Journal of Multivariate Analysis 106 (2012) 212–228

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Covariance selection and multivariate dependence

Bhaskar Bhattacharya

Department of Mathematics, Southern Illinois University, Carbondale, IL 62901-4408, USA

ARTICLE INFO

Article history:
Received 30 June 2010
Available online 18 November 2011

AMS subject classifications:
62H12
62F30
90C25
60E15

Keywords:
Association
Convergence
Covariance structure analysis
Empirical entropy
Fenchel duality
I-projection
Longitudinal data
MTP_2
Multivariate normal distribution
Nonnormal

ABSTRACT

Considering the covariance selection problem of multivariate normal distributions, we show that its Fenchel dual formulation is insightful and allows one to calculate direct estimates under decomposable models. We next generalize the covariance selection to multivariate dependence, which includes MTP_2 and trends in longitudinal studies as special cases. The iterative proportional scaling algorithm, used for estimation in covariance selection problems, may *not* lead to the correct solution under such dependence. Addressing this situation, we present a new algorithm for dependence models and show that it converges correctly using tools from Fenchel duality. We discuss the speed of convergence of the new algorithm. When normality does not hold, we show how to estimate the covariance matrix in an empirical entropy approach. The approaches are compared via simulation and it is shown that the estimators developed here compare favorably with existing ones. The methodology is applied on a real data set involving decreasing CD4+ cell numbers from an AIDS study.

1. Introduction and preliminaries

We consider a $p \times 1$ random vector $\boldsymbol{X}$ having a normal distribution with mean $\mathbf{0}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. The probability density function (pdf) of $\boldsymbol{X}$ is given by

$$
\begin{aligned}
f(\boldsymbol{x}) &= \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}} \\
&= \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma^{ij}x_i x_j}, \qquad (1.1)
\end{aligned}
$$

for $\boldsymbol{x} \in \Re^p$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$, $\boldsymbol{\Sigma}^{-1} = (\sigma^{ij})$. Matrices and vectors are written with bold letters, and probability measures (PMs) are written with capital but not bold letters (P, Q, etc.). The convention of expressing an element of a matrix with subscripts and the elements of its inverse with superscripts will be followed throughout.

The covariance selection model [8] selects some of the σ^{ij} to be equal to zero, which corresponds to the conditional independence of X_i and X_j given X_k, $k \neq i, j$. Given a random sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ from (1.1), the covariance selection problem is to find a maximum likelihood estimate (MLE) of $\boldsymbol{\Sigma}$ subject to the restriction that $\sigma^{ij} = 0$, for some pairs $(i, j), i < j$. Dempster [8] showed that the likelihood equations for this model are

$$
(1)\ \sigma_{ij} = s_{ij}, \quad \forall (i,j) \in \mathcal{F} \cup \{(i,i), \forall i\}, \qquad (2)\ \sigma^{ij} = 0, \quad \forall (i,j) \in \overline{\mathcal{F} \cup \{(i,i), \forall i\}}, \qquad (1.2)
$$

E-mail address: bhaskar@siu.edu.

0047-259X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2011.11.002

where $\boldsymbol{S} = (s_{ij}) = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'/n$ and $\mathcal{F} \subset \mathcal{M} = \{(i,j) : i < j, 1 \leq i,j \leq p\}$. To find the estimates of $\boldsymbol{\Sigma}$ subject to (1.2), Wermuth and Scheidt [26] presented an algorithm with one constraint at a time. Using cliques (one or more edge/constraint) of an undirected graph, Speed and Kiiveri [25] presented two algorithms to solve the same problem. These latter algorithms are generalizations of the iterative proportional scaling (IPS) algorithm [16] for the contingency tables. Chaudhuri et al. [5] have considered estimation of a covariance matrix under the constraint that certain covariances are zero. They presented an iterative conditional fitting algorithm for computing the MLE of the constrained covariance matrix.

In this paper, we first formulate the covariance selection problem as an I-projection problem (defined in (1.5)) subject to (1) of (1.2) in a Fenchel duality framework [19]. The original problem of finding the solution (I-projection) pdf subject to the above constraints is called the *primal* problem. We show that the related (Fenchel) *dual* problem amounts to solving for *scalars* only, and hence is substantially easier to deal with. In particular, we show that the scalar solutions of the dual problems from different constraints *are* the elements of the inverse of the covariance matrix of the solution for the primal problem. Solving the maximum likelihood estimation problem in previous paragraph with an I-projection is not new (e.g., [25]), but the duality approach of this paper to the same problem *is*.

Solving the covariance selection problem also amounts to finding the normal pdf with prescribed marginals. For this purpose, we present a variant of the IPS algorithm using dual solutions (Algorithm 2.4) obtained from solving one constraint at a time. At every step of this algorithm, the solution is a multivariate normal pdf with an updated $\boldsymbol{\Sigma}^{-1}$. After convergence, one can find $\boldsymbol{\Sigma}$ from $\boldsymbol{\Sigma}^{-1}$. It is known that for decomposable models, the IPS algorithm ends in one cycle [13,25]. In dual formulation, not only the finite termination can be proved easily but also direct estimates can be found as shown in Theorems 2.5 and 2.6.

Next we generalize the covariance selection to models in which the covariance matrix belongs to a polyhedral subset of the cone of positive definite matrices. Such constraints could be expressed as

$$\mathcal{C}_\ell = \left\{ \boldsymbol{\Sigma} = (\sigma_{ij}) : \sum_{(i,j)\in\mathcal{I}_\ell} a_{ij}^{\ell}\sigma_{ij} \geq b_\ell \right\}, \quad \ell = 1,\ldots,m, \tag{1.3}$$

where the constants a_{ij}^{ℓ}, b_ℓ are given, and $\mathcal{I}_\ell \subset \mathcal{M}$.

Consider three useful situations below.

(1) A multivariate pdf f is MTP_2 if $f(\boldsymbol{x})f(\boldsymbol{y}) \leq f(\boldsymbol{x}\wedge\boldsymbol{y})f(\boldsymbol{x}\vee\boldsymbol{y})$, where $\wedge, \vee$ mean coordinatewise minimum, maximum respectively [15]. Rinott and Scarsini [23] shows that a multivariate normal $\boldsymbol{X}$ is MTP_2 if and only if $\boldsymbol{\Sigma}^{-1}$ has all nonpositive off-diagonal elements ($\boldsymbol{\Sigma}$ has nonpositive partial covariances). Then $\boldsymbol{\Sigma}$ has nonnegative elements. In this context, a related notion is 'associated' random variables. A vector $\boldsymbol{X} = (X_1,\ldots,X_p)$ is said to be *associated* if for every pair of nondecreasing functions f, g, $\text{cov}[f(\boldsymbol{X}), g(\boldsymbol{X})] \geq 0$. If $\boldsymbol{X}$ is MTP_2, then $\boldsymbol{X}$ is associated. For a multivariate normal $\boldsymbol{X}$, Bølviken [4] establishes the following two basic results on dependence and association among $|X_1|,\ldots,|X_p|$. First, $|X_1|,\ldots,|X_p|$ are positively likelihood ratio dependent in sequence [11] if and only if $\boldsymbol{\Sigma}$ has nonnegative partial covariances. Second, $|X_1|,\ldots,|X_p|$ are associated if $\boldsymbol{\Sigma}$ has nonnegative partial covariances. This result on association gives rise to probability inequalities for regions other than rectangular ones.

(2) The covariances in longitudinal studies are positive and nonincreasing over time. To model such data, it is a common practice to use either an 'unstructured' or a 'patterned' covariance matrix. The unstructured model makes no assumption about the variances and covariances whatsoever. The covariance pattern models: compound symmetry ($\text{cov}(X_{i,j}, X_{i,j+k}) = \sigma^2\rho, \rho \geq 0$), Toeplitz ($\text{cov}(X_{i,j}, X_{i,j+k}) = \sigma^2\rho_k$), autoregressive ($\text{cov}(X_{i,j}, X_{i,j+k}) = \sigma^2\rho^k, \rho \geq 0$), exponential ($\text{cov}(X_{i,j}, X_{i,j+k}) = \sigma^2 e^{-\theta|t_{i,j}-t_{i,j+k}|}, \theta > 0$), etc. are well known, but all assume equal variance σ^2 and make specific assumptions about the covariances. However, in longitudinal studies the variances are rarely constant over time, e.g. compare the baseline and post-baseline measurements [12]. Our approach fits in between these extremes. While we allow unconstrained covariance matrix, we seek its estimate subject to given constraints on its elements.

(3) Other than nonincreasing, other types of constraints on covariances are also of interest. Examples: (a) In an AIDS study, Diggle et al. [9] consider average covariances along the diagonal ($\overline{\sigma}_k = \sum_{j=1}^{p-k}\sigma_{j,j+k}/(p-k), k = 1,\ldots,p-1$). These average covariances are expected to be nonincreasing as one moves away from the diagonal. The example in Section 5 finds restricted estimates of covariances subject to these restrictions. (b) Suppose in a six treatment therapy course, it is of interest to measure change in health outcome between successive therapy treatments, e.g., $X_2 - X_1, X_3 - X_2$, etc., where X_i is the effect of the ith treatment. Assuming the effect of therapy is most at the beginning of the course, it would be reasonable to assume that $\text{var}(X_2 - X_1) \geq \text{var}(X_3 - X_2)$, etc. The restrictions in (a), (b) are not easy to handle using the specific forms of patterned covariance matrices.

To be specific, in the case of a (4×4) covariance matrix, one might be interested in constraints such as $\{\sigma_{12} \geq \sigma_{13} \geq \sigma_{14}, \sigma_{23} \geq \sigma_{24}, \sigma_{23} \geq \sigma_{13}, \sigma_{34} \geq \sigma_{24} \geq \sigma_{14}\}$, $\{(\sigma_{12}+\sigma_{23}+\sigma_{34})/3 \geq (\sigma_{13}+\sigma_{24})/2 \geq \sigma_{14}\}$, or $\{\text{var}(X_1 \pm X_2) \geq \text{var}(X_1 \pm X_3)\}$, etc. without using patterned covariance matrices. Multivariate dependence as expressed in such situations can be expressed in the form of (1.3) by choosing a_{ij}^{ℓ}, b_ℓ appropriately.

The constraints (1.3) are nontrivial extensions of (1) in (1.2) due to constants a_{ij}^{ℓ}, b_ℓ and the presence of inequalities. Thus finding the I-projection subject to (1.3) is an extension to the covariance selection problem. Our solution in Section 2 using Fenchel duality to the covariance selection problem serves as a motivation to the method of Section 3. In the presence of dependences such as (1.3), the iterative algorithms of Speed and Kiiveri [25] and those of Section 2 may *not* converge to the correct solution (see Example 3.7). For this case, we suggest an adjustment to the iterative algorithm of Section 2, and show that this new algorithm (Algorithm 3.8) converges correctly.

The technique of covariance structure analysis is popularly used for multivariate data in behavioral sciences. However, estimation of covariance matrix subject to inequality constraints is not very common. Lee [18] considered penalty function method to find the MLE under normal distribution and also discussed generalized least squares method with inequality constraints. Shaw and Geyer [24] also considered maximum likelihood estimation (and testing) in constrained covariance models using a cutting plane algorithm. However, as shown in Section 5.2, these numerical methods run into difficulties due to necessity to invert matrices and inability to find initial starting values especially at smaller sample sizes. These difficulties are mostly avoided in duality based methods of this paper.

To describe the preliminaries of our approach, consider two probability measures (PMs) P and Q defined on an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$. The *I-divergence* or the *Kullback–Leibler distance* between P and Q is defined [6] as

$$I(P|Q) = \begin{cases} \int \ln(dP/dQ)\, dP, & \text{if } P \ll Q, \\ +\infty, & \text{elsewhere.} \end{cases} \tag{1.4}$$

Although $I(P|Q)$ is not a metric, it is always nonnegative and equals 0 if and only if $P = Q$. Hence it is often interpreted as a measure of 'divergence' or 'distance' between P and Q.

For a given Q and a specified set of PMs $\mathcal{C}$, it is often of interest to find $R \in \mathcal{C}$ which satisfies

$$I(R|Q) = \inf_{P\in\mathcal{C}} \int \ln(dP/dQ)\, dP \quad (<\infty). \tag{1.5}$$

Such an R is called the *I-projection* of Q onto $\mathcal{C}$. Csiszár [7] has shown that R exists uniquely if $\mathcal{C}$ is convex, variation-closed and there exists $P \in \mathcal{C}$ such that $I(P|Q) < \infty$. Csiszár [7] also gives a characterization of R as follows: R is the I-projection of Q onto the convex set $\mathcal{C}$ if and only if

$$I(P|Q) \geq I(P|R) + I(R|Q) \tag{1.6}$$

for every $P \in \mathcal{C}$ (equality holds if R is an algebraic inner point of $\mathcal{C}$).

We will work with the normed, linear vector space $L_1(Q)$ as the primal space since $I(P|Q) < \infty$ implies $dP/dQ \in L_1(Q)$. This would imply that $L_\infty(Q)$ would be the dual space; however it is too restrictive. Hence we define our dual space to be $\overline{M}(\mathcal{X}, \mathcal{B})$, the set of extended-valued, $\mathcal{B}$-measurable functions on $\mathcal{X}$. As we will work with pdf's, relating to the set $\mathcal{C}$ of PMs, we define a set of pdf's $\mathcal{C}_0 = \{f(\boldsymbol{x}) \in L_1(Q) : f(\boldsymbol{x}) = (dP/dQ)(\boldsymbol{x}), P \in \mathcal{C}\}$. Writing $I(P|Q) = \int f(\boldsymbol{x}) \ln f(\boldsymbol{x}) dQ(\boldsymbol{x})$, consider the functionals

$$h(f) = \int f(\boldsymbol{x}) \ln f(\boldsymbol{x}) dQ(\boldsymbol{x}) \quad \text{and} \quad h^*(g) = \ln\left[\int e^{g(x)} dQ(x)\right].$$

The following theorem taken from [2] is modified to suit to our context.

