A partial orthogonalization method for simulating covariance and concentration graph matrices

Irene Córdoba[1] IRENE.CORDOBA@UPM.ES
Gherardo Varando[1,2] GHERARDO.VARANDO@UPM.ES
Concha Bielza[1] MCBIELZA@FI.UPM.ES
Pedro Larrañaga[1] PEDRO.LARRANAGA@FI.UPM.ES
[1]*Department of Artificial Intelligence, Universidad Politécnica de Madrid (Spain)*
[2]*Department of Mathematical Sciences, University of Copenhagen (Denmark)**

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. However, the link strengths in the resulting graphical model, determined by off-diagonal entries in the SPD matrix, are in many scenarios extremely weak. Recovering the structure of the undirected graph thus becomes a challenge, and algorithm validation is notably affected. In this paper, we propose an alternative method which overcomes such problem yet yields a compatible SPD matrix. We generate a partially row-wise-orthogonal matrix factor, where pairwise orthogonal rows correspond to missing edges in the undirected graph. In numerical experiments ranging from moderately dense to sparse scenarios, we obtain that, as the dimension increases, the link strength we simulate is stable with respect to the structure sparsity. Importantly, we show in a real validation setting how structure recovery is greatly improved for all learning algorithms when using our proposed method, thereby producing a more realistic comparison framework.

Keywords: Concentration graph; covariance graph; positive definite matrix simulation; undirected graphical model; algorithm validation.

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 (Cox and Wermuth, 1993; Kauermann, 1996) and concentration graphs (Dempster, 1972; Lauritzen, 1996) 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 (Yuan and Lin, 2007; Xue and Zou, 2012; Ledoit and Wolf, 2012), with parameters fixed to ensure positive definiteness.

$*$. Starting August 2018

The issue of positive definiteness is especially relevant when the structure is randomly generated. One approach to overcome this could be to sample from a matrix distribution with support over the symmetric positive definite matrices compatible with the undirected graph structure. The hyper Wishart distributions (Dawid and Lauritzen, 1993; Letac and Massam, 2007) are the most well-developed in this sense, since they form a conjugate family for Bayesian analysis. However, while sampling algorithms are available for general concentration graphs (Carvalho et al., 2007; Lenkoski, 2013), in covariance graphs they have been developed only in the decomposable case (Khare and Rajaratnam, 2011).

In general, hyper Wishart distributions are rarely used in validation scenarios (Williams et al., 2018), and instead in the literature the most common approach to ensure positive definiteness is to enforce diagonal dominance in the covariance or concentration matrix (Lin et al., 2009; Arvaniti and Claassen, 2914; Stojkovic et al., 2017). However, when the undirected graph is moderately dense, the off-diagonal elements in the generated 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 (Schäfer and Strimmer, 2005a,b; Krämer et al., 2009; Cai et al., 2011). In this paper, we propose an alternative method to overcome this problem based on partial orthogonalizations. In particular, we build a matrix factor where pairwise orthogonal rows correspond to missing edges in the undirected graph. Our method does not suffer from the problem of weak link strengths, as we numerically check in a wide range of sparsity scenarios. We also use our simulation method in a real validation setting and show how the performance is greatly improved for every learning algorithm, thereby potentially changing 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 and covariance graphs, and the main characteristics of diagonally controlled matrices. Next, in Section 3, we present our partial orthogonalization method, analyzing its main properties and our particular implementation. 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 Dawid (1980) 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 $\mathbf{M}$.

2.1 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 (Cox and Wermuth, 1993; Kauermann, 1996),

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

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 (Lauritzen, 1996) 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, $X_i \perp\!\!\!\perp X_j \mid \boldsymbol{X}_{V\setminus\{i,j\}} \iff \omega_{ij} = 0$.

One can always construct multivariate Gaussian distributions belonging to a covariance or concentration graph, for an arbitrary structure G. Furthermore, these models are Markov equivalent, in the sense that they represent the same set of distributions, whenever the respective structures share the same disconnected complete subgraphs (Jensen, 1988; Drton and Richardson, 2008).

2.2 Symmetric positive definite matrices and undirected graphs

The statistical independences implied by both covariance and concentration graph models are explicitly represented in a symmetric positive definite matrix. It is of our interest the problem on how to simulate such kind of matrices, subject to the constraint of being compatible with a given undirected graph. We will abstract ourselves from whether such graph has been randomly generated or pre-specified.

