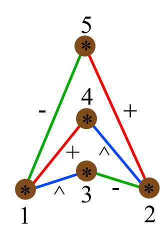
Applying Corollary 3.1, we can describe the set of all sufficient statistics for which the MLE exists. We denote this set by $\mathcal{C}_G$. It is given by the projection of the positive definite cone $\mathbb{S}^m_{\succ 0}$ onto the edge set of the graph G:

$$\mathcal{C}_G := \pi_G(\mathbb{S}^m_{\succ 0}).$$

So $\mathcal{C}_G$ is also a convex cone and shown in dark orange in Figure 1. Moreover, we proved in [20], Proposition 2.1, that the cone of sufficient statistics $\mathcal{C}_G$ is the convex dual to the cone of concentration matrices $\mathcal{K}_G$.

EXAMPLE 3.2. For small-dimensional problems we are able to give a graphical representation of the cone of sufficient statistics $\mathcal{C}_G$. For example, consider the Gaussian graphical model on the bipartite graph $K_{2,3}$ with concentration matrices of the form

$$K = \begin{pmatrix} \lambda_1 & 0 & \lambda_2 & \lambda_3 & \lambda_4 \\ 0 & \lambda_1 & \lambda_4 & \lambda_2 & \lambda_3 \\ \lambda_2 & \lambda_4 & \lambda_1 & 0 & 0 \\ \lambda_3 & \lambda_2 & 0 & \lambda_1 & 0 \\ \lambda_4 & \lambda_3 & 0 & 0 & \lambda_1 \end{pmatrix},$$



Note that in order to reduce the number of parameters and be able to draw $\mathcal{C}_G$ in three-dimensional space, we assume additional equality constraints on the nonzero entries of the concentration matrix, represented by the graph coloring above. Such colored Gaussian graphical models, where the coloring represents equality constraints on the concentration matrix, are called RCON-models and have been introduced in [16].

Without loss of generality we can rescale K and assume that all diagonal entries are one. The cone of concentration matrices $\mathcal{K}_G$ for this model is shown in Figure 2(a). Its algebraic boundary is described by $\{\det(K) = 0\}$ and is shown in Figure 2(b). In this example, the determinant factors into two components, a cylinder and an ellipsoid. Dualizing the boundary of $\mathcal{K}_G$ by the algorithm described in our previous paper ([20], Proposition 2.4) results in the hypersurface shown in Figure 2(e). The double cone is dual to the cylinder in Figure 2(b). By making the double cone transparent as shown in Figure 2(d), we see the enclosed ellipsoid, which is dual to the ellipsoid in Figure 2(b). The cone of sufficient statistics $\mathcal{C}_G$ is shown in Figure 2(c). The MLE exists if and only if the sufficient statistic lies in the interior of this convex body. Using the elimination criterion of Theorem 3.3, we can show that the MLE exists with probability one already for one observation.

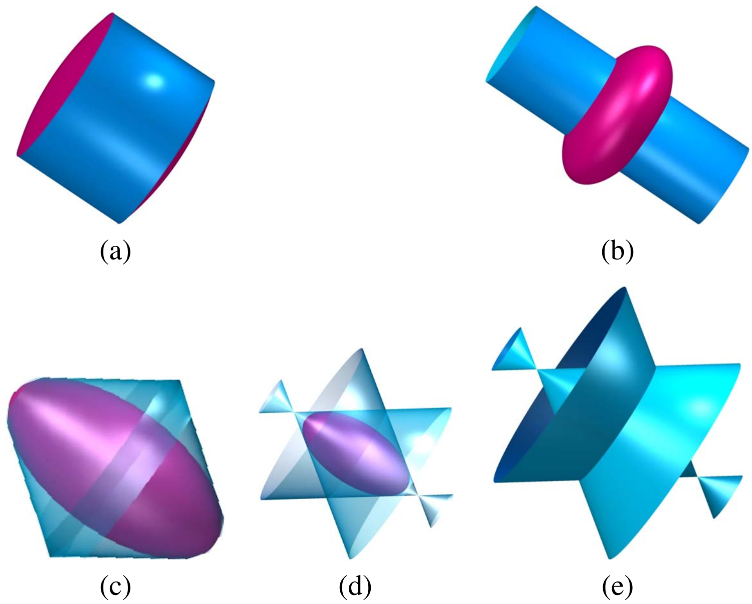


FIG. 2. *These pictures illustrate the convex geometry of maximum likelihood estimation for Gaussian graphical models. The cone of concentration matrices $\mathcal{K}_G$ is shown in* (a), *its algebraic boundary in* (b), *the dual cone of sufficient statistics in* (c) *and its algebraic boundary in* (d) *and* (e), *where* (d) *is the transparent version of* (e).

In this paper, we examine the existence of the MLE for n observations in the range $q \leq n < q^*$, for which the existence of the MLE is not well understood. Geometrically, we look at the manifold of rank n matrices on the boundary of the cone $\mathbb{S}^m_{\succeq 0}$. In general, its projection

$$\pi_G\big(\{M \in \mathbb{S}^m_{\succeq 0} \mid \mathrm{rk}(M) = n\}\big) \tag{2}$$

lies in the topological closure of the cone $\mathcal{C}_G$. The MLE exists with probability one for n observations if and only if the projection (2) lies in the interior of $\mathcal{C}_G$.

Based on the geometric interpretation of maximum likelihood estimation in Gaussian graphical models, we can derive a sufficient condition for the existence of the MLE. The following algebraic elimination criterion can be used as an algorithm to establish existence of the MLE with probability one for n observation.

THEOREM 3.3 (Elimination criterion). *Let $I_{G,n}$ be the elimination ideal obtained from the ideal of* $(n+1) \times (n+1)$*-minors of a symmetric $m \times m$ matrix S of unknowns by eliminating all unknowns corresponding to nonedges of the graph G. If $I_{G,n}$ is the zero ideal, then the MLE exists with probability one for n observations.*

PROOF. The variety corresponding to the ideal of $(n+1) \times (n+1)$-minors of a symmetric $m \times m$ matrix S of unknowns consists of all $m \times m$ matrices of rank at most n. Eliminating all unknowns corresponding to nonedges of the graph G results in the elimination ideal $I_{G,n}$ (see, e.g., [7]) and is geometrically equivalent to a projection onto the cone of sufficient statistics $\mathcal{C}_G$. Let V be the variety corresponding to the elimination ideal $I_{G,n}$. We denote by k its dimension and by μ a k-dimensional Lebesgue measure. The MLE exists with probability one for n observations if

$$\mu(V \cap \partial\mathcal{C}_G) = 0,$$

where $\partial\mathcal{C}_G$ denotes the boundary of the cone of sufficient statistics $\mathcal{C}_G$.

