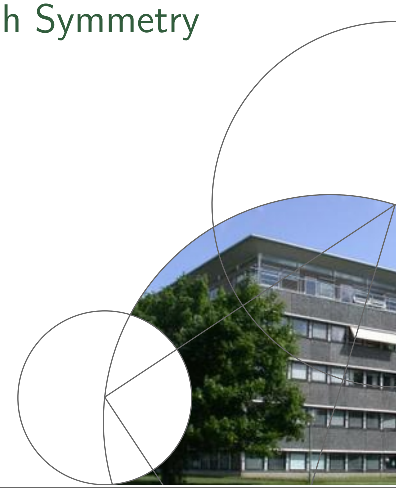
Faculty of Science

# Gaussian Graphical Models with Symmetry

Steffen Lauritzen
Department of Mathematical Sciences







Introduce *symmetry to obtain further parsimony* so models can be well estimated when number of variables $|V|$ higher than number of observed units $n$, $n << |V|$.

Also, sometimes there are *natural and inherent symmetries* in problems under study, e.g. when these involve twins, measurements on right and left sides, dimensions of a starfish, etc.





- Models with symmetry in covariance are classical and admit unified theory (Wilks, 1946; Votaw, 1948; Olkin and Press, 1969; Andersson, 1975; Andersson et al., 1983);
- Stationary autoregressions (circular) (Anderson, 1942; Leipnik, 1947);
- Spatial Markov models (Whittle, 1954; Besag, 1974; Besag and Moran, 1975);





General combinations with conditional independence are more recent:

(Hylleberg et al., 1993; Andersson and Madsen, 1998; Madsen, 2000; Drton and Richardson, 2008; Højsgaard and Lauritzen, 2008; Gehrmann, 2011b; Gottard et al., 2011; Gehrmann, 2011a; Gehrmann and Lauritzen, 2012).

Although *literarure is steadily growing.*

## Gaussian graphical models with symmetry

Several possible types of restriction:

- RCON restricts concentration matrix;
- RCOR restricts partial correlations;
- RCOV restricts covariances
- RCOP has restrictions generated by permutation symmetry.





Empirical concentration matrix of examination marks of 88 students in 5 mathematical subjects.

| | Mechanics | Vectors | Algebra | Analysis | Statistics |
|---|---|---|---|---|---|
| Mechanics | 5.24 | −2.44 | −2.74 | 0.01 | −0.14 |
| Vectors | −2.44 | 10.43 | −4.71 | −0.79 | −0.17 |
| Algebra | −2.74 | −4.71 | 26.95 | −7.05 | −4.70 |
| Analysis | 0.01 | −0.79 | −7.05 | 9.88 | −2.02 |
| Statistics | −0.14 | −0.17 | −4.70 | −2.02 | 6.45 |

.

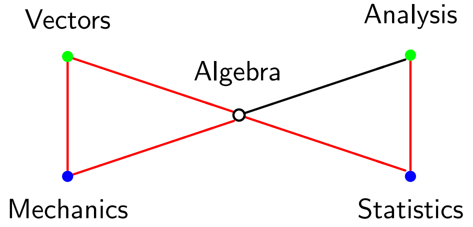
Data reported in Mardia et al. (1979)





## RCON model: Mathematics Marks

Data support model with symmetry restrictions as in figure:




Elements of concentration matrix corresponding to same colours are identical.
*Black or white neutral and corresponding parameters vary freely.*
*RCON model* since restrictions apply to *concentration* matrix





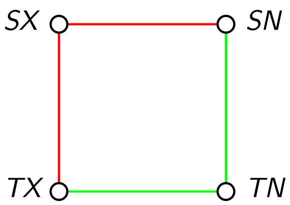
Cox and Wermuth (1993) report data on personality characteristics on 684 students:

Table below shows empirical concentrations ($\times 100$) (on and above diagonal), partial correlations (below diagonal), and standard deviations for personality characteristics of 684 students.

| | *SX* | *SN* | *TX* | *TN* |
|---|---|---|---|---|
| *SX* (State anxiety) | 0.58 | −0.30 | −0.23 | 0.02 |
| *SN* (State anger) | 0.45 | 0.79 | −0.02 | −0.15 |
| *TX* (Trait anxiety) | 0.47 | 0.03 | 0.41 | −0.11 |
| *TN* (Trait anger) | −0.04 | 0.33 | 0.32 | 0.27 |
| Standard deviations | 6.10 | 6.70 | 5.68 | 6.57 |

## RCOR model

Data strongly support conditional independence model displayed below with *partial correlations* strikingly similar in pairs:



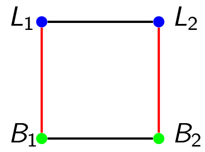

Scales for individual variables may not be compatible. *Partial correlations invariant under changes of scale*, and more meaningful.
Such symmetry models are denoted *RCOR models.*



UNIVERSITY OF COPENHAGEN DEPARTMENT OF MATHEMATICAL SCIENCES

## RCOP model: Frets' heads.

Data from Frets (1921). Length and breadth of the heads of 25 pairs of first and second sons. Data support the model




*Distribution unchanged if sons are switched. RCOP model* as determined by permutation of labels.