Denote as $\mathbb{S}$ the space of symmetric $p \times p$ matrices and as $\mathbb{S}^{>0}$ its subspace of symmetric positive definite matrices. For a fixed undirected graph G let $\mathcal{M}_G$ be the set of matrices $\mathbf{M}$ with zeros in the entries represented by the missing edges in G, that is,

$$\mathcal{M}_G = \{\mathbf{M} \in \mathbb{R}^{p\times p} \text{ s.t. } m_{ij} = m_{ji} = 0 \text{ if } i \not\sim_G j\}.$$

Let $\mathbb{S}(G) = \mathbb{S} \cap \mathcal{M}_G$ and $\mathbb{S}^{>0}(G) = \mathbb{S}^{>0} \cap \mathcal{M}_G$ be the sets of symmetric and symmetric positive definite matrices with undirected graphical constraints.

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

2.3 Diagonally Controlled Matrices

When a matrix $\mathbf{M} \in \mathbb{S}$ satisfies that $m_{ii} > \sum_{j\neq i} |m_{ij}|$ for each $i \in \{1, \ldots, p\}$, then $\mathbf{M}$ belongs to $\mathbb{S}^{>0}$. Thus a simple method to generate a matrix in $\mathbb{S}^{>0}(G)$ consists in generating a random matrix in $\mathbb{S}(G)$ 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 for its simplicity and the ability to control the singularity of the generated matrices, as we will now explain.

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

Input: Undirected graph G
Output: Matrix belonging to $\mathbb{S}^{>0}(G)$
1: $\mathbf{M} \leftarrow$ random $p \times p$ matrix in $\mathbb{S}(G)$
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** $\mathbf{M}$

It is possible to control the minimum eigenvalue of a matrix by varying its diagonal elements (Honorio et al., 2012). In particular, let G be an undirected graph, $\mathbf{M}$ a matrix in $\mathbb{S}(G)$, 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 $\mathbb{S}^{>0}(G)$ and has eigenvalues greater or equal to ϵ, where λ_{min}^{-} 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 follows (Cai et al., 2011). If $\kappa_0 > 1$ is the desired condition number and we already have a matrix $\mathbf{M} \in \mathbb{S}(G)$ with maximum eigenvalue $\lambda_{max} > 0$, then

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

belongs to $\mathbb{S}^{>0}(G)$ and has condition number equal to κ_0. Covariance and concentration matrices with an upper bound on the condition number are attractive in certain estimation scenarios (Joong-Ho et al., 2013).

3. Simulating matrices in $\mathbb{S}^{>0}(G)$ by partial orthogonalization

Let G_p be an Erdös-Rényi (Erdös and Rényi, 1959) random graph over p nodes with edge probability $d \in (0, 1)$ and given G_p let $\mathbf{M} \in \mathbb{S}(G_p)$ be a symmetric random matrix with i.i.d. non-zero off-diagonal entries m_{ij} following a distribution with $0 < \mu < +\infty$ the expected absolute value. If we denote with $\mathbf{M}' \in \mathbb{S}^{>0}(G_p)$ the output of the diagonal dominance method (Section 2.3), we have that for all $1 \leq i \leq p$ and $j \neq i$:

$$r_{ij} = \frac{|m'_{ij}|}{m'_{ii}} < \frac{|m_{ij}|}{\sum_{t \neq i} |m_{it}|} < \frac{|m_{ij}|}{\sum_{t \neq i,j} |m_{it}|} = \frac{|m_{ij}|}{(p-2) \sum_{t \neq i,j} |m_{it}| / (p-2)}. \tag{1}$$

By the strong law of large number and since $|m_{ij}|$ and $\sum_{t \neq i,j} |m_{it}|$ are independent, we have that $r_{ij} \to 0$ almost surely.

In order to overcome such issue, we propose an alternative method which does not rely on diagonal dominance, which we will now describe. If we consider an arbitrary $p \times p$ full rank matrix $\mathbf{Q}$, the product $\mathbf{Q}\mathbf{Q}^t$ is positive definite and symmetric, and therefore lies in $\mathbb{S}^{>0}$. Moreover, $\mathbf{Q}\mathbf{Q}^t$ belongs to $\mathbb{S}^{>0}(G)$ if and only if

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

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 $\mathbf{Q}$.

