Minimax Estimation of Laplacian Constrained Precision Matrices

Jiaxi Ying **José Vinícius de M. Cardoso** **Daniel P. Palomar**

The Hong Kong University of Science and Technology, Hong Kong

Abstract

This paper considers the problem of high-dimensional sparse precision matrix estimation under Laplacian constraints. We prove that the Laplacian constraints bring favorable properties for estimation: the Gaussian maximum likelihood estimator exists and is unique almost surely on the basis of one observation, irrespective of the dimension. We establish the optimal rate of convergence under Frobenius norm by the derivation of the minimax lower and upper bounds. The minimax lower bound is obtained by applying Le Cam-Assouad's method with a novel construction of a subparameter space of multivariate normal distributions. The minimax upper bound is established by designing an adaptive ℓ_1-norm regularized maximum likelihood estimation method and quantifying the rate of convergence. We prove that the proposed estimator attains the optimal rate of convergence with an overwhelming probability. Numerical experiments demonstrate the effectiveness of the proposed estimator.

1 Introduction

Estimating high-dimensional precision matrices is a crucial problem in a number of fields such as bioinformatics, genomics, and finance. For example, knowledge of the precision matrix is useful for clustering and discriminant analysis. Precision matrix is closely related to graphical models which provide a powerful tool in modeling the relationships among a large number of random variables, and have been widely explored in many applications such as gene expression analysis, functional magnetic resonance imaging, risk management, and portfolio allocation (Lauritzen, 1996; Yuan and Lin, 2007; Park et al., 2017; Banerjee et al., 2008; Agrawal et al., 2019).

To avoid the curse of dimensionality, various structural assumptions are imposed on the models such as sparsity, bandability, and total positivity. Accordingly, many regularized or constrained estimators have been proposed to estimate the precision matrices under model assumptions. One popular assumption is sparsity, and the sparse precision matrix estimation has been extensively studied (d'Aspremont et al., 2008; Friedman et al., 2008; Mazumder and Hastie, 2012; Rothman et al., 2008; Ravikumar et al., 2011; Wang et al., 2016; Honorio et al., 2012). Bandability model arises from many applications such as time series data (Bickel and Gel, 2011), and assumes that the entries of the underlying precision matrix decay based on their distance from the diagonal. Tapering methods were developed to estimate bandable precision matrices (Hu and Negahban, 2017) or precision matrices with bandable Cholesky factor (Liu et al., 2020). Total positivity (Fallat et al., 2017; Slawski and Hein, 2015; Soloff et al., 2020; Wang et al., 2020; Lauritzen et al., 2019a,b), as a special form of positive dependence between random variables, has received wide attention in portfolio design (Agrawal et al., 2019) and auction theory (Milgrom and Weber, 1982), where the precision matrix is an M-matrix under multivariate Gaussian distribution.

Recently, there is growing interest in estimating Laplacian constrained precision matrices in the field of signal processing and machine learning over graphs (Dong et al., 2019; Lake and Tenenbaum, 2010; Egilmez et al., 2017; Kumar et al., 2019, 2020; Ying et al., 2020a,b; Cardoso et al., 2020). In this problem, the underlying precision matrix is a Laplacian matrix. By imposing the Laplacian constraints, estimating a precision matrix can be viewed as learning a graph Laplacian, which plays an important role in graph Fourier transform: its eigenvalues and eigenvectors can be interpreted as spectral frequencies and Fourier basis (Shuman et al., 2013). More recently, the

Proceedings of the 24[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130.

penalized maximum likelihood estimation of Laplacian constrained precision matrices has been explored with the convex or nonconvex regularizer (Ortega et al., 2018; Zhao et al., 2019; Ying et al., 2020a,b), and the rate of convergence has been obtained in Ying et al. (2020a). However, the rate optimality is still unknown in the literature. This paper aims to establish the minimax optimal rate of convergence for estimating sparse Laplacian constrained precision matrices.

The minimax risk is one of the most commonly used benchmark to quantify the difficulty of an estimation problem. Minimax rates of convergence have been established for estimating several classes of structured covariance and precision matrices, including sparse, banded covariance matrices (Bickel and Gel, 2011; Cai et al., 2010; Cai and Zhou, 2012; Cai et al., 2016c; Chen et al., 2018) or precision matrices (Cai et al., 2016b; Ren et al., 2015; Hu and Negahban, 2017; Liu et al., 2020). Cai et al. (2016b) established the optimal rates of convergence for estimating sparse precision matrices under a range of losses. Hu and Negahban (2017) established the minimax estimation bounds for estimating banded precision matrices under the spectral norm. Recently, Liu et al. (2020) obtained the optimal rates of convergence for estimating precision matrices with bandable Cholesky factor under the operator norm and the Frobenius norm. More recently, Soloff et al. (2020) established the optimal rates of convergence for estimating the precision matrices with total positivity under the symmetrized Stein loss. See (Cai et al., 2016c; Kim, 2020) for comprehensive reviews on this topic. However, no such theoretical results have been established for the sparse Laplacian constrained precision matrix estimation. It is worth mentioning that the optimal rate of convergence for estimating the sparse Laplacian constrained precision matrices in this paper cannot be derived from the work above, because the parameter space of interest is disjoint with those in the literature.

Our contributions: In this paper, we consider the problem of estimating high-dimensional sparse Laplacian constrained precision matrices. The contributions of this paper are summarized as follows:

- It is well-known that the Gaussian maximum likelihood estimator (MLE) of a general precision matrix does not exist if the number of variables is larger than the number of observations. However, we prove that the Laplacian constraints can reduce the necessary number of observations to make the MLE well-defined: the Gaussian maximum likelihood estimator under Laplacian constraints exists and is unique almost surely with only one observation, regardless of the dimension.
- The minimax lower bound characterizes the fundamental difficulty of an estimation problem, and a desired estimator is expected to achieve the lower bound. In this paper, we obtain the minimax lower bound under Frobenius norm for estimating the sparse Laplacian constrained precision matrices, which can be a benchmark in evaluating rate optimality of the estimators.
- We propose an estimator by solving an adaptive ℓ_1-regularized maximum likelihood estimation problem with Laplacian constraints, and establish its rate of convergence. The upper and lower bounds together yield the optimal rate of convergence, and show that the proposed estimator attains the optimal rate of convergence with an overwhelming probability.
- Numerical experiments demonstrate the effectiveness of the proposed estimator, and show that for the convex formulations, we should use adaptive ℓ_1-norm rather than ℓ_1-norm for edge recovery in our problem.

The remainder of the paper is organized as follows. The background of Gaussian graphical models, and favorable properties of Laplacian constraints are provided in Section 2. We propose a new estimator, and establish the optimal rate of convergence in Section 3. Experimental results are provided in Section 4, and conclusions are drawn in Section 5. An open source R package is available at `https://github.com/mirca/sparseGraph` for the experiments in this paper.