Theorem 1.1. *The functional $h^*(g)$ is convex conjugate to $h(f)$. The primal and dual problems are*

$$\inf_{f\in\mathcal{C}_0} h(f) \quad \text{and} \quad \inf_{g\in\mathcal{C}_0^*} h^*(g) < \infty, \tag{1.7}$$

respectively, where the dual cone $\mathcal{C}_0^*$ is defined as $\mathcal{C}_0^* = \{g : \int g(\boldsymbol{x}) f(\boldsymbol{x}) dQ(\boldsymbol{x}) \geq 0, \forall f \in \mathcal{C}_0\}$. For some $f_0 \in \mathcal{C}_0, g_0 \in \mathcal{C}_0^*$, if $h(f_0) + h^*(g_0) \leq 0$, then f_0, g_0 solve the two optimization problems of (1.7), respectively, and, $h(f_0) = -h^*(g_0)$. Also, then $f_0(\boldsymbol{x}) = e^{g_0(\boldsymbol{x})} / \int e^{g_0(\boldsymbol{x})}\, dQ(\boldsymbol{x})$. □

Depending on the form of $\mathcal{C}$, it may be difficult to find a solution to (1.5). For finite $\mathcal{X}$, Csiszár [7] has shown that if $\mathcal{C}$ can be expressed as $\cap_{i=1}^k \mathcal{C}_i$, where each $\mathcal{C}_i$ is a closed, linear set, then the sequence of cyclic iterated I-projections onto the $\mathcal{C}_i$ converges to the solution of (1.5). Dykstra [10] modified Csiszár 's procedure to encompass the case where each $\mathcal{C}_i$ is an arbitrary closed, convex set. Bhattacharya [1] extended this procedure to the case of infinite dimensions.

When f, g are the pdf's of $P = \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, $Q = \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Gamma})$, respectively, then the I-divergence in (1.4) is given by

$$I(f|g) = I(P|Q) = -\frac{1}{2}\left[\log\det(\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1}) + \mathrm{tr}(\boldsymbol{I} - \boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1})\right]. \tag{1.8}$$

2. Covariance selection model

The primal problem from (1.5) is

$$\inf_{\mathcal{K}_\mathcal{C}} \int f(\boldsymbol{x}) \ln f(\boldsymbol{x}) dQ(\boldsymbol{x}) \tag{2.1}$$

where $\mathcal{K}_{\mathcal{C}} = \{\delta f(\boldsymbol{x}) : \delta \geq 0, f = dP/dQ, P \in \mathcal{C}\}$. We define $\mathcal{K}_{\mathcal{C}}$ this way so that it is a cone of functions, but note that if $\delta \neq 1$, then from (1.4) the integral in (2.1) is $+\infty$; hence we will skip δ in defining $\mathcal{K}_{\mathcal{C}}$ from now on. As solving the covariance selection problem also amounts to finding the normal pdf with prescribed marginals (which corresponds to specifying the covariance matrix only as the mean is set at zero), we formulate the covariance selection problem as finding the I-projection of a given Q onto $\mathcal{C}$, where

$$\mathcal{C} = \{P : E_P(X_iX_j) = s_{ij}, i, j \in \mathcal{I} = \{1, \ldots, a\}\} \tag{2.2}$$

for given s_{ij}. For simplicity in (2.2) and the rest of this section, we have considered the rectangular index sets. The case of more general index sets as in (1.2) would follow from the material of Section 3, which will be apparent later.

Lemma 2.1 shows that when we I-project from $Q = \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Gamma})$ onto $\mathcal{C}$, the solution is a multivariate normal distribution, for which the elements of the inverse of the covariance matrix with indices in $\mathcal{I}$ are updated from those of $\boldsymbol{\Gamma}^{-1}$, but the elements with indices in $\overline{\mathcal{I}}$ are left unchanged as those of $\boldsymbol{\Gamma}^{-1}$. We present a proof of the scaling factor of the I-projection solution to (2.1) using duality (compare with Lemma 2 of [25]) and identify the elements of the solution inverse matrix as scalar solutions to the dual problem.

Let $\boldsymbol{\Gamma} = (\gamma_{ij})$, $\boldsymbol{\Gamma}^{-1} = (\gamma^{ij})$, $s_{\mathcal{I}}^{ij}$ be the (i, j)th element of the inverse of $\mathbf{S}_{\mathcal{I}}$, the matrix with elements of $\mathbf{S}$ restricted to $\mathcal{I}$ and $\gamma_{\mathcal{I}}^{ij}$ be defined in a similar way.

Lemma 2.1. *The solution of* (2.1) *subject to* (2.2) *is given by* $\boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, *where*

(i) $\sigma^{ij} = s_{\mathcal{I}}^{ij} - \gamma_{\mathcal{I}}^{ij} + \gamma^{ij}$, $i, j \in \mathcal{I}$, $\sigma^{ij} = \gamma^{ij}$, *otherwise, and*
(ii) $\boldsymbol{\Sigma}$ *is positive definite.*

Proof. (i) Recall $\mathcal{K}_{\mathcal{C}}$, which is defined below (2.1). Writing $\mathcal{K}_{\mathcal{C}} = \{f(\boldsymbol{x}) : \int (x_ix_j - s_{ij}) f(\boldsymbol{x}) dQ(\boldsymbol{x}) = 0, i, j \in \mathcal{I}\}$, from [19] the dual set is given by $\mathcal{K}_{\mathcal{C}}^* = \left\{\sum_{i,j\in\mathcal{I}} \alpha_{ij}(x_ix_j - s_{ij}) : \alpha_{ij} \in \Re, i, j \in \mathcal{I}\right\}$. The corresponding dual problem (1.7) is, equivalently (ignoring the natural log), expressed as

$$\inf_{\alpha_{ij}\in\Re, i,j\in\mathcal{I}} \int e^{\sum_{i,j\in\mathcal{I}} \alpha_{ij}(x_ix_j - s_{ij})} dQ = \inf_{\alpha_{ij}\in\Re, i,j\in\mathcal{I}} \int e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}}(-2\alpha_{ij})(x_ix_j - s_{ij})} dQ. \tag{2.3}$$

Let $Q_{\mathcal{I}}$ be the $\mathcal{I}$-marginal of Q, $c(\boldsymbol{\Gamma}_{\mathcal{I}}) = \int e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}} \gamma_{\mathcal{I}}^{ij} x_ix_j} d\boldsymbol{x}_{\mathcal{I}}$, $\boldsymbol{x}_{\mathcal{I}} = (x_i,\ i \in \mathcal{I})$ and $b = e^{\frac{1}{2}\sum_{i,j\in\mathcal{I}}(s_{\mathcal{I}}^{ij} - \gamma_{\mathcal{I}}^{ij}) s_{ij}}$. Differentiating the last quantity in (2.3) with respect to α_{ij}, and then replacing $-2\alpha_{ij} = s_{\mathcal{I}}^{ij} - \gamma_{\mathcal{I}}^{ij}$ we get

$$\begin{aligned}
b \int (x_ix_j - s_{ij}) e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}}(s_{\mathcal{I}}^{ij} - \gamma_{\mathcal{I}}^{ij}) x_ix_j} dQ &= b \int (x_ix_j - s_{ij}) e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}}(s_{\mathcal{I}}^{ij} - \gamma_{\mathcal{I}}^{ij}) x_ix_j} dQ_{\mathcal{I}} \\
&= \frac{b}{c(\boldsymbol{\Gamma}_{\mathcal{I}})} \int (x_ix_j - s_{ij}) e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}}(s_{\mathcal{I}}^{ij} - \gamma_{\mathcal{I}}^{ij}) x_ix_j - \frac{1}{2}\sum_{i,j\in\mathcal{I}} \gamma_{\mathcal{I}}^{ij} x_ix_j} d\boldsymbol{x}_{\mathcal{I}} \\
&= \frac{b}{c(\boldsymbol{\Gamma}_{\mathcal{I}})} \int (x_ix_j - s_{ij}) e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}} s_{\mathcal{I}}^{ij} x_ix_j} d\boldsymbol{x}_{\mathcal{I}} \\
&= \frac{b}{c(\boldsymbol{\Gamma}_{\mathcal{I}})}(0) = 0.
\end{aligned}$$

Hence it follows that $\alpha_{ij}^* = -(s_{\mathcal{I}}^{ij} - \gamma_{\mathcal{I}}^{ij})/2$, $i, j \in \mathcal{I}$, $\alpha_{ij}^* = 0$, otherwise, solves the dual problem (2.3). Now, using Theorem 1.1, the solution to (2.1) is given by P^* where

$$\frac{dP^*}{dQ} = \frac{e^{\sum_{i,j\in\mathcal{I}} \alpha_{ij}^*(x_ix_j - s_{ij})}}{\int e^{\sum_{i,j\in\mathcal{I}} \alpha_{ij}^*(x_ix_j - s_{ij})} dQ} = \frac{e^{\sum_{i,j\in\mathcal{I}} \alpha_{ij}^* x_ix_j}}{\int e^{\sum_{i,j\in\mathcal{I}} \alpha_{ij}^* x_ix_j} dQ}. \tag{2.4}$$

Replacing α_{ij}^* by its value, the denominator of the last expression in (2.4) can be simplified as

$$\int e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}}(s_{\mathcal{I}}^{ij} - \gamma_{\mathcal{I}}^{ij}) x_ix_j} dQ_{\mathcal{I}} = \frac{1}{c(\boldsymbol{\Gamma}_{\mathcal{I}})} \int e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}} s_{\mathcal{I}}^{ij} x_ix_j} d\boldsymbol{x}_{\mathcal{I}} = \frac{c(\mathbf{S}_{\mathcal{I}})}{c(\boldsymbol{\Gamma}_{\mathcal{I}})}, \tag{2.5}$$

where $c(\mathbf{S}_{\mathcal{I}}) = \int e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}} s_{\mathcal{I}}^{ij} x_ix_j} d\boldsymbol{x}_{\mathcal{I}}$. Then, using (2.4), (2.5), we can express the solution as

$$f^*(\boldsymbol{x}) = \frac{dP^*}{d\boldsymbol{x}} = \frac{dP^*}{dQ}\frac{dQ}{d\boldsymbol{x}} = \frac{c(\boldsymbol{\Gamma}_{\mathcal{I}})}{c(\mathbf{S}_{\mathcal{I}})c(\boldsymbol{\Gamma})} e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}}((s_{\mathcal{I}}^{ij} - \gamma_{\mathcal{I}}^{ij}) + \gamma^{ij}) x_ix_j - \frac{1}{2}\sum_{(i\notin\mathcal{I})\cup(j\notin\mathcal{I})} \gamma^{ij} x_ix_j},$$

$(c(\boldsymbol{\Gamma}) = \int e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}} \gamma^{ij} x_ix_j} d\boldsymbol{x})$, which is the pdf of an $\boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, where the i, jth entry of $\boldsymbol{\Sigma}^{-1}$ is σ^{ij} as given in the statement of the lemma.

To see that P^* satisfies the constraint, note that for $i, j \in \mathcal{I}$,

$$\begin{aligned}\int x_i x_j \, dP^* &= \frac{c(\mathbf{\Gamma}_{\mathcal{I}})}{c(\mathbf{S}_{\mathcal{I}})} \int x_i x_j e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}}(s^{ij}_{\mathcal{I}}-\gamma^{ij}_{\mathcal{I}})x_i x_j} \, dQ, \\ &= \frac{c(\mathbf{\Gamma}_{\mathcal{I}})}{c(\mathbf{S}_{\mathcal{I}})} \int x_i x_j e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}}(s^{ij}_{\mathcal{I}}-\gamma^{ij}_{\mathcal{I}})x_i x_j} \, dQ_{\mathcal{I}}, \\ &= \frac{1}{c(\mathbf{S}_{\mathcal{I}})} \int x_i x_j e^{-\frac{1}{2}\sum_{i,j\in\mathcal{I}} s^{ij}_{\mathcal{I}} x_i x_j} \, d\boldsymbol{x} = s_{ij}.\end{aligned}$$

(ii) Since we are minimizing a convex function over a closed, convex set, the solution exists. Thus, $\mathbf{\Sigma}^{-1}$ exists and must be positive definite. Hence $\mathbf{\Sigma}$ is positive definite. □

Remark 2.2. When $\mathcal{F} = \phi$ (empty set) in (1.2), $\mathbf{\Gamma} = \boldsymbol{I}$, the identity matrix, we get the solution of the corresponding covariance selection problem, thus the problem of maximum likelihood is solved by an I-projection. The equations (2) in (1.2) is redundant since they are satisfied by the solutions already, as seen from Lemma 2.1. □

Remark 2.3. When $\mathcal{F} = \mathcal{M}$ in (1.2), the solution is the extension of well-known maximum entropy characterization of the multivariate normal distribution (e.g., [14]), where the Lebesgue measure is replaced by a PM Q. □

For the problem of finding the normal pdf whose marginals are known beforehand, let $\boldsymbol{X}_{\mathcal{I}_\ell}$ be a subvector of $\boldsymbol{X}$ with indices taken from $\mathcal{I}_\ell$, where $\mathcal{I}_\ell \subset \{1, \dots, p\}$, $1 \le \ell \le m$, and $\mathbf{S}_\ell$ be the covariance matrix of $\boldsymbol{X}_{\mathcal{I}_\ell}$. We consider the problem in (2.1) with

$$\mathcal{C} = \bigcap_{\ell=1}^{m} \mathcal{C}_\ell, \quad \mathcal{C}_\ell = \{P : E_P(\boldsymbol{X}_{\mathcal{I}_\ell}\boldsymbol{X}^T_{\mathcal{I}_\ell}) = \mathbf{S}_\ell\}, \tag{2.6}$$

$\mathbf{S}_\ell$, $1 \le \ell \le m$, are given, assuming the $\mathcal{C}_\ell$'s are compatible. Considering this situation, we present the following algorithm, which is a dual version of the corresponding iterative proportional fitting algorithm.

