ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3315-6701-9/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICASSP55912.2026.11463936

LEARNING GRAPHICAL MODELS UNDER LOW-RANK FACTOR ANALYSIS STRUCTURE

Rui Zhou, Jiaxi Ying†, Wenqiang Pu*, Licheng Zhao*, and Wenjie Wang‡*

* Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong-Shenzhen
†Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong
‡Theory Lab, Central Research Institute, 2012 Labs, Huawei Technologies Co., Ltd., Hong Kong

ABSTRACT

Graph models have attracted considerable attention for its wide applications across various fields, e.g., finance, social science, and genomics. It can be carried out via the maximum likelihood estimation (MLE) of the precision matrix from the observed data. The sparsity is usually induced into the estimation and has shown its success in many real-world applications. Recently, there are some evidences showing that incorporating low-rank factor analysis (FA) structure into graphical models is a promising approach for dimensional reduction so as to improve the graph learning performance. However, it is to be investigated about how to learn such structured graphical models. In this paper, we propose an MLE problem formulation for learning graphical models considering jointly the low-rank FA structure and sparsity pattern on precision matrices. A block coordinate descent algorithm is further proposed to solve our proposed formulation. The superior performance of our proposed formulation and algorithm are verified through numerical simulations on both synthetic data and real financial data.

Index Terms— Graph learning, low-rank FA, sparsity, non-convex optimization algorithms

1. INTRODUCTION

In the era of big data, Gaussian Graphical Models (GGM) have been gaining increasing popularity in various research areas such as bioinformatics, genomics, and financial analysis [1, 2, 3]. The power of GGM stems from its capability to detect, exhibit, and predict the underlying connections among dependent entities. The performance of GGM is closely related to the accuracy of its internal parameters, so the induced model estimation task becomes crucial. The estimation task is named Gaussian graphical model learning and can be formulated as maximum likelihood estimation (MLE) of the precision matrix based on observed data.

It has been extensively verified that incorporating prior knowledge of sparsity in the precision matrix leads to a remarkable improvement in the performance of GGM. The well-known graphical lasso applied the ℓ_1-norm penalized MLE and imposed sparsity elementwisely on the precision matrix [4]. In later studies, researchers proposed to replace the ℓ_1-norm penalty with nonconvex ones, e.g., capped ℓ_1-penalty [5] and smooth clipped absolute deviation [6]. Parallel to graphical lasso, the Bayesian approach was developed to induce sparsity in the precision matrix estimation by Laplace prior [7]. Extensive results illustrate that this approach can well detect the sparse structure of the precision matrix and reveal the conditional dependency among entities. Interested readers may refer to [8, 9, 10, 11] for detailed descriptions of graph modeling performances under sparsity encouragement.

The low-rank factor analysis (FA) model is a prevailing structure in covariance matrix [12]. The principal assumption is that high-dimensional data samples are driven by low-dimensional factors and uncorrelated noise. Therefore, the covariance matrix of high-dimensional data admits a low-rank plus diagonal structure. This structure meets practical needs in several areas such as finance [13], cognitive radio [14]. A recent study has shown that precision matrices also enjoy a low-rank plus diagonal structure in some applications [15].

In this paper, we investigate the problem of learning graphical models, where the precision matrices follow the low-rank FA structure, i.e., low-rank plus diagonal, and sparsity pattern. The key contributions of this paper include:

- We propose a novel optimization problem formulation for learning the GGM, where the precision matrix is sparse and follows the low-rank plus diagonal structure.
- We design an efficient and provably convergent optimization algorithm based on the block coordinate descent framework for solving the proposed formulation.
- Numerical experiments on synthetic and real financial data reveal that our proposed method outperforms the existing methods.

The paper is organized as follows: Sec. 2 presents the problem, Sec. 3 details the proposed algorithm, Sec. 4 reports numerical experiments, and Sec. 5 concludes the paper.