*Both RCON, RCOV, and RCOR* because all aspects the joint distribution are unaltered when labels are switched.



UNIVERSITY OF COPENHAGEN DEPARTMENT OF MATHEMATICAL SCIENCES

Undirected graph $\mathcal{G} = (V, E)$.

Colouring vertices of $\mathcal{G}$ with different colours induces partitioning of $V$ into *vertex colour classes.*

Colouring edges $E$ partitions $E$ into disjoint *edge colour classes*

$$V = V_1 \cup \cdots \cup V_T, \quad E = E_1 \cup \cdots \cup E_S.$$

$\mathcal{V} = \{V_1, \ldots, V_T\}$ is a *vertex colouring,*

$\mathcal{E} = \{E_1, \ldots, E_S\}$ is an *edge colouring,*

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a *coloured graph.*

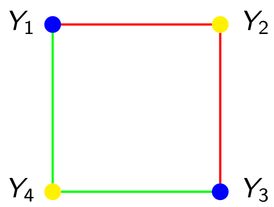


UNIVERSITY OF COPENHAGEN DEPARTMENT OF MATHEMATICAL SCIENCES

## RCON model

1. *Diagonal elements $K$ corresponding to vertices in the same vertex colour class must be identical.*
2. *Off–diagonal entries of $K$ corresponding to edges in the same edge colour class must be identical.*

The set of positive definite matrices which satisfy these restrictions is denoted $\mathcal{S}^+(\mathcal{V}, \mathcal{E}) = \mathcal{S}^+(\mathcal{G})$.




Corresponding RCON model will have concentration matrix

$$K = \begin{pmatrix} k_{11} & k_{12} & 0 & k_{14} \\ k_{21} & k_{22} & k_{23} & 0 \\ 0 & k_{32} & k_{33} & k_{34} \\ k_{41} & 0 & k_{43} & k_{44} \end{pmatrix}.$$





## Likelihood function

Consider a sample $Y^1 = y^1, \ldots, Y^n = y^n$ of $n$ observations of $Y$ and let $W$ denote the matrix of sums of squares and products

$$W = \sum_{\nu=1}^{n} Y^{\nu}(Y^{\nu})^{*}.$$

The log-likelihood function based on the sample is

$$\log L = \frac{n}{2} \log \det(K) - \frac{1}{2} \operatorname{tr}(KW) \qquad (1)$$

Note that the restrictions defined are *linear in the concentration matrix $K$* so *RCON model is linear exponential model.*





## Likelihood equations

For each vertex colour class $u \in \mathcal{V}$ let $T^u$ be the $|V| \times |V|$ diagonal matrix with entries $T^u_{\alpha\alpha} = 1$ if $\alpha \in u$ and 0 otherwise.

Similarly, for each edge colour class $u \in \mathcal{E}$ let $T^u$ have entries $T^u_{\alpha\beta} = 1$ if $\{\alpha, \beta\} \in u$ and 0 otherwise, i.e. the *adjacency matrix* of $u$, e.g.

$$T^{blue} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; T^{red} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Again, $T^u, u \in \mathcal{V} \cup \mathcal{E}$ form a basis for $S(\mathcal{V}, \mathcal{E})$.





Likelihood function then becomes

$$\log L(K) = \frac{n}{2} \log(\det K) - \sum_{u \in \mathcal{V} \cup \mathcal{E}} k_u \operatorname{tr}\{T^u W\}/2.$$

MLE is obtained by *equating canonical sufficient statistics to their expectation,* i.e.

$$\operatorname{tr}(T^u W) = n \operatorname{tr}(T^u K^{-1}), \quad u \in \mathcal{V} \cup \mathcal{E}, \qquad (2)$$

*provided such a solution exists.*

The entries of the information matrix are

$$I(\hat{\theta})_{uv} = f\,\mathrm{tr}(K^u\hat{\Sigma}K^v\hat{\Sigma})/2. \tag{3}$$

The likelihood equations can thus be solved by Newton iteration, provided appropriate starting values can be found.