If $I_{G,n}$ is the zero ideal, then the variety V is full-dimensional, and its dimension $\dim(V) = k = \dim(\mathcal{C}_G)$. So if we assume that $\mu(V \cap \partial\mathcal{C}_G) > 0$, then $\mu(\partial\mathcal{C}_G) > 0$, which is a contradiction to $\dim(\partial\mathcal{C}_G) < k$. $\square$

For small examples, the elimination ideal $I_{G,n}$ can be computed, for example, using `Macaulay2` [13], a software system for research in algebraic geometry. If $I_{G,n}$ is not the zero ideal, then an analysis of polynomial inequalities is required. One needs to carefully examine how the components of V are located. The argument is subtle because the algebraic boundary of $\mathcal{C}_G$ may in fact intersect the interior of $\mathcal{C}_G$. So even if the projection V is a component of the algebraic boundary of $\mathcal{C}_G$, the MLE might still exist with positive probability. We will encounter and describe such an example in detail in Section 6.

4. Bipartite graphs. In this section, we first derive the MLE existence results for bipartite graphs $K_{2,m}$ paralleling the results on cycles proven by Buhl [6]. Let the graph $K_{2,m}$ be labeled as shown in Figure 3. A minimal chordal cover is given in Figure 3 (right). As for cycles, for bipartite graphs $K_{2,m}$ we have $q = 2$ and $q^* = 3$. Therefore only the case of $n = 2$ observations is interesting.

Let X_1 and X_2 denote two independent samples from the distribution $\mathcal{N}_{m+2}(0, \Sigma)$, which obeys the undirected pairwise Markov property on $K_{2,m}$. We denote by X the $(m+2) \times 2$ data matrix consisting of the two samples X_1 and X_2 as columns. The rows of X are denoted by $x_1, \dots, x_{m+2}$. Similarly as for cycles in [6], we will

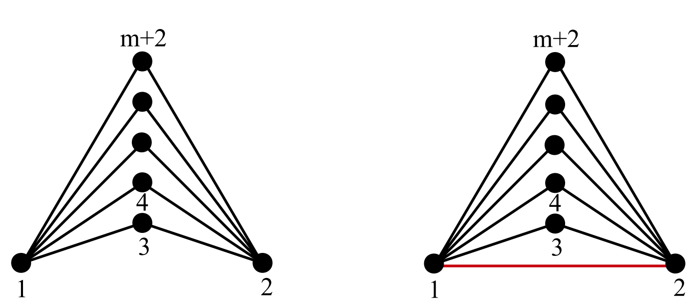


FIG. 3. *Bipartite graph $K_{2,m}$ (left) and minimal chordal cover of $K_{2,m}$ (right).*

describe a criterion on the configuration of data vectors $x_1, \dots, x_{m+2}$ ensuring the existence of the MLE. Our proof is essentially the same argument as used by Buhl [6] for cycles. The following characterization of positive definite matrices of size 3×3 proven in [3] will be helpful in this context.

LEMMA 4.1. *The matrix*

$$\begin{pmatrix} 1 & \cos(\alpha) & \cos(\beta) \\ \cos(\alpha) & 1 & \cos(\gamma) \\ \cos(\beta) & \cos(\gamma) & 1 \end{pmatrix}$$

with $0 < \alpha, \beta, \gamma < \pi$ *is positive definite if and only if*

$$\alpha < \beta + \gamma, \qquad \beta < \alpha + \gamma, \qquad \gamma < \alpha + \beta, \qquad \alpha + \beta + \gamma < 2\pi.$$

PROPOSITION 4.2. *The MLE on the graph* $K_{2,m}$ *exists with probability one for* $n \geq 3$ *observations, and the MLE does not exist for* $n < 2$ *observations. For* $n = 2$ *observations the MLE exists if and only if the lines generated by* x_1 *and* x_2 *are direct neighbors* [*see Figure* 4 (*left*)].

PROOF. Because the problem of existence of the MLE is a positive definite matrix completion problem, we can rescale and rotate the data vectors $x_1, \dots, x_{m+2}$ (i.e., perform an orthogonal transformation) without changing the problem. So without loss of generality we can assume that the vectors $x_1, \dots, x_{m+2} \in \mathbb{R}^2$ have length one, lie in the upper unit half circle and $x_1 = (1, 0)$. We need to prove that the MLE exists if and only if the data configuration is as shown in Figure 4 (middle) or (right).

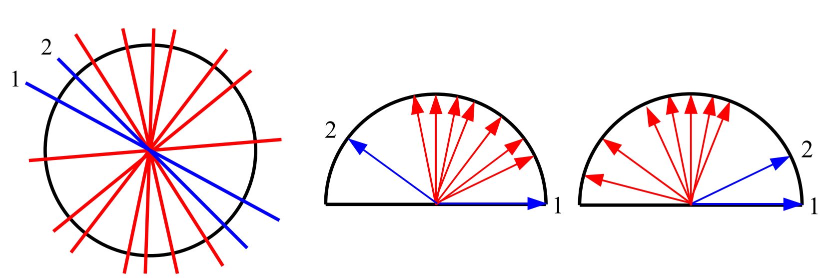


FIG. 4. *The MLE on* $K_{2,m}$ *exists in the following situations. Lines and data vectors corresponding to the variables* 1 *and* 2 *are drawn in blue. Lines and data vectors corresponding to the variables* $3, 4, \dots, m+2$ *are drawn in red.*

Let θ_{ij} denote the angle between vector x_i and x_j. Then the $K_{2,m}$-partial sample covariance matrix $S_{K_{2,m}}$ is of the form

$$\left(\begin{array}{cc|cccc}
1 & \star & \cos(\theta_{13}) & \cos(\theta_{14}) & \cdots & \cos(\theta_{1,m+2}) \\
\star & 1 & \cos(\theta_{23}) & \cos(\theta_{24}) & \cdots & \cos(\theta_{2,m+2}) \\
\hline
\cos(\theta_{13}) & \cos(\theta_{23}) & 1 & \star & \cdots & \star \\
\cos(\theta_{14}) & \cos(\theta_{24}) & \star & 1 & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \star \\
\cos(\theta_{1,m+2}) & \cos(\theta_{2,m+2}) & \star & \cdots & \star & 1
\end{array}\right).$$

We put stars ($\star$) at all positions not corresponding to edges in the graph. The stars represent the entries of the sample covariance matrix which are not part of the sufficient statistics.