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 62201362, 62571347, and 62206182, in part by the Shenzhen Science and Technology Program under Grant JCYJ20250604191208011, in part by the Shenzhen Low-Altitude Airspace Strategic Program Portfolio under Grant Z25306102, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515010154, and in part by SRIBD under Grant J00120260001. *(Correspondence: Wenqiang Pu)*

2. PROBLEM FORMULATION

In this section, we introduce the problem formulation of MLE for modeling the GGM when the precision matrix is sparse and admits the low-rank FA structure. After attempts, we found that directly imposing both structures in the MLE formulation leads to an very intractable problem. Therefore we propose to first reformulate the precision matrix estimation problem with only low-rank FA structure, and then introduce the sparsity in the constraint of the obtained reformulation.

Suppose n observations of p dimensional random variables $\{\boldsymbol{x}_i\}_{i=1}^n$ ($\boldsymbol{x}_i \in \mathbb{R}^p$) are generated from a GGM model, the MLE problem on the precision matrix under the low-rank FA structure can be written as

$$\begin{aligned} \underset{\boldsymbol{M},\boldsymbol{\Psi}}{\mathsf{min}} \quad & \mathrm{Tr}\left(\left(\boldsymbol{M}+\boldsymbol{\Psi}\right)\boldsymbol{S}\right) - \log\det\left(\boldsymbol{M}+\boldsymbol{\Psi}\right) \\ \mathsf{s.t.} \quad & \mathrm{rank}\left(\boldsymbol{M}\right) \leq r, \boldsymbol{M} \succeq \boldsymbol{0}, \boldsymbol{\Psi} \succ \boldsymbol{0}, \end{aligned} \tag{1}$$

where $\boldsymbol{M}$ is the low-rank component with rank r ($r < p$), $\boldsymbol{\Psi} = \mathrm{Diag}\left(\psi_1, \ldots, \psi_n\right)$ is the diagonal component, and $\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T$ is the sample covariance matrix from observations. We define a new variable $\tilde{\boldsymbol{M}}$ as $\tilde{\boldsymbol{M}} = \boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{M}\boldsymbol{\Psi}^{-\frac{1}{2}}$. Then the above problem can be rewritten as

$$\begin{aligned} \underset{\tilde{\boldsymbol{M}},\boldsymbol{\Psi}}{\mathsf{min}} \quad & \mathrm{Tr}\left(\boldsymbol{\Psi}^{\frac{1}{2}}\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)\boldsymbol{\Psi}^{\frac{1}{2}}\boldsymbol{S}\right) - \log\det\boldsymbol{\Psi}^{\frac{1}{2}}\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)\boldsymbol{\Psi}^{\frac{1}{2}} \\ \mathsf{s.t.} \quad & \mathrm{rank}(\tilde{\boldsymbol{M}}) \leq r,\ \tilde{\boldsymbol{M}} \succeq \boldsymbol{0},\ \boldsymbol{\Psi} \succ \boldsymbol{0}. \end{aligned} \tag{2}$$

Now we consider to promote the sparsity in the precision matrix by putting ℓ_1-norm constraint on the off-diagonal elements of $\tilde{\boldsymbol{M}}$, i.e.,

$$\begin{aligned} \underset{\tilde{\boldsymbol{M}},\boldsymbol{\Psi}}{\mathsf{min}} \quad & \mathrm{Tr}\left(\boldsymbol{\Psi}^{\frac{1}{2}}\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)\boldsymbol{\Psi}^{\frac{1}{2}}\boldsymbol{S}\right) - \log\det\boldsymbol{\Psi}^{\frac{1}{2}}\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)\boldsymbol{\Psi}^{\frac{1}{2}} \\ \mathsf{s.t.} \quad & \|\tilde{\boldsymbol{M}}\|_{\mathrm{off},1} \leq \rho,\ \mathrm{rank}(\tilde{\boldsymbol{M}}) \leq r,\ \tilde{\boldsymbol{M}} \succeq \boldsymbol{0},\ \boldsymbol{\Psi} \succ \boldsymbol{0}. \end{aligned} \tag{3}$$

