Existence of the Maximum Likelihood Estimator in Graphical Gaussian Models

Steffen Lauritzen, University of Oxford

Durham Symposium on Mathematical Aspects of Graphical Models

July 8, 2008

$X = (X_v, v \in V) \sim \mathcal{N}_V(0, \Sigma)$ with Σ regular and $K = \Sigma^{-1}$.
Graphical Gaussian Model represented by $\mathcal{G} = (V, E)$, $K \in \mathcal{S}^+(\mathcal{G})$.

$K \in \mathcal{S}^+(\mathcal{G})$ is set of (symmetric) positive definite matrices with

$$k_{\alpha\beta} = 0 \text{ whenever } \alpha \not\sim \beta.$$

How many observations are needed to ensure estimability of K for a given graph $\mathcal{G}$? Equivalently,

for a given sample size, how complex can $\mathcal{G}$ be for K to be estimable?

The log-likelihood function based on a sample of size n is

$$\begin{aligned} \log L(K) &= \frac{n}{2}\log(\det K) - \operatorname{tr}(KW)/2 \\ &= \frac{n}{2}\log(\det K) - \operatorname{tr}\{KW(\mathcal{G})\}/2 \end{aligned}$$

where W is the Wishart matrix of sums of squares and products of the X's and $W(\mathcal{G})$ the *partial matrix* $W(\mathcal{G}) = \{W_c, c \in \mathcal{C}\}$.

$W(\mathcal{G})$ is in the cone of *partially positive semidefinite* (PPS) matrices (W_c all positive semidefinite), denoted $\mathcal{Q}_\mathcal{G}$. The cone of *partially positive definite* (PPD) matrices is denoted $\mathcal{Q}_\mathcal{G}^+$.

If we write the sample as a $|V| \times n$ matrix $\mathbf{X}$ with rows $\mathbf{X}_v, v \in V$ and columns $\mathbf{X}^\nu, \nu = 1, \ldots n$ then $W = \mathbf{X}\mathbf{X}^\top$. Hence $W(\mathcal{G})$ is also in $\mathcal{Q}_\mathcal{G}^e$, the PPS matrices which are also *extendable* to full positive semidefinite matrices (PPSE).

Since the restriction $K \in \mathcal{S}^+(\mathcal{G})$ is *linear* in K, this is the likelihood function of a canonical and linear exponential family with K as the canonical parameter and the partial matrix $W(\mathcal{G})$ as its canonical sufficient statistic.

The exponential family property implies that *the MLE of Σ is the unique element with $K = \Sigma^{-1} \in \mathcal{S}^+(\mathcal{G})$ satisfying*

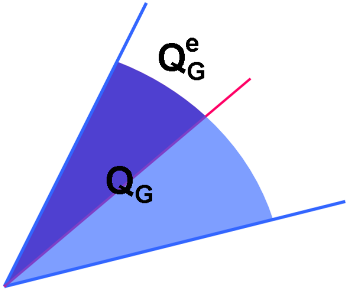
$$n\Sigma(G) = W(\mathcal{G})$$

provided such an element exists.

Standard exponential family theory: *a solution exists if and only if* $W(\mathcal{G})$ *is in the interior* $\mathcal{Q}_{\mathcal{G}}^{e\circ}$ *of the cone* $\mathcal{Q}_{\mathcal{G}}^{e}$ of extendable PPS matrices, which are those which are extendable to PPD matrices.

If $n \geq |V|$, $\mathrm{rank}(\mathbf{X}) = \mathrm{rank}(W) = |V|$ *with probability 1, so* W *is in* $\mathcal{Q}_{\mathcal{G}}^{+}$*, implying that* $W(\mathcal{G})$ *is in* $\mathcal{Q}_{\mathcal{G}}^{e\circ}$.

What happens if $n << |V|$*?*



The cones of extendable and non-extendable PPD matrices.

Matrix completion (Paulsen et al. 1989) is concerned with the question of equality between $\mathcal{Q}^e_{\mathcal{G}}$ and $\mathcal{Q}_{\mathcal{G}}$.