Alternatively, Jensen et al. (1991) described a globally convergent algorithm, iterating one parameter at a time, using Newton's method on the $f$th root of the *reciprocal* likelihood function; in this instance yielding the iterative step

$$\theta_u \leftarrow \theta_u + \frac{\Delta_u}{\mathrm{tr}(K^u\hat{\Sigma}K^u\hat{\Sigma}) + \Delta_u^2/2} \tag{4}$$

with $\Delta_u = \mathrm{tr}(K^u W) - f\,\mathrm{tr}(K^u\Sigma)$.





Fitted concentrations ($\times 1000$) for examination marks assuming the RCON model displayed.

| | Mechanics | Vectors | Algebra | Analysis | Statistics |
|---|---|---|---|---|---|
| Mechanics | 6.30 | −3.38 | −3.38 | 0 | 0 |
| Vectors | −3.38 | 10.29 | −3.38 | 0 | 0 |
| Algebra | −3.38 | −3.38 | 24.21 | −6.65 | −3.38 |
| Analysis | 0 | 0 | −6.65 | 10.29 | −3.38 |
| Statistics | 0 | 0 | −3.38 | −3.38 | 6.30 |

The model displayed earlier yields an excellent fit with a likelihood ratio of $-2\log \mathrm{LR} = 7.2$ on 7 degrees of freedom, when compared to the butterfly model without symmetry.





# RCOR models

1. Diagonal elements of $K$ corresponding to vertices in same vertex colour class must be identical.
2. *partial correlations* along edges in the same edge colour class must be identical.

The set of positive definite matrices which satisfy the restrictions of an $RCOR(\mathcal{V},\mathcal{E})$ model is denoted $\mathcal{R}^+(\mathcal{V},\mathcal{E}) = \mathcal{R}^+(\mathcal{G})$.





Define $A$ as diagonal matrix with

$$a_\alpha = \sqrt{k_{\alpha\alpha}},\ \alpha \in u \in \mathcal{V}$$

We can uniquely represent $K \in \mathcal{R}^+(\mathcal{V},\mathcal{E})$ as

$$K = ACA,$$

where $C$ has all diagonal entries equal to one and off-diagonal entries are negative partial correlations

$$c_{\alpha\beta} = -\rho_{\alpha\beta\,|\,V\setminus\{\alpha,\beta\}} = k_{\alpha\beta}/\sqrt{k_{\alpha\alpha}k_{\beta\beta}} = k_{\alpha\beta}/(a_\alpha a_\beta).$$

*Vertex colour classes restrict $A$, whereas edge colour classes restrict $C$.*

## Likelihood equations

Although restrictions linear in each of $A$ and $C$, they are in general not linear in $K$.

*RCOR models are curved exponential families.*

The likelihood function becomes

$$\log L = \frac{n}{2}\log\det\{C\} + n\sum_{u\in\mathcal{V}}\log a_u \operatorname{tr}(T^u) - \frac{1}{2}\operatorname{tr}\{CAWA\}$$

*$\log L$ concave in $A$ for fixed $C$ and vice versa*, but *not in general jointly.*





Differentiation yields the likelihood equations

$$\operatorname{tr}(T^u AWA) = n\operatorname{tr}(T^u C^{-1}), u \in \mathcal{E};\ \operatorname{tr}(T^u ACAW) = n\operatorname{tr}(T^u), u \in \mathcal{V}.$$

*MLE is not necessarily unique.*

If the MLE exists uniquely, *an alternating algorithm converges to the MLE,* alternating between maximizing in $A$ for fixed $C$ and conversely.





## Anxiety and anger

Fitted concentrations ($\times 100$) (on and above diagonal) and partial correlations (below diagonal) for RCOR model:

| | $SX$ | $SN$ | $TX$ | $TN$ |
|---|---|---|---|---|
| $SX$ (State anxiety) | 0.59 | $-0.31$ | $-0.22$ | 0 |
| $SN$ (State anger) | 0.46 | 0.78 | 0 | -0.15 |
| $TX$ (Trait anxiety) | 0.46 | 0 | 0.40 | -0.10 |
| $TN$ (Trait anger) | 0 | 0.31 | 0.31 | 0.28 |

Fitting the RCOR model yields likelihood ratio $-2\log\mathrm{LR} = 0.22$ on 2 d.o.f. comparing with the model without symmetry.





Let $G$ be permutation matrix for elements of $V$. If $Y \sim \mathcal{N}_{|V|}(0, K)$ then $GY \sim \mathcal{N}_{|V|}(0, GKG^*)$.

Let $\Gamma \subseteq S(V)$ be a subgroup of such permutations.

