International Journal of Approximate Reasoning 125 (2020) 240–250

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

On generating random Gaussian graphical models

Irene Córdoba [a,*], Gherardo Varando [b,*], Concha Bielza [a], Pedro Larrañaga [a]

[a] Universidad Politécnica de Madrid, Madrid, Spain
[b] University of Copenhagen, Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 29 April 2019
Received in revised form 31 May 2020
Accepted 28 July 2020
Available online 12 August 2020

Keywords:
Concentration graph
Covariance graph
Positive definite matrix simulation
Undirected graphical model
Algorithm validation

ABSTRACT

Structure learning methods for covariance and concentration graphs are often validated on synthetic models, usually obtained by randomly generating: (i) an undirected graph, and (ii) a compatible symmetric positive definite (SPD) matrix. In order to ensure positive definiteness in (ii), a dominant diagonal is usually imposed. In this work we investigate different methods to generate random symmetric positive definite matrices with undirected graphical constraints. We show that if the graph is chordal it is possible to sample uniformly from the set of correlation matrices compatible with the graph, while for general undirected graphs we rely on a partial orthogonalization method.

1. Introduction

Structure learning algorithms in graphical models are validated using either benchmark or randomly generated synthetic models from which data is sampled. This allows to evaluate their performance by comparing the recovered graph, obtained by running the algorithm over the generated data, with the known true structure. The synthetic graphical models are typically constructed in a two-step manner: a graph structure is selected at random or chosen so that it is representative of the problem at hand; and, similarly, its parameters are fixed or randomly sampled.

Covariance [6,14] and concentration graphs [9,17] are graphical models where the variables are assumed to follow a multivariate Gaussian distribution, and the structure is directly read off in the covariance or concentration matrix, respectively. Looking at the literature on these models, one finds that typical benchmark structures are Toeplitz, banded, diagonally spiked and block diagonal covariance or concentration matrices [34,33,18], with parameters fixed to ensure positive definiteness.

The issue of positive definiteness is especially relevant when the structure is randomly generated. One approach to ensure it is to sample from a matrix distribution with support over the symmetric positive definite matrices compatible with the undirected graph structure. The hyper Wishart distributions [8,20] are the most developed in this context, since they form a conjugate family for Bayesian analysis. However, while sampling algorithms are available for general concentration graphs [3,19], in covariance graphs they have been developed only in the decomposable case [15].

In general, hyper Wishart distributions are rarely used in validation scenarios [32], and instead in the literature the most common approach to ensure positive definiteness is to enforce diagonal dominance in the covariance or concentration matrix [22,1,28]. However, when the undirected graph is moderately dense, the off-diagonal elements in the generated

* Corresponding authors.
E-mail addresses: irene.cordoba@upm.es (I. Córdoba), gherardo.varando@math.ku.dk (G. Varando), mcbielza@fi.upm.es (C. Bielza), pedro.larranaga@fi.upm.es (P. Larrañaga).

https://doi.org/10.1016/j.ijar.2020.07.007
0888-613X/

matrices, often interpreted as link strengths, are extremely small with respect to the diagonal entries and structure recovery becomes a challenge, thereby compromising the structure learning algorithm validation [26,27,16,2].

In this paper, we propose alternative methods to generate positive definite matrices with undirected graphical constraints: the partial orthogonalization method proposed in [4], uniform sampling when the graph is chordal and a combination of uniform sampling and partial orthogonalization for general graphs. We show that the partial orthogonalization method could suffer from drawbacks similar to the diagonal dominance when the matrix factor is obtained with i.i.d. elements. For this reason we propose to combine uniform sampling for chordal graphs and the partial orthogonalization method.

We also use our simulation method in a validation setting and show how the performance ranking of the various structure learning algorithms changes dramatically, thereby modifying the conclusions drawn if only using diagonally dominant matrices for comparison.

The rest of the paper is organized as follows. Preliminaries are introduced in Section 2, where we briefly overview concentration, covariance graphs and directed graphical models. Next, in Section 3, we present the classical diagonal dominance method, a proposed partial orthogonalization method and the uniform sampling for chordal graphs. Section 4 contains a description of the experiment set-up we have considered, and the interpretation of the results obtained. Finally, in Section 5 we conclude the paper and outline our plans for future research.

2. Preliminaries

In the remainder of the paper, we will use the following notation. We let $X_1, \ldots, X_p$ denote p random variables and $\boldsymbol{X}$ the random vector they form. For each subset $I \subseteq \{1, \ldots, p\}$, $\boldsymbol{X}_I$ will be the subvector of $\boldsymbol{X}$ indexed by I, that is, $(X_i)_{i \in I}$. We follow [7] and abbreviate conditional independence in the joint distribution of $\boldsymbol{X}$ as $\boldsymbol{X}_I \perp\!\!\!\perp \boldsymbol{X}_J \mid \boldsymbol{X}_K$, meaning that $\boldsymbol{X}_I$ is conditionally independent of $\boldsymbol{X}_J$ given $\boldsymbol{X}_K$, with I, J, K pairwise disjoint subsets of indices. Entries in a matrix are denoted with the respective lower case letter, for example, m_{ij} denotes the (i, j) entry in matrix $\mathbf{M}$.

With $\mathbb{S}$ and $\mathbb{S}^p_{>0}$ we denote the sets of symmetric and symmetric positive definite matrices of dimension $p \times p$. We denote the set of symmetric positive definite matrices with unit diagonal as,

$$\mathcal{E}_p = \{\mathbf{M} \in \mathbb{S}^p_{>0} \text{ s.t. } m_{ii} = 1 \quad \text{for } i = 1, \ldots, p\}.$$

The set $\mathcal{E}_p$ is called the elliptope of dimension p [29] and its volume has been obtained by [12] and [21].

With $\mathcal{S}^k_+$ we denote the k-dimensional hemisphere with positive first coordinate,

$$\mathcal{S}^k_+ = \{\boldsymbol{v} \in \mathbb{R}^k \text{ s.t. } ||\boldsymbol{v}||_2 = 1 \text{ and } v_1 > 0\}.$$

We will also use $\mathcal{U}^p$ to denote the set of upper triangular matrices of dimension $p \times p$ with positive diagonal, that is, the sets of Cholesky factors for positive definite matrices. With $\mathcal{U}^p_1 \subset \mathcal{U}^p$ we denote the subset with unit rows, that is the Cholesky factors for correlation matrices.

2.1. Undirected Gaussian graphical models

Covariance and concentration graphs are graphical models where it is assumed that the statistical independences in the distribution of a multivariate Gaussian random vector $\boldsymbol{X} = (X_1, \ldots, X_p)$ can be represented by an undirected graph $G = (V, E)$. Typically, $\boldsymbol{X}$ is assumed to have zero mean for lighter notation, and $V = \{1, \ldots, p\}$ so that it indexes the random vector, that is, $\boldsymbol{X}_V = \boldsymbol{X}$. We will represent the edge set E as a subset of $V \times V$, therefore $(i, j) \in E$ if and only if $(j, i) \in E$.