It always holds that

$$\mathcal{Q}^{e\circ}_{\mathcal{G}} \subseteq \mathcal{Q}^{+}_{\mathcal{G}}.$$

It holds that

$$\mathcal{Q}^{e\circ}_{\mathcal{G}} = \mathcal{Q}^{+}_{\mathcal{G}}$$

if and only if $\mathcal{G}$ is chordal.

It holds that

$$\mathcal{Q}^{e}_{\mathcal{G}} = \mathcal{Q}_{\mathcal{G}}$$

if and only if $\mathcal{G}$ is chordal.

All standard and well-known in a number of contexts.

A non-extendable PPD matrix

For the chordless four-cycle, the matrix below is in $\mathcal{Q}^+_{\mathcal{G}} \setminus \mathcal{Q}^{\text{eo}}_{\mathcal{G}}$ if $|\rho|$ is sufficiently large ($\rho \geq 1/2$):

$$K = \begin{pmatrix} 1 & \rho & * & -\rho \\ \rho & 1 & \rho & * \\ * & \rho & 1 & \rho \\ -\rho & * & \rho & 1 \end{pmatrix}.$$

If there is a strong positive correlation ρ between the pairs of variables (X_1, X_2), (X_2, X_3), and (X_3, X_4), then X_1 and X_4 cannot possibly be strongly negatively correlated.

Very limited results are available on the non-chordal case other than counterexamples such as above.

The MLE exists if and only if $W(\mathcal{G}) \in \mathcal{Q}_{\mathcal{G}}^{eo}$. When is this the case?

If $\mathcal{G}$ chordal, we have $\mathcal{Q}_{\mathcal{G}}^{eo} = \mathcal{Q}_{\mathcal{G}}^{+}$ and hence we just have to ensure that $W(\mathcal{G})$ is PPD.

Thus, in the chordal case MLE exists with probability one if

$$n \geq \max_{C \in \mathcal{C}(\mathcal{G})} |C|$$

and it does not exist if

$$n < \max_{C \in \mathcal{C}(\mathcal{G})} |C|.$$

If the MLE exists for a given graph $\mathcal{G}$, it clearly also exists for any subgraph obtained by removing edges.

So if there is a *chordal cover*, i.e. a graph $\mathcal{G}^* = (V, E^*)$ with $E \subseteq E^*$, and $n \geq \max_{C \in \mathcal{C}(\mathcal{G}^*)} |C|$, the MLE also exists in $\mathcal{G}$.

The *treewidth* $\tau(\mathcal{G})$ of a graph is one less than the smallest maximal clique in a chordal cover as above, i.e.

$$\tau(\mathcal{G}) = \min_{\mathcal{G}^*:\mathcal{G}^* \text{ chordal cover of } \mathcal{G}} \max_{C \in \mathcal{C}(\mathcal{G}^*)} |C| - 1.$$

Thus the *treewidth of a tree is 1.* A chordal graph $\mathcal{G}$ has treewidth is $\tau(\mathcal{G}) = \max_{C \in \mathcal{C}(\mathcal{G})} |C| - 1$.

The *treewidth of the $d \times d$ lattice is d*.

Rephrasing previous remarks we get for a general case that

If $n > \tau(\mathcal{G})$, the MLE exists with probability 1.

Finding the treewidth of a graph is NP-complete, but *deciding for fixed n whether $n > \tau(\mathcal{G})$ is linear* in $|V|$.

And since $\mathcal{Q}_{\mathcal{G}}^{\mathrm{eo}} \subseteq \mathcal{Q}_{\mathcal{G}}^{+}$, it follows that if $W(\mathcal{G})$ is only PPS, the MLE does not exist, i.e.
If $n < \max_{C \in \mathcal{C}(\mathcal{G})} |C|$, the MLE does not exist.

What happens in the gap, i.e. when $\max_{C \in \mathcal{C}(\mathcal{G})} |C| \leq n \leq \tau(\mathcal{G})$?

Only results that I know of are given in Buhl (1993).