where ρ is a tuning parameter controlling the sparsity level in $\tilde{\boldsymbol{M}}$. Note that the estimated precision matrix $\boldsymbol{\Theta}$ will be recovered as $\boldsymbol{\Theta} = \boldsymbol{\Psi}^{\frac{1}{2}}(\tilde{\boldsymbol{M}}+\boldsymbol{I})\boldsymbol{\Psi}^{\frac{1}{2}}$, which means $\Theta_{ij} = 0$ iff $\tilde{M}_{ij} = 0$ ($i, j = 1, \ldots, p$).

Following [16, Sec. IV], we replace the low-rank constraint $\mathrm{rank}(\tilde{\boldsymbol{M}}) \leq r$ and positive semidefinite constraint $\tilde{\boldsymbol{M}} \succeq \boldsymbol{0}$ with $\tilde{\boldsymbol{M}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$, where $\boldsymbol{U} \in \mathbb{R}^{p\times r}$ and $\boldsymbol{\Lambda} = \mathrm{Diag}\left(\lambda_1, \ldots, \lambda_r\right)$ are new dummy variables satisfying $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_r$ and $\boldsymbol{\lambda} = \left[\lambda_1, \ldots, \lambda_r\right]^T \geq \boldsymbol{0}$. Then the hard constraint $\tilde{\boldsymbol{M}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ is further relaxed as the regularization term $\frac{\tau}{2}\|\tilde{\boldsymbol{M}} - \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\|_F^2$ in the objective. Finally the problem formulation is given in (4).

In what follows, we consider solving the Problem (4) to obtain the precision matrix estimate. Although it is still a nonconvex problem, we can develop an efficient algorithm based on the block coordinate descent (BCD) framework [17] as in the next section.

3. PROPOSED ALGORITHM

In this section, we propose the graphical learning with low-rank FA and sparsity structures (GFA) algorithm to solve Problem (4). It is designed by cyclically updating blocks $(\tilde{\boldsymbol{M}}, \boldsymbol{U}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$.

Update of $\tilde{\boldsymbol{M}}$: The sub-problem on $\tilde{\boldsymbol{M}}$ is

$$\begin{aligned} \underset{\tilde{\mathbf{M}}}{\mathsf{min}} \quad & \|\tilde{\boldsymbol{M}} - \boldsymbol{\Omega}\|_F^2 \\ \mathsf{s.t.} \quad & \tilde{\boldsymbol{M}} \succeq \boldsymbol{0},\ \|\tilde{\boldsymbol{M}}\|_{\mathrm{off},1} \leq \rho, \end{aligned} \tag{5}$$

where $\boldsymbol{\Omega} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T - \frac{1}{\tau}\boldsymbol{\Psi}^{\frac{1}{2}}\boldsymbol{S}\boldsymbol{\Psi}^{\frac{1}{2}}$. This is a problem of projection onto the intersection of two convex sets in Hilbert spaces, which is a convex problem and can be solved by several toolboxes, e.g., CVX [18, 19]. For computational efficiency, we propose to solve it by Dykstra's projection algorithm [20]. The details are given in Algorithm 1, where $\prod_{\mathbb{S}_+}(\boldsymbol{Z})$ is projecting a symmetric matrix $\boldsymbol{Z}$ onto the positive semidefinite cone and can be achieved simply by eigenvalue decomposition [21], $\prod_{\ell_1}(\boldsymbol{Z})$ is projecting a symmetric matrix $\boldsymbol{Z}$ onto the specific ℓ_1 norm ball $\{\boldsymbol{Z} \mid \|\boldsymbol{Z}\|_{\mathrm{off},1} \leq \rho\}$ and can be very efficiently obtained by a water-filling-like algorithm [22, Lemma 1][23].