Distribution of $Y$ *invariant under the action of $\Gamma$* if and only if

$$GKG^* = K \text{ for all } G \in \Gamma. \tag{5}$$

Since $G$ satisfies $G^{-1} = G^*$, (5) is equivalent to

$$GK = KG \text{ for all } G \in \Gamma, \tag{6}$$

i.e. that $G$ commutes with $K$.

An *RCOP model generated by* $\Gamma \subseteq \mathrm{Aut}(\mathcal{G})$ is given by assuming

$$K \in \mathcal{S}^+(\mathcal{G}, \Gamma) = \mathcal{S}^+(\mathcal{G}) \cap \mathcal{S}^+(\Gamma)$$

where $\mathcal{S}^+(\Gamma)$ is the set of positive definite matrices satisfying

$$GK = KG \text{ for all } G \in \Gamma.$$





An RCOP model can also be represented by a graph colouring:

If $\mathcal{V}$ denotes the *vertex orbits* of $\Gamma$, i.e. the equivalence classes of

$$\alpha \equiv_\Gamma \beta \iff \beta = G(\alpha) \text{ for some } G \in \Gamma,$$

and similarly $\mathcal{E}$ the *edge orbits*, i.e. the equivalence classes of

$$\{\alpha, \gamma\} \equiv_\Gamma \{\beta, \delta\} \iff \{\beta, \delta\} = \{G(\alpha), G(\gamma)\} \text{ for some } G \in \Gamma,$$

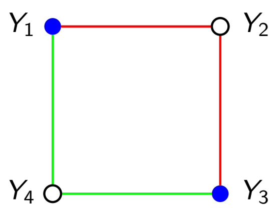
then we have

$$\mathcal{S}^+(\mathcal{G}, \Gamma) = \mathcal{S}^+(\mathcal{V}, \mathcal{E}) = \mathcal{R}^+(\mathcal{V}, \mathcal{E}).$$

*Hence an RCOP model can also be represented as an RCON or an RCOR model* with vertex orbits as vertex colour classes and edge orbits as edge colour classes.



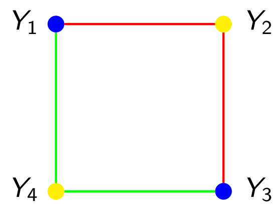





This colouring is *generated by permutations* (13), so model is RCOP.








This graph is *regular* (Siemons, 1983), but symmetry is *not generated by permutations*.

## Likelihood equations

Representing an RCOP model as an RCON model yields the likelihood equations

$$\operatorname{tr}(T^u W) = \operatorname{tr}(T^u K^{-1}), u \in \mathcal{V} \cup \mathcal{E}; \quad \Sigma^{-1} \in \mathcal{S}^+(\mathcal{V}, \mathcal{E}). \quad (7)$$

However, *for RCOP models these equations are equivalent to*

$$\operatorname{tr}(T^\ell \overline{W}) = \operatorname{tr}(T^\ell K^{-1}), \ell \in V \cup E; \quad K \in \mathcal{S}^+(\mathcal{G}), \qquad (8)$$

where

$$\overline{W} = \frac{1}{|\Gamma|} \sum_{G \in \Gamma} GWG^*.$$

Hence, *RCOP models can be fitted by Iterative Proportional Scaling*, replacing $W$ with $\overline{W}$.





## Frets' heads

Observed concentrations ($\times 100$) (on and above diagonal) together with fitted concentrations for RCOP model.

| | $L1$ | $B1$ | $L2$ | $B2$ |
|---|---|---|---|---|
| $L1$ (Length of head of first son) | 3.21 | −1.16 | −0.78 | −1.11 |
| $B1$ (Breadth of head of first son) | −1.71 | 2.21 | −0.50 | 0.48 |
| $L2$ (Length of head of second son) | −1.42 | 0 | 2.67 | −1.89 |
| $B2$ (Breadth of head of second son) | 0 | −1.83 | −1.71 | 3.37 |
| Fitted concentrations | 2.89 | 2.44 | 2.89 | 2.44 |

The likelihood ratio comparing to model without symmetries is equal to $-2 \log \mathrm{LR} = 5.18$ on 5 degrees of freedom.



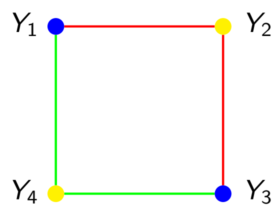


The additional symmetry reduces the number of observations necessary for existence of the MLE.