The graph $K_{2,m}$ can be extended to a chordal graph by adding one edge as shown in Figure 3 (right). So by Theorem 2.2, $S_{K_{2,m}}$ can be extended to a positive definite matrix if and only if the $(1,2)$ entry of $S_{K_{2,m}}$ can be completed in such a way that all the submatrices corresponding to maximal cliques are positive definite. This is equivalent to the existence of $\rho \in \mathbb{R}$ with $0 < \rho < \pi$ such that

$$\begin{pmatrix} 1 & \cos(\rho) & \cos(\theta_{1i}) \\ \cos(\rho) & 1 & \cos(\theta_{2i}) \\ \cos(\theta_{1i}) & \cos(\theta_{2i}) & 1 \end{pmatrix} \succ 0 \qquad \text{for all } i \in \{3, 4, \ldots, m+2\}.$$

By Lemma 4.1 this occurs if and only if

$$\left.\begin{array}{c} \theta_{1i} - \theta_{2i} \\ \theta_{2i} - \theta_{1i} \end{array}\right\} < \rho < \left\{\begin{array}{c} \theta_{1j} + \theta_{2j} \\ 2\pi - \theta_{1j} - \theta_{2j} \end{array}\right. \qquad \text{for all } i, j \in \{3, 4, \ldots, m+2\},$$

which is equivalent to

$$2\theta_{ai} < \theta_{1i} + \theta_{2i} + \theta_{1j} + \theta_{2j} < 2\pi + 2\theta_{ai} \tag{3}$$

for all $a \in \{1, 2\}$, $i, j \in \{3, 4, \ldots, m+2\}$. We distinguish two cases.

Case 1. There is a vector x_j lying between x_1 and x_2, which implies that $\theta_{1j} + \theta_{2j} = \theta_{12}$. If there was a vector x_i, $i \neq j$, which does not lie between x_1 and x_2, then

$$\theta_{1j} + \theta_{2j} + \theta_{1i} + \theta_{2i} = 2\theta_{1i},$$

which is a contradiction to (3). Hence all vectors $x_3, x_4, \ldots x_{m+2}$ lie between x_1 and x_2, in which case

$$\theta_{1i} + \theta_{2i} + \theta_{1j} + \theta_{2j} = 2\theta_{12},$$

and inequality (3) is satisfied.

Case 2. The vectors x_1 and x_2 are direct neighbors, which implies that $\theta_{1i} + \theta_{2i} = \theta_{12} + 2\theta_{2i}$ for all $i \in \{3, 4, \ldots, m+2\}$, in which case inequality (3) is satisfied.

This proves that for two observations, the MLE exists if and only if the data configuration is as shown in Figure 4 (middle) or (right). □

The geometric explanation of what is happening in this example is that the projection of the positive definite matrices of rank 2 intersects the interior and the boundary of the cone of sufficient statistics $\mathcal{C}_G$ with positive measure. The sufficient statistics originating from data vectors, where lines 1 and 2 are neighbors, lie in the interior of $\mathcal{C}_G$. If lines 1 and 2 are not neighbors, the corresponding sufficient statistics lie on the boundary of the cone $\mathcal{C}_G$, and the MLE does not exist. A similar situation is encountered in Example 6.2 and depicted in Figure 8.

It is worth remarking that if the $m+2$ variables are independent, we can compute the probability of existence of the MLE by a combinatorial argument. In this case, the probability that the MLE exists is given by

$$\frac{2m!}{(m+1)!} = \frac{2}{m+1}.$$

A different approach to gaining a better understanding of maximum likelihood estimation in Gaussian graphical models is to study the ML degree of the underlying graph. The map taking a sample covariance matrix S to its maximum likelihood estimate $\hat{\Sigma}$ is an algebraic function, and its degree is the ML degree of the model. See [9], Definition 2.1.4. The ML degree represents the algebraic complexity of the problem of finding the MLE. This suggests that a larger ML degree results in a more difficult MLE existence problem. We proved in [20] that the ML degree is one if and only if the underlying graph is chordal. It is conjectured in [9], Section 7.4, that the ML degree of the cycle grows exponentially in the cycle length. An interesting contrast to the cycle conjecture is the following theorem, where we prove that the ML degree for bipartite graphs $K_{2,m}$ grows linearly in the number of variables.

THEOREM 4.3. *In a Gaussian graphical model with underlying graph $K_{2,m}$ the ML degree is* $2m+1$.

PROOF. Given a generic matrix $S \in \mathbb{S}^{m+2}$, we fix $\Sigma \in \mathbb{S}^{m+2}$ with entries $\Sigma_{ij} = S_{ij}$ for $(i, j) \in E$ and unknowns $\Sigma_{12} = \Sigma_{21} = y$ and $\Sigma_{ij} = z_{ij}$ for all other $(i, j) \notin E$. We denote by $K = \Sigma^{-1}$ the corresponding concentration matrix. The ML degree of $K_{2,m}$ is the number of complex solutions to

$$(\Sigma^{-1})_{ij} = 0 \qquad \text{for all } (i, j) \notin E.$$

Let A denote the set consisting of the two distinguished vertices $\{1, 2\}$, and let $B = V \setminus A$. In the following we will use the block structure

$$\Sigma = \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix}, \qquad K = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{BA} & K_{BB} \end{pmatrix}.$$

For example, for the graph $K_{2,5}$ the corresponding covariance matrix Σ and concentration matrix K are of the form

$$\Sigma = \left(\begin{array}{cc|ccc} 1 & y & S_{13} & S_{14} & S_{15} \\ y & 1 & S_{23} & S_{24} & S_{25} \\ \hline S_{13} & S_{23} & 1 & z_{34} & z_{35} \\ S_{14} & S_{24} & z_{34} & 1 & z_{45} \\ S_{15} & S_{25} & z_{35} & z_{45} & 1 \end{array}\right),$$

$$K = \left(\begin{array}{cc|ccc} K_{11} & 0 & K_{13} & K_{14} & K_{15} \\ 0 & K_{22} & K_{23} & K_{24} & K_{25} \\ \hline K_{13} & K_{23} & K_{33} & 0 & 0 \\ K_{14} & K_{24} & 0 & K_{44} & 0 \\ K_{15} & K_{25} & 0 & 0 & K_{55} \end{array}\right).$$

Note that the block K_{BB} is a diagonal matrix. Hence the Schur complement

$$\Sigma_{BB} - \Sigma_{BA}\Sigma_{AA}^{-1}\Sigma_{AB}$$

is also a diagonal matrix. Writing out the off-diagonal entries of this matrix results in the following expression for the variables z in terms of the variable y:

$$z_{ij} = -\frac{1}{1-y^2}\big(y(S_{1i}S_{2j} + S_{1j}S_{2i}) - S_{1i}S_{1j} - S_{2i}S_{2j}\big).$$