Notation: Both X_{ij} and $[\boldsymbol{X}]_{ij}$ denote the (i,j)-th entry of the matrix $\boldsymbol{X}$. $[p]$ denotes the set $\{1,\ldots,p\}$. $\|\boldsymbol{x}\|$, $\|\boldsymbol{X}\|_{\mathrm{F}}$, and $\|\boldsymbol{X}\|_2$ denote Euclidean norm, Frobenius norm, and operator norm, respectively. Let $\|\boldsymbol{x}\|_{\max} = \max_i |x_i|$ and $\|\boldsymbol{x}\|_{\min} = \min_i |x_i|$. $I\{A\}$ denotes the indicator function of the event A. $\lfloor x \rfloor$ denotes the largest integer less than or equal to x, and $\lceil x \rceil$ denotes the least integer larger than or equal to x. $\mathcal{S}_+^p$ and $\mathcal{S}_{++}^p$ denote the sets of positive semi-definite and positive definite matrices with the size $p \times p$, respectively. For functions $f(n)$ and $g(n)$, we use $f(n) \lesssim g(n)$ if $f(n) \leq cg(n)$, and $f(n) \gtrsim g(n)$ if $f(n) \geq Cg(n)$, for some constants $c, C \in (0, +\infty)$.

2 Laplacian Constrained Gaussian Graphical Models

In this section, we introduce the definition and motivations of Laplacian constrained Gaussian graphical models, and present favorable properties of Laplacian constraints for estimating precision matrices in the high-dimensional regime.

The Laplacian constrained precision matrix is closely related to the Laplacian constrained Gaussian Markov random fields (Ying et al., 2020a). Define an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{W})$, where $\mathcal{V}$ and $\mathcal{E}$ denote the set of nodes and edges, respectively, and each element W_{ij} of $\boldsymbol{W}$ denotes the graph weight between nodes i and j. Then the graph Laplacian $\boldsymbol{L} \in \mathbb{R}^{p\times p}$ is defined by $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, where $\boldsymbol{D}$ is the degree matrix. We say a graph is connected if there is a path from any node to any other node in the graph. It is well-known that the rank of the Laplacian matrix for a connected graph with p nodes is $p-1$. Then the set of all Laplacian matrices that correspond to connected graphs can be formulated as

$$\mathcal{S}_L = \Big\{ \boldsymbol{\Omega} \in \mathbb{R}^{p\times p} \mid \Omega_{ij} = \Omega_{ji} \leq 0, \forall\, i \neq j, \sum_j \Omega_{ij} = 0, \forall\, i \in [p], \operatorname{rank}(\boldsymbol{\Omega}) = p-1 \Big\}. \quad (1)$$

In what follows, we present Laplacian constrained Gaussian Markov random fields, and without loss of generality we assume the random vector $\boldsymbol{x}$ has zero mean.

Definition 1 (Ying et al. (2020a))**.** A zero-mean random vector $\boldsymbol{x} = [x_1, \ldots, x_p]^\top \in V^{p-1}$ is called a Laplacian-structured Gaussian Markov Random Fields (L-GMRF) with parameters $(\mathbf{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} \in \mathcal{S}_L$, if and only if its density function $q_L : V^{p-1} \to \mathbb{R}$ follows

$$q_L(\boldsymbol{x}) = (2\pi)^{-\frac{p-1}{2}} \det{}^\star(\boldsymbol{\Omega})^{\frac{1}{2}} \exp\Big(-\frac{1}{2} \boldsymbol{x}^\top \boldsymbol{\Omega} \boldsymbol{x} \Big), \quad (2)$$

where $\det^\star$ denotes the pseudo determinant defined by the product of nonzero eigenvalues Holbrook (2018), and $V^{p-1} := \{\boldsymbol{x} \in \mathbb{R}^p \mid \mathbf{1}^\top \boldsymbol{x} = 0\}$.

Now the Gaussian maximum likelihood estimator of the Laplacian constrained precision matrix can be formulated as

$$\boldsymbol{\Omega}_O^n := \arg\min_{\boldsymbol{\Omega} \in \mathcal{S}_L} -\log\det{}^\star(\boldsymbol{\Omega}) + \operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{S}), \quad (3)$$

where $\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^\top$ is the sample covariance matrix, and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are *i.i.d.* samples drawn from L-GMRF.

Theorem 1. *The maximum likelihood estimator $\boldsymbol{\Omega}_O^n$ defined in* (3) *exists and is unique almost surely as long as the sample size $n \geq 1$.*

Note that the maximum likelihood estimator in the general Gaussian graphical models does not exist if the number of observations is smaller than the number of variables, i.e., $n < p$. Therefore, the property presented in Theorem 1 is useful in the high-dimensional regime where the number of observations are usually less than the number of nodes.

We finish the section with discussions on the motivations of the Laplacian constrained Gaussian graphical models. Unlike the general Gaussian graphical models, the elements of the Laplacian constrained precision matrix can quantify the expected distance between random variables. More specifically, due to the Laplacian constraints, we can rewrite the probability density function in (2) as

$$q_L(\boldsymbol{x}) = (2\pi)^{-\frac{p-1}{2}} \det{}^\star(\boldsymbol{\Theta})^{\frac{1}{2}} \exp\Big(-\frac{1}{2} \sum_{i\neq j} W_{ij} \left(x_i - x_j\right)^2 \Big),$$

where $W_{ij} = -\Theta_{ij} \geq 0$. We can see that there will be a relatively high probability that $\left(x_i - x_j\right)^2$ is small if the element W_{ij} is large. Such property is desired in modelling smooth graphs where a large graph weight between two nodes represents a significant similarity between their signal values. The Laplacian constrained Gaussian graphical models have been widely explored in graph signal processing (Dong et al., 2016) and semi-supervised learning (Zhu et al., 2003), where the underlying graphs are usually assumed smooth. Some real-world datasets such as stock data fit the Laplacian constrained Gaussian graphical models well (Cardoso et al., 2020; Cardoso and Palomar, 2020).

3 Rate Optimality under Frobenius Norm

In this section, we first propose an adaptive estimator in Section 3.1, and establish the rate of convergence which provides an upper bound for the minimax risk. Then we establish a matching risk lower bound in Section 3.2 by applying testing arguments through the Le Cam-Assouad's method over a constructed sub-parameter space. The upper and lower bounds together yield the optimal rate of convergence for sparse Laplacian constrained precision matrix estimation. At the meantime, we show that the proposed estimator attains the optimal rate of convergence under the Frobenius norm with an overwhelming probability.

3.1 Adaptive estimator and Minimax Upper Bound

To establish the minimax upper bound, we first propose a new method of Adaptive Laplacian constrained Precision matrix Estimation (ALPE), and develop an efficient algorithm to obtain the estimator. Then we establish the rate of convergence of the proposed estimator for estimating sparse Laplacian constrained precision matrices under the Frobenius norm.

3.1.1 Adaptive estimator

Sparsity plays an important role in high-dimensional statistics, which helps avoid overfitting and improve identification of the relationships among data. The ℓ_1-norm is a popular regularizer to impose sparsity. The ℓ_1-norm regularized maximum likelihood estimation of the Laplacian constrained precision matrix can be formulated as

$$\min_{\boldsymbol{\Omega}\in\mathcal{S}_L} -\log\det{}^\star(\boldsymbol{\Omega}) + \operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{S}) + \lambda\sum_{i>j}|\Omega_{ij}|, \qquad (4)$$

where $\lambda > 0$ is the regularization parameter. Notice that for any $\boldsymbol{\Omega} \in \mathcal{S}_L$, $\boldsymbol{\Omega}$ is symmetric, and thus we need only to impose the sparsity on the lower (or upper) triangle part of $\boldsymbol{\Omega}$. It is well-known for graphical lasso that a larger regularization parameter will lead to a larger threshold, and the resultant solution will be sparser, because the elements will be shrunk if its value is less than the threshold. However, unlike graphical lasso, the ℓ_1-norm is not effective in the optimization problem (4) to get a sparse solution (Ying et al., 2020a): empirical results show that the number of nonzero elements grows as the regularization parameter increases, and theoretical results prove that no zero elements will appear in the solution if the regularization parameter is large enough.