Algorithm 2.4. 1. Start with $Q = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{I}) = P_{10}$, $\mathbf{\Sigma}_{1,0} = \boldsymbol{I}$.

2. At nth cycle, ℓth constraint ($1 \le n$, $1 \le \ell \le m$), find the I-projection of $P_{n,\ell-1} = \boldsymbol{N}_p(\mathbf{0}, \mathbf{\Sigma}_{n,\ell-1})$ onto $\mathcal{C}_\ell$, given by $P_{n,\ell} = \boldsymbol{N}_p(\mathbf{0}, \mathbf{\Sigma}_{n,\ell})$. From Lemma 2.1, $\sigma^{ij}_{n,\ell} = s^{ij}_{\mathcal{I}_\ell} - (\mathbf{\Sigma}_{n,\ell-1})^{ij}_{\mathcal{I}_\ell} + \sigma^{ij}_{n,\ell-1}$, $\forall i, j \in \mathcal{I}_\ell$, $\sigma^{ij}_{n,\ell} = \sigma^{ij}_{n,\ell-1}$, otherwise.
3. Replace ℓ by $\ell+1$ (when $\ell = m+1$, replace ℓ by 1, n by $n+1$) and repeat step 2. Repeat steps 2 and 3 until all constraints in $\mathcal{C}$ are approximately satisfied.

The following theorem is a special case of Theorems 3.10 and 3.11 in Section 3, and hence the proof is omitted.

Theorem 2.5. *Algorithm 2.4 converges to the correct solution of (2.1) subject to (2.6). For decomposable constraints, it terminates in one cycle.* □

It is known that with decomposable constraints for contingency tables, the IPS algorithm converges in one cycle [13, Chapter 5]. For normal models with a triangulated and connected graph, Speed and Kiiveri [25] showed that 'decomposability' corresponds to an enumeration of cliques such that for each ℓ, the clique $\mathcal{C}_\ell$ contains a vertex not in $\mathcal{C}_1, \dots, \mathcal{C}_{\ell-1}$, and the IPS algorithm terminates in one cycle here as well.

The following theorem derives the direct estimates for two decomposable cases. Similar results can be found in [20,17] when finding the MLE of the covariance matrix in decomposable cases using algebraic techniques. Suppose the matrix $\mathbf{S}$ is partitioned into $(m \times m)$ matrix blocks, whose i, jth entry is $\mathbf{S}_{ij}$, $1 \le i, j \le m$.

Theorem 2.6. (1) *Consider the I-projection of $Q = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{I})$ onto $\mathcal{C} = \cap_{\ell=1}^m \mathcal{C}_\ell$, $\mathcal{C}_\ell = \{P : E(\boldsymbol{X}_{\mathcal{I}_\ell}\boldsymbol{X}^T_{\mathcal{I}_\ell}) = \mathbf{S}_{\ell\ell}\}$, for $\mathcal{I}_{\ell_1} \cap \mathcal{I}_{\ell_2} = \phi$, $\ell_1 \ne \ell_2$, $\cup_\ell \mathcal{I}_\ell = \{1, \dots, p\}$. Here Algorithm 2.4 terminates in one cycle and the final solution is given by $P_{1m} = \boldsymbol{N}_p(\mathbf{0}, \mathbf{\Sigma}_{1m})$, where $\mathbf{\Sigma}_{1k} = \mathrm{diag}(\mathbf{S}_{11}, \dots, \mathbf{S}_{mm})$, assuming $\mathbf{S}^{-1}_{\ell\ell}$ exist, $\forall \ell$.*

(2) *Consider the I-projection of $Q = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{I})$ onto $\mathcal{C} = \cap_{\ell=1}^{m-1} \mathcal{C}_\ell$, $\mathcal{C}_\ell = \{P : E(\boldsymbol{X}_{\mathcal{I}_\ell}\boldsymbol{X}^T_{\mathcal{I}_\ell}) = \mathbf{S}_{\mathcal{I}_\ell}\}$, where $\mathbf{S}_{\mathcal{I}_\ell} = \begin{pmatrix} \mathbf{S}_{\ell\ell} & \mathbf{S}_{\ell,\ell+1} \\ \mathbf{S}_{\ell+1,\ell} & \mathbf{S}_{\ell+1,\ell+1} \end{pmatrix}$. Here Algorithm 2.4 terminates in one cycle and the final solution is given by $P_{1,m-1} = \boldsymbol{N}_p(\mathbf{0}, \mathbf{\Sigma}_{1,m-1})$, where*

$$\mathbf{\Sigma}_{1,m-1} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \mathbf{S}_{12}\mathbf{S}^{-1}_{22}\mathbf{S}_{23} & \mathbf{S}_{13}\mathbf{S}^{-1}_{33}\mathbf{S}_{34} & \cdots & \mathbf{S}_{1,m-1}\mathbf{S}^{-1}_{m-1,m-1}\mathbf{S}_{m-1,m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \mathbf{S}_{ii} & \mathbf{S}_{i,i+1} & \mathbf{S}_{i,i+1}\mathbf{S}^{-1}_{i+1,i+1}\mathbf{S}_{i+1,i+2} & \cdots & \mathbf{S}_{i,m-1}\mathbf{S}^{-1}_{m-1,m-1}\mathbf{S}_{m-1,m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \mathbf{S}_{mm} \end{pmatrix}, \tag{2.7}$$

assuming all necessary inverses exist.

Proof. (1) We first I-project $Q = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{I})$ onto $\mathcal{C}_1$. Using Theorem 1.1 and Lemma 2.1, the solution is given by dP_{11}/dQ, where $P_{11} = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{11} = \text{diag}(\mathbf{S}_{11}, \boldsymbol{I}))$. Next we I-project P_{11} onto $\mathcal{C}_2$. The corresponding dual problem is:

$$\inf_{\alpha_{ij}, i,j \in \mathcal{I}_2} \int e^{\sum_{\mathcal{I}_2} \alpha_{ij}(x_i x_j - s_{ij})} dP_{11} = \inf_{\alpha_{ij}, i,j \in \mathcal{I}_2} \int e^{\sum_{\mathcal{I}_2} \alpha_{ij}(x_i x_j - s_{ij})} d(P_{11})_{\mathcal{I}_2}. \tag{2.8}$$

Using Lemma 2.1, the solution dP_{12}/dP_{11} is a multivariate normal distribution with elements of inverse covariance matrix given by: $s^{ij}_{\mathcal{I}_2} - \delta^{ij}$, where $\delta^{ij} = 1$, if $i = j$, and $\delta^{ij} = 0$, otherwise. Writing $dP_{12}/dQ = (dP_{12}/dP_{11})(dP_{11}/dQ)$, the solution is given by $P_{12} = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{12} = \text{diag}(\mathbf{S}_{11}, \mathbf{S}_{22}, \boldsymbol{I}))$. This process continues and by induction at the mth step we get the solution as given in the statement of the theorem, which satisfies all the constraints.

(2) As shown in (1), the I-projection of $Q = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{I})$ onto $\mathcal{C}_1$ is given by $P_{11} = \boldsymbol{N}_p\left(\mathbf{0}, \boldsymbol{\Sigma}_{11} = \text{diag}\left(\mathbf{S}_{\mathcal{I}_1}, \boldsymbol{I}\right)\right)$. Next we I-project P_{11} onto $\mathcal{C}_2$. The corresponding dual problem is obtained from (2.8). Using Lemma 2.1, the solution dP_{12}/dP_{11} is a multivariate normal distribution with elements of inverse covariance matrix are given by: $s^{ij}_{\mathcal{I}_2} - s^{ij}_{\mathcal{I}_1 \cap \mathcal{I}_2}$, if $i, j \in \mathcal{I}_1 \cap \mathcal{I}_2$, and, $s^{ij}_{\mathcal{I}_2} - \delta^{ij}$, if $i, j \in \mathcal{I}_2 - \mathcal{I}_1$, where $\delta^{ij} = 1$, if $i = j$, and $\delta^{ij} = 0$, otherwise.

Writing $dP_{12}/dQ = (dP_{12}/dP_{11})(dP_{11}/dQ)$, the solution is given by $P_{12} = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{12})$ where in matrix terms

$$\boldsymbol{\Sigma}_{12}^{-1} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{-1}_{\mathcal{I}_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{-1}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I} \end{pmatrix} + \begin{pmatrix} \mathbf{S}^{-1}_{\mathcal{I}_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{pmatrix}. \tag{2.9}$$

Using expressions of inverses of partitioned matrices $\mathbf{S}_{\mathcal{I}_1}, \mathbf{S}_{\mathcal{I}_2}$ [22] and direct multiplication, one can validate $\boldsymbol{\Sigma}_{12}$ given below,

$$\boldsymbol{\Sigma}_{12} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} & \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{23} & \mathbf{0} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \mathbf{S}_{23} & \mathbf{0} \\ \mathbf{S}_{32}\mathbf{S}_{22}^{-1}\mathbf{S}_{21} & \mathbf{S}_{32} & \mathbf{S}_{33} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I} \end{pmatrix}. \tag{2.10}$$

For projection onto $\mathcal{C}_\ell$, $\ell \geq 3$, we use induction on $q =$ the number of constraints. Assuming the result holds for any $q < m - 1$, to show that it holds for $q + 1$, the (block) matrix identities obtained $\boldsymbol{\Sigma}_{1j}\boldsymbol{\Sigma}_{1j}^{-1} = \boldsymbol{I}$, $1 \leq j \leq q$, are used and with matrix algebra the form of $\boldsymbol{\Sigma}_{1,q+1}$ is verified. It is now obvious that $\boldsymbol{\Sigma}_{1,m-1}$ satisfies $\mathcal{C}_\ell$, $\forall \ell$ (equivalently, one can also check that the dual solutions are zeros starting from the second cycle). Hence the algorithm terminates after the first cycle. □

For decomposable constraints different from those in Theorem 2.6, the form of $\boldsymbol{\Sigma}_{1,m-1}$ would also be different.

3. Multivariate dependence

For $\boldsymbol{X} \sim P = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, the covariance selection model uses constraints (2.2), which corresponds to conditional independence of two components of $\boldsymbol{X}$ with indices in $\bar{\mathcal{I}}$. In this section, we are interested in estimating the covariance matrix subject to dependence such as (1.3). Such constraints can be expressed as

$$\mathcal{C} = \{P : \boldsymbol{A\sigma} \geq \boldsymbol{b}\}, \tag{3.1}$$

where $\boldsymbol{\sigma} = (\sigma_j)$ is a column vector obtained by stacking the elements of $\boldsymbol{\Sigma}$ in a column (ignoring those below the diagonal) of length $r = p(p+1)/2$, and $\boldsymbol{A}$ is an $(m \times r)$ matrix with $\text{rank}(\boldsymbol{A}) = m$ $(<r)$, $\boldsymbol{b}$ is an m-vector of given constants. Several useful cases of this type of dependence are described below (1.3). Recall $\mathcal{K}_{\mathcal{C}} = \{\alpha f(\boldsymbol{x}) : \alpha \geq 0, f = dP/dQ, P \in \mathcal{C}\}$ and let $\mathcal{D}$ be the set of all symmetric matrices.

We consider the problem of finding the I-projection of a given $Q = \boldsymbol{N}_p(\mathbf{0}, \boldsymbol{\Gamma})$ onto $\mathcal{C}$ in (3.1). The next lemma derives the constraint region for the corresponding dual problem.

Lemma 3.1. *For $\mathcal{C}$ in (3.1), the dual cone of $\mathcal{K}_{\mathcal{C}}$ is given by*

$$\mathcal{K}^*_{\mathcal{C}} = \left\{g : g = -\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta x} - c, \boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A}) \in \mathcal{D}, c = c(\boldsymbol{\alpha}, \boldsymbol{b}) \in \Re, \boldsymbol{\alpha} \in \Re^{+p}, \boldsymbol{x} \in \Re^p\right\}. \tag{3.2}$$

Proof. The relation between the elements of $\boldsymbol{\sigma} = (\sigma_j)$ in (3.1) and $\boldsymbol{\Sigma} = (\sigma_{st})$, $s < t$ in (1.3) can be obtained as follows. Let $d_i = p - i + 1$, $1 \leq i \leq p$. To obtain σ_{st} from given σ_j, for any j, $1 \leq j \leq r$, find i $(1 \leq i \leq m)$ such that $\sum_1^{i-1} d_k < j \leq \sum_1^i d_k$ $(\sum_1^0 d_k = 0)$; then define arrays $a(j), b(j)$ such that $a(j) = s = i, b(j) = t = j - \sum_1^{i-1} d_k$. On the other hand, given σ_{st}, one can find σ_j by using $j = \sum_1^{s-1} d_k + t - s + 1$.