Thus, 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 $\mathbf{Q}$. The pseudocode for the described procedure can be found in Algorithm 2.

Algorithm 2 Simulation of a matrix in $\mathbb{S}^{>0}(G)$ using partial orthogonalization

Input: Undirected graph G
Output: Matrix belonging to $\mathbb{S}^{>0}(G)$
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. $i \not\sim_G j$ and $j < i\}$
4: **end for**
5: **return** $\mathbf{Q}\mathbf{Q}^t$

After Algorithm 2 has finished, it outputs a matrix that correctly reflects the graphical structure given by the input graph G. If the entries in matrix $\mathbf{Q}$ are initially simulated as i.i.d. centered subgaussian, then its condition number $\kappa(\mathbf{Q}) \geq p$ with high probability (Rudelson and Vershynin, 2009). Therefore, in such case the condition number of the matrices $\mathbf{Q}\mathbf{Q}^t$ returned by Algorithm 2 will satisfy $\kappa(\mathbf{Q}\mathbf{Q}^t) \geq p^2$ as the graph structure becomes denser, as shown in Figure 1. Although the magnitude of the condition numbers shown are relatively high, this has not been an issue in our numerical experiments (see Section 4).

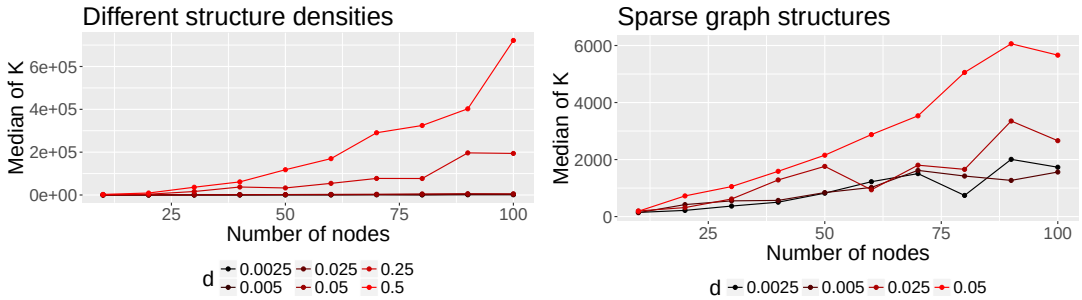

Figure 1: The median of the condition number as a function of the number of variables p for different structure densities d. We can see that the lower bound of p^2 is valid for dense structures ($d = 0.5$ and $d = 0.25$ in the left), whereas sparse graphs yield lower condition numbers (right figure). K: condition number of the matrix.

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. $i \not\sim_G j$ and $j < i\}$. This particularization of the proposed method is reflected in Algorithm 3, where $proj_{\boldsymbol{v}}(\boldsymbol{u})$ denotes the orthogonal projection of a vector $\boldsymbol{u}$ on another vector $\boldsymbol{v}$. The loop in line 3 constructs a set of orthogonal vectors $\tilde{\boldsymbol{q}}_j$ which span the same subspace than the original rows $\boldsymbol{q}_j$ belonging to $i^{\perp}$. This orthogonal base is later used in the loop at line 9 for ensuring that $\boldsymbol{q}_i$ is jointly orthogonal to all the vectors in $i^{\perp}$. Note however that these auxiliary orthogonal set of vectors is discarded in the next iteration, and the original $\boldsymbol{q}_j$ are kept in the factor matrix $\mathbf{Q}$ that will be used in the last computation of line 13.