In covariance graphs, the independences represented are marginal, meaning that whenever there is a missing edge (i, j) in G, the random variables X_i and X_j are marginally independent. More formally, this is called the pairwise Markov property of covariance graphs [6,14],

$$X_i \perp\!\!\!\perp X_j \quad \text{for } i, j \in V \text{ s.t. } i \not\sim_G j,$$

where $i \sim_G j$ is the adjacency relationship on the graph G, that is, $i \sim_G j$ if and only if $(i, j) \in E$. Note further that $X_i \perp\!\!\!\perp X_j$ if and only if $\sigma_{ij} = 0$.

By contrast, in concentration graphs, a missing edge implies a conditional independence; specifically, in this case the pairwise Markov property [17] becomes

$$X_i \perp\!\!\!\perp X_j \mid \boldsymbol{X}_{V \setminus \{i,j\}} \quad \text{for } i, j \in V \text{ s.t. } i \not\sim_G j.$$

In turn, this can be read off in the concentration matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$, that is, $X_i \perp\!\!\!\perp X_j \mid \boldsymbol{X}_{V \setminus \{i,j\}} \iff \omega_{ij} = 0$.

Therefore, the statistical independences implied by both covariance and concentration graph models are in correspondence with zero entries in a symmetric positive definite matrix. Thus, in the following we will focus on how to simulate such kind of matrices. For a fixed undirected graph G let $\mathcal{M}^p(G)$ be the set of matrices with zeros in the entries represented by the missing edges in G, that is,

$$\mathcal{M}^p(G) = \{\mathbf{M} \in \mathbb{R}^{p\times p} : m_{ij} = m_{ji} = 0 \text{ if } (i,j) \notin E\}.$$

Let $\mathbb{S}^p(G) = \mathbb{S}^p \cap \mathcal{M}^p(G)$ and $\mathbb{S}^p_{>0}(G) = \mathbb{S}^{>0} \cap \mathcal{M}^p(G)$ be the sets of symmetric and symmetric positive definite matrices with undirected graphical constraints. Similarly $\mathcal{E}_p(G) = \mathcal{E}_p \cap \mathcal{M}^p(G)$ is the set of correlation matrices with undirected graphical constraints.

Note that the covariance matrix $\mathbf{\Sigma}$ of a Gaussian random vector $\boldsymbol{X}$ whose distribution belongs to a covariance graph with structure G satisfies that $\mathbf{\Sigma} \in \mathbb{S}^p_{>0}(G)$. Analogously, if the distribution belongs to a concentration graph with structure G, then $\mathbf{\Omega} = \mathbf{\Sigma}^{-1} \in \mathbb{S}^p_{>0}(G)$. In either case it is clear that the goal is to simulate elements belonging to $\mathbb{S}^p_{>0}(G)$, or to $\mathcal{E}^p(G)$.

2.2. Cholesky factorization and directed graphical models

If $G = (V = \{1,\dots,p\}, E)$ is an acyclic directed graph and we assume that $1 \prec \cdots \prec p$ is a topological order, that is, $\text{pa}(i) \subseteq \{1,\dots,i-1\}$ for all $i \in V$, then we can define the ordered Markov property for the Bayesian network model,

$$X_i \perp\!\!\!\perp X_j | \boldsymbol{X}_{\text{pa}(i)} \quad \text{for all } i \in V,\ j \notin \text{pa}(i),\ j{<}i. \tag{1}$$

If the ordered Markov property holds for a Gaussian distribution it is equivalent to saying that the coefficient β_{ij} of variable X_j in the regression of X_i on $X_1,\dots,X_{i-1}$ is zero for all $j \notin \text{pa}(i)$. Therefore, the set of edges E in a Gaussian Bayesian network can be expressed as

$$E = \{(j,i) \text{ s.t. } \beta_{ij} \neq 0\}. \tag{2}$$

We can rewrite the above Markov property as a triangular regression system [30]. Specifically, for each $i \in V$, X_i can be written as a regression over its parents,

$$X_i = \sum_{j<i} \beta_{ij} X_j + \varepsilon_i = \sum_{j\in\text{pa}(i)} \beta_{ij} X_j + \varepsilon_i, \tag{3}$$

where $\varepsilon_1,\dots,\varepsilon_p$ is a vector of zero-mean independent Gaussian noise.

We can write Equation (3) in matrix notation as $\boldsymbol{X} = \mathbf{B}\boldsymbol{X} + \boldsymbol{\varepsilon}$, with $\mathbf{B}$ strictly lower triangular, since $1,\dots,p$ is assumed to be a topological order of G. Rearranging the equation we obtain $\boldsymbol{X} = (\mathbf{I}_p - \mathbf{B})^{-1}\boldsymbol{\varepsilon}$. Taking variances on both sides, we arrive at the upper Cholesky factorization of the precision matrix [24]

$$\mathbf{\Sigma}^{-1} = \mathbf{\Omega} = (\mathbf{I}_p - \mathbf{B})^t \mathbf{V}^{-1} (\mathbf{I}_p - \mathbf{B}) = \mathbf{U}\mathbf{U}^t, \tag{4}$$

where $\mathbf{U} = (\mathbf{I}_p - \mathbf{B})^t \sqrt{\mathbf{V}^{-1}} \in \mathcal{U}^p$ and $\mathbf{V}$ is a diagonal matrix with $v_{ii} = \text{var}(\varepsilon_i) = \text{var}(X_i | \boldsymbol{X}_{\text{pa}(i)})$.

The upper Cholesky factorization in Equation (4) is closely related to the classical/lower Cholesky factorization, as follows. Let $\tilde{\mathbf{\Omega}}$ be the matrix obtained from $\mathbf{\Omega}$ by reordering the variables so that they follow the reverse of a perfect/topological ordering, also known as *fill-in free* or *perfect elimination ordering* [see25, for example]. Then if $\tilde{\mathbf{\Omega}} = \mathbf{L}\mathbf{L}^t$ is its standard lower Cholesky decomposition, it can be verified that $\mathbf{L}^t$ is equal to the transpose of $\mathbf{U}$ (Equation (4)) with respect to its anti-diagonal. Furthermore, the parameters of the Gaussian Bayesian network are obtained from $\mathbf{U}$ [31] as

$$\beta_{ij} = -\frac{u_{ji}}{u_{ii}}; \qquad \text{var}(X_i | \boldsymbol{X}_{\text{pa}(i)}) = \frac{1}{u_{ii}^2}. \tag{5}$$

The upper Cholesky factorization in Equation (4) can be used as a parametrization of the inverse covariance matrix for Gaussian distributions satisfying the ordered Markov property, that is, Gaussian Bayesian networks: from Equation (5) we have that, for $j{<}i$,

$$(j,i) \notin E \iff X_i \perp\!\!\!\perp X_j | \boldsymbol{X}_{\text{pa}(i)} \iff \beta_{ij} = 0 \iff u_{ji} = 0. \tag{6}$$

The Gaussian Bayesian network model can thus be expressed as

$$\mathcal{B}(G) = \{\mathbf{\Omega} = \mathbf{\Sigma}^{-1} = \mathbf{U}\mathbf{U}^t \text{ s.t. } \mathbf{U} \in \mathcal{U}^p \text{ and } u_{ji} = 0 \text{ if } (j,i) \notin E\}, \tag{7}$$

where $G = (V, E)$ is an acyclic digraph with $1 \prec \cdots \prec p$ being a topological order of G.