To impose the sparsity effectively, we propose to use the adaptive weights to penalize the different coefficients in the ℓ_1-norm penalty. Then we propose an adaptive ℓ_1-norm regularized maximum likelihood estimation method, which can be formulated as

$$\begin{aligned}\widehat{\boldsymbol{\Omega}} := \arg\min_{\boldsymbol{\Omega}} &-\log\det{}^\star(\boldsymbol{\Omega}) + \operatorname{tr}(\boldsymbol{\Omega}\boldsymbol{S}) \\ &+ \lambda\sum_{i>j}|A_{ij}\Omega_{ij}|, \quad \text{subject to } \boldsymbol{\Omega}\in\mathcal{S}_L,\end{aligned} \qquad (5)$$

where $\boldsymbol{A}$ contains the data-dependent weights.

Now we discuss how to define the weights in $\boldsymbol{A}$. Let $\boldsymbol{\Omega}^\star$ be the underlying true precision matrix. If $\Omega^\star_{ij}$ is small or zero, a big penalty is expected on Ω_{ij} in (5), and thus A_{ij} should be large; Otherwise, A_{ij} should be small. Therefore, we can define

$$A_{ij} = \frac{1}{\left(|\widetilde{\Omega}_{ij}| + \varepsilon\right)^q}, \quad \text{for } i > j, \qquad (6)$$

where $q > 0$, ε is a small positive constant, and $\widetilde{\boldsymbol{\Omega}}$ is an initial consistent estimator, for example we can use the maximum likelihood estimator $\boldsymbol{\Omega}^n_O$ defined in (3). As the sample size grows, $\widetilde{\Omega}_{ij}$ will converge to zero if the underlying $\Omega^\star_{ij} = 0$, and the resultant A_{ij} will become large. Therefore, the proposed estimator $\widehat{\boldsymbol{\Omega}}$ can find out the true supports of the underlying precision matrix if the sample size is large enough, which is demonstrated by numerical simulations in Section 4.

Note that we can also use other decreasing functions in defining A_{ij}. To improve the estimation performance, we can recursively estimate $\widehat{\boldsymbol{\Omega}}$ through the optimization (5), and use the current estimation as $\widetilde{\boldsymbol{\Omega}}$ to update the weights A_{ij}. It is worth emphasizing that (5) is a convex optimization problem, and thus it does not suffer from the issue of multiple local minima, which is different from the nonconvex estimation method in (Ying et al., 2020a) with a concave penalty.

3.1.2 Algorithm

The estimation in (5) is a constrained optimization problem that includes the Laplacian constraints $\boldsymbol{\Omega} \in \mathcal{S}_L$. According to (1), we can see that for any $\boldsymbol{\Omega} \in \mathcal{S}_L$, $\boldsymbol{\Omega}$ is symmetric, and the sum of each row or column is zero. Therefore, there is a linear constraint between diagonal and off-diagonal elements, and the degrees of freedom of $\boldsymbol{\Omega}$ are $p(p-1)/2$ instead of p^2. To handle this linear constraint, we introduce a linear operator.

Definition 2 (Kumar et al. (2020))**.** The linear operator $\mathcal{L} : \mathbb{R}^{p(p-1)/2} \to \mathbb{R}^{p\times p}$, $\boldsymbol{x} \mapsto \mathcal{L}\boldsymbol{x}$, is defined by

$$[\mathcal{L}\boldsymbol{x}]_{ij} = \begin{cases} -x_k & i > j, \\ [\mathcal{L}\boldsymbol{x}]_{ji} & i < j, \\ -\sum_{j\neq i}[\mathcal{L}\boldsymbol{x}]_{ij} & i = j, \end{cases} \qquad (7)$$

where $k = i - j + \frac{j-1}{2}(2p-j)$.

The operator $\mathcal{L}$ defines a linear mapping from a vector $\boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$ to a matrix $\mathcal{L}\boldsymbol{x} \in \mathbb{R}^{p\times p}$ which satisfies $\sum_j[\mathcal{L}\boldsymbol{x}]_{ij} = 0$ for any $i \in [p]$, and $[\mathcal{L}\boldsymbol{x}]_{ij} = [\mathcal{L}\boldsymbol{x}]_{ji}$ for any $i \neq j$. A toy example is given for $\boldsymbol{x} \in \mathbb{R}^3$ as follows,

$$\mathcal{L}\boldsymbol{x} = \begin{bmatrix} \sum_{i=1,2} x_i & -x_1 & -x_2 \\ -x_1 & \sum_{i=1,3} x_i & -x_3 \\ -x_2 & -x_3 & \sum_{i=2,3} x_i \end{bmatrix}.$$

With the usage of the operator $\mathcal{L}$, the proposed estimator defined in (5) can be written as $\widehat{\boldsymbol{\Omega}} = \mathcal{L}\hat{\boldsymbol{w}}$, where $\hat{\boldsymbol{w}}$ is obtained by

$$\begin{aligned}\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} &-\log\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) + \operatorname{tr}(\boldsymbol{S}\mathcal{L}\boldsymbol{w}) \\ &+ \lambda\sum_i a_i w_i, \quad \text{subject to } \boldsymbol{w} \geq \boldsymbol{0},\end{aligned} \qquad (8)$$

where $\boldsymbol{J}$ is a constant matrix with each element equal to $1/p$, and $a_i \geq 0$ for $i \in [p(p-1)/2]$.

Note that the term $\det^\star(\boldsymbol{\Omega})$ in (5) can be written as $\det(\mathcal{L}\boldsymbol{w} + \boldsymbol{J})$ in (8), because there is only one nonzero eigenvalue of the matrix $\boldsymbol{J}$ which is equal to 1, and its eigenvector is orthogonal to the row and column spaces

of $\mathcal{L}\boldsymbol{w}$. There is a rank constraint $\text{rank}(\boldsymbol{\Omega}) = p-1$ in (5) since $\boldsymbol{\Omega} \in \mathcal{S}_L$. We can see that this rank constraint is removed in the formulation (8). The reason is as follows: the rank constraint can be satisfied if $(\mathcal{L}\boldsymbol{w} + \boldsymbol{J}) \in \mathcal{S}_{++}^p$, which can hold directly by the two facts. One fact is that $(\mathcal{L}\boldsymbol{w} + \boldsymbol{J})$ is non-singular for any $\boldsymbol{w}$ in the feasible set of (8); the other one is that $\mathcal{L}\boldsymbol{w}+\boldsymbol{J}$ must be positive semi-definite for any $\boldsymbol{w} \geq \boldsymbol{0}$, because $\mathcal{L}\boldsymbol{w}$ is a diagonally dominant matrix.