Algorithm 1 Dykstra's projection algorithm for solving $\tilde{\mathbf{M}}$ sub-problem (5).

1: Initialize $\tilde{\boldsymbol{M}}^0 = \boldsymbol{\Omega}$, $\boldsymbol{P}^0 = \boldsymbol{Q}^0 = \boldsymbol{0}$,
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: $\quad \boldsymbol{Y}^k = \prod_{\mathbb{S}_+}\left(\tilde{\mathbf{M}}^k + \mathbf{P}^k\right)$;
4: $\quad \boldsymbol{P}^{k+1} = \tilde{\boldsymbol{M}}^k + \boldsymbol{P}^k - \boldsymbol{Y}^k$;
5: $\quad \tilde{\boldsymbol{M}}^{k+1} = \prod_{\ell_1}\left(\boldsymbol{Y}^k + \boldsymbol{Q}^k\right)$;
6: $\quad \boldsymbol{Q}^{k+1} = \boldsymbol{Y}^k + \boldsymbol{Q}^k - \tilde{\boldsymbol{M}}^{k+1}$;
7: $\quad$ Terminate when converges;
8: **end for**
9: Return $\tilde{\boldsymbol{M}}^{k+1}$.

Update of $\boldsymbol{\lambda}$: The sub-problem on $\boldsymbol{\lambda}$ is

$$\underset{\boldsymbol{\lambda}\geq\mathbf{0}}{\mathsf{min}} \quad \frac{\tau}{2}\|\tilde{\mathbf{M}} - \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\|_F^2 - \sum_{i=1}^r \log\left(\lambda_i + 1\right), \tag{6}$$

which can be rewritten and simplified as r sub-problems. For any λ_i, the problem is given as [16]

$$\underset{\lambda_i\geq 0}{\mathsf{min}} \quad \frac{\tau}{2}\left(\lambda_i - b_i\right)^2 - \log\left(\lambda_i + 1\right), \tag{7}$$

where $b_i = [\boldsymbol{U}\tilde{\boldsymbol{M}}\boldsymbol{U}^T]_{ii}$. The closed-form solution of the sub-problem (7) is given in the following lemma.

Lemma 1. *The optimal solution to Problem (7) is*

$$\lambda_i^\star = \frac{1}{2}\left[b_i - 1 + \sqrt{(b_i-1)^2 + 4(b_i + 1/\tau)}\right].$$

$$\begin{array}{ll} \underset{\tilde{\boldsymbol{M}},\boldsymbol{U},\boldsymbol{\Lambda},\boldsymbol{\Psi}}{\mathsf{min}} & \mathrm{Tr}\left(\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)\boldsymbol{\Psi}^{\frac{1}{2}}\boldsymbol{S}\boldsymbol{\Psi}^{\frac{1}{2}}\right)-\sum_{i=1}^{r}\log\left(\lambda_i+1\right)-\log\det\left(\boldsymbol{\Psi}\right)+\frac{\tau}{2}\left\|\tilde{\boldsymbol{M}}-\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\right\|_F^2 \\ \mathsf{s.t.} & \tilde{\boldsymbol{M}}\succeq\boldsymbol{0},\ \|\tilde{\boldsymbol{M}}\|_{\mathrm{off},1}\leq\rho,\ \boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}_r,\ \boldsymbol{\lambda}\geq\boldsymbol{0},\ \boldsymbol{\Psi}=\mathrm{Diag}\left(\psi_1,\ldots,\psi_p\right)\succ\boldsymbol{0}. \end{array} \tag{4}$$

Proof. Taking the gradient of objective of Problem (7) being zero, we obtain the following equation to solve

$$\tau\left(\lambda_i-b_i\right)-\frac{1}{\lambda_i+1}=0.$$

Since $\tau>0$ and $b_i\geq 0$, the above equation must admit two solutions, which are located within region $(-\infty,-1)$ and $(0,+\infty)$, respectively. Note that its objective is a convex function when $\lambda_i\geq 0$. Then the solution within $(0,+\infty)$ must be the optimal solution of Problem (7). □