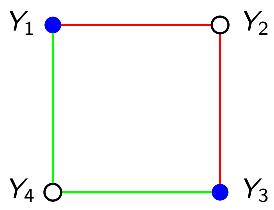
Uhler (2012) uses computational algebraic geometry to investigate all edge regular graphs with four vertices but has *no complete result nor a general principle for calculating the number of necessary observations*.








Symmetry is not generated by permutations, but $n = 1$ *observations is sufficient for existence of the MLE.*




This colouring is generated by permutations (13) but *$n = 2$ observations are necessary to ensure existence of the MLE.*

*For all other RCOP models on this graph, $n = 1$ observation is sufficient.*





Game between *Forecaster* and *Nature*:

Forecaster quotes probability distribution $Q$ for a random quantity $X$. Then Nature reveals $X = x$.

How well did Forecaster do? A *score* is calculated $S(x, Q)$ representing a loss to Forecaster. The function $S(x, Q)$ is a *scoring rule* (Good, 1952; McCarthy, 1956).

A common example of such a scoring rule is the *logarithmic score*

$$S(x, Q) = -\log q(x)$$

where $q(x)$ is the density of $Q$ w.r.t. some fixed measure on $\mathcal{X}$.





We can extend the definition of a scoring rule to $S(P, Q)$ for any probability distribution $P$ as

$$S(P, Q) = \mathbb{E}_{X\sim P}\{S(X, Q)\} = \int S(x, Q)\, P(dx)$$

and further, using the right-hand expression, to $S(\mu, Q)$ for any positive and finite measure. Then $S$ is linear in the first argument.

A scoring rule is *proper* if it encourages honesty, i.e. if the loss is minimized for $Q = P$, i.e. if

$$S(P, P) = \inf_Q S(P, Q).$$

It is *strictly proper* if the minimum is unique.





The *logarithmic score is strictly proper.* Other examples of strictly proper scoring rules include for $\mathcal{X}$ being finite the *Brier score*

$$S(x, Q) = ||q||_2^2 - 2q(x),$$

where $q$ is the pmf of $Q$ and $||q||_2^2 = \sum_x q(x)^2$, and the *spherical score*

$$S(x, Q) = -q(x)/||q||_2.$$

Also, for $\mathcal{X} = \mathbb{R}$, the *Bregman scores* are strictly proper

$$S(x, Q) = \phi'\{q(x)\} + \int \left[\phi\{q(y)\} - q(y)\phi'\{q(y)\}\right] \mu(dy),$$

where $\phi$ is any strictly concave real function.

Every strictly proper scoring rule induces an *entropy function*

$$H(P,P) = S(P,P)$$

and a non-negative *divergence* (Dawid, 1998; Grünwald and Dawid, 2004)

$$D(P,Q) = S(P,Q) - S(P,P) = S(P,Q) - H(P) \geq 0.$$

For the *logarithmic score* we get the *Shannon entropy*

$$H(P) = \mathbb{E}_{X\sim P}\{-\log p(X)\}$$

and the *Kullback–Leibler divergence*

$$D(P,Q)) = \mathbb{E}_{X\sim P}\{-\log q(X) + \log p(X)\} = \mathbb{E}_{X\sim P}\{\log p(X)/q(X)\}.$$





Suppose $\mathcal{X} \subseteq \mathbb{R}^p$ and the density $q = dQ/dx$ of $Q$ satisfies:

$$\mathbb{E}_{X\sim P}\|\nabla \log g(X)\|_p^2 < \infty \text{ for all } P, Q \in \mathcal{P};$$

as well as $g(x) \to 0$ and $\|\nabla g(x)\|_p \to 0$ as $x$ approaches the boundary of $\mathcal{X}$.

Then Hyvärinen (2005) showed that the divergence function

$$D_2(P,Q) = \mathbb{E}_{X\sim P}\|\nabla \log g(x) - \nabla \log f(x)\|_p^2$$

where $f$ is the density of $P$, is induced by the scoring rule

$$S_2(x,Q) = \frac{1}{2}\|\nabla \log q(x)\|_p^2 + \Delta \log q(x).$$

which is *strictly proper* (Dawid and Lauritzen, 2005).





Let $\mathcal{P} = \{Q_\theta, \theta \in \Theta\}$ and $X^1 = x^1, \ldots, X = x^n$ be a sample in $\mathcal{X}$ with empirical distribution $\hat{P}$.

The *score estimator* of $\theta$ is determined as the minimizer

$$\check{\theta} = \arg\min_{\theta\in\Theta} \sum_{i=1}^n S(x^i, Q_\theta) = \arg\min_{\theta\in\Theta} \mathbb{E}_{X\sim\hat{P}}\{S(X,Q)\}.$$

Dawid and Lauritzen (2005) show that *this minimization yields an unbiased estimating equation*

$$\sum_{i=1}^n S'(x^i,\theta) = 0,$$

where $S'(x,\theta)$ is the vector of derivatives of $S(x,Q_\theta)$ w.r.t. $\theta$. Solutions are *M-estimators* (Huber, 1964, 1967) and typically consistent although not efficient.