We may write the constraints in (3.1) as $\mathcal{C} = \cap_{i=1}^m \mathcal{C}_i$, $\mathcal{K}_{\mathcal{C}} = \cap_{i=1}^m \mathcal{K}_{\mathcal{C}_i}$, where

$$\mathcal{C}_i = \left\{P : \sum_{j=1}^r a_{ij}\sigma_j \geq b_i\right\}, \qquad \mathcal{K}_{\mathcal{C}_i} = \{f(\boldsymbol{x}) : P \in \mathcal{C}_i\}, \tag{3.3}$$

where $\boldsymbol{A} = (a_{ij})$, $\sigma_j = \text{cov}(X_{a(j)}, X_{b(j)})$. If the constraints in $\mathcal{C}_i$ involves covariances between the subvectors $\boldsymbol{X}_{\mathcal{G}_i}$ and $\boldsymbol{X}_{\mathcal{H}_i}$ of $\boldsymbol{X}$, where $\boldsymbol{X}_{\mathcal{G}_i}$ ($\boldsymbol{X}_{\mathcal{H}_i}$) contains the X-variates with indices in subset $\mathcal{G}_i$ ($\mathcal{H}_i$) of $\{1, \ldots, p\}$, and $\boldsymbol{A}_i$ is a diagonal matrix which contains the nonzero elements of the ith row of $\boldsymbol{A}$, then one can write from (3.3),

$$\mathcal{K}_{\mathcal{C}_i} = \left\{ f(\boldsymbol{x}) : \int [\boldsymbol{x}'_{\mathcal{G}_i} \boldsymbol{A}_i \boldsymbol{x}_{\mathcal{H}_i} - b_i] f(\boldsymbol{x})\, dQ(\boldsymbol{x}) \geq 0 \right\}. \tag{3.4}$$

(It is not necessary for $\boldsymbol{A}_i$ to be diagonal as above) The dual cone of $\mathcal{K}_{\mathcal{C}_i}$ is

$$\mathcal{K}^*_{\mathcal{C}_i} = \left\{ g : g = \alpha_i [\boldsymbol{x}'_{\mathcal{G}_i} \boldsymbol{A}_i \boldsymbol{x}_{\mathcal{H}_i} - b_i] : \alpha_i \in \Re^+, \boldsymbol{x}_{\mathcal{G}_i} \in \Re^{|\mathcal{G}_i|}, \boldsymbol{x}_{\mathcal{H}_i} \in \Re^{|\mathcal{H}_i|} \right\},$$

where $|\cdot|$ indicates size. Then from [19] the dual of $\mathcal{K}_{\mathcal{C}}$ is $\mathcal{K}^*_{\mathcal{C}} = \oplus_{i=1}^m \mathcal{K}^*_{\mathcal{C}_i}$, which can be written as

$$\begin{aligned} \mathcal{K}^*_{\mathcal{C}} &= \left\{ g(\boldsymbol{\alpha}) : g(\boldsymbol{\alpha}) = \sum_{i=1}^m \alpha_i [\boldsymbol{x}'_{\mathcal{G}_i} \boldsymbol{A}_i \boldsymbol{x}_{\mathcal{H}_i} - b_i] : \alpha_i \in \Re^+, \boldsymbol{x}_{\mathcal{G}_i} \in \Re^{|\mathcal{G}_i|}, \boldsymbol{x}_{\mathcal{H}_i} \in \Re^{|\mathcal{H}_i|}, \forall i \right\} \\ &= \left\{ g(\boldsymbol{\alpha}) : g(\boldsymbol{\alpha}) = \sum_{i=1}^m \alpha_i \boldsymbol{x}'_{\mathcal{G}_i} \boldsymbol{A}_i \boldsymbol{x}_{\mathcal{H}_i} - \sum_{i=1}^m \alpha_i b_i : \alpha_i \in \Re^+, \boldsymbol{x}_{\mathcal{G}_i} \in \Re^{|\mathcal{G}_i|}, \boldsymbol{x}_{\mathcal{H}_i} \in \Re^{|\mathcal{H}_i|}, \forall i \right\} \\ &= \{ g(\boldsymbol{\alpha}) : g(\boldsymbol{\alpha}) = \boldsymbol{x}'[\boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A})]\boldsymbol{x} - c(\boldsymbol{\alpha}, \boldsymbol{b}), \boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A}) \in \mathcal{D}, \boldsymbol{\alpha} \in \Re^{p+}, c(\boldsymbol{\alpha}, \boldsymbol{b}) \in \Re, \boldsymbol{x} \in \Re^p \}, \end{aligned}$$

for some matrix $\boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A})$ obtained from the identity

$$\boldsymbol{x}'[\boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A})]\boldsymbol{x} = -2 \sum_{i=1}^m \alpha_i \boldsymbol{x}'_{\mathcal{G}_i} \boldsymbol{A}_i \boldsymbol{x}_{\mathcal{H}_i} \tag{3.5}$$

and $c(\boldsymbol{\alpha}, \boldsymbol{b}) = \sum_{i=1}^m \alpha_i b_i$. Note that depending on the constraints specified $\boldsymbol{\Delta}$ may have one or more zero rows and columns. □

The next lemma formulates the dual problem.

Lemma 3.2. *Assume $\boldsymbol{T} = (\boldsymbol{\Delta} + \boldsymbol{\Gamma}^{-1})^{-1}$ is positive definite. The dual problem is*

$$\inf_{\boldsymbol{\alpha} \geq \boldsymbol{0}} \left\{ -\frac{1}{2} \ln \det \left(\boldsymbol{\Gamma} \boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A}) + \boldsymbol{I}_p \right) - c(\boldsymbol{\alpha}, \boldsymbol{b}) \right\}, \tag{3.6}$$

where $\boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A})$, $c(\boldsymbol{\alpha}, \boldsymbol{b})$ are defined in Lemma 3.1.

Proof. We will abbreviate as $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A})$, $c = c(\boldsymbol{\alpha}, \boldsymbol{b})$. Using the form of the dual problem $\left[\ln \int e^g\, dQ\right]$ from Theorem 1.1, and the form of the dual solution $g = -\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}\boldsymbol{x} - c$ from Lemma 3.1, we get

$$\begin{aligned} \int e^g\, dQ &= e^{-c} \int e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}\boldsymbol{x}}\, dQ \\ &= e^{-c} \int e^{-\frac{1}{2}(\boldsymbol{x}'\boldsymbol{\Delta}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{\Gamma}^{-1}\boldsymbol{x})} d\boldsymbol{x} / (2\pi)^{p/2} (\det(\boldsymbol{\Gamma}))^{1/2} \\ &= e^{-c} \int e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{T}^{-1}\boldsymbol{x}} d\boldsymbol{x} / (2\pi)^{p/2} (\det(\boldsymbol{\Gamma}))^{1/2} \quad [\text{where } \boldsymbol{T}^{-1} = \boldsymbol{\Delta} + \boldsymbol{\Gamma}^{-1}] \\ &= e^{-c} \left[\det(\boldsymbol{T}\boldsymbol{\Gamma}^{-1})\right]^{1/2} \quad \left[\text{since} \int e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{T}^{-1}\boldsymbol{x}} d\boldsymbol{x} = (2\pi)^{p/2} (\det(\boldsymbol{T}))^{1/2}\right] \\ &= e^{-c} \left[\det\left[\boldsymbol{\Gamma}\left(\boldsymbol{\Delta} + \boldsymbol{\Gamma}^{-1}\right)\right]^{-1}\right]^{1/2} \\ &= e^{-c} \left[\det(\boldsymbol{\Gamma}\boldsymbol{\Delta} + \boldsymbol{I}_p)\right]^{-1/2}. \end{aligned}$$

Since $h^*(g) = \ln \int e^g\, dQ$, we get the expression in (3.6). □

The following theorem characterizes the solution of (2.1) subject to (3.1). It also relates the covariance pattern $\boldsymbol{\Sigma}$ of $\boldsymbol{X}$ specified by $\mathcal{C}$ in (3.1) and the conditional covariance structure $\boldsymbol{\Gamma}^{-1}$ to the conditional covariance structure ($\boldsymbol{T}^{-1}$) of the solution.

Theorem 3.3. *Suppose $\boldsymbol{\alpha}^*$ solves (3.6). Then*

(i) *the solution of (2.1) subject to (3.1) is a multivariate normal distribution given by $\boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{T})$, where $\boldsymbol{T} = [\boldsymbol{\Delta}(\boldsymbol{\alpha}^*, \boldsymbol{A}) + \boldsymbol{\Gamma}^{-1}]^{-1}$, $\boldsymbol{\Delta}(\boldsymbol{\alpha}^*, \boldsymbol{A})$ is obtained from (3.5).*
(ii) *The conditional covariance structure of Q is $\boldsymbol{\Gamma}^{-1}$, whereas that of the solution is $\boldsymbol{T}^{-1} = \boldsymbol{\Delta}(\boldsymbol{\alpha}^*, \boldsymbol{A}) + \boldsymbol{\Gamma}^{-1}$.*

Proof. (i) Using the form of the dual solution $g^* = -\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}(\boldsymbol{\alpha}^*, \boldsymbol{A})\boldsymbol{x} - c(\boldsymbol{\alpha}^*, \boldsymbol{b})$ from Lemma 3.1, we get the solution of (2.1) subject to (3.1) is given by

$$f^*(\boldsymbol{x}) = \frac{dP^*}{d\boldsymbol{x}} = \frac{dP^*}{dQ}\frac{dQ}{d\boldsymbol{x}} = \frac{e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}(\boldsymbol{\alpha}^*,\boldsymbol{A})\boldsymbol{x}-c(\boldsymbol{\alpha}^*,\boldsymbol{b})}}{\int e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}(\boldsymbol{\alpha}^*,\boldsymbol{A})\boldsymbol{x}-c(\boldsymbol{\alpha}^*,\boldsymbol{b})}\,dQ}\,\frac{e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Gamma}^{-1}\boldsymbol{x}}}{c(\boldsymbol{\Gamma})}$$
$$= \frac{e^{-\frac{1}{2}\boldsymbol{x}'\left(\boldsymbol{\Delta}(\boldsymbol{\alpha}^*,\boldsymbol{A})+\boldsymbol{\Gamma}^{-1}\right)\boldsymbol{x}}}{\int e^{-\frac{1}{2}\boldsymbol{x}'\left(\boldsymbol{\Delta}(\boldsymbol{\alpha}^*,\boldsymbol{A})+\boldsymbol{\Gamma}^{-1}\right)\boldsymbol{x}}\,d\boldsymbol{x}},$$

which is the pdf of an $\boldsymbol{N}_p(\boldsymbol{0}, (\boldsymbol{\Delta}(\boldsymbol{\alpha}^*, \boldsymbol{A}) + \boldsymbol{\Gamma}^{-1})^{-1})$ distribution.

(ii) The conditional covariance structure of the solution follows from the fact that when $\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{T})$, then $t^{ij} = \text{cov}(X_i, X_j | X_s, s \neq i, j)$. $\square$

The solution provided by Theorem 3.3 depends on the dual solution $\boldsymbol{\alpha}^*$, which solves (3.6). The following theorem gives a necessary and sufficient condition for $\boldsymbol{\alpha}^*$ so that $g(\boldsymbol{\alpha}^*)$ solves the dual problem.

Let $\boldsymbol{\Delta}_0 = \boldsymbol{\Delta}(\boldsymbol{\alpha}_0, \boldsymbol{A})$, $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A})$, $c_0 = c(\boldsymbol{\alpha}_0, \boldsymbol{b})$, $c = c(\boldsymbol{\alpha}, \boldsymbol{b})$, $\boldsymbol{T}_0 = \boldsymbol{T}(\boldsymbol{\alpha}_0, \boldsymbol{A})$, $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{\alpha}, \boldsymbol{A})$. Let $E_{\boldsymbol{L}}(\cdot)$ refers to the expectation of $\cdot$ with respect to an $\boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{L})$ distribution.

Theorem 3.4. *A necessary and sufficient condition for $\boldsymbol{\alpha}_0$ so that $g(\boldsymbol{\alpha}_0) = -\boldsymbol{x}'\boldsymbol{\Delta}_0\boldsymbol{x}/2 - c_0$ solves the dual problem is*

$$E_{\boldsymbol{T}_0}\left[\boldsymbol{x}'\boldsymbol{\Delta}_0\boldsymbol{x} + 2c_0\right] = 0, \qquad E_{\boldsymbol{T}_0}\left[\boldsymbol{x}'\boldsymbol{\Delta}\boldsymbol{x} + 2c\right] \le 0, \tag{3.7}$$

$\forall g(\boldsymbol{\alpha}) = -\boldsymbol{x}'\boldsymbol{\Delta}\boldsymbol{x}/2 - c \in \mathcal{K}^*_{\mathcal{C}}$ *where* $\boldsymbol{T}_0 = \left[\boldsymbol{\Delta}(\boldsymbol{\alpha}_0, \boldsymbol{A}) + \boldsymbol{\Gamma}^{-1}\right]^{-1}$.

Proof. First we show that a necessary and sufficient condition for g_0 to solve the dual problem in (1.7) is

$$\int g_0 e^{g_0}\,dQ = 0, \qquad \int g e^{g_0}\,dQ \le 0, \quad \forall g \in \mathcal{C}^*.$$

The necessary part has been proved in Lemma 3.2 of [1]. To prove the sufficiency part, note that for $\forall g \in \mathcal{K}^*$, $0 < \delta < 1$, with $h^*(g) = \ln[\int e^g\,dQ]$,

$$h^*(g_0 + \delta(g - g_0)) \le h^*(g_0) + \delta[h^*(g) - h^*(g_0)], \quad \text{or,}$$
$$h^*(g) - h^*(g_0) \ge \frac{1}{\delta}[h^*(g_0 + \delta(g - g_0)) - h^*(g_0)].$$

Considering $\delta \to 0+$, we get $h^*(g) - h^*(g_0) \ge (d/d\delta)h^*(\delta g + (1 - \delta)g_0)|_{\delta=0} \ge 0$.