Update of $\boldsymbol{\Psi}$: The sub-problem on $\boldsymbol{\Psi}$ is

$$\begin{array}{ll} \underset{\boldsymbol{\Psi}}{\mathsf{min}} & \mathrm{Tr}\left(\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)\boldsymbol{\Psi}^{\frac{1}{2}}\boldsymbol{S}\boldsymbol{\Psi}^{\frac{1}{2}}\right)-\log\det\left(\boldsymbol{\Psi}\right) \\ \mathsf{s.t.} & \boldsymbol{\Psi}=\mathrm{Diag}\left(\psi_1,\ldots,\psi_p\right), \end{array} \tag{8}$$

Denote $\boldsymbol{\alpha}=\mathrm{diag}\left(\boldsymbol{\Psi}^{\frac{1}{2}}\right)$ and we can rewrite the above problem as

$$\begin{array}{ll} \underset{\boldsymbol{\alpha}}{\mathsf{min}} & \boldsymbol{\alpha}^T\boldsymbol{\Gamma}\boldsymbol{\alpha}-\sum_{i=1}^{p}\log\alpha_i^2 \\ \mathsf{s.t.} & \boldsymbol{\alpha}\geq\boldsymbol{0}, \end{array} \tag{9}$$

where $\boldsymbol{\Gamma}=\boldsymbol{S}\odot\left(\tilde{\boldsymbol{M}}+\boldsymbol{I}\right)$ is guaranteed to be positive definite since $\boldsymbol{S},\tilde{\boldsymbol{M}}+\boldsymbol{I}\succ\boldsymbol{0}$ [24]. This is a convex problem and can be efficiently solved by a gradient descent method [16, Algorithm 2].

Update of $\boldsymbol{U}$: The sub-problem on $\boldsymbol{U}$ is

$$\begin{array}{ll} \underset{\boldsymbol{U}}{\mathsf{min}} & \|\tilde{\boldsymbol{M}}-\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\|_F^2 \\ \mathsf{s.t.} & \boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}_r, \end{array} \tag{10}$$

The optimal solution of the above problem is r principle eigenvectors of $\tilde{\boldsymbol{M}}$ with the corresponding eigenvalues following the same order of $\boldsymbol{\Lambda}$ [16].

The overall GFA algorithm is summarized in Algorithm 2. The convergence analysis of our proposed FANC algorithm is given in the following.

Proposition 1. *The sequence* $\left\{\tilde{\boldsymbol{M}}^k,\boldsymbol{\lambda}^k,\boldsymbol{\Psi}^k,\boldsymbol{U}^k\right\}$ *generated by Algorithm 2 converges to the set of stationary points of Problem (4).*

Proposition 1 shows that the sequence $\left\{\tilde{\boldsymbol{M}}^k,\boldsymbol{\lambda}^k,\boldsymbol{\Psi}^k,\boldsymbol{U}^k\right\}$ generated by Algorithm 2 has at least one limit point, and each limit point is a stationary point of Problem (4). The convergence proof of Algorithm 2 directly follows from the convergence proof part in [25, Theorem 5]. Note that the Proposition 1 is guaranteed without conditions on the updating order, while different updating orders may lead to different stationary points, owing to the non-convex nature of the problem.

Algorithm 2 GFA algorithm for solving Problem (4).