The optimization (8) is convex, and the objective function is smooth. Therefore, we can establish a sequence by the projected gradient descent which converges to the global minimum $\hat{\boldsymbol{w}}$. More specifically, in the $(t+1)$-th iteration, $\boldsymbol{w}^{t+1}$ is updated by

$$\boldsymbol{w}^{t+1} = \mathcal{P}_+ \left(\boldsymbol{w}^t - \eta \left(\mathcal{L}^* \left(\boldsymbol{S} - \left(\mathcal{L}\boldsymbol{w}^t + \boldsymbol{J} \right)^{-1} \right) + \lambda \boldsymbol{a} \right) \right),$$

where $\mathcal{P}_+(\boldsymbol{a}) = \max(\boldsymbol{a}, \boldsymbol{0})$, and $\mathcal{L}^*$ is the adjoint operator of $\mathcal{L}$, $\mathcal{L}^* : \mathbb{R}^{p \times p} \to \mathbb{R}^{p(p-1)/2}$, defined by

$$[\mathcal{L}^* \boldsymbol{Y}]_k = Y_{i,i} - Y_{i,j} - Y_{j,i} + Y_{j,j}, \tag{9}$$

where $i, j \in [p]$ satisfying $k = i - j + \frac{j-1}{2}(2p-j)$ and $i > j$. The learning rate η can be chosen by backtracking line search. In addition, it is interesting to extend the proposed method to estimate other structured matrices such as Hankel matrices using Hankel linear operator (Cai et al., 2016a; Ying et al., 2018).

3.1.3 Minimax Upper Bound

In this section, we derive the minimax upper bound under the Frobenius norm by establishing the rate of convergence for the proposed estimator ALPE.

We define a parameter space $\mathcal{F}(d_{n,p}, M)$ of the Laplacian constrained precision matrices as

$$\mathcal{F}(d_{n,p}, M) = \left\{ \boldsymbol{\Omega} \in \mathbb{R}^{p \times p} \middle| \max_i \sum_{j=1}^{p} I\left\{\Omega_{ij} \neq 0\right\} \leq d_{n,p} \right\} \\ \left\{ \frac{1}{M} \leq \lambda_2(\boldsymbol{\Omega}) \leq \lambda_{\max}(\boldsymbol{\Omega}) \leq M, \boldsymbol{\Omega} \in \mathcal{S}_L \right\},$$

where M is some universal constant, the sparsity parameter $d_{n,p}$ is allowed to grow as n and p grow, and $\lambda_{\max}(\boldsymbol{\Omega})$ and $\lambda_2(\boldsymbol{\Omega})$ denote the maximum and second smallest eigenvalues of $\boldsymbol{\Omega}$, respectively. The underlying precision matrices considered in this paper belong to $\mathcal{F}(d_{n,p}, M)$. Note that any $\boldsymbol{\Omega} \in \mathcal{F}(d_{n,p}, M)$ denotes the Laplacian matrix of a connected graph which consists of p nodes and has at most $d_{n,p}$ nonzero edges for each node. For any given precision matrix $\boldsymbol{\Omega} \in \mathcal{F}(d_{n,p}, M)$, denote

$$\mathcal{S} = \{(i,j) \,|\, |\Omega_{ij}| > 0 \text{ for } i > j\} \tag{10}$$

as the support set of the nonzero elements in the lower triangle matrix of $\boldsymbol{\Omega}$. Theorem 2 provides an upper bound under the Frobenius norm over the parameter space $\mathcal{F}(d_{n,p}, M)$.

Theorem 2. *Suppose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are i.i.d. samples drawn from a p-dimensional L-GMRF with the parameters $(\boldsymbol{0}, \boldsymbol{\Omega})$, and the sample size n is lower bounded by*

$$n \geq \max\Big(8c_0^{-1} c_d \left(\|\boldsymbol{A}_{\mathcal{S}}\|_{\max} + \|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min} \right)^2 M^2 |\mathcal{S}| \log p, \\ 8c_d \|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}^2 \log p \Big),$$

where $c_0 = 1/\left(8 \left\|\mathcal{L}^*(\mathcal{L}\boldsymbol{w}^\star + \boldsymbol{J})^{-1}\right\|_{\max}^2\right)$ is a constant, and $c_d \geq (d+2)\|\boldsymbol{A}_{\mathcal{S}^c}\|_{\min}^{-2}$ with a constant $d > 0$. Then the ALPE *estimator $\widehat{\boldsymbol{\Omega}}$, defined in (5) with $\lambda = \sqrt{c_0^{-1} c_d \log p / n}$, of the precision matrix $\boldsymbol{\Omega}$ obeys*

$$\inf_{\boldsymbol{\Omega} \in \mathcal{F}(d_{n,p}, M)} \mathbb{P}\left\{ \frac{1}{p} \left\| \widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega} \right\|_{\text{F}}^2 \lesssim \frac{d_{n,p} \log p}{n} \right\} \geq 1 - p^{-d}.$$

Note that d is a user-defined parameter and can be arbitrarily large. A larger d yields a larger probability with which the claims in Theorem 2 hold, but also leads to a more stringent requirement on the number of samples.

Remark 1. Theorem 2 presents the upper bound of the convergence rate of the proposed ALPE estimator under the squared Frobenius norm. By establishing the minimax lower bound in Theorem 3, we show that this rate of convergence cannot be improved by any other estimators, and thus the ALPE estimator attains the optimal rate of convergence.

3.2 Minimax Lower Bound

To obtain the optimal rate of convergence, a key step is to establish the minimax lower bound, and prove that the lower bound matches the upper bound obtained in Theorem 2. In this section, we turn to derive the minimax lower bound for estimating sparse Laplacian constrained precision matrices over the parameter space $\mathcal{F}(d_{n,p}, M)$ under the Frobenius norm. The derivation of the lower bound is based on testing arguments through Le Cam-Assouad's method, and a careful construction of a subset of the parameter space $\mathcal{F}(d_{n,p}, M)$.

Theorem 3. *Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be i.i.d. samples drawn from p-dimensional L-GMRF with the parameters $(\boldsymbol{0}, \boldsymbol{\Omega})$. Assume that there exist universal constants $c > 0$ and $\beta > 1$ such that $p \geq cn^\beta$. The minimax risk for estimating the precision matrix $\boldsymbol{\Omega}$ over the parameter space $\mathcal{F}(d_{n,p}, M)$ with $d_{n,p} = o\left(\sqrt{\frac{n}{\log p}}\right)$ based on the random samples $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ satisfies*

$$\inf_{\widehat{\boldsymbol{\Omega}}} \sup_{\boldsymbol{\Omega} \in \mathcal{F}(d_{n,p}, M)} \mathbb{E} \frac{1}{p} \left\| \widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega} \right\|_{\text{F}}^2 \gtrsim \frac{d_{n,p} \log p}{n}.$$

Remark 2. The upper and lower bounds, established in Theorem 2 and Theorem 3 respectively, match with each other, and thus yield the optimal rate of convergence. Although the optimal rate of convergence for estimating unconstrained sparse precision matrices has been obtained in (Cai et al., 2016b), it is still motivated to establish the minimax rate of convergence for estimating the Laplacian constrained precision matrices, because their parameter spaces are disjoint with each other. It is also interesting to explore the minimax rate of convergence for the case of sparse precision matrices under MTP_2 constraint (Fallat et al., 2017; Lauritzen et al., 2019b) in the future work.