Setting the minor M_{12} of Σ to zero results in the last equation of the form

$$y\det(\Sigma_{BB}) + (\text{polynomial in } z \text{ of degree } m-1) = 0. \tag{4}$$

We note that $\det(\Sigma_{BB})$ is a polynomial in z of degree m, where the degree 0 term is 1. So by multiplying equation (4) with $(1-y^2)^m$, we get a degree $2m+1$ equation in y and therefore $2m+1$ complex solutions for y. For each solution of y we get one solution for the variables z, which proves that the ML degree of $K_{2,m}$ is $2m+1$. □

Bipartite graphs and cycles are classes of graphs with $q = 2$ and $q^* = 3$. What can we say about such graphs in general regarding the existence of the MLE for two observations? A related question has been studied from a purely algebraic point of view in [4]. A cycle-completable graph is defined to be a graph such that every partial matrix M_G has a positive definite completion if and only if M_G is positive definite on all submatrices corresponding to maximal cliques in the graph, and all submatrices corresponding to cycles in the graph can be completed to a positive definite matrix. It is shown in [4] that a graph is cycle-completable if and only if there is a chordal cover with no new 4-clique.

Buhl [6] studied cycles from a more statistical point of view and described a criterion on the data vectors for the existence of the MLE for two observations. Combining the results of [4] and [6], we get the following result:

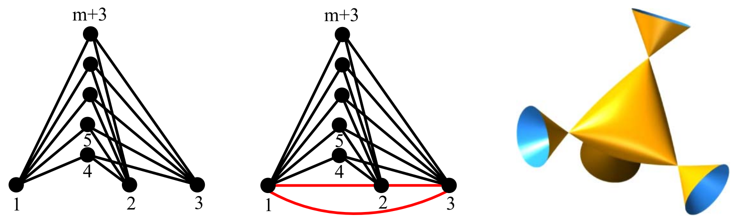


FIG. 5. *Bipartite graph $K_{3,m}$ (left) and minimal chordal cover of $K_{3,m}$ (middle). The tetrahedron-shaped pillow consisting of all positive semidefinite 3×3 matrices with ones on the diagonal is shown in the right figure.*

COROLLARY 4.4. *Let G be a graph with $q = 2$ and $q^* \geq 3$. Then the following statements are equivalent*:

(i) *For $n = 2$ observations, the MLE exists if and only if Buhl's cycle condition is satisfied on every induced cycle.*

(ii) $q^* = 3$.

This result solves the problem of existence of the MLE for all graphs with $q = 2$ and $q^* = 3$. Note that Corollary 4.4 is more general than Proposition 4.2. The proof, however, is more involved and less constructive.

For bipartite graphs $K_{3,m}$ the situation is more complicated and we do not yet have results similar to Proposition 4.2 and Theorem 4.3. We will nevertheless describe some preliminary results.

Let the graph $K_{3,m}$ be labeled as shown in Figure 5. A minimal chordal cover is given in Figure 5 (middle). Hence, $q = 2$ and $q^* = 4$. The convex body shown in Figure 5 (right) consists of all positive semidefinite 3×3 matrices with ones on the diagonal. We call it the tetrahedron-shaped pillow. We will prove that the existence of the MLE is equivalent to a nonempty intersection of such inflated and shifted tetrahedron-shaped pillows.

COROLLARY 4.5. *The MLE on the graph $K_{3,m}$ exists if and only if the m inflated and shifted tetrahedron-shaped pillows corresponding to the maximal cliques in a minimal chordal cover of $K_{3,m}$ shown in Figure 5 (middle) have nonempty intersections.*

PROOF. Applying Theorem 2.2 in a similar way as in the proof of Theorem 4.2, the partial covariance matrix $S_{K_{3,m}}$ can be extended to a positive definite matrix if and only if the entries corresponding to the missing edges $(1, 2)$, $(1, 3)$ and $(2, 3)$ can be completed in such a way that all the submatrices corresponding to maximal cliques in the minimal chordal cover (Figure 5, middle) are positive definite. This is equivalent to the existence of $x, y, z \in \mathbb{R}$ with $-1 < x, y, z < 1$

such that

$$\begin{pmatrix} 1 & s_{1i} & s_{2i} & s_{3i} \\ s_{1i} & 1 & x & y \\ s_{2i} & x & 1 & z \\ s_{3i} & y & z & 1 \end{pmatrix} \succ 0 \qquad \text{for all } i \in \{4, 5, \ldots, m+3\}, \tag{5}$$

where s_{ai}, $a \in \{1, 2, 3\}$, $i \in \{4, 5, \ldots, m + 3\}$ are the sufficient statistics corresponding to edges in the bipartite graph $K_{3,m}$. Using Schur complements and rescaling, (5) holds if and only if

$$\begin{pmatrix} 1 & x_i & y_i \\ x_i & 1 & z_i \\ y_i & z_i & 1 \end{pmatrix} \succ 0 \qquad \text{for all } i \in \{4, 5, \ldots, m+3\}, \tag{6}$$

where

$$x_i = \frac{x - s_{1i}s_{2i}}{\sqrt{1 - s_{1i}^2}\sqrt{1 - s_{2i}^2}}, \qquad y_i = \frac{y - s_{1i}s_{3i}}{\sqrt{1 - s_{1i}^2}\sqrt{1 - s_{3i}^2}},$$

$$z_i = \frac{x - s_{2i}s_{3i}}{\sqrt{1 - s_{2i}^2}\sqrt{1 - s_{3i}^2}}.$$

So the MLE exists if and only if the inflated and shifted tetrahedron-shaped pillows corresponding to the inequalities in (6) have nonempty intersection. □

We used the software package `Macaulay2` to compute the ML degree of $K_{3,m}$ for $m \leq 4$. It is an open problem to find a general formula or a recurrence relation for the ML degree of $K_{l,m}$, where $l \geq 3$.

m	1	2	3	4
ML degree	1	7	57	131

5. Small graphs. In this section we analyze the 3×3 grid in particular and complete the discussion of [20] with the number of observations and the corresponding existence probability of the MLE for all graphs with 5 or less vertices.

The 3×3 grid is shown in Figure 6 (left) and has $q = 2$ and $q^* = 4$. This example represents the starting point of this paper and is the original problem posed by

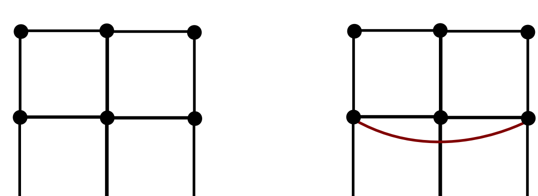


FIG. 6. 3×3 *grid* $\mathcal{H}$ (*left*) *and grid with additional edge* $\mathcal{H}'$ (*right*).