2.3. Markov equivalence between Gaussian graphical models

The intersection between Markov and Bayesian network models occurs at what are called *decomposable*/*chordal*/*triangulated* undirected graphs, or, equivalently, acyclic digraphs with no *v-structures*. An undirected graph G is said to be chordal if all cycles of length at least 4 have a chord. A v-structure in an acyclic digraph G with edge set E, is a configuration where if $(i,j) \in E$, $(k,j) \in E$ and $i \neq k$, then $(i,k) \notin E$ and $(k,i) \notin E$, that is, a v-structure is when two vertices share a

Algorithm 1 Simulation of a matrix in $\mathbb{S}^p_{>0}(G)$ using diagonal dominance.

Input: Undirected graph G
Output: Matrix belonging to $\mathbb{S}^p_{>0}(G)$
1: $\mathbf{M} \leftarrow$ random matrix in $\mathbb{S}^p(G)$
2: **for** $i = 1, \ldots, p$ **do**
3: $\quad m_{ii} \leftarrow \sum_{i \neq j} |m_{ij}|+$ random positive perturbation
4: **end for**
5: **return** $\mathbf{M}$

common child but they are not adjacent. If an acyclic digraph has no v-structures, then for each node the set of parents is completely connected. The skeleton of an acyclic digraph with no v-structures is chordal; and, equivalently, any chordal undirected graph can be oriented into an acyclic digraph with no v-structures, as follows: let $C_1, \ldots, C_k$ denote a perfect sequence of cliques in an undirected chordal graph $G = (V, E)$, and write $H_j = C_1 \cup \ldots \cup C_j$, $R_j = C_j \setminus H_{j-1}$, following Lauritzen [17]. A *perfect ordering*, $v_1 \prec \cdots \prec v_p$, for the vertices of G is formed by first taking the vertices in C_1, then those in R_2, until R_k. This ordering has associated an acyclic directed orientation of G, $G_D = (V, E_D)$, which has no v-structures. In fact, $v_1 \prec \cdots \prec v_p$ is a topological ordering for G_D. Therefore, denoting for $v_i \in V$ as $\mathrm{pr}(v_i) = \{v_1, \ldots, v_{i-1}\}$ and $\mathrm{bd}(v_i) = \{v_j \in V : (v_i, v_j) \in E\}$, then we have

$$|E| = \sum_{i=1}^{p} |\mathrm{bd}(v_i) \cap \mathrm{pr}(v_i)| = \sum_{i=1}^{p} |\mathrm{pa}(v_i)| = |E_D|.$$

In the Gaussian case, this implies that if $1 \prec \cdots \prec p$ is a perfect ordering for G. Therefore, the theory of Section 2.2 for Gaussian Bayesian networks applies and $\mathbf{\Sigma}^{-1} = \mathbf{U}\mathbf{U}^t$ with $\mathbf{U} \in \mathcal{U}^p$ and the same zero pattern as in the upper triangle of $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$,

$$(j, i) \notin E_D \iff u_{ji} = 0 \iff \omega_{ji} = \omega_{ij} = 0 \iff (i, j) \notin E. \tag{8}$$

Thus we have that if G is a chordal undirected graph, then $\mathbb{S}^p_{>0}(G) = \mathcal{B}(G_D)$ [30,23].

3. Methods

3.1. Diagonal dominance

When a matrix $\mathbf{M} \in \mathbb{S}^p$ satisfies that $m_{ii} > \sum_{j \neq i} |m_{ij}|$ for each $i \in \{1, \ldots, p\}$, then $\mathbf{M}$ belongs to $\mathbb{S}^p_{>0}$. Thus a simple method to generate a matrix in $\mathbb{S}^p_{>0}(G)$ consists in generating a random matrix in $\mathbb{S}^p(G)$ and then choosing diagonal elements so the final matrix is diagonally dominant, as in Algorithm 1. The usual approach for generating the initial matrix in line 1 is to use independent and identically distributed (i.i.d.) nonzero entries. The diagonal dominance method has been extensively used in the literature mainly due to its simplicity and the ability to control the singularity of the generated matrices, as we will now explain.

Obviously it is then possible to generate correlation matrices in $\mathbb{S}^p_{>0}(G)$ using Algorithm 1 and then rescaling them to be in $\mathcal{E}^p(G)$.

It is even possible to control the minimum eigenvalue of a matrix by varying its diagonal elements [11]. In particular, let G be an undirected graph, $\mathbf{M}$ a matrix in $\mathbb{S}^p(G)$, and $\epsilon > 0$ the desired lower-bound on the eigenvalues. If λ_{min} is the minimum eigenvalue of $\mathbf{M}$, then $\mathbf{M} + (\lambda^-_{min} + \epsilon)\mathbf{I}_p$ belongs to $\mathbb{S}^p_{>0}(G)$ and has eigenvalues greater or equal to ϵ, where $\lambda^-_{min} = \max(-\lambda_{min}, 0)$ denotes the negative part of λ_{min}.

Similarly, one can control the condition number, that is, the ratio of the largest to smallest eigenvalue, of the generated matrix as in [2]: if $\kappa_0 > 1$ is the desired condition number and we already have a matrix $\mathbf{M} \in \mathbb{S}^p(G)$ with maximum eigenvalue $\lambda_{max} > 0$, then

$$\mathbf{M} + \frac{\lambda_{max} - \kappa_0 \lambda_{min}}{\kappa_0 - 1} \mathbf{I}_p$$

belongs to $\mathbb{S}^p_{>0}(G)$ and has condition number equal to κ_0. Covariance and concentration matrices with an upper bound on the condition number are appealing in certain estimation scenarios [13].