The proof of Theorem 3 consists of three main steps. We first construct a subset $\mathcal{F}_*$ of the parameter space $\mathcal{F}(d_{n,p}, M)$ satisfying that the risk of estimation over $\mathcal{F}_*$ is essentially the same as that of estimation over $\mathcal{F}(d_{n,p}, M)$. Then we apply the Le Cam-Assouad's method (Cai and Zhou, 2012) to the constructed $\mathcal{F}_*$. Finally, we calculate one factor, and bound the total variation affinity between two multivariate normal mixtures. We outline the main procedures of the proof as follows, and the detailed proofs of the technical lemmas can be found in the full paper.

Proof of Theorem 3. The proof contains three steps.

Step 1: Constructing the sub-parameter set $\mathcal{F}_*$. It is usually difficult to derive the minimax lower bound over the original parameter space $\mathcal{F}(d_{n,p}, M)$. However, we can restrict our attention to a subset of the parameter space. Suppose $\mathcal{F}_*$ is a sub-parameter set of $\mathcal{F}(d_{n,p}, M)$, i.e., $\mathcal{F}_* \subset \mathcal{F}(d_{n,p}, M)$. Then one has

$$\inf_{\widehat{\boldsymbol{\Omega}}} \sup_{\boldsymbol{\Omega}\in\mathcal{F}(d_{n,p},M)} \left\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\right\|_{\mathrm{F}} \geq \inf_{\widehat{\boldsymbol{\Omega}}} \sup_{\boldsymbol{\Omega}\in\mathcal{F}_*} \left\|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\right\|_{\mathrm{F}}. \tag{11}$$

It is worth mentioning that the construction of $\mathcal{F}_*$ should be careful: on one hand, $\mathcal{F}_*$ should be simple such that we can apply general minimax lower bound techniques such as Le Cam-Assouad's method to this set; on the other hand, $\mathcal{F}_*$ cannot be too much restrictive, otherwise the lower bound of $\mathcal{F}_*$ will be smaller than that of $\mathcal{F}(d_{n,p}, M)$, and the resultant bound cannot match the upper bound in Theorem 2.

Now we construct the sub-parameter set $\mathcal{F}_*$. Let $m = \lfloor \frac{p}{2} \rfloor$, and $\mathcal{B}$ be the collection of all the vectors $\boldsymbol{\xi} \in \{0,1\}^p$ satisfying $\xi_i = 0$ for $1 \leq i \leq p-m$, and $\xi_i = 0$ or 1 for $p-m+1 \leq i \leq p$ under the constraint that $\|\xi\|_0 = k$. We then collect m row vectors $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m \in \mathcal{B}$ as an $m \times p$ matrix, $\left(\boldsymbol{\xi}_1^\top, \ldots, \boldsymbol{\xi}_m^\top\right) \in \mathcal{B}^m$, with the i-th row of the matrix equal to $\boldsymbol{\xi}_i^\top$ for $i \in [m]$.

We consider a subset $\Xi \subset \mathcal{B}^m$, satisfying that each column sum of the matrix in Ξ is less than or equal to $2k$. Note that each $\boldsymbol{\xi} \in \Xi$ is an $m \times p$ matrix, and we denote $\boldsymbol{\xi} = \left(\boldsymbol{\xi}_1^\top, \ldots, \boldsymbol{\xi}_m^\top\right)$. For each $\boldsymbol{\xi} \in \Xi$, we define a $p \times p$ symmetric matrix $\mathcal{R}_i(\boldsymbol{\xi}_i)$ with the i-th row of $\mathcal{R}_i(\boldsymbol{\xi}_i)$ equal to $\boldsymbol{\xi}_i^\top$, the i-th column equal to $\boldsymbol{\xi}_i$, and the rest of entries 0. Note that for each $\boldsymbol{\xi} \in \Xi$, each row or column sum of the matrix $\sum_{i=1}^m \mathcal{R}_i(\boldsymbol{\xi}_i)$ is less than or equal to $2k$.

Let $\Gamma = \{0,1\}^m$. Define

$$\Theta = \Gamma \otimes \Xi, \tag{12}$$

where for any $\theta \in \Theta$, one has $\theta = (\gamma, \boldsymbol{\xi})$ with $\gamma = (\gamma_1, \ldots, \gamma_m) \in \Gamma$ and $\boldsymbol{\xi} = \left(\boldsymbol{\xi}_1^\top, \ldots, \boldsymbol{\xi}_m^\top\right) \in \Xi$. Now we associate each θ with a precision matrix $\boldsymbol{\Omega}(\theta)$ by

$$\boldsymbol{\Omega}(\theta) = \boldsymbol{L}_o + \epsilon_{n,p} \boldsymbol{L}_b(\theta), \tag{13}$$

where $\boldsymbol{L}_o$ is a constant Laplacian matrix without respect to θ defined by

$$\boldsymbol{L}_o = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}. \tag{14}$$

$\boldsymbol{L}_b(\theta)$ is a Laplacian matrix defined by $\boldsymbol{L}_b(\theta) = \boldsymbol{D}_b(\theta) - \boldsymbol{W}_b(\theta)$, where $\boldsymbol{D}_b(\theta)$ is the degree matrix, and $\boldsymbol{W}_b(\theta)$ is the adjacency matrix constructed by

$$\boldsymbol{W}_b(\theta) = \sum_{i=1}^m \gamma_i \mathcal{R}_i(\boldsymbol{\xi}_i), \quad \theta = (\gamma, \boldsymbol{\xi}) \in \Theta. \tag{15}$$

We define the sub-parameter space $\mathcal{F}_*$ by

$$\mathcal{F}_* = \Big\{ \boldsymbol{\Omega}(\theta) | \boldsymbol{\Omega}(\theta) = \boldsymbol{L}_o + \epsilon_{n,p} \boldsymbol{L}_b(\theta), \theta = (\gamma, \boldsymbol{\xi}) \in \Theta,$$
$$\boldsymbol{L}_b(\theta) = \boldsymbol{D}_b(\theta) - \boldsymbol{W}_b(\theta), \boldsymbol{W}_b(\theta) = \sum_{i=1}^m \gamma_i \mathcal{R}_i(\boldsymbol{\xi}_i) \Big\},$$

where $\epsilon_{n,p} = c_\epsilon \sqrt{\frac{\log p}{n}}$, where $c_\epsilon \in (0, \frac{\beta-1}{8\beta})$ is a constant. Finally we specify the value of k to ensure $\mathcal{F}_* \subset \mathcal{F}(d_{n,p}, M)$. We set $k = \lceil \frac{d_{n,p}-3}{2} \rceil$ to make sure the maximum degree of the graph is less than or equal to $d_{n,p}$.

The insight into the construction of $\mathcal{F}_*$ is the combination of the line graph and bipartite graph, which correspond to $\boldsymbol{L}_o$ and $\boldsymbol{L}_b(\theta)$ in (13), respectively. Notice that the original parameter space $\mathcal{F}(d_{n,p}, M)$ is the collection of the precision matrices of the connected graphs that consist of p nodes and have at most $d_{n,p}$ nonzero edges for each node. To guarantee $\mathcal{F}_* \subset \mathcal{F}(d_{n,p}, M)$, each $\boldsymbol{\Omega}(\theta) \in \mathcal{F}_*$ should also correspond to a connected graph. In the meantime, the risk of estimation over $\mathcal{F}_*$ is required to be essentially the

same as that of estimation over $\mathcal{F}(d_{n,p}, M)$. To make the construction to satisfy the two requirements, we restrict our attention to the connected graphs that can be divided into two sub-graphs: one is a constant line graph that ensures the constructed graph is connected, and the other one is a bipartite graph characterized by Θ which can make the lower bound sharp.