Algorithm 3 Simulation of a matrix in $\mathbb{S}^{>0}(G)$ by modified Gram-Schmidt orthogonalization

Input: Undirected graph G
Output: Matrix belonging to $\mathbb{S}^{>0}(G)$

1: $\mathbf{Q} \leftarrow$ random $p \times p$ matrix
2: **for** $i = 1, \ldots, p$ **do**
3: **for** $j = 1, \ldots, i-1$ **and** $j \not\sim_G i$ **do**
4: $\tilde{\boldsymbol{q}}_j \leftarrow \boldsymbol{q}_j$
5: **for** $k = 1, \ldots, j-1$ **and** $k \not\sim_G i$ **do**
6: $\tilde{\boldsymbol{q}}_j \leftarrow \tilde{\boldsymbol{q}}_j - proj_{\tilde{\boldsymbol{q}}_k}(\tilde{\boldsymbol{q}}_j)$
7: **end for**
8: **end for**
9: **for** $j = 1, \ldots, i-1$ **and** $j \not\sim_G i$ **do**
10: $\boldsymbol{q}_i \leftarrow \boldsymbol{q}_i - proj_{\tilde{\boldsymbol{q}}_j}(\boldsymbol{q}_i)$
11: **end for**
12: **end for**
13: **return** $\mathbf{Q}\mathbf{Q}^t$

The computational complexity of Algorithm 3 is mainly given by the loop in line 3, where an orthogonal base is found for the subspace spanned by nonadjacent rows to i. In the worst case scenario, such row set $i^{\perp}$ has cardinality $i-1$, making the worst-case complexity of the inner loop $O(i^2 p)$, and that of the overall algorithm $O(p^4)$.

4. Numerical experiments

In this section we perform a simulation study to compare our proposed method to generate matrices in $\mathbb{S}^{>0}(G)$ against the diagonal dominance one. We have used random Erdös-Rényi (Erdös and Rényi, 1959) undirected graphs $G = (V, E)$. The size of the vertex set V, p, will take each of the values in the first row of Table 1. The probability of the inclusion of an edge in E, d, will take the values displayed in the second row of Table 1. This probability can be thought of as an indicator of the graph's density, ranging from sparse structures ($d = 0.0025$) to dense ones ($d = 0.5$). In particular, for every $(p, d) \in P \times D$ (Table 1) we generate 10 Erdös-Rényi graphs, $G_1^{p,d}, \ldots, G_{10}^{p,d}$, and we sample 10 matrices in $\mathbb{S}^{>0}(G_n^{p,d})$ ($n \in \{1, \ldots, 10\}$) using our proposed method (Algorithm 2) and diagonal dominance (Algorithm 1). In total we thus sample 100 matrices for every pair of parameters $(p, d) \in P \times D$. Both methods need to generate a matrix with random entries as a first step. In order to generate the initial matrices in both methods, we sample i.i.d. entries following a uniform distribution $\mathrm{U}[0, 1]$.

We compute for every (p, d) in Table 1 the average $\overline{R}$ of the maximum ratio $R = \max_{j \neq i} r_{ij}$, where r_{ij} are defined as in Equation (1) and the dependence on the matrix under consideration has been omitted for notational simplicity. A plot of the behaviour of $\overline{R}$ as a function of the number of variables in the model, for different density values of the graphical structure, is shown in Figure 2. We can observe that our proposed method generates matrices with an asymptotically (in the number of variables p) constant value of $\overline{R}$. On the contrary for undirected graphs whose density is higher than 0.025, which are usually found in applications (Krämer et al., 2009), $\overline{R}$ goes to zero as p increases for matrices simulated using the diagonal dominance method. Only for arguably very

Parameter	Value set
p	$P = \{10, 20, \ldots, 100, 125, 150, 200, 250, 300, 400, 500, 750, 1000\}$
d	$D = \{0.0025, 0.005, 0.025, 0.05, 0.25, 0.5\}$

Table 1: Setting of the numerical experiments for simulating from $\mathbb{S}^{>0}(G)$, with G an undirected graph. Values for the size of the vertex set (p) and the density of the structure (d).

sparse matrices ($d \leq 0.005$), this method is able to avoid such asymptotic behaviour with respect to p, but as we have shown in Section 3 for sufficiently high values of p the same behaviour is to be expected. In particular, since the $\mathrm{U}[0, 1]$ distribution is bounded, we can obtain from Equation 1 that almost surely $r_{ij} \leq 2(p-1)^{-1}d^{-1}$ asymptotically and thus $R = \mathcal{O}(p^{-1})$, thereby allowing for an approximate computation of the number of variables p_0 from where, for a given structure density d_0, R is arbitrarily close to zero.