3.2. Partial orthogonalization

If we consider a full rank matrix $\mathbf{Q} \in \mathbb{R}^{p \times p}$ the product $\mathbf{Q}\mathbf{Q}^t$ is a symmetric positive definite matrix. Moreover, $\mathbf{Q}\mathbf{Q}^t \in \mathbb{S}^p_{>0}(G)$, for a given undirected graph G, if and only if:

$$\boldsymbol{q}_i \perp \boldsymbol{q}_j \quad \text{for } i \not\sim_G j,$$

Algorithm 2 Simulation of a matrix in $\mathcal{E}^p(G)$ using partial orthogonalization.

Input: Undirected graph G
Output: Matrix belonging to $\mathcal{E}^p(G)$
1: $\mathbf{Q} \leftarrow$ random $p \times p$ matrix
2: **for** $i = 1, \ldots, p$ **do**
3: orthogonalize $\boldsymbol{q}_i$ with respect to the span of $\{\boldsymbol{q}_j$ s.t. $i \nsim_G j$ and $j < i\}$
4: normalize $\boldsymbol{q}_i$, $\boldsymbol{q}_i = \boldsymbol{q}_i / \|\boldsymbol{q}_i\|_2$
5: **end for**
6: **return** $\mathbf{Q}\mathbf{Q}^t$

where $\perp$ denotes orthogonality with respect to the standard scalar product on $\mathbb{R}^p$, and $\boldsymbol{q}_i$ is the i-th row of $\mathbf{Q}$.

This fact suggests a very simple idea to generate matrices in $\mathbb{S}^p_{>0}(G)$: given an undirected graph G, we can impose Markov properties for the matrix $\mathbf{Q}\mathbf{Q}^t$ simply by orthogonalizing the respective rows of $\mathbf{Q}$. If moreover we also normalize the rows of $\mathbf{Q}$ we generate a matrix in the elliptope with graphical constrains $\mathcal{E}^p(G)$. The pseudocode for the described procedure can be found in Algorithm 2.

In particular we can use a modified Gram-Schmidt orthogonalization procedure that iteratively orthogonalizes every row $\boldsymbol{q}_i$ with respect to the set of rows $i^{\perp} = \{\boldsymbol{q}_j$ s.t. $i \nsim_G j$ and $j < i\}$.

3.3. Uniform sampling for chordal graphs

When G is a chordal graph, it is possible to sample uniformly from the set $\mathcal{E}^p(G)$ extending the results in [5]. In particular, for an undirected chordal graph G where $1 \prec \cdots \prec p$ is a perfect ordering, we consider the parametrization of $\mathcal{E}^p(G)$ induced by the Cholesky factorization (Section 2.3),

$$\mathcal{E}^p(G) = \{\mathbf{M} = \mathbf{U}\mathbf{U}^t \text{ s.t. } \mathbf{U} \in \mathcal{U}_1^p \text{ and } u_{ij} = 0 \text{ if } (i, j) \notin E\}.$$

Thus, if we further define the set,

$$\mathcal{U}_1^p(G) = \{\mathbf{U} \in \mathcal{U}_1^p \text{ s.t. } u_{ij} = 0 \text{ if } (i, j) \notin E\},$$

then

$$\begin{array}{rccc} \Phi: & \mathcal{U}_1^p(G) & \to & \mathcal{E}^p(G) \\ & \mathbf{U} & \mapsto & \mathbf{U}\mathbf{U}^t \end{array}$$

is a one-to-one parametrization of $\mathcal{E}^p(G)$. The Jacobian of Φ has been obtained by [25] and in [5], as

$$\det\left(J\Phi(\mathbf{U})\right) = 2^p \prod_{i=1}^{p} u_{ii}^{\mathrm{pa}(i)+1} \tag{9}$$

where $\mathrm{pa}(i)$ denotes the set of parents of node i in G_D, the acyclic directed orientation of G (which has $1 \prec \cdots \prec p$ as a topological ordering, see Section 2.3).

To sample from the uniform distribution over $\mathcal{E}^p(G)$ using parametrization Φ, we apply the area formula as Diaconis et al. [10], Theorem 1: we sample matrices in $\mathcal{U}_1^p(G)$ from a density proportional to $\det\left(J\Phi(\mathbf{U})\right)$ and then we apply the parametrization Φ. We observe that the Jacobian of Φ in Equation (9) factorizes across the rows $\boldsymbol{u}_i$ of $\mathbf{U}$, and thus we can sample the rows of $\mathbf{U}$ independently. In particular, for the i-th row of $\mathbf{U}$, we have that

$$\begin{aligned} & u_{ij} = 0 \quad j < i, \\ & u_{ii} > 0, \\ & u_{ik} = 0 \quad k \notin \mathrm{ch}(i), \end{aligned}$$

where $\mathrm{ch}(i)$ denotes the children set of node i in graph G_D. Therefore, for each $i \in \{1, \ldots, p\}$, the vector of non-zero entries in the i-th row of $\mathbf{U}$ has to be sampled in the hemisphere $\mathcal{S}_+^{|\mathrm{ch}(i)|}$ from a density proportional to a power of the first non-zero entry in such row, u_{ii}. This task can be done with the same Metropolis sampling procedure described in detail in [5] and outlined in Algorithm 4 for completeness (with default noise variance σ_ϵ and burn-in time t_b).

3.4. Combining uniform sampling and partial orthogonalization

When the undirected graph G is not chordal it is not possible to direct the edges without creating v-structures. This implies that applying Algorithm 3 to a triangulation of a non-chordal graph G will result in a matrix with more non-zeros entries than the desired ones. To overcome such issue we propose to combine the two approaches in the previous sections and first sample the Cholesky factor as in Algorithm 3 for the triangulated graph, and then apply the partial orthogonalization procedure as in Algorithm 2 to obtain a matrix in $\mathcal{E}^p(G)$.

The method is detailed in Algorithm 5.

For chordal graphs, Algorithm 5 obviously reduces to the uniform sampling of Algorithm 3.

Algorithm 3 Uniform sampling in $\mathcal{E}^p(G)$.

Input: Chordal graph G with $1 \prec \cdots \prec p$ as a perfect ordering
Output: A matrix uniformly sampled in $\mathcal{E}^p(G)$
$G_D \leftarrow$ acyclic directed orientation of G
$\mathbf{U}^n \leftarrow \mathbf{0}_p$
for $i \in \{1, \ldots, p\}$ **do**
  $\boldsymbol{v} \leftarrow \texttt{mh\_sphere}(\alpha = |\operatorname{ch}(i)|, \gamma = |\operatorname{pa}(i)| + 1)$
  $u_{ii} \leftarrow v_1$
  $\boldsymbol{u}_{i\,\operatorname{ch}(i)} \leftarrow \boldsymbol{v}_{-1}$ *// Vector v except its first entry*
end for
return $\Phi(\mathbf{U}) = \mathbf{U}\mathbf{U}^t$

Algorithm 4 `mh_sphere`: Metropolis sampling of a vector $\boldsymbol{v}$ in $\mathcal{S}_+^\alpha$ from $f(\boldsymbol{v}) \propto v_1^\gamma$.