Step 2: Applying the Le Cam-Assound's method to $\mathcal{F}_*$. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be *i.i.d.* samples drawn from L-GMRF with the parameters $(\mathbf{0}, \boldsymbol{\Omega}(\theta))$ with $\theta \in \Theta$, and denote the joint distribution by $\mathbb{P}_\theta$. Applying the Le Cam-Assouad's method (Cai and Zhou, 2012) to the parameter space Θ, we have

$$\inf_{\widehat{\boldsymbol{\Omega}}} \max_{\theta \in \Theta} 2^2 \mathbb{E}_\theta \left\| \widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}(\theta) \right\|_{\mathrm{F}}^2 \geq \frac{m\alpha}{2} \min_{1 \leq i \leq m} \left\| \bar{\mathbb{P}}_{0,i} \wedge \bar{\mathbb{P}}_{1,i} \right\|,$$

where the factor α is defined by

$$\alpha := \min_{\{(\theta,\theta') : H(\gamma(\theta), \gamma(\theta')) \geq 1\}} \frac{\|\boldsymbol{\Omega}(\theta) - \boldsymbol{\Omega}(\theta')\|_{\mathrm{F}}^2}{H(\gamma(\theta), \gamma(\theta'))}, \quad (16)$$

in which $H\left(\gamma(\theta), \gamma(\theta')\right) = \sum_{i=1}^m |\gamma(\theta_i) - \gamma(\theta_i')|$. $\bar{\mathbb{P}}_{0,i}$ and $\bar{\mathbb{P}}_{1,i}$ are the mixture distributions defined by

$$\bar{\mathbb{P}}_{a,i} = \frac{1}{2^{m-1} D_\Xi} \sum_{\theta} \left\{ \mathbb{P}_\theta : \gamma_i(\theta) = a \right\}, \quad (17)$$

where D_Ξ is the cardinality of Ξ, $a \in \{0, 1\}$ and $i \in [m]$. We can see that $\bar{\mathbb{P}}_{a,i}$ is the mixture distribution over all $\mathbb{P}_\theta$ with $\gamma_i(\theta)$ fixed to be a while all other components of θ vary over all possible values.

Step 3: Bounding the per comparison loss α defined in (16) and the affinity $\min_{1 \leq i \leq m} \left\| \bar{\mathbb{P}}_{0,i} \wedge \bar{\mathbb{P}}_{1,i} \right\|$. The bounds for the two terms are provided in the next two lemmas.

Lemma 1. *For α defined in* (16)*, one has*

$$\alpha \geq k \epsilon_{n,p}^2. \quad (18)$$

The key technical difficulty is in bounding the affinity between the Gaussian mixtures.

Lemma 2. *Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be i.i.d. samples drawn from L-GMRF with the parameters $(\mathbf{0}, \boldsymbol{\Omega}(\theta))$ with $\theta \in \Theta$. Then there exists a constant $c > 0$ such that*

$$\min_{1 \leq i \leq m} \left\| \bar{\mathbb{P}}_{0,i} \wedge \bar{\mathbb{P}}_{1,i} \right\| \geq c. \quad (19)$$

Finally, Together with (11), Lemma 1 and Lemma 2, the minimax lower bound for estimating precision matrices over $\mathcal{F}(d_{n,p}, M)$ can be obtained by

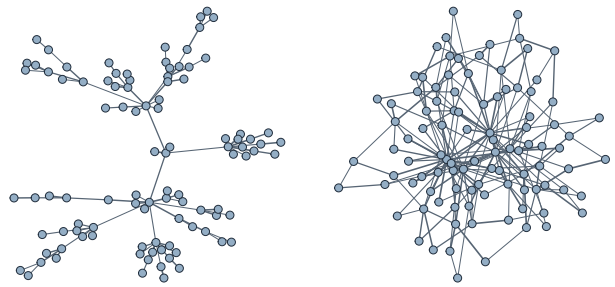
$$\inf_{\widehat{\boldsymbol{\Omega}}} \sup_{\boldsymbol{\Omega} \in \mathcal{F}(d_{n,p}, M)} \mathbb{E} \left\| \widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega} \right\|_{\mathrm{F}}^2 \geq \inf_{\widehat{\boldsymbol{\Omega}}} \sup_{\boldsymbol{\Omega}(\theta) \in \mathcal{F}_*} \mathbb{E}_\theta \left\| \widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}(\theta) \right\|_{\mathrm{F}}^2 \gtrsim \frac{d_{n,p} p \log p}{n},$$

completing the proof. □

4 Numerical Experiments

In this section, we conduct numerical simulations to compare the estimation performance of the estimators including GLE-ADMM (Zhao et al., 2019), NGL-MCP (Ying et al., 2020a), NGL-SCAD (Ying et al., 2020a), and the proposed ALPE. The GLE-ADMM estimates the Laplacian constrained precision matrices by solving the ℓ_1-norm regularized maximum likelihood estimation using alternating direction method of multipliers (ADMM) method, while NGL-MCP and NGL-SCAD use the nonconvex penalties MCP and SCAD instead, and solve the optimization problem via the majorization-minimization method.

We randomly generate the Barabasi-Albert graphs of degree one, also known as tree graphs, and Barabasi-Albert graphs of degree two (Albert and Barabási, 2002) as the underlying ground-truth graphs. The number of nodes p is equal to 100 in the generated graphs, and the graph weights associated with edges are uniformly sampled from $U(2, 5)$. The underlying Laplacian constrained precision matrix in the L-GMRF model can be obtained by $\boldsymbol{\Omega} = \boldsymbol{D} - \boldsymbol{W}$, where $\boldsymbol{D}$ and $\boldsymbol{W}$ are the degree matrix and adjacency matrix of the generated graph, respectively. Then the samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are independently drawn from L-GMRF with the parameters $(\mathbf{0}, \boldsymbol{\Omega})$, and the sample covariance matrix can be computed by $\boldsymbol{S} = \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^\top$.



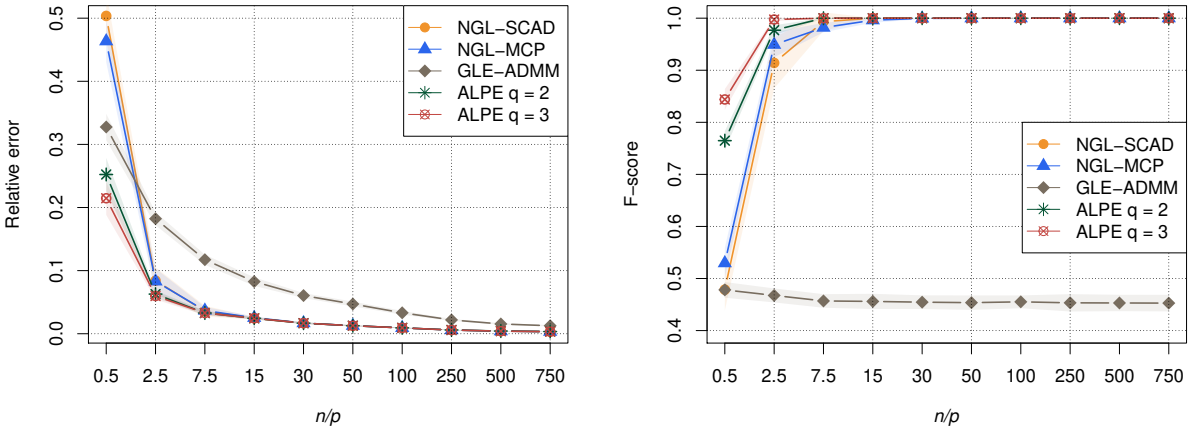
(a) (b)

Figure 1: Illustration of the generated graphs: (a) Barabasi-Albert graphs of degree one, and (b) Barabasi-Albert graphs of degree two.