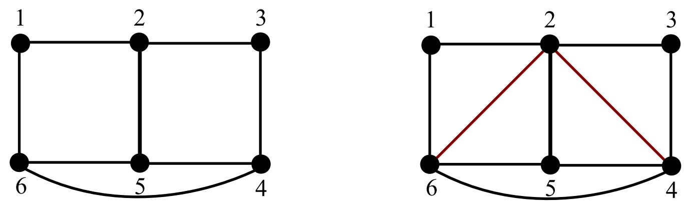


FIG. 7. *Graph $\mathcal{G}$ (left) and minimal chordal cover of $\mathcal{G}$ (right).*

Steffen Lauritzen during his lecture at the "Durham Symposium on Mathematical Aspects of Graphical Models" in 2008. As a preparation, we first discuss the existence of the MLE for the graph $\mathcal{G}$ on six vertices shown in Figure 7. The graph $\mathcal{G}$ also has $q = 2$ and $q^* = 4$, and is the first example for which we can prove that the bound $n \geq q^*$ for the existence of the MLE with probability one is not tight and that the MLE can exist with probability one, even when the number of observations equals the treewidth.

THEOREM 5.1. *The MLE on the graph $\mathcal{G}$ (Figure 7, left) exists with probability one for $n = 3$ observations.*

PROOF. We compute the ideal $I_{\mathcal{G},3}$ by eliminating the variables $s_{13}, s_{15}, s_{16}, s_{24}, s_{26}, s_{34}, s_{35}$ from the ideal of 4×4 minors of the matrix S given in (7). This results in the zero ideal, which by Theorem 3.3 completes the proof. $\square$

REMARK 5.2. Theorem 5.1 is equivalent to the following purely algebraic statement. Let

$$S = \begin{pmatrix} 1 & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ s_{12} & 1 & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{13} & s_{23} & 1 & s_{34} & s_{35} & s_{36} \\ s_{14} & s_{24} & s_{34} & 1 & s_{45} & s_{46} \\ s_{15} & s_{25} & s_{35} & s_{45} & 1 & s_{56} \\ s_{16} & s_{26} & s_{36} & s_{46} & s_{56} & 1 \end{pmatrix} \in \mathbb{S}^6_{\succeq 0} \tag{7}$$

with $\operatorname{rank}(S) = 3$. Then there exist $x, y, a, b, c, d, e \in \mathbb{R}$ such that

$$S' = \begin{pmatrix} 1 & s_{12} & a & s_{14} & b & c \\ s_{12} & 1 & s_{23} & x & s_{25} & y \\ a & s_{23} & 1 & d & e & s_{36} \\ s_{14} & x & d & 1 & s_{45} & s_{46} \\ b & s_{25} & e & s_{45} & 1 & s_{56} \\ c & y & s_{36} & s_{46} & s_{56} & 1 \end{pmatrix} \in \mathbb{S}^6_{\succ 0}.$$

So any partial matrix of rank 3 with specified entries at all positions corresponding to edges in $\mathcal{G}$ can be completed to a positive definite matrix.

COROLLARY 5.3. *Let $\mathcal{H}$ be the 3×3-grid shown in Figure 6. Then the MLE on $\mathcal{H}$ exists with probability one for $n \geq 3$ observations, and the MLE does not exist for $n < 2$ observations.*

PROOF. First note that Groebner bases computations are extremely memory intensive and the elimination ideal $I_{\mathcal{H},3}$ cannot be computed directly due to insufficient memory. We solve this problem by gluing together smaller graphs. The probability of existence of the MLE for the 3×3 grid $\mathcal{H}$ is at least as large as the existence probability when the underlying graph is $\mathcal{H}'$. The graph $\mathcal{H}'$ is a clique sum of two graphs of the form $\mathcal{G}$, for which the MLE existence probability is one for $n \geq 3$. □

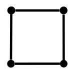
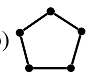
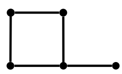
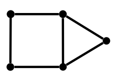
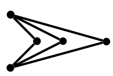
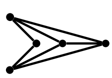
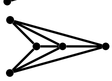
This example shows that although we are not able to compute the elimination ideal for large graphs directly, the algebraic elimination criterion (Theorem 3.3) is still useful also in this situation. We can study small graphs with the elimination criterion and glue them together using clique sums to build larger graphs.

For two observations on the 3×3 grid, the cycle conditions are necessary but not sufficient for the existence of the MLE (Corollary 4.4). Unlike for bipartite graphs $K_{2,m}$, the existence of the MLE does not only depend on the ordering of the lines corresponding to the data vectors in $\mathbb{R}^2$. By simulations with the `Matlab` software `cvx` [12], one can easily find orderings for which the MLE sometimes exists and sometimes does not. Finding a necessary and sufficient criterion for the existence of the MLE for two observations remains an open problem.

We now complete the discussion of [20] with the number of observations and the corresponding existence probability of the MLE for all graphs with 5 or less vertices. All nonchordal graphs with 5 or less vertices are shown in Table 1. The 4-cycle and 5-cycle in (a) and (b) are covered by Buhl's results [6]. The graphs in (c) and (d) are clique sums of two graphs and therefore completable if and only if the submatrices corresponding to the two subgraphs are completable. Graph (e) is the bipartite graph $K_{2,3}$ and covered by Theorem 4.2. For the graph in (f) $q = 3$ and $q^* = 4$. Applying the elimination criterion from Theorem 3.3 shows that three observations are sufficient for the existence of the MLE. The last example, the 5-wheel in graph (g), is also covered by Buhl's results [6].

6. Colored Gaussian graphical models. For some applications, symmetries in the underlying Gaussian graphical model can be assumed. Adding symmetry to the conditional independence restrictions of a graphical model reduces the number of parameters and in some cases also the number of observations needed for the existence of the MLE. The symmetry restrictions can be represented by a graph coloring, where edges, or vertices, respectively, have the same coloring if the corresponding elements of the concentration matrix are equal. Such models are called RCON-models [16]. We discussed such a model earlier in Example 3.2.