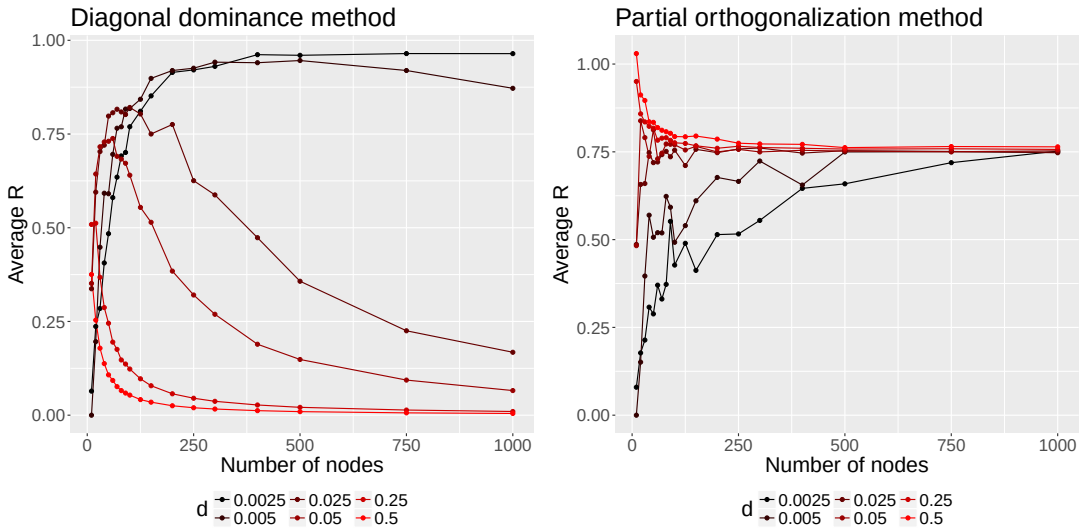

Figure 2: The average of R as a function of the number of variables p for different structure densities d.

The above mentioned conclusions are complementarily drawn from Figure 3, where we have jointly plotted the performance of both methods for the two extreme values we have considered for the structure density: 0.0025 (very sparse) and 0.5 (very dense). We also show as a shade one standard deviation on either side of the mean. We can observe that in the sparse scenario both methods perform reasonably well, with the diagonal dominance method being more stable in terms of the standard deviation; however, we must also point out that as p increases, our method becomes more robust, being equally stable for structures of a thousand vertices. By contrast, in the dense case the diagonal dominance method performance is terribly affected early, being almost zero for $p > 125$, approximately. Our proposed method, however, manages to achieve a reasonable value for the average ratio, and the constant behaviour in p can be clearly observed. We also obtain that our method

is more stable for dense structures than in the sparse case. In some sense, this is not surprising, because the more missing edges in the undirected graph of the model, the more orthogonalizations in Algorithm 2, therefore the numerical stability is reduced, even though competitive results are equally obtained.

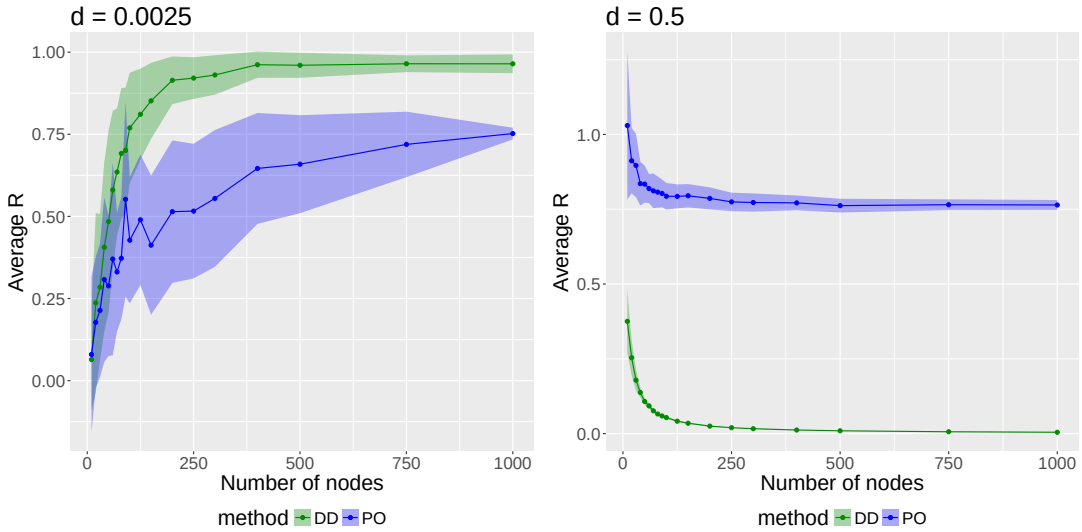

Figure 3: The average of R as a function of number of variables p, for the two extreme values of d: very sparse matrices (left) and very dense matrices (right). Standard deviation of the average is shown. DD: Diagonal dominance method; PO: Partial orthogonalization method.