If $S(x,Q) = -\log q(x)$ is the logarithmic score, the equation is the *likelihood equation* and the score estimator is the *maximum likelihood estimator.*

The *score matching estimator* (Hyvärinen, 2005) is the estimator corresponding to the scoring rule

$$S_2(x,Q) = \frac{1}{2}\|\nabla \log q(x)\|_p^2 + \Delta \log q(x).$$

Note that $S_2(x,Q)$ can be calculated if we only know $q$ up to an unknown proportionality factor.

Hence, if *$q(x\,|\,\theta) = c(\theta)h(x)$, we do not need a simple expression for the constant $c(\theta)$.*

For a Gaussian distribution we have

$$\log q(y) = c - y^\top K y/2$$

and hence

$$\nabla \log q(y) = -Ky$$

$$\Delta \log q(y) = -\operatorname{tr}(K)$$

$$\sum_{i=1}^n S_2(y_i, Q) = \operatorname{tr} K^2 W/2 - n \operatorname{tr}(K).$$





For a Gaussian linear concentration model with $L$ being a $d$-dimensional subspace of $\mathcal{S}^p$ the symmetric $p \times p$ matrices and $I_p \in L$, we then get the estimating equation

$$\Pi_L(K \circ W) = nI_p$$

where $A \circ B = (AB^\top + BA^\top)/2$ is the *Jordan product* (Albert, 1946) of the symmetric matrices $A$ and $B$.

Note that *this equation is linear* in $K$ and thus it has a unique solution if and only if the map $K \to K \circ W$ has trivial kernel. Even when there is a unique solution $\check{K}$, *$\check{K}$ may not be positive semidefinite*.





## Some questions

1. For which $L$ is the map $Q_W : K \to \Pi_L\{(KW + WK)/2\}$ invertible with probability one?
2. If $Q_W$ is invertible, when is $Q_W^{-1}$ positive definite with probability one?
3. If $Q_W$ is invertible, when is $\check{K} = Q_W^{-1}(nI_p)$ positive definite with high probability?





For a basis $T^u, u = 1, \dots d$ for $L$ we have the MLE equations

$$\operatorname{tr}(T^u W) = n \operatorname{tr}(T^u K^{-1}), u = 1, \dots, d$$

whereas the SME equations are

$$\operatorname{tr}(T^u W K) = n \operatorname{tr}(T^u), u = 1, \dots, d.$$

*If the SME exists, then the MLE also exists,* i.e. if $K \to K \circ W$ has trivial kernel the MLE exists, but not conversely (Forbes and Lauritzen, 2014).

Suppose that $L$ is closed under the Jordan product or, equivalently, $\Theta = L \cap \mathcal{S}^p_+$ is closed under inversion (Jensen, 1988). Includes all models determined by group invariance (Andersson, 1975).

For such models the MLE and the SME coincide (Forbes and Lauritzen, 2014). More precisely:

*If the subspace $L$ is a Jordan subalgebra, the score matching estimator is equal to the maximum likelihood estimator and*

$$\hat{K} = \check{K} = \{\Pi_L(W)\}^{-1},$$

*provided $\Pi_L(W)$ is invertible.*





In particular this implies that *the map $\Pi_L : \mathcal{S}^p \to L$ is Löwner positive,* i.e. maps non-negative definite elements into non-negative definite elements.

Indeed it holds for any idempotent linear map $\Pi : \mathcal{S}^p \to L = \mathrm{range}(\Pi)$ that

*$\Pi$ is Löwner positive if and only if $L$ is a Jordan subalgebra of $\mathcal{S}^p$.*

(Effros and Størmer, 1979; Fuglede and Jensen, 2013).





Observing $y = (y^1, \dots, y^n)$, the score matching equation has a unique solution iff the quadratic form

$$D_2(K) = \sum_{i=1}^{n} \|Ky^i\|^2$$

is positive definite on $L$. If $T^1, \dots, T^d$ is an orthogonal basis for $L$, the matrix for this quadratic form is $M(y) = \{m_{uv}(y)\}$ where

$$m_{uv}(y) = \mathrm{tr}(T^u W T^v)$$

and hence *$D_2$ is positive definite if and only if $\det M(y) > 0$.*





This determinant is a polynomial in $y$; hence *either $\det M(y) = 0$ for all $y$ or $\det M(y) > 0$ almost everywhere* (Okamoto, 1973).