From Lemma 3.2, we get

$$\int \left[-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}_0\boldsymbol{x} - c_0\right]\frac{e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}_0\boldsymbol{x}-c_0}}{\int e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}_0\boldsymbol{x}-c_0}\,dQ}\,dQ = 0,$$
$$\int \left[-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}\boldsymbol{x} - c\right]\frac{e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}_0\boldsymbol{x}-c_0}}{\int e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}_0\boldsymbol{x}-c_0}\,dQ}\,dQ \le 0, \quad \forall \boldsymbol{\alpha} \text{ with } g(\boldsymbol{\alpha}) \in \mathcal{K}^*_{\mathcal{C}}.$$

After multiplying both of these equations by appropriate normalizing constants, we get

$$\int \left[-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}_0\boldsymbol{x} - c_0\right]\frac{e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{T}_0^{-1}\boldsymbol{x}}}{\int e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{T}_0^{-1}\boldsymbol{x}}d\boldsymbol{x}}\,d\boldsymbol{x} = 0,$$
$$\int \left[-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Delta}\boldsymbol{x} - c\right]\frac{e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{T}_0^{-1}\boldsymbol{x}}}{\int e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{T}_0^{-1}\boldsymbol{x}}d\boldsymbol{x}}\,d\boldsymbol{x} \le 0, \quad \forall \boldsymbol{\alpha}, g(\boldsymbol{\alpha}) \in \mathcal{K}^*_{\mathcal{C}}.$$

from which the conditions in (3.7) follow. $\square$

Corollary 3.5. *If $b_i = 0$, $\forall i$, then $c_0, c = 0$ and the conditions in (3.7) reduce to*

$$E_{\boldsymbol{T}_0}(\boldsymbol{x}'\boldsymbol{\Delta}_0\boldsymbol{x}) = 0, \qquad E_{\boldsymbol{T}_0}(\boldsymbol{x}'\boldsymbol{\Delta}\boldsymbol{x}) \ge 0. \quad \square \tag{3.8}$$

Lemma 3.6 addresses the applications discussed in Section 1. In part (b), for MTP_2 normal distributions with given constraint $\mathcal{C}$, it relates the conditional covariance structure of the solution ($\boldsymbol{T}^{-1}$) to the conditional covariance structure of the distribution we project from ($\boldsymbol{\Gamma}^{-1}$). Part (c) is useful when considering the nonincreasing covariances in longitudinal data analysis.

Lemma 3.6. (a) *For coordinates (i, j), the conditional covariance structure of $\boldsymbol{T}$ $(t^{i,j})$ is updated from that of $\boldsymbol{\Gamma}$ $(\gamma^{i,j})$ if and only if the pair (i, j) is used in the description of constraints $\mathcal{C}$.*
(b) *$\boldsymbol{Y} = (Y_1, \ldots, Y_p) \sim R = \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{T})$ be the I-projection of $\boldsymbol{U} = (U_1, \ldots, U_p) \sim Q = \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Gamma})$ onto $\mathcal{C} = \{P : E_P(X_iX_j) \geq 0, \forall i \neq j\}$, the class of all* MTP$_2$ *normal distributions. Then the conditional covariances are related as*

$$\text{cov}(Y_i, Y_j|Y_s, s \neq i, j) \geq \text{cov}(U_i, U_j|U_s, s \neq i, j), \quad \forall i \neq j.$$

(c) *The I-projection of $Q = \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Gamma})$ onto $\mathcal{C} = \{P : \sum_{i=1}^{k}\sum_{j=d+1}^{p} a_{i,j-d}\sigma_{ij} \geq 0\}, 1 \leq k \leq p, k+1 \leq d \leq p$, $\boldsymbol{A}$ is the $k \times (p-d)$ matrix of given constants, is $\boldsymbol{N}(\boldsymbol{0}, \boldsymbol{T})$, with the conditional covariance structure given by $\boldsymbol{T}^{-1} = \boldsymbol{\Delta} + \boldsymbol{\Gamma}^{-1}$, where $\boldsymbol{\Delta} = (\boldsymbol{0}, \boldsymbol{0}, -2\alpha^*\boldsymbol{A}/\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}/-2\alpha^*\boldsymbol{A}', \boldsymbol{0}, \boldsymbol{0})$, where α^* solves the related dual problem.*

Proof. (a) This follows by inspecting the proofs of Lemma 3.1 and Theorem 3.3.

(b) If $Q \in \mathcal{C}$, then the I-projection of Q onto $\mathcal{C}$ is itself. If $Q \notin \mathcal{C}$, then the solution is $\boldsymbol{N}(\boldsymbol{0}, (\boldsymbol{\Delta}_a + \boldsymbol{\Gamma}^{-1})^{-1})$, where $\boldsymbol{\alpha}^*$ solves $\inf_{\boldsymbol{\alpha}\geq\boldsymbol{0}} \left\{-\frac{1}{2}\ln\det\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A}) + \boldsymbol{I}_p\right)\right\}$ since $c(\boldsymbol{\alpha}, \boldsymbol{b}) = 0$. Using the identity (3.5), here $\boldsymbol{\Delta}(\boldsymbol{\alpha}, \boldsymbol{A})_{ij} = \alpha_{ij}, i \neq j, = 0, i = j$. Then the conditional covariance of the solution $(\boldsymbol{Y})$ is given by $\text{cov}(Y_i, Y_j|Y_s, s \neq i, j) = \gamma^{ij} + \alpha_{ij}, \forall i \neq j$. Since $\alpha_{ij} \geq 0, \forall i \neq j$, the result follows. □

(c) Since the dual space is $\mathcal{K}^*_{\mathcal{C}} = \{g(\alpha) : g(\alpha) = -2\alpha\sum_{i=1}^{k}\sum_{j=d+1}^{p} a_{i,j-d}x_ix_j\}$, the form of $\boldsymbol{\Delta}$ is as indicated. □

Lemma 3.6 shows that the coordinates involved in the constraints affect the same coordinates of the conditional covariance structure. However for the solution covariance matrix, the coordinates other than those involved in the definition of $\mathcal{C}$ are also affected.

In the presence of multiple constraints, i.e. when $\mathcal{C} = \cap_{\ell=1}^{k}\mathcal{C}_\ell$ where $\mathcal{C}_\ell = \{P : \sum_{i=1}^{c_\ell}\sum_{j=d_\ell+1}^{p} a_{ij}\sigma_{ij} \leq b_\ell\}, b_\ell < 0$, $1 \leq c_\ell \leq p, c_\ell + 1 \leq d_\ell \leq p$, the solution has a conditional covariance pattern of more complex nature, which is a sum of k dependence structures $(\boldsymbol{\Delta})$ shown in Lemma 3.6(c). □

In general, when the constraints are compatible and overlapping, an iterative algorithm would be needed to find a solution. However, the IPS algorithms (used for equality constraints) may *not* converge to the correct solution as shown in the following example. Thus some adjustment is necessary.

Example 3.7. We consider the I-projection of $\boldsymbol{X} \sim \boldsymbol{N}_p(0, \boldsymbol{\Gamma})$ onto $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2$ where $\mathcal{C}_1 = \{P : \sigma_{12} \geq \sigma_{13}\}, \mathcal{C}_2 = \{P : \sigma_{23} \geq \sigma_{13}\}$. The IPS algorithm produces the estimate of $\boldsymbol{T}^\dagger$ after 25 cycles. However, the correct solution is $\boldsymbol{T}^*$, obtained after one cycle by using Algorithm 3.8, where

$$\boldsymbol{\Gamma} = \begin{pmatrix} 7 & 1 & 5 \\ 1 & 8 & 2 \\ 5 & 2 & 4 \end{pmatrix}, \qquad \boldsymbol{T}^\dagger = \begin{pmatrix} 5.742 & 4.514 & 4.516 \\ 4.514 & 10.141 & 4.516 \\ 4.516 & 4.516 & 3.911 \end{pmatrix}, \qquad \boldsymbol{T}^* = \begin{pmatrix} 5 & 4 & 4 \\ 4 & 11.057 & 4.486 \\ 4 & 4.486 & 3.629 \end{pmatrix}. \quad \square$$

Before we describe the new algorithm for dependence models (1.3), we discuss the interplay between the primal and dual problems at the (n, i)th step. Suppose at the completion of the $(n, i-1)$th step, the solution is of the form (using Theorem 3.3),

$$\frac{dP_{n,i-1}}{dQ} = c_{n,i-1}e^{-\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x}/2},$$

or, $P_{n,i-1} = \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{T}_{n,i-1})$, $\boldsymbol{T}^{-1}_{n,i-1} = \boldsymbol{\Delta}_{n,i-1} + \boldsymbol{\Gamma}^{-1}$, $\boldsymbol{\Delta}_{n,i}$ be the matrix obtained from the identity (3.5) and $c_{n,i-1}$ is a normalizing constant. At the (n, i)th step, we consider the problem of finding the I-projection of $P_{n,i-1}$ onto $\mathcal{C}_i$. To begin, an adjustment is made by removing the effects of the ith constraint from the previous cycle. Thus we form S_{ni}, where

$$\frac{dS_{ni}}{dQ} = e^{-\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x}/2-\alpha_{n-1,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}. \tag{3.9}$$

Consider the I-projection of S_{ni} onto $\mathcal{C}_i$, given by $P_{n,i}$ so that

$$\frac{dP_{n,i}}{dS_{n,i}} = \frac{e^{\alpha_{n,i}\left(\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right)}}{\int e^{\alpha_{n,i}\left(\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right)}dS_{n,i}}, \tag{3.10}$$

or, $P_{n,i} = \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{T}_{n,i})$, $\boldsymbol{T}^{-1}_{n,i} = \boldsymbol{\Delta}_{n,i} + \boldsymbol{\Gamma}^{-1}$, where using (3.9) the dual problem is expressed as

$$\inf_{\delta\in\Re^+}\int e^{\delta\left(\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right)}dS_{ni} = \inf_{\delta\in\Re^+}\int e^{-\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x}/2-(\alpha_{n-1,i}-\delta)\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}dQ, \tag{3.11}$$

$\alpha_{n,i} = \max\{0, \delta_{n,i}\}$, $\delta_{n,i} = \delta$ solves

$$\int\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right]e^{\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x}-(\alpha_{n-1,i}-\delta)\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}dQ = 0. \tag{3.12}$$

Multiplying (3.9) and (3.10), we get

$$\frac{dP_{ni}}{dQ}(x) = c_{n,i}e^{-\boldsymbol{x}'\boldsymbol{\Delta}_{n,i}\boldsymbol{x}/2},$$

where $c_{n,i}$ is a normalizing constant, $\boldsymbol{\Delta}_{n,i} = ((\boldsymbol{\Delta}_{n,i})_{s,t})$, and for $1 \le j \le r$,

$$(\boldsymbol{\Delta}_{n,i})_{s,t} = \begin{cases} (\boldsymbol{\Delta}_{n,i-1})_{s,t} - (\alpha_{n-1,i} - \alpha_{n,i})a_{ij}, & \text{for } a(j) = s,\ b(j) = t, \\ (\boldsymbol{\Delta}_{n,i-1})_{s,t}, & \text{otherwise.} \end{cases} \tag{3.13}$$

Thus the new iterative algorithm can be stated as follows.

Algorithm 3.8. 1. Start with $n = 1$, $i = 1$ with $\boldsymbol{\Delta}_{1,0} = \boldsymbol{0}$. Set $\alpha_{0,i} = 0$, $\forall i$.
2. For nth cycle, ith step, find $\alpha_{n,i} = \max\{0, \delta_{n,i}\}$, where $\delta_{n,i}$ solves (3.12). Construct the current estimate $\boldsymbol{\Delta}_{n,i} = ((\boldsymbol{\Delta}_{n,i})_{s,t})$ as described in (3.13).
3. Replace i by $i + 1$ $(m + 1 = 1)$. Go to 2. Stop when all the constraints are approximately satisfied.

Remark 3.9. To prove that Algorithm 3.8 converges to the correct solution, we need to assume that each of the following three quantities in A, B, C below are finite (see [1] for details). While writing a computer program, these assumptions can be verified by calculating them in the program itself. It would take rather unusual set of circumstances for these assumptions to be violated, and we find it extremely difficult to construct examples of such behavior.

$$\begin{aligned}(\text{A})\ \sup_{n,i}\int dS_{n,i} &= \sup_{n,i}\int e^{-\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x}/2-\alpha_{n-1,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}dQ \\ &= \sup_{n,i} e^{\alpha_{n-1,i}b_i}(2\pi)^{p/2}\det(\boldsymbol{T}^{\dagger}_{n,i-1}),\end{aligned}$$

where $(\boldsymbol{T}^{\dagger}_{n,i-1})^{-1} = \boldsymbol{\Delta}^{\dagger}_{n,i-1} + \boldsymbol{\Gamma}^{-1}$, $\boldsymbol{\Delta}^{\dagger}_{n,i-1}$ is obtained from the identity $\boldsymbol{x}'\boldsymbol{\Delta}^{\dagger}_{n,i-1}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x} - \alpha_{n-1,i}\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}$.

$$\begin{aligned}(\text{B})\ \sup_{n,i}\int \ln\left(\frac{dS_{n,i}}{dQ}\right)dP_{n,i} &= \sup_{n,i}\int\left(\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x} - \alpha_{n-1,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right]\right)dP_{n,i} \\ &= \sup_{n,i} E_{\boldsymbol{T}_{n,i}}\left(\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x}\right) \\ &= \sup_{n,i}\operatorname{tr}\left(\boldsymbol{T}_{n,i}\left(\boldsymbol{T}^{-1}_{n,i-1} - \boldsymbol{\Gamma}^{-1}\right)\right),\end{aligned}$$

where $\boldsymbol{T}^{-1}_{n,i} = \boldsymbol{\Delta}_{n,i} + \boldsymbol{\Gamma}^{-1}$, since $\int\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right]dP_{n,i} = 0$, and, $E_{\boldsymbol{T}_{n,i}}\left(\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x}\right) = E_{\boldsymbol{T}_{n,i}}\left[\operatorname{tr}\left(\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x}\right)\right] = E_{\boldsymbol{T}_{n,i}}\left[\operatorname{tr}\left(\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\right)\right] = \operatorname{tr}\left[E_{\boldsymbol{T}_{n,i}}\left(\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\right)\right] = \operatorname{tr}\left(\boldsymbol{T}_{n,i}\boldsymbol{\Delta}_{n,i-1}\right) = \operatorname{tr}\left(\boldsymbol{T}_{n,i}\left(\boldsymbol{T}^{-1}_{n,i-1} - \boldsymbol{\Gamma}^{-1}\right)\right)$.

$$\begin{aligned}(\text{C})\ \sup_{n,i}\int e^{\sum_{i=1}^{t} y_{n,i}}dQ &= \sup_{n,i} e^{\sum\limits_{i=1}^{t}\alpha_{n,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}dS_{n,i} \\ &= \sup_{n,i} E_{\boldsymbol{T}_{n,i}}\left(\boldsymbol{x}'\boldsymbol{\Delta}_{n,i-1}\boldsymbol{x}\right) \\ &= \sup_{n,i}\operatorname{tr}\left(\boldsymbol{T}_{n,i}\left(\boldsymbol{T}^{-1}_{n,i-1} - \boldsymbol{\Gamma}^{-1}\right)\right). \quad \square\end{aligned}$$

When solving (2.1) under constraints (1.3), let the solution be $\boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{T}^*)$, where $\boldsymbol{T}^{*-1} = \boldsymbol{\Delta}^* + \boldsymbol{\Gamma}^{-1}$. The following theorem validates that Algorithm 3.8 converges correctly, and a sketch of its proof is given in the Appendix (for details see [1]).