Input: Dimension α of the sphere and power γ of the density
Output: A vector sampled in $\mathcal{S}_+^\alpha$
$\boldsymbol{v}_0 \leftarrow$ random standard multivariate Gaussian observation of dimension $\alpha + 1$
$v_{01} \leftarrow |v_{01}|$
$\boldsymbol{v}_0 \leftarrow$ normalize $\boldsymbol{v}_0$
for $t = 0, \ldots, t_b + 1$ **do**
  for $j = 1, \ldots, \alpha + 1$ **do**
    $\epsilon_j \leftarrow$ random Gaussian observation with zero mean and variance σ_ϵ^2
  end for
  $\boldsymbol{v}' \leftarrow \boldsymbol{v}_t + \boldsymbol{\epsilon}$
  $\boldsymbol{v}' \leftarrow$ normalize $\boldsymbol{v}'$, $\boldsymbol{v}' = \boldsymbol{v}'/\|\boldsymbol{v}'\|_2$
  $\delta \leftarrow$ random uniform observation on $[0, 1]$
  if $v_1' \geq 0$ **and** $\delta \leq (v_1'/v_{t1})^\gamma$ **then**
    $\boldsymbol{v}_t \leftarrow \boldsymbol{v}'$
  end if
end for
return $\boldsymbol{v}_{t_b+1}$

Algorithm 5 Simulation of a matrix in $\mathcal{E}^p(G)$ combining uniform sampling with respect to a triangulation and partial orthogonalization.

Input: Undirected graph G
Output: Matrix belonging to $\mathcal{E}^p(G)$
$G' \leftarrow$ triangulation of G following a perfect ordering defined by permutation σ
$G_D \leftarrow$ acyclic directed orientation of G'
$\mathbf{U} \leftarrow \mathbf{0}$
for $i = 1, \ldots, p$ **do**
  $\boldsymbol{v} \leftarrow \texttt{mh\_sphere}(\alpha = |\operatorname{ch}(\sigma(i))|, \gamma = |\operatorname{pa}(\sigma(i))| + 1)$
  $u_{ii} \leftarrow v_1$
  $\boldsymbol{u}_{i\,\operatorname{ch}(\sigma(i))} \leftarrow \boldsymbol{v}_{-1}$ *// Vector v except its first entry*
end for
$\mathbf{Q} \leftarrow$ permute rows and columns in $\mathbf{U}$ with σ^{-1} *// Revert the perfect ordering to retrieve original ordering of the nodes in G*
for $i = 1, \ldots, p$ **do**
  orthogonalize $\boldsymbol{q}_i$ with respect to the span of $\{\boldsymbol{q}_j$ s.t. $i \nsim_G j$ and $j < i\}$
  normalize $\boldsymbol{q}_i$, $\boldsymbol{q}_i = \boldsymbol{q}_i/\|\boldsymbol{q}_i\|_2$
end for
return $\mathbf{Q}\mathbf{Q}^t$

4. Experiments

In this section we report the results of numerical experiments performed to explore the behaviour of the methods presented. The implementation of the methods in the previous sections can be found in the R package **gmat**.[1] The partial orthogonalization procedure has been implemented in C for improved performance. The experiments can be reproduced following the instructions and using the code available at the repository https://github.com/irenecrsn/ggmsim.

4.1. Three variables

We consider the simple chordal graph $G = (\{1, 2, 3\}, \{\{1, 2\}, \{2, 3\}\})$ over three variables depicted in Fig. 1, and we analyze graphically how the three proposed methods behave.

We sample 5000 correlation matrices from $\mathcal{E}^3(G)$ using the diagonal dominance method (Algorithm 1), the partial orthogonalization method (Algorithm 2) and the uniform sampling (Algorithm 3). We use independent standard Gaussian

[1] Version in development: https://github.com/irenecrsn/gmat.

Fig. 1. Chordal undirected graph with three variables.

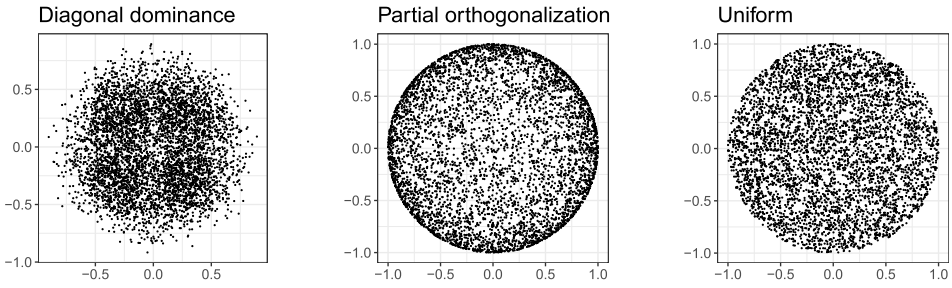

Fig. 2. Scatter plot of the two non-zero entries for correlation matrices sampled from $\mathcal{E}^3(G)$, with G as in Fig. 1.

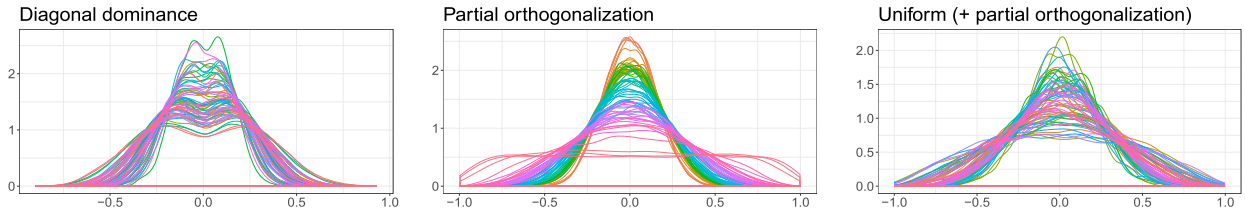

Fig. 3. Marginal densities of the non-zero entries of matrices sampled from $\mathcal{E}^{50}(G)$; where G is a random graph with 50 vertices and probability of edges 0.05. The first entry in the lower triangle (2, 1) corresponds to the red colour, while the last entry in the last row of the lower triangle (50, 49) corresponds to the pink colour. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

random variables to initialize the random matrices in both Algorithms 1 and 2. Matrices in $\mathcal{E}^3(G)$ have two non-zero upper triangular entries $(1,2)$ and $(2,3)$, and moreover $\mathcal{E}^3(G)$ can be represented as the interior of the two dimensional unit ball:

$$\mathcal{E}^3(G) = \left\{ \begin{pmatrix} 1 & x & 0 \\ x & 1 & y \\ 0 & y & 1 \end{pmatrix} \text{ s.t. } x^2 + y^2 < 1 \right\} \simeq \{(x, y) \in \mathbb{R}^2 \text{ s.t. } x^2 + y^2 < 1\}$$

The scatter plot of the two non-zero upper triangular entries for the three sampling methods is shown in Fig. 2.