1: Initialize $\tilde{\boldsymbol{M}}^0$, $\boldsymbol{\lambda}^0$,$\boldsymbol{\Psi}^0$, $\boldsymbol{U}^0$.
2: **for** $k=0,1,2,\ldots$ **do**
3: Update $\tilde{\boldsymbol{M}}^{k+1}$ via Algorithm 1;
4: Update $\boldsymbol{\lambda}^{k+1}$ as in Lemma 1;
5: Update $\boldsymbol{\Psi}^{k+1}$ via [16, Algorithm 2];
6: Update $\boldsymbol{U}^{k+1}$ as [16, Lemma 2];
7: Terminate when converges;
8: **end for**
9: Return $\boldsymbol{\Theta}=(\boldsymbol{\Psi}^{k+1})^{\frac{1}{2}}\left(\tilde{\boldsymbol{M}}^{k+1}+\boldsymbol{I}\right)(\boldsymbol{\Psi}^{k+1})^{\frac{1}{2}}$.

4. NUMERICAL EXPERIMENTS

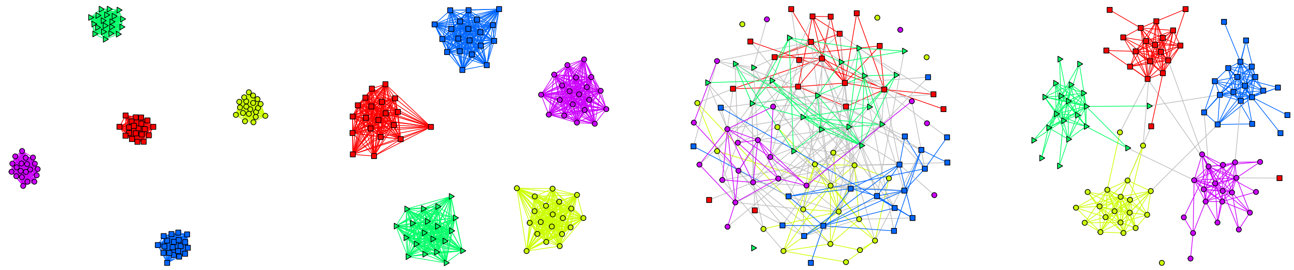
Here we perform numerical experiments to show the performance of our proposed formulation and algorithm, using both synthetic data and real financial data. The experiments also includes several graph learning benchmarks for comparison.

4.1. Synthetic Data

The simulation is set with $p=100$, $r=5$, and $n=400$, meaning the sample size is not much larger than the number of variables. We generate the precision matrix $\boldsymbol{\Theta}$ as in [16, Sec. VI-A], resulting in r separate groups in the true graph. Data $\{\boldsymbol{x}_i\}_{i=1}^n$ is sampled independently from $\mathcal{N}(\boldsymbol{0},(\boldsymbol{\Theta}^\star)^{-1})$. For the GFA method, we set $\tau=10$, $\rho=200$, and $r=5$. For GLasso[1] [4], we do not penalize the diagonal and tune the regularization parameter until some points are isolated. For Bagus[2] [7], the parameter is tuned using the Bayesian information criterion as recommended.

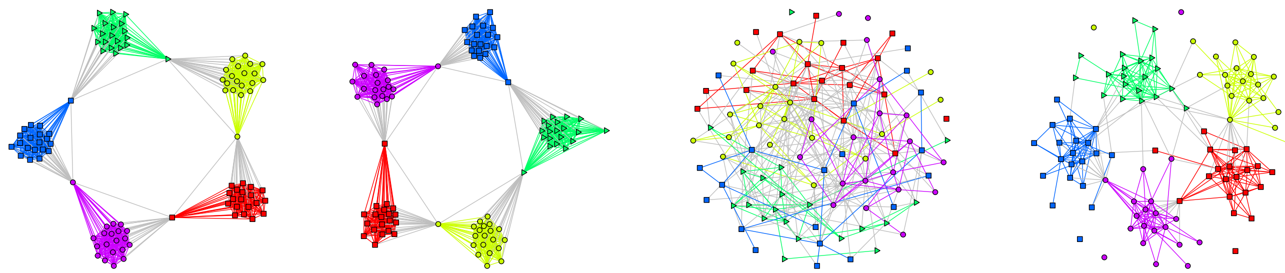
We consider two types of graphs and present both true graphs and results of learned graphs in Fig. 1 and Fig. 2. In Fig. 1, we use the aforementioned setting thus points can be divided into $r=5$ separate groups. The five separate groups are labelled with different colors. The Bagus method can substantially recover five groups with a few miss-connecting edges, and the Glasso method cannot identify the five groups. They fail to distinguish five groups because they only consider to promote sparsity without the low-rank FA structure

[1]Glasso: https://cran.r-project.org/web/packages/glasso.
[2]Bagus: https://github.com/garyganuiuc/SSLasso.