Theorem 3.10. *The sequence of matrices $\boldsymbol{\Delta}_{n,i}$ produced by Algorithm 3.8 converges to the matrix $\boldsymbol{\Delta}^*$ as $n \to \infty$.* $\square$

Algorithm 2.4 can be seen as a special case of Algorithm 3.8 by appropriate choice of constants and with equality.

Finite termination. To discuss finite termination under Algorithm 3.8 for constraints (3.1), first recall that the coordinates in $\mathcal{C}_i$ are from the index set $\mathcal{I}_i$. We define the constraints $\mathcal{C}_i$ in (3.3) to be *decomposable* if $\mathcal{I}_i$'s are disjoint, $\forall i$.

Theorem 3.11. *When $\mathcal{C}_i$'s are decomposable, Algorithm 3.8 terminates in one cycle.*

Proof. At the first cycle, the ith step, we I-project $P_{1,i-1}$ onto $\mathcal{C}_i$ and the corresponding dual problem is

$$\begin{aligned}\inf_{\alpha_i\in\Re^+}\int e^{\alpha_i\left(\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right)}dP_{1,i-1}(\boldsymbol{x}) &= \inf_{\alpha_i\in\Re^+}\int e^{\alpha_i\left(\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right)}d(P_{1,i-1}(\boldsymbol{x}))_{\mathcal{I}_i} \\ &= \inf_{\alpha_i\in\Re^+}\int e^{\alpha_i\left(\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right)}d(Q(\boldsymbol{x}))_{\mathcal{I}_i},\end{aligned}$$

as the integral only involves the variables with coordinates from $\mathcal{I}_i$. The solution is given by $\alpha_{1i} = \max\{0, \delta_{1i}\}$, $\delta = \delta_i$ solves

$$\int \left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right] e^{\delta\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right]}\, d(Q(\boldsymbol{x}))_{\mathcal{I}_i} = 0. \tag{3.14}$$

At the ith step, the solution is

$$\frac{dP_{1,i}}{dQ} = c_{1,i}e^{\sum_{\ell=1}^{i}\alpha_{1\ell}\left[\boldsymbol{x}'_{\mathcal{G}_\ell}\boldsymbol{A}_\ell\boldsymbol{x}_{\mathcal{H}_\ell}\right]} = c_{1,i}e^{-\boldsymbol{x}'\boldsymbol{\Delta}_{1,i}^{-1}\boldsymbol{x}/2},$$

where $c_{1,i-1}$ is a normalizing constant, and $\boldsymbol{\Delta}_{1,i}^{-1}$ is obtained from the identity $\sum_{\ell=1}^{i}\alpha_{1\ell}\left[\boldsymbol{x}'_{\mathcal{G}_\ell}\boldsymbol{A}_\ell\boldsymbol{x}_{\mathcal{H}_\ell}\right] = -\boldsymbol{x}'\boldsymbol{\Delta}_{1,i}^{-1}\boldsymbol{x}/2$. Using induction, at the mth step, the solution is

$$\frac{dP_{1,m}}{dQ} = c_{1,m}e^{-\sum_{\ell=1}^{m}\alpha_{1\ell}\left[\boldsymbol{x}'_{\mathcal{G}_\ell}\boldsymbol{A}_\ell\boldsymbol{x}_{\mathcal{H}_\ell}\right]} = c_{1,m}e^{-\boldsymbol{x}'\boldsymbol{\Delta}_{1,m}^{-1}\boldsymbol{x}/2},$$

where $c_{1,m}$ is a normalizing constant. To begin the second cycle, we first form $dS_{21}/dQ = e^{-\sum_{\ell=2}^{m}\alpha_{1\ell}\left[\boldsymbol{x}'_{\mathcal{G}_\ell}\boldsymbol{A}_\ell\boldsymbol{x}_{\mathcal{H}_\ell}\right]}$, after removing the effect of $\mathcal{C}_1$ from the first cycle. The corresponding dual problem is

$$\begin{aligned}\inf_{\alpha_1\in\Re^+}\int e^{\alpha_1\left(\boldsymbol{x}'_{\mathcal{G}_1}\boldsymbol{A}_1\boldsymbol{x}_{\mathcal{H}_1}-b_1\right)}dS_{2,1}(\boldsymbol{x}) &= \inf_{\alpha_1\in\Re^+}\int e^{\alpha_1\left[\boldsymbol{x}'_{\mathcal{G}_1}\boldsymbol{A}_1\boldsymbol{x}_{\mathcal{H}_1}-b_1\right]+\sum_{\ell=2}^{m}\alpha_{1\ell}\left[\boldsymbol{x}'_{\mathcal{G}_\ell}\boldsymbol{A}_\ell\boldsymbol{x}_{\mathcal{H}_\ell}-b_\ell\right]}dQ(\boldsymbol{x})\\ &= \inf_{\alpha_1\in\Re^+}\int e^{\alpha_1\left(\boldsymbol{x}'_{\mathcal{G}_1}\boldsymbol{A}_1\boldsymbol{x}_{\mathcal{H}_1}-b_1\right)}d(Q(\boldsymbol{x}))_{\mathcal{I}_1},\end{aligned}$$

the solution of which is α_{11}, and hence we get $P_{21} = P_{11}$. This continues for $\ell \geq 2$, and this proves the result. □

4. Empirical entropy estimation

Often it may not be appropriate to assume that the population is multivariate normal. For such cases, we consider an approach based on empirical entropy [21], where an estimate of the underlying distribution is obtained by maximizing a nonparametric entropy under the desired constraints.

We associate a weight w_k with the kth sample observation $\boldsymbol{Y}^{(k)}$, $k = 1, \ldots, n$. As our interest is on $\boldsymbol{\Sigma}$, we assumed here $\boldsymbol{\mu} = \boldsymbol{0}$. Thus we solve the problem

$$\max -\sum_{k=1}^{n} nw_k \ln w_k \quad \left[\text{or equivalently, } \min \sum_{k=1}^{n} nw_k \ln\left(\frac{w_k}{1/n}\right)\right] \tag{4.1}$$

subject to

$$w_k \geq 0, \quad \forall k, \sum_{k=1}^{n} w_k = 1,$$

$$\sum_{i,j\in\mathcal{I}_\ell} a_{ij}^{\ell}\left[\sum_{k=1}^{n} w_k(Y_i^{(k)}-\mu_i)(Y_j^{(k)}-\mu_j)\right] \geq b_\ell, \quad 1 \leq \ell \leq m. \tag{4.2}$$

Writing (4.2) as $\mathcal{K}_\ell = \sum_{k=1}^{n} g_{\ell k}w_k \geq 0$, where

$$g_{\ell k} = \sum_{i,j\in\mathcal{I}_\ell} a_{ij}^{\ell}(Y_i^{(k)}-\mu_i)(Y_j^{(k)}-\mu_j) - b_\ell, \quad 1 \leq \ell \leq m, \tag{4.3}$$

the problem (4.1) reduces to a special case of [3]. In this case, the dual problem is equivalent to

$$\begin{aligned}\inf_{y\in\left(\bigcap_{i=1}^{m}\mathcal{K}_i\right)^*}\sum_{i=1}^{n}\frac{1}{n}e^{y} &= \inf_{y\in(\mathcal{K}_1^*+\cdots+\mathcal{K}_m^*)}\sum_{i=1}^{n}\frac{1}{n}e^{y}\\ &= \inf_{y_\ell\in\mathcal{K}_\ell^*,1\leq\ell\leq m}\sum_{i=1}^{n}\frac{1}{n}e^{y_1(k)+\cdots+y_m(k)},\end{aligned}$$

where $\mathcal{K}_\ell^* = \{\alpha_\ell g_{\ell k} : \alpha_\ell \geq 0\}$. The following algorithm can be used to solve this problem.

Algorithm 4.1. 1. Begin with $n = 1$, $i = 1$, $\alpha_{0i} = 0$, $\forall i$.
2. At nth cycle, ith step, let $\alpha = \alpha^*$ solve

$$\inf_{\alpha} \sum_{k=1}^{n} g_{\ell k} e^{\sum_{\ell=1}^{i-1} \alpha_{n,\ell} g_{\ell k} + \alpha g_{ik} + \sum_{\ell=i+1}^{m} \alpha_{n-1,\ell} g_{\ell k}},$$

and let $\alpha_{ni} = \max\{0, \alpha^*\}$.

3. If $i < m$, replace i by $i+1$. When $i = m$, replace n with $n+1$, and set $i = 1$. Go to step 2 and continue until all constraints are approximately satisfied.

5. Simulation studies and example

5.1. Maximum likelihood and generalized least squares methods

Lee [18] considered covariance structure analysis under inequality constraints of parameters by using a penalty function approach as follows. Let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, $\boldsymbol{\theta}$ is $q \times 1$ which satisfies $r \le q$ inequality constraints $h_i(\boldsymbol{\theta}) \ge 0$, $1 \le i \le r$. Assuming a sample from the multivariate normal population with covariance matrix $\boldsymbol{\Sigma}$, the constrained MLE (provided it exists) is $\hat{\boldsymbol{\theta}}$ that satisfies $h_i(\hat{\boldsymbol{\theta}}) \ge 0$, $\forall i$ and minimizes

$$F(\boldsymbol{\theta}) = \log|\boldsymbol{\Sigma}| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log|\mathbf{S}| - p,$$

where $\mathbf{S}$ is the sample covariance matrix. For $k \ge 1$, by differentiating the maximum likelihood penalty function $F_k(\boldsymbol{\theta}) = F(\boldsymbol{\theta}) + c_k \sum_{t=1}^{r}[-\ln(h_t(\boldsymbol{\theta}))]$ with respect to θ_i one finds

$$\dot{F}_k(i) = \mathrm{tr}\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}_i\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma} - \mathbf{S}) - c_k \sum_{t=1}^{r} \frac{\dot{h}_t(i)}{h_t(\boldsymbol{\theta})},$$

where $\dot{\boldsymbol{\Sigma}}_i = \partial\boldsymbol{\Sigma}/\partial\theta_i$, $\dot{h}_t(i) = \partial h_t/\partial\theta_i$, and c_k is a decreasing sequence of positive real numbers. The iteration uses scoring algorithm, which at the kth step replaces $\boldsymbol{\theta}$ by $\boldsymbol{\theta} + \Delta\boldsymbol{\theta}$, where $\Delta\boldsymbol{\theta} = -\alpha \boldsymbol{I}_k^{-1}\dot{\boldsymbol{F}}_k$, α is a stepsize parameter, and $\boldsymbol{I}_k$ is the information matrix with (i, j) entry given by

$$I_k(i, j) = \mathrm{tr}\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}_i\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}_j + c_k \sum_{t=1}^{r} \frac{\dot{h}_t(i)\dot{h}_t(j)}{h_t(\boldsymbol{\theta})^2}.$$

The constrained generalized least squares estimator (GLE) of $\boldsymbol{\theta}$ (provided it exists) is $\tilde{\boldsymbol{\theta}}$ that satisfies $h_i(\tilde{\boldsymbol{\theta}}) \ge 0$, $\forall i$ and minimizes

$$Q(\boldsymbol{\theta}) = \frac{1}{2}\mathrm{tr}[(\mathbf{S} - \boldsymbol{\Sigma})\boldsymbol{V}]^2$$

where $\boldsymbol{V}$ is a weight matrix that converges to $\boldsymbol{\Sigma}^{-1}$ in probability. Here we take $\boldsymbol{V} = \boldsymbol{\Sigma}^{-1}$. In the kth iteration, one replaces $\boldsymbol{\theta}$ by $\boldsymbol{\theta} + \Delta\boldsymbol{\theta}$, where $\Delta\boldsymbol{\theta} = -\alpha U_k^{-1}\dot{Q}_k$, where

$$U_k(i, j) = \mathrm{tr}\mathbf{S}^{-1}\dot{\boldsymbol{\Sigma}}_i\mathbf{S}^{-1}\dot{\boldsymbol{\Sigma}}_j + c_k \sum_{t=1}^{r} \frac{\dot{h}_t(i)\dot{h}_t(j)}{h_t(\boldsymbol{\theta})^2},$$

$$\dot{Q}_k(i) = \mathrm{tr}\mathbf{S}^{-1}\dot{\boldsymbol{\Sigma}}_i\mathbf{S}^{-1}(\boldsymbol{\Sigma} - \mathbf{S}) - c_k \sum_{t=1}^{r} \frac{\dot{h}_t(i)}{h_t(\boldsymbol{\theta})}.$$

If the root mean squares of $\Delta\boldsymbol{\theta}$ and the gradient vector are small enough, the process is terminated; otherwise c_k is replaced by $c_k/2$ and the unconstrained minimization is repeated.