We can see that, as expected, the uniform sampling method obtains a uniform distribution over $\mathcal{E}^3(G)$ while the diagonal dominance method and the partial orthogonalization methods have somehow the opposite behaviour. Matrices sampled with partial orthogonalization tend to have large off-diagonal values, while the diagonal dominance method produces matrices with smaller values for the off-diagonal entries.

4.2. Marginal distribution of matrix entries

We investigate here the marginal distribution of non-zeros matrix entries sampled from $\mathcal{E}^p(G)$ with the different methods, for both chordal and non-chordal graphs.

We generate a random undirected graph G over 50 vertices using the Erdős-Rényi model with a probability of edges equal to 0.05. We sample 5000 matrices from $\mathcal{E}^{50}(G)$ using Algorithms 1, 2 and 5. We then plot the marginal densities of the non-zero entries for the three methods. The results are shown in Fig. 3.

We also consider G', the triangulation of G and we generate again 5000 matrices in $\mathcal{E}^{50}(G')$ using the three methods. Plots of the marginal densities of the non-zero entries are shown in Fig. 4.

From both Figs. 3 and 4 we can observe that the diagonal dominance method produces matrices with off-diagonal entries more concentrated around 0, as also pointed-out in [4]. In apparent contrast to the finding in [4], also the partial orthogonalization method seems to produce matrices with entries more concentrated around 0. Intuitively this can be seen as a consequence of the fact that vectors of independent random components are approximately orthogonal in high-dimensions. To further prove this problem of the partial orthogonalization algorithm, we simulate 5000 matrices from $\mathcal{E}^{50}(G_{chain})$, where $G_{chain} = (\{1, \ldots, 50\}, \{\{1, 2\}, \{2, 3\}, \ldots, \{49, 50\}\})$ (see Fig. 5).

As usual we plot the marginal densities of the 49 non-zero entries of the generated matrices with the three different methods (G_{chain} is chordal and thus we can sample uniformly) (see Fig. 6).

We observe that for this graph the distribution induced on the matrix entries is completely different for the three methods. In particular it is interesting to note that the partial orthogonalization method produces matrices $\mathbf{M} \in \mathcal{E}^{50}(G_{chain})$ with the first non-zero entries $m_{1,2}, m_{2,3}, m_{3,4}, \ldots$ more centred around 0 than the last entries $\ldots, m_{48,49}, m_{49,50}$. On the

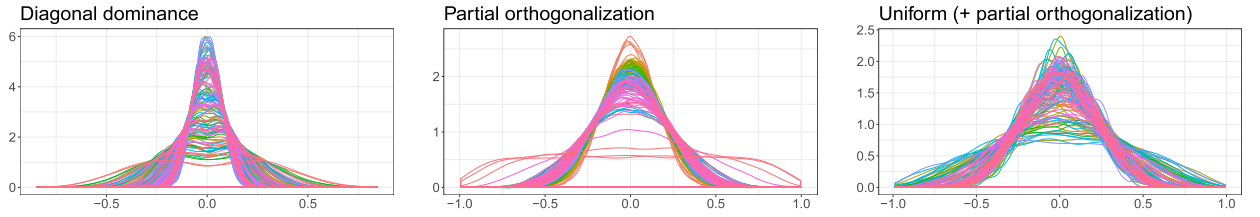

Fig. 4. Marginal densities of the non-zero entries of matrices sampled from $\mathcal{E}^{50}(G')$; where G' is the chordal graphs obtained as the triangulation of a random graph with 50 vertices and probability of edges 0.05. The first entry in the lower triangle (2, 1) corresponds to the red colour, while the last entry in the last row of the lower triangle (50, 49) corresponds to the pink colour.

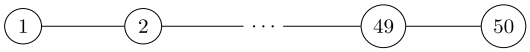

Fig. 5. Chordal undirected graph G_{chain} with 50 variables and 49 edges.

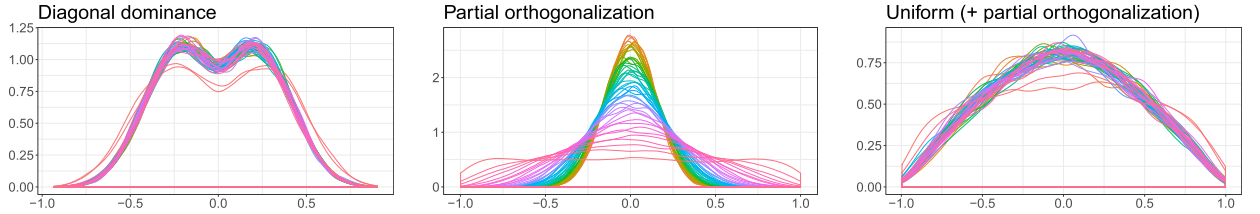

Fig. 6. Marginal densities of the non-zero entries of matrices sampled from $\mathcal{E}^{50}(G_{chain})$. The first entry in the lower triangle, (2, 1), corresponds to the red colour, while the last entry in the last row of the lower triangle, (49, 48), corresponds to the pink colour.

contrary the uniform sampling, correctly produces matrices with the same marginal densities for the entries. This behaviour of the partial orthogonalization procedure is due to the i.i.d. sampling of the elements of factor $\mathbf{Q}$ in Algorithm 2 and not to the orthogonalization part, that instead mitigates this fact (the first entries of the matrix $m_{1,2}, m_{2,3}$ are the ones where no-orthogonalization is applied by Algorithm 2). We remark that such problem for the partial orthogonalization procedure applied to a random matrix $\mathbf{Q}$ with i.i.d. entries can be disturbing since it introduces some asymmetries in the distribution of the matrices that are absent in graph G.

4.3. Validation of structure learning algorithms

The main motivation for the proposed method are the observations that can be found in the literature on covariance and concentration graphs regarding the difficulties of validating the performance of structure learning algorithms [26,16,2]. We have selected the work of Krämer et al. [16, page 7], who highlight how they obtain significantly poorer graph recovery results as the density d of the graphs grows. They simulate the corresponding concentration graph models using the diagonal dominance method, so we have replicated their experiments but using instead as true models those generated with our proposed method.

The results can be seen in Figs. 7 and 8, where we have plotted the true positive rate (TPR, also called power by Krämer et al. [16]) and the positive predictive value (PPV) or precision for $p = 100$ and their sparsest ($d = 0.05$) and densest ($d = 0.25$) scenarios, using matrices simulated with the diagonal dominance (Algorithm 1) and our proposed method (Algorithm 5). The different structure learning methods are the same as those studied by Krämer et al. [16]. Note that in this work we correctly define, in the computations for TPR and PPV, the indefinite fraction $\frac{0}{0}$ to be equal to 1. For some learning methods such as `shrink` and `pls` this drastically affects their curve when comparing to [4].