(a) True graph. (b) GFA (proposed). (c) GLasso [4]. (d) Bagus [7].

Fig. 1: The true and learned graphs with data from separate groups.



(a) True graph. (b) GFA (proposed). (c) GLasso [4]. (d) Bagus [7].

Fig. 2: The true and learned graphs with data from connected groups.

information in estimation. It is significant that our proposed GFA method almost perfectly recovers the graph. It is reasonable because the GFA method is able to make use of the low-rank FA structure, while other methods can not. In Fig. 2, we generate the data using the connected graphs by manually setting r inter-group points. The parameter settings of these methods are the same as in Fig. 1. The results are consistent and our proposed GFA method is the only one successfully distinguishing five groups and those inter-group points.

4.2. Real-World Data

We now use the real financial data to check the performance of our proposed GFA method. Since the true precision matrix of real-world data is unknown, the quality of the estimated precision matrix is evaluated through backtesting the global minimum variance portfolio (GMVP), i.e., $\boldsymbol{w}^{\star} = \hat{\boldsymbol{\Theta}}\mathbf{1}/\left(\mathbf{1}^T\hat{\boldsymbol{\Theta}}\mathbf{1}\right)$ [26]. We conduct backtests following [16, Sec. VI-B], with a lookback window of 60 days and a testing window of 20 days. For GFA, parameters are set to $r = 5$, $\tau = 1$, and $\rho = 100$. GLasso and Bagus parameters are set as in Sec. 4.1.

We present, in Fig. 3, the boxplot of the annual volatility from backtest results. The GMVP using simply sample covariance matrix (SCM) estimate is also included for comparison. It is clear that the SCM method performs the worst, i.e., the highest annual volatility, because the number of samples (60) is not significantly larger than the number of stocks (50). The GLasso and Bagus methods perform better than the SCM method, which may be explained as the advantage of

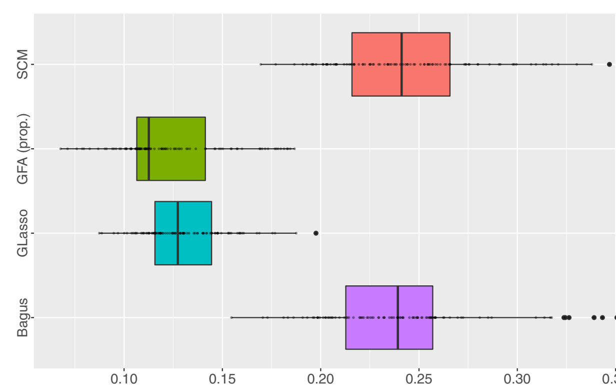


Fig. 3: Boxplot of annual volatility (the lower the better).

promoting sparsity. Significantly, our proposed GFA method is able to outperform the other methods and achieve the lowest volatility. It is because GFA is able to simultaneously consider low-rank FA and sparsity structures.

5. CONCLUSION

In this paper, we have considered the graphical model estimation problem with the low-rank factor analysis structure. We have formulated the problem and proposed a provably convergent GFA algorithm based on the block coordinate descent method. Numerical experiments have shown the efficiency of our algorithm and the superiority of our formulation.

6. REFERENCES

[1] H. Li and J. Gui, "Gradient directed regularization for sparse gaussian concentration graphs, with applications to inference of genetic networks," *Biostatistics*, vol. 7, no. 2, pp. 302–317, 2006.