TABLE 1
This table shows the number of observations (obs.) and the corresponding MLE existence probability for all nonchordal graphs on 5 or fewer vertices

Graph G	**1 obs.**	**2 obs.**	**3 obs.**	**≥4 obs.**
(a) 	No	$p \in (0, 1)$	$p = 1$	$p = 1$
(b) 	No	$p \in (0, 1)$	$p = 1$	$p = 1$
(c) 	No	$p \in (0, 1)$	$p = 1$	$p = 1$
(d) 	No	No	$p = 1$	$p = 1$
(e) 	No	$p \in (0, 1)$	$p = 1$	$p = 1$
(f) 	No	No	$p = 1$	$p = 1$
(g) 	No	No	$p \in (0, 1)$	$p = 1$

We denote the uncolored graph by G and the colored graph by $\mathcal{G}$. Note that in this section the graph G does not contain any self-loops. Let the vertices be colored with p different colors and the edges with q different colors:

$$V = V_1 \sqcup V_2 \sqcup \cdots \sqcup V_p, \qquad p \leq |V|,$$
$$E = E_1 \sqcup E_2 \sqcup \cdots \sqcup E_q, \qquad q \leq |E|.$$

Then the set of all concentration matrices $\mathcal{K}_{\mathcal{G}}$ consists of all positive definite matrices satisfying:

- $K_{\alpha\beta} = 0$ for any pair of vertices α, β that do not form an edge in G.
- $K_{\alpha\alpha} = K_{\beta\beta}$ for any pair of vertices α, β in a common vertex color class V_i.
- $K_{\alpha\beta} = K_{\gamma\delta}$ for any pair of edges (α, β), (γ, δ) in a common edge color class E_j.

This means that also for RCON-models the set $\mathcal{K}_{\mathcal{G}}$ is defined by linear equations on the concentration matrix K. So the geometry of maximum likelihood estimation is the same as that explained in Section 3, and it is straightforward to derive the equivalent of Theorem 2.1 for colored Gaussian graphical models.

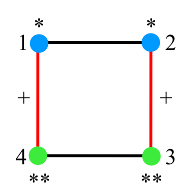
THEOREM 6.1. *In a colored Gaussian graphical model on $\mathcal{G}$ the MLE of the covariance matrix Σ exists if and only if there is a positive definite matrix $\tilde{\Sigma}$ such* *that*

$$\sum_{\alpha\in V_i} \tilde{\Sigma}_{\alpha\alpha} = \sum_{\alpha\in V_i} S_{\alpha\alpha} \quad \textit{and} \quad \sum_{(\alpha,\beta)\in E_j} \tilde{\Sigma}_{\alpha\beta} = \sum_{(\alpha,\beta)\in E_j} S_{\alpha\beta}$$

for all vertex color classes $V_1, \dots, V_p$ and all edge color classes $E_1, \dots, E_q$. Then the MLE $\hat{\Sigma}$ is the unique completion with $(\hat{\Sigma}^{-1})_{\alpha\alpha} = (\hat{\Sigma}^{-1})_{\beta\beta}$ for any pair of vertices α, β in a common vertex color class V_i, $(\hat{\Sigma}^{-1})_{\alpha\beta} = (\hat{\Sigma}^{-1})_{\gamma\delta}$ for any pair of edges (α, β), (γ, δ) in a common edge color class E_j, and $(\hat{\Sigma}^{-1})_{\alpha\beta} = 0$ for all $(\alpha, \beta) \notin E$.

EXAMPLE 6.2 (Frets's heads). We revisit the heredity study of head dimensions known as *Frets's heads* reported in [10]. Part of the original data are the length and breadth of the heads of 25 pairs of first and second sons. This data set was also discussed in [18, 20]. The data supports the following colored Gaussian graphical model, where the joint distribution remains the same when the two sons are exchanged:

$$K = \begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix},$$



In this graph, variable 1 corresponds to the length of the first son's head, variable 2 to the length of the second son's head, variable 3 to the breadth of the second son's head and variable 4 to the breadth of the first son's head. Color classes consisting only of one edge (or vertex) are displayed in black.

Given a sample covariance matrix $S = (s_{ij})$, the five sufficient statistics for this model according to the graph coloring are

$$t_1 = s_{11} + s_{22}, \qquad t_2 = s_{33} + s_{44}, \qquad t_3 = 2s_{12},$$
$$t_4 = 2(s_{23} + s_{14}), \qquad t_5 = 2s_{34}.$$

The algebraic boundary of the cone of sufficient statistics $\mathcal{C}_{\mathcal{G}}$ is computed in [20] and given by the polynomial

$$H_{\mathcal{G}} = (t_1 - t_3) \cdot (t_1 + t_3) \cdot (t_2 - t_5) \cdot (t_2 + t_5) \times (4t_2^2 t_3^2 - 4t_1 t_2 t_4^2 + t_4^4 + 8t_1 t_2 t_3 t_5 - 4t_3 t_4^2 t_5 + 4t_1^2 t_5^2).$$

For two observations the elimination ideal $I_{\mathcal{G},2}$ is the zero ideal. Therefore, the MLE exists with probability 1 for two or more observations in this model. For one observation we get

$$I_{\mathcal{G},1} = \langle 4t_2^2 t_3^2 - 4t_1 t_2 t_4^2 + t_4^4 + 8t_1 t_2 t_3 t_5 - 4t_3 t_4^2 t_5 + 4t_1^2 t_5^2 \rangle,$$

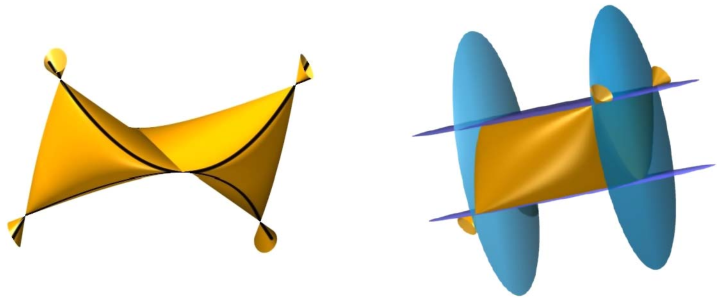


FIG. 8. *All possible sufficient statistics from one observation are shown on the left. The cone of sufficient statistics is shown on the right.*

which corresponds to one of the components of the algebraic boundary of the cone of sufficient statistics. In this example, the algebraic boundary of the cone of sufficient statistics intersects its interior. This is illustrated in Figure 8. In order to get a graphical representation in three-dimensional space, we fixed t_3 and t_5. The variety corresponding to $I_{\mathcal{G},1}$ is shown on the left. We call this hypersurface the bow tie. The cone of sufficient statistics $\mathcal{C}_{\mathcal{G}}$ is the convex hull of the bow tie and shown in Figure 8 (right). Its boundary consists of four planes corresponding to the components $t_1 - t_3$, $t_1 + t_3$, $t_2 - t_5$ and $t_2 + t_5$ shown in blue, and the bows of the bow tie shown in yellow. The black curves show where the planes touch the bow tie. Note that the upper and lower two triangles of the bow tie lie in the interior of $\mathcal{C}_{\mathcal{G}}$. Only the two bows are part of the boundary of $\mathcal{C}_{\mathcal{G}}$. So the MLE exists if the sufficient statistic lies on one of the triangles of the bow tie, and it does not exist if the sufficient statistic lies on one of the bows of the bow tie. Consequently, for one observation the MLE exists with probability strictly between 0 and 1.