Note that there is a significant improvement in the densest case ($d = 0.25$) when using our method (Algorithm 5). All the learning algorithms are close to zero TPR for every sample size when validating on diagonally dominant matrices, which highlights a poor performance (the high PPVs are thus not significant). However, when using matrices obtained via partial orthogonalization, some methods (`lasso` and `adalasso`) are able to achieve a TPR of 0.5 approximately. Importantly, partial least squares regression (`pls`) and the shrinkage estimator (`shrink`) greatly improve, whereas when only using diagonal dominance one could erroneously conclude that those methods are not well fitted for dense structure scenarios. In the sparsest scenario ($d = 0.05$) we observe that the PPV for partial least squares extremely drops when using our proposed simulation method, while the other algorithms rank similarly using either one. This behaviour is expected: the densest scenario ($d = 0.25$) is not intrinsically difficult, but it indeed poses special difficulties when using diagonally dominant matrices, because correlations are in general small (Figs. 3 – 6) and therefore structure recovery amounts to discriminating an absent edge from an extremely small entry, which is a significantly hard task.

The most important conclusion to draw from these results is that sampling procedures for covariance or concentration graph matrices highly influence how the respective structure learning algorithms are ranked. However, it would be incorrect

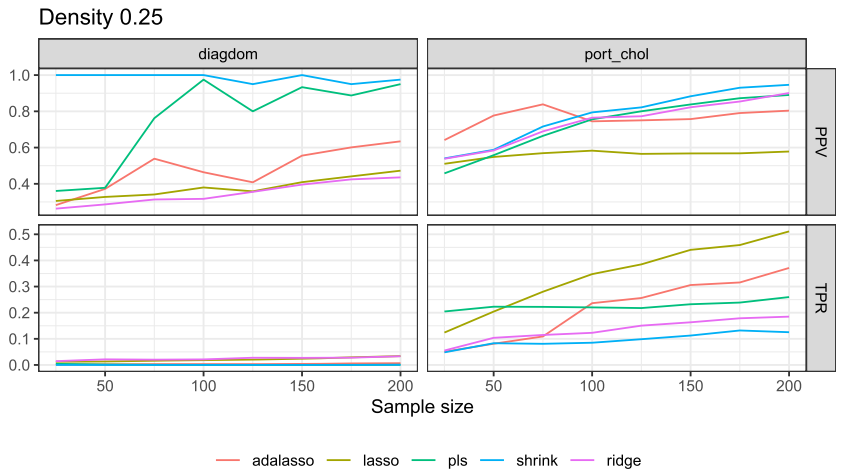

Fig. 7. True positive rate (TPR) and positive predictive value (PPV) of the structure learning algorithms for concentration graphs validated in [16], for the highest density, 0.25. The number of variables (vertices in the undirected graph and dimension of the generated matrices) is fixed at 100. `adalasso`: Adaptive l_1 regularization; `lasso`: l_1 regularization; `pls`: partial least squares regression; `shrink`: shrinkage estimator of Schäfer and Strimmer [27]; `ridge`: l_2 regularization; `diagdom`: Diagonal dominance sampling method; `port_chol`: Uniform sampling with partial orthogonalization of the Cholesky factor.

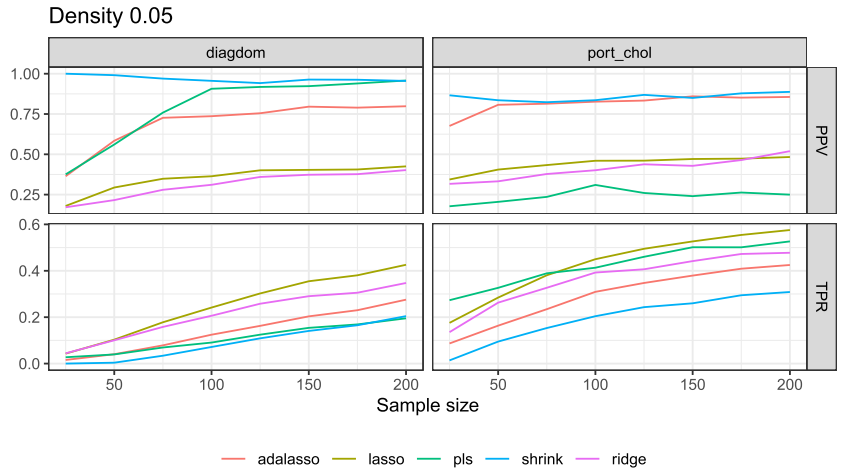

Fig. 8. True positive rate (TPR) and positive predictive value (PPV) of the structure learning algorithms for concentration graphs validated in [16], for the lowest density, 0.05, that is, the sparsest scenario. The number of variables (vertices in the undirected graph and dimension of the generated matrices) is fixed at 100. `adalasso`: Adaptive l_1 regularization; `lasso`: l_1 regularization; `pls`: partial least squares regression; `shrink`: shrinkage estimator of Schäfer and Strimmer [27]; `ridge`: l_2 regularization; `diagdom`: Diagonal dominance sampling method; `port_chol`: Uniform sampling with partial orthogonalization of the Cholesky factor.

to claim that one of the simulation methods is superior to the other. Indeed, we want to highlight the importance of choosing the correct simulation method for each numerical validation scenario. If we want to assess performance for a wide range of covariance or concentration chordal graphs, then our proposed method (Algorithm 5) should be used, since it guarantees uniform sampling and therefore unbiased validation. On the contrary, if we know that our models exhibit small correlations, then using diagonally dominant concentration graph matrices would be justified because they exhibit such property. In doubt, we argue for the use of our method because it samples from a wider space range (Fig. 2), while the diagonal dominance method is largely biased towards matrices away from the space frontier.

5. Conclusions

In this work we introduced two methods to sample from the set $\mathcal{E}^p(G)$ of correlation matrices with undirected graphical constraints, a general partial orthogonalization procedure and a uniform sampling method when the graph G is chordal. We showed with some numerical experiments that both the partial orthogonalization method and the classical diagonal dominance procedure suffer from some drawbacks in effectively exploring the space of correlation matrices with undirected graphical constraints. For chordal graphs, it is possible to sample from the uniform distribution easily, extending a method to

sample correlation matrices uniformly; while for non-chordal graphs we propose to combine the uniform sampling method and the partial orthogonalization by firstly sampling a Cholesky factor related to the triangulated graph and then applying the partial orthogonalization to remove the non-zeros entries related to the edges added in the triangulation. The proposed method has shown to be helpful in the validation of structure learning algorithms overcoming the problems of the diagonal dominance method. The main direction for future research is to investigate how to sample uniformly form the space $\mathcal{E}^p(G)$ for a non-chordal graph G.