5.2. Simulation: comparing estimators

As the MLE ($\hat{\boldsymbol{\Sigma}}_M$) and the I-projection estimator ($\hat{\boldsymbol{\Sigma}}_I$, obtained from Algorithm 3.8) are based on the normality assumption, whereas the generalized least squares ($\hat{\boldsymbol{\Sigma}}_G$) and empirical entropy estimators ($\hat{\boldsymbol{\Sigma}}_E$, obtained from Algorithm 4.1) are not, one would like to compare the performances of all four estimators under normality and non-normality. Our simulation follows the excellent choices of symmetric and skewed distributions of [5]. We simulated $M = 10\,000$ samples with sizes $n = 10, 20, 30, 40, 50, 60, 80, 100$ from a multivariate normal distribution, a multivariate t distribution with 5 degrees of freedom (t_5), a noncentral distribution with 5 degrees of freedom (nct_5), and a standardized multivariate lognormal distribution. For each of these distributions, parameter values are so chosen that the resulting covariance matrix is given by

$$\boldsymbol{\Sigma} = (\sigma_{ij}) = \begin{pmatrix} 1 & 0.6 & 0.3 & 0.1 \\ 0.6 & 1 & 0.4 & 0.2 \\ 0.3 & 0.4 & 1 & 0.2 \\ 0.1 & 0.2 & 0.2 & 1 \end{pmatrix}.$$

The constraints of interest are (1) $\sigma_{12} \ge \sigma_{13}$, (2) $\sigma_{13} \ge \sigma_{14}$, (3) $\sigma_{23} \ge \sigma_{24}$. To obtain $\hat{\boldsymbol{\Sigma}}_I$, we I-project from $Q = \boldsymbol{N}_4(\mathbf{0}, \boldsymbol{\Gamma})$, $\boldsymbol{\Gamma} = \mathbf{S}$ onto these constraint regions.

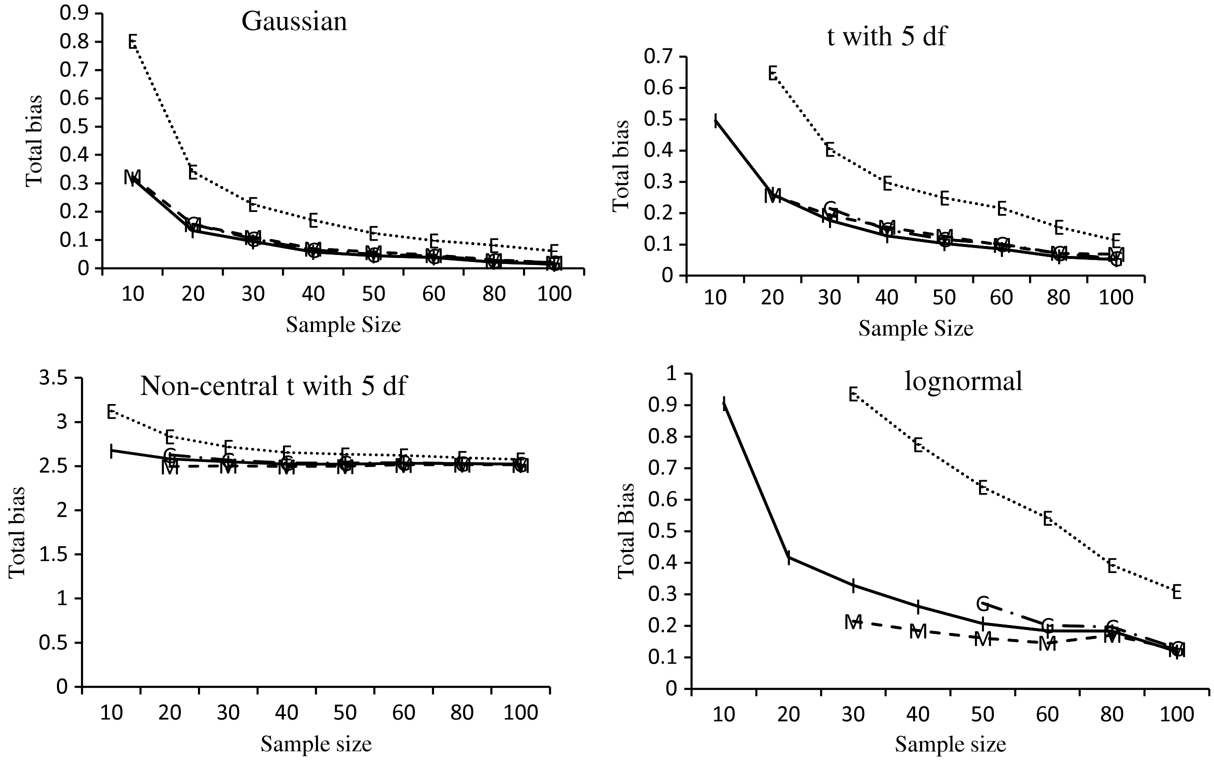

Fig. 1. Simulated total biases of MLE (M: dashed line), GLE (G: dash-dot-dashed line), I-projection (I: solid line), empirical entropy (E: dotted line) for different distributions at different sample sizes.

For the multivariate normal we used mean zero and the above covariance matrix $\mathbf{\Sigma}$. For the t_5 distribution, we chose mean zero and dispersion matrix $\boldsymbol{D} = (3/5)\mathbf{\Sigma}$, which yield covariance matrix $\mathbf{\Sigma}$. For noncentral t_5, samples were taken as $\boldsymbol{Y} = \sqrt{5}\boldsymbol{Z}/s$, where $\boldsymbol{Z} \sim \boldsymbol{N}_4(\boldsymbol{\mu}, \boldsymbol{D})$, $\boldsymbol{D} = (3/5)\mathbf{\Sigma} - [1 - (8/3\pi)]\boldsymbol{\mu}\boldsymbol{\mu}'$, $\boldsymbol{\mu} = (1, 1, 1, 1)'$, $s^2 \sim \chi^2_5$. The multivariate lognormal distribution is obtained as $\boldsymbol{Y} = (e^{\boldsymbol{Z}} - \sqrt{e})/\sqrt{e(e-1)}$, where $\boldsymbol{Z} \sim \boldsymbol{N}_4(\mathbf{0}, \boldsymbol{D})$ with entries of $\boldsymbol{D}$ as $d_{ij} = \log\{1 + (e-1)\sigma_{ij}\}$.

Figs. 1 and 2 present the simulation results on total bias and total root mean square error for the four estimation methods, where

$$\text{bias}(n) = \frac{1}{M}\sum_{i \geq j}\left|\sum_{t=1}^{M}(\hat{\sigma}_{ij}^{(n)} - \sigma_{ij})\right|, \qquad \text{RMSE}(n) = \sqrt{\sum_{i \geq j}\frac{1}{M}\sum_{t=1}^{M}(\hat{\sigma}_{ij}^{(n)} - \sigma_{ij})^2},$$

where $\hat{\sigma}_{ij}^{(n)}$ is obtained from each method separately.

For smaller sample sizes, we experienced difficulties with the MLE, GLE procedures, resulting from our inability to find suitable starting values and hence inverting the information matrices. We also faced difficulties with the empirical entropy procedure at smaller sample sizes for some nonnormal distributions. In Fig. 1, the bias of $\hat{\mathbf{\Sigma}}_E$ is larger than all other methods, substantially for lognormal. Note that only $\hat{\mathbf{\Sigma}}_I$ is available at all sample sizes, and its bias is not farther from the lowest available. The RMSEs in Fig. 2 of all four methods did not vary as much as their biases in Fig. 1, often the empirical entropy method performing best. Under normality, all four methods performed very close to one another, but under nonnormality, MLE is outperformed by others. The $\hat{\mathbf{\Sigma}}_G$ is only available for $n \geq 50$ under lognormal. The estimator $\hat{\mathbf{\Sigma}}_I$ is available at all sample sizes considered and performed close to the best in this simulation.

5.3. Simulation: speed of convergence

When the constraints are *not* decomposable, some information about the speed of convergence is needed before one can be comfortable using the procedure. A modest simulation study was run to determine the number of cycles needed for convergence for the cases of $p = 3, 4$. Out of a total of five models as shown in Table 1, we have chosen three with patterned covariance structures as in an AR(1) model with $\rho = 0.1, 0.5, 0.9$, and two are not AR(1) but decreasing (positive) covariances when away from the diagonal.

The order restrictions of interest are: for $p = 3$, $\{\sigma_{12} \geq \sigma_{13}, \sigma_{23} \geq \sigma_{13}\}$, and for $p = 4$, $\{\sigma_{12} \geq \sigma_{13} \geq \sigma_{14}, \sigma_{23} \geq \sigma_{24}, \sigma_{23} \geq \sigma_{13}, \sigma_{34} \geq \sigma_{24} \geq \sigma_{14}\}$, that is, $m = 2$ and $m = 6$ respectively. Three sample sizes 5, 10, 15 were studied for each model. At each combination of sample size and covariance matrix, 1000 independent random samples from multivariate normal distribution were generated and Algorithm 3.8 was used to estimate the covariance matrix subject to the order restrictions. The algorithm was considered to have converged when the I-divergence between two successive iterates is less than 0.001.

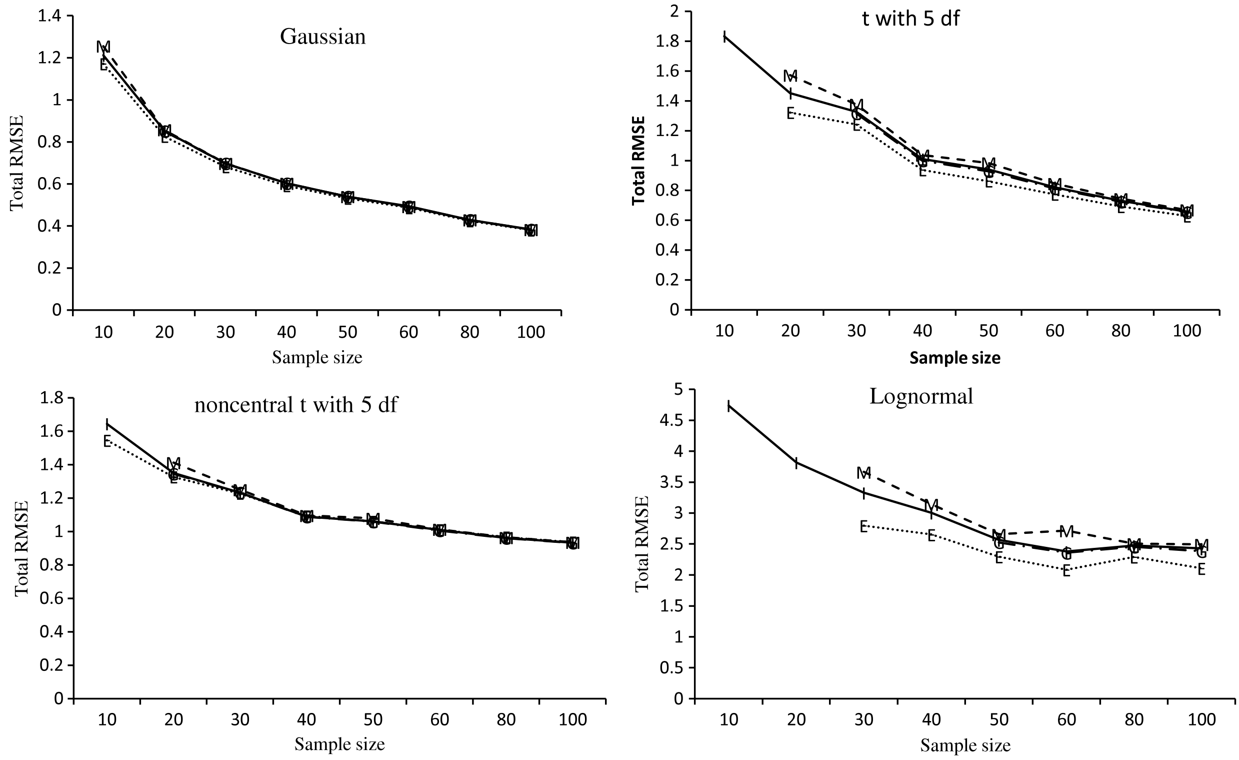

Fig. 2. Simulated total root mean squared errors of MLE (M: dashed line), GLE (G: dash-dot-dashed line), I-projection (I: solid line), empirical entropy (E: dotted line) for different distributions at different sample sizes.

Table 1
Covariance patterns used in the simulation study for Table 2.

p	Covariance patterns				
	1	2	3	4	5
3	$\sigma_{ij} = 0.1^{\lvert i-j \rvert}$	$\sigma_{ij} = 0.5^{\lvert i-j \rvert}$	$\sigma_{ij} = 0.9^{\lvert i-j \rvert}$	$\begin{pmatrix} 1 & 0.3 & 0.2 \\ 0.3 & 1 & 0.5 \\ 0.2 & 0.3 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.7 & 0.67 \\ 0.7 & 1 & 0.7 \\ 0.67 & 0.7 & 1 \end{pmatrix}$
4	$\sigma_{ij} = 0.1^{\lvert i-j \rvert}$	$\sigma_{ij} = 0.5^{\lvert i-j \rvert}$	$\sigma_{ij} = 0.9^{\lvert i-j \rvert}$	$\begin{pmatrix} 1 & 0.3 & 0.2 & 0.1 \\ 0.3 & 1 & 0.3 & 0.2 \\ 0.2 & 0.3 & 1 & 0.3 \\ 0.1 & 0.2 & 0.3 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.3 & 0.2 & 0.18 \\ 0.3 & 1 & 0.5 & 0.48 \\ 0.2 & 0.3 & 1 & 0.52 \\ 0.18 & 0.48 & 0.52 & 1 \end{pmatrix}$

The results are summarized in Table 2. When $\mathbf{S}$ satisfies the order restrictions, it is the restricted estimate. However, as can be seen in most cases the number of times that $\mathbf{S}$ does not conform to the order restrictions is quite large, more so for $p = 4$. The results show that Algorithm 3.8 generally converges in a small number of cycles, being faster for $p = 3$ than for $p = 4$. For both cases of p, the convergence is faster for larger sample sizes. The last column shows that when the estimators disagree, the restricted estimate is closer to the true $\boldsymbol{\Sigma}$ than $\mathbf{S}$ is, most of the time. The distribution of the number of cycles is such that the convergence is attained in five or less number of cycles for the $p = 3$ case, and eight or less number of cycles for the $p = 4$ case. Thus Algorithm 3.8 is a viable estimation procedure for estimating dependent covariance matrices in the cases studied.