To compare the performance of each estimator, we compute the relative error (RE) and F-score (FS) which are defined as

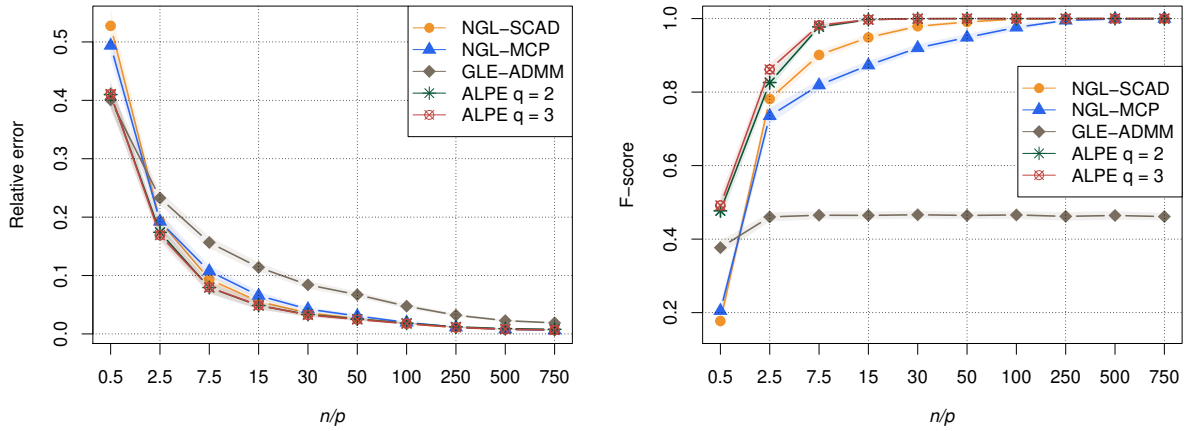
$$\mathsf{RE} = \frac{\left\| \widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega} \right\|_{\mathrm{F}}}{\|\boldsymbol{\Omega}\|_{\mathrm{F}}}, \quad \text{and} \quad \mathsf{FS} = \frac{2\mathsf{tp}}{2\mathsf{tp} + \mathsf{fp} + \mathsf{fn}}, \quad (20)$$

where $\widehat{\boldsymbol{\Omega}}$ and $\boldsymbol{\Omega}$ are the estimated and true Laplacian constrained precision matrices, respectively. In the definition of F-score, the tp denotes true positive (i.e., the case that there is an actual edge and the algorithm detects it), fp denotes false positive (i.e., the case that



(a) (b)

Figure 2: Performance comparisons under (a) Relative error, (b) F-score with different sample size ratios n/p in learning Barabasi-Albert graphs of degree one with 100 nodes.



(a) (b)

Figure 3: Performance comparisons under (a) Relative error, (b) F-score with different sample size ratios n/p in learning Barabasi-Albert graphs of degree two with 100 nodes.

there is no actual edge but algorithm detects one), and fn denotes false negative (i.e., the case that the algorithm failed to detect an actual edge). The F-score takes values in $[0, 1]$, and 1 indicates that all the connections and disconnections between nodes in the underlying ground-truth graphs are correctly identified. We define the true positive rate (TPR) and false positive rate (FPR) as

$$\mathsf{TPR} = \frac{\mathsf{tp}}{\mathsf{tp} + \mathsf{fn}}, \quad \text{and} \quad \mathsf{FPR} = \frac{\mathsf{fp}}{\mathsf{fp} + \mathsf{tn}}.$$

The curves in Figures 2, 3 and 4 are the results of an average of 50 Monte Carlo realizations and the shaded areas represent the one-standard deviation confidence interval. The regularization parameter λ is fine-tuned for each method.

Figure 2(a) shows that the relative errors of all the estimators decrease as the sample size ratio n/p increases, and the proposed ALPE can always achieve a lower estimation error than the compared GLE-ADMM, NGL-MCP and NGL-SCAD under different sample size ratio. Figure 2(b) shows that the NGL-MCP, NGL-SCAD and ALPE achieve the F-score of 1, and thus can identify all the graph edges correctly, if the sample size is large enough. Moreover, the ALPE requires less samples than NGL-MCP and NGL-SCAD to obtain perfect graph structures.

It is observed in Figure 2(b) that the GLE-ADMM with the ℓ_1-norm cannot recover the edges correctly even given a large number of observations, implying that the GLE-ADMM is not subset selection consistent. Therefore, for convex formulations, we should use adaptive ℓ_1-norm rather than ℓ_1-norm to identify the graph structures. Similar conclusions can also be made through Figure 3 in learning Barabasi-Albert graphs of degree two.

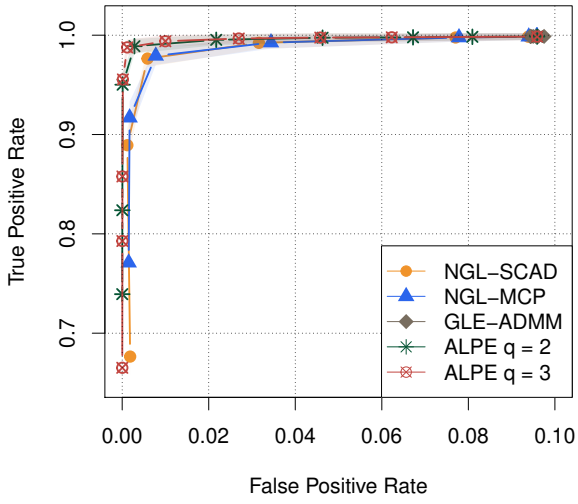


Figure 4: ROC curves for different methods in learning Barabasi-Albert graphs of degree two with 100 nodes and 800 observations. To generate the ROC curves, we set a range of regularization parameters from 0 to 1 for each method.

Figure 4 illustrates the receiver operating characteristic (ROC) curves of different methods in learning Barabasi-Albert graphs of degree two. The ROC curves can evaluate the support recovery performance of the underlying precision matrices. Figure 4 shows that our proposed estimator ALPE can achieve higher true positive rates than all other state-of-the-art methods while keeping low false positive rates, and thus can achieve a better performance in support selection.

5 Conclusions

In this paper, we have proved that the Gaussian maximum likelihood estimator of precision matrices under Laplacian constraints exists and is unique almost surely as long as there is one observation. We have established the optimal rate of convergence under the Frobenius norm for estimating the sparse Laplacian constrained precision matrices. We have obtained the minimax lower bound through the application of Le Cam-Assouad's method to a carefully constructed sub-parameter space. We have proposed an adaptive estimator ALPE, and have proved that ALPE attains the optimal rate of convergence with an overwhelming probability. Numerical simulations have shown that for convex formulations, we should use adaptive ℓ_1-norm rather than ℓ_1-norm for support selection.