CRediT authorship contribution statement

Irene Córdoba: Conceptualization, Formal analysis, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. **Gherardo Varando:** Conceptualization, Formal analysis, Methodology, Software, Writing - original draft, Writing - review & editing. **Concha Bielza:** Supervision, Writing - review & editing. **Pedro Larrañaga:** Supervision, Writing - review & editing.

Declaration of competing interest

There are no conflicts of interest.

Acknowledgements

This work has been partially supported by the Spanish Ministry of Science, Innovation and Universities through the PID2019-109247GB-I00 project. Irene Córdoba has been supported by the predoctoral grant FPU15/03797 from the Spanish Ministry of Science, Innovation and Universities. Gherardo Varando has been supported by a research grant (13358) from Villum Fonden.

References

[1] E. Arvaniti, M. Claassen, Markov network structure learning via ensemble-of-forests models, in: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2014, pp. 42–51.
[2] T. Cai, W. Liu, X. Luo, A constrained l_1 minimization approach to sparse precision matrix estimation, J. Am. Stat. Assoc. 106 (494) (2011) 594–607.
[3] C.M. Carvalho, H. Massam, M. West, Simulation of hyper-inverse Wishart distributions in graphical models, Biometrika 94 (3) (2007) 647–659.
[4] I. Córdoba, G. Varando, C. Bielza, P. Larrañaga, A partial orthogonalization method for simulating covariance and concentration graph matrices, in: V. Kratochvíl, M. Studený (Eds.), Proceedings of the Ninth International Conference on Probabilistic Graphical Models, PMLR, 11-14 September 2018, Prague, Czech Republic, in: Proceedings of Machine Learning Research, vol. 72, 2018, pp. 61–72.
[5] I. Córdoba, G. Varando, C. Bielza, P. Larrañaga, A fast Metropolis-Hastings method for generating random correlation matrices, in: H. Yin, D. Camacho, P. Novais, A.J. Tallón-Ballesteros (Eds.), Intelligent Data Engineering and Automated Learning – IDEAL 2018, Springer International Publishing, Cham, 2018, pp. 117–124.
[6] D.R. Cox, N. Wermuth, Linear dependencies represented by chain graphs, Stat. Sci. 8 (3) (08 1993) 204–218.
[7] A.P. Dawid, Conditional independence for statistical operations, Ann. Stat. 8 (3) (1980) 598–617.
[8] A.P. Dawid, S.L. Lauritzen, Hyper Markov laws in the statistical analysis of decomposable graphical models, Ann. Stat. 21 (3) (1993) 1272–1317.
[9] A.P. Dempster, Covariance selection, Biometrics 28 (1) (1972) 157–175.
[10] P. Diaconis, S. Holmes, M. Shahshahani, Sampling from a Manifold, Collections. Institute of Mathematical Statistics, vol. 10, 2013, pp. 102–125.
[11] J. Honorio, D. Samaras, I. Rish, G. Cecchi, Variable selection for Gaussian graphical models, in: Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, PMLR, 2012, pp. 538–546.
[12] H. Joe, Generating random correlation matrices based on partial correlations, J. Multivar. Anal. 97 (10) (2006) 2177–2189.
[13] W. Joong-Ho, L. Johan, K. Seung-Jean, R. Bala, Condition-number-regularized covariance estimation, J. R. Stat. Soc., Ser. B, Stat. Methodol. 75 (3) (2013) 427–450.
[14] G. Kauermann, On a dualization of graphical Gaussian models, Scand. J. Stat. 23 (1) (1996) 105–116.
[15] K. Khare, B. Rajaratnam, Wishart distributions for decomposable covariance graph models, Ann. Stat. 39 (1) (2011) 514–555.
[16] N. Krämer, J. Schäfer, A.-L. Boulesteix, Regularized estimation of large-scale gene association networks using graphical Gaussian models, BMC Bioinform. 10 (1) (2009) 384.
[17] S.L. Lauritzen, Graphical Models, Oxford University Press, 1996.
[18] O. Ledoit, M. Wolf, Nonlinear shrinkage estimation of large-dimensional covariance matrices, Ann. Stat. 40 (2) (2012) 1024–1060.
[19] A. Lenkoski, A direct sampler for G-Wishart variates, Stat 2 (1) (2013) 119–128.
[20] G. Letac, H. Massam, Wishart distributions for decomposable graphs, Ann. Stat. 35 (3) (2007) 1278–1323.
[21] D. Lewandowski, D. Kurowicka, H. Joe, Generating random correlation matrices based on vines and extended onion method, J. Multivar. Anal. 100 (9) (2009) 1989–2001.
[22] Y. Lin, S. Zhu, D. Lee, B. Taskar, Learning sparse Markov network structure via ensemble-of-trees models, in: Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, PMLR, Florida, 2009, pp. 360–367.
[23] V.I. Paulsen, S.C. Power, R.R. Smith, Schur products and matrix completions, J. Funct. Anal. 85 (1) (1989) 151–178.
[24] M. Pourahmadi, Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation, Biometrika 86 (3) (1999) 677–690.
[25] A. Roverato, Cholesky decomposition of a hyper inverse Wishart matrix, Biometrika 87 (1) (2000) 99–112.
[26] J. Schäfer, K. Strimmer, An empirical Bayes approach to inferring large-scale gene association networks, Bioinformatics 21 (6) (2005) 754–764.
[27] J. Schäfer, K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, Stat. Appl. Genet. Mol. Biol. 4 (1) (2005).
[28] I. Stojkovic, V. Jelisavcic, V. Milutinovic, Z. Obradovic, Fast sparse Gaussian Markov random fields learning based on Cholesky factorization, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017, pp. 2758–2764.

[29] J.A. Tropp, Simplicial faces of the set of correlation matrices, Discrete Comput. Geom. 60 (2018) 512–529.
[30] N. Wermuth, Linear recursive equations, covariance selection, and path analysis, J. Am. Stat. Assoc. 75 (372) (1980) 963–972.
[31] N. Wermuth, D. Cox, G.M. Marchetti, Covariance chains, Bernoulli 12 (5) (10 2006) 841–862.
[32] D.R. Williams, J. Piironen, A. Vehtari, P. Rast, Bayesian estimation of Gaussian graphical models with predictive covariance selection, https://arxiv.org/abs/1801.05725, 2018.
[33] L. Xue, H. Zou, Regularized rank-based estimation of high-dimensional nonparanormal graphical models, Ann. Stat. 40 (5) (2012) 2541–2571.
[34] M. Yuan, Y. Lin, Model selection and estimation in the Gaussian graphical model, Biometrika 94 (1) (2007) 19–35.