Note that $W$ has rank $n$ with probability one in

$$m_{uv}(y) = \mathrm{tr}(T^u W T^v)$$

Say $L$ is *n-estimable* if there is a $y = (y^1, \dots, y^n) \in \mathbb{R}^{p \times n}$ such that $\det M(y) > 0$.

For $n \geq p$, $W$ is positive definite with probability one and hence $M(y)$ is positive definite and any $L$ is $n$-estimable.

Assume $n < p$. Let $\Delta_k = k(k+1)/2$ be the *triangular* numbers.

Then, *if $d = \dim L > \Delta_p - \Delta_{p-n}$, $L$ is not $n$-estimable* (Forbes and Lauritzen, 2014).

*The converse is false:*

$$L = \left\{ \begin{pmatrix} a & c & 0 & f \\ c & b & -f & 0 \\ 0 & -f & a & c \\ f & 0 & c & b \end{pmatrix} : a, b, c, f \in \mathbb{R} \right\},$$

is not 1-estimable although we have $p = 4$ and $d = 4$ and thus

$$\Delta_p - \Delta_{p-n} = \Delta_4 - \Delta_3 = 4 = d.$$

This is an example of a Jordan subalgebra (Jensen, 1988) and — as Jensen — *we conclude that also the MLE fails to exist.*

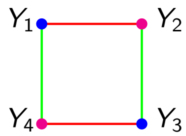




*Modify to get coloured graphical model*

$$L = \left\{ \begin{pmatrix} a & c & 0 & f \\ c & b & f & 0 \\ 0 & f & a & c \\ f & 0 & c & b \end{pmatrix} : a, b, c, f \in \mathbb{R} \right\},$$

*This is 1-estimable* as $\det M(y) = 4y_1y_2y_3y_4$.




This is not a Jordan subalgebra but *we conclude that also the MLE exists.*





Minimum score for the SME is very easy to calculate

$$\sum_{i=1}^{n} S_2(y_i, Q_{\check{K}}) = \operatorname{tr} \check{K}^2 W/2 - n \operatorname{tr}(\check{K}) = -n \operatorname{tr}(\check{K})/2.$$

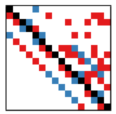
This makes sense even if $\check{K}$ is not positive definite.
So identify concentration graph by minimizing a penalised version of the optimal score:

$$\tilde{S}(\mathcal{G}) = (|V| + |E|)\sqrt{p} \log\log(np)/(2n) - \operatorname{tr}(\check{K}_{\mathcal{G}}).$$

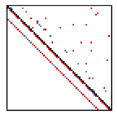
This is *extremely fast*. For example, using this on an $s \times s$ lattice so $p = s^2$ it took for $s = 100$, i.e. $p = 10000$ and $n = 100000$ 10 seconds to identify the lattice structure correctly. Note the concentration matrix is then $10000 \times 10000$, so is rather big...



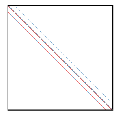




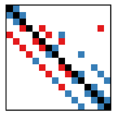
: $p = 16, n = 1 \times p$



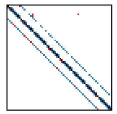
: $p = 64, n = 1 \times p$



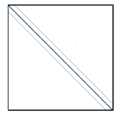
: $p = 256, n = 1 \times p$



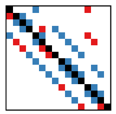
: $p = 16, n = 5 \times p$



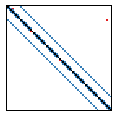
: $p = 64, n = 5 \times p$



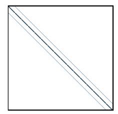
: $p = 256, n = 5 \times p$



: $p = 16, n = 10 \times p$



: $p = 64, n = 10 \times p$



: $p = 256, n = 10 \times p$

Albert, A. A. (1946). On Jordan algebras of linear transformations. *Transactions of the American Mathematical Society*, 59:524–555.

Anderson, R. L. (1942). Distribution of the serial correlation coefficient. *The Annals of Mathematical Statistics*, 13:1–13.

Andersson, S. A. (1975). Invariant normal models. *The Annals of Statistics*, 3:132–154.

Andersson, S. A., Brøns, H., and Jensen, S. T. (1983). Distribution of eigenvalues in multivariate statistical analysis. *The Annals of Statistics*, 11:392–415.

Andersson, S. A. and Madsen, J. (1998). Symmetry and lattice conditional independence in a multivariate normal distribution. *The Annals of Statistics*, 26:525–572.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36:302–309.