Acknowledgments

We would like to thank the anonymous reviewers for their helpful comments. This work was supported by the Hong Kong GRF 16207019 research grant.

References

R. Agrawal, U. Roy, and C. Uhler. Covariance matrix estimation under total positivity for portfolio selection. *arXiv preprint arXiv:1909.04222*, 2019.

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74 (1):47, 2002.

O. Banerjee, L. E. Ghaoui, and A. d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research*, 9(Mar):485–516, 2008.

P. J. Bickel and Y. R. Gel. Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):711–728, 2011.

J.-F. Cai, X. Qu, W. Xu, and G.-B. Ye. Robust recovery of complex exponential signals from random Gaussian projections via low rank Hankel matrix reconstruction. *Applied and Computational Harmonic Analysis*, 41(2):470–490, 2016a.

T. T. Cai and H. H. Zhou. Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics*, 40(5):2389–2420, 2012.

T. T. Cai, C.-H. Zhang, H. H. Zhou, et al. Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38(4):2118–2144, 2010.

T. T. Cai, W. Liu, and H. H. Zhou. Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *The Annals of Statistics*, 44(2): 455–488, 2016b.

T. T. Cai, Z. Ren, H. H. Zhou, et al. Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics*, 10(1):1–59, 2016c.

J. V. d. M. Cardoso and D. P. Palomar. Learning undirected graphs in financial markets. In *2020 54th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2020.

J. V. d. M. Cardoso, J. Ying, and D. P. Palomar. Algorithms for learning graphs in financial markets. *arXiv preprint arXiv:2012.15410*, 2020.

M. Chen, C. Gao, Z. Ren, et al. Robust covariance and scatter matrix estimation under huber's contamination model. *Annals of Statistics*, 46(5):1932–1960, 2018.

A. d'Aspremont, O. Banerjee, and L. El Ghaoui. First-order methods for sparse covariance selection. *SIAM Journal on Matrix Analysis and Applications*, 30(1): 56–66, 2008.

X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst. Learning Laplacian matrix in smooth graph signal representations. *IEEE Transactions on Signal Processing*, 64(23):6160–6173, 2016.

X. Dong, D. Thanou, M. Rabbat, and P. Frossard. Learning graphs from data: A signal representation perspective. *IEEE Signal Processing Magazine*, 36 (3):44–63, 2019.

H. E. Egilmez, E. Pavez, and A. Ortega. Graph learning from data under Laplacian and structural constraints. *IEEE Journal of Selected Topics in Signal Processing*, 11(6):825–841, 2017.

S. Fallat, S. Lauritzen, K. Sadeghi, C. Uhler, N. Wermuth, P. Zwiernik, et al. Total positivity in Markov structures. *The Annals of Statistics*, 45 (3):1152–1184, 2017.

J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

A. Holbrook. Differentiating the pseudo determinant. *Linear Algebra and its Applications*, 548:293 – 304, 2018.

J. Honorio, D. Samaras, I. Rish, and G. Cecchi. Variable selection for gaussian graphical models. In *Artificial Intelligence and Statistics*, pages 538–546, 2012.

A. Hu and S. Negahban. Minimax estimation of bandable precision matrices. In *Advances in Neural Information Processing Systems*, pages 4888–4896, 2017.

A. K. Kim. Obtaining minimax lower bounds: a review. *Journal of the Korean Statistical Society*, pages 1–29, 2020.

S. Kumar, J. Ying, J. V. de Miranda Cardoso, and D. P. Palomar. Structured graph learning via Laplacian spectral constraints. In *Advances in Neural Information Processing Systems*, pages 11647–11658, 2019.

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*, 21(22):1–60, 2020.

B. Lake and J. Tenenbaum. Discovering structure by learning sparse graphs. In *Proceedings of the 33rd Annual Cognitive Science Conference*, pages 778—783, 2010.

S. Lauritzen, C. Uhler, and P. Zwiernik. Total positivity in exponential families with application to binary variables. *arXiv preprint arXiv:1905.00516*, 2019a.

S. Lauritzen, C. Uhler, P. Zwiernik, et al. Maximum likelihood estimation in Gaussian models under total positivity. *The Annals of Statistics*, 47(4):1835–1863, 2019b.

S. L. Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

Y. Liu, Z. Ren, et al. Minimax estimation of large precision matrices with bandable cholesky factor. *The Annals of Statistics*, 48(4):2428–2454, 2020.

R. Mazumder and T. Hastie. The graphical lasso: New insights and alternatives. *Electronic Journal of Statistics*, 6:2125, 2012.

P. R. Milgrom and R. J. Weber. A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society*, pages 1089–1122, 1982.

A. Ortega, P. Frossard, J. Kovaṽcević, J. M. Moura, and P. Vandergheynst. Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106(5):808–828, 2018.

Y. Park, D. Hallac, S. Boyd, and J. Leskovec. Learning the network structure of heterogeneous data via pairwise exponential Markov random fields. *Proceedings of Machine Learning Research*, 54:1302, 2017.

P. Ravikumar, M. J. Wainwright, G. Raskutti, B. Yu, et al. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980, 2011.

Z. Ren, T. Sun, C.-H. Zhang, H. H. Zhou, et al. Asymptotic normality and optimalities in estimation of large gaussian graphical models. *Annals of Statistics*, 43(3):991–1026, 2015.

A. J. Rothman, P. J. Bickel, E. Levina, J. Zhu, et al. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515, 2008.

D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3): 83–98, 2013.

M. Slawski and M. Hein. Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields. *Linear Algebra and its Applications*, 473:145–179, 2015.

J. A. Soloff, A. Guntuboyina, and M. I. Jordan. Covariance estimation with nonnegative partial correlations. *arXiv preprint arXiv:2007.15252*, 2020.

L. Wang, X. Ren, and Q. Gu. Precision matrix estimation in high dimensional gaussian graphical models with faster rates. In *Artificial Intelligence and Statistics*, pages 177–185, 2016.

Y. Wang, U. Roy, and C. Uhler. Learning high-dimensional gaussian graphical models under total positivity without adjustment of tuning parameters. In *International Conference on Artificial Intelligence and Statistics*, pages 2698–2708. PMLR, 2020.

J. Ying, J.-F. Cai, D. Guo, G. Tang, Z. Chen, and X. Qu. Vandermonde factorization of Hankel matrix for complex exponential signal recovery—application in fast NMR spectroscopy. *IEEE Transactions on Signal Processing*, 66(21):5520–5533, 2018.

J. Ying, J. V. d. M. Cardoso, and D. P. Palomar. Does the ℓ_1-norm learn a sparse graph under laplacian constrained graphical models? *arXiv preprint arXiv:2006.14925*, 2020a.

J. Ying, J. V. de Miranda Cardoso, and D. P. Palomar. Nonconvex sparse graph learning under Laplacian constrained graphical model. In *Advances in Neural Information Processing Systems*, 2020b.

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

L. Zhao, Y. Wang, S. Kumar, and D. P. Palomar. Optimization algorithms for graph Laplacian estimation via ADMM and MM. *IEEE Transactions on Signal Processing*, 67(16):4231–4244, 2019.

X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*, pages 912–919, 2003.