Example: the four-cycle has treewidth 2, so if $n > 2$, the MLE exists. If $n = 1$ it does not exist. Buhl (1993) shows that *if $n = 2$, the MLE exists with a probability which is strictly between 0 and 1.*

The above result is easily modified to the p-cycle which has the same treewidth, and can easily be modified to yield full clarity for wheels and, say, the octahedron (Buhl 1993).

The 3×3 lattice has treewidth 3, so MLE exists for $n > 3$ and since the clique size is 2, so $n = 1$ is not enough. But *what happens for $n = 2$ and $n = 3$?* Still open.

We again write the sample as a $|V| \times n$ matrix $\mathbf{X}$ so $W = \mathbf{X}\mathbf{X}^\top$.

The problem of existence/extendability is invariant under rescaling of each X-variable with a constant, i.e. we can pre- and post-multiply W with a diagonal matrix A:

$$\mathbf{X} \to A\mathbf{X}, \text{ or } W \to AWA, \text{ where } A \text{ is diagonal,}$$

expressed both in X-space and in W-space, implying that the problem naturally lives in $\mathbb{RP}^{n-1}$, the *$n-1$-dimensional real projective space.*

Similarly, in X-space, we can post-multiply $\mathbf{X}$ with an orthogonal matrix U since

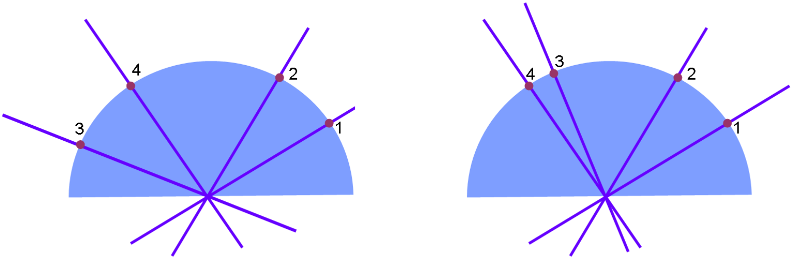
$$\mathbf{X} \to \mathbf{X}U, \text{ or } W \to \mathbf{X}UU^\top\mathbf{X}^\top = W.$$

The four-cycle

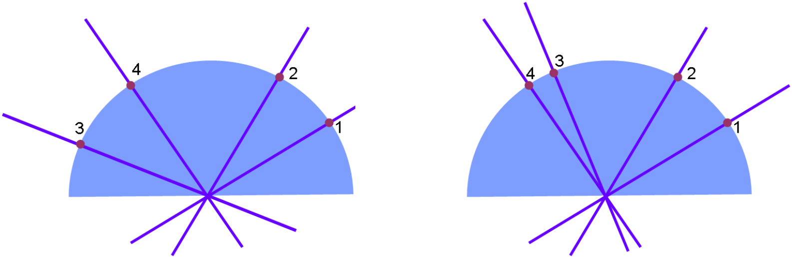
The geometric representation of this particular example for $n = 2$ is illustrative. Then $\mathbf{X}$ is a 4×2 matrix

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{22} \\ x_{21} & x_{22} \\ x_{21} & x_{22} \\ x_{21} & x_{22} \end{pmatrix}.$$

Each row of X generates a line in $\mathbf{R}^2$ through the origin, i.e. a point in $\mathbf{RP}^1$. The question of existence is determined be the relative position of these lines.

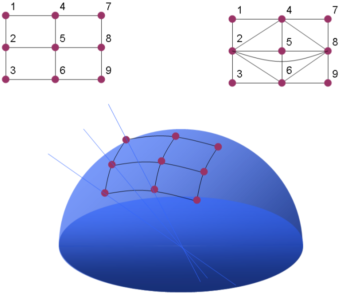


Observations are angles $\cos(\theta_{uv}) = x_u x_v / \sqrt{x_u^2 + x_v^2}$ between neighbours $u \sim v$ in graph.



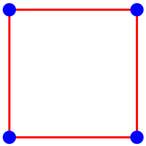
MLE exists in situation to the left, but *it does not exist in the situation to the right* Buhl (1993).