We have also measured the execution time of the two approaches. For this, we have sampled 5000 matrices for the different density values in Table 1, and for a number of variables ranging from 10 to 200, in different step sizes. This experiment has been executed on a machine equipped with Intel Core i7-5820k, 3.30 GHz$\times$12 and 64 GB of RAM. The results are shown in Figure 4. The diagonal dominance method is few orders of magnitude faster than our proposed method, which is somewhat expected given its relative simplicity. We observe how the computational cost of the partial orthogonalization method depends on the structure density. For small values of d the undirected graph contains a lot of disconnected vertices and thus we repeat the loop in line 3 of Algorithm 3 for many matrix rows, being closer to the worst case scenario of $\mathcal{O}(p^4)$. In practice, however, when validating structure learning algorithms only few matrices need to be generated, which is completely affordable in all scenarios of Figure 4. For example, the time needed to generate 10 matrices with 200 nodes in the worst case is approximately five seconds.

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 (Schäfer and Strimmer, 2005a; Krämer et al., 2009; Cai et al., 2011). In particular, Krämer et al. (2009) obtain significantly poorer graph recovery results as the density of the graphs grow. They simulate the corresponding concentration graph models using the diagonal dominance method, so we have replicated their experiments but instead using as true models those generated with our proposed method. The results can be seen in Figure 5, where we

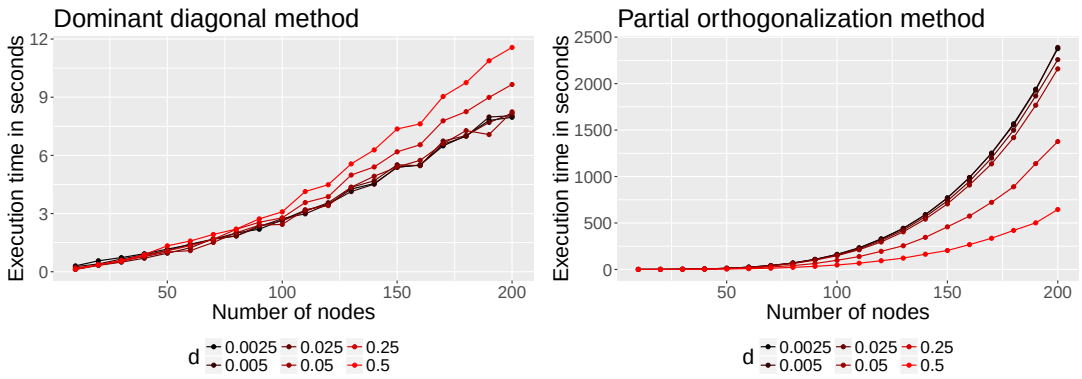

Figure 4: Execution time to simulate 5000 matrices.

have plotted the true positive rate (also called power by Krämer et al. (2009)) and discovery rates for $p = 100$ and their most dense scenario, $d = 0.25$, when using matrices simulated with the diagonal dominance method and our proposal. The different structure learning methods appearing are the same under validation by Krämer et al. (2009). As can be observed, there is significant improvement when using our method: all of the learning algorithms are close to zero true positive rate for every sample size when validating on diagonally dominant matrices, whereas when using matrices obtained via partial orthogonalization, some methods are able to achieve a true positive rate of 0.45 approximately. Furthermore, all true discovery rates are also higher when using matrices simulated by partial orthogonalization. Importantly, partial least squares regression performs reasonably good, whereas when only using diagonal dominance one could erroneously conclude that the method is not well fitted for dense structure scenarios. This small real example already serves to highlight the practical application and usefulness of our proposed method.