[2] Y. Park, D. Hallac, S. Boyd, and J. Leskovec, "Learning the network structure of heterogeneous data via pairwise exponential markov random fields," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1302–1310.

[3] J. V. de Miranda Cardoso, J. Ying, and D. Palomar, "Graphical models in heavy-tailed markets," *Advances in Neural Information Processing Systems*, vol. 34, pp. 19 989–20 001, 2021.

[4] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[5] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization." *Journal of Machine Learning Research*, vol. 11, no. 3, 2010.

[6] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[7] L. Gan, N. N. Narisetty, and F. Liang, "Bayesian regularization for graphical models with unequal shrinkage," *Journal of the American Statistical Association*, vol. 114, no. 527, pp. 1218–1231, 2019.

[8] W. Huang, T. A. Bolton, J. D. Medaglia, D. S. Bassett, A. Ribeiro, and D. Van De Ville, "A graph signal processing perspective on functional brain imaging," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 868–885, 2018.

[9] Y. He and H.-T. Wai, "Detecting central nodes from low-rank excited graph signals via structured factor analysis," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2416–2430, 2022.

[10] N. Deb, A. Kuceyeski, and S. Basu, "Regularized estimation of sparse spectral precision matrices," 2024. [Online]. Available: https://arxiv.org/abs/2401.11128

[11] H. Shu, Z. Chen, Y. Zhang, and H. Zhu, "Nodewise loreg: Nodewise l_0-penalized regression for high-dimensional sparse precision matrix estimation," 2024. [Online]. Available: https://arxiv.org/abs/2406.06481

[12] R. S. Tsay, *Analysis of Financial Time Series*. John wiley & sons, 2005.

[13] Y. Feng and D. P. Palomar, "A signal processing perspective on financial engineering," *Foundations and Trends® in Signal Processing*, vol. 9, no. 1–2, pp. 1–231, 2016.

[14] R. Zhou, W. Pu, M.-Y. You, and Q. Shi, "A robust cooperative sensing approach for incomplete and contaminated data," *IEEE Transactions on Signal Processing*, vol. 72, pp. 3945–3957, 2024.

[15] N. K. Chandra, P. Mueller, and A. Sarkar, "Bayesian scalable precision factor analysis for gaussian graphical models," *Bayesian Analysis*, vol. 1, no. 1, pp. 1–29, 2024.

[16] R. Zhou, J. Ying, and D. P. Palomar, "Covariance matrix estimation under low-rank factor model with nonnegative correlations," *IEEE Transactions on Signal Processing*, vol. 70, pp. 4020–4030, 2022.

[17] S. J. Wright, "Coordinate descent algorithms," *Mathematical Programming*, vol. 151, no. 1, pp. 3–34, 2015.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[19] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming," 2014.

[20] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in hilbert spaces," in *Advances in order restricted statistical inference*. Springer, 1986, pp. 28–47.

[21] N. J. Higham, "Computing a nearest symmetric positive semidefinite matrix," *Linear algebra and its applications*, vol. 103, pp. 103–118, 1988.

[22] D. P. Palomar, "Convex primal decomposition for multicarrier linear mimo transceivers," *IEEE Transactions on Signal Processing*, vol. 53, no. 12, pp. 4661–4674, 2005.

[23] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the ℓ_1-ball for learning in high dimensions," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 272–279.

[24] F. Zhang, *The Schur Complement and Its Applications*. Springer Science & Business Media, 2006.

[25] S. Kumar, J. Ying, J. V. de Miranda Cardoso, and D. P. Palomar, "A unified framework for structured graph learning via spectral constraints." *Journal of Machine Learning Research*, vol. 21, no. 22, pp. 1–60, 2020.

[26] R. Agrawal, U. Roy, and C. Uhler, "Covariance matrix estimation under total positivity for portfolio selection," *Journal of Financial Econometrics*, 2020, nbaa018.