3×3 lattice for $n = 3$



$n = 4$ observations is sufficient. *What is the condition on the angles between graph neighbours for the existence of 9 vectors in higher dimension with same angles?*

Less observations are needed when symmetry is imposed. How much does this help?



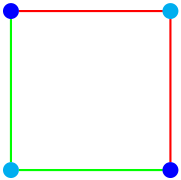
$n = 1$ is sufficient for existence of the MLE! In Højsgaard and Lauritzen (2008) but also classic as it is a circular autoregression of order 1.

$$\hat{\sigma}_{11} = \hat{\sigma}_{22} = \hat{\sigma}_{33} = \hat{\sigma}_{44} = (x_1^2 + x_2^2 + x_3^2 + x_4^2)/4,$$

$$\hat{\sigma}_{12} = \hat{\sigma}_{23} = \hat{\sigma}_{34} = \hat{\sigma}_{41} = (x_1x_2 + x_2x_3 + x_3x_4 + x_4x_1)/4,$$

$$\hat{\sigma}_{13} = \hat{\sigma}_{24} = (\sqrt{1 + 8r^2} - 1)/2,$$

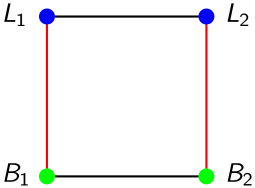
where $r = (x_1x_2 + x_2y_3 + x_3x_4 + x_4x_1)/(x_1^2 + x_2^2 + x_3^2 + x_4^2)$.



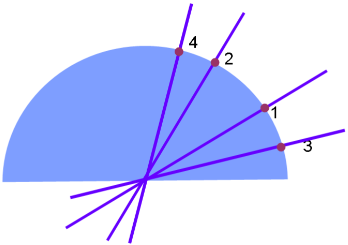
Both RCON and RCOP but not generated by permutation symmetry: Not clear what the condition is for existence.

Frets' heads

Symmetry between the two sons. *RCOP model* as determined by permutation of variable labels and illustrated in figure below



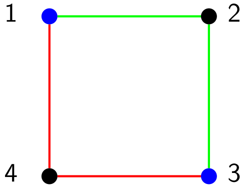
$n = 1$ is sufficient for existence of the MLE!



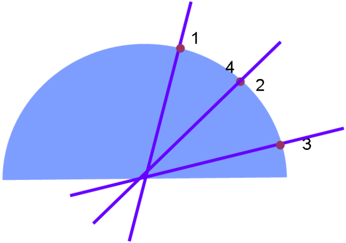
$$\mathbf{X} = \begin{pmatrix} l_1 & l_2 \\ l_2 & l_1 \\ b_2 & b_1 \\ b_1 & b_2 \end{pmatrix}$$

Use the group and the geometry!

Interchanging 1 and 3



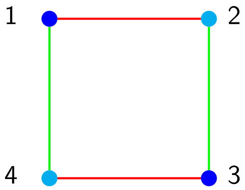
$n = 1$ is sufficient for existence of the MLE!



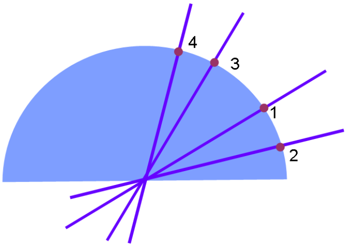
$$\mathbf{X} = \begin{pmatrix} x_1 & x_3 \\ x_2 & x_2 \\ x_3 & x_1 \\ x_4 & x_4 \end{pmatrix}$$

Use the group and the geometry!

Simultaneously interchanging 1 with 3 and 2 with 4



$n = 1$ is sufficient for existence of the MLE!



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

Use the group and the geometry!

Buhl, S. L.: 1993, On the existence of maximum likelihood estimators for graphical Gaussian models, *Scandinavian Journal of Statistics* **20**, 263–270.

Højsgaard, S. and Lauritzen, S. L.: 2008, Graphical Gaussian models with edge and vertex symmetries, *Journal of the Royal Statistical Society, Series B* **68**, in press.

Paulsen, V. I., Power, S. C. and Smith, R. R.: 1989, Schur products and matrix completions, *Journal of Functional Analysis* **85**, 151–178.