All the code has been implemented in R (R Core Team, 2018). Algorithm 3 has been implemented directly in C for improved efficiency. We provide an R package, `gmat`, with such implementation, which contains both our method and the dominant diagonal one, available on CRAN[1]. We have also published[2] the R scripts used for generating the data and figures described throughout this section. Thus, all the above described experiments can be replicated.

5. Conclusions and future research

We have proposed a method for generating covariance and concentration matrices subject to the graphical constraints imposed by an undirected graph. The method is an alternative to the most commonly employed approach of imposing a dominant diagonal. As we have shown, the off-diagonal entries in diagonally dominant matrices suffer from a penalization effect that is worsened as the dimension increases. We have empirically shown how our method overcomes the structure recovery difficulties found when validating learning algorithms with the diagonal dominance method.

We have planned several lines of future research. Since we have obtained very promising results when using our method in a real validation scenario, it would be very interesting to explore how other performance measures, and other structure learning algorithms, are also affected. From the

1. `https://CRAN.R-project.org/package=gmat`
2. `https://github.com/irenecrsn/spdug`

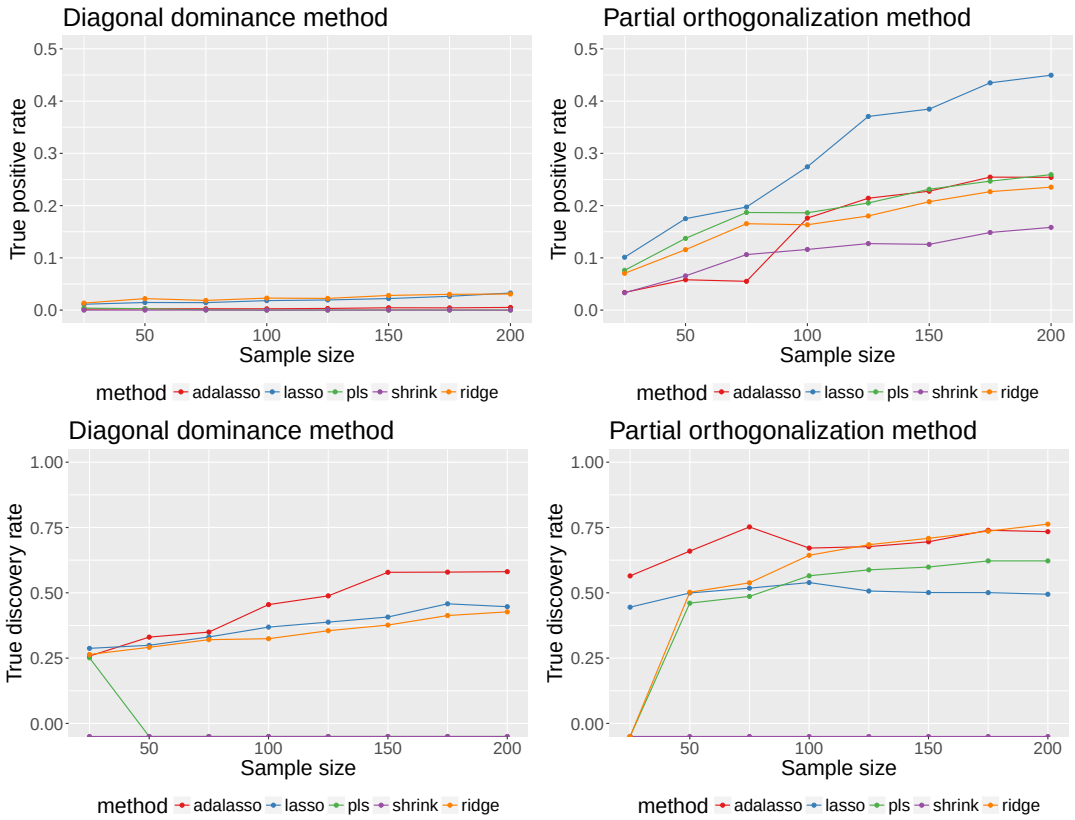

Figure 5: True positive rate and true discovery rate of the structure learning algorithms for concentration graphs validated in (Krämer et al., 2009). The number of variables (vertices in the undirected graph and dimension of the generated matrices) is fixed at 100. `adalasso`: Adaptive l_1 regularization; `lasso`: l_1 regularization; `pls`: partial least squares regression; `shrink`: shrinkage estimator of Schäfer and Strimmer (2005b); `ridge`: l_2 regularization.