A different approach is to run simulations, for example, using `cvx`. We can generate vectors of length four and compute the MLE by solving a convex optimization problem. If `cvx` finds a solution, the MLE exists. For this example, however, `cvx` sometimes does not find a solution, which supports the hypothesis that the MLE exists with probability strictly between 0 and 1 for one observation. In the following, we give a formal proof by characterizing the set of vectors in $\mathbb{R}^4$ for which the MLE exists/does not exist.

For this example, we can exactly characterize not just the sufficient statistics, but also the observations, for which the MLE exists. In other words, we can characterize the observations whose sufficient statistics lie on the triangles of the bow tie. First, note that by exchanging variables 1 and 2 and simultaneously exchanging variables 3 and 4, we get the same model. This means that from one observation $X_1 = (x_1, x_2, x_3, x_4)$ we can generate a second observation $X_2 = (x_2, x_1, x_4, x_3)$.

So the resulting data matrix is given by

$$X = \begin{pmatrix} x_1 & x_2 \\ x_2 & x_1 \\ x_3 & x_4 \\ x_4 & x_3 \end{pmatrix}.$$

Applying Buhl's result about two observations on a Gaussian cycle [6], the MLE exists if and only if the lines corresponding to the vectors

$$y_1 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad y_2 = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}, \qquad y_3 = \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}, \qquad y_4 = \begin{pmatrix} x_4 \\ x_3 \end{pmatrix}$$

are not graph consecutive. This is the case if and only if

$$|x_1| > |x_2| \text{ and } |x_3| > |x_4| \quad \text{or} \quad |x_1| < |x_2| \text{ and } |x_3| < |x_4|. \tag{8}$$

Hence, the MLE for one observation exists if and only if the data is inconsistent, meaning that the head of the first (second) son is longer than the head of the second (first) son, but the breadth is smaller. In this situation the corresponding sufficient statistics lie on the triangles of the bow tie in Figure 8. Otherwise the corresponding sufficient statistics lie on the bows of the bow tie. If Σ is diagonal, the MLE exists with probability 0.5, since all configurations in (8) have the same probability.

In our previous paper [20] we found the defining polynomial $\mathcal{H}_{\mathcal{G}}$ of the cone of sufficient statistics for all colored Gaussian graphical models on the 4-cycle, which have the property that edges in the same color class connect the same vertex color classes. Such models have been studied in [16] and are of special interest, because they are invariant under rescaling of variables in the same vertex color class. In Tables 2 and 3, we complete the discussion of [20] with the number of observations and the corresponding existence probability of the MLE.

For every colored 4-cycle, we computed the elimination ideal $I_{\mathcal{G},n}$ for $n = 1, 2, 3$. If it is the zero ideal, we know from Theorem 3.3 that the MLE exists with probability one. If $I_{\mathcal{G},n}$ is nonzero, we run simulations using `cvx`. If we find examples for which the MLE exists and other examples for which the MLE does not exist, it indicates that the MLE exists with probability strictly between 0 and 1 for n observations. In cases where simulations do not yield any counterexamples, we need to prove that the MLE does indeed not exist by carefully analyzing the components corresponding to the ideal $I_{\mathcal{G},n}$. This is the case for one observation on the graphs (9), (11), (14) and (17). Note that the graphical models (9) and (11) are sub-models of (14) and (17). So if we prove that the MLE does not exist for one observation on the graphs (9) and (11), this follows also for the graphs (14) and (17).

If the cone $\mathcal{C}_{\mathcal{G}}$ for the graphs (9) and (11) is a basic open semialgebraic set (see, e.g., [1]), then $\mathcal{C}_{\mathcal{G}}$ does not meet its algebraic boundary, and the MLE does not exist for one observation. So we end with the following conjecture which would answer the question marks in Table 2:

TABLE 2
Results on the number of observations and the MLE existence probability for all colored Gaussian graphical models with some symmetry restrictions (namely, edges in the same color class connect the same vertex color classes) on the 4-cycle

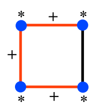
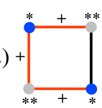
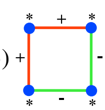
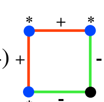
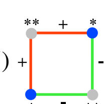
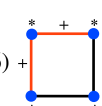
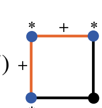
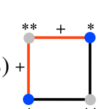
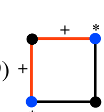
Graph	K	**1 obs.**	**2 obs.**	**≥3 obs.**
(1) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_2 \\ \lambda_2 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(2) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_3 \\ \lambda_3 & 0 & \lambda_3 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(3) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_3 \\ \lambda_2 & 0 & \lambda_3 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(4) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_4 \\ \lambda_3 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(5) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_4 \\ \lambda_3 & 0 & \lambda_4 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(6) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_4 \\ \lambda_2 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(7) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_3 & 0 & \lambda_5 & \lambda_1 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$
(8) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_3 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$
(9) 	$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_4 \\ \lambda_4 & \lambda_2 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_3 & \lambda_6 \\ \lambda_4 & 0 & \lambda_6 & \lambda_2 \end{pmatrix}$	No?	$p = 1$	$p = 1$

TABLE 2
(*Continued*)

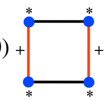
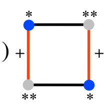
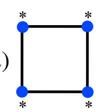
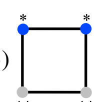
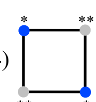
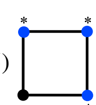
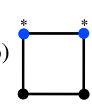
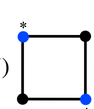
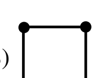
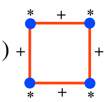
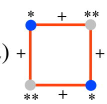
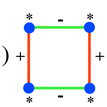
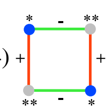
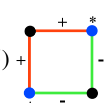
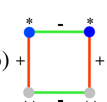
Graph	K	**1 obs.**	**2 obs.**	**≥3 obs.**
(10) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_3 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_4 \\ \lambda_3 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(11) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	No?	$p = 1$	$p = 1$
(12) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_5 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_4 \\ \lambda_5 & 0 & \lambda_4 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(13) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_6 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_6 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$
(14) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_6 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_6 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	No?	$p = 1$	$p = 1$
(15) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_6 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_5 \\ \lambda_6 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$
(16) 	$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_7 \\ \lambda_4 & \lambda_1 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_2 & \lambda_6 \\ \lambda_7 & 0 & \lambda_6 & \lambda_3 \end{pmatrix}$	No	$p = 1$	$p = 1$
(17) 	$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_7 \\ \lambda_4 & \lambda_2 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_1 & \lambda_6 \\ \lambda_7 & 0 & \lambda_6 & \lambda_3 \end{pmatrix}$	No?	$p = 1$	$p = 1$
(18) 	$\begin{pmatrix} \lambda_1 & \lambda_5 & 0 & \lambda_8 \\ \lambda_5 & \lambda_2 & \lambda_6 & 0 \\ 0 & \lambda_6 & \lambda_3 & \lambda_7 \\ \lambda_8 & 0 & \lambda_7 & \lambda_4 \end{pmatrix}$	No	$p \in (0, 1)$	$p = 1$