UNIVERSITY OF COPENHAGEN DEPARTMENT OF MATHEMATICAL SCIENCES

Besag, J. E. and Moran, P. A. P. (1975). On the estimation and testing of spatial interaction in Gaussian lattice processes. *Biometrika*, 62:555–562.

Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statistical Science*, 8:204–218; 247–277.

Dawid, A. P. (1998). Coherent measures of discrepancy, uncertainty and dependence, with applications to Bayesian predictive experimental design. Technical Report 139, Department of Statistical Science, University College London.

Dawid, A. P. and Lauritzen, S. L. (2005). The geometry of decision theory. In *Proceedings of the Second International Symposium on Information Geometry and its Applications*, pages 22–28. University of Tokyo.

Drton, M. and Richardson, T. S. (2008). Binary models for

Steffen Lauritzen — Gaussian Graphical Models with Symmetry — Swiss Winterschool 2015, Lecture 3
Slide 52/52

UNIVERSITY OF COPENHAGEN DEPARTMENT OF MATHEMATICAL SCIENCES

marginal independence. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 41(2):287–309.

Effros, E. D. and Størmer, E. (1979). Positive projections and Jordan structure in operator algebras. *Mathematica Scandinavia*, 45:127–138.

Forbes, P. G. M. and Lauritzen, S. (2014). Linear estimating equations for exponential families with application to Gaussian linear concentration models. *Linear Algebra and its Applications*. published online.

Frets, G. P. (1921). Heredity of head form in man. *Genetica*, 3:193–400.

Fuglede, B. and Jensen, S. T. (2013). Positive projections of symmetric matrices and Jordan algebras. *Expositiones Mathematicae*, 31:295–303.

Gehrmann, H. (2011a). *Graphical Gaussian Models with Symmetries*. PhD thesis, Department of Statistics, University of Oxford.

Steffen Lauritzen — Gaussian Graphical Models with Symmetry — Swiss Winterschool 2015, Lecture 3
Slide 52/52

UNIVERSITY OF COPENHAGEN DEPARTMENT OF MATHEMATICAL SCIENCES

Gehrmann, H. (2011b). Lattices of graphical Gaussian models with symmetries. *Symmetry*, 3(3):653–679.

Gehrmann, H. and Lauritzen, S. L. (2012). Estimation of means in graphical Gaussian models with symmetries. *The Annals of Statistics*, 40. To appear.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B*, 14:107–114.

Gottard, A., Marchetti, G. M., and Agresti, A. (2011). Quasi-symmetric graphical log-linear models. *Scandinavian Journal of Statistics*, 38(3):447–465.

Grünwald, P. D. and Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Annals of Statistics*, 32:1367–1433.

Højsgaard, S. and Lauritzen, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society, Series B*, 70:1005–1027.

Steffen Lauritzen — Gaussian Graphical Models with Symmetry — Swiss Winterschool 2015, Lecture 3
Slide 52/52

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Applied Statistics*, 35(1):73–101.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In Cam, L. M. L. and Neyman, J., editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 221–223, Berkeley, CA. University of California Press.

Hylleberg, B., Jensen, M., and Ørnbøl, E. (1993). Graphical symmetry models. Master's thesis, Aalborg University, Aalborg, Denmark.

Hyvärinen, A. (2005). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6:695–709.

Jensen, S. T. (1988). Covariance hypotheses which are linear in both the covariance and the inverse covariance. *The Annals of Statistics*, 116:302–322.





Jensen, S. T., Johansen, S., and Lauritzen, S. L. (1991). Globally convergent algorithms for maximizing a likelihood function. *Biometrika*, 78:867–877.

Leipnik, R. B. (1947). Distribution of the serial correlation coefficent in a circularly correlated universe. *The Annals of Mathematical Statistics*, 18:80–87.

Madsen, J. (2000). Invariant normal models with recursive graphical Markov structure. *The Annals of Statistics*, 28(4):1150–1178.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press, London.

McCarthy, J. (1956). Measures of the value of information. *Proceedings of the National Academy of Sciences*, 42:654–655.

Okamoto, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *The Annals of Statistics*, 1:763–765.





Olkin, I. and Press, S. J. (1969). Testing and estimation for a circular stationary model. *The Annals of Mathematical Statistics*, 40:1358–1373.

Siemons, J. (1983). Automorphism groups of graphs. *Arch. Math. (Basel)*, 41:379–384.

Uhler, C. (2012). Geometry of maximum likelihood estimation in Gaussian graphical models. *Annals of Statistics*, 40:238–261.

Votaw, D. F. (1948). Testing compound symmetry in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 19:447–473.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41:439–449.

Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 17:257–281.