computational point of view, exploring alternatives to the modified Gram-Schmidt orthogonalization or taking into account special structures in the graph topology could reduce the complexity of our approach. Further theoretical results on the distribution over $\mathbb{S}^{>0}(G)$, for a graph structure G, induced by our method would help to gain insight in properties we have empirically observed, such as the asymptotic stability in the problem dimension or the relationship with other matrix distributions such as the hyper Wishart family.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Economy, Industry and Competitiveness through the Cajal Blue Brain (C080020-09; the Spanish partner of the Blue Brain initiative from EPFL) and TIN2016-79684-P projects; by the Regional Government of Madrid through

the S2013/ICE-2845-CASI-CAM-CM project; by Fundación BBVA grants to Scientific Research Teams in Big Data 2016. Irene Córdoba has been supported by the predoctoral grant FPU15/03797 from the Spanish Ministry of Education, Culture and Sports. Gherardo Varando has been partially supported by a research grant (13358) from VILLUM FONDEN.

References

E. Arvaniti and M. Claassen. Markov network structure learning via ensemble-of-forests models. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pages 42–51. AUAI Press, 2914.

T. Cai, W. Liu, and X. Luo. A constrained l_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

C. M. Carvalho, H. Massam, and M. West. Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*, 94(3):647–659, 2007.

D. R. Cox and N. Wermuth. Linear dependencies represented by chain graphs. *Statistical Science*, 8(3):204–218, 08 1993.

A. P. Dawid. Conditional independence for statistical operations. *The Annals of Statistics*, 8(3): 598–617, 1980.

A. P. Dawid and S. L. Lauritzen. Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21(3):1272–1317, 1993.

A. P. Dempster. Covariance selection. *Biometrics*, 28(1):157–175, 1972.

M. Drton and T. S. Richardson. Graphical methods for efficient likelihood inference in Gaussian covariance models. *Journal of Machine Learning Research*, 9:893–914, 2008.

P. Erdös and A. Rényi. On random graphs, I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

J. Honorio, D. Samaras, I. Rish, and G. Cecchi. Variable selection for Gaussian graphical models. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, pages 538–546, La Palma, Canary Islands, 2012. PMLR.

S. T. Jensen. Covariance hypotheses which are linear in both the covariance and the inverse covariance. *The Annals of Statistics*, 16(1):302–322, 1988.

W. Joong-Ho, L. Johan, K. Seung-Jean, and R. Bala. Condition-number-regularized covariance estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3): 427–450, 2013.

G. Kauermann. On a dualization of graphical Gaussian models. *Scandinavian Journal of Statistics*, 23(1):105–116, 1996.

K. Khare and B. Rajaratnam. Wishart distributions for decomposable covariance graph models. *The Annals of Statistics*, 39(1):514–555, 2011.

N. Krämer, J. Schäfer, and A.-L. Boulesteix. Regularized estimation of large-scale gene association networks using graphical Gaussian models. *BMC Bioinformatics*, 10(1):384, 2009.

S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

O. Ledoit and M. Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 04 2012.

A. Lenkoski. A direct sampler for G-Wishart variates. *Stat*, 2(1):119–128, 2013.

G. Letac and H. Massam. Wishart distributions for decomposable graphs. *The Annals of Statistics*, 35(3):1278–1323, 2007.

Y. Lin, S. Zhu, D. Lee, and B. Taskar. Learning sparse markov network structure via ensemble-of-trees models. In *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, pages 360–367, Florida, USA, 2009. PMLR.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, 62(12):1707–1739, 2009.

J. Schäfer and K. Strimmer. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21(6):754–764, 2005a.

J. Schäfer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1), 2005b.

I. Stojkovic, V. Jelisavcic, V. Milutinovic, and 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, IJCAI-17*, pages 2758–2764, 2017.

D. R. Williams, J. Piironen, A. Vehtari, and P. Rast. Bayesian estimation of Gaussian graphical models with projection predictive selection. *ArXiv e-prints*, 2018.

L. Xue and H. Zou. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40(5):2541–2571, 2012.

M. Yuan and Y. Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.