TABLE 3

All RCOP-models (introduced in [16]), that is, graphs with an additional permutation property on the 4-cycle

Graph	K	1 obs.	2 obs.	≥3 obs.
(1) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_2 \\ \lambda_2 & \lambda_1 & \lambda_2 & 0 \\ 0 & \lambda_2 & \lambda_1 & \lambda_2 \\ \lambda_2 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(2) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_3 \\ \lambda_3 & 0 & \lambda_3 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(3) 	$\begin{pmatrix} \lambda_1 & \lambda_2 & 0 & \lambda_3 \\ \lambda_2 & \lambda_1 & \lambda_3 & 0 \\ 0 & \lambda_3 & \lambda_1 & \lambda_2 \\ \lambda_3 & 0 & \lambda_2 & \lambda_1 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(4) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_2 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_1 & \lambda_3 \\ \lambda_4 & 0 & \lambda_3 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(5) 	$\begin{pmatrix} \lambda_1 & \lambda_4 & 0 & \lambda_4 \\ \lambda_4 & \lambda_2 & \lambda_5 & 0 \\ 0 & \lambda_5 & \lambda_3 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p = 1$	$p = 1$	$p = 1$
(6) 	$\begin{pmatrix} \lambda_1 & \lambda_3 & 0 & \lambda_4 \\ \lambda_3 & \lambda_1 & \lambda_4 & 0 \\ 0 & \lambda_4 & \lambda_2 & \lambda_5 \\ \lambda_4 & 0 & \lambda_5 & \lambda_2 \end{pmatrix}$	$p \in (0, 1)$	$p = 1$	$p = 1$

CONJECTURE 6.3. *The cones $\mathcal{C}_{\mathcal{G}}$ corresponding to the graphs* (9) *and* (11) *are basic open semialgebraic sets.*

7. Conclusion. In this paper, we explained the geometry of maximum likelihood estimation in Gaussian graphical models. The geometric picture can be translated into an algebraic criterion (Theorem 3.3), which allows us to find exact lower bounds on the number of observations needed for the existence of the MLE (with probability 1). Theorem 3.3 holds for any Gaussian graphical model. However, the practical implementation of Theorem 3.3 is based on Groebner bases computations, which are extremely memory intensive. Theorem 5.1 and Corollary 5.3 show the power but also the limitations of computational algebraic geometry. We are, in practice, only able to apply the algebraic elimination criterion directly to very small graphs. One way of getting results for larger graphs is to find a clique decomposition into small subgraphs, which can be handled individually. A different future line of research is to use the small examples to understand the existence of the MLE asymptotically. If we fix a class of graphs, for example, cycles or grids, what can we say about the existence of the MLE as the number of vertices tends to infinity? Medium-sized graphs, however, remain untouched by both approaches, and finding the minimum number of observations needed for the existence of the MLE for such graphs is an interesting open problem.

Acknowledgments. I wish to thank Bernd Sturmfels for many helpful discussions and Steffen Lauritzen for introducing me to the problem of the existence of the MLE in Gaussian graphical models. I would also like to thank two referees who provided helpful comments on the original version of this paper.

REFERENCES

[1] ACQUISTAPACE, F., BROGLIA, F. and VÉLEZ, M. P. (1999). Basicness of semialgebraic sets. *Geom. Dedicata* **78** 229–240. MR1725377

[2] BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. MR0489333

[3] BARRETT, W., JOHNSON, C. R. and TARAZAGA, P. (1993). The real positive definite completion problem for a simple cycle. *Linear Algebra Appl.* **192** 3–31. MR1236734

[4] BARRETT, W. W., JOHNSON, C. R. and LOEWY, R. (1996). The real positive definite completion problem: Cycle completability. *Mem. Amer. Math. Soc.* **122** viii+69. MR1342017

[5] BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **9**. IMS, Hayward, CA. MR0882001

[6] BUHL, S. L. (1993). On the existence of maximum likelihood estimators for graphical Gaussian models. *Scand. J. Stat.* **20** 263–270. MR1241392

[7] COX, D., LITTLE, J. and O'SHEA, D. (1997). *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, New York.

[8] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175.

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

[10] FRETS, G. P. (1921). Heredity of head form in man. *Genetica* **3** 193–400.

[11] GEHRMANN, H. and LAURITZEN, S. L. (2011). Estimation of means in graphical Gaussian models with symmetries. Preprint. Available at http://arxiv.org/abs/1101.3709.

[12] GRANT, M. and BOYD, S. CVX, a Matlab software for disciplined convex programming. Available at http://cvxr.com/cvx/.

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

[14] GRONE, R., JOHNSON, C. R., DE SÁ, E. M. and WOLKOWICZ, H. (1984). Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl.* **58** 109–124. MR0739282

[15] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294

[16] HØJSGAARD, S. and LAURITZEN, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 1005–1027. MR2530327

[17] LAURITZEN, S. L. (1996). *Graphical Models*. Clarendon, Oxford. MR1419991

[18] MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press, London.

[19] SCHÄFER, J. and STRIMMER, K. (2005). Learning large-scale graphical Gaussian models from genomic data. In *Science of Complex Networks*: *From Biology to the Internet and WWW*. The American Institute of Physics, College Park, MD.

[20] STURMFELS, B. and UHLER, C. (2010). Multivariate Gaussian, semidefinite matrix completion, and convex algebraic geometry. *Ann*. *Inst*. *Statist*. *Math*. **62** 603–638. MR2652308

[21] WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester. MR1112133

[22] WU, X., YE, Y. and SUBRAMANIAN, K. R. (2003). Interactive analysis of gene interactions using graphical Gaussian model. *ACM SIGKDD Workshop on Data Mining in Bioinformatics* **3** 63–69.

INSTITUTE OF SCIENCE
AND TECHNOLOGY AUSTRIA
AM CAMPUS 1
A-3400 KLOSTERNEUBURG
AUSTRIA
E-MAIL: caroline.uhler@ist.ac.at
URL: http://www.carolineuhler.com