5.4. Example

The human immune deficiency virus (HIV) causes AIDS by reducing a person's ability to fight infection. HIV attacks the CD4+ cells which orchestrate the body's immunoresponse to infectious agents. CD4+ cells decrease in number with time from infection so that an infected person's CD4+ cell number can be used to monitor disease progression. Diggle et al. [9] observed that there is substantial positive correlation on data that are one year apart, and the degree of correlation goes down as one moves farther from another in time. Correlation depends more strongly on the time between observations than on their absolute values.

Diggle et al. [9] suggested that assuming stationarity, a single correlation (covariance) estimate can be obtained for each distinct value of the time separation or lag (k) by considering averages along the diagonal ($\overline{\sigma}_k = \sum_{j=1}^{7-k} \sigma_{j,j+k}/(7-k)$, $k = 1, \ldots, 6$). We like to estimate the covariance matrix subject to the restriction that these averages follow the

Table 2
Results from the simulation study of convergence rates for the AR(1) and other models (1000 independent trials for each n = sample size, $\mathbf{\Sigma}$ = covariance pattern 1 to 5 combinations); see Table 1 for covariance patterns.

n	$\mathbf{\Sigma}$	3 × 3						4 × 4						
		# times **S** does not conform	No. of cycles to converge				# times $\mathbf{\Sigma}^*$ closer than **S**	# times **S** does not conform	No. of cycles to converge					# times $\mathbf{\Sigma}^*$ closer than **S**
			1	≤2	≤3	≥4			1	≤2	≤3	≤4	≤5	
5	1	624	28	504	613	11	623	974	31	300	612	756	218	956
5	2	469	31	403	461	8	469	937	90	487	715	809	128	890
5	3	526	36	476	524	2	526	927	77	475	718	814	113	902
5	4	607	27	494	591	16	607	971	38	358	656	781	190	948
5	5	642	19	517	630	12	642	986	53	326	623	767	219	920
10	1	603	40	483	599	4	603	973	47	419	806	917	56	956
10	2	368	64	335	366	2	364	875	160	621	821	864	11	788
10	3	468	61	440	468	0	468	859	152	655	803	845	14	806
10	4	582	45	482	576	6	582	949	62	492	828	918	31	921
10	5	637	55	531	630	7	636	984	53	458	839	945	39	921
15	1	591	74	515	589	2	591	972	63	525	862	958	14	944
15	2	288	55	274	288	0	288	848	201	706	840	848	0	750
15	3	378	59	362	378	0	378	828	198	720	813	828	0	746
15	4	555	73	500	553	2	554	951	84	610	892	942	9	915
15	5	621	69	538	618	3	621	983	56	492	901	969	14	912

Table 3
Estimates of covariance matrix for CD4+ residuals.

$t_{ik} \setminus t_{ij}$	Observed							Restricted						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	1	0.66	0.56	0.41	0.29	0.48	0.89	0.675	0.758	0.584	0.458	0.383	0.590	0.625
2	0.66	1	0.49	0.47	0.39	0.52	0.48	0.758	1.405	0.773	0.797	0.769	0.985	0.661
3	0.56	0.49	1	0.51	0.51	0.51	0.44	0.584	0.773	1.188	0.710	0.736	0.797	0.508
4	0.41	0.47	0.51	1	0.68	0.65	0.61	0.458	0.797	0.710	1.230	0.943	0.960	0.689
5	0.29	0.39	0.51	0.68	1	0.75	0.7	0.383	0.769	0.736	0.943	1.298	1.091	0.813
6	0.48	0.52	0.51	0.65	0.75	1	0.75	0.590	0.985	0.797	0.960	1.091	1.413	0.886
7	0.89	0.48	0.44	0.61	0.7	0.75	1	0.625	0.661	0.508	0.689	0.813	0.886	0.769

Table 4
Estimates of average covariances at lags (k) for CD4+ residuals.

k	1	2	3	4	5	6
$\hat{\bar{\sigma}}_k$	0.64	0.578	0.48	0.42	0.48	0.89
$\hat{\bar{\sigma}}^*_k$	0.860	0.778	0.678	0.625	0.625	0.625

nonincreasing order (conjecture of Diggle et al. [9])

$$\bar{\sigma}_1 \geq \bar{\sigma}_2 \geq \cdots \geq \bar{\sigma}_6. \tag{5.1}$$

The estimated correlation matrix for the CD4+ data is presented in the left panel of Table 3. The correlations show some tendency to decrease with time, but remain substantial at all lags. The observed values ($\hat{\bar{\sigma}}_k$) are reported in Table 4. However, these observed averages do not follow (5.1), possibly due to sampling variability. To estimate the covariances subject to (5.1), we apply Algorithm 3.8 on the observed covariance matrix. Here $p = 7$ and $m = 5$. The algorithm converges in two cycles. The restricted estimates of covariances are given in the right panel of Table 3 and the corresponding averages along the diagonal are in Table 4. It is interesting to see in the final solution that not only the covariances which were violating the restrictions are changed but also the ones which were satisfying the restrictions.

6. Final comments

In this paper, we have considered constrained estimation of the covariance matrix using Fenchel duality tools. Motivated by the simplicity of the dual solutions for the covariance selection, we have investigated the duality approach in the general case of multivariate dependence of (1.3). The duality approach has produced characterizations of multivariate normal distributions in terms of its conditional covariance structures. Simulations show that two estimators from the suggested method compare favorably with existing ones.

Acknowledgments

The author thanks two referees and the associate editor whose comments have helped to improve the paper. The work of this paper is supported partly by a grant from the National Science Foundation (DMS 1007528).

Appendix

Proof of Theorem 3.10. Here

$$\frac{dP_{nm}}{dQ} = \frac{e^{\sum_{i=1}^m \alpha_{n,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}}{\int e^{\sum_{i=1}^m \alpha_{n,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}\,dQ}, \qquad y_{n,i} = \alpha_{n,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right].$$

Since,

$$\begin{aligned} I(P_{ni}|S_{ni}) &= \inf_{P\in\mathcal{C}_i} I(P|S_{ni}) \\ &= -\inf_{y\in\mathcal{K}^*_{\mathcal{C}_i}} \ln\int e^y dS_{ni} = -\inf_{y\in\mathcal{K}^*_{\mathcal{C}_i}} \ln\int e^{y+y_{n,i-1}-y_{n-1,i}} dS_{n,i-1} \\ &\ge -\ln\int e^{y_{n,i-1}} dS_{n,i-1} = -\inf_{y\in\mathcal{K}^*_{\mathcal{C}_{i-1}}} \ln\int e^y dS_{n,i-1} = I(P_{n,i-1}|S_{n,i-1}) \end{aligned}$$

so $I(P_{ni}|S_{ni})$ is nondecreasing, which is also bounded above. Similarly for $i = 1$. Hence $I(P_{n,i}|P_{n,i-1}) \to 0$ as $n \to \infty$, for each i. Using this fact and proceeding as in [1] we find that Theorem 3.3 implies

$$\text{(A1)}\quad \lim_{n\to\infty}\sum_{i=1}^m \alpha_{n,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right] dP_{n,m} = \lim_{n\to\infty}\left(h\left(\frac{dP_{n,m}}{dQ}\right) + h^*\left(\sum_{i=1}^m y_{n,i}\right)\right) = 0.$$

Since using assumption C, $e^{\sum_{i=1}^m y_{n,i}} = e^{\sum_{i=1}^m \alpha_{n,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}$ are uniformly integrable, given any sequence of increasing numbers, there is a subsequence $\{n_j\}$ and α_i^* such that $\alpha_{n_j,i} \to \alpha_i^*$ as $\{n_j\} \to \infty$. As $\alpha_{n_j,i} \ge 0$, it follows that $\alpha_i^* \ge 0$, $\forall i$. Then

$$\sum_{i=1}^m \alpha_{n_j,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right] \to \sum_{i=1}^m \alpha_i^*\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right] = -\boldsymbol{x}'\boldsymbol{\Delta}^{*-1}\boldsymbol{x}/2 - c = y_0,$$

say, as $j \to \infty$. Then

$$\text{(A2)}\quad \frac{dP_{n_j m}}{dQ} = \frac{e^{\sum_{i=1}^m \alpha_{n_j,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}}{\int e^{\sum_{i=1}^m \alpha_{n_j,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}\,dQ} \to \frac{x_0}{\int x_0\,dQ},$$

where $x_0 = e^{y_0}$ as $j \to \infty$. Since f is convex and $\mathcal{C}$ is a closed, convex cone, the solution to $\inf_{x\in\mathcal{C}} f(x)$ exists uniquely. So if the sequence $dP_{n,m}/dQ$ has many limit points, they must be equal (a.e. Q).

Using lower-semicontinuity of h, (A1) and (A2), it follows that

$$h\left(\frac{x_0}{\int x_0\,dQ}\right) + h^*(y_0) \le \liminf_{n_j\to\infty}\left[h\left(\frac{e^{\sum_{i=1}^m \alpha_{n_j,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}}{\int e^{\sum_{i=1}^m \alpha_{n_j,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}-b_i\right]}\,dQ}\right) + h^*\left(\sum_{i=1}^m \alpha_{n_j,i}\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i} - b_i\right]\right)\right] = 0.$$

So $x_0/\int x_0\,dQ$, y_0 are (unique) solutions to the primal and dual problems, respectively. It also follows from [1] that $x_0 \in \cap_{i=1}^m \mathcal{K}_{\mathcal{C}_i}$ and $y_0 \in \bigoplus_{i=1}^m \mathcal{K}^*_{\mathcal{C}_i}$. Here $dP_{n,i}/dQ$ converges to $x_0 = dP^*/dQ$. Hence $\boldsymbol{\Delta}_{n,i} \to \boldsymbol{\Delta}^*$ where $\boldsymbol{\Delta}^*$ is obtained from the identity $-\boldsymbol{x}'\boldsymbol{\Delta}^{*-1}\boldsymbol{x}/2 = \sum_i \alpha_i^*\left[\boldsymbol{x}'_{\mathcal{G}_i}\boldsymbol{A}_i\boldsymbol{x}_{\mathcal{H}_i}\right]$. This completes the proof. □

References

[1] B. Bhattacharya, An iterative procedure for general probability measures to obtain I-projection onto intersection of convex sets, Annals of Statistics 34 (2006) 878–902.
[2] B. Bhattacharya, R. Dykstra, A general duality approach to I-projections, Journal of Statistical Planning and Inference 47 (1995) 203–216.
[3] B. Bhattacharya, R. Dykstra, A Fenchel duality aspect of iterative *I*-projection procedures, Annals of the Institute of Statistical Mathematics 49 (1997) 435–446.
[4] E. Bølviken, Probability inequalities for the multivariate normal with non-negative partial correlations, Scandinavian Journal of Statistics 9 (1982) 49–58.

[5] S. Chaudhuri, M. Drton, T.S. Richardson, Biometrika 94 (2007) 199–216.
[6] T.M. Cover, J. Thomas, Elements of Information Theory, Wiley, New York, 1991.
[7] I. Csiszár, I-divergence geometry of probability distributions and minimization problems, Annals of Probability 3 (1975) 146–159.
[8] A.P. Dempster, Covariance selection, Biometrics 28 (1972) 157–175.
[9] P. Diggle, P. Heagerty, K.-Y. Liang, S. Zeger, Analysis of Longitudinal Data, second ed., Oxford University Press, 2002.
[10] R. Dykstra, An iterative procedure for obtaining I-projections onto the intersection of convex sets, Annals of Probability 13 (1985) 975–984.
[11] R. Dykstra, J.E. Hewett, W.A. Thompson Jr., Events which are almost independent, Annals of Statistics 1 (1973) 674–681.
[12] G.M. Fitzmaurice, N.M. Laird, J.H. Ware, Applied Longitudinal Analysis, in: Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, NJ, 2004.
[13] S.J. Haberman, The Analysis of Frequency Data, University of Chicago Press, 1974.
[14] J.N. Kapur, Maximum Entropy Models in Engineering, Wiley, New York, 1989.
[15] S. Karlin, Y. Rinott, Classes of orderings of measures and related correlation inequality, I. Multivariate totally positive distributions, Journal of Multivariate Analysis 10 (1980) 467–498.
[16] S. Kullback, Information Theory and Statistics, Wiley, New York, 1959.
[17] S.L. Lauritzen, Graphical Models, Oxford University Press, 1996.
[18] S.-Y. Lee, Estimation of covariance structure models with parameters subject to functional restraints, Psychometrika 45 (1980) 309–324.
[19] D.G. Luenberger, Optimization by Vector Space Methods, Wiley, New York, 1969.
[20] H. Massam, E. Neher, Estimation and testing for lattice conditional independence models on Euclidean Jordan algebras, Annals of Statistics 26 (1998) 1051–1082.
[21] A.B. Owen, Empirical Likelihood, Chapman and Hall, Boca Raton, FL, 2001.
[22] C.R. Rao, Linear Statistical Inference and its Applications, Wiley, New York, 1973.
[23] Y. Rinott, M. Scarsini, Total positivity order and the normal distribution, Journal of Multivariate Analysis 97 (2006) 1251–1261.
[24] F.H. Shaw, C.J. Geyer, Estimation and testing in constrained covariance component models, Biometrika 84 (1997) 95–102.
[25] T.P. Speed, H.T. Kiiveri, Gaussian Markov distributions over finite graphs, Annals of Statistics 14 (1986) 138–150.
[26] N. Wermuth, E. Scheidt, Fitting a covariance selection model to a matrix, in: Algorithm AS105, Applied Statistics 26 (1977) 88–92.